# Regular Discussion 7 Solutions

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**.
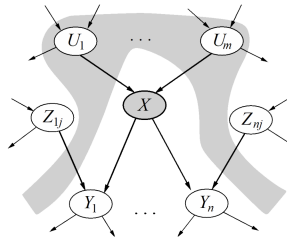

- **Each node is conditionally independent of all other variables given its Markov blanket.** A variable's Markov blanket consists of parents, children, children's other parents.
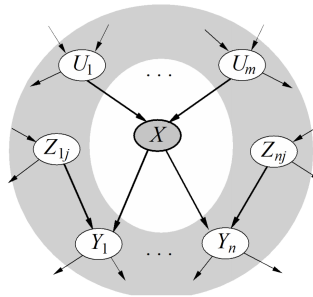
Variable Elimination

An alternate approach to inference by enumeration, where we eliminate variables one by one in order to reduce the maximum size of any factor generated.

Suppose we have the following Bayes Net with factors $P(T), P(C|T), P(S|T)$, and $P(E|C,S)$, and we want to compute $P(T| + e)$.
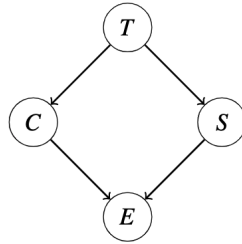
We proceed by eliminating the variables C and S:

1. Join (multiply) all the factors involving C, forming $P(C, +e|T, S) = P(C|T) \cdot P(+e|C, S)$.

2. Sum out C from this new factor, leaving us with a new factor $P(+e|T, S)$.

3. Join all factors involving S, forming $P(+e, S|T) = P(S|T) \cdot P(+e|T, S)$.

4. Sum out S, yielding $P(+e|T)$.

5. Join the remaining two factors $P(T)$ and $P(+e|T)$ and normalize to get $P(T| + e)$.

Sampling

Suppose we want to evaluate $P(Q|E)$ where Q are the query variables and E are the evidence variables.

Prior Sampling: Draw samples from the Bayes net by sampling the parents and then sampling the children given the parents. $P(Q|E) \approx \frac{count(QandE)}{count(E)}$.

Rejection Sampling: Like prior sampling, but ignore all samples that are inconsistent with the evidence.

Likelihood Weighting: Fix the evidence variables, and weight each sample by the probability of the evidence variables given their parents.

Gibbs Sampling:

1. Fix evidence.

2. Initialize other variables randomly

3. Repeat:
   (a) Choose non-evidence variable X.
   (b) Resample X from $P(X|markovblanket(X))$

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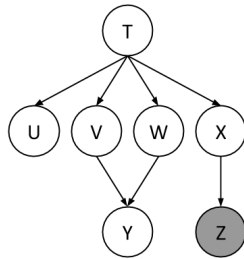
Complete the following description of the factors generated in this process:

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 $\sum_u P(U = u|t) = 1$. 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:

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

3

(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).

Q2. Bayes' Nets Sampling

Assume the following Bayes' net, and the corresponding distributions over the variables in the Bayes' net:

$$A \rightarrow B \rightarrow C \rightarrow D$$

$P(A)$	
$-a$	$3/4$
$+a$	$1/4$

$P(B\|A)$		
$-a$	$-b$	$2/3$
$-a$	$+b$	$1/3$
$+a$	$-b$	$4/5$
$+a$	$+b$	$1/5$

$P(C\|B)$		
$-b$	$-c$	$1/4$
$-b$	$+c$	$3/4$
$+b$	$-c$	$1/2$
$+b$	$+c$	$1/2$

$P(D\|C)$		
$-c$	$-d$	$1/8$
$-c$	$+d$	$7/8$
$+c$	$-d$	$5/6$
$+c$	$+d$	$1/6$
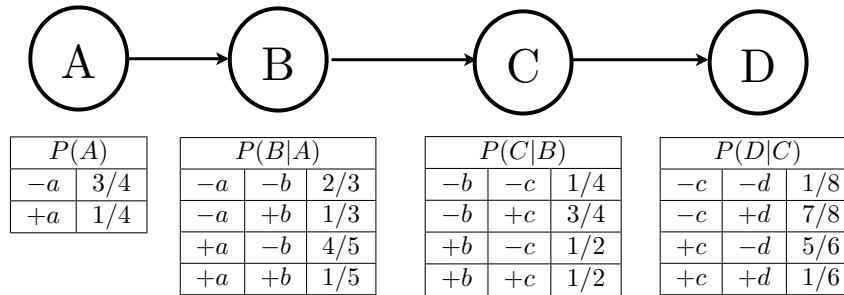
(a) You are given the following samples:

$+a$	$+b$	$-c$	$-d$
$+a$	$-b$	$+c$	$-d$
$-a$	$+b$	$+c$	$-d$
$-a$	$-b$	$+c$	$-d$

$+a$	$-b$	$-c$	$+d$
$+a$	$+b$	$+c$	$-d$
$-a$	$+b$	$-c$	$+d$
$-a$	$-b$	$+c$	$-d$

(i) Assume that these samples came from performing Prior Sampling, and calculate the sample estimate of $P(+c)$.

5/8

(ii) Now we will estimate $P(+c \mid +a, -d)$. Above, clearly cross out the samples that would **not** be used when doing Rejection Sampling for this task, and write down the sample estimate of $P(+c \mid +a, -d)$ below.

2/3

(b) Using Likelihood Weighting Sampling to estimate $P(-a \mid +b, -d)$, the following samples were obtained. Fill in the weight of each sample in the corresponding row.

Sample   **Weight**

$-a$ $+b$ $+c$ $-d$ $P(+b \mid -a)P(-d \mid +c) = 1/3 * 5/6 = 5/18 = 0.277$

$+a$ $+b$ $+c$ $-d$ $P(+b \mid +a)P(-d \mid +c) = 1/5 * 5/6 = 5/30 = 1/6 = 0.17$

$+a$ $+b$ $-c$ $-d$ $P(+b \mid +a)P(-d \mid -c) = 1/5 * 1/8 = 1/40 = 0.025$

$-a$ $+b$ $-c$ $-d$ $P(+b \mid -a)P(-d \mid -c) = 1/3 * 1/8 = 1/24 = 0.042$

(c) From the weighted samples in the previous question, estimate $P(-a \mid +b, -d)$.

$\frac{5/18 + 1/24}{5/18 + 5/30 + 1/40 + 1/24} = 0.625$

(d) Which query is better suited for likelihood weighting, $P(D \mid A)$ or $P(A \mid D)$? Justify your answer in one sentence.

$P(D \mid A)$ is better suited for likelihood weighting sampling, because likelihood weighting conditions only on upstream evidence.

(e) Recall that during Gibbs Sampling, samples are generated through an iterative process.

Assume that the only evidence that is available is $A = +a$. Clearly fill in the circle(s) of the sequence(s) below that could have been generated by Gibbs Sampling.

	Sequence 1			
$1:$	$+a$	$-b$	$-c$	$+d$
$2:$	$+a$	$-b$	$-c$	$+d$
$3:$	$+a$	$-b$	$+c$	$+d$

	Sequence 2			
$1:$	$+a$	$-b$	$-c$	$+d$
$2:$	$+a$	$-b$	$-c$	$-d$
$3:$	$-a$	$-b$	$-c$	$+d$

	Sequence 3			
$1:$	$+a$	$-b$	$-c$	$+d$
$2:$	$+a$	$-b$	$-c$	$-d$
$3:$	$+a$	$+b$	$-c$	$-d$

	Sequence 4			
$1:$	$+a$	$-b$	$-c$	$+d$
$2:$	$+a$	$-b$	$-c$	$-d$
$3:$	$+a$	$+b$	$-c$	$+d$

Gibbs sampling updates one variable at a time and never changes the evidence.

The first and third sequences have at most one variable change per row, and hence could have been generated from Gibbs sampling. In sequence 2, the evidence variable is changed. In sequence 4, the second and third samples have both B and D changing.