CS 561 Artificial Intelligence
Lecture # 2-5

Reasoning with uncertainty

Rashmi Dutta Baruah
Department of Computer Science & Engineering
IIT Guwahati

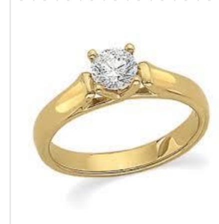
Outline

- Background
  - Knowledge
  - Representation
  - Reasoning
- Uncertainty
  - What is uncertainty?
  - Reasons of Uncertainty
- Reasoning under uncertainty: How probability theory can be used?
- Probability Theory Introduction
  - Basic Probability terminologies
  - Bayes' Rule and its application

Background

The stolen ring.

Mary noticed that her wedding diamond ring is missing from the hotel room. Find out who is the thief considering the following information:



The thief had long brown hair and wearing black shoes.
A person has long black hair if he/she is staying in room 100.
A person has short brown hair if he/she is staying in room 102.
A person has long brown hair if he/she is staying in room 205.
A person has long brown hair if he/she is staying in room 210.
A person is in room 205 if he/she wore black coat.
A person is in room 102 if he/she wore blue shirt.
A person is in room 210 if she wore red gown.
A person wore blue shirt if he was wearing a black tie.
A person wore a red gown if she is bridesmaid.
A person wore black shoes if she was wearing a silver bracelet.
A person wore black shoes if he was wearing a black tie.

James was wearing black coat.
Joe was wearing black shoes.
Jenny was wearing silver bracelet.
Jenny is bridesmaid.
Joy is bridesmaid.
Jacy is bridesmaid.

JENNY

SWI-Prolog (Multi-threaded, version 7.2.2)

File Edit Settings Run Debug Help

```
Welcome to SWI-Prolog (Multi-threaded, 32 bits, Version 7.2.2
Copyright (c) 1990-2015 University of Amsterdam, VU Amsterdam
SWI-Prolog comes with ABSOLUTELY NO WARRANTY. This is free sc
and you are welcome to redistribute it under certain conditic
Please visit http://www.swi-prolog.org for details.

For help, use ?- help(Topic). or ?- apropos(Word).

1 ?- thief(X).
X = 'Jenny' .

2 ?- trace.
true.

[trace] 2 ?- thief(X).
   Call: (7) thief(_G1510) ? creep
   Call: (8) has(_G1510, 'Long', 'Brown') ? creep
   Call: (9) in_room(_G1510, '205') ? creep
   Call: (10) wears(_G1510, 'Black', 'Coat') ? creep
   Exit: (10) wears('James', 'Black', 'Coat') ? creep
   Exit: (9) in_room('James', '205') ? creep
   Exit: (8) has('James', 'Long', 'Brown') ? creep
   Call: (8) wears('James', 'Black', 'Shoes') ? creep
   Call: (9) wears('James', 'Silver', 'Bracelet') ? creep
   Fail: (9) wears('James', 'Silver', 'Bracelet') ? creep
   Redo: (8) wears('James', 'Black', 'Shoes') ? creep
   Call: (9) wears('James', 'Black', 'Tie') ? creep
   Fail: (9) wears('James', 'Black', 'Tie') ? creep
   Redo: (8) wears('James', 'Black', 'Shoes') ? creep
   Fail: (8) wears('James', 'Black', 'Shoes') ? creep
   Redo: (8) has(_G1510, 'Long', 'Brown') ? creep
   Call: (9) in_room(_G1510, '210') ? creep
   Call: (10) wears(_G1510, 'Red', 'Gown') ? creep
   Call: (11) bridesmaid(_G1510) ? creep
   Exit: (11) bridesmaid('Jenny') ? creep
   Exit: (10) wears('Jenny', 'Red', 'Gown') ? creep
   Exit: (9) in_room('Jenny', '210') ? creep
   Exit: (8) has('Jenny', 'Long', 'Brown') ? creep
   Call: (8) wears('Jenny', 'Black', 'Shoes') ? creep
   Call: (9) wears('Jenny', 'Silver', 'Bracelet') ? creep
   Exit: (9) wears('Jenny', 'Silver', 'Bracelet') ? creep
   Exit: (8) wears('Jenny', 'Black', 'Shoes') ? creep
   Exit: (7) thief('Jenny') ? creep
X = 'Jenny' .

[trace] 3 ?-
```

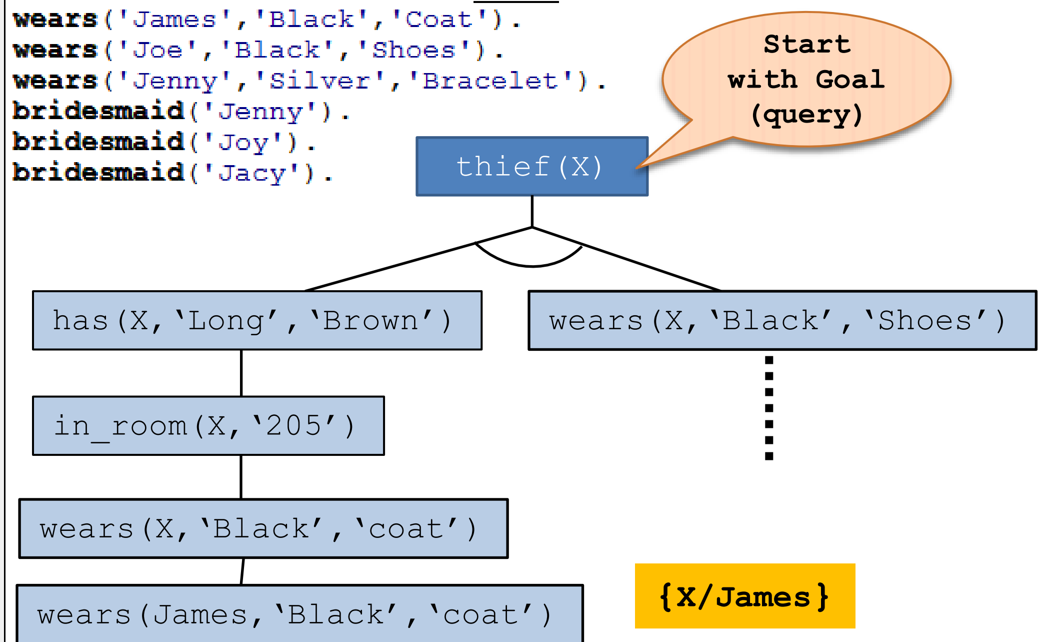
Backward chaining

stolen_ring.pl

File Edit Browse Compile Prolog Pce Help

stolen_ring.pl

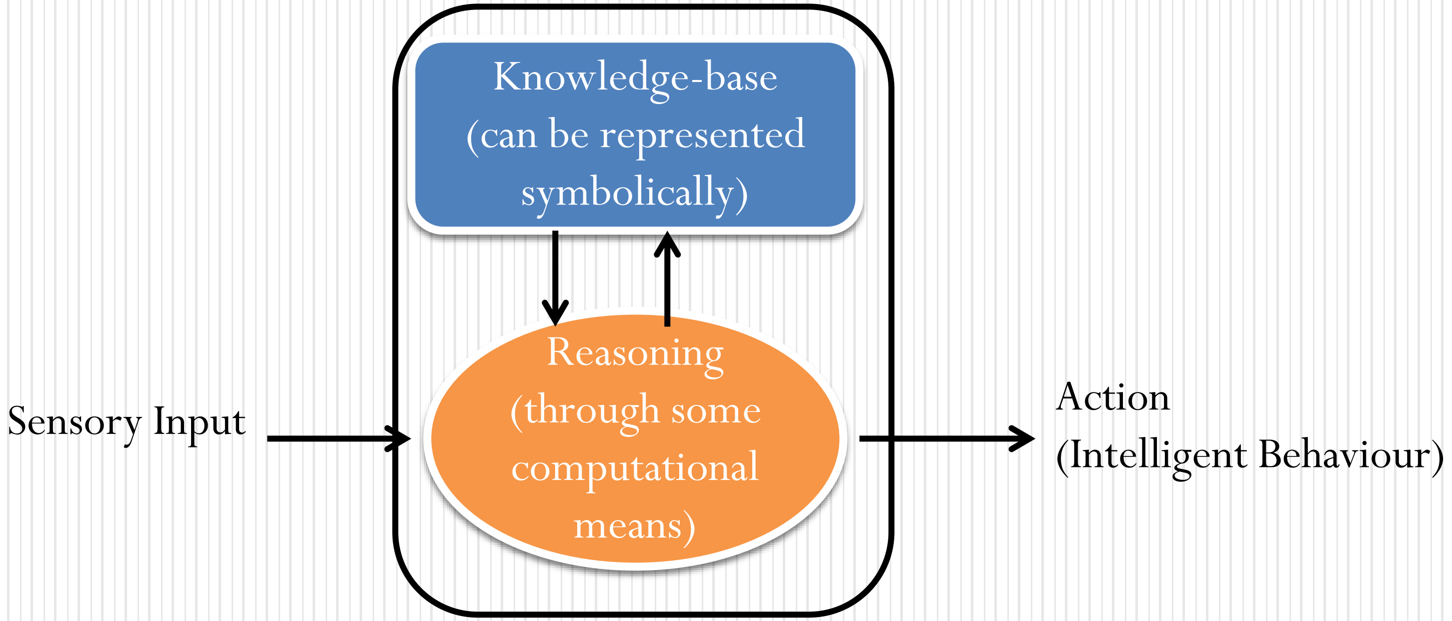
```prolog
thief(X):- has(X,'Long','Brown'), wears(X,'Black','Shoes').
has(X,'Long','Black'):- in_room(X, '100').
has(X,'Short','Brown'):- in_room(X, '102').
has(X,'Long','Brown'):- in_room(X, '205').
has(X,'Long','Brown'):- in_room(X, '210').
in_room(X,'205'):- wears(X,'Black','Coat').
in_room(X,'102'):- wears(X,'Blue','Shirt').
in_room(X,'210'):- wears(X,'Red','Gown').
wears(X,'Blue','Shirt'):- wears(X,'Black','Tie').
wears(X,'Red','Gown'):- bridesmaid(X).
wears(X,'Black','Shoes'):- wears(X,'Silver','Bracelet').
wears(X,'Black','Shoes'):- wears(X,'Black','Tie').
wears('James','Black','Coat').
wears('Joe','Black','Shoes').
wears('Jenny','Silver','Bracelet').
bridesmaid('Jenny').
bridesmaid('Joy').
bridesmaid('Jacy').
```



user:thief/1: (loaded) static, 1 clause, number_of_rules(1)

Background

- Knowledge, Representation, and Reasoning
  - How an agent uses what it knows in deciding what to do?



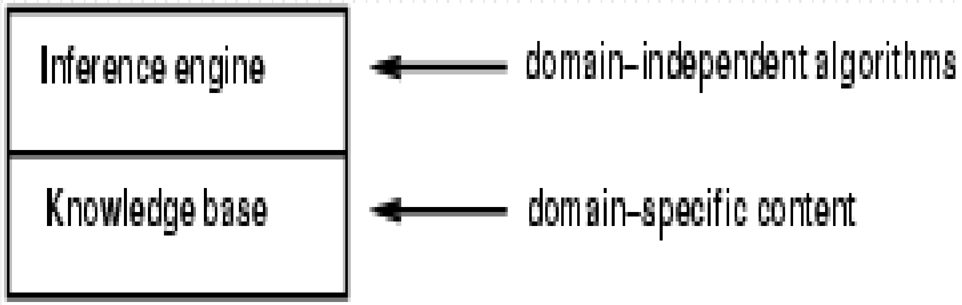
Knowledge-based Agent

Background

- Knowledge
  - collection of propositions believed by an agent
- Representation
  - with formal symbols
- Reasoning
  - formal manipulation of symbols representing believed propositions to produce representation of new ones.
- Role of Logic
  - AI community borrowed tools and techniques of formal symbolic logic for knowledge representation and reasoning.

Knowledge-Based Agents

- KB = knowledge base
  - A set of sentences or facts
  - e.g., a set of statements in a logic language
- Inference
  - Deriving new sentences from old
  - e.g., using a set of logical statements to infer new ones



- A simple KB agent
  - Agent is told or perceives new evidence (TELLs Knowledge base)
    - E.g., A is true
  - Agent selects action, this requires reasoning (ASKs Knowledge base)
    - While reasoning agent may infer new facts to add to the KB
    - E.g., KB = { A -> (B OR C) }, then given A and not C we can infer that B is true
    - B is now added to the KB even though it was not explicitly asserted, i.e., the agent inferred B
  - Agent executes action and also tells KB which action was chosen.

Uncertainty

- Uncertainty (from dictionary): the state of being unsure of something
- Uncertainty that may occur in Knowledge-based agents
  - Uncertainty in data (facts)
    - Imprecise, inaccurate and unreliable data
    - Missing data
    - Example: Medical domain
      - Patient's weight is 45 kg vs. 45.25 kg
      - Patient's weight is 43.444 kg vs. 45 kg (former is more precise not accurate if a person's actual weight is 44.9 kg)
      - Patient's weight is 45.25 kg, measured again – it is 44.95 kg.
      - Medication requires two tests, however results of only one test is available

Uncertainty

- Uncertainty in knowledge (rules)
  - Vagueness in rules
    - Example: If the person is overweight then they usually have large waistline.
  - Not enough rules to cover the problem space (theoretical or practical ignorance, lack of available theory to describe a situation)
  - Rules may be contradictory (different evidences suggesting same diagnosis)
- Issues:
  - How to represent uncertain data and knowledge?
  - How to draw inference using uncertain data and knowledge?

After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease and that the test is 99% accurate (i.e., the probability of testing positive when you do have the disease is 0.99, as is the probability of testing negative when you don't have the disease). The good news is that this is a rare disease, striking only 1 in 100,000 people of your age. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

Approaches to Handle Uncertainty

- The real world is not certain, mostly the facts that are encountered are not absolutely *true* or *false*.
- Formal symbolic logic (Propositional Logic, First order logic) represents knowledge that is either true or false. It is unable to handle uncertainty.

- Approaches to handle uncertainty
  - Probability theory- deals with incompleteness (ignorance about the world)
  - Other approaches
    - Fuzzy Logic – deals with vagueness
    - Dempster-Shafer Theory of Evidence

Reasoning with Uncertainty: Introduction

- Reasoning or Inference (dictionary meaning)
  - thinking that is coherent and logical (reasoning)
  - the reasoning involved in drawing a conclusion or making a logical judgment on the basis of circumstantial evidence and prior conclusions rather than on the basis of direct observation (inference).

- How reasoning with uncertainty relates to Probability Theory?
  - Probability provides a way of summarizing the uncertainty that comes from ignorance, quantifies the *degree of belief*

Reasoning with Uncertainty: Introduction

- Example: Medical Diagnosis

Patient Eric with symptoms : Pale yellowish skin, No temperature, Brown hair

The doctor knows that people with yellow skin (Jaundice) typically have Hepatitis, and people with Hepatitis typically have a temperature. This is all information he has that is relevant to the problem.

Issue 1: Should he proceed under the assumption that Eric has Hepatitis? (inference or reasoning)

Issue 2: How to represent doctor's information? (representation)

Rules: ***Yellow skin*** $\Rightarrow$ ***Hepatitis*** (Is it true for all patients?)

Not all patients have pale yellowish skin due to Hepatitis (could be anemia). So, the rule is not true it could be ***Yellow skin*** $\Rightarrow$ ***Hepatitis*** $\vee$ ***Anemia*** $\vee$ …..

Hepatitis $\Rightarrow$ ***Temperature*** (Is it true for all patients?)

Other Issues: Are all observations relevant (brown hair) to Hepatitis?

Reasoning with uncertainty: Introduction

- Probability theory provides a way of quantifying uncertainty in agent's knowledge (facts that are not necessarily true or false)

The doctor knows (through Medical textbooks) that 90% people with yellow skin (Jaundice) have Hepatitis, and 80% people with Hepatitis typically have a temperature. This is all information he has that is relevant to the problem.

Should he proceed under the assumption that Eric has Hepatitis? (Inference or reasoning is still the issue)

Basic Probability Terminologies

- Probability : a measure of belief (as opposed to being a frequency) – Bayesian Probability or Subjective Probability
- Example: Suppose there are three agents, Alice, Bob and Chris, and one die has been tossed.
  - Alice observes that the outcome is a "6" and tells Bob that the outcome is even but Chris knows nothing about the outcome
  - Alice has probability (degree of belief) about the outcome of the toss to be "6" is "1", Bob has probability of "1/3" (Bob believes Alice and considers that all events outcome are equally likely), Chris has probability "1/6".

Basic Probability Terminologies

- Axioms of Probability
  - Suppose P is a function from propositions (or events) into real numbers that satisfies the following three axioms of probability:
  - Axiom 1: $0 \leq P(\alpha)$ for any proposition α i.e. the belief in any proposition cannot be negative.
  - Axiom 2: $P(\tau) = 1$ if τ is a tautology i.e. if τ is true in all possible worlds, its probability is 1.
  - Axiom 3: $P(\alpha \vee \beta) = P(\alpha) + P(\beta)$ if α and β are contradictory propositions i.e. if $\sim(\alpha \wedge \beta)$ is a tautology. If two propositions can not both be true (they are mutually exclusive) then the probability of their disjunction is the sum of their probabilities.

Basic Probability Terminologies

- Conditional Probability
- We do not only want to know the prior probability of some proposition, but we want to know how this belief is updated when an agent observes new evidence.
- The **unconditional or prior probability** refers to the degree of belief in proposition in the absence of any other information (or when agent has not observed anything).
- The measure of belief in proposition α based on proposition β is called **conditional (or posterior)** probability of α given β and written as $P(\alpha|\beta)$**.** β is also referred to as **evidence**.
- For an agent the conjunction of all his observations of the world is evidence.
- When taking decision an agent has to condition on **all** the evidence it has observed.

Basic Probability Terminologies

- Conditional Probability
- Conditional probabilities are defined in terms of unconditional probabilities
  - $P(\alpha|\beta) = \frac{P(\alpha \wedge \beta)}{P(\beta)}$ which holds whenever $P(\beta) > 0$
  - It can also be written in the form of **product rule**
  - $P(\alpha \wedge \beta) = P(\alpha|\beta)\ P(\beta)$ for α and β to be true , we need β to be true and also need α to be true given β.

Basic Probability Terminologies

- Joint Probability Distribution
  - Random variable (neither random nor variable) : function from some discrete domain (in our case) → [0,1]
  - Domain of random variable is the set of all possible values that it can take on.
  - Example: random variable **Weather** and its domain is { **sunny, rain, cloudy, snow**}
  - *Weather(sunny)* = *0.2*, this is usually written as
    - *P(Weather = sunny) = 0.2* or *P(sunny) = 0.2*
  - **Probability Distribution:** probability assignment of all possible values for a random variable

Basic Probability Terminologies

- Joint Probability Distribution
  - **Probability Distribution:** probability assignment of all possible values for a random variable

    $P(Weather = sunny) = 0.6$

    $P(Weather = rain) = 0.1$

    $P(Weather = cloudy) = 0.29$

    $P(Weather = snow) = 0.01$

    $\mathbf{P}(Weather) = <0.6, 0 .1, 0.29, 0.01>$, **P** indicates that the result is a vector of numbers and defines a probability distribution
  - For multiple random variables, we use the term **Joint Probability Distribution:** probability assignment to all combinations of the values of the random variables

Basic Probability Terminologies

- Joint Probability Distribution
  - Example: **Dentistry domain** with two propositional random variables : ***Cavity*** and ***Toothache***
  - $Cavity = \{cavity, \sim cavity\}$: Does the patient have a cavity or not?
  - $Toothache = \{toothache, \sim toothache\}$: Does the patient have toothache?

	toothache	*~toothache*
cavity	0.04	0.06
~cavity	0.01	0.89

- Inference with Joint Probability Distribution: a simple method of probabilistic inference
  - Joint probability distribution table can be used as a "knowledge base" for answering any query
  - Example: What is the probability of cavities?
  - Answer: The probability of cavities is 0.1 (add elements of *cavity* row i.e. cavities with and without toothache) obtained by Marginalization (or summing out)
  - $\boldsymbol{P(Y) = \sum_{z \in Z} P(Y, z)}$

Basic Probability Terminologies

	toothache	*~toothache*
cavity	0.04	0.06
~cavity	0.01	0.89

- Inference with Joint Probability Distribution
  - Example: What is the probability that if a patient comes with a toothache has a cavity?
  - Answer: $P(cavity \mid toothache) = P(cavity \wedge toothache) / P(toothache) = 0.04/0.05 = 0.8$
  - So, the probability that someone has cavity is 0.1 but if we know that he/she has toothache then the probability increases to 0.8.

Problem: the size of the table

Bayes' Rule and its Use

- Product rule:
  - $P(a \wedge b) = P(a|b)\ P(b)$, it can also be written as
  - $P(a \wedge b) = P(b|a)\ P(a)$
- Equating the two RHS and dividing by $P(a)$, we get the equation known as Bayes' Rule:
  - $P(b|a) = P(a|b)P(b)\ /\ P(a)$
- Bayes' Rule is useful for assessing diagnostic probability from causal probability

  cause effect

  - $P(disease|symptom) = P(symptom|disease)\ P\ (disease)\ /\ P(symptom)$
  - Example: *disease* = measles , *symptom* = high fever
    - *P(disease|symptom)* may be different in India vs US : diagnostic direction
    - *P(symptom | disease)* should be same : causal direction
    - So, it is more useful to learn relationships in causal direction and use it to compute the diagnostic probabilities.

Bayes' Rule and its Use

- Conditioning (can be used to determine *P(symptom)*)
  - $P(a) = P(a \wedge b) + P(a \wedge \sim b) = P(a|b)\,P(b) + P(a|\sim b)P(\sim b)$
- Independence
  - $P(a \wedge b) = P(a).P(b)$
  - $P(a|b) = P(a)$: knowing that *b* is true does not give us any more information about the truth of a
  - $P(b|a) = P(b)$
- Independence is helpful in efficient probabilistic reasoning.
- Conditional Independence
  - a and b are conditionally independent given c
  - $P(a|b, c) = P(a|c)$
  - $P(b|a,c) = P(b|c)$
  - $P(a \wedge b \mid c) = P(a|c).\,P(b|c)$

Using Bayes' Rule: combining evidence

- Example: Dentistry domain
  - *Toothache*
  - *Cavity*
  - *Xray-Spot*
- given variables are not independent but *toothache* and *xrayspot* are conditionally independent given *cavity*
- **Combining evidence**
  - $\mathbf{P}(Cavity \mid toothache\ , xrayspot) = \dfrac{\mathbf{P}(toothache,\ xrayspot \mid Cavity)\ \ \mathbf{P}(Cavity)}{P(toothache,\ xrayspot)}$
  - Assuming that *xrayspot* and *toothache* are conditionally independent
  - $\mathbf{P}(Cavity \mid toothache\ , xrayspot) =$
$$\frac{\mathbf{P}(toothache \mid Cavity)\ \mathbf{P}(xrayspot \mid Cavity)\ \ \mathbf{P}(Cavity)}{P(toothache,\ xrayspot)}$$

Using Bayes' Rule: combining evidence

- **Combining evidence**
  - $\mathbf{P}(Cavity \mid toothache, xrayspot) =$

$$\frac{\mathbf{P}(toothache \mid Cavity)\ \mathbf{P}(xrayspot \mid Cavity)\ \ \mathbf{P}(Cavity)}{P(toothache,\ xrayspot)}$$

  - We started with prior probability of someone having cavity and as the evidence arrives (xrayspot) we multiplied a factor to and so on as new evidences arrive.

- Normalizing constant ($P(toothache, xrayspot)$)

$$P(cavity \mid toothache, xrayspot) + P(\sim cavity \mid toothache, xrayspot) = 1$$

$$\frac{P(toothache \mid cavity)\ P(xrayspot \mid cavity)\ \ P(cavity)}{P(toothache,\ xrayspot)} +$$

$$\frac{P(toothache \mid \sim cavity)\ P(xrayspot \mid \sim cavity)\ \ P(\sim cavity)}{P(toothache,\ xrayspot)} = 1$$

$$P(toothache \mid cavity)\ P(xrayspot \mid cavity)\ \ P(cavity) + P(toothache \mid \sim cavity)$$
$$P(xrayspot \mid \sim cavity)\ \ P(\sim cavity) = P(toothache,\ xrayspot)$$

Conditional Independence

- Given a cause (disease) that influences a number of effects (symptoms), which are conditionally independent, the full joint distribution can be written as

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\ \Pi_i\ \mathbf{P}(Effect_i \mid Cause)$$

  - this is called the *naïve Bayes model*
    - it makes the simplifying assumption that *all* effects are conditionally independent
    - it is naïve in that it is applied to many problems although the effect variables are not precisely conditionally independent given the cause variable
    - nevertheless, such systems often work well in practice

What did we discuss in L2-3?

- What is knowledge, representation, and reasoning?
- What is uncertainty and reasoning under uncertainty?
- Under what situations does logic fail and how Probability theory can be useful in such situations?
- What kind of uncertainty is handled by Bayesian Probability?
- How inference can be performed with Joint probability distribution tables?
- How Bayes' rule can be useful for inference?