Bayseain network

Drawbacks of bayes theorem

- P(H|E) can be calculated
- If P(H|E1,E2…En) is to be calculated complexity for joint probability goes up to 2^n for n different propositions
- Too many probabilities required
- Space needed to store all probabilities
- Time required to calculate probability

Improvement over bayes

- Attaching certainty factor to rules
- Bayesian network
- Dempster-Shafer theory
- Fuzzy logic

Joint Distribution

- k random variables $X_1, \ldots, X_k$
- The joint distribution of these variables is a table in which each entry gives the probability of one combination of values of $X_1, \ldots, X_k$
- Example:

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

P(¬Cavity∧Toothache)

P(Cavity∧¬Toothache)

Inference using Joint Distribution Says It All

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

- P(Toothache) = P((Toothache ∧Cavity) v (Toothache∧¬Cavity))

  = P(Toothache ∧Cavity) + P(Toothache∧¬Cavity)

  = 0.04 + 0.01 = 0.05

- P(Toothache v Cavity)

  = P((Toothache ∧Cavity) v (Toothache∧¬Cavity)

  v (¬Toothache ∧Cavity))

  = 0.04 + 0.01 + 0.06 = 0.11

Joint Distribution Says It All

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

- P(Toothache) = ??

- P(Toothache v Cavity) = ??

Conditional Probability

- Definition:
  $P(A|B) = P(A \wedge B) / P(B)$
- Read $P(A|B)$: probability of A given B

- can also write this as:
  $P(A \wedge B) = P(A|B)\, P(B)$
- called the **product rule**

Example

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

- P(Cavity|Toothache) = P(Cavity∧Toothache) / P(Toothache)
  - P(Cavity∧Toothache) = ?
  - P(Toothache) = ?
  - P(Cavity|Toothache) = 0.04/0.05 = 0.8

Full joint distribution

- In an uncertain domain given variables x1,x2,…..xn in sample space .
- If no other information is available their full joint distribution will require 2^n entries
- Full joint probability is given as
  - $\boldsymbol{P}(X_1, \ldots, X_n) = \pi_{i=1} \boldsymbol{P}(X_i \mid X_1, \ldots, X_{i-1})$
  - (chain rule)

Bayesian Network

- To describe a real world it is not necessary to use a large joint probability table in which probability of all combination of events is stated
- Most of events are conditionally independent of each other
- Network representation of knowledge is used to graphically exhibit interdependence between piece of knowledge
- It shoes degree of belief of proposition

Bayesian Network

- Bayesian networks (BNs), also known as *belief networks* (or Bayes nets for short), belong to the family of probabilistic *graphical models (GMs)*.
- These graphical structures are used to represent knowledge about an uncertain domain.
- Each node in the graph represents a random variable
- Edges between the nodes represent probabilistic dependencies among the corresponding random variables.

Bayesian Network

- BNs correspond to graphical model structure known as a *directed acyclic graph (DAG)* which is popular in statistics, the machine learning, and the artificial intelligence societies.
- They enable an effective representation and computation of the joint probability distribution (JPD) over a set of random variables

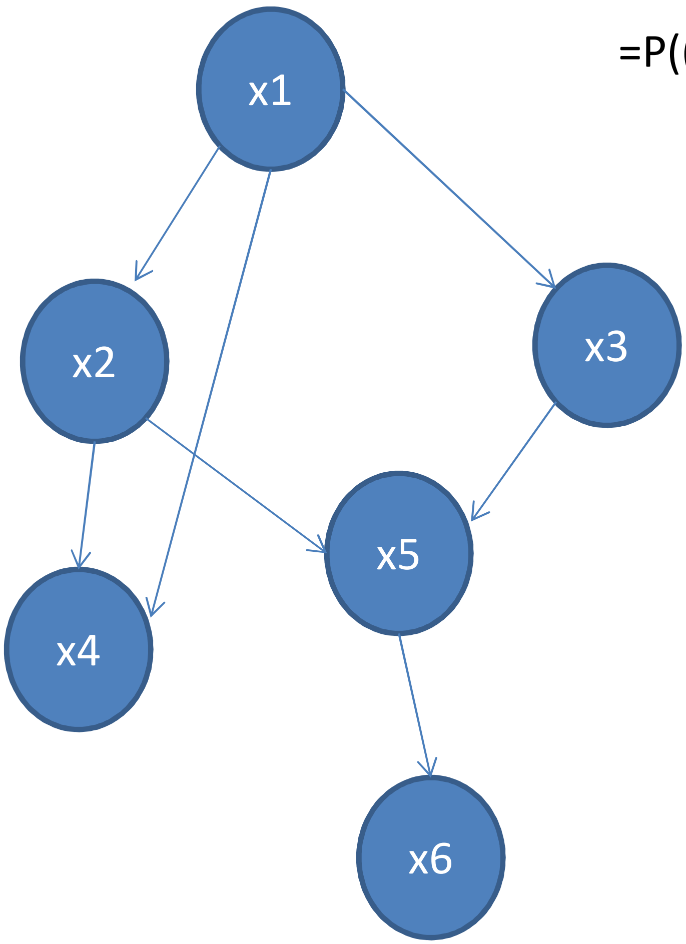
Inference from BN

- Inferencing from BN can be done by propagating probabilities of given and related nodes through network to one or more conclusion nodes
- Representation of uncertain knowledge through proposition x1,x2…xn by joint probability P(x1,x2…xn) require 2^n entries
- JPD can be reduced by considering the inter casual relations amongst events through a directed graph
- Strength of dependencies is qunatified by conditional probability

Inference

P(x1,x2,x3,x4,x5,x6)=P(x1|x2,x3,x4,x5,x6)(P(x2,x3,x4,x5,x6)

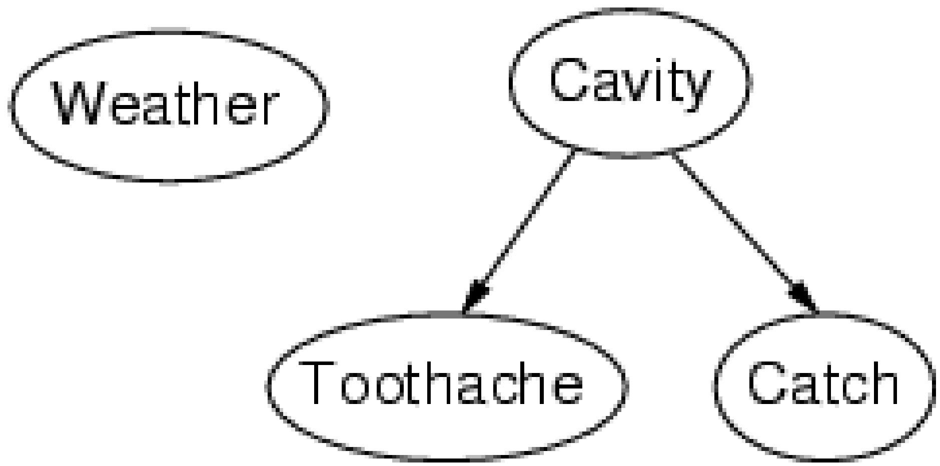
=P((x1)P(x2|x1)P(x3|x1)P(x4|x2)P(x5|x2x3)P(x6|x5)



Once such network is built, inferencing can be done by propagating beliefs

Example

- Topology of network encodes conditional independence assertions:



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

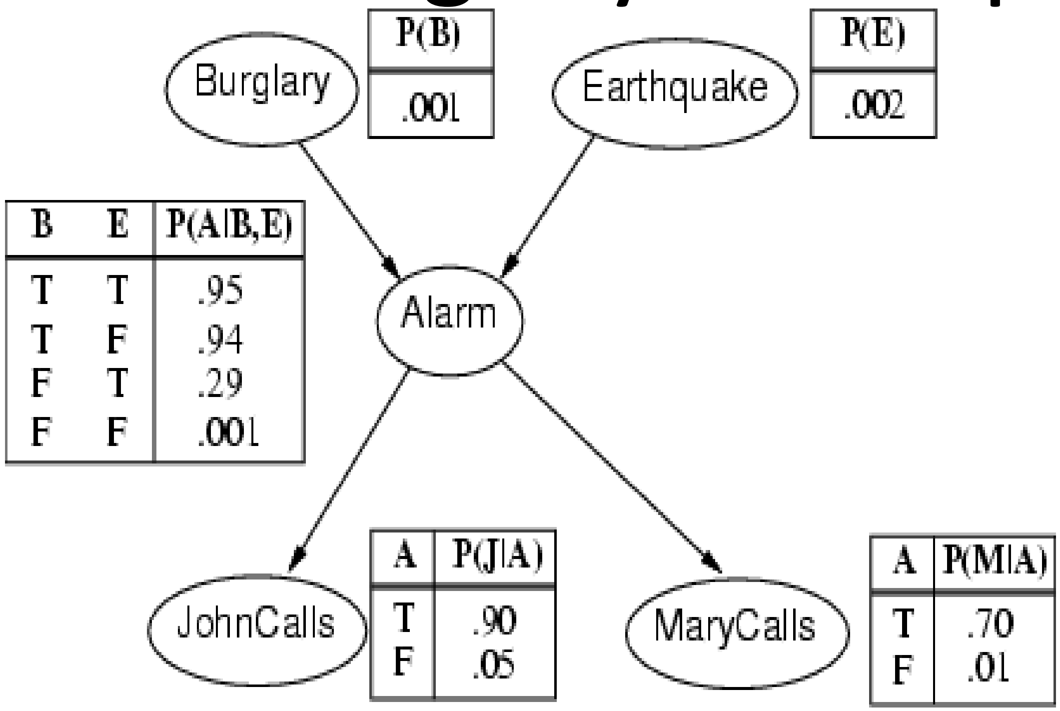
Example

- 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: *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 t

Burglary alarm problem



P(JC,MC,A,B,E)=P(JC|MC,A,B,E)P(MC,A,B,E)

=P(JC|A)P(MC|A,B,E)P(A,B,E)

=P(JC|A)P(MC|A)P(A|B,E)P(B,E)

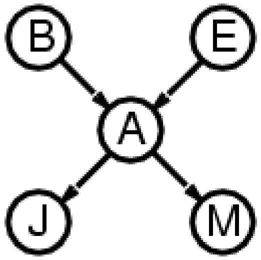
=P(JC|A)P(MC|A)P(A|B,E)P(B|E)P(E)

= P(JC|A)P(MC|A)P(A|B,E)P(B)P(E)

=0.90*0.70*0.95*0.001*0.002

Compactness

- A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values

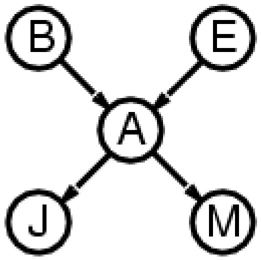


- Each row requires one number p for X_i = *true* (the number for X_i = *false* is just 1-p)
- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers
- I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution
- For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5-1 = 31$)

Semantic

The full joint distribution is defined as the product of the local conditional distributions:

$$\mathbf{P}\,(X_1, \ldots, X_n) = \pi_{i=1}\,\mathbf{P}\,(X_i \mid Parents(X_i))$$



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

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

A Bayesian Network has two parts:

1) <u>qualitative part</u>
    - the structure
    - directed acyclic graph (DAG)
    - vertices represent variables
    - edges represent relations between variables

2) <u>quantitative part</u>
    - the strength of relationship between variables
    - conditional probability function