# Recitation 5: Review for Quiz 1

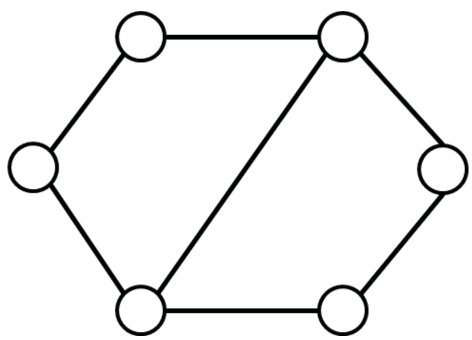
**Exercise 1:**

Please provide answer in True or False for the following questions. Provide explanations for your answers.

(a) Let $G$ be a chordal graph. As we know, one way to obtain chordal graph is to *triangulate* a given graph. Therefore, all chordal graphs must have at least one triangle.

**Solution: False**. A tree graph is also chordal, but there is no triangle in it.

(b) The following undirected graphical model can be obtained by moralization of a directed graphical model. If yes, provide that directed graphical model. If no such directed graphical model exists, explain why.

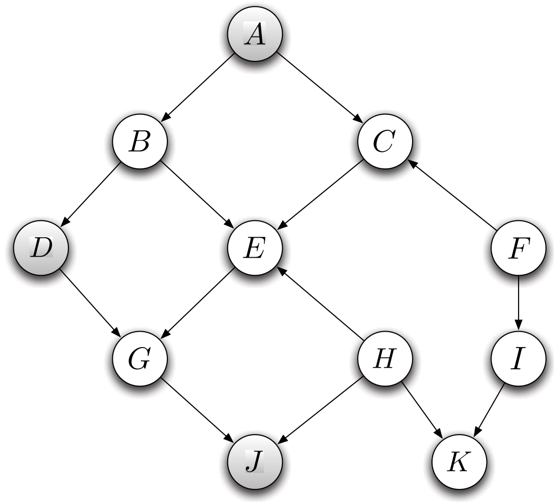


**Solution: False**. Assume it can be moralized from some directed graph. Moralizing a V-structure will result in a triangle. Since the above graph doesn't contain any triangles, all the edges in the undirected graph must be in the directed graph. Since we have a size 4 loop and the directed graph should be acyclic, there must be a V-structure in the directed graph, and thus the moralization process would add chord to the loop.

(c) An undirected graphical model over $N$ variables without triangles cannot have more than $N^2$ factors in its factorization (assume $p_{\mathbf{x}}(\mathbf{x}) > 0$ for all $\mathbf{x}$).

**Solution: True**. According to Hammersley-Clifford theorem, we have potential functions for cliques. If there is no triangle in the graph, the maximal cliques are edges, and for an undirected graph with $N$ nodes, there are at most $N(N-1) < N^2$ edges, thus we cannot have more than $N^2$ factors.

(d) Look at the following graph and answer the questions:
1. Is $B \perp C | A, D, J$?
2. Is $H \perp F | A, D, J$?
3. Moralize the graph.



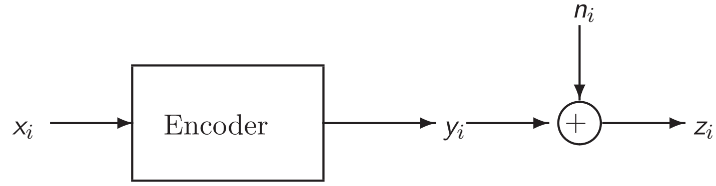

**Solution:** To answer this question, we'll apply Bayes Ball algorithm. Nodes A, D, J have primary shade, while B, C, E, F, G, H have secondary shade.
1. No, we can find a path: B→E→C
2. No, we can find a path: H→E→C→F
3. To moralize the graph, we need to add edges between the following pairs of variables: (A, F), (B, C), (B, H), (C, H), (D, E), (G, H), (H, I).

**Exercise 2:**

In digital communications, it is common to encode data to protect it against errors. A channel encoder takes as input a block of bits $x_1x_2 \dots x_n$, and output a codeword $y_1y_2, \dots y_m$. The mapping from the $x_1x_2 \dots x_n$ to the $y_1y_2, \dots y_m$ is injective, in other words, for any valid codeword $y_1y_2, \dots y_m$, the decoder is able to uniquely identify the original message $x_1x_2 \dots x_n$.

The codeword is sent through a communication channel, and the channel output at the receiver side, a sequence $z_1z_2, \dots z_m$, will be a noise distorted version of $y_1y_2, \dots y_m$.

For concreteness, the channel we use is described as follows: $z_i = y_i + n_i$, where $n_i$ is i.i.d. and Gaussian with mean 0 and variance 1.

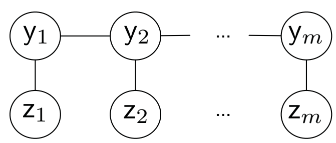



We would like to recover the original message $x_1x_2 \dots x_n$ from the received sequence $z_1z_2 \dots z_m$. Since the mapping from $x_1x_2 \dots x_n$ to the $y_1y_2, \dots y_m$ is injective, we can equivalently estimate $y_1y_2 \dots y_m$.

Our goal is minimize the probability of error, i.e., minimize $\Pr[y_1 y_2 \ldots y_m \neq \hat{y}_1 \hat{y}_2 \ldots \hat{y_m}]$, where $\hat{y}_i$ denotes the estimation of $y_i$, based on the received sequence $z_1 z_2 \ldots z_m$. In this exercise, we describe two very simple encoding schemes.

(a) A repetition code takes one random bit $x_1$ that is equally likely to be 0 or 1, and outputs a codeword $y$, where $y_1 = y_2 = \ldots = y_m = x_1$.
Write down a graphical model for $y_1 y_2 \ldots y_m$ and $z_1 z_2 \ldots z_m$. Specify all the potential functions.

**Solution:**




$$\begin{aligned}
\phi_{y_i}(y_i) &= \frac{1}{2}, \text{ for } y = 0, 1, i = 1, \ldots, m; \\
\psi_{y_i, y_{i+1}}(y_i, y_{i+1}) &= \mathbf{1}[y_i = y_{i+1}]; \\
\psi_{y_i, z_i}(y_i, z_i) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_i - z_i)^2}{2}}, \text{ for } i = 1, \ldots, m,
\end{aligned}$$

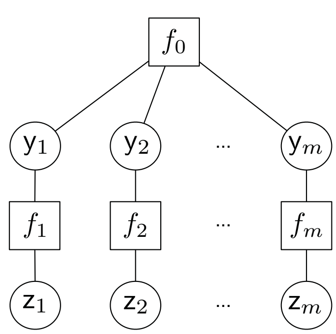
Any tree graph for all the variables $y_1, \ldots, y_m$ is a solution, with similar potential for the consistency constraints.

(b) A single parity check (SPC) code takes as input a block of bits $x_1 x_2 \ldots x_n$, where the input bits $x_i$ are i.i.d. with equal probability of being 0 or 1, and outputs codeword $y_1 y_2 \ldots y_m$. Here $m = n + 1$, $y_i = x_i$ for $1 \leq i \leq n$, and $y_m$ is chosen so that

$$\sum_{i=1}^{m} y_i = 0 (\text{mod } 2)$$

Provide a graphical model for $y_1 y_2 \ldots y_m$ and $z_1 z_2 \ldots z_m$ $(m = n + 1)$. Specify all the potential functions.

**Solution:**



$$f_0(y_1, y_2, \ldots, y_m) = \mathbf{1}[\sum_{i=1}^{m+1} y_i = 0(\text{mod } 2)]$$

$$f_i(y_i, z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_i - z_i)^2}{2}}$$

<u>**Exercise 3:**</u>
Suppose we have a Gaussian random vector $\mathbf{x}$ with information matrix

$$\mathbf{J} = \begin{bmatrix} 5 & 1 & 0 & 0 & 1 \\ 1 & 5 & 0 & 0 & 1 \\ 0 & 0 & 5 & 0 & 1 \\ 0 & 0 & 0 & 5 & 1 \\ 1 & 1 & 1 & 1 & 5 \end{bmatrix}.$$

Draw the undirected graphical model for: (a) $p_{x_1,x_2,x_3,x_4,x_5}$, (b) $p_{x_1,x_2,x_3,x_4|x_5}$, and (c) $p_{x_1,x_2,x_3,x_4}$

**Solution:**

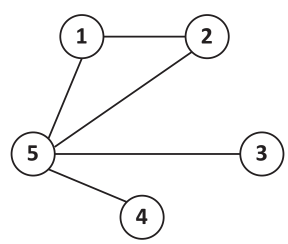



Figure 1: Graphical Model for $p_{x_1,x_2,x_3,x_4,x_5}$ can be drawn by directly looking at the J matrix

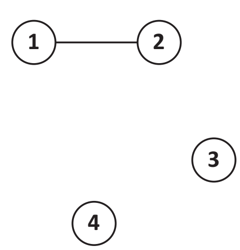



Figure 2: Graphical Model for $p_{x_1,x_2,x_3,x_4|x_5}$ can be drawn by deleting variable node 5 and all edges connecting to it.

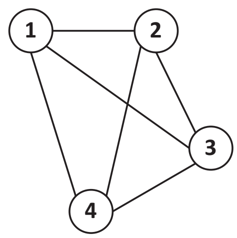
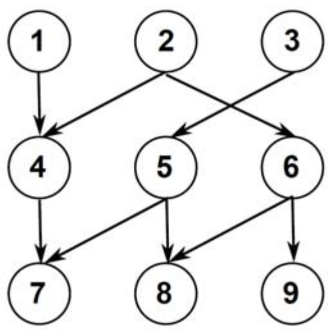



Figure 3: Graphical Model for $p_{x_1,x_2,x_3,x_4|x_5}$ can be drawn by deleting variable node 5 and connecting all its neighbours.

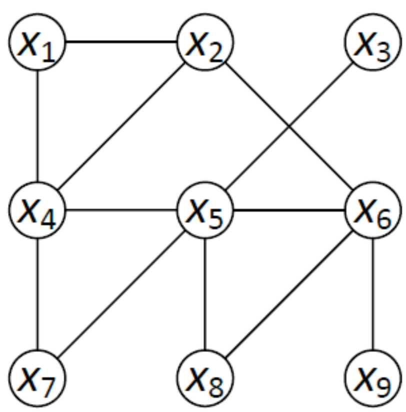
**Exercise 4:**

Consider the directed graph below and answer the following questions:




(a) Moralize the directed graph and consider the resulting undirected graph. Is it a chordal graph? If so, provide an elimination ordering which doesn't add any edges. If not, provide an ordering that adds as few edges as possible.

**Solution:** The moralized graph is:



It's not a chordal graph, because $\mathsf{x}_2$, $\mathsf{x}_4$, $\mathsf{x}_5$, and $\mathsf{x}_6$ form a cycle without a chord.

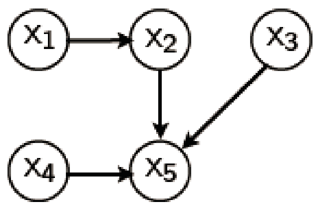
To obtain an elimination ordering that adds as few edges as possible, we can first remove $\mathsf{x}_1$, $\mathsf{x}_3$, $\mathsf{x}_7$, $\mathsf{x}_8$, $\mathsf{x}_9$ in any order without adding an edge. Then we have a loop of size 4 and removing these nodes in any order will result in one added edge.

(b) Consider running elimination algorithm on the original directed graph, using conditional probabilities directly as potentials. We will only perform 1 step, eliminating $\mathsf{x}_4$. Write down the potential functions created in this elimination step, as well as the computational complexity of this step. Assume all the variables take K possible values.

**Solution:** Eliminating $\mathsf{x}_4$ generates a factor over variables $\mathsf{x}_1$, $\mathsf{x}_2$, $\mathsf{x}_5$, and $\mathsf{x}_7$ and costs $O(K^5)$.

**<u>Exercise 5:</u>**

Consider the following directed graphical model over 5 K-ary variables (each variable takes K possible values)



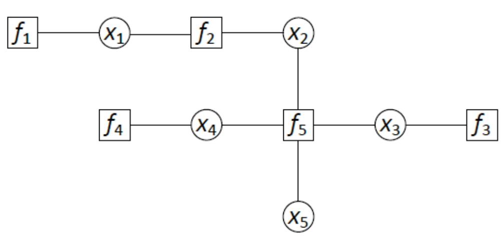

(a) Draw a factor graph that represents this distribution. Specify the factors.

**Solution:**

$$p_{\mathsf{x}_1,\mathsf{x}_2,\mathsf{x}_3,\mathsf{x}_4,\mathsf{x}_5} = p_{\mathsf{x}_1} p_{\mathsf{x}_2|\mathsf{x}_1} p_{\mathsf{x}_3} p_{\mathsf{x}_4} p_{\mathsf{x}_5|\mathsf{x}_2,\mathsf{x}_3,\mathsf{x}_4}$$

We can design a factor graph with the following factors:

$$\begin{aligned} f_1(x_1) &= p_{\mathsf{x}_1}(x_1) \\ f_2(x_1, x_2) &= p_{\mathsf{x}_2|\mathsf{x}_1}(x_2|x_1) \\ f_3(x_3) &= p_{\mathsf{x}_3}(x_3) \\ f_4(x_4) &= p_{\mathsf{x}_4}(x_4) \\ f_5(x_2, x_3, x_4, x_5) &= p_{\mathsf{x}_5|\mathsf{x}_2,\mathsf{x}_3,\mathsf{x}_4}(x_5|x_2, x_3, x_4) \end{aligned}$$



Note the distribution can be represented by a factor graph in many ways. But there should always be a factor that connects $\mathbf{x}_1$ and $\mathbf{x}_2$ and one that connects $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$, and $\mathbf{x}_5$.

(b) Suppose we run the sum-product algorithm on the resulting factor graph. What is the operational cost (additions, multiplications) needed to evaluate the message sent to $\mathbf{x}_3$ from the largest associated factor? You can assume that any other messages that you may rely upon have already been computed.

**Solution:** The largest factor associated with $\mathbf{x}_3$ is $f_5$, which has four neighbours: $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$, and $\mathbf{x}_5$. The message from $f_5$ to $\mathbf{x}_3$ can be calculated by:

$$m_{f_5 \to \mathbf{x}_3}(x_3) = \sum_{\mathbf{x}_2, \mathbf{x}_4, \mathbf{x}_5} f_5(x_2, x_3, x_4, x_5) m_{\mathbf{x}_2 \to f_5}(x_2) m_{\mathbf{x}_4 \to f_5}(x_4) m_{\mathbf{x}_5 \to f_5}(x_5)$$

The calculation requires $O(K^4)$ operations.

**<u>Exercise 6:</u>**

Suppose a collection of discrete random variables of interest $x_n$, $n = 0, 1, 2, \ldots, N$, forms a Markov chain characterized by a transition sdistribution $p(\cdot|\cdot)$ and initial distribution $\pi(\cdot)$. Specifically, for $n = 1, 2, \ldots, N$

$$\Pr[x_0 = a] = \pi(a)$$
$$\Pr[x_n = a \mid x_{n-1} = b] = p(a \mid b)$$

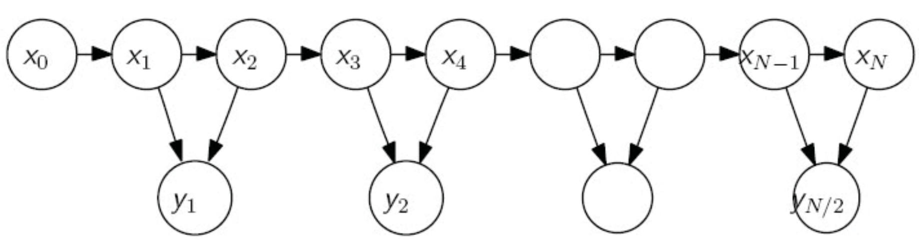
Furthermore, consider a second set of discrete random variables $y_n$, $n = 1, 2, ..., N/K$ of the process $x_n$ characterized by a distribution $q(\cdot|\cdot)$ where $K$ is an integer subsampling factor such that $N/K$ is an integer. Specifically,

$$\Pr[y_n = c \mid x_{nK} = b_0, x_{nK-1} = b_1, ..., x_{nK-K+1} = b_{K-1}] = q(c \mid b_0, b_1, ..., b_{K-1})$$

For parts (a) and (b), we restrict our attention to the case $K = 2$.

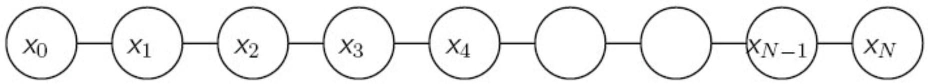
(a) Sketch the directed graph with nodes corresponding to the joint collection of random variables $x_0, x_1, ..., x_N, y_1, y_2, ..., y_{N/2}$.

**Solution:** Sketch of the directed graph:



(b) Suppose we observe $y_1, y_2, ..., y_{N/2}$. Sketch the undirected graph that represents the model for the remaining latent variables, i.e., the $(N+1)$-node undirected graph for the $x_0, x_1, ..., x_N$ alone. Identify all the maximal cliques in this graph and express the corresponding clique potentials in terms of $\pi(\cdot)$, $p(\cdot|\cdot)$ and $q(\cdot|\cdot)$.

**Solution:** Sketch of the undirected graph:




When the graph in part (a) is converted to an undirected graph, the dependencies that observing $y_1, y_2, ..., y_{N/2}$ may introduce in their two parent nodes are already present in the undirected graph, so we have a chain graph after removing the observed nodes.

More formally, the joint distribution $p_{x_0,...,x_N,y_1,...,y_{N/K}}$ factors as

$$p_{x_0,...,x_N,y_1,...,y_{N/K}} = \pi(x_0) \prod_{i=1}^{N} p(x_i|x_{i-1}) \prod_{i=1}^{N/K} q(y_i|x_{iK}, \ldots, x_{iK-K+1})$$

to which the conditional density $p_{x_0,...,x_N|y_1,...,y_{N/K}}$ is proportional, so the conditional density factors the same way.

The key point is that because $K = 2$, the $q(y_i|x_{iK}, \ldots, x_{iK-K+1})$ terms can all be absorbed into the appropriate factor containing $p(x_i|x_{i-1})$, so all factors are in terms of $x_{i-1}, x_i$ pairs ($i = 1, ..., N$), which are the maximal cliques, and the graph is a chain. The clique potentials break into several cases.

$$\psi_{0,1}(x_0, x_1) = \pi(x_0)p(x_1|x_0).$$

Next, for edges between $x_{2i-1}$ and $x_{2i}$, we get

$$\psi_{2i-1,2i} = p(x_{2i}|x_{2i-1})q(y_i|x_{2i}, x_{2i-1}).$$

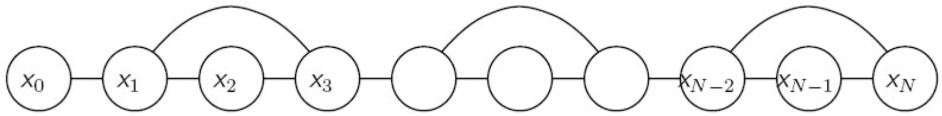
Finally, for edges between $x_{2i}$ and $x_{2i+1}$, we get

$$\psi_{2i,2i+1} = p(x_{2i+1}|x_{2i}).$$

(c) Now consider the case of $K = 3$. If the $y_1, y_2, ..., y_{N/3}$ are again observed, determine the size of the largest clique in the undirected graph for the remaining latent variables $x_0, x_1, ..., x_N$.

**Solution:** The size of the largest clique is 3.

When the directed graph is converted to an undirected graph this time, the observations now introduce extra dependencies beyond those between neighboring parent variables. We will show this for the general $K$.




From the joint distribution

$$p_{x_0,\ldots,x_N,y_1,\ldots,y_{N/K}} = \pi(x_0)\prod_{i=1}^{N} p(x_i|x_{i-1}) \prod_{i=1}^{N/K} q(y_i|x_{iK},\ldots,x_{iK-K+1}).$$

to which the conditional density $p_{x_0,\ldots,x_N|y_1,\ldots,y_{N/K}}$ is proportional, we see that we need to include potentials for the $p(x_i|x_{i-1})$ terms and for the $q(y_i|x_{iK},\ldots,x_{iK-K+1})$ terms. Using Hammersley-Clifford, we know that one way to construct a graph so that $p_{x_0,\ldots,x_N|y_1,\ldots,y_{N/K}}$ is Markov on the constructed graph is by including cliques for all the factors. In our case, this means that we include a clique (i.e., an edge) between $x_i$ and $x_{i-1}$. Also, we add cliques connecting all variables in the set $x_{iK},\ldots,x_{iK-K+1}$, since these variables appear together in a factor.

The maximum clique size is thus at least $K$. We only need to show that no larger cliques exists in the graph. Looking at the graph, this is obvious. To argue this formally, assume that $K \geq 2$. Then, to have a clique of size $K+1$, at the very least we need to have three adjacent vertices with degree at least $K$. However, all vertices in the graph have degree $K-1$, except for $x_1, x_K, x_{K+1}, x_{2K}, x_{2K+1}, \ldots$. These vertices all have degree $K$, but in the subgraph induced by $x_K, x_{K+1}, x_{2K}, x_{2K+1}, \ldots$, no three vertices are adjacent — $x_K$ is only connected to $x_{K+1}$, $x_{2K}$ is only connected to $x_{2K+1}$, and so on. Thus, the maximum clique size is exactly $K$.

6.438 Algorithms for Inference
Fall 2014