SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

Bayes & Uncertainty II

PROF LIM KWAN HUI

50.021 Artificial Intelligence

The following notes are compiled from various sources such as textbooks, lecture materials, Web resources and are shared for academic purposes only, intended for use by students registered for a specific course. In the interest of brevity, every source is not cited. The compiler of these notes gratefully acknowledges all such sources.

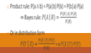
Recap: Bayes Rule

- Product rule: P(a ∧ b) = P(a|b) P(b) = P(b|a) P(a)

  ⇒ Bayes rule: $P(A \mid B) = \dfrac{P(B \mid A)\, P(A)}{P(B)}$

- Or in distribution form

  $$P(Y \mid X) = \frac{P(X \mid Y)\, P(Y)}{P(X)} = \alpha\, P(X \mid Y)\, P(Y)$$



Recap: Bayes Rule

- Product rule: P(a ∧ b) = P(a|b) P(b) = P(b|a) P(a)

  ⇒ Bayes rule: $P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$

- Or in distribution form

  $$P(Y \mid X) = \frac{P(X \mid Y)\, P(Y)}{P(X)} = \alpha\, P(X \mid Y)\, P(Y)$$

- Useful for assessing diagnostic probability from causal probability:
  - $P(Cause \mid Effect) = \frac{P(Effect \mid Cause)\, P(Cause)}{P(Effect)}$

- E.g., let M be meningitis, S be stiff neck:
  - $P(m \mid s) = \frac{P(s \mid m)\, P(m)}{P(s)} = \frac{0.8\ X\ 0.0001}{0.1}$ = 0.0008
  - Note: posterior probability of meningitis still very small!

Exercise

- Given that the test for a disease is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative if you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people.
  - What is it good news that the disease is rare?
  - What are the actual chances of getting the disease, if the test was positive?

Exercise

- Given that the test for a disease is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative if you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people.
  - What is it a good news that the disease is rare?
    - Since the posterior probability of having the disease depends on the prior probability times the probability of testing positive, a single positive test does not make a big difference to the very low prior.
    - E.g., P(d|+) = P(+|d) P(d) / P(+), the posterior P(d|+) depends on prior P(d)

Exercise

- Given that the test for a disease is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative if you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people.
  - What are the actual chances of getting the disease, if the test was positive?
    - Given P(+|d) = 0.99, P(-| !d) = 0.99, and P(d) = 0.0001
      - Hence, P(-|d) = 0.01, and P(+| !d) = 0.01

Exercise

- Given that the test for a disease is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative if you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people.
  - What are the actual chances of getting the disease, if the test was positive?
    - Given P(+|d) = 0.99, P(-| !d) = 0.99, and P(d) = 0.0001
      - Hence, P(-|d) = 0.01, and P(+| !d) = 0.01
    - P(d|+) = P(+|d)P(d) / P(+)

      = P(+|d)P(d) / [P(+, d) + P(+, !d)]

      = P(+|d)P(d) / [P(+|d)P(d) + P(+| !d)P(!d)]

      = 0.99 x 0.0001 / [0.99 x 0.0001 + 0.01 x 0.9999]

      = 0.000099 / [0.000099 + 0.009999]

      = ~0.0098

Why Bayesian networks?

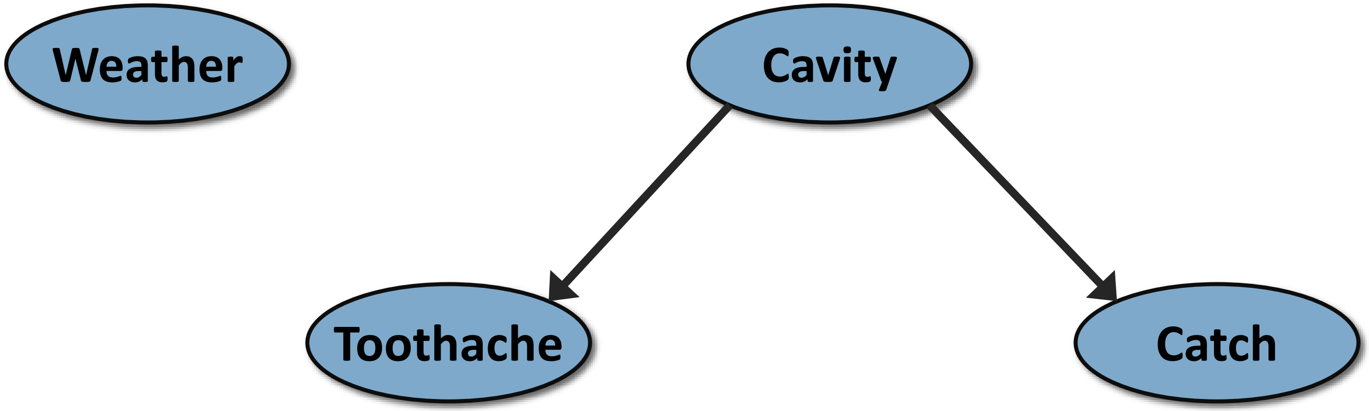
- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time
- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions

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 variable
  - a set of arcs, representing "directly influences"
  - a conditional distribution for each node given its parents: $P(X_i|Parents(X_i))$
- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values

Example: Toothache

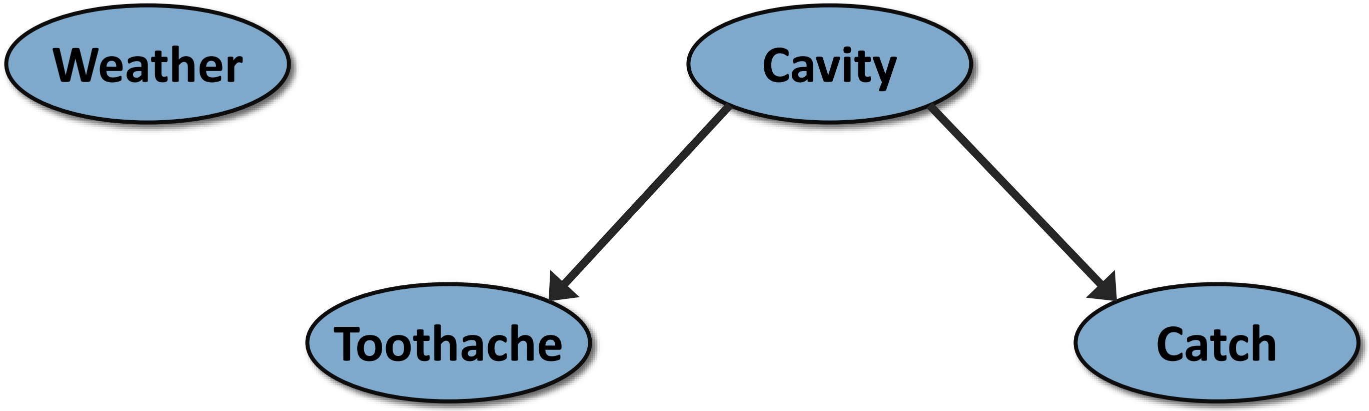
- Topology of network encodes conditional independence assertions:



- Weather is independent of the other variables
- Toothache and Catch are conditionally independent given Cavity

Example: Toothache

- Originally, we had: P(Toothache,Catch,Cavity,Weather)



- Now, we have:
  - P(Weather) P(Toothache,Catch,Cavity)

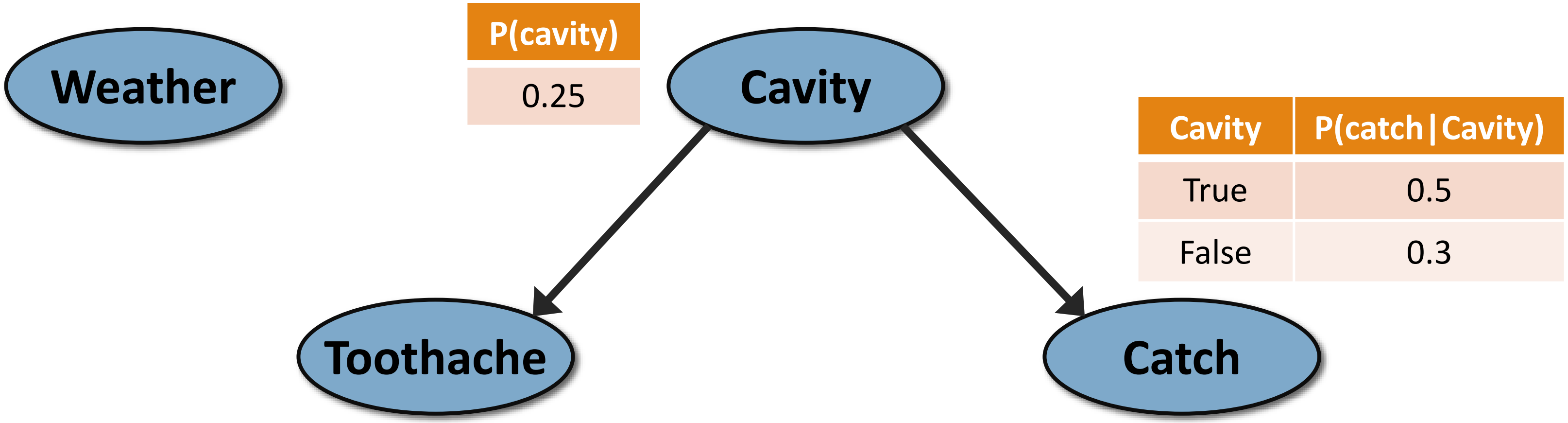
    = P(Weather) P(Toothache | Catch,Cavity) P(Catch,Cavity)

    = P(Weather) P(Toothache | Catch,Cavity) P(Catch | Cavity) P(Cavity)

    = P(Weather) P(Toothache | Cavity) P(Catch | Cavity) P(Cavity)

Example: Toothache

- Originally, we had: P(Toothache,Catch,Cavity,Weather)



P(cavity)
0.25

Cavity	P(catch|Cavity)
True	0.5
False	0.3

- Now, we have:
  - P(Weather) P(Toothache,Catch,Cavity)

    = P(Weather) P(Toothache | Catch,Cavity) P(Catch,Cavity)

    = P(Weather) P(Toothache | Catch,Cavity) P(Catch | Cavity) P(Cavity)

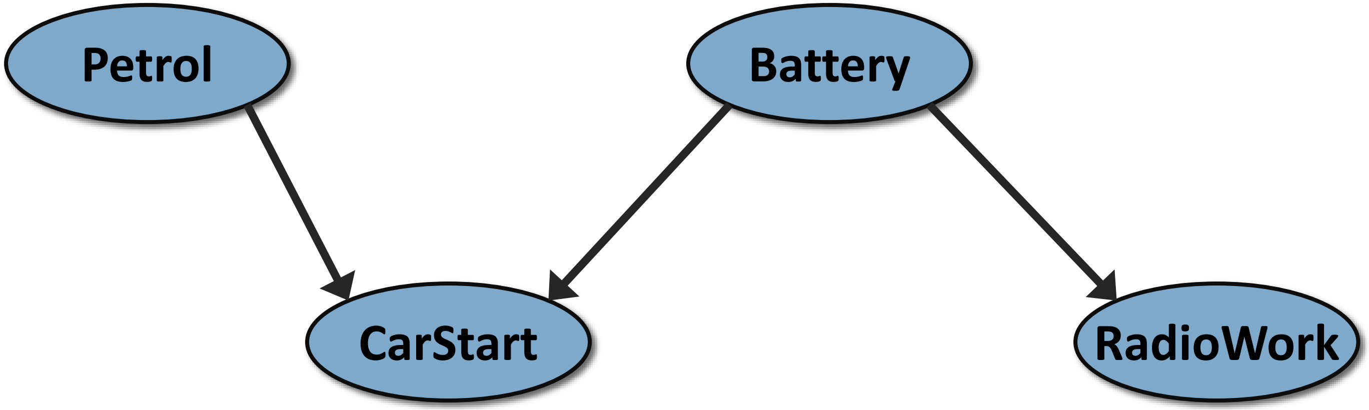
    = P(Weather) P(Toothache | Cavity) P(Catch | Cavity) P(Cavity)

Exercise: Simple Car Diagnosis

- Consider a simple car diagnosis where you need to diagnose why your car is not starting. A lack of petrol or low battery could cause the car to not start. Similarly, low battery might cause the radio to not work. Draw out the Bayesian network corresponding to this setup.

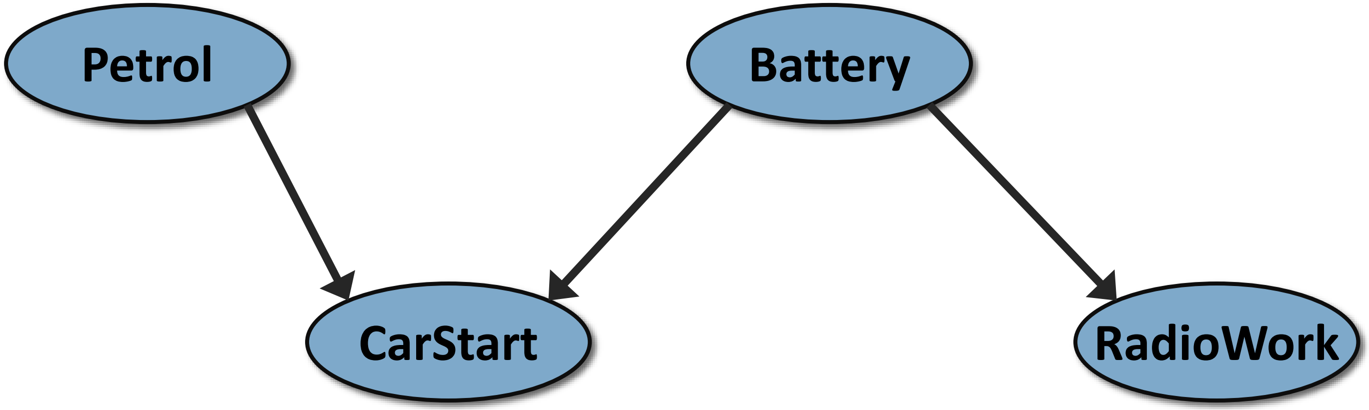
Exercise: Simple Car Diagnosis

- Consider a simple car diagnosis where you need to diagnose why your car is not starting. A lack of petrol or low battery could cause the car to not start. Similarly, low battery might cause the radio to not work. Draw out the Bayesian network corresponding to this setup.



Exercise: Simple Car Diagnosis

- What conditional probability tables would you need to provide use this network, and what is the minimum number of entries you would need to specify in these tables?

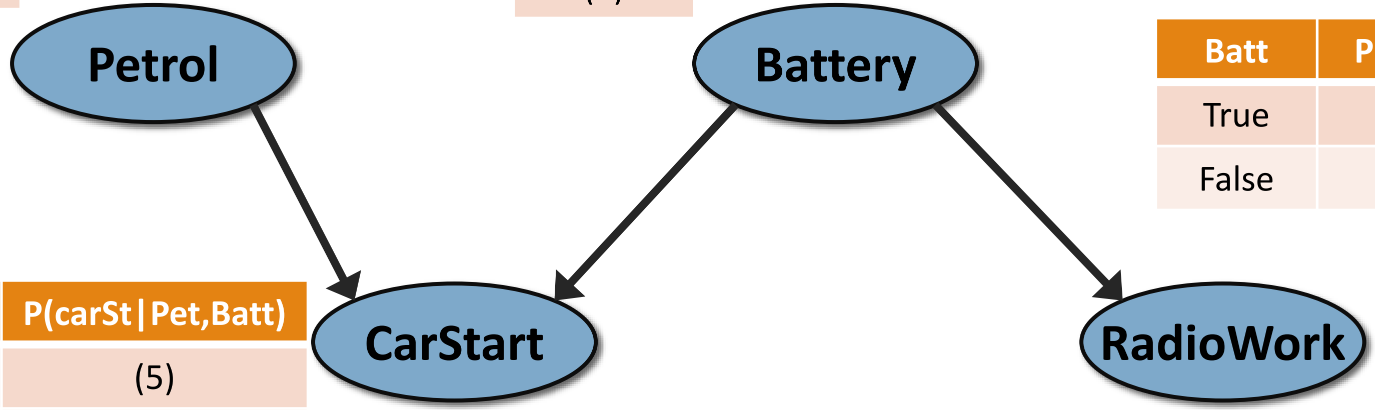


Exercise: Simple Car Diagnosis

- What conditional probability tables would you need to provide use this network, and what is the minimum number of entries you would need to specify in these tables?

P(petrol)
(1)

P(batt)
(2)



Batt	P(radio|Batt)
True	(3)
False	(4)

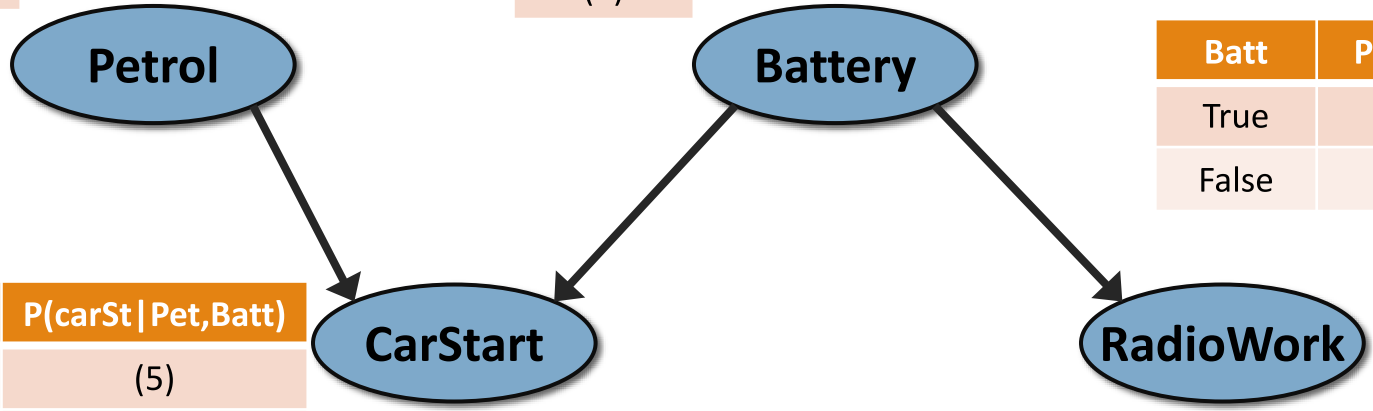
Petr	Batt	P(carSt|Pet,Batt)
True	True	(5)
True	False	(6)
False	True	(7)
False	False	(8)

Exercise: Simple Car Diagnosis

- Assume that the car does not start, but the radio works. How would you calculate the probability that the car is out of petrol?

P(petrol)
(1)

P(batt)
(2)



Batt	P(radio|Batt)
True	(3)
False	(4)

Petr	Batt	P(carSt|Pet,Batt)
True	True	(5)
True	False	(6)
False	True	(7)
False	False	(8)

Exercise: Simple Car Diagnosis

- Assume that the car does not start, but the radio works. How would you calculate the probability that the car is out of petrol?
  - We have the following r.v.: P, C, B, R
  - Given evidence C = !c (i.e., false), R = r, with query P = !p (does not have petrol):
  - P(!p | !c, r) = ?

Exercise: Simple Car Diagnosis

- Assume that the car does not start, but the radio works. How would you calculate the probability that the car is out of petrol?
  - We have the following r.v.: P, C, B, R
  - Given evidence C = !c (i.e., false), R = r, with query P = !p (does not have petrol):

  - $P(!p \mid !c, r) = P(!p, !c, r) / P(!c, r)$

$= \Sigma_B P(!p, !c, B, r) / \Sigma_B \Sigma_P P(P, !c, B, r)$

$= \Sigma_B P(!p) P(!c \mid !p, B) P(B) P(r \mid B) / \Sigma_B \Sigma_P P(P) P(!c \mid P, B) P(B) P(r \mid B)$

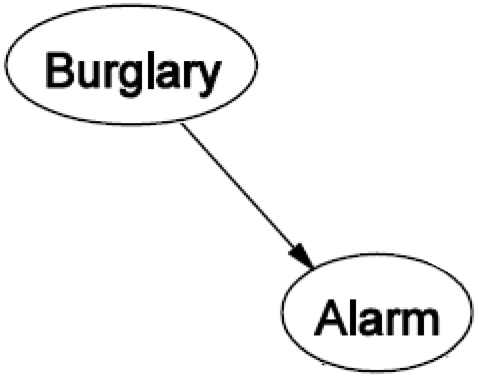
$= P(!p) \Sigma_B P(!c \mid !p, B) P(B) P(r \mid B) / \Sigma_B P(B) \Sigma_P P(P) P(!c \mid P, B) P(r \mid B)$

Example: Earthquake

- Scenario: I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?
- Variables: Burglar, Earthquake, Alarm, JohnCalls, MaryCalls
- Network topology reflects “causal” knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call

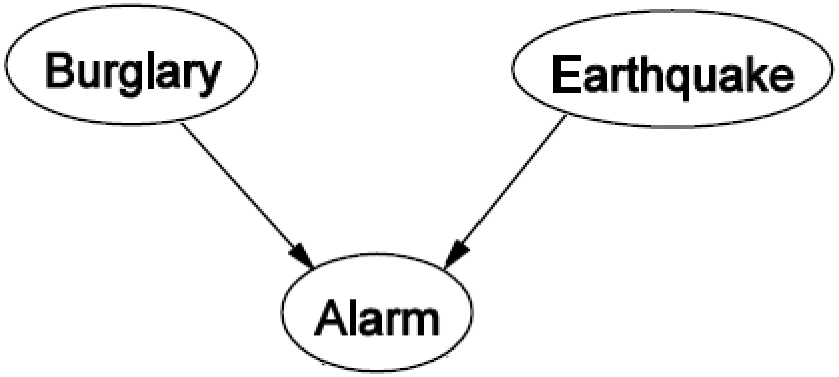
Example: Earthquake

- A burglar can set the alarm off



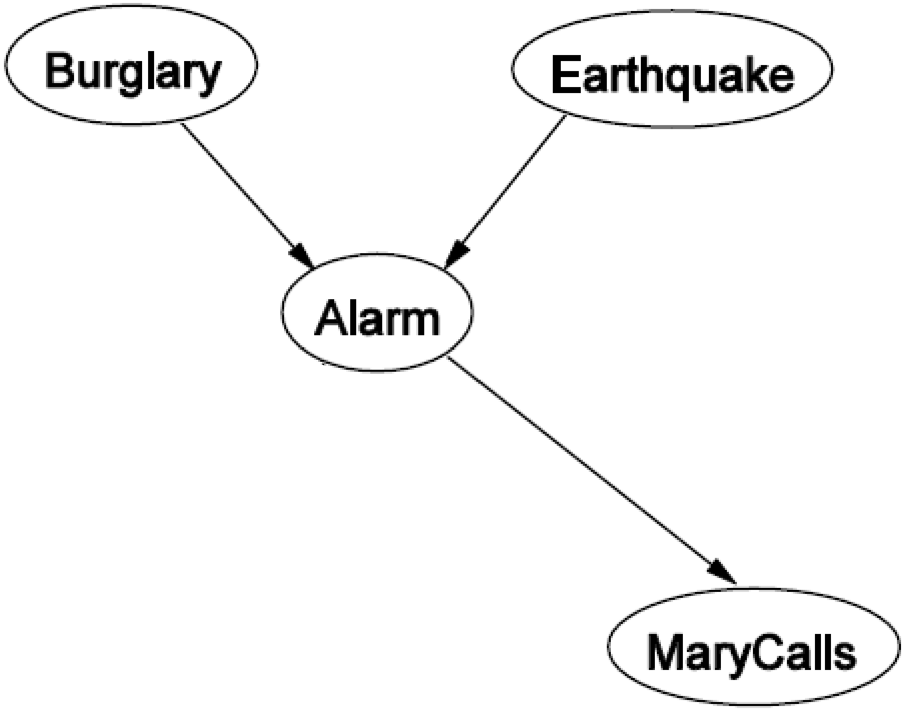
Example: Earthquake

- An earthquake can set the alarm off



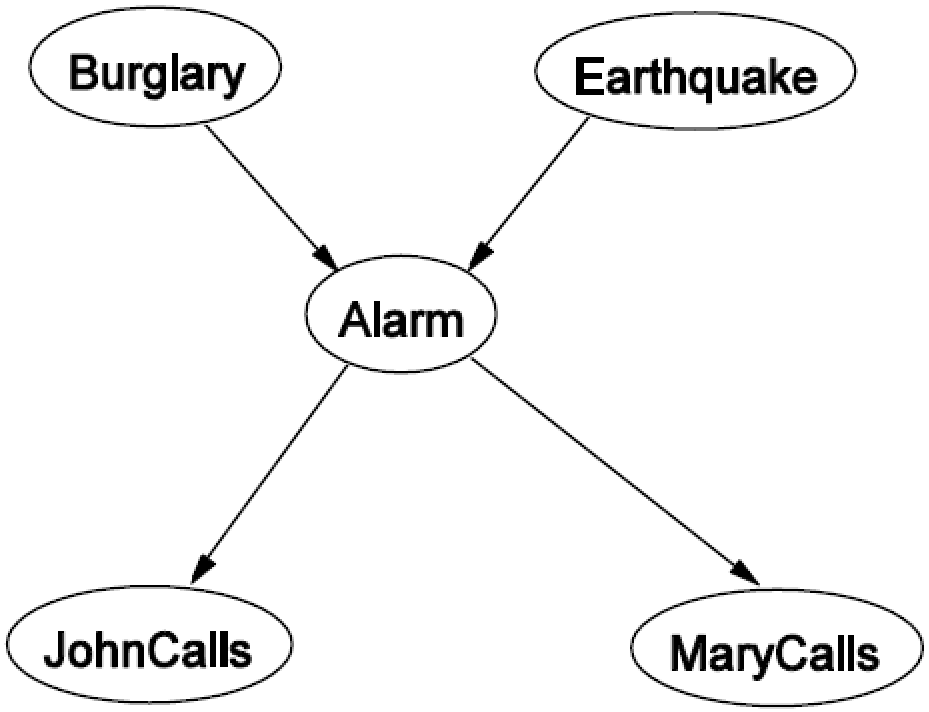
Example: Earthquake

- The alarm can cause Mary to call

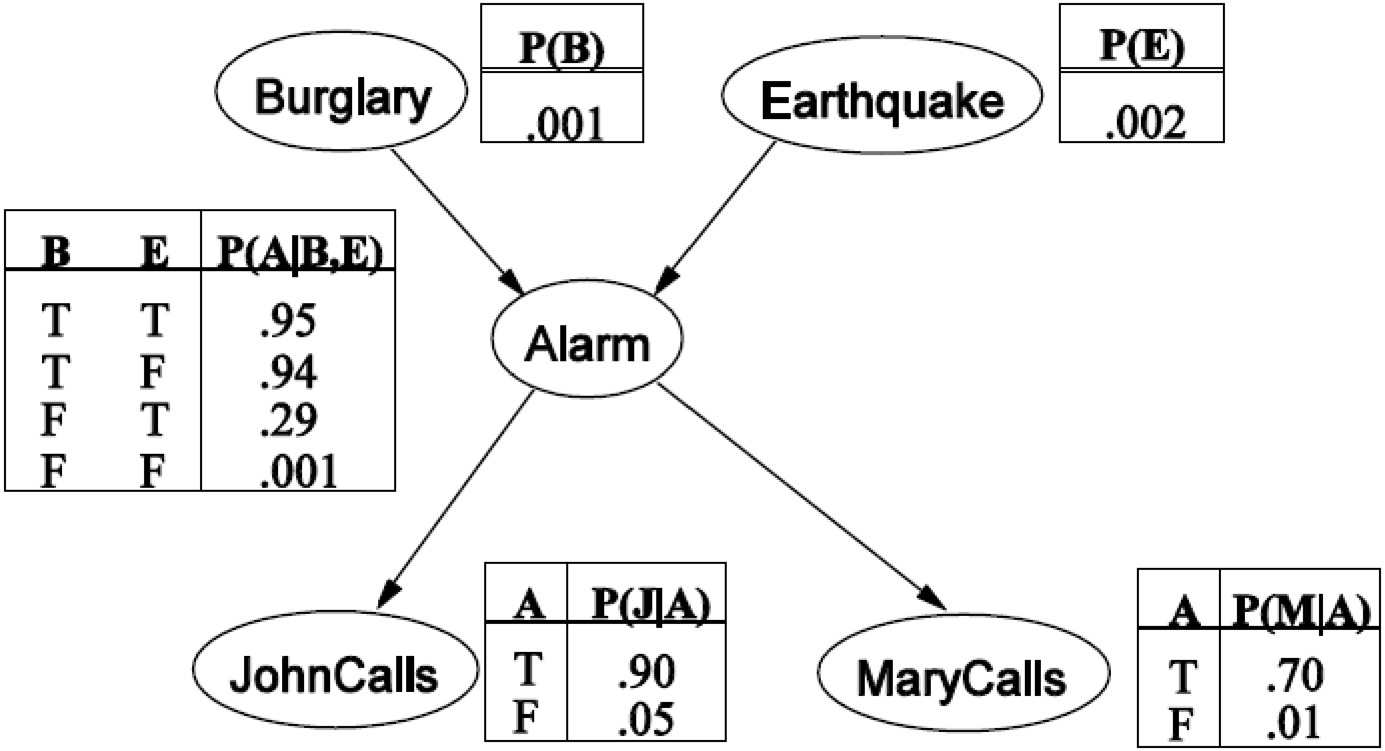


Example: Earthquake

- The alarm can cause John to call



Example: Earthquake



P(B)
.001

P(E)
.002

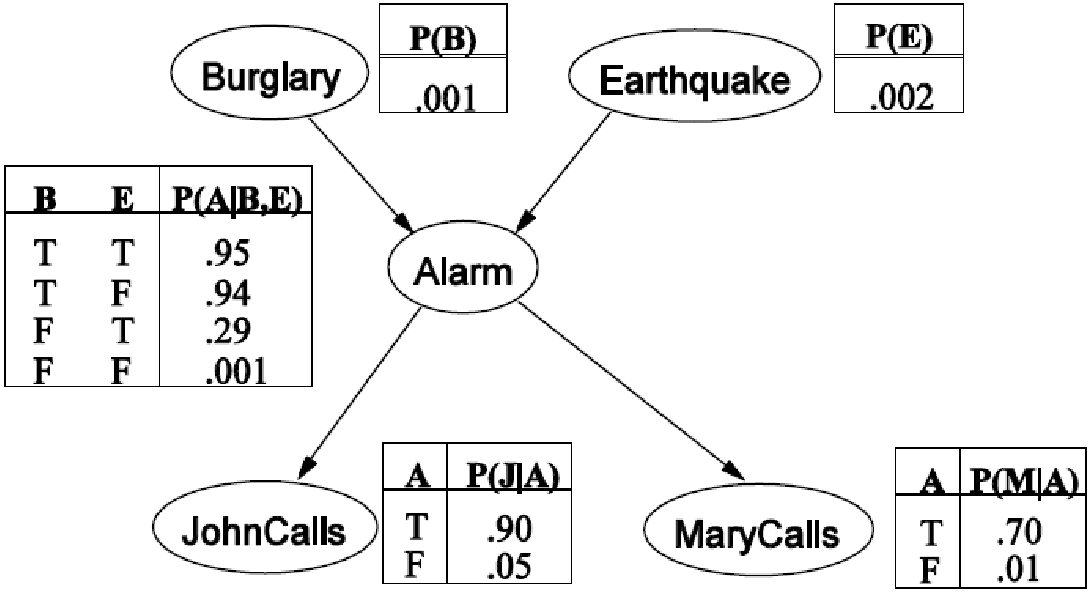
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Exercise: Earthquake

- You are at work and hear on the radio that a minor earthquake has just occurred near your home. John tried to call you, but there has been no call from Mary. How would you calculate the probability of a burglary having occurred given the above information?



P(B)
.001

P(E)
.002

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

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Exercise: Earthquake

- You are at work and hear on the radio that a ***minor earthquake*** has just occurred near your home. ***John tried to call*** you, but there has been ***no call from Mary***. How would you calculate the ***probability of a burglary*** having occurred given the above information?
  - Given evidence E = e, J = j and M = !m, what is the probability of the query B = b, i.e., P(b|e,j,!m)?
  - $P(b|e,j,!m) = \alpha\, P(b,e,j,!m)$
    $= \alpha\, \Sigma_A\, P(b,e,A,j,!m)$
    $= \alpha\, \Sigma_A\, P(b)\, P(e)\, P(A|e,b)\, P(j|A)\, P(!m|A)$
  - Where $\alpha = 1/\Sigma_A\Sigma_B P(B, e, A, j, !m)$
    $= 1/\Sigma_A\Sigma_B P(e)\, P(B)\, P(A|e,B)\, P(j|A)\, P(!m|A)$
    $= 1/P(e)\Sigma_A P(j|A)\, P(!m|A)\, \Sigma_B\, P(B)\, P(A|e,B)$

Global semantics

- Global semantics defines the full joint distribution as the product of the local conditional distributions:

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

- e.g., $\mathbf{P}(j \land m \land a \land \neg b \land \neg e)$

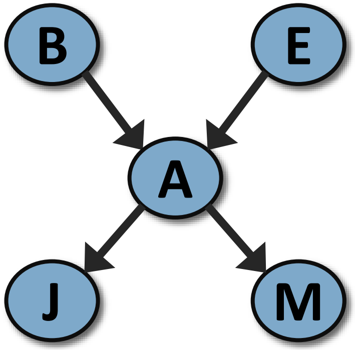
$$= \mathbf{P}\,(j \mid a)\, \mathbf{P}\,(m \mid a)\, \mathbf{P}\,(a \mid \neg b, \neg e)\, \mathbf{P}\,(\neg b)\, \mathbf{P}\,(\neg e)$$

Inference by enumeration

- Slightly intelligent way to sum out variables from the joint without actually constructing its explicit representation

- Simple query on the burglary network:
  - P(B|j,m)

$$= P(B, j, m) / P(j,m)$$
$$= \alpha\, P(B, j, m)$$
$$= \alpha \sum_e \sum_a P(B, e, a, j, m)$$

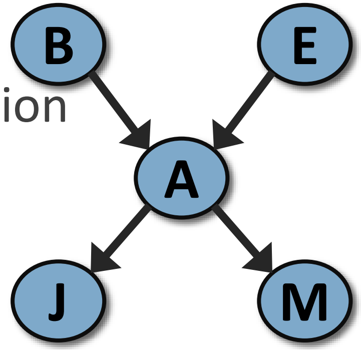


- Rewrite full joint entries using product of CPT entries:
  - P(B|j,m)

$$= \alpha \sum_e \sum_a P(B)P(e)P(a|B, e)P(j|a)P(m|a)$$
$$= \alpha\, P(B) \sum_e P(e) \sum_a P(a|B, e)P(j|a)P(m|a)$$

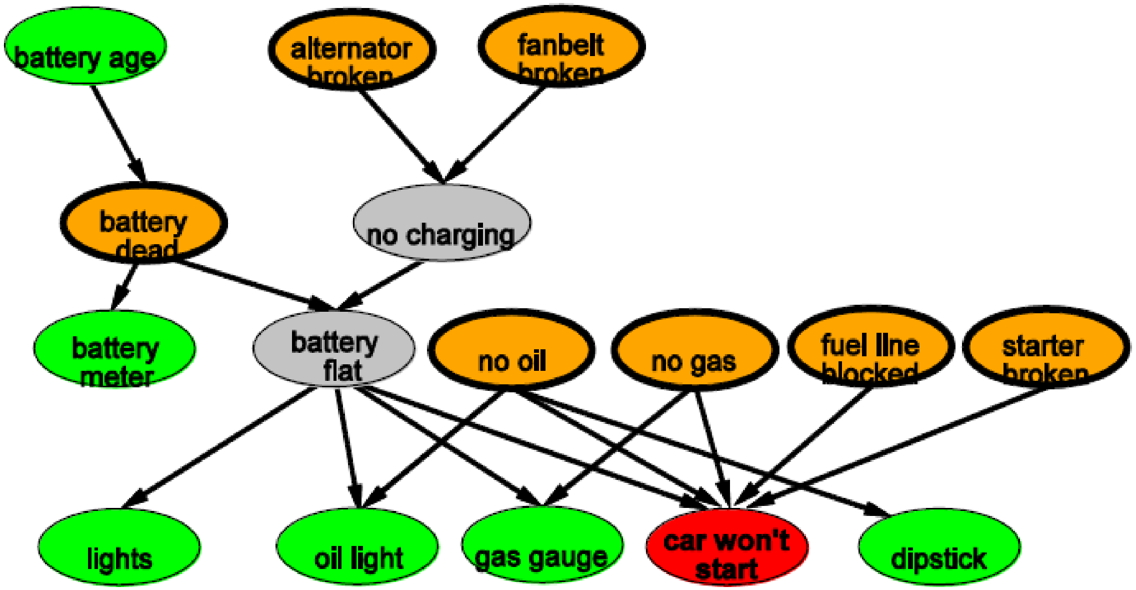
Compactness

- A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values
- Each row requires one number p for X_i = true
  - (the number for X_i = false is just 1 - p)
- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers
- I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution
- For burglary net, 1 + 1 + 4 + 2 + 2=10 numbers (vs. $2^5 = 32$)

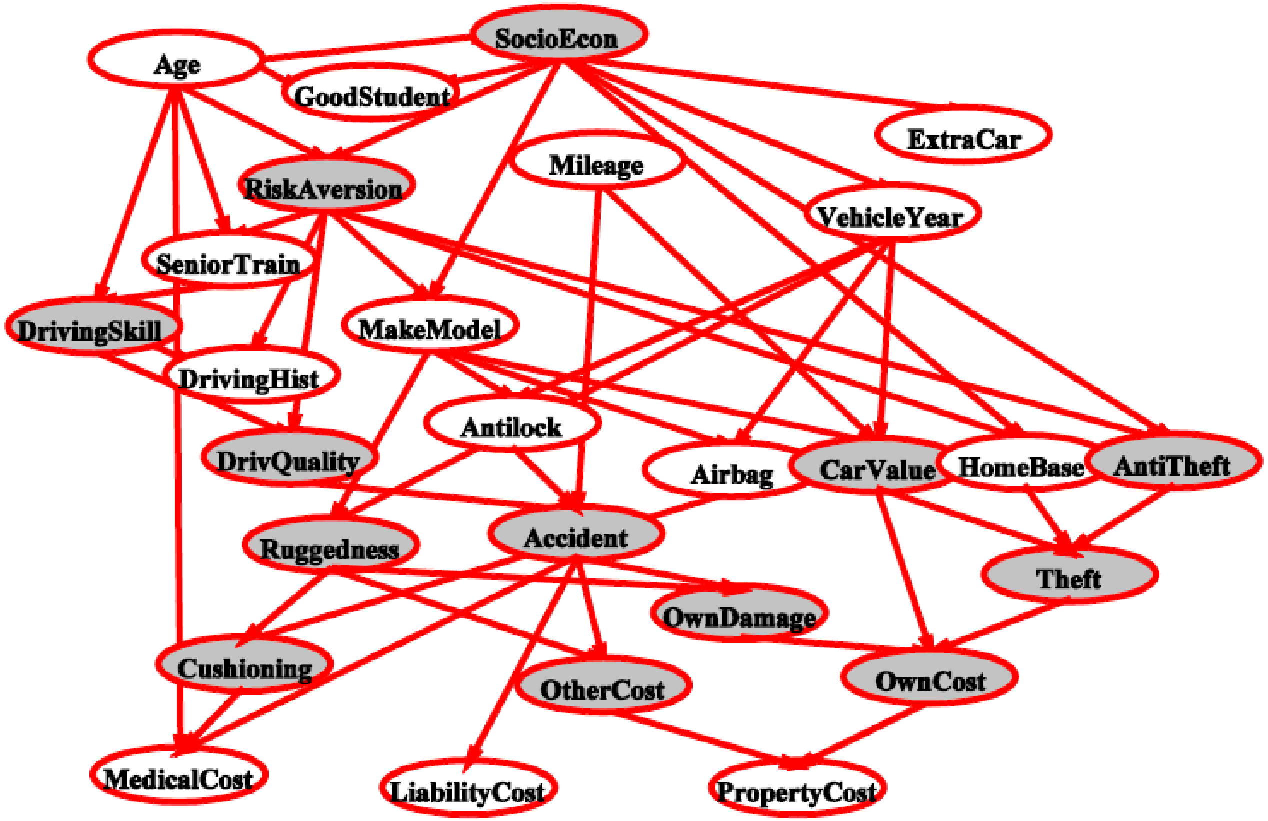


Example: Car diagnosis

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



Example: Car Insurance



Next

- Learn about the Naïve Bayes Classifier and its application to text