Stat 246 Homework V - Solution

CA: Darongsae Kwon

1 [60 points]

(a)

The joint distribution of X_0, X_1, X_2 has the following form.

$$P(x_0, x_1, x_2) = \frac{1}{Z} \exp\left(\sum_{c \in \mathcal{C}} \psi_c(x_c)\right),$$

where $\mathcal{C} = \{(0,1),(1,2)\}$, the set of all maximal cliques. The normalizing constant Z is

$$\begin{aligned} Z &= \sum_{x_0}\sum_{x_1}\sum_{x_2} \exp\left(\alpha x_0 x_1 + \alpha x_1 x_2\right) \\ &= \sum_{x_0}\sum_{x_2}\{1 + \exp(\alpha x_0 + \alpha x_2)\} \\ &= \sum_{x_0}\{2 + \exp(\alpha x_0) + \exp(\alpha x_0 + \alpha)\} \\ &= 5 + 2\exp(\alpha) + \exp(2\alpha). \end{aligned}$$

The marginal distribution on X_0 is

$$\begin{aligned} P(x_0) &= \sum_{x_1}\sum_{x_2} P(x_0, x_1, x_2) \\ &= \sum_{x_1}\sum_{x_2} \exp\left(\alpha x_0 x_1 + \alpha x_1 x_2\right)/Z \\ &= \frac{2 + \exp(\alpha x_0) + \exp(\alpha(x_0 + 1))}{5 + 2\exp(\alpha) + \exp(2\alpha)}. \end{aligned}$$

(b)

We have the joint distribution of X_0, X_1, X_2 given by

$$\begin{aligned} P(x_0, x_1, x_2) &= \frac{\exp(\alpha x_0 x_1 + \alpha x_1 x_2)}{5 + 2\exp(\alpha) + \exp(2\alpha)} \\ &= \frac{\exp(\alpha(x_0 x_1 + x_1 x_2 - 2))}{5\exp(-2\alpha) + 2\exp(-\alpha) + 1}. \end{aligned}$$

Therefore,

$$
\begin{aligned}
P(x_0, x_1, x_2) &\rightarrow \frac{1}{8} \quad \text{as } \alpha \rightarrow 0, \\
P(x_0, x_1, x_2) &\rightarrow \begin{cases} 1\,, & \text{if } x_0 = x_1 = x_2 = 1, \\ 0\,, & \text{otherwise,} \end{cases} \qquad \text{as } \alpha \rightarrow \infty.
\end{aligned}
$$

(c)

Suppose the rows and columns of $M = \begin{pmatrix} 1 & 1 \\ 1 & e^{\alpha} \end{pmatrix}$ are indexed by $0, 1$, that is, $M_{0,0} = M_{0,1} = M_{1,0} = 1$ and $M_{1,1} = e^{\alpha}$. Then we have $M_{X_i, X_{i+1}} = \exp(\alpha X_i X_{i+1})$. Since $(AB)_{i,j} = \sum_k A_{i,k} B_{k,j}$ for any A, B, we have

$$
\begin{aligned}
Z &= \sum_{x_0, x_1, x_2} \exp(\alpha x_0 x_1 + \alpha x_1 x_2) \\
&= \sum_{x_0, x_2} \sum_{x_1} M_{x_0, x_1} M_{x_1, x_2} \\
&= \sum_{x_0, x_2} M^2_{x_0, x_2}
\end{aligned}
$$

which is the sum of all entries of M^2.

(d)

For the graph G over n variables $X_0, \ldots, X_{n-1}$ with edges only between consecutive nodes, the joint distribution of $X_0, \ldots, X_{n-1}$ is

$$
p(x_0, \ldots, x_{n-1}) = \frac{1}{Z} \exp(\alpha x_0 x_1 + \ldots \alpha x_{n-2} x_{n-1}) = \frac{1}{Z} \prod_{k=0}^{n-2} \exp(\alpha x_k x_{k+1}) = \frac{1}{Z} \prod_{k=0}^{n-2} M_{x_k, x_{k+1}}.
$$

Then the normalizing constant Z is

$$
\begin{aligned}
Z &= \sum_{x_0, \ldots, x_{n-1}} \prod_{k=0}^{n-2} M_{x_k, x_{k+1}} \\
&= \sum_{x_0, x_{n-1}} \sum_{x_1, \ldots, x_{n-2}} \prod_{k=0}^{n-2} M_{x_k, x_{k+1}} \\
&= \sum_{x_0, x_{n-1}} M^{n-1}_{x_0, x_{n-1}},
\end{aligned}
$$

since

$$
\begin{aligned}
\sum_{x_1,\ldots,x_{n-2}} \prod_{k=0}^{n-2} M_{x_k,x_{k+1}} &= \sum_{x_2,\ldots,x_{n-2}} \prod_{k=2}^{n-2} M_{x_k,x_{k+1}} \underbrace{\sum_{x_1} M_{x_0,x_1} M_{x_1,x_2}}_{M^2_{x_0,x_2}} \\
&= \sum_{x_3,\ldots,x_{n-2}} \prod_{k=3}^{n-2} M_{x_k,x_{k+1}} \underbrace{\sum_{x_2} M^2_{x_0,x_2} M_{x_2,x_3}}_{M^3_{x_0,x_3}} \\
&\quad \vdots \\
&= M^{n-1}_{x_0,x_{n-1}}.
\end{aligned}
$$

Therefore, the normalizing constant Z is the sum of all entries of M^{n-1}.

The marginal distribution on X_0 is obtained as follows.

$$
\begin{aligned}
P(x_0) &= \sum_{x_1,\ldots,x_{n-1}} P(x_0,\ldots,x_{n-1}) \\
&= \sum_{x_1,\ldots,x_{n-1}} \prod_{k=0}^{n-2} M_{x_k,x_{k+1}}/Z \\
&= \sum_{x_{n-1}} \sum_{x_1,\ldots,x_{n-2}} \prod_{k=0}^{n-2} M_{x_k,x_{k+1}}/Z \\
&= \sum_{x_{n-1}} M^{n-1}_{x_0,x_{n-1}}/Z \\
&= (M^{n-1}_{x_0,0} + M^{n-1}_{x_0,1}) / \sum_{y,y'=0,1} M^{n-1}_{y,y'}.
\end{aligned}
$$

(e)

For any joint distribution $p(x_1,\ldots,x_n)$ on n variables,

$$
\begin{aligned}
p(x_1,\ldots,x_n) &= \frac{p(x_1,\ldots,x_n)}{p(x_1,\ldots,x_{n-1})} \frac{p(x_1,\ldots,x_{n-1})}{p(x_1,\ldots,x_{n-2})} \cdots \frac{p(x_1,x_2)}{p(x_1)} \\
&= p(x_n \mid x_1,\ldots,x_{n-1}) p(x_{n-1} \mid x_1,\ldots,x_{n-2}) \ldots p(x_2 \mid x_1) p(x_1).
\end{aligned}
$$

Suppose the joint distribution is given as $p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{k=1}^{n-1} M_{x_k, x_{k+1}}$ as in (d). Then

$$\begin{aligned}
p(x_n \mid x_1, \ldots, x_{n-1}) &= \frac{p(x_1, \ldots, x_{n-1}, x_n)}{p(x_1, \ldots, x_{n-1})} \\
&= \frac{\prod_{k=1}^{n-1} M_{x_k, x_{k+1}}/Z}{\sum_{x_n} \prod_{k=1}^{n-1} M_{x_k, x_{k+1}}/Z} \\
&= \frac{M_{x_{n-1}, x_n} \prod_{k=1}^{n-2} M_{x_k, x_{k+1}}}{\sum_{x_n} M_{x_{n-1}, x_n} \prod_{k=1}^{n-2} M_{x_k, x_{k+1}}} \\
&= \frac{M_{x_{n-1}, x_n}}{\sum_{x_n} M_{x_{n-1}, x_n}} \\
&= \frac{\exp(\alpha x_{n-1} x_n)}{1 + \exp(\alpha x_{n-1})}.
\end{aligned}$$

Since this is a function of x_{n-1} and x_n only,

$$p(x_n \mid x_1, \ldots, x_{n-1}) = p(x_n \mid x_{n-1}).$$

Similarly for $i = 2, \ldots, n-1$,

$$\begin{aligned}
p(x_i \mid x_1, \ldots, x_{i-1}) &= \frac{p(x_1, \ldots, x_{i-1}, x_i)}{p(x_1, \ldots, x_{i-1})} \\
&= \frac{\sum_{x_{i+1}, \ldots, x_n} \prod_{k=1}^{n-1} M_{x_k, x_{k+1}}/Z}{\sum_{x_i, x_{i+1}, \ldots, x_n} \prod_{k=1}^{n-1} M_{x_k, x_{k+1}}/Z} \\
&= \frac{\prod_{k=1}^{i-1} M_{x_k, x_{k+1}} \sum_{x_{i+1}, \ldots, x_n} \prod_{k=i}^{n-1} M_{x_k, x_{k+1}}}{\prod_{k=1}^{i-2} M_{x_k, x_{k+1}} \sum_{x_i, x_{i+1}, \ldots, x_n} \prod_{k=i-1}^{n-1} M_{x_k, x_{k+1}}} \\
&= \frac{M_{x_{i-1}, x_i} \sum_{x_{i+1}, \ldots, x_n} \prod_{k=i}^{n-1} M_{x_k, x_{k+1}}}{\sum_{x_i} M_{x_{i-1}, x_i} \sum_{x_{i+1}, \ldots, x_n} \prod_{k=i}^{n-1} M_{x_k, x_{k+1}}} \\
&= \frac{M_{x_{i-1}, x_i} \sum_{x_n} M^{n-i}_{x_i, x_n}}{\sum_{x_i} M_{x_{i-1}, x_i} \sum_{x_n} M^{n-i}_{x_i, x_n}}, \quad \left(\text{since } \textstyle\sum_{x_{i+1}, \ldots, x_{n-1}} \prod_{k=i}^{n-1} M_{x_k, x_{k+1}} = M^{n-i}_{x_i, x_n}\right)
\end{aligned}$$

which is a function of x_{i-1} and x_i only, so we have

$$p(x_i \mid x_1, \ldots, x_{i-1}) = p(x_i \mid x_{i-1}),$$

and

$$p(x_1, \ldots, x_n) = p(x_n \mid x_{n-1}) p(x_{n-1} \mid x_{n-2}) \ldots p(x_2 \mid x_1) p(x_1).$$

Thus, this joint distribution defines a Markov Chain. However, since each transition probability $p(x_i \mid x_{i-1}) := P(X_i = x_i \mid X_{i-1} = x_{i-1})$ depends on M^{n-i} which is different for each i, the transition probabilities are not the same at each step, so the Markov Chain is not homogeneous.

Note: We can also use the conditional independence properties defined by the graph to show that the joint distribution defines a Markov Chain. Since the joint distribution is defined as the product of clique functions and the normalizing constant, it has a Markov property with respect to the graph, that is, if C separates A from B, then X_A and X_B are conditionally independent given X_C. Since node $i-1$ separates node i from nodes $1, \ldots, i-2$, X_i is conditionally independent of $X_1, \ldots, X_{i-2}$ given X_{i-1}, so $p(x_i \mid x_1, \ldots, x_{i-1} = p(x_i \mid x_{i-1})$. However, you still need to calculate the transition probabilities to see if the Markov Chain is homogeneous.

(f)

Without clique functions for singletons, we have

$$\begin{aligned} P(x_0, x_1, x_2) &= \exp\left(\alpha x_0 x_1 + \alpha x_1 x_2\right)/Z_0 \\ &= \exp\left(-x^t \Gamma_0 x\right)/Z_0, \end{aligned}$$

where $x = (x_0, x_1, x_2)^t$ and $\Gamma_0 = \begin{pmatrix} 0 & -\alpha/2 & 0 \\ -\alpha/2 & 0 & -\alpha/2 \\ 0 & -\alpha/2 & 0 \end{pmatrix}$. Since Γ_0 is not a positive definite matrix, we cannot compute Z_0.

If we add a clique function for each singleton as given, then

$$\begin{aligned} P(x_0, x_1, x_2) &= \exp\left(\alpha x_0 x_1 + \alpha x_1 x_2 - \beta x_0^2 - \beta x_1^2 - \beta x_2^2\right)/Z \\ &= \exp\left(-x^t \Gamma x\right)/Z, \end{aligned}$$

where $\Gamma = \begin{pmatrix} \beta & -\alpha/2 & 0 \\ -\alpha/2 & \beta & -\alpha/2 \\ 0 & -\alpha/2 & \beta \end{pmatrix}$. Note that the eigenvalues of Γ are $\beta, \beta \pm \alpha/\sqrt{2}$. Thus, to make Γ positive definite, we need to assume that $\beta - \alpha/\sqrt{2} > 0$, that is, $\beta^2 > \alpha^2/2$. Then this joint distribution is a multivariate normal distribution with mean $\mu = (0,0,0)$ and the covariance matrix $\Sigma = \Gamma^{-1}/2$, so the normalizing constant is

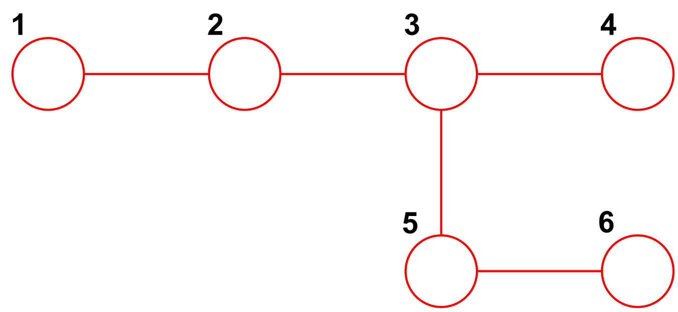
$$Z = |2\pi\Sigma|^{1/2} = |2\pi\Gamma^{-1}/2|^{1/2} = \pi^{3/2}|\Gamma|^{-1/2} = \frac{\pi\sqrt{\pi}}{\sqrt{\beta(\beta + \alpha/\sqrt{2})(\beta - \alpha/\sqrt{2})}} = \frac{\pi\sqrt{\pi}}{\sqrt{\beta(\beta^2 - \alpha^2/2)}}.$$

The joint distribution is then,

$$P(x_0, x_1, x_2) = \frac{\sqrt{\beta(\beta^2 - \alpha^2/2)}}{\pi\sqrt{\pi}} \exp\left(\alpha x_0 x_1 + \alpha x_1 x_2 - \beta x_0^2 - \beta x_1^2 - \beta x_2^2\right).$$

2 [40 points]

(a)



(b)

Since the maximal cliques are $\mathcal{C} = \{(1,2),(2,3),(3,4),(3,5),(5,6)\}$, the joint distribution is

$$\begin{aligned} p(x) &= \frac{1}{Z}\exp\left\{\sum_{c\in\mathcal{C}}\psi_c(x_c)\right\} \\ &= \frac{1}{Z}\exp\{\psi_{12}(x_1,x_2)+\psi_{23}(x_2,x_3)+\psi_{34}(x_3,x_4)+\psi_{35}(x_3,x_5)+\psi_{56}(x_5,x_6)\}. \end{aligned}$$

(c)

The normalizing constant is

$$Z = \sum_{x_1}\ldots\sum_{x_6}\exp\{\psi_{12}(x_1,x_2)+\psi_{23}(x_2,x_3)+\psi_{34}(x_3,x_4)+\psi_{35}(x_3,x_5)+\psi_{56}(x_5,x_6)\}.$$

$\exp\{\psi_{12}(x_1,x_2)+\ldots+\psi_{56}(x_5,x_6)\}$ can be represented as a set of numbers, one for each possible value for $x = (x_1,\ldots,x_6)$. Since there are 6 variables each with K states, there are K^6 values for x, so the summation over all sequences x takes $K^6 - 1 = O(K^6)$ operations.

(d)

Let $f(\bullet) = \exp\{\psi(\bullet)\}$. We can rearrange the order of the summations and the multiplications as follows.

$$
\begin{aligned}
Z &= \sum_{x_1}\cdots\sum_{x_6}\exp\{\psi_{12}(x_1,x_2)+\psi_{23}(x_2,x_3)+\psi_{34}(x_3,x_4)+\psi_{35}(x_3,x_5)+\psi_{45}(x_4,x_5)\} \\
&= \sum_{x_1}\cdots\sum_{x_6} f_{12}(x_1,x_2)f_{23}(x_2,x_3)f_{34}(x_3,x_4)f_{35}(x_3,x_5)f_{56}(x_5,x_6) \\
&= \underbrace{\sum_{x_1}\underbrace{\sum_{x_2}\underbrace{f_{12}(x_1,x_2)\underbrace{\sum_{x_3}\underbrace{\underbrace{f_{23}(x_2,x_3)\underbrace{\sum_{x_4} f_{34}(x_3,x_4)}_{O(K^2)}}_{O(K^2)}\underbrace{\sum_{x_5}\underbrace{f_{35}(x_3,x_5)\underbrace{\sum_{x_6} f_{56}(x_5,x_6)}_{O(K^2)}}_{O(K^2)}}_{O(K^2)}}_{O(K^2)}}_{O(K^2)}}_{O(K^2)}}_{O(K^2)}}_{O(K)}
\end{aligned}
$$

Note that you can start with $\sum_{x_6} f_{56}(x_5,x_6)$ or $\sum_{x_4} f_{34}(x_3,x_4)$ or $\sum_{x_1} f_{12}(x_1,x_2)$, i.e. start by removing a node that has only one neighbor.

Using the above reordered expression, the computational cost is as follows.

The summation over x_6 involves only the function $f_{56}(x_5,x_6)$, which is a table of $K \times K$ numbers. We have to sum this table over x_6 for each values of x_5 and so this summation has $K(K-1)$ cost.

The resulting vector of K numbers $\sum_{x_6} f_{56}(x_5,x_6)$, $x_5 = 1,\ldots,K$, is multiplied by the matrix of numbers $f_{35}(x_3,x_5)$ and so this has K^2 cost.

Since there are 5 summations and 4 multiplications of these kinds and the last summation over x_1 takes only $K-1$ operations, the total cost of evaluating Z is about $9K^2$.

Note: For a chain of N nodes, evaluating the normalizing constant has $O(NK^2)$ cost which is linear in the length of the chain. *See p.396.* Although the graph given here is not a chain, the calculation is done in a similar way. As long as the graph maintains a tree shape as it grows to N nodes, the evaluation of the normalizing constant takes $O(NK^2)$ operations.