Bayes Networks

Probability Independence
Assume you're given three random Boolean variables:

Toothache
Cavity
Catch

where Catch is whether the dentist's tool catches a tooth. If we know the distributions of each variable, we can calculate the joint distribution:

P(Cavity, Toothache, Catch)

Assume we have the following values in the distribution:

	toothache		!toothache	
	catch	!catch	catch	!catch
cavity	0.108	0.012	0.072	0.008
!cavity	0.016	0.064	0.144	0.576

P(toothache or cavity) = events of toothache + events of cavity
= .108 + .012 + .016 + .064 + .072 + .008 = .28

P(cavity) = .108 + .012 + .072 + .008

P(cavity | toothache) = P(cavity, toothache)/P(toothache)
=(.108 + .012) /(.108 + .012 + .016 + .064) = .6

P(!cavity | toothache) = (.016 + .064)/(.108 + .016 + .012 + .064) = .4

1/P(toothache) is a normalization constant so that P(Cavity | toothache) distribution sums to 1.

Example:
P(cavity| catch, toothache) = P(cavity, catch, toothache)/P(catch, toothache)
= .108 / (.108 + .016) = .87

not cavity = .016 / (.108 + .016) = .13

The cavity, not cavity needs to sum to 1, and we get that by normalizing.

There is a dependency between these variables, but add a fourth variable, Weather with four values {sun, rain, cloudy, snow}.

We could ask:
P(cloudy | toothache, cavity, catch).

But, dental health doesn't influence the weather, and so

P(cloudy | toothache, cavity, catch) = P(cloudy)

The observed data doesn't influence our belief about the weather. These variables are independent. Independence means:

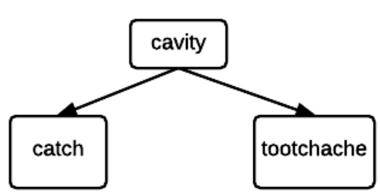
Given random variables a and b that are independent,
P(a | b) = P(a) or P(b | a) = P(b)
and
P(a, b) = P(a)P(b)
Independence reduces the amount of information needed to specify the joint distribution.

Returning to Bayes Rule…
In our Bayes Rule example, we had one symptom, and were asking P(cause | effect), where the effect is the observed symptom. But, what if we want P(cause | effect_1, effect_2, effect_3, … effect_n), where there are multiple pieces of evidence. Identifying independence can simplify our calculations. It requires knowledge of the domain to identify independence.

Example: In the cavity, toothache, catch example, are toothache and catch independent? Not really, if the probe catches a tooth, it probably has a cavity, and that causes a toothache.

But, they are independent in the presence or absence of a cavity. They are connected through cavity, but neither variable has a direct effect on the other. They are conditionally independent.

Graphically, you can think of these three variables like this:



Where the directed edge shows cause -> effect. A single cause can influence a large number of effects, where each of the effects is conditionally independent.

P(cavity | toothache, catch) = αP(toothache | cavity)P(catch | cavity)P(cavity) where

$\alpha = 1/P(\text{toothache})P(\text{catch})$

$P(\text{cause} \mid \text{effect}_i \ldots \text{effect}_n) = \Pi_i P(\text{effect}_i \mid \text{cause})P(\text{cause}) / \Pi i P(\text{effect}_i)$

Bayesian Networks

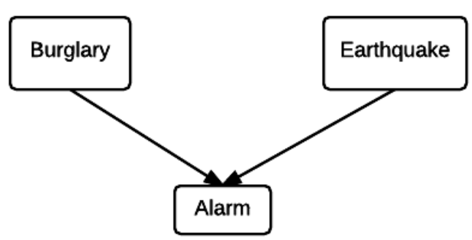
Bayesian Networks, or Bayes nets, provide a systematic way to represent independent and conditionally independent relationships and dependencies. They are used to capturing uncertain knowledge and probabilistic inference.

A Bayes net is a directed graph where each node has probability information.

Properties of Bayes nets

1. Nodes in the network are random variables, discrete or continuous.
2. Directed linked connect pairs of nodes. An arrow from node X to node Y means that X is the parent of Y.
3. Each node X_i has a conditional probability distribution, $P(X_i \mid \text{Parents}(X_i))$ that quantifies the effect of the parents on the node.
4. The graph has no directed cycles.
5. An arrow between node X and node Y means that X has a direct effect on Y.

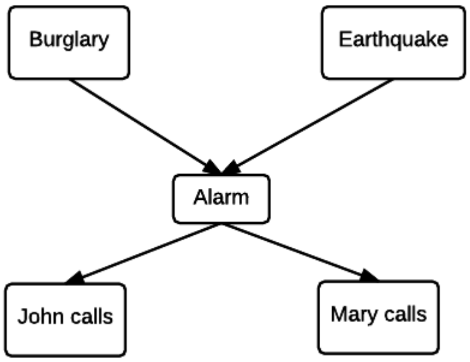
Bayes net example:
A new burglar alarm has been installed and it's fairly reliable, but occasionally goes off when there is an earthquake. The directed graph of Burglary, Alarm, and Earthquake looks like this:



Burglary and earthquake directly affect whether the alarm goes off. There is a dependent relationship, and we could ask P(alarm | burglary) and P(alarm | earthquake) and P(alarm | burglary, earthquake).

You also have two neighbors, John and Mary, who have promised to call you if they hear the alarm. John is reliable and nearly always calls when the alarm sounds, and sometimes he even gets confused and calls when the phone rings. Mary, on the other hand, likes loud music and often misses the alarm and doesn't call.

We can add two more nodes for John Calls and Mary Calls, where those nodes are both dependent on the Alarm node.



There are assumptions in this topology:

1. John and Mary don't see the burglaries or feel the earthquakes, they only respond to the alarm.
2. John and Mary don't talk to each other before calling.

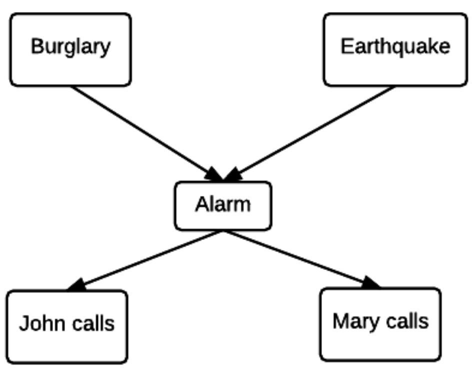
These assumptions are debatable, but the big tasks of designing the Bayes net is deciding what the nodes and links should be.

We can assign some numbers to the nodes.

P(B) = 0.001

P(E) = 0.002

These numbers are all courtesy of UCLA, where probability of earthquake higher than probability of burglary.



B	E	P(a)	P(!a)
t	t	0.95	0.05
t	f	0.94	0.06
f	t	0.29	0.71
f	f	0.001	0.999

A	P(j)
t	0.90
f	0.05

A	P(m)
t	0.70
f	0.01

Example: When there is a burglary, the alarm is 94% reliable. The probability that the alarm will go off is 0.94.

We can use the network to answer questions.

1. The alarm has sounded, but neither a burglary nor an earthquake has occurred and both John and Mary call.

P(j, m, a, !b, !e) = P(j | a) * P(m | a) * P(a | !b, !e) * P(!b) * P(!e)
= 0.90 * 0.70 * 0.001 * 0.999 * 0.998
=0.000628

2. What is the probability that an alarm has gone off given that John has called.

   P(a | j)
3.