Q2. [10 pts] Bayes Net and Sampling

(a) [2 pts] A, B, C are discrete random variables. Given $A \perp\!\!\!\perp B|C$, which of the following equations must hold?

- ■ $P(A|B,C)P(B|A,C) = P(A,B|C)$
- ☐ $P(A,B,C) = P(A)P(B)P(A,B|C)$
- ■ $P(A|C) = \frac{P(A)P(C|A)}{P(C)}$
- ☐ $P(A,B|C) = P(A,B)$
- ○ None of the above.

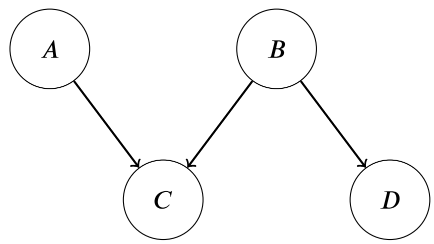
The first option is correct. We can simplify the left side by $P(A|B,C) = P(A|C)$ and $P(B|A,C) = P(B|C)$, then it follows from conditional independence definition.

The second option is incorrect. The left side is equal to $P(A,B|C)P(C)$, so this is true only if $P(A)P(B) = P(C)$ which doesn't make sense.

The third option is correct. It follows from Bayes theorem and doesn't assume any conditions.

The last option is incorrect. The joint probability of A, B can very much depend on C.

Consider the following Bayes Net involving binary random variables A, B, C, D. The relevant probability tables are given.



$P(A)$	
$A = 0$	0.5
$A = 1$	0.5

$P(B)$	
$B = 0$	0.5
$B = 1$	0.5

A	B	C	$P(C\|A,B)$
0	0	0	0.6
0	0	1	0.4
0	1	0	0.4
0	1	1	0.6
1	0	0	0.8
1	0	1	?
1	1	1	0.6
1	1	0	0.4

B	D	$P(D\|B)$
0	0	0.6
0	1	0.4
1	0	0.4
1	1	0.6

(b) **(i)** [1 pt] Calculate $P(C = 1|A = 1, B = 0)$ (the ? entry in the table).

1-0.8=0.2

(ii) [1 pt] Calculate $P(A = 1|B = 0, D = 1)$.

$= P(A = 1) = 0.5$

(c) Instead of calculating the exact quantity, suppose we want to estimate $P(C = 1|D = 1)$ using different sampling methods.

(i) [1 pt] In this subpart we use rejection sampling. Which of the following is a valid topological order and is most efficient for rejection sampling to estimate $P(C = 1|D = 1)$?

- ○ A, B, C, D
- ○ B, A, D, C
- ● B, D, A, C
- ○ D, C, B, A

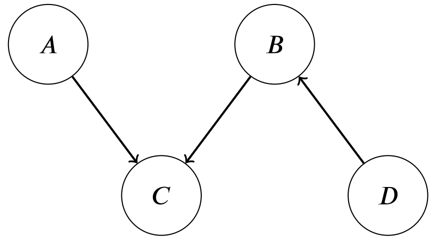
(ii) [1 pt] In this subpart we use likelihood weighting. What is the weight of the sample $(A = 0, B = 0, C = 0, D = 1)$?

0.4

(iii) [2 pts] In this subpart we use Gibbs sampling. We initialize $A = 0$, $B = 0$, $C = 0$, $D = 1$, and choose to re-sample A. What is the probability that we still get $A = 0$ after re-sampling?

$$P(A = 0 | B = 0, C = 0, D = 1) = \frac{P(A=0,B=0,C=0)}{\sum_a P(A=a,B=0,C=0)} = \frac{0.5*0.5*0.6}{0.5*0.5*0.6+0.5*0.5*0.8} = 3/7$$

(d) [2 pts] We reverse the arrow between B and D to create a new Bayes Net (shown below).



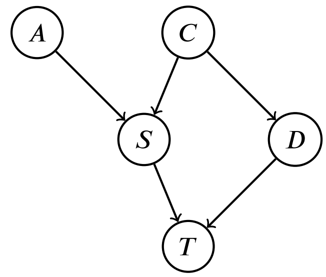
Which of the following statements are true?

- ● The set of joint distributions $P(A, B, C, D)$ that can be modeled by the two Bayes nets are the same.
- ○ The set of joint distributions that can be modeled by the old Bayes net is a subset of the set of joint distributions that can be modeled by the new Bayes net.
- ○ The set of joint distributions that can be modeled by the new Bayes net is a subset of the set of joint distributions that can be modeled by the old Bayes net.
- ○ None of the above.

Q2. [13 pts] Bayesian Networks: Across the Spider-Verse

Miles is curious about the probability that he arrives to school on time. The factors involved with him arriving to school on time can be represented by the following Bayes Net (assume that each variable is a binary variable):

- A: Sets alarm
- S: Over sleeps
- D: Dad is late
- C: Fighting crime last night
- T: Arrives at school on time



(a) [7 pts] Miles wants to calculate $P(C|T)$ using variable elimination. Assume he eliminates variables in alphabetical order (A, D, S).

(i) [1 pt] What factors does he have available at the start?

$P(A), P(C), P(S|A,C), P(D|C), P(T|S,D)$

(ii) [1 pt] First, he eliminates A, and get the new factor

$f_1(C,S) = \sum_A P(A) * P(S|A,C)$

Write out the remaining factors

$P(C), P(D|C), P(T|S,D), f_1(C,S)$

(iii) [1 pt] Then, he eliminates D, and get the new factor

$f_2(C,S,T) = \sum_D P(D|C) * P(T|S,D)$

Write out the remaining factors

$P(C), f_1(C,S), f_2(C,S,T)$

(iv) [1 pt] Then, he eliminates S, and get the new factor

$f_3(C,T) = \sum_S f_1(C,S) * f_2(C,S,T)$

Write out the remaining factors

$P(C), f_3(C,T)$

(v) [1 pt] Finally, join any remaining factors to calculate

$f_4(C,T) = P(C) * f_3(C,T)$

(vi) [1 pt] How can he use this to calculate $P(C=+c|T=-t)$? Your answer should be in terms of f_4.

$P(C=+c|T=-t) = \frac{f_4(+c,-t)}{f_4(+c,-t)+f_4(-c,-t)}$

(vii) [1 pt] Order factors f_1, f_2, and f_3 in increasing order of size. f_1 < f_3 < f_2

The following CPTs correspond to the Bayes Net above:

A	C	S	$P(S\|A,C)$
$+a$	$+c$	$+s$	11/16
$+a$	$+c$	$-s$	5/16
$+a$	$-c$	$+s$	1/8
$+a$	$-c$	$-s$	7/8
$-a$	$+c$	$+s$	9/10
$-a$	$+c$	$-s$	1/10
$-a$	$-c$	$+s$	1/5
$-a$	$-c$	$-s$	4/5

T	D	S	$P(T\|S,D)$
$+t$	$+d$	$+s$	1/20
$+t$	$+d$	$-s$	2/5
$+t$	$-d$	$+s$	1/5
$+t$	$-d$	$-s$	1
$-t$	$+d$	$+s$	19/20
$-t$	$+d$	$-s$	3/5
$-t$	$-d$	$+s$	4/5
$-t$	$-d$	$-s$	0

C	D	$P(D\|C)$
$+c$	$+d$	1/4
$+c$	$-d$	3/4
$-c$	$+d$	0
$-c$	$-d$	1

C	$P(A)$
$-c$	4/5
$+c$	1/5

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

Miles is now interested in calculating $P(C = +c|T = -t)$ via sampling. He generates the following random samples (assume the variables were generated from left to right):

Sample	A	C	S	D	T
1	$+a$	$-c$	$-s$	$-d$	$+t$
2	$+a$	$-c$	$+s$	$-d$	$-t$
3	$+a$	$+c$	$+s$	$-d$	$-t$
4	$-a$	$-c$	$-s$	$+d$	$-t$
5	$-a$	$-c$	$+s$	$+d$	$-t$

(b) [3 pts] Assuming Miles uses prior sampling:

(i) [1 pt] Bubble in the samples that Miles uses to calculate the final probability.

☐ 1 ■ 2 ■ 3 ■ 4 ■ 5

(ii) [2 pts] What is the probability he calculates via prior sampling?

$P(C = +c|T = -t) =$ 1/4

(c) [3 pts] Now assuming Miles uses likelihood weighting:

(i) [1 pt] What weight does Miles assign to each sample?

Sample 1: 0 Sample 2: 4/5 Sample 3: 4/5

Sample 4: 3/5 Sample 5: 19/20

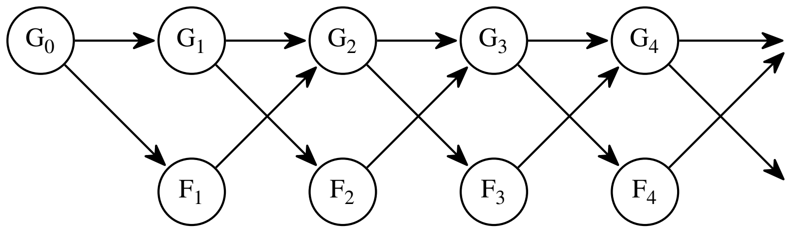
What final probability does he calculate?

(ii) [2 pts] What is the probability he calculates via likelihood weighting?

$P(C = +c|T = -t) =$ 16/63

Q6. [10 pts] Dynamic Bayes' Nets: Book Club

Each week t ($t \geq 0$), the members of a book club decide on the genre of the book G_t to read that week: romance ($+g_t$) or science fiction ($-g_t$). Each week's choice is influenced by the genre chosen in the previous week. In addition, after week t ($t \geq 0$), they collect feedback (F_{t+1}) from the members, which can be overall positive ($+f_{t+1}$) or negative ($-f_{t+1}$), and that influences the choice at week $t+2$. The situation is shown in the Bayes' net below:



(a) [1 pt] Recall that the joint distribution up to time T for the "standard" HMM model with state X_t and evidence E_t is given by $P(X_{0:T}, E_{1:T}) = P(X_0)\prod_{t=1}^{T} P(X_t \mid X_{t-1})P(E_t \mid X_t)$.

Write an expression for the joint distribution in the model shown above.

$P(F_{1:T}, G_{0:T}) =$ $P(G_0)P(F_1 \mid G_0)P(G_1 \mid G_0)\prod_{t=2}^{T} P(F_t \mid G_{t-1})P(G_t \mid G_{t-1}F_{t-1})$

This is just the ordinary expression for the joint distribution of a Bayes net. Recall that in a Bayes net, multiplying all of the conditional probability tables together (one per node) results in the joint distribution. The tricky part to writing this expression is getting the first couple of steps right ($t = 0$ and $t = 1$).

(b) [2 pts] Which of the following Markov assumptions are implied by the network structure, assuming $t \geq 2$?

- ☐ $P(G_t \mid G_{0:t-1}, F_{1:t-1}) = P(G_t \mid F_{t-1})$
- ☐ $P(G_t \mid G_{0:t-1}, F_{1:t-1}) = P(G_t \mid G_{t-1})$
- ☐ $P(G_t, F_t \mid G_{0:t-1}, F_{1:t-1}) = P(G_t, F_t \mid G_{t-1})$
- ■ $P(G_t, F_t \mid G_{0:t-1}, F_{1:t-1}) = P(G_t, F_t \mid G_{t-1}, F_{t-1})$
- ☐ $P(F_t \mid G_{0:t-1}, F_{1:t-1}) = P(F_t \mid F_{t-1})$
- ■ $P(F_t \mid G_{0:t-1}, F_{1:t-1}) = P(F_t \mid G_{t-1}, F_{t-2})$
- ○ None of the above

These can all be worked out precisely using independence of non-descendants given parents. For the last one, note that the F_{t-2} is superfluous, but the equation is still correct.

The conditional probability tables for G_t and F_t are shown below, where a, b, c, d, p, q are constants between 0 and 1.

F_t	G_{t-1}	$P(F_t \mid G_{t-1})$
+	+	p
−	+	$1-p$
+	−	q
−	−	$1-q$

G_t	F_{t-1}	G_{t-1}	$P(G_t \mid F_{t-1}, G_{t-1})$
+	+	+	a
−	+	+	$1-a$
+	+	−	b
−	+	−	$1-b$
+	−	+	c
−	−	+	$1-c$
+	−	−	d
−	−	−	$1-d$

Let's consider this model as a Markov chain with state variables (F_t, G_t).

(c) [2 pts] Write out the transition probabilities below.

Your answer should be an expression, possibly in terms of a, b, c, d, p, and q.

$x_1 = P(+f_t, +g_t \mid +f_{t-1}, +g_{t-1}) =$ ap

$x_2 = P(+f_t, +g_t \mid +f_{t-1}, -g_{t-1}) =$ bq

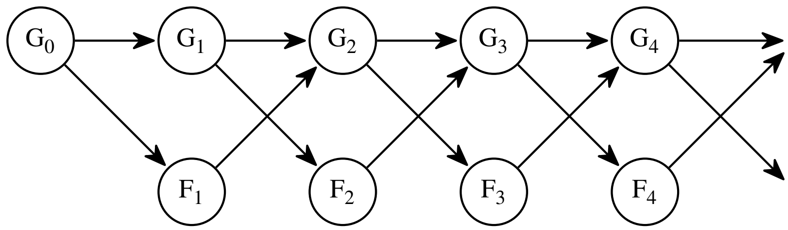
$x_3 = P(+f_t, +g_t \mid -f_{t-1}, +g_{t-1}) =$ cp

$x_4 = P(+f_t, +g_t \mid -f_{t-1}, -g_{t-1}) =$ dq

These entries follow from the fact that $P(F_t, G_t \mid F_{t-1}, G_{t-1}) = P(F_t \mid G_{t-1})P(G_t \mid F_{t-1}, G_{t-1})$. You can think of this expression as joining together two factors (CPTs) of the Bayes' net.

Intuitively, you can also derive these expressions by considering the transition dynamics of this Markov chain. For example, consider deriving x_1. At time $t-1$, we have state $+f_{t-1}, +g_{t-1}$, and we want to know how likely it is for the state at time t to be $+f_t, +g_t$. In order for this transition to happen, $+f_{t-1}$ must transition to $+f_t$. The probability that this happens is governed by $P(F_t|G_{t-1})$, and we look up the value in the table $P(+f_t|+g_{t-1})$ (since we know $+g_{t-1}$ at time $t-1$) to find p. Then, $+g_{t-1}$ must transition to $+g_t$. This transition is governed by the $P(G_t|F_{t-1}, G_{t-1})$, and substituting in the desired values $P(+g_t|+f_{t-1}, +g_{t-1})$ lets us look up the value a in the table. Since both of these transitions must happen, the probability of the overall transition is ap. The same approach can be used to derive the other three expressions.

The Bayes' net, repeated for your convenience:



(d) [2 pts] As t goes to infinity, the stationary distribution of this Markov chain is shown below, where y_1, y_2, y_3, y_4 are constants between 0 and 1.

$$y_1 = P(+f_\infty, +g_\infty)$$
$$y_2 = P(+f_\infty, -g_\infty)$$
$$y_3 = P(-f_\infty, +g_\infty)$$
$$y_4 = P(-f_\infty, -g_\infty)$$

Write an equation that must be true if this is a stationary distribution.

Your answer should be an equation, possibly in terms of x_1, x_2, x_3, x_4 (from the previous part), y_1, y_2, y_3, y_4.

$$x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 = y_1$$

Recall that in the equilibrium distribution, the probability of being in a state no longer depends on the time step. In other words, if the probabilities of being in the four states at time ∞ are y_1, y_2, y_3, y_4, then at the next time step $\infty + 1$, after the transition dynamics are applied once, the probabilities of being in the four states are still y_1, y_2, y_3, y_4.

From the previous subpart, we have the probabilities of transitioning into state $(+f, +g)$ from each of the four possible states. Therefore, we can write an equilibrium equation about the probability of being in $(+f, +g)$, which is y_1.

$x_1y_1 =$ probability of starting in $(+f, +g)$, then transitioning to $(+f, +g)$. Note that x_1 is the probability of the transition and y_1 is the probability of starting in the state.

$x_2y_2 =$ probability of starting in $(+f, -g)$ and transitioning to $(+f, +g)$.

$x_3y_3 =$ probability of starting in $(-f, +g)$ and transitioning to $(+f, +g)$.

$x_4y_4 =$ probability of starting in $(-f, -g)$ and transitioning to $(+f, +g)$.

Since these are the only four ways to transition into $(+f, +g)$, if we sum them up, we should get the probability that we're in $(+f, +g)$ on the next time step, and because this is an equilibrium distribution, this should be equal to the probability that we're in $(+f, +g)$ right now.

(e) [3 pts] Select all true statements.

- ■ There are values of a, b, c, d such that the feedback model defined by p, q is irrelevant to the stationary distribution of the genre G.
- ☐ Even if $P(F_t, G_t)$ reaches a unique stationary distribution, it is possible that the marginal probabilities $P(F_t)$ and $P(G_t)$ do not reach unique stationary distributions.
- ☐ For any values of a, b, c, d, p, q, there is at least one stationary distribution.
- ○ None of the above

True. If $a = c = 1$ and $b = d = 0$, then G_t is simply copied from G_{t-1} regardless of F. In that case the stationary distribution is the same as $P(G_0)$, and p and q are irrelevant for computing the stationary distribution.

False. $P(F_t)$ and $P(G_t)$ can be computed directly from $P(F_t, G_t)$, so if the latter is constant then so are the former.

False. If $a = c = 0$ and $b = d = 1$, then G_t is always the opposite of G_{t-1}, i.e., a deterministic cycle, and there is no stationary distribution.