ISTA 410/510 Midterm I (take home, same structure as assignments)

For contribution to the final grade, due dates, current late policy, and instructions for handing the assignment in, see the assignment web page.

Please create a PDF document with your answers and/or the results of any programs that you write. You should also hand in your programs. (This assignment does **not** require any programing, but if you choose to write code for some part of it, you should hand it in).

Grad students should hand in work for 10 points, undergraduates 6 points. The main different between the (*) questions is the degree of specificity of the question (grad students need to develop assumptions for their approach) rather than technical difficulty. So undergraduates might consider swapping some questions.

1. For each of the factorizations below, **first** reorder the factors so that all variables first appear as not being conditioned on ($). In other words, they first appear to the left of the "|" . This is one canonical way to write the chain rule, i.e., P(A,B,C)=P(A)P(B|A)P(C|A,B). (The answer is not necessarily unique). **Second**, identify any differences between this way of writing the factorization and the canonical form of the chain rule ($). This identifies where the factorization restricts the distribution from the general case. **Third**, draw graph representing the factorization ($).

   (a) p(A,B,C,D,E)=p(A|D,E)p(B|D)p(C)p(D|C)p(E)

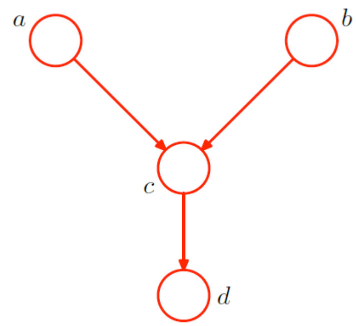
   (b) p(A,B,C,D,E)=p(A|C)p(E|B)p(C)p(B)p(D|A,E)

   (c) p(A,B,C,D,E)=p(C|A,B,D,E)p(A|B,D)p(D|B)p(B)p(E|A,B,D)

2. For **each** of your graphs in the previous question, use d-separation to argue whether or not i) C and B must be conditionally independent give D ($), and ii) E and C must be conditionally independent given B ($). [Note that "must be" is asking about properties that must be true, regardless of the distribution. There may be some specific distributions which have extra conditional independencies not evident by d-separation, but there will be other distributions where this would not be the case].

3. (*, From Bishop) ($)

Figure 8.54 Example of a graphical model used to explore the conditional independence properties of the head-to-head path a–c–b when a descendant of c, namely the node d, is observed.



8.10 ($\star$) Consider the directed graph shown in Figure 8.54 in which none of the variables is observed. Show that $a \perp\!\!\!\perp b \mid \emptyset$. Suppose we now observe the variable d. Show that in general $a \not\perp\!\!\!\perp b \mid d$.

4. (** From Bishop)

8.11 ($\star\star$) Consider the example of the car fuel system shown in Figure 8.21, and suppose that instead of observing the state of the fuel gauge G directly, the gauge is seen by the driver D who reports to us the reading on the gauge. This report is either that the gauge shows full $D = 1$ or that it shows empty $D = 0$. Our driver is a bit unreliable, as expressed through the following probabilities

$$p(D = 1|G = 1) = 0.9 \quad (8.105)$$

$$p(D = 0|G = 0) = 0.9. \quad (8.106)$$

Suppose that the driver tells us that the fuel gauge shows empty, in other words that we observe $D = 0$. Evaluate the probability that the tank is empty given only this observation. Similarly, evaluate the corresponding probability given also the observation that the battery is flat, and note that this second probability is lower. Discuss the intuition behind this result, and relate the result to Figure 8.54.

5. (++ (a,b,c) =1; (d,e,f)=1; (g,h)=1). Supposed that your undergraduate **grades** (G) influence the **letters** (L) that professors write for you (higher grade, better letter). Another thing that influences those letters is your **interactions** (I) with those professors (better interaction, better letter). Now suppose that the **rank** (R) of the graduate program you get into is influenced by your **GRE test (T)**, your grades, and your letter (better letter increases probability that rank will be higher). Finally, supposed that whether or not you get an academic **job** (J) depends on the graduate program rank (better rank increases your chances), and the **state** (S) where the job is located (political comment goes here).

(a) Draw a directed graphical model for this problem ($).

(b) Write down an expression for the probability distribution that corresponds to your model ($).

(c) Argue that getting the job is independent of your grades, *conditioned on your grad program* ($). (Yup, eventually they are forgotten about!).

(d) Are GRE scores independent of grades in your model? Provide a formal argument based on rules learned in class ($).

(e) Are GRE scores independent of the quality of the letter in your model? Provide an argument based on rules learned in class ($).

(f) Are GRE scores independent of the quality of the letter *given the grad program*? Provide an argument based on rules learned in class ($).

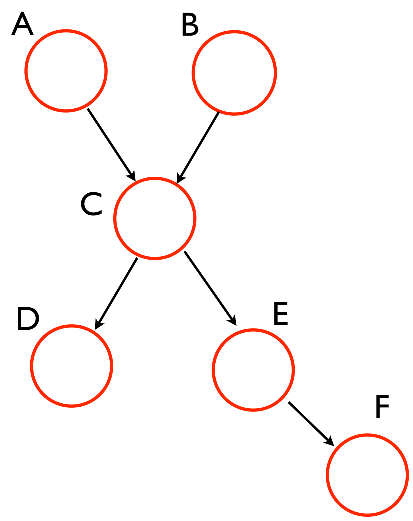
(g) Now suppose that getting the job *also* depends on their interaction with their professors (better interaction, higher chance). In other words, the interaction with the professor might have revealed some qualities have which may also help them in the interview. Is getting the job *still* independent of their grades *given their grad program*? Provide an argument based on rules learned in class ($).

If you argue "no", then explain it in English as well ($). In particular address whether learning that your friend had really high grades increases or decreases the amount of money you are willing to bet that they get the job.

(h) Using the extended model from (g), suppose you now have also come across the letter (the professor forgot to shred the draft). Conditioned on your knowledge of both the *grad program* that your friend is in, and the *letter*, are their grades independent of them getting a job? ($). Provide an argument based on rules learned in class. ($).

If you argue "no", then explain it in English as well ($). In particular address whether learning that your friend had really high grades increases or decreases the amount of money you are willing to bet that they get the job.

6. Consider this graph:



a) Using the abstract definition of what a directed graphical model means in terms of local conditional independencies, write down all the local conditional independencies that it expresses ($). You can make use of the compact representation for X being conditionally independent of multiple other by separating them with commas. (This is valid because conditioning on them considered as a group implies doing so for each one in particular). Also, you do not need to write down independencies that are symmetric; one of the two will suffice.

b) Use this to derive an expression for the probability distribution expressed ($).

c) Write down the probability distribution available by our first definition of a directed graphical model that specified a particular factorization of the distribution ($). Is is the same? ($). Do you expect it to be? ($).

7. (*) Provide a graphical model for the following problem. Consider that there is a prior probability where in the sky you might see an asteroid, as most of them are in solar orbit in the same plane as the other planets, and so you will find them near the “ecliptic”. Further, the location of the moon in the sky influences detections. Ditto for the weather. Even if there is a signal, the detector might miss it. On the other hand, the detector hallucinates detections (noise) detections. Provide a graphical model for detections of moving objects that might be asteroids, together with the relevant variables ($). Note that part of the problem is to figure out what they are. Be very clear in your writeup what the variables represent ($). Finally, ensure that you have a generative model by explaining how you get real data samples by ancestral sampling ($).

8. (*) Provide a graphical model for the following problem. Consider a dancer with LED lights attached to appropriate parts of their body including their center of mass, their forehead, and major joints (shoulders, knees, elbows, ankles). Consider a camera function with unknown parameters which have prior probability distributions. This camera model mediates how points in 3D become points in images. Notice that you may not see an LED due to occlusion. Provide a graphical model for this situation that includes the data we plan to observe (location of LED lights in the image) ($). Explain your notation and assumptions clearly ($) — there are many possible answers, but we would like a model that makes sense and is consistent with your graphical model. Finally, ensure that you have a generative model by explaining how you get real data samples by ancestral sampling ($).

9. (*) Consider people a potential drug cure for some affliction. Consider the binary variable D (whether a person takes the drug), the binary variable C (whether they get cured), and a binary variable H (whether they are health conscious). Whether they get cured is a (statistically) a function both of whether they took the drug, and whether they are health conscious. Further, the probability they took the drug can depend on whether they are health conscious. Provide a Bayes net for the statistical dependencies of the variables ($). Suppose we want to know whether the drug is helpful or not. Explain why this is not the right graph, and provide one that is better ($). Now consider determining whether a doctor should prescribe the drug (without trying to figure out whether the patient is health conscious or not). Create numbers that shows that the answer is different for the two graphs. In particular, you should be able to show that the drug can be bad for both groups (health conscious or not), but using the first graph can suggest that the drug should be prescribed.

10. (**) Any exercise in Koller and Friedman chapter 3 that you find interesting.

11. (**) Same as (13) excluding the one that you have already done.

12. (**) Same as (14) excluding ones that you have already done.