Assignment 2

Due: October 22nd at 8pm EST

Instructions: This is an individual assignment, not group work. Though you may discuss the problems with your classmates, you must solve the problems and write the solutions independently. As stated in the syllabus, copying code from a classmate or the internet (even with minor changes) constitutes plagiarism. You are required to submit your answers in pdf form (use LaTeX) in a file called `<your-UNI>-hw2.pdf` to courseworks. The code for the programming assignment should be appended at the end of this pdf. Late submissions will be penalized, except in extenuating circumstances such as medical or family emergency. Submissions submitted 0-24 hours late will be penalized 10%, 24-48 hours late by 20%, 48-72 hours late by 30%, and later than 72 hours by 100% (i.e., zero credit). Each question is worth 5 points, for a total of 30 points.

Problem 1

Let $X \sim N(\mu, \Sigma)$ where $X = (X_1, ..., X_p)$. We saw in class that marginal independence in the multivariate normal model corresponds to zeros in the covariance matrix, i.e., $X_i \perp\!\!\!\perp X_j$ if and only if $\Sigma_{ij} = 0$. We also saw that conditional independence where the conditioning set is "everything else" corresponds to zeros in the precision matrix: $X_i \perp\!\!\!\perp X_j | X \setminus \{X_i, X_j\}$ if and only if $K_{ij} = 0$. Let $X_S \subset X \setminus \{X_i, X_j\}$ be some arbitrary subset of the variables. Explain why the following is true: $X_i \perp\!\!\!\perp X_j | X_S$ if and only if $(\Sigma_{\{ijS\},\{ijS\}})^{-1}_{ij} = 0$ (Hint: that conditional independence statement amounts to a marginal independence $X_i \perp\!\!\!\perp X_j$ in the conditional distribution where $X_S = x_S$. What is that conditional distribution?)

Problem 2

Show that the binary Ising model (pairwise MRF) implies that

$$P(X_i = 1 | X_{-i} = x_{-i}) = \frac{1}{1 + \exp(-\theta_{i0} - \sum_{i \sim k} \theta_{ik} x_k)}$$

where X_{-i} denotes all the vertices except X_i. That is, the Ising model implies a logistic model for each vertex conditional on all the rest. (This is excersize 17.11 in Hastie et al.)

Problem 3

Use the following code to simulate data from a given MRF independence model:

```
set.seed(123)
( K <- cbind(c(10,7,7,0),c(7,20,0,7),c(7,0,30,7),c(0,7,7,40)) )
data <- as.data.frame(mvrnorm(n=10000,mu=c(0,0,0,0),Sigma=solve(K)))
colnames(data) <- c("X1","X2","X3","X4")
```

What are the conditional independencies that are representing in this precision matrix? What is the corresponding graph? Verify the conditional independence constraints by using linear regression. Explain all this.
Next, use the `gRim` package to fit the model, i.e., estimate the precision matrix subject to the graph constraints.

```
library(gRim)
glist <- list( *insert list of edges here* )
ddd <- cov.wt(data, method="ML")
fit <- ggmfit(ddd$cov, ddd$n.obs, glist) # Estimate parameters using IPF
fit$K # estimated precision matrix
```

Did it work? How do you know?

Problem 4

Consider the Gaussian Bayesian Network model with the following covariance matrix:

```
set.seed(123)
( Sig <- cbind(c(3,-1.4,0,0),c(-1.4,3,1.4,1.4),c(0,1.4,3,0),c(0,1.4,0,3)) )
data <- as.data.frame(mvrnorm(n=10000,mu=c(0,0,0,0),Sigma=Sig))
colnames(data) <- c("X1","X2","X3","X4")
```

and the DAG $\mathcal{G}$ with edges $X_1 \to X_2 \leftarrow X_3$ and $X_4 \to X_2$.

a) What correlation constraints does this model represent? Estimate the correlation matrix.

b) Consider also the moralized graph $\mathcal{G}^m$ and what the corresponding precision matrix K would look like. What are the partial correlation constraints represented in K? How does this make sense with respect to Σ above?

c) Following steps similar to the previous problem, estimate the corresponding precision matrix K from this data (using `ggmfit`). Take the inverse and compare to the true covariance matrix.

Problem 5

Use `dagitty` to simulate 10000 observations from this graph:

```
g <- dagitty( "dag{ x <- u1; u1 -> m <- u2 ; u2 -> y }" )
```

Here U_1, U_2 represent unmeasured variables. Estimate the effect of X on Y adjusting for M in a linear regression, obtaining a 95% confidence interval for the effect. Then estimate the same effect (and confidence interval) using the correct sufficient adjustment set that you can obtain from `dagitty`. What conclusion should be drawn from this example?

Problem 6

Construct the DAG in Figure 1 as a daggity object. Simulate 10000 observations from this graph as you did on the last homework. Estimate the effect of E on F and the effect of B on A using backdoor adjustment and linear regression. If there is more than one sufficient adjustment set, try each of the ones identified by `dagitty` and compare them. Are the point estimates similar? Do the estimates have similar variance (or confidence interval length)? Compare also these estimates against an approach which simply adjusts for *all* other variables in the graph. How are the results different (if they are) and what is the explanation?

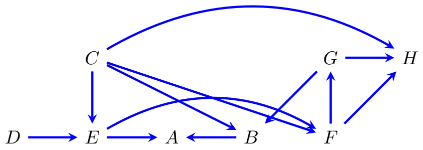

Figure 1