Uncertain Knowledge and Reasoning

By Dr. Sonali Patil

Comparing abduction, deduction, and induction

Deduction:	major premise:	All balls in the box are black	**A => B** **A** **---------** **B**
	minor premise:	These balls are from the box	
	conclusion:	These balls are black	
Abduction:	rule:	All balls in the box are black	**A => B** **B** **-------------** **Possibly A**
	observation:	These balls are black	
	explanation:	These balls are from the box	
Induction:	case:	These balls are from the box	**Whenever A then B** **-------------** **Possibly A => B**
	observation:	These balls are black	
	hypothesized rule:	All ball in the box are black	

Deduction reasons from causes to effects
Abduction reasons from effects to causes
Induction reasons from specific cases to general rules

Reasoning Under Uncertainty

- Uncertainty
- Sources of uncertainty
- Methods to handle Uncertainty
- Probability Theory
- Uncertainty and Rational Decisions
- Basic Probability Notations

Uncertainty

Reasoning under uncertainty.

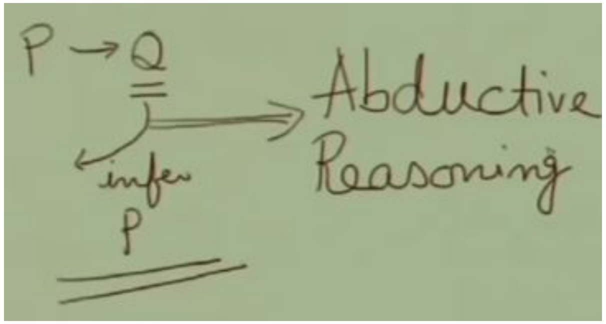
- There are different types of uncertainties. What are the different ways in which you can deal with that?

- The doorbell problem
  - The doorbell rang at 12 O'clock at midnight.
  - Que to ans
    - was someone there at the door?
    - Mohan was sleeping in the room. Did Mohan wake up when the doorbell rang?
    - My fact is that the doorbell rang at 12 O'clock in the midnight. Therefore if we place the propositions in the logic form

- Proposition 1: $AtDoor(x) \Rightarrow Doorbell$
- Proposition 2: $Doorbell \Rightarrow Wake(Mohan)$

Uncertainty

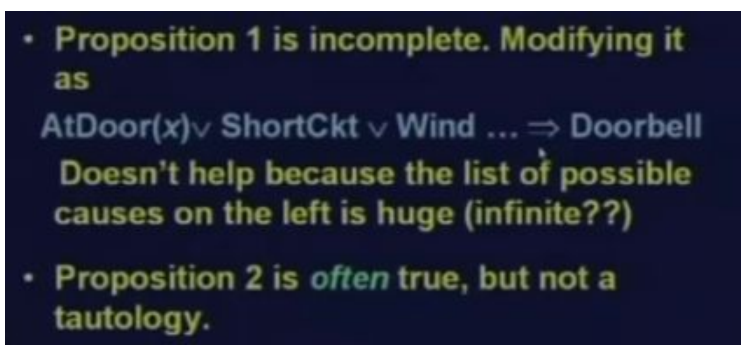
- Given Doorbell, can we say AtDoor(x), because AtDoor(x) $\rightarrow$ Doorbell?
- Can we say that there is some one at Door? We can using the deductive reasoning/normal implication (p implies q, if p true…Q is necessarily true, if p false…q may be or may not be true)
- Abductive Reasoning (p implies q and we find q is true then we infer p. Most of the time right, but may not always). Other reasons, though rare
  - Short Circuit
  - Wind
  - Dog or other Animal pressed the button



Uncertainty

- Given Doorbell, can we say Wake(Mohan), because Doorbell $\rightarrow$ Wake(Mohan)?
- Using, Deductive Reasoning. Yes, if proposition 2 is always true
- However always this may not be true (May be tired and in sound sleep)

Hence, we cannot answer either of Questions with certainty.



Uncertainty

Planning Example

- Let action A(t) denote leaving for the airport t minutes before the flight
  - For a given value of t, will A(t) get me there on time?
- Problems:
  - Partial observability (roads, other drivers' plans, etc.)
  - Noisy sensors (traffic reports)
  - Uncertainty in action outcomes (flat tire, etc.)
  - Immense complexity of modelling and predicting traffic

Uncertainty

- Diagnosis always involves uncertainty.
- Eg:

  Dental diagnosis: (toothache)

  Toothache $\longrightarrow$ *Cavity*

  Its wrong as not all people with toothaches have cavity. It may be due other problems

  Toothache $\longrightarrow$ *Cavity V Gum Problem V Abscess…….*

  In order to complete the list, we have to add an almost unlimited list of possible problems

The causal rule for this:

  Cavity $\longrightarrow$ *Toothache*

  This is not also the right one. Not all cavities cause pain

Uncertainty

- Trying to cope up with domains like Medical diagnosis fails for 3 main reasons:

✔ Laziness:

Too hard to list out all antecedence & consequents needed to ensure an exception-less rule and too hard to use such rules.

✔ Theoretical ignorance:

The domains like Medical science has No complete theory for the domain.

✔ Practical ignorance:

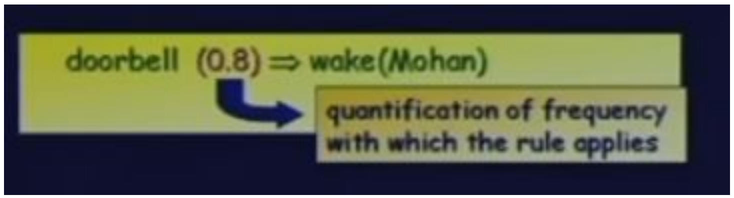
Even all rules are known – uncertain about a particular patient. Not all test have been or can be run.

Uncertainty

- The problems like Doorbell/diagnosis are very common in real world
- In AI, we need to reason under such circumstances
- We solve such problems by proper modelling of Uncertainty and impreciseness and developing appropriate reasoning techniques

Sources of uncertainty

- Implications may be weak



- Imprecise language like often, rarely, sometimes
  - Need to quantify these terms of frequencies
  - Need to design rules for reasoning with these frequencies

- Precise information (input) may be too complex
  - Too many antecedents or consequents

- Incomplete Knowledge
  - We may not know or guess all the possible antecedents or consequents
  - The bell rang due to some spooky reason

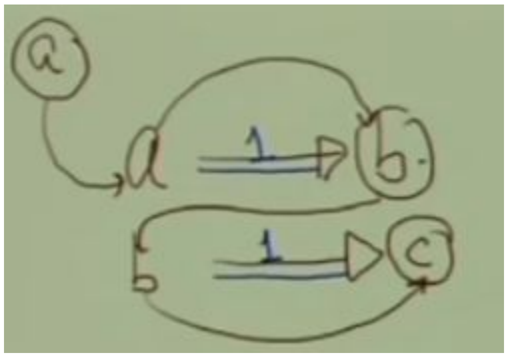
Sources of uncertainty

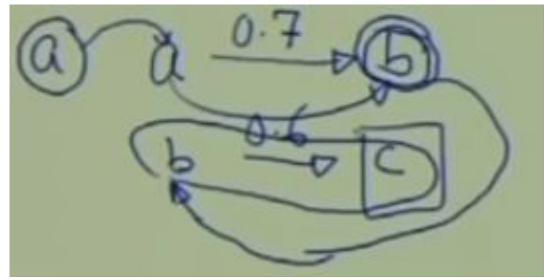
- Conflicting Information
  - Patient-complicated symptoms-two diff doctors-may be possible they differ in there diagnosis if the symptoms do not lead to a vary obvious disease

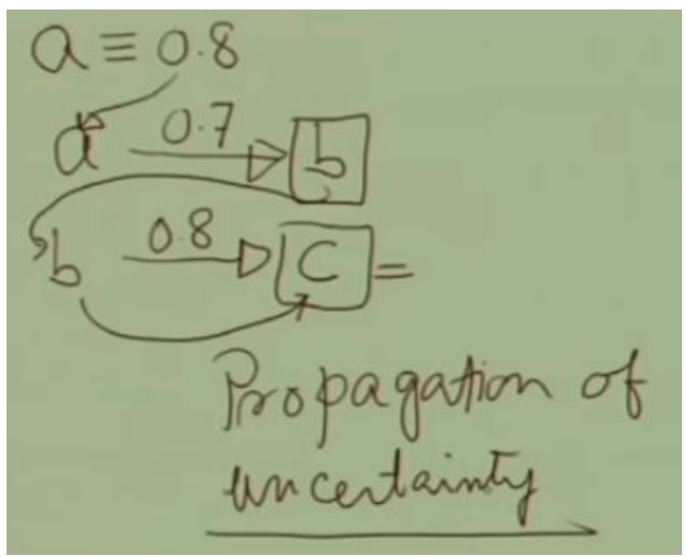
Experts often provide conflicting information:
quantification of measure of belief

- Propagation of Uncertainties
  - In absence of interdependencies of propagation of uncertain knowledge the uncertainty of the conclusions increases

Tomorrow(sunny) [0.6], Tomorrow(warm) [0.8]
Tomorrow(sunny) $\wedge$ Tomorrow(warm) [?]


a
a
1
b.
b
1
c


a
a
0.7
b
0.6
b
c


a ≡ 0.8
a
0.7
b
b
0.8
C
Propagation of uncertainty

Sources of uncertainty

- Uncertain **inputs**
  - Missing data
  - Noisy data
- Uncertain **knowledge**
  - Multiple causes lead to multiple effects
  - Incomplete enumeration of conditions or effects
  - Incomplete knowledge of causality in the domain
  - Probabilistic/stochastic effects
- Uncertain **outputs**
  - Abduction and induction are inherently uncertain
  - Default reasoning, even in deductive fashion, is uncertain
  - Incomplete deductive inference may be uncertain

□ Probabilistic reasoning only gives probabilistic results (summarizes uncertainty from various sources)

Methods of handling Uncertainty

- Fuzzy Logic
  - Logic that extends traditional 2-valued logic to be a continuous logic (values from 0 to 1)
    - while this early on was developed to handle natural language ambiguities such as “you are *very* tall” it instead is more successfully applied to device controllers
- Probabilistic Reasoning
  - Using probabilities as part of the data and using Bayes theorem or variants to reason over what is most likely
- Hidden Markov Models
  - A variant of probabilistic reasoning where internal states are not observable (so they are called hidden)
- Certainty Factors and Qualitative Fuzzy Logics
  - More ad hoc approaches (non formal) that might be more flexible or at least more human-like (MYCIN expert system)
- Neural Networks

Uncertainty tradeoffs

- **Bayesian networks:** Nice theoretical properties combined with efficient reasoning make BNs very popular; limited expressiveness, knowledge, engineering challenges may limit uses
- **Non-monotonic logic:** Represent commonsense reasoning, but can be computationally very expensive
- **Certainty factors:** Not semantically well founded
- **Fuzzy reasoning:** Semantics are unclear (fuzzy!), but has proved very useful for commercial applications

Probability Theory

- Deals with degrees of belief
- Provides a way of summarizing the uncertainty that comes from our laziness & ignorance thereby solving the qualification problem (specifying all exceptions)
  - A90 will take us to airport on time, as long as car doesn't break down or run out of gas, does not indulge into accident, no accidents on bridge, plane doesn't live early, no meteorite hits the car, and ….)
- Toothache $\xrightarrow{0.8}$ cavity
  - The probability that the patient has a cavity, given that she has a toothache is 0.8

Probability Theory

- Consider previous statement: “The probability that the patient has a cavity, given that she has a toothache is 0.8”
- If we later learn that patient has a history of gum disease we can say “The probability that the patient has a cavity, given that she has a toothache and a history of gum disease, is 0.4”
- If further we gather evidence, we can say “The probability that the patient has a cavity, given all we know now, is almost zero”
- Above three statements do not contradict each other; each is a separate assertion about a difference knowledge state

Uncertainty and rational decisions

- “Say A90 has 92% chance of catching our flight. Is it rational choice? Not necessarily
- A180 has higher probability of reaching. If its vital to not miss the flight, then its worth risking the longer wait time at airport
- A1440 almost guarantees reaching on time but I’d have to stay overnight in the airport (intolerable wait and may be unpleasant diet of airport food)
- To make choices, an agent must have preferences between different possible outcomes of various plans
- **Utility Theory** is used to represent & reason with preferences

Uncertainty and rational decisions

- **Utility Theory**
  - Every state has a degree of usefulness or utility, to an agent and the agent will prefer states with higher utility
  - The utility state is relative to agent
  - Ex. Consider the state in which White has checkmated Black in chess. Here, Utility is high for agent playing White but low for agent playing Black
  - A Utility function can account for any set of preferences- quirky or typical, noble or perverse

Uncertainty and rational decisions

- **Decision Theory**
  - Preferences as expressed by utilities, are combined with probabilities in general theory of rational decisions

  Decision Theory = Probability Theory + Utility Theory
- **Maximum Expected Utility(MEU)**
  - An agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all possible outcomes of the action. This is principle of MEU
  - Here the term expected is not vague. Its average or statistical mean of outcomes weighted by the probability of outcome
- The basic difference between A decision-theoretic agent & other agents is that the former's belief state represents not just the possibilities for world states but also their probabilities

Uncertainty and rational decisions

```
function DT-AGENT(percept) returns an action
   persistent: belief_state, probabilistic beliefs about the current state of the world
               action, the agent's action

   update belief_state based on action and percept
   calculate outcome probabilities for actions,
       given action descriptions and current belief_state
   select action with highest expected utility
       given probabilities of outcomes and utility information
   return action
```

A decision-theoretic agent that selects rational actions.

Uncertainty and rational decisions summary

- **Rational** behavior:
  - For each possible action, identify the possible outcomes
  - Compute the **probability** of each outcome
  - Compute the **utility** of each outcome
  - Compute the probability-weighted **(expected) utility** over possible outcomes for each action
  - Select the action with the highest expected utility (principle of **Maximum Expected Utility**)

Bayesian reasoning

- Probability theory
- Bayesian inference
  - Use probability theory and information about independence
  - Reason diagnostically (from evidence (effects) to conclusions (causes)) or causally (from causes to effects)
- Bayesian networks
  - Compact representation of probability distribution over a set of propositional random variables
  - Take advantage of independence relationships

Basic Probability Notation

- A **random variable** is a variable whose possible values are the numerical outcomes of a random experiment.
  - It is a function which associates a unique numerical value with every outcome of an experiment.
  - Its value varies with every trial of the experiment.
  - It describes an outcome that cannot be determined in advance
  - It is Boolean , Discrete or continuous
  - Ex. Roll of a die, number of emails received in a day etc.
- The **sample space S** of the random variable X is the set of all possible worlds
  - The possible worlds are mutually exclusive & exhaustive (at a time one possible outcome and all possible outcomes are in the S)
  - Tossing a coin: S={H,T}
  - Tossing two coins simultaneously S={HH, HT, TH, TT}
  - Rolling a die: S={ 1,2,3,4,5,6}

Basic Probability Notation

- An **atomic event** is a complete **specification of the state** of the world about which the agent is uncertain.
- Eg:

  Cavity & Toothache has four distinct atomic events.

  $Cavity = False \wedge Toothache = True$

  $Cavity = True \wedge Toothache = True$

  $Cavity = False \wedge Toothache = False$

  $Cavity = True \wedge Toothache = False$

Basic Probability Notation

- The sample space is denoted by Ω (upper case omega) and elements in sample space are denoted by ω (lower case omega)

- $P(\omega)$ -> Probability of occurance of ω

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega \text{ and } \sum_{\omega \in \Omega} P(\omega) = 1 \, .$$

- Probabilistic assertions & queries are not about particular possible worlds, but about sets of them
  - The two dice add upto 11, Doubles are rolled; Picking ace from pack of cards, number of email> 100 in a day, etc.
  - These sets are called **events.** Event is subset of ω
  - Events are described by proposition in common language

Basic Probability Notation

– The probability associated with the proposition is defined to be the sum of the probabilities of the world in which the proposition holds

$$\text{For any proposition } \phi,\ P(\phi) = \sum_{\omega \in \phi} P(\omega)\ .$$

– ϕ is getting odd number after rolling the dice

S={1, 2, 3, 4, 5, 6}, ϕ ={1, 3, 5}

P(Odd)=P(1)+ P(3) + P(5)= 1/6 + 1/6 + 1/6 = ½

Unconditional or Prior Probabilities

- Degree of belief in proposition in the absence of any other information/evidence
- P(Fever)=0.1
  - The probability that the patient has fever is 0.1(in absence of any other information
- A die is rolled, P(odd), P(even) indicated the probability of getting the odd number and the probability of getting the even number on the rolled dice respectively. Both of these are prior probabilities
- When a pair of dice rolled simultaneously, the possible outcomes are 36. P(doubles), P(Total=15) are prior probabilities

Unconditional or Prior Probabilities

- The random variables here Fever, Doubles, Odd, Even are **Discrete Random variables** as they take finite number of distinct values
- The Boolean random variables have values True or false ex. P(cavity)
- A **continuous random variable** is a random variable that takes infinite number of distinct values
  - EX. $P(Temp=x) = Uniform_{[18C,26C]}(x)$
  - Expresses that the temperature id distributed uniformly between 10 and 26 degrees
  - This is called Probability Density Function

Conditional or Posterior Probabilities

- Let A be an event in the world and B be another event. Suppose that events A and B are not mutually exclusive, but occur conditionally on the occurrence of the other. The probability that event A will occur if event B occurs is called the conditional probability. Conditional probability is denoted mathematically as p(A|B) in which the vertical bar represents GIVEN and the complete probability expression is interpreted as "Conditional probability of event A occurring given that event B has occurred".

$$p(A|B) \square \frac{\text{the number of times A and B can occur}}{\text{the number of times B can occur}}$$

Conditional or Posterior Probabilities

- The number of times A and B can occur, or the probability that both A and B will occur, is called the **joint probability** of A and B. It is represented mathematically as p(A∩B). The number of ways B can occur is the probability of B, p(B), and thus

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

- The eq. of conditional can also be written in the form

$$p(A \cap B) = p(A|B)\ p(B) \rightarrow$$ **Product Rule**

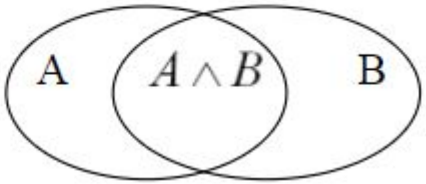
- Similarly, the conditional probability of event B occurring given that event A has occurred equals

$$p(B|A) = \frac{p(B \cap A)}{p(A)}$$

$$p(B \cap A) = p(B|A)\, p(A) \rightarrow$$ **Product Rule**

Probability Axioms

- All probabilities are between 0 & 1
  - $0 \leq P(A) \leq 1$
- Necessarily True propositions have probability 1 and necessarily false propositions have probability 0
  - (P(true) = 1 and P(false) = 0)
- Probability of disjunction
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$ → **Inclusion-Exclusion Principle**



- These axioms often called as Kolmogorov's axiom

Probability Axioms

- From Axioms we can derive other properties
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$ Substitute $B = \neg A$
  - $P(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A)$

    $1 = P(A) + P(\neg A) - 0$

    $\mathbf{P(\neg A) = 1 - P(A)}$

    $\mathbf{P(A) = 1 - P(\neg A)}$

- A and B mutually exclusive □ $P(A \vee B) = P(A) + P(B)$

  $P(e_1 \vee e_2 \vee e_3 \vee \ldots e_n) = P(e_1) + P(e_2) + P(e_3) + \ldots + P(e_n)$

  The probability of a proposition **a** is equal to the sum of the probabilities of the atomic events in which **a** holds

  e(a) – the set of atomic events in which **a** holds

  $$P(a) = \sum_{e_i \in e(a)} P(e_i)$$

Inference using Full Joint Distribution

Probability distribution $\mathbf{P}$**(Cavity, Toothache)**

	Toothache	~ Toothache
Cavity	0.04	0.06
~ Cavity	0.01	0.89

Sum of all entries =1

P(Cavity) = 0.04 + 0.06 = 0.1 (using Axioms)

P(Cavity $\vee$ Toothache) = 0.04 + 0.01 + 0.06 = 0.11

P(Cavity|Toothache) =
P(Cavity $\wedge$ Toothache)/P(Toothache)
= 0.04 / (0.04 + 0.01)
= 0.8

- Obtain P(~ cavity), P(Toothache), P(~ Toothache), P(cavity| ~ toothache) P(~ cavity| toothache), P(~ cavity| ~ toothache)

Inference using Full Joint Distribution

Start with the joint distribution $\mathbf{P}(Cavity, Catch, Toothache)$:

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

For any proposition ϕ, sum the atomic events where it is true:

$P(\phi) = \sum_{\omega:\omega \models \phi} P(\omega)$

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

$$P(cavity \vee toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

$$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

- Process- Marginalization or summing out.

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

You can also compute conditional probabilities:

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

$$P(cavity | toothache) = ?$$

Inference using Full Joint Distribution

P(cavity|toothache)= P(cavity^ Toothache)/ P(Toothache)
= (0.108+0.012)/(0.108+0.012+0.016+0.064)
= 0.6

- Observe, P(cavity|toothache)+P(~ cavity|toothache) =0.6 + 0.4=1 as it should be
- 1/P(toothache) remains constant no matter which value of cavity we calculate. Such constants in probability are called as normalization constant

Inference using Full Joint Distribution

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Denominator can be viewed as a normalization constant α

$$\begin{aligned}
\mathbf{P}(Cavity|toothache) &= \alpha\,\mathbf{P}(Cavity, toothache) \\
&= \alpha\,[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \\
&= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle] \\
&= \alpha\,\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle
\end{aligned}$$

General idea: compute distribution on query variable by fixing evidence variables and summing over hidden variables

Independence

A and B are independent iff

$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$

- Independence is simplifying the modelling assumption
- Variable represented for probability are

  P(Weather, toothache , catch , cavity)
- It can be deduced as

P(weather= cloudy) P(toothache ,catch ,cavity)

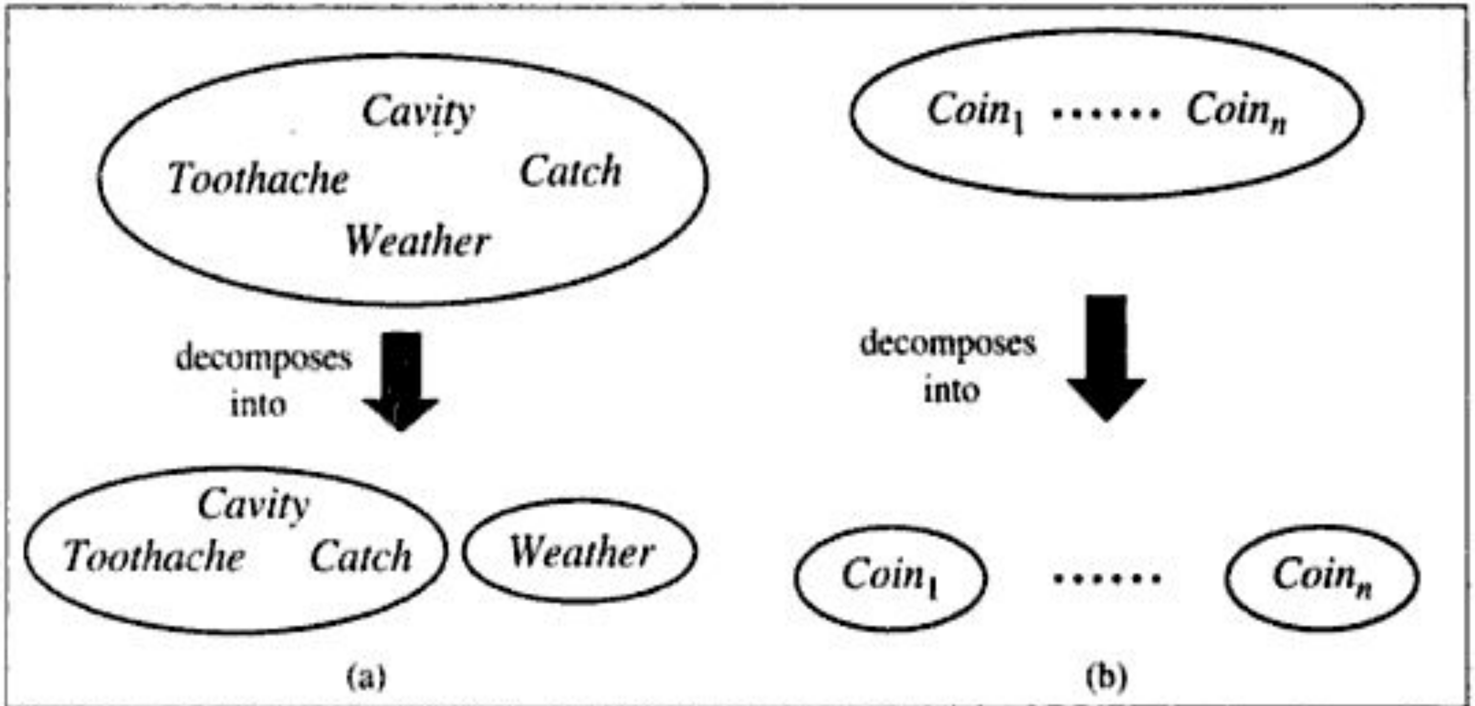


Figure 13.5 Two examples of factoring a large joint distribution into smaller distributions, using absolute independence. (a) Weather and dental problems are independent. (b) Coin flips are independent.

32 entries reduced to 12; for n independent biased coins, $2^n \to n$

Absolute independence powerful but rare

How to verify Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P(W)$

W	P
sun	0.6
rain	0.4

- Given a joint distribution $P_1(T,W)$ how to verify T and W are independent or not
- Build marginals for each of the variables. Her two variables so two marginals

How to verify Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P_2(T, W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$P(W)$

W	P
sun	0.6
rain	0.4

- Calculate another distribution P2(T,W) as P(T)*P(W)
- If P1(T,W)= P2(T,W)… T and W are independent

Example independence

- N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

$P(X_2)$

H	0.5
T	0.5

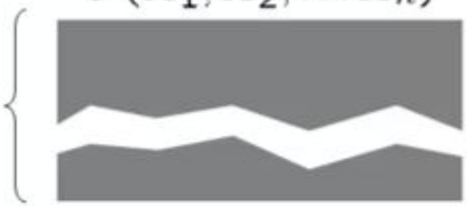
$\dots$

$P(X_n)$

H	0.5
T	0.5

$P(X_1, X_2, \dots X_n)$

2^n



Bayesian or Bayes Rule

- Let A be an event in the world and B be another event.

  Hence from product rule

  $P(A \wedge B) = P(A|B) * P(B)$

  $P(A \wedge B) = P(B|A) * P(A)$

- LHS are same. Equating the RHS of both equations yields

$$P(A|B) = P(B \mid A) * P(A) / P(B)$$

$$P(B|A) = P(A \mid B) * P(B) / P(A)$$

Bayes rule/
Bayesian rule

where:
p(A|B) is the conditional probability that event A occurs given that event B has occurred;
p(B|A) is the conditional probability of event B occurring given that event A has occurred;
p(A) is the probability of event A occurring;
p(B) is the probability of event B occurring.

Bayesian or Bayes Rule
(Hypothesis-Evidence)

Likelihood of Evidence given Hypothesis

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

Posterior Probability of Hypothesis given evidence

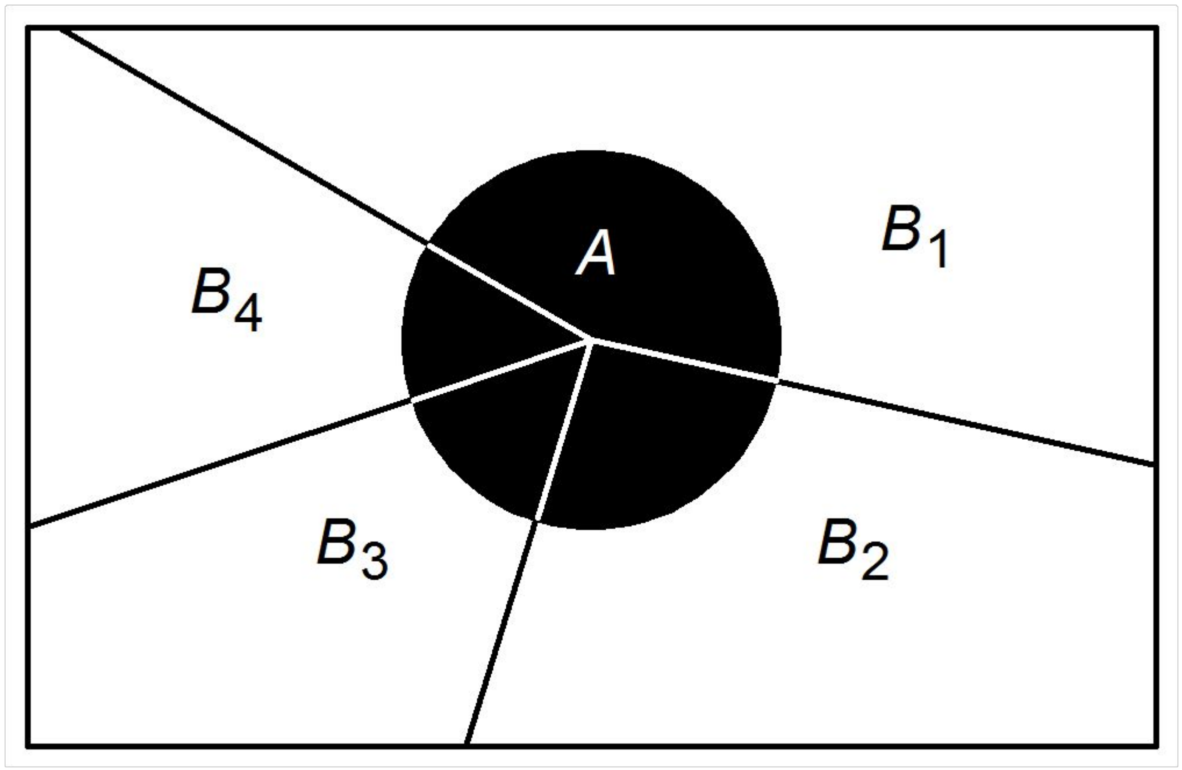
Normalizing Constant

Prior Probability of Hypothesis

Bayesian or Bayes Rule

The Joint Probability

$$\sum_{i=1}^{n} p(A \cap B_i) = \sum_{i=1}^{n} p(A|B_i) \times p(B_i)$$



Bayesian or Bayes Rule

- If the occurrence of event A depends on only two mutually exclusive events, B and NOT B , we obtain:

$$p(A) = p(A|B) \times p(B) + p(A|\neg B) \times p(\neg B)$$

where is the logical function NOT.

Similarly,

$$p(B) = p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)$$

Substituting this equation into the Bayesian rule yields:

$$\boxed{p(A|B) = \frac{p(B|A) \times p(A)}{p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)}}$$

Bayesian or Bayes Rule

- The Bayesian rule expressed in terms of hypotheses and evidence looks like this:

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E|H) \times p(H) + p(E|\neg H) \times p(\neg H)}$$

where:

p(H) is the prior probability of hypothesis H being true;

p(E|H) is the probability that hypothesis H being true will result in evidence E;

p(H) is the prior probability of hypothesis H being false;

p(E| H) is the probability of finding evidence E even when hypothesis H is false.

Example: Bayes Rule

$$P(\text{Cancer} \mid \text{Test+}) = \frac{P(\text{Test+} \mid \text{Cancer})P(\text{Cancer})}{P(\text{Test +})}$$

$P(\text{Test+} \mid \text{Cancer}) = 0.9$ $\quad P(\text{Test}- \mid \text{Cancer}) = 0.1$

$P(\text{Test+} \mid \text{No Cancer}) = 0.01$ $\quad P(\text{Test}- \mid \text{Cancer}) = 0.99$

$P(\text{Cancer}) = 0.0001$

$$\begin{aligned} P(\text{Test+}) &= P(\text{Test+} \mid \text{Cancer})P(\text{Cancer}) + \\ &\quad P(\text{Test+} \mid \text{No cancer})P(\text{No Cancer}) \\ &= 0.9\text{x}0.0001 + 0.01\text{x}(1-0.0001) \\ &= 0.0010899 \end{aligned}$$

Example: Bayes Rule

$$P(\text{Cancer} \mid \text{Test+}) = \frac{P(\text{Test+} \mid \text{Cancer})P(\text{Cancer})}{P(\text{Test +})}$$

$P(\text{Test+} \mid \text{Cancer}) = 0.9$

$P(\text{Test+} \mid \text{No Cancer}) = 0.01$

$P(\text{Cancer}) = 0.0001$

$P(\text{Test+}) = 0.0010899$

$P(\text{Cancer} \mid \text{Test+}) = 0.9 \text{x} 0.0001/0.0010899$

$= 0.08$

Example: Bayes' rule

- Disease Meningitis:
- It Cause patient to have stiff neck- 50% of the time.
- Prior Probability that Patients has meningitis is 1/50000.
- Prior Probability that patient has stiff neck is 1/20.


Severe headache


Stiff neck


Dislike of bright lights


Fever/vomiting


Drowsy and less responsive/ vacant


Rash (develops anywhere on body)

- Let s be stiff neck & m be Meningitis.

$$P(s|m) = 0.5$$

$$P(m) = 1/50000$$

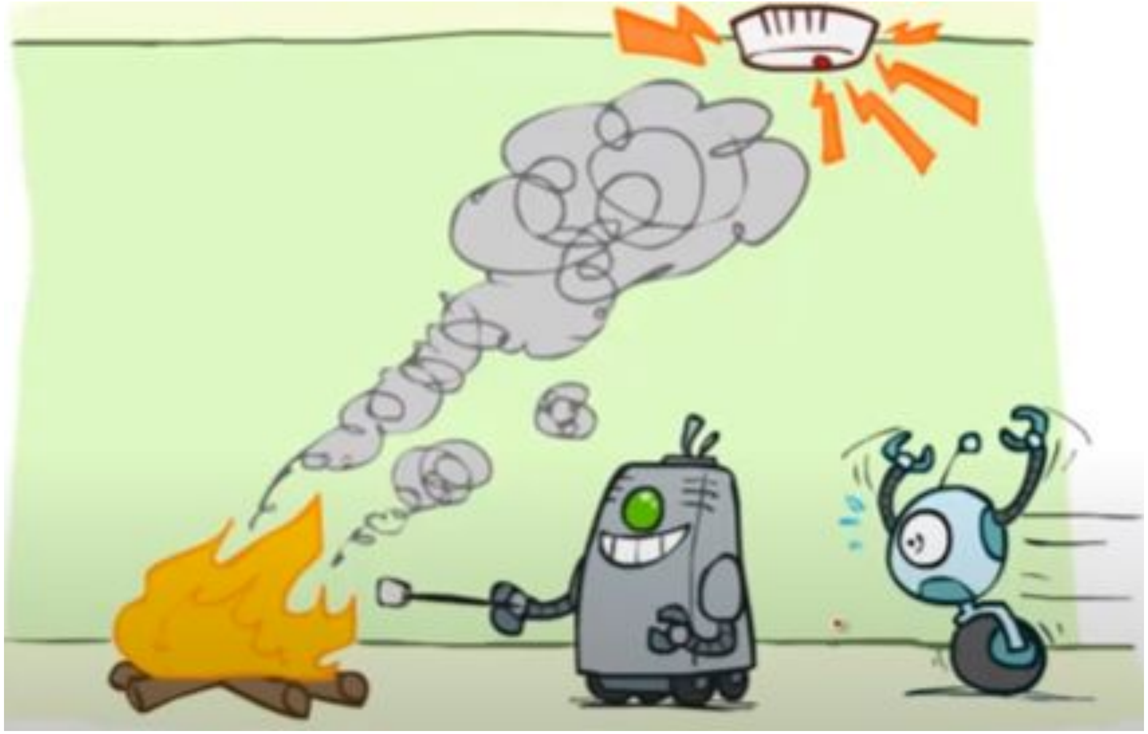
$$P(s)= 1/20$$

$$P(m|s) = \frac{P(s|m)\ P(m)}{P(s)}$$

$$= \frac{0.5 * 1/50000}{1/20} = 0.0002.$$

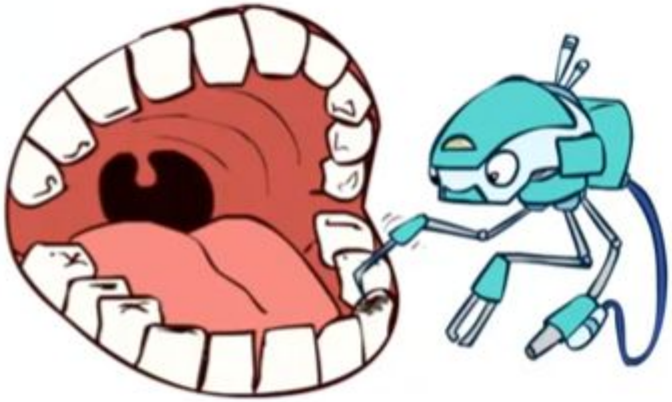
- 1 in 5000 patients with stiff neck to have Meningitis.

Conditional Independence



Conditional Independence

- P(Toothache, Cavity, Catch)
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)
- The same independence holds if I don't have a cavity:
  - P(+catch | +toothache, -cavity) = P(+catch| -cavity)
- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)
- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)
  - One can be derived from the other easily



Conditional Independence

- Unconditional (absolute) independence very rare (why?)
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.
- X is conditionally independent of Y given Z $X \perp\!\!\!\perp Y | Z$

  if and only if:

  $$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

  or, equivalently, if and only if

  $$\forall x, y, z : P(x|z, y) = P(x|z)$$

Conditional Independence

$\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ independent entries (or: parameters)

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

(1) $P(catch|toothache, cavity) = P(catch|cavity)$

The same independence holds if I haven't got a cavity:

(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

So $Catch$ is conditionally independent of $Toothache$ given $Cavity$:

$\mathbf{P}(Catch|Toothache, Cavity) = \mathbf{P}(Catch|Cavity)$

Equivalent statements:

$\mathbf{P}(Toothache|Catch, Cavity) = \mathbf{P}(Toothache|Cavity)$

$\mathbf{P}(Toothache, Catch|Cavity) = \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)$

Note that (conditional) independence is symmetric!

Conditional Independence

Write out full joint distribution using chain rule:

$$\begin{aligned}
&\mathbf{P}(Toothache, Catch, Cavity) \\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch, Cavity) \\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity) \\
&= \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}$$

I.e., 2 + 2 + 1 = 5 independent numbers (equations 1 and 2 remove 2)

In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in n to linear in n.

Conditional independence is our most basic and robust form of knowledge about uncertain environments.

But how then, do we compute e.g. $\mathbf{P}(Cavity|Toothache)$?

Conditional Independence

- What about this domain:
  - Traffic
  - Umbrella
  - Raining

Conditional Independence

- What about this domain:
  - Fire
  - Smoke
  - Alarm

Conditional Independence & Chain Rule

- Chain rule: $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

- Trivial decomposition:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

- With assumption of conditional independence:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

- Bayes'nets / graphical models help us express conditional independence assumptions

Probabilistic reasoning- Bayesian network

- **Bayesian network is a systematic way to represent independence relationships explicitly.**
- Data structure to represent the dependencies among variables.
- **Directed graph** – each node is annotated with quantitative probability information.

Specification of Bayesian network

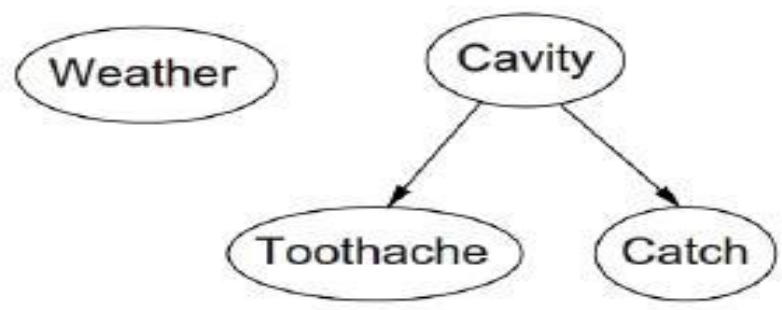
1. A set of random variables makes up the node.
2. A set of directed links or arrows connects pair of nodes.

X is the Parent of Y.

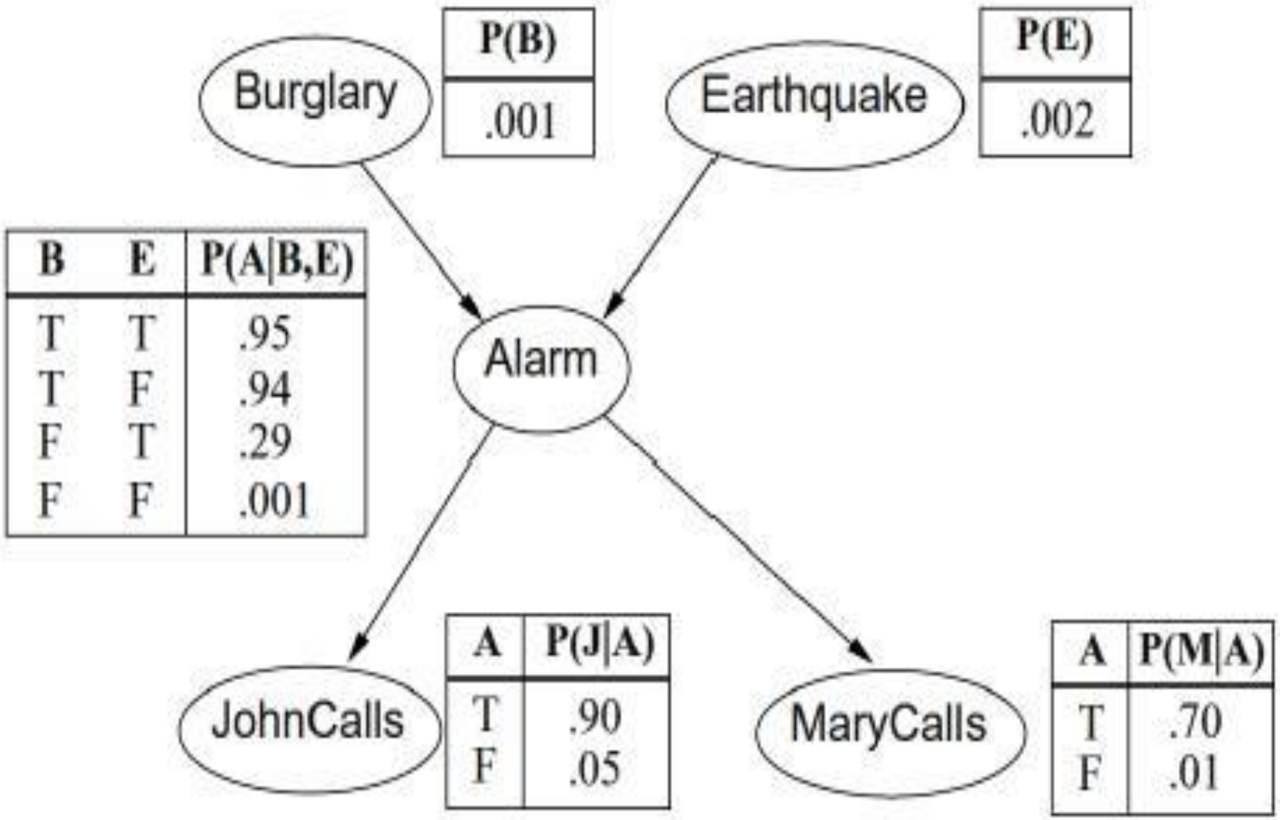
3. Each node X_i has a conditional probability distribution $P(X_i|Parent(X_i))$
4. No directed cycles.

Topology

- Nodes & Links – specifies the conditional independence relationships.
- Variables *Weather* ,*Toothache* ,*Catch* ,*Cavity* .



Burglar Alarm Example



Burglary

Earth quake

Burglary

P(B)
. 001

Earth Quake

P(E)
.002

John calls

A	P(J)
T	.90
F	.05

Alarm

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

Mary Calls

A	P(M)
T	.70
F	.01

Semantics of Bayesian network

- 2 ways to understand:

1. To see the network as a representation of the joint probability distribution.
2. View as an encoding of a collection of conditional independence statements.

Full joint distribution

- Joint distribution is the probability of a conjunction of assignment to each variables.

$$P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n).$$

Example

- Calculate probability that alarm sounds but neither a burglary nor an earth quake has occurred and both John & Mary call.

$$
\begin{aligned}
&P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) \\
&\quad = P(j|a)P(m|a)P(a|\neg b \wedge \neg e)P(\neg b)P(\neg e) \\
&\quad = 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.00062 .
\end{aligned}
$$

Node ordering

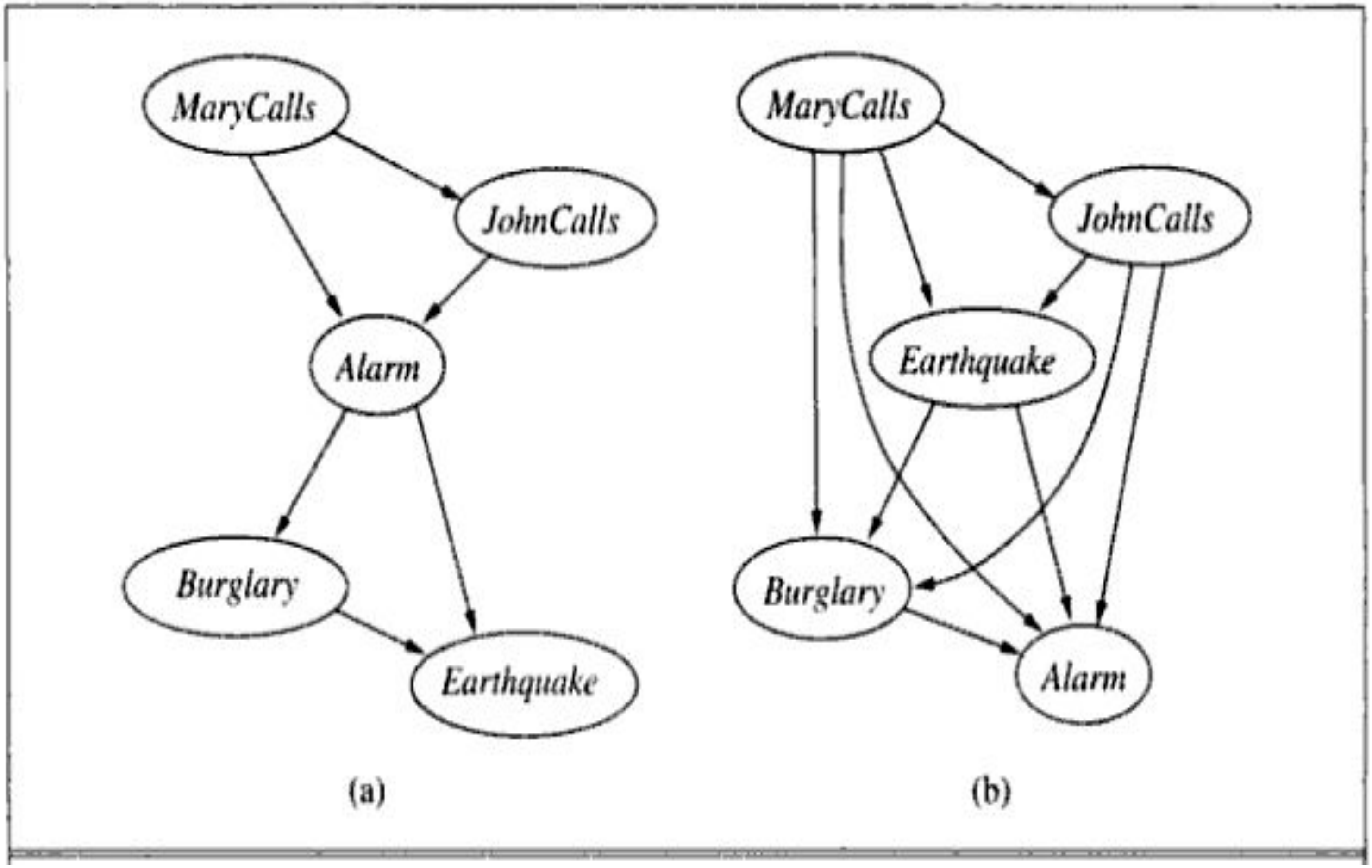
- Adding *MaryCalls*: No parents.
- Adding *JohnCalls*: If Mary calls, that probably means the alarm has gone off, which of course would make it more likely that John calls. Therefore, *JohnCalls* needs *MaryCalls* as a parent
- Adding *Alarm*: Clearly, if both call, it is more likely that the alarm has gone off than if just one or neither call, so we need both *MaryCalls* and *JohnCalls* as parents.
- Adding *Burglary*: If we know the alarm state, then the call from John or Mary might give us information about our phone ringing or Mary's music, but not about burglary:

  $$\mathbf{P}(Burglary \mid Alarm, JohnCalls, MaryCalls) = \mathbf{P}(Burglary \mid Alarm)\,.$$

  Hence we need just *Alarm* as parent.

- Adding *Earthquake*: if the alarm is on, it is more likely that there has been an earthquake. (The alarm is an earthquake detector of sorts.) But if we know that there has been a burglary, then that explains the alarm, and the probability of an earthquake would be only slightly above normal. Hence, we need both *Alarm* and *Burglary* as parents.

Node ordering



Inference in Bayesian Network

- Exact Inference.
- Approximate Inference.
- Exact Inference:

  - Compute posterior Probability for a set of query, given observed event.

posterior probabilities $P(X|e)$

Eg: Burglary:

Observed event- John calls=true & Mary calls=true

$P(\text{Burglary}|\text{Johncalls}=\text{true},\text{Marycalls}=\text{true})$

$= \langle 0.284, 0.716 \rangle$

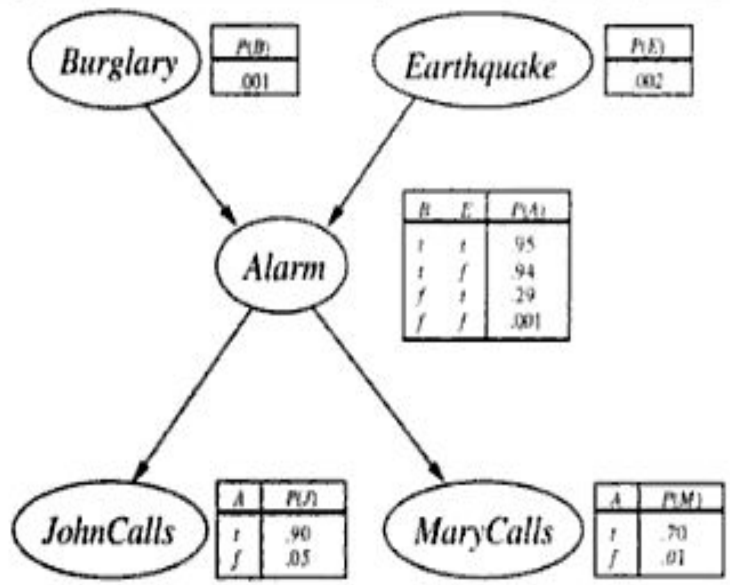


Figure 14.2 A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters B, E, A, J, and M stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

Inference by Enumeration

- Conditional probability By Full-joint diatribution. Query P(X|e) is

$$\mathbf{P}(X|\mathbf{e}) = \alpha\,\mathbf{P}(X,\mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X,\mathbf{e},\mathbf{y})\,.$$

Eg:

P(Burglary|Johncalls=true,Marycalls=true).

Hidden variable are Alarm & Earth quake.

$$\mathbf{P}(B|j,m) = \alpha\,\mathbf{P}(B,j,m) = \alpha \sum_{e} \sum_{a} \mathbf{P}(B,e,a,j,m)\,.$$

- Applying Semantics of Bayseian network:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i)) .$$

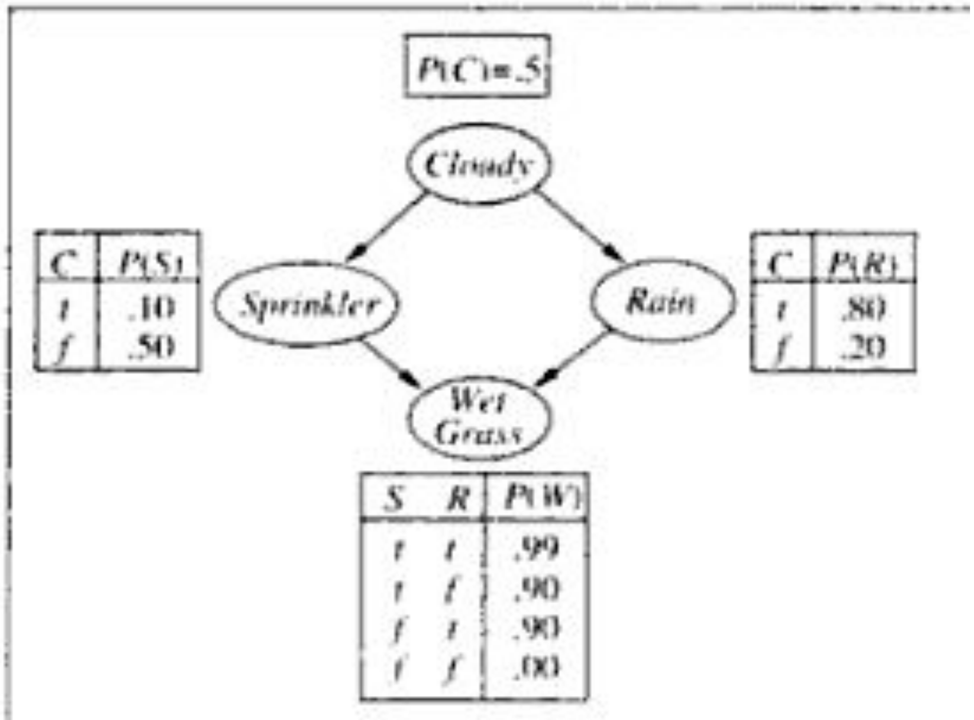
- P(b) - $P(b|j,m) = \alpha \sum_e \sum_a P(b)P(e)P(a|b,e)P(j|a)P(m|a)$. ed outw

- By stuc

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

$$\mathbf{P}(B|j,m) = \alpha \langle 0.00059224, 0.0014919 \rangle \approx \langle 0.284, 0.716 \rangle .$$

Multiply connected network



(a)

Approximate inference

- Exact inference is not applicable in multiply connected network.
- Monte Carlo algorithm is used to provide approximate answers.(Samples).
- 2 ways to calculate:

1. Direct Sampling method.
2. Markov chain simulation.

Direct Sampling

- Generation of samples from a known probability distribution.
- Eg: **Assuming an ordering**

  [Cloudy, Sprinkler, Rain, Wet Grass]

1. Sample from $\mathbf{P}(Cloudy) = \langle 0.5, 0.5 \rangle$; suppose this returns *true*.
2. Sample from $\mathbf{P}(Sprinkler \mid Cloudy = true) = \langle 0.1, 0.9 \rangle$; suppose this returns *false*.
3. Sample from $\mathbf{P}(Rain \mid Cloudy = true) = \langle 0.8, 0.2 \rangle$; suppose this returns *true*.
4. Sample from $\mathbf{P}(WetGrass \mid Sprinkler = false, Rain = true) = \langle 0.9, 0.1 \rangle$; suppose this returns *true*.

In this case, PRIOR-SAMPLE returns the event [*true*, *false*, *true*, *true*].

Rejection sampling

- Conditional probability P(X|e).
- Estimate P(*Rain* |*Sprinkler =true*) using 100 samples.
- 73 have *sprinkler = False* are rejected,27 have *sprinkler= True.*
- Of the 27 , 8 have *Rain = true* & 19 have *Rain=False*.

$$\mathbf{P}(Rain|Sprinkler = true) \approx \textsc{Normalize}(\langle 8, 19 \rangle) = \langle 0.296, 0.704 \rangle .$$

Probabilistic Reasoning

- Till now Static world ,Dynamic aspects of the problem.
- **State & Observation**:

  X_{t} - Set of unobserved state variable.

  E_{t} - Set of observed evidence.

  e_{t} - Set of Values.
- Eg: Umbrella problem

  t – set of start state

  $R_{0}, R_{1}, R_{2..}$ Set of State Variable.

  $U_{1}, U_{2...}$ Evidence Variable.

Stationary Process & Markov Assumption.

- Set of variable- unbounded , state & evidence changes over time.
- 2 problems:

  1. unbounded num of conditional probability table.(each variable)

  2. Unbounded num of parents.

Solutions

- Stationary Process- Changes in the world –caused by a stationary process.

  Eg: $P(U_t | Parents(U_t))$ – same for all t
- Markov assumption – Handling the infinite number of parents. Current state depends on finite history of previous states.
- Markov process or chain:
  - First order Markov Process
  - Second order Markov Process.

- First order Markov Process:

  - Current state depends on the previous state & not on early states.

    $P(X_t \mid X_{0:t-1}) = P(X_t \mid X_{t-1})$ – Transition Model

- Second order Markov Process:

  - Depends on 2 previous states.

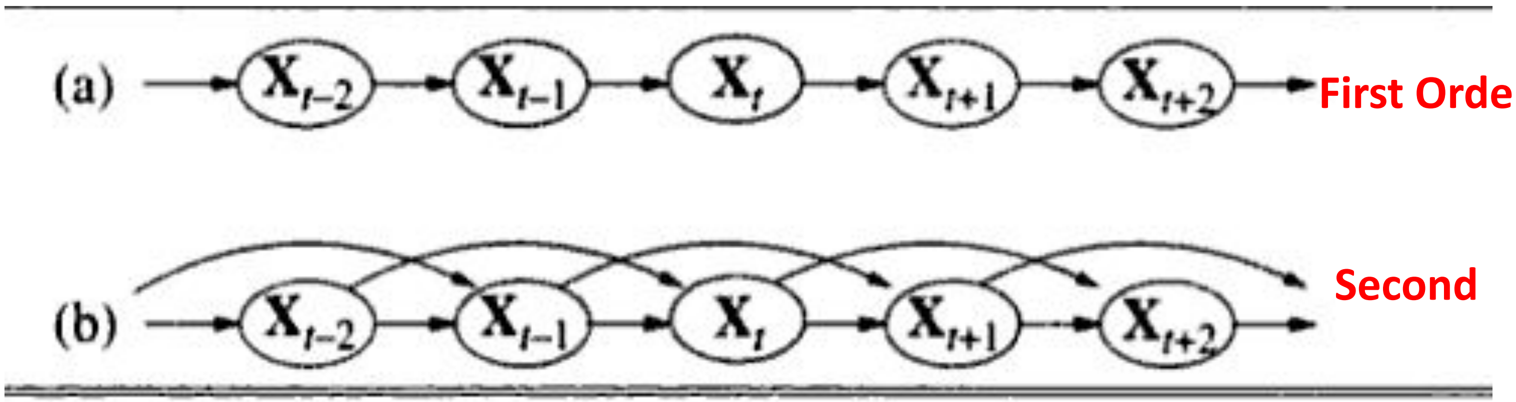
    $P(X_t \mid X_{t-2}, X_{t-1})$

- Restrict the parent of the evidence variable E_t.

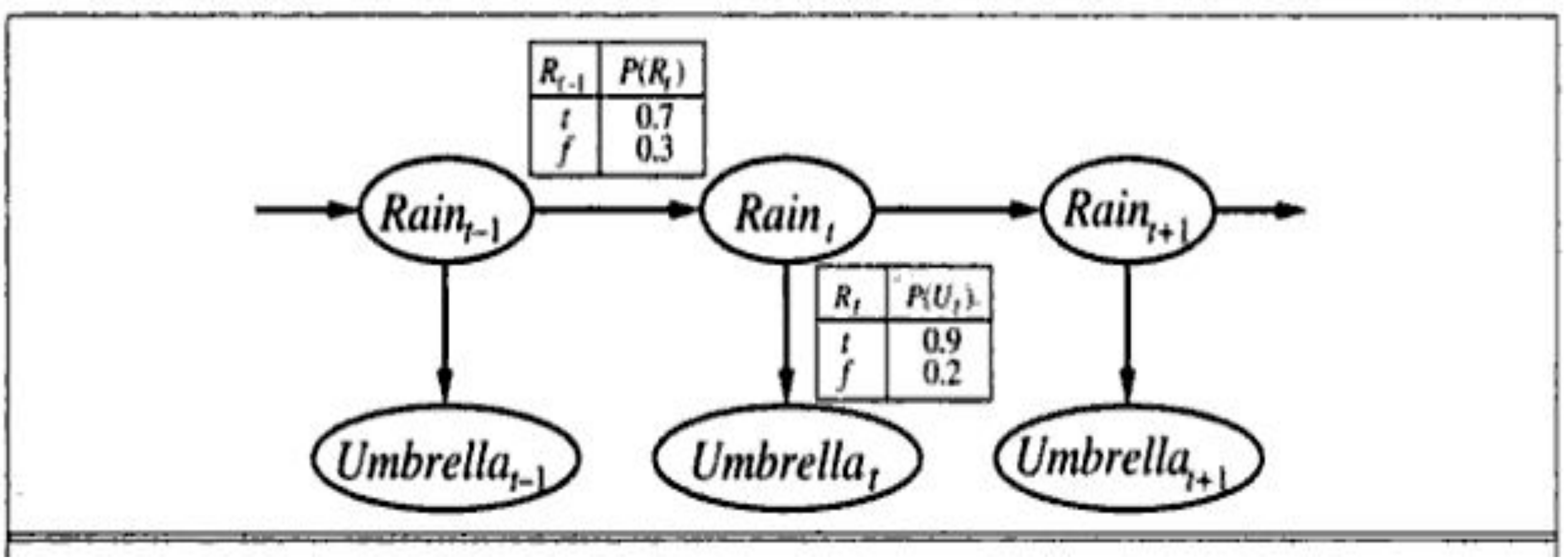
  $P(E_t \mid X_{0:t}, E_{0:t-1}) = P(E_t \mid X_t)$

  - Sensor Model.

Example



(a)
X_{t-2}
X_{t-1}
X_t
X_{t+1}
X_{t+2}
First Order
(b)
X_{t-2}
X_{t-1}
X_t
X_{t+1}
X_{t+2}
Second



R_{t-1} | $P(R_t)$
t | 0.7
f | 0.3
$Rain_{t-1}$
$Rain_t$
$Rain_{t+1}$
R_t | $P(U_t)$
t | 0.9
f | 0.2
$Umbrella_{t-1}$
$Umbrella_t$
$Umbrella_{t+1}$

Approximate in predicting

- 2 solutions:

1. Increasing the order of the Markov process model. For example, we could make a second-order model by adding $Rain_{t-2}$ as a parent of $Rain_t$, which might give slightly more accurate predictions (for example, in Palo Alto it very rarely rains more than two days in a row).

2. Increasing the set of state variables. For example, we could add $Season_t$ to allow us to incorporate historical records of rainy seasons, or we could add $Temperature_t$, $Humidity_t$ and $Pressure_t$ to allow us to use a physical model of rainy conditions.

Temporal Model

- HTM **Hierarchical temporal memory** is a biomimetic model based on the **memory-prediction theory** of brain function described by Jeff Hawkins in his book *On Intelligence*.

Inference in Temporal Model

- Basic inference tasks.

 Filling & Monitoring: (Observation of previous states).

  Prediction: (Future state).

  Smoothing or Hindsight: (Past state - Observation)

  Most Likely explanation: Sequence of states-generated through Observation.

Rain [True ,True ,False, True]``

- **Filtering:**

$$P(X_{t+1} \mid e_{1:t+1}) = f\, P(e_{t+1,} P(X_t \mid e_{1:t})\,)$$

- **Prediction:**

$$P(X_{t+k+1} \mid e_{1:t}) = \sum P(X_{t+k+1} \mid X_{t+k})\, P(X_{t+k} \mid e_{1:t})$$

- **Smoothing:**

$$P(X_k \mid e_{1:t}) \quad 1 \leq k \leq t$$

$$P(X_k \mid e_{1:k})\, P(X_{k+1:t} \mid X_k)$$

$$= f_{1:k}\ b_{k+1:t}$$

- Smoothing:

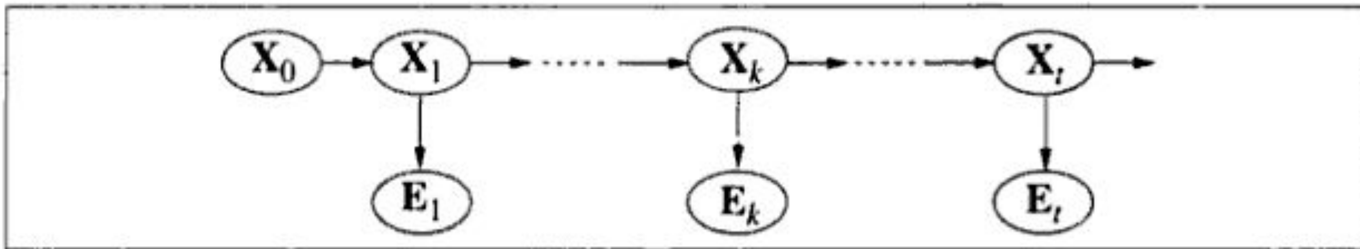
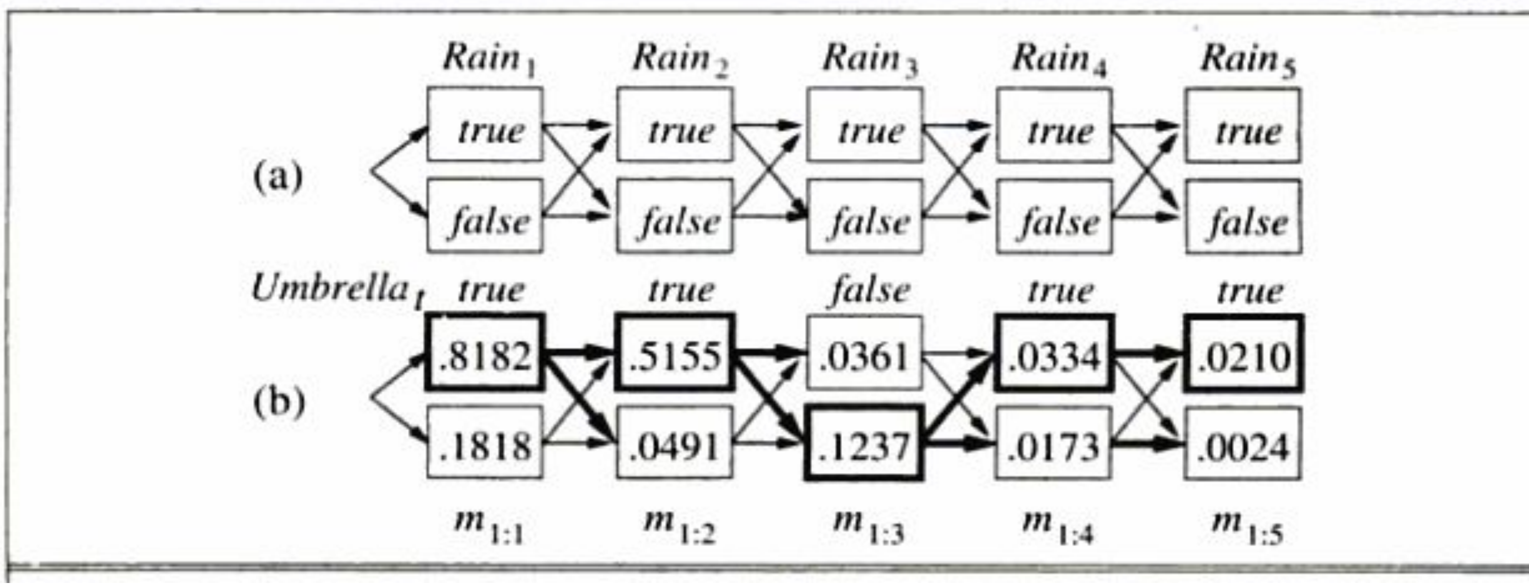


Figure 15.3 Smoothing computes $\mathbf{P}(\mathbf{X}_k|\mathbf{e}_{1:t})$, the posterior distribution of the state at some past time k given a complete sequence of observations from 1 to t.

- Finding the most likely sequence:



Hidden Markov Models

- State of the process :- Single discrete random variable.
- Simplified matrix algorithm:

$$\mathbf{T}_{ij} = P(X_t = j \mid X_{t-1} = i) \; .$$

State *i* to state *j.*

- *Eg: **Umbrella world**: (**Transition Model**)*

$$\mathbf{T} = \mathbf{P}(X_t \mid X_{t-1}) = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}$$

- **Sensor Model**:

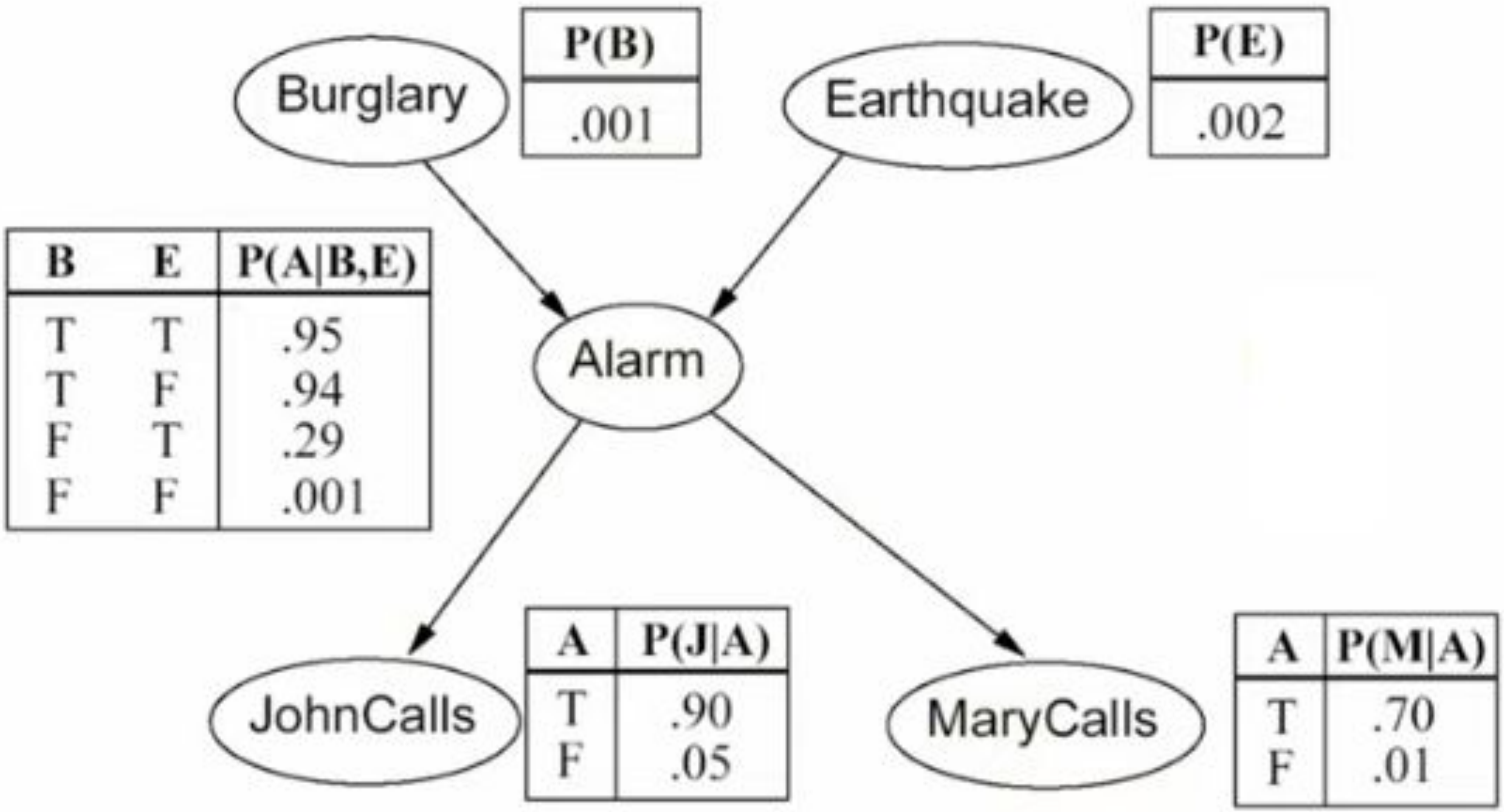
Diagonal Matrix O_t .

Eg:

Umbrella world, U_1 = true.

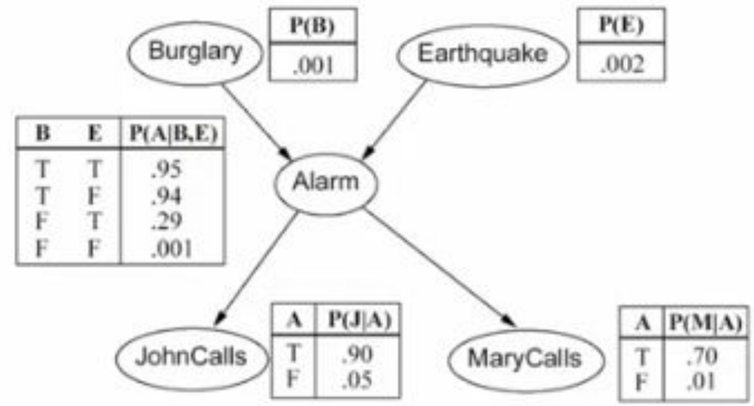
$$\begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \end{bmatrix}$$

- You have a new burglar alarm installed at home.
- It is fairly reliable at detecting burglary, but also sometimes responds to minor earthquakes.
- You have two neighbors, John and Merry , who promised to call you at work whe they hear the alarm.
- John always calls when he hears the alarm, but sometimes confuses telephone ringing with the alarm and calls too.
- Merry likes loud music and sometimes misses the alarm.
- Given the evidence of who has or has not called, we would like to estimate the probability of a burglary.


Burglary
P(B)
.001
Earthquake
P(E)
.002
B E P(A|B,E)
T T .95
T F .94
F T .29
F F .001
Alarm
JohnCalls
A P(J|A)
T .90
F .05
MaryCalls
A P(M|A)
T .70
F .01

BAYESIAN BELIEF NETWORKS – EXAMPLE – 1

1. What is the probability that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both John and Merry call?



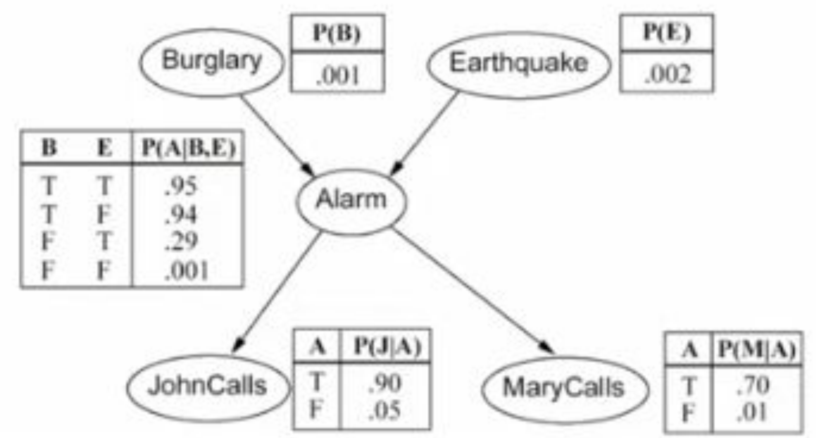
Solution:

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

$$= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998$$

$$= 0.00062$$

2. What is the probability that John call?



Solution:

$P(j) = P(j \mid a)\,P(a) + P(j \mid \neg a)\,P(\neg a)$

$= P(j|a)\{P(a|b,e)*P(b,e)+P(a|\neg b,e)*P(\neg b,e)+P(a|b,\neg e)*P(b,\neg e)+P(a|\neg b,\neg e)*P(\neg b,\neg e)\}$

$+ P(j|\neg a)\{P(\neg a|b,e)*P(b,e)+P(\neg a|\neg b,e)*P(\neg b,e)+P(\neg a|b,\neg e)*P(b,\neg e)+P(\neg a|\neg b,\neg e)*$

$P(\neg b,\neg e)\}$

$= 0.90 * 0.00252 + 0.05 * 0.9974 = 0.0521$