CS 181: Bayesian Networks

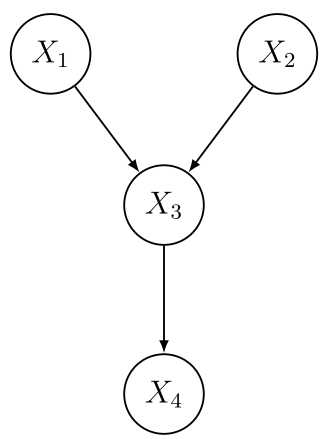
Week of: April 10, 2017
Harvard University

1 Section Objectives

- To understand key concepts including independence assumptions, d-separation, and inference.
- To understand model design choices related to building Bayesian networks.

2 Introduction

A Bayesian network is a graphical model that represents random variables and their dependencies using a directed acyclic graph. Bayesian networks are useful because they allow us to efficiently model joint distributions over many variables by taking advantage of the local dependencies between variables. With Bayesian networks, we can easily reason about conditional independence and perform inference on large joint distributions.



Modeling the joint distribution $p(X_1, X_2, X_3, X_4)$ using the dependencies between X_1, X_3 and X_2, X_3 and X_3, X_4.

3 Network Basics

A patient goes to the doctor for a medical condition, and the doctor suspects 3 diseases as the cause of the condition. The 3 diseases are D_1, D_2, and D_3, and they are independent from each other (given no other observations). There are 4 symptoms S_1, S_2, S_3, and S_4, and the doctor wants to check for presence in order to find the most probable cause. S_1 can be caused by D_1, S_2 can be caused by D_1 and D_2, S_3 can be caused by D_1 and D_3, and S_4 can be caused by D_3. Assume all random variables are Bernoulli, i.e. the patient has the disease/symptom or not.

- **Q:** Draw a Bayesian network for this problem.

- **Q:** Write down the expression for the joint probability distribution given this network.

- **Q:** How many parameters are required to describe this joint distribution?

- **Q:** How many parameters would be required to represent the CPTs in a Bayesian network if there were no conditional independences between variables?

- **Q:** What is an example of the 'explaining away' phenomenon in the compact Bayesian Network?

- **Q:** What diseases do we gain information about when observing the fourth symptom ($S_4 = true$)?

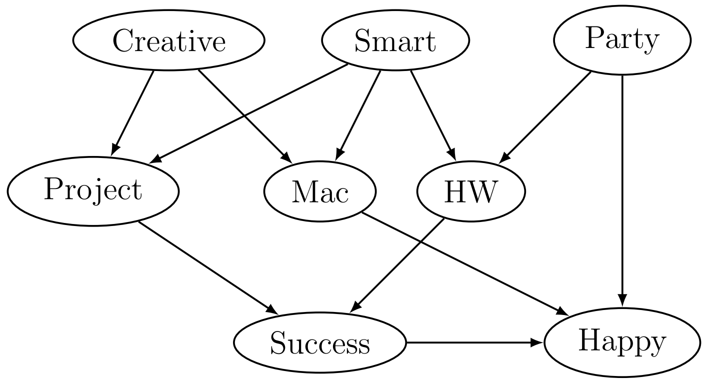
- **Q:** Suppose we know that the third symptom is present ($S_3 = true$). What does observing the fourth symptom ($S_4 = true$) tell us now?

4 D-Separation

As part of a comprehensive study of the role of CS 181 on people's happiness, we have been collecting important data from students. In an entirely optional survey that all students are required to complete, we ask the following highly objective questions:

Do you party frequently [Party: Yes/No]?
Are you smart [Smart: Yes/No]?
Are you creative [Creative: Yes/No]? (Please only answer Yes or No)
Did you do well on all your homework assignments? [HW: Yes/No]
Do you use a Mac? [Mac: Yes/No]
Did your last major project succeed? [Project: Yes/No]
Did you succeed in your most important class? [Success: Yes/No]
Are you currently Happy? [Happy: Yes/No]

After consulting behavioral psychologists we build the following model:



- **Q:** True or False: *Party* is independent of *Success* given HW.

- **Q:** True or False: *Creative* is independent of *Happy* given Mac.

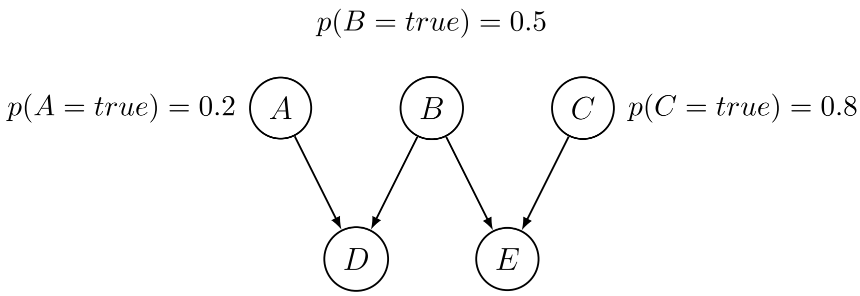
- **Q:** True or False: *Party* is independent of *Smart* given *Success*.

- **Q:** True or False: *Party* is independent of *Creative* given *Happy*.

- **Q:** True or False: *Party* is independent of *Creative* given *Success*, *Project* and *Smart*.

5 Inference

Consider the following Bayesian network, where all variables are Bernoulli.



A	B	$p(D = true \mid A, B)$
F	F	0.9
F	T	0.6
T	F	0.5
T	T	0.1

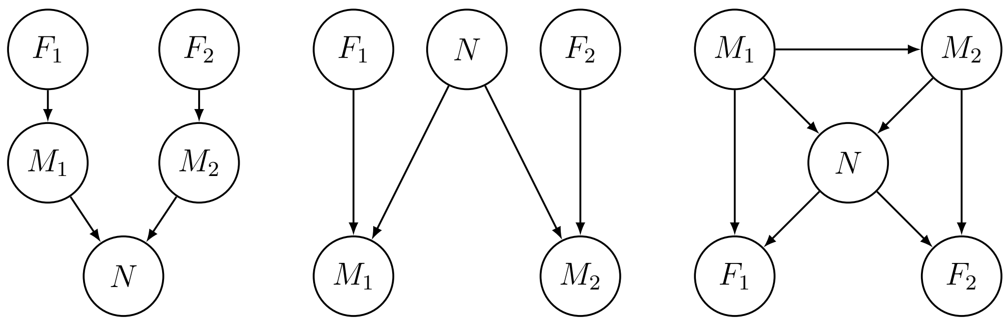
B	C	$p(E = true \mid B, C)$
F	F	0.2
F	T	0.4
T	F	0.8
T	T	0.3

- **Q:** What is the probability that all five variables are simultaneously $false$?

- **Q:** What is the probability that A is $false$ given that the remaining variables are all known to be $true$?

6 Reasoning about the Correctness of Networks

Two astronomers in different parts of the world make measurements M_1 and M_2 of the number of stars N in some small region of the sky. Each telescope may be badly out of focus (events F_1 and F_2). Consider the following Bayesian networks.



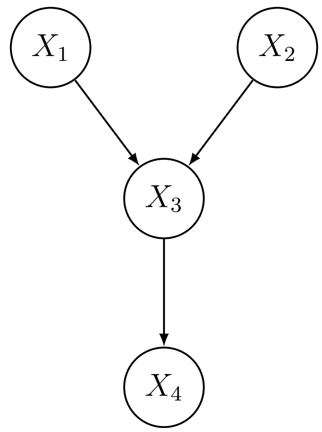
(i) (ii) (iii)

- **Q:** Which of these Bayesian networks correctly represents the distribution?

- **Q:** Of the correct networks, why might one be preferred?

7 Variable Elimination

In this section, we discuss an exact inference algorithm called variable elimination. Consider the Bayesian network we saw in lecture:



Assume that all of the random variables are Bernoulli, meaning their domain is $\{0, 1\}$, and thus the domain size $k = 2$. In this network, we can encode the joint distribution as

$$p(x_1, x_2, x_3, x_4) = p(x_3|x_1, x_2)p(x_4|x_3)p(x_1)p(x_2) \tag{1}$$

If we wanted to calculate the marginal distribution of X_4, we could naively marginalize out the other variables, giving us

$$p(x_4) = \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_3|x_1, x_2)p(x_4|x_3)p(x_1)p(x_2) \tag{2}$$

Calculating this naively requires multiplying 4 values for each of the 8 possible combinations of x_1, x_2, x_3. In general, if there were many variables then the number of combinations would grow exponentially in the number of variables!

However, note that because we have a compact, Bayesian net representation, we can calculate the marginal distribution more efficiently. By reordering the sums and eliminating one variable at a time, we derive the variable elimination procedure. For example, we can calculate the joint distribution as:

$$p(x_4) = \sum_{x_1,x_2,x_3} p(x_3|x_1,x_2)p(x_4|x_3)p(x_1)p(x_2) \tag{3}$$

$$= \sum_{x_2,x_3} p(x_4|x_3)p(x_2) \sum_{x_1} p(x_3|x_1,x_2)p(x_1) \tag{4}$$

$$= \sum_{x_3} p(x_4|x_3) \sum_{x_2} p(x_2)p(x_3|x_2) \tag{5}$$

$$= \sum_{x_3} p(x_4|x_3)p(x_3) \tag{6}$$

$$= p(x_4) \tag{7}$$

Here, we eliminate x_1, then x_2, then x_3. This is working in 'leaves first' order towards the query, x_4. Alternatively, we could have eliminated variables in a different order, as follows:

$$p(x_4) = \sum_{x_1,x_2,x_3} p(x_3|x_1,x_2)p(x_4|x_3)p(x_1)p(x_2) \tag{8}$$

$$= \sum_{x_1,x_2} p(x_1)p(x_2) \sum_{x_3} p(x_3|x_1,x_2)p(x_4|x_3) \tag{9}$$

$$= \sum_{x_1} p(x_1) \sum_{x_2} p(x_2)p(x_4|x_1,x_2) \tag{10}$$

$$= \sum_{x_1} p(x_1)p(x_4|x_1) \tag{11}$$

$$= p(x_4) \tag{12}$$

Here, we eliminate x_3 then x_3 then x_1. For the following questions, assume the following CPTs:

x_1	$p(x_1)$
0	0.3
1	0.7

x_2	$p(x_2)$
0	0.6
1	0.4

x_3	x_1	x_2	$p(x_3\|x_1,x_2)$
0	0	0	0.5
0	0	1	0.2
0	1	0	0.9
0	1	1	0.5
1	0	0	0.5
1	0	1	0.8
1	1	0	0.1
1	1	1	0.5

x_4	x_3	$p(x_4\|x_3)$
0	0	0.7
0	1	0.1
1	0	0.3
1	1	0.9

- **Q:** Following the first ordering, first use variable elimination on X_1 to compute the CPT for $p(X_3|X_2)$. This represents the first intermediate term. Draw the resulting Bayesian network.

- **Q:** Using the second ordering, we would first use variable elimination on X_3 to compute the probability table for $p(X_4|X_1, X_2)$. Compute the CPT and draw the resulting Bayesian network.

- **Q:** How many sum-product calculations do each of these variable elimination orders require? Which one is preferable?