Stat 316 Stochastic Processes on Graphs

Bethe-Peierls approximation: an informal introduction

Andrea Montanari *Lecture 11 - 29/10/2007*

Bethe-Peierls approximation [Bet35] reduces the problem of computing partition functions and expectation values to the one of solving a set of non-linear equations. In general, this 'reduction' involves an uncontrolled error. We are interested here in mean-field models, for which Bethe-Peierls can be proved to be asymptotically exact in the large system limit. In fact we already saw an example of this type, namely the ferromagnetic Ising model on random regular graphs.

In the context of mean field spin glasses, the method was further refined by Mézard, Parisi and Virasoro to include 'replica symmetry breaking' effects. In such applications, it is referred to as the 'cavity method.' A closely related approach is provided by the so-called TAP (Thouless-Anderson-Palmer) equations [MPV87].

This lecture focuses on general ideas, while next will discuss a more formal justification.

1 Bethe equations

We consider a general model on the graph $G = (V, E)$ with variables $x_i \in \mathcal{X}$, and a distribution

$$\mu(x) = \frac{1}{Z} \prod_{(ij)\in E} \psi_{ij}(x_i, x_j)\,. \tag{1}$$

Given $(i, j) \in E$, define $\mu^{(ij)}(\,\cdot\,)$ to be the modified distribution whereby the contribution of edge (i, j) has been 'taken out.' Explicitly

$$\mu^{(ij)}(x) \equiv \frac{1}{Z^{(ij)}} \prod_{(kl)\in E\setminus(ij)} \psi_{kl}(x_k, x_l)\,. \tag{2}$$

Further, for $i \in V$ we let $\mu^{(i)}(\,\cdot\,)$ be the modified measure in which all the edges incident on i have been removed:

$$\mu^{(i)}(x) \equiv \frac{1}{Z^{(i)}} \prod_{(kl)\in E,\ i\notin(kl)} \psi_{kl}(x_k, x_l)\,. \tag{3}$$

Given $U \subseteq V$, I shall denote by μ_U (respectively, by $\mu_U^{(ij)}$, $\mu_U^{(i)}$) the marginal distribution of $x_U \equiv \{x_i : i \in U\}$ when x is distributed according to μ (respectively $\mu^{(ij)}$, $\mu^{(i)}$).

The first step consists in deriving a set of exact equations relating the marginals of $\mu^{(i)}$ to the ones of $\mu^{(ij)}$. In order to write such equations, it is convenient to adopt the following notation. Whenever f and g are two non-negative functions on the same domain, I'll write $f(x) \cong g(x)$ if they differ by an overall normalization. We then have

$$\mu_{ij}^{(ij)}(x_i, x_j) \cong \sum_{x_{\partial i\setminus i,j}} \mu_{\partial i}^{(i)}(x_{\partial i}) \prod_{l\in\partial i\setminus j} \psi_{il}(x_i, x_l)\,. \tag{4}$$

Of course this does not solve the problem because we have more variables than unknowns. The Bethe-Peierls method consists in writing a set of mean field equations for the 'modified marginals'

$$\nu_{i\to j}(x_i) \equiv \mu_i^{(ij)}(x_i)\,. \tag{5}$$

We then have

$$\mu_{ij}^{(ij)}(x_i, x_j) = \nu_{i\to j}(x_i)\nu_{j\to i}(x_j)\,. \tag{6}$$

The crucial approximation consists in assuming that

$$\mu^{(i)}(x_{\partial i}) = \prod_{l\in\partial i} \nu_{l\to i}(x_l) + \mathsf{ERR}\,. \tag{7}$$

where ERR is an error term that is assumed to be small.

Bethe-Peierls equations are obtained by plugging the last expressions in Eq. (4) and neglecting the error term. If $\nu_{j\to i}(x_j) > 0$, we get

$$\nu_{i\to j}(x_i) \cong \prod_{l\in\partial i\setminus j} \sum_{x_l\in\mathcal{X}} \psi_{il}(x_i,x_l)\nu_{l\to i}(x_l)\,. \tag{8}$$

Bethe-Peierls method consists in solving these equations for the 'messages' (or 'cavity fields') $\{\nu_{i\to j}\}$ and then using them for estimating the marginals of μ. For instance

$$\mu_i(x_i) \cong \prod_{j\in\partial i} \sum_{x_j} \psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\,. \tag{9}$$

Exercise 1: Assume G to be a tree. Prove that the error term in Eq. (7) vanishes in this case, and that Bethe-Peierls equations have a unique solution corresponding to the actual marginals of μ.

2 Bethe free entropy

Since within Bethe approximation all marginals can be expressed in terms of the messages $\{\nu_{i\to j}\}$ that solve Bethe equations (8), it is perhaps not surprising that the free entropy can be expressed in terms of the same messages as well. It is more surprising that a simple expression exists

$$\Phi\{\nu\} = -\sum_{(ij)\in G} \log\Big\{\sum_{x_i,x_j}\psi_{ij}(x_i,x_j)\nu_{i\to j}(x_i)\nu_{j\to i}(x_j)\Big\} + \sum_{i\in V}\log\Big\{\sum_{x_i}\prod_{j\in\partial i}\sum_{x_j}\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\Big\}\,. \tag{10}$$

In general we shall regard this as a function from the set of possible messages to reals $\Phi : \{\nu_{i\to j}\} \to \mathbb{R}$. This function is called the *Bethe free entropy*.

It is easy to prove that this expression is exact if G is a tree. More precisely, if G is a tree and ν^* is the unique solution of the Bethe equations (8), then $\log Z = \Phi\{\nu^*\}$. There are many ways of proving this fact. A simple one consists in progressively disconnecting G in a recursive fashion.

We start form a simple remark. If $f_a(x) \cong p(x)$, $a \in \{1,2,3\}$ and some probability distribution p, then

$$\log\left\{\sum_x \frac{f_1(x)f_2(x)}{f_3(x)}\right\} = \log\left\{\sum_x f_1(x)\right\} + \log\left\{\sum_x f_2(x)\right\} - \log\left\{\sum_x f_3(x)\right\}\,. \tag{11}$$

Let us then describe the first step of the recursion. Let $(ij)\in E$ be an edge in G. Denote by $Z_{i\to j}(x_i)$ the constrained partition function for the subtree rooted at i and not including j, whereby we force x_i to take the value in argument. Then we obviously have

$$Z = \sum_{x_i,x_j} Z_{i\to j}(x_i)\psi_{ij}(x_i,x_j)Z_{j\to i}(x_j) \tag{12}$$

$$= \sum_{x_i,x_j} \frac{\big\{Z_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\big\}\big\{\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)Z_{j\to i}(x_j)\big\}}{\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)}\,. \tag{13}$$

It is easy to see that, if G is a tree, $Z_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j) \cong \mu_{ij}(x_i,x_j)$, $\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)Z_{j\to i}(x_j) \cong \mu_{ij}(x_i,x_j)$, $\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j) \cong \mu_{ij}(x_i,x_j)$. By the above simple remark

$$\log Z = \log\Big\{\sum_{x_i,x_j} Z_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\Big\} + \log\Big\{\sum_{x_i,x_j}\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)Z_{j\to i}(x_j)\Big\}$$

$$-\log\Big\{\sum_{x_i,x_j}\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\Big\}\,. \tag{14}$$

At this point we can interpret the first two terms as log partition functions for models of reduced size. The associated graphs are subtrees rooted at i and j.

If we repeat recursively this operation, we end up with a term of the form $-\log\{\sum_{x_i,x_j} \nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)\nu_{j\to i}(x_j)\}$ for each edge. The residual system is composed of disconnected 'stars' centered at the vertices of G. The corresponding log-partition function is given by the second sum in Eq. (10).

One important property of the Bethe free entropy is that its stationary points are solutions of the Bethe equations (8). This is proved by simple calculus

$$\frac{\partial \Phi\{\nu\}}{\partial \nu_{j\to i}(x_j)} = -\frac{\sum_{x_i} \nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)}{\sum_{x_i',x_j'} \nu_{i\to j}(x_i')\nu_{j\to i}(x_j')\psi_{ij}(x_i',x_j')} + \frac{\sum_{x_i} \mathsf{T}\nu_{i\to j}(x_i)\psi_{ij}(x_i,x_j)}{\sum_{x_i',x_j'} \mathsf{T}\nu_{i\to j}(x_i')\nu_{j\to i}(x_j')\psi_{ij}(x_i',x_j')} \tag{15}$$

where we defined

$$\mathsf{T}\nu_{i\to j}(x_i) \equiv \frac{1}{z_{i\to j}} \prod_{l\in\partial i\setminus j} \sum_{x_l} \psi_{il}(x_i,x_l)\nu_{l\to i}(x_l)\,. \tag{16}$$

Exercise 2: Prove that this implies that Bethe equations have at least one solution.

3 Example: Ising models

As an illustration, we consider again the Ising model, that gives a distribution over $x \in \{+1,-1\}^V$, of the form

$$\mu(x) = \frac{1}{Z}\exp\Big\{\beta \sum_{(ij)\in E} J_{ij}x_ix_j\Big\}\,. \tag{17}$$

Here we don't assume any more ferromagnetic interactions. The 'couplings' $\{J_{ij}\}$ are generic real numbers. The messages $\nu_{i\to j}$ can then be encoded effectively through the following log-likelihood ratios ('cavity fields')

$$h_{i\to j} = \frac{1}{2}\log\frac{\nu_{i\to j}(+1)}{\nu_{i\to j}(-1)}\,. \tag{18}$$

The Bethe equations then reduce to the ones below

$$h_{i\to j} = \sum_{l\in\partial i\setminus j} \text{atanh}\,\{\tanh\beta J_{il}\tanh h_{l\to i}\}\,. \tag{19}$$

In terms of the cavity fields $\{h_{i\to j}\}$, we can compute the local magnetizations $m_i = \mathbb{E}\{x_i\}$ using Eq. (9):

$$m_i = \tanh\left\{\sum_{l\in\partial i} \text{atanh}\,\{\tanh\beta J_{il}\tanh h_{l\to i}\}\right\}\,. \tag{20}$$

As a particular case, let us reconsider a regular graph with $J_{ij} = +1$ for each edge $(ij)\in E$. Then these equations admit a solution of the form $h_{i\to j} = h$, where h solves

$$h = (k-1)\text{atanh}\,\{\tanh\beta J_{il}\tanh h\}\,. \tag{21}$$

For the ferromagnetic Ising model on a random regular graph, we indeed proved that the marginals can be computed by solving this equation.

Exercise 3: Write the Bethe free entropy (10) in the case of Ising models. Show that, in the case of regular graphs, under the assumption $h_{i\to j} = h$, it reduces to the form we proved a few lectures ago.

Exercise 4: Exhibit an example of regular graph G, with degree larger than 2, such that the associated ferromagnetic Ising model does not have phase transitions. What happens within Bethe-Peierls approximation?

4 Fully connected limit

When G is the complete graph, Bethe equations can often be simplified. Here we'll consider two such examples: the Curie-Weiss model (that we already considered at the beginning of the course) and the Sherrington-Kirkpatrick model.

4.1 Curie-Weiss model

We already encountered the Curie-Weiss model. This is defined by the general form (17) whereby G is the complete graph and $J_{ij} = 1/N$ for all (i, j). It follows from Eq. (19) that

$$h_{i\to j} = \frac{\beta}{N} \sum_{l\in V\setminus i,j} \tanh h_{l\to i} + O(1/N)\,. \tag{22}$$

Therefore $h_{i\to j} = h + O(1/N)$ where h solves

$$h = \beta \tanh h\,. \tag{23}$$

By Eq. (9), the local magnetization is given by $m = \tanh h$ and thus solves the equation

$$m = \tanh \beta m\,, \tag{24}$$

which we proved at the beginning of the course.

4.2 Sherrington-Kirkpatrick model

In the Sherrington-Kirkpatrick model, G is the complete graph and the couplings are $J_{ij} = J'_{ij}/\sqrt{N}$ with J'_{ij} iid normal random variables of mean 0 and variance 1. Hereafter I'll drop the prime.

By expanding again Eq. (19), we get

$$h_{i\to j} = \sum_{l\in V\setminus i,j} \frac{\beta J_{il}}{\sqrt{N}} \tanh h_{l\to i} + O(N^{-1/2})\,. \tag{25}$$

The relation to local magnetizations is not as simple as for the Curie-Weiss model. Expanding Eq. (9), we get

$$\begin{aligned} \operatorname{atanh} m_i &= \sum_{l\in V\setminus i,j} \frac{\beta J_{il}}{\sqrt{N}} \tanh h_{l\to i} + O(N^{-1/2}) = && (26)\\ &= h_{i\to j} + \frac{\beta J_{ij}}{\sqrt{N}} m_j + O(N^{-1/2})\,. && (27)\end{aligned}$$

Substituting in (both sides of!) Eq. (25), and neglecting terms of order $N^{-1/2}$, we get the so-called TAP equations

$$\operatorname{atanh} m_i = \frac{\beta}{\sqrt{N}} \sum_{l\in V\setminus i} J_{il} m_l - m_i \sum_{l\in V\setminus i} \frac{\beta^2 J_{il}^2}{N}(1 - m_l^2)\,. \tag{28}$$

References

[Bet35] Hans A. Bethe. Statistical theory of superlattices. *Proc. Roy. Soc. London A*, 150:552–558, 1935.

[MPV87] Marc Mézard, Giorgio Parisi, and Miguel A. Virasoro. *Spin Glass Theory and Beyond*. World Scientific, Singapore, 1987.