Reasoning

- **Reasoning** is the act of deriving a conclusion from certain premises using a given methodology.
  - Any knowledge system must reason, if it is required to do something which has not been told explicitly .
  - For reasoning, the system must find out what it needs to know from what it already knows.
  - **Example :**

    If we know : *Robins are birds.*

    All birds have wings

    Then if we ask: *Do robins have wings?*

    To answer this question - some reasoning must go.

Uncertainity in reasoning

- In Expert Systems, we must often attempt to draw correct conclusions from poorly formed and uncertain evidence using unsound inference rules.

- This is not an impossible task; we do it successfully in almost every aspect of our daily survival.

Define…

- Uncertainty can be defined as the lack of the exact knowledge that would enable us to reach a perfectly reliable conclusion.
  - This is because information available to us can be in its imperfect, such as inconsistent, incomplete, or unsure, or all three.
- An example: unknown data or imprecise language

Uncertainty Handling Methods

- Abductive reasoning
- Property inheritance
- Fuzzy logic
- Certainty Factor (CF)

E.g. Suppose:
If x is a bird then x flies

Abductive reasoning would say that
“All fly things are birds”

By property inheritance
“All birds can fly”
but, remember the case that
Penguin cannot fly?

Statistical reasoning:

- Bayes theorem
- Dempster-Shafer theory

Probabilistic Reasoning

- Probabilistic reasoning stands as a potent method of handling uncertainty in the realm of Artificial Intelligence (AI).

- **Effective implementation of probabilistic reasoning in AI necessitates a thoughtful and sound strategy:**

1. Defining the Problem: Identify and clearly define the problem to be solved.
2. Data Collection: Assemble high-quality, relevant data as the backbone of the probabilistic reasoning model.
3. Model Selection: Depending upon the problem, select the appropriate probabilistic model such as Bayesian Networks, Markov Models, etc.
4. Training the Model: Input the collected data into the selected model and train it.
5. Testing and Refinement: Test the predictions and outcomes, refine the model accordingly and reassess it.
6. Robust Implementation: Factor in supporting the model to make decisions in real-time in production environments.

Bayesian Approach

- Bayesian approach (or Bayes theorem) is based on formal probability theory.
- It provides a way of computing the probability of a hypothesis (without sampling) following from a particular piece of evidence, given only the probabilities with which the evidence follows from actual cause.
- To use this approach, reliable statistical data that define the prior probabilities for each hypothesis must be available
  - As these requirements are rarely satisfied on real-world problem, so only a few systems have been built based on bayesian reasoning

- Bayes' Theorem is an important tool in understanding what we really know, given the evidence and other information we have. It helps incorporate "conditional probabilities" into our conclusions.

  Prior probability: The prior probability of an event is probability computed before observing new information.

- 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 is a probability of occurring an event when another event has already happened.

- Bayes' Theorem *tells us quantitatively how to update our prior information, given new evidence*.

Bayesian Inference: The explanation with the highest posterior probability

1. Definition of conditional probability

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

2. Bayes' Theorem

$$\Pr(H \mid D) = \frac{\Pr(H)\Pr(D \mid H)}{\Pr(D)}$$

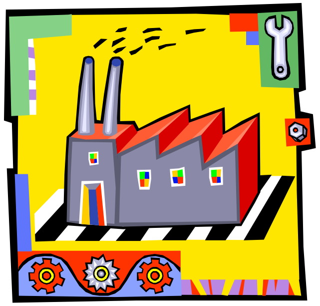
Prior probability, the probability of the hypothesis on previous knowledge

Likelihood function, probability of the data given the hypothesis

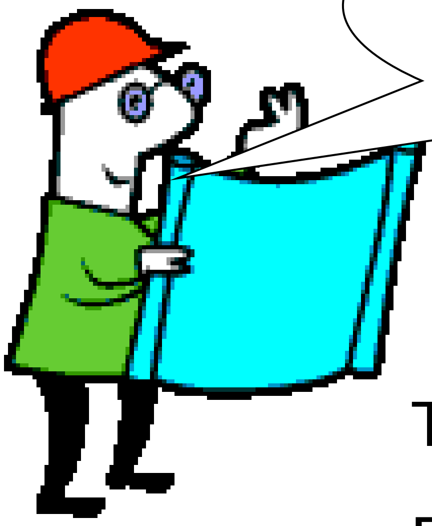
Posterior probability, the probability of the hypothesis given the data

Unconditional probability of the data, a normalizing constant ensuring the posterior probabilities sum to 1.00

Application of Bayes' Theorem



- Consider a manufacturing firm that receives shipment of parts from two suppliers.
- Let A_1 denote the event that a part is received from supplier 1;
- A_2 is the event the part is received from supplier 2



Thus:

$P(A_1) = .65$ and $P(A_2) = .35$

Quality levels differ between suppliers

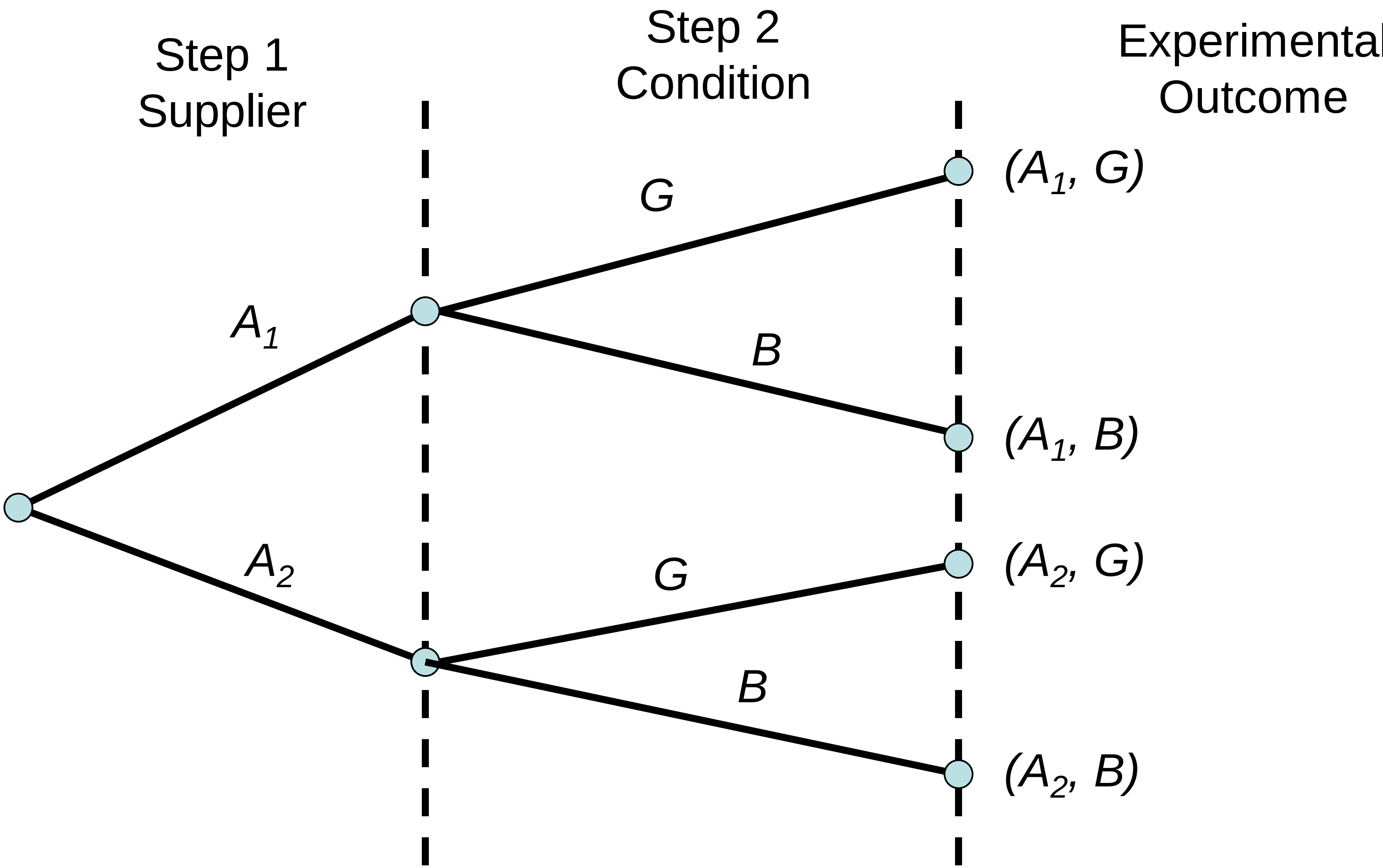
	Percentage Good Parts	Percentage Bad Parts
Supplier 1	98	2
Supplier 2	95	5

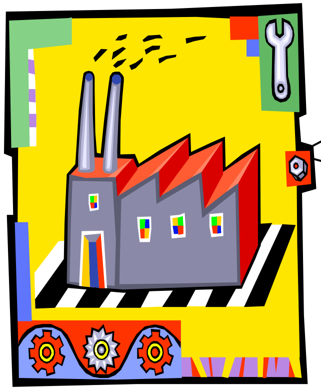
Let G denote that a part is good and B denote the event that a part is bad. Thus we have the following conditional probabilities:

$$P(G \mid A_1) = .98 \quad \text{and} \quad P(B \mid A_1) = .02$$

$$P(G \mid A_2) = .95 \quad \text{and} \quad P(B \mid A_2) = .05$$

Tree Diagram for Two-Supplier Example

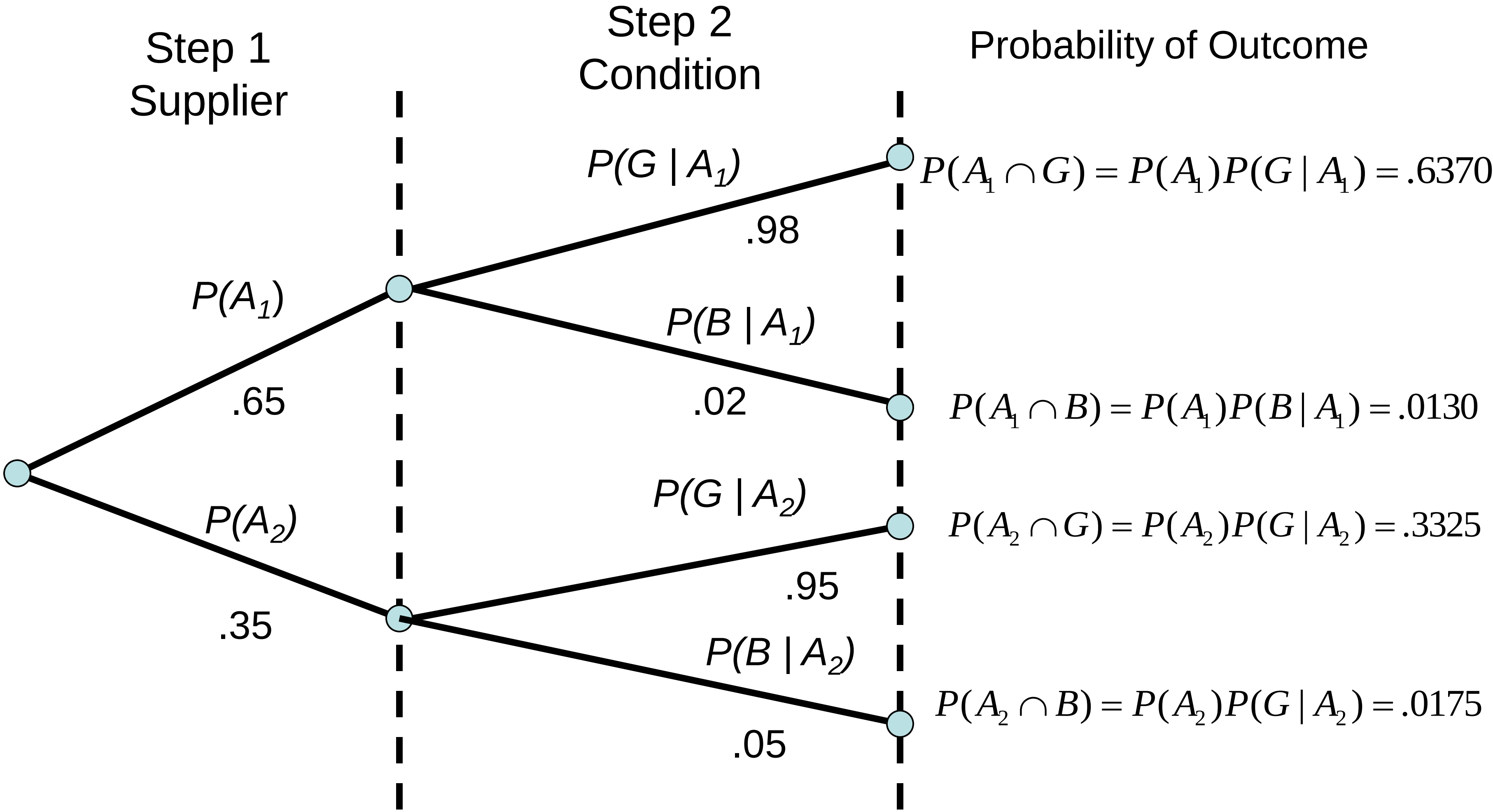


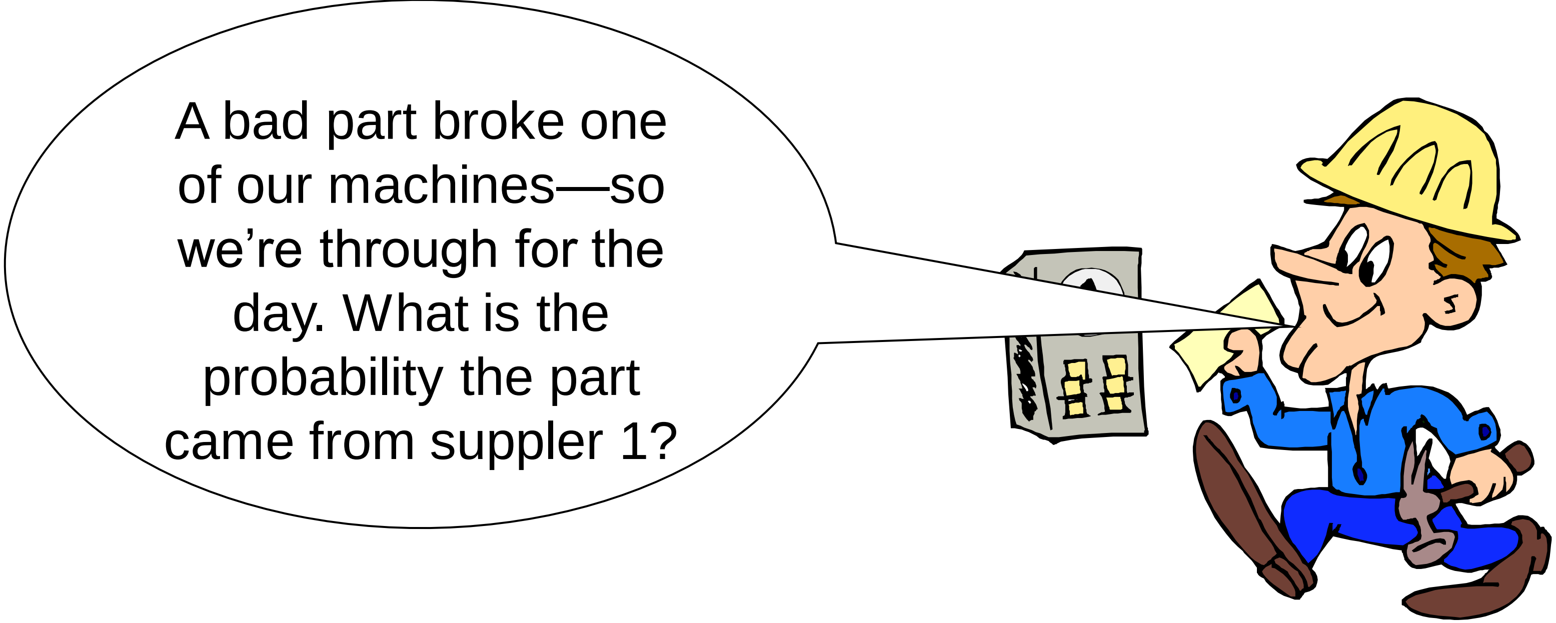


Each of the experimental outcomes is the intersection of 2 events. For example, the probability of selecting a part from supplier 1 that is good is given by:

$$P(A_1, G) = P(A_1 \cap G) = P(A_1)P(G \mid A_1)$$

Probability Tree for Two-Supplier Example



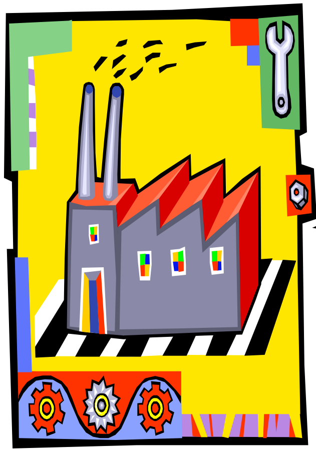


We know from the law of conditional probability that:

$$P(A_1 \mid B) = \frac{P(A_1 \cap B)}{P(B)} \qquad (1)$$

Observe from the probability tree that:

$$P(A_1 \cap B) = P(A_1)P(B \mid A_1) \qquad (2)$$



The probability of selecting a bad part is found by adding together the probability of selecting a bad part from supplier 1 and the probability of selecting bad part from supplier 2.

That is:

$$
\begin{aligned}
P(B) &= P(A_1 \cap B) + P(A_2 \cap B) \\
&= P(A_1)P(B \mid A_1) + P(A_2)P(B / A_2)
\end{aligned}
\qquad (3)
$$

Bayes' Theorem for 2 events

By substituting equations (2) and (3) into (1), and writing a similar result for $P(B \mid A_2)$, we obtain Bayes' theorem for the 2 event case:

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)}$$

$$P(A_2 \mid B) = \frac{P(A_2)P(B \mid A_2)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)}$$

Cont….

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)}$$

$$= \frac{(.65)(.02)}{(.65)(.02) + (.35)(.05)} = \frac{.0130}{.0305} = .4262$$

$$P(A_2 \mid B) = \frac{P(A_2)P(B \mid A_2)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)}$$

$$= \frac{(.35)(.05)}{(.65)(.02) + (.35)(.05)} = \frac{.0175}{.0305} = .5738$$

Bayes' Theorem

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \ldots + P(A_n)P(B \mid A_n)}$$

Tabular Approach to Bayes' Theorem—2-Supplier Problem

(1) Events A_i	(2) Prior Probabilities $P(A_i)$	(3) Conditional Probabilities $P(B \mid A_1)$	(4) Joint Probabilities $P(A_i \cap B)$	(5) Posterior Probabilities $P(A_i \mid B)$
A_1	.65	.02	.0130	.0130/.0305 =.4262
A_2	.35	.05	.0175	.0175/.0305 =.5738
	1.00		P(B)=.0305	**1.0000**

Bayesian Reasoning

ASSUMPTIONS

100 out of 10,000 women aged forty who participate in a routine screening have cancer

80 of every 100 women with cancer will get positive tests

950 out of 9,900 women without cancer will also get positive tests

PROBLEM

If 10,000 women in this age group undergo a routine screening, about what fraction of women with positive tests will actually have cancer?

Bayesian Reasoning

Before the screening:

100 women with cancer

9,900 women without cancer

After the screening:

A = 80 women with cancer and positive test

B = 20 women with cancer and negative test

C = 950 women without cancer and positive test

D = 8,950 women without cancer and negative test

Proportion of cancer patients with positive results, within the group of ALL patients with positive results:

$A/(A+C) = 80/(80+950) = 80/1030 = 0.078 = 7.8\%$

Compact Formulation

C = cancer present, T = positive test

p(A|B) = probability of A given B, ~ = not

PRIORS

PRIOR PROBABILITY

p(C) = 1%

CONDITIONAL PROBABILITIES

p(T|C) = 80%

p(T|~C) = 9.6%

POSTERIOR PROBABILITY (or REVISED PROBABILITY)

p(C|T) = ?

-----> Bayes' theorem

$$p(C|T) = \frac{\overbrace{p(T|C)*p(C)}^{A}}{\underbrace{P(T|C)*p(C)}_{A} + \underbrace{p(T|\sim C)*p(\sim C)}_{C}}$$

Bayesian Reasoning

Prior Probabilities:

100/10,000 = 1/100 = 1% = p(C)

9,900/10,000 = 99/100 = 99% = p(~C)

Conditional Probabilities:

A = 80/10,000 = (80/100)*(1/100) = p(T|C)*p(C) = 0.008

B = 20/10,000 = (20/100)*(1/100) = p(~T|C)*p(C) = 0.002

C = 950/10,000 = (950/9900)*(99/100) = p(T|~C)*p(~C) = 0.095

D = 8,950/10,000 = (8950/9900*(99/100) = p(~T|~C) *p(~C) = 0.895

Rate of cancer patients with positive results, within the group of ALL patients with positive results:

A/(A+C) = 0.008/(0.008+0.095) = 0.008/0.103 = 0.078 = 7.8%

Advantages:

Most significant is their sound theoretical foundation in probability theory.

Most mature uncertainty reasoning methods

Well defined semantics for decision making

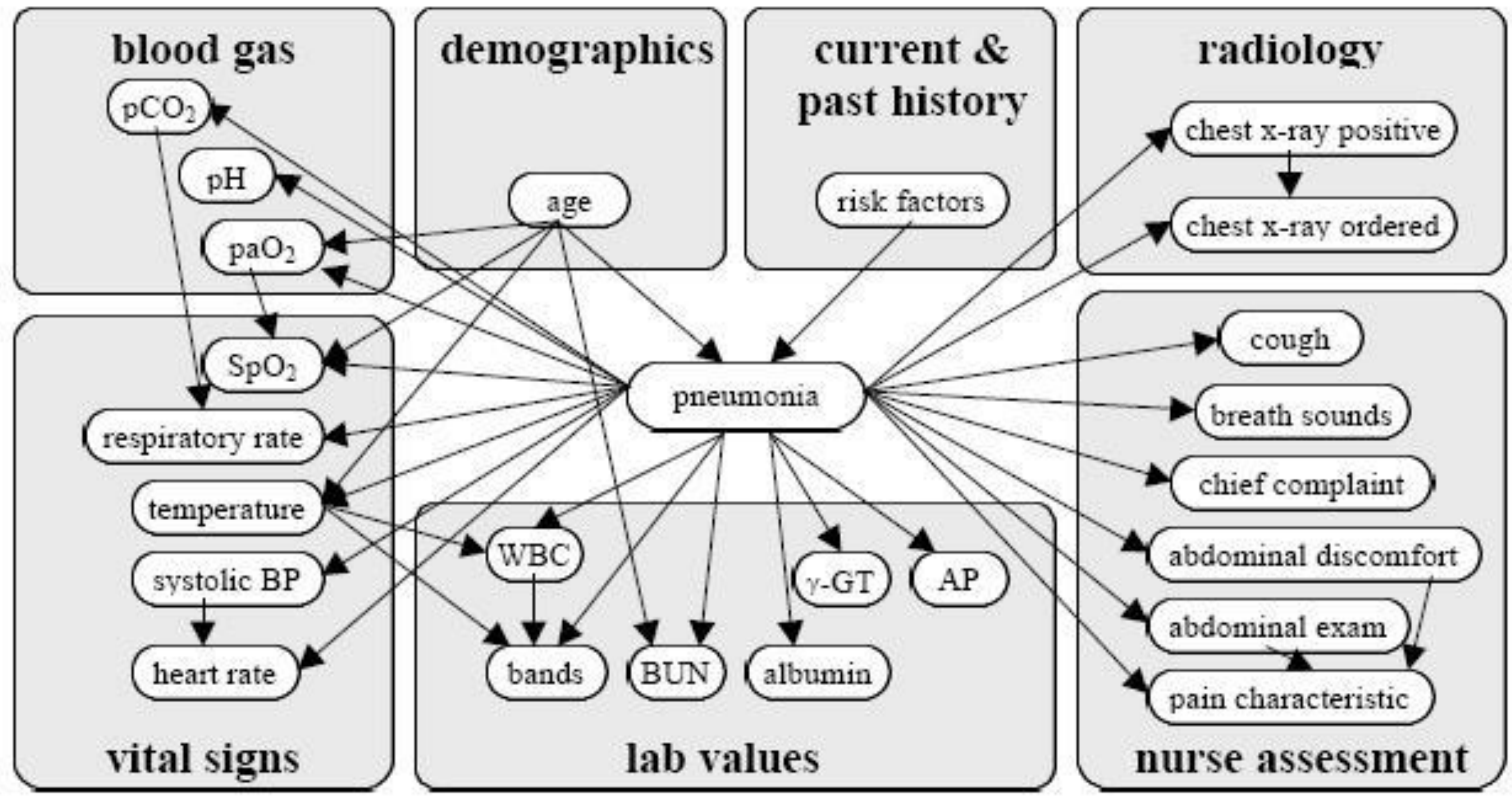
Main disadvantage:

They require a significant amount of probability data to construct a knowledge base.

Bayesian Networks

- Bayesian networks help us reason with uncertainty
- They are used in many applications eg.:
  - Spam filtering / Text mining
  - Speech recognition
  - Robotics
  - Diagnostic systems
  - Syndromic surveillance

Bayesian Networks (An Example)



From: Aronsky, D. and Haug, P.J., Diagnosing community-acquired pneumonia with a Bayesian network, In: *Proceedings of the Fall Symposium of the American Medical Informatics Association,* (1998) 632-636.

Bayesian networks

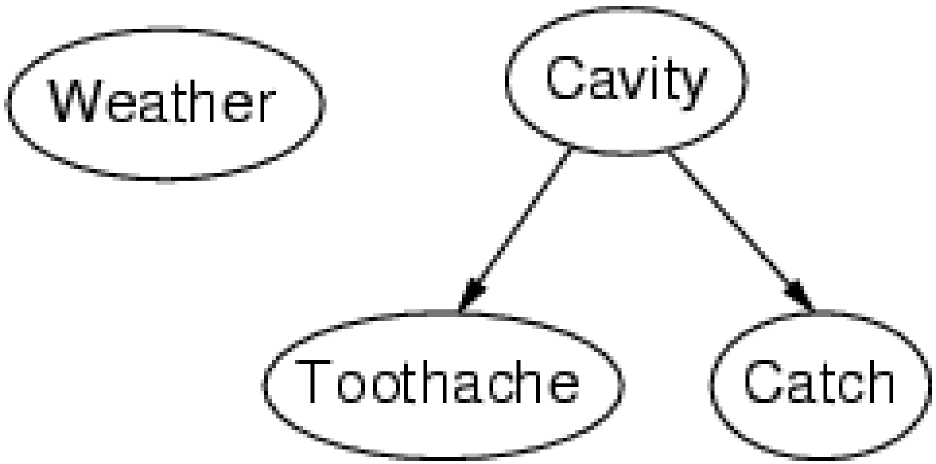
- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions
- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link ≈ "directly influences")
  - a conditional distribution for each node given its parents:

$$\mathbf{P}\,(X_i \mid Parents\,(X_i))$$

- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values
- A node is independent of its nondescendents given its parents.

Example

- Topology of network encodes conditional independence assertions:

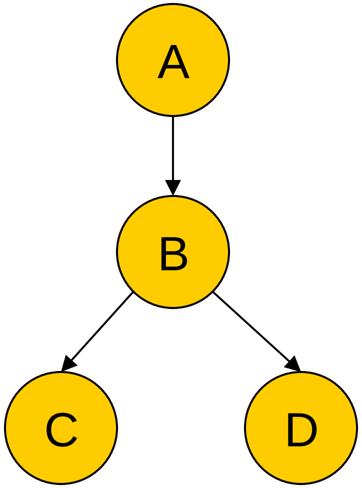


- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent given *Cavity*

A Bayesian Network

A Bayesian network is made up of:

1. A Directed Acyclic Graph



2. A set of tables for each node in the graph

A	P(A)
false	0.6
true	0.4

A	B	P(B|A)
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

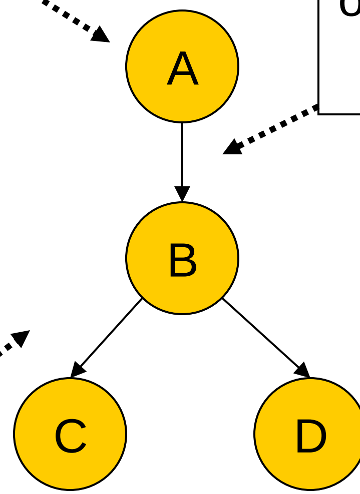
B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

A Directed Acyclic Graph

Each node in the graph is a random variable

A node *X* is a parent of another node *Y* if there is an arrow from node *X* to node *Y* eg. *A* is a parent of *B*



Informally, an arrow from node *X* to node *Y* means *X* has a direct influence on *Y*

A Set of Tables for Each Node

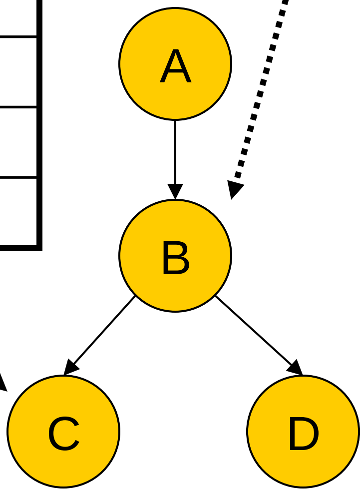
A	P(A)
false	0.6
true	0.4

A	B	P(B|A)
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

Each node X_i has a conditional probability distribution P(X_i | Parents(X_i)) that quantifies the effect of the parents on the node

The parameters are the probabilities in these conditional probability tables (CPTs)

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1



B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

A Set of Tables for Each Node

Conditional Probability Distribution
for C given B

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

For a given combination of values of the parents (B in this example), the entries for P(C=true | B) and P(C=false | B) must add up to 1

eg. P(C=true | B=false) + P(C=false |B=false)=1

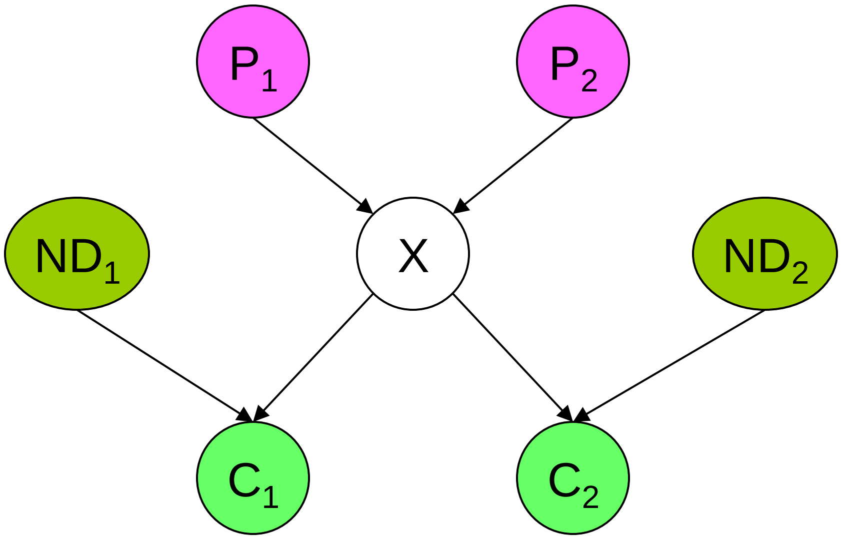
Bayesian Networks

Two important properties:

1. Encodes the conditional independence relationships between the variables in the graph structure
2. Is a compact representation of the joint probability distribution over the variables

Conditional Independence

The Markov condition: given its parents (P_1, P_2), a node (X) is conditionally independent of its non-descendants (ND_1, ND_2)



The Joint Probability Distribution

Due to the Markov condition, we can compute the joint probability distribution over all the variables $X_1, \ldots, X_n$ in the Bayesian net using the formula:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i \mid Parents(X_i))$$

Where Parents(X_i) means the values of the Parents of the node X_i with respect to the graph

Using a Bayesian Network Example

Using the network in the example, suppose you want to calculate:

P(A = true, B = true, C = true, D = true)

= P(A = true) * P(B = true | A = true) *

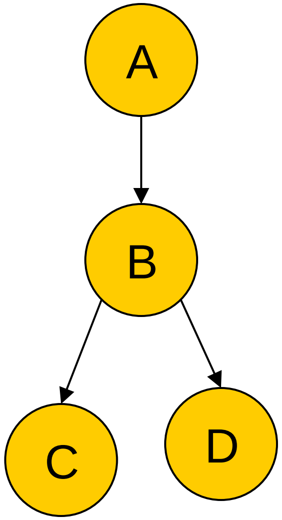
P(C = true | B = true) P(D = true | B = true)

= (0.4)*(0.3)*(0.1)*(0.95)

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

A	B	P(B|A)
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95



Using a Bayesian Network Example

Using the network in the example, suppose you want to calculate:

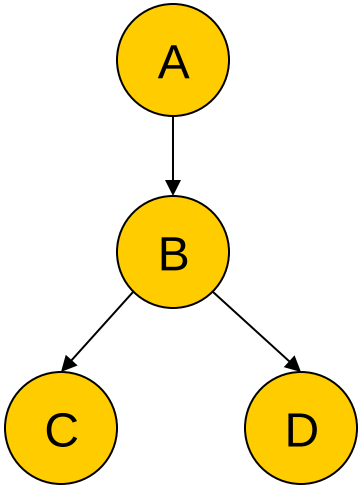
$$P(A = \text{true}, B = \text{true}, C = \text{true}, D = \text{true})$$

$$= P(A = \text{true}) * P(B = \text{true} \mid A = \text{true}) * P(C = \text{true} \mid B = \text{true})\ P(D = \text{true} \mid B = \text{true})$$

$$= (0.4)*(0.3)*(0.1)*(0.95)$$

This is from the graph structure

These numbers are from the conditional probability tables



How is the Bayesian network created?

1. Get an expert to design it
   - Expert must determine the structure of the Bayesian network
     - This is best done by modeling direct causes of a variable as its parents
   - Expert must determine the values of the CPT entries
     - These values could come from the expert's informed opinion
     - Or an external source eg. census information
     - Or they are estimated from data
     - Or a combination of the above

2. Learn it from data
   - This is a much better option but it usually requires a large amount of data
   - This is where Bayesian statistics comes in!