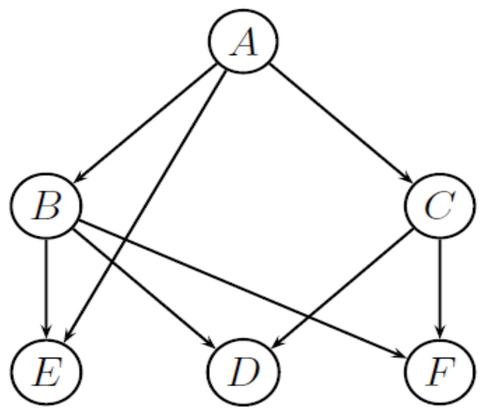
Homework VI-KEY

1. A graph with no links is a trivial D-Map. True/False

True. A graph is said to be D- Map of a distribution if all CI satisfied by distribution is reflected on the graph, of course graph with no links will reflect any conditional independency.

2. Consider the Bayesian network given below



a. Is A conditionally independent of D give {B,C}.

   Yes,

b. Is E marginally independent of F

   No

c. Which edge would you delete to make A independent of C.

   The edge A->C

3. Evaluate the distribution p(a), p(b|c) and p(c|a) corresponding to the joint distribution given in the Table. Hence show by direct evaluation that p(a,b,c) = p(a) p(c|a) p(b|c). Draw the corresponding directed graph.

a	b	c	p(a, b, c)
0	0	0	0.192
0	0	1	0.144
0	1	0	0.048
0	1	1	0.216
1	0	0	0.192
1	0	1	0.064

1	1	0	0.048
1	1	1	0.096

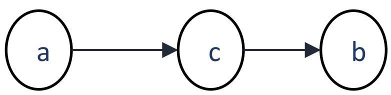
Tables for p(a), p(c|a), p(b|c) evaluated by marginalizing and conditioning the joint distribution from the given table.

a	p(a)
0	0.6
1	0.4

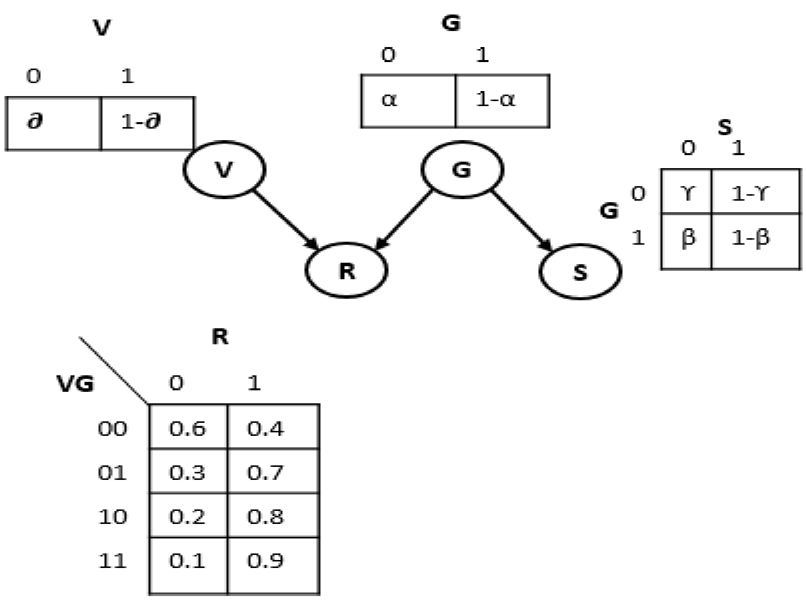
a	c	p(c|a)
0	0	0.4
0	1	0.6
1	0	0.6
1	1	0.4

c	b	p(b|c)
0	0	0.8
0	1	0.2
1	0	0.4
1	1	0.6

Multiplying the three distribution together we recover the joint distribution p(a, b, c)given in the table, thereby allowing us to verify the validity of the decomposition p(a, b, c) = p(a) * p(c|a) * p(b|c). We can express the distibution using the graph.



4. Consider the directed graphical model in following figure with 4 binary variables.



a. Write down the expression for P(S=1|V=1) in terms of α, β, ϒ, ∂.

b. Write down the expression for P(S=1|V=0) . Is it the same or different to P(S=1|V=1? Explain why.

c. Find the maximum likelihood estimate of α, β, ϒ using the following dataset, where each row is a training case.

V	G	R	S
1	1	1	1
1	1	0	1
1	0	0	0

a. The expression for P(S =1 | V=1) is

$$P(S = 1 \mid V = 1) = \frac{P(S = 1, V = 1)}{P(V = 1)}$$

$$= \frac{1}{P(V = 1)} \sum_{R=0}^{1} \sum_{G=0}^{1} P(V = 1) * P(G) * P(R|V = 1, G) * P(S = 1|G)$$

$$= \sum_{RG} P(G) * P(R|V = 1, G) * P(S = 1|G)$$

$$= \sum_{G} P(G) * P(S = 1|G) \sum_{R} P(R|V = 1, G)$$

$$= \sum_{G} P(G) * P(S = 1|G) * 1$$

$$= P(G = 0) * P(S = 1 \mid G = 0) + P(G = 1) * P(S = 1 \mid G = 1)$$

$$= \alpha(1 - \gamma) + (1 - \alpha) * (1 - \beta)$$

$$= 1 - \alpha\gamma + \alpha\beta - \beta$$

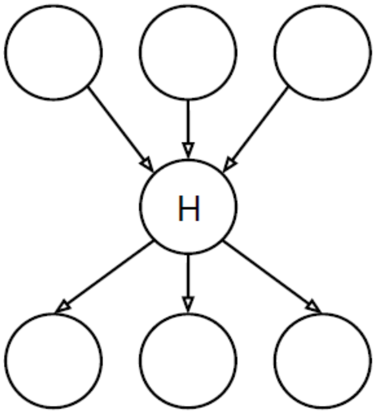
b. We find P(S=1 | V=1) and P(S=1 | V=0) are same because they are independent of V

c. MLE can be estimated by couting events. Thus $\alpha = \frac{1}{3}$ $\beta = 0$ and $\gamma = 1$

V and S are independent given nothing
V and S are independent given G
V and S are independent given R&G
V and S are dependent given R.

5. Hidden variables in DGMs:

a. Consider the following graphical model, where we number nodes left to right, top to bottom.



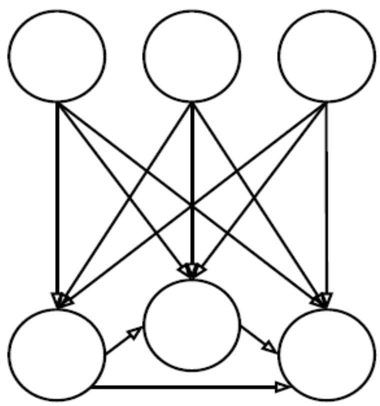
The graph defines the joint as

$$P(X_1, X_2, X_3, X_4, X_5, X_6) = \sum_h P(X_1)P(X_2)P(X_3)p(H = h|X_1X_2X_3)P(X_4|H = h)P(X_5|H = h)P(X_6|H = h)$$

where we have marginalized over the hidden variable H.

Assuming all nodes are binary, how many parameters does this model have?

b. Consider the following graph and its joint distribution (again we number nodes from left to right and from top to bottom)



$$P(X_1, X_2, X_3, X_4, X_5, X_6) = P(X_1)P(X_2)P(X_3)\, P(X_4|X_1, X_2, X_3)P(X_5|X_1, X_2, X_3, X_4)P(X_6|X_1, X_2, X_3, X_4, X_5)$$

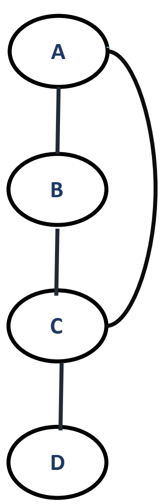
Assuming all nodes are binary, how many parameters does this model have?

a. For the graph on the left, the CPDs for nodes 1,2,3 have 1 free parameter each (since they are Bernoulli). p(H|X1:3) has 8 free parameters, one per conditioning case. p(Xi|H) for i = 4 : 6 are 2 * 2 tables, but due to the sum to one constraint, only have 2 free parameters. Hence in total there are 3 * 1 + 8 + 3 * 2 = 17.

b. For the graph on the right, the CPDs for nodes 1,2,3 have 1 free parameter each (since they are Bernoulli). $p(X_4|X_{1:3})$ has 8 parameters, one per conditioning case. $p(X_5|X_{1:4})$ has 16 parameters. $p(X_6|X_{1:5})$ has 32 parameters. In total there are 3 + 8 + 16 + 32 = 59 parameters.

6. What is the complexity of computing $P(E = e)$ using variable elimination in the following Bayesian network along the ordering (A, B, C, D) The edges in the Bayesian network are $A \rightarrow B, A \rightarrow C, B \rightarrow C, C \rightarrow D \; and \; D \rightarrow E.$

The functions after instantiating evidence variable are P(B|A), P(C|A, B), P(D|C) and function of D. The induced graph along the ordering ABCD is shown below. The number of children for A, B, C and D are 2, 1, 1 and 0 respectively, so the width of the tree is 2. The complexity of the variable elimination algorithm is O(n exp(w+1), where n is no of non-evidence variables and w is the width of the ordering. Therefore the complexity is O(4 exp(3)).



7. What is the complexity of computing $P(E = e)$ using variable elimination in the following Bayesian network along the ordering (B, C, D, A). The edges in the Bayesian network are $A \rightarrow B, B \rightarrow C, C \rightarrow D \; and \; D \rightarrow E.$

The functions after instantiating evidence variable are P(B|A), P(C|B), P(D|C), P(A) and function of D. The induced graph along the ordering BCDA is shown below. The number of children for B, C, D and A are 2, 2, 1 and 0 respectively, so the width of the tree is 2. The complexity of the variable elimination algorithm is O(n exp(w+1), where n is no of non-evidence variables and w is the width of the ordering. Therefore the complexity is O(4 exp(3)).