Probabilistic Reasoning

Unit # 2

Chain Rule

- Given variables $X_1, X_2, \ldots.., X_n$, the chain rule is
  $P(x_1, x_2, \ldots.., x_n) = P(x_n \mid x_{n-1}, \ldots.., x_2, x_1) \ldots\ldots\ldots$
  $P(x_2 \mid x_1)\, P(x_1)$

Example
(Source Neapolitan, 2009)

- Let S = Sex, H = Height and W = Wage

Case	Sex	Height (inches)	Wage ($)
1	female	64	30,000
2	female	64	30,000
3	female	64	40,000
4	female	64	40,000
5	female	68	30,000
6	female	68	40,000
7	male	64	40,000
8	male	64	50,000
9	male	68	40,000
10	male	68	50,000
11	male	70	40,000
12	male	70	50,000

Example (Cont'd)

s	$P(s)$
female	1/2
male	1/2

h	$P(h)$
64	1/2
68	1/3
70	1/6

w	$P(w)$
30,000	1/4
40,000	1/2
50,000	1/4

The joint distribution of S and H is as follows:

s	h	$P(s,h)$
female	64	1/3
female	68	1/6
female	70	0
male	64	1/6
male	68	1/6
male	70	1/6

Example (Cont'd)
Marginal and Joint Distribution

s \ h	64	68	70	Distribution of S
female	$1/3$	$1/6$	0	$1/2$
male	$1/6$	$1/6$	$1/6$	$1/2$
Distribution of H	$1/2$	$1/3$	$1/6$	

Example (Cont'd)
Chain Rule

$$P(s,h,w) = P(w|h,s)P(h|s)P(s).$$

There are eight combinations of values of the three random variables. The table that follows shows that the equality holds for two of the combination.

s	h	w	$P(s,h,w)$	$P(w\|h,s)P(h\|s)P(s)$
female	64	30,000	$\frac{1}{6}$	$\left(\frac{1}{2}\right)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right)=\frac{1}{6}$
female	64	40,000	$\frac{1}{12}$	$\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)\left(\frac{1}{2}\right)=\frac{1}{12}$

It is left as an exercise to show that the equality holds for the other six combinations.

Exercise 1
(Source: Neapolitan, 2009)

- Are H and W independent?
- Are H and W conditionally independent given S?

Exercise 2
(Source: Neapolitan, 2009)

- A forgetful nurse is supposed to give Mr. Nguyen a pill each day. The probability that the nurse will forget to give the pill on a given day is 0.3. If Mr. Nguyen receives the pill, the probability he will die is 0.1. If he does not receive the pill, the probability he will die is 0.8. Mr. Nguyen died today.
- Use Bayes' Theorem to compute the probability that the nurse forgot to give him the pill.

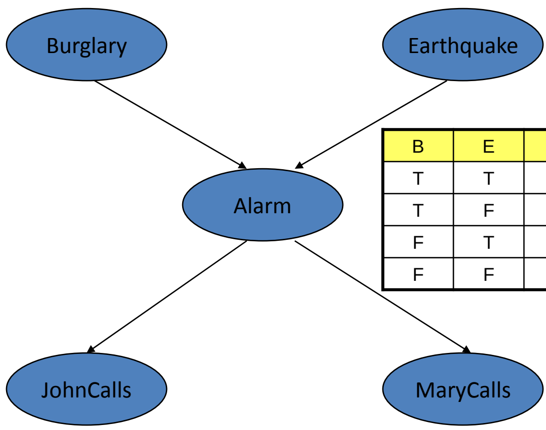
Exercise 3 (Source: Neapolitan, 2009)

- An oil well might be drilled on Professor Neapolitan's farm in Texas. Based on what has happened on similar farms, we judge the probability of oil being present to be 0.5, the probability of only natural gas being present to be 0.2, and the probability of neither being present to be 0.3.
- If oil is present, a geological test will give a positive result with probability 0.9; if only natural gas is present, it will give a positive result with probability 0.3; and if neither is present, the test will be positive with probability 0.1.
- Suppose the test comes back positive. Use Bayes' Theorem to compute the probability that oil is present.

Earthquake Example (Pearl)

P(B)
0.1

P(E)
0.02



B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|A)
T	0.9
F	0.05

A	P(M|A)
T	0.7
F	0.01

Bayesian Networks

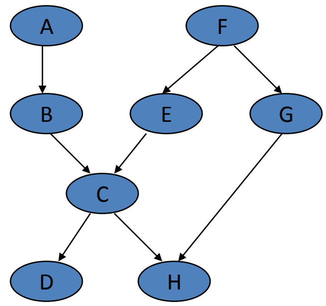
- Two problems with using full joint distributions for probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to estimate anything empirically about more than a few variables at a time
- Bayesian Networks are a technique for describing complex joint distributions using a bunch of simple, local distributions
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions

Bayesian Networks

- A BN is a Directed Acyclic Graph (DAG) in which:
  - A set of random variables makes up the nodes in the network.
  - A set of directed links or arrows connects pairs of nodes.
  - Each node has a conditional probability table that quantifies the effects the parents have on the node.
- The intuitive meaning of an arrow from a parent to a child is that the parent directly influences the child.
- The direction of this influence is often taken to represent casual influence.
- These influences are quantified by conditional probabilities.

Bayesian Networks (Cont'd)

- A problem domain is modeled by a list of variables $X_1, X_2, \ldots, X_n$.
- Knowledge about the problem domain is represented by a joint probability $P(X_1, X_2, \ldots, X_n)$.
- General probability distribution of 8 variables with 2 states each has 2^8 = 256 possible values and $2^8 - 1$ probabilities need to be specified.
- Assumes that each node is conditionally independent of all its non-descendants given its parents.
- Product of all conditional probabilities is the joint probability of all variables.
  - $P(X_1, X_2, \ldots, X_n = \prod P(X_i \mid \text{parents}(X_i))$



18 probabilities are required to specify the joint distribution