Section Handout 8 Solutions

CS188 Spring 2019 Section 8: Bayes Nets (Inference and Sampling)

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. $P(Q|E) \approx \frac{count(Q and E)}{count(E)}$.

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

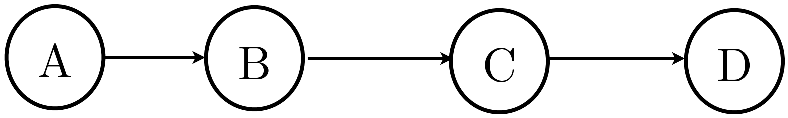
Likelihood Weighting: Fix the evidence variables, and weight each sample by the probability of the evidence variables given their parents.

Gibbs Sampling:

1. Fix evidence.
2. Initialize other variables randomly
3. Repeat:
   (a) Choose non-evidence variable X.
   (b) Resample X from $P(X|markovblanket(X))$

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:

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

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

(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.

(f) Let us suppose that you are doing Gibbs sampling and you begin with the initial sample $(+a, -b, -c, +d)$. We will sample a variable given the others, in this order: A, B, C, D.

(i) Calculate the probability table for $P(A| -b, -c, +d)$.

$$\begin{aligned} P(A|-b,-c,+d) &= \frac{P(A,-b,-c,+d)}{\sum_a P(A=a,-b,-c,+d)} \\ &= \frac{P(A)P(-b|A)P(-c|-b)P(+d|-c)}{\sum_a P(A=a)P(-b|A=a)P(-c|-b)P(+d|-c)} \\ &= \frac{P(A)P(-b|A)}{\sum_a P(A=a)P(-b|A=a)} \end{aligned}$$

So $P(A=+a|-b,-c,+d) = \frac{\frac{1}{4}\cdot\frac{4}{5}}{\frac{1}{4}\cdot\frac{4}{5}+\frac{3}{4}\cdot\frac{2}{3}} = \frac{2}{7}$ and $P(A=-a|-b,-c,+d) = \frac{\frac{3}{4}\cdot\frac{2}{3}}{\frac{1}{4}\cdot\frac{4}{5}+\frac{3}{4}\cdot\frac{2}{3}} = \frac{5}{7}$.

(ii) Let us suppose that in sampling $P(A|-b,-c,+d)$ we had sampled $-a$, in sampling $P(B|-a,-c,+d)$ we had sampled $-b$, and in sampling $P(C|-a,-b,+d)$ we had sampled $+c$. Calculate the probability table for $P(D|-a,-b,+c)$.

$$\begin{aligned} P(D|-a,-b,+c) &= \frac{P(-a,-b,-c,D)}{\sum_d P(-a,-b,-c,D=d)} \\ &= \frac{P(-a)P(-b|-a)P(+c|-b)P(D|+c)}{\sum_d P(-a)P(-b|-a)P(+c|-b)P(D=d|+c)} \\ &= \frac{P(D|+c)}{\sum_d P(D=d|+c)} \end{aligned}$$

So $P(D=+d|-a,-b,+c) = \frac{\frac{1}{6}}{\frac{5}{6}+\frac{1}{6}} = \frac{1}{6}$ And $P(D=-d|-a,-b,+c) = \frac{\frac{5}{6}}{\frac{5}{6}+\frac{1}{6}} = \frac{5}{6}$.

(g) Let us consider Gibbs sampling for deterministic CPTs. Consider the distribution over binary variables X, Y, Z factorized as $P(X)P(Y)P(Z|X,Y)$, where $Z = X xor Y$, and both X and Y have uniform prior distributions. Suppose we want to query $P(X|-z)$. Let the initial sample be $+x, +y, -z$.

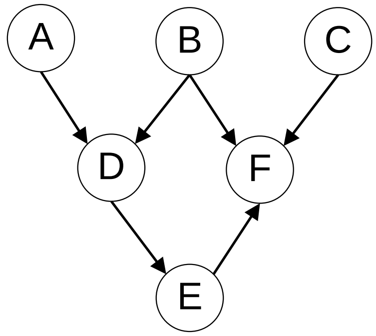
(i) What is the probability table $P(X|-z)$ we would get if we used single-variable Gibbs sampling? What is the problem with this approach? From the samples we would see that $P(X=+x|-z)=1$ and $P(X=-x|-z)=0$. This is because the third variable will be completely determined by the other two in Gibbs sampling for this probability model, so the Markov Chain will not be able to visit every possible variable assignment; we need the Markov Chain to be *ergodic*.

(ii) Now let us do block Gibbs sampling, where we sample X and Y simultaneously. What is the probability table $P(X|-z)$ we would get with block Gibbs sampling?
From the samples we would see that $P(X=+x|-z)=0.5$ and $P(X=-x|-z)=0.5$ as desired.

Q2. Bayes Nets

(a) For the following graphs, explicitly state the minimum size set of edges that must be removed such that the corresponding independence relations are guaranteed to be true.

Marked the removed edges with an 'X' on the graphs.

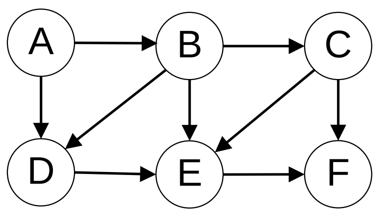


$A \perp\!\!\!\perp B | F$

$A \perp\!\!\!\perp F | D$

$B \perp\!\!\!\perp C$

(i) AD. Nodes A, B, and F form a V-structure. In order to make sure that $A \perp\!\!\!\perp B|F$, we would have to remove AD, or DE or EF. Removing DE or EF does not guarantee that $A \perp\!\!\!\perp F|D$, so to remove the minimal amount of edges, we remove AD.



$A \perp\!\!\!\perp D | B$

$A \perp\!\!\!\perp F | C$

$C \perp\!\!\!\perp D | B$

(ii) AD, (EF OR AB)

We need to remove AD because there is a direct edge between A and D which does not guarantee independence between A and D, no matter what we condition on. In order for $A \perp\!\!\!\perp F|C$, since A is a non-descendent of F, we would need C to be the only parent that separates F from A. This can be achieved by either removing EF or removing AB. To enforce $C \perp\!\!\!\perp D|B$ does not require us to remove any edges.