Bayes Nets

- The **prior** of a random variable is its distribution before any information is observed. The **likelihood** of an observation is its distribution given the values of the random variables it depends on. The **posterior** of a random variable is its distribution given some observation. **Bayes' rule** relates these three quantities: $P(B|A) \propto P(B)P(A|B)$; the posterior is proportional to the prior times the likelihood.

- The **chain rule** of conditional probability:
  $P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i|X_1, ..., X_{i-1})$.

- Two random variables X and Y are **conditionally independent** given a random variable Z if $P(X|Y, Z) = P(X|Z)$ (or equivalently, if $P(Y|X, Z) = P(Y|Z)$).

- A **Bayes net** is a directed acyclic graph in which each node represents a random variable and has a CPT for that variable given the variables in its parent nodes. The purpose of Bayes nets is to compress the representation of a distribution by taking advantage of conditional independence, because the full joint distribution of all the variables is exponentially large. Instead of specifying all d^n values in the joint distribution (where d is the number of values in the domain of a random variable and n is the number of random variables), it only needs to specify at most $n \cdot d^k$ values (where k is the maximum number of parents per node).

- A node's **Markov blanket** is its parents, its children, and its children's parents. A node is conditionally independent of all other nodes given its Markov blanket.

- A **query variable** is one whose distribution is to be computed. An **evidence variable** is one whose value is known. A **hidden variable** is a non-query variable whose value is not known, so it has to be marginalized out when computing the query distribution.

- The **enumeration algorithm** for inference in Bayes nets uses recursive depth-first enumeration to compute probabilities by multiplying and summing entries from the nodes' CPTs. Its running time is $O(d^n)$.

- The **variable-elimination algorithm** for inference in Bayes nets improves on enumeration by computing with **factors** (which are similar to CPTs) to avoid recomputation. The **pointwise product** of a set of factors multiplies corresponding entries:

A	B	$f_0(A,B)$
0	0	0.01
0	1	0.02
1	0	0.03
1	1	0.04

$\times$

B	C	$f_1(B,C)$
0	0	0.11
0	1	0.12
1	0	0.13
1	1	0.14

$\times$

A	$f_2(A)$
0	0.21
1	0.22

$=$

A	B	C	$f_3(A,B,C)$
0	0	0	$0.01 \times 0.11 \times 0.21$
0	0	1	$0.01 \times 0.12 \times 0.21$
0	1	0	$0.02 \times 0.13 \times 0.21$
0	1	1	$0.02 \times 0.14 \times 0.21$
1	0	0	$0.03 \times 0.11 \times 0.22$
1	0	1	$0.03 \times 0.12 \times 0.22$
1	1	0	$0.04 \times 0.13 \times 0.22$
1	1	1	$0.04 \times 0.14 \times 0.22$

A variable can be **summed out** of a factor by summing each set of entries in which the only thing that varies is the value of that variable. For example, if we want to sum out C from the factor f_3 we just computed:

A	B	C	$f_3(A,B,C)$
0	0	0	0.000231
0	0	1	0.000252
0	1	0	0.000546
0	1	1	0.000588
1	0	0	0.000726
1	0	1	0.000792
1	1	0	0.001144
1	1	1	0.001232

$\rightarrow$

A	B	$f_4(A,B)$
0	0	$0.000231 + 0.000252$
0	1	$0.000546 + 0.000588$
1	0	$0.000726 + 0.000792$
1	1	$0.001144 + 0.001232$

Exercises

1. (AIMA 14.11) In your local nuclear power station, there is an alarm that senses when a temperature gauge exceeds a given threshold. The gauge measures the temperature of the core. Consider the Boolean variables A (alarm sounds), F_A (alarm is faulty), and F_G (gauge is faulty) and the multivalued nodes G (gauge reading) and T (actual core temperature).

   (a) Draw a Bayesian network for this domain, given that the gauge is more likely to fail when the core temperature gets too high.

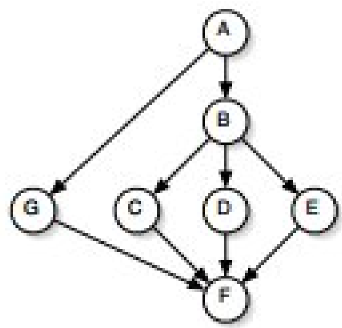
   (b) Is your network a polytree? Why or why not? (A network is called a polytree if there is at most one undirected path between any two of its nodes.)

   (c) Suppose there are just two possible actual and measured temperatures, normal and high; the probability that the gauge gives the correct temperature is x when it is working, but y when it is faulty. Give the conditional probability table associated with G.

   (d) Suppose the alarm works correctly unless it is faulty, in which case it never sounds. Give the conditional probability table associated with A.

   (e) Suppose the alarm and gauge are working and the alarm sounds. Calculate an expression for the probability that the temperature of the core is too high, in terms of the various conditional probabilities in the network.

2. Consider the network shown below:



(a) Assuming the nodes are binary, how many parameters are required to specify the CPTs?

(b) Is F independent of A given B?

(c) Is G independent of E given A and F?

(d) Is B independent of F given C, D, and E?

(e) Which variables are irrelevant to the query $\Pr(d|c)$? (d and c are specific values of variables D and C)

(f) Give an expression for $\Pr(d|c)$ in terms of parameters stored in the network.

(g) Which factors are created by variable elimination using order A, B, E, F, G?

(h) Is there another elimination order with a smaller largest factor?