Rylan Schaeffer

6.438 Problem Set 1

1.1

a) Suppose $x \perp\!\!\!\perp y \mid z$. By defn of C.I.

$$p(x, y \mid z) = p(x \mid z)\, p(y \mid z)$$

$$\Rightarrow p(x, y, z) = \underbrace{p(x \mid z)}_{h(x,z)}\ \underbrace{p(y \mid z)\, p(z)}_{g(y,z)}$$

Suppose $p(x, y, z) = h(x, z)\, g(y, z)$. Define $h(x, z) = p(x \mid z)$ and

$g(y, z) = p(y \mid z)\, p(z)$. By chain rule

$$p(x, y, z) = p(x \mid y, z)\, p(y \mid z)\, p(z)$$

$$= p(x \mid z)\, p(y \mid z)\, p(z)$$

$$\Rightarrow x \perp\!\!\!\perp y \mid z$$

b) We want pairwise but not mutual independence for 3 events A, B, C

$$P(A, B, C) = P(A)\, P(B \mid A)\, P(C \mid A, B) \quad \text{by chain rule}$$

$$= P(A)\, P(B)\, P(C \mid A, B) \quad \text{by pairwise indep.}$$

$$\neq P(A)\, P(B)\, P(C)$$

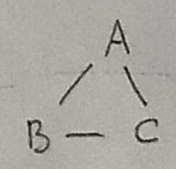
$$\Rightarrow P(C \mid A, B) \neq P(C)$$

Let $A, B \sim \text{Bern}(p = 0.5)$. Defince $C = A \text{ XOR } B$. Then

A	B	C
0	0	1
0	1	0
1	0	0
1	1	1

$$\Rightarrow \left.\begin{array}{ll} P(C=0) = \frac{1}{2} & \text{and } P(C=1) = \frac{1}{2} \\ P(C=0 \mid A) = \frac{1}{2} & \text{and } P(C=1 \mid A) = \frac{1}{2} \\ P(C=0 \mid B) = \frac{1}{2} & \text{and } P(C=1 \mid B) = \frac{1}{2} \end{array}\right\} \Rightarrow \text{C is pairwise independent}$$

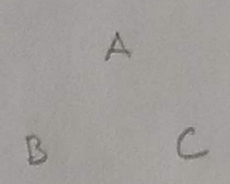
Undirected Graph:



1.1

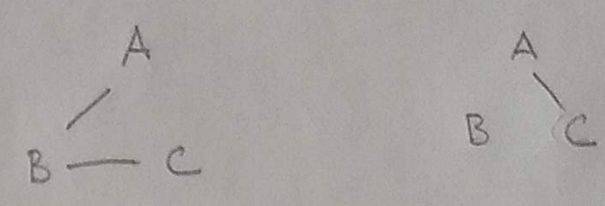
b cont) Removing all edges $\Rightarrow$ all independent

WRONG



Removing 1 or 2 edges breaks symmetry

WRONG



1.2

a) Symmetry:

$$P(x_A, x_B | x_C) = P(x_A | x_C) P(x_B | x_C)$$
$$= P(x_B | x_C) P(x_A | x_C)$$
$$= P(x_B, x_A | x_C) \quad \Rightarrow \quad x_B \perp\!\!\!\perp x_A | x_C$$

b) Decomposition:

$$P(x_A, x_B, x_D | x_C) = P(x_A | x_C) P(x_B, x_D | x_C)$$
$$\sum_{x_D} \qquad = \sum_{x_D}$$
$$P(x_A, x_B | x_C) = P(x_A | x_C) \sum_{x_D} P(x_B, x_D | x_C)$$
$$= P(x_A | x_C) P(x_B | x_C) \quad \Rightarrow \quad x_A \perp\!\!\!\perp x_B | x_C$$

Redo instead marginalizing over $x_B \Rightarrow x_A \perp\!\!\!\perp x_D | x_C$

c) Weak Union

$$P(x_A, x_B | x_C, x_D) = P(x_A | x_B, x_C, x_D) P(x_B | x_C, x_D) \qquad \text{(Chain Rule)}$$
$$= P(x_A | x_C, x_D) P(x_B | x_C, x_D) \qquad \text{(CI from (b))}$$
$$\Rightarrow \quad x_A \perp\!\!\!\perp x_B | x_C, x_D$$

d) Contraction:

$$P(x_A, x_B, x_C | x_D) = P(x_A | x_B | x_D) P(x_C | x_A, x_B, x_D) \qquad \text{(Chain Rule)}$$
$$= P(x_A | x_D) P(x_B | x_D) P(x_C | x_A, x_B, x_D) \qquad \text{CI 1}$$
$$= P(x_A | x_D) P(x_B | x_D) P(x_C | x_B, x_D) \qquad \text{CI 2}$$
$$= P(x_A | x_D) P(x_{B \cup C} | x_D)$$
$$\Rightarrow \quad x_A \perp\!\!\!\perp x_{B \cup C} | x_D$$

Rylan Schaeffer

6.438 Problem Set 1

1.3.1

a) i. True. Any path connecting B-G must pass H. H is not observed & has V-structure i.e. $G \to H \leftarrow F$, blocking the path

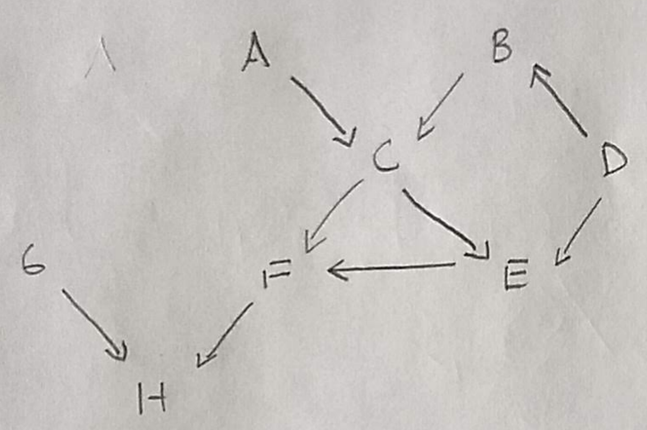
ii. False. Path $C \leftarrow B \to D$ and B is unobserved $\Rightarrow$ path unblocked $\Rightarrow$ not CI

iii. False. Path $C \leftarrow B \to D$ and B is unobserved $\Rightarrow$ path unblocked $\Rightarrow$ not CI

iv. True Path $B \to D \to E \leftarrow C \to F \to H$ has a $\to E \leftarrow$ but E is unobserved $\Rightarrow$ path is blocked

Paths $B \to C \to F \to H$, $B \to C \to E \to F \to H$, $B \to D \to E \to F \to H$ are all blocked by conditioning on C, F

b)



The only graph w/ same V-structures & skeleton as Fig 1.3.1 is w/ the B-D edge swapped

c) i. <u>Necessary</u>: Let G_1, G_2 be I-equivalent. For contradiction, add a directed edge to G_1. from $X_i \to X_j$. Then $X_{\pi_j}^{G_1} \neq X_{\pi_j}^{G_2}$ (i.e. node j has different parents). $\Rightarrow$

$$P(X_j^{G_1} \mid X_{\pi_j}^{G_1}) \neq P(X_j^{G_2} \mid X_{\pi_j}^{G_2})$$

$\Rightarrow G_1, G_2$ are not I-equivalent.

remove edge $\Rightarrow$ diff d-separation

<u>Insufficient</u>: $A \to B \leftarrow C$ and $A \to B \to C$ have same skeleton but different independencies i.e. $A \perp\!\!\!\perp C \mid B$ on the right, but $A \not\perp\!\!\!\perp C \mid B$ on the left

1.3

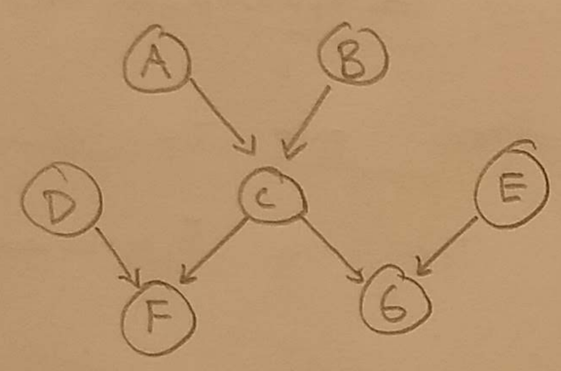
c) ii. Earlier part (i) showed that if G_1 & G_2 I-equivalent $\Rightarrow$ G_1 & G_2 have same skeleton

Claim: Suppose G_1 & G_2 are I equivalent. Here I show G_1 & G_2 must have same immoralities

<u>Proof:</u> For each node X_i, consider its Markov boundary X_{B_i}. There are 2 sources of immoralities

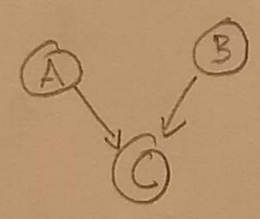
1. Multiple parents
2. Children w/ other additional parents

e.g.



Changing direction of parents induces conditional indep & removes pairwise indep $\Rightarrow$ violating I-equivalence

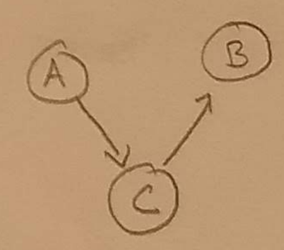
e.g. Before



$A \perp\!\!\!\perp B$

$A \not\perp\!\!\!\perp B | C$

After

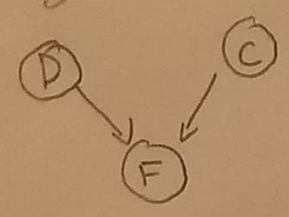


$A \not\perp\!\!\!\perp B$

$A \perp\!\!\!\perp B | C$

Similarly, changing direction of (shared) children violates indep

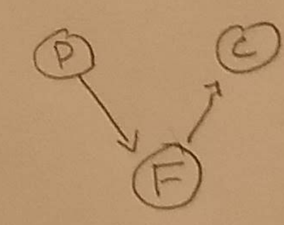
e.g. Before



$D \perp\!\!\!\perp C$

$D \not\perp\!\!\!\perp C | F$

After



$D \not\perp\!\!\!\perp C$

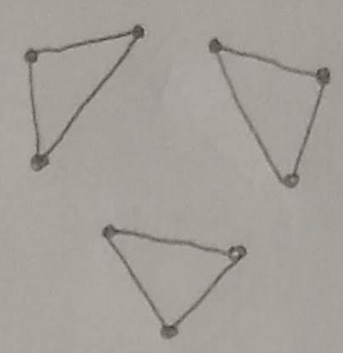
$D \perp\!\!\!\perp C | F$

$\Rightarrow$ For each node, its Markov boundary cannot change immoralities

$\Rightarrow$ If each Markov boundary can't change immoralities, G_1 & G_2 must have same immoralities

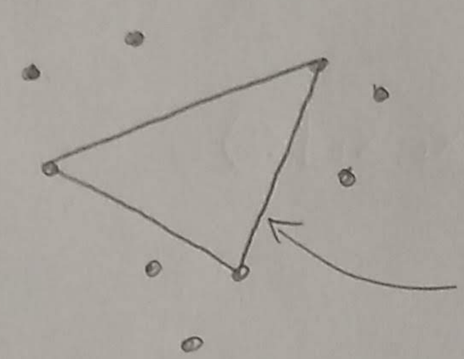
1.4

a) Consider undirected graph of disjoint triangles e.g.



Take its complement i.e. fully connected - triangle edges.

Any maximal clique can only include 1 node in each triangle (since nodes in same triangle are <u>not</u> connected) e.g.



The # of factor nodes = # of maximal cliques

Here, the number of maximal cliques is $\dfrac{\binom{3}{1}\binom{3}{1}\cdots\binom{3}{1}}{n/3}$

So the number of maximal cliques is $3^{n/3}$

Equivalently $(\sqrt[3]{3})^n$

Let $c = \sqrt[3]{3} > 1$. Then the aforementioned undirected graph has at least c^n vertices.

1.4

b) First, I claim no cycles $\Rightarrow$ original graph is <u>chordal</u> & <u>not $\tilde{K}_4$</u>

Proof by Contrapositive:

<u>CP Claim</u>: Not chordal or has $\tilde{K}_4$ $\Rightarrow$ cycle exists

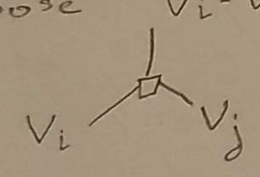
<u>Proof</u>: Consider not chordal

$\Rightarrow \exists$ cycle w/ len ≥ 4 s.t. no edge joins any 2 non-adj nodes. Say cycle is $x_1 - x_2 - \dots x_N - x_1$ $N \geq 4$

① Suppose $\forall i,j \in \{1, \dots N\}$ have no other connections.

Then factor graph $V_i \not\text{---} V_j$ & cycle preserved $\Rightarrow$ cycle exists

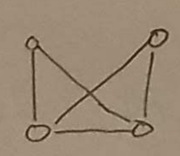
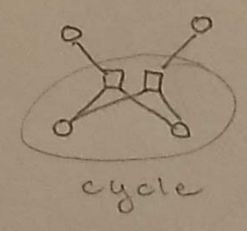
② Suppose $V_i - V_j$ belong to shared clique, Then

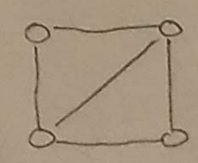
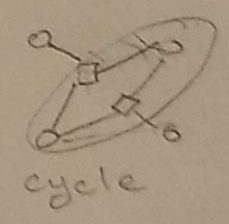


& cycle preserved $\Rightarrow$ cycle exists

③ Suppose $V_i - V_j$ belong to non-shared cliques. Then again the factor graph $V_i \not\text{---} V_j$ & cycle preserved $\Rightarrow$ cycle exists

Consider $\tilde{K}_4$

$\Rightarrow$  $\Rightarrow$ 

or  $\Rightarrow$ 

Algorithm: Check for chordality: $O(V)$

Check for $\tilde{K}_4$: $O(V^4)$

If graph chordal & not $\tilde{K}_4$, return True. Else return False

Total time $O(V^4 + V) = O(V^4)$ $\Leftarrow$ polynomial time

1.5

a. Consider soln to 1.1b: 3 binary r.v. s.t. $x \oplus y \oplus z = 0$ (XOR)

$I(p) = \{ x \perp\!\!\!\perp y,\ x \perp\!\!\!\perp z,\ y \perp\!\!\!\perp z \}$

Importantly, $x \perp\!\!\!\perp y | z,\ x \perp\!\!\!\perp z | y,\ y \perp\!\!\!\perp z | x \notin I(p)$

Directed Graphs	Violation
$x \quad y \quad z$	$x \perp\!\!\!\perp y \mid z$
$x \to y \quad z$	$x \not\perp\!\!\!\perp y$
$x \to y \to z$	$x \not\perp\!\!\!\perp y$
$x \to y,\ x \to z$	$y \not\perp\!\!\!\perp z$
$x \to y \to z,\ x \to z$	$x \not\perp\!\!\!\perp y$
Undirected	
$x \quad y \quad z$	$x \perp\!\!\!\perp y \mid z$
$x - y \quad z$	$x \perp\!\!\!\perp y$
$x - y - z$	$x \perp\!\!\!\perp y$
$x - y - z,\ x - z$	$x \perp\!\!\!\perp y$

No directed or undirected graph has $I(G) = I(p)$

1.5

b) $X \to Y \leftarrow Z$ here, $X \perp\!\!\!\perp Z$ but $X \not\perp\!\!\!\perp Z \mid Y$

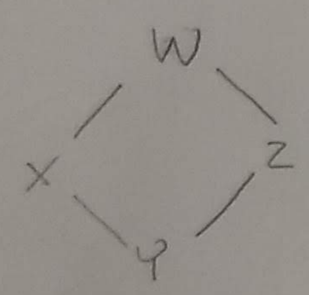
undirected possibilities:

$X \to Y \quad Z$ doesn't capture $X \not\perp\!\!\!\perp Z \mid Y$

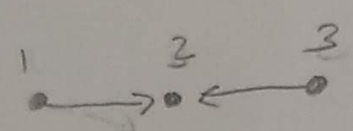
$X - Y - Z$ or $X \to Y \leftarrow Z$ doesn't capture $X \perp\!\!\!\perp Z \mid Y$

$X - Y - Z$ $\quad$ $X - Y - Z$ doesn't capture $X \perp\!\!\!\perp Z$

c)



some node in cycle <u>must</u> have immorality



In undirected graph $1 \perp\!\!\!\perp 3 \mid 2, 4$

In directed graph $1 \not\perp\!\!\!\perp 3 \mid 2, 4$

g) i) <u>No</u>. First, note if 2 undirected graphs have different vertices $\Rightarrow$ different distributions.

Second, try removing/adding edge $i-j$. In one graph, $x_i \perp\!\!\!\perp x_j \mid$ everything else but in the other $x_i \not\perp\!\!\!\perp x_j \mid$ everything else

The only way to bridge this difference is to ensure both undirected graphs have same edges

Same Nodes + Same Edges $\Rightarrow$ same graph

ii) Yes. $A \to B \to C$, $A \leftarrow B \to C$, $A \leftarrow B \leftarrow C$ all express the same (conditional) independencies

1.6

a) $x - t$

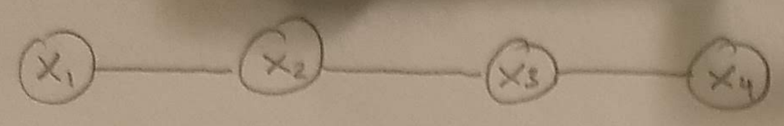
Define
$$\left.\begin{aligned} \Psi(x) &= 1 \\ \Psi(t) &= p(t) \\ \Psi(x,t) &= p(x|t) \end{aligned}\right\} \Rightarrow \Psi(x,t)\,\Psi(x)\,\Psi(t) = P(x,t)$$

b)
$$\begin{tikzcd} & y_1 \\ x - t \arrow[ur, dash] \arrow[r, dash] \arrow[dr, dash] & \vdots \\ & y_N \end{tikzcd}$$

Define
$$\left.\begin{aligned} \Psi(x) &= 1 \\ \Psi(t) &= p(t) \\ \Psi(x,t) &= p(x|t) \\ \Psi(y,t) &= \exp\left(\theta_t^T y\right) \\ \Psi(y) &= 1 \end{aligned}\right\} \Rightarrow \Psi(x,t)\,\Psi(x)\,\Psi(y)\,\Psi(t,y) = P(x,y,t)$$

1.7

a)



$$\Lambda = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & \varnothing & \varnothing & & \\ 2 & \varnothing & \varnothing & \varnothing & \\ 3 & & \varnothing & \varnothing & \varnothing \\ 4 & & & \varnothing & \varnothing \end{array}$$

$\varnothing$ means non-zero

(Conditional) Independencies

$1 \perp\!\!\!\perp 3$	$1 \perp\!\!\!\perp 4 \mid 2$	$2 \perp\!\!\!\perp 4 \mid 1$
$1 \perp\!\!\!\perp 4$	$1 \perp\!\!\!\perp 4 \mid 3$	
$2 \perp\!\!\!\perp 4$	$1 \perp\!\!\!\perp 3 \mid 4$	

Testing other conditional independencies. Let $\Lambda_{A,B|C}$ denote covariance of conditional distribution $p(x_A, x_B \mid x_C)$

$$\Lambda_{1,2|3} = \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & \varnothing & \varnothing \\ 2 & \varnothing & \varnothing \end{array} - \begin{array}{c|c} & 3 \\ \hline 1 & \\ 2 & \varnothing \end{array} \; {}_3[\varnothing]^{-1}{}_3 \begin{array}{cc} 1 & 2 \\ \hline & \varnothing \end{array} \Rightarrow \text{Not Diag} \Rightarrow 1 \not\perp\!\!\!\perp 2 \mid 3$$

$$\Lambda_{1,2|4} = \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & \varnothing & \varnothing \\ 2 & \varnothing & \varnothing \end{array} - \begin{array}{c|c} & 4 \\ \hline 1 & \\ 2 & \end{array} \; {}_4[\varnothing]^{-1} [\quad] \Rightarrow \text{Not Diag} \Rightarrow 1 \not\perp\!\!\!\perp 2 \mid 4$$

By symmetry, we also have $4 \not\perp\!\!\!\perp 3 \mid 2$ and $4 \not\perp\!\!\!\perp 3 \mid 1$

$$\Lambda_{2,3|1} = \begin{array}{c|cc} & 2 & 3 \\ \hline 2 & \varnothing & \varnothing \\ 3 & \varnothing & \varnothing \end{array} - \begin{array}{c|c} & 1 \\ \hline 2 & \varnothing \\ 3 & \end{array} \; {}_1[\varnothing]^{-1}{}_1 \begin{array}{cc} 2 & 3 \\ \hline \varnothing & \end{array} \Rightarrow \text{Not Diag} \Rightarrow 2 \not\perp\!\!\!\perp 3 \mid 1$$

By symmetry, we also have $2 \not\perp\!\!\!\perp 3 \mid 4$

$$\Lambda_{1,3|2} = \begin{array}{c|cc} & 1 & 3 \\ \hline 1 & \varnothing & \\ 3 & & \varnothing \end{array} - \begin{array}{c|c} & 2 \\ \hline 1 & \varnothing \\ 3 & \varnothing \end{array} \; {}_2[\varnothing]^{-1} [\varnothing \;\; \varnothing] \Rightarrow \text{Not Diag} \Rightarrow 1 \not\perp\!\!\!\perp 3 \mid 2$$

$$\Lambda_{1,4|2} = \begin{array}{c|cc} & 1 & 4 \\ \hline 1 & \varnothing & \\ 4 & & \varnothing \end{array} - \begin{array}{c|c} & 2 \\ \hline 1 & \varnothing \\ 4 & \end{array} \; {}_2[\varnothing]^{-1} [\varnothing \;\;] \Rightarrow \text{Diag} \Rightarrow 1 \perp\!\!\!\perp 4 \mid 2$$

Not shown: $1 \perp\!\!\!\perp 3 \mid 4$ & $2 \perp\!\!\!\perp 4 \mid 1$

By symmetry, we also have $2 \not\perp\!\!\!\perp 4 \mid 3$ and $1 \perp\!\!\!\perp 4 \mid 3$

$$\Lambda_{1,2|3,4} = \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & \varnothing & \varnothing \\ 2 & \varnothing & \varnothing \end{array} - \begin{array}{c|c} & 3 \;\; 4 \\ \hline 1 & \\ 2 & \varnothing \quad \end{array} \; {}_{3,4}\begin{bmatrix} \varnothing & \varnothing \\ \varnothing & \varnothing \end{bmatrix}^{-1}_{3,4} \begin{bmatrix} & \varnothing \\ & \end{bmatrix} \Rightarrow \text{Not Diag} \Rightarrow 1 \not\perp\!\!\!\perp 2 \mid 3, 4$$

$$\Rightarrow 3 \not\perp\!\!\!\perp 4 \mid 1, 2$$

$$\Lambda_{1,3|2,4} = \begin{array}{c|cc} & 1 & 3 \\ \hline 1 & \varnothing & \\ 3 & & \varnothing \end{array} - \begin{array}{c|cc} & 2 & 4 \\ \hline 1 & \varnothing & \\ 3 & \varnothing & \varnothing \end{array} \; {}_{2,4}\begin{bmatrix} \varnothing & \\ & \varnothing \end{bmatrix}^{-1}_{2,4} \begin{bmatrix} \varnothing & \varnothing \\ & \varnothing \end{bmatrix} \Rightarrow \text{Not Diag} \Rightarrow 1 \not\perp\!\!\!\perp 3 \mid 2, 4$$

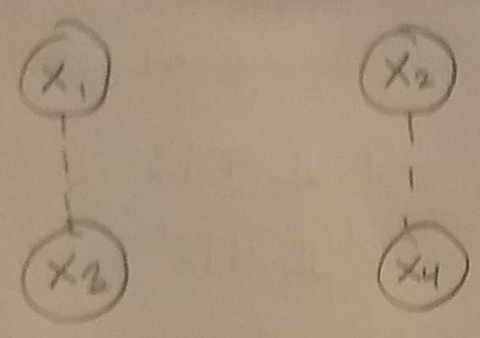
$$\Rightarrow 2 \not\perp\!\!\!\perp 4 \mid 1, 3$$

$$\Lambda_{1,4|2,3} = \begin{array}{c|cc} & 1 & 4 \\ \hline 1 & \varnothing & \\ 4 & & \varnothing \end{array} - \begin{array}{c|cc} & 2 & 3 \\ \hline 1 & \varnothing & \\ 4 & & \varnothing \end{array} \; {}_{2,3}\begin{bmatrix} \varnothing & \varnothing \\ \varnothing & \varnothing \end{bmatrix}^{-1}_{2,3} \begin{bmatrix} \varnothing & \\ & \varnothing \end{bmatrix} \Rightarrow \text{Not Diag} \Rightarrow 1 \not\perp\!\!\!\perp 4 \mid 2, 3$$

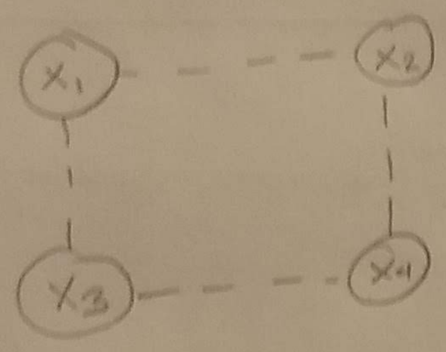
Not shown: $2 \not\perp\!\!\!\perp 3 \mid 1, 4$

1.7

b)



c)



1.8

a) $Y|X \sim N(X\beta, \sigma^2 I)$

i. $L(\beta) = \log p(Y|X, \beta, \sigma^2 I)$

$$= -\frac{1}{2} \log 2\pi |\sigma^2 I| - \frac{1}{2}(Y - X\beta)^T (\sigma^2 I)^{-1} (Y - X\beta)$$

ii. $\nabla_\beta L(\beta) = 0 \Rightarrow 0 = -\frac{1}{2} X^T (Y - X\beta)$

$$= X^T Y - X^T X \beta$$

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T Y}$$

The ML estimate is the same as the least-square estimate.

This shows minimizing MSE $\Leftrightarrow$ maximum likelihood w/ iid Gaussian noise

1.8

b) $L(\mu, \Lambda) = \prod_{i=1}^{n} p(x_i | \mu, \Lambda)$

$$\log L(\mu, \Lambda) = \sum_{i=1}^{n} \log \left(\frac{1}{\sqrt{2\pi |\Lambda|}} \exp\left(-\frac{1}{2} (x_i - \mu)^T \Lambda^{-1} (x_i - \mu) \right) \right)$$

$$= -\frac{n}{2} \log 2\pi |\Lambda| - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^T \Lambda^{-1} (x_i - \mu)$$

$$\nabla_\mu \log L(\mu, \Lambda) = 0 = -\sum_{i=1}^{n} \Lambda^{-1} (x_i - \mu)$$

$$n\mu = \sum_{i=1}^{n} x_i$$

$$\boxed{\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

$$\nabla_{\Lambda^{-1}} \log L(\mu, \Lambda) = 0 = \frac{n}{2} \nabla_{\Lambda^{-1}} \log |\Lambda^{-1}| - \frac{1}{2} \sum_{i=1}^{n} \nabla_{\Lambda^{-1}} \operatorname{Tr}\left[(x_i - \mu)(x_i - \mu)^T \Lambda^{-1} \right]$$

$$n\Lambda = \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T$$

$$\boxed{\hat{\Lambda} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T}$$

Computational

(a) i. $\beta_1 = 1$

$\beta_2 = 0$

$\beta_3 = 1$

β_1 & β_3 are most important

ii. $\beta_{2|\pi_2} = 0.8(1) + 0.8(1) = 1.6$

$\beta_{1|\pi_1} = 1$

$\beta_{3|\pi_3} = 1$

The results differ from a(i) b/c X_1 & X_3 sever any impact X_2 has on Y

(b) Model $\hat{Y} = X_1\beta_1 + X_2\beta_2 + X_3\beta_3$

$\Rightarrow \beta_1 = -1.49, \quad \beta_2 = 2.19, \quad \beta_3 = -1.02$

Model $\hat{Y} = X_2\beta_2$

$\Rightarrow \beta_2 = 0.65$

Model $\hat{Y} = X_1\beta_1 + X_2\beta_2$

$\Rightarrow \beta_1 = -1.49 \qquad \beta_2 = 1.69$

Model $\hat{Y} = X_2\beta_2 + X_3\beta_3$

$\Rightarrow \beta_2 = 1.15 \qquad \beta_3 = -1.03$

```
import pandas as pd
from sklearn.linear_model import LinearRegression


data = pd.read_csv('data.csv')


linear_model = LinearRegression(fit_intercept=False, copy_X=True)


columns_in_regressions = [
	['X_1', 'X_2', 'X_3'],
	['X_2'],
	['X_1', 'X_2'],
	['X_3', 'X_2']
]

for columns_in_regression in columns_in_regressions:
	linear_model.fit(data[columns_in_regression], data['Y'])
	print('Included variables: ', columns_in_regression)
	print('Parameter Estimates: ', linear_model.coef_, '\n')
```