Analysis of Exchange Ratio for Exchange Monte Carlo Method

Kenji Nagata, and, Sumio Watanabe, *Member, IEEE*

***Abstract*— The exchange Monte Carlo method was proposed as an improved algorithm of Markov Chain Monte Carlo method and its effectiveness has been shown in many fields. In the exchange Monte Carlo method, the setting of temperatures is important to make the algorithm efficient because this setting controls the exchange ratio, with which the position exchange between two sequences is accepted. However, the mathematical foundation of exchange MC method has not yet been established. In this paper, we rigorously prove the mathematical relation between the symmetrized Kullback divergence and the exchange ratio, by which the optimal setting of temperatures is devised.**

I. INTRODUCTION

The Markov Chain Monte Carlo (MCMC) method is well known as an algorithm to generate the sample sequence which converges to a target distribution. This algorithm is widely used in many fields such as statistics, bioinformatics and information science. However, it requires huge computational cost to generate the sample sequence, in particular, in the case that the target distribution has high potential barriers [6] and that the ground state of target distribution is not one point but an analytic set [9].

Recently, various improvements of MCMC method have been developed based on the idea of extended ensemble methods, which are surveyed in [7]. These methods give us a general strategy to overcome the problem of huge computational cost. An exchange Monte Carlo (MC) method is well known as one of the extended ensemble methods [6]. This method is to generate the sample sequence from a joint distribution, which consists of many distributions with different temperatures. Its algorithm is based on two steps of MCMC simulations. One is the conventional update of MCMC simulation for each distribution. The other is the exchange process between two sequences with a certain probability. As for the exchange MC method, its effectiveness has been shown in an optimization problem [5][10], a protein-folding problem [11] and the Bayesian learning in hierarchical learning machines [8].

When we design the exchange MC method, the setting of temperature is very important [4]. The values of temperature have close relation to the exchange ratio and its average, with which the exchange between two sequence is accepted. In order to make the exchange MC method efficient, the exchange ratio needs to be not low and not too high. Therefore, the optimal setting of temperature enables us to design the efficient exchange MC method. The symmetrized Kullback divergence between two distributions with different temperature is used as a criterion for the setting of temperature because this Kullback divergence has relation to the average exchange ratio [7]. Based on this fact, the design method for setting of temperature has been proposed. However, this method needs some previous simulations. Moreover, the mathematical relation between the symmetrized Kullback divergence and the average exchange ratio has not been clarified.

In this paper, we mathematically calculate the symmetrized Kullback divergence and the average exchange ratio, and clarify the relation between the symmetrized Kullback divergence and the average exchange ratio.

This paper consists of six chapters. In Chapter II and III, we explain the framework of exchange MC method and the design of exchange MC method. In Chapter IV, the main result of this paper is described. Discussion and Conclusion are followed in Chapter V and VI.

II. EXCHANGE MONTE CARLO METHOD

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

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

where $Z(n)$ is the normalization constant. The exchange MC method treats a compound system which consists of non-interacting K sample sequences of the system concerned. The k-th sample sequence $\{w_k\}$ converges in law to the random variable which is subject to the following probability distribution

$$p(w|t_k) = \frac{1}{Z(nt_k)} \exp(-nt_k H(w))\varphi(w) \quad (1 \le l \le L),$$

where $t_1 < t_2 < \cdots < t_L$. Given a set of the temperatures $\{t\} = \{t_1, \cdots, t_K\}$, the simultaneous distribution for finding $\{w\} = \{w_1, w_2, \cdots, w_K\}$ is expressed as a simple product formula

$$p(\{w\}) = \prod_{k=1}^{K} p(w_k|t_k). \tag{1}$$

The exchange MC method is based on two types of updating in constructing a Markov chain. One is conventional updates of MCMC simulation such as Gibbs sampler and Metropolis algorithm for each target distribution $p(w_k|t_k)$. The other is the position exchange between two sequences, that is,

Kenji Nagata is with the Department of Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, MailBox R2-5,4259 Nagatsuda, Midori-ku, Yokohama, 226-8503 Japan (phone.+81-45-924-5018, e-mail kenji.nagata@cs.pi.titech.ac.jp).

Sumio Watanabe is with the P&I Lab.,Tokyo Institute of Technology, MailBox R2-5,4259 Nagatsuda, Midori-ku, Yokohama,226-8503@Japan (phone.+81-45-924-5017, e-mail swatanab@pi.titech.ac.jp).

$\{w_k, w_{k+1}\} \to \{w_{k+1}, w_k\}$. The transition probability u is defined as

$$\begin{aligned} u &= \min(1, r), \\ r &= \frac{p(w_{k+1}|t_k)p(w_k|t_{k+1})}{p(w_k|t_k)p(w_{k+1}|t_{k+1})} \\ &= \exp(n(t_{k+1} - t_k)(H(w_{k+1}) - H(w_k))). \end{aligned} \quad (2)$$

Hereafter, we call u exchange ratio. Under these updates, the simultaneous distribution of Eq.(1) is invariant because these updates satisfy the detailed balance condition for the distribution of Eq.(1) [6].

Consequently, the following two steps are carried out alternately:

1) Each sequence is generated simultaneously and independently for a few iteration by conventional MCMC method.
2) Two positions are exchanged with the exchange ratio u.

The advantage of exchange MC method is to make the convergence of sample sequence earlier comparing the conventional MCMC method. A disadvantage of the conventional MCMC method is that it requires huge computational cost to generate sample sequence from the target distribution. The reason is that the sample sequence is hard to converge in law to the target distribution, in particular, in the case that the target distribution has high potential barriers and that the ground state of target distribution is not one point but an analytic set. The exchange MC method can realize the efficient sampling by preparing a simple distribution such as a normal distribution, which is easy for sample sequence to converge. In practical, we set the temperature of target distribution as $t_K = 1$, and that of simple distribution as $t_1 = 0$.

III. Design of Exchange Monte Carlo Method

When we design the exchange MC method, the setting of temperature is very important to make the exchange MC method efficient. As we can see in Eq.(2), temperature has close relation to the exchange ratio. Therefore, temperature is a very important parameter in adjusting the exchange ratio and its average.

For the efficient exchange MC method, the set of temperature needs to optimize so that the average exchange ratio for two adjacent distribution become not low and not too high. In order to carry out the efficient exchange MC method, the time for a sample to move from end to end (from t_1 to t_K) in the space of temperature is good to be short. Therefore, it is not efficient for the average exchange ratio to be low. On the contrary, to make the average exchange ratio high, the range of temperature has to be very small, that is, the total number K of temperature has to be large. Therefore, this setting is not also efficient because it needs huge cost to generate the sample from each distribution. In practical, the set of temperature is configured so that the average exchange ratio becomes equal for all combinations of distributions.

As the criterion for setting of temperature, The following symmetrized Kullback divergence $I(t_k, t_{k+1})$ is used [7],

$$\begin{aligned} I(t_k, t_{k+1}) &= \int p(w_k|t_k) \log \frac{p(w_k|t_k)}{p(w_k|t_{k+1})} dw_k \\ &+ \int p(w_{k+1}|t_{k+1}) \log \frac{p(w_{k+1}|t_{k+1})}{p(w_{k+1}|t_k)} dw_{k+1}. \end{aligned}$$

This function has the following property,

$$E[\log r] = -I(t_k, t_{k+1}),$$

where $E[\log r]$ means the average of $\log r$ over the joint distribution $p(w_k|t_k) \times p(w_{k+1}|t_{k+1})$. Moreover, when the free energy $F(nt)$ is defined as follows,

$$F(nt) = -\log \int \exp(-ntH(w))\varphi(w)dw,$$

the following equation is satisfied in small range of temperature,

$$I(t_k, t_{k+1}) = \frac{\partial^2 F}{\partial t^2}\Big|_{t=t_k} (t_{k+1} - t_k)^2.$$

Hence, the set of temperature can be set so that the symmetrized Kullback divergence becomes constant by setting the range of temperature in inverse proportion to $\sqrt{\partial^2 F / \partial t^2}$. However, the mathematic definition of average exchange ratio $J(t_k, t_{k+1})$ is as follows,

$$\begin{aligned} J(t_k, t_{k+1}) &= E[u] \\ &= \int\int u P(w_k|t_k) P(w_{k+1}|t_{k+1}) dw_k dw_{k+1}, \end{aligned}$$

whose property is not clarified. Therefore, the mathematical relation between the symmetrized Kullback divergence and the average exchange ratio has not been analytically clarified.

In this paper, we show the analytical results for the symmetrized Kullback divergence and for the average exchange ratio in the low temperature limit, that is, $n \to \infty$. These results reveal the mathematical relation between the symmetrized Kullback divergence and the average exchange ratio, and can be used as the criterion for setting of temperature.

IV. Main Result

In this section, we consider the exchange MC method between the following two distributions,

$$\begin{aligned} p_1(w) &= \frac{1}{Z(nt)} \exp(-ntH(w))\varphi(w) \\ p_2(w) &= \frac{1}{Z(n(t+\Delta t))} \exp(-n(t+\Delta t)H(w))\varphi(w). \end{aligned}$$

For these distributions, the symmetrized Kullback divergence I and the average exchange ratio J are rewritten as follows,

$$\begin{aligned} I &= \int p_1(w_1) \log \frac{p_1(w_1)}{p_2(w_1)} dw_1 \\ &\quad + \int p_2(w_2) \log \frac{p_2(w_2)}{p_1(w_2)} dw_2 \\ J &= \int\int u p_1(w_1) p_2(w_2) dw_1 dw_2, \end{aligned}$$

where u is a function of w_1 and w_2 as Eq.(2).

In general, the distribution $p_1(w)$ and $p_2(w)$ are not asymptotic to the normal distribution as $n \to \infty$ because the property that the hessian of energy function $H(w)$ is positive is not satisfied. We can assume $H(w) \geq 0$ and $H(w_0) = 0 (\exists w_0)$ without loss of generality. The zeta function of $H(w)$ and $\varphi(w)$ is defined by

$$\zeta(z) = \int H(w)^z \varphi(w) dw,$$

where z is a one complex variable. $\zeta(z)$ is a holomorphic function in the region of $Re(z) > 0$, and can be analytically continued to the meromorphic function on the entire complex plane, whose poles are all real, negative, and rational numbers [3] [13]. We also define the rational number $-\lambda$ as the largest pole of zeta function $\zeta(z)$ and the natural number m as its order.

Then, the following theorem about the symmetrized Kullback divergence can be obtained.

Theorem 1: The symmetrized Kullback divergence converges to the following equation as $n \to \infty$,

$$I = \lambda \left(\frac{\Delta t}{t}\right)^2 \left(1 - \frac{\Delta t}{t} + O\left(\left(\frac{\Delta t}{t}\right)^2\right)\right).$$

(proof) The symmetrized Kullback divergence I is equal to,

$$I = n\Delta t \left\{ \int H(w) p_1(w) dw - \int H(w) p_2(w) dw \right\},$$

because of the following equations,

$$\begin{aligned} \log \frac{p_1(w)}{p_2(w)} &= \log Z(n(t+\Delta t)) - \log Z(nt) + n\Delta t H(w), \\ \log \frac{p_2(w)}{p_1(w)} &= \log Z(nt) - \log Z(n(t+\Delta t))) - n\Delta t H(w). \end{aligned}$$

Therefore, by defining

$$K(p) = n\Delta t \int H(w) p(w) dw,$$

we obtain $I = K(p_1) - K(p_2)$. Firstly, we analyze the functional $K(p_1)$. The functional $K(p_1)$ is expressed by using the Dirac delta function $\delta(s)$ as follows,

$$\begin{aligned} K(p_1) &= n\Delta t \int H(w) p_1(w) dw \\ &= n\Delta t \frac{\int H(w) \exp(-ntH(w)) \varphi(w) dw}{\int \exp(-ntH(w)) \varphi(w) dw} \\ &= n\Delta t \frac{\int_0^\infty s e^{-nts} ds \int \delta(s - H(w)) \varphi(w) dw}{\int_0^\infty e^{-nts} ds \int \delta(s - H(w)) \varphi(w) dw}. \end{aligned}$$

The integration for w in the above equation is well known as the state density function $V(s)$. This function has the following asymptotic form for $s \to 0$ [12][13],

$$\begin{aligned} V(s) &= \int_0^\infty \delta(s - H(w)) \varphi(w) dw \\ &\cong c s^{\lambda - 1} (-\log s)^{m-1}, \end{aligned} \qquad (3)$$

where the real number c is the function of $H(w)$ and $\varphi(w)$. From this equation and by putting $s' = nts$,

$$\begin{aligned} K(p_1) &= n\Delta t \frac{\int \left(\frac{s'}{nt}\right)^\lambda e^{-s'} (\log nt - \log s')^{m-1} \frac{ds'}{nt}}{\int \left(\frac{s'}{nt}\right)^{\lambda-1} e^{-s'} (\log nt - \log s')^{m-1} \frac{ds'}{nt}} \\ &= \frac{\Delta t}{t} \frac{\int e^{-s'} s'^\lambda \left(1 + O\left(\frac{1}{\log nt}\right)\right) ds'}{\int e^{-s'} s'^{\lambda-1} \left(1 + O\left(\frac{1}{\log nt}\right)\right) ds'} \\ &= \frac{\Delta t}{t} \frac{\Gamma(\lambda+1) + O\left(\frac{1}{\log nt}\right)}{\Gamma(\lambda) + O\left(\frac{1}{\log nt}\right)} \\ &\to \frac{\Delta t}{t} \lambda, \end{aligned}$$

where $\Gamma(\lambda)$ is the gamma function. In the same way, the functional $K(p_2)$ is given by

$$\begin{aligned} K(p_2) &= n\Delta t \int H(w) p_2(w) dw \\ &\to \frac{\Delta t}{t + \Delta t} \lambda. \end{aligned}$$

Consequently, the symmetrized Kullback divergence I for $n \to \infty$ is obtained as follows,

$$\begin{aligned} I &= \lambda \left(\frac{\Delta t}{t} - \frac{\Delta t}{t + \Delta t}\right) \\ &= \lambda \frac{\Delta t^2}{t(t+\Delta t)} \\ &= \lambda \left(\frac{\Delta t}{t}\right)^2 \left(1 - \frac{\Delta t}{t} + O\left(\left(\frac{\Delta t}{t}\right)^2\right)\right), \end{aligned}$$

which completes the theorem.(Q.E.D)

From Theorem 1, we can make the symmetrized Kullback divergence for arbitrary temperature t constant by the temperature setting that the value $\frac{\Delta t}{t}$ becomes constant, that is, the set $\{t_k\}$ of temperature is set as geometrical progression. In particular, if the value $\frac{\Delta t}{t}$ is small, the symmetrized Kullback divergence can be made constant value a by setting Δt as follows,

$$\Delta t = t \sqrt{\frac{a}{\lambda}}.$$

However, it is not clear that the average exchange ratio becomes equal for arbitrary temperature t by the above temperature setting. Moreover, the relation betweeen the values of symmetrized Kullback divergence and of average exchange ratio is not also clarified.

Next, we analyze the average exchange ratio.

Theorem 2: The average exchange ratio J converges to the following equation as $n \to \infty$,

$$J = 1 - \frac{|\Delta t|}{t} \frac{\Gamma(\lambda + \frac{1}{2})}{\sqrt{\pi}\Gamma(\lambda)} + O\left(\left(\frac{\Delta t}{t}\right)^2\right).$$

(proof)In the case that $\Delta t \geq 0$, the average exchange ratio J is expressed by using the definition of exchange ratio u as follows,

$$\begin{aligned} J &= \int\int_{H(w_1)<H(w_2)} p_1(w_1)p_2(w_2)dw_1dw_2 \\ &\quad + \int\int_{H(w_1)\geq H(w_2)} rp_1(w_1)p_2(w_2)dw_1dw_2 \\ &= 2\int\int_{H(w_1)<H(w_2)} p_1(w_1)p_2(w_2)dw_1dw_2 \\ &= 2\int\int_{H(w_1)<H(w_2)} \frac{e^{-ntH(w_1)}\varphi(w_1)}{Z(nt)} \\ &\quad \times \frac{e^{-n(t+\Delta t)H(w_2)}\varphi(w_2)}{Z(n(t+\Delta t))}dw_1dw_2. \end{aligned}$$

In the case that $\Delta t < 0$, J is given by

$$\begin{aligned} J &= 2\int\int_{H(w_1)>H(w_2)} \frac{e^{-ntH(w_1)}\varphi(w_1)}{Z(nt)} \\ &\quad \times \frac{e^{-n(t+\Delta t)H(w_2)}\varphi(w_2)}{Z(n(t+\Delta t))}dw_1dw_2. \end{aligned}$$

Thus, we analyze in the case that $\Delta t \geq 0$. In the same way as the analysis of Theorem 1, the following equation holds,

$$Z(nt) \to \frac{c(\log nt)^{m-1}}{(nt)^\lambda}\Gamma(\lambda).$$

Hence, by defining

$$\begin{aligned} J^* &= \int\int_{H(w_1)<H(w_2)} e^{-ntH(w_1)}\varphi(w_1) \\ &\quad \times e^{-n(t+\Delta t)H(w_2)}\varphi(w_2)dw_1dw_2, \end{aligned}$$

we obtain

$$J = 2\frac{J^*}{Z(nt)Z(n(t+\Delta t))}.$$

The function J^* is expressed by using the Dirac delta function,

$$\begin{aligned} J^* &= \int_0^\infty ds_2 \int_0^{s_2} ds_1 e^{-nts_1}e^{-n(t+\Delta t)s_2} \\ &\quad \times \int \delta(s_1 - H(w_1))\varphi(w_1)dw_1 \\ &\quad \times \int \delta(s_2 - H(w_2))\varphi(w_2)dw_2 \\ &= \int_0^\infty ds_2 \int_0^{s_2} ds_1 e^{-nts_1}e^{-n(t+\Delta t)s_2} \\ &\quad \times cs_1^{\lambda-1}(-\log s_1)^{m-1}cs_2^{\lambda-1}(-\log s_2)^{m-1}. \end{aligned}$$

By putting $s_1 = s_1's_2$ and $s_2' = nts_2$, it follows that,

$$\begin{aligned} J^* &= \int_0^\infty ds_2 \int_0^1 ds_1' e^{-nts_1's_2}e^{-n(t+\Delta t)s_2} \\ &\quad \times cs_1'^{\lambda-1}(-\log s_1's_2)^{m-1}cs_2^{2\lambda-1}(-\log s_2)^{m-1} \\ &= \int_0^\infty \frac{ds_2'}{nt}\int_0^1 ds_1' e^{-s_1's_2'}e^{-\left(1+\frac{\Delta t}{t}\right)s_2'} \\ &\quad \times cs_1'^{\lambda-1}(\log nt - \log s_1's_2')^{m-1} \\ &\quad \times c\left(\frac{s_2'}{nt}\right)^{2\lambda-1}(\log nt - \log s_2')^{m-1} \\ &\to \frac{c^2(\log nt)^{2(m-1)}}{(nt)^{2\lambda}}\Bigg\{O\left(\left(\frac{\Delta t}{t}\right)^2\right) \\ &\quad + \int_0^\infty ds_2' \int_0^1 ds_1' e^{-(1+s_1')s_2'}s_1'^{\lambda-1}s_2'^{2\lambda-1} \\ &\quad - \frac{\Delta t}{t}\int_0^\infty ds_2' \int_0^1 ds_1' e^{-(1+s_1')s_2'}s_1'^{\lambda-1}s_2'^{2\lambda}\Bigg\}. \end{aligned}$$

In the above analysis, we use the Taylor expansion of $e^{-\left(1+\frac{\Delta t}{t}\right)s_2'}$ for $s_2' = 0$ as follows,

$$e^{-s_2'}\left(1 - \frac{\Delta t}{t}s_2' + O\left(\left(\frac{\Delta t}{t}\right)^2\right)\right).$$

Also by putting $s_2'' = (1+s_1')s_2'$,

$$\begin{aligned} J^* &= \frac{c^2(\log nt)^{2(m-1)}}{(nt)^{2\lambda}}\Bigg\{O\left(\left(\frac{\Delta t}{t}\right)^2\right) \\ &\quad + \int_0^\infty e^{-s_2''}(s_2'')^{2\lambda-1}ds_2'' \int_0^1 \frac{(s_1')^{\lambda-1}}{(1+s_1')^{2\lambda}}ds_1' \\ &\quad - \frac{\Delta t}{t}\int_0^\infty e^{-s_2''}(s_2'')^{2\lambda}ds_2'' \int_0^1 \frac{(s_1')^{\lambda-1}}{(1+s_1')^{2\lambda+1}}ds_1'\Bigg\} \end{aligned}$$

By using the following equations,

$$\begin{aligned} \int_0^1 \frac{(s_1')^{\lambda-1}}{(1+s_1')^{2\lambda}}ds &= \frac{\Gamma(\lambda)^2}{2\Gamma(2\lambda)} \\ \int_0^1 \frac{(s_1')^{\lambda-1}}{(1+s_1')^{2\lambda+1}}ds &= \frac{\lambda\Gamma(\lambda)^2}{2\Gamma(2\lambda+1)} + \frac{1}{2\lambda 4^\lambda}, \end{aligned}$$

the function J^* is given by

$$\begin{aligned} J^* &= \frac{(c)^2(\log nt)^{2(m-1)}}{(nt)^{2\lambda}}\Bigg\{\frac{\Gamma(\lambda)^2}{2} \\ &\quad - \frac{\Delta t}{t}\left(\frac{\lambda\Gamma(\lambda)^2}{2} + \frac{\Gamma(2\lambda)}{4^\lambda}\right) + O\left(\left(\frac{\Delta t}{t}\right)^2\right)\Bigg\}. \end{aligned}$$

Consequently, the average exchange ratio J is,

$$\begin{aligned} J &= 2\frac{J^*}{Z(nt)Z(n(t+\Delta t))} \\ &\to \left(1+\frac{\Delta t}{t}\right)^\lambda \\ &\quad \times \left(1 - \frac{\Delta t}{t}\left(\lambda + \frac{\Gamma(\lambda+\frac{1}{2})}{\sqrt{\pi}\Gamma(\lambda)}\right) + O\left(\left(\frac{\Delta t}{t}\right)^2\right)\right) \\ &\cong 1 - \frac{\Delta t}{t}\frac{\Gamma(\lambda+\frac{1}{2})}{\sqrt{\pi}\Gamma(\lambda)} + O\left(\left(\frac{\Delta t}{t}\right)^2\right). \qquad (4) \end{aligned}$$

In the case that $\Delta t < 0$, the function J can be calculated by the same analysis as,

$$J \cong 1 + \frac{\Delta t}{t}\frac{\Gamma(\lambda+\frac{1}{2})}{\sqrt{\pi}\Gamma(\lambda)} + O\left(\left(\frac{\Delta t}{t}\right)^2\right). \quad (5)$$

From eq.(4) and (5), we prove this theorem. (Q.E.D)

From Theorem 2, the average exchange ratio becomes constant for arbitrary temperature t by setting the value $\frac{\Delta t}{t}$ constant, that is to say, It is clarified that the exchange ratio becomes constant if we set the temperature in order to make the symmetrized Kullback divergence constant. In particular, in small $\frac{\Delta t}{t}$, when the value of symmetrized Kullback divergence is a, the value of average exchange ratio is as follows,

$$J = 1 - \sqrt{\frac{a}{\lambda}}\frac{\Gamma(\lambda+\frac{1}{2})}{\sqrt{\pi}\Gamma(\lambda)},$$

which depends on not only the value a but also λ. Moreover, comparing these theorems, we can see the difference that the average exchange ratio is expressed by the linear function of Δt which has the maximum value 1 at $\Delta t = 0$, while the symmetrized Kullback divergence is the quadratic which has the minimum value 0 at $\Delta t = 0$.

V. Discussion

In this paper, we analyzed the symmetrized Kullback divergence and the average exchange ratio in the low temperature limit, and clarified the mathematical relation between two functions. As results, the following three properties are clarified,

1) When the symmetrized Kullback divergence for arbitrary temperature t is constant, the average exchange ratio is also constant.
2) Then, the set of temperature $\{t_k\}$ is set as geometrical progression.
3) The symmetrized Kullback divergence and the average exchange ratio has the difference between linear and quadratic.

Let us discuss two points in association to this paper.

Firstly, we discuss the relation between the shape of target distribution and the setting of temperature. As we can see in Theorem 2, once the set of temperature $\{t_k\}$ is determined, the value of average exchange ratio depends on the value λ. The average exchange ratio becomes small for the distribution with large value λ because the coefficient of linear term for the average exchange ratio is the monotonically increasing function for λ. In fact, when the function $A(\lambda)$ is defined by,

$$A(\lambda) = \frac{\Gamma(\lambda+\frac{1}{2})}{\sqrt{\pi}\Gamma(\lambda)},$$

it is satisfied that,

$$\frac{dA(\lambda)}{d\lambda} = \frac{\Gamma(\lambda+\frac{1}{2})}{\sqrt{\pi}\Gamma(\lambda)}\left(\psi(\lambda+\frac{1}{2}) - \psi(\lambda)\right) > 0.$$

Therefore, for the target distribution with small value λ, the exchange MC method can work more efficiently. On the other hand, by comparing two distributions whose ground state are one point and analytic set in the sample space, the latter distribution is well known to have smaller value λ than the former distribution [2][14]. Consequently, the exchange MC method can work efficiently for the target distribution with the energy function whose ground state is analytic set. As an example of such distributions, the Bayesian posterior distribution in hierarchical learning machines such as neural networks and normal mixtures is well known, and this theorem shows the availability of exchange MC method for the Bayesian learning in hierarchical learning machines.

Secondly, we discuss the design of exchange MC method. This theorem gives us the design method of optimal temperature in order to make the average exchange ratio constant. However, the optimum value of average exchange ratio is not clarified, which leads to the design of optimal number K of temperature. Moreover, since the exchange MC method includes the algorithm of conventional MCMC method, the design of conventional MCMC method should be considered in the future.

VI. Conclusion

In this paper, we analytically calculated the symmetrized Kullback divergence and the average of exchange ratio, and clarified the relation between the symmetrized Kullback divergence and the exchange ratio. As results, the following properties are clarified,

1) When the symmetrized Kullback divergence for arbitrary temperature t is constant, the average exchange ratio is also constant.
2) Then, the set of temperature $\{t_k\}$ is set as geometrical progression.
3) The symmetrized Kullback divergence and the average exchange ratio has the difference between linear and quadratic.

As the future works, verifying the theoretical result obtained by this study by some experiments, constructing the design of exchange MC method, and applying these results to the practical problems such as the Bayesian learning should be addressed.

VII. Acknowledgement

This work was supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for JSPS Research Fellows 18-5809 and for Scientific Research 18009007.

References

[1] H.Alzer, “On some inequalities for the Gamma and Psi functions,” Mathematics of computation, 66 (217), pp.373-389, 1997.
[2] M.Aoyagi, S.Watanabe, “Stochastic complexities of reduced rank regression in Bayesian estimation,” Neural Networks, Vol.18, No.7, pp.924-933, 2005.
[3] M. F. Atiyah, “Resolution of singularities and division of distributions,” Commun. Pure Appl. Math., vol. 13, pp. 145-150, 1970.
[4] K.Hukushima, “Domain-Wall Free-Energy of Spin Glass Models: Numerical Method and Boundary Conditions,” Physical Review E 60, pp.3606-3613, 1999.

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

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

[7] Y.Iba, “Extended Ensemble Monte Carlo”, International Journal of Modern Physics, C12, pp.623-656, 2001.

[8] K.Nagata, S.Watanabe, “The Exchange Monte Carlo Method for Bayesian Learning in Singular Learning Machines,” Proceedings of World Congress on Computational Intelligence 2006 (WCCI2006), pp.6383-6389, 2006.

[9] N.Nakano, K.Takahashi, S.Watanabe, “On the Evaluation Criterion of the MCMC Method in Singular Learning Machines” , Trans. of IEICE, Vol.J88-D-2, No.10, pp.2011-2020, 2005.

[10] K.Pinn, C.Wieczerkowski, “Number of magic squares from parallel tempering Monte Carlo,” Int. J. Mod. Phys. C9, 541, 1998.

[11] Y.Sugita, Y.Okamoto, “Replica-exchange molecular dynamics method for protain folding,” Chemical Physics Letters, 314, pp.141-151, 1999.

[12] S.Watanabe, “Algebraic analysis for nonidentifiable learning machines,” Neural Computation, Vol.13, No.4, pp.899-933, 2001.

[13] S.Watanabe, “Algebraic geometrical methods for hierarchical learning machines,” Neural Networks, Vol.14, No.8, pp.1049-1060, 2001.

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