PROBABILISTIC GRAPHICAL MODELS

SESSION # 16 : Some Problems

SRINATH NAIDU

srinath.naidu@pilani.bits-pilani.ac.in

BITS Pilani

Pilani | Dubai | Goa | Hyderabad

The instructor is gratefully acknowledging the authors who made their course materials freely available online.

Problem 1

Consider the following Gibbs distribution

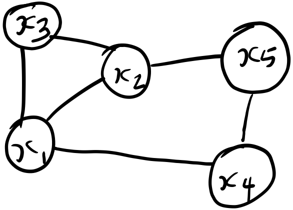
$$P(x_1, x_2, \dots x_5) = \phi_1(x_1, x_2)\,\phi_2(x_1, x_3)\,\phi_3(x_1, x_4)\,\phi_4(x_2, x_3)\,\phi_4(x_2, x_5)\,\phi_5(x_4, x_5)$$

1) Visualise this distribution as an undirected graph

2) Do we have $x_3 \perp x_4 \mid x_1, x_2$?

Answer - Problem 1

We draw a vertex for each variable x_i. There is an edge between two variables if the variables occur together in a factor. This gives us



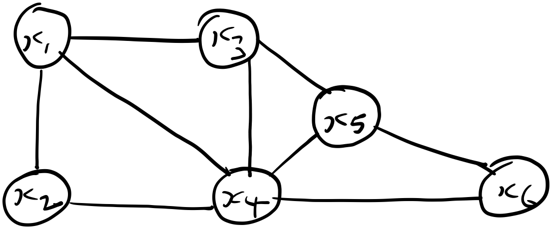
Answer - Problem 1

Given x_1 and x_2 there is no active path between x_3 and x_4.

Therefore $x_3 \perp x_4 \mid x_1, x_2$

Problem 2

Consider the undirected graph G below:



a) what is the set of Gibbs distributions induced by the graph?

b) if P factorizes according to G, does $P(x_3 / x_2, x_4) = P(x_3 / x_4)$ hold

Answer - Problem 2

a) The maximal cliques in the graph are (x_1, x_2, x_4), (x_1, x_3, x_4), (x_3, x_4, x_5), (x_4, x_5, x_6)

The Gibbs distribution induced by G is

$$p(x_1, x_2, x_3, x_4, x_5, x_6) \propto \phi_1(x_1, x_2, x_4)\,\phi_2(x_1, x_3, x_4)\,\phi_3(x_3, x_4, x_5)\,\phi_4(x_4, x_5, x_6)$$

Answer - Problem 2

To answer part (b) we need to check if x_3 is conditionally independent of x_2 given x_4. In graph G, there exists an active path between x_3 and x_2 even when x_4 is observed. Therefore,

$$P(x_3 \mid x_2, x_4) \neq P(x_3 \mid x_4).$$

Problem 3

Assume that you have the following Markov Blankets for all variables $x_1, x_2, x_3, x_4, y_1, y_2, \dots y_4$

$MB(x_1) = \{x_2, y_1\}$, $MB(x_2) = \{x_1, x_3, y_2\}$, $MB(x_3) = \{x_2, x_4, y_3\}$

$MB(x_4) = \{x_3, y_4\}$, $MB(y_1) = \{x_1\}$, $MB(y_2) = \{x_2\}$

$MB(y_3) = \{x_3\}$, $MB(y_4) = \{x_4\}$

Let p be the corresponding pdf. How does p factorize?

(Assume p is $+$ve).

Answer - Problem 3

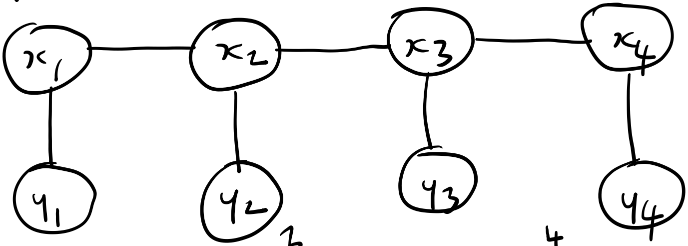
The key idea is that in undirected graphical models the Markov Blanket for a node is its set of neighbours

Given all the Markov Blankets we know what local Markov property p must satisfy

For positive distributions we have an equivalence between p satisfying the local Markov property and p factorizing over the graph

Answer - Problem 3

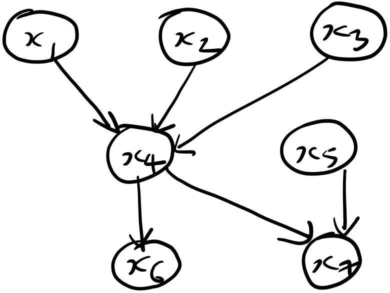
The graph satisfying the Markov Blanket relationships is



The factorization is $\frac{1}{Z}\prod_{i=1}^{3} m(x_i, x_{i+1}) \prod_{i=1}^{4} g(x_i, y_i)$

Problem 4

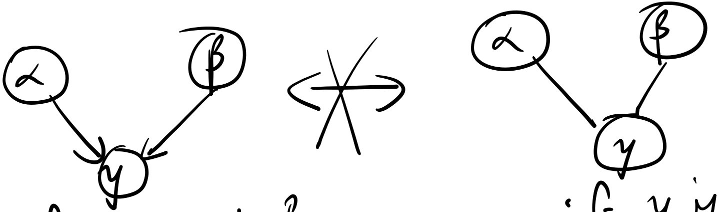
For distributions that factorize over the graph below, find the minimal undirected I-map.



Answer - Problem 4

We construct a ~~moralized graph~~ where the unmarried parents of a vertex are married.

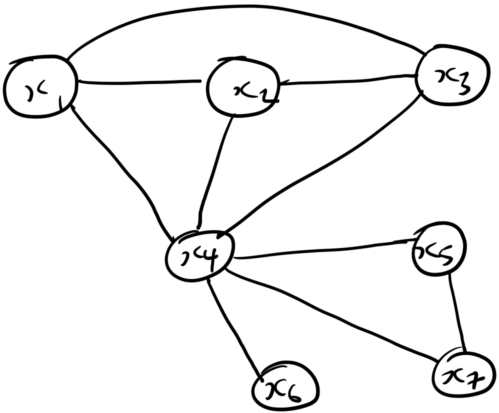
Why do we do this?



if y is specified, α and β are not independent in the directed graph

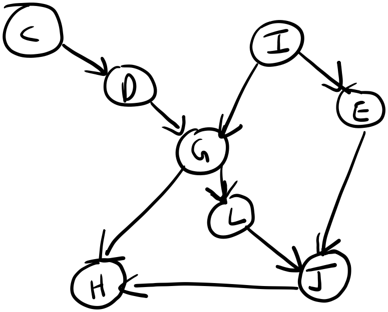
∴ If y is specified α and β are independent

Answer - Problem 4



Problem 5

Consider a DAG G associated with a student's performance in some course



C = coherence
I = intelligence
D = difficulty
G = grade
L = letter of recommendation
J = job
E = exam score
H = happy

Problem 5

Construct a clique tree for this example with respect to an instance of variable elimination where Φ is the set of conditional probabilities $P(x_i / pa_G(x_i))$ associated to G, $Z = \{C, D, I, G, E, L, H\}$ and $<$ is the ordering C, D, I, H, G, E, L

Answer- Problem 5

We first run variable elimination on this DAG for the set of factors $\phi = \{P(C), P(D/C), P(I), P(G/D,I), P(E/I), P(L/G), P(J/E,L), P(H/G,J)\}$

and the ordering C, D, I, H, G, E, L. We then compute

(a) $\psi_1(C,D) = P(C)\,P(D/C)$, $\tau_1(D) = \sum_C \psi_1(C,D)$

(b) $\psi_2(G,D,I) = \tau_1(D)\,P(G/D,I)$, $\tau_2(G,I) = \sum_D \psi_2(G,D,I)$

Answer - Problem 5

(c) $\psi_3(G, E, I) = \tau_2(G, I)\, P(I)\, P(E/I)$, $\tau_3(G, E) = \sum_{I} \psi_3(G, E, I)$

(d) $\psi_4(H, G, J) = P(H/G, J)$, $\tau_4(G, J) = \sum_{H} \psi_4(H, G, J)$

(e) $\psi_5(G, J, E, L) = \tau_4(G, J)\, \tau_3(G, E)\, P(L/G)$, $\tau_5(J, E, L) = \sum_{G} \psi_5(G, J, E$

(f) $\psi_6(J, E, L) = \tau_5(J, E, L)\, P(J/E, L)$, $\tau_6(J, L) = \sum_{E} \psi_6(J, E, L)$

(g) $\psi_7(J, L) = \tau_6(J, L)$, $\tau_7(J) = \sum_{L} \psi_7(J, L)$

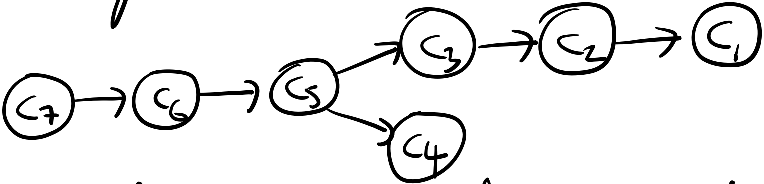
Answer - Problem 5

We then create one clique for each φ_i consisting of the variables in $Scope(\varphi_i)$

(a) $C_1 = \{C, D\}$
(b) $C_2 = \{D, G, J\}$
(c) $C_3 = \{G, E, I\}$
(d) $C_4 = \{H, G, J\}$
(e) $C_5 = \{G, J, E, L\}$
(f) $C_6 = \{J, E, L\}$
(g) $C_7 = \{J, L\}$

Answer - Problem 5

We then create a clique tree by connecting cliques C_i and C_j by the edges $C_i \leftarrow C_j$ if ϕ_j is defined by τ_i. We get the following clique tree:



Then we specify a directed clique tree, taking C_7 as the root

Problem 6

Let ϕ denote the set of factors in the previous problem. Consider the clique tree

$$C_1 - C_2 - C_3 - C_5 - C_4$$

where $C_1 = \{C, D\}$, $C_2 = \{D, I, G\}$, $C_3 = \{G, E, I\}$,

$C_4 = \{G, H, J\}$, $C_5 = \{G, J, L, E\}$

Assign each factor a clique: $\alpha(P(C)) = C_1$, $\alpha(P(D/C)) = C_1$,

$\alpha(P(G/I, D)) = C_2$, $\alpha(P(I)) = C_3$, $\alpha(P(E/I)) = C_3$, $\alpha(P(H/G, J)) = C_4$

Problem 6

$\alpha(P(L/G)) = C_5$, $\alpha(P(J/L, E)) = C_5$.

(a) What are the initial potentials of the given clique tree?

(b) With C_5 as root compute $\beta_r(C_1)$ using sum-product messaging.

Answer - Problem 6

(a) The initial potential ψ_i for clique C_i is obtained by multiplying together all the factors assigned to C_i

$$\psi_1 = P(C)\,P(D/C)$$

$$\psi_2 = P(G/I,D)$$

$$\psi_3 = P(I)\,P(E/I)$$

$$\psi_4 = P(H/G,J)$$

$$\psi_5 = P(L/G)\,P(J/L,E)$$

Answer - Problem 6

(b) We fix C_5 to be the root and reorient edges of the clique tree so that C_5 is the root

$$C_1 \leftarrow C_2 \leftarrow C_3 \leftarrow C_5 \rightarrow C_4$$

The sepsets are respectively $S_{1,2} = \{D\}$, $S_{2,3} = \{G, I\}$, $S_{3,5} = \{G, E\}$ $S_{4,5} = \{G, J\}$

Answer - Problem 6

The message $\delta_{i \to j}$ is computed by summing out the variables in C_i that are not in $S_{i,j}$ which labels the edge between C_i and C_j

$$\delta_{1 \to 2}(D) = \sum_{C} \psi_1(C, D)$$

$$\delta_{2 \to 3}(G, I) = \sum_{D} \psi_2(G, I, D)\, \delta_{1 \to 2}(D)$$

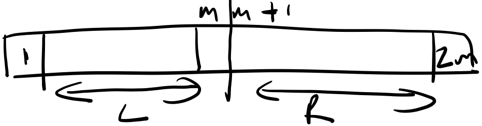
$$\delta_{3 \to 5}(G, E) = \sum_{I} \psi_3(E, I)\, \delta_{2 \to 3}(G, I)$$

$$\delta_{4 \to 5}(G, J) = \sum_{H} \psi_4(G, H, J)$$

Answer - Problem 6

Hence the beliefs $\beta_5(G, J, L, E) = \psi_5(G, J, L, E)\, \delta_{3\to5}(G, J)\, \delta_{4\to5}(G, J)$

We can only compute β at the root since it has received all its messages



Problem 7

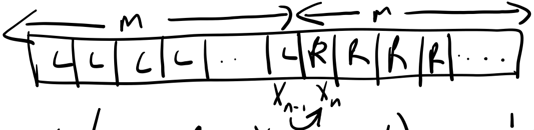
A mouse moves along a tiled corridor with $2m$ tiles where $m > 1$. When $i \neq 1, 2m$, it moves to tile $i+1$ or $i-1$ with equal probability. From the boundary tiles it moves right or left with probability 1.

Each time the mouse moves to a tile $i \leq m$ or $i > m$ an electronic device outputs a signal L or R respectively.

Can the generated sequence of signals L and R be described as a Markov chain with states L and R?

Answer - Problem 7

To see if the sequence is a Markov chain we need to check if the next state is determined only by the current state

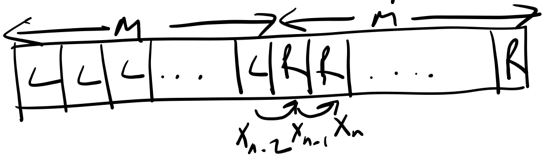


$$P(X_{n+1}=L \mid X_n=R,\ X_{n-1}=L) = \frac{1}{2}$$

Since we are on the border, going left will generate L and going right will generate R

Answer - Problem 7

what about $P(X_{n+1}=L / X_n = R, X_{n-1}=R, X_{n-2}=L)$?



We are now deep into R territory so that whether we move left or right from X_n we will not see a L.

$\therefore P(X_{n+1}=L / X_n=R, X_{n-1}=R, X_{n-2}=L) = 0$

Problem 8

Consider the same scenario as in <u>Problem 7</u>
except that the device outputs L or R when the
mouse moves to tile 1 or tile 2m, and not when
$i \leq m$ or $i > m$.
Can the generated sequence of L and R now be
described as a Markov chain with states L and R?

Answer - Problem 8

In this case the sequence of states can be described as a Markov chain.

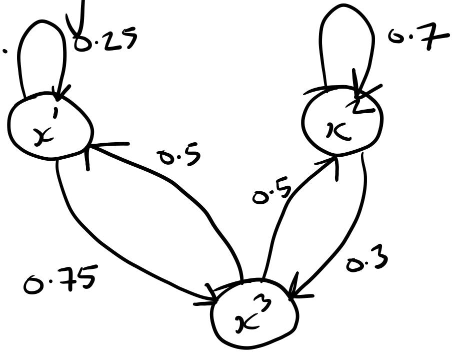
At any time t given that the mouse has just moved to position 1, the probability of any event that concerns its position in future depends only on its current state

$$P(X_{n+1} = x_{n+1} / X_n = L, X_{n-1} = x_{n-1}, \dots X_1 = x_1)$$

$$= P(X_{n+1} = x_{n+1} / X_n = L)$$ and similarly for $X_n = R$

Problem 9

Find the stationary distribution for the following Markov chain.



Answer - Problem 9

For a stationary distribution $p^{t}(x') \approx p^{t+1}(x')$

where $p^{t+1}(x') = \sum_{x \in Val(X)} p^{t}(x) T(x \to x')$

$$\left.\begin{aligned} \pi(x^1) &= 0.25\,\pi(x^1) + 0.5\,\pi(x^3) \\ \pi(x^2) &= 0.7\,\pi(x^2) + 0.5\,\pi(x^3) \\ \pi(x^3) &= 0.75\,\pi(x^1) + 0.3\,\pi(x^2) \end{aligned}\right\} \text{three transition equations}$$

$$\pi(x^1) + \pi(x^2) + \pi(x^3) = 1$$

Answer: Problem 9

We can set up the problem in matrix terms as follows:

$$\begin{bmatrix} 0.25 & 0 & 0.5 \\ 0 & 0.7 & 0.5 \\ 0.75 & 0.3 & 0 \end{bmatrix} \begin{bmatrix} \pi(x^1) \\ \pi(x^2) \\ \pi(x^3) \end{bmatrix} = \begin{bmatrix} \pi(x^1) \\ \pi(x^2) \\ \pi(x^3) \end{bmatrix} \Rightarrow A x = x$$

We need to fit the eigenvector to the matrix above which corresponds to eigenvalue 1. Verify that any matrix whose columns add up to 1 will have eigenvalue 1 innovate achieve lead

Answer - Problem 9

$Ax = \lambda x \Rightarrow \det(A - \lambda I) = 0 \Rightarrow \det(A - I) = 0$ for $\lambda = 1$

$(A - I)x = 0$

$$\begin{bmatrix} -0.75 & 0 & 0.5 \\ 0 & -0.3 & 0.5 \\ 0.75 & 0.3 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

After Gaussian elimination

we get ...

Answer - Problem 9

$$\begin{bmatrix} -0.75 & 0 & 0.5 \\ 0 & -0.3 & 0.5 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

x_3 is a free variable; x_1 and x_2 are pivot variables

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_3/3 \\ 5x_3/3 \\ x_3 \end{bmatrix}$$

Answer - Problem 9

How do we decide the value of x_3?

$$\begin{bmatrix} \pi(x^1) \\ \pi(x^2) \\ \pi(x^3) \end{bmatrix} = \begin{bmatrix} 2\alpha/3 \\ 5\alpha/3 \\ \alpha \end{bmatrix}$$

$$\pi(x^1) + \pi(x^2) + \pi(x^3) = 1$$

$$\frac{7\alpha}{3} + \alpha = 1$$

$$\Rightarrow \alpha = \frac{3}{10}$$

$$\begin{bmatrix} \pi(x^1) \\ \pi(x^2) \\ \pi(x^3) \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.5 \\ 0.3 \end{bmatrix}$$

Problem 10

Build a Chow-Liu tree for the following empirical data

A	B	C	D
0	0	1	0
0	0	1	1
0	1	0	0
1	0	0	1
0	0	1	1

Answer - Problem 10

Obtain weights for every pair of vertices where the weight = mutual information

For example

$$I_{A,B} = \sum_{A,B} p(A,B) \log_2 \frac{p(A,B)}{p(A)\,p(B)}$$

The various probabilities are determined empirically

Answer - Problem 10

Let us calculate the edge weight between A and B.

$$I_{A,B} = \sum_{M} p(A,B) \log_2 \frac{p(A,B)}{p(A)\,p(B)}$$

We have $p(A=0, B=0) = \frac{3}{5}$ $\quad p(A=0, B=1) = \frac{1}{5}$

$p(A=1, B=0) = \frac{1}{5}$

$p(A=0) = \frac{4}{5}$

$p(B=0) = \frac{4}{5}$

innovate achieve lead

Answer - Problem 10

$$I(A,B) = p_{A,B}(0,0)\log_2\frac{p_{AB}(0,0)}{p_A(0)p_B(0)} + p_{AB}(0,1)\log_2\frac{p_{AB}(0,1)}{p_A(0)p_B(1)}$$

$$+ p_{AB}(1,0)\log_2\frac{p_{AB}(1,0)}{p_A(1)p_B(0)}$$

$$= \frac{3}{5}\log_2\frac{\frac{3}{5}}{\left(\frac{4}{5}\right)\left(\frac{4}{5}\right)} + \frac{1}{5}\log_2\frac{\frac{1}{5}}{\left(\frac{4}{5}\right)\left(\frac{1}{5}\right)} + \frac{1}{5}\log_2\frac{\frac{1}{5}}{\frac{4}{5}\,\frac{1}{5}}$$

$$= 0.07$$

Answer - Problem 1c

Similarly for $I_{A,C}$

$P(A=0, C=0) = \frac{1}{5}$

$P(A=0, C=1) = \frac{3}{5}$

$P(A=1, C=0) = \frac{1}{5}$

$P(A=1, C=1) = \frac{0}{5}$

$P(A=0) = \frac{4}{5}$

$P(C=0) = \frac{2}{5}$

Answer - Problem 10

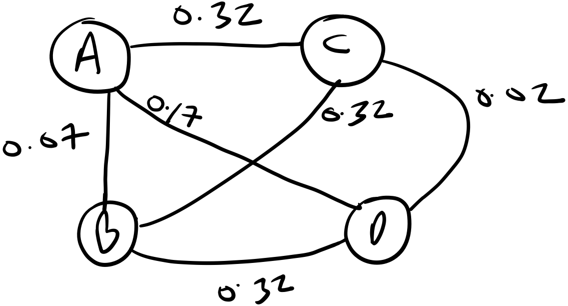
$$I_{A,C} = P_{AC}(0,0)\log_2\frac{P_{AC}(0,0)}{P_A(0)P_C(0)} + P_{AC}(0,1)\log_2\frac{P_{AC}(0,1)}{P_A(0)P_C(1)} + P_{AC}(1,0)\log_2\frac{P_{AC}(1,0)}{P_A(1)P_C(0)}$$

$$= \frac{1}{5}\log\frac{\frac{1}{5}}{\frac{8}{25}} + \frac{3}{5}\log_2\frac{\frac{3}{5}}{\left(\frac{4}{5}\right)\left(\frac{3}{5}\right)} + \frac{1}{5}\log_2\frac{\frac{1}{5}}{\left(\frac{1}{5}\right)\left(\frac{2}{5}\right)}$$

$$= 0.32$$

Answer - Problem 10

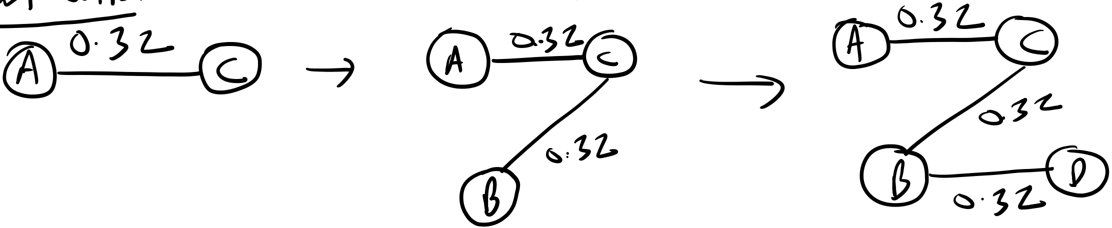
We get the following graph:



Answer - Problem 10

To construct maximum weight spanning tree:

start with



Each time we add the max weight edge that does not create a cycle

Problem 11

Prior information about the parameters of a biased coin with sample space $[x^1, x^2]$ suggests that the parameter vector (θ^1, θ^2) obeys a Dirichlet distribution with parameters 2 and 3 respectively. Let θ_1 stand for the outcome x^1 and θ_2 for the outcome x^2. What is $P[X[1] = x^1]$? What is the smallest number of samples M that we need in order to conclude that $P(X[M+1] = x^1 / D) = \frac{1}{2}$ where $D: X[1], X[2] \ldots X[M]$?

Answer – Problem 11

We have $P(X[1] = x^1) = \frac{\alpha_1}{\alpha_1 + \alpha_2} = \frac{2}{2+3} = \frac{2}{5}$

Further $P(X[M+1] = x^1 / D) = \frac{M[1] + \alpha_1}{M + \alpha_1 + \alpha_2}$ where $M = M[1] + M[2]$

We need $P(X[M+1] = x^1 / D) = \frac{1}{2}$

This means

$$\frac{M[1] + 2}{M[1] + M[2] + 5} = \frac{1}{2}$$

$$\Rightarrow \quad 2M[1] + 4 = M[1] + M[2] + 5$$

Answer - Problem 11

This gives $M[1] = M[2] + 1$.

The smallest possible value for $M[2] = 0$ which gives

$M[1] = 1$

Thus $M = M[1] + M[2] = 1$ is the smallest number of samples we need in order to conclude that

$$P(X[M+1] = x^1 / D) = \frac{1}{2}$$