UNIT-II

Dr.Y. SANGEETHA
Associate Professor

Uncertain and probabilistic reasoning - Basic Probability Notation

Uncertain Knowledge

- Agents don't have complete knowledge about the world.
- Agents need to make decisions based on their uncertainty.
- An agent needs to reason about its uncertainty.

Diagnosing a dental patient‘s toothache

Let us try to apply propositional logic:

Toothache ⇒ Cavity

Hmm, is it really true?

– not all patients with toothaches have cavities; some of them have gum disease, an abscess, or other problems

Toothache ⇒ Cavity ∨ GumProblem ∨ Abscess ∨ …

We could try turning the rule into a causal rule:

Cavity ⇒ Toothache

But this is not right either – not all cavities cause pain

The only way to fix the rule is to make it logically exhaustive!

Sources of Uncertainty

- **Uncertain data**
  - missing data, unreliable, ambiguous, imprecise representation, inconsistent, subjective, derived from defaults, noisy…
- **Uncertain knowledge representation**
  - restricted model of the real system
  - limited expressiveness of the representation mechanism
- **inference process**
  - Derived result is formally correct, but wrong in the real world
  - New conclusions are not well-founded (eg, inductive reasoning)
  - Incomplete, default reasoning methods

Handling Uncertain Knowledge

Problems using first-order logic for diagnosis:

Laziness:
Too much work to make complete rules.
Too much work to use them

Theoretical ignorance:
Medical Science has no complete theory for the domain

Practical ignorance:
We can’t run all tests anyway

Probability can be used to *summarize* the laziness and ignorance !

Probability

- Probability can be defined as a chance that an uncertain event will occur.
- It is the numerical measure of the likelihood that an event will occur. The value of probability always remains between 0 and 1 that represent ideal uncertainties.

- $0 \leq P(A) \leq 1$,   where P(A) is the probability of an event A.
- $P(A) = 0$, indicates total uncertainty in an event A.
- $P(A) = 1$, indicates total certainty in an event A.

$$\textbf{Probability of occurrence} = \frac{\text{Number of desired outcomes}}{\text{Total number of outcomes}}$$

- Probability is all about the possibility of various outcomes. The set of all possible outcomes is called the **sample space**.
- Only one outcome in the sample space is possible at a time, and the sample space must contain all possible values.

sample space - Ω (capital omega)

specific outcome – ω

We represent the probability of an event ω as $P(\omega)$.

- The two basic axioms of probability are:

  - $0 \leq P(\omega) \leq 1$

  - $\sum_{\omega} P(\omega) = 1$

- the probability of any event has to be between 0 (impossible) and 1 (certain),
- the sum of the probabilities of all events should be 1.

Uncertainty

Let action A_t = leave for airport t minutes before flight

Will A_t get me there on time?

Problems:

partial observability (road state, other drivers' plans, etc.)

1. noisy sensors (traffic reports)
2. uncertainty in action outcomes (flat tire, etc.)
3. immense complexity of modeling and predicting traffic

"A_{25} will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc "

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport ...)

The main tool for dealing with degrees of belief is **probability theory**, which assigns to each sentence a numerical degree of belief between 0 and 1.

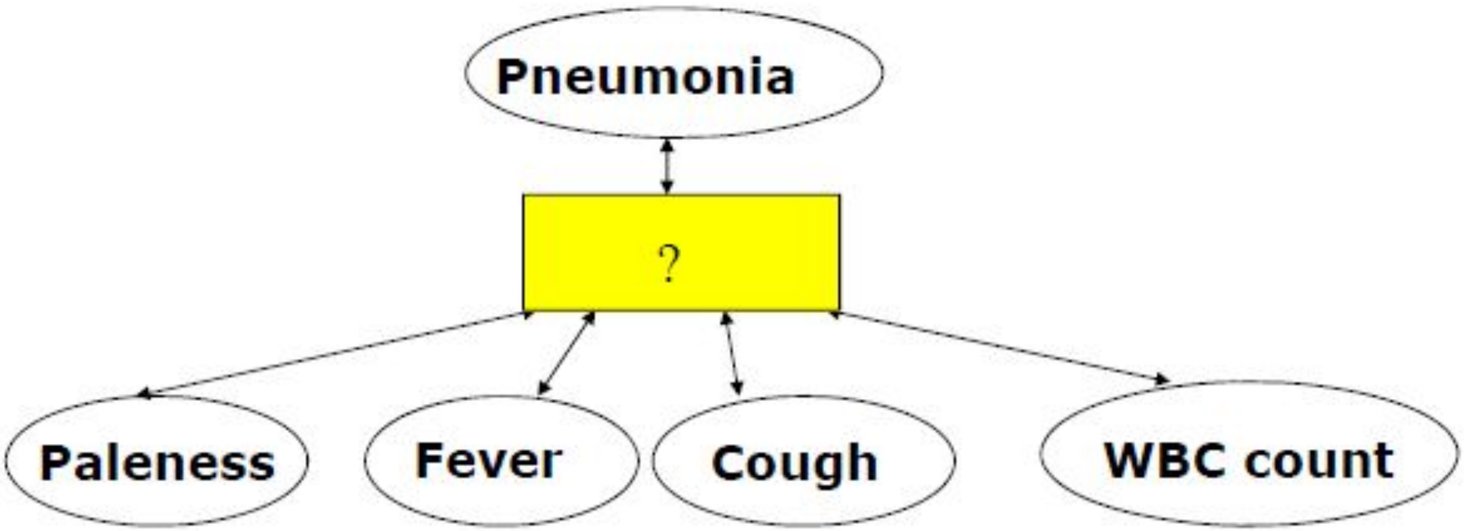
Handling uncertain knowledge

- The sentence itself is *in fact* either **true** or **false** .
- A degree of belief **is different from a degree of truth** .
- A probability of 0.8 does not mean “80% true”, but rather an 80% degree of belief that something is true.

Modeling the uncertainty.

Key challenges:

- How to represent the relations in the presence of uncertainty?
- How to manipulate such knowledge to make inferences?
  - **Humans can reason with uncertainty.**



Methods for representing uncertainty

Probability theory

- A well defined theory for modeling and reasoning in the presence of uncertainty
- A natural choice to replace certainty factors

Facts (propositional statements)

- Are represented via **random variables** with two or more values

  Example: *Pneumonia* is a random variable

  values: ***True*** **and** ***False***

- Each value can be achieved **with some probability:**

$$P(Pneumonia = True) = 0.001$$

$$P(WBCcount = high) = 0.005$$

Methods for representing uncertainty

Probabilistic extension of propositional logic

- **Propositions:**
  - statements about the world
  - Statements are represented by the assignment of values to **random variables**
- **Random variables:**

! – **Boolean** $\underbrace{Pneumonia}_{\textbf{Random variable}}$ is either $\underbrace{True, False}_{\textbf{Values}}$

! – **Multi-valued** $\underbrace{Pain}_{\textbf{Random variable}}$ is one of $\underbrace{\{Nopain, Mild, Moderate, Severe\}}_{\textbf{Values}}$

– **Continuous** $\underbrace{HeartRate}_{\textbf{Random variable}}$ is a value in $\underbrace{<0;180>}_{\textbf{Values}}$

Unconditional or Prior Probability

$P(A)$ denotes the unconditional probability or prior probability that A will appear *in the absence of any other information*, for example:

$$P(Cavity) = 0.1$$

$Cavity$ is a proposition. We obtain prior probabilities from statistical analysis or general rules.

In Bayesian statistical inference, the prior probability is the probability of an event before new data is collected. we have an initial belief, known as a prior, which we update as we gain additional information.

In general, a random variable can take on $true$ and $false$ values, as well as other values:

$$P(Weather = Sunny) = 0.7$$
$$P(Weather = Rain) = 0.2$$
$$P(Weather = Cloudy) = 0.08$$
$$P(Weather = Snow) = 0.02$$
$$P(Headache = true) = 0.1$$

Conditional or Posterior Probability

- **Posterior Probability:** The probability that is calculated after all evidence or information has taken into account. It is a combination of prior probability and new information.

Conditional Probability

New information can change the probability.

Example: The probability of a cavity increases if we know the patient has a toothache.

If additional information is available, we can no longer use the prior probabilities!

$P(A \mid B)$ is the conditional or posterior probability of A given that *all we know* is B:

$$P(Cavity \mid Toothache) = 0.8$$

$\mathbf{P}(X \mid Y)$ is the table of all conditional probabilities over all values of X and Y.

$\mathbf{P}(Weather \mid Headache)$ is a 4×2 table of conditional probabilities of all combinations of the values of a set of random variables.

	$Headache = true$	$Headache = false$
$Weather = Sunny$	$P(W = Sunny \mid Headache)$	$P(W = Sunny \mid \neg Headache)$
$Weather = Rain$		
$Weather = Cloudy$		
$Weather = Snow$		

Conditional probabilities result from unconditional probabilities (if $P(B) > 0$) (per definition):

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

- Product rule: $P(A \wedge B) = P(A \mid B)P(B)$
- Similarly: $P(A \wedge B) = P(B \mid A)P(A)$

Joint Probability

The agent assigns probabilities to every proposition in the domain.

An atomic event is an assignment of values to all random variables $X_1, \ldots, X_n$ (= complete specification of a state).

Example: Let X and Y be boolean variables. Then we have the following 4 atomic events: $X \wedge Y$, $X \wedge \neg Y$, $\neg X \wedge Y$, $\neg X \wedge \neg Y$.

The joint probability distribution $P(X_1, \ldots, X_n)$ assigns a probability to every *atomic event*.

	$Toothache$	$\neg Toothache$
$Cavity$	0.04	0.06
$\neg Cavity$	0.01	0.89

Probability distribution

Defines probability for **all possible value assignments**

Example 1:

$$P(Pneumonia = True) = 0.001$$
$$P(Pneumonia = False) = 0.999$$

$Pneumonia$	$\mathbf{P}(Pneumonia)$
$True$	0.001
$False$	0.999

$$P(Pneumonia = True) + P(Pneumonia = False) = 1$$

Probabilities sum to 1 !!!

Example 2:

$$P(WBCcount = high) = 0.005$$
$$P(WBCcount = normal) = 0.993$$
$$P(WBCcount = high) = 0.002$$

$WBCcount$	$\mathbf{P}(WBCcount)$
$high$	0.005
$normal$	0.993
low	0.002

Joint probability distribution

Joint probability distribution (for a set variables)

- Defines probabilities for **all possible assignments of values to variables in the set**

Example: variables *Pneumonia* and *WBCcount*

$\mathbf{P}(pneumonia, WBCcount)$

Is represented by 2×3 array(matrix)

Pneumonia \ *WBCcount*	*high*	*normal*	*low*
True	0.0008	0.0001	0.0001
False	0.0042	0.9929	0.0019

Full joint distribution

- **the joint distribution for all variables in the problem**
  - It defines the complete probability model for the problem

Example: pneumonia diagnosis

- **Variables:** *Pneumonia, Fever, Paleness, WBCcount, Cough*
- Full joint probability: P(*Pneumonia, Fever, Paleness, WBCcount, Cough)*
  - defines the probability for all possible assignments of values to these variables

$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = T, Paleness = T)$$

$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = T, Paleness = F)$$

$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = F, Paleness = T)$$

… etc

- **How many probabilities are there?**

All relevant probabilities can be computed using the joint probability by expressing them as a disjunction of atomic events.

Examples:

$$\begin{aligned} P(Cavity \vee Toothache) = {} & P(Cavity \wedge Toothache) \\ & + P(\neg Cavity \wedge Toothache) \\ & + P(Cavity \wedge \neg Toothache) \end{aligned}$$

We obtain unconditional probabilities by adding across a row or column:

$$P(Cavity) = P(Cavity \wedge Toothache) + P(Cavity \wedge \neg Toothache)$$

$$P(Cavity \mid Toothache) = \frac{P(Cavity \wedge Toothache)}{P(Toothache)} = \frac{0.04}{0.04 + 0.01} = 0.80$$

Bayes' Rule

We know (product rule):

$$P(A \wedge B) = P(A \mid B)P(B) \text{ and } P(A \wedge B) = P(B \mid A)P(A)$$

By equating the right-hand sides, we get

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

$$\boxed{\Rightarrow P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}}$$

where *A* and *B* are events, *P(A|B)* is the conditional probability that event *A* occurs given that event *B* has already occurred (*P(B|A)* has the same meaning but with the roles of *A* and *B* reversed) and *P(A)* and *P(B)* are the probabilities of event *A* and event *B* occurring respectively.

Applying Bayes' Rule

$$
\begin{aligned}
P(Toothache \mid Cavity) &= 0.4 \\
P(Cavity) &= 0.1 \\
P(Toothache) &= 0.05 \\
P(Cavity \mid Toothache) &= \frac{0.4 \times 0.1}{0.05} = 0.8
\end{aligned}
$$

Why don't we try to assess $P(Cavity \mid Toothache)$ directly?

$P(Toothache \mid Cavity)$ (causal) is more robust than $P(Cavity \mid Toothache)$ (diagnostic):

Bayes' Rule

- Useful for assessing diagnostic probability from causal probability

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

Medical diagnosis

- from past cases we know P(symptoms|disease), P(disease), P(symptoms)
- for a new patient we know symptoms and looking for diagnosis P(disease|symptoms)

Bayes' in Action

Example:

A doctor knows that the disease meningitis causes the patient to have a stiff neck, say, 50% of the time. The doctor also knows some unconditional facts: the prior probability that a patient has meningitis is 1/50,000, and the prior probability that any patient has a stiff neck is 1/20. Let *s be the proposition that the patient has a stiff neck and m be the proposition that the patient has meningitis.*

Bayes' rule (cont'd)

P(StiffNeck=true | Meningitis=true) = 0.5

P(Meningitis=true) = 1/50000

P(StiffNeck=true) = 1/20

P(Meningitis=true | StiffNeck=true)

= P(StiffNeck=true | Meningitis=true) P(Meningitis=true) / P(StiffNeck=true)

= (0.5) * (1/50000) / (1/20)

= 0.0002

That is, we expect only 1 in 5000 patients with a stiff neck to have meningitis.

Applications

- **In finance**, for example, Bayes' theorem can be used to rate the risk of lending money to potential borrowers.
- **In medicine**, the theorem can be used to determine the accuracy of medical test results by taking into consideration how likely any given person is to have a disease and the general accuracy of the test.

Application of Bayes' theorem in Artificial intelligence

- **Following are some applications of Bayes' theorem:**
- It is used to calculate the next step of the robot when the already executed step is given.
- Bayes' theorem is helpful in weather forecasting.

Representing Knowledge in an Uncertain Domain

- if a patient has a liver disorder

What could be the cause of this liver disorder?

gallstones could be a cause
a history of hepatitis could be another
it could be alcoholism or many others

Unobservable

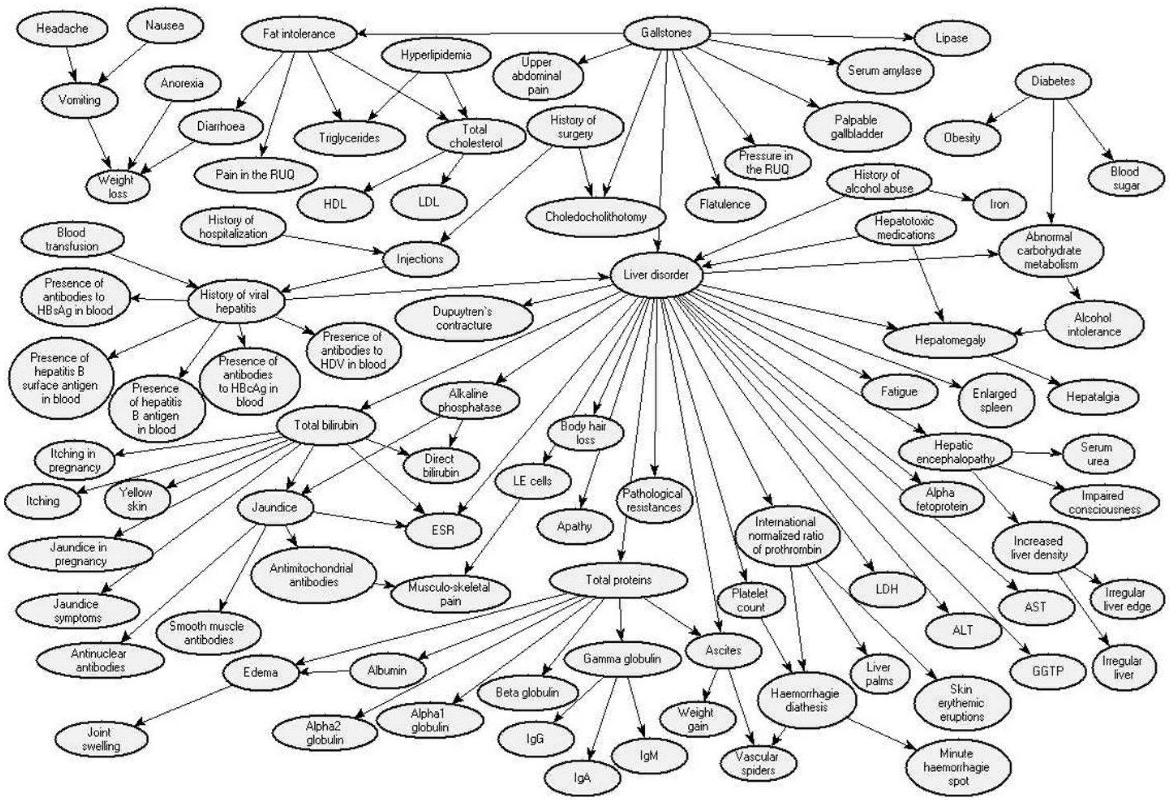
what does liver disorder cause?

It cause fatigue, body hair loss, enlarged spleen, etc

Therefore A Bayes Network can be used for cause and effect based on probability to explain a specific case, given a set of known probabilities.

Bayesian Network

- A Bayesian Network is a network that can explain quite complicated structures, like in our example of the cause of a liver disorder.



A Bayesian Network Model for Diagnosis of Liver Disorders

- A Bayesian Network is composed of nodes, where the nodes correspond to events that you might or might not know.
- These nodes called random variables are connected by arrows, and if there is an arrow from X to Y, X is said to be parent to Y.
- Each node Xi has a **conditional** probability distribution P(Xi|Parents(Xi)).
- Bayes Networks define the **probability distribution** over graphs of random variables.

Bayesian Networks

- A Bayesian network specifies a joint distribution in a structured form
- Represent dependence/independence via a directed graph
  - Nodes = random variables
  - Edges = direct dependence
- Structure of the graph ⇔ Conditional independence relations

In general,

$$p(X_1, X_2, \ldots X_N) = \prod p(X_i \mid \text{parents}(X_i))$$

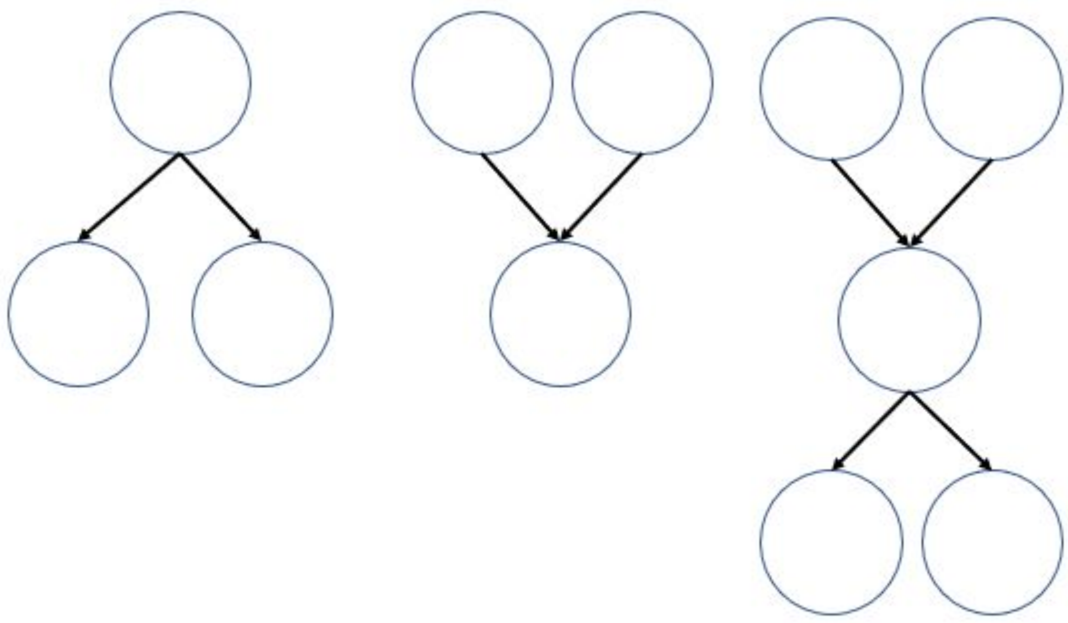
The full joint distribution

The graph-structured approximation

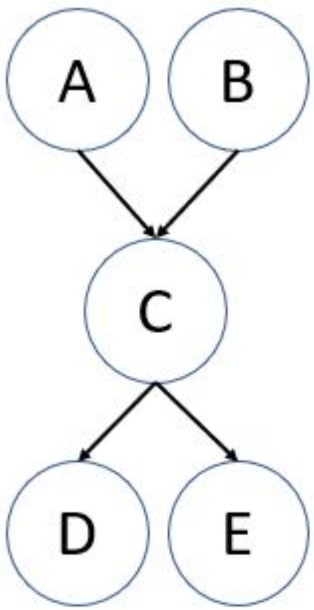
- Requires that graph is acyclic (no directed cycles)
- 2 components to a Bayesian network
  - The graph structure (conditional independence assumptions)
  - The numerical probabilities (for each variable given its parents)

Types of Bayesian networks

There are many different types of Bayes networks (see below)



Let us consider the example of last one



A and B are only dependent on their own variable

So Distribution is P(A) and P(B),

C is conditioned on A and B

so we have P(C|A,B)

E are conditioned on C

P(D|C), P(E|C).

$P(A,B,C,D,E) = P(A)*P(B)*P(C|A,B)*P(D|C)*P(E|C)$

Example (Perls' example)

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes the alarm is set off by minor earthquakes. Is there a burglar?

- John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm.
- Mary likes rather loud music and sometimes misses the alarm.

- Variables: *Burglary*, *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

Constructing a Bayesian Network: Step 1

- Order the variables in terms of causality (may be a partial order)

  e.g., {E, B} -> {A} -> {J, M}

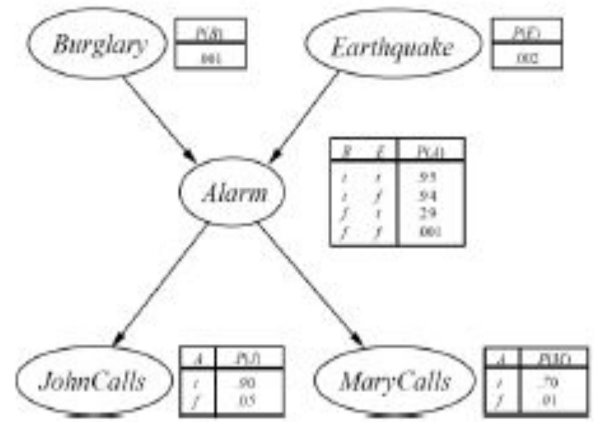
- $P(J, M, A, E, B) = P(J, M \mid A, E, B)\ P(A \mid E, B)\ P(E, B)$

$$\approx P(J, M \mid A) \qquad P(A \mid E, B)\ P(E)\ P(B)$$

$$\approx P(J \mid A)\ P(M \mid A)\ P(A \mid E, B)\ P(E)\ P(B)$$

These CI assumptions are reflected in the graph structure of the Bayesian network

Constructing this Bayesian Network: Step 2



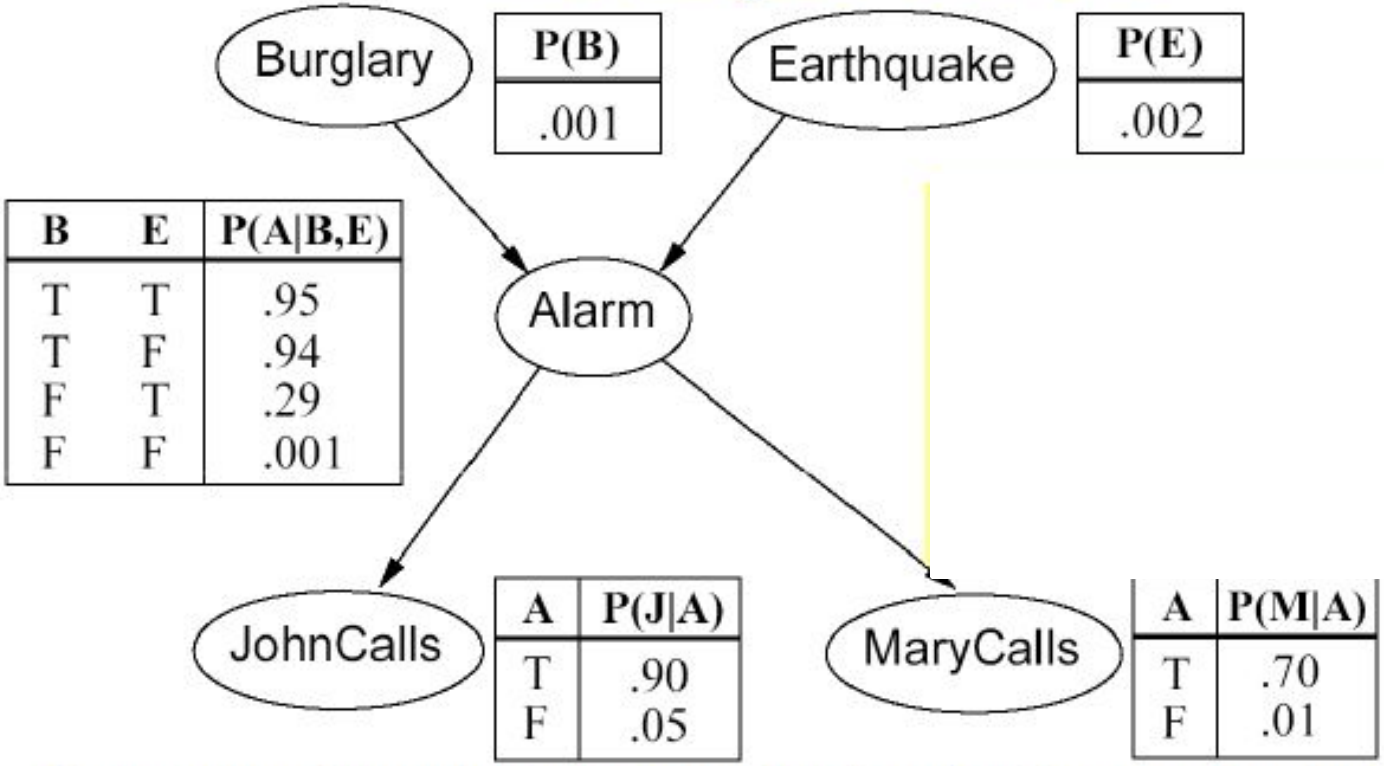
- P(J, M, A, E, B) =
  P(J | A) P(M | A) P(A | E, B) P(E) P(B)

- There are 3 conditional probability tables (CPDs) to be determined:
  P(J | A), P(M | A), P(A | E, B)
  - Requiring 2 + 2 + 4 = 8 probabilities

- And 2 marginal probabilities P(E), P(B) -> 2 more probabilities

- Where do these probabilities come from?
  - Expert knowledge
  - From data (relative frequency estimates)

Example cont'd



P(B)
.001

P(E)
.002

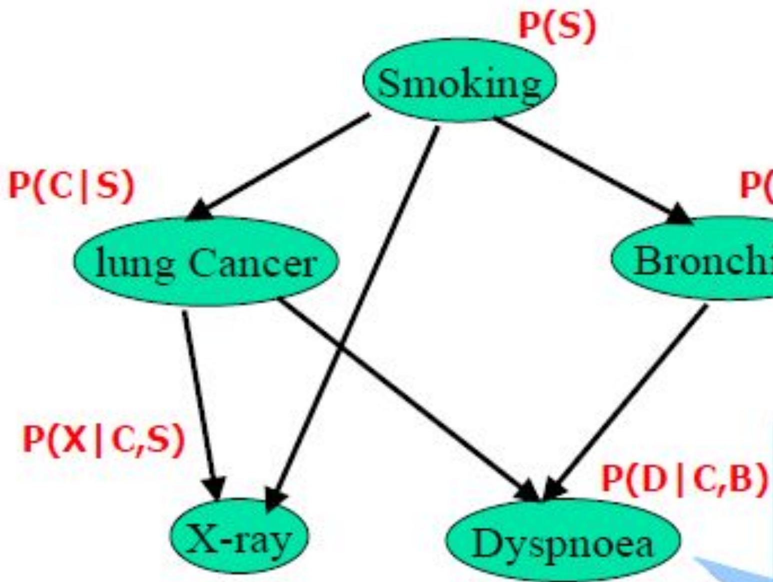
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

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

The topology shows that burglary and earthquakes directly affect the probability of alarm, but whether Mary or John call depends only on the alarm.

Bayesian Network: $\mathbf{BN} = (\mathbf{G}, \boldsymbol{\Theta})$



G - directed acyclic graph (DAG)
nodes – random variables
edges – direct dependencies

Θ - set of parameters in all conditional probability distributions (CPDs)

CPD:

C	B	D=0	D=1
0	0	0.1	0.9
0	1	0.7	0.3
1	0	0.8	0.2
1	1	0.9	0.1

CPD of node X: P(X | parents(X))

Compact representation of joint distribution in a **product form** (chain rule):

$$P(S, C, B, X, D) = P(S)\, P(C|S)\, P(B|S)\, P(X|C,S)\, P(D|C,B)$$

$1+2+2+4+4=13$ **parameters instead of** $2^5 = 32$

Semantics

Suppose we have the variables $X_n, \ldots, X_1$.

The probability for them to have the values $x_n, \ldots, x_1$, respectively, is $P(x_n, \ldots, x_1)$:

$$
\begin{aligned}
&= P(x_n, \ldots, x_1) \\
&= P(x_n \mid x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1) \\
&= P(x_n \mid x_{n-1}, \ldots, x_1) P(x_{n-1} \mid x_{n-2}, \ldots, x_1) P(x_{n-2}, \ldots, x_1) \\
&= \ldots \\
&= \prod_{i=1}^{n} P(x_i \mid x_{i-1}, \ldots, x_1) = \prod_{i=1}^{n} P(x_i \mid parents(x_i))
\end{aligned}
$$

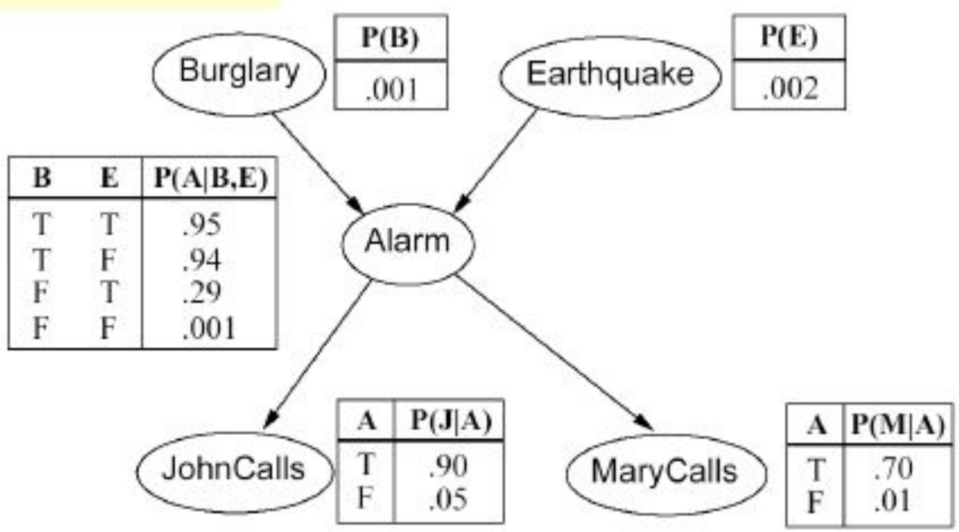
$P(x_n, \ldots, x_1)$: is short for $P(X_n = x_n, \ldots, X_1 = x_1)$

e.g.,

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

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

$= \ldots$



P(B)
.001

P(E)
.002

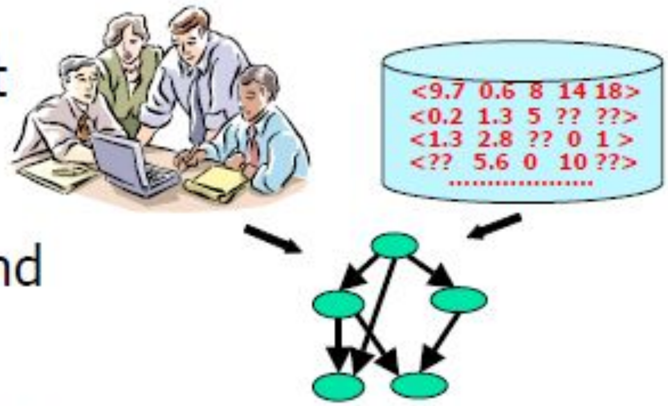
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

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

Learning Bayesian Networks

- Combining domain expert knowledge with data



- Efficient representation and inference
- Incremental learning: P(H) ↗ or ↘
- Handling missing data: **<1.3 2.8 ?? 0 1 >**
- Learning causal relationships: 

What are BNs useful for?

- Diagnosis: $P(\text{cause}|\text{symptom})=?$
- Prediction: $P(\text{symptom}|\text{cause})=?$
- Classification: $\max\limits_{class} P(\text{class}|\text{data})$
- Decision-making (given a cost function)