Stat Learning HW3

Yaniv Bronshtein

11/17/2021

Import necessary libraries

```
library(gRbase)
library(gRain)
library(gRim)
library(ggm)
library(bnlearn)
```

```
##
## Attaching package: 'bnlearn'

## The following objects are masked from 'package:gRbase':
##
##     ancestors, children, parents
```

```
library(glasso)
```

Function to determine if DAG is acyclic

```
is.acyclic <- function(A){
  ## A is an adjacency matrix
  ## i.e. a_ij = 1 if the edge i->j *is* in the graph
  ##      a_ij = 0 if the edge i->j *is not* in the graph
  if (nrow(A) == 1)
    return(TRUE)
  ## compute transitive closure
  H <- A
  diag(H) <- 1
  repeat {
    HH <- sign(H %*% H)
    if (all(HH == H))
      break
    else H <- HH
  }
  diag(H) <- 0
  ## h_ij =1 if there is a directed path from i to j
  l <- H[lower.tri(H)]
  u <- t(H)[lower.tri(t(H))]
  com <- (l & u)
  all(!com)
}
```

Question 2

2. This problem uses a dataset gene.txt containing expression measurements of 9 genes. The names of the genes are included in the dataset. It also uses a R function is.acyclic(). Here is how you will use this function to tell whether a directed graph is acyclic or not. Represent a directed graph G as an adjacency matrix A: if there is an edge from node i to node j, then A[i, j] = 1, otherwise A[i, j] = 0. Note that the adjacency matrix A is not symmetric for a directed graph. The function is.acyclic() takes the adjacency matrix A as the input and returns whether it is acyclic or not. Both the dataset and the R function can be downloaded from the Homework folder. The R file also contains some examples on how to use the is.acyclic() function.

Extract the subset dataframe

```
df <- read.table('../data/gene.txt', header = TRUE)
df_subset <- df[,c(3,6,4,1)]
```

(a) Write your own code try all possible DAGs on the four genes GAL1, GAL2, GAL3, GAL7. There are 6 pairs of nodes, and between each pair of nodes, there can be three options: no edge, or an edge with either direction. Therefore, there are in total 36 = 729 directed graphs over 4 nodes. For each of them, check whether the graph is acyclic using the is.acyclic() function. If it is indeed a DAG, estimated the model parameters, and report the BIC.

```
bases <- c('1','2','3')
all <- expand.grid(rep(list(bases), 6))

allAMat <- vector(mode='list',length = nrow(all))

BIC_scores <- NULL
dag <- empty.graph(names(df_subset))

genes <- c('GAL1','GAL2','GAL3','GAL7')
for (i in (1:nrow(all))){
  A <- matrix(0, 4, 4)

  A[1,2] <- +(all$Var1[i] == '2')
  A[2,1] <- +(all$Var1[i] == '3')
  A[1,3] <- +(all$Var2[i] == '2')
  A[3,1] <- +(all$Var2[i] == '3')
  A[1,4] <- +(all$Var3[i] == '2')
  A[4,1] <- +(all$Var3[i] == '3')
  A[2,3] <- +(all$Var4[i] == '2')
  A[3,2] <- +(all$Var4[i] == '3')
  A[2,4] <- +(all$Var5[i] == '2')
  A[4,2] <- +(all$Var5[i] == '3')
  A[3,4] <- +(all$Var6[i] == '2')
  A[4,3] <- +(all$Var6[i] == '3')
  rownames(A) <- genes
  colnames(A) <- genes
  allAMat[[i]] <- A
  if (is.acyclic(allAMat[[i]])){
     amat(dag) <- allAMat[[i]]
     BIC_val <- score(dag, df_subset)
     BIC_scores <- c(BIC_scores, BIC_val)
  }

}
```

(b) Report the models with three smallest BIC (there may be multiple models with the same BIC, report all of them). Plot the corresponding DAGs. [You can try to use the R code at the end of the slides to plot the DAGs. Or since there are only 4 nodes, you can draw by hand.]

```
bic_report <- data.frame(cbind(BIC_scores, Graph = seq(length(allAMat))))
```

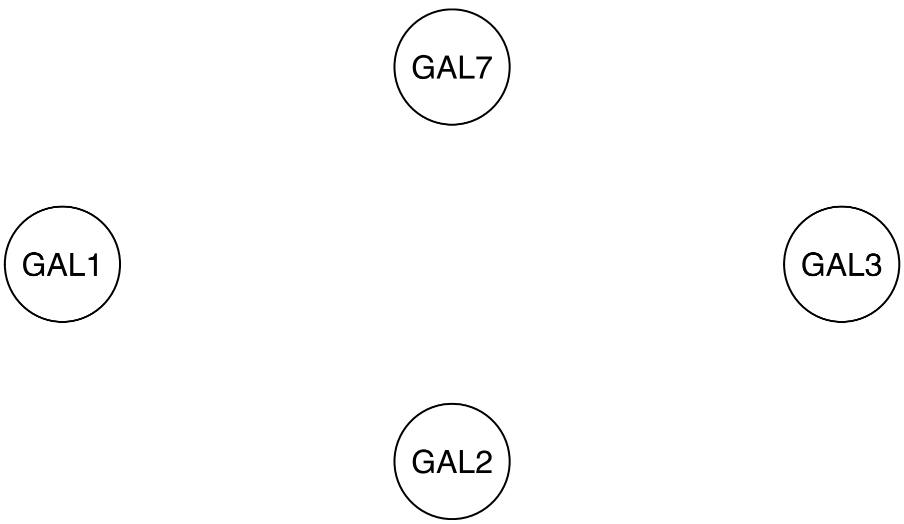
```
## Warning in cbind(BIC_scores, Graph = seq(length(allAMat))): number of rows of
## result is not a multiple of vector length (arg 1)
```

```
bic_report <- bic_report[order(bic_report[,1]),]
head(bic_report)
```

```
##     BIC_scores Graph
## 1    -970.1622     1
## 544  -970.1622   544
## 4    -961.3429     4
## 547  -961.3429   547
## 7    -961.3429     7
## 550  -961.3429   550
```

Note: There is a tie which is why 4 plots were displayed Plot1

```
amat(dag) <- allAMat[[bic_report[1,2]]]
plot(dag)
```

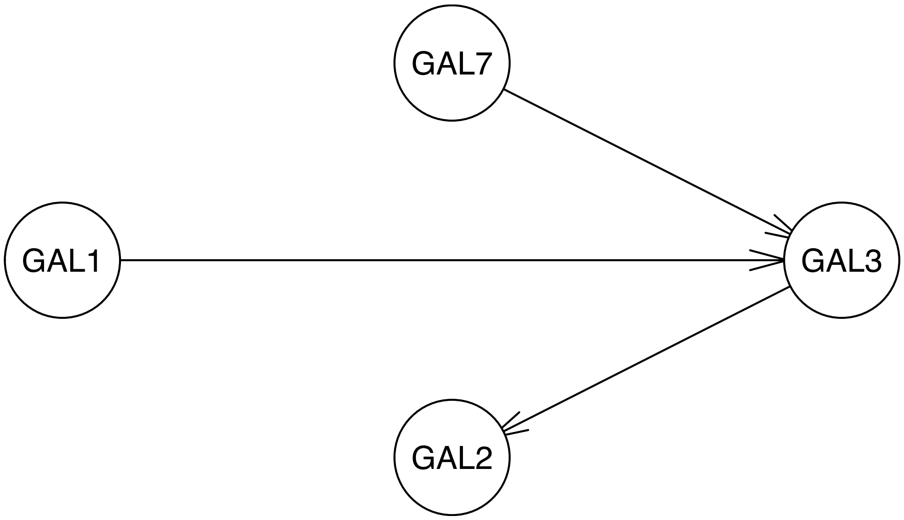



```
cat('BIC:', bic_report[1,1],'\n')
```

```
## BIC: -970.1622
```

Plot2

```
amat(dag) <- allAMat[[bic_report[2,2]]]
plot(dag)
```




```
cat('BIC:', bic_report[2,1],'\n')
```

```
## BIC: -970.1622
```

Plot3

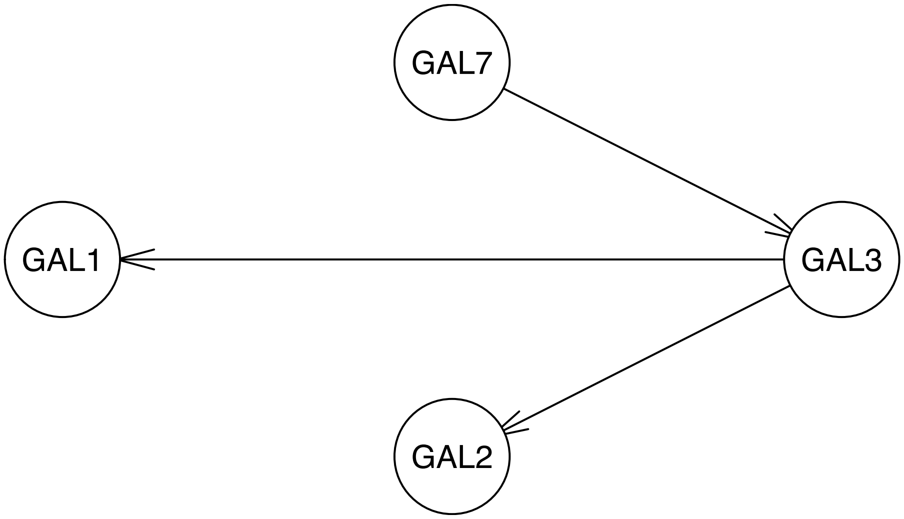
```
amat(dag) <- allAMat[[bic_report[3,2]]]
plot(dag)
```


```
cat('BIC:', bic_report[3,1],'\n')
```

```
## BIC: -961.3429
```

Plot4

```
amat(dag) <- allAMat[[bic_report[4,2]]]
plot(dag)
```




```
cat('BIC:', bic_report[4,1],'\n')
```

```
## BIC: -961.3429
```

Question 3

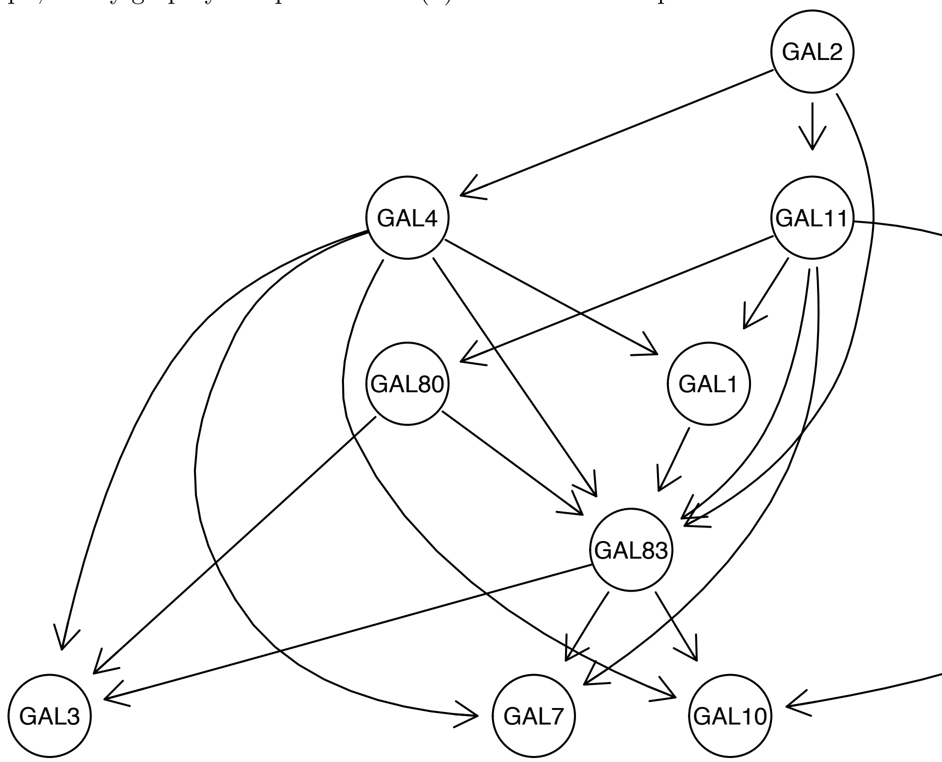
(a) Write you own code to implement the step-wise algorithm for estimating the graph structure. Write a generic code which takes a data matrix of any size as the input (so that it won't take any additional effort if you want to use it for the in-class workshop).

```
# Function to implement stepwise algorithm
stepwise_alg <- function(df) {
  cols <- colnames(df)
  n <- length(cols)
  A <- matrix(0, n, n)
  rownames(A) <- cols
  colnames(A) <- cols

  dag <- empty.graph(cols)
  amat(dag) <- A
  min_bic <- score(dag, df)

  smaller_BIC_exists = TRUE
  while(smaller_BIC_exists) {
    bic_Vals <- list()
    graphs <- list()

    edge_idx <- which(A==1, arr.ind=TRUE)

    len_edge_idx <- length(edge_idx)
    if (len_edge_idx > 0) {
      for (i in 1:(len_edge_idx/2)) {
        currA <- A
        currA[edge_idx[i,1], edge_idx[i,2]] <- 0

          dag <- empty.graph(cols)
          amat(dag) <- currA
```

```
        curr_bic <- score(dag, df)
      bic_Vals <- c(bic_Vals, curr_bic)
      graphs[[length(graphs)+1]] <- currA
    }
  }

  no_edge_idx <- which(A==0, arr.ind=TRUE)
  len_no_edge_idx <- length(no_edge_idx)
  if (len_no_edge_idx > 0) {
    for (i in 1:(len_no_edge_idx/2)) {
      if(no_edge_idx[i,1] != no_edge_idx[i,2]) {
        currA <- A
        currA[no_edge_idx[i,1], no_edge_idx[i,2]] <- 1

        if (is.acyclic(currA)) {
          dag <- empty.graph(names(df))
            amat(dag) <- currA
            curr_bic <- score(dag, df)
          bic_Vals <- c(bic_Vals, curr_bic)
          graphs[[length(graphs)+1]] <- currA
        }
      }
    }
  }

  if (len_edge_idx > 0) {

    for (i in 1:(len_edge_idx/2)) {
      currA <- A
      currA[edge_idx[i,1], edge_idx[i,2]] <- 0
      currA[edge_idx[i,2], edge_idx[i,1]] <- 1

      if (is.acyclic(currA)) {
        dag <- empty.graph(names(df))
          amat(dag) <- currA
          curr_bic <- score(dag, df)
        bic_Vals <- c(bic_Vals, curr_bic)
        graphs[[length(graphs)+1]] <- currA
      }
    }
  }
  unlisted_bic_Vals <- unlist(bic_Vals)
  this_min_bic <- min(unlisted_bic_Vals)
  this_min_graph <- which.min(unlisted_bic_Vals)

  if (this_min_bic < min_bic) {
    min_bic <- this_min_bic
    A <- graphs[[this_min_graph]]
  }
  else {
    smaller_BIC_exists <- FALSE
  }
}
```

```
  graph_and_bic <- list(A, min_bic)
  return(graph_and_bic)
}
```

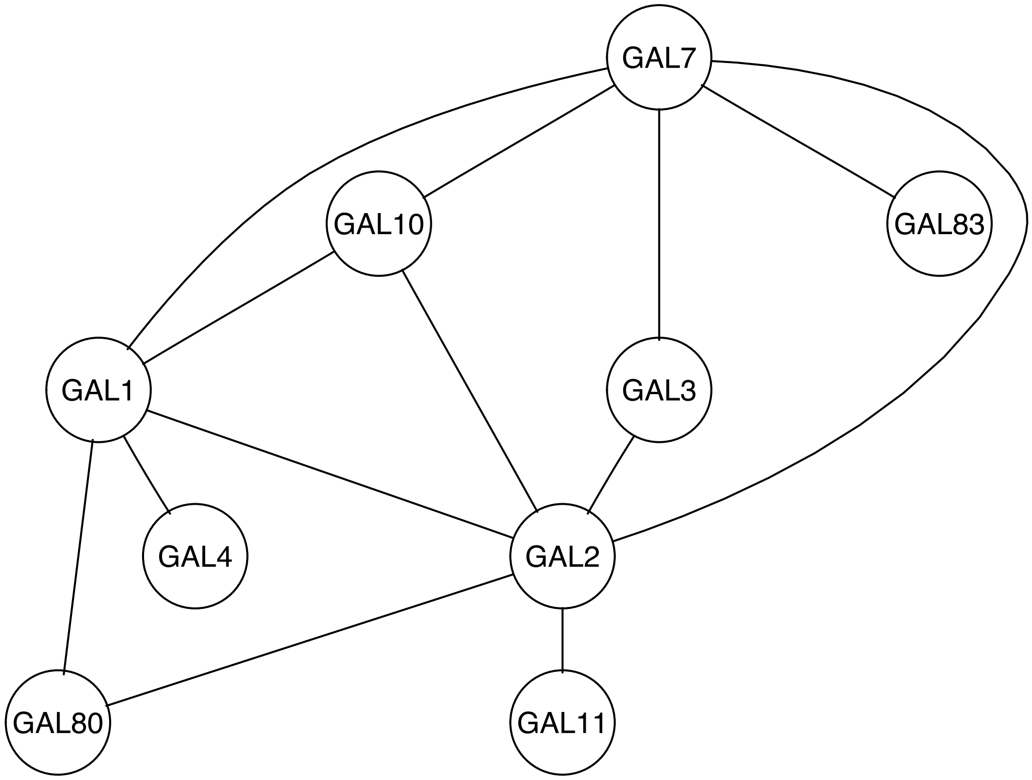
b) Try your program on the full gene.txt dataset. You will need an initial graph to start your algorithm, which can be the empty graph, full graph, or any graph you report in Part (b) of Problem 2. Report the DAG you found with the smallest BIC.



Question 5

(a) Use the graphical lasso to estimate the undirected graph for the gene.txt data. Try different choices of the tuning parameter lambda. Report and plot the undirected graphs you found.

```
lasso_model <- glasso(cov(df),rho=.05)

theta <- lasso_model$wi
cols <- names(df)
colnames(theta) <- cols
rownames(theta) <- cols
adj <- (theta != 0)
adj <- adj*1
diag(adj) <- 0
g1 <- as(adj, "graphNEL")

plot(g1)
```

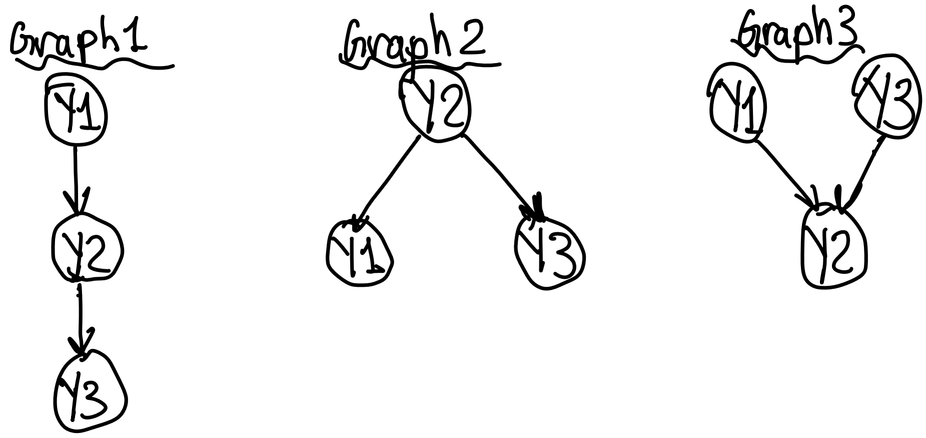



```r
cat('Rho:', .05,'\n')
```

```
## Rho: 0.05
```

Analysis: The greater the value of the constant Rho, the more disconnected the graph becomes

1. This problem considers the 3 DAGs talked about during lecture.



a). Write down the regression representation of the Gaussian graphical model for each of the given graphs

In the lecture, we used the factorized joint distribution:

$$p(x) = \prod_{k=1}^{K} p(x_k \mid pa_k)$$

The conditional distribution of x_k given its parents pa_k:

$$x_k \mid pa_k \sim N\left(\sum_{j \in pa_k} w_{kj} x_j + b_k, v_k\right), \quad 1 \leq k \leq K$$

And we rewrote as:

$$x_k = \sum_{j \in pa_k} w_{kj} x_j + b_k + \sqrt{v_k}\,\epsilon_k$$

where $\epsilon_1, \ldots, \epsilon_K$ are i.i.d $N(0,1)$

Thus, for Graph 1:

$$Y_1 = b_1 + \sqrt{V_1}\,\epsilon_1$$
$$Y_2 = W_{21} Y_1 + b_2 + \sqrt{V_2}\,\epsilon_2$$
$$Y_3 = W_{32} Y_2 + b_3 + \sqrt{V_3}\,\epsilon_3$$

For Graph2:

$$Y_1 = W_{12} Y_2 + b_1 + \sqrt{V_1}\,\epsilon_1$$
$$Y_2 = b_2 + \sqrt{V_2}\,\epsilon_2$$
$$Y_3 = W_{32} Y_2 + b_3 + \sqrt{V_3}\,\epsilon_3$$

For Graph3:

$$Y_1 = b_1 + \sqrt{V_1}\,\epsilon_1$$
$$Y_2 = W_{21} Y_1 + W_{23} Y_3 + b_2 + \sqrt{V_2}\,\epsilon_2$$
$$Y_3 = b_3 + \sqrt{V_3}\,\epsilon_3$$

b) Write down the covariance matrix of the joint distribution corresponding to each of these three distributions

set Bias = 0:

For Graph 1:

$$Y_1 = b_1 + \sqrt{V_1}\,\epsilon_1$$
$$Y_2 = W_{21} Y_1 + b_2 + \sqrt{V_2}\,\epsilon_2$$
$$Y_3 = W_{32} Y_2 + b_3 + \sqrt{V_3}\,\epsilon_3$$
$$\Downarrow$$

$$Y_1 = \sqrt{V_1}\,\epsilon_1$$

$$-w_2 Y_1 + Y_2 = \sqrt{V_2}\,\epsilon_2$$

$$-w_{32} Y_2 + Y_3 = \sqrt{V_3}\,\epsilon_3$$

In Matrix Form:

$$\begin{bmatrix} 1 & 0 & 0 \\ -w_{21} & 1 & 0 \\ 0 & -w_{32}Y_2 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} \sqrt{V_1}\,\epsilon_1 \\ \sqrt{V_2}\,\epsilon_2 \\ \sqrt{V_3}\,\epsilon_3 \end{bmatrix}$$

$$\Downarrow \qquad \Downarrow \qquad \Downarrow$$

$$A * Y = \epsilon$$

$$Var(cX) = c^2 Var(X)$$

$$Y = A^{-1}\epsilon$$

$$cov(Y) = A^{-1} cov(\epsilon) (A^{-1})^T$$

$$cov(Y) = cov\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ w_{21} & 1 & 0 \\ (w_{21}w_{32}) & w_{32} & 1 \end{bmatrix} \begin{bmatrix} V_1 & 0 & 0 \\ 0 & V_2 & 0 \\ 0 & 0 & V_3 \end{bmatrix} \begin{bmatrix} 1 & w_{21} & (w_{21}w_{32}) \\ 0 & 1 & w_{32} \\ 0 & 0 & 1 \end{bmatrix}$$

Multiply first two matrices:

$$= \begin{bmatrix} V_1 & 0 & 0 \\ w_{21}V_1 & V_2 & 0 \\ (w_{21}w_{32}V_1) & (w_{32}V_2) & V_3 \end{bmatrix} \begin{bmatrix} 1 & w_{21} & (w_{21}w_{32}) \\ 0 & 1 & w_{32} \\ 0 & 0 & 1 \end{bmatrix}$$

$$\boxed{= \begin{bmatrix} V_1 & V_1 w_{21} & V_1 w_{21} w_{32} \\ w_{21}V_1 & w_{21}^2 V_1 + V_2 & w_{21}^2 V_1 w_{32} + V_2 w_{32} \\ w_{21}w_{32}V_1 & w_{21}^2 w_{32} V_1 + w_{32} V_2 & w_{21}^2 w_{32}^2 V_1 + w_{32}^2 V_2 + V_3 \end{bmatrix}}$$

For Graph2:

Set Bias=0

$$
\begin{aligned}
Y_1 &= w_{12}Y_2 + b_1 + \sqrt{v_1}\,\epsilon_1 \\
Y_2 &= b_2 + \sqrt{v_2}\,\epsilon_2 \\
Y_3 &= w_{32}Y_2 + b_3 + \sqrt{v_3}\,\epsilon_3
\end{aligned}
\Longrightarrow
\begin{aligned}
Y_1 - w_{12}Y_2 &= \sqrt{v_1}\,\epsilon_1 \\
Y_2 &= \sqrt{v_2}\,\epsilon_2 \\
-w_{32}Y_2 + Y_3 &= \sqrt{v_3}\,\epsilon_3
\end{aligned}
$$

In Matrix Form:

$$
\underbrace{\begin{bmatrix} 1 & -w_{12} & 0 \\ 0 & 1 & 0 \\ 0 & -w_{32} & 1 \end{bmatrix}}_{A}
\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}}_{Y}
=
\underbrace{\begin{bmatrix} \sqrt{v_1}\,\epsilon_1 \\ \sqrt{v_2}\,\epsilon_2 \\ \sqrt{v_3}\,\epsilon_3 \end{bmatrix}}_{\epsilon}
$$

$$
\begin{aligned}
AY &= \epsilon \\
Y &= A^{-1}\epsilon \\
Cov(Y) &= A^{-1} Cov(\epsilon) (A^{-1})^T
\end{aligned}
$$

$$
Cov(Y) = \begin{bmatrix} 1 & w_{12} & 0 \\ 0 & 1 & 0 \\ 0 & w_{32} & 1 \end{bmatrix}
\begin{bmatrix} v_1 & 0 & 0 \\ 0 & v_2 & 0 \\ 0 & 0 & v_3 \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ w_{12} & 1 & w_{32} \\ 0 & 0 & 1 \end{bmatrix}
$$

$$
= \begin{bmatrix} v_1 & w_{12}v_2 & 0 \\ 0 & v_2 & 0 \\ 0 & w_{32}v_2 & v_3 \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ w_{12} & 1 & w_{32} \\ 0 & 0 & 1 \end{bmatrix}
$$

$$= \boxed{\begin{bmatrix} V_1 + W_{12}^2 V_2 & W_{12} V_2^2 & W_{12} V_2 W_{32} \\ V_2 W_{12} & V_2 & V_2 W_{32} \\ W_{32} V_2 W_{12} & W_{32} V_2 & W_{32}^2 V_2 + V_3 \end{bmatrix}}$$

cov

| by

For Graph 3:

$Y_1 = b_1 + \sqrt{V_1}\,\epsilon_1$

$Y_2 = W_{21} Y_1 + W_{23} Y_3 + b_2 + \sqrt{V_2}\,\epsilon_2$

$Y_3 = b_3 + \sqrt{V_3}\,\epsilon_3$

Set bias $= 0$

$$\begin{aligned} Y_1 \qquad\qquad\qquad\qquad &= \sqrt{V_1}\,\epsilon_1 \\ -W_{21} Y_1 \quad Y_2 \quad W_{23} Y_3 &= \sqrt{V_2}\,\epsilon_2 \\ Y_3 &= \sqrt{V_3}\,\epsilon_3 \end{aligned}$$

In Matrix Form:

$$\begin{bmatrix} 1 & 0 & 0 \\ -W_{21} & 1 & W_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} \sqrt{V_1}\,\epsilon_1 \\ \sqrt{V_2}\,\epsilon_2 \\ \sqrt{V_3}\,\epsilon_3 \end{bmatrix}$$

$\Downarrow$

$AY = \epsilon$

$\mathrm{cov}(Y) = A^{-1}\,\mathrm{cov}(\epsilon)\,(A^{-1})^T$

$$\mathrm{Cov}(Y) = \begin{bmatrix} 1 & 0 & 0 \\ W_{21} & 1 & W_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} V_1 & 0 & 0 \\ 0 & V_2 & 0 \\ 0 & 0 & V_3 \end{bmatrix} \begin{bmatrix} 1 & W_{21} & 0 \\ 0 & 1 & 0 \\ 0 & W_{23} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} V_1 & 0 & 0 \\ W_{21}V_1 & V_2 & W_{23}V_3 \\ 0 & 0 & V_3 \end{bmatrix} \begin{bmatrix} 1 & W_{21} & 0 \\ 0 & 1 & 0 \\ 0 & W_{23} & 1 \end{bmatrix}$$

$$= \boxed{\begin{bmatrix} V_1 & V_1 W_{21} & 0 \\ W_{21}V_1 & W_{21}^2 V_1 + V_2 + W_{23}^2 V_3 & W_{23}V_3 \\ 0 & V_3 W_{23} & V_3 \end{bmatrix}}$$

c). Are any of these graphical models equivalent? (Equivalent if any Matrix under one model can also be generated from the other model selecting suitable parameters and vice versa]. Prove your finding

Let's try Graph1 ⟺ Graph2 :

- $V_{(G1)1} = V_{(G2)1} + V_{(G2)2} W_{(G2)12}^2$
- $V_{(G1)2} = V_{(G2)} - V_{(G1)1} W_{(G1)21}^2$
- $V_{(G1)3} = V_{(G2)3} + V_{(G2)2} W_{(G2)32}^2 - V_{(G1)1} W_{(G1)21}^2 W_{(G1)32}^2 - V_{(G1)2} W_{(G1)32}^2$
- $W_{(G1)32} = \frac{V_{(G2)2} W_{(G2)12} W_{(G2)32}}{V_{(G1)1} W_{(G1)21}}$
- $W_{(G1)21} = \frac{V_{(G2)2} W_{(G2)12}}{V_{(G1)1}}$

The number of equations is less than the number of unknowns so we cannot prove that Graph1 ⟺ Graph2.

Lets try Graph1 ⟺ Graph3 :

- $V_{(G1)1} = V_{(G3)1}$
- $V_{(G1)2} = V_{(G3)1} W_{(G3)21}^2 + V_{(G3)2} + V_{(G3)3} W_{(G3)}^2 - V_{(G1)1} W_{(G1)}^2$
- $V_{(G1)3} = V_{(G3)3} - V_{(G1)1} W_{(G1)21}^2 W_{(G1)32}^2 - V_{(G1)2} W_{(G1)32}^2$
- $W_{(G1)21} = \frac{V_{(G3)2} W_{(G3)12}}{V_{(G1)1}}$
- $W_{(G1)32} = \frac{V_{(G3)2} W_{(G3)32} - V_{(G1)1} W_{(G1)21}^2 W_{(G1)32}}{V_{(G1)2}}$

The same problem emerges as when we tried to prove Graph1 ⟺ Graph2. Thus Graph1 ⟺ Graph3

Finally let's try Graph2 <=> Graph3

- $V_{(G2)1} = V_{(G3)1} - V_{(G2)2} W_{(G2)12}^2$
- $V_{(G2)2} = \dfrac{V_{(G3)1} W_{(G3)21}}{W_{(G2)12}}$
- $V_{(G2)3} = V_{(G3)3} - V_{(G2)2} W_{(G2)32}^2$
- $W_{(G2)12} = \dfrac{V_{(G3)1} W_{(G3)21}}{V_{(G2)2}}$
- $W_{(G2)32} = \dfrac{W_{(G3)23} V_{(G3)3}}{V_{(G2)2}}$

The number of equations is equal to the number of unknowns, so there exists a unique combination of the variables such that Graph 2 can be transformed into Graph 3 and vice versa.

4. Let $x = (x_1, \ldots, x_p)'$ be a random vector with multivariate normal distribution $N(\mu, \Sigma)$. Partition x into two subvectors x_1 and x_2, and partition the mean vector and covariance matrix accordingly as:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

a). Define $z_2 = x_2 - [\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)]$ Show that $E[z_2] = 0$

$$E[z_2] = E[x_2 - [\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)]]$$

$$= E[x_2] - E[\mu_2] - \Sigma_{21}\Sigma_{11}^{-1}(E[x_1] - E[\mu_1])$$

$$= \cancelto{0}{\mu_2 - \mu_2} - \Sigma_{21}\Sigma_{11}^{-1}\cancelto{0}{(\mu_1 - \mu_1)}$$

$$= 0$$

Thus, we have proven that $E[z_2] = 0$

b). Show that $cov(x_1, Z_2) = 0$ and $Var(Z_2) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$

$$Cov(X_1, Z_2) = Cov(X_1, X_2 - [\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)])$$

$$= Cov(X_1, X_2) - \cancelto{0}{Cov(X_1, \mu_2)} - Cov(X_1, \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1))$$

(From Matrix) (Distrib Covariance)

$$= \Sigma_{12} - \Sigma_{21}\Sigma_{11}^{-1}(Cov(X_1, X_1) - \cancelto{0}{Cov(X_1, \mu_1)})$$

(Variance)

$$= \Sigma_{12} - \Sigma_{21}\Sigma_{11}^{-1}(\Sigma_{11})$$

$$= \Sigma_{12} - \Sigma_{21} I$$

$$= \Sigma_{12} - \Sigma_{21}$$

$$= 0$$

$$Var(Z_2) = Var[x_2 - [\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)]]$$

$$= Var[X_2] - \cancelto{0}{Var[\mu_2]} - \Sigma_{21}\Sigma_{11}^{-1} Var(X_1 - \mu_1)(\Sigma_{21}\Sigma_{11}^{-1})^T$$

$$= Var[X_2] - \Sigma_{21}\Sigma_{11}^{-1} Var[X_1] (\Sigma_{21}\Sigma_{11}^{-1})^T$$

$$= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{11}(\Sigma_{21}\Sigma_{11}^{-1})^T$$

$$= \Sigma_{22} - \Sigma_{21} I (\Sigma_{21}\Sigma_{11}^{-1})^T$$

Thus, $Var(Z_2) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$

c). Conclude from part (a) and (b) that

$$E(X_2|X_1) = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)$$

$$Var(X_2|X_1) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

$$Z_2 = X_2 - [\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)]$$

$$X_2 = Z_2 + [\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)]$$

$$E[X_2|X_1] = E(Z_2 + [\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)] \,|\, X_1)$$

$$= \cancel{E(Z_2|X_1)}^{0} + E(\mu_2|X_1) + E(\Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)|X_1)$$

$$= \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(E(X_1|X_1) - E(\mu_1|X_1))$$

$$= \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)$$

$$Var(X_2|X_1) = Var(Z_2 + [\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)] \,|\, X_1)$$

$$= Var(Z_2|X_1) + 0$$

$$= Var(Z_2)$$

$$= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \quad \text{(From Part b)}$$

d). Let $\mu=0$

From Part (b) we know:

$$\operatorname{cov}\begin{pmatrix} X_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}$$

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} I & 0 \\ \Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$\Sigma = \underbrace{\begin{pmatrix} I & 0 \\ \Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}}_{M_1}\underbrace{\begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}}_{M_2}\underbrace{\begin{pmatrix} I & \Sigma_{11}^{-1}\Sigma_{21} \\ 0 & I \end{pmatrix}}_{M_3}$$

$$(M_1M_2M_3)^{-1} = M_3^{-1}M_2^{-1}M_1^{-1}$$

$$\Sigma^{-1} = \begin{pmatrix} I & \Sigma_{11}^{-1}\Sigma_{21} \\ 0 & I \end{pmatrix}^{-1}\begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}^{-1}\begin{pmatrix} I & 0 \\ \Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} I & -\Sigma_{11}^{-1}\Sigma_{21} \\ 0 & I \end{pmatrix}\begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix}\begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}$$

e). Let $\theta = \Sigma^{-1}$ be the Precision Matrix, and denote its elements by θ_{jk}. Show that for any two components X_j and X_k of X, it holds that

$$\mathrm{Cor}(X_j, X_k \mid X \setminus \{X_j, X_k\}) = \frac{-\theta_{jk}}{\sqrt{\theta_{jj}\theta_{kk}}}$$

Also known as Partial Correlation of X_j and X_k.
Tells us whether X_j and X_k are conditional independent given the rest of X depends on whether θ_{jk} is 0.

We know $X \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$

$$\mu_2 = \begin{pmatrix} \mu_j \\ \mu_k \end{pmatrix} \quad \text{and} \quad \Sigma_{22} \begin{pmatrix} Var(X_j) & Cov(X_j, X_k) \\ Cov(X_k, X_j) & Var(X_k) \end{pmatrix}$$

We have shown $Var(X_2|X_1) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \Sigma_{jk}^{-1}$

We have also shown

$$\Sigma^{-1} = \begin{pmatrix} I & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{11}^{-1} & \left(-\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}\right) \\ 0 & (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}$$

$$\begin{pmatrix} \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}^{-1})^{-1}\Sigma_{21}\Sigma_{11}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \\ (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{21}\Sigma_{11}^{-1} & \hat{\Sigma}_j \end{pmatrix}$$

$\theta = \Sigma^{-1}$ so

$$\begin{pmatrix} \theta_{jj} & \theta_{jk} \\ \theta_{kj} & \theta_{kk} \end{pmatrix} = \begin{pmatrix} Var(X_j) & Cov(X_j, X_k) \\ Cov(X_k, X_j) & Var(X_k) \end{pmatrix}$$

$$Cov(X_j, X_k) = -\frac{\theta_{jk}}{\theta_{kk}} Var(X_j) = Cov(X_k, X_j) = -\frac{\theta_{jk}}{\theta_{jj}} Var(X_k)$$

Thus, $Cor(X_2, X_1) = Cor(X_j, X_k \mid X \setminus \{X_j, X_k\}) = \dfrac{Cov(X_j, X_k)}{\sqrt{Var(X_j) Var(X_k)}}$