Maximum Generating Rate of Variable-Length Markov Random Sequences

OSAMU UCHIDA
Department of Human and Information Science
Tokai University
1117 Kitakaname, Hiratsuka, Kanagawa, 259-1292
JAPAN

Abstract: - In this paper, we investigate the maximum generating rate of the variable-length Markov random sequence. Our results are generalizations of Han's results on the variable-length intrinsic randomness.

Key-Words: - Random number generation, general source, Markov process, divergence distance, variational distance, variable-length random number

1 Introduction

In 1995, Vembu and Verdu [6] considered the following problem, called the intrinsic randomness problem:

> At what rate can we generate fair random bits using the given general source $\boldsymbol{X}$ with arbitrary small (but nonzero) tolerance?

and clarified that the supremum of achievable *fixed-length* intrinsic randomness rates is equal to the spectral inf-entropy rate $\underline{H}(\boldsymbol{X})$ (see [5]) of the source $\boldsymbol{X}$, and the supremum of achievable *variable-length* intrinsic randomness rates is equal to $\liminf_{n\to\infty}\frac{1}{n}H(X^n)$. Latterly, the results of variable-length intrinsic randomness are generalized to the case with *countably infinite* source alphabet by Han [2].

On the other hand, Han and Uchida [4] considered the problem of variable-length *nonuniform* random number generation and showed that an optimal source code with cost can be regarded as a variable-length nonuniform random number generator. In this paper, we establish formulae for the optimal generating rate of the *variable-length Markov* random sequence. Our results are generalizations of the results of Han [2] on the variable-length intrinsic randomness.

2 Preliminaries

Let $\mathcal{X}$ be a countably infinite alphabet and $\mathcal{Y} = \{0,1,\cdots,K-1\}$ be a finite alphabet, called source alphabet and code alphabet, respectively. Let $\boldsymbol{X} = \left\{X^n = (X_1^{(n)}, X_2^{(n)}, \cdots, X_n^{(n)})\right\}_{n=1}^{\infty}$ be the general source (see, e.g., [5, 2]), where each component random variable $X_i^{(n)}$ $(1 \le i \le n)$ takes values in $\mathcal{X}$. The stochastic matrix and initional probability distribution of target process $\boldsymbol{Y}$ is denoted by $Q = \{Q(k|j)\}_{j,k\in\mathcal{Y}}$ and $\boldsymbol{q} = \{q(k)\}_{k\in\mathcal{Y}}$, respectively[1]. With any nonnegative integer m we define the random variable $Y^{(m)}$ taking values in $\mathcal{Y}^m$ by

$$\Pr\{Y^{(m)} = (y_1,\cdots,y_m)\} = q(y_1)\prod_{i=2}^{m} Q(y_i|y_{i-1})$$

where m is called the length of $Y^{(m)}$ and $Y^{(0)}$ denotes the constant random variable that coincides with the null string Λ with probability 1. From now on, we call $Y^{(m)}$ the Markov random sequence of length m. Moreover, given a nonnegative integer-valued random variable I, we call $Y^{(I)}$ the *variable-length Markov random sequence*.

Let $\mathcal{Y}^*$ be the set of all finite strings (including the null string Λ) taken from $\mathcal{Y}$. Given a

[1] We assume that the process $\boldsymbol{Y}$ is irreducible and $Q(k|j) > 0$ $(\forall j, k \in \mathcal{Y})$.

variable-length mapping $\varphi_n : \mathcal{X}^n \to \mathcal{Y}^*$, we define the set $\mathcal{D}_m$ for any nonnegative integer m by

$$\mathcal{D}_m = \{\boldsymbol{x} \in \mathcal{X}^n \mid l(\varphi_n(\boldsymbol{x})) = m\},$$

where $l(\boldsymbol{y})$ denotes the length of $\boldsymbol{y} \in \mathcal{Y}^*$ and we put

$$\mathcal{J}(\varphi_n) = \{m \mid \Pr\{X^n \in \mathcal{D}_m\} > 0\}.$$

For any $m \in \mathcal{J}(\varphi_n)$, we define X_m^n as the random variable taking values in $\mathcal{D}_m$ with the distribution given by

$$P_{X_m^n}(\boldsymbol{x}) = \frac{P_{X^n}(\boldsymbol{x})}{\Pr\{X^n \in \mathcal{D}_m\}} \quad (\boldsymbol{x} \in \mathcal{D}_m).$$

Let us now consider to construct a mapping $\varphi_n : \mathcal{X}^n \to \mathcal{Y}^*$ for all $m \in \mathcal{J}(\varphi_n)$ such that $\varphi_n(X_m^n)$ asymptotically approximates the Markov random sequence of length m. That is to say, we consider the problem of constructing $\varphi_n : \mathcal{X}^n \to \mathcal{Y}^*$ such that $\varphi_n(X^n)$ asymptotically approximates a variable-length Markov random sequence. The average length per source letter of the variable-length Markov random sequence generated by φ_n is given by

$$\frac{1}{n}E\{l(\varphi_n(X^n))\} = \frac{1}{n}\sum_{m \in \mathcal{J}(\varphi_n)} m \Pr\{X^n \in \mathcal{D}_m\}$$

which we call the generating rate of the variable-length Markov random sequence. In the following section, we consider to generate a variable-length Markov random sequence with as large generating rate as possible by transforming the coin random number $\boldsymbol{X} = \{X^n\}_{n=1}^\infty$. (In this paper, all the logarithms are taken to the base K, and we assume that $0 \log 0 = 0$).

3 Generating Rate of the Variable-Length Markov Random Number

In this section, we investigate the maximum generating rate of the variable-length Markov random sequence.

3.1 Case with the Divergence Distance

First, we formulate the problem as follows. (Hereafter, we use the denotation $D(X||Y)$ as the divergence distance $D(P_X||P_Y) \equiv \sum_{z \in \mathcal{Z}} P_X(z) \log \frac{P_X(z)}{P_Y(z)}$).

Definition 1 : R is called an *achievable variable-length Markov random sequence generating rate* for the source $\boldsymbol{X}$ if there exists a mapping $\varphi_n : \mathcal{X}^n \to \mathcal{Y}^*$ such that

$$\liminf_{n\to\infty} \frac{1}{n}E\{l(\varphi_n(X^n))\} \geq R,$$

$$\lim_{n\to\infty} \sup_{m \in \mathcal{J}(\varphi_n)} D(\varphi_n(X_m^n)||Y^{(m)}) = 0.$$

Moreover, we define the *supremum achievable variable-length random number generating rate* $S^+(\boldsymbol{X})$ by the supremum of achievable variable-length Markov random sequence generating rates.

With this definition, we have the following first main theorem.

Theorem 1 : For any general source $\boldsymbol{X} = \{X^n\}_{n=1}^\infty$, we have

$$S^+(\boldsymbol{X}) = \frac{1}{H(\boldsymbol{Y})} \liminf_{n\to\infty} \frac{1}{n} H(X^n).$$

where $H(\boldsymbol{Y})$ is the entropy rate of the target process $\boldsymbol{Y}$.

First, to prove Theorem 1, we ready one lemma.

Lemma 1 : Let $\{\mathcal{W}_n\}_{n=1}^\infty$ be a sequence of finite sets and $R > 0, a > 0$ be any constants. Suppose that the probabilities of the random variable W_n taking values in $\mathcal{W}_n$ satisfy the condition

$$P_{W_n}(w) \leq K^{-n(a+\gamma)R} \quad (\forall w \in \mathcal{W}_n),$$

where $\gamma > 0$ is an arbitrary small constant. Then, there exists a mapping $\varphi_n : \mathcal{W}_n \to \mathcal{Y}^{\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor}$ such that

$$\begin{aligned} &D(\varphi_n(W_n)||Y^{(\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor)}) \\ &\quad \leq naR\Big\{K^{-\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor(H(\boldsymbol{Y})-\varepsilon)} + K^{-n\gamma R} \\ &\qquad + K^{-\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor c}\Big\},\end{aligned}$$

where $\varepsilon > 0$ is an arbitrarily small constant and

$$c = \varepsilon - \frac{K^2 \log\left(\left\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\right\rfloor + 1\right)}{\left\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\right\rfloor}.$$

Proof of Lemma 1:

Given $\boldsymbol{y} \in \mathcal{Y}^*$, let $n(v|u)$ be the number of transitions from the letter u to the letter v in $\boldsymbol{y}$ with the cyclic convention that y_n precedes y_1. Let $P(u,v) = \frac{n(v|u)}{n}$ and $p(u) = \sum_{v\in\mathcal{Y}} P(u,v)$. Then, given a sequence $\boldsymbol{y} \in \mathcal{Y}^*$, we define its Markov type as the empirical distribution on $\mathcal{Y}\times\mathcal{Y}$ given by $\{Q(u,v)\}_{u,v\in\mathcal{Y}}$. Moreover, we define the conditional empirical divergence as

$$D(P||Q|p) = \sum_{u,v\in\mathcal{Y}} p(u)P(v|u)\log\frac{P(v|u)}{Q(v|u)},$$

where $P(v|u) = \frac{P(u,v)}{p(u)}$.

Now, we set

$$T_n^\varepsilon = \left\{\boldsymbol{y} \in \mathcal{Y}^{\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor} \mid D(P||Q|p) \le \varepsilon\right\},$$

where $\varepsilon > 0$ is an arbitrary small constant (T_n^ε is called the $\varepsilon-$ typical set of sequences with length $\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor$). It can be shown (see, e.g., [1, 7]) that

$$\Pr\{Y_n \notin T_n^\varepsilon\} \le K^{-\lfloor naR\frac{1}{H(\boldsymbol{Y})+1}\rfloor c}, \qquad (1)$$

where

$$c = \varepsilon - \frac{K^2 \log\left(\left\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\right\rfloor + 1\right)}{\left\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\right\rfloor}$$

and $Y_n = Y^{(\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor)}$. Next, we number all the elements of T_n^ε in order as

$$T_n^\varepsilon = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_{M_n}\} \quad (M_n = |T_n^\varepsilon|).$$

From the basic property of the typical sequence

$$\begin{aligned} P_{Y_n}(\boldsymbol{y}) &\ge K^{-\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor(H(\boldsymbol{Y})+\varepsilon)} \\ &\ge K^{-naR} \quad (\boldsymbol{y} \in T_n^\varepsilon), \qquad (2) \end{aligned}$$

we have

$$1 \ge \sum_{\boldsymbol{y}\in T_n^\varepsilon} P_{Y_n}(\boldsymbol{y}) \ge \sum_{\boldsymbol{y}\in T_n^\varepsilon} K^{-naR} = M_n K^{-naR},$$

and hence,

$$M_n \le K^{naR}. \qquad (3)$$

Let us now construct the mapping $\varphi_n : \mathcal{W}_n \to \mathcal{Y}^{\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor}$ as follows. First, for $\boldsymbol{y}_1$, select a subset $A(1)$ of $\mathcal{W}_n$ so as to satisfy the conditions:

$$\sum_{w\in A(1)} P_{W_n}(w) \le P_{Y_n}(\boldsymbol{y}_1)$$

and, for any $w' \in \mathcal{W}_n - A(1)$,

$$P_{Y_n}(\boldsymbol{y}_1) < \sum_{w\in A(1)} P_{W_n}(w) + P_{W_n}(w').$$

Next, for $\boldsymbol{y}_2$, select a subset $A(2) \subset \mathcal{W}_n - A(1)$ so as to satisfy the conditions:

$$\sum_{w\in A(2)} P_{W_n}(w) \le P_{Y_n}(\boldsymbol{y}_2)$$

and, for any $w' \in \mathcal{W}_n - A(1)\cup A(2)$,

$$P_{Y_n}(\boldsymbol{y}_2) < \sum_{w\in A(2)} P_{W_n}(w) + P_{W_n}(w').$$

In an analogous manner, we define the subsets $A(3), \cdots, A(M_n - 1)$. Moreover, we define

$$A(M_n) = \mathcal{W}_n - \bigcup_{i=1}^{M_n-1} A(i).$$

Now, define the mapping $\varphi_n : \mathcal{W}_n \to \mathcal{Y}^{\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor}$ as

$$\varphi_n(w) = \boldsymbol{y}_i \ \text{ for } \ w \in A(i) \quad (i = 1, 2, \cdots, M_n)$$

and set $\tilde{Y}_n = \varphi_n(W_n)$. Then, it can be easily shown that

$$P_{\tilde{Y}_n}(\boldsymbol{y}_i) \le P_{Y_n}(\boldsymbol{y}_i) \quad (i = 1, 2, \cdots, M_n - 1), \qquad (4)$$

$$
\begin{aligned}
&P_{\tilde{Y}_n}(\boldsymbol{y}_{M_n}) \\
&\quad\le P_{Y_n}(\boldsymbol{y}_{M_n}) + (M_n - 1)K^{-n(a+\gamma)R} \\
&\qquad + \Pr\{Y_n \notin T_n^{\varepsilon}\} \\
&\quad\le P_{Y_n}(\boldsymbol{y}_{M_n}) + K^{-n\gamma R} \\
&\qquad + K^{-\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor c} \qquad (5)
\end{aligned}
$$

where the last inequality follows from (1) and (3). Therefore, from (2),(4),(5) and the basic property of the typical sequence

$$
P_{Y_n}(\boldsymbol{y}) \le K^{-\lfloor naR\frac{1}{H(\boldsymbol{Y})+1}\rfloor(H(\boldsymbol{Y})-\varepsilon)} \qquad (\forall \boldsymbol{y} \in T_n^{\varepsilon}),
$$

we conclude that

$$
\begin{aligned}
&D(\varphi_n(W_n)||Y_{\boldsymbol{Y}}^{(\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor)}) \\
&= \sum_{\boldsymbol{y}\in\mathcal{Y}^{\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor}} P_{\tilde{Y}_n}(\boldsymbol{y}) \log \frac{P_{\tilde{Y}_n}(\boldsymbol{y})}{P_{Y_n}(\boldsymbol{y})} \\
&= \sum_{i=1}^{M_n} P_{\tilde{Y}_n}(\boldsymbol{y}_i) \log \frac{P_{\tilde{Y}_n}(\boldsymbol{y}_i)}{P_{Y_n}(\boldsymbol{y}_i)} \\
&\le P_{\tilde{Y}_n}(\boldsymbol{y}_{M_n}) \log \frac{P_{\tilde{Y}_n}(\boldsymbol{y}_{M_n})}{P_{Y_n}(\boldsymbol{y}_{M_n})} \\
&\le P_{\tilde{Y}_n}(\boldsymbol{y}_{M_n}) \log \frac{1}{P_{Y_n}(\boldsymbol{y}_{M_n})} \\
&\le naR\left\{K^{-\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor(H(\boldsymbol{Y})-\varepsilon)}\right. \\
&\qquad \left. + K^{-n\gamma R} + K^{-\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor c}\right\}.
\end{aligned}
$$

□

Proof of Theorem 1:

i)Direct Part:

The fundamental way of this proof is equivalent to the way of the proof of the direct part of [2, Theorem 3.1].

Letting $\gamma > 0$ be an arbitrarily small constant, we partition the interval $[0, +\infty)$ into the subintervals as

$$
I_j = [R_j, R_{j+1}) \quad (j = 0, 1, 2, \cdots)
$$

where $R_j = 3\gamma j$. According to this interval partition, we divide the set $\mathcal{X}^n$ into mutually disjoint subsets as follows

$$
S_n^{(j)} = \left\{\boldsymbol{x} \in \mathcal{X}^n \mid \frac{1}{n}\log\frac{1}{P_{X^n}(\boldsymbol{x})} \in I_j\right\} \qquad (j = 0, 1, 2, \cdots).
$$

Next, divide $J = \{0, 1, \cdots\}$ into the following two subsets:

$$
J_1 = \left\{j \ge 1 \mid \Pr\left\{X^n \in S_n^{(j)}\right\} \ge K^{-n\gamma R_j}\right\}, \qquad (6)
$$

$$
J_2 = \{0\} \cup \left\{j \ge 1 \mid \Pr\left\{X^n \in S_n^{(j)}\right\} < K^{-n\gamma R_j}\right\}
$$

and, for each $j \in J_1$ define the random variable $\tilde{X}_j^n$ taking values in $S_n^{(j)}$ by

$$
P_{\tilde{X}_j^n}(\boldsymbol{x}) = \frac{P_{X^n}(\boldsymbol{x})}{\Pr\left\{X^n \in S_n^{(j)}\right\}} \quad (\boldsymbol{x} \in S_n^{(j)}). \quad (7)
$$

Since $\boldsymbol{x} \in S_n^{(j)}$ implies $P_{X^n}(\boldsymbol{x}) \le K^{-nR_j}$, it follows from (6) and(7) that, for all $\boldsymbol{x} \in S_n^{(j)}$ $(j \in J_1)$

$$
P_{\tilde{X}_j^n}(\boldsymbol{x}) \le K^{-n(1-\gamma)R_j}.
$$

Then, by means of Lemma 1 with $R = R_j, a = 1 - 2\gamma, \mathcal{W}_n = S_n^{(j)}, W_n = \tilde{X}_j^n$, there exists a mapping $\varphi_n^{(j)} : S_n^{(j)} \to \mathcal{Y}^{\lfloor n(1-2\gamma)R_j\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor}$ such that

$$
\begin{aligned}
&D(\varphi_n^{(j)}(\tilde{X}_j^n)||Y^{(\lfloor n(1-2\gamma)R_j\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor)}) \\
&\quad\le n(1-2\gamma)R_j\left\{K^{-\lfloor n(1-2\gamma)R_j\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor(H(\boldsymbol{Y})-\varepsilon)}\right. \\
&\qquad \left. + K^{-n\gamma R_j} + K^{-\lfloor n(1-2\gamma)R_j\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor c}\right\} \\
&\qquad\qquad (j \in J_1) \quad (8)
\end{aligned}
$$

where

$$
c = \varepsilon - \frac{K^2\log\left(\left\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\right\rfloor + 1\right)}{\left\lfloor naR\frac{1}{H(\boldsymbol{Y})+\varepsilon}\right\rfloor}.
$$

Now, define the variable-length encoder $\varphi_n : \mathcal{X}^n \to \mathcal{Y}^*$ as

$$
\varphi_n(\boldsymbol{x}) = \begin{cases} \varphi_n^{(j)}(\boldsymbol{x}) & \text{for } \boldsymbol{x} \in S_n^{(j)} \ (\exists j \in J_1), \\ \Lambda & \text{otherwise} \end{cases}
$$

where Λ is the null string of length 0. Then, (8) is rewritten as

$$
\begin{aligned}
&D(\varphi_n(\tilde{X}_j^n)||Y^{(\lfloor n(1-2\gamma)R_j\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor)}) \\
&\quad\le n(1-2\gamma)R_j\left\{K^{-\lfloor n(1-2\gamma)R_j\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor(H(\boldsymbol{Y})-\varepsilon)}\right. \\
&\qquad \left. + K^{-n\gamma R_j} + K^{-\lfloor n(1-2\gamma)R_j\frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor c}\right\} \\
&\qquad\qquad (j \in J_1) \quad (9)
\end{aligned}
$$

Since $R_j = 3\gamma j$, we observe here that for each $j \in J_1$

$$(1-2\gamma)R_{j+1} > (1-2\gamma)R_j > 0$$

holds, which means that the length $\lfloor n(1-2\gamma)R_j \frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor$ of the ranges $\mathcal{Y}^{\lfloor n(1-2\gamma)R_j \frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor}$ of the mappings $\varphi_n^{(j)}$ are all different for all sufficiently learge n ($\forall j \in J_1$).

Moreover, if we put $C_n = \varphi_n^{-1}(\Lambda)$ and define the random variable X_0^n taking values in C_n by

$$P_{X_0^n}(\boldsymbol{x}) = \frac{P_{X^n}(\boldsymbol{x})}{\Pr\{X^n \in C_n\}} \quad (\boldsymbol{x} \in C_n)$$

it follows that

$$D(\varphi_n(X_0^n)||Y^{(0)}) = 0 \quad (\forall n = 1, 2, \cdots). \tag{10}$$

Define

$$\mathcal{J}(\varphi_n) = \{0\} \cup \left\{ \left\lfloor n(1-2\gamma)R_j \frac{1}{H(\boldsymbol{Y})+\varepsilon} \right\rfloor \mid j \in J_1 \right\}$$

and for each $m \in \mathcal{J}(\varphi_n) - \{0\}$ define the random variable X_m^n to be

$$X_m^n = \tilde{X}_j^n \quad \left(m = \left\lfloor n(1-2\gamma)R_j \frac{1}{H(\boldsymbol{Y})+\varepsilon} \right\rfloor \right).$$

Then, (9) can be rewritten in the following form:

$$\begin{aligned} D(\varphi_n(X_m^n)||Y^{(m)}) &\le n(1-2\gamma)R_{j_m} \cdot \\ &\quad \left\{ K^{-\lfloor n(1-2\gamma)R_{j_m} \frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor (H(\boldsymbol{Y})-\varepsilon)} \right. \\ &\quad \left. + K^{-n\gamma R_{j_m}} + K^{-\lfloor n(1-2\gamma)R_{j_m} \frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor c} \right\} \\ &\quad (m \in \mathcal{J}(\varphi_n) - \{0\}) \end{aligned} \tag{11}$$

where $j = j_m$ is uniquely determined by $m \in \mathcal{J}(\varphi_n) - \{0\}$ owing to the equation $m = \lfloor n(1-2\gamma)R_j \frac{1}{H(\boldsymbol{Y})+\varepsilon}\rfloor$. Then, summarizing (10) and (11), we have

$$\lim_{n\to\infty} \sup_{m \in \mathcal{J}(\varphi_n)} D(\varphi_n(X_m^n)||Y^{(m)}) = 0. \tag{12}$$

On the other hand, from the fact that

$$\begin{aligned} &E\{l(\varphi_n(X^n))\} \\ &= \sum_{m \in \mathcal{J}(\varphi_n)} m \Pr\{\varphi_n(X^n) = m\} \\ &= \sum_{j \in J_1} \left\lfloor n(1-2\gamma)R_j \frac{1}{H(\boldsymbol{Y})+\varepsilon} \right\rfloor \cdot \\ &\quad \Pr\left\{ X^n \in S_n^{(j)} \right\}, \end{aligned}$$

we have

$$\begin{aligned} &\frac{1}{n} E\{l(\varphi_n(X^n))\} \\ &\ge \frac{1}{H(\boldsymbol{Y})+\varepsilon} \cdot \\ &\quad \left[\frac{1-2\gamma}{n} H(X^n) - \frac{6\gamma(1-2\gamma)K^{-3n\gamma^2}}{(1-K^{-3n\gamma^2})^2} \right. \\ &\quad \left. -6\gamma(1-2\gamma) - \frac{1}{n} \right] \end{aligned} \tag{13}$$

in a same fashion as in the proof of the converse part of [2, Theorem3.1]. Taking $\liminf_{n\to\infty}$ on the both sides of (13), we fave

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{n} E\{l(\varphi_n(X^n))\} \\ &\ge \frac{1}{H(\boldsymbol{Y})+\varepsilon} \cdot \\ &\quad \left[\liminf_{n\to\infty} \frac{1}{n} H(X^n) - 2\gamma \left(\liminf_{n\to\infty} \frac{1}{n} H(X^n) \right) \right. \\ &\quad \left. -6\gamma(1-2\gamma) \right]. \end{aligned} \tag{14}$$

Since $\gamma > 0$ and $\varepsilon > 0$ are arbitrary small, expressions (12) and (14) imply that any rate R such that

$$R < \frac{1}{H(\boldsymbol{Y})} \liminf_{n\to\infty} \frac{1}{n} H(X^n)$$

is achievable. □

ii)Converse Part:

In view of Pinsker's inequality

$$\frac{\log e}{2} \left(d(\varphi_n(X_m^n), Y^{(m)}) \right)^2 \le D(\varphi_n(X_m^n)||Y^{(m)}),$$

the converse part of Theorem 2 of the following subsection implies the converse part of Theorem 1. □

Remark 1 : In the case where $Y^{(m)}$ is uniform random number, Theorem 1 coincides with [2, Theorem 3.1].

3.2 Case with the Variational Distance

The variational distance $d(X, Y)$ between two random variables X, Y taking values in a finite set $\mathcal{Z}$ is defined by

$$d(X, Y) = \sum_{z \in \mathcal{Z}} |P_X(z) - P_Y(z)|.$$

Definition 2 : R is called an *achievable variable-length Markov random sequence generating rate* for the source $\boldsymbol{X}$ if there exists a mapping $\varphi_n : \mathcal{X}^n \to \mathcal{Y}^*$ such that

$$\liminf_{n \to \infty} \frac{1}{n} E\{l(\varphi_n(X^n))\} \geq R,$$

$$\lim_{n \to \infty} \sup_{m \in \mathcal{J}(\varphi_n)} d(\varphi_n(X_m^n), Y^{(m)}) = 0.$$

Moreover, we define the *supremum achievable variable-length Markov random sequence generating rate* $S(\boldsymbol{X})$ by the supremum of achievable variable-length Markov random sequence generating rates.

With this definition, we have the following second main theorem.

Theorem 2 : For any general source $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$, we have

$$S(\boldsymbol{X}) = \frac{1}{H(\boldsymbol{Y})} \liminf_{n \to \infty} \frac{1}{n} H(X^n).$$

Proof:
i)Direct Part:
In view of Pinsker's inequality

$$\frac{\log e}{2} \left(d(\varphi_n(X_m^n), Y^{(m)}) \right)^2 \leq D(\varphi_n(X_m^n) || Y^{(m)}),$$

the achievability of a rate R in the sense of Definition 1 implies that of the rate R of Definition 2. Then, from Theorem 1, we conclude that any rate R such that

$$R < \frac{1}{H(\boldsymbol{Y})} \liminf_{n \to \infty} \frac{1}{n} H(X^n)$$

is achievable. □

ii)Converse Part:
From the fact that $H(Y^{(m)}) \geq mH(\boldsymbol{Y})$ (see, e.g., [3, Remark 3.4]), we can show the converse part of Theorem 2 in an entirely same manner as in the proof of the converse part of [2, Theorem 4.1]. □

Remark 2 : In the case where $Y^{(m)}$ is uniform random number, Theorem 2 coincides with [2, Theorem 4.1].

References

[1] L. D. Davisson, G. Longo and A. Sgarro, "The error exponent for the noiseless encoding of finite ergodic Markov sources," IEEE Trans. Inform. Theory, vol. IT-27, pp.431–438, July 1981.

[2] T. S. Han, "Theorems on the variable-length intrinsic randomness," *IEEE Trans. Inform. Theory*, Vol. 46, pp.2108–2116, Sept. 2000.

[3] T. S. Han and K. Kobayashi, *Mathematics of Information and Coding*, American Math. Soc., 2002.

[4] T. S. Han and O. Uchida, "Source code with cost as a nonuniform random number generator," *IEEE Trans. Inform. Theory*, vol. 46, pp.712–717, March 2000.

[5] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, pp.752–772, May 1993.

[6] S. Vembu and S. Verdu, "Generating random bits from an arbitrary source: fundamental limits," *IEEE Trans. Inform. Theory*, vol. 41, pp.1322–1332, Sept. 1995.

[7] S. Natarajan, "Large deviations, hypotheses testing, and source coding for finite Marko chains," IEEE Trans. Inform. Theory, vol. IT-31, pp.360–365, May 1985.