مجلّه‌ی پژوهش‌های آماری ایران

J. Statist. Res. Iran **7** (2010): 103–120

پژوهشکده‌ی آمار دوره‌ی ۷، شماره‌ی ۲، پاییز و زمستان ۱۳۸۹، صص ۱۰۳–۱۲۰

Taylor Expansion for the Entropy Rate of Hidden Markov Chains

Gh. Yari* and Z. Nikooravesh

Iran University of Science and Technology

Abstract. We study the entropy rate of a hidden Markov process, defined by observing the output of a symmetric channel whose input is a first order Markov process. Although this definition is very simple, obtaining the exact amount of entropy rate in calculation is an open problem. We introduce some probability matrices based on Markov chain's and channel's parameters. Then, we try to obtain an estimate for the entropy rate of hidden Markov chain by matrix algebra and its spectral representation. To do so, we use the Taylor expansion, and calculate some estimates for the first and the second terms, for the entropy rate of the hidden Markov process and its binary version, respectively. For small ε (channel's parameter), the entropy rate has $o(\varepsilon^2)$, as a maximum error, when it is calculated by the first term of Taylor expansion and it has $o(\varepsilon^3)$, as a maximum error, when it is calculated by the second term.

Keywords. Entropy rate; hidden Markov process; Taylor approximation; spectral representation.

MSC 2010: 60J05, 54C70.

1 Introduction

Let $X = \{X_k\}_{k\geqslant 1}$ be a first-order stationary Markov process on $\{0, 1, \ldots, m-1\}$, with transition matrix $\mathbf{P} = \{p_{ab}\}$ such that for every $k \geqslant 1$, $p_{ab} = P_X(X_k = b|X_{k-1} = a)$ where $a, b \in \{0, 1, \ldots, m-1\}$. Consider also a noise

* Corresponding author

process $E = \{E_k\}_{k\geqslant 1}$, independent of X, such that $P(E_i = l) = \varepsilon_l$ where $l \in \{0, 1, \ldots, m-1\}$ and $\sum_{l=0}^{m-1} \varepsilon_l = 1$. Now, define the process $Z = \{Z_k\}_{k\geqslant 1}$, with

$$Z_k = X_k \oplus E_k, \qquad k \geqslant 1, \tag{1}$$

where $\oplus$ denotes addition modulo m.

Consider a stochastic process $\{Y_k\}_{k\geqslant 1}$ with state space $\mathbf{S}$. The Shannon entropy rate of the stochastic process $\{Y_k\}_{k\geqslant 1}$ is (Cover, 2006)

$$H(\mathcal{Y}) = \lim_{n\to\infty} \frac{H(Y_1, Y_2, \ldots, Y_n)}{n}, \tag{2}$$

where Y_t is a random variable demonstrating the state at time t, and $H(Y_1, Y_2, \ldots, Y_n)$ is the joint entropy of $(Y_1, Y_2, \ldots, Y_n)$ with the joint distribution $P(y_1, y_2, \ldots, y_n)$ where

$$\begin{aligned} H(Y_1, Y_2, \ldots, Y_n) &= -\sum_{y_1\in\mathbf{S}}\sum_{y_2\in\mathbf{S}}\cdots\sum_{y_n\in\mathbf{S}} P(y_1, y_2, \ldots, y_n) \log P(y_1, y_2, \ldots, y_n) \\ &= -\mathbf{E}_{Y_1,Y_2,\ldots,Y_n} \log P(Y_1, Y_2, \ldots, Y_n). \end{aligned} \tag{3}$$

Shannon (1948) proved the convergence in probability $\frac{1}{n} \log P(y_1, y_2, \ldots, y_n)$ to $H(\mathcal{Y})$.

In the rest of the paper we deal with the entropy rate for hidden Markov process $\{Z\}_{i\geqslant 1}$ of (1) as a function of

$$\mathbf{\Upsilon} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{m-1}].$$

The process $\{Z_k\}_{k\geqslant 1}$ is a stochastic process, also it is an example of a hidden Markov process. Indeed, a hidden Markov process can be seen as a process resulting from observing any discrete-time, finite state homogeneous Markov chain through a discrete-time memoryless and noisy channel (Drake, 1965). Some applications include automatic character recognition (Raviv, 1967), speech recognition (Jelinek, 1975), communications and information theory (Bahi, 1974), DNA sequencing (Churchill, 1989), and others.

Despite the simplicity of the definition of hidden Markov processes is misleading, and the extensive researches have been done on their properties and applications, some questions on fundamental properties of the processes remain open, even for the simple case (1). Some of these questions concern the performance of filtering , denoising (Ordentlich, 2006), and compression on hidden Markov sources. In all these cases, algorithms exist that achieve optimal performance (e.g., minimal residual noise or code length), even universally (without knowledge of the process parameters). However, in general, the optimal value of the performance of interest for each of the problems has not been explicitly characterized. In the case of compression, the problem of interest is the determination of the entropy rate $H(\mathcal{Z})$ of the process Z as an explicit expression in the parameters of the process.

Computing the Shannon entropy (here it is called entropy) of a hidden Markov process was studied by Blackwell (1957), which is based on the intrinsic complexity of expressing the hidden Markov process entropy as a function of the process parameters.

Since computing the exact amount of entropy rate of any process is still impossible, some researchers have tried to find some proper bounds or acceptable estimation for this. Ordentlith (2004) obtained upper and lower bounds for binary markov chain and symmetric binary channel, based on their parameters. Also Ordentlich (2006) presented a different method for analyzing the entropy rate of the hidden Markov process. In this method, they used log likelihood ratio function. Jacquent (2004) defined stochastic matrices with Markov chain's probability transitions. Then, he estimated the entropy rate of binary hidden Markov chain via symmetric channel by using matrix algebra and Taylor expansion. Zuk (2005) showed formulas for higher-order coefficients of the Taylor expansion in the symmetric case for binary Hidden Markov Chain. Han (2006) generalized Zuk's results to a natural class of hidden Markov chains called "Black Holes". Jacquent (2008) explained that the maximum error for estimating the entropy rate by the first term of Taylor expansion is the square of channel's parameter. Luo (2009) acquired a fixed point functional equation via log likelihood function for obtaining the entropy rate. But there was a problem. There was no explicit answer to this equation. So he had to use numerical method to solve this problem.

Following the Zuk's and Jacquent's works, mentioned above, we are going to obtain the first order term of Taylor expansion for the entropy rate of a hidden Markov chain with arbitrary finite state space, and the second order term for the entropy rate of a binary hidden Markov chain. Our study will focus on the estimation of the entropy rate of the hidden Markov chain, where the channel parameters (noises) ε_is are small. The paper is organized as follows: In Section 2 we explain some required preliminaries. We show $P_Z(z_1^n)$ is the product of some probability matrices. In Sections 3 and 4, we outline the analysis of our main results. In theorems 1 and 2, we present as explicit first and second order terms of Taylor expansion of $H(\mathcal{Z})$ near $\boldsymbol{\Upsilon} = 0$, as a function of the parameters p_{ab}, respectively. We show that the linear terms of the expansion can be expressed as a Kullback-Liebler divergence.

2 Preliminaries

For any sequence $\{Y_k\}_{k\geqslant 1}$ we denote a finite sub-sequence $Y_i, Y_{i+1}, \ldots, Y_j$, $j \geqslant i$ of $\{Y_k\}_{k\geqslant 1}$ by Y_i^j. Using our assumptions on the processes X and E, one can obtain easily

$$P(Z_1^n, E_n) = \sum_{k=0}^{m-1} P(E_n) P_X(Z_n \oplus E_n | Z_{n-1} \oplus k) P(Z_1^{n-1}, E_{n-1} = k). \quad (4)$$

Let

$$\mathbf{P}_n := [P(Z_1^n, E_n = 0), P(Z_1^n, E_n = 1), \ldots, P(Z_1^n, E_n = m-1)], \quad (5)$$

and get $\mathbf{M}(Z_{n-1}, Z_n)$ as a probability matrix with dimension $m \times m$ and entries $\varepsilon_{j-1} P_X(Z_n \oplus (j-1) | Z_{n-1} \oplus (i-1))$ in ith row and jth column. So it is easy to show

$$\mathbf{P}_n = \mathbf{P}_{n-1} \mathbf{M}(Z_{n-1}, Z_n), \qquad n > 1, \quad (6)$$

and

$$P_Z(Z_1^n) = \mathbf{P}_1 \mathbf{M}(Z_1, Z_2) \cdots \mathbf{M}(Z_{n-1}, Z_n) \mathbf{1}^t, \quad (7)$$

where $\mathbf{1} = [1, 1, \ldots, 1]_{1\times m}$ and superscript t denotes transposition.

We construct these matrices for a given realization z_1^n of Z_1^n. Using the notation $\mathbf{M}_i = \mathbf{M}(z_i, z_{i+1})$ we get

$$\mathbf{M}_i = \begin{bmatrix} P_X(z_{i+1}|z_i) & 0 & \cdots & 0 \\ P_X(z_{i+1}|z_i \oplus 1) & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ P_X(z_{i+1}|z_i \oplus m-1) & 0 & \cdots & 0 \end{bmatrix}$$

$$+ \varepsilon_1 \begin{bmatrix} -P_X(z_{i+1}|z_i) & P_X(z_{i+1} \oplus 1|z_i) & 0 & \cdots & 0 \\ -P_X(z_{i+1}|z_i \oplus 1) & P_X(z_{i+1} \oplus 1|z_i \oplus 1) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -P_X(z_{i+1}|z_i \oplus m-1) & P_X(z_{i+1} \oplus 1|z_i \oplus m-1) & 0 & \cdots & 0 \end{bmatrix}$$

$$+\varepsilon_2 \begin{bmatrix} -P_X(z_{i+1}|z_i) & 0 & P_X(z_{i+1}\oplus 2|z_i) & 0 & \cdots & 0 \\ -P_X(z_{i+1}|z_i\oplus 1) & 0 & P_X(z_{i+1}\oplus 2|z_i\oplus 1) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ -P_X(z_{i+1}|z_i\oplus m-1) & 0 & P_X(z_{i+1}\oplus 2|z_i\oplus m-1) & 0 & \cdots & 0 \end{bmatrix}$$

$$+\cdots$$

$$+\varepsilon_{m-1} \begin{bmatrix} -P_X(z_{i+1}|z_i) & 0 & \cdots & 0 & P_X(z_{i+1}\oplus m-1|z_i) \\ -P_X(z_{i+1}|z_i\oplus 1) & 0 & \cdots & 0 & P_X(z_{i+1}\oplus m-1|z_i\oplus 1) \\ \vdots & \vdots & \vdots & & \vdots \\ -P_X(z_{i+1}|z_i\oplus m-1) & 0 & \cdots & 0 & P_X(z_{i+1}\oplus m-1|z_i\oplus m-1) \end{bmatrix}$$

$$:= \mathbf{M}_i^{(0)} + \varepsilon_1 \mathbf{M}_i^{(1)} + \cdots + \varepsilon_{m-1}\mathbf{M}_i^{(m-1)}, \qquad 1 \leqslant i \leqslant n-1. \tag{8}$$

Similarly, one can show

$$\begin{aligned} \mathbf{P}_1 &= [P_X(z_1), 0, \cdots, 0] + \varepsilon_1[-P_X(z_1), P_X(z_1\oplus 1), 0, \ldots, 0] \\ &\quad + \varepsilon_2[-P_X(z_1), 0, P_X(z_1\oplus 2), 0, \ldots, 0] + \cdots + \\ &\quad + \varepsilon_{m-1}[-P_X(z_1), 0, \ldots, 0, P_X(z_1\oplus m-1)] \\ &:= \mathbf{P}_1^{(0)} + \varepsilon_1\mathbf{P}_1^{(1)} + \cdots + \varepsilon_{m-1}\mathbf{P}_1^{(m-1)}, \end{aligned} \tag{9}$$

and

$$\begin{aligned} P_Z(z_1^n) &= \mathbf{P}_1\mathbf{M}_1\mathbf{M}_2\cdots\mathbf{M}_{n-1}\mathbf{1}^t \\ &= \left(\mathbf{P}_1^{(0)} + \sum_{i=1}^{m-1}\varepsilon_i\mathbf{P}_1^{(i)}\right)\prod_{k=1}^{n-1}\left(\mathbf{M}_k^{(0)} + \sum_{i=1}^{m-1}\varepsilon_i\mathbf{M}_k^{(i)}\right)\mathbf{1}^t. \end{aligned} \tag{10}$$

3 The First Order Term in the Entropy Rate of a Hidden Markov Chain

The following formula will be useful in computing the Shannon entropy of Z

$$R_n(s, \mathbf{\Upsilon}) = \sum_{z_1^n} P_Z^s(z_1^n), \tag{11}$$

where the exponent s of P_Z is a complex variable, and the summation is over all n_tuples of $\{0, 1, \dots, m-1\}$. Note that by the definition (1), for $\mathbf{\Upsilon} = 0$, we can write

$$R_n(s, 0) = \sum_{z_1^n} P_X^s(z_1^n). \tag{12}$$

Now by differentiating from both sides of (11) and (12) with respect to s, we have

$$\frac{\partial}{\partial s} R_n(s, \mathbf{\Upsilon}) = \sum_{z_1^n} P_Z^s(z_1^n) \log P_Z(z_1^n) \tag{13}$$

and

$$\frac{\partial}{\partial s} R_n(s, 0) = \sum_{z_1^n} P_X^s(z_1^n) \log P_X(z_1^n). \tag{14}$$

The rewrite (13) and (14), by the definition of the entropy of a subsequence from a stochastic process, follows

$$H_n(Z_1^n) = -\frac{\partial}{\partial s} R_n(s, \mathbf{\Upsilon})|_{s=1}, \tag{15}$$

and

$$H_n(X_1^n) = -\frac{\partial}{\partial s} R_n(s, 0)|_{s=1}, \tag{16}$$

where $H_n(Y_1^n)$, for any random variable $\{Y_i\}_{i \geqslant 1}$, is Shannon entropy of subsequence $\{Y_1, Y_2, \dots, Y_n\}$.

3.1 Taylor Expansion

Using Taylor expansion near $\mathbf{\Upsilon} = 0$, we have

$$R_n(s, \mathbf{\Upsilon}) = R_n(s, 0, \dots, 0) + \sum_{k=1}^{m-1} \varepsilon_k \frac{\partial}{\partial \varepsilon_k} R_n(s, \mathbf{\Upsilon})|_{\mathbf{\Upsilon}=0} + o(\varepsilon_{max}^2), \tag{17}$$

where $\varepsilon_{max} = \max\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{m-1}\}$. Differentiate from both sides of above formula with respect to s at $s = 1$ implies

$$H_n(Z_1^n) = H_n(X_1^n) - \sum_{k=1}^{m-1} \varepsilon_k \frac{\partial^2}{\partial s \partial \varepsilon_k} R_n(s, \mathbf{\Upsilon})|_{\mathbf{\Upsilon}=0, s=1} + o(\varepsilon_{max}^2), \tag{18}$$

For our aims, we must compute $\frac{\partial^2}{\partial s \partial \varepsilon_k} R_n(s, \boldsymbol{\Upsilon})|_{\boldsymbol{\Upsilon}=0,s=1}$. So we compute firstly

$$\frac{\partial}{\partial \varepsilon_k} R_n(s, \boldsymbol{\Upsilon})|_{\boldsymbol{\Upsilon}=0} = \sum_{z_1^n} s P_Z^{s-1}(z_1^n) \frac{\partial}{\partial \varepsilon_k} P_Z(z_1^n)|_{\boldsymbol{\Upsilon}=0}. \tag{19}$$

To compute $\frac{\partial}{\partial \varepsilon_k} R_n(s, \boldsymbol{\Upsilon})$ at $\boldsymbol{\Upsilon} = 0$ we may calculate

$$\begin{aligned}\frac{\partial}{\partial \varepsilon_k} P_Z(z_1^n)|_{\boldsymbol{\Upsilon}=0} &= \mathbf{P}_1^{(k)} \prod_{j=1}^{n-1} \mathbf{M}_j^{(0)} \\ &+ \mathbf{P}_1^{(0)} \sum_{i=1}^{n-2} \prod_{j=1}^{i-1} \mathbf{M}_j^{(0)} \mathbf{M}_i^{(k)} \prod_{j=i+1}^{n-1} \mathbf{M}_j^{(0)} \\ &+ \mathbf{P}_1^{(0)} \prod_{j=1}^{n-2} \mathbf{M}_j^{(0)} \mathbf{M}_{n-1}^{(k)}. \end{aligned} \tag{20}$$

Using (8) and (9), the derivative of $P_Z(z_1^n)$ at $\boldsymbol{\Upsilon} = 0$ can be calculated as

$$\begin{aligned}\frac{\partial}{\partial \varepsilon_k} P_Z(z_1^n)|_{\boldsymbol{\Upsilon}=0} &= -n P_X(z_1^n) + P_X(z_1 \oplus k \ z_2^n) \\ &+ \sum_{i=1}^{n-2} P_X(z_1^i \ z_{i+1} \oplus k \ z_{i+2}^n) + P_X(z_1^{n-1} z_n \oplus k). \end{aligned} \tag{21}$$

3.2 Computation of the Entropy Rate

Now we can compute $H(\mathcal{Z})$ (the entropy rate of the hidden Markov chain $\{Z_i\}_{i \geqslant 1}$), i.e.,

$$H(\mathcal{Z}) = \lim_{n \to \infty} \frac{1}{n} H_n(Z_1^n). \tag{22}$$

Theorem 1. *Suppose $p_{ab} > 0$ for any $a, b \in \{0, 1, \ldots, m-1\}$. The first order term in the entropy rate of the hidden Markov chain Z is*

$$H(\mathcal{Z}) = H(\mathcal{X}) + \sum_{k=1}^{m-1} \varepsilon_k \mathbf{D}(P_X(z_1 z_2 z_3) || P_X(z_1 \ z_2 \oplus m-k \ z_3)) + o(\varepsilon_{max}^2), \tag{23}$$

where $\mathbf{D}(P_X(z_1 z_2 z_3) || P_X(z_1 \ z_2 \oplus m - k \ z_3))$ is the relative entropy between

$P_X(z_1z_2z_3)$ *and* $P_X(z_1\ z_2\oplus m-k\ z_3)$, *such that:*

$$\mathbf{D}(P_X(z_1z_2z_3)||P_X(z_1\ z_2\oplus m-k\ z_3)) = \sum_{z_1z_2z_3} P_X(z_1z_2z_3) \times \log\frac{P_X(z_1z_2z_3)}{P_X(z_1\ z_2\oplus m-k\ z_3)}. \quad (24)$$

Proof. First we define some new matrices $\mathbf{P}(s), \mathbf{Q}_{1k}(s), \mathbf{Q}_{2k}(s), \Pi(s)$ and $\Pi_k(s)$ with entries

$$\{\mathbf{P}(s)\}_{ij} = p_{ij}^s, \quad \{\mathbf{Q}_{1k}(s)\}_{ij} = p_{ij}p_{i\ \ j\oplus m-k}^{s-1}, \quad \{\mathbf{Q}_{2k}(s)\}_{ij} = p_{ij}p_{i\oplus m-k\ \ j}^{s-1}, \quad (25)$$

where $p_{ij} = P_X(X_d = j|X_{d-1} = i)$ for any $d \geqslant 1$ and $0 \leqslant i, j \leqslant m-1$, also

$$\{\mathbf{\Pi}(s)\}_i = p_i^s, \qquad \{\mathbf{\Pi}_k(s)\}_i = p_i^{s-1}p_{i\oplus k}, \quad (26)$$

where $p_i = P_X(i)$ for any $0 \leqslant i \leqslant m-1$.

We imply $P_Z(Z_1^n) = P_X(Z_1^n)$ while $\mathbf{\Upsilon} = 0$. Using (19) and (20)

$$\frac{\partial}{\partial\varepsilon_k}R_n(s,\mathbf{\Upsilon})|_{\mathbf{\Upsilon}=0} = \sum_{z_1^n} s[-nP_X^s(z_1^n) + P_X^{s-1}(z_1^n)P_X(z_1\oplus k\ z_2^n) + \sum_{i=1}^{n-2} P_X^{s-1}(z_1^n)P_X(z_1^i z_{i+1}\oplus k\ z_{i+2}^n) + P_X^{s-1}(z_1^n)P_X(z_1^{n-1}z_n\oplus k)]. \quad (27)$$

From (25) and (26), we obtain

$$\frac{\partial}{\partial\varepsilon_k}R_n(s,\mathbf{\Upsilon})|_{\mathbf{\Upsilon}=0} = s[-n\Pi(s)\mathbf{P}^{n-1}(s) + \Pi_k(s)\mathbf{Q}_{2k}(s)\mathbf{P}^{n-2}(s) + \sum_{i=1}^{n-2}\Pi(s)\mathbf{P}^{i-1}(s)\mathbf{Q}_{1k}(s)\mathbf{Q}_{2k}(s)\mathbf{P}^{n-i-2}(s) + \Pi(s)\mathbf{P}^{i-2}(s)\mathbf{Q}_{1k}(s)]\mathbf{1}^t, \quad (28)$$

$$\frac{\partial}{\partial\varepsilon_k}R_n(s,\mathbf{\Upsilon})|_{\mathbf{\Upsilon}=0} = s[(\Pi_k(s)\mathbf{Q}_{2k}(s) - \Pi(s)\mathbf{P}(s))\mathbf{P}^{n-2}(s) + \Pi(s)\sum_{i=1}^{n-2}\mathbf{P}^{i-1}(s)(\mathbf{Q}_{1k}(s)\mathbf{Q}_{2k}(s) - \mathbf{P}^2(s))\mathbf{P}^{n-i-2}(s) + \Pi(s)\mathbf{P}^{n-2}(s)(\mathbf{Q}_{1k}(s) - \mathbf{P}(s))]\mathbf{1}^t. \quad (29)$$

Now by using the spectral representation of the matrix $\mathbf{P}(s)$, we can differentiate from both sides of (29) respect to s at $s = 1$

$$\lim_{n\to\infty} \frac{1}{n} \cdot \frac{\partial^2}{\partial\varepsilon\partial s} R_n(s,\varepsilon)|_{\varepsilon=0,s=1} = \Pi(1)\frac{\partial}{\partial s}(\mathbf{Q}_k(s) - \mathbf{P}^2(s))|_{s=1}\mathbf{1}^t + o(1). \quad (30)$$

More details are mentioned in Appendix A.

The derivation of the (i,j)th entry of $\mathbf{Q}_k(s) - \mathbf{P}^2(s)$ respect to s at $s = 1$ is

$$\begin{aligned}\frac{\partial}{\partial s}\{\mathbf{Q}_k(s) - \mathbf{P}^2(s)\}_{ij}|_{s=1} &= \frac{\partial}{\partial s}\sum_{r=0}^{m-1}(p_{ir}p^{s-1}_{r\ j\oplus m-k}p_{rj}p^{s-1}_{r\oplus m-k\ j} - p^s_{ir}p^s_{rj})|_{s=1}\\ &= \sum_{r=0}^{m-1} p_{ir}p_{rj}\frac{\log p_{i\ r\oplus m-k}p_{r\oplus m-k\ j}}{\log p_{ir}p_{rj}}. \end{aligned} \quad (31)$$

Therefore

$$\Pi(1)\frac{\partial}{\partial s}(\mathbf{Q}_k(s) - \mathbf{P}^2(s))|_{s=1}\mathbf{1}^t = \sum_{j=0}^{m-1}\sum_{i=0}^{m-1}\sum_{r=0}^{m-1} p_i p_{ir}p_{rj}\frac{\log p_{i\ r\oplus m-k}p_{r\oplus m-k\ j}}{\log p_{ir}p_{rj}}, \quad (32)$$

and according to (51) we have

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \cdot \frac{\partial^2}{\partial\varepsilon_k\partial s} R_n(s,\Upsilon)|_{\Upsilon=0,s=1} &= \sum_{z_1z_2z_3} P(z_1z_2z_3)\frac{\log P(z_1\ z_2\oplus m-k\ z_3)}{\log P(z_1z_2z_3)}\\ &= -\mathbf{D}(P(z_1z_2z_3)||P(z_1\ z_2\oplus m-k\ z_3)).\end{aligned} \quad (33)$$

Using (15), (16), (17), (22) and (33), the proof is completed. □

4 The Second Order Term in the Entropy Rate of a Binary Hidden Markov Chain

In this section we consider a binary hidden Markov chain, i.e., $m = 2$ in Theorem 1. Using Theorem 1, the first order term in the entropy rate of the binary hidden Markov chain $\{Z_i\}_{i\geqslant1}$ is

$$H(\mathcal{Z}) = H(\mathcal{X}) + \varepsilon\mathbf{D}(P_X(z_1z_2z_3)||P_X(z_1\bar{z}_2z_3)) + o(\varepsilon^2), \quad (34)$$

where $\varepsilon = P(E_i = 1)$ and $\bar{z} = z \oplus 1$ (the boolean complement of a binary variable z). To compute the second order term in the entropy rate of the binary hidden Markov chain $\{Z_i\}_{i\geqslant 1}$ we use Taylor expansion near $\varepsilon = 0$, therefore

$$R_n(s,\varepsilon) = R_n(s,0) + \varepsilon\frac{\partial}{\partial\varepsilon}R_n(s,\varepsilon)|_{\varepsilon=0} + \frac{\varepsilon^2}{2}\frac{\partial^2}{\partial\varepsilon^2}R_n(s,\varepsilon)|_{\varepsilon=0} + o(\frac{\varepsilon^3}{6}\frac{\partial^3}{\partial\varepsilon^3}R_n(s,\varepsilon')), \tag{35}$$

and

$$H(Z_1^n) = H(X_1^n) - \varepsilon\frac{\partial^2}{\partial s\partial\varepsilon}R_n(s,\varepsilon)|_{\varepsilon=0} - \frac{\varepsilon^2}{2}\cdot\frac{\partial^3}{\partial s\partial\varepsilon^2}R_n(s,\varepsilon)|_{\varepsilon=0} + o(\frac{\varepsilon^3}{6}\frac{\partial^4}{\partial s\partial\varepsilon^3}R_n(s,\varepsilon')). \tag{36}$$

The equality (11) follows

$$\frac{\partial}{\partial\varepsilon}R_n(s,\varepsilon)|_{\varepsilon=0} = \sum_{z_1^n} sP_Z^{s-1}(z_1^n)\frac{\partial}{\partial\varepsilon}P_Z(z_1^n)|_{\varepsilon=0}, \tag{37}$$

and

$$\frac{\partial^2}{\partial\varepsilon^2}R_n(s,\varepsilon)|_{\varepsilon=0} = \sum_{z_1^n} s(s-1)P_Z^{s-2}(z_1^n)\left(\frac{\partial}{\partial\varepsilon}P_Z(z_1^n)|_{\varepsilon=0}\right)^2 + \sum_{z_1^n} sP_Z^{s-1}(z_1^n)\frac{\partial^2}{\partial\varepsilon^2}P_Z(z_1^n)|_{\varepsilon=0}. \tag{38}$$

Similar to (21), one can show

$$\frac{\partial}{\partial\varepsilon}P_Z(z_1^n)|_{\varepsilon=0} = -nP_X(z_1^n) + \sum_{i=1}^n P_X(z_1^n \oplus e_i), \tag{39}$$

and

$$\frac{\partial^2}{\partial\varepsilon^2}P_Z(z_1^n)|_{\varepsilon=0} = 2\Bigg[\frac{n(n-1)}{2}P_X(z_1^n) - (n-1)\sum_{i=1}^n P_X(z_1^n \oplus e_i) + \sum_{i=1}^n\sum_{j=i+1}^n P_X(z_1^n \oplus e_i \oplus e_j)\Bigg], \tag{40}$$

where e_i denotes ith column of an identity matrix of order n. To compute $\frac{\partial}{\partial\varepsilon}R_n(s,\varepsilon)|_{\varepsilon=0}$ and $\frac{\partial^2}{\partial\varepsilon^2}R_n(s,\varepsilon)|_{\varepsilon=0}$ at $s=1$ we define

$$\begin{aligned}
\Pi_1(s) &= \begin{bmatrix} p_1p_0^{s-1} & p_0p_1^{s-1} \end{bmatrix}, \qquad \overline{\Pi_1}(s) = \begin{bmatrix} p_0p_1^{s-1} & p_1p_0^{s-1} \end{bmatrix},\\
\Pi_2(s) &= \begin{bmatrix} p_1^2p_0^{s-2} & p_0^2p_1^{s-2} \end{bmatrix}, \qquad \overline{\Pi_2}(s) = \begin{bmatrix} p_0^2p_1^{s-2} & p_1^2p_0^{s-2} \end{bmatrix},\\
\mathbf{Q}_1(s) &= \begin{bmatrix} p_{00}p_{01}^{s-1} & p_{01}p_{00}^{s-1}\\ p_{10}p_{11}^{s-1} & p_{11}p_{10}^{s-1} \end{bmatrix}, \quad \mathbf{Q}_2(s) = \begin{bmatrix} p_{00}p_{10}^{s-1} & p_{01}p_{11}^{s-1}\\ p_{10}p_{00}^{s-1} & p_{11}p_{01}^{s-1} \end{bmatrix},\\
\mathbf{Q}_1^{(2)}(s) &= \begin{bmatrix} p_{00}^2p_{01}^{s-2} & p_{01}^2p_{00}^{s-2}\\ p_{10}^2p_{11}^{s-2} & P_{11}^2P_{10}^{s-2} \end{bmatrix}, \mathbf{Q}_2^{(2)}(s) = \begin{bmatrix} p_{00}^2p_{10}^{s-2} & p_{01}^2p_{11}^{s-2}\\ p_{10}^2p_{00}^{s-2} & P_{11}^2P_{01}^{s-2} \end{bmatrix},\\
\mathbf{Q}_3(s) &= \begin{bmatrix} p_{00}p_{11}^{s-1} & p_{01}p_{10}^{s-1}\\ p_{10}p_{01}^{s-1} & p_{11}p_{00}^{s-1} \end{bmatrix}, \quad \mathbf{Q}_4(s) = \begin{bmatrix} p_{00}p_{11}p_{01}^{s-2} & p_{01}p_{10}p_{00}^{s-2}\\ p_{10}p_{01}p_{11}^{s-2} & P_{11}P_{00}^2P_{10}^{s-2} \end{bmatrix}.
\end{aligned} \tag{41}$$

In the rest of this paper we need to use $\frac{\partial}{\partial\varepsilon}R_n(s,\varepsilon)|_{\varepsilon=0}$ and $\frac{\partial^2}{\partial\varepsilon^2}R_n(s,\varepsilon)|_{\varepsilon=0}$, which will compute in Appendix B1. With spectral representation of the matrix $\mathbf{P}(s)$, we have

$$\lim_{n\to\infty}\frac{1}{n}\cdot\frac{\partial^2}{\partial\varepsilon\partial s}R_n(s,\varepsilon)|_{\varepsilon=0,s=1} = \mathbf{\Pi}(1)\frac{\partial}{\partial s}(\mathbf{Q}(s)-\mathbf{P}^2(s))|_{s=1}\mathbf{1}^t + o(1), \tag{42}$$

and

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\cdot\frac{\partial^3}{\partial\varepsilon^2\partial s}R_n(s,\varepsilon)|_{\varepsilon=0,s=1} &= -2\mathbf{\Pi}(1)\frac{\partial}{\partial s}(\mathbf{Q}(s)-\mathbf{P}^2(s))|_{s=1}\\
&+ \mathbf{\Pi}(1)\frac{\partial}{\partial s}(\mathbf{Q}_1(s)\mathbf{Q}_3(s)\mathbf{Q}_2(s)-\mathbf{P}^3(s))|_{s=1}\mathbf{1}^t\\
&+ \mathbf{\Pi}(1)(\mathbf{Q}_1^{(2)}(1)\mathbf{Q}_2^{(2)}(1)+2\mathbf{Q}_4(1))\mathbf{1}^t + o(1).
\end{aligned} \tag{43}$$

For more details, see Appendix B2.

It is easy to show that $\frac{\partial}{\partial s}(\mathbf{Q}(s)-\mathbf{P}^2(s))|_{s=1}$ is equal to

$$\begin{bmatrix} p_{00}^2\log\frac{p_{01}p_{10}}{p_{00}^2} + p_{01}p_{10}\log\frac{p_{00}^2}{p_{01}p_{10}} & p_{00}p_{01}\log\frac{p_{01}p_{11}}{p_{00}p_{01}} + p_{01}p_{11}\log\frac{p_{00}p_{01}}{p_{01}p_{11}}\\ p_{10}p_{00}\log\frac{p_{11}p_{10}}{p_{10}p_{00}} + p_{11}p_{10}\log\frac{p_{10}p_{00}}{p_{11}p_{10}} & p_{10}p_{01}\log\frac{p_{11}^2}{p_{10}p_{01}} + p_{11}^2\log\frac{p_{10}p_{01}}{p_{11}^2} \end{bmatrix}. \tag{44}$$

It follows

$$\Pi(1)\frac{\partial}{\partial s}(\mathbf{Q}(s)-\mathbf{P}^2(s))|_{s=1}\mathbf{1}^t = \sum_{z_1z_2z_3} P_X(z_1z_2z_3)\log\frac{P_X(z_1z_2z_3)}{P_X(z_1\bar{z}_2z_3)} = \mathbf{D}(P_X(z_1z_2z_3)||P_X(z_1\bar{z}_2z_3)). \quad (45)$$

If

$$\mathbf{M} := \frac{\partial}{\partial s}(\mathbf{Q}_1(s)\mathbf{Q}_3(s)\mathbf{Q}_2(s)-\mathbf{P}^3(s))|_{s=1}, \quad (46)$$

with entries

$$\mathbf{M}_{00} = p_{00}^3\log\frac{p_{01}p_{11}p_{10}}{p_{00}^3} + 2p_{00}p_{01}p_{10}\log\frac{p_{00}p_{01}p_{10}}{p_{00}p_{10}p_{01}} + p_{11}p_{01}p_{10}\log\frac{p_{00}^3}{p_{11}p_{01}p_{10}},$$

$$\mathbf{M}_{01} = p_{00}^2p_{01}\log\frac{p_{11}^2p_{01}}{p_{00}^2p_{01}} + p_{01}^2p_{10}\log\frac{p_{00}p_{01}p_{10}}{p_{01}^2p_{10}} + p_{00}p_{01}p_{11}\log\frac{p_{01}^2p_{10}}{p_{00}p_{01}p_{11}} + p_{01}p_{11}^2\log\frac{p_{00}^2p_{01}}{p_{01}p_{11}^2},$$

$$\mathbf{M}_{10} = p_{10}p_{00}^2\log\frac{p_{10}p_{11}^2}{p_{10}p_{00}^2} + p_{00}p_{10}p_{11}\log\frac{p_{10}^2p_{01}}{p_{00}p_{10}p_{11}} + p_{01}p_{10}^2\log\frac{p_{00}p_{11}p_{10}}{p_{01}p_{10}^2} + p_{10}p_{11}^2\log\frac{p_{00}^2p_{10}}{p_{10}p_{11}^2},$$

$$\mathbf{M}_{11} = p_{10}p_{00}p_{01}\log\frac{p_{11}^3}{p_{10}p_{00}p_{01}} + 2p_{11}p_{10}p_{01}\log\frac{p_{10}p_{01}p_{11}}{p_{11}p_{10}p_{01}} + p_{11}^3\log\frac{p_{01}p_{10}p_{00}}{p_{11}^3},$$

then

$$\Pi(1)\frac{\partial}{\partial s}(\mathbf{Q}_1(s)\mathbf{Q}_3(s)\mathbf{Q}_2(s)-\mathbf{P}^3(s))|_{s=1}\mathbf{1}^t = \sum_{z_1z_2z_3z_4} P_X(z_1z_2z_3z_4) \times\log\frac{P_X(z_1z_2z_3z_4)}{P_X(z_1\bar{z}_2\bar{z}_3z_4)} = \mathbf{D}(P_X(z_1z_2z_3z_4)||P_X(z_1\bar{z}_2\bar{z}_3z_4)). \quad (47)$$

At the end, using (36), (42), (43), (45) and (47), we have the following theorem.

Theorem 2. *Suppose $p_{ab} > 0$ for any $a, b \in \{0, 1\}$. The second order term in the entropy rate of the process Z is,*

$$H(\mathcal{Z}) = H(\mathcal{X}) + (\varepsilon - \varepsilon^2)\mathbf{D}(P_X(z_1z_2z_3)||P_X(z_1\overline{z}_2z_3)) + \frac{\varepsilon^2}{2}\mathbf{D}(P_X(z_1z_2z_3z_4)||P_X(z_1\overline{z}_2\overline{z}_3z_4)) + \frac{\varepsilon^2}{2}\mathbf{\Pi}(1)(\mathbf{Q}_1^{(2)}(1)\mathbf{Q}_2^{(2)}(1) + 2\mathbf{Q}_4(1))\mathbf{1}^t + o(\varepsilon^3). \quad (48)$$

Conclusions

We studied the entropy rate of the hidden Markov chain defined as the output of a symmetric channel whose input is a Markov chain. We used the Taylor expansion to compute the first order term in the entropy rate of the hidden Markov chain and the second order term in the entropy rate of a binary hidden Markov chain. We will try to obtain the higher order term in the entropy rate of the general hidden Markov processes.

References

Bahi, L.R., Cocke, J., Jelinek, F. and Raviv, J. (1974). Optimal decoding of linear codes for minimizing symbol error rate. *J. IEEE Trans. Inform. Theory* **IT-20**, 284-287.

Blackwell, D. (1957). The entropy of functions of finite-state Markov chains. *First Prague Conf. Information Theory, Statistical Decision Functions, Random Processes* 13-20.

Bremaud, P. (1998). *Markov Chains.* Springer Verlag, NewYork, Berlin, Heidelberg.

Churchill, G.A. (1989). Stochastic models for heterogeneous DNA sequences. *J. Bull. Math. Biol.* **51**, 79-94.

Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory.* Jone Wiley and Sons, Inc., New York.

Drake, A.W. (1965). Observation of a Markov source through a noisy channel. *Proc. IEEE Symp. Signal Transmission and Processing*, Columbia Univ., New York, 12-18.

Han, G. and Marcus, B. (2006). Analyticity of entropy rate of hidden Markov chains. *J. IEEE Trans. Inform. Theory*, **52**, 5251-5266.

Jacquent, P., Seroussi, G. and Szpankowski, W. (2004). On the entropy of a hidden Markov process. *Data Compression Conference (DCC'04)*, 362-371.

Jacquent, P., Seroussi, G. and Szpankowski, W. (2008). On the entropy of a hidden Markov process. *J. Theoretical Computer Science*, **395**, 203-219.

Jelinek, F., Bahi, L.R. and Mercer, R.L. (1975). Design of a linguistic statistical decoder for recognition of continuous speech. *J. IEEE Trans. Inform. Theory* **IT-21**, 250-256.

Karlin, S. and Taylor, H. (1975). *A First Course in Stochastic Processes*, Academic Press, New York.

Luo, J. and Guo, D. (2009). On the entropy rate of hidden markov processes observed through arbitrary memoryless channels. *J. IEEE Trans. Inform. Theory* **55**, 1460-1467.

Ordentlich, E. and Weissman, T. (2004). New bounds on the entropy rate pf hidden Markov process. *Information theory workshop San Antonio*, 117-122.

Ordentlich, E. and Weissman, T. (2006). On the optimality of symbol by symbol filtering and denoising. *J. IEEE Trans. Inform. Theory* **52**, 19-40.

Raviv, J. (1967). Decision making in Markov chains applied to the problem of pattern recognition. *J. IEEE Trans. Inform. Theory* **IT-3**, 536-551.

Shannon, C. (1948). A mathematical theory of communication. *J. Bell Syst. Tech.* **27**, 379-423, 623-656.

Zuk, O., Kanter, I. and Domany, E. (2005). Asymptotics of the entropy rate for a hidden Markov process. *J. Stat. Phys.* **121**, 343-360.

Appendix

Appendix A

We use the spectral representation (Karlin, 1975) of the matrix $\mathbf{P}(s)$. Since $p_{ab} > 0$, for any $a, b \in \{0, 1, \ldots, m-1\}$, the Perron-Frobenius theorem (Bremaud, 1998) applies. So there exists a real eigenvalue $\lambda_1(s)$ with algebraic geometric multiplicity one such that $\lambda_1(s) > 0$, and $\lambda_1(s) > |\lambda_j(s)|$ for any other eigenvalue $\lambda_j(s)$. Moreover the left eigenvector $l_1(s)$ and the right eigenvector $r_1(s)$ associated with $\lambda_1(s)$ can be chosen positive and such that $l_1(s)r_1^t(s) = 1$.

Let $\lambda_2(s), \lambda_3(s), \ldots, \lambda_m(s)$ be the eigenvalues of the $\mathbf{P}(s)$ other than λ_1 ordered in such a way that $\lambda_1(s) > |\lambda_2(s)| > |\lambda_3(s)| > \cdots > |\lambda_m(s)|$ and we know that the vectors r_1 and l_1 are real-valued with nonnegative components.

The matrix spectral representation yields

$$\mathbf{P}^k(s) = \lambda_1^k(s)(r_1^t(s)l_1(s)) + o(|\lambda_2|^k). \tag{49}$$

Since $\mathbf{P}(1)$ is a positive stochastic matrix, we have $\lambda_1(1) = 1$, so$|\lambda_2| < 1$.

Note that $(r_1^t(s)l_1(s))$ is an outer product resulting in a $m \times m$ matrix. Define $\mathbf{Q}_k(s) = \mathbf{Q}_{1k}(s)\mathbf{Q}_{2k}(s)$. It follows immediately that $\mathbf{Q}_k(1) = \mathbf{P}^2(1)$ and $\mathbf{Q}_{1k}(1) = \mathbf{Q}_{2k}(1) = \mathbf{P}(1)$. Therefore, by differentiating $\frac{\partial}{\partial\varepsilon}R_n(s,\varepsilon)|_{\boldsymbol{\Upsilon}=0}$ respect to s at $s = 1$, the only terms do not vanish in the derivative are those involving the derivative of $\mathbf{Q}_k(s)-\mathbf{P}^2(s)$, $\mathbf{Q}_{1k}(s)-\mathbf{P}(s)$, $\mathbf{Q}_{2k}(s)-\mathbf{P}(s)$. Now, differentiating from both sides of (29) respect to s at $s = 1$, and simplifying power sums, one can obtain:

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\cdot\frac{\partial^2}{\partial\varepsilon_k\partial s}R_n(s,\boldsymbol{\Upsilon})|_{\boldsymbol{\Upsilon}=0,s=1} &= \Pi(1)(r_1^t(1)l_1(1))\frac{\partial}{\partial s}\\ &\times(\mathbf{Q}_k(s)-\mathbf{P}^2(s))|_{s=1}(r_1^t(1)l_1(1))\mathbf{1}^t + o(1).\end{aligned} \tag{50}$$

For the transition probability matrix $\mathbf{P} = \mathbf{P}(1)$ we have $l_1(1) = \Pi(1)$ and $r_1(1) = [1, 1, \ldots, 1]$.
Thus $\Pi(1)r_1^t(1) = l_1(1)\mathbf{1}^t = 1$ and (50) are simplified to

$$\lim_{n\to\infty}\frac{1}{n}\cdot\frac{\partial^2}{\partial\varepsilon_k\partial s}R_n(s,\boldsymbol{\Upsilon})|_{\boldsymbol{\Upsilon}=0,s=1} = \Pi(1)\frac{\partial}{\partial s}(\mathbf{Q}_k(s)-\mathbf{P}^2(s))|_{s=1}\mathbf{1}^t+o(1). \tag{51}$$

Appendix B

Due to the definition of $\mathbf{Q}_1(s)$ and $\mathbf{Q}_2(s)$ in (41) we have

$$\begin{aligned}\frac{\partial}{\partial\varepsilon}R_n(s,\varepsilon)|_{\varepsilon=0} &= s(\overline{\Pi}(s)\mathbf{Q}_2(s) - \Pi(s)\mathbf{P}(\mathbf{s}))\mathbf{P}^{n-2}\mathbf{1}^t\\ &+ s\Pi(s)\sum_{i=1}^{n-2}\mathbf{P}^{i-1}(s)(\mathbf{Q}_1(s)\mathbf{Q}_2(s) - \mathbf{P}^2(s))\mathbf{P}^{n-i-2}(s)\mathbf{1}^t\\ &+ s\Pi(s)\mathbf{P}^{n-2}(s)(\mathbf{Q}_1(s) - \mathbf{P}(s))\mathbf{1}^t,\end{aligned} \tag{52}$$

and

$$
\begin{aligned}
\frac{\partial^2}{\partial\varepsilon^2}R_n(s,\varepsilon)|_{\varepsilon=0} = {} & s(s-1)[n^2\Pi(s)\mathbf{P}^{n-1}(s) - 2n\overline{\Pi_1}(s)\mathbf{Q}_2(s)\mathbf{P}^{n-2}(s) \\
& - 2n\Pi(s)\mathbf{P}^{n-2}(s)\mathbf{Q}_1(s) + \Pi(s)\mathbf{P}^{n-2}(s)\mathbf{Q}_1^{(2)}(s) \\
& - 2n\sum_{i=2}^{n-1}\Pi(s)\mathbf{P}^{i-2}(s)\mathbf{Q}_1(s)\mathbf{Q}_2(s)\mathbf{p}^{n-i-1}(s) \\
& + \overline{\Pi_2}(s)\mathbf{Q}_2^{(2)}\mathbf{P}^{n-2}(s) \\
& + \sum_{i=2}^{n-1}\Pi(s)\mathbf{P}^{i-2}(s)\mathbf{Q}_1^{(2)}(s)\mathbf{Q}_2^{(2)}(s)\mathbf{p}^{n-i-1}(s) \\
& + 2\Pi_1(s)\mathbf{Q}_4(s)\mathbf{Q}_2(s)\mathbf{P}^{n-3}(s) \\
& + 2\sum_{i=3}^{n-1}\overline{\Pi_1}(s)\mathbf{Q}_2(s)\mathbf{P}^{i-3}(s)\mathbf{Q}_1(s)\mathbf{Q}_2(s)\mathbf{p}^{n-i-1}(s) \\
& + 2\sum_{i=2}^{n-3}\sum_{j=i+2}^{n-1}\Pi(s)\mathbf{P}^{i-2}(s)\mathbf{Q}_1(s)\mathbf{Q}_2(s)\mathbf{p}^{j-i-2}(s) \\
& \times \mathbf{Q}_1(s)\mathbf{Q}_2(s)\mathbf{p}^{n-j-1}(s) \\
& + 2\Pi(s)\mathbf{P}^{n-3}(s)\mathbf{Q}_1(s)\mathbf{Q}_4(s) + 2\overline{\Pi_1}(s)\mathbf{Q}_2(s)\mathbf{P}^{n-3}(s)\mathbf{Q}_1(s) \\
& + 2\sum_{i=2}^{n-2}\Pi(s)\mathbf{P}^{i-2}(s)\mathbf{Q}_1(s)\mathbf{Q}_4(s)\mathbf{Q}_2(s)\mathbf{p}^{n-i-2}(s) \\
& + 2\sum_{i=2}^{n-2}\Pi(s)\mathbf{P}^{i-2}(s)\mathbf{Q}_1(s)\mathbf{Q}_2(s)\mathbf{p}^{n-i-2}(s)\mathbf{Q}_1(s)]\mathbf{1}^t \\
& + 2s[\frac{n(n-1)}{2}\Pi(s)\mathbf{P}^{n-1}(s) - (n-1)\overline{\Pi_1}(s)\mathbf{Q}_2(s)\mathbf{P}^{n-2}(s) \\
& + \sum_{i=2}^{n-2}\Pi(s)\mathbf{P}^{i-2}(s)\mathbf{Q}_1(s)\mathbf{Q}_3(s)\mathbf{Q}_2(s)\mathbf{p}^{n-i-2}(s) \\
& - (n-1)\Pi(s)\mathbf{P}^{n-2}(s)\mathbf{Q}_1(s) \\
& + \sum_{i=3}^{n-1}\overline{\Pi_1}(s)\mathbf{Q}_2(s)\mathbf{P}^{i-3}(s)\mathbf{Q}_1(s)\mathbf{Q}_2(s)\mathbf{p}^{n-i-1}(s)
\end{aligned}
$$

$$
\begin{aligned}
&+ \Pi(s)\mathbf{P}^{n-3}(s)\mathbf{Q}_1(s)\mathbf{Q}_3(s) \\
&+ \sum_{i=2}^{n-2} \Pi(s)\mathbf{P}^{i-2}(s)\mathbf{Q}_1(s)\mathbf{Q}_2(s)\mathbf{p}^{n-i-2}(s)\mathbf{Q}_1(s) \\
&+ \overline{\Pi_1}(s)\mathbf{Q}_3(s)\mathbf{Q}_2(s)\mathbf{P}^{n-3}(s) \\
&+ \sum_{i=2}^{n-3}\sum_{j=i+2}^{n-1} \Pi(s)\mathbf{P}^{i-2}(s)\mathbf{Q}_1(s)\mathbf{Q}_2(s)\mathbf{p}^{j-i-2}(s) \\
&\times \mathbf{Q}_1(s)\mathbf{Q}_2(s)\mathbf{p}^{n-j-1}(s) \\
&- (n-1)\sum_{i=2}^{n-1} \Pi(s)\mathbf{P}^{i-2}(s)\mathbf{Q}_1(s)\mathbf{Q}_2(s)\mathbf{P}^{n-i-1}(s) \\
&+ \overline{\Pi_1}(s)\mathbf{Q}_2(s)\mathbf{P}^{n-3}(s)\mathbf{Q}_1(s)]\mathbf{1}^t. \qquad (53)
\end{aligned}
$$

Appendix B2

With spectral representation of the matrix $\mathbf{P}(s)$ as introduced in Appendix A, and with (52) and (53), we have

$$
\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\cdot\frac{\partial^2}{\partial\varepsilon\partial s}R_n(s,\varepsilon)|_{\varepsilon=0,s=1} &= \Pi(1)(r_1^t(1)l_1(1))\frac{\partial}{\partial s} \\
&\times(\mathbf{Q}(s)-\mathbf{P}^2(s))|_{s=1}(r_1^t(1)l_1(1))\mathbf{1}^t + o(1),
\end{aligned} \qquad (54)
$$

and

$$
\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\cdot\frac{\partial^3}{\partial\varepsilon^2\partial s}R_n(s,\varepsilon)|_{\varepsilon=0,s=1} &= -2\Pi(1)(r_1^t(1)l_1(1))\frac{\partial}{\partial s} \\
&\times(\mathbf{Q}(s)-\mathbf{P}^2(s))|_{s=1}(r_1^t(1)l_1(1))\mathbf{1}^t \\
&+ \mathbf{\Pi}(1)(r_1^t(1)l_1(1))(\mathbf{Q}_1^{(2)}(1)\mathbf{Q}_2^{(2)}(1) \\
&+ 2\mathbf{Q}_4(1))(r_1^t(1)l_1(1))\mathbf{1}^t \\
&+ \Pi(1)(r_1^t(1)l_1(1))\frac{\partial}{\partial s}(\mathbf{Q}_1(s)\mathbf{Q}_3(s)\mathbf{Q}_2(s) \\
&- \mathbf{P}^3(s))|_{s=1}(r_1^t(1)l_1(1))\mathbf{1}^t + o(1). \qquad (55)
\end{aligned}
$$

Gh. Yari
School of Mathematics,
Iran University of Science and Technology,
Tehran, Iran.
email: *yari@iust.ac.ir*

Z. Nikooravesh
School of Mathematics,
Iran University of Science and Technology,
Tehran, Iran.
email: *z_nikooravesk@iust.as.ir*