Bayes Nets for representing and reasoning about uncertainty

Andrew W. Moore

16-731 Spring 1998

Handout AWM4

See Russell and Norvig, Chapter 15

Probability

By far the most universally accepted and used formalism for uncertainty.

Well developed semantics and theory and proofs and examples and software and education.

But how good is it for representing large amounts of general purpose knowledge about an uncertain world?

Are there alternatives?

What are the computational issues?

We'll come back to this at the end. But for now let's steam ahead with the glamorous world of .... *probability*

The fundamental rules of probability

Let A be an event, the occurence of which we are uncertain about.

The axioms of probability:

0 <= P(A) <= 1

P(True) = 1 , P(False) = 0

P(A or B) = P(A) + P(B) - P(A and B)

Another fact about probability theory.

P(~A) = 1 - P(A)

This can be deduced from the axioms we just saw:

0 <= P(A) <= 1

P(True) = 1 , P(False) = 0

P(A or B) = P(A) + P(B) - P(A and B)

How????

Another fact

P(B) = P(B ^ A) + P(B ^ ~A)

Why?

More generally (prove by induction):

If we know that exactly one of A1 , A2 ..., An are true (i.e . P(A1 or A2 or ... An) = 1, and forall i,j unequal, P(Ai and Aj) = 0) THEN we know:

P(B) = P(B ^ A1) + P(B ^ A2) + ... P(B ^An)

Example

My mood can take one of two values: Happy , Sad.

The weather can take one of three values: Rainy , Sunny Cloudy .

My knowledge base sa ys..

P(Mood=Happy ^ W eather=Rainy) = 0.2

P(Mood=Happy ^ W eather=Sunny) = 0.1

P(Mood=Happy ^ W eather=Cloudy) = 0.4

Can I compute P(Mood=Happy)?

Can I compute P(Moody=Sad)?

Can I compute P(W eather=Rainy) ?

What' s so great about these axioms?

The axioms of probability mean you ma y not represent the following knowledge:

P(A) = 0.4

P(B) = 0.3

P(A ^ B) = 0.0

P(A $\vee$ B) = 0.8

Would you ever w ant to do that?

Difficult philosophical question.

Pragmatic answer: if you disobey the la ws of probability in your knowledge representation, you can make suboptimal decisions . See R&N page 421 for more details .

Conditional probability

P(A | B) denotes the probability of A given that B' s occurence is known.

P(Ca vity) = 0.04

“There' s a 4% c hance at any time that you ha ve a ca vity”

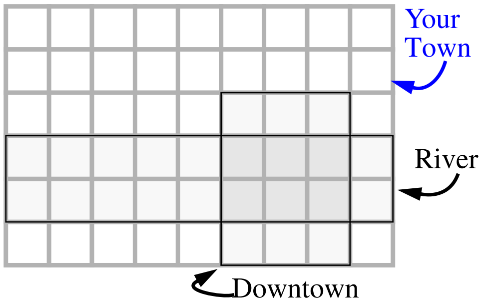
P(Ca vity | T oothac he) = 0.8

“If you ha ve a toothac he , there' s an 80% chance you ha ve a ca vity”

P(B|A) = 1 is equivalent to A => B

P(B|A) = 0.95 is a little bit like a “soft fuzzy” version of A=>B .

Lunar Lander Example



A lunar lander crashes somewhere in your town (one of the cells at random on the grid). The crash point is uniformly random.

R is the event that it crashes in the river.

D is the event that it crashes downtown.

- What are P(R) P(D) P(D ^ R) ?
- What is P(D | R) ?
- What is P(R | D) ?
- What is P(R ^ D) / P(D) ?

Are your eyebrows waggling with interest?

Conditional Probability

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

Useful to remember ("the chain rule"):

P(A ^ B) = P(A | B) P(B). Don't just remember this algebraically. Understand it! Live it!

Exercise: Prove that P(A ^ B) <= P(A) for any events A and B.

P(Welsh) = 0.001

P(Sensible | Welsh) = 0.5

P(LeekEater | Sensible ^ Welsh) = 0.1

What is

P(Welsh ^ Sensible ^ Leekeater)?

Combining what we know ...

If we know that exactly one of A1 , A2 ..., An are true (i.e. P(A1 or A2 or ... An) = 1, and forall i,j unequal, P(Ai and Aj) = 0) THEN we know:

P(B) = P(B ^ A1) + P(B ^ A2) + ... P(B ^An)

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

So P(A^B) = P(B|A) P(A)

So...

If we know that exactly one of A1 , A2 ..., An are true (i.e. P(A1 or A2 or ... An) = 1, and forall i,j unequal, P(Ai and Aj) = 0) THEN we know:

P(B) = P(B|A1)P(A1) + P(B|A2)P(A2) + ...
... P(B|An) P(An)

Bayes' Rule

Simplest form:

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

More general form:

$$P(A \mid B, X) = \frac{P(A \mid X)\,P(B \mid A, X)}{P(B \mid X)}$$

Medical Diagnosis

A doctor knows that meningitis causes a stiff neck 50% of the time.

The doctor knows that if a person is randomly selected from the US population, there's a 1/50,000 chance the person will have meningitis.

The doctor knows that if a person is randomly selected from the US population, there's a 5% chance the person will have a stiff neck.

You walk into the doctor complaining of the **symptom** of a stiff neck. What's the probability that the **underlying cause** is meningitis?

Generalizing Bayes Rule

If we know that exactly one of A1 , A2 ..., An are true (i.e. P(A1 or A2 or ... An) = 1, and forall i,j unequal, P(Ai and Aj) = 0) THEN we know:

P(B) = P(B|A1)P(A1) + P(B|A2)P(A2) + ...

... P(B|An) P(An)

and in general

P(B|X) = P(B|A1,X)P(A1,X) + ...

... P(B|An,X) P(An,X)

so

$$P(Ak \mid B,X) = \frac{P(Ak|X)\, P(B \mid Ak,X)}{\sum_i P(Ai|X)\, P(B \mid Ai,X)}$$

Example

I have three identical boxes labeled H1, H2 and H3.

Into H1 I place 1 black bead, 3 white beads.

Into H2 I place 2 black beads, 2 white beads.

Into H3 I place 4 black beads, no white beads.

I draw a box at random. I remove a bead at random from that box. What can I deduce from the color of the bead as to which box I drew?

If I replace the bead, then redraw another bead at random from the same box, how well can I predict its color before drawing it?

These two questions are the foundations of **reasoning with uncertainty** and **machine learning**.

Bayesian Ranting

A nice way to look at this:

- H1 , H2 and H3 were my prior models of the world.
- The fact that P(H1) = 1/3, P(H2) = 1/3, P(H3) = 1/3 was my **prior distribution**.
- The color of the bead was a piece of **evidence** about the true model of the world.
- The use of bayes' rule was a piece of probabilistic inference, giving me a **posterior distribution** on possible worlds.
- Learning is prior + evidence ---> posterior
- A piece of evidence decreases my ignorance about the world.
- Distributions are good ways of describing your state of knowledge. Knowledge that includes uncertainty measure can mean much better decision-making.

Another example

Suppose we will wish to reason about flying ability , birdhood and youth in animals .

(Note: a contrived example! Bayesian and probabilistic applications are not usually in the zoological field).
We can set up or knowledge base as a probability distribution before we receive any information about an animal.

Bird	Flier	Young	Prob
T	T	T	0
T	T	F	0.2
T	F	T	0.04
T	F	F	0.01
F	T	T	0.01
F	T	F	0.01
F	F	T	0.23
F	F	F	0.5

Each row is a prior hypothesis about the state of the animal.

Marginalizing

This is a Joint Probability Distribution :

Bird	Flier	Young	Prob
T	T	T	0
T	T	F	0.2
T	F	T	0.04
T	F	F	0.01
F	T	T	0.01
F	T	F	0.01
F	F	T	0.23
F	F	F	0.5

Its probabilities must add up to 1. It defines all basic conjunctive probabilities , e.g: P(B ^ ~F ^ Y) = 0.04

We can compute marginal probabilities (probabilities for subsets of variables taking specified values) easily ...
e.g. P(B ^ ~Y)

We can compute conditional probabilities with ease . E.G .:
P(Y |B ^ ~F) =

Conditional probabilities from the joint distribution

Let x1 , x2 ... be values True or False .

$$P(X_1=x_1, X_2=x_2, \ldots X_n=x_n \mid X_{n+1}=x_{n+1}, \ldots X_{n+k}=x_{n+k}) = \frac{\text{Sum of all entries for which X1=x1, ... X(n+k) = x(n+k)}}{\text{Sum of all entries for which X(n+1)=x(n+1)....X(n+k) = x(n+k)}}$$

Bird	Flier	Young	Prob
T	T	T	0
T	T	F	0.2
T	F	T	0.04
T	F	F	0.01
F	T	T	0.01
F	T	F	0.01
F	F	T	0.23
F	F	F	0.5

P(B=False | F=True ,Y=False) =

$$\frac{\text{P(B=False,F=True,Y=False)}}{\text{P(B=True,F=True,Y=False)+P(B=True,F=False,Y=False)}} = \frac{0.01}{0.21} = \frac{1}{21}$$

The joint probability distribution

Contains all the domain knowledge you'll need for any conditional probability .

"What's the probability that it's young given that it either flies or is a bird"

"I took two animals . I don't know what they were , but one of them is known to fly , and the other is young . If I give you one at random, what's the chance it is a bird?"

Joint probability distributions are a great way of describing knowledge . Or so it seems .

Question: What's the big problem with this?

Using fewer numbers

Suppose there are two events:

- M : Manuela teaches the AI core lecture (otherwise it's Andrew)
- S : It is sunny

The joint p.d.f for these events contain four entries.

If we want to build the joint p.d.f we'll have to invent those four numbers. OR WILL 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.

What extra assumption can you make?

Independence

"The sunshine levels do not depend on and do not influence who is teaching."

This can be specified very simply:

P(S | M) = P(S)

This is a powerful statement!

It required extra domain knowledge. A different kind of knowledge than p.d.f.s. It needed an understanding of causation.

From P(S | M) = P(S), the rules of probabilty imply: *(can you prove these?)*

- P(~S | M) = P(~S)
- P(M | S) = P(M)
- P(M ^ S) = P(M) P(S)
- P(~M^S) = P(~M)P(S), P(M^~S) = P(M)P(~S), P(~M^~S) = P(~M)P(~S)

Independence

We've stated:

P(M) = 0.6

P(S) = 0.3

P(S | M) = P(S)

From these two numbers, and an independence assumption, we can derive the full joint pdf.

M	S	Prob
T	T	
T	F	
F	T	
F	F	

And since we now have the joint pdf, we can make any queries we like.

A more interesting case

Suppose there are three events:

- M : Manuela teaches the AI core lecture (otherwise it's Andrew)
- S : It is sunny
- L : The lecturer arrives slightly late.

Assume both lecturers are sometimes delayed by bad weather. And Andrew is more likely to arrive late than Manuela.

Let's begin with writing down knowledge we're happy about:

P(S | M) = P(S), P(S) = 0.3, P(M) = 0.6

Now, lateness is not independent of the weather and is not independent of the lecturer. We must choose what else we need to write down.

We know the joint pdf of S and M, so let's just write down P(L | S=x, M=y) in the 4 cases.

A more interesting case

- M : Manuela teaches the AI core lecture (otherwise it's Andrew)
- S : It is sunny
- L : The lecturer arrives slightly late.

Assume both lecturers are sometimes delayed by bad weather. And Andrew is more likely to arrive late than Manuela.

P(S | M) = P(S)
P(S) = 0.3
P(M) = 0.6
P(L | M ^ S) = 0.05
P(L | M ^ ~S) = 0.1
P(L | ~M ^ S) = 0.1
P(L | ~M ^ ~S) = 0.2

Now we can derive a full joint p.d.f with a "mere" six numbers instead of eight. (Savings are larger for larger numbers of variables).

Question: Express P(L=x ^ M=y ^ S=z) in terms that only need the above expressions, where x, y and z may each be True or False.

An even cuter trick

Suppose we have these three events:

- M : Lecture taught by Manuela
- L : Lecturer arrives late
- R : Lecture concerns robots

Suppose Andrew has a higher chance of being late than Manuela.

Suppose Andrew has a higher chance of giving robotics lectures than Manuela.

What kind of independence can we find?

How about:

- P(L | M) = P(L) ?
- P(R | M) = P(R) ?
- P(L | R) = P(L) ?

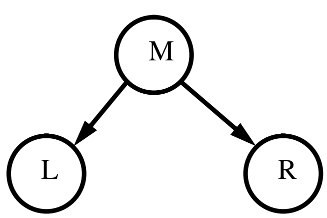
Conditional independence

Once you know who the lecturer is, then whether they arrive late doesn't affect whether the lecture concerns robots.

P(R | M , L) = P(R | M)
and
P(R | ~M , L) = P(R | ~M)
We express this in the following way:

"R and L are conditionally independent given M"

Which is also notated by the following diagram.



Given knowledge of M, 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

forall x,y ,z in {T,F}:

P(R=x | M=y ^ L=z) = P(R = x | M = y)

More generally:

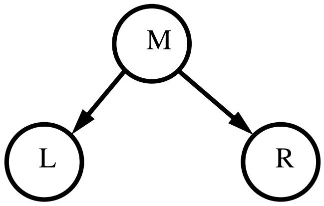
Let S1 and S2 and S3 be sets of variables.

Set-of-variables S1 and set-of-variables S2 are conditionally independent given S3 if forall assignments of values to the variables in the sets,

P(S1's assignments | S2's assignments and S3's assignments) =

P(S1's assigns | S3's assigns)

Conditional Independence

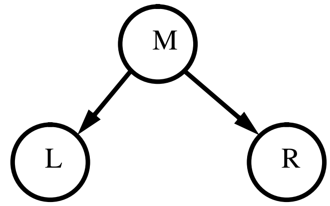


We can write down P(M). And then, since we know L is only directly influenced by M, we can write down the values of P(L|M) and P(L|~M) and know we've fully specified L's beha vior. Ditto for R.

- P(M) = 0.6
- P(L | M) = 0.085
- P(L | ~M) = 0.17
- P(R | M) = 0.3
- P(R | ~M) = 0.6

'R and L conditionally independent given M'

Conditional Independence



P(M) = 0.6
P(L | M) = 0.085
P(L | ~M) = 0.17
P(R | M) = 0.3
P(R | ~M) = 0.6

Conditional Independence:
P(R|M , L) = P(R|M) ,
P(R | ~M , L) = P(R | ~M)

Once again, we can obtain any member of the joint prob dist that we desire:
P(L=x ^ R=y ^ M=z) =

A Bayes Net

Suppose:
T : The lecture started by 10:35
L : The lecturer arrives late
R : The lecture concerns robots
M : The lecturer is Manuela
S : It is sunny

- T only directly influenced by L (i.e. T is conditionally independent of R,M,S given L)
- L only directly infuenced by M and S (i.e. L is conditionally independent of R given M & S)
- R only directly influenced by M (i.e. R is conditionally independent of L, S, given M)
- M and S are independent

Bayes Nets Formalized

A Ba yes net (also called a belief network) is an augmented directed acyclic graph, represented by the pair V , E where:

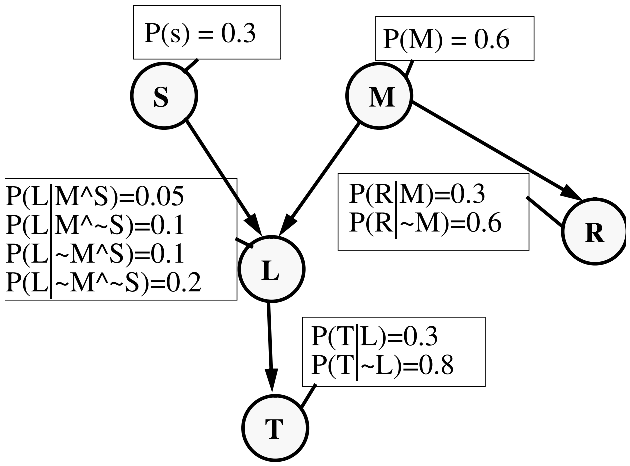
- V is a set of vertices.
- E is a set of directed edges joining vertices. No loops of any length are allowed.

Eac h 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 all possible combinations of parental values.

The is an edge from X to Y if and only if X has a direct influence on Y.

A Bayes Net



- Two unconnected variables still can affect each other.
- Each node is conditionally independent of anyone earlier in the tree, given its parents.
- You can deduce many other conditional independence relations from a Bayes net. It's complicated though.

Building a Bayes Net

We will place an ordering on nodes (call them X1 , X2 ... Xn), such that:

- X1 has no parents,
- Parents(Xi) is a subset of {X1, X2 , ... Xi-1}

This can always be done for any acyclic graph. There are usually multple solutions.

To make a Bayes Net, follow these rules:

1. Choose a set of relevant variables.
2. Choose an ordering for them X1 .. Xn
3. While there are variables left:
   1. Pick Xi and add a node to the network
   2. Set Parents(Xi) to be a minimal set of already-added nodes such that we have conditional independence of Xi and all other members of {X1 .. Xi-1 } given Parents(Xi)
   3. Define the conditional prob. table of P(Xi=x | Assignments of Parents(Xi)).

Example Bayes Net Building

Suppose we're building a nuclear power station. There are the following random variables:

GRL : Gauge Reads Low.
CTL : Core temperature is low.
FG : Gauge is faulty.
FA : Alarm is faulty.
AS : Alarm sounds

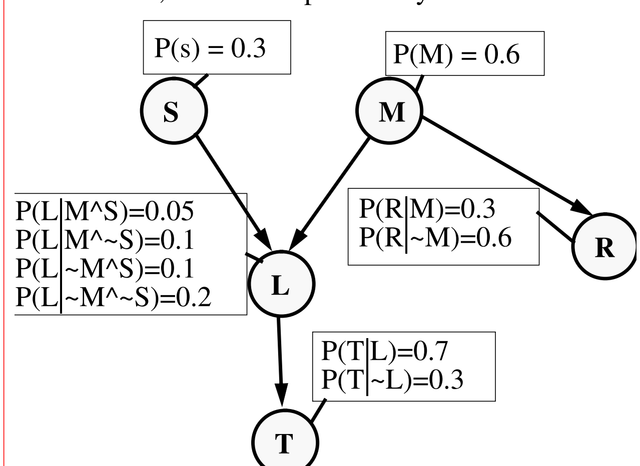
If alarm working properly, the alarm is meant to sound if the gauge stops reading a low temp.

If gauge working properly, the gauge is meant to read the temp of the core.

Computing with a Bayes Net

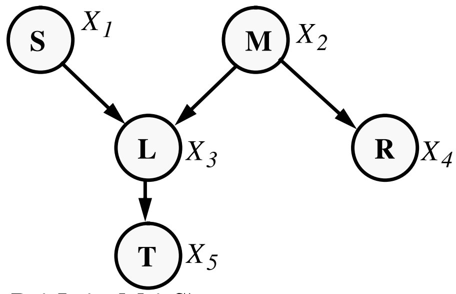
The first thing we might want to do is compute an entry in a joint prob. table:

Given an assignment of truth values to our variables, what is the probability? E.G:



What is P(S ^ ~M ^ L ^ ~R ^ T)?

Computing with Bayes Net



P(T ^ ~R ^ L ^ ~M ^ S) =

P(T | ~R ^ L ^ ~M ^ S) * P(~R ^ L ^ ~M ^ S) =

P(T | L) * P(~R ^ L ^ ~M ^ S) =

P(T | L) * P(~R | L^~M^S) * P(L^~M^S) =

P(T|L)*P(~R|~M)*P(L^~M^S) =

P(T|L)*P(~R|~M)*P(L|~M^S)*P(~M^S) =

P(T|L)*P(~R|~M)*P(L|~M^S)*P(~MS)*P(S)=

P(T|L)*P(~R|~M)*P(L|~M^S)*P(~M)*P(S).

The general case

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

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

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

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

=

$$\prod_{i=1}^{n} P((X_i = x_i)|((X_{i-1} = x_{i-1}) \wedge ...(X_1 = x_1)))$$

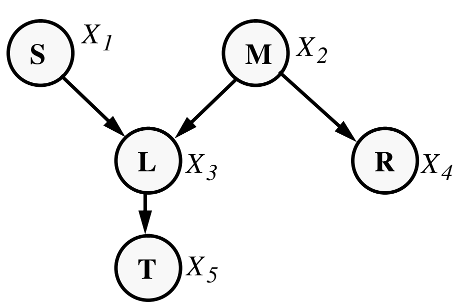
=

$$\prod_{i=1}^{n} P((X_i = x_i)|\text{Assignments of Parents}(X_i))$$

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

Where are we now?

- We ha ve a methodology for building Ba yes nets .
- We don't require exponential storage to hold our probability table . Only exponential in the maximum number of parents of any node .
- We can compute probabilities of any given assignment of truth values to the variables . And we can do it in time linear with the number of nodes .
- So we can also compute answers to any questions .



E.G . What could we do to compute P(R | T , ~S)?

The sad, bad, news

We don't just w ant the simple queries . We are often interested in conditional probabilities:

P(Some variable | Some other variables values)

Conditional probabilities computed using the automatic procedure of enumerating all matc hing cases in the joint are expensive:

Exponential in the number of variables.

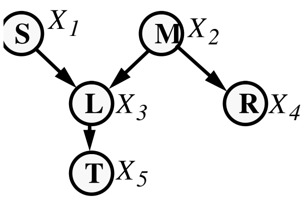
But perhaps there are faster w ays of querying Ba yes nets?

Sadder and worse news:

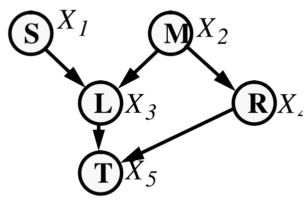
General querying of Bayes nets is NP-complete.

Bayes nets queries: Practical algorithms

A poly-tree is a directed acyclic graph in which no two nodes have more than one path between them.



A poly tree

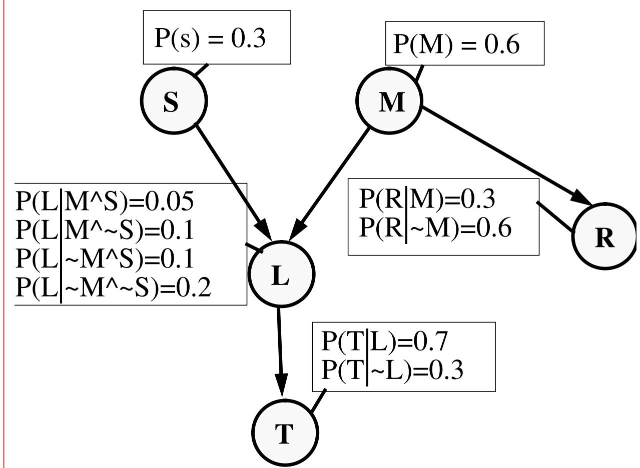


Not a poly tree
(but still a legal Bayes net)

- If net is a poly-tree, there is a linear-time algorithm (which we won't discuss).
- The best general-case algorithm converts a general net to a poly-tree (often at huge expense) and calls the poly-tree algorithm.
- Another popular, practical approach that is gaining popularity (doesn't assume polytree):

Stochastic Simulation. Lets look at it.

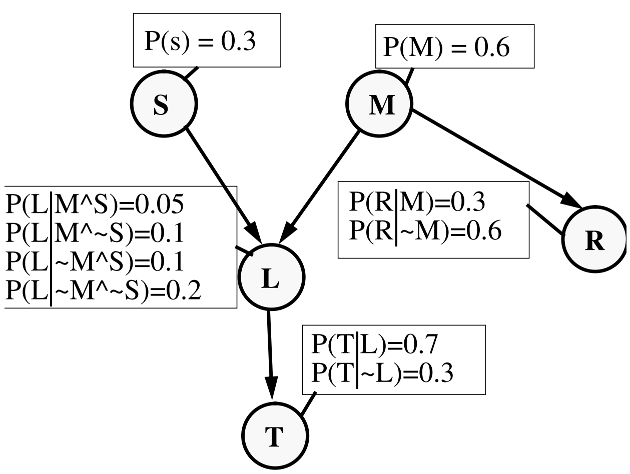
Sampling from the Joint Distribution



It's pretty easy to generate a set of variable-assignments at random with the same probability as the underlying joint distribution.

How?

Sampling from the Joint Distribution



- 1. Randomly choose S. S = True with prob 0.3
- 2. Randomly choose M. M = True with prob 0.6
- 3. Randomly choose L. The probability that L is true depends on the assignments of S and M. E.G. if steps 1 and 2 had produced S=True, M=False, then probability that L is true is 0.1.
- 4. Randomly choose R. Probability depends on M.
- 5. Randomly choose T. Probability depends on T.

Now S, M, L, R, T are a random sample from the joint distribution.

A general sampling algorithm

Let's generalize the example on the previous slide to a general Bayes Net.

As in Slide 24, call the variables $X_1 .. X_n$, where *Parents*(X_i) must be a subset of $\{ X_1 .. X_{i-1} \}$.

For $i = 1$ to n:

1. Find parents, if any, of X_i. Assume $n(i)$ parents. Call them $X_{p(i,1)}, X_{p(i,2)} \ldots X_{p(i,n(i))}$.
2. Recall the values that those parents were randomly given: $x_{p(i,1)}, x_{p(i,2)} \ldots x_{p(i,n(i))}$.
3. Look up in the lookup-table for:
   $P(X_i{=}True \mid X_{p(i,1)}{=}x_{p(i,1)}, X_{p(i,2)}{=}x_{p(i,2)} \ldots X_{p(i,n(i))}{=}x_{p(i,n(i))})$
4. Randomly set $x_i = True$ according to this probability

$x_1, x_2, \ldots x_n$ are now a sample from the joint distribution of $X_1, X_2, \ldots X_n$.

Stochastic Simulation Example

Someone wants to know P(R | T ^ ~S).

This is the same as

P(R = True | T = True ^ S = False)

We'll do lots of random samplings and count the number of occurences of the following:

- Nc : Number of random samplings in which T=True and S=False .
- Ns : Number of random samplings in which R=True and T=True and S=False .
- N : Number of random samplings .

Now, if N is big enough, Nc/N is a good estimate of P(T=True and S=False).

Also if N is big enough, Ns/N is a good estimate of P(R=True and T=True and S=False).

P(R|T^~S) = P(R^T^~S) / P(T^~S), so Ns/Nc can be a good estimate of P(R|T^~S).

Stochastic Simulation: General Algorithm

Someone wants to know P(E1| E2). where E1 is a set of variable assignments and E2 is a set of variable assignments (e.g. E1 = "R=True"; E2 = "T=True and S = False")

We'll do lots of random samplings and count the number of occurences of the following:

- Nc : Number of random samplings in which the assignments in E2 are satisfied.
- Ns : Number of random samplings in which the assignments in E1 are satisfied AND the assignments in E2 are satisfied.
- N : Number of random samplings .

Now, if N is big enough, Nc/N is a good estimate of P(E2).

Also if N is big enough, Ns/N is a good estimate of P(E1 and E2).

P(E1|E2) = P(E1^E2) / P(E2), so Ns/Nc can be a good estimate of P(E1 | E2).

Likelihood weighting

Problem with Stochastic Sampling:

With lots of constraints in E, or unlikely events in E, then most of the simulations will be thrown away (they'll have no effect on Nc , or Ns).

Imagine we're part way through our simulation.

In E2 we have the constraint Xi = v

We're just about to generate a value for Xi at random. Given the values assigned to the parents , we see that P(Xi = v | parents) = p .

Now, we know that with stochastic sampling:

- we'll generate "Xi = v" proportion p of the time , and procede .
- And we'll generate a different value proportion 1-p of the time , and the simulation will be wasted.

Instead, always generate Xi = v , but weight the answer by weight "p" to compensate .

Likelihood weighting

Algorithm:

1. Set Nc := 0, Ns := 0
2. Generate a random assignment of all variables that matches E2. This process returns a weight w.
3. Define w to be the probability that this assignment would have been generated instead of an unmatching assignment during its generation in the original algorithm.

   Fact: w is the product of all likelihood factors involved in the generation.
4. If the values in our assignment equal the values in E:
   - Nc := Ns + w
   - If our sample matches E1 then

     Ns := Ns + w
5. Goto 2

Again, Ns / Nc estimates P(E1 | E2)

Case Study I

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.

Questions

- What are the strengths of probabilistic networks compared with propositional logic?
- What are the weaknesses of probabilistic networks compared with propositional logic?
- What are the strengths of probabilistic networks compared with predicate logic?
- What are the weaknesses of probabilistic networks compared with predicate logic?
- (How) could predicate logic and probabilistic networks be combined?
- How do probabilistic methods compare with non-monotonic representations of uncertainty?

What you should know

- Elementary manipulation of probabilities.
- Using Bayes rule to update probabilities of hypothesized models.
- The joint probability distribution and how to marginalize it to derive arbitrary conditional probabilities.
- 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
- The slow (exponential) method for computing arbitrary conditional probabilities.
- The stochastic simulation method and likelihood weighting.

In those cases where your lecturer or these handouts are too incomprehensible, consult the uncertainty handout. Also try Chap 14 and 15 of Russell and Norvik. Elementary probability is handled in numerous "Elementary probability" texts.