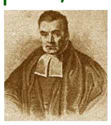
Probabilistic Belief States and Bayesian Networks

(Where we exploit the sparseness of direct interactions among components of a world)

R&N: Chap. 14, Sect. 14.1–4



Probabilistic Belief

- Consider a world where a dentist agent D meets with a new patient P
- D is interested in only whether P has a cavity; so, a state is described with a single proposition – Cavity
- Before observing P, D does not know if P has a cavity, but from years of practice, he believes Cavity with some probability p and ¬Cavity with probability 1-p
- The proposition is now a boolean random variable and (Cavity, p) is a **probabilistic belief**

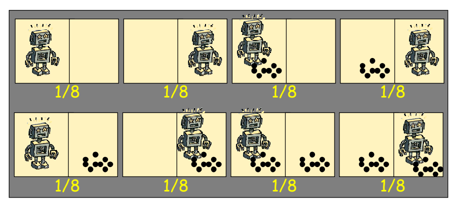
Probabilistic Belief State

- The world has only two possible states, which are respectively described by Cavity and ¬Cavity
- The **probabilistic belief state** of an agent is a probabilistic distribution over all the states that the agent thinks possible
- In the dentist example, D's belief state is:

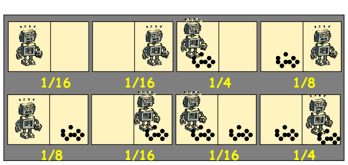
Cavity	¬ Cavity
p	1-p

Vacuum Robot

If the robot has no idea what the state of the world is, and thinks that all states are equally probable (using the "principle of indifference"), its belief state is:



How are beliefs and belief states related?



How a belief affects the entire belief state and the other beliefs?

(Clean(R_1), 5/16)
(Clean(R_2), 0.5)
(In(Robot,R_1), 0.5)
(In(Robot,R_2), 0.5)

It is usually more convenient to deal with individual beliefs than entire belief states, e.g.:

- The robot may choose to execute Suck(R_2) only if Clean(R_2) has low probability

- The robot may directly observe whether Clean(R_1) or Clean(R_2)

Back to the dentist example ...

- We now represent the world of the dentist D using three propositions – Cavity, Toothache, and PCatch
- D's belief state consists of 2^3 = 8 states each with some probability:
  {Cavity∧Toothache∧PCatch,
  ¬Cavity∧Toothache∧PCatch,
  Cavity∧¬Toothache∧PCatch,...}

The belief state is defined by the full joint probability of the propositions

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Probabilistic Inference

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

P(Cavity ∨Toothache) = 0.108 + 0.012 + ...
= 0.28

Probabilistic Inference

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

P(Cavity) = 0.108 + 0.012 + 0.072 + 0.008
= 0.2

Probabilistic Inference

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Marginalization: $P(c) = \Sigma_t \Sigma_{pc} P(c \wedge t \wedge pc)$
using the conventions that c = Cavity or ¬Cavity and that Σ_t is the sum over t = {Toothache, ¬Toothache}

Conditional Probability

- $P(A \wedge B) = P(A|B)\,P(B)$
  $= P(B|A)\,P(A)$
  P(A|B) is the posterior probability of A given B

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

P(Cavity|Toothache) = P(Cavity∧Toothache)/P(Toothache)
= (0.108+0.012)/(0.108+0.012+0.016+0.064) = 0.6

Interpretation: After observing Toothache, the patient is no longer an "average" one, and the prior probabilities of Cavity is no longer valid

P(Cavity|Toothache) is calculated by keeping the ratios of the probabilities of the 4 cases unchanged, and normalizing their sum to 1

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

P(Cavity|Toothache) = P(Cavity∧Toothache)/P(Toothache)
= (0.108+0.012)/(0.108+0.012+0.016+0.064) = 0.6
P(¬Cavity|Toothache)=P(¬Cavity∧Toothache)/P(Toothache)
= (0.016+0.064)/(0.108+0.012+0.016+0.064) = 0.4

P(c|Toochache) = α P(c ∧Toothache)
= α Σ_{pc} P(c ∧Toothache ∧ pc)
= α [(0.108, 0.016) + (0.012, 0.064)]
= α (0.12, 0.08) = (0.6, 0.4)

normalization constant

Conditional Probability

- P(A∧B) = P(A|B) P(B)
  = P(B|A) P(A)
- P(A∧B∧C) = P(A|B,C) P(B∧C)
  = P(A|B,C) P(B|C) P(C)
- P(Cavity) = $\Sigma_t \Sigma_{pc}$ P(Cavity∧t∧pc)
  = $\Sigma_t \Sigma_{pc}$ P(Cavity|t,pc) P(t∧pc)
- P(c) = $\Sigma_t \Sigma_{pc}$ P(c∧t∧pc)
  = $\Sigma_t \Sigma_{pc}$ P(c|t,pc)P(t∧pc)

Independence

- Two random variables A and B are independent if
  P(A∧B) = P(A) P(B)
  hence if P(A|B) = P(A)
- Two random variables A and B are independent given C, if
  P(A∧B|C) = P(A|C) P(B|C)
  hence if P(A|B,C) = P(A|C)

Updating the Belief State

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

- Let D now observe Toothache with probability 0.8 (e.g., “the patient says so”)
- How should D update its belief state?

Updating the Belief State

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

- Let E be the evidence such that P(Toothache|E) = 0.8
- We want to compute P(c∧t∧pc|E) = P(c∧pc|t,E) P(t|E)
- Since E is not directly related to the cavity or the probe catch, we consider that c and pc are independent of E given t, hence: P(c∧pc|t,E) = P(c∧pc|t)

Updating the Belief State

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	~~0.108~~ 0.432	~~0.012~~ 0.048	~~0.072~~ 0.018	~~0.008~~ 0.002
¬Cavity	~~0.016~~ 0.064	~~0.064~~ 0.256	~~0.144~~ 0.036	~~0.576~~ 0.144

- Let E be the evidence such that P(Toothache|E) = 0.8
- … t∧pc|E) = P(c∧pc|t,E) P(t|E)
- Since E is not directly related to the cavity or the probe catch, we … e independent of E given t, he … t)

To get these 4 probabilities we normalize their sum to 0.8

To get these 4 probabilities we normalize their sum to 0.2

Issues

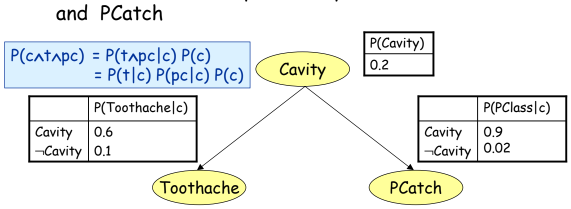
- If a state is described by n propositions, then a belief state contains 2^n states (possibly, some have probability 0)
- → Modeling difficulty: many numbers must be entered in the first place
- → Computational issue: memory size and time

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

- Toothache and PCatch are independent given Cavity (or ¬Cavity), but this relation is hidden in the numbers ! [Verify this]
- Bayesian networks explicitly represent independence among propositions to reduce the number of probabilities defining a belief state

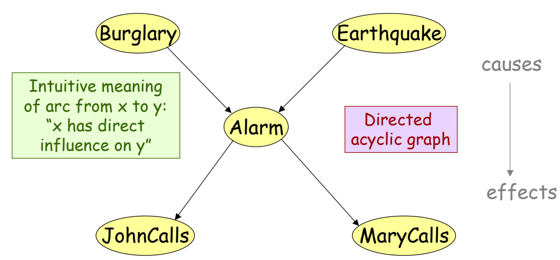
Bayesian Network

- Notice that Cavity is the "cause" of both Toothache and PCatch, and represent the causality links explicitly
- Give the prior probability distribution of Cavity
- Give the conditional probability tables of Toothache and PCatch

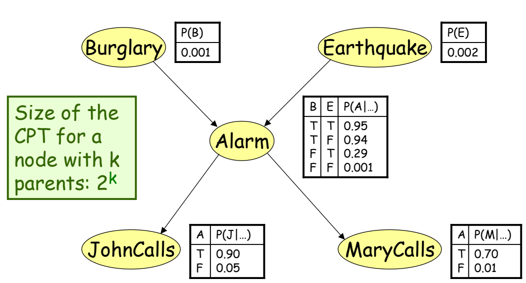


5 probabilities, instead of 7

A More Complex BN

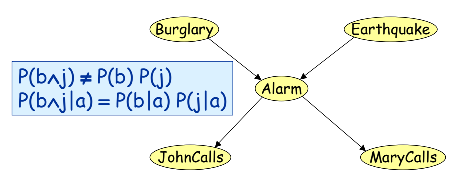


A More Complex BN



10 probabilities, instead of 31

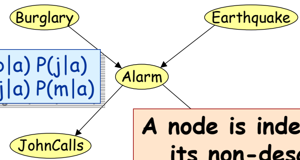
What does the BN encode?



Each of the beliefs JohnCalls and MaryCalls is independent of Burglary and Earthquake given Alarm or ¬Alarm

For example, John does not observe any burglaries directly

What does the BN encode?



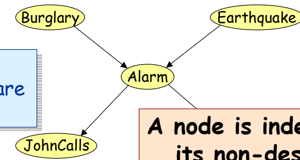
$P(b \wedge j|a) = P(b|a)\,P(j|a)$
$P(j \wedge m|a) = P(j|a)\,P(m|a)$

A node is independent of its non-descendants given its parents

The beliefs JohnCalls and MaryCalls are independent given Alarm or ¬Alarm

For instance, the reasons why John and Mary may not call if there is an alarm are unrelated

What does the BN encode?



Burglary and Earthquake are independent

A node is independent of its non-descendants given its parents

The beliefs JohnCalls and MaryCalls are independent given Alarm or ¬Alarm

For instance, the reasons why John and Mary may not call if there is an alarm are unrelated

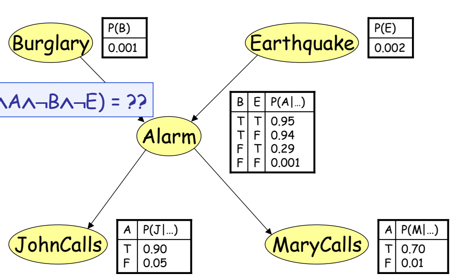
Locally Structured World

- A world is locally structured (or sparse) if each of its components interacts directly with relatively few other components
- In a sparse world, the CPTs are small and the BN contains much fewer probabilities than the full joint distribution
- If the # of entries in each CPT is bounded by a constant, i.e., O(1), then the # of probabilities in a BN is linear in n – the # of propositions – instead of 2^n for the joint distribution

But does a BN represent a belief state?

In other words, can we compute the full joint distribution of the propositions from it?

Calculation of Joint Probability



P(B)
0.001

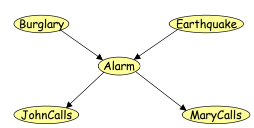
P(E)
0.002

B	E	P(A|…)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|…)
T	0.90
F	0.05

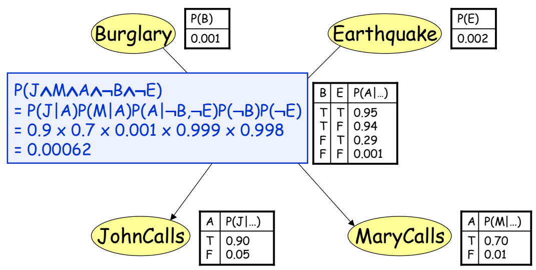
A	P(M|…)
T	0.70
F	0.01

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



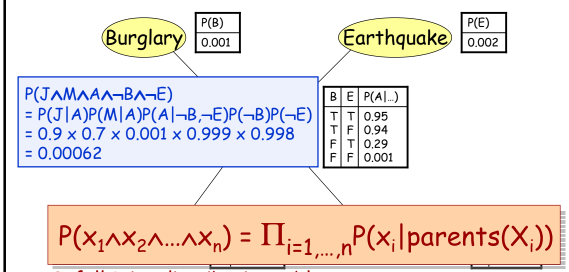
- $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$
  $= P(J \wedge M|A,\neg B,\neg E) \times P(A \wedge \neg B \wedge \neg E)$
  $= P(J|A,\neg B,\neg E) \times P(M|A,\neg B,\neg E) \times P(A \wedge \neg B \wedge \neg E)$
  (J and M are independent given A)
- $P(J|A,\neg B,\neg E) = P(J|A)$
  (J and $\neg B \wedge \neg E$ are independent given A)
- $P(M|A,\neg B,\neg E) = P(M|A)$
- $P(A \wedge \neg B \wedge \neg E) = P(A|\neg B,\neg E) \times P(\neg B|\neg E) \times P(\neg E)$
  $= P(A|\neg B,\neg E) \times P(\neg B) \times P(\neg E)$
  (¬B and ¬E are independent)
- $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) = P(J|A)P(M|A)P(A|\neg B,\neg E)P(\neg B)P(\neg E)$

Calculation of Joint Probability



$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$
$= P(J|A)P(M|A)P(A|\neg B,\neg E)P(\neg B)P(\neg E)$
$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$
$= 0.00062$

Calculation of Joint Probability

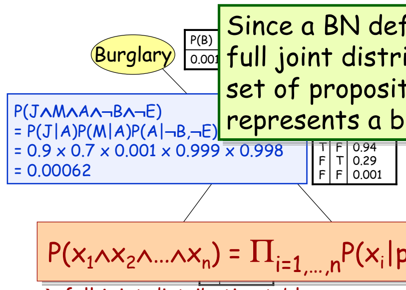


$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$
$= P(J|A)P(M|A)P(A|\neg B,\neg E)P(\neg B)P(\neg E)$
$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$
$= 0.00062$

$P(x_1 \wedge x_2 \wedge \ldots \wedge x_n) = \Pi_{i=1,\ldots,n} P(x_i|\text{parents}(X_i))$

→ full joint distribution table

Calculation of Joint Probability



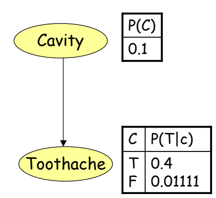
Since a BN defines the full joint distribution of a set of propositions, it represents a belief state

$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$
$= P(J|A)P(M|A)P(A|\neg B,\neg E)$
$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$
$= 0.00062$

$P(x_1 \wedge x_2 \wedge \ldots \wedge x_n) = \Pi_{i=1,\ldots,n} P(x_i|\text{parents}(X_i))$

→ full joint distribution table

Querying the BN



- The BN gives P(t|c)
- What about P(c|t)?
- P(Cavity|t)
  = P(Cavity ∧ t)/P(t)
  = P(t|Cavity) P(Cavity) / P(t)
  [Bayes' rule]
- P(c|t) = α P(t|c) P(c)
- Querying a BN is just applying the trivial Bayes' rule on a larger scale

Querying the BN

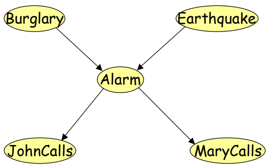
- New evidence E indicates that JohnCalls with some probability p
- We would like to know the posterior probability of the other beliefs, e.g. P(Burglary|E)
- $P(B|E) = P(B \wedge J|E) + P(B \wedge \neg J|E)$
  $= P(B|J,E)\, P(J|E) + P(B|\neg J,E)\, P(\neg J|E)$
  $= P(B|J)\, P(J|E) + P(B|\neg J)\, P(\neg J|E)$
  $= p\, P(B|J) + (1\text{-}p)\, P(B|\neg J)$
- We need to compute P(B|J) and P(B|¬J)

Querying the BN

- $P(b|J) = \alpha P(b \wedge J)$
  $= \alpha \Sigma_m \Sigma_a \Sigma_e P(b \wedge J \wedge m \wedge a \wedge e)$ [marginalization]
  $= \alpha \Sigma_m \Sigma_a \Sigma_e P(b)P(e)P(a|b,e)P(J|a)P(m|a)$ [BN]
  $= \alpha P(b) \Sigma_e P(e) \Sigma_a P(a|b,e)P(J|a) \Sigma_m P(m|a)$ [re-ordering]
- Depth-first evaluation of P(b|J) leads to computing each of the 4 following products twice:
  P(J|A) P(M|A), P(J|A) P(¬M|A), P(J|¬A) P(M|¬A), P(J|¬A) P(¬M|¬A)
- Bottom-up (right-to-left) computation + caching – e.g., variable elimination algorithm (see R&N) – avoids such repetition
- For singly connected BN, the computation takes time linear in the total number of CPT entries (→ time linear in the # propositions if CPT's size is bounded)

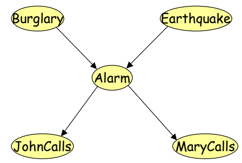
Singly Connected BN

A BN is singly connected if there is at most one undirected path between any two nodes



is singly connected

Comparison to Classical Logic

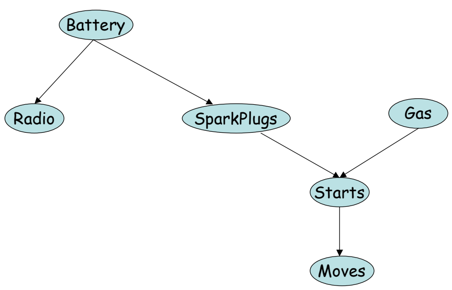


Burglary → Alarm
Earthquake → Alarm
Alarm → JohnCalls
Alarm → MaryCalls

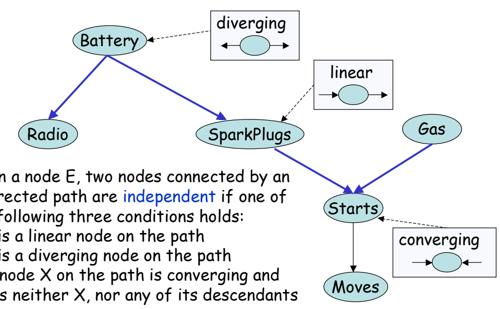
If the agent observes ¬JohnCalls, it infers ¬Alarm, ¬MaryCalls, ¬Burglary, and ¬Earthquake

If it observes JohnCalls, then it infers nothing

More Complicated Singly-Connected Belief Net



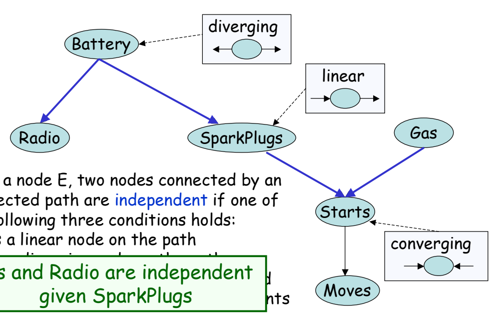
Independence Relations



Given a node E, two nodes connected by an undirected path are independent if one of the following three conditions holds:

- E is a linear node on the path
- E is a diverging node on the path
- A node X on the path is converging and E is neither X, nor any of its descendants

Independence Relations

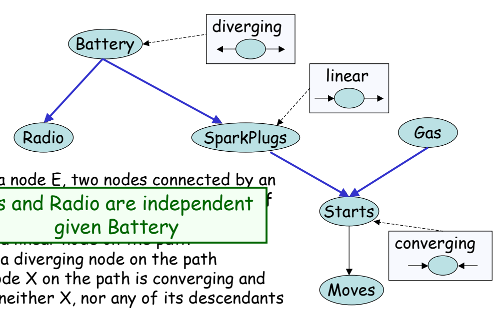


Given a node E, two nodes connected by an undirected path are independent if one of the following three conditions holds:

- E is a linear node on the path

Gas and Radio are independent given SparkPlugs

Independence Relations

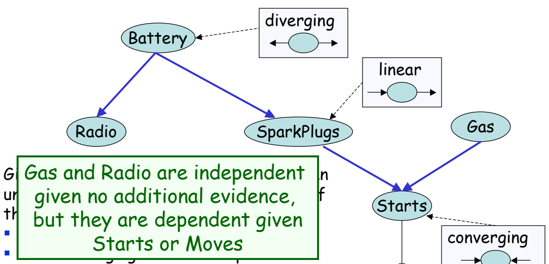


Given a node E, two nodes connected by an

Gas and Radio are independent given Battery

- E is a diverging node on the path
- A node X on the path is converging and E is neither X, nor any of its descendants

Independence Relations

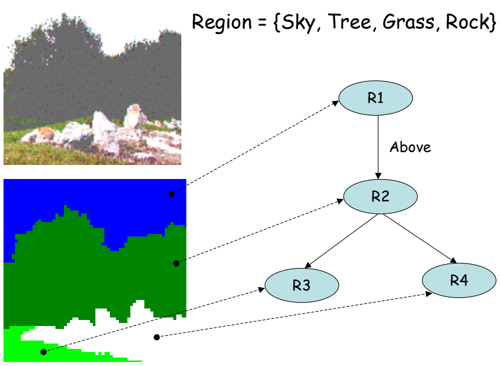


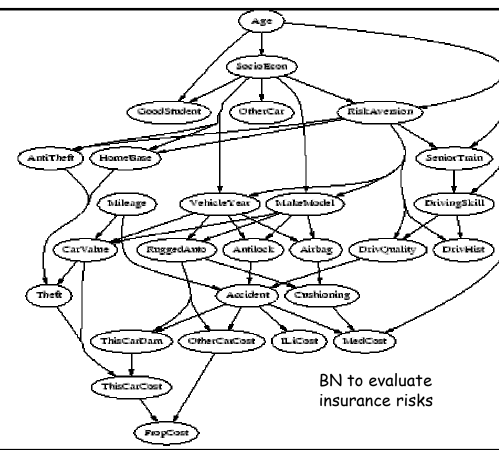
- A node X on the path is converging and E is neither X, nor any of its descendants

Some Applications of BN

- Medical diagnosis, e.g., lymph-node diseases
- Troubleshooting of hardware/software systems
- Fraud/uncollectible debt detection
- Data mining
- Analysis of genetic sequences
- Data interpretation, computer vision, image understanding

Region = {Sky, Tree, Grass, Rock}





BN to evaluate insurance risks