Module 10

Reasoning with Uncertainty - Probabilistic reasoning

10.1 Instructional Objective

- The students should understand the role of uncertainty in knowledge representation
- Students should learn the use of probability theory to represent uncertainty
- Students should understand the basic of probability theory, including
    - Probability distributions
    - Joint probability
    - Marginal probability
    - Conditional probability
    - Independence
    - Conditional independence
- Should learn inference mechanisms in probability theory including
    - Bayes rule
    - Product rule
- Should be able to convert natural language statements into probabilistic statements and apply inference rules
- Students should understand Bayesian networks as a data structure to represent conditional independence
- Should understand the syntax and semantics of Bayes net
- Should understand inferencing mechanisms in Bayes net
- Should understand efficient inferencing techniques like variable ordering
- Should understand the concept of d-separation
- Should understand inference mechanism for the special case of polytrees
- Students should have idea about approximate inference techniques in Bayesian networks

At the end of this lesson the student should be able to do the following:

- **Represent a problem in terms of probabilistic statemenst**
- Apply Bayes rule and product rule for inferencing
- **Represent a problem using Bayes net**
- **Perform probabilistic inferencing using Bayes net.**

Lesson
26

Reasoning with Uncertain information

10. 2 Probabilistic Reasoning

Using logic to represent and reason we can represent knowledge about the world with facts and rules, like the following ones:

bird(tweety).
fly(X) :- bird(X).

We can also use a theorem-prover to reason about the world and deduct new facts about the world, for e.g.,
?- fly(tweety).
Yes

However, this often does not work outside of toy domains - non-tautologous certain rules are hard to find.

A way to handle knowledge representation in real problems is to extend logic by using certainty factors.

In other words, replace
IF condition THEN fact
with
IF condition with certainty *x* THEN fact with certainty *f(x)*

Unfortunately cannot really adapt logical inference to probabilistic inference, since the latter is not context-free.

Replacing rules with conditional probabilities makes inferencing simpler.

Replace
smoking -> lung cancer
or
lotsofconditions, smoking -> lung cancer
with
P(lung cancer | smoking) = 0.6

Uncertainty is represented explicitly and quantitatively within probability theory, a formalism that has been developed over centuries.

A probabilistic model describes the world in terms of a set S of possible states - the sample space. We don't know the true state of the world, so we (somehow) come up with a probability distribution over S which gives the probability of any state being the true one. The world usually described by a set of variables or attributes.

Consider the probabilistic model of a fictitious medical expert system. The 'world' is described by 8 binary valued variables:

Visit to Asia? A
Tuberculosis? T
Either tub. or lung cancer? E
Lung cancer? L
Smoking? S
Bronchitis? B
Dyspnoea? D
Positive X-ray? X

We have $2^8 = 256$ possible states or configurations and so 256 probabilities to find.

10.3 Review of Probability Theory

The primitives in probabilistic reasoning are *random variables*. Just like primitives in Propositional Logic are propositions. A random variable is not in fact a variable, but a function from a sample space S to another space, often the real numbers.

For example, let the random variable Sum (representing outcome of two die throws) be defined thus:
Sum(die1, die2) = die1 +die2

Each random variable has an associated probability distribution determined by the underlying distribution on the sample space

Continuing our example : P(Sum = 2) = 1/36,
P(Sum = 3) = 2/36, . . . , P(Sum = 12) = 1/36

Consdier the probabilistic model of the fictitious medical expert system mentioned before. The sample space is described by 8 binary valued variables.

Visit to Asia? A
Tuberculosis? T
Either tub. or lung cancer? E
Lung cancer? L
Smoking? S
Bronchitis? B
Dyspnoea? D
Positive X-ray? X

There are $2^8 = 256$ events in the sample space. Each event is determined by a joint instantiation of all of the variables.

S = {(A = f, T = f,E = f,L = f, S = f,B = f,D = f,X = f),
(A = f, T = f,E = f,L = f, S = f,B = f,D = f,X = t), . . .
(A = t, T = t,E = t,L = t, S = t,B = t,D = t,X = t)}

Since S is defined in terms of joint instantations, any distribution defined on it is called a joint distribution. ll underlying distributions will be joint distributions in this module. The variables {A,T,E, L,S,B,D,X} are in fact random variables, which ‘project’ values.

L(A = f, T = f,E = f,L = f, S = f,B = f,D = f,X = f) = f
L(A = f, T = f,E = f,L = f, S = f,B = f,D = f,X = t) = f
L(A = t, T = t,E = t,L = t, S = t,B = t,D = t,X = t) = t

Each of the random variables {A,T,E,L,S,B,D,X} has its own distribution, determined by the underlying joint distribution. This is known as the margin distribution. For example, the distribution for L is denoted P(L), and this distribution is defined by the two probabilities P(L = f) and P(L = t). For example,

P(L = f)
= P(A = f, T = f,E = f,L = f, S = f,B = f,D = f,X = f)
+ P(A = f, T = f,E = f,L = f, S = f,B = f,D = f,X = t)
+ P(A = f, T = f,E = f,L = f, S = f,B = f,D = t,X = f)
. . .
P(A = t, T = t,E = t,L = f, S = t,B = t,D = t,X = t)

P(L) is an example of a marginal distribution.

Here’s a joint distribution over two binary value variables A and B

	A=0	A=1
B=0	0.2	0.3
B=1	0.4	0.1

We get the marginal distribution over B by simply adding up the different possible values of A for any value of B (and put the result in the “margin”).

	A=0	A=1	
B=0	0.2	0.3	0.5 (= 0.2+0.3)
B=1	0.4	0.1	0.5 (=0.4 + 0.1)

In general, given a joint distribution over a set of variables, we can get the marginal distribution over a subset by simply summing out those variables not in the subset.

In the medical expert system case, we can get the marginal distribution over, say, A,D by simply summing out the other variables:

$$P(A,D) = \sum_{T,E,L,S,B,X} P(A,T,E,L,S,B,D,X)$$

However, computing marginals is not an easy task always. For example,
P(A = t,D = f)
= P(A = t, T = f,E = f,L = f, S = f,B = f,D = f,X = f)
+ P(A = t, T = f,E = f,L = f, S = f,B = f,D = f,X = t)
+ P(A = t, T = f,E = f,L = f, S = f,B = t,D = f,X = f)
+ P(A = t, T = f,E = f,L = f, S = f,B = t,D = f,X = t)
. . .
P(A = t, T = t,E = t,L = t, S = t,B = t,D = f,X = t)

This has 64 summands! Each of whose value needs to be estimated from empirical data. For the estimates to be of good quality, each of the instances that appear in the summands should appear sufficiently large number of times in the empirical data. Often such a large amount of data is not available.

However, computation can be simplified for certain special but common conditions. This is the condition of *independence* of variables.

Two random variables A and B are independent iff

P(A,B) = P(A)P(B)

i.e. can get the joint from the marginals

This is quite a strong statement: It means for **any** value *x* of A and **any** value *y* of B

P(A = *x*,B = *y*) = P(A = *x*)P(B = *y*)

Note that the independence of two random variables is a property of a the underlying

$$P(A,B) = P(A)P(B) \qquad P'(A,B) \neq P'(A)P'(B)$$

probability distribution. We can have

Conditional probability is defined as:

$$P(A|B) \stackrel{\text{def}}{=} \frac{P(A,B)}{P(B)}$$

It means for any value *x* of A and any value *y* of B

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

If A and B are independent then

$$P(A|B) = P(A)$$

Conditional probabilities can represent causal relationships in both directions.

From cause to (probable) effects

$$\begin{array}{rcl} Car\_start = f & \leftarrow & Cold\_battery = t \\ P(Car\_start = f & | & Cold\_battery = t) = 0.8 \end{array}$$

From effect to (probable) cause

$$\begin{array}{rcl} Cold\_battery = t & \leftarrow & Car\_start = f \\ P(Cold\_battery = t & | & Car\_start = f) = 0.7 \end{array}$$