Homework 7
Due Thursday November 29

(1) Create a simulated dataset as follows. Create a 20 by 20 matrix E. The diagonal elements should be zero. Make each upper diagonal matrix a 1 or 0 by generating a random Bernoulli with success probability .05. Now let $b = 40$ and define Σ^{-1} using this code:

```
B    = b*E
a    = apply(B,1,sum)
a    = max(a) + 1
a    = rep(a,nrow(E))
Sinv = B + diag(a)
```

Draw the graph. The following code might be helpful for drawing a graph:

```
draw = function(E){
     par(pty="s")
     n = nrow(E)
     angle = seq(0,2*pi,length=n+1)
     angle = angle[1:n]
     x = cos(angle)
     y = sin(angle)
     plot(x,y,pch=20,lwd=3,xlab="",ylab="",
          xaxt="n",yaxt="n",xlim=c(-1.5,1.5),ylim=c(-1.5,1.5))
     for(i in 1:n){
          text(1.2*cos(angle[i]),1.2*sin(angle[i]),paste("X",i),lwd=3,font=2)
          }
     for(i in 1:(n-1)){
          for(j in (i+1):n){
               if(E[i,j] == 1)lines(c(x[i],x[j]),c(y[i],y[j]),lwd=3)
               }
          }
     return(NULL)
     }
```

Now generate 100 random vectors from a $N(0, \Sigma)$. To do this, use the following fact: if $Z \sim N(0, I)$ then $X = \Sigma^{1/2} Z \sim N(0, \Sigma)$. To compute the square root of a matrix use:

```
e = eigen(A)
V = e$vectors
s = V %*% sqrt(diag(e$values)) %*% t(V)
```

Estimate the graph from your data. Use the SIN method and the lasso method and compare your answers. To assess the variability of the estimator draw 10 bootstrap samples and draw the graphs from these bootstrap samples.

(2) Consider random variables (X_1, X_2, X_3). In each of the following cases, draw a graph with the fewest possible number of edges that has the given independence relations.

(a) $X_1 \amalg X_3 \mid X_2$.

(b) $X_1 \amalg X_2 \mid X_3$ and $X_1 \amalg X_3 \mid X_2$.

(c) $X_1 \amalg X_2 \mid X_3$ and $X_1 \amalg X_3 \mid X_2$ and $X_2 \amalg X_3 \mid X_1$.

(3) Consider random variables (X_1, X_2, X_3, X_4). In each of the following cases, draw a graph with the fewest possible number of edges that has the given independence relations.

(a) $X_1 \amalg X_3 \mid X_2, X_4$ and $X_1 \amalg X_4 \mid X_2, X_3$ and $X_2 \amalg X_4 \mid X_1, X_3$.

(b) $X_1 \amalg X_2 \mid X_3, X_4$ and $X_1 \amalg X_3 \mid X_2, X_4$ and $X_2 \amalg X_3 \mid X_1, X_4$.

(c) $X_1 \amalg X_3 \mid X_2, X_4$ and $X_2 \amalg X_4 \mid X_1, X_3$.

(4) Construct a distribution on three variables that cannot be represented by an undirected graph. Construct a distribution on four variables that cannot be represented by a directed graph.

(5) Get the undirected graph data from the course website. There are 5 binary variables. Fit an undirected graph using loglinear models. Approximate the distribution of the data with a Normal. Now estimate the graph using SIN. (In other words, just use the sample covariance matrix.)

(6) Write down the conditional independencies from Figures 1-4.

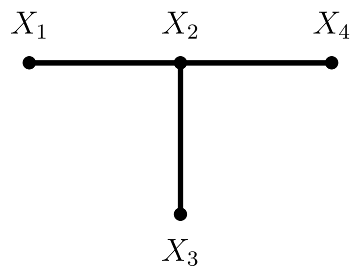

Figure 1:

Figure 2:

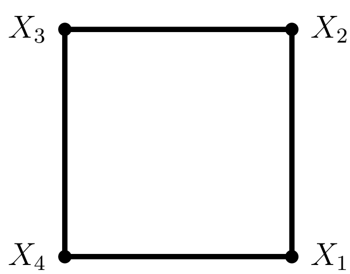

Figure 3:

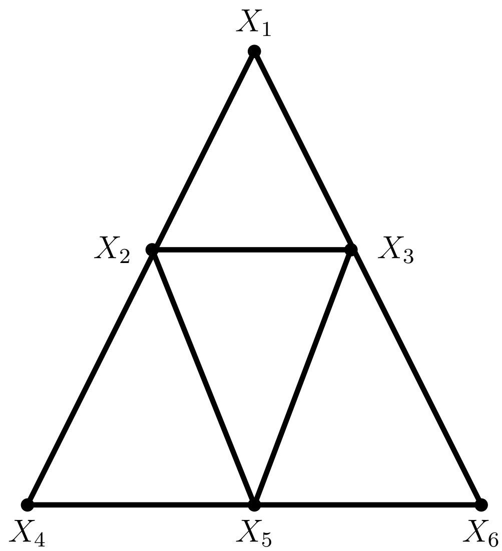

Figure 4: