Regular Discussion 7 Solutions

Probability

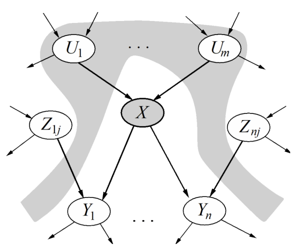
A **random variable** represents an event whose outcome is unknown. A **probability distribution** is an assignment of weights to outcomes A **joint distribution** over discrete random variables is a table of probabilities which captures the likelihood of each possible **outcome**, also known as an **assignment** of values to the random variables.

To write that random variables X and Y are **marginally independent**, we write $X \perp\!\!\!\perp Y$. To write that random variables X and Y are **conditionally independent** given another random variable Z, we write $X \perp\!\!\!\perp Y|Z$.

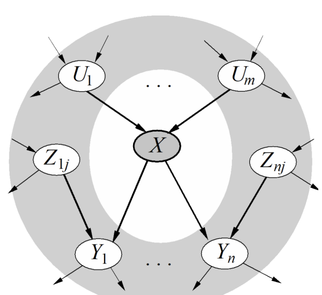
Bayesian Network Representation

In a Bayesian network, rather than storing information in a giant table, probabilities are instead distributed across a large number of smaller local probability tables along with a **directed acyclic graph** (DAG) which captures the relationships between variables. Thus, if we have a node representing variable X, we store $P(X|A_1, A_2, ..., A_N)$, where $A_1, ..., A_N$ are the parents of X.

- **Each node is conditionally independent of all its ancestor nodes (non-descendents) in the graph, given all of its parents**.

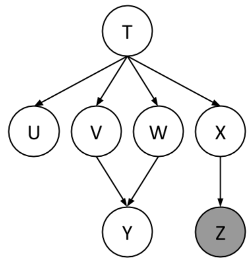


- **Each node is conditionally independent of all other variables given its Markov blanket.** A variable's Markov blanket consists of parents, children, children's other parents.



1 Variable Elimination

Using the same Bayes Net (shown below), we want to compute $P(Y \mid +z)$. All variables have binary domains. Assume we run variable elimination to compute the answer to this query, with the following variable elimination ordering: X, T, U, V, W.



Complete the following description of the factors generated in this process:

After inserting evidence, we have the following factors to start out with:

$$P(T), P(U|T), P(V|T), P(W|T), P(X|T), P(Y|V,W), P(+z|X)$$

(a) When eliminating X we generate a new factor f_1 as follows, which leaves us with the factors:

$$f_1(+z|T) = \sum_x P(x|T)P(+z|x) \qquad P(T), P(U|T), P(V|T), P(W|T), P(Y|V,W), f_1(+z|T)$$

(b) When eliminating T we generate a new factor f_2 as follows, which leaves us with the factors:

$$f_2(U,V,W,+z) = \sum_t P(t)P(U|t)P(V|t)P(W|t)f_1(+z|t) \qquad P(Y|V,W), f_2(U,V,W,+z)$$

(c) When eliminating U we generate a new factor f_3 as follows, which leaves us with the factors:

$$f_3(V,W,+z) = \sum_u f_2(u,V,W,+z) \qquad P(Y|V,W), f_3(V,W,+z)$$

Note that U could have just been deleted from the original graph, because $\sum_u P(U|t) = 1$. We can see this in the graph: we can remove any leaf node that is not a query variable or an evidence variable.

(d) When eliminating V we generate a new factor f_4 as follows, which leaves us with the factors:

$$f_4(W,Y,+z) = \sum_v f_3(v,W,+z)P(Y|v,W) \qquad f_4(W,Y,+z)$$

(e) When eliminating W we generate a new factor f_5 as follows, which leaves us with the factors:

$$f_5(Y,+z) = \sum_w f_4(w,Y,+z) \qquad f_5(Y,+z)$$

(f) How would you obtain $P(Y \mid +z)$ from the factors left above:
Simply renormalize $f_5(Y, +z)$ to obtain $P(Y \mid +z)$. Concretely,

$$P(y \mid +z) = \frac{f_5(y, +z)}{\sum_{y'} f_5(y', +z)}$$

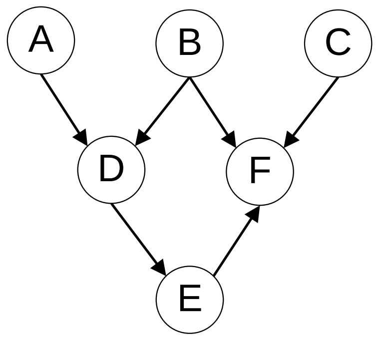
(g) What is the size of the largest factor that gets generated during the above process?
$f_2(U, V, W, +z)$. This contains 3 unconditioned variables, so it will have $2^3 = 8$ probability entries (U, V, W are binary variables, and we only need to store the probability for $+z$ for each possible setting of these variables).

(m) Does there exist a better elimination ordering (one which generates smaller largest factors)?
Yes. One such ordering is X, U, T, V, W. All factors generated with this ordering contain at most 2 unconditioned variables, so the tables will have at most $2^2 = 4$ probability entries (as all variables are binary).

2 Bayes Nets

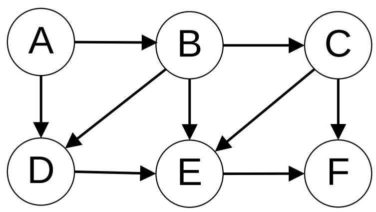
(a) For the following graphs, explicitly state the minimum size set of edges that must be removed such that the corresponding independence relations are guaranteed to be true.

Marked the removed edges with an 'X' on the graphs.



$$A \perp\!\!\!\perp B \mid F$$
$$A \perp\!\!\!\perp F \mid D$$
$$B \perp\!\!\!\perp C$$

(i) AD



$$A \perp\!\!\!\perp D \mid B$$
$$A \perp\!\!\!\perp F \mid C$$
$$C \perp\!\!\!\perp D \mid B$$

(ii) AD, (EF OR AB)

(b) You're performing variable elimination over a Bayes Net with variables A, B, C, D, E. So far, you've finished joining over (but not summing out) C, when you realize you've lost the original Bayes Net!

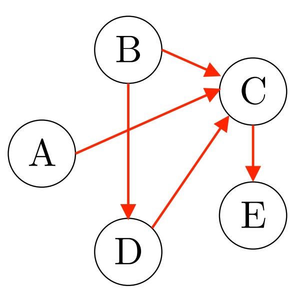
Your current factors are $f(A), f(B), f(B, D), f(A, B, C, D, E)$. Note: these are factors, NOT joint distributions. You don't know which variables are conditioned or unconditioned.

(i) What's the smallest number of edges that could have been in the original Bayes Net? Draw out one such Bayes Net below.

Number of edges = 5

The original Bayes net must have had 5 factors, 1 for each node. f(A) and f(B) must have corresponded to nodes A and B, and indicate that neither A nor B have any parents. f(B, D), then, must correspond to node D, and indicates that D has only B as a parent. Since there is only one factor left, f(A, B, C, D, E), for the nodes C and E, those two nodes must have been joined while you were joining C. This implies two things: 1) E must have had C as a parent, and 2) every other node must have been a parent of either C or E.

The below figure is one possible solution that uses the fewest possible edges to satisfy the above.



(ii) What's the largest number of edges that could have been in the original Bayes Net? Draw out one such Bayes Net below.

Number of edges = 8

The constraints are the same as outlined in part i). To maximize the number of edges, we make each of A, B, and D a parent of both C and E, as opposed to a parent of one of them.

The below figure is the only possible solution.