1. In a second-order Markov chain, each state depends on the two previous states, i.e.,

$$P\left[X^{(t+1)} = S_k | X^{(t)} = S_j, X^{(t-1)} = S_i, \cdots\right] = P\left[X^{(t+1)} = S_k | X^{(t)} = S_j, X^{(t-1)} = S_i\right].$$

Show that the second-order Markov chain can always be converted to a first-order Markov chain. (Hint: by redesign the states)

A1:If State={A,C,T,G,...} , we can redesign $Z^t = \{X^t, X^{t-1}\}$

To prove the $\mathrm{P}(Z^t|Z^{t-1})$ *is exist.*

I show the $\mathrm{P}(Z^t|Z^{t-1})$ *by simple permutation at here.*(just {A,C,T,G})

$t \backslash t-1$	$AA\ AC\ AT\ AG$	CA	$\ldots$	$\ldots\ GG$
AA CA TA GA	**A**	**0**	**0**	**0**
AC $\vdots$ $\vdots$	**0**	**C**	**0**	**0**
$\vdots$ $\vdots$	**0**	**0**	**T**	**0**
$\vdots$ GG	**0**	**0**	**0**	**G**

Eg: $\mathrm{P}(Z^t = CA|Z^{t-1} = AT) = \mathrm{P}(\mathrm{C}|\mathrm{A})$

2. Prove that a Bayesian network must be a Directed Acyclic Graph (DAG).

A2:At Bayesian Network , we can let $\mathrm{P}(X_1, X_2, \ldots, X_N)$ be factorized into $\prod_{i=1}^{N} P(X_i|Parient(X_i))$. And draw a factorization as a graph.

(if the node X_i don't have the $Parients$. We can use the notation $P(X_i)$)

I assume the graph of Bayesian Network is a Directed **cyclic** Graph , and then I got a contradiction because the factorization can't be finished...(every node has a parent but I just have N node.)

So the graph of Bayesian Network is a Directed A**cyclic** Graph.

3. Given random variables A, B, C, and D, answer true or false and justify your answer:

(a) $\{A\} \perp\!\!\!\perp \{B\}|\{C\}$ implies $\{A\} \perp\!\!\!\perp \{B\}$;

(b) $\{A\} \perp\!\!\!\perp \{B\}$ implies $\{A\} \perp\!\!\!\perp \{B\}|\{C\}$;

(c) $\{A\} \perp\!\!\!\perp \{B, C\}|\{D\}$ implies $\{A\} \perp\!\!\!\perp \{B\}|\{D\}$.

A3(a): false , because we just need to consider the situations

1.tail to tail -> P(A,B)=$\int P(A,B,C)d_c = \int P(A|C)P(B|C)P(C)d_c \neq P(A)P(B)$

2.head to tail -> P(A,B)=$\int P(A,B,C)d_c = \int P(B|C)P(C|A)P(A)d_c \neq P(A)P(B)$

A3(b): false , because we just need to consider the situation

1. head to head -> P(A,B|C)=$\frac{\mathrm{P(A,B,C)}}{P(C)} = \frac{\frac{P(A,B,C)}{P(A,B)} * P(A,B)}{P(C)} = \frac{P(C|A,B) * P(A) * P(B)}{P(C)} \neq P(A|C)P(B|C)$

A3(c): yes , because we just need to consider the situations

1. tail to tail -> P(A,B|D)= $\frac{\mathrm{P(A,B,D)}}{P(D)} = \frac{\mathrm{P(A,B|D)P(D)}}{P(D)} = \frac{\mathrm{P(A|D)P(B|D)P(D)}}{P(D)} = \mathrm{P(A|D)P(B|D)}$

2. head to tail ->P(A,B|D)= $\frac{\mathrm{P(A,B,D)}}{P(D)} = \frac{\mathrm{P(A|B,D)P(B,D)}}{P(D)} = \frac{\mathrm{P(A|D)P(B,D)}}{P(D)} = \frac{\mathrm{P(D|A)P(A)P(B|D)}}{P(D)} = \mathrm{P(B|D)P(A|D)}$

4. Given a Hidden Markov model with time homogeneous Gaussian emission probability $P[\boldsymbol{x}^{(t)}|z_i^{(t)}, \theta_i] = \frac{1}{(2\pi)^{d/2} det(\boldsymbol{\Sigma}_i)^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}^{(t)}-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}^{(t)}-\boldsymbol{\mu}_i)} dx$, where $\theta_i = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. Consider the problem finding $\Theta = (\boldsymbol{\pi}^{(1)}, \boldsymbol{A}, \{\theta_k\}_{k=1}^K)$ using the EM algorithm. Show that maximizing $\mathcal{Q}(\Theta; \Theta^{old})$ in the M-step gives $\boldsymbol{\mu}_i = \frac{\sum_{t=1}^T \gamma_i^{(t)} \boldsymbol{x}^{(t)}}{\sum_{t=1}^T \gamma_i^{(t)}}$ and $\boldsymbol{\Sigma}_i = \frac{\sum_{t=1}^T \gamma_i^{(t)} (\boldsymbol{x}^{(t)}-\boldsymbol{\mu}_i)(\boldsymbol{x}^{(t)}-\boldsymbol{\mu}_i)^\top}{\sum_{t=1}^T \gamma_i^{(t)}}$.

A4: let $\gamma_i = \begin{cases} 1 & x^t \ \ is \ \ state \ \ i \\ 0 & \qquad Others \end{cases}$, and $N_i = \sum_{t=1}^T {\gamma_i}^t$

To Max $\log \mathrm{P(X|}Z_i, \Theta_i)$=log$\left(\prod_{t=1}^T \left((\frac{1}{(2\pi)^{\frac{d}{2}} * \det(\Sigma_i)^{\frac{1}{2}}}) e^{-\frac{1}{2}(x^t - u_i)^T {\Sigma_i}^{-1}(x^t - u_i)} \right)^{{\gamma_i}^t} \right)$

$$= -\frac{N_i d}{2}\log(2\pi) - \frac{N_i}{2}\log\left(det(\Sigma_i)\right) - \frac{1}{2}\sum_{t=1}^T {\gamma_i}^t \, (x^t - u_i)^T {\Sigma_i}^{-1}(x^t - u_i)$$

$$\frac{d \log \mathrm{P(X|}Z_k, \Theta_k)}{du} = -\sum_{t=1}^T {\gamma_i}^t \, (x^t -)^T {\Sigma_i}^{-1} = 0 \quad \rightarrow \quad \boldsymbol{u_i} = \frac{\boldsymbol{\sum_{t=1}^T {\gamma_i}^t x^t}}{\boldsymbol{\sum_{t=1}^T {\gamma_i}^t}}$$

$$\frac{d \log \mathrm{P(X|}Z_i, \Theta_i)}{d{\Sigma_i}^{-1}} = \frac{N_i}{2}\Sigma_i - \frac{1}{2}\sum_{t=1}^T {\gamma_i}^t \, (x^t - u_i)(x^t - u_i)^T = 0$$

$$\rightarrow \ \boldsymbol{\Sigma_i} = \frac{\boldsymbol{\sum_{t=1}^T {\gamma_i}^t \, (x^t - u_i)(x^t - u_i)^T}}{\boldsymbol{N_i}} = \frac{\boldsymbol{\sum_{t=1}^T {\gamma_i}^t \, (x^t - u_i)(x^t - u_i)^T}}{\boldsymbol{\sum_{t=1}^T {\gamma_i}^t}}$$