

# 1 Sampling

We will look over different methods of sampling and how to query using these samples.

(a) Sampling from a Single Variable Distribution

	A	P(A)
Distribution	+a	0.7
	-a	0.3

	1	2	3	4
Random draws	0.4	0.6	0.9	0.7

Given the above random draws from the uniform distribution  $[0,1)$ , what will the samples be?

+a, +a, -a, -a

(b) Sampling from a Conditional Distribution

	A	B	$P(B A)$
Distribution	+a	+b	0.7
	+a	-b	0.3
	-a	+b	0.6
	-a	-b	0.4

	1	2	3	4
Random draws	0.4	0.6	0.9	0.7

Given the above random draws from the uniform distribution  $[0,1)$ , what will the samples be for A = -a?

+b, -b, -b, -b

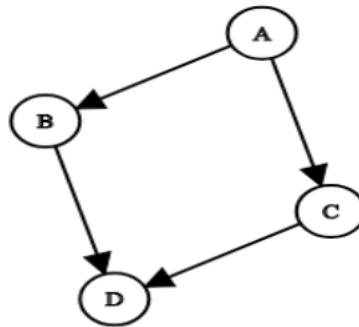
## 2 Sampling from Bayes Nets

A	P(A)
+a	0.8
-a	0.2

A	B	P(B A)
+a	+b	0.8
+a	-b	0.2
-a	+b	0.5
-a	-b	0.5

A	C	P(C A)
+a	+c	0.7
+a	-c	0.3
-a	+c	0.9
-a	-c	0.1

B	C	D	P(D B,C)
+b	+c	+d	0.3
+b	+c	-d	0.7
+b	-c	+d	0.1
+b	-c	-d	0.9
-b	+c	+d	0.2
-b	+c	-d	0.8
-b	-c	+d	0.9
-b	-c	-d	0.1



Random draws	1	2	3	4	5	6	7	8
	0.3	0.6	0.05	0.9	0.68	0.45	0.38	0.2

(a) Prior Sampling

List samples drawn using the above random draws

$(+a, +b, +c, -d), (+a, +b, +c, +d)$

Answer the following queries:

P(+d):

0.5

Can we answer the query  $P(+a | -c)$ ?

No. We have no samples with  $C = -c$ .

(b) Sampling using Rejection Sampling : Rejecting sample once an evidence variable has been sampled to take on a value inconsistent with the evidence of the query

List samples drawn using the above random draws for the query  $P(-d | -b)$ :

$(+a, -b, +c, -d)$

What is a drawback of rejection sampling?

Imagine we have large bayes net with the evidence at the bottom. Let's say while sampling, we traverse the bayes net to reach the evidence only to get an inconsistent value. This waste a lot of time.

(c) Sampling using Likelihood Sampling:

List the weighted samples drawn using the above random draws for the query  $P(-d | -a, -b)$ :

$0.1(+c, -d), 0.1()etc$

Answering queries using weighted samples. We are given the following weighted samples:

A	B	C	D	weight
+a	-b	-c	-d	0.5
+a	+b	+c	-d	0.3
+a	-b	+c	-d	0.1
+a	-b	+c	+d	0.4
-a	-b	-c	+d	0.1

$P(-d)$ :

$P(+a | +c, -d)$ :

(d) Gibbs sampling.

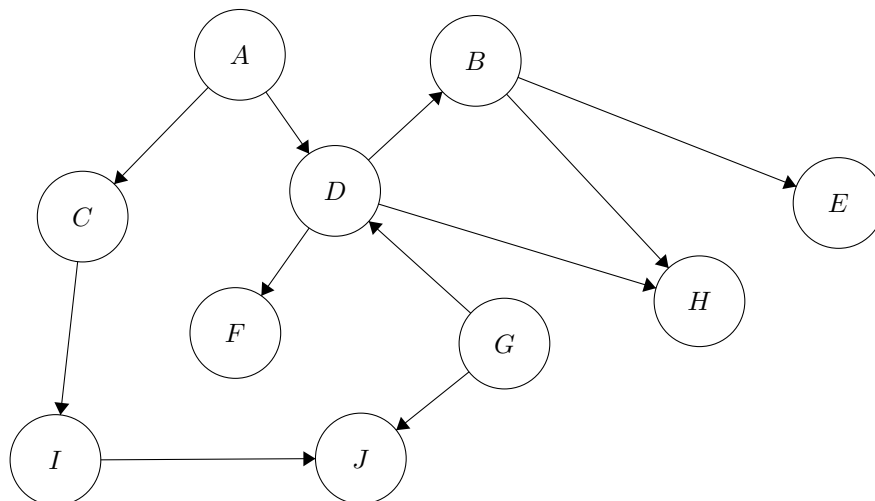
Why is Gibbs sampling better than Likelihood sampling?

Evidence influences the choice of downstream variables, but not upstream ones. We would like to consider evidence when we sample every variable. The evidence is taken into account for every variable in gibbs sampling.

Let's say we initialize the sample to  $(-a, +b, +c, +d)$ . We then resample A first. What is the probability that the sample equals  $(+a, +b, +c, +d)$ ?

### 3 Markov Blankets

Consider the following Bayes net:



- (a) What does the Markov blanket of a node consist of and what is its significance?

The Markov blanket of a node  $A$  consists of all the parents of  $A$ , all the children of  $A$ , and all of  $A$ 's children's other parents. Formally, we have:

$$\text{MB}(A) = \text{Parents}(A) \cup \text{Children}(A) \cup \left( \bigcup \{ \text{Parents}(C) \mid \forall C \in \text{Children}(A) \} \right)$$

The Markov blanket is significant in that  $A$  is conditionally independent of all other variables given its Markov Blanket.

- (b) Is  $C$  independent of  $D$  given  $A$  and  $I$ ?

Yes, because:  $\text{MB}(C) = \{A, I\}$

- (c) Give the Markov blankets for  $B$ ,  $D$ , and  $G$ .

$$\text{MB}(B) = \{D, E, H\}$$

$$\text{MB}(D) = \{A, F, G, H, B\}$$

$$\text{MB}(G) = \{A, D, I, J\}$$

## 4 Variable Elimination and Variable Ordering

Sometimes we want to find the probability distribution of a certain variable but we don't have its direct probability values. We do, however, have the distribution of joint or conditional probabilities containing that variable. We can find the probability by summing over the joint probability of all of our given variables, by finding a way to express the joint probability in terms of distributions we know. This is called variable elimination.

$$P(A_i) = \sum_{A_1} \sum_{A_2} \cdots \sum_{A_{i-1}} \sum_{A_{i+1}} \cdots \sum_{A_n} P(A_1, A_2, \dots, A_n)$$

The trick is to figure out how to right  $P(A_1, A_2, \dots, A_n)$  as a product of known distributions using properties like chain rule or Bayes' rule. For  $h$  known distributions, we eventually find an identity:

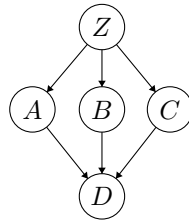
$$P(A_1, A_2, \dots, A_n) = \prod_{k=1}^{k=h} P(A_{j_{k,1}}, A_{j_{k,2}}, \dots, A_{j_{k,m_k}} \mid A_{i_{k,1}}, A_{i_{k,2}}, \dots, A_{i_{k,n_k}})^{-1^{b_k}}$$

Now, we plug it back in to the original summation:

$$P(A_i) = \sum_{A_1} \sum_{A_2} \cdots \sum_{A_{i-1}} \sum_{A_{i+1}} \cdots \sum_{A_n} \prod_{k=1}^{k=h} P(A_{j_{k,1}}, A_{j_{k,2}}, \dots, A_{j_{k,m_k}} \mid A_{i_{k,1}}, A_{i_{k,2}}, \dots, A_{i_{k,n_k}})^{-1^{b_k}}$$

Now, recall back to recitation 9 Magnetic Factors. How can you optimize this expression to minimize computation costs? (i.e. minimize the number of multiplication operations needed to be performed). To do this, we take the factors that do not contain the variable that their innermost summation is iterating over, and move it up the summation chain until it is in front of a summation operation that iterates over one of the variables it contains. We do this for all factors. Now, we recall that summation is an associative operation. This means, we can reorder the summations and still achieve the same result. Therefore, we can further optimize computation by ordering the summations in such a way that that maximizes amount of factors factored out to the left.

- (a) Now, let's say we have the Bayes Net:



With the distribution tables for probabilities  $P(D | A, B, C)$ ,  $P(Z)$ ,  $P(A | Z)$ ,  $P(B | Z)$ ,  $P(C | Z)$ , and we want to find  $P(D)$ . Given that we only have that  $Z$  influences all the variables (so all variables are independent given  $Z$ ), determine the optimal ordering of variables we need to sum together to achieve  $P(D)$ .

First we find the joint distribution in terms of known distributions:

$$P(D, A, B, C, Z) = P(D | Z, A, B, C)P(A, B, C | Z)P(Z)$$

By chain rule; and since the variables  $A, B, C$  are independent given  $Z$ , we have:

$$= P(D | Z, A, B, C)P(A | Z)P(B | Z)P(C | Z)P(Z)$$

Since  $D$  is independent of  $Z$  given  $A, B, C$ , we have

$$= P(D | A, B, C)P(A | Z)P(B | Z)P(C | Z)P(Z)$$

Now, we do variable elimination:

$$P(D) = \sum_A \sum_B \sum_C \sum_Z P(D, A, B, C, Z) = \sum_A \sum_B \sum_C \sum_Z P(D | A, B, C)P(A | Z)P(B | Z)P(C | Z)P(Z)$$

Now, we rearrange the summations to get the optimal factoring:

$$= \sum_Z P(Z) \sum_A P(A | Z) \sum_B P(B | Z) \sum_C P(C | Z)P(D | A, B, C)$$