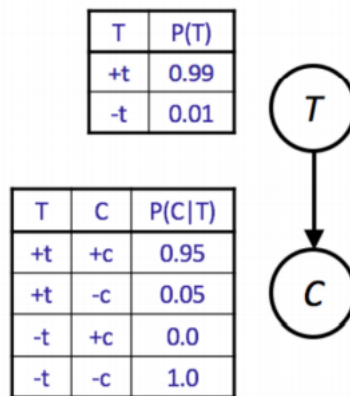


Bayesian Network Representation

In a Bayesian network, rather than storing information in a giant table, probabilities are instead distributed across a large number of smaller local probability tables along with a **directed acyclic graph** (DAG) which captures the relationships between variables. Thus, if we have a node representing variable X , we store $P(X|A_1, A_2, \dots, A_N)$, where A_1, \dots, A_N are the parents of X .



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

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

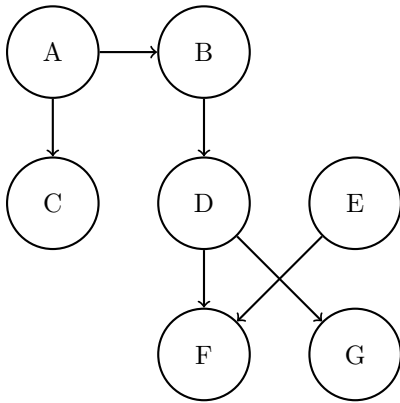
Likelihood Weighting: Fix the evidence variables, and weight each sample by the probability of the evidence variables given their parents.

Gibbs Sampling:

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

1 Bayes' Nets: Representation and Independence

Parts (a) and (b) pertain to the following Bayes' Net.



- (a) Express the joint probability distribution as a product of terms representing individual conditional probabilities tables associated with the Bayes Net.

$$P(A)P(C|A)P(B|A)P(D|B)P(E)P(F|D, E)P(G|D)$$

- (b) Assume each node can take on 4 values. How many entries do the factors at A, D, and F have?

A: 4

D: 4²

F: 4³

Consider the following probability distribution tables. The joint distribution $P(A, B, C, D)$ is equal to the product of these probability distribution tables.

	A	B	$P(B A)$		B	C	$P(C B)$		C	D	$P(D C)$
A	$P(A)$	+a	+b	0.9	+b	+c	0.8	+c	+d	0.25	
+a	0.8	+a	-b	0.1	+b	-c	0.2	+c	-d	0.75	
-a	0.2	-a	+b	0.6	-b	+c	0.8	-c	+d	0.5	
		-a	-b	0.4	-b	-c	0.2	-c	-d	0.5	

(c) State all non-conditional independence assumptions that are implied by the probability distribution tables.

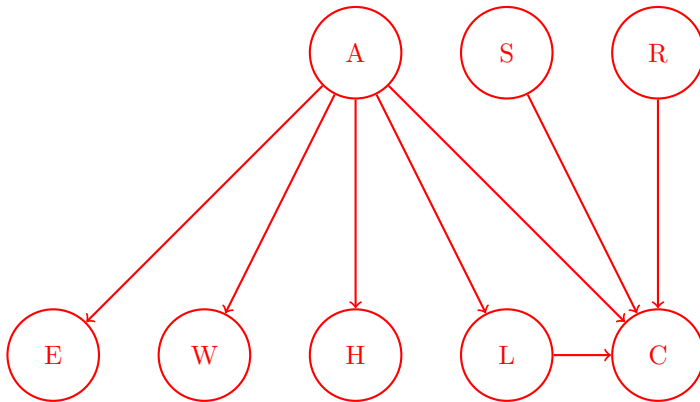
From the tables, we have $A \perp\!\!\!\perp B$ and $C \perp\!\!\!\perp D$. Then, we have every remaining pair of variables: $A \perp\!\!\!\perp C, A \perp\!\!\!\perp D, B \perp\!\!\!\perp C, B \perp\!\!\!\perp D$

You are building advanced safety features for cars that can warn a driver if they are falling asleep (A) and also calculate the probability of a crash (C) in real time. You have at your disposal 6 sensors (random variables):

- E : whether the driver's eyes are open or closed
- W : whether the steering wheel is being touched or not
- L : whether the car is in the lane or not
- S : whether the car is speeding or not
- H : whether the driver's heart rate is somewhat elevated or resting
- R : whether the car radar detects a close object or not

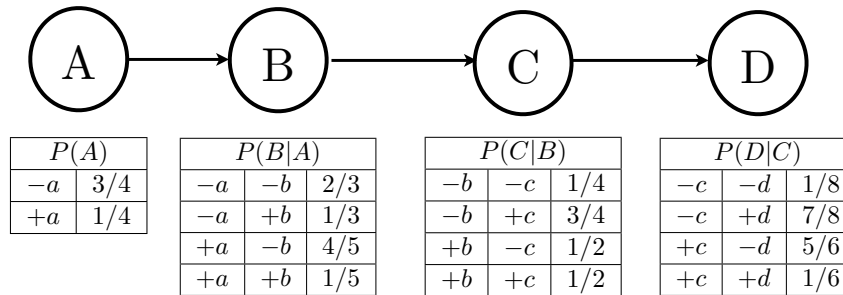
A influences $\{E, W, H, L, C\}$. C is influenced by $\{A, S, L, R\}$.

(d) Draw the Bayes Net associated with the description above by adding edges between the provided nodes where appropriate.



Q2. Bayes' Nets Sampling

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



(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
$-a \quad +b \quad +c \quad -d$	$\underline{P(+b \mid -a)P(-d \mid +c) = 1/3 * 5/6 = 5/18 = 0.277}$
$+a \quad +b \quad +c \quad -d$	$\underline{P(+b \mid +a)P(-d \mid +c) = 1/5 * 5/6 = 5/30 = 1/6 = 0.17}$
$+a \quad +b \quad -c \quad -d$	$\underline{P(+b \mid +a)P(-d \mid -c) = 1/5 * 1/8 = 1/40 = 0.025}$
$-a \quad +b \quad -c \quad -d$	$\underline{P(+b \mid -a)P(-d \mid -c) = 1/3 * 1/8 = 1/24 = 0.042}$

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

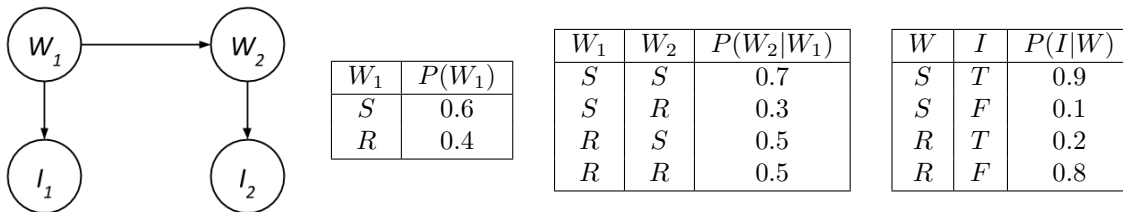
Sequence 4				
1 :	$+a$	$-b$	$-c$	$+d$
2 :	$+a$	$-b$	$-c$	$-d$
3 :	$+a$	$+b$	$-c$	$+d$

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.

3 Sampling and Dynamic Bayes Nets

We would like to analyze people’s ice cream eating habits on sunny and rainy days. Suppose we consider the weather, along with a person’s ice-cream eating, over the span of two days. We’ll have four random variables: W_1 and W_2 stand for the weather on days 1 and 2, which can either be rainy R or sunny S, and the variables I_1 and I_2 represent whether or not the person ate ice cream on days 1 and 2, and take values T (for truly eating ice cream) or F. We can model this as the following Bayes Net with these probabilities.



Suppose we produce the following samples of (W_1, I_1, W_2, I_2) from the ice-cream model:

R, F, R, F R, F, R, F S, F, S, T S, T, S, T S, T, R, F
R, F, R, T S, T, S, T S, T, S, T S, T, R, F R, F, S, T

1. What is $\hat{P}(W_2 = R)$, the probability that sampling assigns to the event $W_2 = R$?
Number of samples in which $W_2 = R$: 5. Total number of samples: 10. Answer $5/10 = 0.5$.
2. Cross off samples above which are rejected by rejection sampling if we’re computing $P(W_2|I_1 = T, I_2 = F)$.

Rejection sampling seems to be wasting a lot of effort, so we decide to switch to likelihood weighting. Assume we generate the following six samples given the evidence $I_1 = T$ and $I_2 = F$:

$$(W_1, I_1, W_2, I_2) = \left\{ (S, T, R, F), (R, T, R, F), (S, T, R, F), (S, T, S, F), (S, T, S, F), (R, T, S, F) \right\}$$

3. What is the weight of the first sample (S, T, R, F) above?

The weight given to a sample in likelihood weighting is

$$\prod_{\text{Evidence variables } e} \Pr(e|\text{Parents}(e)).$$

In this case, the evidence is $I_1 = T, I_2 = F$. The weight of the first sample is therefore

$$w = \Pr(I_1 = T|W_1 = S) \cdot \Pr(I_2 = F|W_2 = R) = 0.9 \cdot 0.8 = 0.72$$

4. Use likelihood weighting to estimate $P(W_2|I_1 = T, I_2 = F)$.

The sample weights are given by

(W_1, I_1, W_2, I_2)	w	(W_1, I_1, W_2, I_2)	w
S, T, R, F	0.72	S, T, S, F	0.09
R, T, R, F	0.16	S, T, S, F	0.09
S, T, R, F	0.72	R, T, S, F	0.02

To compute the probabilities, we thus normalize the weights and find

$$\hat{P}(W_2 = \mathbf{R} | I_1 = \mathbf{T}, I_2 = \mathbf{F}) = \frac{0.72 + 0.16 + 0.72}{0.72 + 0.16 + 0.72 + 0.09 + 0.09 + 0.02} = 0.889$$

$$\hat{P}(W_2 = \mathbf{S} | I_1 = \mathbf{T}, I_2 = \mathbf{F}) = 1 - 0.889 = 0.111.$$