Course on probabilistic graphical models
Master MVA
Practice exercises 3

These exercises are not meant to provide an exhaustive coverage of the material to review for the final exam. To some extend they focus more specifically on material that is not covered in the homeworks. 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 style. Some exercises are easy, and a few can be much harder so don't be discouraged if you find some of them difficult. They are primarily designed to help you review and consolidate your understanding of the course.

Trees, entropies and polytopes

The main goal of this problem is to practice the connexion between canonical and moments parameters in the context trees, manipulate the entropy for trees and introduce some concepts like the marginal polytope and the local polytope which are relevant for variational inference methods.

In the entire problem, $[K]$ denotes the set $\{1, \dots, K\}$.

Preliminary questions

1. Show that if $X_1 \perp\!\!\!\perp X_3 \mid X_2$, then the joint distribution of (X_1, X_2, X_3) is such that

$$p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3)\,\frac{p(x_1, x_2)}{p(x_1)p(x_2)}\,\frac{p(x_2, x_3)}{p(x_2)p(x_3)}.$$

We have

$$p(x_1, x_2, x_3) = p(x_1, x_2)p(x_3|x_2) = \frac{p(x_1, x_2)p(x_2, x_3)}{p(x_2)},$$

hence the result.

2. Deduce from this formula that if $X_1 \perp\!\!\!\perp X_3 \mid X_2$, then

$$H(X_1, X_2, X_3) = H(X_1) + H(X_2) + H(X_3) - I(X_1, X_2) - I(X_2, X_3).$$

We have

$$H(X_1, X_2, X_3) = -\mathbb{E}[\log p(X_1, X_2, X_3)] = -\sum_{i=1}^{3}\mathbb{E}[\log p(X_i)] - \mathbb{E}\Big[\log \frac{p(X_1, X_2)}{p(X_1)p(X_2)}\Big] - \mathbb{E}\Big[\log \frac{p(X_2, X_3)}{p(X_2)p(X_3)}\Big],$$

which proves the result.

Part A

Consider in all this part a distribution $p(x_1, \dots, x_d)$ that factorizes according to an **undirected tree** $T = (V, E)$. Without loss of generality, we will assume that X_i is a binary indicator vector of dimension K, i.e. that X_i takes values in $\{0,1\}^K \cap \{x \mid \sum_{k=1}^K x_k = 1\}$.

1. Show that the joint distribution on $(X_1, \dots, X_d)$, i.e. $p(x_1, \dots, x_d)$, can be always be expressed as an explicit function of the marginal distributions $\big(p_i(x_i)\big)_{i \in V}$ and of the pairwise distributions $\big(p_{ij}(x_i, x_j)\big)_{\{i,j\} \in E}$ only. (Hint: Use the relation between undirected trees and directed trees and use the form of the factorization in a directed tree).

   See solution of question 2

2. Give an expression of $p(x_1, \dots, x_d)$ as a function of $(p_i)_{i \in V}$ and $(p_{ij}/(p_i p_j))_{\{i,j\} \in E}$ that does not rely on some arbitrary orientation of the tree. (Hint: generalize the formula for $p(x_1, x_2, x_3)$ in preliminary question 1 and prove that it is valid for all trees by induction).

   One can show that for a tree we always have

   $$p(x_1, \dots, x_n) = \prod_{i \in V} p(x_i) \prod_{\{i,j\} \in E} \frac{p(x_i, x_j)}{p(x_i) p(x_j)}.$$

   The proof is by induction on the number of leaves, by stripping one leaf at a time.

3. Deduce from the previous formula that the entropy $H(X_1, \dots, X_d)$ for a distribution that factorizes according to a tree can be expressed as a function of $H(X_i)$ for all $i \in V$ and of all $I(X_i, X_j)$ for all $\{i, j\} \in E$.

   Taking the log and then expectation of the previous formula we get

   $$H(X_1, \dots, X_d) = \sum_{i \in V} H(X_i) - \sum_{\{i,j\} \in E} I(X_i, X_j).$$

4. Show that it can alternatively be expressed as a function of $H(X_i)$ for $i \in V$, $H(X_i, X_j)$ for $\{i, j\} \in E$ and of the degrees d_i of all nodes.

   Using that $I(X_i, X_j) = H(X_i) + H(X_j) - H(X_i, X_j)$ we get

   $$H(X_1, \dots, X_d) = \sum_{i \in V} (1 - d_i) H(X_i) + \sum_{\{i,j\} \in E} H(X_i, X_j),$$

   where d_i is the degree of node i.

5. If the distribution $p(x_1, \dots, x_d) > 0$, show that it can be written in exponential family form. What are the elements of the vector $\phi(x)$ of sufficient statistics? What are the components of the vector η of natural parameters? (Hint: introduce indicator variables).

   $$\log p(x_1, \dots, x_n) = \sum_{i \in V} \sum_{k \in [K]} \eta_{ik} x_{ik} + \sum_{\{i,j\} \in E} \sum_{k,l \in [K]} \eta_{ijkl} x_{ik} x_{jl} - A(\eta).$$

6. What are the components of the vector μ of moment parameters in the exponential family of the previous question?

   The vector of moment parameters is formed of the

   $$\mu_{ik} = \mathbb{P}(X_{ik} = 1) \quad \textit{and the} \quad \mu_{ijkl} = \mathbb{P}(X_{ik} = 1, X_{jl} = 1),$$

   respectively for all $i \in V, k \in [K]$ and for all $\{i, j\} \in E$ and $k, l \in [K]$.

7. Show that given a vector of moment parameters μ, the corresponding vector of natural parameters η can be expressed in closed form as a simple function of μ. Explicit the function.

   Given the answer to question 2, we can write the same distribution as

$$p(x_1, \ldots, x_n) = \left(\prod_{i \in V} \prod_{k \in [K]} \mu_{ik}^{x_{ik}}\right)\left(\prod_{\{i,j\} \in E} \sum_{k,l \in [K]} \left(\frac{\mu_{ijkl}}{\mu_{ik}\mu_{jl}}\right)^{x_{ik}x_{jl}}\right).$$

   So that we can identify

$$\eta_{ik} = \log \mu_{ik} \qquad \text{and} \qquad \eta_{ijkl} = \log \frac{\mu_{ijkl}}{\mu_{ik}\mu_{jl}}.$$

8. Show that the entropy $H(X_1, \ldots, X_d)$ can be expressed as a function of μ only. Explicit that function.

$$\widetilde{H}(\mu) = -\sum_{i \in V} \sum_{k \in [K]} \mu_{ik} \log \mu_{ik} - \sum_{\{i,j\} \in E} \sum_{k,l \in [K]} \mu_{ijkl} \log \frac{\mu_{ijkl}}{\mu_{ik}\mu_{jl}}.$$

Part B

We consider the graphical model $\mathcal{P}_G$ consisting of all distributions over $\mathcal{X} := \mathcal{X}_1 \times \ldots \times \mathcal{X}_d$ with $|\mathcal{X}_i| = K_i < \infty$ that factorize according to some undirected graph $G = (V, E)$. For convenience, we assume that $K_i = K$ for all i and identify $\mathcal{X}_i$ with a copy of $\{0,1\}^K \cap \{x \mid \sum_{k=1}^K x_k = 1\}$.

Consider the set $\mathcal{H}_G$ defined by

$$\mathcal{H}_G := \{\mu = \big((\mu_i)_{i \in V}, (\mu_{ij})_{\{i,j\} \in E}\big)^\top \mid \mu_i \in [0,1]^K, \mu_{ij} \in [0,1]^{K \times K}, \sum_{k=1}^K [\mu_i]_k = 1, \sum_{1 \le k,l \le K} [\mu_{ij}]_{kl} = 1\}.$$

The *local polytope* of $\mathcal{P}_G$ is the polytope $\mathcal{L}_G$ defined by

$$\mathcal{L}_G := \Big\{\mu \in \mathcal{H}_G \mid \forall \{i,j\} \in E, \sum_{l=1}^K [\mu_{ij}]_{kl} = [\mu_i]_k, \sum_{k=1}^K [\mu_{ij}]_{kl} = [\mu_j]_l\Big\}.$$

The *moment polytope* (or *marginal polytope*) is the set $\mathcal{M}_G$ defined by

$$\mathcal{M}_G = \{\mu \in \mathbb{R}^{Kd + K^2(d-1)} \mid \exists p \in \triangle_{\mathcal{X}}, \quad \mathbb{E}_p[\phi(X)] = \mu\},$$

where $\triangle_{\mathcal{X}}$ is the set of all probability distributions over $\mathcal{X}$, and

$$\phi(x) = \big((x_{ik})_{i \in V, k \in [K]}, (x_{ik}x_{jl})_{\{i,j\} \in E,\, k,l \in [K]}\big)^\top.$$

1. Show that for any graph G, we have $\mathcal{M}_G \subset \mathcal{L}_G$.

   Any $\mu \in \mathcal{M}_G$ must satisfy the marginalization constraints of $\mathcal{H}_G$ and of $\mathcal{L}_G$, since it is a consistent set of marginal distributions, which shows the result.

2. If $G = T$ is a tree, then use the results of part A to show that for any element $\tau \in \mathcal{L}_T$ such that all components of τ are strictly positive, it is possible to construct a distribution in $\mathcal{P}_T$ whose moments are equal to τ.

   Set $\mu = \tau$ in the first expression of the solution of question 7. This shows that there exists a tree distribution with exactly τ as a vector of moments.

3. Use the result of the previous question to show that for a tree T, we must have $\mathcal{M}_T = \mathcal{L}_T$.

   The moments that are strictly positive are dense in the set of all moments, and so the inclusion

$$\mathcal{L}_T \subset \mathcal{M}_T$$

   follows from the previous question by closure.