Q1. [18 pts] Markov Decision Processes

(a) [4 pts] Write out the equations to be used to compute Q_i^* from R, T, V_{i-1}^*, γ and to compute V_i^* from R, T, Q_i^*, γ.

$$Q_i^*(s,a) = \sum_{s'} T(s,a,s') \left(R(s,a,s') + \gamma V_{i-1}^*(s') \right)$$

$$V_i^*(s) = \max_a Q_i^*(s,a)$$

(b) [10 pts] Consider the MDP with transition model and reward function as given in the table below. Assume the discount factor $\gamma = 1$, i.e., no discounting. Fill in the values for $V_0^*, V_1^*, V_2^*, Q_1^*, Q_2^*$ in the graph below.
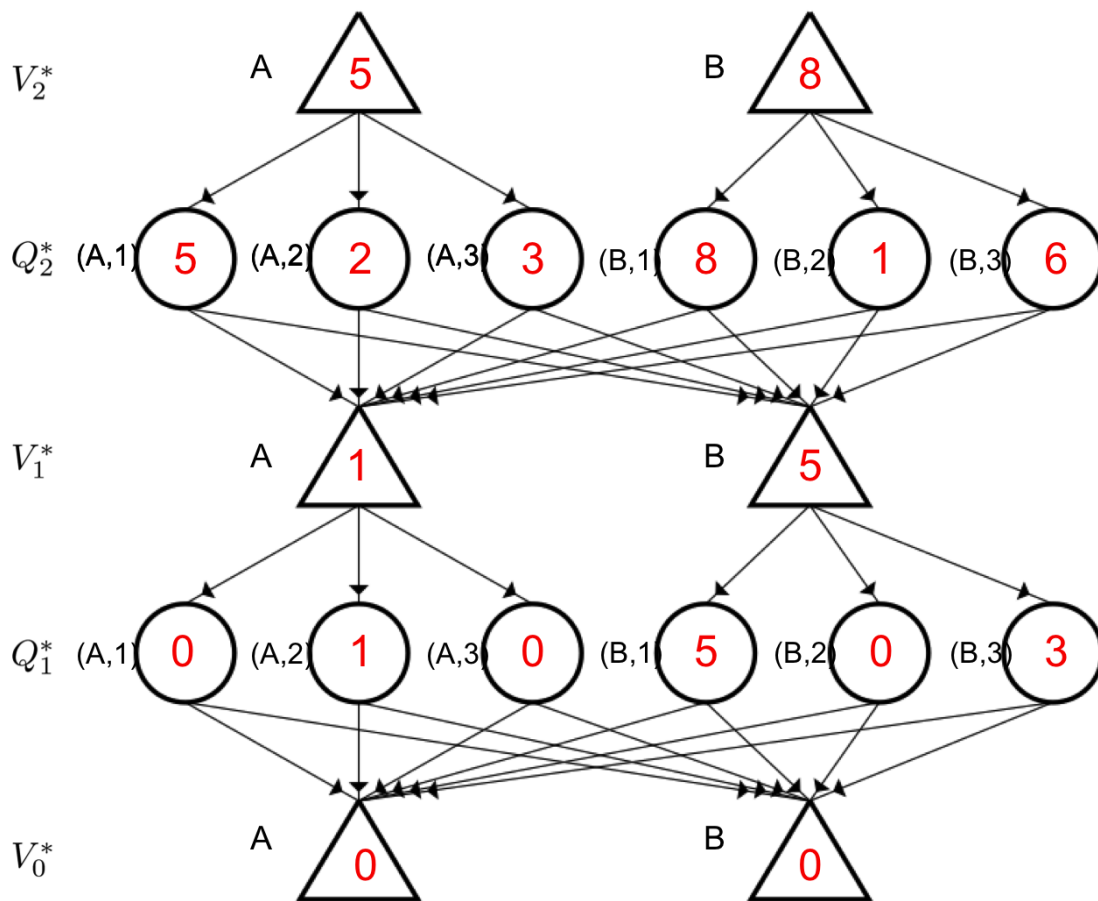
s	a	s'	T(s,a,s')	R(s,a,s')
A	1	A	0	0
A	1	B	1	0
A	2	A	1	1
A	2	B	0	0
A	3	A	0.5	0
A	3	B	0.5	0

s	a	s'	T(s,a,s')	R(s,a,s')
B	1	A	0.5	10
B	1	B	0.5	0
B	2	A	1	0
B	2	B	0	0
B	3	A	0.5	2
B	3	B	0.5	4

(c) [4 pts] Let $\pi_i^*(s)$ be the optimal action in state s with i time steps to go. Fill in the following tables:

s	$\pi_1^*(s)$
A	2
B	1

s	$\pi_2^*(s)$
A	1
B	1

V_2^* A △ 5   B △ 8

Q_2^* (A,1) 5 (A,2) 2 (A,3) 3 (B,1) 8 (B,2) 1 (B,3) 6

V_1^* A △ 1   B △ 5

Q_1^* (A,1) 0 (A,2) 1 (A,3) 0 (B,1) 5 (B,2) 0 (B,3) 3

V_0^* A △ 0   B △ 0

Q2. [23 pts] MDPs and Utility: Short Questions

Each True/False question is worth 2 points. Leaving a question blank is worth 0 points. **Answering incorrectly is worth −2 points.**

For the questions that are not True/False, answer as concisely as possible (and no points are subtracted for a wrong answer to these).

(a) **Utility.**

(i) [*true* or *false*] If an agent has the preference relationship $(A \succ B) \wedge (B \succ C) \wedge (C \succ A)$ then this agent can be induced to give away all of its money.

For most utility functions over money the answer would be true, but there are some special utility functions for which it would not be true. As we did not specify a utility function over money, technically the statement is actually false. The fact that a few special utility functions make this statement false is not at all the angle we intended to test you on when making this question. We accepted any answer.

(ii) [*true* or *false*] Assume Agent 1 has a utility function U_1 and Agent 2 has a utility function U_2. If $U_1 = k_1 U_2 + k_2$ with $k_1 > 0, k_2 > 0$ then Agent 1 and Agent 2 have the same preferences.

For any a, b: $U_2(a) > U_2(b)$ equivalent to $k_1 U_2(a) > k_1 U_2(b)$ since $k_1 > 0$

$k_1 U_2(a) > k_1 U_2(b)$ equivalent to $k_1 U_2(a) + k_2 > k_1 U_2(b) + k_2$ for any k_2

(b) **Insurance.** Some useful numbers: $\log(101) \approx 4.6151, \log(71) \approx 4.2627$.

PacBaby just found a $100 bill—it is the only thing she owns. Ghosts are nice enough not to kill PacBaby,

but when they find PacBaby they will steal all her money. The probability of the ghosts finding PacBaby is 20%. PacBaby's utility function is $U(x) = \log(1 + x)$ (this is the natural logarithm, i.e., $\log e^x = x$), where x is the total monetary value she owns. When PacBaby gets to keep the \$100 (ghosts don't find her) her utility is $U(\$100) = \log(101)$. When PacBaby loses the \$100 (per the ghosts taking it from her) her utility is $U(\$0) = \log(1 + 0) = 0$.

(i) [2 pts] What is the expected utility for PacBaby?

$0.8 * \log(101) + 0.2 * \log(1) = 0.8 * \log(101) + 0.2 * 0 \approx 3.6921$

(ii) [4 pts] Pacgressive offers theft insurance: if PacBaby pays an insurance premium of \$30, then they will reimburse PacBaby \$70 if the ghosts steal all her money (after paying \$30 in insurance, she would only have \$70 left). What is the expected utility for PacBaby if she takes insurance? For PacBaby to maximize her expected utility should she take this insurance?

When taking insurance, PacBaby's expected utility equals $0.8 \log(1 + 70) + 0.2 \log(1 + 70) = \log(71) \approx 4.2627$. Yes, PacBaby should take the insurance.

(iii) [2 pts] In the above scenario, what is the expected <u>monetary</u> value of selling the insurance from Pacgressive's point of view?

The expected monetary value equals $0.8 * 30 + 0.2 * (-40) = 16$.

(c) **MDPs.**

(i) [*true* or *false*] If the only difference between two MDPs is the value of the discount factor then they must have the same optimal policy.
A counterexample suffices to show the statement is false. Consider an MDP with two sink states. Transitioning into sink state A gives a reward of 1, transitioning into sink state B gives a reward of 10. All other transitions have zero rewards. Let A be one step North from the start state. Let B be two steps South from the start state. Assume actions always succeed. Then if the discount factor $\gamma < 0.1$ the optimal policy takes the agent one step North from the start state into A, if the discount factor $\gamma > 0.1$ the optimal policy takes the agent two steps South from the start state into B.

(ii) [*true* or *false*] When using features to represent the Q-function (rather than having a tabular representation) it is possible that Q-learning does not find the optimal Q-function Q^*.
Whenever the optimal Q-function, Q^*, cannot be represented as a weighted combination of features, then the feature-based representation would not even have the expressiveness to find the optimal Q-function, Q^*.

(iii) [*true* or *false*] For an infinite horizon MDP with a finite number of states and actions and with a discount factor γ, with $0 < \gamma < 1$, value iteration is guaranteed to converge.

(d) [5 pts] Recall that for a deterministic policy π where $\pi(s)$ is the action to be taken in state s we have that the value of the policy satisfies the following equations:
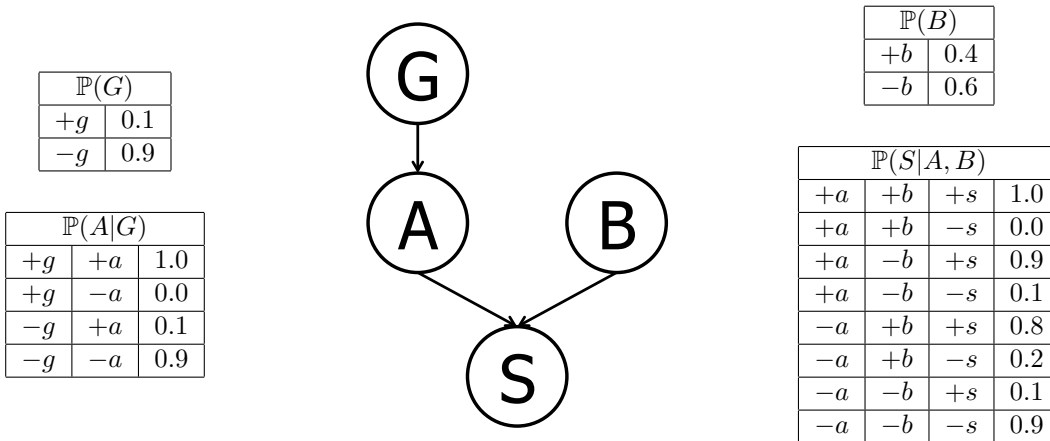
$$V^\pi(s) = \sum_{s'} T(s, \pi(s), s') \left(R(s, \pi(s), s') + \gamma V^\pi(s') \right).$$

Now assume we have a stochastic policy π where $\pi(s, a) = P(a|s)$ is equal to the probability of taking action a when in state s. Write the equivalent of the above equation for the value of this stochastic policy.

$$V^\pi(s) = \sum_a \pi(s, a) \sum_{s'} T(s, a, s') \left(R(s, a, s') + \gamma V^\pi(s') \right)$$

Q3. [14 pts] Bayes' Nets Representation and Probability

Suppose that a patient can have a symptom (S) that can be caused by two different diseases $(A$ and $B)$. It is known that the variation of gene G plays a big role in the manifestation of disease A. The Bayes' Net and corresponding conditional probability tables for this situation are shown below. For each part, you may leave your answer as an arithmetic expression.

$\mathbb{P}(G)$	
$+g$	0.1
$-g$	0.9

$\mathbb{P}(A\vert G)$		
$+g$	$+a$	1.0
$+g$	$-a$	0.0
$-g$	$+a$	0.1
$-g$	$-a$	0.9

$\mathbb{P}(B)$	
$+b$	0.4
$-b$	0.6

$\mathbb{P}(S\vert A,B)$			
$+a$	$+b$	$+s$	1.0
$+a$	$+b$	$-s$	0.0
$+a$	$-b$	$+s$	0.9
$+a$	$-b$	$-s$	0.1
$-a$	$+b$	$+s$	0.8
$-a$	$+b$	$-s$	0.2
$-a$	$-b$	$+s$	0.1
$-a$	$-b$	$-s$	0.9

(Bayes' Net: $G \to A$, $A \to S$, $B \to S$)

(a) [2 pts] Compute the following entry from the joint distribution:

$\mathbb{P}(+g, +a, +b, +s) =$

$\mathbb{P}(+g)\mathbb{P}(+a\vert+g)\mathbb{P}(+b)\mathbb{P}(+s\vert+b,+a) = (0.1)(1.0)(0.4)(1.0) = 0.04$
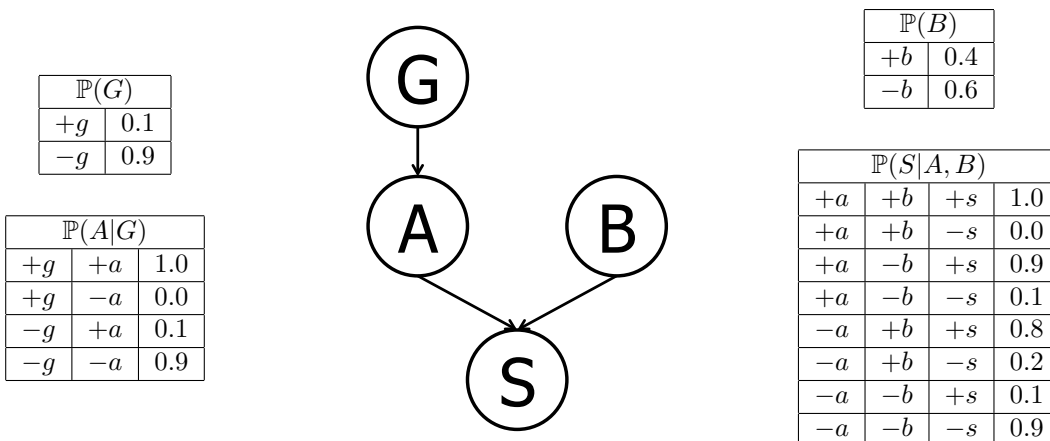
(b) [2 pts] What is the probability that a patient has disease A?

$\mathbb{P}(+a) = \mathbb{P}(+a\vert+g)\mathbb{P}(+g) + \mathbb{P}(+a\vert-g)\mathbb{P}(-g) = (1.0)(0.1) + (0.1)(0.9) = 0.19$

(c) [2 pts] What is the probability that a patient has disease A given that they have disease B?

$\mathbb{P}(+a\vert+b) = \mathbb{P}(+a) = 0.19$ The first equality holds true as we have $A \perp\!\!\!\perp B$, which can be inferred from the graph of the Bayes' net.

The figures and table below are identical to the ones on the previous page and are repeated here for your convenience.

$\mathbb{P}(B)$	
$+b$	0.4
$-b$	0.6

$\mathbb{P}(G)$	
$+g$	0.1
$-g$	0.9

$\mathbb{P}(A\|G)$		
$+g$	$+a$	1.0
$+g$	$-a$	0.0
$-g$	$+a$	0.1
$-g$	$-a$	0.9

$\mathbb{P}(S\|A,B)$			
$+a$	$+b$	$+s$	1.0
$+a$	$+b$	$-s$	0.0
$+a$	$-b$	$+s$	0.9
$+a$	$-b$	$-s$	0.1
$-a$	$+b$	$+s$	0.8
$-a$	$+b$	$-s$	0.2
$-a$	$-b$	$+s$	0.1
$-a$	$-b$	$-s$	0.9

(d) [4 pts] What is the probability that a patient has disease A given that they have symptom S and disease B?

$$\mathbb{P}(+a|+s,+b) = \frac{\mathbb{P}(+a,+b,+s)}{\mathbb{P}(+a,+b,+s)+\mathbb{P}(-a,+b,+s)} = \frac{\mathbb{P}(+a)\mathbb{P}(+b)\mathbb{P}(+s|+a,+b)}{\mathbb{P}(+a)\mathbb{P}(+b)\mathbb{P}(+s|+a,+b)+\mathbb{P}(-a)\mathbb{P}(+b)\mathbb{P}(+s|-a,+b)}$$
$$= \frac{(0.19)(0.4)(1.0)}{(0.19)(0.4)(1.0)+(0.81)(0.4)(0.8)} = \frac{0.076}{0.076+0.2592} \approx 0.2267$$

(e) [2 pts] What is the probability that a patient has the disease carrying gene variation G given that they have disease A?

$$\mathbb{P}(+g|+a) = \frac{\mathbb{P}(+g)\mathbb{P}(+a|+g)}{\mathbb{P}(+g)\mathbb{P}(+a|+g)+\mathbb{P}(-g)\mathbb{P}(+a|-g)} = \frac{(0.1)(1.0)}{(0.1)(1.0)+(0.9)(0.1)} = \frac{0.1}{0.1+0.09} = 0.5263$$

(f) [2 pts] What is the probability that a patient has the disease carrying gene variation G given that they have disease B?

$\mathbb{P}(+g|+b) = \mathbb{P}(+g) = 0.1$ The first equality holds true as we have $G \perp\!\!\!\perp B$, which can be inferred from the graph of the Bayes' net.

Q4. [16 pts] D-Separation

(a) [16 pts] Based only on the structure of the (new) Bayes' Net given below, circle whether the following conditional independence assertions are guaranteed to be true, guaranteed to be false, or cannot be determined by the structure alone.
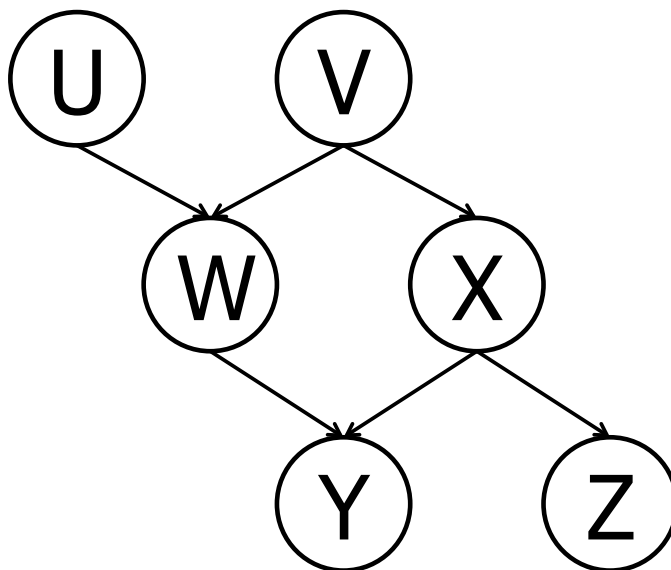
|			
$U \perp\!\!\!\perp V$	Guaranteed true	Cannot be determined	Guaranteed false
$U \perp\!\!\!\perp V \mid W$	Guaranteed true	Cannot be determined	Guaranteed false
$U \perp\!\!\!\perp V \mid Y$	Guaranteed true	Cannot be determined	Guaranteed false
$U \perp\!\!\!\perp Z \mid W$	Guaranteed true	Cannot be determined	Guaranteed false
$U \perp\!\!\!\perp Z \mid V, Y$	Guaranteed true	Cannot be determined	Guaranteed false
$U \perp\!\!\!\perp Z \mid X, W$	Guaranteed true	Cannot be determined	Guaranteed false
$W \perp\!\!\!\perp X \mid Z$	Guaranteed true	Cannot be determined	Guaranteed false
$V \perp\!\!\!\perp Z \mid X$	Guaranteed true	Cannot be determined	Guaranteed false

Q5. [22 pts] Variable Elimination

(a) [10 pts] For the Bayes' net below, we are given the query $P(Z \mid +y)$. All variables have binary domains. Assume we run variable elimination to compute the answer to this query, with the following variable elimination ordering: U, V, W, T, X.
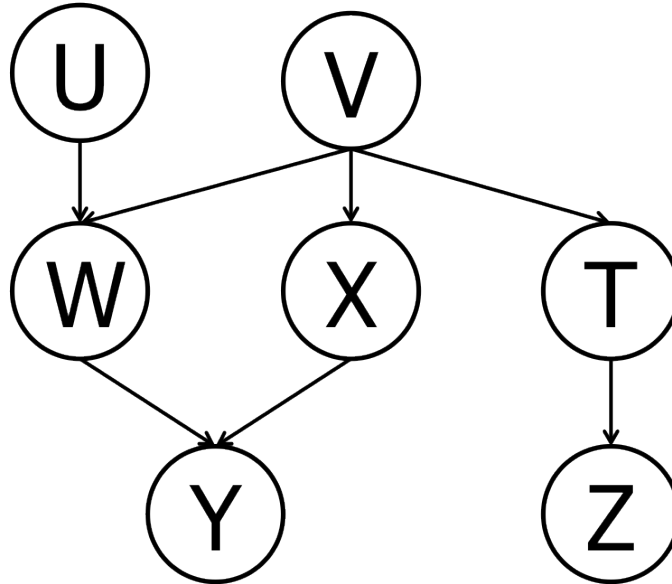
Complete the following description of the factors generated in this process:

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

$$P(U), P(V), P(W|U,V), P(X|V), P(T|V), P(+y|W,X), P(Z|T).$$

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

$$f_1(V, W) = \sum_u P(u)P(W|u, V).$$

This leaves us with the factors:

$$P(V), P(X|V), P(T|V), P(+y|W,X), P(Z|T), f_1(V, W).$$

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

$$f_2(T, W, X) = \sum_v P(v)P(X|v)P(T|v)f_1(v, W).$$

This leaves us with the factors:

$$P(+y|W,X), P(Z|T), f_2(T, W, X).$$

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

$$f_3(T, X, +y) = \sum_w P(+y|w, X) f_2(T, w, X).$$

This leaves us with the factors:

$$P(Z|T), f_3(T, X, +y).$$

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

$$f_4(X, +y, Z) = \sum_t P(Z|t) f_3(t, X, +y).$$

This leaves us with the factor:

$$f_4(X, +y, Z).$$

When eliminating X we generate a new factor f_5 as follows:

$$f_5(+y, Z) = \sum_x f_4(x, +y, Z).$$

This leaves us with the factor:

$$f_5(+y, Z)$$

.

(b) [2 pts] Briefly explain how $P(Z \mid +y)$ can be computed from f_5.

Simply renormalize f_5 to obtain $P(Z \mid +y)$. Concretely, $P(z \mid +y) = \frac{f_5(z, +y)}{\sum_{z'} f_5(z', y)}$.

(c) [2 pts] Amongst $f_1, f_2, \ldots, f_5$, which is the largest factor generated? (Assume all variables have binary domains.) How large is this factor?

$f_2(T, W, X)$ is the largest factor generated. It has 3 variables, hence $2^3 = 8$ entries.

(d) [8 pts] Find a variable elimination ordering for the same query, i.e., for $P(Z \mid y)$, 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.
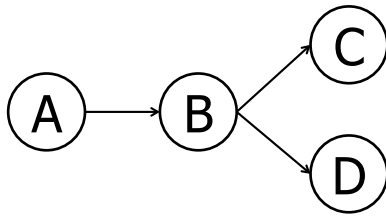
Note: in the naive ordering we used earlier, the first line in this table would have had the following two entries: U, $f_1(V, W)$.

Note: multiple orderings are possible.

Variable Eliminated	Factor Generated
T	$f_1(Z, V)$
X	$f_2(+y, W, V)$
W	$f_3(+y, U, V)$
U	$f_4(+y, V)$
V	$f_5(+y, Z)$

Q6. [7 pts] Bayes' Nets Sampling

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

A	$\mathbb{P}(A)$
$+a$	$1/5$
$-a$	$4/5$

B	C	$\mathbb{P}(C|B)$
$+b$	$+c$	$1/4$
$+b$	$-c$	$3/4$
$-b$	$+c$	$2/5$
$-b$	$-c$	$3/5$

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

B	D	$\mathbb{P}(D|B)$
$+b$	$+d$	$1/2$
$+b$	$-d$	$1/2$
$-b$	$+d$	$4/5$
$-b$	$-d$	$1/5$

(a) [2 pts] Your task is now to estimate $\mathbb{P}(+b| - a, -c, -d)$ using rejection sampling. Below are some samples that have been produced by prior sampling (that is, the rejection stage in rejection sampling hasn't happened yet). Cross out the samples that would be rejected by rejection sampling:

$$-a \quad -b \quad +c \quad +d \qquad -a \quad -b \quad -c \quad -d$$
$$+a \quad -b \quad -c \quad +d \qquad -a \quad +b \quad +c \quad +d$$
$$-a \quad -b \quad +c \quad -d \qquad +a \quad -b \quad -c \quad -d$$

(b) [1 pt] Using those samples, what value would you estimate for $\mathbb{P}(+b| - a, -c, -d)$ using rejection sampling?

0

(c) [4 pts] Using the following samples (which were generated using likelihood weighting), estimate $\mathbb{P}(+b \mid -a, -c, -d)$ using likelihood weighting, or state why it cannot be computed.

$$-a \quad -b \quad -c \quad -d$$
$$-a \quad +b \quad -c \quad -d$$
$$-a \quad -b \quad -c \quad -d$$

We compute the weights of each solution, which are the product of the probabilities of the evidence variables conditioned on their parents.

$$w_1 = w_3 = P(-a)P(-c \mid -b)P(-d \mid -b) = 4/5 * 3/5 * 1/5 = 12/125$$
$$w_2 = P(-a)P(-c \mid +b)P(-d \mid +b) = 4/5 * 3/4 * 1/2 = 12/40$$

so normalizing, we have $(w_2)/(w_2 + w_1 + w_3) = \frac{12/40}{12/40 + 12/125 + 12/125} = \frac{25}{41} \approx 0.609756.$