Review exercises on probabilistic graphical models
AMMI 2018-2019

These exercises are not meant to provide an exhaustive coverage of the material to review for the final exam. Also, all these exercises should not be taken as representative of the difficulty of the questions posed at the exam, although several questions of the exam are likely to have a similar level.

Factorization and exponential families

Let $G = (V, E)$ be a directed graph, Π_j the set of parents of node j in G and

$$p(x) = \prod_{j=1}^{d} p(x_j \mid x_{\Pi_j})$$

a probability distribution that factorizes in that graph on a discrete random variable $X = (X_1, X_2, \ldots, X_d)$. Assume that X_j takes values in $\{1, \ldots, K_j\}$. Write the distribution in exponential form.

For all x_j, we can write $\log(\mathbb{P}(X_j = x_j \mid X_{\Pi_j} = x_{\Pi_j}))$ as a vector of size $K_j \times \prod_{i \in \Pi_j} K_i$. We call this vector η_j. We denote by $\phi(x_j)$ the one-hot encoding of $(x_j = k, x_{\Pi_j} = \mathbf{K})$. We then have that $\langle \eta_j, \phi(x_j) \rangle = \log(\mathbb{P}(X_j = x_j \mid X_{\Pi_j} = x_{\Pi_j}))$. Concatenating all η_j into one vector η and doing the same with $\phi(x_j)$, we recover that $\langle \phi(x), \eta \rangle = \sum_j \log(\mathbb{P}(X_j = x_j \mid X_{\Pi_j} = x_{\Pi_j}))$.

Finally,

$$\exp(\langle \phi(x), \eta \rangle) = \exp(\sum_j \log(\mathbb{P}(X_j = x_j \mid X_{\Pi_j} = x_{\Pi_j}))) = \prod_j \mathbb{P}(X_j = x_j \mid X_{\Pi_j} = x_{\Pi_j}) = p(x)$$

Marginalization

Let $G = (V, E)$ be an undirected graph, with $|V| = d$. Let $p(x_V)$ be a distribution that factorizes according to that graph G. Consider the distribution $p(x_{V \setminus \{i\}}) = \sum_{x_i} p(x_V)$ on the $d-1$ nodes in $V \setminus \{i\}$.

- Assume that i had only a single neighbor. What is the undirected graph with nodes $V' = V \setminus \{i\}$ that has the smallest possible number of nodes and such that $p(x_{V \setminus \{i\}})$ factorizes according to that graph?

  We have:

$$p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}, i \in C} \psi_C(x_C) \prod_{C \in \mathcal{C}, i \notin C} \psi_C(x_C)$$

  Hence:

$$p(x_{V \setminus \{i\}}) = \frac{1}{Z} \left(\sum_{x_i} \prod_{C \in \mathcal{C}, i \in C} \psi_C(x_C) \right) \prod_{C \in \mathcal{C}, i \notin C} \psi_C(x_C)$$

  In the single neighbor case, we have that

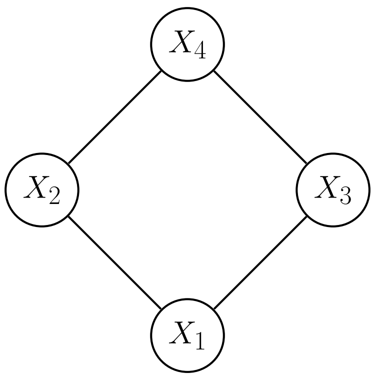
$$\{C \in \mathcal{C}, i \in C\} = \{(i, j), (i)\}$$

So $\sum_{x_i} \prod_{C\in\mathcal{C}, i\in C} \psi_C(x_C)$ is a function of x_j which can be integrated into the potential $\psi_j(x_j)$. The marginal probability is then factorized under the graph induced by G on V'.

- Answer the same question but for a general node i.
  We now have that $\sum_{x_i} \prod_{C\in\mathcal{C}, i\in C} \psi_C(x_C)$ is a function of all the neighbors of x_i (since if x_j is not a neighbor of x_i, it cannot be in a clique with x_i). So we add a potential $\psi_{\mathcal{N}(i)}(x_{\mathcal{N}(i)})$ and obtain that $p(x'_V)$ factorizes in the graph induced by G on V', where we added edges to put all of the neighbors of x_i in a clique.

Directed vs undirected graph

Consider the undirected graph $G = (V, E)$ with $V = \{1, 2, 3, 4\}$ and $E = \big\{\{1, 2\}, \{1, 3\}, \{2, 4\}, \{3, 4\}\big\}$. Let $\mathcal{L}(G)$ be the set of distributions that factorize according to G



1. Draw the graph

2. Prove that there exists no directed graph G' on 4 nodes such that $\mathcal{L}(G) = \mathcal{L}(G')$
   The original graph only implies the two following conditional independence relations :

   - $X_1 \perp\!\!\!\perp X_4 \ | \ X_{2,3}$ (1)
   - $X_2 \perp\!\!\!\perp X_3 \ | \ X_{1,4}$ (2)

   We will use in the following that for two undirected graphs G, G', $G \neq G' \implies \mathcal{L}(G) \neq \mathcal{L}(G')$.

   - $0, 1$ and 2 edges : the directed graphs are necessarily trees, hence they are equivalent to their undirected version, but these undirected versions are not equal to the graph under consideration.
   - 3 edges:
     - the undirected graph is a star, in which case, up to relabeling we can put X_1 in the center. Then X_1 and X_4 are not independent given the two others.
     - the undirected graph is a triangle, then one node is independent from the 3 others, which is not true in the original graph.
     - the undirected graph is a chain. Then its directed version has at most 1 v-structure. In this case, there is conditional independence of the two end nodes given one of the two in the middle, which is not true in our original graph.
   - 4 edges:
     - the undirected graph is a cycle. Then as in the case of the chain, presence of v-structure in the directed graph gives unwanted independence properties. If there are no v-structure, each node has at most 1 parent which leads to a cycle.
     - the undirected graph has a star : see previous case.

  - the undirected graph has a triangle: unwanted independence by conditioning on the connexion from the triangle to the last node.
- 5 and 6 edges:
  - There is necessarily a rectangle. See previous case

Gaussian Markov Chain

The following exercise does not require to do any tedious calculations. If you get into complicated calculations, it means you have the wrong approach...

Assume that the ε_i are i.i.d. with $\varepsilon_i \sim \mathcal{N}(0,1)$ and let $X_1 = \varepsilon_1,\ X_2 = \rho X_1 + \varepsilon_2,\ X_3 = \rho X_2 + \varepsilon_3$.

(a) What is the precision matrix of the joint distribution of (X_1, X_2, X_3)?

(b) Compute $\mathbb{E}[X_2 \mid X_1, X_3]$ and $\mathrm{Var}(X_2 \mid X_1, X_3)$.

The joint density is

$$\propto \exp -\frac{1}{2}\big[x_1^2 + (x_2 - \rho x_1)^2 + (x_3 - \rho x_2)^2\big],$$

which reveals by identification with

$$\exp(-\frac{1}{2}x^\top \Lambda x + \eta^\top x - A(\eta, \Lambda))$$

that

$$\Lambda = \begin{bmatrix} 1+\rho^2 & -\rho & 0 \\ -\rho & 1+\rho^2 & -\rho \\ 0 & -\rho & 1 \end{bmatrix} \quad \textit{and} \quad \eta = \mu = 0.$$

So that

$$Var(X_2 \mid X_1, X_3) = \Lambda_{22}^{-1} = \frac{1}{1+\rho^2}$$

and

$$\mathbb{E}[X_2 \mid X_1, X_3] = \eta_2 - \Lambda_{22}^{-1}\Lambda_{2,(1,3)}(X_1, X_3)^\top = \frac{\rho}{1+\rho^2}(X_1 + X_3).$$

The computation of the conditional mean variance can also be obtained from completing the square in the expression of the density:

$$p(x_2|x_1, x_3) \propto \exp -\frac{1}{2}\big[(1+\rho^2)x_2^2 - 2\rho x_1 x_2 - 2\rho x_3 x_2\big] \propto \exp -\frac{1}{2}(1+\rho^2)\big(x_2 - 2\tfrac{\rho}{1+\rho^2}(x_1 + x_3)\big)^2.$$