LETTER

Theory and Experiments of Exchange Ratio for Exchange Monte Carlo Method

Kenji Nagata

Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology
MailBox R2-5, 4259, Nagatsuta-chou, Midori-ku, Yokohama, 226-8503, Japan
E-mail: kenji.nagata@cs.pi.titech.ac.jp

Sumio Watanabe

P&I. Lab., Tokyo Institute of Technology
MailBox R2-5, 4259, Nagatsuta-chou, Midori-ku, Yokohama, 226-8503, Japan
E-mail: swatanab@cs.pi.titech.ac.jp

(Submitted on December 14, 2007)

Abstract – In hierarchical learning machines such as neural networks, Bayesian learning provides better generalization performance than maximum likelihood estimation. However, its accurate approximation using the Markov chain Monte Carlo (MCMC) method requires a huge computational cost. The exchange Monte Carlo (EMC) method was proposed as an improvement on the MCMC method. Although it has been shown to be effective not only in Bayesian learning but also in many fields, the mathematical foundation of the EMC method has not yet been established. In our previous work, we derived the asymptotic behavior of the average exchange ratio, which is used as a criterion for designing the EMC method. In this paper, we verify the accuracy of our theoretical result by the simulation of Bayesian learning in linear neural networks, and propose the method to check the convergence of EMC method based on our theoretical result.

Keywords – Markov Chain Monte Carlo Method, Exchange Monte Carlo Method, Exchange Ration.

1. Introduction

A lot of learning machines with hierarchical structures such as neural networks and hidden Markov models are widely used for pattern recognition, gene analysis and many other applications. For these hierarchical learning machines, Bayesian learning is proven to provide better generalization performance than maximum likelihood estimation [1] [14] [15].

In Bayesian learning, we need to compute the expectation over a Bayesian posterior distribution, which cannot be performed exactly. Therefore, Bayesian learning requires some approximation methods. One of the well-known approximation methods is the Markov Chain Monte Carlo (MCMC) method. The MCMC method is a well-known algorithm for generating a sample sequence which converges to a target distribution. However, it requires a huge computational cost because of slow convergence of a sample sequence, in particular, in the Bayesian posterior distribution for a hierarchical learning machine.

Recently, various improvements of the MCMC method have been developed, based on the idea of an extended ensemble. These improvements are surveyed in [7]. The multicanonical method [3] and the simulated tempering [9] are examples of this extended ensemble methods. This idea gives us a general strategy to overcome the problem of slow convergence of a conventional MCMC method. The exchange Monte Carlo (EMC) method is well known as an extended ensemble method [5]. This method involves generating the sample sequence from a joint distribution, which consists of many distributions with different temperatures. Its algorithm is based on two MCMC simulation steps. One step is the conventional update of the MCMC simulation for each distribution. The other is a probabilistic exchange process between two neighboring sequences. The EMC method has been successfully applied to not only Bayesian learning in hierarchical learning machines [8] but also an optimization problem [6] [12] and a protein-folding problem [13].

When we design the EMC method, the temperature settings have an important effect on the efficiency of the algorithm [4]. The temperature values have a close relation to the exchange ratio and its average, which is the acceptance ratio of the exchange process. Although the average exchange ratio is used as a criterion for the setting of temperatures, the mathematical property of the average exchange ratio has not been established. In our previous work, we derived an analytic result concerning the average exchange ratio, and propose the optimal temperature setting for the EMC method.

In this paper, we verify our theoretical result by comparing the theoretical value of average exchange ratio with the experimental value for the simulation of Bayesian learning in linear neural network. In addition, we propose the method for checking the convergence of the EMC method based on our theoretical result.

This paper consists of six sections. In section 2, we outline the EMC method and explain its design. In section 3, the main result of the analysis of the EMC method is presented. In section 4, we describe the experimental results. The discussion is followed in section 5 and the conclusion in section 6.

2. Background

2.1 The Exchange Monte Carlo Method

In this section, we introduce the well-known EMC method.

Suppose that $w \in R^d$ and our aim is to generate a sample sequence from the following target probability distribution with energy function $f(w)$ and probability distribution $\varphi(w)$,

$$p(w) = \frac{1}{Z(n)} exp(-nf(w))\varphi(w),$$

where $Z(n)$ is the normalization constant. In Bayesian learning, for example, we need to generate a sample sequence from the Bayesian posterior distribution in order to estimate the expectation. Then, the number n, the function $f(w)$, and the probability distribution $\varphi(w)$ respectively correspond to the number of training data, Kullback information between a true distribution and a learning machine, and the prior distribution of parameters. The EMC method treats a compound system, which consists of K non-interacting sample sequences from the system concerned. The k-th sample sequence $\{w_k\}$ converges to the following probability distribution:

$$p(w \mid t_k) = \frac{1}{Z(nt_k)} exp(-nt_k f(w))\varphi(w) \quad (1 \le k \le K),$$

where $t_1 < t_2 < \cdots < t_K$ is called "inverse temperatures" (in this paper, we call "temperatures" hereafter). Given a set of temperatures $\{t\} = \{t_1, \cdots, t_K\}$, the joint distribution for $W = \{w_1, \cdots, w_K\}$ can be expressed by a simple product formula,

$$p(W) = \prod_{k=1}^{K} p(w_k \mid t_k).$$

The EMC method is based on two types of updating for construction of a Markov chain. One is the conventional update of an MCMC simulation such as the Metropolis algorithm for each target distribution $p(w_k \mid t_k)$. The other is a position exchange between two neighboring sequences, that is, $\{w_k, w_{k+!}\} \to \{w_{k+1}, w_k\}$. The transition probability u is determined by the detailed balance condition for the joint distribution $p(W)$ as follows,

$$u = min(1, r)$$
$$r = \frac{p(w_{k+1} \mid t_k)p(w_k \mid t_{k+1})}{p(w_k \mid t_k)p(w_{k+1} \mid t_{k+1})} = exp(n(t_{k+1} - t_k)(f(w_{k+1}) - f(w_k))). \quad (1)$$

Henceforth, we call u "exchange ratio".

Consequently, the following two steps are carried out alternately:

1. Each sequence is generated simultaneously and independently for a few iterations by the conventional MCMC method.

2. Two positions are exchanged using the exchange ratio u .

The advantage of the EMC method is that it results in accelerated convergence of the sample sequence compared to the conventional MCMC method. The conventional MCMC method of generating a sample sequence from the target distribution is very expensive because this algorithm is based on local updating. The EMC method can realize efficient sampling by preparing a simple distribution such as a normal distribution, which makes convergence of the sample sequence straightforward. In practice, we set the temperature of the target distribution to be $t_K = 1$, and that of the simple distribution to be $t_1 = 0$.

2.2 Exchange ratio of the EMC method

When the EMC method is implemented, its efficiency depends strongly on the temperature setting. As we can see from Eq.(1), the temperature is closely related to the exchange ratio, so it is a very important parameter for adjusting the exchange ratio and its average.

For the efficient EMC method, each sample needs to wander over the whole temperature region. In addition, the time for a sample to move from end to end (from t_1 to t_K) is good to be short. Therefore, it is inefficient for the interval of neighboring temperatures to be large, which leads to the low average exchange ratio. On the contrary, in order to make the average exchange ratio higher, the interval of neighboring temperatures has to be smaller, which involve the huge total number K of temperatures. Therefore, this arrangement is also inefficient because it is very expensive to generate a sample sequence from each distribution. Consequently, the set of temperatures needs to be optimized so that the average exchange ratios for any neighboring temperatures becomes not low and not too high.

In this way, the average exchange ratio is used as a criterion for the setting of temperatures. However, the theoretical property of the average exchange ratio has not been clarified. Hence, in order to obtain the value of average exchange ratio, we have to simulate EMC method. Moreover, there is the problem that the accuracy of experimental average exchange ratio is not clear because EMC method is based on a probabilistic algorithm.

In our previous work, we analytically clarified the asymptotic behavior of average exchange ratio in the low temperature limit, that is, as $n \to \infty$ [10]. This result gives us a criterion for optimizing the set of temperatures and for checking the convergence of EMC simulations. However, the accuracy of this theoretical result has not been clarified experimentally.

In this paper, we verify the accuracy of our result by comparing the theoretical value of average exchange ratio with the experimental value, and propose the method for checking the convergence of EMC method based on our theoretical result..

3. Our Previous Work

In this section, we present the theoretical result of our previous study [10] [11].

We assume that $t > 0$ and $t + \Delta t > 0$, and consider the EMC method applied to the following two distributions,

$$p_1(w) = \frac{1}{Z(nt)} exp\left(-ntf(w)\right)\varphi(w)$$

$$p_2(w) = \frac{1}{Z(nt)} exp\left(-n(t+\Delta t)f(w)\right)\varphi(w),$$

where the interval Δt of two temperatures is not necessarily small. As we can see from Eq.(1), the exchange ratio u is a function of each sample w_1 and w_2 . Hence, we define the average exchange ratio J as the expectation of exchange ratio over the joint distribution $p_1(w) \times p_2(w)$ as follows,

$$J = \iint u p_1(w_1) p_2(w_2) dw_1 dw_2 .$$

Note that the average exchange ratio J is a function of the temperature t and the interval Δt of two temperatures.

In a lot of learning machines with hierarchical structures, the Bayesian posterior distribution does not converge to the normal distribution as $n \to \infty$ because the Hessian of log likelihood $f(w)$ is not positive definite. We can assume $f(w) \geq 0$ and $f(w_0) = 0\ (\exists w_0)$ without loss of generality.

The zeta function of $f(w)$ and $\varphi(w)$ is defined by

$$\xi(z)=\int f(w)^z \varphi(w)dw,$$

where z is a one complex variable. Then $\xi(z)$ is a holomorphic function in the region of $Re(z)>0$, and can be analytically continued to a meromorphic function on the entire complex plane, whose poles are all real, negative, and rational numbers [2]. We also define the rational number $-\lambda$ as the largest pole of the zeta function $\xi(z)$ and the natural number m as its order. If the Hessian matrix $\left(\frac{\partial^2 f}{\partial w_i \partial w_j}\right)$ is positive definite for arbitrary w, it follows that $\lambda=\frac{d}{2}, m=1$. Otherwise, λ and m can be calculated by using the resolution of singularities in algebraic geometry [2]. In fact, there are some studies which calculate the values λ and m for a certain energy function $f(w)$ and a probability distribution $\varphi(w)$ [15].

Then, the following lemma holds [14].

Lemma 1 The state density function $V(s)(s>0)$ has the following asymptotic expansion for $s \to 0$,

$$\begin{aligned}V(s)&=\int \delta(s-f(w))\varphi(w)dw\\ &\cong cs^{\lambda-1}(-\log s)^{m-1},\end{aligned}$$

where c is a constant.

By using this lemma, we proved the following theorem about the average exchange ratio [11].

Theorem 1 The average exchange ratio J converges in the following way as $n \to \infty$:

$$J \to \begin{cases}\left(1+\dfrac{\Delta t}{t}\right)^{\lambda}\dfrac{2\Gamma(2\lambda)}{\Gamma(\lambda)^2} g\left(\lambda,\dfrac{\Delta t}{t}\right) & (\text{if } \Delta t \geq 0)\\ \left(1-\dfrac{\Delta t}{t+\Delta t}\right)^{\lambda}\dfrac{2\Gamma(2\lambda)}{\Gamma(\lambda)^2} g\left(\lambda,-\dfrac{\Delta t}{t+\Delta t}\right) & (\text{if } \Delta t < 0),\end{cases}$$

where the function $g\left(\lambda,\frac{\Delta t}{t}\right)$ is defined by

$$g\left(\lambda,\frac{\Delta t}{t}\right)=\int_0^1 \frac{s^{\lambda-1}}{\left(1+\frac{\Delta t}{t}+s\right)^{2\lambda}}ds.$$

The condition $n \to \infty$ means that the coefficients, nt and $n(t+\Delta t)$, of exponential part in the distributions $p_1(w)$ and $p_2(w)$ also go to infinity. Hence, if the value t is equal to or smaller than order of $\frac{1}{n}$, Theorem 1 cannot be applied to these distributions.

Although the range of temperature which our theorem is applicable is limited, we clarify the average exchange ratio quantitatively. According to Theorem 1, the average exchange ratio is expressed by a function of the term $\frac{\Delta t}{t}$ or $\frac{\Delta t}{t+\Delta t}$. Hence, in order to make the average exchange ratio constant over the various temperatures, the set of temperatures should be set as the values $\frac{\Delta t}{t}$ are constant over the various temperatures. Then, the set $\{t_k\}$ of temperatures becomes a geometric progression. Consequently, we can see by Theorem 1 that the set of temperatures should be set as a geometric progression in order to make the average exchange ratio constant over the various temperatures.

4. Experiments

In this section, we show the experimental results to verify the accuracy of our theoretical study.

4.1 Setting

Setting of target distribution: Let w be set as $\{w=A,B\}$, where A and B are respectively an $H \times M$ matrix and an $N \times H$ matrix. Then, the dimension d of w is $d=(M+N)H$. In our experiment, we considered the sampling from the following target distribution,

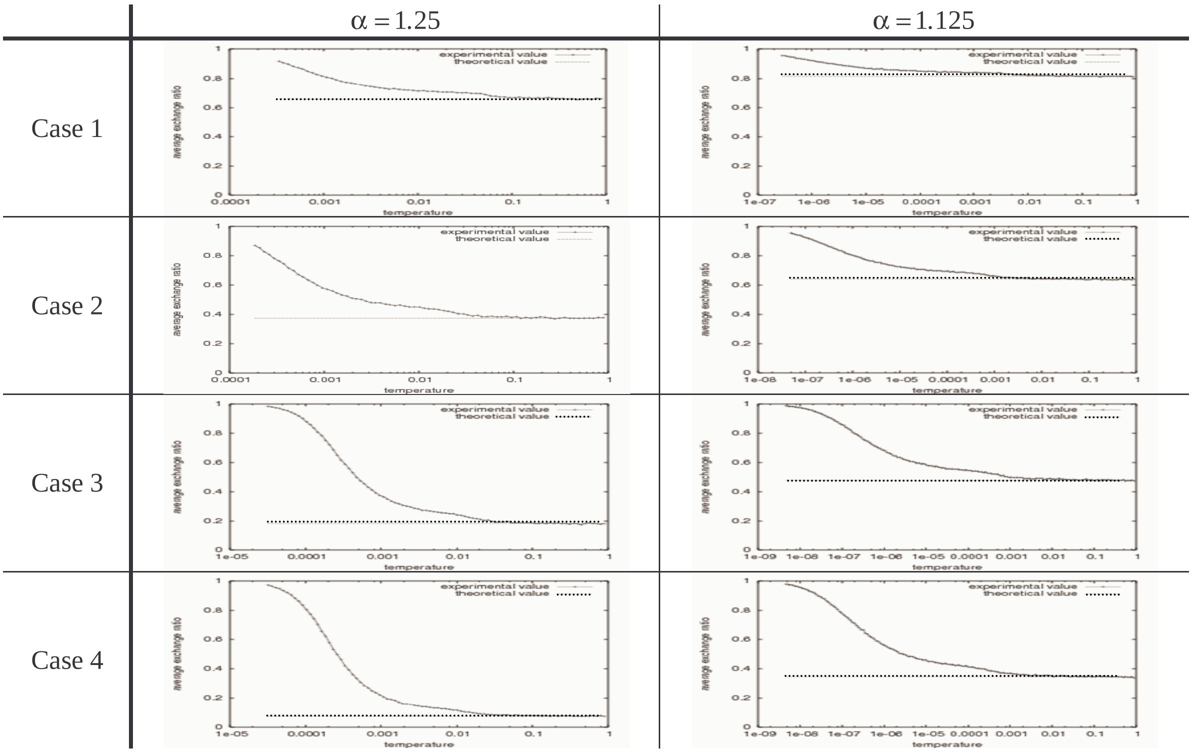

Figure 1. Theoretical value and experimental value of the average exchange ratio against the temperatures.

$$p(w) = \frac{1}{Z(n)} exp(-nf(w))\varphi(w).$$

Let the function $f(w)$ be given by

$$f(w) = \|BA - B_0 A_0\|^2,$$

where A_0 and B_0 are respectively an $H_0 \times M$ matrix and an $N \times H_0$ matrix. This setting corresponds to Bayesian learning in the reduced rank regression [1]. Here, M, N, H_0, and H respectively show the numbers of input units, output units, hidden units for a true structure, and hidden units for a learning machine. The number n was set as 1000 and the probability distribution $\varphi(w)$ was defined by the d -dimensional normal distribution whose mean and variance are respectively 0 and 10. The elements of matrix A_0 and B_0 were randomly chosen from the standard normal distribution.

In our experiment, we simulated the following four cases,

1. $M = N = 5, H_n = 1, H = 3$
2. $M = N = 10, H_n = 2, H = 6$
3. $M = N = 15, H_n = 3, H = 9$
4. $M = N = 20, H_n = 4, H = 12.$

In these cases, the value can be analytically calculated as follows [1],

$$\lambda = \frac{-(H_0 + H)^2 - M^2 - N^2}{8} + \frac{2(H_0 + H)M + 2(H_0 + H)N + 2MN}{8},$$

and each value becomes 8, 32, 72 and 128.

Setting of EMC method: Our theoretical result claims that the set of temperatures should be set as a geometric progression in order to make the average exchange ratio constant over the various temperatures. Therefore, in our experiment, we set each temperature $\{t_1,\cdots,t_K\}$ as follows,

$$t_k = \begin{cases} 0 & (\,if\ k=1) \\ \alpha^{-K+k} & (\,otherwise\,), \end{cases}$$

where α was set as 1.25 and 1.125. The total number K of temperatures was set as shown in Table 1.

Table 1. Setting of total number K of temperatures

	Case 1	Case 2	Case 3	Case 4
$\alpha=1.25$	70	75	90	90
$\alpha=1.125$	130	145	165	165

As a criterion for the iteration of EMC, we define "Monte Carlo Step (MCS)". Let 1 MCS means once simulating Step 1 and Step 2 for EMC algorithm. In Step 1, we used the Metropolis algorithm as a conventional MCMC method. The iteration for Step 1 was set as one. In Step 2, the rule for selecting exchange pairs was

$$\{(w_1,w_2),(w_3,w_4),\cdots\} \quad (\,if\ \ MCS\ is\ odd\,),$$
$$\{(w_2,w_3),(w_4,w_5),\cdots\} \quad (\,if\ \ MCS\ is\ even).$$

Average exchange ratio: In these settings, the value of $\frac{\Delta t}{t}$ is $\frac{\Delta t}{t}=\alpha-1.$ From the value of $\frac{\Delta t}{t}$ and λ, we calculate the theoretical value of average exchange ratio using numerical integration. Table 2 shows the theoretical value of average exchange ratio in each case.

Table 2. Theoretical value of average exchange ratio.

	Case 1	Case 2	Case 3	Case 4
$\alpha=1.25$	0.660716	0.374445	0.181824	0.074821
$\alpha=1.125$	0.816654	0.638955	0.480643	0.346667

The function $g\left(\lambda,\frac{\Delta t}{t}\right)$ was calculated by using the numerical integration.

For numerically calculating the average exchange ratio, let the average exchange ratio $J(t_k)\,\{k:1\le k\le K\}$ be defined as follows,

$$J(t_k)=\frac{1}{MCS}\sum_{i=1}^{MCS} r_i(t_k)$$

$$r(t_k)=\begin{cases}1 & (\text{if exchange between } w_k \text{ and } w_{k+1} \text{ is accepted}) \\ 0 & (\text{otherwise}).\end{cases}$$

4.2 Experimental result

Under these settings, we simulated some experiments.

Firstly, we verified the accuracy of our theorem. Figure 1 shows the experimental result of average exchange ratio. The horizontal axis shows the value of temperature and the vertical one the average exchange ratio. In this experiment, we set MCS 100000. In these figures, there are horizontal lines, which show the theoretical values of average exchange ratio. As we above mentioned, our theorem is applicable to the distributions which have temperatures larger than order of $\frac{1}{n}$. Hence, the experimental value of average exchange ratio can be seen to be equal to the theoretical value for large value of temperatures, in the range [0.05:1.0] in the case that $\alpha=1.25$, and in the range [0.001:1.0] in the case that $\alpha=1.125$. On the contrary, in the other range of each case, the experimental value of average exchange ratio is larger than the theoretical value.

Secondly, we studied the behavior of average exchange ratio as MCS increases. Figure 2 shows the average exchange ratios for some temperatures. The value of each temperature is in the range [0.05:1.0] in the case that

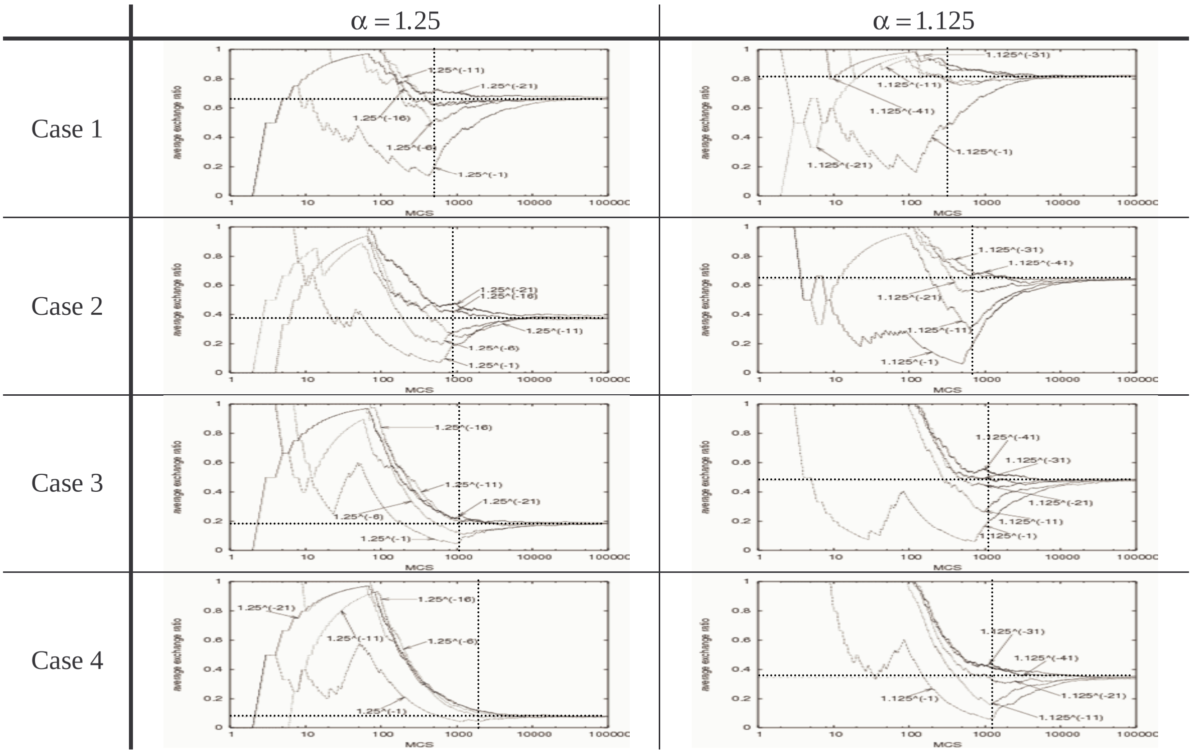

Figure 2. Convergence of the average exchange ratio. The horizontal lines in these figures show the theoretical value of average exchange ratio.

$\alpha = 1.25$ and in the range [0.001:1.0] in the case that $\alpha = 1.125$. This means that all the values of average exchange ratios in Figure 2 finally converge to the theoretical value because of the above experimental result. The horizontal axis shows the value of MCS and the vertical one the average exchange ratio. The horizontal lines on these figures show the theoretical values. From these figures, we can see that any average exchange ratios begin to converge to the theoretical value in a certain MCS, which is illustrated by the vertical line in these figures.

In order to clarify this point more clearly, we compared the average exchange ratio with the value of function $f(w)$. Figure 3 shows the comparison between the average exchange ratio and the function $f(w)$. The horizontal axis shows the value of MCS and vertical ones the average exchange ratio of the temperature $t = \alpha^{-1}$ and the value of function $f(w)$ of temperature $t = 1.0$. By comparing these functions, the MCS when the value of function $f(w)$ converges and the MCS when the average exchange ratio begins to converge are almost equal. Therefore, we can check the convergence of function $f(w)$ by monitoring the value of average exchange ratio. This fact can be used as a criterion for checking the convergence of EMC method.

5. Discussion

In this paper, we clarified the accuracy of our theoretical result by comparing the theoretical value of average exchange ratio to the experimental value and proposed the method to check the convergence of EMC method.

In the first experiment, we verified the fact that, in large value of temperatures, the experimental value of average exchange ratio is almost equal to the theoretical value, and that the average exchange ratios are almost constant over the various temperatures by setting the temperatures as a geometric progression. In addition, such range of temperatures depends on the value of α. This is because of the asymptotic property of Theorem 1, that is, the influence of lower terms of the asymptotic form of the average exchange ratio. This influence can be seen to be large as the interval Δt of temperature becomes large by this experiment.

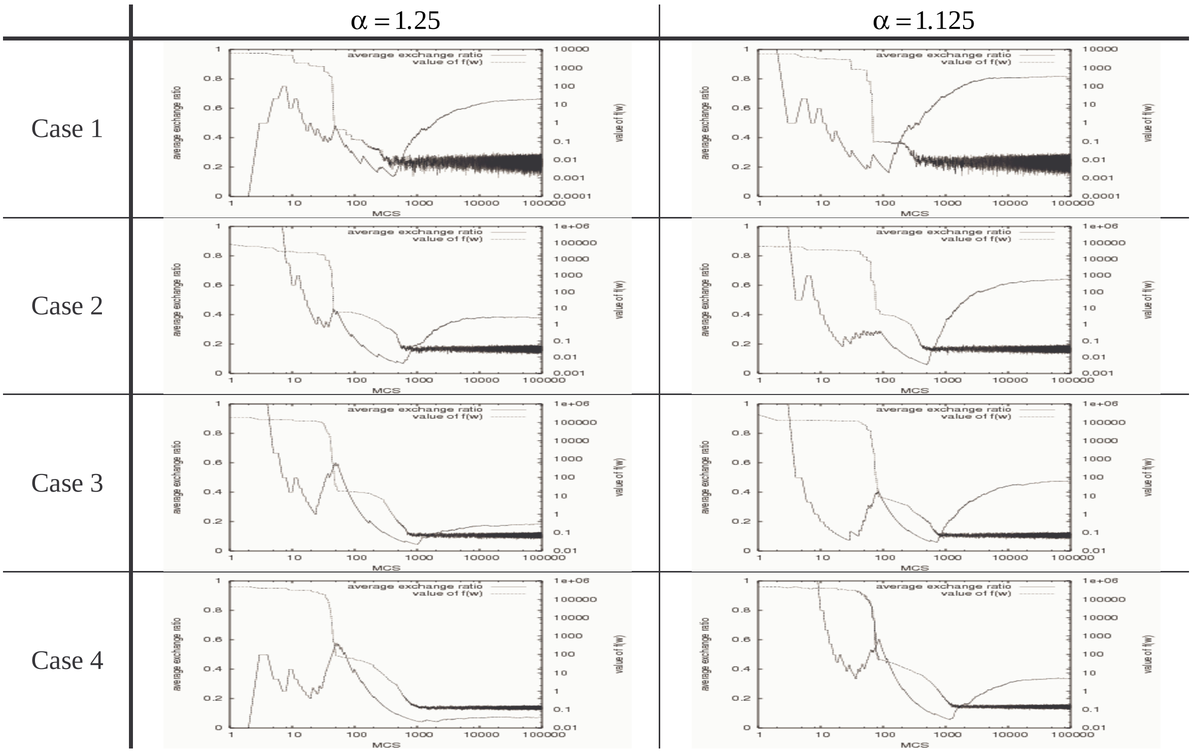

Figure 3. A comparison between the average exchange ratio and the value of function $f(w)$ against MCS.

On the contrary, in the small value of temperature, the average exchange ratio is larger than the theoretical value. In general, the behavior of average exchange ratio depends on the two distributions, $p(w|t=0)$ and $p(w|t=1)$. For our experience, if the peak(s) of the distribution $p(w|t)$ change as the value t of temperature increases from 0 to 1, the average exchange ratio rapidly decreases in certain temperature(s) by setting the temperatures as a geometric progression, which leads to inefficient EMC method. In our experiment, there is no temperature where the average exchange ratio rapidly decreases. This means that EMC method can work efficiently for Bayesian learning in reduced rank regression.

In the second experiment, we verified that the experimental values of average exchange ratios for any temperatures begin to converge to the theoretical value in a certain MCS, and that the value of function $f(w)$ also converges in this MCS. Based on these facts, we proposed the method to check the convergence of EMC method.

When discussing the convergence of EMC method, there are the following two problems. One is how many samples should be rejected in order to reduce the influence of initial value. The other is how many samples should be generated in order to approximate a target distribution accurately. These two problems are very important to generate a sample sequence from a target distribution accurately. Our proposed method, which is to monitor the average exchange ratios for some temperatures, is to overcome the problem 1. In general, a method to monitor the value of function $f(w)$ is often used in order to address the problem 1. However, it is not easy to check the convergence by this method because the value of function $f(w)$ after a sample converges is generally unknown. On the contrary, since the theoretical value of average exchange ratio is clarified, it is easy to check the convergence by our proposed method. Moreover, although the theoretical value cannot be calculated because the value of λ is unknown, our proposed method can be applied to checking the convergence by using the property that the average exchange ratios are almost constant over the various temperatures by setting the temperatures as a geometric progression. On the other hand, as a method to address the problem 2, a method to count the total samples which move from t_1 to t_K in a temperature space is often used. This method is considered to have close relation to the average exchange ratio for all temperatures. Hence, our theoretical result can be applied to addressing the problem 2, which should be addressed as a future work.

6. Conclusion

In this paper, we clarified the accuracy of our theoretical result by comparing the theoretical value of average exchange ratio to the experimental value. As a result, the following properties are verified that the experimental value of average exchange ratio is almost equal to the theoretical value, that the average exchange ratios are almost constant over the various temperatures by setting the temperatures as a geometrical progression, and that the experimental values of average exchange ratios for any temperatures begin to converge to the theoretical value in a certain MCS. Moreover, from these properties, we proposed the method to check the convergence of EMC method. As the future works, constructing the design of EMC method and applying these results to the practical problem should be addressed.

Acknowledgment: This work was partially supported by the Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for JSPS Fellows 18-5809, and for Scientific Research 18079007, 2007.

References

[1] M. Aoyagi, S. Watanabe, “Stochastic complexities of reduced rank regression in Bayesian estimation,” Neural Networks, Vol.18, No.7, pp.924-933, 2005.

[2] M. F. Atiyah, “Resolution of singularities and division of distributions,” Communications of Pure and Applied Mathematics, Vol.13, pp.145-150, 1970.

[3] B. A. Berg, T. Neuhaus, “Multicanonical algorithms for first order phase transitions,” Physics Letter B, Vol.267, No.2, pp.249-253, 1991.

[4] K. Hukushima, “Domain Wall Free Energy of Spin Glass Models: Numerical Method and Boundary Conditions,” Physical Review E, Vol.60, pp.3606-3613, 1999.

[5] K. Hukushima, K.Nemoto, “Exchange Monte Carlo Method and Application to Spin Glass Simulations,” Journal of Physical Society of Japan, Vol.65, No.6, pp.1604-1608, 1996.

[6] K. Hukushima, “Extended ensemble Monte Carlo approach to hardly relaxing problems,” Computer Physics Communications, Vol.147, pp.77-82, 2002.

[7] Y. Iba, “Extended Ensemble Monte Carlo,” International Journal of Modern Physics C, Vol.12, pp.623-656, 2001.

[8] F. Liang, “An effective Bayesian neural network classifier with a comparison study to support vector machine,” Neural Computation, Vol.15, pp.1959-1989, 2003.

[9] E. Marinari, G. Parisi, “Simulated tempering: a new Monte Carlo scheme,” Europhysics Letters, Vol.19, No.6, pp.451-455, 1992.

[10] K. Nagata, S. Watanabe, “Analysis of Exchange Ratio for Exchange Monte Carlo Method,” in Proc. of The First IEEE Symposium on Foundation of Computational Intelligence (FOCI'07), pp.434-439, 2007.

[11] K. Nagata, S. Watanabe, “Algebraic Geometric Study of Exchange Monte Carlo Method,” in Proc. of International Conference on Artificial Neural Networks (ICANN2007), to appear.

[12] K. Pinn, C. Wieczerkowski, “Number of Magic Squares from Parallel Tempering Monte Carlo,” International Journal of modern Physics, Vol.9, No.4, pp.541-546, 1998.

[13] Y. Sugita, Y. Okamoto, “Replica-exchange molecular dynamics method for protein folding,” Chemical Physics Letters, Vol.314, No.1-2, pp.141-151, 1999.

[14] S. Watanabe, “Algebraic Analysis for Nonidentifiable Learning Machines,” Neural Computation, Vol.13, pp.899-933, 2001.

[15] K. Yamazaki, S. Watanabe, “Singularities in mixture models and upper bounds of stochastic complexity,” Neural networks, Vol.16, No.7, pp.1029-1038, 2003.

Kenji Nagata received the B.E. degree from Dept. of Computer Science, Tokyo Institute of Technology, and the M.E. degree from Dept. of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Japan. He is currently a PhD student of Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology. His research interest includes learning theory with hierarchical learning machine and the theory and application of Markov chain Monte Carlo method.

Sumio Watanabe was born in Japan, 1959. He received the Ph. D. degree in applied electronics in 1993 from Tokyo Institute of Technology, Japan. Dr. Watanabe is currently a professor at Precision and Intelligence Laboratory in Tokyo Institute of Technology. He is interested in mathematical learning theory.