

3.5)

Given to compute $P(Y|X=0)$, $P(Y|X=1)$,
 $P(Z_1=1|Y)$, $P(Z_2=1|Y)$

$$P(Y|X=0) = P(Y_1, Y_2, Y_3 | X=0)$$

$$= P(Y_1 | X=0) P(Y_2, Y_3 | X=0, Y_1) \quad [\text{Product rule with evidence}]$$

conditionally

Now from rule ② of d-separation Y_2, Y_3 are independent

of Y_1 given X .

Hence

$$= P(Y_1 | X=0) P(Y_2, Y_3 | X=0)$$

$$= P(Y_1 | X=0) P(Y_2 | X=0) P(Y_3 | X=0, Y_2) \quad [\text{product rule with evidence}].$$

Again from rule ② of d-separation,

Y_2, Y_3 are conditionally independent given X .

$$= P(Y_1 | X=0) P(Y_2 | X=0) P(Y_3 | X=0) \quad - ①$$

Similarly

$$P(Y|X=1) = P(Y_1 | X=1) P(Y_2 | X=1) P(Y_3 | X=1) \quad - ②$$

$$P(Z_1=1|Y) = P(Z_1=1|Y_1, Y_2, Y_3) \Rightarrow \text{CPT table} \quad - ③$$

$$P(Z_2=1|Y) = P(Z_2=1|Y_1, Y_2, Y_3) \Rightarrow \text{CPT table.} \quad - ④$$

Computing ^{value} first two rows of the table.

First row

y_1	y_2	y_3	y
0	0	0	1

$$P(y|x=0) = P(y_1=0|x=0) P(y_2=0|x=0) P(y_3=0|x=0) \quad [\text{from eq ①}]$$

$$= P(y_1=0|x=0) P(y_2=0|x=0) P(y_3=0|x=0)$$

$$= (1-0.25)(1-0.5)(1-0.25)$$

$$= 0.25 \times 0.5 \times 0.25$$

$$= 0.09375$$

$$P(y|x=1) = P(y_1=0|x=1) P(y_2=0|x=1) P(y_3=0|x=1) \quad [\text{from eq ②}]$$

$$= (1-0.5)(1-0.25)(1-0.25)$$

$$= 0.5 \times 0.25 \times 0.25$$

$$= 0.09375$$

$$P(z_1=1|y_1=0, y_2=0, y_3=0)$$

$$= 0.9 \quad (\text{from CPT table})$$

$$P(z_2=1|y_1=0, y_2=0, y_3=0) = 0.1 \quad (\text{from CPT table})$$

Second row

y_1	y_2	y_3	y
1	0	0	2

$$P(y|x=0) = P(y_1=1|x=0) P(y_2=0|x=0) P(y_3=0|x=0)$$

$$= 0.75(1-0.5)(1-0.25)$$

$$= 0.75 \times 0.5 \times 0.25$$

$$= 0.28125$$

$$\begin{aligned}
 P(Y|X=1) &= P(Y_1=1|X=1)P(Y_2=0|X=1)P(Y_3=0|X=1) \\
 &= 0.5 \times (1-0.25) \times (1-0.25) \\
 &= 0.5 \times 0.75 \times 0.75 \\
 &= 0.09375
 \end{aligned}$$

$$P(Z_1=1|Y_1=1, Y_2=0, Y_3=0) = 0.8 \text{ (CPT table).}$$

$$P(Z_2=1|Y_1=1, Y_2=0, Y_3=0) = 0.2 \text{ (CPT table).}$$

Table

Y_1	Y_2	Y_3	Y	$P(Y X=0)$	$P(Y X=1)$	$P(Z_1=1 Y)$	$P(Z_2=1 Y)$
0	0	0	1	0.09375	0.09375	0.9	0.1
1	0	0	2	0.28125	0.09375	0.8	0.2
0	1	0	3	0.09375	0.03125	0.7	0.3
0	0	1	4	0.03125	0.03125	0.6	0.4
1	1	0	5	0.28125	0.03125	0.5	0.5
1	0	1	6	0.09375	0.28125	0.4	0.6
0	1	1	7	0.03125	0.09375	0.3	0.7
1	1	1	8	0.09375	0.09375	0.2	0.8

3.6)

a) To show

$$\sum_z P(Z=z | B_1, B_2, \dots, B_n) = 1 \text{ for sum of } z \text{ over } [-\alpha, \alpha]$$

$$= \sum_{z=-\infty}^{\infty} P(Z=z | B_1, B_2, \dots, B_n)$$

$$= \sum_{z=-\infty}^{\infty} \left(\frac{1-\alpha}{1+\alpha} \right) \alpha^{|z-f(B)|}$$

if Z varies from $-\infty$ to ∞ then $Z-f(B)$ will also

vary from $-\infty$ to ∞ , hence we can replace Z with

y where $y = Z-f(B)$

$$= \sum_{y=-\infty}^{\infty} \left(\frac{1-\alpha}{1+\alpha} \right) \alpha^{|y|}$$

$$= \sum_{y=-\infty}^{-1} \left(\frac{1-\alpha}{1+\alpha} \right) \alpha^{|y|} + \sum_{y=1}^{\infty} \left(\frac{1-\alpha}{1+\alpha} \right) \alpha^{|y|} + \underbrace{\left(\frac{1-\alpha}{1+\alpha} \right) \alpha^{|y|}}_{L(y=0)}$$

$$= \frac{1-\alpha}{1+\alpha} + \underbrace{\sum_{y=-\infty}^{-1} \left(\frac{1-\alpha}{1+\alpha} \right) \alpha^{|y|}}_{(1)} + \underbrace{\sum_{y=1}^{\infty} \left(\frac{1-\alpha}{1+\alpha} \right) \alpha^{|y|}}_{(2)}$$

Now

(1) & (2) are the same because because in both the cases $|y|$ will vary from 1 to ∞ .

Hence we get

$$= \frac{1-\alpha}{1+\alpha} + 2 \sum_{y=1}^{\infty} \left(\frac{1-\alpha}{1+\alpha} \right) \alpha^y$$

$$= \frac{1-\alpha}{1+\alpha} + 2 \left(\frac{1-\alpha}{1+\alpha} \right) \left[\sum_{y=1}^{\infty} \alpha^y \right] \Rightarrow \text{Infinite GP}$$

$$= \frac{1-\alpha}{1+\alpha} + 2 \left(\frac{1-\alpha}{1+\alpha} \right) \left(\frac{\alpha}{1-\alpha} \right)$$

• ~~$\frac{2\alpha}{1-\alpha}$~~

$$= \frac{1-\alpha}{1+\alpha} + \frac{2\alpha}{1+\alpha}$$

$$= \frac{1+\alpha}{1+\alpha} = 1.$$

Hence $P(Z=z | B_1, B_2, \dots, B_n) = 1$ when $z \in [-\infty, \infty]$.

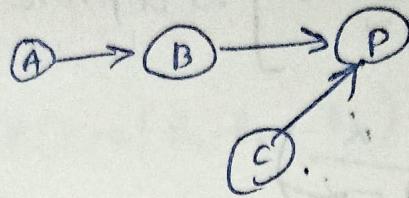
c) I have plotted the values of probability in logspace for 1 million experiments. for $P(B_i=1 | z=128)$ for $i \in \{2^{15}, 8, 10^4\}$.

I have also printed the last 15 probability values for each $i \in \{2^{15}, 8, 10^4\}$, we can clearly see that the most significant bits are the same in all the values, and also the plots appear to be converged. Hence we can confidently say that estimates have converged to good degree of precision.

3.7)

a) To compute

$$P(B|A, C, D)$$



$$P(B|A, C, D) \stackrel{\substack{E \\ Y}}{\underset{\substack{x \\ D}}{\approx}} \left[P(x|y, E) = \frac{P(y|x, E) P(x|E)}{P(y|E)} \right]$$

[Bayes rule with evidence]

$$\frac{P(A|B, C, D) P(B|C, D)}{P(A)}$$

$$= \frac{P(D|B, A, C) P(B|A, C)}{P(D|A, C)}$$

$$= \frac{P(D|B, C) P(B|A)}{\sum_b P(D, B=b|A, C)}$$

$P(D|B, C) = P(D|B|C)$
using d-separation rule ①
 $P(B|A, C) = P(B|A)$
using d-separation rule ③

$$= \frac{P(D|B, C) P(B|A)}{\sum_b P(B=b|A, C) P(D|B=b, A, C)}$$

Product rule
with evidence

$$= \frac{P(D|B,C) P(B|A)}{\sum_b P(B=b|A) P(D|B=b,C)} \left\{ \begin{array}{l} P(B=b|A,c) = P(B=b|A) \\ \text{using d-separation rule D} \\ P(D|B=b,C) = P(D|B=b,A,C) \\ \text{using d-separation rule ①} \end{array} \right.$$

$$= \frac{P(D|B,C) P(B|A)}{\sum_b P(B=b|A) P(D|B=b,C)} - \text{eq } ①$$

we can compute $P(B|A)$ $P(D|B,C)$ from CPT tables.

b) $P(B | \{A, C, D\}, \{E, F\})$

B is conditionally independent of $\{E, F\}$ given $\{A, C, D\}$.

All paths from node E to B is blocked by node ' C '
d-separation rule ②.

All paths from node F to B is blocked by node ' A' , d-separation rule ②.

Hence $P(B|A, C, D, E, F) = P(B|A, C, D)$

$$= \frac{P(D|B,C)P(B|A)}{\sum_b P(B=b|A) P(D|B=b,C)}$$

c) To compute:

$$P(B, E, F | A_1, C_1, D)$$

$$= P(B | A_1, C_1, D) P(E, F | A_1, C_1, D, B) \quad (\text{Product rule with evidence})$$

$$= P(B | A_1, C_1, D) P(E | A_1, B_1, C_1, D) P(F | A_1, C_1, D, B, E) \quad ["]$$

$$= P(B | A_1, C_1, D) P(E | A_1, B_1, C_1, D) P(F | A_1, B_1, C_1, D, E) \quad ["]$$

$P(B | A_1, C_1, D)$ has already been computed in (i) [eqn ①]

$$P(E | A_1, B_1, C_1, D) = P(E | C) \Rightarrow \text{CPT}$$

using d-separation rule ②

A, B, D are blocked by C .

$$P(F | A_1, B_1, C_1, D, E) = P(F | A) \Rightarrow \text{CPT}$$

using d-separation rule ②

B, C, D, E are blocked by A .

$$= P(B | A_1, C_1, D) P(E | C) P(F | A).$$

3.8)

a) Given 'T' samples $\{q_{t+1}, x_t, y_t, z_t\}_{t=1}^T$

We need to estimate $P(Q=q | E=e)$

from the formula of likelihood weighting we have

$$P(\theta_1 = q_1 | E=e) = \frac{\sum_{t=1}^T I(q_1, q_{1t}) P(E=e | q_{1t}, q_{2t}, y_{t1}, z_t)}{\sum_{t=1}^T P(E=e | q_{1t}, q_{2t}, y_{t1}, z_t)}$$

$$\text{Now } P(E | x_1, \theta_1, y_{12}) = P(E | Y_{12})$$

All paths from x_1, θ_1 are blocked by Y_{12} according
using d-separation rule ①.

$$= \frac{\sum_{t=1}^T I(q_1, q_{1t}) \cdot P(E=e | y_{t1}, z_t)}{\sum_{t=1}^T P(E=e | y_{t1}, z_t)}.$$

we can estimate $P(E=e | y_{t1}, z_t)$ from CPT table.

b) Given 'T' samples of $q_{1t}, q_{2t}, x_t, y_{t1}, z_t \}_{t=1}^T$

To estimate $P(\theta_1 = q'_1 | E_1 = e_1, E_2 = e_2)$, similar to bit a) we have

$$P(\theta_1 = q'_1 | E_1 = e_1, E_2 = e_2)$$

$$P(\theta_1 = q'_1, \theta_2 = q'_2 | E_1 = e_1, E_2 = e_2)$$

$$= \frac{\sum_{t=1}^T I(q'_1, q_{1t}) I(q'_2, q_{2t}) P(E_1 = e_1, E_2 = e_2 | q_{1t}, q_{2t}, x_t, y_{t1}, z_t)}{\sum_{t=1}^T P(E_1 = e_1, E_2 = e_2 | q_{1t}, q_{2t}, x_t, y_{t1}, z_t)}$$

Now

$$P(E_1 = e_1, E_2 = e_2 | q_{1t}, q_{2t}, \eta_t, y_t, z_t)$$

$$= P(E_1 = e_1 | q_{1t}, q_{2t}, \eta_t, y_t, z_t) P(E_2 = e_2 | q_{1t}, q_{2t}, \eta_t, y_t, z_t, E_1 = e_1)$$

E_1 is conditionally independent of

q_{2t}, z_t, y given q_1, x using d-separation rule.

Also E_2 is conditionally independent of q_2, x, y, q_1 given $E_1; z$

Hence we get

$$P(q_1 = q_1, q_2 = q_2 | E_1 = e_1, E_2 = e_2)$$

$$= \sum_{t=1}^T I(q_1, q_{1t}) I(q_2, q_{2t}) \cdot P(E_1 = e_1 | q_{1t}, x = x_t) \\ P(E_2 = e_2 | E_1 = e_1, z = z_t)$$

$$P(E_1 = e_1 | q_{1t}, x = x_t) P(E_2 = e_2 | E_1 = e_1, z = z_t)$$

[values of $P(E_1 = e_1 | q_{1t}, x = x_t)$
and $P(E_2 = e_2 | E_1 = e_1, z = z_t)$ can be
computed from the CPT table].

3.1)

a) To compute

$P(X_{t+1} = j | X_1 = i) = [A^t]_{ij}$ where A^t is
the power t^{th} of matrix A.

We will prove this using induction,

\Rightarrow for $t=1$

$$\text{A } P(X_2 = j | X_1 = i) = [A]_{ij} \text{ which is true}$$

as given $A_{ij} = P(X_{t+1} = j | X_t = i)$

$$\Rightarrow P(X_2 = j | X_1 = i) = A_{ij}$$

\Rightarrow Assume true for $t=K$ to be true.

i.e.,

$$\text{A } P(X_{t+1} = j | X_1 = i) = [A^t]_{ij}$$

\Rightarrow Now to prove $t=K+1$ to be true that is $t+1$ to be true

true

$$P(X_{t+1+1} = j | X_1 = i) \Rightarrow [A^{t+1}]_{ij}$$

LHS

$$P(X_{t+2} = j | X_1 = i)$$

$$= \sum_{\gamma=1}^m P(X_{t+2} = j, X_{t+1} = \gamma | X_1 = i) \quad \begin{array}{l} \text{Marginalization} \\ X_t \in \{1, 2, \dots, m\} \\ \text{given} \end{array}$$

$$= \sum_{\gamma=1}^m P(X_{t+1} = \gamma | X_1 = i) P(X_{t+2} = j | X_1 = i, X_{t+1} = \gamma) \quad [\text{product rule}].$$

$$= \sum_{\tau=1}^m P(X_{t+1}=\tau | X_1=i) P(X_{t+2}=j | X_{t+1}=\tau)$$

X_{t+2} is conditionally independent of X_1 given X_{t+1} [d-separation rule ①].

$$= \sum_{\tau=1}^m [A^t]_{in} \cdot P(X_{t+2}=j | X_{t+1}=\tau)$$

$\underbrace{\quad}_{\text{from induction step}}$

$$= \sum_{\tau=1}^m [A^t]_{in} [A]_{n\sigma j} \rightarrow \text{eq ②}$$

$\underbrace{\quad}_{\text{given.}}$

basically eq ②, we are multiplying i^{th} row of A^t with j^{th} column of A , by definition of matrix multiplication we know that

$$\text{def. } [AB]_{ij} = \sum_{r=1}^K [A]_{ir} [B]_{rj}$$

where $A \Rightarrow n \times K$

$B \Rightarrow K \times y$

Hence,

$$\text{eq ② implies } \sum_{\tau=1}^m [A^t]_{in} [A]_{n\sigma j} = [A^{t+1}]_{ij}$$

Hence our claim holds true for $t+1(K+1)$ as well.

Hence by induction hypothesis we have

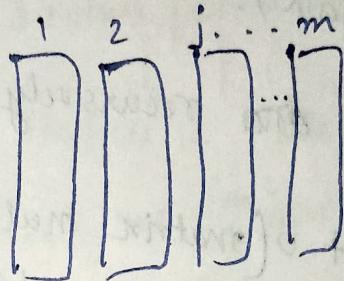
$P(X_{t+1} = j | X_1 = i) = [A^t]_{ij}$ where A^t is t^{th} power of matrix A .

b) We need to find the computational complexity of computing

$$P(X_{t+1} = j | X_1 = i) = [A^t]_{ij}$$

Hence to find ij^{th} term of A^t . Consider 'A'

$A \text{ matrix}$
 $\text{in } i$



to find A^2 matrix i^{th} row we need to multiply the i^{th} row with every column.

Computation to multiply with 'i' column is 'm'.

To compute every entry of i^{th} row $\Rightarrow m^2$.

Now to compute $[A^t]_{ij}$ we need $[A^{t-1}]_{j^{th} \text{ row}}^{i^{th} \text{ row}}$ and multiply it with j^{th} column of $A \Rightarrow O(m)$. ③

To compute A^2 i^{th} row it took m^2 steps.

A^3 i^{th} row it takes $2m^2$ steps.

Hence A^{t-1} i^{th} row it takes $(t-1)m^2$ steps - ④

Total complexity = $O((t-1)m^2 + m) = O(m^2 t)$.

From ③ & ④

(C) Another way is to compute (A^t) directly and then take the $i^{th} j^{th}$ entry of the matrix.

A^t can be computed by finding $A^{t/2}$ then squaring it i.e.,

$$A^t = A^{t/2} \cdot A^{t/2}$$

(if t is odd we may need to multiply it with A' again).

Consider this as a recursive algorithm

$$T(t) = T(t/2) + O(\text{matrix multiplication complexity}).$$

Now to multiply two matrices $A \Rightarrow n \times k$, $B \Rightarrow k \times y$ it takes $O(nyk)$ because each entry takes $O(k)$ and we have ny entries to compute. Hence $O(nyk)$.

In our case $n=y=k=m$. Hence matrix multiplication

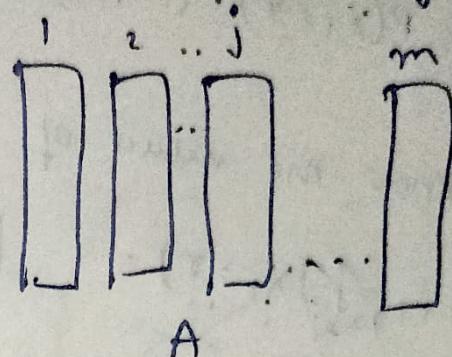
$$\text{complexity} = O(m^3)$$

$$T(t) = T(t/2) + O(m^3)$$

$$T(t) = m^3 \log_2 t \Rightarrow \text{then we pick } (ij)^{\text{th}} \text{ entry.}$$

Hence At computation complexity is $O(m^3 \log_2 t)$.

d) From 1b) to compute $(A^2)_{ij}$ we take i th row of A and multiply it with all columns of A .



Now instead of multiply all entries of i th row (ie. m terms) we can only take s entries (which are non-zero) and multiply with corresponding indices in the column.

Hence to compute the i th row of A^2 we have $O(sm)$ complexity. Hence ~~too~~ to compute.

A^{+1} i th row we have $O(smt-1)$.

Total complexity = $O(smt-1) + O(m)$

L) to compute $(A^2)_{ij}$
from i th row A^{+1} and j th column of A .

$$= O(smt-1) + O(m) = O(smt).$$

$$e) P(X_1=i | X_{T+1}=j) = \frac{P(X_{T+1}=j | X_1=i)P(X_1=i)}{P(X_{T+1}=j)}$$

$$= \frac{P(X_{T+1}=j | X_1=i)P(X_1=i)}{\sum_{x_1=1}^m P(X_{T+1}=j, X_1=x_1)} \quad [\text{Marginalization}].$$

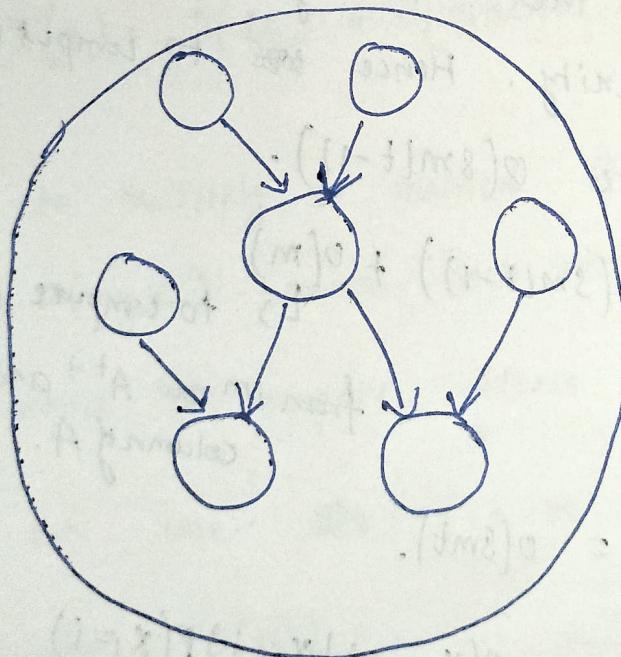
$$= \frac{P(x_{T+1} = j | x_1 = i) P(x_1 = i)}{\sum_{r=1}^m P(x_1 = r) P(x_{T+1} = j | x_1 = r)} \quad [\text{product rule}].$$

\Rightarrow we know the values of $P(x_1 = r)$ (prior probabilities)
 $P(x_{T+1} = j | x_1 = r) = [A^t]_{rj}$ where A^t is ~~the~~ t^{th} power of A .

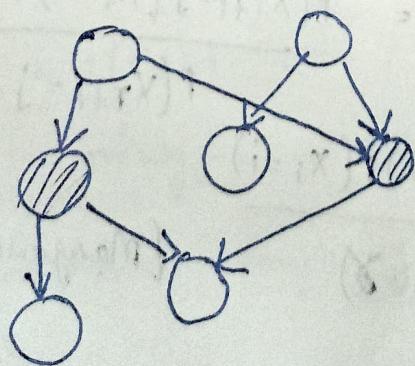
$$= \frac{[A^t]_{ij} P(x_1 = i)}{\sum_{r=1}^m P(x_1 = r) [A^t]_{rj}}$$

= .

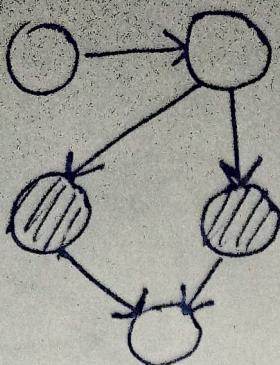
3.3)
B200



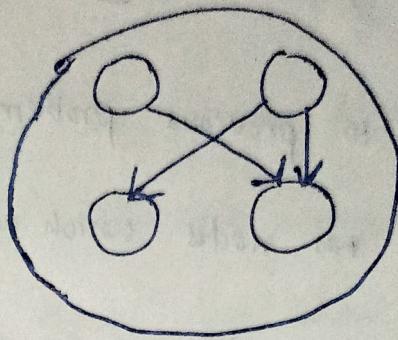
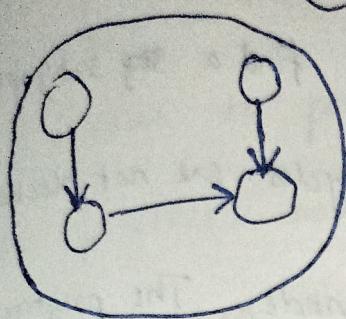
Polytree
(No cycles).



nodes which are shaded can be clustered, then it is a polytree.

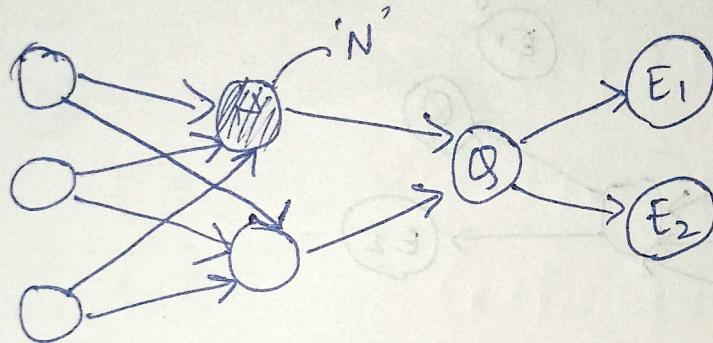


nodes which are shaded can be clustered, then it's a polytree.

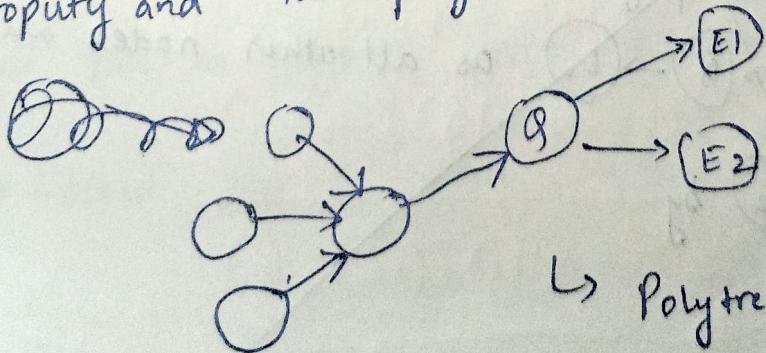


Polytree (No cycles).

3.4) a)

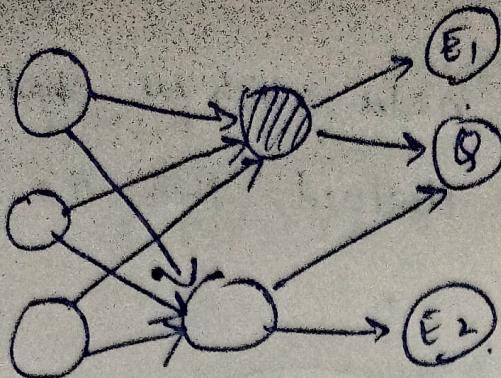


We cannot find a ~~subgraph~~ subgraph which is a polytree as there are nodes which are not blocked by E_1, E_2 using d-separation. and they are causing cycles. Hence we shade node 'N'. and then we ^{can} remove it by cutset property and run polytree

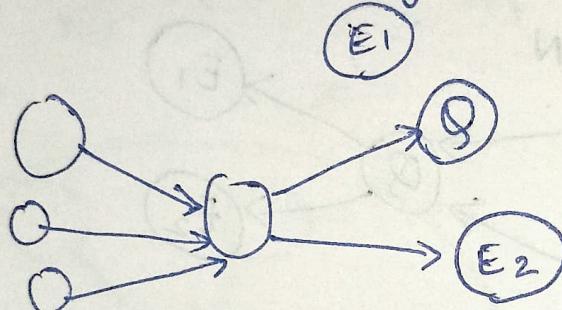


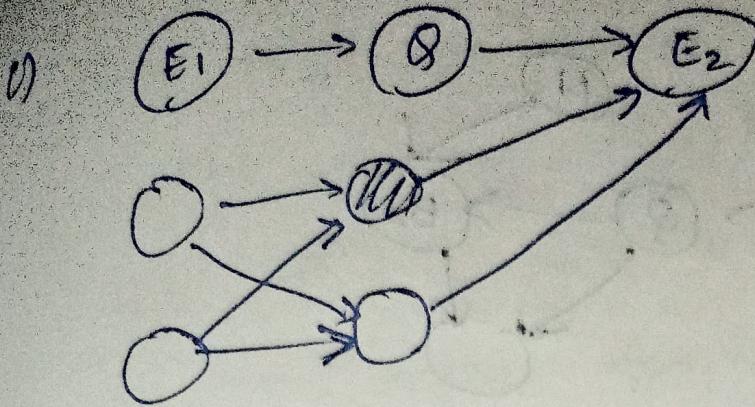
↳ Polytree.

b)

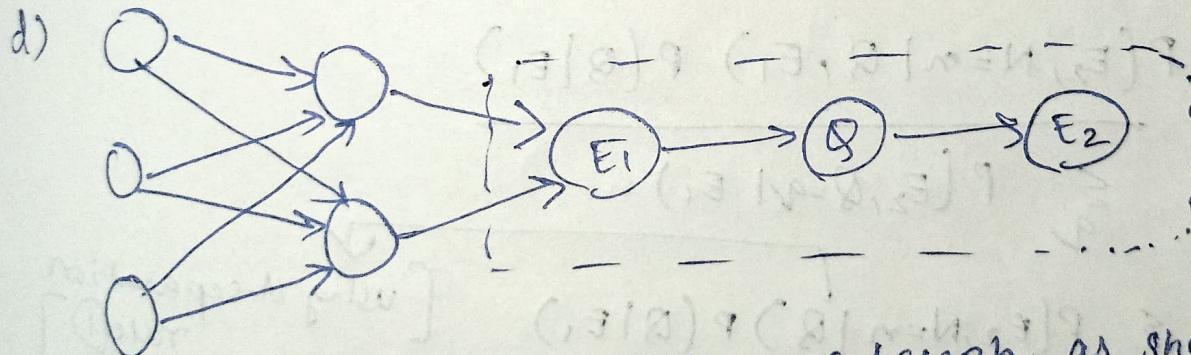
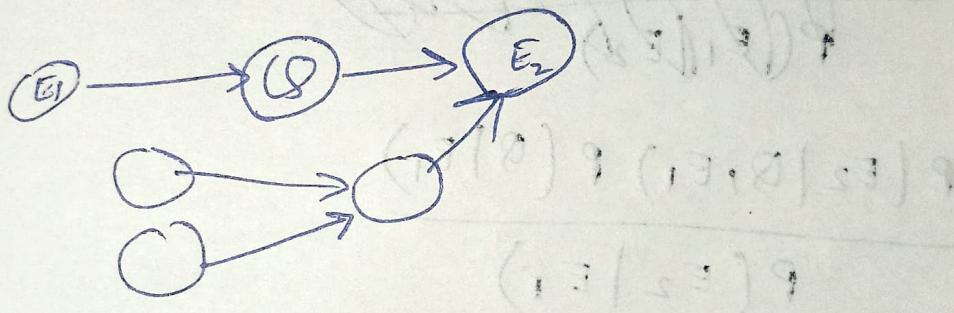


Similar to previous problem we cannot find a ~~big~~ subgraph as polytree, as nodes which are causing cycles are not blocked by E_1, E_2 . Hence we shade one node. The polytree obtained by cutset property is

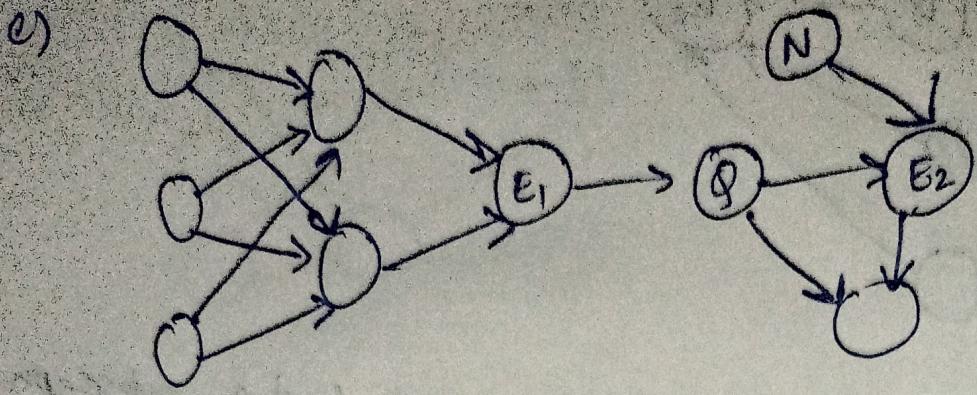




similar to previous problem we cannot find a subgraph as polytree, as nodes ~~whose~~ which are causing cycles are not blocked by E_1, E_2 . Hence we shade one node. The polytree obtained by subset property.



We can consider the subgraph as shown in the dotted line as polytree, remaining part of the graph is blocked by E_1 node using d-separation.



we need $P(Q|E_1, E_2)$, I have marked parent of E_2 as 'N'.

Now $P(Q|E_1, E_2)$ [Bayes rule $P(X|Y, E) = \frac{P(Y|X, E)P(X|E)}{P(Y|E)}$]

$$\frac{P(Q|E_1, E_2) P(E_1, E_2)}{P(E_1, E_2)}$$

$$= \frac{P(E_2|Q, E_1) P(Q|E_1)}{P(E_2|E_1)}$$

$$= \sum_n P(E_2, N=n|Q, E_1) P(Q|E_1)$$

$$\sum_q P(E_2, Q=q|E_1)$$

$$= \sum_n P(E_2, N=n|Q) P(Q|E_1) \quad \text{[using d-separation rule ①]}$$

$$\sum_q P(Q=q|E_1) P(E_2|E_1, Q=q)$$

ALL VALUES
CAN BE
COMPUTED
FROM
(CPT) = $\sum_n P(E_2, N=n|Q) P(Q|E_1)$

$$\sum_q P(Q=q|E_1) P(E_2|Q=q)$$

$$= \sum_n P(E_2, N=n|Q) P(Q|E_1)$$

$$\sum_q \sum_n P(Q=q|E_1) P(E_2, N=n|Q=q)$$

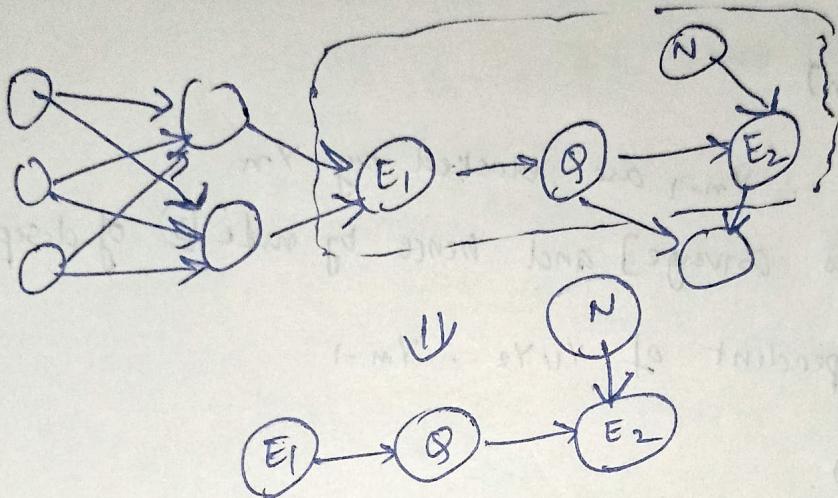
d-separation rule ①

$$\sum_n P(E_2, N=n|Q) P(Q|E_1)$$

$$\sum_q \sum_n P(Q=q|E_1) P(N=n|Q=q)$$

From eq(5) we can see that

the value of $P(Q|E_1, E_2)$ only depends on node 'N', E_1, E_2
since including node 'N' does not cause any cycles we
can run polytree algorithm on below dotted subgraph.



3.2) a) $P(Y_1|X_1) = \sum_{x_0} P(Y_1, X_0=x_0 | X_1) \quad [\text{marginalization}]$

$$= \sum_{x_0} P(X_0=x_0 | X_1) P(Y_1 | X_1, X_0=x_0)$$

$$= \sum_{x_0} P(X_0=x_0) P(Y_1 | X_1, X_0=x_0) \quad [\text{d-separation rule (3)}]$$

$P(X_0=x_0), P(Y_1 | X_1, X_0=x_0)$ can be computed from
CPT.

b) $P(Y_1) = \sum_{x_0} \sum_{x_1} P(Y_1, X_0=x_0, X_1=x_1) \quad [\text{marginalization}]$

$$= \sum_{x_0} \sum_{x_1} P(X_0=x_0) P(Y_1, X_1=x_1 | X_0=x_0) \quad [\text{product rule}]$$

$$= \sum_{x_0} \sum_{x_1} P(X_0=x_0) P(X_1=x_1 | X_0=x_0) P(Y_1 | X_0=x_0, X_1=x_1) \quad [\text{product rule}]$$

$$= \sum_{x_0} \sum_{x_1} P(X_0=x_0) P(X_1=x_1) P(Y_1 | X_0=x_0, X_1=x_1) \quad \rightarrow [\text{d-separation rule (3)}]$$

$$= \sum_{x_0} \sum_{x_1} P(x_0=x_0) P(x_1=x_1) P(y_1 | x_0=x_0, x_1=x_1)$$

↓ ↓

All these values can be obtained from CPT.

c) $P(x_n | y_1, y_2, \dots, y_{n-1})$

$$= P(x_n)$$

Since y_1, y_2, \dots, y_{n-1} are blocked by y_n

[edges converge] and hence by rule ③ of d-separation

x_n is independent of y_1, y_2, \dots, y_{n-1}

$$= P(x_n).$$

=

d) $P(y_n | x_n, y_1, y_2, \dots, y_{n-1})$

$$= \sum_x P(y_n, x_{n-1}=x | x_n, y_1, y_2, \dots, y_{n-1}) \quad [\text{Marginalization}]$$

$$= \sum_x P(x_{n-1}=x | x_n, y_1, \dots, y_{n-1}) P(y_n | y_1, y_2, \dots, y_{n-1}, x_n, x_{n-1})$$

⇒ x_{n-1} is conditionally independent of y_n given y_1, y_2, \dots, y_{n-1}

⇒ ~~Eq 8.8~~ [d-separation rule ③, blocked by y_n]

y_n is independent of y_1, y_2, \dots, y_{n-1} given x_{n-1}, x_n .

x_{n-1} blocks by rule ② edges converge of
d-separation.

$$= \sum_n P(X_{n+1} = n | Y_1, Y_2, \dots, Y_{n-1}) P(Y_n | X_{n+1} = n)$$

↑
 value given
 ↓
 value can be obtained from
 CPT table.

e) $P(Y_n | Y_1, Y_2, \dots, Y_{n-1})$

$$= \sum_{n_{\text{obs}}} \sum_n P(Y_n, X_n = n, X_{n+1} = n_{\text{obs}} | Y_1, Y_2, \dots, Y_{n-1})$$

$$= \sum_{n_{\text{obs}}} \sum_n P(X_{n+1} = n_{\text{obs}} | Y_1, Y_2, \dots, Y_{n-1}) \cdot P(Y_n | X_{n+1} = n_{\text{obs}} | Y_1, Y_2, \dots, Y_{n-1}) \quad [\text{product rule}]$$

$$= \sum_{n_{\text{obs}}} \sum_n P(X_{n+1} = n_{\text{obs}} | Y_1, Y_2, \dots, Y_{n-1})$$

$$\quad \quad \quad P(X_{n+1} = n_{\text{obs}} | Y_1, Y_2, \dots, Y_{n-1}) P(Y_n | Y_1, Y_2, \dots, Y_{n-1}, X_{n+1})$$

X_{n+1} is independent of X_{n+1} given Y_1, Y_2, \dots, Y_{n-1}
 as by d-separation rule ③ [Y_n converges]

$$= \sum_{n_{\text{obs}}} \sum_n P(X_{n+1} = n | Y_1, Y_2, \dots, Y_{n-1}) P(X_n = n_{\text{obs}}) P(Y_n | Y_1, Y_2, \dots, Y_{n-1})$$

Y_n is independent of Y_1, Y_2, \dots, Y_{n-1} given X_{n+1}, X_n
 Hence we get as X_{n+1} blocks other nodes by rule ②
 of d-separation

$$= \sum_{n_{\text{obs}}} \sum_n P(X_{n+1} = n | Y_1, Y_2, \dots, Y_{n-1}) P(X_n = n_{\text{obs}}) P(Y_n | X_{n+1} = n_{\text{obs}}, X_n = n_{\text{obs}})$$

↓
 given
 ↓
 CPT
 ↓
 CPT