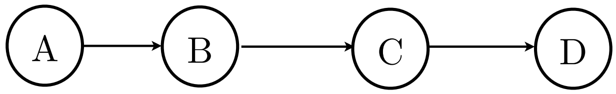
CS 188
Summer 2022

Regular Discussion 5A Solutions

Q1. Bayes' Nets Sampling

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



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

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

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

$P(D \mid C)$		
$-c$	$-d$	1/8
$-c$	$+d$	7/8
$+c$	$-d$	5/6
$+c$	$+d$	1/6

(a) You are given the following samples:

$+a$	$+b$	$-c$	$-d$
$+a$	$-b$	$+c$	$-d$
$\boxed{-a}$	$\boxed{+b}$	$\boxed{+c}$	$\boxed{-d}$
$\boxed{-a}$	$\boxed{-b}$	$\boxed{+c}$	$\boxed{-d}$

$\boxed{+a}$	$\boxed{-b}$	$\boxed{-c}$	$\boxed{+d}$
$+a$	$+b$	$+c$	$-d$
$\boxed{-a}$	$\boxed{+b}$	$\boxed{-c}$	$\boxed{+d}$
$\boxed{-a}$	$\boxed{-b}$	$\boxed{+c}$	$\boxed{-d}$

(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 :	$+a$	$+b$	$-c$	$-d$

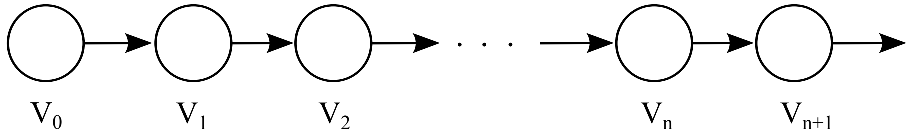
Sequence 4

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.

2 HMMs



Consider a Markov model like the one above. For the first three parts of this problem, assume the domain of our variables is {a, b}. Fill in the table below with the probability of being in each state after a large number of transitions, when $P(V_n) = P(V_{n+1})$. If the values never reach a point when $P(V_n) = P(V_{n+1})$, write 'None'.

(a) In the left part of the table, assume that we start with a uniform distribution ($P(V_0 = \text{a}) = P(V_0 = \text{b}) = 0.5$). In the right part of the table, assume that we start with the distribution that has $P(V_0 = \text{a}) = 1.0$.

<table>
<tr><th rowspan="2">Transition Probabilities</th><th colspan="2">$P(V_n)$ given that $P(V_0)$ is uniform</th><th colspan="2">$P(V_n)$ given that $P(V_0 = \text{a}) = 1.0$</th></tr>
<tr><th>$V_n = \text{a}$</th><th>$V_n = \text{b}$</th><th>$V_n = \text{a}$</th><th>$V_n = \text{b}$</th></tr>
<tr><td><table><tr><th rowspan="2">V_{i-1}</th><th colspan="2">$P(V_i \mid V_{i-1})$</th></tr><tr><th>$V_i = \text{a}$</th><th>$V_i = \text{b}$</th></tr><tr><td>a</td><td>0.5</td><td>0.5</td></tr><tr><td>b</td><td>0.5</td><td>0.5</td></tr></table></td><td>0.5</td><td>0.5</td><td>0.5</td><td>0.5</td></tr>
<tr><td><table><tr><th rowspan="2">V_{i-1}</th><th colspan="2">$P(V_i \mid V_{i-1})$</th></tr><tr><th>$V_i = \text{a}$</th><th>$V_i = \text{b}$</th></tr><tr><td>a</td><td>0.9</td><td>0.1</td></tr><tr><td>b</td><td>0.3</td><td>0.7</td></tr></table></td><td>0.75</td><td>0.25</td><td>0.75</td><td>0.25</td></tr>
<tr><td><table><tr><th rowspan="2">V_{i-1}</th><th colspan="2">$P(V_i \mid V_{i-1})$</th></tr><tr><th>$V_i = \text{a}$</th><th>$V_i = \text{b}$</th></tr><tr><td>a</td><td>0.0</td><td>1.0</td></tr><tr><td>b</td><td>1.0</td><td>0.0</td></tr></table></td><td>0.5</td><td>0.5</td><td>None</td><td>None</td></tr>
</table>

(b) For this part our variables have the domain {a, b, c}. Fill in the table at the bottom with the probability of being in each state after a large number of transitions, when $P(V_n) = P(V_{n+1})$. In the left part of the table, assume that we start with a uniform distribution ($P(V_0 = \text{a}) = P(V_0 = \text{b}) = P(V_0 = \text{c}) = \frac{1}{3}$). In the right part of the table, assume that we start with the distribution that has $P(V_0 = \text{a}) = 1.0$.

	$P(V_i \mid V_{i-1})$		
V_{i-1}	$V_i = \text{a}$	$V_i = \text{b}$	$V_i = \text{c}$
a	0.5	0.5	0.0
b	0.5	0.5	0.0
c	0.0	0.0	1.0

$P(V_n)$ given that $P(V_0)$ is uniform			$P(V_n)$ given that $P(V_0 = \text{a}) = 1.0$		
a	b	c	a	b	c
0.33	0.33	0.33	0.5	0.5	0.0

(c) Now we will consider a Hidden Markov Model, and look at properties of the Viterbi Algorithm. The Viterbi algorithm finds the most probable sequence of hidden states $X_{1:S}$ given a sequence of observations $y_{1:S}$. Recall that for the canonical HMM structure, the Viterbi algorithm performs the following update at each time step:

$$m_t[x_t] = P(y_t|x_t) \max_{x_{t-1}}[P(x_t|x_{t-1})m_{t-1}[x_{t-1}]]$$

Assume we have an HMM where:

- The hidden variable X can take on H values
- The (observed) emission variable Y can take on E values
- Our sequence has S steps

(d) **(i)** What is the run time of the Viterbi algorithm?

○ $O(SEH)$ ○ $O(SEH^2)$ ● $O(SH^2)$
○ $O(SH)$ ○ $O(EH)$ ○ $O(EH^2)$
○ $O(SH^2 + SEH)$

Ignoring the storage of the emission probabilities, $P(Y_t|X_t)$, and the transition probabilities, $P(X_t|X_{t-1})$, what are the storage requirements of the Viterbi algorithm?

○ $O(S)$ ○ $O(E)$ ○ $O(H)$
● $O(SH)$ ○ $O(SE)$ ○ $O(EH)$
○ $O(S+H)$ ○ $O(S+E)$ ○ $O(E+H)$
○ $O(SEH)$

Now, assume that most of the transitions in our HMM have probability zero. In particular, suppose that for any given hidden state value, there are only K possible next state values for which the transition probability is non-zero. To exploit this sparsity, we change the Viterbi Algorithm to only consider the non-zero transition edges during each max computation inside each update. You can think of this as the Viterbi algorithm ignoring edges that correspond to zero probability transitions in the transition lattice diagram.

(ii) What is the run time of this modified algorithm?

○ $O(SEH)$ ○ $O(SEH^2)$ ○ $O(SH^2)$
○ $O(SH)$ ○ $O(EH)$ ○ $O(EH^2)$
○ $O(SH^2 + SEH)$
○ $O(SEK)$ ○ $O(SEHK)$ ● $O(SHK)$
○ $O(SK)$ ○ $O(EK)$ ○ $O(EHK)$
○ $O(SK + SEK)$ ○ $O(SHK + SEK)$

Ignoring the storage of the emission probabilities, $P(Y_t|X_t)$, and the transition probabilities, $P(X_t|X_{t-1})$, what are the storage requirements of this modified Viterbi algorithm?

○ $O(S)$ ○ $O(E)$ ○ $O(H)$
● $O(SH)$ ○ $O(SE)$ ○ $O(EH)$
○ $O(S+H)$ ○ $O(S+E)$ ○ $O(E+H)$
○ $O(SEH)$
○ $O(K)$ ○ $O(SK)$ ○ $O(EK)$
○ $O(HK)$ ○ $O(S+K)$ ○ $O(E+K)$
○ $O(H+K)$ ○ $O(SEK)$