

CS188 Spring 2019 Section 7: Probability and Bayes Nets
(Representation and Inference)

Probability

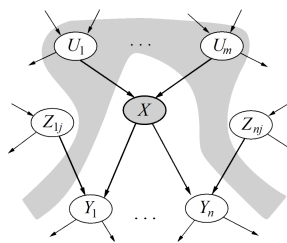
A **random variable** represents an event whose outcome is unknown. A **probability distribution** is an assignment of weights to outcomes. A **joint distribution** over discrete random variables is a table of probabilities which captures the likelihood of each possible **outcome**, also known as an **assignment** of values to the random variables.

To write that random variables X and Y are **marginally independent**, we write $X \perp\!\!\!\perp Y$. To write that random variables X and Y are **conditionally independent** given another random variable Z , we write $X \perp\!\!\!\perp Y | Z$.

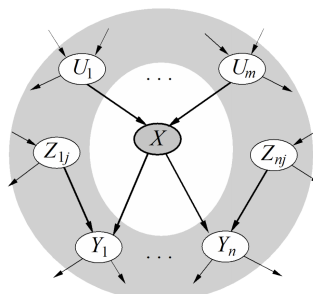
Bayesian Network Representation

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 .

- Each node is conditionally independent of all its ancestor nodes (non-descendants) in the graph, given all of its parents.



- Each node is conditionally independent of all other variables given its Markov blanket. A variable's Markov blanket consists of parents, children, children's other parents.



Bayesian Network Inference

Inference by enumeration

Given a joint PDF, we can trivially perform compute any desired probability distribution $P(Q_1 \dots Q_k | e_1 \dots e_k)$ using a simple and intuitive procedure known as **inference by enumeration**, for which we define three types of variables we will be dealing with:

1. **Query variables** Q_i , which are unknown and appear on the left side of the probability distribution we are trying to compute.
2. **Evidence variables** e_i , which are observed variables whose values are known and appear on the right side of the probability distribution we are trying to compute.
3. **Hidden variables**, which are values present in the overall joint distribution but not in the distribution we are currently trying to compute.

In this procedure, we collect all the rows consistent with the observed evidence variables, sum out all the hidden variables, and finally normalize the table so that it is a probability distribution (i.e. values sum to 1).

Inference by elimination

Let us suppose we have a joint distribution $p(X_1, X_2, X_3, X_4) = p(X_1)p(X_2|X_1)p(X_3|X_1)p(X_4|X_2, X_3)$. We want to compute $p(X_2|X_4 = x_4)$. The query variable is X_2 , the evidence variable is X_4 , and the hidden variables are X_1 and X_3 . The **elimination ordering** we will use is to first eliminate X_3 , then X_1 . When we eliminate a variable X_i , we will create a **factor** f_{X_i} , which is an unnormalized probability distribution. Here are the steps.

1. Write out the marginal distribution between the query and evidence variables $p(X_2, X_4 = x_4)$ by summing over the values of the hidden variables.

$$p(X_2, X_4 = x_4) = \sum_{x_1} \sum_{x_3} p(X_1 = x_1)p(X_2|X_1 = x_1)p(X_3 = x_3|X_1 = x_1)p(X_4 = x_4|X_2, X_3 = x_3)$$

2. Distributed the summations according to the elimination ordering.

$$p(X_2, X_4 = x_4) = \sum_{x_1} p(X_1 = x_1)p(X_2|X_1 = x_1) \sum_{x_3} p(X_3 = x_3|X_1 = x_1)p(X_4 = x_4|X_2, X_3 = x_3)$$

3. Eliminate variables by summing over the conditional probability distributions containing the variable.

- (a) Eliminate X_3 to create factor $f_{X_3}(X_4 = x_4|X_1 = x_1, X_2) = \sum_{x_3} p(X_3 = x_3|X_1 = x_1)p(X_4 = x_4|X_2, X_3 = x_3)$ to obtain

$$p(X_2, X_4 = x_4) = \sum_{x_1} p(X_1 = x_1)p(X_2|X_1 = x_1)f_{X_3}(X_4 = x_4|X_1 = x_1, X_2).$$

- (b) Eliminate X_1 to create factor $f_{X_1}(X_2, X_4 = x_4) = \sum_{x_1} p(X_1 = x_1)p(X_2|X_1 = x_1)f_{X_3}(X_4 = x_4|X_1 = x_1, X_2)$ to obtain

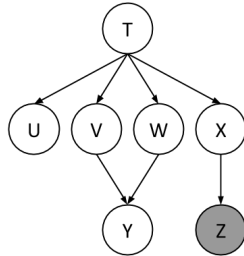
$$p(X_2, X_4 = x_4) = f_{X_1}(X_2, X_4 = x_4).$$

4. Obtain the conditional probability.

$$p(X_2 = x_2|X_4 = x_4) = \frac{f_{X_1}(X_2 = x_2, X_4 = x_4)}{\sum_{x'_2} f_{X_1}(X_2 = x'_2, X_4 = x_4)}.$$

1 Variable Elimination

Using the same Bayes Net (shown below), we want to compute $P(Y \mid +z)$. All variables have binary domains. Assume we run variable elimination to compute the answer to this query, with the following variable elimination ordering: X, T, U, V, W .



Complete the following description of the factors generated in this process:

After inserting evidence, we have the following factors to start out with:

$$P(T), P(U|T), P(V|T), P(W|T), P(X|T), P(Y|V, W), P(+z|X)$$

- (a) When eliminating X we generate a new factor f_1 as follows, which leaves us with the factors:

$$f_1(+z|T) = \sum_x P(x|T)P(+z|x) \quad P(T), P(U|T), P(V|T), P(W|T), P(Y|V, W), f_1(+z|T)$$

- (b) When eliminating T we generate a new factor f_2 as follows, which leaves us with the factors:

- (c) When eliminating U we generate a new factor f_3 as follows, which leaves us with the factors:

- (d) When eliminating V we generate a new factor f_4 as follows, which leaves us with the factors:

- (e) When eliminating W we generate a new factor f_5 as follows, which leaves us with the factors:

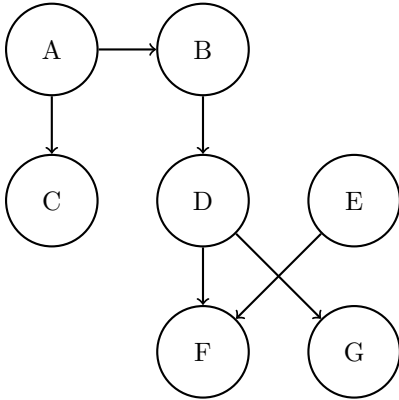
- (f) How would you obtain $P(Y \mid +z)$ from the factors left above:

- (g) What is the size of the largest factor that gets generated during the above process?

- (m) Does there exist a better elimination ordering (one which generates smaller largest factors)?

Q2. Bayes' Nets: Representation and Independence

Parts (a), (b), and (c) 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.

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

A: _____

D: _____

F: _____

(c) Mark the statements that are guaranteed to be true. Recall that every variable is conditionally independent of its non-descendants given its parents, and every variable is conditionally independent of all other variables given its Markov blanket.

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

$F \perp\!\!\!\perp G|D$

$A \perp\!\!\!\perp F$

$B \perp\!\!\!\perp F|D$

$D \perp\!\!\!\perp E|F$

$C \perp\!\!\!\perp G$

$E \perp\!\!\!\perp A|D$

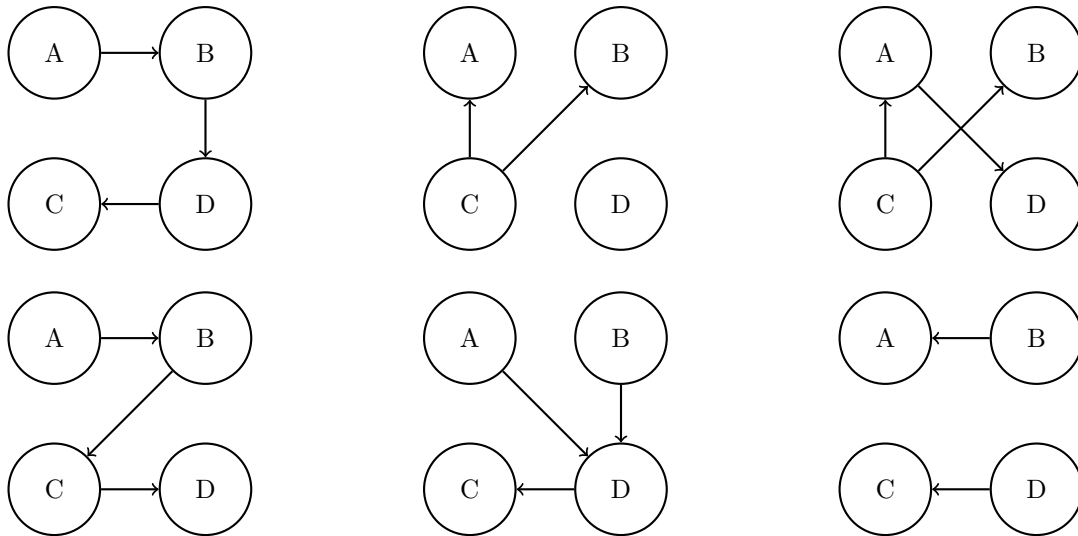
$D \perp\!\!\!\perp E$

Parts (d) and (e) pertain to 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

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

(e) Circle all the Bayes net(s) that can represent a distribution that is consistent with the tables given.

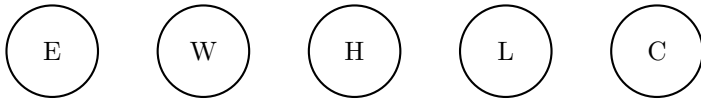
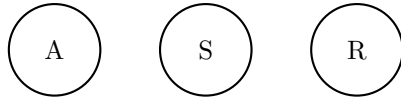


You are building advanced safety features for cars that can warn a driver if they are falling asleep (A) and also calculate the probability of a crash (C) in real time. You have at your disposal 6 sensors (random variables):

- 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 $\{A, S, L, R\}$.

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



(g) Mark all the independence assumptions that must be true.

- | | |
|---|---|
| <input type="checkbox"/> $E \perp\!\!\!\perp S$ | <input type="checkbox"/> $L \perp\!\!\!\perp R C$ |
| <input type="checkbox"/> $W \perp\!\!\!\perp H A$ | <input type="checkbox"/> $W \perp\!\!\!\perp R$ |
| <input type="checkbox"/> $S \perp\!\!\!\perp R$ | <input type="checkbox"/> $A \perp\!\!\!\perp C$ |
| <input type="checkbox"/> $E \perp\!\!\!\perp L$ | <input type="checkbox"/> $E \perp\!\!\!\perp C L$ |

(h) The car's sensors tell you that the car is in the lane ($L = +l$) and that the car is not speeding ($S = -s$). Now you would like to calculate the probability of crashing, $P(C|+l, -s)$. We will use the variable elimination ordering R, A, E, W, H . Write down the largest factor generated during variable elimination. Box your answer.

(i) Write down a more efficient variable elimination ordering, i.e. one whose largest factor is smaller than the one generated in the previous question.