Discrete Math Final
Bayesian Belief Networks

John Mill
IS/CS
Spokane Falls CC

May 30, 2012

1 Introduction

One of the reasons we collect statistics is to help us make predictions about the future. In this exercise we will discover one way to efficiently calculate predictions about the future using multiple evidences.

2 Motivation Exercises

Please answer the following questions as we explore the necessicty of Bayes nets.

1. Imagine we have some population $\mathcal{P}$ of which two percent have cancer. Now imagine we have some individual, x from $\mathcal{P}$ or a similarly distributed population. Find $P(x = \textit{has cancer})$, the probability that x has cancer?

2. Exercise 1 shows us that we really need more information in order to provide more useful or detailed diagnosis. Imagine that we have a test, T, that we have administered to $\mathcal{P}$. $T : x \in \mathcal{P} \rightarrow \{1, -1\}$ where 1 indicates that x has tested positive for cancer and -1 indicates that x has tested negative for cancer. Further imagine that T is perfectly accurate. Using the definition of conditional probability, a) find the probability that x has cancer given a positive result from T and b) find the probability that x has cancer given that T returns a negative result.

   Since T is perfectly accurate, the set of people with no cancer and who test positive for it is empty.

3. Hopefully, exercise 2 shows you that with more information, we can provide more accurate predictions. Unfortunately, we are rarely have perfect tests. Imagine T is a bit more realistic. Let us assume that on people with cancer, T is 96 percent accurate. On people without cancer, T is 99 percent accurate. Again, a) find the probability that x has cancer given a positive result from T and b) find the probability that x has cancer given that T returns a negative result.

4. Even given the slightly messy results of the previous example, having some evidence about cancer allowed us to make more accurate predictions. So, what is better than one piece of evidence? How about N pieces of evidence. Imagine we have some condition A we wish to predict and we have evidence $X_1, X_2, ..., X_N$. Compute $P(A|X_1, X_2, ..., X_N)$ using the definition of conditional probability.

5. $P(X_1, X_2, ..., X_N)$ is called a joint probability distribution over the variables $X_1, X_2, ..., X_N$. If the variables are Boolean valued, we need a truth table with 2^N entries. There has to be an

easier way!!! There is. But first, we need some intermediate steps. Show that $P(A, B, C, D) = P(A|B, C, D)P(B|C, D)P(C|D)P(D)$.

6. We have converted a joint probability distribution to the product of conditional probabilities. But we can even make that easier. We will need to introduce the idea of conditional independence. Regular independence says that if you have two events A and B, they are independent iff $P(A, B) = P(A)P(B)$. With conditional independence, we say that two events A and B are *conditionally independent* with respect to event Z iff $P(A, B|Z) = P(A|Z)P(B|Z)$. Now, show that if A and B are conditionally independent with respect to Z, then $P(A|B, Z) = P(A|Z)$.

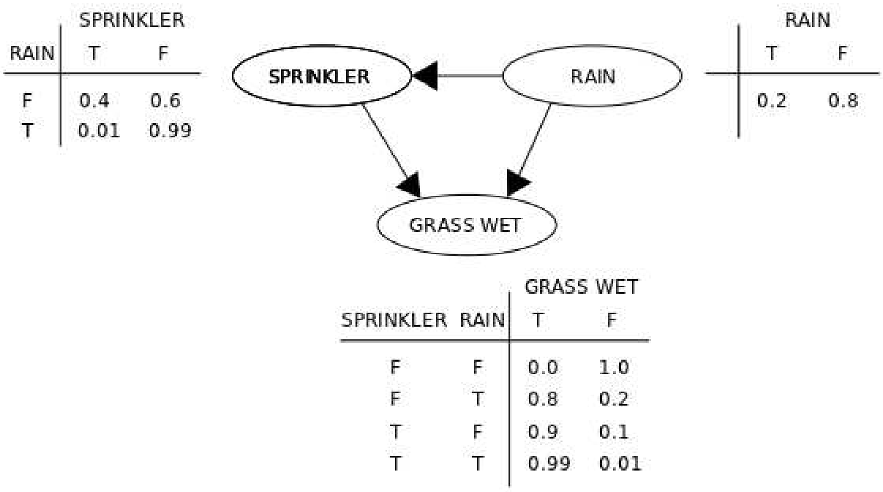
3 Bayesian Belief Networks

So we have learned that having more evidence allows us to make more accurate predictions. However, more evidence leads to more difficult computations of joint probability distributions. To simplify the computation of joint probability distributions, we have learned we can re-write them as a product of conditional probabilities. Further we can reduce the complexity of conditional probabilities with conditional independence. We can remove conditionally independent variables from the condition variables.

A Bayesian belief network is a directed acyclic graph (DAG) to encode conditional dependence. In a belief network there is a vertex for each variable in the joint probability distribution. The set of edges indicate conditional dependence.

3.1 Reasoning with a Bayesian Belief Net

Consider the following image (which I stole from Wikipedia).



This network encodes a belief network for the state of the lawn. We use G for Grass is wet, S for Sprinkler is on, and R for raining.

The process of reasoning with a Bayes net is now pretty simple. When a question is posed of the network, you first convert the question into joint probability distributions. Then convert the joint probability distributions into a chain of conditionals. Then solve the conditionals using the information encoded into the Bayes net.

For example, let us imagine we wish to know the probability that it is raining given that the grass is wet.

$$
\begin{aligned}
P(R=t|G=t) &= \frac{P(R=t,G=t)}{P(G=t)} \\
&= \frac{\begin{array}{c} P(R=t,G=t,S=t)+ \\ P(R=t,G=t,S=f) \end{array}}{\begin{array}{c} P(R=t,G=t,S=t)+ \\ P(R=t,G=t,S=f)+ \\ P(R=f,G=t,S=t)+ \\ P(R=f,G=t,S=f) \end{array}} \\
&= \frac{\begin{array}{c} P(G=t|S=t,R=t)P(S=t|R=t)P(R=t)+ \\ P(G=t|S=f,R=t)P(S=f|R=t)P(R=t) \end{array}}{\begin{array}{c} P(G=t|S=t,R=t)P(S=t|R=t)P(R=t)+ \\ P(G=t|S=f,R=t)P(S=f|R=t)P(R=t)+ \\ P(G=t|S=t,R=f)P(S=t|R=f)P(R=f)+ \\ P(G=t|S=f,R=f)P(S=f|R=f)P(R=f) \end{array}} \\
&= \frac{\begin{array}{c} .99 \times .01 \times .2+ \\ .8 \times .99 \times .2 \end{array}}{\begin{array}{c} .99 \times .01 \times .2+ \\ .8 \times .99 \times .2+ \\ .9 \times .4 \times .8+ \\ 0 \times .6 \times .8 \end{array}} \\
&= \frac{.00198 + .1584}{.00198 + .1584 + .288 + 0} \\
&= \frac{.16038}{.44838} \\
&\approx .357
\end{aligned}
$$

Considering that it only rains 20 percent of the time, when the grass is wet, it is likely due to the sprinkler. Now, I know that this feels like a great deal of work, but let us try this process on a bigger example....

3.2 Exercises

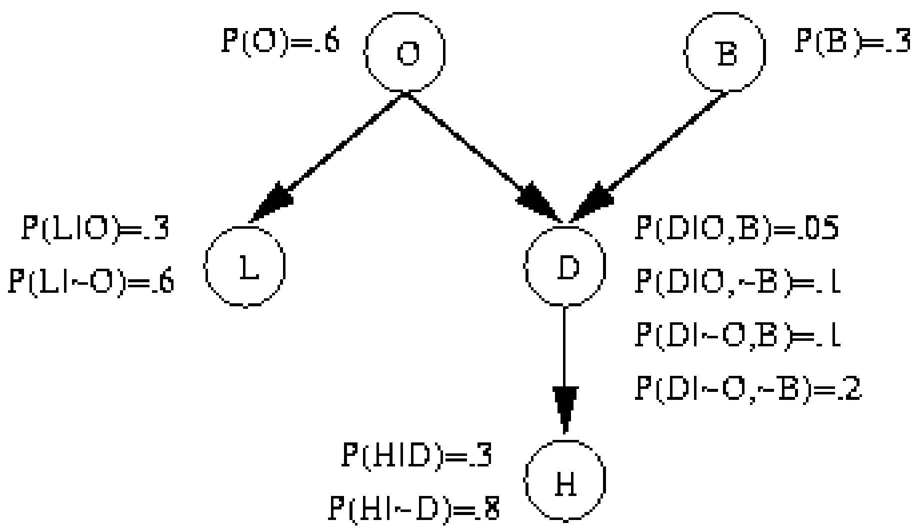
Please answer the following questions.

1. Example From (Charniak, 1991)

   Consider the problem domain in which when I go home I want to know if someone in my family is home before I go in. Let's say I know the following information: (1) When my wife leaves the house, she often (but not always) turns on the outside light. (She also sometimes turns the light on when she's expecting a guest.) (2) When nobody is home, the dog is often left outside. (3) If the dog has bowel-troubles, it is also often left outside. (4) If the dog is outside, I will probably hear it barking (though it might not bark, or I might hear a different dog barking and think it's my dog). Given this information, define the following five Boolean random variables:

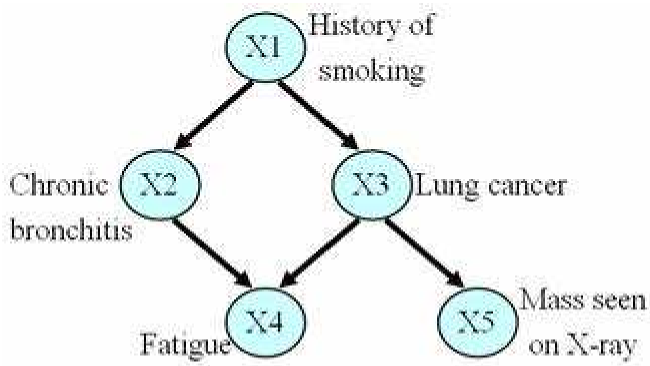
   O: Everyone is Out of the house
   L: The Light is on
   D: The Dog is outside
   B: The dog has Bowel troubles

H: I can Hear the dog barking



What is the probability everyone is out of the house given that the lights are on, the dog has bowel trouble, and I can hear the dog barking?

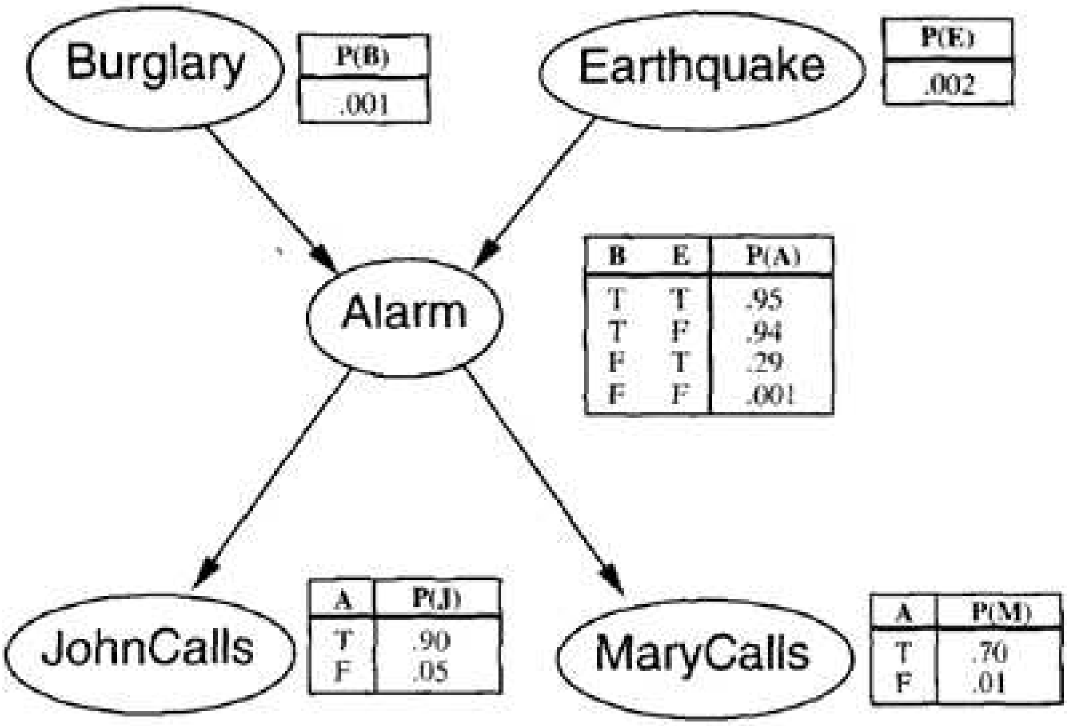
2. Consider this example.....




```
P(X1=no)=0.8
P(X1=yes)=0.2

P(X2=absent | X1=no)=0.95
P(X2=absent | X1=yes)=0.75
P(X2=present | X1=no)=0.05
P(X2=present | X1=yes)=0.25

P(X3=absent | X1=no)=0.99995
P(X3=absent | X1=yes)=0.997
P(X3=present | X1=no)=0.00005
P(X3=present | X1=yes)=0.003

P(X4=absent | X2=absent, X3=absent)=0.95
P(X4=absent | X2=absent, X3=present)=0.5
```

```
P(X4=absent | X2=present, X3=absent)=0.9
P(X4=absent | X2=present, X3=present)=0.25
P(X4=present | X2=absent, X3=absent)=0.05
P(X4=present | X2=absent, X3=present)=0.5
P(X4=present | X2=present, X3=absent)=0.1
P(X4=present | X2=present, X3=present)=0.75

P(X5=absent | X3=absent)=0.98
P(X5=absent | X3=present)=0.4
P(X5=present | X3=absent)=0.02
P(X5=present | X3=present)=0.6
```

Mr Bean has a history of smoking, no history of bronchitis, no fatigue, and a mass seen on his x-ray. Given all that, what is the probability that Mr. Bean has cancer?

3. 

Oh no! John has called and Mary has not! In addition, we know that there is no burglary, and that the alarm was active. Given that, what is the probability that an earth quake has occured?