Introdução à Gestão de Conhecimento

- Bayesian Networks

Objectives

Review on Probability Theory
Bayesian Networks

Ways to deal with Uncertainty

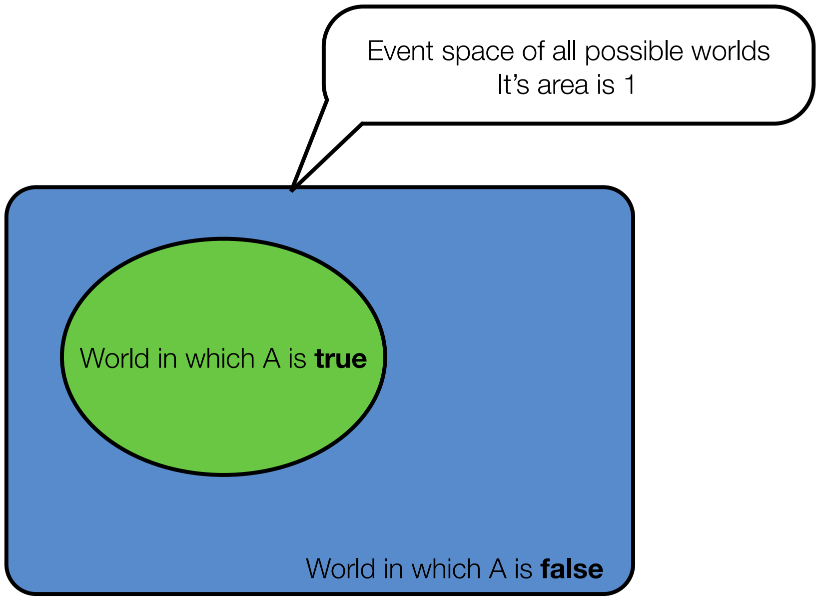
- Three-valued logic: True / False / Maybe
- Fuzzy logic (truth values between 0 and 1)
- Non-monotonic reasoning
- Dempster-Shafer theory (and an extension known as quasi-Bayesian theory)
- Possibilistic Logic
- Probability

Discrete Random Variables

- A is a Boolean-valued random variable if A denotes an event, and there is some degree of uncertainty as to whether A occurs.
- Examples
    - A = The US president in 2023 will be male
    - A = You wake up tomorrow with a headache
    - A = You have Ebola

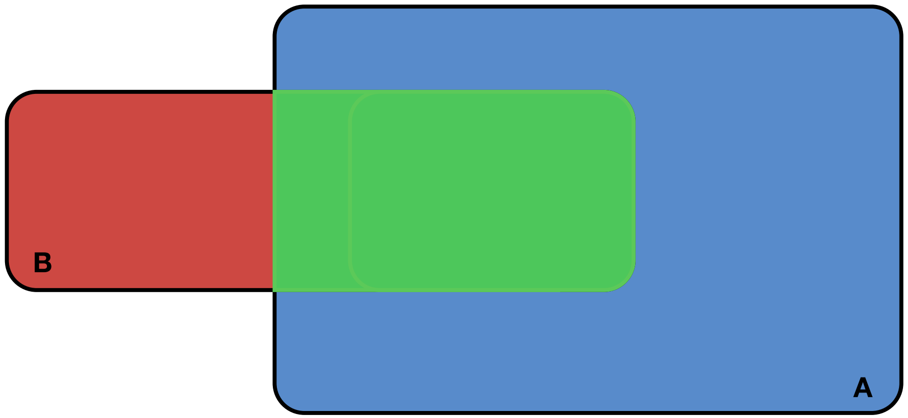
Probabilities

- We write P(A) as “the fraction of possible worlds in which A is true”
- P(A) = Area of the green oval



Interpreting the axioms

- Axioms:
  - 0<=P(A)<=1
  - P(True) = 1
  - P(False) = 0
  - P(A or B)=P(A) + P(B) - P(A and B)



Theorems from the Axioms

- 0<=P(A)<=1, P(True)=1, P(False)=0
- P(A or B) = P(A) + P(B) - P(A and B)
- From these we can prove:
  - **P(not A) = P(~A) = 1-P(A)**

- 0<= P(A) <=1, P(True)=1, P(False)=0
- P(A or B) =P(A) + P(B) - P(A and B)
- From these we can prove:
  - **P(A) = P(A ^ B) + P(A ^ ~B)**

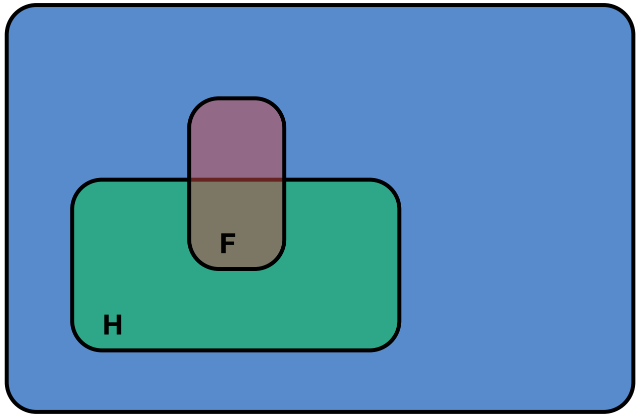
Conditional Probability

- $P(A|B)$ = Fraction of worlds in which B is true that also have A true

  H = "Have a headache"
  F = "Coming down with Flu"

  $P(H) = 1/10$
  $P(F) = 1/40$
  $P(H|F) = 1/2$

- "Headaches are rare and flu is rarer, but if you're coming down with 'flu there's a 50-50 chance you'll have a headache."



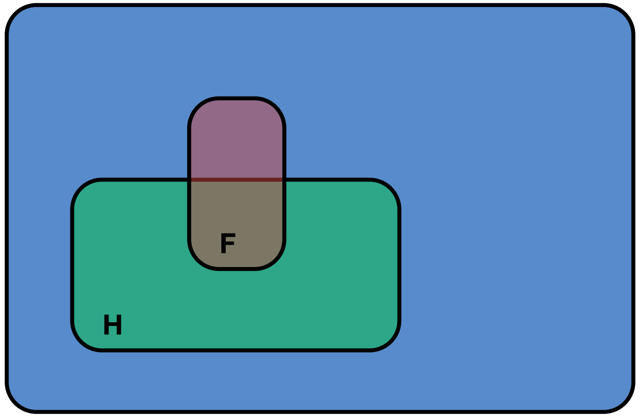
Conditional Probability

P(H|F) = Fraction of flu-inflicted worlds in which you have a headache

$$= \frac{\text{\#worlds with flu and headache}}{\text{\#worlds with flu}}$$

$$= \frac{\text{Area of “H and F” region}}{\text{Area of “F” region}}$$

$$= \frac{P(H \wedge F)}{P(F)}$$



Conditional Probability

- Definition of Conditional Probability

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

- Corollary: The Chain Rule

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

- **The Bayes Rule**:

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

Conditional Probability

- Suppose A can take on more than 2 values
- A is a random variable with arity k if it can take on exactly one value out of $\{v_1, v_2, .. v_k\}$
- Thus...

  $P(A=v_i \wedge A=v_j)=0$ if $i \neq j$
  $P(A=v_1 \vee A=v_2 \vee A=v_k) = 1$

- From the previous axioms:

  $P(B \wedge [A=v_1 \vee A=v_2 \vee A=v_i]) = \sum P(B \wedge A=v_j)$

- And thus we can prove

  $P(B) = \sum P(B \wedge A = v_j)$

More General Forms of Bayes Rule

- $$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\sim A)P(\sim A)}$$

- $$P(A|B \wedge X) = \frac{P(B|A \wedge X)P(A \wedge X)}{P(B \wedge X)}$$

- $$P(A=v_i|B) = \frac{P(B|A=v_i)P(A=v_i)}{\sum P(B|A=v_k)P(A=v_k)}$$

Bayes Rule - Example

- Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?
- **Solution**: The sample space is defined by two mutually-exclusive events - it rains or it does not rain. Additionally, a third event occurs when the weatherman predicts rain. Notation for these events appears below.
- Event **A1**. It rains on Marie's wedding.
- Event **A2**. It does not rain on Marie's wedding
- Event **B**. The weatherman predicts rain.
- In terms of probabilities, we know the following:

  P(A1) = 5/365 =0.0136985 [It rains 5 days out of the year.]

  P(A2) = 360/365 = 0.9863014 [It does not rain 360 days out of the year.]

  P(B | A1) = 0.9 [When it rains, the weatherman predicts rain 90% of the time.]

  P(B | A2) = 0.1 [When it does not rain, the weatherman predicts rain 10% of the time.]

- We want to know P(A1 | B), the probability it will rain on the day of Marie's wedding, given a forecast for rain by the weatherman. The answer can be determined from Bayes' theorem, as shown below.

  $$P(A1 \mid B) = \frac{P(B \mid A1) * P(A1)}{P(A1)\,P(B \mid A1) + P(A2)\,P(B \mid A2)}.$$

  P(A1 | B) = (0.014)(0.9) / [(0.014)(0.9) + (0.986)(0.1)]

  P(A1 | B) = 0.111

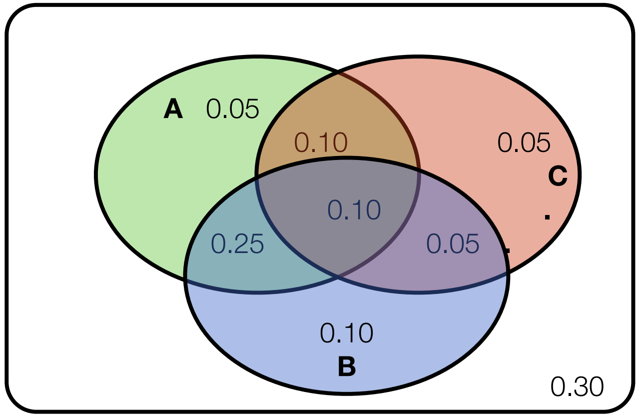
- Note the somewhat unintuitive result. Even when the weatherman predicts rain, it only rains only about 11% of the time. Despite the weatherman's gloomy prediction, there is a good chance that Marie will not get rained on at her wedding.

The Joint Distribution

- Recipe for making a joint distribution of M variables:

1. Make a truth table listing all combinations of values of your variables (if there are M Boolean variables then the table will have 2^M rows).
2. For each combination of values, say how probable it is.
3. If you subscribe to the axioms of probability, those numbers must sum to 1.

A	B	C	Prob
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10



Using the Joint Distribution

- Once you have the JD you can ask for the probability of any logical expression involving your attribute
- $P(E) = \sum_{\text{rows matching E}} P(\text{row})$
- P(poor male) = 0.46
- P(poor) = 0.75

gender	hours/ worked	wealth	Prob
Female	<40	poor	0.25
		rich	0.03
	>40	poor	0.04
		rich	0.01
Male	<40	poor	0.33
		rich	0.10
	>40	poor	0.13
		rich	0.11

Inference with the Joint Distribution

- $P(E_1|E_2) = \dfrac{P(E_1 \wedge E_2)}{P(E_2)} =$

$$= \frac{\sum_{\text{rows matching } E_1 \text{ and } E_2} P(\text{row})}{\sum_{\text{rows matching } E_2} P(\text{row})}$$

- P(Male | Poor) = 0.46 / 0.75 = 0.61

gender	hours/ worked	wealth	Prob
Female	<40	poor	0.25
		rich	0.03
	>40	poor	0.04
		rich	0.01
Male	<40	poor	0.33
		rich	0.10
	>40	poor	0.13
		rich	0.11

Joint distributions

- Good news

  Once you have a joint distribution, you can ask important questions about stuff that involves a lot of uncertainty

- Bad news

  Impossible to create for more than about ten attributes because there are so many numbers needed when you build it.

Independence

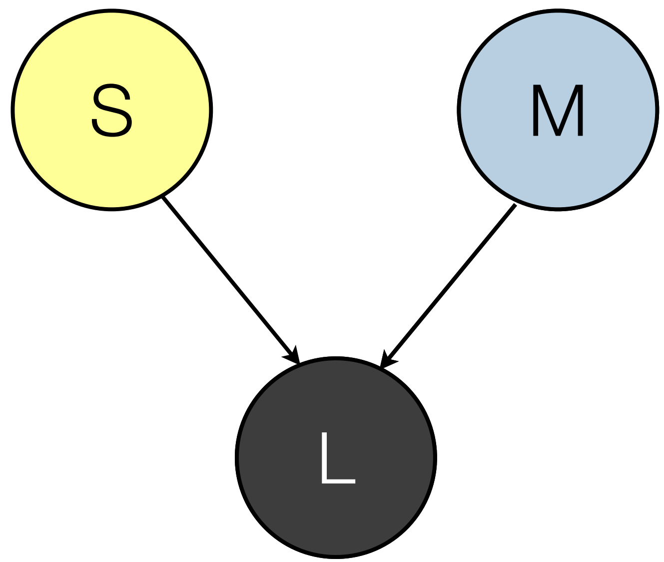
- Suppose there are two events:
  - M: class subject is Maths
  - S: It is Sunny
- The joint probability distribution function (p.d.f.) for these events contain 4 entries
- If we want to build the joint p.d.f. we'll have to invent those four numbers. Do we?
  - We don't have to specify with bottom level conjunctive events such as P(~M^S) IF...
  - ...instead it may sometimes be more convenient for us to specify things like: P(M), P(S).
- But just P(M) and P(S) don't derive the joint distribution. So you can't answer all questions.

Independence

- "The sunshine levels do not depend on and do not influence the subject"
- This can be specified very simply:
  - $P(S \mid M) = P(S)$ **This is a powerful statement!**
- It required extra domain knowledge. A different kind of knowledge than numerical probabilities. It needed an understanding of causation.
- From $P(S \mid M) = P(S)$, the rules of probability imply:
- $P(\sim S \mid M) = P(\sim S)$
- $P(M \mid S)=P(M)$
- $P(M \wedge S)=P(M).P(S)$
- $P(\sim M \wedge S) = P(\sim M)\, P(S)$, $P(M \wedge \sim S) = P(M)P(\sim S)$, $P(\sim M \wedge \sim S) = P(\sim M)P(\sim S)$

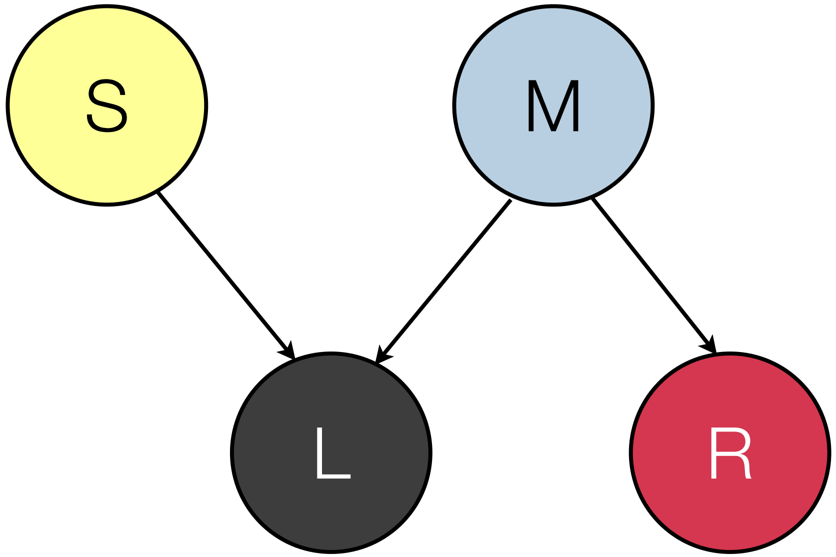
A bit of notation

- Assume a new event
  - L : The lecturer arrives slightly late.
- $P(S \mid M) = P(S)$
  $P(S) = 0.3$
  $P(M) = 0.6$
- $P(L \mid M \wedge S) = 0.05$
  $P(L \mid M \wedge {\sim}S) = 0.1$
  $P(L \mid {\sim}M \wedge S) = 0.1$
  $P(L \mid {\sim}M \wedge {\sim}S) = 0.2$



Conditional Independence

- Assume a new event
  - R: we will do Revisions in class
- $P(R \mid M,L) = P(R \mid M)$
  $P(R \mid \sim M,L) = P(R \mid \sim M)$
- "R and L are conditionally independent given M"
- Given knowledge of M and S, knowing anything else in the diagram won't help us with L, etc.



Conditional Independence formalized

- R and L are conditionally independent given M if for all x,y,z in {True,False}:
    - P(R=x | M=y ^ L=z) = P(R=x | M=y)
- More generally:
    - Let S1 and S2 and S3 be sets of variables.
    - Set-of-variables S1 and set-of-variables S2 are conditionally independent given S3 if for all assignments of values to the variables in the sets,

      P(S1’s assignments | S2’s assignments & S3’s assignments)
      = P(S1’s assignments | S3’s assignments)
    - **BUT**

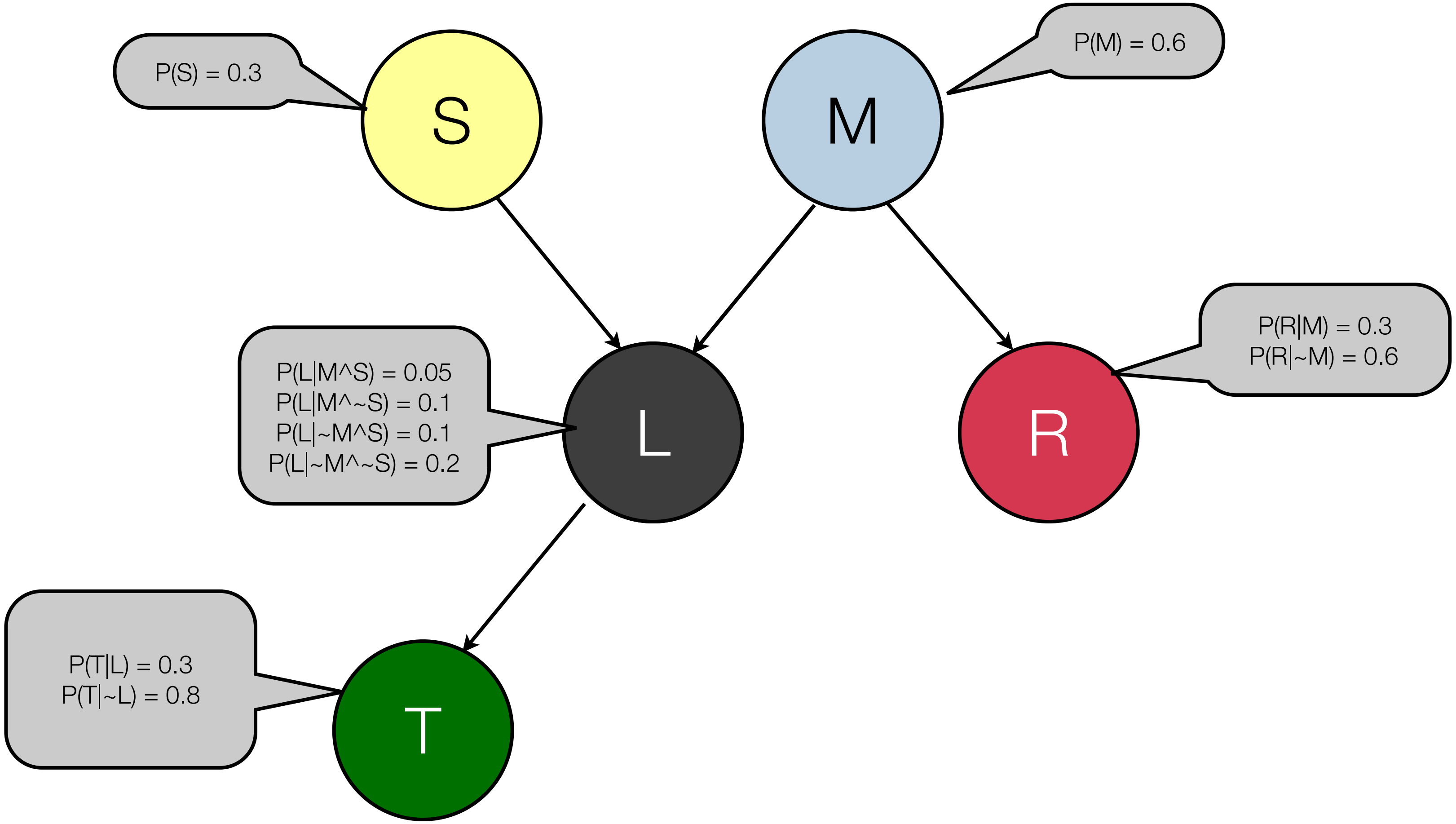
      P(S1’s assignments | S3’s assignments)
      ≠ P(S1’s assignments | S2’s assignments & S3’s assignments)

Bayes Nets Formalized

- A Bayes net (also called a belief network) is an augmented directed acyclic graph, represented by the pair V , E where:
  - V is a set of vertices.
  - E is a set of directed edges joining vertices. No loops of any length are allowed.
- Each vertex in V contains the following information:
  - The name of a random variable
  - A probability distribution table indicating how the probability of this variable's values depends on all possible combinations of parental values.

Computing a Joint Entry

- Assume a new event T (class starts at 9h15)
- How to compute an entry in a joint distribution?
  - What is P(S ^ ~M ^ L ^ ~R ^ T)?



Computing a Joint Entry

- $P(T \wedge \sim R \wedge L \wedge \sim M \wedge S) =$

  $P(T \mid \sim R \wedge L \wedge \sim M \wedge S) * P(\sim R \wedge L \wedge \sim M \wedge S) =$

  $P(T \mid L) * P(\sim R \wedge L \wedge \sim M \wedge S) =$

  $P(T \mid L) * P(\sim R \mid L \wedge \sim M \wedge S) * P(L \wedge \sim M \wedge S) =$

  $P(T \mid L) * P(\sim R \mid \sim M) * P(L \wedge \sim M \wedge S) =$

  $P(T \mid L) * P(\sim R \mid \sim M) * P(L \mid \sim M \wedge S) * P(\sim M \wedge S) =$

  $P(T \mid L) * P(\sim R \mid \sim M) * P(L \mid \sim M \wedge S) * P(\sim M \mid S) * P(S) =$

  $P(T \mid L) * P(\sim R \mid \sim M) * P(L \mid \sim M \wedge S) * P(\sim M) * P(S)$

The general case

- $P(X_1=x_1 \wedge X_2=x_2 \wedge \ldots.X_{n-1}=x_{n-1} \wedge X_n=x_n) =$
  $P(X_n=x_n \wedge X_{n-1}=x_{n-1} \wedge \ldots.X_2=x_2 \wedge X_1=x_1) =$
  $P(X_n=x_n \mid X_{n-1}=x_{n-1} \wedge \ldots.X_2=x_2 \wedge X_1=x_1) * P(X_{n-1}=x_{n-1} \wedge\ldots. X_2=x_2 \wedge X_1=x_1) =$
  $P(X_n=x_n \mid X_{n-1}=x_{n-1} \wedge \ldots.X_2=x_2 \wedge X_1=x_1) * P(X_{n-1}=x_{n-1} \mid\ldots. X_2=x_2 \wedge X_1=x_1) *$
  $P(X_{n-2}=x_{n-2} \wedge\ldots. X_2=x_2 \wedge X_1=x_1) =$

  ...

  $=\prod P((X_i =x_i) \mid ((X_{i-1}=x_{i-1})\wedge(X_1 =x_1)))$

  $=\prod P((X_i =x_i) \mid \text{Assignments of Parents}(X_i))$

- So any entry in joint pdf table can be computed. And so **any conditional probability** can be computed.

Case Study’s

- Pathfinder system. (Heckerman 1991, Probabilistic Similarity Networks, MIT Press, Cambridge MA).
  - Diagnostic system for lymph-node diseases.
  - 60 diseases and 100 symptoms and test-results.
  - 14,000 probabilities
  - Expert consulted to make net.
    - 8 hours to determine variables.
    - 35 hours for net topology.
    - 40 hours for probability table values.
  - Apparently, the experts found it quite easy to invent the causal links and probabilities.
  - Pathfinder is now outperforming the world experts in diagnosis. Being extended to several dozen other medical domains.

What you should know

- The meanings and importance of independence and conditional independence.
- The definition of a Bayes net.
- Computing probabilities of assignments of variables (i.e. members of the joint p.d.f.) with a Bayes net.

Exercise

- For example, suppose that we are interested in diagnosing cancer in patients who visit a chest clinic.
  - Let A represent the event "Person has cancer"
  - Let B represent the event "Person is a smoker"
- Suppose we know the probability of the prior event A is 0.1 on the basis of past data (10% of patients entering the clinic turn out to have cancer). Thus:
  - $P(A)=0.1$
- We want to compute the probability of the posterior event $P(A|B)$.
- It is difficult to find this out directly. However, we are likely to know $P(B)$ by considering the percentage of patients who smoke suppose $P(B)=0.5$. We are also likely to know $P(B|A)$ by checking from our records the proportion of smokers among those diagnosed with cancer. Suppose $P(B|A)=0.8$.
- Use Bayes' rule to compute $P(A|B)$

Exercise

- Doing it using Erlang:
- First we represent the basic probability facts as Erlang functions:

```
p({patient, cancer}) -> 0.1;
p({patient, smoker}) -> 0.5.
```

- We then represent the conditional probabilities

```
cp({patient, smoker}, {patient, cancer}) -> 0.8;
cp(_, _) -> none.
```

- Next we write query rules that can be used to find various probabilities. There are three rules, covering the case where the probability is known, the conditional probability is known, or we can compute the conditional probability using Bayes theorem.

```
getp({A, B}) when is_tuple(A), is_tuple(B) ->
     case cp(A, B) of
        none ->
            Pba = cp(B, A),
            Pa = getp(A),
            Pb = getp(B),
            Pba * Pa / Pb;
        X ->
            X
     end;
getp(A) -> p(A).
```

- We now have a simple system for representing and querying knowledge expressed as probabilities.

```
1> bayes:getp({{patient, cancer}, {patient, smoker}});
```

Exercise 2

- Consider the next probabilities:

  p(bonus) = 0.6

  p(not_bonus) = 0.4

  cp(money ,bonus) = 0.8

  cp(money, not_bonus) = 0.3

  cp(hawaii, money) = 0.7

  cp(Hawaii, not_money) = 0.1

  cp(san_francisco, money) = 0.2

  cp(san_francisco, not_money) = 0.5

  cp(wierd_people, san_francisco) = 0.95

  cp(wierd_people, not_san_francisco) = 0.6

  cp(surfing, Hawaii) = 0.75

  cp(surfing, not_Hawaii) = 0.2

  P(surfing) = ?