Assignment 1

Yuki Joyama (yj2803)

Problem 1

(a) $X_2 \perp\!\!\!\perp X_3, X_7, X_8, X_9, X_{10}|X_1$
$X_3 \perp\!\!\!\perp X_2|X_1$
$X_4 \perp\!\!\!\perp X_1, X_7, X_8, X_{10}|X_2, X_3$
$X_5 \perp\!\!\!\perp X_1, X_2, X_3, X_7, X_8, X_9, X_{10}|X_4$
$X_6 \perp\!\!\!\perp X_1, X_2, X_3, X_4, X_7, X_8, X_9, X_{10}|X_5$
$X_7 \perp\!\!\!\perp X_1, X_2, X_4, X_5, X_6|X_3$
$X_8 \perp\!\!\!\perp X_1, X_2, X_4, X_5, X_6|X_3, X_7$
$X_9 \perp\!\!\!\perp X_1, X_2, X_3, X_5, X_6, X_7|X_4, X_8, X_{10}$
$X_{10} \perp\!\!\!\perp X_1, X_2, X_3, X_4, X_5, X_6, X_7|X_8$

(b) $X_2 \perp_d X_9|X_4$: False. X_4 is a collider and it is in the conditioning set so the path is active.
$X_7 \perp_d X_5|\{X_3, X_8\}$: True. X_4 is a collider and it is not in the conditioning set so the path is non-active. And X_9 is a collider and also not in the conditioning set so the path is non-active. There is no active path between X_7 and X_5.
$\{X_2, X_4\} \perp_d X_7|\{X_6, X_9, X_{10}\}$: False. X_9 is a collider and it is in the conditioning set so the path is active.

Problem 2

We want to know if X_i is d-connected to $X\backslash\{Mb(X_i, \mathcal{G}), X_i\}$ given $Mb(X_i, \mathcal{G})$. By definition of Markov blanket wrt $\mathcal{G}$, $Mb(X_i, \mathcal{G}) \equiv Pa(X_i) \cup Ch(X_i) \cup Pa(Ch(X_i))$. Any path from X_i to $X\backslash\{Mb(X_i, \mathcal{G}), X_i\}$ go through $Pa(X_i)$, $Ch(X_i)$ (where there is no $Pa(Ch(X_i))$), or $Ch(X_i)$ and $Pa(Ch(X_i))$. In the first case, $Pa(X_i)$ cannot be a collider and it is in the conditioning set so the path is not active. In the second case, $Ch(X_i)$ cannot be a collider and it is in the conditioning set so the path is not active. In the third case, $Ch(X_i)$ is a collider and it is in the conditioning set so the path is active. However, $Pa(Ch(X_i))$ cannot be a collider and it is in the conditioning set so this path is not active. Therefore, X_i cannot be d-connected to $X\backslash\{Mb(X_i, \mathcal{G}), X_i\}$ and $X_i \perp_d X\backslash\{Mb(X_i, \mathcal{G}), X_i\}|Mb(X_i, \mathcal{G})$ holds. By global Markov property, $X_i \perp_d X\backslash\{Mb(X_i, \mathcal{G}), X_i\}|Mb(X_i, \mathcal{G}) \Longrightarrow X_i \perp\!\!\!\perp X\backslash\{Mb(X_i, \mathcal{G}), X_i\}|Mb(X_i, \mathcal{G})$.

Problem 3

Agreements: $D \perp\!\!\!\perp \{A, C\}|B$
Disagreements: $A \perp\!\!\!\perp C|\phi$ (a), $C \perp\!\!\!\perp \{A, D\}|B$ (b)

Problem 4

(a) Unshielded colliders in Figure 3: $C \rightarrow E \leftarrow A$, $C \rightarrow B \leftarrow A$, $D \rightarrow E \leftarrow B$
No DAGs in the option matches these three structures.

(b) An unshielded collider in a DAG will violate the Markov equivalence with respect to the chain DAG. Taken this into account, we can consider two cases that will be Markov equivalent to the chain DAG:
1.Flip all the arrows to left
2.For $X_i (i = 2, 3, ..., p - 1)$, flip all the arrows to left before X_i
There are p-2 possible X_is that can be the pivot of the arrows. Therefore, $p - 2 + 1 = p - 1$ DAGs are Markov equivalent to the chain DAG.

Problem 5

(a)

```
## $paths
##  [1] "C -> B -> A <- E -> F -> G -> H" "C -> B -> A <- E -> F -> H"
##  [3] "C -> B <- G -> H"                "C -> B <- G <- F -> H"
##  [5] "C -> E -> A <- B <- G -> H"      "C -> E -> A <- B <- G <- F -> H"
##  [7] "C -> E -> F -> G -> H"           "C -> E -> F -> H"
##  [9] "C -> F -> G -> H"                "C -> F -> H"
## [11] "C -> F <- E -> A <- B <- G -> H" "C -> H"
##
## $open
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
```

(b)

```
## E and G are not d-separated given A and B.
```

(c)

```
## A _||_ C | B, E
## A _||_ D | C, E
## A _||_ D | B, E
## A _||_ F | C, E, G
## A _||_ F | B, E
## A _||_ G | B, C, F
## A _||_ G | B, E
## A _||_ H | C, F, G
## A _||_ H | C, E, G
## A _||_ H | B, C, F
## A _||_ H | B, E
## B _||_ D | C, E
## B _||_ D | C, F
## B _||_ D | C, G
## B _||_ E | C, F
## B _||_ E | C, G
## B _||_ F | C, G
## B _||_ H | C, G
## C _||_ D
## C _||_ G | F
## D _||_ F | C, E
## D _||_ G | F
## D _||_ G | C, E
```

```
## D _||_ H | C, F
## D _||_ H | C, E
## E _||_ G | F
## E _||_ H | C, F
```

With the option “type = all.pairs”:

```
## A _||_ C | B, E
## A _||_ C | B, D, E
## A _||_ C | B, E, F
## A _||_ C | B, D, E, F
## A _||_ C | B, E, G
## A _||_ C | B, D, E, G
## A _||_ C | B, E, F, G
## A _||_ C | B, D, E, F, G
## A _||_ C | B, E, H
## A _||_ C | B, D, E, H
## A _||_ C | B, E, F, H
## A _||_ C | B, D, E, F, H
## A _||_ C | B, E, G, H
## A _||_ C | B, D, E, G, H
## A _||_ C | B, E, F, G, H
## A _||_ C | B, D, E, F, G, H
## A _||_ D | B, E
## A _||_ D | C, E
## A _||_ D | B, C, E
## A _||_ D | B, E, F
## A _||_ D | C, E, F
## A _||_ D | B, C, E, F
## A _||_ D | B, E, G
## A _||_ D | C, E, G
## A _||_ D | B, C, E, G
## A _||_ D | B, E, F, G
## A _||_ D | C, E, F, G
## A _||_ D | B, C, E, F, G
## A _||_ D | B, E, H
## A _||_ D | C, E, H
## A _||_ D | B, C, E, H
## A _||_ D | B, E, F, H
## A _||_ D | C, E, F, H
## A _||_ D | B, C, E, F, H
## A _||_ D | B, E, G, H
## A _||_ D | C, E, G, H
## A _||_ D | B, C, E, G, H
## A _||_ D | B, E, F, G, H
## A _||_ D | C, E, F, G, H
## A _||_ D | B, C, E, F, G, H
## A _||_ F | B, E
## A _||_ F | B, C, E
## A _||_ F | B, D, E
## A _||_ F | B, C, D, E
## A _||_ F | B, E, G
## A _||_ F | C, E, G
## A _||_ F | B, C, E, G
```

```
## A _||_ F | B, D, E, G
## A _||_ F | C, D, E, G
## A _||_ F | B, C, D, E, G
## A _||_ F | B, E, H
## A _||_ F | B, C, E, H
## A _||_ F | B, D, E, H
## A _||_ F | B, C, D, E, H
## A _||_ F | B, E, G, H
## A _||_ F | C, E, G, H
## A _||_ F | B, C, E, G, H
## A _||_ F | B, D, E, G, H
## A _||_ F | C, D, E, G, H
## A _||_ F | B, C, D, E, G, H
## A _||_ G | B, E
## A _||_ G | B, C, E
## A _||_ G | B, D, E
## A _||_ G | B, C, D, E
## A _||_ G | B, C, F
## A _||_ G | B, C, D, F
## A _||_ G | B, E, F
## A _||_ G | B, C, E, F
## A _||_ G | B, D, E, F
## A _||_ G | B, C, D, E, F
## A _||_ G | B, E, H
## A _||_ G | B, C, E, H
## A _||_ G | B, D, E, H
## A _||_ G | B, C, D, E, H
## A _||_ G | B, C, F, H
## A _||_ G | B, C, D, F, H
## A _||_ G | B, E, F, H
## A _||_ G | B, C, E, F, H
## A _||_ G | B, D, E, F, H
## A _||_ G | B, C, D, E, F, H
## A _||_ H | B, E
## A _||_ H | B, C, E
## A _||_ H | B, D, E
## A _||_ H | B, C, D, E
## A _||_ H | B, C, F
## A _||_ H | B, C, D, F
## A _||_ H | B, E, F
## A _||_ H | B, C, E, F
## A _||_ H | B, D, E, F
## A _||_ H | B, C, D, E, F
## A _||_ H | B, E, G
## A _||_ H | C, E, G
## A _||_ H | B, C, E, G
## A _||_ H | B, D, E, G
## A _||_ H | C, D, E, G
## A _||_ H | B, C, D, E, G
## A _||_ H | C, F, G
## A _||_ H | B, C, F, G
## A _||_ H | C, D, F, G
## A _||_ H | B, C, D, F, G
## A _||_ H | B, E, F, G
```

```
## A _||_ H | C, E, F, G
## A _||_ H | B, C, E, F, G
## A _||_ H | B, D, E, F, G
## A _||_ H | C, D, E, F, G
## A _||_ H | B, C, D, E, F, G
## B _||_ D | C, E
## B _||_ D | A, C, E
## B _||_ D | C, F
## B _||_ D | C, E, F
## B _||_ D | A, C, E, F
## B _||_ D | C, G
## B _||_ D | C, E, G
## B _||_ D | A, C, E, G
## B _||_ D | C, F, G
## B _||_ D | C, E, F, G
## B _||_ D | A, C, E, F, G
## B _||_ D | C, E, H
## B _||_ D | A, C, E, H
## B _||_ D | C, F, H
## B _||_ D | C, E, F, H
## B _||_ D | A, C, E, F, H
## B _||_ D | C, G, H
## B _||_ D | C, E, G, H
## B _||_ D | A, C, E, G, H
## B _||_ D | C, F, G, H
## B _||_ D | C, E, F, G, H
## B _||_ D | A, C, E, F, G, H
## B _||_ E | C, F
## B _||_ E | C, D, F
## B _||_ E | C, G
## B _||_ E | C, D, G
## B _||_ E | C, F, G
## B _||_ E | C, D, F, G
## B _||_ E | C, F, H
## B _||_ E | C, D, F, H
## B _||_ E | C, G, H
## B _||_ E | C, D, G, H
## B _||_ E | C, F, G, H
## B _||_ E | C, D, F, G, H
## B _||_ F | C, G
## B _||_ F | C, D, G
## B _||_ F | C, E, G
## B _||_ F | A, C, E, G
## B _||_ F | C, D, E, G
## B _||_ F | A, C, D, E, G
## B _||_ F | C, G, H
## B _||_ F | C, D, G, H
## B _||_ F | C, E, G, H
## B _||_ F | A, C, E, G, H
## B _||_ F | C, D, E, G, H
## B _||_ F | A, C, D, E, G, H
## B _||_ H | C, G
## B _||_ H | C, D, G
## B _||_ H | C, E, G
```

```
## B _||_ H | A, C, E, G
## B _||_ H | C, D, E, G
## B _||_ H | A, C, D, E, G
## B _||_ H | C, F, G
## B _||_ H | A, C, F, G
## B _||_ H | C, D, F, G
## B _||_ H | A, C, D, F, G
## B _||_ H | C, E, F, G
## B _||_ H | A, C, E, F, G
## B _||_ H | C, D, E, F, G
## B _||_ H | A, C, D, E, F, G
## C _||_ D
## C _||_ G | F
## C _||_ G | D, F
## C _||_ G | E, F
## C _||_ G | D, E, F
## D _||_ F | C, E
## D _||_ F | A, C, E
## D _||_ F | B, C, E
## D _||_ F | A, B, C, E
## D _||_ F | C, E, G
## D _||_ F | A, C, E, G
## D _||_ F | B, C, E, G
## D _||_ F | A, B, C, E, G
## D _||_ F | C, E, H
## D _||_ F | A, C, E, H
## D _||_ F | B, C, E, H
## D _||_ F | A, B, C, E, H
## D _||_ F | C, E, G, H
## D _||_ F | A, C, E, G, H
## D _||_ F | B, C, E, G, H
## D _||_ F | A, B, C, E, G, H
## D _||_ G | C, E
## D _||_ G | A, C, E
## D _||_ G | B, C, E
## D _||_ G | A, B, C, E
## D _||_ G | F
## D _||_ G | C, F
## D _||_ G | B, C, F
## D _||_ G | A, B, C, F
## D _||_ G | E, F
## D _||_ G | C, E, F
## D _||_ G | A, C, E, F
## D _||_ G | B, C, E, F
## D _||_ G | A, B, C, E, F
## D _||_ G | C, E, H
## D _||_ G | A, C, E, H
## D _||_ G | B, C, E, H
## D _||_ G | A, B, C, E, H
## D _||_ G | C, F, H
## D _||_ G | B, C, F, H
## D _||_ G | A, B, C, F, H
## D _||_ G | C, E, F, H
## D _||_ G | A, C, E, F, H
```

```
## D _||_ G | B, C, E, F, H
## D _||_ G | A, B, C, E, F, H
## D _||_ H | C, E
## D _||_ H | A, C, E
## D _||_ H | B, C, E
## D _||_ H | A, B, C, E
## D _||_ H | C, F
## D _||_ H | B, C, F
## D _||_ H | A, B, C, F
## D _||_ H | C, E, F
## D _||_ H | A, C, E, F
## D _||_ H | B, C, E, F
## D _||_ H | A, B, C, E, F
## D _||_ H | C, E, G
## D _||_ H | A, C, E, G
## D _||_ H | B, C, E, G
## D _||_ H | A, B, C, E, G
## D _||_ H | C, F, G
## D _||_ H | A, C, F, G
## D _||_ H | B, C, F, G
## D _||_ H | A, B, C, F, G
## D _||_ H | C, E, F, G
## D _||_ H | A, C, E, F, G
## D _||_ H | B, C, E, F, G
## D _||_ H | A, B, C, E, F, G
## E _||_ G | F
## E _||_ G | C, F
## E _||_ G | B, C, F
## E _||_ G | A, B, C, F
## E _||_ G | D, F
## E _||_ G | C, D, F
## E _||_ G | B, C, D, F
## E _||_ G | A, B, C, D, F
## E _||_ G | C, F, H
## E _||_ G | B, C, F, H
## E _||_ G | A, B, C, F, H
## E _||_ G | C, D, F, H
## E _||_ G | B, C, D, F, H
## E _||_ G | A, B, C, D, F, H
## E _||_ H | C, F
## E _||_ H | B, C, F
## E _||_ H | A, B, C, F
## E _||_ H | C, D, F
## E _||_ H | B, C, D, F
## E _||_ H | A, B, C, D, F
## E _||_ H | C, F, G
## E _||_ H | A, C, F, G
## E _||_ H | B, C, F, G
## E _||_ H | A, B, C, F, G
## E _||_ H | C, D, F, G
## E _||_ H | A, C, D, F, G
## E _||_ H | B, C, D, F, G
## E _||_ H | A, B, C, D, F, G
```

According to the documentation, the default for this function is type = “missing.edge”. This returns a list of conditional independencies with minimal testable implication per missing edge while type = “all.pairs” returns all implied conditional independencies between two variables. This is why the first one is shorter than the second one.

(d) The summary of the simulated data (N=10000) is shown below.

```
##        A                   B                   C                   D            
##  Min.   :-3.689914   Min.   :-3.66834   Min.   :-3.715549   Min.   :-3.397235  
##  1st Qu.:-0.684691   1st Qu.:-0.66291   1st Qu.:-0.678795   1st Qu.:-0.682320  
##  Median :-0.011717   Median : 0.02416   Median :-0.010833   Median :-0.004992  
##  Mean   :-0.005188   Mean   : 0.01402   Mean   :-0.002295   Mean   :-0.014104  
##  3rd Qu.: 0.664292   3rd Qu.: 0.68337   3rd Qu.: 0.674964   3rd Qu.: 0.656542  
##  Max.   : 3.359318   Max.   : 3.78853   Max.   : 3.388888   Max.   : 3.638291  
##        E                   F                   G            
##  Min.   :-4.148542   Min.   :-3.733913   Min.   :-3.489782  
##  1st Qu.:-0.672298   1st Qu.:-0.672419   1st Qu.:-0.671242  
##  Median : 0.023623   Median : 0.002315   Median : 0.016195  
##  Mean   : 0.004979   Mean   : 0.001992   Mean   : 0.005013  
##  3rd Qu.: 0.670245   3rd Qu.: 0.681429   3rd Qu.: 0.684476  
##  Max.   : 4.160425   Max.   : 3.991859   Max.   : 3.415378  
##        H            
##  Min.   :-3.482478  
##  1st Qu.:-0.655090  
##  Median : 0.005633  
##  Mean   : 0.006265  
##  3rd Qu.: 0.660801  
##  Max.   : 3.735608
```

Markov blanket for vertex B are:

```
## [1] "C" "G" "A" "E"
```

Let’s check the linear regression of $B \sim Mb(B, \mathcal{G})$ + remaining covariates using this simulated data.

```
## 
## Call:
## lm(formula = B ~ A + C + D + E + F + G + H, data = sim)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -2.5674 -0.5001  0.0082  0.5023  3.0091 
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  0.0127456  0.0074923   1.701   0.0889 .  
## A            0.3360154  0.0079161  42.447  < 2e-16 ***
## C           -0.2107823  0.0081714 -25.795  < 2e-16 ***
## D            0.0008288  0.0078329   0.106   0.9157    
## E            0.0404660  0.0083848   4.826 1.41e-06 ***
## F            0.0028081  0.0104751   0.268   0.7887    
## G            0.4678821  0.0107319  43.597  < 2e-16 ***
```

```
## H           -0.0003218  0.0093214  -0.035   0.9725
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7491 on 9992 degrees of freedom
## Multiple R-squared:  0.4435, Adjusted R-squared:  0.4431
## F-statistic:  1138 on 7 and 9992 DF,  p-value: < 2.2e-16
```

The coefficients for variables outside of the Markov blanket (D, F, H) have p-values greater than 0.05, indicating their independence from B. Meanwhile, the coefficients for variables within the Markov blanket have p-values less than 0.05, confirming that the Markov blanket property holds for vertex B.

Extra Credit

I will examine if X_i is d-connected to X_j given $Pa(X_i, \mathcal{G}) \cup Pa(X_j, \mathcal{G})$.
Any path from X_i to X_j must go through at least one vertex because X_i and X_j are not adjacent to each other. When $X_i \leftarrow ... \leftarrow X_j$, the vertex to the right of X_i is a parent of X_i. This cannot be a collider and it is in the conditioning set so the path is not active. When $X_i \rightarrow ... \leftarrow X_j$, no vertex between X_i and X_j can be $Pa(X_i, \mathcal{G})$ or $Pa(X_j, \mathcal{G})$ and at least one vertex will be a collider. This collider is not in the conditioning set so the path is not active. When $X_i \rightarrow ... \rightarrow X_j$, the vertex to the left of X_j is a parent of X_j. This cannot be a collider and it is in the conditioning set so the path is not active. The path between X_i and X_j is not active in all the three cases, so we can say that in any DAG with X_i not adjacent to X_j, necessarily $X_i \perp_d X_j | Pa(X_i, \mathcal{G}) \cup Pa(X_j, \mathcal{G})$ holds.

Code

```r
# problem 5
library(dagitty)

# construct fig 5 DAG
g <- dagitty('dag {
    D [pos="0,0"]
    E [pos="1,0"]
    C [pos="1,-1"]
    A [pos="2,0"]
    B [pos="3,0"]
    F [pos="4,0"]
    G [pos="4,-1"]
    H [pos="5,-1"]

    D -> E -> A <- B <- G -> H
    C -> E -> F -> G
    C -> H
    C -> B
    C -> F -> H
}')

# a: path from C to H
paths(g, "C", "H")
```

```r
# b: d-separation between E and G given A and B
if(dseparated(g, "E", "G", c("A", "B"))){
  message("E"," and ", "G"," are d-separated given A and B.")
} else {
  message("E"," and ", "G", " are not d-separated given A and B.")
}

# c: list the conditional independencies relationships implied by the model
impliedConditionalIndependencies(g)
impliedConditionalIndependencies(g, type = "all.pairs")

set.seed(2024)
# d: simulate data from this DAG, which associates the DAG with a linear structural equation model
# path coefficient (-0.7, 0.7), sample size = 10000
sim <- simulateSEM(
  g,
  b.default = NULL,
  b.lower = -0.7,
  b.upper = 0.7,
  N = 10000
)

summary(sim)

markovBlanket(g, 'B')

# construct a linear model
lm_b = lm(B ~ A + C + D + E + F + G + H, sim)
summary(lm_b)
```