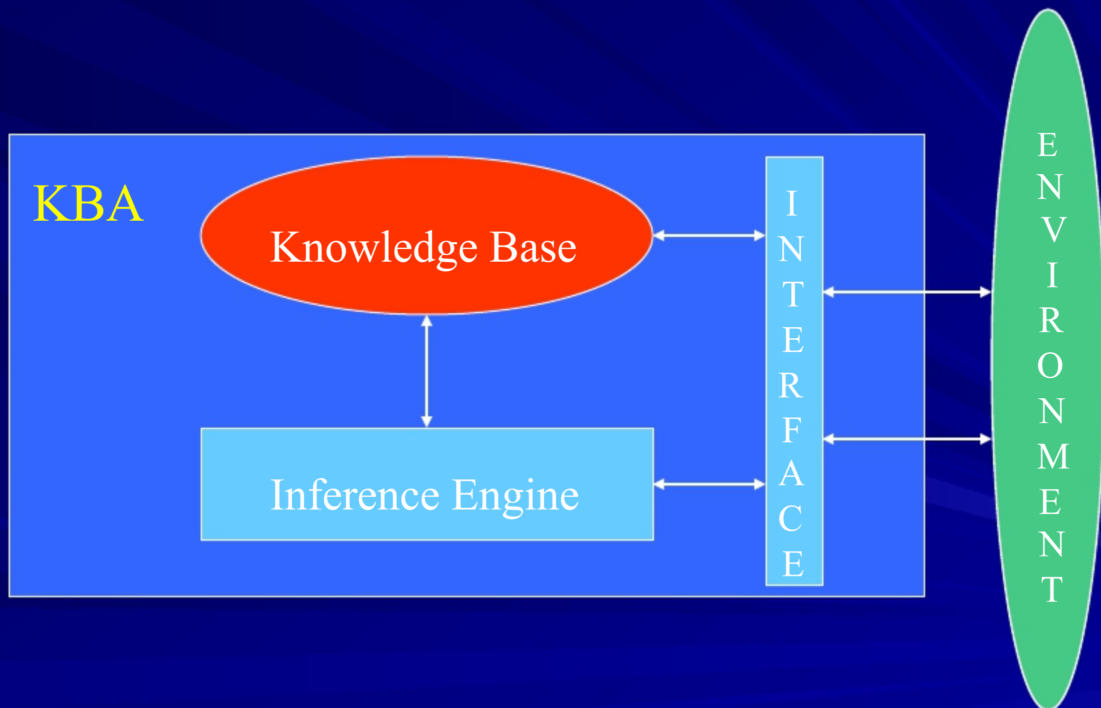
# Bayesian Networks: Knowledge Representation and Inference

© Dr. Leonardo Garrido
Departamento de Computación, Región Norte
Tecnológico de Monterrey, Campus Monterrey
E-mail: leonardo.garrido@itesm.mx
LinkedIn page: https://www.linkedin.com/in/drleonardogarrido/
Tec page: http://homepages.mty.itesm.mx/lgarrido

# Knowledge-Based Agents

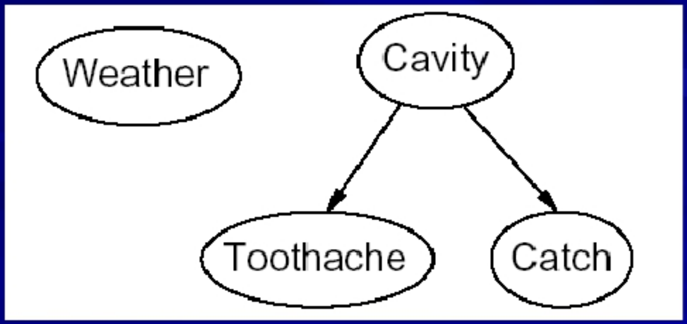


# What is a Bayesian Net?

- It is a data structure.
- It represents the dependencies among variables
- It gives a concise specification of any full joint probability distribution.
- It takes advantage of independence and conditional independence relationships among variables.
- It sometimes is also called belief network, probabilistic network, causal network, or knowledge map.
- Graphical models in statistics include Bayesian nets.
- Decision networks or influence diagrams are extensions of Bayesian nets.

# A simple example

- Weather is independent of the other variables
- The conditional independence of Toothache and Catch given Cavity is indicated by the absence of a link between Toothache and Catch.
- Cavity is a direct cause of Toothache and Catch, whereas no direct causal relationship exists between Toothache and Catch.



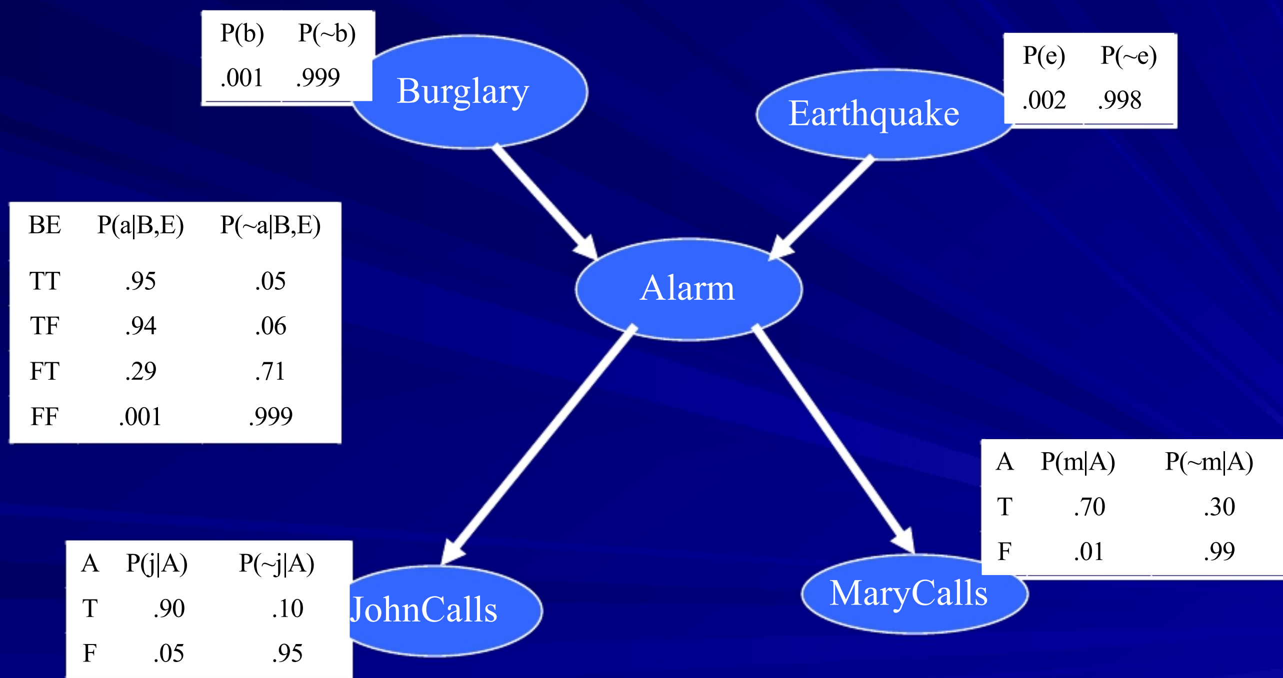
# Another example: Predicting an Earthquake

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

# The classic net




| P(b) | P(~b) |
|---|---|
| .001 | .999 |

| P(e) | P(~e) |
|---|---|
| .002 | .998 |

| BE | P(a\|B,E) | P(~a\|B,E) |
|---|---|---|
| TT | .95 | .05 |
| TF | .94 | .06 |
| FT | .29 | .71 |
| FF | .001 | .999 |

| A | P(j\|A) | P(~j\|A) |
|---|---|---|
| T | .90 | .10 |
| F | .05 | .95 |

| A | P(m\|A) | P(~m\|A) |
|---|---|---|
| T | .70 | .30 |
| F | .01 | .99 |

# Bayesian Networks

- A Bayesian network is a graph in which each node is annotated with quantitative probability information:
  - A set of random variables makes up the nodes of the network (discrete or continuous).
  - A set of directed links or arrows connects pairs of nodes (if there is a node from node X to node Y, X is said to be a parent of Y).
  - Each node $X_i$ has a conditional probability distribution $P(Xi \mid \text{Parents}\ (X_i))$.
  - The graph is directed and acyclic (it has no directed cycles).

# Semantics

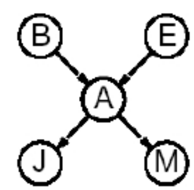
- A Bayesian Network provides a complete description of the domain.
- Every entry in the full joint probability distribution can be calculated from the information in the network.

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

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

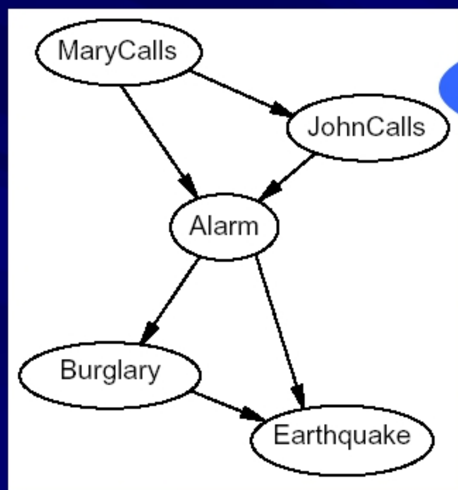
e.g., $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

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

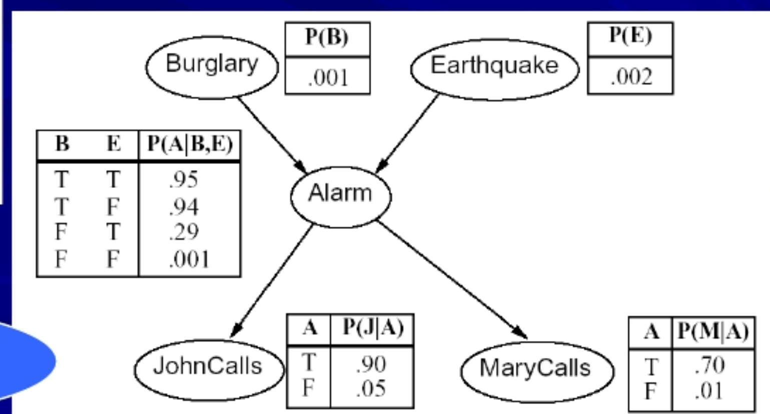



= 0.90 X 0.70 X 0.001 X 0.999 X 0.998 = 0.00062

# Comparing topological designs




It needs 13 probabilities




| P(B) |
|---|
| .001 |

| P(E) |
|---|
| .002 |

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

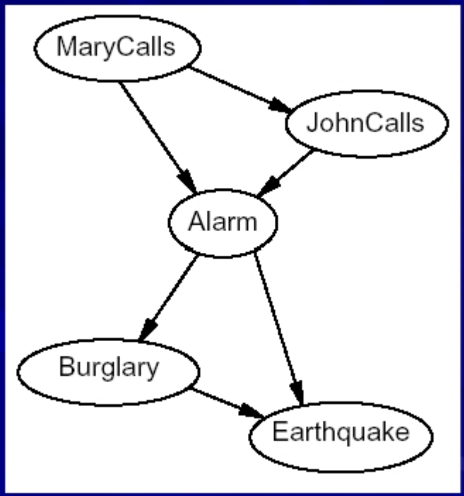
| A | P(J\|A) |
|---|---|
| T | .90 |
| F | .05 |

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

It needs 10 probabilities

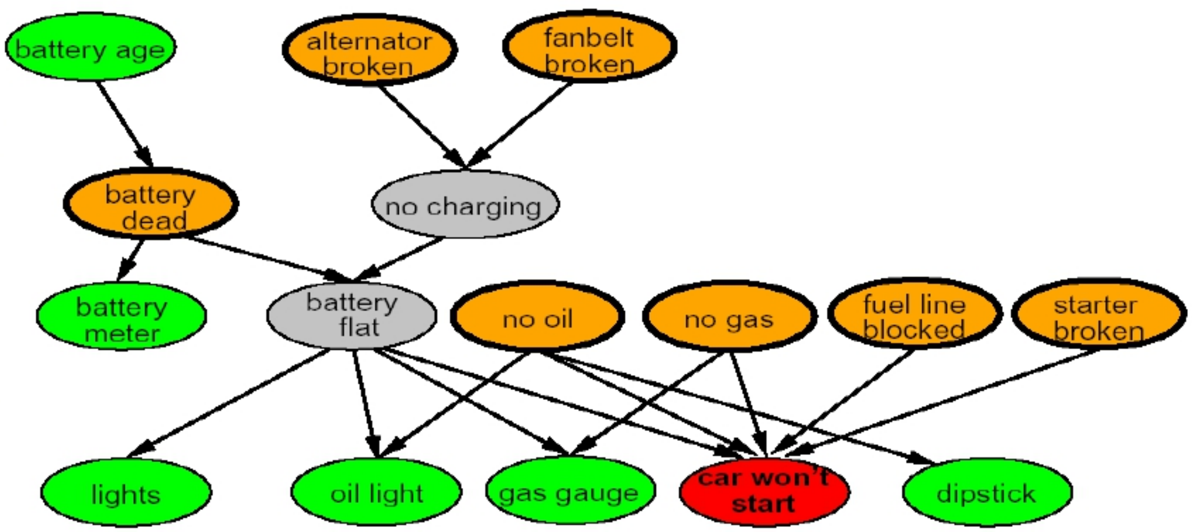
# Are non-causal models better?

- Deciding conditional independence is hard in non causal directions.
- Assessing conditional probabilities is hard in non causal directions.
- Causal models and conditional independence seem to be hardwired in humans!

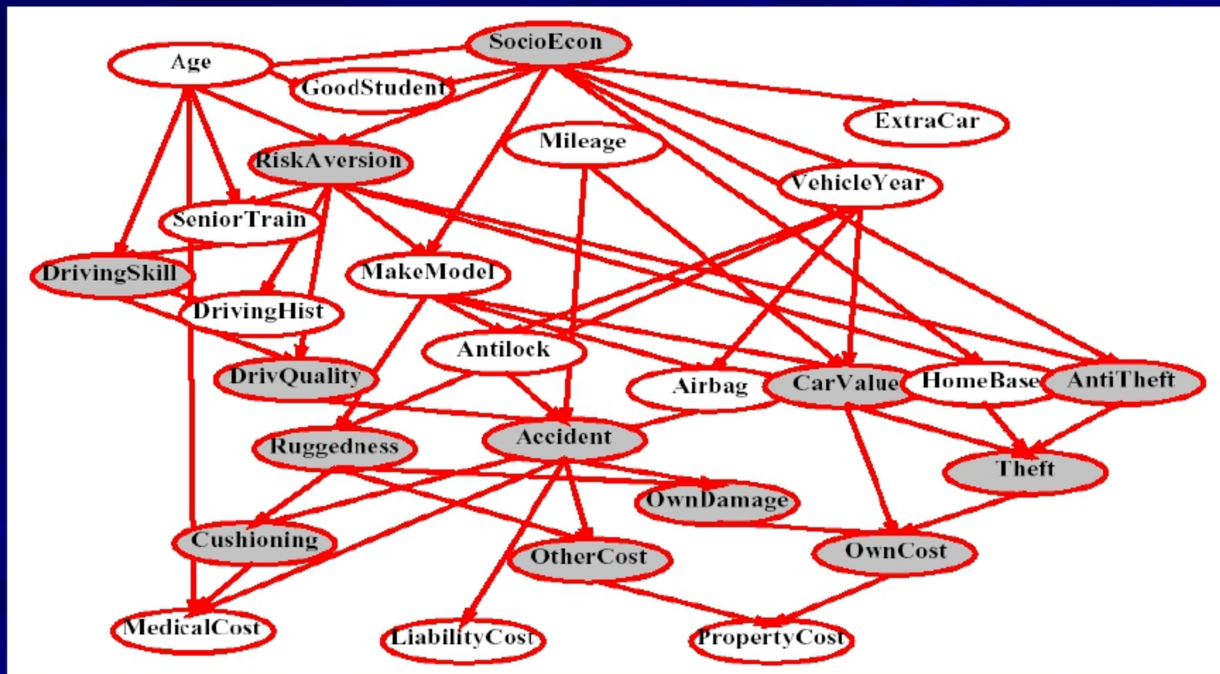


# Another example: Car diagnosis

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



# Another example: Car insurance



# The most simple net

- Only two nodes:
  - Enfermedad (Cavity)
  - Sintoma (Toothache)

- $P(C) = 0.01$, $P(T|C) = 0.9$, $P(T|-C) = 0.2$

- Bayes Network?
- Full Joint Distribution?
- $P(C,T) = P(C)*P(T|C) = 0.01*0.9 = 0.009$
- $P(C,-T) = P(C)*P(-T|C) = 0.01*0.1 = 0.001$

  $P(-C,T) = P(-C)*P(T|-C) = 0.99*0.2 = 0.198$

  $P(-C,-T) = P(-C)*P(-T|-C) = 0.99*0.8 = 0.792$

# The most simple net

- Only two nodes:
  - Enfermedad (Cavity)
  - Sintoma (Toothache)
- $P(C) = 0.01$, $P(T|C) = 0.9$, $P(T|-C) = 0.2$
- Bayes Network?
- $P(C|T) = P(C,T)/P(T) = P(C,T)/[P(T,C)+P(T,-C)]$
- $=0.009/(0.009+0.198) = 0.04347$
- 

# The most simple net

- Only two nodes:
  - Enfermedad (Cavity)
  - Sintoma (Toothache)

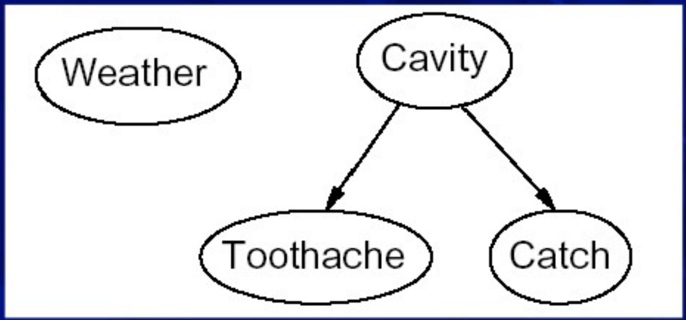
- $P(C) = 0.01$, $P(T|C) = 0.9$, $P(T|-C) = 0.2$

- Bayes Network?

- $P(C|-T) = P(C,-T)/P(-T) = P(C,-T)/[P(-T,C)+P(-T,-C)]$
- $= 0.001/(0.001+0.792) = 0.00126$
-

# A little bit more complicate net

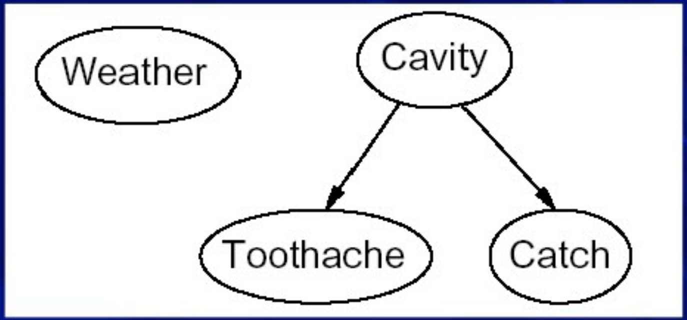
- Three nodes:
  - $P(C) = 0.01$
  - $P(T|C) = 0.9$,
  - $P(T|-C) = 0.2$
  - $P(Ch|C) = 0.9$,
  - $P(Ch|-C) = 0.2$




- $P(C|T,Ch) = ?$
- $P(-C|T,Ch) = ?$

# A little bit more complicate net

- Three nodes:
- P(C) = 0.01
- P(T|C) = 0.9,
- P(T|-C) = 0.2
- P(Ch|C) = 0.9,
- P(Ch|-C) = 0.2




- P(C|T,Ch) = P(T,Ch|C) P(C) / P(T,Ch)
- = P(T|C) P(Ch|C) P(C) / P(T,Ch)

  = 0.9*0.9*0.01/P(T,Ch) = 0.0081 → 0.1698

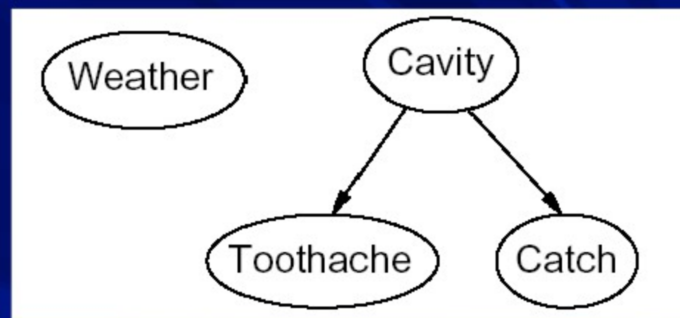
P(-C|T,Ch) = P(T,Ch|-C) P(-C) / P(T,Ch)

= P(T|-C) P(Ch|-C) P(-C) / P(T,Ch)

= 0.2*0.2*0.99/P(T,Ch) = 0.0396 → 0.8301

# A little bit more complicate net

- Three nodes:
  - $P(C) = 0.01$
  - $P(T|C) = 0.9$,
  - $P(T|-C) = 0.2$
  - $P(Ch|C) = 0.9$,
  - $P(Ch|-C) = 0.2$



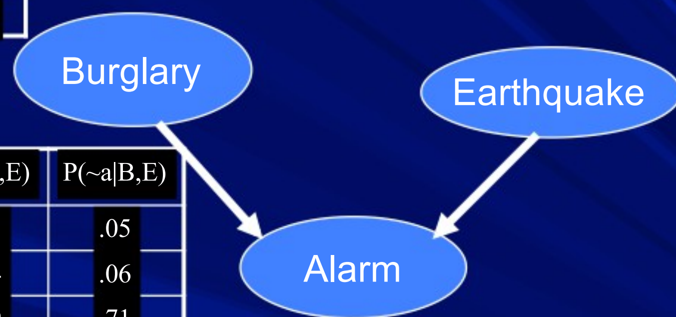

- $P(C|T,-Ch) = P(T,-Ch|C)\ P(C)\ /\ P(T,-Ch)$
- $= P(T|C)\ P(-Ch|C)\ P(C)\ /\ P(T,-Ch)$
- $= 0.9*0.1*0.01/P(T,Ch) = 0.0009 \rightarrow 0.0056$
- $P(-C|T,-Ch) = P(T,-Ch|-C)\ P(-C)\ /\ P(T,-Ch)$

  $= P(T|-C)\ P(-Ch|-C)\ P(-C)\ /\ P(T,-Ch)$

  $= 0.2*0.8*0.99/P(T,Ch) = 0.1584 \rightarrow 0.9943$

# Explaining away

| P(b) | P(~b) |
|---|---|
| .001 | .999 |



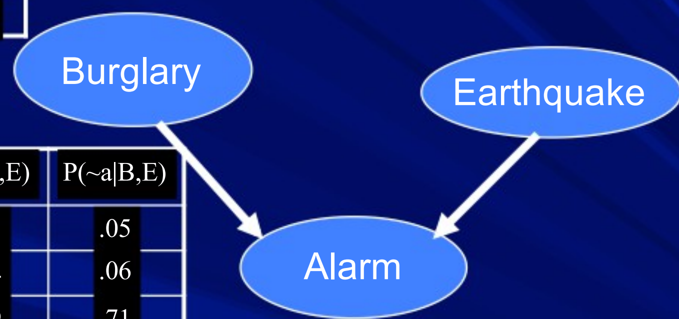

| P(e) | P(~e) |
|---|---|
| .002 | .998 |

| B E | P(a\|B,E) | P(~a\|B,E) |
|---|---|---|
| T T | .95 | .05 |
| T F | .94 | .06 |
| FT | .29 | .71 |
| F F | .001 | .999 |

- What is the probability P(b|e) = P(b) !!! = 0.001

# Explaining away

| P(b) | P(~b) |
|---|---|
| .001 | .999 |




| P(e) | P(~e) |
|---|---|
| .002 | .998 |

| B E | P(a\|B,E) | P(~a\|B,E) |
|---|---|---|
| T T | .95 | .05 |
| T F | .94 | .06 |
| FT | .29 | .71 |
| F F | .001 | .999 |

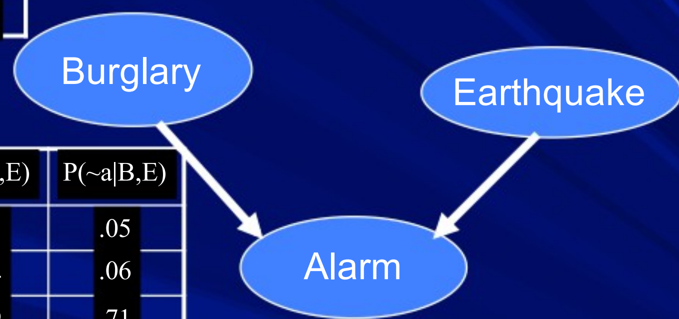
- $P(b|a,e) = P(a|b,e)\ P(b|e)\ P(e)\ /P(a|e)\ P(e) = P(a|b,e)\ P(b|e)\ /P(a|e)$
- $= 0.95 * 0.001\ /\ P(a|e)$
- $= 0.00095\ /\ [P(a|e,b)\ P(b) + P(a|e,\text{-}b)\ P(\text{-}b)]$

$= 0.00095\ /\ 0.95*0.001 + 0.29*0.999$

$= 0.00095\ /\ 0.29066 = 0.003268$

# Explaining away

| P(b) | P(~b) |
|---|---|
| .001 | .999 |




| P(e) | P(~e) |
|---|---|
| .002 | .998 |

| B E | P(a\|B,E) | P(~a\|B,E) |
|---|---|---|
| T T | .95 | .05 |
| T F | .94 | .06 |
| FT | .29 | .71 |
| F F | .001 | .999 |

- $P(b|a) = P(a|b)\, P(b) / P(a)$
- $P(a) = P(a|b,e)P(b,e) + P(a|b,-e)P(b,-e) + P(a|-b,e)P(-b,e) + P(a|-b,-e)P(-b,-c)$
- $= 0.95*0.000002 + 0.94*0.000998 + 0.29*0.001998 + 0.001*0.9970$

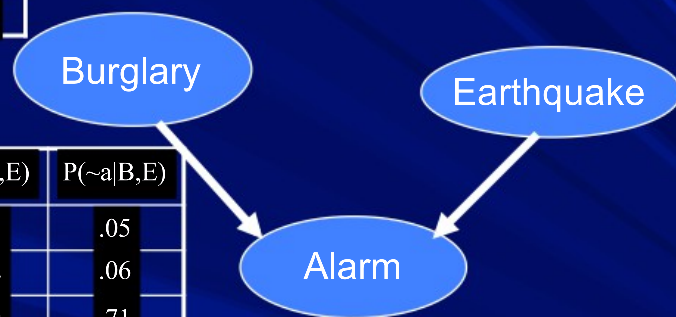
  $= 0.0025$

  $P(a|b) = P(a|b,e)\, P(e) + P(a|b,-e)\, P(-e) = 0.95*0.002 + 0.94*0.998 = 0.94002$

  $P(b|a) = 0.94002*.001/0.0025 = 0.3760$

# Explaining away

| P(b) | P(~b) |
|---|---|
| .001 | .999 |




| P(e) | P(~e) |
|---|---|
| .002 | .998 |

| B E | P(a\|B,E) | P(~a\|B,E) |
|---|---|---|
| T T | .95 | .05 |
| T F | .94 | .06 |
| FT | .29 | .71 |
| F F | .001 | .999 |

- P(b|a,-e) = P(a|b,-e) P(b|-e) P(-e) /P(a|-e) P(-e) = P(a|b,-e) P(b|-e) /P(a|-e)
- = 0.94 * 0.001 / P(a|-e)
- = 0.00094 / [P(a|-e,b) P(b) + P(a|-e,-b) P(-b)]

  = 0.00094 / 0.94*0.001 + 0.001*0.999

  = 0.00094 / 0.29066 = 0.4847