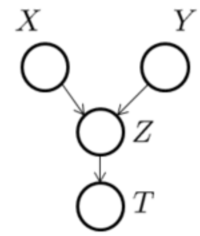
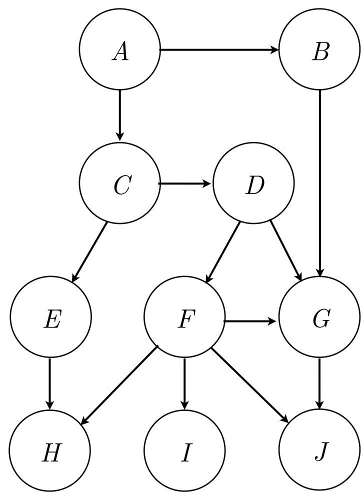
IFT6269-A2019 | **Hwk 3** | Name:
Prof: Simon Lacoste-Julien | Due date: Nov 3, 2020 | Student id:

Note: There are 10 bonus points in this assignment. The maximum number of points achievable for this assignment is 80. However, your score will be capped at 70 and grading will be over 70.

1. **DGM (5 points)**
   Consider the directed graphical model G on the right. Write down the implied factorization for any joint distribution $p \in \mathcal{L}(G)$. Is it true that $X \perp\!\!\!\perp Y \mid T$ for any $p \in \mathcal{L}(G)$? Prove or disprove.



2. **D-separation in DGM (5 points)**
   Indicate (yes or no) which conditional independence statements are true?



(a) $C \perp B$?
(b) $C \perp B \mid A$?
(c) $C \perp B \mid A, J$?
(d) $C \perp B \mid A, J, D$?
(e) $C \perp G$?
(f) $C \perp G \mid B$?
(g) $C \perp G \mid B, D$?
(h) $C \perp G \mid B, D, H$?
(i) $C \perp G \mid B, D, H, E$?
(j) $B \perp I \mid J$?

3. **Positive interactions in-V-structure (10 points)**
   Let X, Y, Z be binary random variables with a joint distribution parametrized according to the graph: $X \to Z \leftarrow Y$. We define the following:

$$\alpha := P(X = 1), \quad \beta := P(X = 1 \mid Z = 1), \quad \gamma := P(X = 1 \mid Z = 1, Y = 1)$$

   (a) For all the following cases, provide examples of joint probability tables (and calculate the quantities a, b, c), so that each of the following conditions are (individually) true:
   (i) $\gamma < \alpha$
   (ii) $\alpha < \gamma < \beta$
   (iii) $\beta < \alpha < \gamma$.
   (b) Think of X and Y as causes and Z as a common effect, for the previous three cases, briefly state (in a sentence or two) why the claims are true for your examples.

   Please use the following format to represent the joint distributions you choose for the cases in part *a)*. Study the expected format carefully and note how this is sufficient to represent the joint distribution. Also, use the corresponding boxes to write the values of α, β and γ for each of the three sub-questions.

X	Y	$P(X,Y)$
1	1	
1	0	
0	1	
0	0	

X	Y	$P(Z=1 \mid X,Y)$
1	1	
1	0	
0	1	
0	0	

α	β	γ

4. **Equivalence of directed tree DGM with undirected tree UGM (10 points)**
Let G be a directed tree and G' its corresponding undirected tree, i.e., the orientation of edges is ignored. Prove that $\mathcal{L}(G) = \mathcal{L}(G')$. *Hint: Recall that by the definition of a directed tree, G does not contain any v-structure.*

5. **Hammersley-Clifford counter-example (10 points)**
In class, we mentioned that the strict positivity of the joint distribution was crucial in the Hammersley-Clifford theorem. Here is a counter-example (4.4 in Koller & Friedman) that shows the problems when we have zero probabilities.

Consider a joint distribution p over four binary random variables: X_1, X_2, X_3 and X_4 which gives probability $\frac{1}{8}$ to each of the following eight configurations, and zero to all others:

$$(0,0,0,0) \quad (1,0,0,0) \quad (1,1,0,0) \quad (1,1,1,0) \quad (0,0,0,1) \quad (0,0,1,1) \quad (0,1,1,1) \quad (1,1,1,1)$$

Let G be the usual four nodes undirected graph X_1—X_2—X_3—X_4—X_1. One can show that p satisfies the global Markov property with respect to this graph G because of trivial deterministic relationships. For example, if we condition on $X_2 = 0$ and $X_4 = 0$, then the only value of X_3 with non-zero probability is $X_3 = 0$, and thus $X_3|(X_2 = 0, X_4 = 0)$ being a deterministic random variable, it is trivially conditionally independent to X_1. By (painfully) going through all other possibilities, we get similar situations: $X_2 = 0$ and $X_4 = 1$ forces $X_1 = 0$, etc.

Prove that the distribution p **cannot** factorize according to G, and thus $p \notin \mathcal{L}(G)$. *Hint:* argue by contradiction.

6. **Bizarre conditional independence properties (10 points)**
Let (X, Y, Z) be a random vector with a finite sample space. Consider the following statement:

$$\text{“If } X \perp\!\!\!\perp Y \mid Z \text{ and } X \perp\!\!\!\perp Y \text{ then } (X \perp\!\!\!\perp Z \text{ or } Y \perp\!\!\!\perp Z).\text{”}$$

   (a) Is this true if one assumes that Z is a binary variable? Prove or disprove.

   (b) Is the statement true in general? Prove or disprove.

7. **EM and Gaussian mixtures (4 points)**

Derive the form of the **M-step updates** for the parameters $\{\pi_k, \boldsymbol{\mu}_k, \sigma_k\}_{k=1}^K$ of a Gaussian mixture model in which the covariance matrices are proportional to the identity:

$$p(x) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(x \mid \boldsymbol{\mu}_k, \sigma_k \mathbf{I})$$

8. **Implementation: EM and Gaussian mixtures (26 points)**

Follow the instructions in this Colab notebook:

https://colab.research.google.com/drive/1mtyFbcWZAuoNV8zBicoH1Y0Bc_Sv2c7G