

# Linear Response Formula and Generalized Belief Propagation for Probabilistic Inference

Kazuyuki Tanaka <sup>1</sup>

Graduate School of Information Sciences, Tohoku University,  
Aramaki-aza-aoba 09, Aoba-ku, Sendai 980-8579, Japan

Linear response formulas for the generalized belief propagation in approximate inference are derived by using a cluster variation method. The linear response formulas can give us the marginal probability between every pair of nodes and we obtain good accuracy in some examples of practical probabilistic inferences.

## 1 Introduction

In probabilistic inference for signal processing, medical diagnosis, code theory, digital communication, machine learning and so on, belief propagation (BP) is one of powerful approximate methods to calculate a belief for each node within a practical computational time[1]. Recently, it has been pointed out that the BP has some mathematical structures in common with advanced mean-field (MF) methods in the statistical-mechanics[2]. Yedidia *et al.*[3] have pointed out that the update rule of generalized belief propagation (GBP) is equivalent to a cluster variation method (CVM) that is one of the advanced MF methods.

However, it is difficult to calculate the marginal probability in a pair of nodes, which are not connected directly to each other, by means of the BP, because the recursion formulas in the BP are constructed only from some marginal probabilities. In Refs.[4, 5], it was mentioned that the correlations in every pair of nodes in a probabilistic model can be calculated by combining the native MF approximation with a linear response (LR) theory. The present author extended the naive MF approximation in their framework to the CVM[6] for the case that each node has two states and derived the marginal probability in every pair of nodes as an inverse of matrix. On the other hand, Welling and Teh[7] constructed an update rule of the GBP to calculate pairwise marginal probabilities by using the LR theory.

In this paper, we derive the general formula for the marginal probability in every pair of nodes for any probabilistic models by combining the CVM with the LR theory. for the case that each node has

---

<sup>1</sup>E-mail: kazu@statp.is.tohoku.ac.jp

multi- states. The framework of the present paper is a generalization of [6] and [7].

## 2 Linear Response Theory

We consider a probabilistic model consisting of some nodes. A random variable  $x_i$  is associated with every node  $i$ . The set of all the nodes is denoted by  $\Omega$  and the number of all the nodes is expressed in terms of notation  $|\Omega|$ . The set of all the random variables  $x_i$  is denoted by the notation  $\mathbf{x}$ , such that  $\mathbf{x} \equiv \{x_i | i \in \Omega\}$ . For the probability  $P(\mathbf{x})$  for the set of nodes,  $\Omega$ , we have the following equality:

$$\lim_{h \rightarrow 0} \left\{ \frac{\partial}{\partial h} \left( \frac{\sum_{\mathbf{z}} \delta_{n,z_j} \exp(h\delta_{m,z_i}) P(\mathbf{z})}{\sum_{\mathbf{z}} \exp(h\delta_{m,z_i}) P(\mathbf{z})} \right) \right\} = P_{ij}(m, n) - P_i(m)P_j(n), \quad (1)$$

where  $P_i(m) \equiv \sum_{\mathbf{z}} \delta_{m,z_i} P(\mathbf{z})$  and  $P_{ij}(m, n) \equiv \sum_{\mathbf{z}} \delta_{m,z_i} \delta_{n,z_j} P(\mathbf{z})$ . The notation  $\sum_{\mathbf{z}}$  is the summation over all the possible configurations with respect to the set of random variables  $\mathbf{z} \equiv \{z_i | i \in \Omega\}$ . Thus,  $P_{ij}(m, n) - P_i(m)P_j(n)$  in nodes  $i$  and  $j$  is regarded as a differentiate with respect to  $h$  of the average of  $\delta_{n,x_j}$  in the probabilistic model specified by the factor  $\exp(h\delta_{m,x_i})P(\mathbf{x})$ .

## 3 Cluster Variation Method

In order to explain the framework of CVM, we should define some notations for clusters. Cluster is a set of nodes. When a node  $i$  belongs to a cluster  $\gamma$ , we call  $i$  an element of  $\gamma$  and we express it in terms of the notation  $i \in \gamma$ . When all the nodes in a cluster  $\gamma'$  belong to a cluster  $\gamma$ , we call  $\gamma'$  a subcluster of  $\gamma$ . When a cluster  $\gamma'$  is a subcluster of a cluster  $\gamma$ , we use the notation  $\gamma' \leq \gamma$ . We express  $\gamma' < \gamma$  when a cluster  $\gamma'$  is a proper subcluster of  $\gamma$ .

First of all, we have to specify a set of basic clusters. Every basic cluster must not be a subcluster of another element in the set of basic clusters. We denote the set of basic clusters by  $B$ . We consider such a set  $C$  of clusters that a cluster is in  $C$  if and only if it is a cluster in  $B$  or is the cluster of the common nodes of two or more clusters in  $B$ , excluding the empty cluster  $\mathbf{0}$ . The set of all the clusters, which belong to the set  $C$  and do not belong to the set  $B$ , is denoted by the notation  $C \setminus B$ . A set of random variables  $x_i$  associated with nodes  $i$  belonging to a cluster  $\gamma$  is denoted by  $\mathbf{x}_\gamma \equiv \{x_i | i \in \gamma\}$ . The notation  $\mathbf{x}_{\Omega \setminus \gamma} \equiv \{x_i | i \in \Omega \setminus \gamma\}$  is a set of random variables  $x_i$  associated with nodes  $i$  not belonging to the cluster  $\gamma$ .

We consider the following representation

$$P(\mathbf{x}) \equiv \frac{\prod_{\{\gamma | \gamma \in C\}} W_\gamma(\mathbf{x}_\gamma)^{-\mu(\gamma)}}{\sum_{\mathbf{z}} \prod_{\{\gamma | \gamma \in C\}} W_\gamma(\mathbf{z}_\gamma)^{-\mu(\gamma)}}, \quad (2)$$

where

$$\mu(\alpha) \equiv -1 - \sum_{\{\gamma | \gamma > \alpha, \gamma \in C\}} \mu(\gamma) \quad (\alpha \in C). \quad (3)$$

Each function  $W_\alpha(\mathbf{x}_\alpha)$  satisfies the normalization condition  $\sum_{\mathbf{z}_\alpha} W_\alpha(\mathbf{z}_\alpha) = 1$ . The function  $\mu(\gamma)$  ( $\gamma \in C$ ) is referred as a Möbius function in the set theory.

In order to calculate the correlation in every pair of nodes  $i$  and  $j$  ( $i \in \Omega, j \in \Omega$ ), we consider the probabilistic model under an infinitesimal parameter  $h_i$  defined by

$$P^{(i,m)}(\mathbf{x}) \equiv \frac{\exp(h_i \delta_{m,x_i}) P(\mathbf{x})}{\sum_{\mathbf{z}} \exp(h_i \delta_{m,z_i}) P(\mathbf{z})}. \quad (4)$$

The left-hand side of Eq.(1) is corresponding to the differential of the marginal probability

$$P_j^{(i,m)}(n) \equiv \sum_{\mathbf{z}} \delta_{n,z_j} P^{(i,m)}(\mathbf{z}) \quad (5)$$

in the probabilistic model (4). In the present section, we give the general method for calculating the marginal probability  $P_j^{(i,m)}(n)$  for any pair of nodes  $i$  and  $j$  by using the CVM. The differential of the average can be obtained by using the LR theory. The calculation of the differential by means of the CVM will be explained in the next section.

In the CVM[6], the marginal probabilities  $\{P_\gamma^{(i,m)} | \gamma \in C\}$  are approximately expressed as follows:

$$P_\gamma^{(i,m)}(\mathbf{x}_\gamma) \simeq \frac{W_\gamma(\mathbf{x}_\gamma) \exp\left(\sum_{\{\alpha | \alpha \leq \gamma, \alpha \in C \setminus B\}} \lambda_{\alpha,\gamma}^{(i,m)}(\mathbf{x}_\alpha)\right)}{\sum_{\mathbf{z}_\gamma} W_\gamma(\mathbf{z}_\gamma) \exp\left(\sum_{\{j | \alpha \leq \gamma, \alpha \in C \setminus B\}} \lambda_{\alpha,\gamma}^{(i,m)}(\mathbf{z}_\alpha)\right)} \quad (\gamma \in C), \quad (6)$$

where  $\lambda_{\alpha,\gamma}^{(i,m)}(\mathbf{x}_\alpha)$  are determined so as to satisfy the following equations:

$$h_i \delta_{x_i,m} \delta_{i,\alpha} = - \sum_{\{\gamma | \gamma \geq \alpha, \gamma \in C\}} \mu(\gamma) \lambda_{\alpha,\gamma}^{(i,m)}(\mathbf{x}_\alpha) \quad (\alpha \in C \setminus B), \quad (7)$$

$$P_\alpha^{(i,m)}(\mathbf{x}_\alpha) = \sum_{\mathbf{z}_\gamma} \delta_{\mathbf{x}_\alpha, \mathbf{z}_\alpha} P_\gamma^{(i,m)}(\mathbf{z}_\gamma) \quad (\gamma > \alpha, \gamma \in C, \alpha \in C \setminus B). \quad (8)$$

In the similar arguments, we obtain the marginal probabilities of the probabilistic model given in Eq.(2) as follows:

$$P_\gamma(\mathbf{x}_\gamma) \equiv \sum_{\mathbf{z}} \delta_{\mathbf{x}_\alpha, \mathbf{z}_\alpha} P(\mathbf{z}) \simeq \frac{W_\gamma(\mathbf{x}_\gamma) \exp\left(\sum_{\{\alpha | \alpha \leq \gamma, \alpha \in C \setminus B\}} \lambda_{\alpha,\gamma}(\mathbf{x}_\alpha)\right)}{\sum_{\mathbf{z}_\gamma} W_\gamma(\mathbf{z}_\gamma) \exp\left(\sum_{\{j | \alpha \leq \gamma, \alpha \in C \setminus B\}} \lambda_{\alpha,\gamma}(\mathbf{z}_\alpha)\right)} \quad (\gamma \in C), \quad (9)$$

where  $\lambda_{\alpha,\gamma}(\mathbf{x}_\alpha)$  are determined so as to satisfy the following equations:

$$0 = - \sum_{\{\gamma | \gamma \geq \alpha, \gamma \in C\}} \mu(\gamma) \lambda_{\alpha,\gamma}(\mathbf{x}_\alpha) \quad (\alpha \in C \setminus B), \quad (10)$$

$$P_\alpha(\mathbf{x}_\alpha) = \sum_{\mathbf{z}_\gamma} \delta_{\mathbf{x}_\alpha, \mathbf{z}_\alpha} P_\gamma(\mathbf{z}_\gamma) \quad (\gamma > \alpha, \gamma \in C, \alpha \in C \setminus B). \quad (11)$$

## 4 Linear Response in Cluster Variation Method.

In this section, we give a formula to calculate a marginal probability in every pair of nodes by means of the CVM and the LR theory. As shown in Eq.(1), for the probabilistic model defined by Eq.(4), the average of  $x_j$  for each node  $j$  is defined by

$$P_j^{(i,m)}(n) \equiv \sum_{\mathbf{z}} \delta_{n,z_j} P^{(i,m)}(\mathbf{z}) \quad (j \in \Omega). \quad (12)$$

In order to obtain the marginal probability  $P_{ij}(m, n)$ , we wish to compute the derivative of Eq.(12) with respect to  $h_i$ . The quantity  $\lim_{h_i \rightarrow 0} \frac{\partial P_j^{(i,m)}(n)}{\partial h_i}$  is corresponding to the first order terms in expanding the marginal probability  $P_j^{(i,m)}(n)$ , in powers of  $h_i$ . Hence, the marginal probability  $P_{ij}(m, n)$  in a pair of nodes  $i$  and  $j$  is related to the derivative of  $P_j^{(i,m)}(n)$  with respect to the infinitesimal parameter  $h_i$  at a node  $i$  as follows:

$$P_j^{(i,m)}(n) - P_j(n) = (P_{ij}(m, n) - P_i(m)P_j(n))h_i. \quad (13)$$

The marginal probabilities  $P_{ij}(m, n)$  and  $P_i(m)$ . We have to consider the quantities  $P_j^{(i,m)}(n) - P_j(n)$  and  $P_{ij}(m, n) - P_i(m)P_j(n)$  separately for the following four cases:

**Case 1.** Both the nodes  $i$  and  $j$  belong to the set  $C \cap \Omega$ .

**Case 2.** The node  $i$  belongs to the set  $C \cap \Omega$  and  $j$  belongs to the set  $\Omega \setminus (C \cap \Omega)$ .

**Case 3.** Both the nodes  $i$  and  $j$  belong to the set  $\Omega \setminus (C \cap \Omega)$ .

Here the set  $\Omega \setminus (C \cap \Omega)$  consists of all the nodes that belong to the set  $\Omega$  and do not belong to the set  $C$ .

First, we consider the case of  $i, j \in C \cap \Omega$  (Case 1). Expanding  $P_\alpha^{(i,m)}(\mathbf{x}_\alpha)$  ( $\alpha \leq \gamma$ ,  $\alpha \in C \setminus B$ ,  $\gamma \in C$ ) in powers of  $h_i$  and  $\lambda_{\beta,\gamma}^{(i,m)}(\mathbf{z}_\beta) - \lambda_{\beta,\gamma}(\mathbf{z}_\beta)$ , and retaining only the first order terms, we get

$$P_\alpha^{(i,m)}(\mathbf{x}_\alpha) - P_\alpha(\mathbf{x}_\alpha) = \sum_{\{\beta|\beta \leq \alpha, \beta \in C\}} \sum_{\mathbf{z}_\beta} \langle \alpha, \mathbf{x}_\alpha | \mathbf{A}_\gamma | \beta, \mathbf{z}_\beta \rangle \left( \lambda_{\beta,\gamma}^{(i,m)}(\mathbf{z}_\beta) - \lambda_{\beta,\gamma}(\mathbf{z}_\beta) \right) \quad (i \in C \cap \Omega), \quad (14)$$

where

$$\langle \alpha, \mathbf{x}_\alpha | \mathbf{A}_\gamma | \beta, \mathbf{x}_\beta \rangle \equiv P_{\alpha \cup \beta}(\mathbf{x}_{\alpha \cup \beta}) - P_\alpha(\mathbf{x}_\alpha)P_\beta(\mathbf{x}_\beta) \quad (\gamma \in C, \alpha \in C \setminus B, \beta \in C \setminus B, \alpha \cup \beta \leq \gamma), \quad (15)$$

We have to remark that the nodes  $j$  and  $k$  in the quantity  $\langle j | \mathbf{A}_\gamma | k \rangle$  belong not only to the cluster  $\gamma$  but also to subclusters of the cluster  $\gamma$  ( $\gamma \in C$ ), such that  $j, k \leq \gamma$  and  $j, k \in C \setminus B$ . For a fixed  $\gamma$ , we regard Eq.(14) as a system of linear equations for unknowns  $\lambda_{k,\gamma}^{(i,m)}(x_k) - \lambda_{k,\gamma}(x_k)$  ( $k \leq \gamma$ ,  $k \in C$ ,  $\gamma \in C$ ), and express the solution as follows:

$$\lambda_{\alpha,\gamma}^{(i,m)}(\mathbf{x}_\alpha) - \lambda_{\alpha,\gamma}(\mathbf{x}_\alpha) = \sum_{\{\beta|\beta \leq \alpha, \beta \in C\}} \sum_{\mathbf{z}_\beta} \langle \alpha, \mathbf{x}_\alpha | \mathbf{A}_\gamma^{-1} | \beta, \mathbf{z}_\beta \rangle \left( P_\beta^{(i,m)}(\mathbf{z}_\beta) - P_\beta(\mathbf{z}_\beta) \right). \quad (16)$$

Here the matrices  $\langle j | \mathbf{A}_\gamma^{-1} | k \rangle$  are the matrix elements of the inverse of the matrix  $\mathbf{A}_\gamma$ . Substituting Eq.(16) into Eqs.(7) and (10) and writing  $P_\beta^{(i,m)}(\mathbf{z}_\beta) - P_\beta(\mathbf{z}_\beta)$  which are determined consistently by Eq.(8) and (11), as  $P^{(i,m)}(\mathbf{z}_\beta) - P(\mathbf{z}_\beta)$ , we can express the result as

$$\sum_{\{\beta|\beta \in C \setminus B\}} \sum_{\mathbf{z}_\beta} \langle \alpha, \mathbf{x}_\alpha | \mathbf{G} | \beta, \mathbf{z}_\beta \rangle \left( P_\beta^{(i,m)}(\mathbf{z}_\beta) - P_\beta(\mathbf{z}_\beta) \right) = h_i \delta_{x_i, m} \delta_{i, \alpha}, \quad (17)$$

where

$$\langle \alpha, \mathbf{x}_\alpha | \mathbf{G} | \beta, \mathbf{x}_\beta \rangle \equiv - \sum_{\{\gamma | \gamma \geq \alpha, \gamma \geq \beta, \gamma \in C\}} \mu(\gamma) \langle \alpha, \mathbf{x}_\alpha | \mathbf{A}_\gamma^{-1} | \beta, \mathbf{x}_\beta \rangle. \quad (18)$$

From Eqs.(13) and (17), we obtain the quantity  $P_{ij}(m, n) - P_i(m)P_j(n)$  in pairs of nodes  $i$  and  $j$  for Case 1 as follows:

$$P_{ij}(m, n) - P_i(m)P_j(n) = \langle i, m | \mathbf{G}^{-1} | j, n \rangle \quad (i, j \in \Omega \cap C). \quad (19)$$

Next, we consider the case of  $i \in \Omega \setminus (C \cap \Omega)$  and  $j \in C \cap \Omega$  (Case 2). Expanding  $P_\alpha^{(i,m)}(\mathbf{x}_\alpha)$  ( $\alpha \leq \gamma$ ,  $\alpha \in C \setminus B$ ,  $\gamma \in C$ ) in powers of  $h_i$  and  $\lambda_{\beta,\gamma}^{(i,m)}(\mathbf{z}_\beta) - \lambda_{\beta,\gamma}(\mathbf{z}_\beta)$ , and retaining only the first order terms, we get

$$\begin{aligned} P_\alpha^{(i,m)}(\mathbf{x}_\alpha) - P_\alpha(\mathbf{x}_\alpha) &= \langle i, x_i | \mathbf{K}_\gamma | \alpha, \mathbf{x}_\alpha \rangle h_i \delta_{i \in \gamma} \\ &+ \sum_{\{\beta | \beta \leq \alpha, \beta \in C\}} \sum_{\mathbf{z}_\beta} \langle \alpha, \mathbf{x}_\alpha | \mathbf{A}_\gamma | \beta, \mathbf{z}_\beta \rangle \left( \lambda_{\beta,\gamma}^{(i,m)}(\mathbf{z}_\beta) - \lambda_{\beta,\gamma}(\mathbf{z}_\beta) \right) \\ &\quad (\gamma \in C, j \in C, j \leq \gamma, i \in \Omega \setminus (C \cap \Omega)). \end{aligned} \quad (20)$$

where  $\delta_{i \in \gamma} \equiv 1$  for  $i \in \gamma$  and  $\delta_{i \in \gamma} \equiv 0$  for  $i \notin \gamma$ . We can solve Eq.(20) with respect to  $\lambda_{\beta,\gamma}^{(i,m)}(\mathbf{z}_\beta) - \lambda_{\beta,\gamma}(\mathbf{z}_\beta)$ . By substituting it to Eq.(7) and (10) and by using Eqs.(13), we obtain

$$P_{ij}(m, n) - P_i(m)P_j(n) = \sum_{\{\alpha | \alpha \in C \setminus B\}} \sum_{\mathbf{z}_\alpha} \langle i, m | \mathbf{G}^{-1} | \alpha, \mathbf{z}_\alpha \rangle \langle \alpha, \mathbf{z}_\alpha | \mathbf{R} | j, n \rangle \quad (i \in \Omega \setminus (C \cap \Omega), j \in C \cap \Omega) \quad (21)$$

where

$$\begin{aligned} \langle j | \mathbf{R} | k \rangle = \langle k | \mathbf{R} | j \rangle &\equiv \sum_{\{\beta | \beta \leq \alpha, \beta \in C \setminus B\}} \langle k, x_k | \mathbf{A}_\alpha^{-1} | \beta, \mathbf{x}_\beta \rangle \langle \beta, \mathbf{x}_\beta | \mathbf{K}_\alpha | j, x_j \rangle \\ &\quad (k \in C \cap \Omega, k \in \alpha, j \in \Omega \setminus (C \cap \Omega), j \in \alpha, \alpha \in B), \end{aligned} \quad (22)$$

$$\langle \alpha, x_\alpha | \mathbf{K}_\gamma | \beta, x_\beta \rangle \equiv P_{\alpha \cup \beta}(\mathbf{z}_{\alpha \cup \beta}) - P_\alpha(\mathbf{x}_\alpha) P_\beta(\mathbf{x}_\beta) \quad (\alpha \in C \cup \Omega, \beta \in C \cup \Omega, \alpha \leq \gamma, \beta \leq \gamma), \quad (23)$$

and  $P_{\gamma'}(\mathbf{x}_{\gamma'}) \equiv \sum_{\mathbf{z}_\gamma} \delta_{\mathbf{x}_{\gamma'}, \mathbf{z}_\gamma} P_\gamma(\mathbf{z}_\gamma)$  ( $\gamma \in C$ ,  $\gamma' \leq \gamma$ ).

Finally, we consider the case of  $i, j \in \Omega \setminus (C \cap \Omega)$  (Case 3). Expanding  $P_\alpha^{(i,m)}(\mathbf{x}_\alpha)$  ( $\alpha \leq \gamma$ ,  $\alpha \in C \setminus B$ ,  $\gamma \in C$ ) in powers of  $h_i$  and  $\lambda_{\beta,\gamma}^{(i,m)}(\mathbf{z}_\beta) - \lambda_{\beta,\gamma}(\mathbf{z}_\beta)$ , and retaining only the first order terms, we get

$$\begin{aligned} P_\alpha^{(i,m)}(\mathbf{x}_\alpha) - P_\alpha(\mathbf{x}_\alpha) &= \sum_{\{\beta | \beta \leq \alpha, \beta \in C\}} \sum_{\mathbf{z}_\beta} \langle \alpha, \mathbf{x}_\alpha | \mathbf{K}_\gamma | \beta, \mathbf{z}_\beta \rangle \left( \lambda_{\beta,\gamma}^{(i,m)}(\mathbf{z}_\beta) - \lambda_{\beta,\gamma}(\mathbf{z}_\beta) \right) \\ &\quad (\alpha \in B, i, j \in \alpha, i, j \in \Omega \setminus (C \cap \Omega)). \end{aligned} \quad (24)$$

We solve Eq.(20) with respect to  $\lambda_{\beta,\gamma}^{(i,m)}(\mathbf{z}_\beta) - \lambda_{\beta,\gamma}(\mathbf{z}_\beta)$ . By substituting it to Eq.(24) and by using Eqs.(21) and (13), we obtain

$$\begin{aligned} P_{ij}(m, n) - P_i(m)P_j(n) &= \sum_{\{\alpha | \alpha \in C \setminus B\}} \sum_{\{\beta | \beta \in C \setminus B\}} \sum_{\mathbf{z}_\alpha} \sum_{\mathbf{z}_\beta} \langle i, m | \mathbf{R} | \alpha, \mathbf{z}_\alpha \rangle \langle \alpha, \mathbf{z}_\alpha | \mathbf{G}^{-1} | \beta, \mathbf{z}_\beta \rangle \langle \beta, \mathbf{z}_\beta | \mathbf{R} | j, n \rangle \\ &\quad (i, j \in \Omega \setminus (C \cap \Omega)). \end{aligned} \quad (25)$$

The marginal probabilities  $\{P_\gamma | \gamma \in C\}$  for the probabilistic model  $P(\mathbf{x})$  are obtained from the GBP. The update rule of GBP is constructed as an iterative procedure by using Eqs.(9)-(11). By substituting the numerical results of  $\{P_\gamma | \gamma \in C\}$  to Eqs.CVMMatrixX) and (22) with Eqs.(15) and (23), the matrices  $\mathbf{G}$  and  $\mathbf{R}$  can be calculated.

## 5 Concluding Remarks

In the present paper, we have derived a general formula for pairwise marginal probability in any probabilistic model. The final results are given as compact forms given in Eqs.(19), (21) and (25). For an example “Asia” that is available at a website [8], we consider the conditional probability  $\Pr\{X_T = P|X_D = \text{Present}\} \equiv \Pr\{X_T = P, X_D = P\}/\Pr\{X_D = P\}$ , where the notation “P”, “T” and “D” denote Present, Tuberculosis, and Dyspnea, respectively. The nodes “T” and “D” are not connected directly to each other. By using the present method. we obtain  $\Pr\{X_D = P\} = 0.4393$  and  $\Pr\{X_T = P, X_D = P\} = 0.0082$  and then the value of the conditional probability  $\Pr\{X_T = P|X_D = P\} = 0.0187$ . In contrast, the corresponding value in the exact calculation for  $\Pr\{X_T = P|X_D = P\}$  is 0.0188.

The present framework is based on the one in [6]. In [6], the present author imposed an assumption that we have  $C \cap \Omega = C \setminus B$  and considered only the LR of the first moment of  $x_i$  ( $i \in \Omega$ ). If we do not impose the above assumption, we have to consider all moments of  $x_i, x_i x_j, x_i x_j x_k$  ( $i, j, k \in C \cup \Omega$ ) and so on. In the present work, we consider only the LR of the marginal probability  $P_\gamma^{m,i}(\mathbf{x}_\gamma)$  ( $\gamma \in C$ ). In the present framework, we do not need such an assumption and give simpler formulas than in Ref.[6]. The graph that satisfies the assumption  $C \cap \Omega = C \setminus B$  is called *factor graph*. Welling and Teh[7] also investigated the LR for the probabilistic models on factor graphs by using the GBP and derived the update rule for the marginal probability in every pair of nodes. They expanded the update rule of GBP in powers of *messages*. In the present paper, the author has derived the more compact LR formula in the CVM for more general graphical models. In the CVM, the approximate marginal probabilities given in Eqs.(6) and (7) are derived by minimization of an approximate free energy expressed in terms of  $\{P_\gamma^{i,m}(\mathbf{x}_\gamma)|\gamma \in C\}$  under the reducibilities (8)[6]. The present author has expanded the approximate marginal probabilities (6) in powers of not *messages* but Lagrange multipliers for reducibilities (8) in order to obtain the LR in the CVM. Both formulations should give us the same results in the probabilistic models for factor graphs.

## Acknowledgements

This work was partly supported by the Grants-In-Aid (No.14084203) for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## References

- [1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.
- [2] M. Opper and D. Saad (eds), *Advanced Mean Field Methods — Theory and Practice —*, MIT Press, 2001.
- [3] J. S. Yedidia, W. T. Freeman and Y. Weiss: Generalized belief propagation, *Advances in Neural Information Processing Systems*, vol.13, pp.689-695, MIT Press, 2001.

- [4] H. J. Kappen and F. B. Rodríguez: Efficient learning in Boltzmann machines using linear response theory, *Neural Computation*, vol.10, pp.1137-1156, 1998.
- [5] T. Tanaka: Mean field theory of Boltzmann machine learning, *Phys. Rev. E*, vol.58, pp.2302-2310, 1998.
- [6] K. Tanaka: Probabilistic inference by means of cluster variation method and linear response theory, *IEICE Transactions on Information and Systems*, vol.E86-D, pp.1228-1242, 2003.
- [7] M. Welling and Y. W. Teh: Linear response for approximate inference, *Advances in Neural Information Processing Systems*, vol.16, MIT Press, 2003, to appear.
- [8] <http://www.norsys.com/networklibrary.html>.