Assignment 1

Due: October 1st at 8pm EST

Instructions: This homework requires some calculation, short proofs, and programming in R. 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>-hw1.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). The first 6 questions are worth 5 points each, and the programming assignment is worth 10 points, for a total of 40 points.

Problem 1

Consider the DAG in Figure 1.

a) List all the independencies that comprise the local Markov property for this DAG.

b) Is $X_2 \perp_d X_9 | X_4$? Is $X_7 \perp_d X_5 | \{X_3, X_8\}$? Is $\{X_2, X_4\} \perp_d X_7 | \{X_6, X_9, X_{10}\}$?

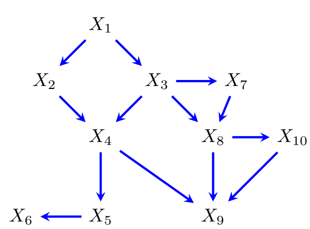

Figure 1

Problem 2

In class we defined the Markov blanket of a vertex X_i in DAG $\mathcal{G}$. Prove, using d-separation, that this set satisfies $X_i \perp\!\!\!\perp X \setminus \{\text{Mb}(X_i, \mathcal{G}), X_i\} | \, \text{Mb}(X_i, \mathcal{G})$.

Problem 3

Consider an undirected graph $\mathcal{G}$ which satisfies the local Markov property for UGs. Prove that it also satisfies the pairwise Markov property using the graphoid axioms. That is, assume that $X_i \perp\!\!\!\perp X \setminus \text{Cl}(X_i, \mathcal{G}) | \, \text{Ne}(X_i, \mathcal{G})$ and that $X_i \not\sim X_j$ in $\mathcal{G}$, and derive that $X_i \perp\!\!\!\perp X_j | X \setminus \{X_i, X_j\}$. (Hint: this is quite easy if you pick the right graphoid axiom!)

Problem 4

a) Consider the DAG $\mathcal{G}$ in Figure 2. Which of the DAGs in Figure 3 are Markov equivalent to $\mathcal{G}$?

b) How many DAGs are Markov equivalent to the "chain DAG" $X_1 \to X_2 \to \cdots \to X_p$ (excluding itself)?

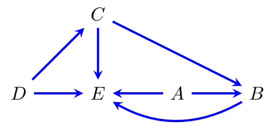

Figure 2

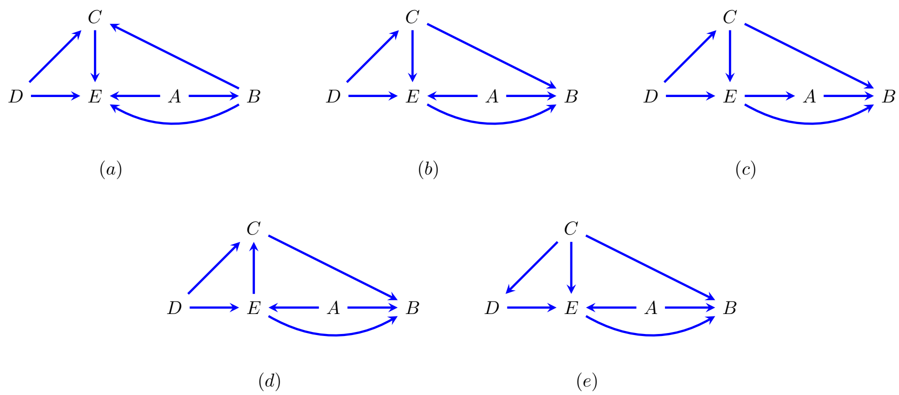

Figure 3

Problem 5

Consider the DAG $\mathcal{G}$ in Figure 1 and it's moralized graph $\mathcal{G}^m$.

a) How many factor potentials appear in the factorization property for $\mathcal{G}^m$, when the factorization is by *maximal* cliques?

b) If each factor is parameterized by one parameter for each variable which appears in its scope, plus one parameter for each pair in the scope, and one for each three-way interaction (each triple in the scope), and so on... how many parameters are there in total in this representation of the joint distribution? (You can assume that the single-variable parameters appear only once for each variable, so you're not "double counting" when the same variable appears in multiple cliques.)

c) How many parameters would be required to specify the joint without a graphical model, assuming binary variables?

d) How many parameters would be required to specify the joint using the DAG factorization, assuming binary variables?

e) How many factor potentials appear when the MRF factorization is by pairwise cliques?

f) How many parameters are there total in this latter representation of the joint distribution, with same scheme as above?

Problem 6

Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be three disjoint sets of nodes in a DAG $\mathcal{G}$. Prove that $\mathbf{C}$ d-separates $\mathbf{A}$ from $\mathbf{B}$ in $\mathcal{G}$ if and only if $\mathbf{C}$ separates $\mathbf{A}$ from $\mathbf{B}$ in $(\mathcal{G}_{\text{An}(\mathbf{A},\mathbf{B},\mathbf{C})})^m$. (Recall that we use the convention that $X_i \in \text{An}(X_i, \mathcal{G})$.)

Problem 7

In this problem you'll get acquainted with using the R package "dagitty". Being by installing the package and load it into your R session. (You can find basic instructions and tutorials for daggity online: `http://dagitty.net/primer/`)

Construct the DAG in Figure 4 as a daggity object. For all tasks below, copy and paste the output of your code into your pdf.

a) Using the `paths()` function, list all paths from C to H.

b) Use the `dseparated()` function to determine whether $E \perp_d G | A, B$.

c) Using the `impliedConditionalIndependencies()` function, list the conditional independencies relationships implied by the model. Try this also with the option "type = all.pairs". (This might take a while to run.) What explains the difference between these two lists, why is the first one shorter than the second? Consult the documentation of the package.

d) Simulate data from this DAG using the `simulateSEM()` function, which associates the DAG with a linear structural equation model. Set the path coefficient range to $(0.4, 0.7)$ and the sample size to 10000. Verify with linear regression that the Markov blanket property holds for vertex B, i.e., that $\text{Mb}(B, \mathcal{G})$ (use the `markovBlanket()` function) renders the rest of the variables independent of B. You can do this by examining the p-values (or confidence intervals) for all covariates outside the Markov blanket in the regression of $B \sim \text{Mb}(B, \mathcal{G}) +$ remaining covariates.

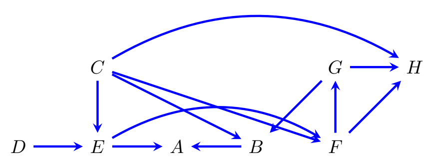

Figure 4