CS 188 Summer 2019 | Introduction to Artificial Intelligence

Written HW 4 Sol.

Self-assessment due: Tuesday 7/30/2018 at 11:59pm (submit via Gradescope)

Q1. Probability

(a) For the following questions, you will be given a set of probability tables and a set of conditional independence assumptions. Given these tables and independence assumptions, write an expression for the requested probability tables. Keep in mind that your expressions cannot contain any probabilities other than the given probability tables. If it is not possible, mark "Not possible."

(i) Using probability tables $\mathbf{P}(\mathbf{A}), \mathbf{P}(\mathbf{A} \mid \mathbf{C}), \mathbf{P}(\mathbf{B} \mid \mathbf{C}), \mathbf{P}(\mathbf{C} \mid \mathbf{A}, \mathbf{B})$ and no conditional independence assumptions, write an expression to calculate the table $\mathbf{P}(\mathbf{A}, \mathbf{B} \mid \mathbf{C})$.

$\mathbf{P}(\mathbf{A}, \mathbf{B} \mid \mathbf{C}) =$ ________________ ● Not possible.

(ii) Using probability tables $\mathbf{P}(\mathbf{A}), \mathbf{P}(\mathbf{A} \mid \mathbf{C}), \mathbf{P}(\mathbf{B} \mid \mathbf{A}), \mathbf{P}(\mathbf{C} \mid \mathbf{A}, \mathbf{B})$ and no conditional independence assumptions, write an expression to calculate the table $\mathbf{P}(\mathbf{B} \mid \mathbf{A}, \mathbf{C})$.

$\mathbf{P}(\mathbf{B} \mid \mathbf{A}, \mathbf{C}) = \frac{P(A)\ P(B|A)\ P(C|A,B)}{\sum_b P(A)\ P(B|A)\ P(C|A,B)}$ ○ Not possible.

(iii) Using probability tables $\mathbf{P}(\mathbf{A} \mid \mathbf{B}), \mathbf{P}(\mathbf{B}), \mathbf{P}(\mathbf{B} \mid \mathbf{A}, \mathbf{C}), \mathbf{P}(\mathbf{C} \mid \mathbf{A})$ and conditional independence assumption $\mathbf{A} \perp\!\!\!\perp \mathbf{B}$, write an expression to calculate the table $\mathbf{P}(\mathbf{C})$.

$\mathbf{P}(\mathbf{C}) = \sum_a P(A \mid B)\ P(C \mid A)$ ○ Not possible.

(iv) Using probability tables $\mathbf{P}(\mathbf{A} \mid \mathbf{B}, \mathbf{C}), \mathbf{P}(\mathbf{B}), \mathbf{P}(\mathbf{B} \mid \mathbf{A}, \mathbf{C}), \mathbf{P}(\mathbf{C} \mid \mathbf{B}, \mathbf{A})$ and conditional independence assumption $\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C}$, write an expression for $\mathbf{P}(\mathbf{A}, \mathbf{B}, \mathbf{C})$.

$\mathbf{P}(\mathbf{A}, \mathbf{B}, \mathbf{C}) =$ ________________ ● Not possible.

(b) For each of the following equations, select the *minimal set* of conditional independence assumptions necessary for the equation to be true.

(i) $\mathbf{P}(\mathbf{A}, \mathbf{C}) = \mathbf{P}(\mathbf{A} \mid \mathbf{B})\ \mathbf{P}(\mathbf{C})$

- ■ $A \perp\!\!\!\perp B$
- ☐ $A \perp\!\!\!\perp B \mid C$
- ■ $A \perp\!\!\!\perp C$
- ☐ $A \perp\!\!\!\perp C \mid B$
- ☐ $B \perp\!\!\!\perp C$
- ☐ $B \perp\!\!\!\perp C \mid A$
- ☐ No independence assumptions needed.

(ii) $\mathbf{P}(\mathbf{A} \mid \mathbf{B}, \mathbf{C}) = \frac{\mathbf{P}(\mathbf{A})\ \mathbf{P}(\mathbf{B}|\mathbf{A})\ \mathbf{P}(\mathbf{C}|\mathbf{A})}{\mathbf{P}(\mathbf{B}|\mathbf{C})\ \mathbf{P}(\mathbf{C})}$

- ☐ $A \perp\!\!\!\perp B$
- ☐ $A \perp\!\!\!\perp B \mid C$
- ☐ $A \perp\!\!\!\perp C$
- ☐ $A \perp\!\!\!\perp C \mid B$
- ☐ $B \perp\!\!\!\perp C$
- ■ $B \perp\!\!\!\perp C \mid A$
- ☐ No independence assumptions needed.

(iii) $\mathbf{P}(\mathbf{A}, \mathbf{B}) = \sum_{\mathbf{c}} \mathbf{P}(\mathbf{A} \mid \mathbf{B}, \mathbf{c})\ \mathbf{P}(\mathbf{B} \mid \mathbf{c})\ \mathbf{P}(\mathbf{c})$

- ☐ $A \perp\!\!\!\perp B$
- ☐ $A \perp\!\!\!\perp B \mid C$
- ☐ $A \perp\!\!\!\perp C$
- ☐ $A \perp\!\!\!\perp C \mid B$
- ☐ $B \perp\!\!\!\perp C$
- ☐ $B \perp\!\!\!\perp C \mid A$
- ■ No independence assumptions needed.

(iv) $\mathbf{P}(\mathbf{A}, \mathbf{B} \mid \mathbf{C}, \mathbf{D}) = \mathbf{P}(\mathbf{A} \mid \mathbf{C}, \mathbf{D})\ \mathbf{P}(\mathbf{B} \mid \mathbf{A}, \mathbf{C}, \mathbf{D})$

- ☐ $A \perp\!\!\!\perp B$
- ☐ $A \perp\!\!\!\perp B \mid C$
- ☐ $A \perp\!\!\!\perp B \mid D$
- ☐ $C \perp\!\!\!\perp D$
- ☐ $C \perp\!\!\!\perp D \mid A$
- ☐ $C \perp\!\!\!\perp D \mid B$
- ■ No independence assumptions needed.

(c) **(i)** Mark **all** expressions that are equal to $\mathbf{P(A \mid B)}$, **given no independence assumptions**.

- ☐ $\sum_c P(A \mid B, c)$
- ■ $\sum_c P(A, c \mid B)$
- ☐ $\frac{P(B|A)\ P(A|C)}{\sum_c P(B,c)}$
- ■ $\frac{\sum_c P(A,B,c)}{\sum_c P(B,c)}$
- ☐ $\frac{P(A,C|B)}{P(C|B)}$
- ☐ $\frac{P(A|C,B)\ P(C|A,B)}{P(C|B)}$
- ☐ None of the provided options.

(ii) Mark **all** expressions that are equal to $\mathbf{P(A, B, C)}$, given that $\mathbf{A \perp\!\!\!\perp B}$.

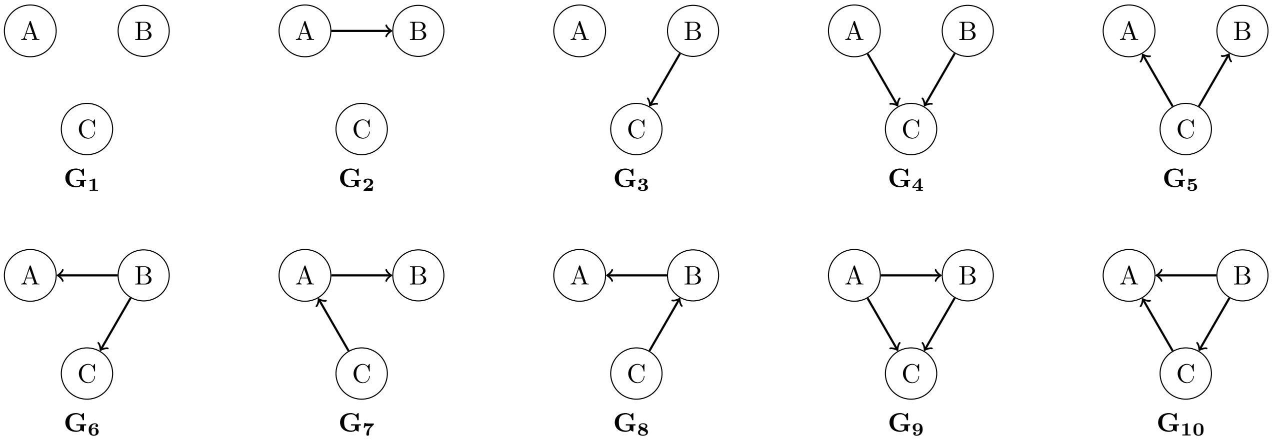
- ☐ $P(A \mid C)\ P(C \mid B)\ P(B)$
- ■ $P(A)\ P(B)\ P(C \mid A, B)$
- ☐ $P(C)\ P(A \mid C)\ P(B \mid C)$
- ☐ $P(A)\ P(C \mid A)\ P(B \mid C)$
- ■ $P(A)\ P(B \mid A)\ P(C \mid A, B)$
- ■ $P(A, C)\ P(B \mid A, C)$
- ☐ None of the provided options.

(iii) Mark **all** expressions that are equal to $\mathbf{P(A, B \mid C)}$, given that $\mathbf{A \perp\!\!\!\perp B \mid C}$.

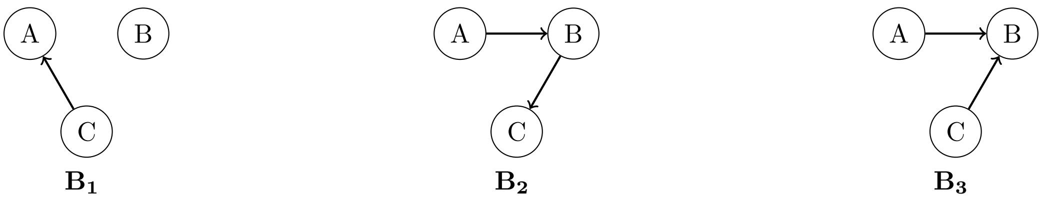
- ■ $P(A \mid C)\ P(B \mid C)$
- ☐ $\frac{P(A)\ P(B|A)\ P(C|A,B)}{\sum_c P(A,B,c)}$
- ☐ $P(A \mid B)\ P(B \mid C)$
- ☐ $\frac{P(C)\ P(B|C)\ P(A|C)}{P(C|A,B)}$
- ☐ $\frac{\sum_c P(A,B,c)}{P(C)}$
- ■ $\frac{P(C,A|B)\ P(B)}{P(C)}$
- ☐ None of the provided options.

Q2. Bayes' Nets: Representation

Assume we are given the following ten Bayes' nets, labeled $\mathbf{G_1}$ to $\mathbf{G_{10}}$:



Assume we are also given the following three Bayes' nets, labeled $\mathbf{B_1}$ to $\mathbf{B_3}$:



Before we go into the questions, let's enumerate all of the (conditional) independence assumptions encoded in all the Bayes' nets above. They are:

- $\mathbf{G_1}$: $AB; AB|C; AC; AC|B; BC; BC|A$
- $\mathbf{G_2}$: $AC; AC|B; BC; BC|A$
- $\mathbf{G_3}$: $AB; AB|C; AC; AC|B$
- $\mathbf{G_4}$: AB
- $\mathbf{G_5}$: $AB|C$
- $\mathbf{G_6}$: $AC|B$
- $\mathbf{G_7}$: $BC|A$
- $\mathbf{G_8}$: $AC|B$
- $\mathbf{G_9}$: $\emptyset$
- $\mathbf{G_{10}}$: $\emptyset$
- $\mathbf{B_1}$: $AB; AB|C; BC; BC|A$
- $\mathbf{B_2}$: $AC|B$
- $\mathbf{B_3}$: AC

(a) Assume we know that a joint distribution $\mathbf{d_1}$ (over $\mathbf{A}, \mathbf{B}, \mathbf{C}$) can be represented by Bayes' net $\mathbf{B_1}$. Mark all of the following Bayes' nets that are guaranteed to be able to represent $\mathbf{d_1}$.

☐ $\mathbf{G_1}$ ☐ $\mathbf{G_2}$ ☐ $\mathbf{G_3}$ ■ $\mathbf{G_4}$ ■ $\mathbf{G_5}$

☐ $\mathbf{G_6}$ ■ $\mathbf{G_7}$ ☐ $\mathbf{G_8}$ ■ $\mathbf{G_9}$ ■ $\mathbf{G_{10}}$

☐ None of the above.

Since $\mathbf{B_1}$ can represent $\mathbf{d_1}$, we know that $\mathbf{d_1}$ must satisfy the assumptions that $\mathbf{B_1}$ follows, which are: $AB; AB|C; BC; BC|A$. We cannot assume that $\mathbf{d_1}$ satisfies the other two assumptions, which are AC and $AC|B$, and so a Bayes' net that makes at least one of these two extra assumptions will not be guaranteed to be able to represent $\mathbf{d_1}$. This eliminates the choices $\mathbf{G_1}, \mathbf{G_2}, \mathbf{G_3}, \mathbf{G_6}, \mathbf{G_8}$. The other choices $\mathbf{G_4}, \mathbf{G_5}, \mathbf{G_7}, \mathbf{G_9}, \mathbf{G_{10}}$ are guaranteed to be able to represent $\mathbf{d_1}$ because they do not make any additional independence assumptions that $\mathbf{B_1}$ makes.

(b) Assume we know that a joint distribution $\mathbf{d_2}$ (over $\mathbf{A}, \mathbf{B}, \mathbf{C}$) can be represented by Bayes' net $\mathbf{B_2}$. Mark all of the following Bayes' nets that are guaranteed to be able to represent $\mathbf{d_2}$.

☐ $\mathbf{G_1}$ ☐ $\mathbf{G_2}$ ☐ $\mathbf{G_3}$ ☐ $\mathbf{G_4}$ ☐ $\mathbf{G_5}$

■ $\mathbf{G_6}$ ☐ $\mathbf{G_7}$ ■ $\mathbf{G_8}$ ■ $\mathbf{G_9}$ ■ $\mathbf{G_{10}}$

☐ None of the above.

Since $\mathbf{B_2}$ can represent $\mathbf{d_2}$, we know that $\mathbf{d_2}$ must satisfy the assumptions that $\mathbf{B_2}$ follows, which is just: $AC|B$. We cannot assume that $\mathbf{d_2}$ satisfies any other assumptions, and so a Bayes' net that makes at least one other extra assumptions will not be guaranteed to be able to represent $\mathbf{d_2}$. This eliminates the choices $\mathbf{G_1}, \mathbf{G_2}, \mathbf{G_3}, \mathbf{G_4}, \mathbf{G_5}, \mathbf{G_7}$. The other choices $\mathbf{G_6}, \mathbf{G_8}, \mathbf{G_9}, \mathbf{G_{10}}$ are guaranteed to be able to represent $\mathbf{d_2}$ because they do not make any additional independence assumptions that $\mathbf{B_2}$ makes.

(c) Assume we know that a joint distribution $\mathbf{d_3}$ (over $\mathbf{A}, \mathbf{B}, \mathbf{C}$) ***cannot*** be represented by Bayes' net $\mathbf{B_3}$. Mark all of the following Bayes' nets that are guaranteed to be able to represent $\mathbf{d_3}$.

☐ $\mathbf{G_1}$ ☐ $\mathbf{G_2}$ ☐ $\mathbf{G_3}$ ☐ $\mathbf{G_4}$ ☐ $\mathbf{G_5}$

☐ $\mathbf{G_6}$ ☐ $\mathbf{G_7}$ ☐ $\mathbf{G_8}$ ■ $\mathbf{G_9}$ ■ $\mathbf{G_{10}}$

☐ None of the above.

Since $\mathbf{B_3}$ cannot represent $\mathbf{d_3}$, we know that $\mathbf{d_3}$ is unable to satisfy at least one of the assumptions that $\mathbf{B_3}$ follows. Since $\mathbf{B_3}$ only makes one independence assumption, which is AC, we know that $\mathbf{d_3}$ does not satisfy AC. However, we can't claim anything about whether or not $\mathbf{d_3}$ makes any of the other independence assumptions. $\mathbf{d_3}$ might not make any (conditional) independence assumptions at all, and so only the Bayes' nets that don't make any assumptions will be guaranteed to be able to represent $\mathbf{d_3}$. Hence, the answers are the fully connected Bayes' nets, which are $\mathbf{G_9}, \mathbf{G_{10}}$.

(d) Assume we know that a joint distribution $\mathbf{d_4}$ (over $\mathbf{A}, \mathbf{B}, \mathbf{C}$) can be represented by Bayes' nets $\mathbf{B_1}$, $\mathbf{B_2}$, and $\mathbf{B_3}$. Mark all of the following Bayes' nets that are guaranteed to be able to represent $\mathbf{d_4}$.

■ $\mathbf{G_1}$ ■ $\mathbf{G_2}$ ■ $\mathbf{G_3}$ ■ $\mathbf{G_4}$ ■ $\mathbf{G_5}$

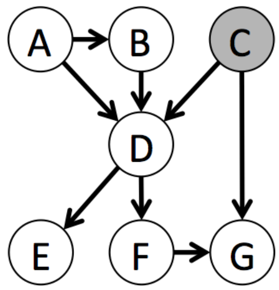
■ $\mathbf{G_6}$ ■ $\mathbf{G_7}$ ■ $\mathbf{G_8}$ ■ $\mathbf{G_9}$ ■ $\mathbf{G_{10}}$

☐ None of the above.

Since $\mathbf{B_1}, \mathbf{B_2}, \mathbf{B_3}$ can represent $\mathbf{d_4}$, we know that $\mathbf{d_4}$ must satisfy the assumptions that $\mathbf{B_1}, \mathbf{B_2}, \mathbf{B_3}$ make. The union of assumptions made by these Bayes' nets are: $AB; AB|C; BC; BC|A, AC, AC|B$. Note that this set of assumptions encompasses all the possible assumptions that you can make with 3 random variables, so any Bayes' net over $\mathbf{A}, \mathbf{B}, \mathbf{C}$ will be able to represent $\mathbf{d_4}$.

Q3. Variable Elimination

(a) For the Bayes' net below, we are given the query $P(A, E \mid +c)$. All variables have binary domains. Assume we run variable elimination to compute the answer to this query, with the following variable elimination ordering: B, D, G, F.



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

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

$$P(A), P(B|A), P(+c), P(D|A, B, +c), P(E|D), P(F|D), P(G|+c, F)$$

When eliminating B we generate a new factor f_1 as follows:

$$f_1(A, +c, D) = \sum_b P(b|A)P(D|A, b, +c)$$

This leaves us with the factors:

$$P(A), P(+c), P(E|D), P(F|D), P(G|+c, F), f_1(A, +c, D)$$

When eliminating D we generate a new factor f_2 as follows:

$$f_2(A, +c, E, F) = \sum_d P(E|d)P(F|d)f_1(A, +c, d)$$

This leaves us with the factors:

$$P(A), P(+c), P(G|+c, F), f_2(A, +c, E, F)$$

When eliminating G we generate a new factor f_3 as follows:

$$f_3(+c, F) = \sum_g P(g|+c, F)$$

This leaves us with the factors:

$$P(A), P(+c), f_2(A, +c, E, F), f_3(+c, F)$$

Let's make sure to account for error propagation in our grading of this one.

When eliminating F we generate a new factor f_4 as follows:

$$f_4(A, +c, E) = \sum_f f_2(A, +c, E, f) f_3(+c, f)$$

This leaves us with the factors:

$$P(A), P(+c), f_4(A, +c, E)$$

(b) Write a formula to compute $P(A, E \mid +c)$ from the remaining factors.

$P(A, E \mid +c) = \frac{P(A)P(+c)f_4(A,+c,E)}{\sum_{a,e} P(a)P(+c)f_4(a,+c,e)}$ or alternatively: $P(A, E \mid +c) \propto P(A)P(+c)f_4(A, +c, E)$ and include statement that says renormalization is needed to obtain $P(A, E \mid +c)$.

(c) Among f_1, f_2, f_3, f_4, which is the largest factor generated, and how large is it? Assume all variables have binary domains and measure the size of each factor by the number of rows in the table that would represent the factor.

$f_2(A, +c, E, F)$ is the largest factor generated. It has 3 non-instantiated variables, hence $2^3 = 8$ entries.

(d) Find a variable elimination ordering for the same query, i.e., for $P(A, E \mid +c)$, for which the maximum size factor generated along the way is smallest. Hint: the maximum size factor generated in your solution should have only 2 variables, for a size of $2^2 = 4$ table. Fill in the variable elimination ordering and the factors generated into the table below.

Variable Eliminated	Factor Generated
B	$f_1(A, +c, D)$
G	$f_2(+c, F)$
F	$f_3(+c, D)$
D	$f_4(A, +c, E)$

For example, in the naive ordering we used earlier, the first row in this table would have had the following two entries: B, $f_1(A, +c, D)$.

Note: multiple orderings are possible. An ordering is good if it eliminates all non-query variables (B, D, F, G) and its largest factor has only two variables.

Q4. Bayes Nets: Sampling

Consider the following Bayes Net, where we have observed that $B = +b$ and $D = +d$.

$P(A)$	
$+a$	0.5
$-a$	0.5

$P(B \mid A)$		
$+a$	$+b$	0.8
$+a$	$-b$	0.2
$-a$	$+b$	0.4
$-a$	$-b$	0.6

$P(C \mid B)$		
$+b$	$+c$	0.1
$+b$	$-c$	0.9
$-b$	$+c$	0.7
$-b$	$-c$	0.3

$P(D \mid A, C)$			
$+a$	$+c$	$+d$	0.6
$+a$	$+c$	$-d$	0.4
$+a$	$-c$	$+d$	0.1
$+a$	$-c$	$-d$	0.9
$-a$	$+c$	$+d$	0.2
$-a$	$+c$	$-d$	0.8
$-a$	$-c$	$+d$	0.5
$-a$	$-c$	$-d$	0.5

(a) Consider doing Gibbs sampling for this example. Assume that we have initialized all variables to the values $+a, +b, +c, +d$. We then unassign the variable C, such that we have $A = +a$, $B = +b$, $C = ?$, $D = +d$. Calculate the probabilities for new values of C at this stage of the Gibbs sampling procedure.

$P(C = +c$ at the next step of Gibbs sampling$) = \frac{0.1 \cdot 0.6}{0.1 \cdot 0.6 + 0.9 \cdot 0.1} = \boxed{\frac{2}{5}}$

$P(C = -c$ at the next step of Gibbs sampling$) = \frac{0.9 \cdot 0.1}{0.1 \cdot 0.6 + 0.9 \cdot 0.1} = \boxed{\frac{3}{5}}$

(b) Consider a sampling scheme that is a hybrid of rejection sampling and likelihood-weighted sampling. Under this scheme, we first perform rejection sampling for the variables A and B. We then take the sampled values for A and B and extend the sample to include values for variables C and D, using likelihood-weighted sampling.

(i) Below is a list of candidate samples. Mark the samples that would be rejected by the rejection sampling portion of the hybrid scheme.

■ $-a$ $-b$
☐ $+a$ $+b$
■ $+a$ $-b$
☐ $-a$ $+b$

(ii) To decouple from part (i), you now receive a *new* set of samples shown below. Fill in the weights for these samples under our hybrid scheme.

				Weight
$-a$	$+b$	$-c$	$+d$	0.5
$+a$	$+b$	$-c$	$+d$	0.1
$+a$	$+b$	$-c$	$+d$	0.1
$-a$	$+b$	$+c$	$+d$	0.2
$+a$	$+b$	$+c$	$+d$	0.6

(iii) Use the weighted samples from part (ii) to calculate an estimate for $P(+a \mid +b, +d)$.

The estimate of $P(+a \mid +b, +d)$ is $\frac{0.1 + 0.1 + 0.6}{0.5 + 0.1 + 0.1 + 0.2 + 0.6} = \boxed{\frac{8}{15}}$

(c) We now attempt to design an alternative hybrid sampling scheme that combines elements of likelihood-weighted and rejection sampling. For each proposed scheme, indicate whether it is valid, i.e. whether the weighted samples it produces correctly approximate the distribution $P(A, C \mid +b, +d)$.

(i) *First collect a likelihood-weighted sample for the variables A and B. Then switch to rejection sampling for the variables C and D. In case of rejection, the values of A and B and the sample weight are* ***thrown away****. Sampling then restarts from node* ***A****.*

● Valid ○ Invalid

(ii) *First collect a likelihood-weighted sample for the variables A and B. Then switch to rejection sampling for the variables C and D. In case of rejection, the values of A and B and the sample weight are* ***retained****. Sampling then restarts from node* ***C****.*

○ Valid ● Invalid

The sampling procedure in part (i) is the correct way of combining likelihood-weighted and rejection sampling: any time a node gets rejected, the sample must be thrown out in its entirety. In part (ii), however, the evidence that $D = +d$ has no effect on which values of A are sampled or on the sample weights. This means that values for A would be sampled according to $P(A| + b)$, not $P(A| + b, +d)$.

As an extreme case, suppose node D had a different probability table where $P(+d| + a) = 0$. Following the procedure from part (ii), we might start by sampling $(+a, +b)$ and assigning a weight according to $P(+b| + a)$. However, when we move on to rejection sampling we will be forced to continuously reject all possible values because our evidence $+d$ is inconsistent with our the assignment of $A = +a$. This means that the procedure from part (ii) is flawed to the extent that it might fail to generate a sample altogether!