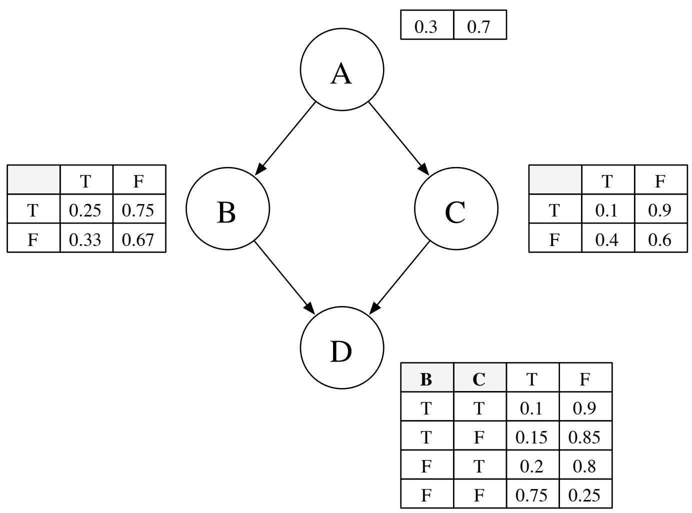
Inference in Bayesian Networks

Ajit Singh

December 2, 2005

1 Basic Idea

The basic idea behind inference in Bayesian networks is to observe that when a joint distribution is factored, each of the factors can be distributed across a sum. For example, on the following network



$$P(A) = \sum_B \sum_C \sum_D P(A, B, C, D) \tag{1}$$

$$= \sum_B \sum_C \sum_D P(A)P(B|A)P(C|A)P(D|B, C) \tag{2}$$

$$= \sum_B P(A)P(B|A) \sum_C P(C|A) \sum_D P(D|B, C) \tag{3}$$

While these three equations all compute the same marginal distribution, equation 2 requires first multiplying all the local conditional probabilities together to recover the joint.

Multiplying all these tables together gives us a single table of size exponential in the number of variables n. If we distribute the factors, as in equation 3, we only multiply a few factors at a time and the memory (and time) savings can be substantial.

2 Operations on Factors

2.1 Factor Notation

A convenient way to view a probability distribution in a Bayesian network is as a factor, or function mapping an assignment of variables to a real number. In our 4-node network, the local conditional probability table $P(D|B,C)$ can also be denoted $f(B,C,D)$. For example, $f(B=F,C=T,D=F)=0.8$ and $f(B=F,C=T,D=T)=0.2$.

Formally, let the n variables in the network be denoted $X_1, X_2, \ldots, X_n$. Let factors be denoted by lower case letters. Let $\textsc{Val}(X_i)$ denote the values that X_i takes.

Multiply Factors : If $X, Y \subseteq \{X_1, \ldots, X_n\}$ and f, g, h are factors then $f(X \cup Y) = g(X)h(Y)$. This is similar to relational join in databases.

Evidence : Think of it as partial assignment. If we observe that $X_i = T$ then in all the functions that involve X_i, we fix the value of X_i to true.

Marginalization : If $X = \{X_{i_1}, X_{i_2}, \ldots, X_{i_k}\} \subseteq \{X_1, \ldots, X_n\}$ then $g(X - \{X_{i_1}\}) = \sum_{x_{i_1} \in \textsc{Val}(X_{i_1})} f(X_{i_1} = x_{i_1}, X_{i_2}, \ldots, X_{i_k})$.