Solutions for Week 4

Bruno Loureiro and Luca Saglietti

25.03.2021

1 Exercise 4.1

Write the following problems (a) in terms of a probability distribution and (b) in terms of a graphical model by drawing a (small) example of the corresponding factor graph.

(1) **p-spin model**

One model that is commonly studied in physics is the so-called Ising 3-spin model. The Hamiltonian of this model is written as

$$\mathcal{H}(\{S_i\}_{i=1}^N) = -\sum_{(ijk)\in E} J_{ijk} S_i S_j S_k - \sum_{i=1}^N h_i S_i \tag{1.1}$$

where E is a given set of (unordered) triplets $i \neq j \neq k$, J_{ijk} is the interaction strength for the triplet $(ijk) \in E$, and h_i is a magnetic field on spin i. The spins are Ising, which in physics means $S_i \in \{+1, -1\}$.

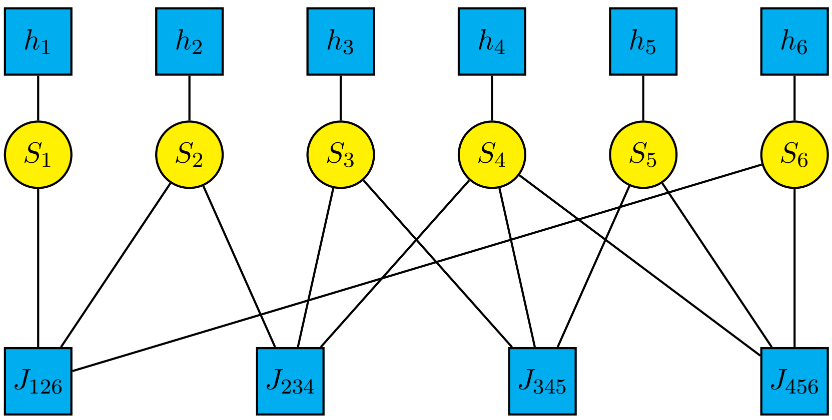
Solution: Since the Hamiltonian is given, in this case the probability distribution is simply given by the Boltzmann distribution:

$$\mathbb{P}_\beta(\boldsymbol{S} = \boldsymbol{s}) = \frac{1}{\mathcal{Z}_\beta} e^{-\beta\mathcal{H}(\boldsymbol{s})} = \prod_{i=1}^N e^{\beta h_i s_i} \prod_{(ijk)\in E} e^{\beta J_{ijk} s_i s_j s_k} \tag{1.2}$$

Note we have two flavours of factor nodes in this model: first, a local interaction term for each spin variable s_i, $i \in \{1, \cdots, N\}$; second, a three-body interaction term which couple triplets $(ijk) \in E$ of spins. In the notation of the lectures:

$$g_i(s_i) = e^{\beta h_i s_i}, \qquad f_{(ijk)}(s_i, s_j, s_k) = e^{\beta J_{ijk} s_i s_j s_k} \tag{1.3}$$

Now let's explicitly draw an example of factor graph. Consider for instance $N = 6$ spins and the triplet set $E = \{(126), (234), (345), (456)\}$. Its factor graph is given by:



(2) **Independent set problem**

Independent set is a problem defined and studied in combinatorics and graph theory. Given a (unweighted, undirected) graph $G(V, E)$, an independent set $S \subseteq V$ is defined as a subset of nodes such that if $i \in S$ then for all $j \in \partial i$ we have $j \notin S$. In other words in for all $(ij) \in E$ only i or j can belong to the independent set.

(a) Write a probability distribution that is uniform over all independent sets on a given graph.

Solution: Let $N = |V|$ denote the number of nodes in G. One way of parametrising a subset of nodes $S \subset V$ is to assign to every node $i = 1, \cdots, N$ a binary variable:

$$\sigma_i^S = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases} . \tag{1.4}$$

which indicates whether node i belongs to S. Similarly, to every edge $(ij) \in E$, define a function:

$$f_{(ij)}(\sigma_i^S, \sigma_j^S) = \mathbb{I}\left((\sigma_i^S, \sigma_j^S) \neq (1,1)\right). \tag{1.5}$$

Or in words: $f_{(ij)}$ is one if at most one of the nodes i, j connected by the edge (ij) belong to S. With these two definitions, we can characterise an independent subset $S \subset V$ as:

$$S \text{ is independent} \qquad \Leftrightarrow \qquad \text{for all distinct } i, j \in S, \qquad f_{(ij)}(\sigma_i, \sigma_j) = 1 \tag{1.6}$$

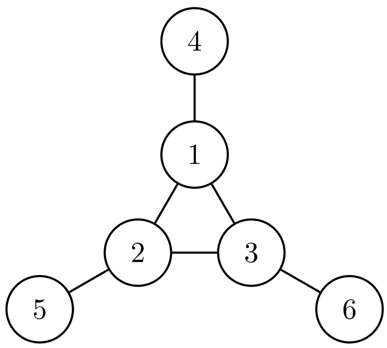
Or in words: an independent set is such that none of the nodes belonging to it are connected by an edge of the graph. For a given set of nodes $\boldsymbol{\sigma} \in \{0,1\}^N$ the uniform probability measure over independent sets is given by:

$$\mathbb{P}(\boldsymbol{\sigma}) = \frac{1}{\mathcal{Z}_G} \prod_{(ij)\in E} \mathbb{I}\left((\sigma_i, \sigma_j) \neq (1,1)\right) \tag{1.7}$$

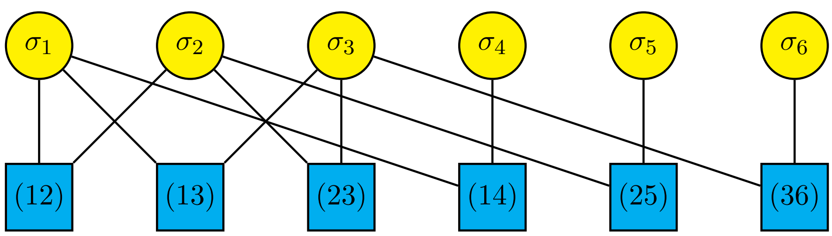
where:

$$\mathcal{Z}_G = \sum_{\boldsymbol{\sigma}\in\{0,1\}^N} \prod_{(ij)\in E} \mathbb{I}\left((\sigma_i, \sigma_j) \neq (1,1)\right) = \text{number of independent sets in } G \tag{1.8}$$

As an example, consider the following graph G:



The factor graph associated with the independent set measure is given by:



(b) Write a probability distribution that gives a larger weight to larger independent sets, where the size of an independent set is simply $|S|$.

Solution: Note that the size of a set $|S|$ can be expressed in terms of the variables σ_i^S as:

$$|S| = \sum_{i=1}^{N} \sigma_i^S \tag{1.9}$$

To assign a larger weight to independent sets which are larger, we just need to multiply our density by any positive increasing function $g(|S|)$:

$$\mathbb{P}(\boldsymbol{\sigma}) = \frac{1}{\tilde{\mathcal{Z}}_G} g\left(\sum_{i=1}^{N} \sigma_i\right) \prod_{(ij)\in E} \mathbb{I}\left((\sigma_i, \sigma_j) \neq (1,1)\right) \tag{1.10}$$

For example, we can choose $g(x) = e^{hx}$ for $h > 0$ to get:

$$\mathbb{P}(\boldsymbol{\sigma}) = \frac{1}{\tilde{\mathcal{Z}}_G} \prod_{i=1}^{N} e^{h\sigma_i} \prod_{(ij)\in E} \mathbb{I}\left((\sigma_i, \sigma_j) \neq (1,1)\right) \tag{1.11}$$

Note that this would introduce a local factor node to variable node in the factor graph.

(3) **Matching problem**

Matching is another classical problem of graph theory. It is related to a dimer problem in statistical physics. Given a (unweighted, undirected) graph $G(V, E)$ a matching $M \subseteq E$ is defined as a subset of edges such that if $(ij) \in M$ then no other edge that contains node i or j can be in M. In other words a matching is a subset of edges such that no two edges of the set share a node.

(a) Write a probability distribution that is uniform over all matchings on a given graph.

Solution: The construction of the factor graph for matching is very similar to the one for the independent set, with the crucial difference that the variable nodes are the edges of G, instead of the nodes. As before, we start by assigning a binary variable to each edge of G which identifies whether it belongs or not to M:

$$s_{(ij)} = \begin{cases} 1 & \text{if } (ij) \in M \\ 0 & \text{otherwise} \end{cases} \tag{1.12}$$

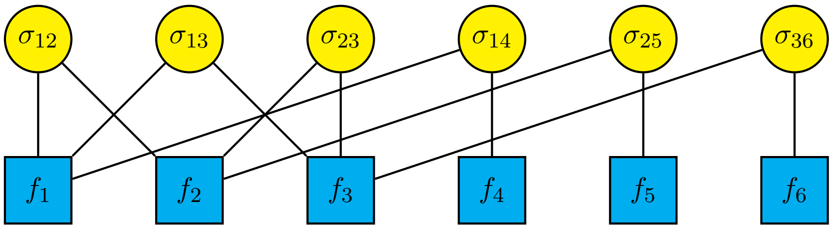
Let $N = |V|$. As before, for every node $i = 1, \cdots, N$ we assign a function which is zero if the node is attached to two edges belonging to M:

$$f_i\left(\{s_{(ij)}\}_{j\in\partial i}\right) = \mathbb{I}\left(\sum_{j\in\partial i} s_{(ij)} \leq 1\right) \tag{1.13}$$

Note that with this definition we allow for nodes to be unpaired. If we would like only perfect matchings (i.e. when all eges are paired), we would impose equality. The uniform measure over all matchings in G can therefore be written as:

$$\mathbb{P}(\boldsymbol{s}) = \frac{1}{\mathcal{Z}_G} \prod_{i=1}^{N} \mathbb{I}\left(\sum_{j\in\partial i} s_{(ij)} \leq 1\right) \tag{1.14}$$

where the partition function $\mathcal{Z}_G$ counts the total number of matching sets M in G. Note that different from (2), $\boldsymbol{s} \in \mathbb{R}^{|E|}$. Note that the uniform measure assigns the same weight to large matchings (i.e. when as many edges as possible are matched) and smaller matchings (e.g. when only half of the edges are matched) . To illustrate the factor graph of the matching problem, consider the same graph as in problem (2). The associated factor graph is given by:



(b) Write a probability distribution that that gives a larger weight to larger matchings, where the size of a matching is simply $|M|$.

Solution: As before, we can write the size of a matching set as a function of $\boldsymbol{s}$:

$$|M| = \sum_{(ij)\in E} s_{(ij)} \tag{1.15}$$

Therefore, to assign a bigger weight to larger matchings, we just need to multiply the measure by any positive increasing function $g(|M|)$:

$$\mathbb{P}(\boldsymbol{s}) = \frac{1}{\mathcal{Z}_G} g\left(\sum_{(ij)\in E} s_{(ij)}\right) \prod_{i=1}^{N} \mathbb{I}\left(\sum_{j\in\partial i} s_{(ij)} \leq 1\right) \tag{1.16}$$

This is the softer way to encourage a perfect matching than to impose equality at the factor function f_i.

2 Exercise 4.2:

Show that the BP equations we derived in the lecture

$$\chi^{j\to a}_{s_j} = \frac{1}{Z^{j\to a}} g_j(s_j) \prod_{b\in\partial j\setminus a} \psi^{b\to j}_{s_j} \tag{2.1}$$

$$\psi^{a\to i}_{s_i} = \frac{1}{Z^{a\to i}} \sum_{\{s_j\}_{j\in\partial a\setminus i}} f_a\left(\{s_j\}_{j\in\partial a}\right) \prod_{j\in\partial a\setminus i} \chi^{j\to a}_{s_j} \tag{2.2}$$

are stationarity conditions of the Bethe free entropy (†) under the constraint that both $\sum_s \psi^{a\to i}_s = 1$ and $\sum_s \chi^{i\to a}_s = 1$ for all $(ia) \in E$.

Solution: Recall that the Bethe free entropy is given by:

$$
\begin{aligned}
N\Phi_{\text{Bethe}} &= \log Z = \sum_{i=1}^{N} \log Z^i + \sum_{a=1}^{M} \log Z^a - \sum_{(ia)} \log Z^{ia} \\
Z^i &\equiv \sum_s g_i(s) \prod_{a\in\partial i} \psi_s^{a\to i} \\
Z^a &\equiv \sum_{\{s_i\}_{i\in\partial a}} f_a\left(\{s_i\}_{i\in\partial a}\right) \prod_{i\in\partial a} \chi_{s_i}^{i\to a} \\
Z^{ia} &\equiv \sum_s \chi_s^{i\to a} \psi_s^{a\to i}
\end{aligned}
\tag{2.3}
$$

To show that the BP equations can be obtained as a stationary condition from the Bethe free entropy, we need to show that the conditions:

$$
\frac{\partial \Phi_{\text{Bethe}}}{\partial \chi_{s_j}^{j\to a}} \overset{!}{=} 0, \qquad \frac{\partial \Phi_{\text{Bethe}}}{\partial \psi_{s_i}^{a\to i}} \overset{!}{=} 0 \tag{2.4}
$$

lead to the BP equations. First, note that since the messages are independent, we have:

$$
\frac{\partial \chi_{s_j}^{j\to a}}{\partial \psi_{s_i}^{b\to i}} = 0, \qquad \frac{\partial \chi_{s_j}^{j\to a}}{\partial \chi_{s_i}^{i\to b}} = \delta_{ab}\delta_{ij}, \qquad \frac{\partial \psi_{s_j}^{a\to j}}{\partial \psi_{s_i}^{b\to i}} = \delta_{ab}\delta_{ij} \tag{2.5}
$$

for any variable $i, j \in V$ and factor $a, b \in F$ nodes. Therefore, the derivative of the partition functions with respect to the messages are given by:

$$
\frac{\partial Z^i}{\partial \chi_{s_j}^{j\to a}} = 0, \qquad \frac{\partial Z^i}{\partial \psi_{s_j}^{a\to j}} = \delta_{ij}\; g_j(s_j) \prod_{c\in\partial j\setminus a} \psi_{s_j}^{c\to j}
$$

$$
\frac{\partial Z^a}{\partial \chi_{s_j}^{j\to b}} = \delta_{ab} \sum_{\{s_k\}_{k\in\partial a\setminus j}} f_a\left(\{s_k\}_{k\in\partial a\setminus j}\right) \prod_{i\in\partial a\setminus j} \chi_{s_i}^{i\to a}, \qquad \frac{\partial Z^a}{\partial \psi_{s_j}^{b\to j}} = 0
$$

$$
\frac{\partial Z^{ai}}{\partial \chi_{s_j}^{j\to b}} = \delta_{ab}\delta_{ij}\; \psi_{s_i}^{a\to i}, \qquad \frac{\partial Z^{ai}}{\partial \psi_{s_j}^{b\to j}} = \delta_{ab}\delta_{ij}\; \chi_{s_i}^{i\to a}
$$

Therefore, taking the derivative of the Bethe free entropy with respect to $\chi_{s_j}^{j\to a}$:

$$
\begin{aligned}
N\frac{\partial \Phi_{\text{Bethe}}}{\partial \chi_{s_j}^{j\to a}} &\overset{(a)}{=} \frac{1}{\partial \chi_{s_j}^{j\to a}} \log Z^a - \frac{1}{\partial \chi_{s_j}^{j\to a}} \log Z^{ja} \\
&= \frac{1}{Z^a} \sum_{\{s_k\}_{k\in\partial a\setminus j}} f_a\left(\{s_k\}_{k\in\partial a\setminus j}\right) \prod_{i\in\partial a\setminus j} \chi_{s_i}^{i\to a} - \frac{1}{Z^{ja}} \psi_{s_j}^{a\to j}
\end{aligned}
\tag{2.6}
$$

where in (a) we used the fact that only the i and (ia) factors of the sum contribute. Setting this to zero using the normalisation condition to show that $Z^{j\to a} = Z^{ia}/Z^a$ give us the update equation for $\psi^{a\to j}$:

$$
\psi_{s_j}^{a\to j} = \frac{1}{Z^{a\to j}} \sum_{\{s_k\}_{k\in\partial a\setminus j}} f_a\left(\{s_k\}_{k\in\partial a\setminus j}\right) \prod_{i\in\partial a\setminus j} \chi_{s_i}^{i\to a} \tag{2.7}
$$

Similarly, the derivative of the Bethe free entropy with respect to ${\psi s_j}^{a\to j}$ reads:

$$
\begin{aligned}
N\frac{\partial\Phi_{\text{Bethe}}}{\partial\psi_{s_j}^{a\to j}} &\overset{(a)}{=} \frac{1}{\partial\psi_{s_j}^{a\to j}}\log Z^j - \frac{1}{\partial\psi_{s_j}^{a\to j}}\log Z^{ja} \\
&= \frac{1}{Z^j}g_j(s_j)\prod_{c\in\partial j\setminus a}\psi_{s_j}^{c\to j} - \frac{1}{Z^{ja}}\chi_{s_j}^{j\to a}
\end{aligned}
\tag{2.8}
$$

Setting this to zero and using the normalisation condition for $\chi_{s_j}^{j\to a}$ leads to the BP equation for $\chi_{s_j}^{j\to a}$:

$$
\chi_{s_j}^{j\to a} = \frac{1}{Z^{j\to a}}g_j(s_j)\prod_{c\in\partial j\setminus a}\psi_{s_j}^{c\to j} \tag{2.9}
$$