Thomas von Belle   Statistical Genomics   1/4

Exam: 06-04-2018   52766582

Problem 1

a)

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$ ✓

b) - Single edge deletion is always valid, since we have 8 edges we can reach 8 neighbor graphs by single edge deletion. ✓

- Single edge addition, we have to be careful, since we cannot create cycles. We can add the following edges:

$A \to D$ $\quad D \to A$ $\quad E \to C$ $\quad F \to C$

$A \to E$ $\quad E \to A$ $\quad F \to A$

$A \to F$ $\quad E \to B$ $\quad F \to B$

This totals 10 edge additions. ✓

- Single edge reversal, we again should be careful for cycles but we can reverse:

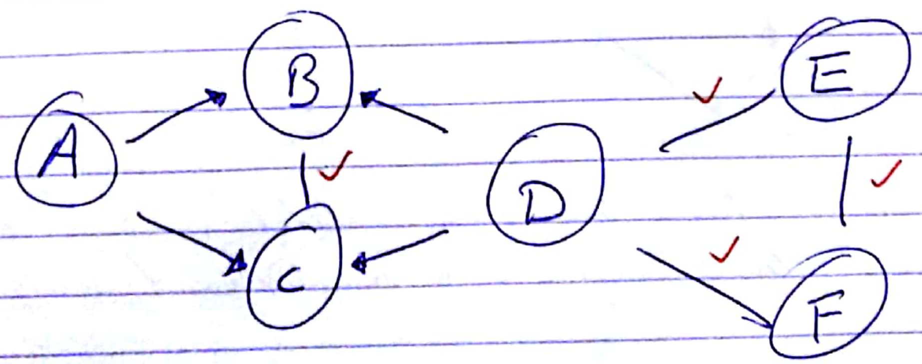
$A \to B$ $\quad D \to B$ $\quad F \to E$

$B \to C$ $\quad D \to E$

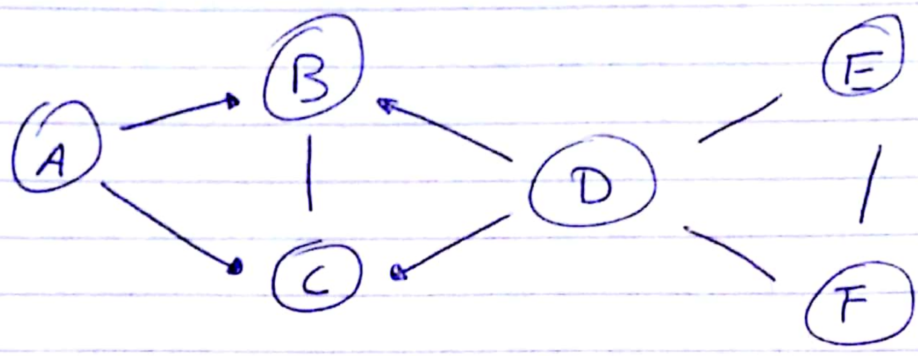
This totals 5 edge reversals ✓

In total we can reach $8 + 10 + 5 = 23$ neighbor graphs. ✓

c) Gv :

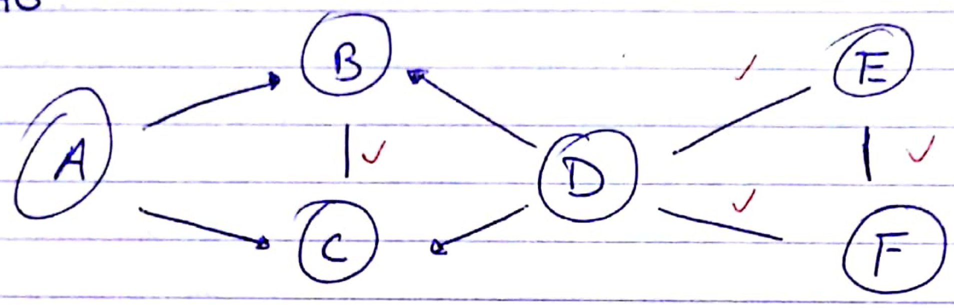


G*



f)

CPDAG



d) $\binom{4}{4} + \binom{4}{3} + \binom{4}{2} + \binom{4}{1} + \binom{4}{0} = \underbrace{1 + 4 + 6 + 4 + 1}_{=16} = 15$

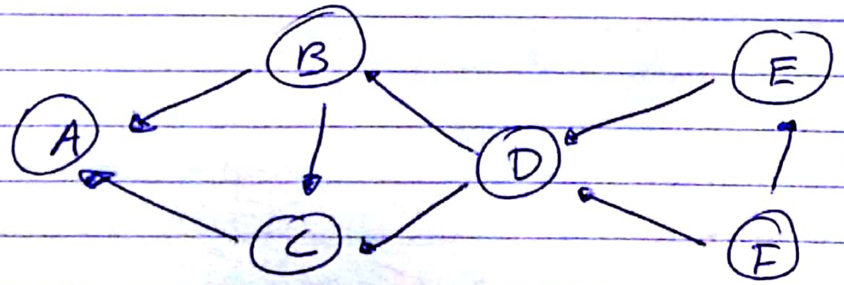
equivalence classes.

This is since we have 4 directional edges in the CPDAG. hence we can swap 4 out of those or any 3 or any 2 or just 1 and we will end up in a different equivalence class.

CYCLES ? −2P

NEW V-STRUCTURES ?

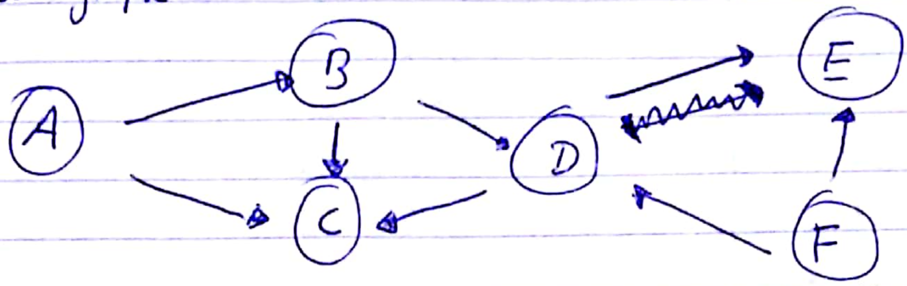
e) Consider the graph:



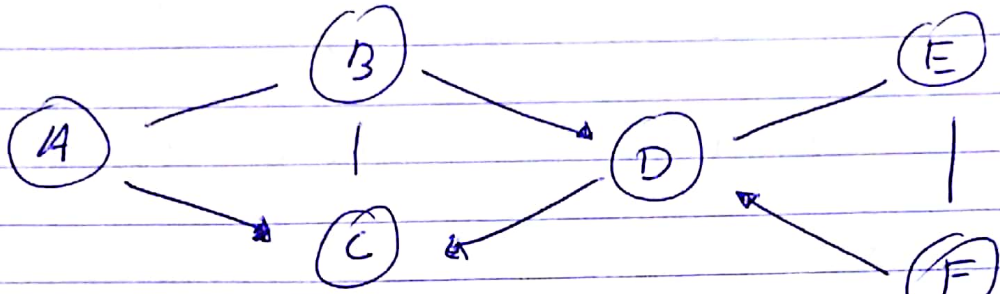
This graph has the same skeleton, but no v-structures Therefore, there exists such a DAG. ✓

Scanned by CamScanner

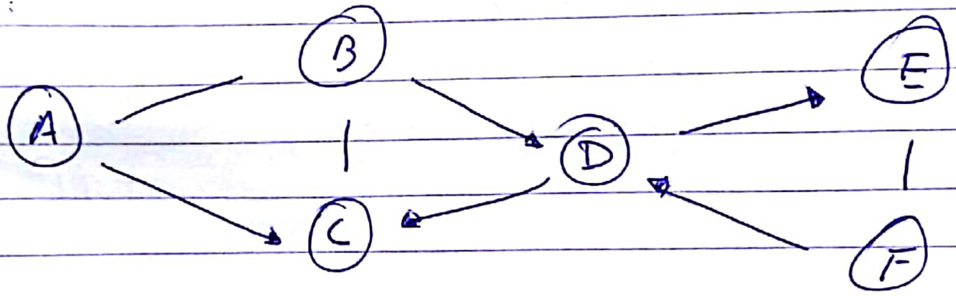
f) The graph:



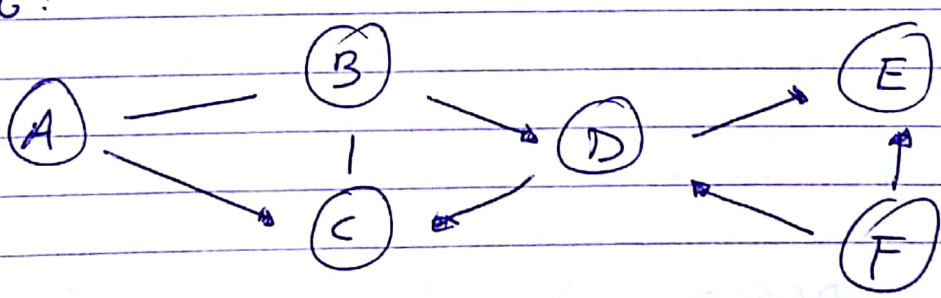
has G_V:



G_{V+}:



CPDAG:



Therefore, the edge $F \to E$ is compelled.

g)

$$MB(D) = \{\underbrace{E, F}_{\text{parents}}, \underbrace{B, C}_{\text{children}}, \underbrace{A}_{\text{co-parent}}\}$$

h) $A \to B \leftarrow D \leftarrow E$ blocked collides on B
$A \to C \leftarrow D \leftarrow E$ blocked collides on C
$A \to C \leftarrow B \leftarrow D \leftarrow E$ blocked " C
$A \to B \to C \leftarrow D \leftarrow E$ blocked " C
$A \to B \leftarrow D \leftarrow F \to E$ blocked " B
$A \to C \leftarrow D \leftarrow F \to E$ blocked " C
$A \to B \to C \leftarrow D \leftarrow F \to E$ blocked " C
$A \to C \leftarrow B \leftarrow D \leftarrow F \to E$ blocked " C

✓

i) Conditional on $Z = \{C\}$ the following paths are open:
$A \to C \leftarrow D \leftarrow E$
$A \to C \leftarrow B \leftarrow D \leftarrow E$
$A \to B \to C \leftarrow D \leftarrow E$
$A \to C \leftarrow D \leftarrow F \to E$
$A \to B \to C \leftarrow D \leftarrow F \to E$
$A \to C \leftarrow B \leftarrow D \leftarrow F \to E$.

2 MISSING
−1P

j) factorisation:

$$\mathbb{P}(A,B,C,D,E,F) = \mathbb{P}(A)\cdot\mathbb{P}(B|A,D)\cdot\mathbb{P}(C|A,B,D)\cdot\mathbb{P}(D|E,F)\cdot\mathbb{P}(E|F)\cdot\mathbb{P}(F)$$

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

$$\mathbb{P}(D,E,F) = \mathbb{P}(D|E,F)\cdot\mathbb{P}(E|F)\cdot\mathbb{P}(F).$$ WHY?
−1P

Therefore,

$$\mathbb{P}(A,B,C|D,E,F) = \mathbb{P}(A)\cdot\mathbb{P}(B|A,D)\cdot\mathbb{P}(C|A,B,D)$$

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

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

□

26/30

2/4

Thomas van Belle   Statistical Genomics

Exam: 06-04-2018   S2766586.

Problem 21

a)

$Q(G_1, G_2) = 0$ ✓ $Q(G_2, G_1) = 0$ ✓

$Q(G_1, G_3) = 1$ ✓   ③ $Q(G_3, G_1) = \cancel{0}\ \frac{1}{2}$ ✓

$Q(G_2, G_3) = 1$ ✓ $Q(G_3, G_2) = \frac{1}{2}$ ✓

here $|N(G_1)| = 1$, $|N(G_2)| = 1$, $|N(G_3)| = 2$.

-1

$$A(G_3, G_1) = \min\left\{1, \frac{P(D|G_1)\cdot P(G_1)}{P(D|G_3)\cdot P(G_3)} \cdot \left(\frac{Q(G_3,G_1)}{Q(G_1,G_3)}\right)\right\}$$

$$= \min\left\{1, \frac{20a \cdot 0{,}4 \cdot 1/2}{a \cdot 0{,}4 \cdot 1}\right\}$$

$$= \min\{1, 10\} = 1$$ -1

$$A(G_1, G_3) = \min\left\{1, \frac{P(D|G_3)\cdot P(G_3)}{P(D|G_1)\cdot P(G_1)} \cdot \left(\frac{Q(G_1,G_3)}{Q(G_3,G_1)}\right)\right\}$$

$$= \min\left\{1, \frac{a \cdot 0{,}4 \cdot 1}{20 \cdot a \cdot 0{,}4 \cdot 1/2}\right\} = \frac{1}{10}$$ -1

$$A(G_3, G_2) = \min\left\{1, \frac{P(D|G_2)\cdot P(G_2)}{P(D|G_3)\cdot P(G_3)} \cdot \left(\frac{Q(G_3,G_2)}{Q(G_2,G_3)}\right)\right\}$$

$$= \min\left\{1, \frac{20a \cdot 0{,}2 \cdot \frac{1}{2}}{a \cdot 0{,}4 \cdot 1}\right\}$$

$$= \min\left\{1, \frac{2a}{0{,}4a}\right\} = 1$$

OVERALL

$-2P$

-1

$$A(G_2, G_3) = \min\left\{1, \frac{\mathbb{P}(D|G_3)\cdot\mathbb{P}(G_3)}{\mathbb{P}(D|G_2)\cdot\mathbb{P}(G_2)}\cdot\frac{Q(G_2,G_3)}{Q(G_3,G_2)}\right\}$$

-1

$$= \min\left\{1, \frac{a \cdot 0{,}4 \cdot 1}{20a \cdot 0{,}2 \cdot \frac{1}{2}}\right\}$$

$$= \min\left\{1, \frac{1}{20}\cdot 2 \cdot 2\right\}$$

$$= \frac{1}{5}.$$

Therefore, since $T(G, G^*) = Q(G, G^*) \cdot A(G, G^*)$

We obtain:

$$T^{(a)} = \begin{bmatrix} 9/10 & 0 & \frac{1}{10} \\ 0 & \frac{4}{5} & \frac{1}{5} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

$\boxed{8/10}$

b) $Q(G_1, G_2) = \frac{1}{2}$ ✓ $\quad Q(G_2, G_1) = \frac{1}{2}$ ✓

$Q(G_1, G_3) = \frac{1}{2}$ ✓ $\quad Q(G_3, G_1) = \frac{1}{2}$ ✓

$Q(G_2, G_3) = \frac{1}{2}$ ✓ $\quad Q(G_3, G_2) = \frac{1}{2}$. ✓

here

$|N(G_1)| = |N(G_2)| = |N(G_3)| = 2.$

Scanned by CamScanner

Then,

$$A(G_1, G_2) = \min\left\{1, \frac{P(D|G_2) \cdot P(G_2)}{P(D|G_1) \cdot P(G_1)}\right\}$$

$$= \min\left\{1, \frac{0,2}{0,4}\right\} = \frac{1}{2}.$$

$$A(G_2, G_1] = \min\left\{1, \frac{0,4}{0,2}\right\} = 1$$

$$A(G_1, G_3) = \min\left\{1, \frac{P(D|G_3) \cdot P(G_3)}{P(D|G_1) \cdot P(G_1)}\right\}$$

$$= \min\left\{1, \frac{a \cdot 0,4}{20a \cdot 0,4}\right\} = \frac{1}{20}$$

$$A(G_3, G_1) = \min\{1, 20\} = 1$$

$$A(G_2, G_3) = \min\left\{1, \frac{P(D|G_3) \cdot P(G_3)}{P(D|G_2) \cdot P(G_2)}\right\}$$

$$= \min\left\{1, \frac{a \cdot 0,4}{20a \cdot 0,2}\right\} = \frac{1}{10}$$

$$A(G_3, G_2) = \min\{1, 10\} = 1.$$

Hence,

(b)

$$\underline{T} = \begin{bmatrix} 29/40 & \frac{1}{4} & \frac{1}{40} \\ \frac{1}{2} & 9/20 & \frac{1}{20} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$ ✓

$\boxed{10/10}$

Scanned by CamScanner

c)

$\pi^{(a)} \cdot T^{(a)} = \pi^{(a)}$

where

$\pi^{(a)} = (p_1, p_2, p_3)$

Then, we get:

FOLLOW-UP MISTAKE

$T^{(a)}$ WRONG

APPROACH OKAY

$p_1 \cdot 9/10 + p_3 \cdot \frac{1}{2} = p_1 \Rightarrow p_1 = 5 p_3$

$p_2 \cdot 4/5 + p_3 \cdot \frac{1}{2} = p_2 \Rightarrow p_2 = \frac{5}{2} p_3$

$p_1 \cdot \frac{1}{10} + p_2 \cdot \frac{1}{5} = p_3 \Rightarrow \frac{1}{2} p_3 + \frac{1}{2} p_3 = p_3$ &.

& $p_1 + p_2 + p_3 = 1$

$5p_3 + 5/2\, p_3 + p_3 = 1 \Rightarrow 17/2 \cdot p_3 = 1$

$\Rightarrow p_3 = 2/17$

$\pi^{(a)} = (10/17,\ 5/17,\ 2/17)$

$\pi^{(b)} \cdot T^{(b)} = \pi^{(b)}$

$\pi^{(b)} = (p_1, p_2, p_3)$.

$\boxed{5/5}$

Then,

$p_1 \cdot 29/40 + \frac{1}{2} p_2 + \frac{1}{2} p_3 = p_1$

$p_1 \cdot 11/4 + p_2 \cdot 9/20 + p_3 \cdot 1/20 = p_2$

$p_1 \cdot \frac{1}{40} + p_2 \cdot \frac{1}{20} = p_3$ See last page!

~~$p_2 = \frac{20}{44} \cdot p_1 + \frac{1}{11} \cdot p_3 \Rightarrow \frac{1}{2} p_1 + \frac{20}{88} \cdot p_1 + \frac{1}{22} p_3 = p_3$~~

~~$\Rightarrow \frac{21}{22} p_3 = \frac{8}{11} p_1$~~

~~$p_3 = \frac{16}{21} \cdot p_1$~~

~~hence $p_2 = \frac{20}{44} p_1 + \frac{1}{11} \cdot (\frac{16}{21} p_1) = \frac{11}{21} p_1$~~

~~$p_1 + p_2 + p_3 = 1 \Rightarrow p_1 + \frac{11}{21} p_1 + \frac{16}{21} p_1 = 1$~~

~~$\Rightarrow \frac{16}{7} p_1 = 1 \Rightarrow p_1 = 7/16, \quad \pi^{(b)} = (7/16,\ 11/48,\ 1/3)$~~

Scanned by CamScanner

Thomas ... Bella Statistical Genomics
Exam: 06-04-2018 s2766582.

Problem 3)

a) $\mathbb{E}(X_1) = \mathbb{E}(2+\varepsilon_1) = 2 + \mathbb{E}(\varepsilon_1) = 2$

$\mathbb{E}(X_2) = \mathbb{E}(-X_1+\varepsilon_2) = \mathbb{E}(-X_1) + \mathbb{E}(\varepsilon_2)$
$= \mathbb{E}(-X_1) + 0$
$= -2$

$\mathbb{E}(X_3) = \mathbb{E}(2X_2+\varepsilon_3)$
$= \mathbb{E}(2X_2) + \mathbb{E}(\varepsilon_3)$
$= 2\mathbb{E}(X_2) + 0 = -4$

hence, $\mu = \begin{bmatrix} 2 \\ -2 \\ -4 \end{bmatrix}$ ✓

$Var(X_1) = Var(2+\varepsilon_1) = Var(\varepsilon_1) = 1$

~~$Var(X_2) = Var(-X_1+\varepsilon_2)$~~
~~$= \mathbb{E}(X_2^2) - \mathbb{E}(X_2)^2$~~

~~$\mathbb{E}(X_2^2) = \mathbb{E}((-X_1+\varepsilon_2)^2) = \mathbb{E}(X_1^2 - 2X_1\varepsilon_2 + \varepsilon_2^2)$~~
~~$= \mathbb{E}(X_1^2) - 2\mathbb{E}(X_1\varepsilon_2) + \mathbb{E}(\varepsilon_2^2)$~~
~~$= Var(X_1) + \mathbb{E}(X_1)^2$~~

See next page!

~~$cov(X_1, X_2) = \mathbb{E}((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2)))$~~

~~$= \mathbb{E}(X_1X_2 - X_2\mathbb{E}(X_1) - X_1\mathbb{E}(X_2) + \mathbb{E}(X_1)\mathbb{E}(X_2))$~~
~~$= \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = \mathbb{E}(X_1X_2) + 4$~~
~~$= \mathbb{E}(X_1(-X_1+\varepsilon_2)) \overset{+4}{=} \mathbb{E}((2+\varepsilon_1)(-2-\varepsilon_1+\varepsilon_2)) + 4$~~
~~$= \mathbb{E}(-4 - \varepsilon_1^2 - 2\varepsilon_1 + 2\varepsilon_2 + \varepsilon_1\varepsilon_2) + 4$~~
~~$= -1$~~

$$Var(X_2) = var(-X_1 + \varepsilon_2) = var(\varepsilon_2 - \varepsilon_1 - 2)$$
$$= var(\varepsilon_2 - \varepsilon_1)$$
$$= var(\varepsilon_2) + var(\varepsilon_1) = 2.$$

$$Var(X_3) = Var(2 \cdot X_2 + \varepsilon_3) = 4 \cdot var(X_2) + var(\varepsilon_3)$$
$$= 4 \cdot 2 + 1$$
$$= 9.$$

$$Cov(X_1, X_2) = cov(2 + \varepsilon_1, \varepsilon_2 - \varepsilon_1 - 2)$$
$$= cov(\varepsilon_1, \varepsilon_2 - \varepsilon_1)$$
$$= cov(\varepsilon_1, \varepsilon_2) - cov(\varepsilon_1, \varepsilon_1)$$
$$= 0 - var(\varepsilon_1) = -1 = cov(X_2, X_1)$$

$$\cancel{Cov(X_2, X_3) = Cov(\varepsilon_2 - \varepsilon_1 - 2, 2\varepsilon_2 + \varepsilon_3 - 2\varepsilon_1 - 4)}$$
$$\cancel{= Cov(\varepsilon_2 - \varepsilon_1, 2\varepsilon_2 + \varepsilon_3 - 2\varepsilon_1)}$$
$$\cancel{= Cov(\varepsilon_2, 2\varepsilon_2) + Cov(}$$

$$Cov(X_1, X_3) = Cov(X_1, 2X_2 + \varepsilon_3) = cov(X_1, 2X_2) + cov(X_1, \varepsilon_3)$$
$$= 2Cov(X_1, X_2) + cov(\varepsilon_1, \varepsilon_3) = 2 \cdot cov(X_1, X_2)$$
$$= -2 = cov(X_3, X_1)$$

$$Cov(X_2, X_3) = Cov(-X_1 + \varepsilon_2, X_3) = -cov(X_1, X_3) + cov(\varepsilon_2, X_3)$$
$$= \cancel{cov}\ 2 + cov(\varepsilon_2, -2X_1 + 2\varepsilon_2 + \varepsilon_3)$$
$$= 2 + 2 \cdot var(\varepsilon_2) + cov(\varepsilon_2, \varepsilon_3) - cov(\varepsilon_2, 2X_1)$$
$$= 4 - 2cov(\varepsilon_2, X_1)$$
$$= 4 - 2 \cdot cov(\varepsilon_2, 2 + \varepsilon_1) = 4 = cov(X_3, X_2)$$

hence $\Sigma = \begin{bmatrix} 1 & -1 & -2 \\ -1 & 2 & 4 \\ -2 & 4 & 9 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -2 \\ -1 & 2 & 4 \\ -2 & 4 & 9 \end{bmatrix}.$

10/10

b) $\mathbb{E}(X_3) = -4, \quad Var(X_3) = 9$

$$\mathbb{E}(X_2 \mid X_3 = a) = \mathbb{E}(X_2) \overset{+}{-} \frac{Cov(X_2, X_3)}{Var(X_3)} \cdot (a - \mathbb{E}(X_3))$$

$$= -2 - \frac{4}{9} \cdot (a+4)$$

$$= -2 - \frac{16}{9} - \frac{4a}{9} = -\left(\frac{34}{9} + \frac{4a}{9}\right) = -\frac{34+4a}{9}$$

$$Var(X_2 \mid X_3 = a) = \left(1 - \frac{Cov(X_2, X_3)^2}{Var(X_2)\, var(X_3)}\right) \cdot var(X_2)$$

$$= \left(1 - \frac{16}{18}\right) \cdot 2 = \frac{2}{9}$$

$$\mathbb{E}(X_1 \mid X_2 = a) = \mathbb{E}(X_1) \overset{+}{-} \frac{Cov(X_1, X_2)}{Var(X_2)} \cdot (a - \mathbb{E}(X_2))$$

$$= 2 - \frac{-1}{2} \cdot (a+2)$$

$$= 2 + \frac{a}{2} + 1 = 3 + \frac{a}{2}.$$

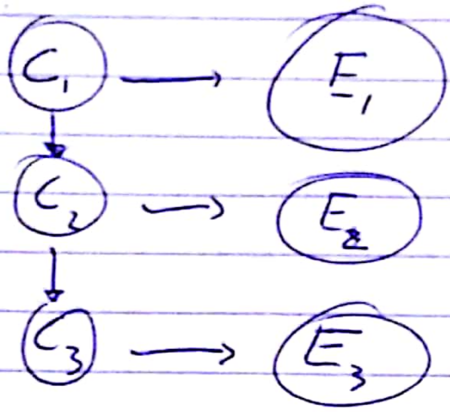
$$Var(X_1 \mid X_2 = a) = \left(1 - \frac{Cov(X_1, X_2)^2}{Var(X_1)\, Var(X_2)}\right) \cdot var(X_1)$$

$$= \left(1 - \frac{1}{2}\right) \cdot 1 = \frac{1}{2}.$$

Continues on last page.

Problem 4)

a) DAG:



$$\mathbb{P}(C_1, C_2, C_3, E_1, E_2, E_3) = \mathbb{P}(C_1) \cdot \mathbb{P}(C_2 | C_1) \cdot \mathbb{P}(C_3 | C_2) \cdot \mathbb{P}(E_1 | C_1) \cdot \mathbb{P}(E_2 | C_2) \cdot \mathbb{P}(E_3 | C_3)$$

5

b) $\mathbb{P}(E_2 = 1 \mid C_1 = 1, C_2 = 1, C_3 = 1, E_1 = 1, E_3 = 1)$

$= \mathbb{P}(E_2 = 1 \mid C_2 = 1) = 0,1$

2/2

- $\mathbb{P}(E_2 = 1 \mid C_1 = 1, C_3 = 1, E_1 = 1, E_3 = 1)$

$= \mathbb{P}(E_2 = 1 \mid C_2 = 1) \cdot \mathbb{P}(C_2 = 1 \mid C_1 = 1) + \mathbb{P}(E_2 = 1 \mid C_2 = 2) \cdot \mathbb{P}(C_2 = 2 \mid C_1 = 1)$

ALSO DEPENDS ON C3

$= 0,1 \cdot 0,8 + 0,9 \cdot 0,2$

$= 0,08 + 0,18 = 0,26$

4/8

5/5

EX4

6/10

11/15

Statistical genomics — Thomas van Belle 4/4

Exam: 06-04-2018 — S2766582

Problem ~~3~~ 2:

c) Continued:

$$P_1 \cdot {}^{29}/_{40} + \frac{1}{2} P_2 + \frac{1}{2} P_3 = P_1$$

$$P_1 \cdot {}^{1}/_{4} + {}^{9}/_{20} P_2 + {}^{1}/_{20} \cdot P_3 = P_2$$

$$P_1 \cdot {}^{1}/_{40} + P_2 \cdot {}^{1}/_{20} = P_3$$

Thus, $P_1 \cdot {}^{1}/_{4} + {}^{9}/_{20} P_2 + \frac{1}{80} P_1 + \frac{1}{40} P_2 = P_2$

$$\frac{21}{80} P_1 + {}^{19}/_{40} P_2 = P_2$$

$$\frac{21}{80} P_1 = \frac{21}{40} P_2$$

$$P_1 = 2 P_2.$$

$$P_1 + P_2 + P_3 = 1 \Rightarrow 2P_2 + P_2 + \frac{1}{40} P_1 + \frac{1}{20} P_2 = 1$$

$$\Rightarrow 2P_2 + P_2 + \frac{1}{20} P_2 + \frac{1}{20} P_2 = 1$$

$$\Rightarrow \frac{62}{20} P_2 = 1.$$

$$P_2 = \frac{20}{62} = \frac{10}{31}.$$

Hence, $\pi^{(b)} = \left(\frac{20}{31}, \frac{10}{31}, \frac{1}{31} \right)$ ✓

Problem 3

b) continued:

~~The~~ regression equations are given by:

$$X_3 = -4 + \tilde{\varepsilon}_3 \quad ✓$$

$$X_2 = -\frac{34}{9} - \frac{4}{9} \cdot X_3 + \tilde{\varepsilon}_2$$

$$X_1 = 3 + \frac{1}{2} X_2 + \tilde{\varepsilon}_1$$

~~Where~~ $\tilde{\varepsilon}_3 \sim N(0, 3^2)$

Thus, $\mathbb{E}(X_3) = -4$

$$\mathbb{E}(X_2) = -\frac{34}{9} - \frac{4}{9} \cdot -4 = -2$$

$$\mathbb{E}(X_1) = 3 + \frac{1}{2} \cdot -2 = 2$$

Also, $Var(X_3) = var(\tilde{\varepsilon}_3) = 9.$ ✓

$$Var(X_2) = \frac{16}{81} \cdot Var(X_3) + Var(\tilde{\varepsilon}_2)$$

$$= \frac{16}{9} + var(\tilde{\varepsilon}_2) = 2$$

hence $var(\tilde{\varepsilon}_2) = 2 - \frac{16}{9} = \frac{2}{9}$ ✓

So, $\tilde{\varepsilon}_2 \sim N\left(0, \left(\sqrt{2}/3\right)^2\right).$

$$Var(X_1) = \frac{1}{4} \cdot Var(X_2) + Var(\tilde{\varepsilon}_1)$$

$$= \frac{1}{4} \cdot 2 + var(\tilde{\varepsilon}_1) = 1$$

hence, $Var(\tilde{\varepsilon}_1) = \frac{1}{2},$ ✓

so $\tilde{\varepsilon}_1 \sim N\left(0, \left(\frac{1}{\sqrt{2}}\right)^2\right).$

Scanned by CamScanner

hence,

$$X_3 = -4 + \tilde{\varepsilon}_3$$ ✓

$$X_2 = -\frac{34}{9} - \frac{4}{9} X_3 + \tilde{\varepsilon}_2$$

$$X_1 = 3 + \frac{1}{2} X_2 + \tilde{\varepsilon}_1$$

With, $\tilde{\varepsilon}_3 \sim \mathcal{N}(0, 3^2)$

$\tilde{\varepsilon}_2 \sim \mathcal{N}(0, (\sqrt{2}/3)^2)$ ✓

$\tilde{\varepsilon}_1 \sim \mathcal{N}(0, (1/\sqrt{2})^2)$.

ONE SYSTEMATIC MISTAKE IN EQUATION

−2P

$\boxed{8/10}$