Algebraic Geometric Study of Exchange Monte Carlo Method

Kenji Nagata[1] and Sumio Watanabe[2]

[1] Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, MailBox R2-5,4259, Nagatsuta, Midori-ku, Yokohama, 226-8503 Japan
kenji.nagata@cs.pi.titech.ac.jp
[2] P&I Lab.,Tokyo Institute of Technology
swatanab@pi.titech.ac.jp

Abstract. In hierarchical learning machines such as neural networks, Bayesian learning provides better generalization performance than maximum likelihood estimation. However, its accurate approximation using Markov chain Monte Carlo (MCMC) method requires huge computational cost. The exchange Monte Carlo (EMC) method was proposed as an improved algorithm of MCMC method. Although its effectiveness has been shown not only in Bayesian learning but also in many fields, the mathematical foundation of EMC method has not yet been established. In this paper, we clarify the asymptotic behavior of symmetrized Kullback divergence and average exchange ratio, which are used as criteria for designing the EMC method.

1 Introduction

A lot of learning machines with hierarchical structures such as neural networks, 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 [12][13].

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

Recently, various improvements of MCMC methods have been developed based on the idea of extended ensemble, which are surveyed in [6]. Multicanonical method [2] and simulated tempering [8] belong to this category called the extended ensemble method. This idea gives us a general strategy to overcome the problem of huge computational cost. The exchange Monte Carlo (EMC) method is well known as one of the extended ensemble methods [4]. This method is to

generate a sample sequence from the 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 neighboring sequences with a certain probability. The EMC method has been successfully applied to not only Bayesian neural network learning in the literature [7] but also an optimization problem [5][10] and a protein-folding problem [11].

When we design the EMC method, the setting of temperature is very important to make this algorithm efficient [3]. The values of temperatures have close relation to the exchange ratio and its average, which is the acceptance ratio of exchange process. The symmetrized Kullback divergence between two neighboring distributions is used as a criterion for setting temperatures because this Kullback divergence has relation to the average exchange ratio [6]. However, the mathematical relation between the symmetrized Kullback divergence and the average exchange ratio has not been clarified. Moreover, there are two types of exchange ratio, Metropolis type and heat bath type. It often becomes a problem which type should be used.

In this paper, we mathematically clarify the symmetrized Kullback divergence and the average exchange ratio for Metropolis type and for heat bath type. This analytic result gives us the mathematical relation between the symmetrized Kullback divergence and the average exchange ratio, the optimal setting of temperatures, and the properties of the average exchange ratio for Metropolis type and for heat bath type.

This paper consists of five chapters. In Chapter 2, we explain the framework of EMC method and the design of EMC method. In Chapter 3, the main result of analysis for the EMC method is described. Discussion and Conclusion are followed in Chapter 4 and 5.

2 Background

2.1 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 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. In Bayesian learning, in order to calculate the expectation over the Bayesian posterior distribution, a sample sequence from the posterior distribution is needed. Then, the number n, the function $H(w)$ and the probability distribution $\varphi(w)$ respectively correspond to the number of training data, the log likelihood and the prior distribution. 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

$$p(w|t_k) = \frac{1}{Z(nt_k)} \exp(-nt_k H(w))\varphi(w) \quad (1 \leq k \leq K),$$

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

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

The EMC method is based on two types of updating in constructing a Markov chain. One is the conventional updates based on the Metropolis algorithm for each target distribution $p(w_k|t_k)$. The other is the position exchange between two sequences, that is, $\{w_k, w_{k+1}\} \to \{w_{k+1}, w_k\}$. The transition probability u is determined by the detailed balance condition for the joint distribution (1). We have two types of position exchange. One is the Metropolis type as follows,

$$\begin{aligned} u_1 &= \min(1, r) \\ r &= \frac{p(w_{k+1}|t_k)p(w_k|t_{k+1})}{p(w_k|t_k)p(w_{k+1}|t_{k+1})} = \exp(n(t_{k+1} - t_k)(H(w_{k+1}) - H(w_k))). \end{aligned} \tag{2}$$

The other is the heat bath type,

$$\begin{aligned} u_2 &= \frac{p(w_{k+1}|t_k)p(w_k|t_{k+1})}{p(w_k|t_k)p(w_{k+1}|t_{k+1}) + p(w_{k+1}|t_k)p(w_k|t_{k+1})} \\ &= \frac{1}{2}\left\{1 + \tanh\left(\frac{1}{2}n\left(t_{k+1} - t_k\right)\left(H(w_{k+1}) - H(w_k)\right)\right)\right\}. \end{aligned} \tag{3}$$

Hereafter, we call u_1 and u_2 exchange ratios. Under these updates, the joint distribution of Eq.(1) is invariant because these updates satisfy the detailed balance condition for the distribution of Eq.(1) [4].

Consequently, the following two steps are carried out alternately:

1. Each sequence is generated simultaneously and independently for a few iteration by a conventional MCMC method.
2. Two positions are exchanged with the exchange ratio u_1 or u_2.

The advantage of EMC method is to accelerate the convergence of sample sequence comparing the conventional MCMC method. The conventional MCMC method requires huge computational cost to generate a sample sequence from the target distribution because this algorithm is based on local updating. The EMC 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 $\varphi(w)$ is sought to be a simple distribution that is easy to sample from and $t_1 = 0$.

2.2 Exchange Ratios for EMC Method

When we design the EMC method, the setting of temperatures is very important to make the EMC method efficient. As we can see in Eq.(2) and Eq.(3), temperature has close relation to the exchange ratio. Therefore, temperatures are very important parameters in adjusting the exchange ratio and its average.

For the efficient EMC method, each sample needs to wander over the whole temperature region. Moreover, 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 not efficient 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 high, the interval of neighboring temperatures has to be very small, that is, the total number K of temperature needs to be large. Therefore, this setting is not also efficient because it needs huge cost to generate a sample from each distribution. Consequently, the set of temperatures needs to optimize so that the average exchange ratios for any neighboring temperatures become not low and not too high.

As a criterion for the setting of temperature, the following symmetrized Kullback divergence $I(t_k, t_{k+1})$ is used [6],

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

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 by

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

the following equation is satisfied in small interval of temperatures, $I(t_k, t_{k+1}) = \frac{\partial^2 F}{\partial t^2}\Big|_{t=t_k} (t_{k+1} - t_k)^2$. By using these properties, the implementation approach of temperatures based on the value of $\sqrt{\partial^2 F/\partial t^2}$ has been proposed.

As above mentioned, the average exchange ratio, the symmetrized Kullback divergence and the free energy are used as criteria for the setting of temperatures. However, the theoretical properties of these functions are not clarified. Hence, in practical, the simulations of EMC method have to be carried out in order to obtain the value of each function. However, since it is typically difficult to check the convergence of EMC simulation, the accuracy of experimental value of each function is not clear.

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, $n \to \infty$. This result gives us the criteria for the setting of temperature and for checking the convergence of EMC simulations.

3 Main Result

In this section, we show the main result of this paper. We assume that $t > 0$ and $t + \Delta t > 0$, and consider the EMC method between the two distributions, $p_1(w) = p(w|t)$ and $p_2(w) = p(w|t + \Delta t)$, where the real number Δt is not necessarily small. For these distributions, the symmetrized Kullback divergence I is rewritten as follows,

$$I = \int p_1(w_1) \log \frac{p_1(w_1)}{p_2(w_1)} dw_1 + \int p_2(w_2) \log \frac{p_2(w_2)}{p_1(w_2)} dw_2.$$

As we can see in Eq.(2) and Eq.(3), the exchange ratios u_1 and u_2 are functions of w_1 and w_2. Hence, we define the average exchange ratios J_1 and J_2 as the expectations of exchange ratios u_1 and u_2 over the joint distribution $p_1(w_1) \times p_2(w_2)$ as follows,

$$J_1 = \int \int u_1 p_1(w_1) p_2(w_2) dw_1 dw_2$$
$$J_2 = \int \int u_2 p_1(w_1) p_2(w_2) dw_1 dw_2.$$

In a lot of learning machines such as neural networks, normal mixtures and hidden Markov models, the Bayesian posterior distribution does not converge to the normal distribution as $n \to \infty$ because the Hessian of log likelihood $H(w)$ is not positive definite. 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. Then $\zeta(z)$ 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 [1]. We also define the rational number $-\lambda$ as the largest pole of zeta function $\zeta(z)$ and the natural number m as its order. If the Hessian matrix $\left(\frac{\partial^2 H(w)}{\partial w_i \partial w_j} \right)$ is positive definite for arbitrary w, it holds that $\lambda = \frac{d}{2}$, $m = 1$. Otherwise, λ and m can be calculated by using the resolution of singularities in algebraic geometry [1]. In fact, there are some studies to calculate the value λ and m for a certain energy function $H(w)$ and a probability distribution $\varphi(w)$ [13].

Then, the following lemma holds [12].

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

$$V(s) = \int \delta(s - H(w)) \varphi(w) dw \cong c s^{\lambda - 1} (-\log s)^{m-1},$$

where c is a constant.

In our previous work, we analyzed the symmetrized Kullback divergence and the average exchange ratio for Metropolis type [9]. However, the condition

$\Delta t \ll t$ is assumed in this analysis. In this paper, we analyze these functions without this assumption.

Firstly, we show the theorem about the symmetrized Kullback divergence and the average exchange ratio for Metropolis type.

Theorem 1. *The symmetrized Kullback divergence I and the average exchange ratio J_1 for Metropolis type respectively converge to the following values as $n \to \infty$,*

$$I \to \lambda \frac{(\Delta t)^2}{t(t+\Delta t)}$$
$$J_1 \to \begin{cases} \left(1+\frac{\Delta t}{t}\right)^{\lambda} \frac{2\Gamma(2\lambda)}{\Gamma(\lambda)^2} A(\lambda, \frac{\Delta t}{t}) & \textit{(if } \Delta t \geq 0\textit{)} \\ \left(1-\frac{\Delta t}{t+\Delta t}\right)^{\lambda} \frac{2\Gamma(2\lambda)}{\Gamma(\lambda)^2} A(\lambda, -\frac{\Delta t}{t+\Delta t}) & \textit{(if } \Delta t < 0\textit{)}, \end{cases}$$

where $A(\lambda, \frac{\Delta t}{t})$ is defined by

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

(outline of the proof) This theorem is simply proven by using the almost same analysis of [9] without applying the Taylor expansion about $\frac{\Delta t}{t} = 0$. (Q.E.D)

From Theorem 1, we can make the symmetrized Kullback divergence constant over the various temperatures by the temperature setting that the value $\frac{\Delta t}{t}$ or $\frac{\Delta t}{t+\Delta t}$ is constant, that is, the set $\{t_k\}$ of temperature is set as geometric progression. Moreover, under this setting, the average exchange ratio for Metropolis type becomes constant over the various temperatures.

Secondly, we reveal the average exchange ratio J_2 for heat bath type. We show its proof in Appendix.

Theorem 2. *The average exchange ratio J_2 for heat bath type converges to the following value as $n \to \infty$,*

$$J_2 \to \begin{cases} \frac{1}{2}\left(1+\left(1+\frac{\Delta t}{t}\right)^{\lambda} \frac{B\left(\lambda, \frac{\Delta t}{t}\right)}{\Gamma(\lambda)^2}\right) & \textit{(if } \Delta t \geq 0\textit{)} \\ \frac{1}{2}\left(1+\left(1-\frac{\Delta t}{t+\Delta t}\right)^{\lambda} \frac{B\left(\lambda, -\frac{\Delta t}{t+\Delta t}\right)}{\Gamma(\lambda)^2}\right) & \textit{(if } \Delta t < 0\textit{)}, \end{cases}$$

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

$$B\left(\lambda, \frac{\Delta t}{t}\right) = \int_0^\infty ds_1 \int_0^\infty ds_2 \tanh\left(\frac{1}{2}\frac{\Delta t}{t}(s_2-s_1)\right) e^{-s_1} e^{-(1+\frac{\Delta t}{t})s_2} s_1^{\lambda-1} s_2^{\lambda-1}.$$

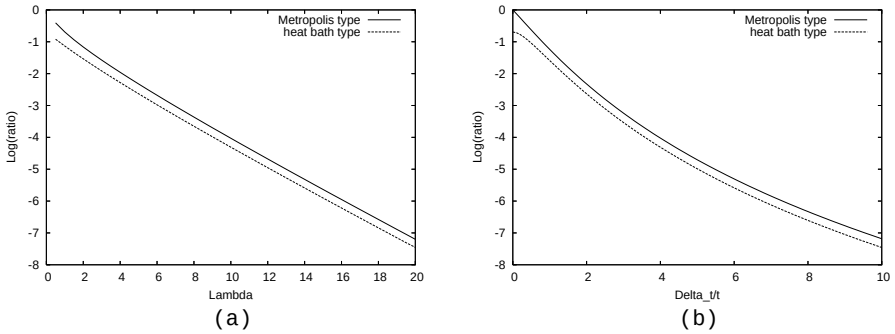

Fig. 1. The theoretical value of average exchange ratio (a) for the value λ with $\frac{\Delta t}{t} = 2.0$, and (b) for the value $\frac{\Delta t}{t}$ with $\lambda = 5.0$. The solid line is for Metropolis type and the dashed line is for heat bath type.

Theorem 2 claims that the average exchange ratio for heat bath type is also constant over the various temperatures by setting the value $\frac{\Delta t}{t}$ constant. Consequently, if we set the value $\frac{\Delta t}{t}$ constant, the symmetrized Kullback divergence and the average exchange ratio for both type are clarified to be constant over the various temperatures.

4 Discussion

In this paper, we analyzed the symmetrized Kullback divergence and the average exchange ratio for Metropolis type and for heat bath type in the low temperature limit. As above mentioned, the EMC simulations have to be carried out in order to obtain the value of these functions. However, the accuracy of these experimental values are not clear because of the difficulty of checking the convergence of EMC simulations. Our result gives us the accurate value of these functions without the EMC simulations, which leads to the criteria for the setting of temperature. Moreover, since the accuracy of the experimental values of these functions can be evaluated by using our result, we can check the convergence of EMC simulation, which is very important for the EMC method.

Let us discuss three points in association to this paper.

Firstly, we discuss the relation between the shape of target distribution and setting of temperature. Figure 1(a) shows the relation between the value λ and the theoretical value of average exchange ratio for each type with $\frac{\Delta t}{t} = 2.0$. According to this figure, the EMC method works efficiently for the target distribution with small value λ, which requires high average exchange ratio. On the other hand, by comparing two distributions whose ground states 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 [13]. Consequently, the EMC method works efficiently for the target distribution with the energy function

whose ground state is an 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 EMC method for the Bayesian learning in hierarchical learning machines.

Secondly, we discuss the properties of average exchange ratio for both type. When we design the EMC method, which type should be used as the exchange ratio, Metropolis type or heat bath type, often becomes a problem. By comparing each exchange ratio, the computational cost is almost equal. Therefore, we should select the type to make the average exchange ratio high for the same setting of temperature. Figure 1(b) shows the relation between the value $\frac{\Delta t}{t}$ and the theoretical value of average exchange ratio for each type with $\lambda = 5.0$. According to Figure 1(a) and 1(b), the average exchange ratio for Metropolis type is clarified to be higher than that for heat bath type in the low temperature limit. This claims the effectiveness of Metropolis type for the EMC method.

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

5 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 a result, the following properties are clarified,

1. When the symmetrized Kullback divergence for arbitrary temperature t is constant, the average exchange ratio is also constant over the various temperatures.
2. Then, the set of temperature $\{t_k\}$ is set as geometric progression.

As the future works, verifying the theoretical result in this study by some experiments, constructing the implementation approach of EMC method, and applying these results to the practical problems such as the Bayesian learning should be addressed.

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 18079007.

References

1. M. F. Atiyah, "Resolution of singularities and division of distributions," Communications of Pure and Applied Mathematics, Vol.13, pp.145-150, 1970.
2. B. A. Berg, T. Neuhaus, "Multicanonical algorithms for first order phase transitions," Physics Letter B, Vol.267, No.2, pp.249-253, 1991.
3. K. Hukushima, "Domain Wall Free Energy of Spin Glass Models: Numerical Method and Boundary Conditions," Physical Review E, Vol.60, pp.3606-3613, 1999.
4. 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.
5. K. Hukushima, "Extended ensemble Monte Carlo approach to hardly relaxing problems," Computer Physics Communications, Vol.147, pp.77-82, 2002.
6. Y. Iba, "Extended Ensemble Monte Carlo," International Journal of Modern Physics C, Vol.12, pp.623-656, 2001.
7. F. Liang, "An effective Bayesian neural network classifier with a comparison study to support vector machine," Neural Computation, Vol.15, pp.1959-1989, 2003.
8. E. Marinari, G. Parisi, "Simulated tempering: a new Monte Carlo scheme," Europhysics Letters, Vol.19, No.6, pp.451-455, 1992.
9. 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.
10. 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.
11. Y. Sugita, Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," Chemical Physics Letters, Vol.314, No.1-2, pp.141-151, 1999.
12. S. Watanabe, "Algebraic Analysis for Nonidentifiable Learning Machines," Neural Computation, Vol.13, pp.899-933, 2001.
13. K. Yamazaki, S. Watanabe, "Singularities in mixture models and upper bounds of stochastic complexity," Neural networks, Vol.16, No.7, pp.1029-1038, 2003.

Appendix: Proof of Theorem 2

In Appendix, we show the proof of Theorem 2 about the average exchange ratio for heat bath type.

(outline of the proof) The average exchange ratio J_2 is expressed by the definition of exchange ratio u_2 as follows,

$$J_2 = \frac{1}{2}\left(1 + \int dw_1 \int dw_2 \tanh\left(\frac{n\Delta t}{2}(H(w_2) - H(w_1))\right)\right.$$
$$\left.\times \frac{e^{-ntH(w_1)}\varphi(w_1)}{Z(nt)} \frac{e^{-n(t+\Delta t)H(w_2)}\varphi(w_2)}{Z(n(t+\Delta t))}\right).$$

Then, the normalization constant $Z(nt)$ is given from Lemma 1,

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

Hence, by defining

$$J_2^* = \int dw_1 \int dw_2 \tanh\left(\frac{n\Delta t}{2}(H(w_2) - H(w_1))\right) \times e^{-ntH(w_1)}\varphi(w_1)e^{-n(t+\Delta t)H(w_2)}\varphi(w_2),$$

we obtain

$$J_2 = \frac{1}{2}\left(1 + \frac{J_2^*}{Z(nt)Z(n(t+\Delta t))}\right).$$

The function J_2^* is expressed by using the Dirac delta function as follows,

$$\begin{aligned} J_2^* &= \int_0^\infty ds_1 \int_0^\infty ds_2 \tanh\left(\frac{n\Delta t}{2}(s_2 - s_1)\right) e^{-nts_1}e^{-n(t+\Delta t)s_2} \\ &\quad \times \int dw_1 \delta(s_1 - H(w_1))\varphi(w_1) \int dw_2 \delta(s_2 - H(w_2))\varphi(w_2) \\ &\cong \int_0^\infty ds_1 \int_0^\infty ds_2 \tanh\left(\frac{n\Delta t}{2}(s_2 - s_1)\right) e^{-nts_1}e^{-n(t+\Delta t)s_2} \\ &\quad \times cs_1^{\lambda-1}(-\log s_1)^{m-1} cs_2^{\lambda-1}(-\log s_2)^{m-1}. \end{aligned}$$

In the case that $\Delta t \geq 0$, the function J_2^* is given by putting $s_1' = nts_1$ and $s_2' = nts_2$ as follows,

$$J_2^* \cong \frac{c^2(\log nt)^{2(m-1)}}{(nt)^{2\lambda}} B\left(\lambda, \frac{\Delta t}{t}\right).$$

In the case that $\Delta t < 0$, the function J_2^* is given by putting $s_1' = n(t+\Delta t)s_1$ and $s_2' = n(t+\Delta t)s_2$ as follows,

$$J_2^* \cong \frac{c^2(\log n(t+\Delta t))^{2(m-1)}}{(n(t+\Delta t))^{2\lambda}} B\left(\lambda, -\frac{\Delta t}{t+\Delta t}\right).$$

Consequently, the average exchange ratio J_2 is given by

$$\begin{aligned} J_2 &= \frac{1}{2}\left(1 + \frac{J_2^*}{Z(nt)Z(n(t+\Delta t)}\right) \\ &\to \begin{cases} \frac{1}{2}\left(1 + \left(1 + \frac{\Delta t}{t}\right)^\lambda \frac{B(\lambda, \frac{\Delta t}{t})}{\Gamma(\lambda)^2}\right) & (\text{if } \Delta t \geq 0) \\ \frac{1}{2}\left(1 + \left(1 - \frac{\Delta t}{t+\Delta t}\right)^\lambda \frac{B(\lambda, -\frac{\Delta t}{t+\Delta t})}{\Gamma(\lambda)^2}\right) & (\text{if } \Delta t < 0), \end{cases} \end{aligned}$$

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