4.1)

CSE 250A - Homework 4

a) $\mathcal{L} = \log P(data) =$

$\log\left(\prod_{n=1}^{N} P(X_n)\right) =$

$\sum_{d=1}^{2\cdot D} \sum_{n=1}^{C_d} \log(P(X=d)) =$

$$\boxed{\sum_{d=1}^{2\cdot D} C_d \cdot \log(P_d)}$$

b) $\hookrightarrow$ Use from part a $\quad \sum_{d=1}^{2\cdot D} P_d - 1 = 0$

$\mathcal{L} = C_d \cdot \log(P_d) - \lambda \cdot \left(\sum_{d=1}^{2\cdot D} P_d - 1\right)$

$\frac{\partial \mathcal{L}}{\partial P_d} = \frac{\partial}{\partial P_d}\left(C_d \cdot \log(P_d) - \lambda \cdot \left(\sum_{d=1}^{2\cdot D} P_d - 1\right)\right) =$

$\frac{C_d}{P_d} - \lambda \cdot 1 + \lambda \cdot 0 =$

$\frac{C_d}{P_d} - \lambda$

$\frac{C_d}{P_d} - \lambda = 0$

$\frac{C_d}{P_d} = \lambda$

$C_d = \lambda \cdot P_d$

$P_d = \frac{C_d}{\lambda}$

Recall that $\sum_{d=1}^{2\cdot D} P_d = 1$

$\sum_{d=1}^{2\cdot D} P_d = \sum_{d=1}^{2\cdot D} \frac{C_d}{\lambda} = 1 \longrightarrow \lambda = \sum_{d=1}^{2\cdot D} C_d$

$P_d = \frac{C_d}{\lambda}$

$$\boxed{P_d = \frac{C_d}{\sum_{d=1}^{2\cdot D} C_d}}$$

$\begin{cases} -1 & \text{when } d \text{ is odd} \\ 1 & \text{when } d \text{ is even} \end{cases}$

split the summation

odd $2 \cdot d - 1$

even $2 \cdot d$

c) $$\sum_{d=1}^{2 \cdot D} (-1)^d \cdot P_d = 0$$

$$\sum_{d=1}^{2 \cdot D} (-1)^d \cdot P_d =$$

$$\sum_{d=1}^{D} \underbrace{(-1)^{2 \cdot d - 1}}_{\text{always } -1} \cdot P_{2 \cdot d - 1} + \sum_{d=1}^{D} \underbrace{(-1)^{2 \cdot d}}_{\text{always } 1} \cdot P_{2d} =$$

$$(-1) \cdot \sum_{d=1}^{D} P_{2 \cdot d - 1} + (1) \cdot \sum_{d=1}^{D} P_{2 \cdot d} =$$

$$-\sum_{d=1}^{D} P_{2 \cdot d - 1} + \sum_{d=1}^{D} P_{2 \cdot d} = 0$$

$$\sum_{d=1}^{D} P_{2 \cdot d} = \sum_{d=1}^{D} P_{2 \cdot d - 1}$$

Thus, $P(X \in \{2, 4, \ldots, 2 \cdot D\}) = P(X \in \{1, 3, \ldots, 2 \cdot D - 1\})$ when $\sum_{d=1}^{2 \cdot D} (-1)^d \cdot P_d = 0$

d) One of our constraints is $\sum_{d=1}^{2 \cdot D} P_d = 1$. From part c, we also know that $\sum_{d=1}^{D} P_{2 \cdot d} = \sum_{d=1}^{D} P_{2 \cdot d - 1}$. Thus, this implies that $\sum_{d=1}^{D} P_{2 \cdot d} = \frac{1}{2}$ and $\sum_{d=1}^{D} P_{2 \cdot d - 1} = \frac{1}{2}$. From part a, we have $\sum_{d=1}^{2 \cdot D} C_d \cdot \log(P_d)$. Let's split this into evens and odds too: $\sum_{d=1}^{D} C_{2 \cdot d - 1} \cdot \log(P_{2 \cdot d - 1}) + \sum_{d=1}^{D} C_{2 \cdot d} \cdot \log(P_{2 \cdot d})$. Overall, we want to solve for

$$\max_{P_d} \sum_{d=1}^{2 \cdot D} C_d \cdot \log(P_d) \quad \text{s.t.} \quad \sum_{d=1}^{2 \cdot D} P_d = 1 \quad \text{and} \quad \sum_{d=1}^{2 \cdot D} (-1)^d \cdot P_d = 0$$

Rewrite this on the next page.

Using $\sum_{d=1}^{2\cdot D} (-1)^d \cdot p_d = 0$, convert

$\sum_{d=1}^{2\cdot D} p_d = 1$ into $\sum_{d=1}^{D} p_{2\cdot d-1} = \frac{1}{2}$ and $\sum_{d=1}^{D} p_{2\cdot d} = \frac{1}{2}$

Thus, we can split this maximization problem into 2 different problems.

$$\max_{p_{2\cdot d-1}} \sum_{d=1}^{D} c_{2\cdot d-1} \cdot \operatorname{Log}(p_{2\cdot d-1}) \quad \text{s.t.} \sum_{d=1}^{D} p_{2\cdot d-1} = \frac{1}{2}$$

$$\max_{p_{2\cdot d}} \sum_{d=1}^{D} c_{2\cdot d} \cdot \operatorname{Log}(p_{2\cdot d}) \quad \text{s.t.} \sum_{d=1}^{D} p_{2\cdot d} = \frac{1}{2}$$

Recall from part b that to maximize $\mathcal{L}$, we got $p_d = \frac{c_d}{\sum_{d=1}^{2D} c_d}$

Plug these 2 problems into the maximize equation above

$$\boxed{p_{2\cdot d-1} = \frac{c_{2\cdot d-1}}{\sum_{d=1}^{D} c_{2\cdot d-1}} \qquad p_{2\cdot d} = \frac{c_{2\cdot d}}{\sum_{d=1}^{D} c_{2\cdot d}}}$$

4.2)

b) Recall from part b of question 4.1 that $p_d = \frac{c_d}{\sum_{d=1}^{2D} c_d}$ maximizes the log-likelihood.

$$P_1(X_1 = x_1) = \frac{P_1(X_1 = x_1)}{P_1(X_1)} = \frac{COUNT_1(X_1 = x_1)}{COUNT_1(X_1)} = \frac{COUNT(X_1 = x_1)}{T}$$

$$COUNT_1(X_1) = \sum_{t=1}^{T} 1 = T$$

$$P_{i+1}(X_{i+1} = x_{i+1} \mid X_i = x_i) = \frac{P_i(X_{i+1} = x_{i+1} \mid X_i = x_i)}{P_i(X_i = x_i)} =$$

$$\frac{COUNT_i(X_i = x_i, X_{i+1} = x_{i+1})}{COUNT_i(X_i = x_i)}$$ For all i from 1 to $n-1$

$$\boxed{P_1(X_1 = x_1) = \frac{COUNT(X_1 = x_1)}{T}}$$

$$\boxed{P_{i+1}(X_{i+1} = x_{i+1} \mid X_i = x_i) = \frac{COUNT_i(X_i = x_i, X_{i+1} = x_{i+1})}{COUNT_i(X_i = x_i)}}$$ For all i from 1 to $n-1$

a) $$P_n(X_n = x_n) = \frac{P_n(X_n = x_n)}{P_n(X_n)} = \frac{COUNT_n(X_n = x_n)}{COUNT_n(X_n)} = \frac{COUNT(X_n = x_n)}{T}$$

$$P_i(X_i = x_i \mid X_{i+1} = x_{i+1}) = \frac{P_i(X_i = x_i \mid X_{i+1} = x_{i+1})}{P_{i+1}(X_{i+1} = x_{i+1})} =$$

$$\frac{COUNT_i(X_i = x_i, X_{i+1} = x_{i+1})}{COUNT_{i+1}(X_{i+1} = x_{i+1})}$$ For all i from 1 to $n-1$

$$\boxed{P_n(X_n = x_n) = \frac{COUNT(X_n = x_n)}{T}}$$

$$\boxed{P_i(X_i = x_i \mid X_{i+1} = x_{i+1}) = \frac{COUNT_i(X_i = x_i, X_{i+1} = x_{i+1})}{COUNT_{i+1}(X_{i+1} = x_{i+1})}}$$ For all i from 1 to $n-1$

c) $P(G_1=g_1)=\left(\prod_{i=1}^{n-1} P_i(X_i=x_i \mid X_{i+1}=x_{i+1})\right)\cdot P_n(X_n=x_n)$

$P(G_2=g_2)=\left(\prod_{i=1}^{n-1} P_{i+1}(X_{i+1}=x_{i+1} \mid X_i=x_i)\right)\cdot P_1(X_1=x_1)$

Expand each

$$P(G_1=g_1)=\left(\prod_{i=1}^{n-1}\frac{COUNT_i(X_i=x_i, X_{i+1}=x_{i+1})}{COUNT_{i+1}(X_{i+1}=x_{i+1})}\right)\cdot\frac{COUNT_n(X_n=x_n)}{T}=$$

$$\left(\prod_{i=1}^{n-2}\frac{COUNT_i(X_i=x_i, X_{i+1}=x_{i+1})}{COUNT_{i+1}(X_{i+1}=x_{i+1})}\right)\cdot\frac{COUNT_{n-1}(X_{n-1}=x_{n-1}, X_n=x_n)}{\cancel{COUNT_n(X_n=x_n)}}\cdot\frac{\cancel{COUNT_n(X_n=x_n)}}{T}$$

$$\left(\frac{\prod_{i=1}^{n-1}COUNT_i(X_i=x_i, X_{i+1}=x_{i+1})}{\prod_{i=1}^{n-2}COUNT_{i+1}(X_{i+1}=x_{i+1})}\right)\cdot\frac{1}{T}=$$

$$\left(\frac{\prod_{i=1}^{n-1}COUNT_i(X_i=x_i, X_{i+1}=x_{i+1})}{\prod_{i=1}^{n-2}COUNT_{i+1}(X_{i+1}=x_{i+1})}\right)\cdot\frac{COUNT_1(X_1=x_1)}{COUNT_1(X_1=x_1)}\cdot\frac{1}{T}=$$

$$\left(\frac{\prod_{i=1}^{n-1}COUNT_i(X_i=x_i, X_{i+1}=x_{i+1})}{\prod_{i=0}^{n-2}COUNT_{i+1}(X_{i+1}=x_{i+1})}\right)\frac{COUNT_1(X_1=x_1)}{T}\cdot\frac{1}{T}=$$

$$\left(\frac{\prod_{i=1}^{n-1}COUNT_i(X_i=x_i, X_{i+1}=x_{i+1})}{\prod_{i=1}^{n-1}COUNT_i(X_i=x_i)}\right)\cdot\frac{COUNT_1(X_1=x_1)}{T}=$$

$$\left(\prod_{i=1}^{n-1}\frac{COUNT_i(X_i=x_i, X_{i+1}=x_{i+1})}{COUNT_i(X_i=x_i)}\right)\cdot\frac{COUNT_1(X_1=x_1)}{T}=$$

$$\left(\prod_{i=1}^{n-1} P_{i+1}(X_{i+1}=x_{i+1} \mid X_i=x_i)\right)\cdot P(X_1=x_1)=$$

$$P(G_2=g_2)$$

Thus, $P(G_1=g_1)=P(G_2=g_2)$

d) I see something other than looking at nodes X_3 and X_{n-2}.
Note that there are two nodes X_1 and X_n with only one parent, unlike X_n in G_1 and X_1 in G_2.
If we were to compute the joint distribution for G_3, we would need both

$P(X_1=x_1)=\frac{COUNT_1(X_1=x_1)}{T}$ and $P(X_n=x_n)=\frac{COUNT(X_n=x_n)}{T}$

In the joint distribution, there would then be at least 2 Ts in the denominator, whereas the joint distributions for G_1 and G_2 only have one T in the denominator.

Therefore, $P(G_3=g_3) \neq P(G_1=g_1)$ and
$P(G_3=g_3) \neq P(G_2=g_2)$