Reasoning Uncertainty

Deduction vs. Induction

DEDUCTION

- Commonly associated with *formal logic*
- Involves reasoning from known premises to a conclusion
- The conclusions reached are inevitable, certain, inescapable

INDUCTION

- Commonly known as *informal logic* or *everyday argument*
- Involves drawing uncertain inferences based on probabilistic reasoning
- The conclusions reached are probable, reasonable, plausible, believable

Reasoning under Uncertainty

- Reasoning under uncertainty is central to creating machines that exhibit intelligent behavior, one of the most-studied problems in artificial intelligence (AI).

- Classical FOL cannot handle uncertainty

$$\forall p \; Symptom(p, Toothache) \Rightarrow Disease(p, Cavity)$$

• Toothache can be caused in many other cases and we may include all other cases.

$$\forall p \; Disease(p, Cavity) \Rightarrow Symptom(p, Toothache)$$

• Cavity does not always cause toothache.

Predicting versus Diagnosing

- Probabilistic reasoning can be used for predicting outcomes (*from cause to effect*)
    - Given that I have a cavity, what is the chance that I will have toothache?
- Probabilistic reasoning can also be used for diagnosis (*from effect to cause*)
    - Given that I am having toothache, what is the chance that it is being caused by a cavity?

We need a methodology for reasoning that can work both ways.

Reasoning under uncertainty classified into two paradigms:
Bayesian and non-Bayesian.

Probability Axioms

- $0 \le p(A) \le 1$
- $p(\text{true}) = 1$, $p(\text{false}) = 0$
- **Independent events** A and B: $p(A \cap B) = p(A)\, p(B)$.
- The events $E_1, E_2, \ldots, E_n$ in a sample space S, are independent if

  $$p(E_{i1} \cap \ldots \cap E_{ik}) = p(E_{i1}) \ldots p(E_{ik})$$

  for each subset $\{i_1, \ldots, i_k) \subseteq \{1, \ldots, n\}, 1 \le k \le n, n \ge 1$.
- Events A and B are mutually exclusive: $p(A \cup B) = p(A) + p(B)$
- When A and B are not mutually exclusive:

  $p(A \cup B) = p(A) + p(B) - p(A \cap B)$
- This is also called **Law of Addition.**

Conditional Probabilities

The probability of an event A, given B occurred, is called a conditional probability and indicated by

$$P(A \mid B)$$

The conditional probability is defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \text{ for } P(B) \neq 0.$$

- **Multiplicative Law** of probability for two events is then defined as

  $$P(A \cap B) = P(A \mid B)\, P(B)$$

  which is equivalent to the following

  $$P(A \cap B) = P(B \mid A)\, P(A)$$

- **Generalized Multiplicative Law**

  $$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1 \mid A_2 \cap \ldots \cap A_n)\ P(A_2 \mid A_3 \cap \ldots \cap A_n) \ldots\ P(A_{n-1} \mid A_n)\ P(A_n)$$

Bayes' Theorem

From conditional probability:

$$P(H \mid E) = \frac{P(H \cap E)}{P(E)},$$

Furthermore, we have, $$P(E \mid H) = \frac{P(E \cap H)}{P(H)}$$

So,

$$P(E \mid H)P(H) = P(H \mid E)P(E) = P(H \cap E)$$

Thus

$$P(H \mid E) = \frac{P(E \mid H)\, P(H)}{P(E)}$$

In real-life practice, the probability $P(H \mid E)$ cannot always be found in the literature or obtained from statistical analysis. The conditional probabilities $P(E \mid H)$ however often are easier to obtain from the probabilities $P(E)$, $P(H)$ and $P(E \mid H)$;

Bayesian Approach

- Bayesian approaches represent knowledge about a set of state variables as probabilities.

- State variables can be either discrete or continuous.

- Variables are initialized a prior (starting estimate) and then updated using Bayes' rule when new evidence is received.

In Bayesian model assumes that all variables are **uncorrelated**.

Rather than reasoning about the truth or falsity of a proposition, reason about the belief that a proposition or event is true or false

• For each primitive proposition or event, attach a **degree of belief** to the sentence.

• Use **probability theory** as a formal means of dealing with degrees of belief.

Beliefs

- We use probability to describe the world and existing uncertainty
- Agents will have *beliefs* based on their current state of knowledge
  - E.g. *P(Some day AI agents will rule the world)=0.2* reflects a personal belief, based on one's state of knowledge about current AI, technology trends etc.
- Different agents may hold different beliefs, as these are *subjective*
- Beliefs may change over time as agents get new evidence
- *Prior (unconditional) beliefs* denote belief prior to the arrival of any new evidence.

Bayesian network is based on Joint probability distribution and conditional probability.

Bayesian belief network dealing with probabilistic events and solve problem which has uncertainty.

Belief nets use an acyclic directed graph to represent joint probability distribution

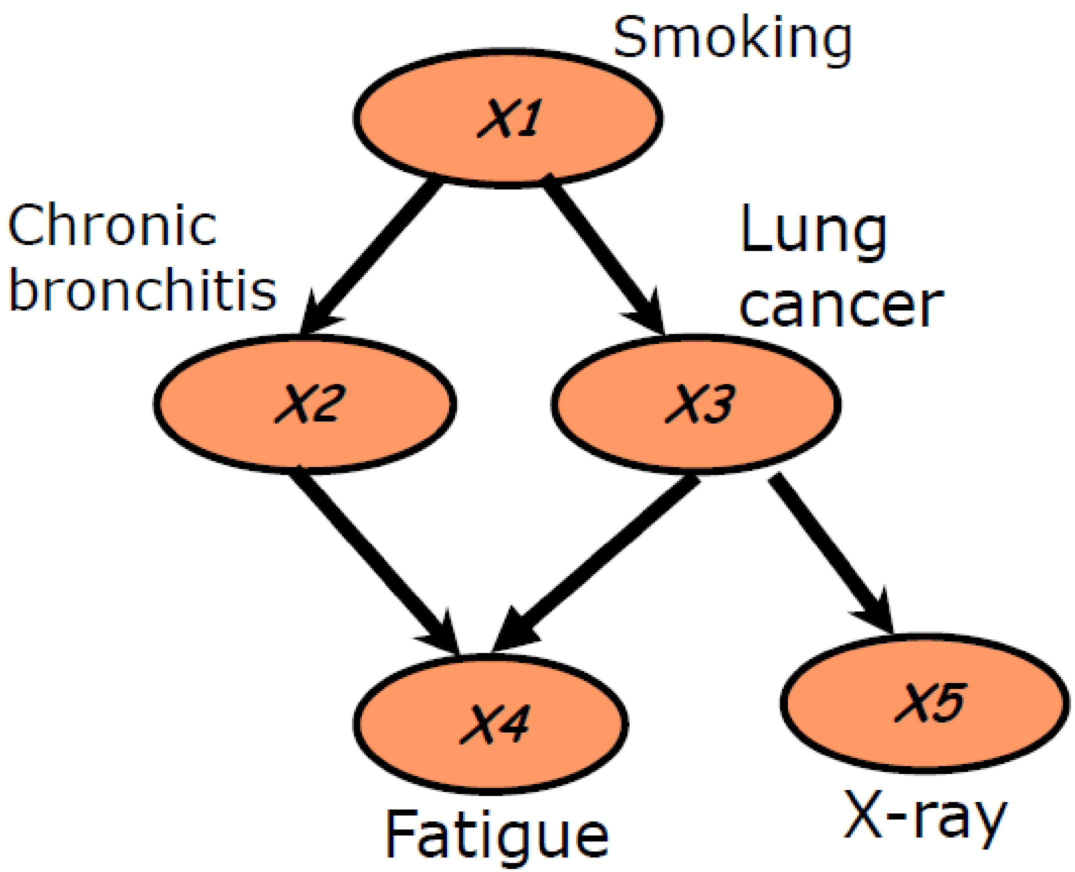
Belief Networks

A belief network is a graph with the following:

- Nodes: Set of random variables
- Directed links: The intuitive meaning of a link from node X to node Y is that X has a direct influence on Y

Each node has a conditional probability table that quantifies the effects that the parent have on the node.

The graph has no directed cycles. It is a *directed acyclic graph* (DAG).



$$P(X1, X2, X3, X4, X5)$$
$$= P(X1)P(X2 \mid X1)P(X3 \mid X1)P(X4 \mid X2, X3)P(X5 \mid X3)$$

Joint probability distribution

- If we have variables x_1, x_2, x_3,....., x_n, then Joint probability distribution.
- $\mathbf{p[x_1, x_2, x_3, \ldots\ldots, x_n]}$
  $\mathbf{= p[x_1 | x_2, x_3, \ldots\ldots, x_n] p[x_2, x_3, \ldots\ldots, x_n]}$
  $\mathbf{= p[x_1 | x_2, x_3, \ldots\ldots, x_n] p[x_2 | x_3, \ldots\ldots, x_n] \ldots. p[x_{n-1} | x_n] p[x_n].}$

$$P(X_i | X_{i-1}, \ldots\ldots\ldots, X_1) = P(X_i \,|\, Parents(X_i))$$

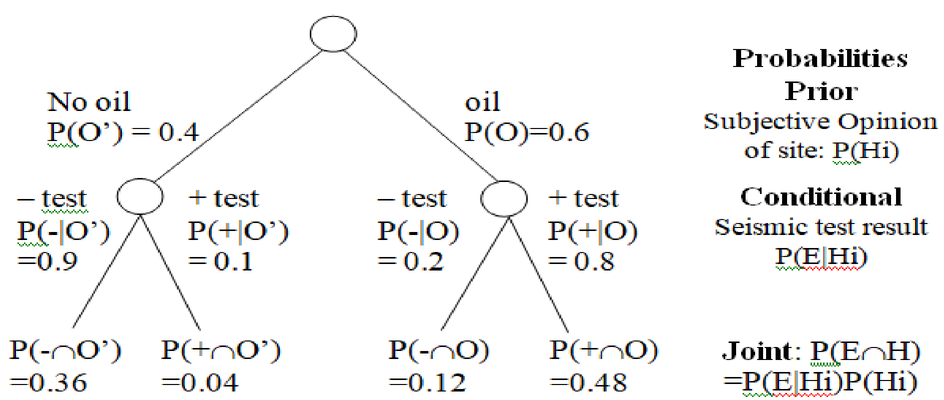
The basic task of a belief network

- Compute the posterior probability for a query variable given an observed event
- i.e., an **assignment of values** to a set of **evidence variables** while other variables are **not assigned values** (the so-called **hidden variables**).

- Example: Suppose the prospector believes that there is a better than 50-50 chance of finding oil, and assumes P(O) = 0.6 and P(O’) = 0.4
- Using the **seismic survey** technique, we obtain the following conditional probabilities, where + means a positive outcome and - is a negative outcome
- P(+ | O) = 0.8 P(- | O) = 0.2 (false -)
- P(+ | O’) = 0.1 (false +) P(- | O’) = 0.9
- Using the prior and conditional probabilities,

we can construct the initial

probability tree.

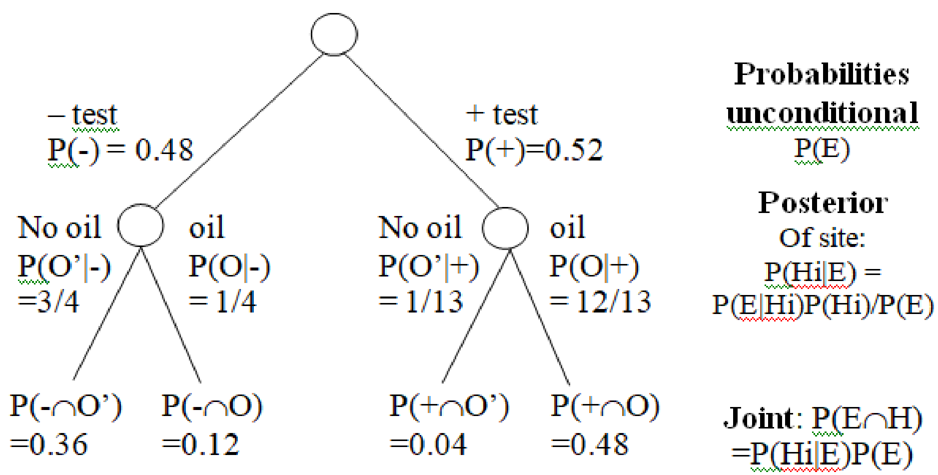


Initial probability tree for oil exploration

- Using Addition law to calculate the total probability of a + and a - test

$$P(+) = P(+ \cap O) + P(+ \cap O') = 0.48 + 0.04 = 0.52$$
$$P(-) = P(- \cap O) + P(- \cap O') = 0.12 + 0.36 = 0.48$$

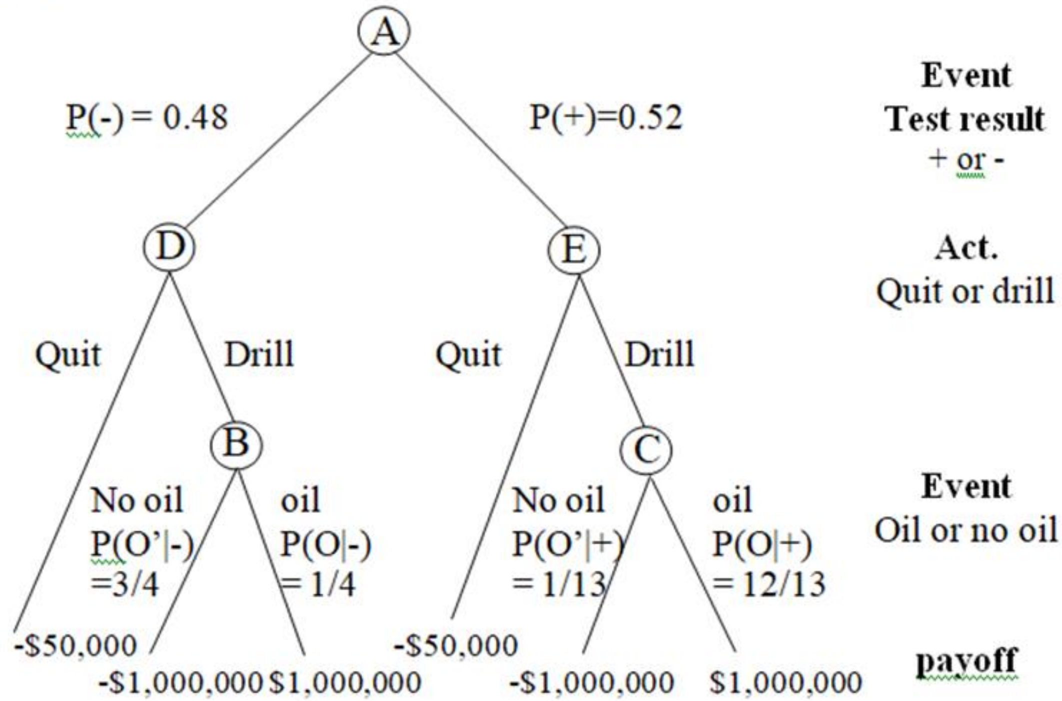
- P(+) and P(-) are unconditional probabilities that can now be used to calculate the posterior probabilities at the site, as shown below.



Revised probability tree for oil exploration

The assumed amounts are:
Oil lease, if successful: $1,250,000
Drilling expense: -$200,000
Seismic survey: -$50,000

Thus if oil is found, the payoff is
$1,250,000 - $200,000 - $50,000 = $1,000,000
while a decision to quit after the seismic test result gives a payoff of -$50,000.



- In order for the prospector to make the best decision, the expected payoff Is calculated at event node A.

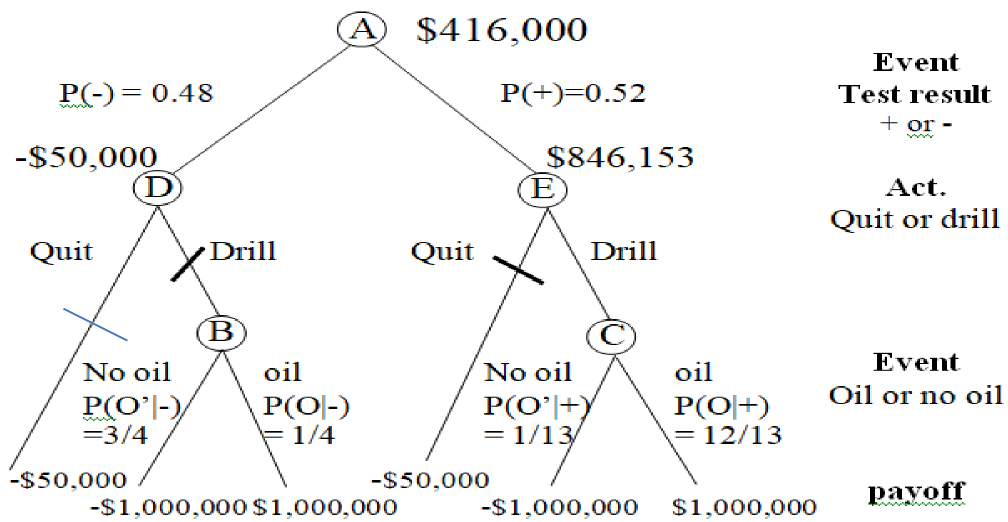
- To compute the expected payoff at A, work backward from the leaves. This process is called **backward induction**.

• The decision tree shows the optimal strategy for the prospector.
• If the seismic test is positive, the site should be drilled, otherwise, the site should be abandoned.
• The decision tree is an example of hypothetical reasoning or “what if” type of situations.
• By exploring alternate paths of action, we can prune paths that do not lead to optimal payoffs.

- The expected payoff from an event node is the sum of the payoffs times the probabilities leading to the payoffs.

Expected payoff at node C
$846,153 = ($1,000,000) (12/13) - ($1,000,000) (1/13)

Expected payoff at node B
-$500,000 = ($1,000,000) (1/4) - ($1,000,000) (3/4)



Complete Bayesian Decision tree for oil exploration
Using backward induction

Bayes' rule and knowledge-based systems

- Knowledge-base IF-THEN format:
- IF X is true THEN Y can be concluded with probability p
- If we observe that X is true, then we can conclude that Y exist with the specified probability.
- For example
- IF the patient has a cold
- THEN the patient will sneeze (0.75)
- But what if we reason abductively and observe Y (i.e., the patient sneezes) while knowing nothing about X (i.e., the patient has a cold)?
- What can we conclude about it?
- Bayes' Theorem describes how we can derive a probability for X.

- Y (denotes some piece of evidence (typically referred to as E) and X denotes some hypothesis (H) given

$$P(H \mid E) = \frac{P(E \mid H)\, P(H)}{P(E)}$$

or

$$P(H \mid E) = \frac{P(E \mid H)\, P(H)}{P(E \mid H)P(H) + P(E \mid H')P(H')}$$

- The probability of sneezing is the sum of the conditional probability that he sneezes when with a cold and the conditional probability without cold.

- P(H) = P(has a cold) = 0.2
- P(E | H) =P(observed sneezing | has a cold) = 0.75
- P(E | H’) = P(was observed sneezing | does not have a cold) = 0.2

Then

P(E) = P(observed sneezing) = (0.75)(0.2) + (0.2)(0.8) = 0.31

and

- $$P(H \mid E) = P(\text{has a cold} \mid \text{observed sneezing}) = \frac{(0.75)(0.2)}{(0.31)}$$

probability of having a cold given that he sneezes = 0.48387

$$P(H \mid E') = \frac{P(E' \mid H)P(H)}{P(E')}$$

$$= \frac{(1-0.75)\,(0.2)}{(1 - 0.31)}$$

= 0.07246 probability of having a cold would be if was not sneezing:

Propagation of Belief

- We have only considered when each piece of evidence affects only one hypothesis.
- To deal with real-world problems, consider "m" hypotheses H_1, H_2, ... H_m and "n" pieces of evidence E_1, ..., E_n.

$$P(H_i \mid E_{j_1} \cap E_{j_2} \cap \ldots \cap E_{j_k}) = \frac{P(E_{j_1} \cap E_{j_2} \cap \ldots \cap E_{j_k} \mid H_i)\, P(H_i)}{P(E_{j_1} \cap E_{j_2} \cap \ldots \cap E_{j_k})}$$

$$= \frac{P(E_{j_1} \mid H_i) P(E_{j_2} \mid H_i) \ldots P(E_{j_k} \mid H_i) P(H_i)}{\sum_{l=1}^{m} P(E_{j_1} \mid H_l) P(E_{j_2} \mid H_l) \ldots P(E_{j_k} \mid H_l) P(H_l)}$$

where $\{j_1, \ldots, j_k) \subseteq \{1, \ldots, n\}$

- This probability is called the posterior probability of hypothesis H_i from observing evidence E_{j1}, E_{j2}, ..., E_{jk}.

- Assumptions:
- The hypotheses $H_1, \ldots, H_m$, $m \geq 1$, are mutually exclusive.
- Furthermore, the hypotheses $H_1, \ldots, H_m$ are collectively exhaustive.
- The pieces of evidence $E_1, \ldots, E_n$, $n \geq 1$, are conditionally independent given any hypothesis H_i, $1 \leq i \leq m$.
- The events $E_1, E_2, \ldots, E_n$, are *conditionally independent* given an event H if

$$P(E_{j_1} \cap \ldots \cap E_{j_k} \mid H) = P(E_{j_1} \mid H) \ldots P(E_{j_k} \mid H)$$

for each subset $\{j_1, \ldots, j_k\} \subseteq \{1, \ldots, n\}$.

- Often causes great difficulties for probabilistic based methods.

- Consider three mutually exclusive and exhaustive hypotheses with values.
- H_1, the patient, has a cold;
- H_2, the patient has an allergy; and
- H_3, the patient has a sensitivity to light
- With prior probabilities, $p(H_i)$'s, and two conditionally independent pieces of evidence
- E_1, the

	i = 1 (cold)	i = 2 (allergy)	i = 3 (light sensitive)
$P(H_i)$	0.6	0.3	0.1
$P(E_1 \mid H_i)$	0.3	0.8	0.3
$P(E_2 \mid H_i)$	0.6	0.9	0.0

If we observe evidence E_1 (e.g., the patient sneezes), we can compute posterior probabilities for the hypotheses:

$$P(H_1 \mid E_1) = \frac{(0.3)(0.6)}{(0.3)(0.6) + (0.8)(0.3) + (0.3)(0.1)} = 0.4$$

$$P(H_2 \mid E_1) = \frac{(0.8)(0.3)}{(0.3)(0.6) + (0.8)(0.3) + (0.3)(0.1)} = 0.53$$

$$P(H_3 \mid E_1) = \frac{(0.3)(0.1)}{(0.3)(0.6) + (0.8)(0.3) + (0.3)(0.1)} = 0.06$$

- The belief in hypotheses H_1 and H_3 have both decreased while the belief in hypothesis H_2 has increased after observing E_1.

- If E_2 (e.g., the patient coughs) is now observed, new posterior probabilities

$$P(H_1 \mid E_1 \cap E_2) = \frac{(0.3)(0.6)(0.6)}{(0.3)(0.6)(0.6) + (0.8)(0.9)(0.3) + (0.3)(0.0)(0.1)} = 0.33$$

$$P(H_2 \mid E_1 \cap E_2) = \frac{(0.8)(0.9)(0.3)}{(0.3)(0.6)(0.6) + (0.8)(0.9)(0.3) + (0.3)(0.0)(0.1)} = 0.67$$

$$P(H_3 \mid E_1 \cap E_2) = \frac{(0.3)(0.0)(0.1)}{(0.3)(0.6)(0.6) + (0.8)(0.9)(0.3) + (0.3)(0.0)(0.1)} = 0.0$$

Burglar Alarm at home

- Harry installed a new burglar alarm at his home to detect burglary.
- The alarm reliably responds at detecting a burglary but also responds for minor earthquakes.
- Harry has two neighbors John and Marya, who have taken a responsibility to inform Harry at work when they hear the alarm.
- John always calls Harry when he hears the alarm, but sometimes he got confused with the phone ringing and calls at that time too.
- On the other hand, Mary likes to listen to high music, so sometimes she misses to hear the alarm.
- We would like to compute the probability of Burglary Alarm.

Problem:

- Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and John and Mary both called the Harry.
- <u>Events occurring in this network:</u>

  Burglary (B)
  Earthquake(E)
  Alarm(A)
  John Calls(J)
  Mary calls(M)

- write the events of problem statement in terms of probability:
  p[J, M, A, B, E],

Write probability statement using joint probability distribution:

p[J, M, A, B, E]= p[J | M, A, B, E]. p[M, A, B, E]

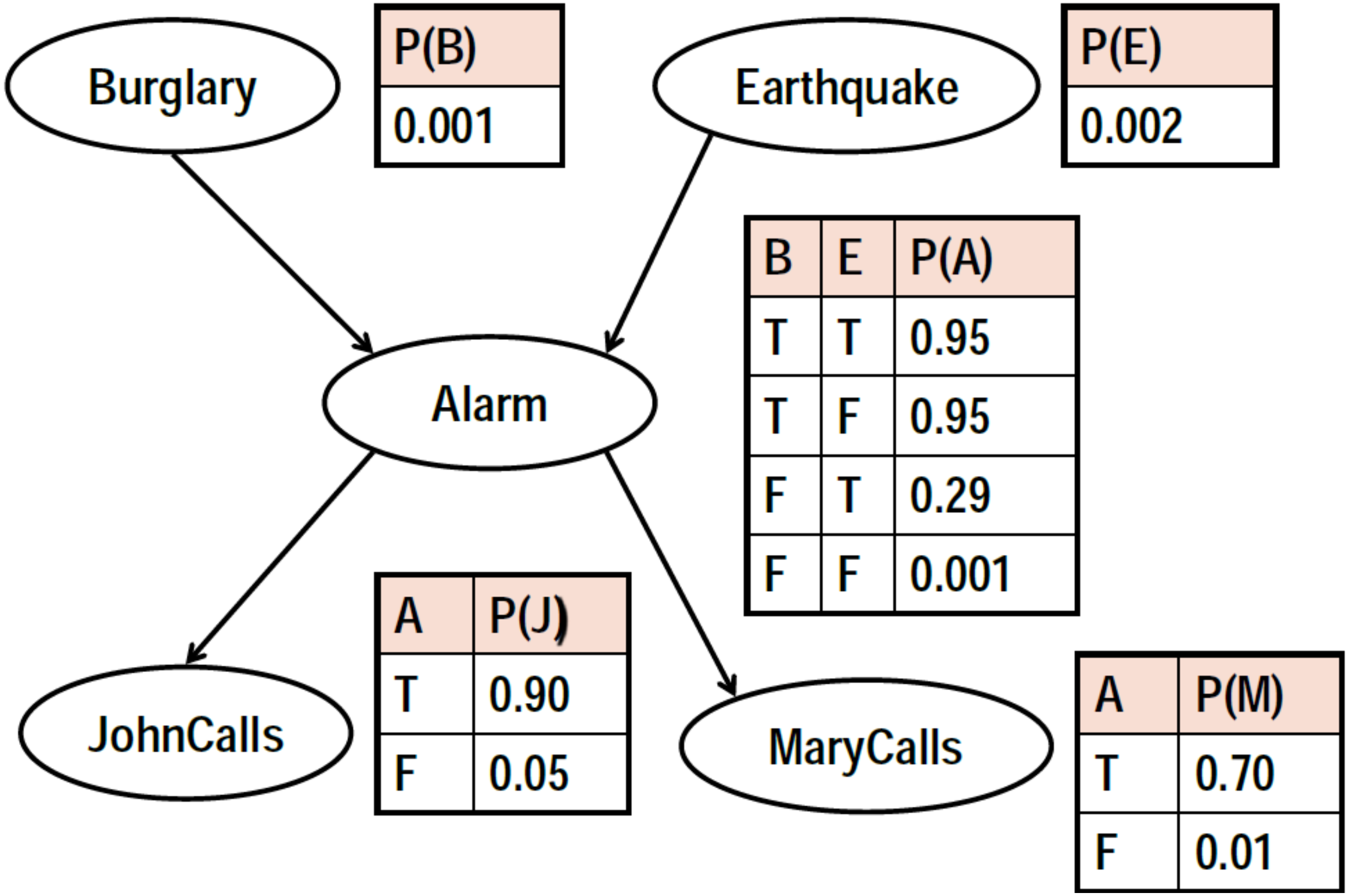
= p[J | M, A, B, E]. p[M | A, B, E]. p[A, B, E]

= .

= .

= .

p(B= True) = 0.001, probability of burglary.
p(B= False)= 0.999, probability of no burglary.
p(E= True)= 0.002, probability of a minor earthquake
p(E= False)= 0.998, probability that an earthquake not occurred.



P(B)
0.001

P(E)
0.002

B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

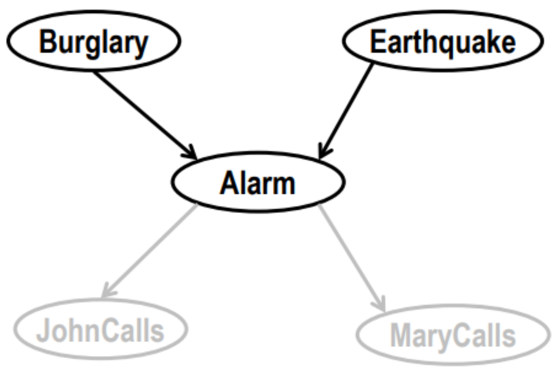
A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

The joint probability distribution

- A generic entry in the joint probability distribution $P(x_1, \ldots, x_n)$ is given by:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i))$$

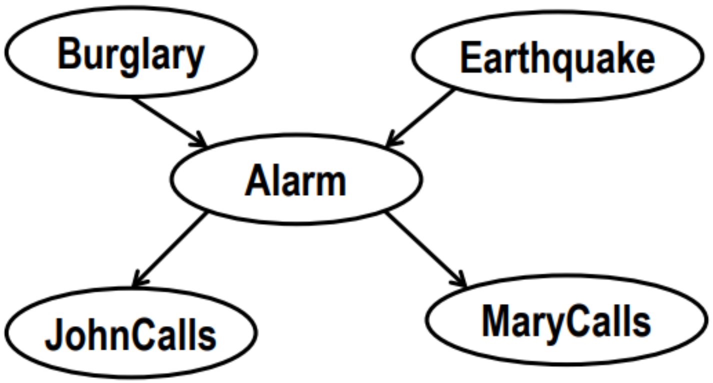


B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

The joint probability distribution

- Probability of the event that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both Mary and John call:

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$$
$$= P(J \mid A)\ P(M \mid A)\ P(A \mid \neg B \wedge \neg E)\ P(\neg B)\ P(\neg E)$$
$$= 0.9 \text{ X } 0.7 \text{ X } 0.001 \text{ X } 0.999 \text{ X } 0.998$$
$$= 0.00062$$



B	E	P(A)
T	T	0.95
T	F	0.95
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(E)
0.002

P(B)
0.001

The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$$
\begin{aligned}
P(B) &= 0.001 \\
P(B') &= 1 - P(B) = 0.999 \\
P(E) &= 0.002 \\
P(E') &= 1 - P(E) = 0.998
\end{aligned}
$$

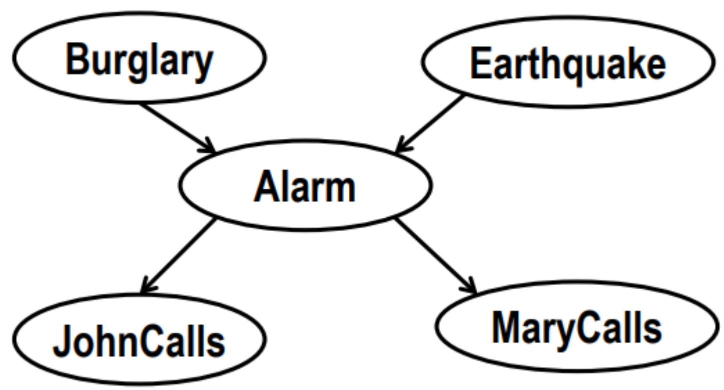
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$$P(A) = P(AB'E') + P(AB'E) + P(ABE') + P(ABE)$$

$$= P(A \mid B'E').P(B'E') + P(A \mid B'E).P(B'E) + P(A \mid BE').P(BE') + P(A \mid BE).P(BE)$$

$$= 0.001 \times 0.999 \times 0.998 + 0.29 \times 0.999 \times 0.002 + 0.95 \times 0.001 \times 0.998 + 0.95 \times 0.001 \times 0.0$$

$$= 0.001 + 0.0006 + 0.0009 = 0.0025$$

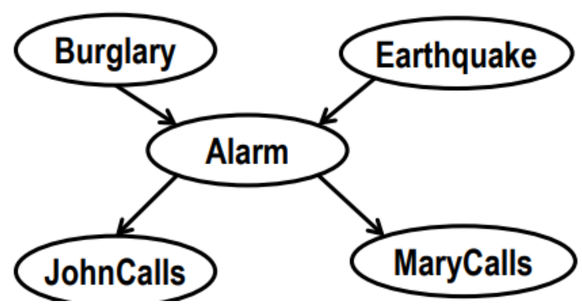
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution: *Find P(J)*

$$
\begin{aligned}
P(J) &= P(JA) + P(JA') \\
&= P(J \mid A).P(A) + P(J \mid A').P(A') \\
&= 0.9 \times 0.0025 + 0.05 \times (1 - 0.0025) \\
&= 0.052125 \\
P(AB) &= P(ABE) + P(ABE') = 0.95 \times 0.001 \times 0.002 + 0.95 \times 0.001 \times 0.998 \\
&= 0.00095
\end{aligned}
$$

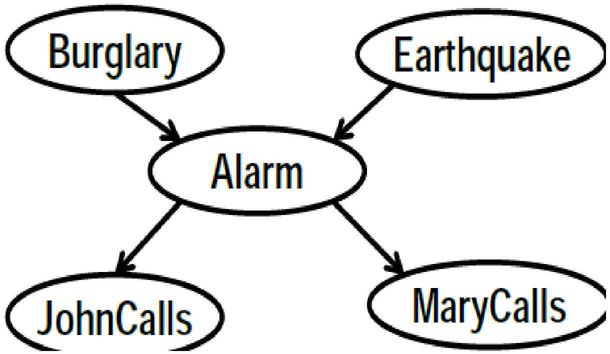
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution: *Find* P(A′B) *and* P(AE)

$$
\begin{aligned}
P(A'B) &= P(A'BE) + P(A'BE') \\
&= P(A' \mid BE).P(BE) + P(A' \mid BE').P(BE') \\
&= (1 - 0.95) \times 0.001 \times 0.002 \\
&+ (1 - 0.95) \times 0.001 \times 0.998 \\
&= 0.00005
\end{aligned}
$$

$$
\begin{aligned}
P(AE) &= P(AEB) + P(AEB') \\
&= 0.95 \times 0.001 \times 0.002 + 0.29 \times 0.999 \times 0.002 = 0.00058
\end{aligned}
$$

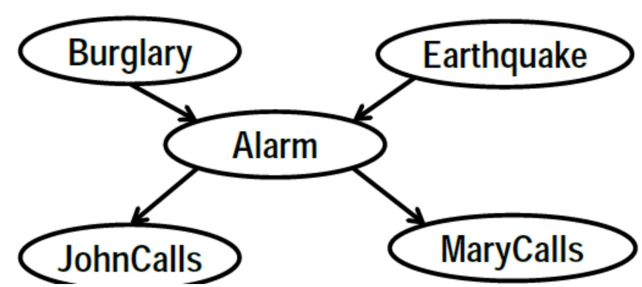
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution

$$P(AE') = P(AE'B) + P(AE'B')$$

$$= 0.95 \times 0.001 \times 0.998 + 0.001 \times 0.999 \times 0.998$$

$$= 0.001945$$

$$P(A'E') = P(A'E'B) + P(A'E'B')$$

$$= P(A' \mid BE').P(BE') + P(A' \mid B'E').P(B'E')$$

$$= (1 - 0.95) \times 0.001 \times 0.998 + (1 - 0.001) \times 0.999 \times 0.998 = 0.996$$

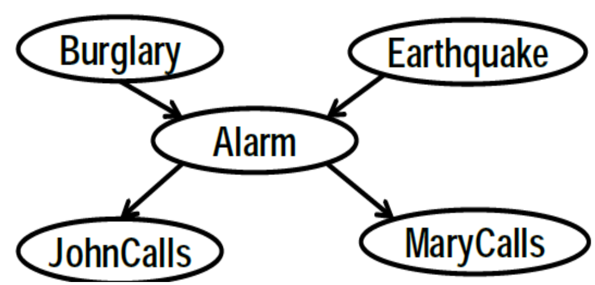
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution: *Find* P(JB)

$$
\begin{aligned}
P(JB) &= P(JBA) + P(JBA') \\
&= P(J \mid AB).P(AB) + P(J \mid A'B).P(A'B) \\
&= P(J \mid A).P(AB) + P(J \mid A').P(A'B) \\
&= 0.9 \times 0.00095 + 0.05 \times 0.00005 \\
&= 0.00086
\end{aligned}
$$

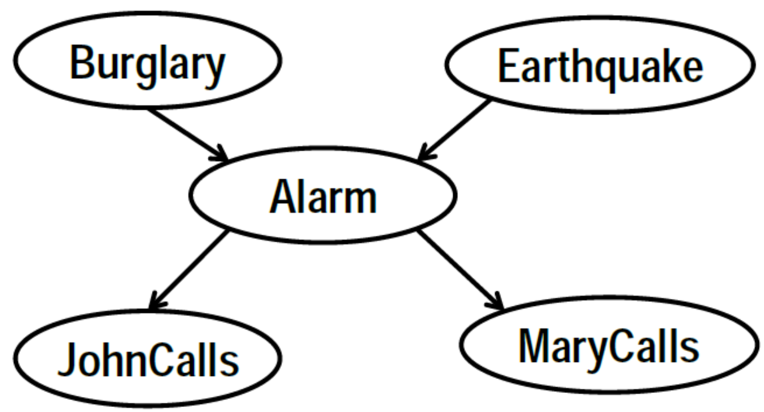
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$$P(J \mid B) = P(JB) / P(B) = 0.00086 / 0.001 = 0.86$$

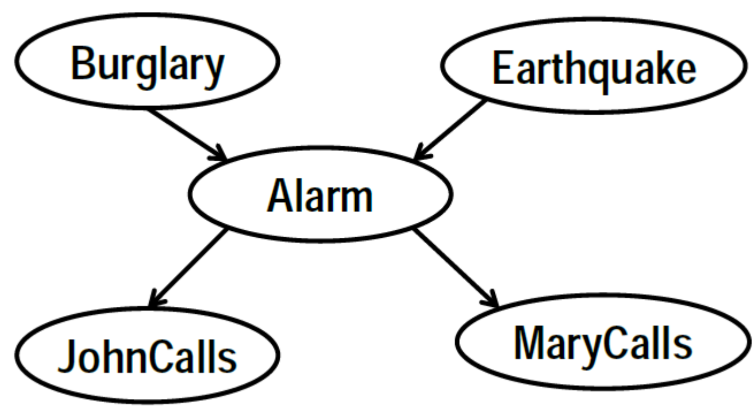
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution

$$
\begin{aligned}
P(MB) &= P(MBA) + P(MBA') \\
&= P(M \mid AB).P(AB) + P(M \mid A'B).P(A'B) \\
&= P(M \mid A).P(AB) + P(M \mid A').P(A'B) \\
&= 0.7 \times 0.00095 + 0.01 \times 0.00005 \\
&= 0.00067
\end{aligned}
$$

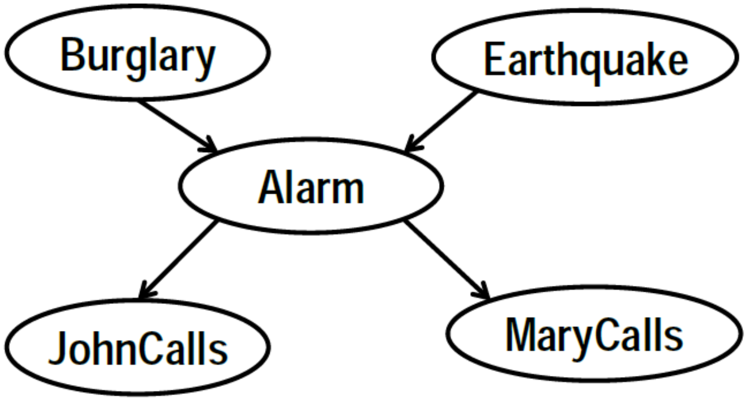
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution

$$P(M \mid B) = P(MB) / P(B) = 0.00067 / 0.001 = 0.67$$

$$P(B \mid J) = P(JB) / P(J) = 0.00086 / 0.052125 = 0.016$$

$$P(B \mid A) = P(AB) / P(A) = 0.00095 / 0.0025 = 0.38$$

$$P(B \mid AE) = P(ABE) / P(AE) = [\, P(A \mid BE).P(BE) \,] / P(AE)$$

$$= [\, 0.95 \times 0.001 \times 0.002 \,] / 0.00058$$

$$= 0.003$$

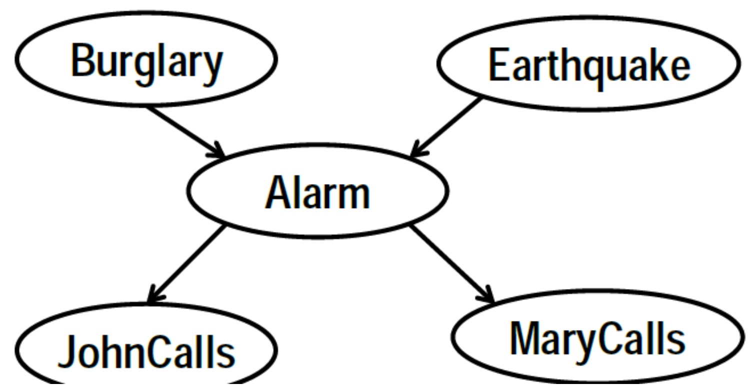
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$$P(AJE') = P(J \mid AE').P(AE') = P(J \mid A).P(AE')$$

$$= 0.9 \times 0.001945 = 0.00175$$

$$P(A'JE') = P(J \mid A'E').P(A'E') = P(J \mid A').P(A'E')$$

$$= 0.05 \times 0.996 = 0.0498$$

$$P(JE') = P(AJE') + P(A'JE') = 0.00175 + 0.0498 = 0.05155$$

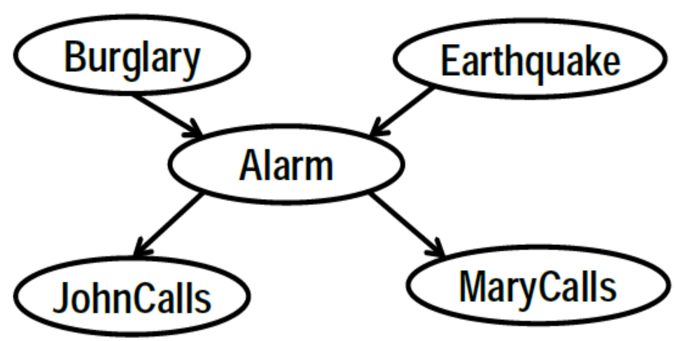
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



$$P(A \mid JE') = P(AJE') / P(JE') = 0.00175 / 0.05155 = 0.03$$

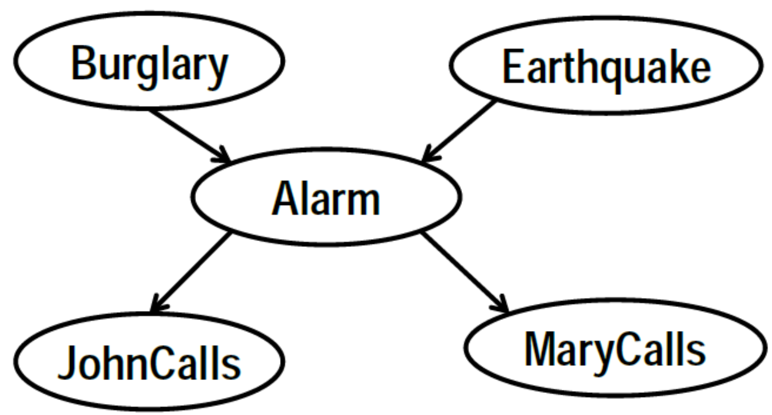
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



$$P(BJE') = P(BJE'A) + P(BJE'A')$$

$$= P(J \mid ABE').P(ABE') + P(J \mid A'BE').P(A'BE')$$

$$= P(J \mid A).P(ABE') + P(J \mid A').P(A'BE')$$

$$= 0.9 \times 0.95 \times 0.001 \times 0.998 + 0.05 \times (1 - 0.95) \times 0.001 \times 0.998$$

$$= 0.000856$$

$$P(B \mid JE') = P(BJE') / P(JE') = 0.000856 / 0.05155 = 0.017$$

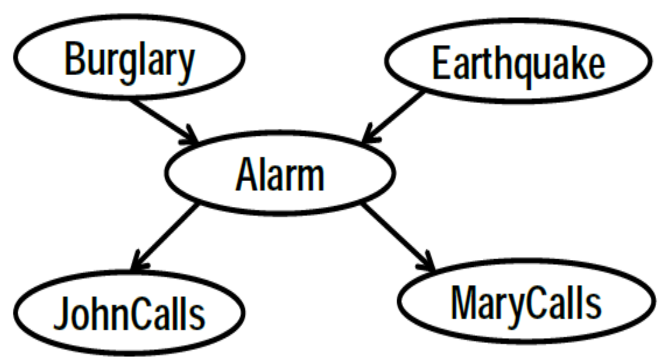
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

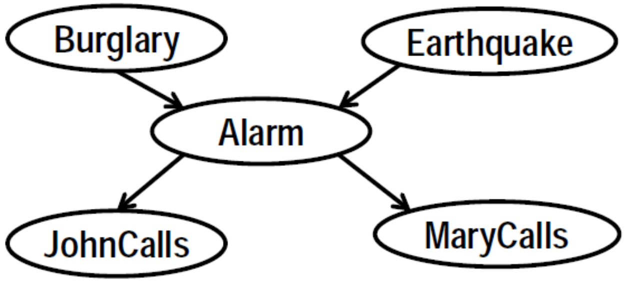
P(E)
0.002

P(B)
0.001



Inferences using belief networks

- Diagnostic inferences (from effects to causes)
  - Given that JohnCalls, infer that
    P(Burglary | JohnCalls) = 0.016



- Causal inferences (from causes to effects)
  - Given Burglary, infer that
    P(JohnCalls | Burglary) = 0.86
    P(MaryCalls | Burglary) = 0.67

Advantages and disadvantages of Bayesian methods

- Most significant is their sound theoretical foundation in probability theory. Thus, they are currently the most mature of all of the uncertainty reasoning methods.

- Require a significant amount of sample sizes so that the probabilities obtained are accurate and sufficient data to construct a knowledge base.

- Furthermore, human experts are normally uncertain, whether the values consistent and comprehensive?

MYCIN

- Developed at Stanford University in 1972
- Regarded as the first true "expert system"
- **Assist physicians in the treatment of blood infections**
- Many revisions and extensions to MYCIN over the years

- Physician wishes to specify an "antimicrobial agent" - basically an antibiotic - to kill bacteria or arrest their growth

The Decision Process

- There are four questions in the process of deciding on treatment:
  - **Does the patient have a significant infection?**
  - **What are the organism(s) involved?**
  - **What set of drugs might be appropriate to treat the infection?**
  - **What is the best choice of drug or combination of drugs to treat the infection?**

- Physician wishes to specify an “antimicrobial agent” - basically an antibiotic - to kill bacteria or arrest their growth
- Some agents are poisonous!
- No agent is effective against all bacteria
- Most physicians are not expert in the field of antibiotics

MYCIN Components

- **KNOWLEDGE BASE:**
  - facts and knowledge about the domain
- **DYNAMIC PATIENT DATABASE:**
  - information about a particular case
- **CONSULTATION PROGRAM:**
  - asks questions, gives advice on a particular case
- **EXPLANATION PROGRAM:**
  - answers questions and justifies advice
- Basic rule structure in MYCIN is:

  if **condition$_1$ *and....and* condition$_m$ *hold***

  then draw **conclusion$_1$ *and....and* condition$_n$**
- Rules written in the LISP programming language
- Rules can include certainty factors to help weight the conclusions drawn

An Example Rule

IF:(1) The stain of the organism is Gram negative, and

(2) The morphology of the organism is rod, and

(3) The aerobicity of the organism is aerobic

THEN:

There is strongly suggestive evidence (0.8) that the class of the organism is *Enterobacteria*

Calculating Certainty

- Rule certainties are regarded as probabilities
- Therefore must apply the rules of probability in combining rules
- Multiplying probabilities which are less than certain results in lower and lower certainty!
- Eg $0.8 \times 0.6 = 0.48$

Utility Theory

- Utility theory in AI provides a **mathematical** framework for understanding how AI systems make choices among different options based on their perceived value or utility.
- Utility theory offers a framework for making decisions in situations of ambiguity by putting utilities(values) on several possible results.
- It is very useful in optimising and modeling decision-making processes by considering uncertain and probabilistic outcomes in different situations.
- It allows one to assign subjective values or preferences to different outcomes and helps make optimal choices based on these values.
- Utility theory is widely used in various AI applications such as **game theory**, **economics**, **robotics**, and **recommendation systems**, among others.

- In a **recommendation system**, the utility could describe the level of user satisfaction with a particular recommendation.
- In a **robotics application**, a utility could represent the cost or risk of different actions.
- The AI system can then choose the action with the **highest expected utility**.
- Basic notation commonly used in utility theory:

Let x represents an outcome or option

Let $U(x)$ denote the utility function, which maps x to its utility value

Let $p(x)$ denote the probability of outcome x occurring.

Let $E[U(x)]$ denote the expected utility of outcome x, which is the sum of the utility values of all possible outcomes weighted by their respective probabilities.

$$E[U(x)] = \sum_i p(x_i) \, . \, U(x_i)$$

Maximum Expected Utility

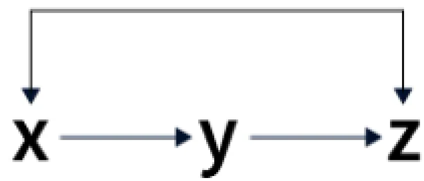
- MEU is a decision-making principle that suggests choosing the option that maximizes the expected utility.
- In other words, an AI system should select the option that is expected to yield the highest utility value, taking into account the probabilities of different outcomes.
- Utility theory in AI is based on a set of axioms or principles that define the properties of a rational utility function.
- Key utility theory axioms:
- <u>Orderability</u>

  A rational utility function should allow for comparing different outcomes based on their utility values. In other words, if $U(x)>U(y)$, then outcome x is preferred to outcome y.

- **Transitivity**

  If outcome *x* is preferred to outcome *y*, and outcome *y* is preferred to outcome *z*, then outcome *x* should be preferred to outcome *z*.

- This axiom ensures that the preferences modeled by the utility function are **consistent** and do **not lead to contradictions**.



Continuity

• Small changes in the probabilities of outcomes should result in small changes in the expected utility.

• This axiom ensures that the utility function is **smooth** and **well-behaved** and that small changes in probabilities do not result in abrupt changes in decision-making.

- **Substitutability**

  If two outcomes, *x* and *y*, are equally preferred, then we should equally prefer any combination of *x* and *y*.

  This axiom allows for substituting equally preferred outcomes without affecting the decision-making process.

- **Monotonicity**

  If the probability of an outcome increases, its expected utility should also increase.

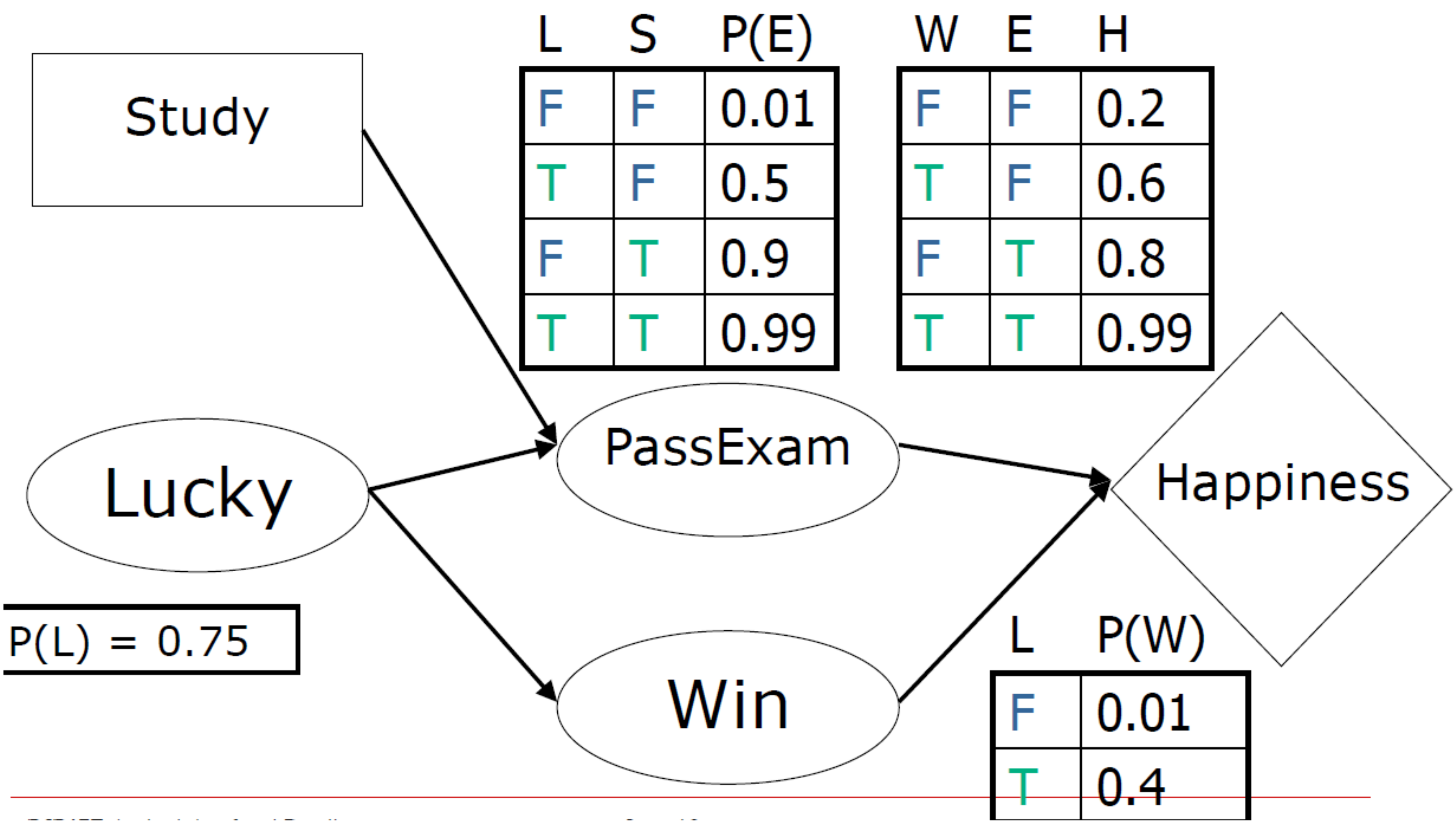
  This axiom ensures that an increase in the likelihood of an outcome increases its perceived value or utility.

- **Decomposability**

  The utility function should be able to represent preferences over multiple attributes or features of an outcome in a **decomposable manner**.

  This allows for modeling complex decision problems with multiple dimensions or criteria.

Assignment



L	S	P(E)
F	F	0.01
T	F	0.5
F	T	0.9
T	T	0.99

W	E	H
F	F	0.2
T	F	0.6
F	T	0.8
T	T	0.99

L	P(W)
F	0.01
T	0.4