PAPER Special Section on Information Theory and Its Applications

Performance Analysis of the Interval Algorithm for Random Number Generation in the Case of Markov Coin Tossing*

Yasutada OOHAMA[†a)], ***Senior Member***

SUMMARY In this paper we analyze the interval algorithm for random number generation proposed by Han and Hoshi in the case of Markov coin tossing. Using the expression of real numbers on the interval [0,1), we first establish an explicit representation of the interval algorithm with the representation of real numbers on the interval [0,1) based one number systems. Next, using the expression of the interval algorithm, we give a rigorous analysis of the interval algorithm. We discuss the difference between the expected number of the coin tosses in the interval algorithm and their upper bound derived by Han and Hoshi and show that it can be characterized explicitly with the established expression of the interval algorithm.

key words: *random number generation, interval algorithm, Markov coin tossing, number systems, performance analysis*

1. Introduction

Simulation problems of generating random sequences from a prescribed information source by using a random sequence from a given information source are called the random number generation. In the random number generation random sequences from a prescribed information sources are called the *target* random sequences which we wish to *produce* and the random sequence from given information sources are called the *coin* random sequences that the target random sequences are *made from*.

There have been several works on the random number generation in the field of computer science and information theory. Some interesting relations between random number generation and information theory have been found in the papers of Elias [1] and Knuth and Yao [2].

Han and Hoshi [3] studied a variable-to-fixed random number generation problem. They studied the method of generating target random sequences of *fixed length* from a prescribed information source by using coin random sequences of *variable length* from a given information source. They proposed a simple algorithm called the interval algorithm and obtained results for its performance analysis.

When *coin* random sequences are from a stationary memoryless source, Han and Hoshi [3] established an upper bound of the average length of coin random sequences necessary to create target random sequences. The derived bound is characterized with a fraction of two entropies of given and prescribed sources and is shown to be asymptotically optimal for large length of output sequences. They further studied an extended case, where coin random sequences are from a stationary Markov information source. We hereafter call the stationary Markov information sources which outputs coin random sequences the Markov coin tossing. Han and Hoshi [3] also investigated a random number generation problem of generating a prescribed target random process using a given coin random process. Watanabe and Han [4] investigated this random generation problem by the information spectrum approach [5].

In [6], the author studied the performance analysis of the interval algorithm for random number generation proposed by Han and Hoshi [3]. Using representation of real numbers, the author refined Han and Hoshi's performance analysis of the interval algorithm. In the above work the author treated the problem that we wish to generate a target random variable by using a coin random sequence from a stationary memoryless source.

In this paper we analyze the interval algorithm for random number generation proposed by Han and Hoshi [3] in the case of Markov coin tossing. We extend the method developded by the author [6] to this case, deriving several explicit results.

As a theoretical extension we have an importance on the study of the random number generation problem in the case of Markov coin tossing. We also have a practical importance on this study. From a practical point of view information resources which output coin random sequence must be easily accessible and available. On the other hand, information resources in the real world that we can easily access to utilize include several data such as text data, digitally processed audio, image or video data. Most of them have memory and are mathematically modeled by Markov information sources. Hence, considering applications of the random number generation in practical situations such that we only have a few choices of information resources available as generators of coin random sequences, we inevitably face to the study of the random number generation in the case of Markov coin tossing.

In this paper we derive explicit results on the performance analysis of the interval algorithm for random number generation using an expression of real numbers in the unit interval [0,1). On the expression of real numbers in the unit interval, we establish a kind of generalized number system based on the stochastic structure of the coin random pro-

Manuscript received January 25, 2020.
Manuscript revised June 1, 2020.

[†]The author is with The University of Electro-Communications, Chofu-shi, 182-8585 Japan.

*This work was presented in part at the Symposium on Nonlinear Theory and Its Applications (NOLTA2016), Yugawara, Japan, Nov. 27–30, 2016.

a) E-mail: oohama@uec.ac.jp

DOI: 10.1587/transfun.2020TAP0008

cess. Using the above representation of real numbers on the interval, we find an explicit expression of the interval algorithm. We further present a rigorous analysis of the interval algorithm using the expression of the algorithm.

We discuss the difference between the expected number of the coin tosses in the interval algorithm and their upper bound derived by Han and Hoshi and show that it can be characterized explicitly with the established expression of the interval algorithm.

An explicit representation of the interval algorithm developed by the author [6] can be extended to the case of Markov coin tossing. However, this case yields some specific difficulty in the performance analysis of the interval algorithm. To state this difficulty we define a map φ representing the interval algorithm. We further define a random variable S which generates the target random variable X by φ, that is, $\varphi(S) = X$. Precise definitions of those quantities will be stated in Sects. 2 and 4. Performance of the interval algorithm is measured by an expected number of coin tossing denoted by $\bar{L}$. In the case where coin random sequences are from a stationary memoryless source we have $H(S) = \bar{L}H$, where H is the entropy rate of the descrete memoryless source. In this case the performance analysis for the interval algorithm is reduced to an evaluation of $H(S)$. However, as stated in [3], this equality does not hold in general in the case of Markov coin tossing. In this paper we present a class of stationary Markov information sources having a *symmetrical property* on their stochastic matrices. We prove that for Markov information sources belonging to this class the above equality holds. For Markov information sources not belonging to this class, another method of evaluating $\bar{L}$ will be necessary.

The results of this paper were presented in part at [7], where several arguments are omitted because of page constraint. Furthermore, it contains a mistake. In this paper we provide those arguments and give a complete proof of our main result on the performance analysis of interval algorithm. We also fix the above mistake in [7].

2. Interval Algorithm for Random Number Generation

Let X be random variables taking values in a finite set $\mathcal{X} := \{0, 1, \cdots, N-1\}$. Let $p_X := \{p_X(x)\}_{x\in\mathcal{X}}$ be a probability distribution of X. Let $\{Y_t\}_{t=1}^{\infty}$ be a stationary Markov source. For each $t = 1, 2, \cdots$, Y_t takes values in a finite set $\mathcal{Y} := \{0, 1, \cdots, M-1\}$. The stationary Markov source $\{Y_t\}_{t=1}^{\infty}$ is specified with the $M \times M$ stochastic matrix denoted by $P = [P_{ij}]$, where

$$P_{ij} = \Pr\{Y_{t+1} = j | Y_t = i\}, \text{ for } t = 1, 2, \cdots.$$

We also write $P_{ij}, (i, j) \times \mathcal{Y}^2$ as $P_{ij} = p_Y(j|i)$. Let $\mathcal{Y}^*$ denote the set of all finite sequence emitted from the above information source. We write a string from information source as $y_l^m := y_l\, y_{l+1} \cdots y_m \in \mathcal{Y}^*$. If $l > m$, the string y_l^m means *null* string denoted by λ. When $l = 1$, we frequently omit the suffix 1 of y_1^m and write $y^m = y_1 y_2 \cdots y_m$. Let $p_Y(y_l^m)$ denote the probability of y_l^m. Since the information source is a stationary Markov source, we have

$$p_Y(y_l^m) = p_Y(y_l) P_{y_l y_{l+1}} \cdots P_{y_{m-1} y_m}.$$

Here $\{p_Y(a)\}_{a\in\mathcal{Y}}$ is a stationary distribution computed from P. The probability of the null string λ assumes to be one.

In this paper we deal with the variable to fixed random number generation problem of generating target random variable X by using the coin random sequence $Y_1 Y_2 \cdots Y_i \cdots$ from a stationary Markov information sources $\{Y_t\}_{t=1}^{\infty}$. A formal definition of the variable to fixed random number generation problem is the following. Repeated tosses of the coin random variable Y produces random sequence $Y_1, Y_2, \cdots$ from a Markov source. The coin toss terminates at some finite time L to generate a random variable X with a prescribed distribution p_X. L is a random variable specified in terms of a deterministic two valued function such that $f(Y^i) =$ 'Continue' for $1 \le i \le L-1$ and $f(Y^L) =$'Stop'. The output X is expressed as $X = \psi(Y^L)$ with some deterministic function ψ.

For the given generating algorithm (f, ψ) of random number generation let $\mathcal{S}_x, x \in \mathcal{X}$ be a set of all input strings $y^l \in \mathcal{Y}^*$ that generate x. It is obvious that $\mathcal{S}_x, x \in \mathcal{X}$ are disjoint. Set

$$\mathcal{S} := \sum_{x\in\mathcal{X}} \mathcal{S}_x,$$

where we have used the notation '$\sum$' for the sum of disjoint sets instead of '$\cup$'. Hereafter, to distinguish the sum of disjoint sets from the union of sets, we use the notation '+' or '$\sum$' for the sum of *disjoint* sets.

In the above random number generation problem Han and Hoshi [3] proposed a simple algorithm called interval algorithm and evaluated its performance. Let $I = [0, 1)$. Define the cumulative probabilities for p_Y by

$$c_Y(0) := 0,$$
$$c_Y(y) := \sum_{i<y} p_Y(i), 1 \le y \le M-1.$$

Using these probabilities, define the decomposition of I by

$$I_Y(y) := [c_Y(y), c_Y(y) + p_Y(y)).$$

For p_X, we use the same notations and definitions as those for p_Y. For given $y_1 \in \mathcal{Y}$, define the cumulative probabilities for $p_Y(\cdot|y_1) = \{p_Y(y_2|y_1)\}_{y_2\in\mathcal{Y}}$ by

$$c_Y(0|y_1) := 0,$$
$$c_Y(y_2|y_1) := \sum_{i<y_2} p_Y(i|y_1), 1 \le y_2 \le M-1.$$

For $k = 1, 2, \cdots$, and any string $y^k = y_1 y_2 \cdots y_k \in \mathcal{Y}^k$, define the semi-open interval $I_Y(y^k) := [L_Y(y^k), U_Y(y^k))$ by the following recursions:

$$\left.\begin{aligned} &L_Y(y_1) = c_Y(y_1),\\ &U_Y(y_1) = c_Y(y_1) + p_Y(y_1)\\ &L_Y(y^i) = L_Y(y^{i-1}) + p_Y(y^{i-1}) c_Y(y_i|y_{i-1}),\\ &U_Y(y^i) = L_Y(y^i) + p_Y(y^i), \text{ for } 2 \le i \le k. \end{aligned}\right\} \quad (1)$$

The procedure of computing upper and lower end points of the interval corresponding to a given sequence is equivalent to the encoding algorithm in the arithmetic coding. On intervals generated by the above recursion we have the following property.

Property 1: For any $n \geq 2$, any $a^n \in \mathcal{Y}^n$, we have that for any $1 \leq m \leq n-1$,

$$[L_Y(a^m), L_Y(a^n)) = \sum_{k=m+1}^{n} \sum_{y<a_k} I_Y(a^{k-1}y), \tag{2}$$

$$[U_Y(a^n), U_Y(a^m)) = \sum_{k=m+1}^{n} \sum_{y>a_k} I_Y(a^{k-1}y). \tag{3}$$

Proof of Property 1 is given in Appendix. This property will be a basis of a key important result, which yields an explicit representation of the interval algorithm. We derive this key result in the next section.

Interval algorithm by Han and Hoshi [3] can be stated in the following.

Interval Algorithm (Han and Hoshi [3]):

1) Set $i = k = 1$, $y_0 = \lambda$.
2) Given y_{i-1}, generate a letter $y_i \in \mathcal{Y}$ according to the transition probability $p_Y(y_i|y_{i-1})$ of the coin random variable. Here for $i = 1$, the quantity $p_Y(y_1|y_0) = p_Y(y_1|\lambda) = p_Y(y_1)$ is the stationary probability of the coin random variable.
3) Compute $I_Y(y^i) = [L_Y(y^i), U_Y(y^i))$ according to the recursion (1).
4) If $I_Y(y^i) \subseteq I_X(x)$ for some $x \in \mathcal{X}$, then output x as the value of target random variable X and stop the algorithm.
5) Set $i = k + 1$ and go to 2).

In the above interval algorithm the target random variable X can exactly be produced.

3. An Explicit Representation of the Interval Algorithm

In this section we give two expressions of real numbers in the interval $I = [0, 1)$ on the number system. There is some complementary relation between the above two expressions. Using those expressions we give an explicit form of the interval algorithm.

3.1 Representation of Real Numbers

For $z \in [0, 1)$, define the sequence $\{a_i\}_{i=1}^{\infty} \in \mathcal{Y}^*$ such that

$$z \in I_Y(a^i), i = 1, 2, \cdots .$$

It can easily be verified that using $a_1, a_2, \cdots$, z can be expressed in the following manner:

$$z = \sum_{k\geq 1} p_Y(a^{k-1}) \sum_{a<a_k} p_Y(a|a_{k-1})$$

$$= \sum_{k\geq 1} p_Y(a^{k-1}) c_Y(a_k|a_{k-1}).$$

Here we assume that $a_0 = \lambda$ for $k = 1$. The same rule of notation will be used in the subsequent arguments. We call the above expression *the p_Y-ary representation of the real number z* and write as

$$z = 0.a_1a_2a_3\cdots . \tag{4}$$

In the above expression, if we wish to express z with the sum of the number having the expression

$$0.a_1a_2a_3\cdots a_t00\cdots$$

and the other remaining term, we write

$$z = 0.a_1a_2\cdots a_t \; + \; 0.0_{a_1}0_{a_2}\cdots 0_{a_t}a_{t+1}\cdots , \tag{5}$$

where the second term is defined by

$$0.0_{a_1}0_{a_2}\cdots 0_{a_t}a_{t+1}\cdots := \sum_{k\geq t+1} p_Y(a^{k-1}) c_Y(a_k|a_{k-1}).$$

Next, for $z \in [0, 1)$, set $\bar{z} = 1 - z$. Using the sequence $\{a_i\}_{i\geq 1}$ appearing in the p_Y-ary representation of the real number z, $\bar{z}$ has an expression

$$\bar{z} = \sum_{k\geq 1} p_Y(a^{k-1}) \sum_{a>a_k} p_Y(a|a_{k-1}).$$

Then, adopting the notation

$$c_Y(\bar{a}|a_{k-1}) := \sum_{i>a} p_Y(i|a_{k-1}) ,$$

we obtain the following expression

$$\bar{z} = \sum_{k\geq 1} p_Y(a^{k-1}) c_Y(\bar{a}_k|a_{k-1}).$$

We call the above expression *the p_Y-ary co-representation of the real number z* and write as

$$\bar{z} = 0.\bar{a}_1\bar{a}_2\bar{a}_3\cdots . \tag{6}$$

Let $z^{(n)}$ denote the real number which is obtained by rounding off z to n-digits in the p_Y-ary representation, that is,

$$z^{(n)} := 0.a_1a_2\cdots a_n.$$

Similarly, let $\bar{z}^{(n)}$ denote the real number which is obtained by rounding off $\bar{z}$ to n-digits in the p_Y-ary co-representation, that is,

$$\bar{z}^{(n)} := 0.\bar{a}_1\bar{a}_2\cdots \bar{a}_n.$$

It can easily be verified that the p_Y-ary representation and the p_Y-ary co-representation of the real number z satisfy the following.

Property 2:

a) For any i, $z \in I_Y(a^i)$.
b) $c_Y(a_i|a_{i-1}) + c_Y(\bar{a}_i|a_{i-1}) = 1 - p_Y(a_i|a_{i-1})$.

c) For $z = 0.a_1a_2\cdots a_n\cdots \in [0,1)$, we have

$$z^{(n)} + \bar{z}^{(n)} = 1 - p_Y(a^n).$$

From Properties 1 and 2, we have the following lemma.

Lemma 1: We assume that z has the following p_Y-ary expression:

$$z = 0.a_1a_2\cdots a_n\cdots \in [0,1).$$

Then for any $m \geq 1$, we have the following:

$$[L_Y(a^m), z) = \sum_{k\geq m+1}\sum_{y<a_k} I_Y(a^{k-1}y), \tag{7}$$

$$[z, U_Y(a^m)) = \sum_{k\geq m+1}\sum_{y>a_k} I_Y(a^{k-1}y). \tag{8}$$

Proof: By Property 1, we have

$$[L_Y(a^m), z^{(n)}) = [L_Y(a^m), L_Y(a^n))$$

$$= \sum_{k=m+1}^{n}\sum_{y<a_k} I_Y(a^{k-1}y), \tag{9}$$

$$[z^{(n)} + p_Y(a^n), U_Y(a^m)) = [U_Y(a^n), U_Y(a^m))$$

$$= \sum_{k=m+1}^{n}\sum_{y>a_k} I_Y(a^{k-1}y). \tag{10}$$

Note that

$$\lim_{n\to\infty} z^{(n)} = \lim_{n\to\infty}(z^{(n)} + p_Y(a^n)) = z.$$

Hence by letting $n \to \infty$ in (9) and (10), we have

$$[L_Y(a^m), z) = \sum_{k\geq m+1}\sum_{y<a_k} I_Y(a^{k-1}y),$$

$$[z, U_Y(a^m)) = \sum_{k\geq m+1}\sum_{y>a_k} I_Y(a^{k-1}y),$$

completing the proof. □

Lemma 1 plays an important role in deriving an explicit representation of the interval algorithm. The detail of derivation is stated in Sect. 5.

Kanaya [8], Oohama et al. [9] point out that the p_Y-ary representation has a close connection with the arithmetic coding and the Markov shift. In the following we explain this connection. Let $\mathcal{A}$ be a set of $y^2 \in \mathcal{Y}^2$ such that $p_Y(y^2) > 0$. Note that

$$\sum_{y^2\in\mathcal{A}} I_Y(y^2) = I, \sum_{y\in\mathcal{Y}} I_Y(y) = I.$$

Define $\tau_Y : I \to I$ and $\phi_Y : I \to \mathcal{Y}$ by

$$\tau_Y(z) = (p_Y(y_1|y_2))^{-1}\left(z - L_Y(y^2)\right) + L_Y(y_1),$$
$$\text{for } y^2 \in \mathcal{A} \text{ and } z \in I_Y(y^2),$$
$$\phi_Y(z) = y, \text{ for } y \in \mathcal{Y} \text{ and } z \in I_Y(y).$$

The map τ_Y is called the Markov shift in the terminology of ergodic theory since it can be regarded as a shift on the Markov process specified with P. As an example of (τ_Y, ϕ_Y), we consider the case where $M = |\mathcal{Y}| = 3$ and

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}. \tag{11}$$

In this example $\mathcal{A} = \mathcal{Y}^2 - \{(0,0)\}$. The stationary distribution is $(p_Y(0), p_Y(1), p_Y(2)) = (0.2, 0.4, 0.4)$. The maps τ_Y and ϕ_Y for P given by (11) are shown in Fig. 1. Let $z \in [0,1)$ be an initial value. We consider the sequence $\phi_Y(z)\phi_Y(\tau_Y(z))\cdots\phi_Y(\tau_Y^{k-1}(z))$ generated by the initial value z, the map τ_Y and the quantizer ϕ_Y. Then, we have the following property.

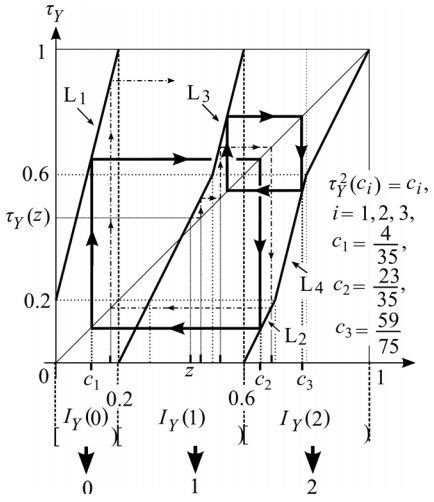


Fig. 1 The maps τ_Y and ϕ_Y for P given by (11). The quantities $c_1 = 4/35$, $c_2 = 23/35$, and $c_3 = 59/75$ satisfies $\tau_Y^2(c_i) = c_i, i = 1,2,3$.

Property 3 (Kanaya [8], Oohama et al. [9]):

a) $I_Y(a_1a_2\cdots a_k)$ is equal to the set of initial values z generating $\phi_Y(z)\phi_Y(\tau_Y(z))\cdots\phi_Y(\tau_Y^{k-1}(z)) = a_1a_2\cdots a_k$.
b) The sequence $\{\phi_Y(\tau_Y^{k-1}(z))\}_{k=0}^{\infty}$ coincides with the p_Y-ary representation of z.
c) The procedure of producing sequence using iteration of τ_Y and quantization by ϕ_Y is equivalent to the decoding process in the arithmetic coding.

The followings are two examples of p_Y-ary representations of $z \in I$.

Example 1: We consider the example where P is given by (11). The map τ_Y is shown in Fig. 1. In this figure the quantities $c_1 = 4/35$, $c_2 = 23/35$, and $c_3 = 59/75$ satisfies $\tau_Y^2(c_i) = c_i, i = 1,2,3$. The line segments $\mathrm{L}_i, i = 1,2$ are related to the computation of $c_i, i = 1,2$. Those are explicitly given by

$$\mathrm{L}_1 : \tau_Y(z) = 4z + 0.2 \text{ for } z \in [0, 0.2),$$
$$\mathrm{L}_2 : \tau_Y(z) = 2z - 1.2 \text{ for } z \in [0.6, 0.7).$$

The line segments $\mathrm{L}_i, i = 3,4$ are related to the computation

of c_3. Those are explicitly given by

$$\mathrm{L}_3 : \tau_Y(z) = 4z - 1.4 \text{ for } z \in [0.5, 0.6),$$
$$\mathrm{L}_4 : \tau_Y(z) = 4z - 2.6 \text{ for } z \in [0.7, 0.8).$$

It can be seen from Fig. 1 that we have

$$\left.\begin{aligned} \phi_Y(c_1) = 0, \phi_Y(\tau_Y(c_1)) = 2, \\ \phi_Y(c_2) = 2, \phi_Y(\tau_Y(c_2)) = 0, \\ \phi_Y(c_3) = 2, \phi_Y(\tau_Y(c_3)) = 1. \end{aligned}\right\} \tag{12}$$

Then by (12) and Property 3 parts a) and b), the p_Y-ary representations of $c_i, i = 1, 2, 3$ are

$$c_1 = 0.02020202\cdots,\ c_2 = 0.20202020\cdots,$$
$$c_3 = 0.21212121\cdots.$$

Example 2: We consider the case where $M = |\mathcal{Y}| = 3$ and

$$P = \begin{bmatrix} 0.25 & 0.25 & 0.5 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}. \tag{13}$$

In this example $\mathcal{A} = \mathcal{Y}^2$. The stationary distribution is $(p_Y(0), p_Y(1), p_Y(2)) = (1/4, 1/3, 5/12)$. The maps τ_Y and ϕ_Y for P given by (13) are shown in Fig. 2. In this figure the quantities $c'_1 = 1/7$, $c'_2 = 7/9$ satisfies $\tau_Y^2(c'_i) = c'_i, i = 1, 2$. In Fig. 2, the line segments $\mathrm{L}_i, i = 1, 2$ are related to the computation of c'_1. Those are explicitly given by

$$\mathrm{L}_1 : \tau_Y(z) = (10/3)z + 1/6 \text{ for } z \in [1/8, 1/4),$$
$$\mathrm{L}_2 : \tau_Y(z) = (12/5)z - 7/5 \text{ for } z \in [7/12, 11/16).$$

The line segments $\mathrm{L}_i, i = 3, 4$ are related to the computation of c'_2. Those are explicitly given by

$$\mathrm{L}_3 : \tau_Y(z) = 5z - 23/12 \text{ for } z \in [1/2, 7/12),$$

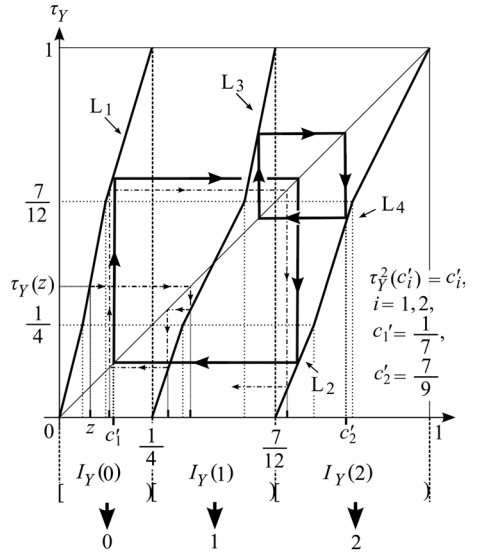


Fig. 2 The maps τ_Y and ϕ_Y for P given by (13). The quantities $c'_1 = 1/7$, $c'_2 = 7/9$ satisfy $\tau_Y^2(c'_i) = c'_i, i = 1, 2$.

$$\mathrm{L}_4 : \tau_Y(z) = (16/5)z - 39/20 \text{ for } z \in [11/16, 19/24),$$

It can be seen from Fig. 2 that we have

$$\left.\begin{aligned} \phi_Y(c'_1) = 0, \phi_Y(\tau_Y(c'_1)) = 2, \\ \phi_Y(c'_2) = 2, \phi_Y(\tau_Y(c'_2)) = 1. \end{aligned}\right\} \tag{14}$$

Then by (14) and Property 3 parts a) and b), the p_Y-ary representations of $c'_i, i = 1, 2$ are

$$c'_1 = 0.02020202\cdots,\ c'_2 = 0.21212121\cdots.$$

3.2 An Explicit Representation of the Interval Algorithm

In this subsection, we give an explicit form of the interval algorithm by using the p_Y-ary representation and p_Y-ary co-representation of the real number in the interval $I = [0, 1)$. It can easily be seen from the definition of the interval algorithm the interval $I_X(x) = [L_X(x), U_X(x))$ corresponding to the target random number $x \in \mathcal{X}$ has a form of a disjoint sum of the intervals $I_Y(\cdot)$. In our previous work we obtained an explicit form of the disjoint sum in the case where the source $\{Y_t\}_{t=1}^{\infty}$ representing coin tossings is a discrete memoryless source. In the present case where $\{Y_t\}_{t=1}^{\infty}$ is a stationary Markov source the same result holds. This result is as follows.

Theorem 1: For $x \in \mathcal{X}$, let $I_X(x) = [L_X(x), U_X(x))$ be an interval corresponding to the target random variable X taking values in $\mathcal{X}$. Suppose that lower and upper endpoints $L_X(x)$ and $U_X(x)$ have the following p_Y-ary representation and p_Y-ary co-representations:

$$L_X(x) = 0.a_1a_2\cdots, \quad \overline{L_X(x)} = 0.\bar{a}_1\bar{a}_2\cdots,$$
$$U_X(x) = 0.b_1b_2\cdots.$$

For each $x \in \mathcal{X}$, there exists an integer $t = t(x)$ such that representations of $L_X(x)$ and $U_X(x)$ have first different values at the t-th place at their p_Y-ary representations. Then, we have

$$p_X(x) = p_Y(a^{t-1})\Bigg[\sum_{a_t<a<b_t} p_Y(a|a_{t-1}) + \sum_{k\ge t+1}\Big\{p_Y(a_t^{k-1}|a_{t-1})c_Y(\bar{a}_k|a_{k-1}) + p_Y(b_t^{k-1}|a_{t-1})c_Y(b_k|b_{k-1})\Big\}\Bigg], \tag{15}$$

where

$$\sum_{a_t<a<b_t} p_Y(a|a_{t-1}) = 0$$

when $b_t = a_t + 1$. Furthermore, we have the following description of $I_X(x)$ with the disjoint sum of intervals corresponding to the target random sequences in the interval algorithm:

$$I_X(x) = \sum_{a_t<y<b_t} I_Y(a^{t-1}y)$$

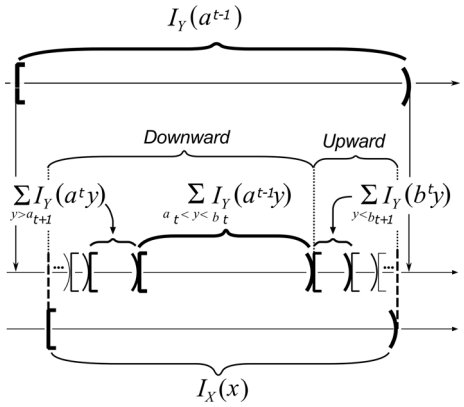


Fig. 3 Upward and downward sequences of intervals.

$$+\sum_{k\geq t+1}\left\{\sum_{y>a_k} I_Y(a^{k-1}y)+\sum_{y<b_k} I_Y(b^{k-1}y)\right\}. \tag{16}$$

Proof of the equality (15) in Theorem 1 is quite parallel with that of the similar equality with respect to $p_X(x), x \in \mathcal{X}$ in [6]. For the equality (16) in Theorem 1, we give a simple and rigorous proof of this equality without depending on the equality (15). Lemma 1 is a key result for the proof. This lemma together with some simple observations on the p_Y-representations of two endpoints $L_X(x)$ and $U_X(x)$ of $I_X(x)$, $x \in \mathcal{X}$ yields (16). The detail of the proof of Theorem 1 is given in Sect. 5.

It can be seen from the above presentation that the interval $\sum_{a_t<y<b_t} I_Y(a^{t-1}y)$ is in the middle of the interval $I_X(x)$ and that the sequence of intervals $\{\sum_{y>a_{k+1}} I_Y(a^k y)\}_{k\geq t}$ entirely covers the lower part of the interval $I_X(x)$. Those intervals are called *downward sequences* in Han and Hoshi [3]. We also know that the sequence of intervals $\{\sum_{y<b_k} I_Y(b^{k-1}y)\}_{k\geq t+1}$ in the third term in the right member of the above equation entirely covers the upper part of $I_X(x)$. This sequence of the intervals are called *upward sequence* in Han and Hoshi [3]. The result of Theorem 1 can be regarded as giving an explicit form of upward/downward sequences of intervals in the interval algorithm. Those sequences of intervals is shown in Fig. 3.

Based on the expression of $p_X(x), x \in \mathcal{X}$ in Theorem 1, set

$$\mathcal{D}_{t,x} := \left\{y^t : y^{t-1} = a^{t-1}, a_t < y_t < b_t\right\}. \tag{17}$$

Furthermore, for $l \geq t+1$, set

$$\mathcal{D}_{l,x} := \left\{y^l : y^{l-1} = a^{l-1}, a_l < y_l\right\}, \tag{18}$$

$$\mathcal{U}_{l,x} := \left\{y^l : y^{t-1} = a^{t-1}, y_t^{l-1} = b_t^{l-1}, y_l < b_l\right\}. \tag{19}$$

Then, we have the following.

$$S = \sum_{x\in\mathcal{X}}\left\{\mathcal{D}_{t,x}+\sum_{l\geq t+1}\mathcal{D}_{l,x}+\sum_{l\geq t+1}\mathcal{U}_{l,x}\right\}. \tag{20}$$

For $x \in \mathcal{X}$, define

$$\mathcal{D}_x := \mathcal{D}_{t,x}+\sum_{l\geq t+1}\mathcal{D}_{l,x},\ \mathcal{U}_x := \sum_{l\geq t+1}\mathcal{U}_{l,x}.$$

It is obvious that $\mathcal{S}_x = \mathcal{D}_x + \mathcal{U}_x, x \in \mathcal{X}$. In the remaining part of this section we present two examples of random number generation. For each example, we compute $\mathcal{S}_x$, $\mathcal{D}_x$, and $\mathcal{U}_x$ for $x \in \mathcal{X}$.

Example 3: We consider the case where $M = 3$, $N = 4$. The target random variable X has the following distribution:

$$\begin{aligned} p_X &= (p_X(0), p_X(1), p_X(2), p_X(3)) \\ &= (4/35, 19/35, 68/525, /16/75). \end{aligned}$$

We assume that the stationary Markov process $\{Y_t\}_{t=1}$ specified with (11) in Example 1 generates coin random sequences. The p_Y-array representation for this example is discussed in Example 1. In the random number generation problem treated here the choice of p_X is closely related to the periodic point of the map τ_Y defining the Markov shift. In fact, we have $L_X(1) = 4/35 = c_1$, $L_X(2) = 23/35 = c_2$, $L_X(3) = 59/75 = c_3$, where $c_i, i = 1, 2, 3$ are the same quantity as those in Example 1. Those are the periodic points of τ_Y satisfying $\tau_Y^2(c_i) = c_i$, $i = 1, 2, 3$. Using the p_Y-array representations of $c_i, i = 1, 2, 3$ in Example 1, we have

$$\begin{aligned} &L_X(1) = 0.02020202\cdots,\ L_X(2) = 0.20202020\cdots, \\ &L_X(3) = 0.21212121\cdots. \end{aligned}$$

Applying the formula (16) on $I_X(x), x \in \mathcal{X}$ in Theorem 1 to the present example, we have the following:

$$\begin{aligned} I_X(0) &= \sum_{k\geq 0}\sum_{y<2} I_Y(0[20]^k y), \\ I_X(1) &= I_Y(1) + \sum_{k\geq 1}\sum_{y>0} I_Y([02]^k y) \\ &\quad + \sum_{k\geq 1}\sum_{y<2} I_Y([20]^k y), \\ I_X(2) &= \sum_{k\geq 0}\sum_{y>0} I_Y(2[02]^k y) \\ &\quad + \sum_{k\geq 1}\left\{\sum_{y<2} I_Y(2[12]^k 1y) + I_Y(2[12]^{k+1}0)\right\}, \\ I_X(3) &= \sum_{k\geq 0} I_Y(2[12]^k 2). \end{aligned}$$

Hence the sets $\mathcal{S}_x$, $\mathcal{D}_x$, and $\mathcal{U}_x$ for $x = 0, 1, 2, 3$ are

$$\begin{aligned} &\mathcal{S}_0 = \mathcal{U}_0 = \left\{0[20]^k 0, 0[20]^k 1\right\}_{k\geq 0}, \\ &\mathcal{S}_1 = \mathcal{D}_1 + \mathcal{U}_1, \\ &\quad \begin{cases} \mathcal{D}_1 = \{1\} + \left\{[02]^k 1, [02]^k 2\right\}_{k\geq 1}, \\ \mathcal{U}_1 = \left\{[20]^k 0, [20]^k 1\right\}_{k\geq 1}, \end{cases} \\ &\mathcal{S}_2 = \mathcal{D}_2 + \mathcal{U}_2, \\ &\quad \begin{cases} \mathcal{D}_2 = \left\{2[02]^k 1, 2[02]^k 2\right\}_{k\geq 1}, \\ \mathcal{U}_2 = \left\{2[12]^k 10, 2[12]^k 11, 2[12]^{k+1}0\right\}_{k\geq 0}, \end{cases} \end{aligned}$$

$$\mathcal{S}_3 = \mathcal{D}_3 = \left\{2[12]^k 2\right\}_{k\geq 0}.$$

Example 4: We consider the case where $M = 3$, $N = 3$. The target random variable X has the following distribution:

$$p_X = (p_X(0), p_X(1), p_X(2)) = (1/7, 40/63, 2/9).$$

We assume that the stationary Markov process $\{Y_t\}_{t=1}$ specified with (13) in Example 2 generates coin random sequences. The p_Y-array representation for this example is discussed in Example 2. In the random number generation problem treated here the choice of p_X is closely related to the periodic point of the map τ_Y defining the Markov shift. In fact, we have $L_X(1) = 1/7 = c'_1$, $L_X(2) = 7/9 = c'_2$, where $c'_i, i = 1, 2$ are the same quantity as those in Example 2. Those are the periodic points of τ_Y satisfying $\tau_Y^2(c'_i) = c'_i$, $i = 1, 2$. Using the p_Y-array representations of $c'_i, i = 1, 2$ in Example 2, we have

$$L_X(1) = 0.02020202\cdots,\ L_X(2) = 0.21212121\cdots.$$

Applying the formula (16) on $I_X(x), x \in \mathcal{X}$ in Theorem 1 to the present example, we have the following:

$$\begin{aligned}
I_X(0) &= \sum_{k\geq 0}\sum_{y<2} I_Y(0[20]^k y),\\
I_X(1) &= I_Y(1) + \sum_{k\geq 1}\sum_{y>0} I_Y([02]^k y)\\
&\quad + \sum_{k\geq 0}\left\{I_Y(2[12]^k 0) + \sum_{y<2} I_Y([21]^{k+1} y)\right\},\\
I_X(2) &= \sum_{k\geq 0} I_Y(2[12]^k 2).
\end{aligned}$$

Hence the sets $\mathcal{S}_x$, $\mathcal{D}_x$, and $\mathcal{U}_x$ for $x = 0, 1, 2$ are

$$\begin{aligned}
&\mathcal{S}_0 = \mathcal{U}_0 = \left\{0[20]^k 0, 0[20]^k 1\right\}_{k\geq 0},\\
&\mathcal{S}_1 = \mathcal{D}_1 + \mathcal{U}_1,\\
&\begin{cases}\mathcal{D}_1 = \{1\} + \left\{[02]^k 1, [02]^k 2\right\}_{k\geq 1},\\ \mathcal{U}_1 = \left\{2[12]^k 0, [21]^{k+1}0, [21]^{k+1}1\right\}_{k\geq 0},\end{cases}\\
&\mathcal{S}_2 = \mathcal{D}_2 = \left\{2[12]^k 2\right\}_{k\geq 0}.
\end{aligned}$$

4. Performance Analysis of the Interval Algorithm

In this section we present a rigorous performance analysis of the interval algorithm using the expression of the interval algorithm we gave in the previous section.

4.1 Some Preliminaries

We define several quantities necessary for describing our result on the performance analysis of the interval algorithm. Let $S \in \mathcal{S}$ be a random variable with the distribution

$$\Pr\left\{S = y^l \in \mathcal{S}\right\} = p_Y(y_1)p_Y(y_2|y_1)\cdots p_Y(y_l|y_{l-1}).$$

For $y^l \in \mathcal{S}$ define the map $\varphi : \mathcal{S} \to \mathcal{X}$ such that

$$\varphi(y^l) := x \text{ if } y^l \in \mathcal{D}_{l,x} \text{ or } y^l \in \mathcal{U}_{l,x}. \tag{21}$$

Define $\varphi_1 : \mathcal{S} \to \{0, 1\}$ by

$$\varphi_1(y^l) := \begin{cases} 0 & \text{if } y^l \in \mathcal{D}_{l,x}\\ 1 & \text{if } y^l \in \mathcal{U}_{l,x}.\end{cases} \tag{22}$$

Set $V := \varphi_1(\mathcal{S})$. Furthermore, define the map $\varphi_2 : \mathcal{S} \to \mathcal{Y}^2$ by $\varphi_2(y^l) = (y_{l-1}, y_l)$. Set $W := \varphi_2(\mathcal{S})$. For each $(a', a) \in \mathcal{Y} \times (\mathcal{Y} - \{0\})$, consider the set of integers l that satisfy $y^{l+1} \in \mathcal{D}_{l+1,x}$ and $(y_l, y_{l+1}) = (a', a)$. Let $l_{1,a',a}, l_{2,a',a}, \cdots$ be its elements arranged in the increasing order. By definition it is obvious that

$$t - 1 \leq l_{1,a',a} < l_{2,a',a} < \cdots < l_{k,a',a} < l_{k+1,a',a} < \cdots.$$

Similarly, for each $(b', b) \in \mathcal{Y} \times (\mathcal{Y} - \{M - 1\})$, consider the set of integers l satisfying $y^{l+1} \in \mathcal{U}_{l+1,x}$ and $(y_l, y_{l+1}) = (b', b)$. Let $\tilde{l}_{1,b',b}, \tilde{l}_{2,b',b}, \cdots$ be its elements arranged in the increasing order. By definition it is obvious that

$$t \leq \tilde{l}_{1,b',b} < \tilde{l}_{2,b',b} < \cdots < \tilde{l}_{k,b',b} < \tilde{l}_{k+1,b',b} < \cdots.$$

Let

$$p_{S|VWX}(y^{l_{k,a',a}+1}|0, a', a, x), k = 1, 2, \cdots,$$

denote conditional probabilities of $S = y^{l_{k,a',a}+1}$ for given $V = 0, W = (a', a)$, and $X = x$. Let $p_{S|VWX}(\cdot|0, a', a, x)$ denote the probability distribution which consists of those probabilities. Similarly, let

$$p_{S|VWX}(y^{\tilde{l}_{k,b',b}+1}|1, b', b, x), k = 1, 2, \cdots,$$

denote conditional probabilities of $S = y^{\tilde{l}_{k,b',b}+1}$ for given $V = 1, W = (b', b)$, and $X = x$. Let $p_{S|VWX}(\cdot|1, b', b, x)$ denote the probability distribution which consists of those probabilities. In the remaining part of this subsection we compute the above two probability distributions, which will be useful for later arguments on the performance analysis of the interval algorithm. By the expression of $p_X(x)$ using the coin random sequences we obtain

$$\begin{aligned}
&\Pr\left\{S = y^{l_{k,a',a}+1}, V = 0, W = (a', a), X = x\right\}\\
&= \Pr\left\{Y^{t-1} = a^{t-1}, Y_t^{l_{k,a',a+1}} = a_t^{l_{k,a',a}} a\right\}\\
&= p_Y(a_t^{l_{k,a',a}} a|a^{t-1})p_Y(a^{t-1})\\
&\overset{(a)}{=} p_Y(a_t^{l_{k,a',a}} a|a_{t-1})p_Y(a^{t-1}),
\end{aligned} \tag{23}$$

where if $l_{1,a',a} = t - 1$, we define $a_t^{l_{1,a',a}} = \lambda$. Step (a) follows from the Markov property of coin random sequences. Similarly, we obtain

$$\begin{aligned}
&\Pr\left\{S = y^{\tilde{l}_{k,b',b}+1}, V = 1, W = (b', b), X = x\right\}\\
&= p_Y(b_t^{\tilde{l}_{k,b',b}} b|a_{t-1})p_Y(a^{t-1}).
\end{aligned} \tag{24}$$

Set

$$\eta_0(a', a, x|a_{t-1}) := \sum_{k\geq 1} p_Y(a_t^{l_{k,a',a}} a|a_{t-1}), \tag{25}$$

$$\eta_1(b', b, x|a_{t-1}) := \sum_{k\geq 1} p_Y(b_t^{\tilde{l}_{k,b',b}} b|a_{t-1}). \tag{26}$$

From (23) and (25), we have

$$\begin{aligned}
&\Pr\{V = 0, W = (a', a), X = x\} = \sum_{k\geq 1} 1 \\
&\times \Pr\left\{S = y^{l_{k,a',a}+1}, V = 0, W = (a', a), X = x\right\} \\
&= \sum_{k\geq 1} p_Y(a_t^{l_{k,a',a}} a|a_{t-1}) p_Y(a^{t-1}) \\
&= \eta_0(a', a, x|a_{t-1}) p_Y(a^{t-1}).
\end{aligned} \tag{27}$$

Similarly, from (24), and (26), we have

$$\begin{aligned}
&\Pr\{V = 1, W = (b', b), X = x\} \\
&= \eta_1(b', b, x|a_{t-1}) p_Y(a^{t-1}).
\end{aligned} \tag{28}$$

From (23) and (24), we have

$$\begin{aligned}
&p_{S|VWX}(y^{l_{k,a',a}+1}|0, a', a, x) \\
&= \Pr\left\{S = y^{l_{k,a',a}+1} \middle| V = 0, W = (a', a), X = x\right\} \\
&= \frac{p_Y(a_t^{l_{k,a',a}} a|a_{t-1})}{\eta_0(a', a, x|a_{t-1})}.
\end{aligned} \tag{29}$$

Similarly, from (24) and (28), we have

$$\begin{aligned}
&p_{S|VWX}(y^{\tilde{l}_{k,b',b}+1}|1, b', b, x) \\
&= \frac{p_Y(b_t^{\tilde{l}_{k,b',b}} b|a_{t-1})}{\eta_1(b', b, x|a_{t-1})}.
\end{aligned} \tag{30}$$

Define two probability distributions on positive integers by

$$\begin{aligned}
&p_Y^{(0)}(\cdot|a', a, x, a_{t-1}) \\
&:= \left(p_Y(a_t^{l_{k,a',a}} a|a_{t-1})/\eta_0(a', a, x|a_{t-1})\right)_{k=1,2,\cdots}, \\
&p_Y^{(1)}(\cdot|b', b, x, a_{t-1}) \\
&:= \left(p_Y(b_t^{\tilde{l}_{k,b',b}} b|a_{t-1})/\eta_1(b', b, x|a_{t-1})\right)_{k=1,2,\cdots}.
\end{aligned}$$

Then we have

$$p_{S|VWX}(\cdot|0, a', a, x) = p_Y^{(0)}(\cdot|a', a, x, a_{t-1}), \tag{31}$$

$$p_{S|VWX}(\cdot|1, b', b, x) = p_Y^{(1)}(\cdot|b', b, x, a_{t-1}). \tag{32}$$

4.2 Performance Evaluation of the Interval Algorithm

In this subsection we state our main result on the performance analysis of the interval algorithm. In the following arguments, $H(\cdot)$ designates the entropy of a probability distribution or a random variable and $D(\cdot\|\cdot)$ designates the Kullback-Leibler divergence between two probability distributions.

For each $i \in \mathcal{Y}$, let $Y_2(i)$ be a random variable with the distribution $\{P_{ij}\}_{j=0}^{M-1}$. Entropy rate of $\{Y_t\}_{t=1}^{\infty}$ is the following:

$$\begin{aligned}
H(Y_2|Y_1) &= \sum_{i=0}^{M-1} p_Y(i) \sum_{j=0}^{M-1} P_{ij}[-\log P_{ij}] \\
&= \sum_{i=0}^{M-1} p_Y(i) H(Y_2(i)).
\end{aligned}$$

Define

$$H_{\min}(Y_2(\cdot)) := \min_{0\leq i\leq M-1} H(Y_2(i)), \tag{33}$$

$$H_{\max}(Y_2(\cdot)) := \max_{0\leq i\leq M-1} H(Y_2(i)). \tag{34}$$

Then we have

$$H_{\min}(Y_2(\cdot)) \leq H(Y_2|Y_1) \leq H_{\max}(Y_2(\cdot)).$$

Here we have a certain nontrivial class of information sources where the above two bounds $H_{\min}(Y_2(\cdot))$ and $H_{\max}(Y_2(\cdot))$ match. For given $Y_1 = y_1 \in \mathcal{Y}$, we define a probability distribution Q_{y_1} by

$$Q_{y_1} := p_Y(\cdot|y_1) = \{P_{y_1 y_2}\}_{y_2\in\mathcal{Y}}$$

Let $\mathscr{S}(\mathcal{Y})$ denote the representation of the symmetric group of permutations of $\mathcal{Y}$ by the $|\mathcal{Y}| \times |\mathcal{Y}|$ permutation matrix. We consider the following condition.

Condition: We call that the stochastic matrix P satisfies a symmetrical property if for any $y_1, y_1' \in \mathcal{Y}$, there exists $\Pi \in \mathscr{S}(\mathcal{Y})$ such that $Q_{y_1'} = Q_{y_1}\Pi$.

Then we have the following.

Lemma 2: If the stochastic matrix P of the stationary Markov information source $\{Y_t\}_{t=1,2,\cdots}$ satisfies a symmetrical property, we have

$$H_{\min}(Y_2(\cdot)) = H(Y_2|Y_1) = H_{\max}(Y_2(\cdot)).$$

Proof: Let $i_{\min} \in \mathcal{Y}$ be the symbol i such that it attains $H_{\min}(Y_2(\cdot))$ defined by (33). Similarly, let $i_{\max} \in \mathcal{Y}$ be the symbol i such that it attains $H_{\max}(Y_2(\cdot))$ defined by (34). Since P satisfies a symmetrical property, we have that

$$Q_{i_{\min}} = Q_{i_{\max}}\Pi \text{ for some } \Pi \in \mathscr{S}(\mathcal{Y}). \tag{35}$$

Then we have the following chain of equalities:

$$\begin{aligned}
H_{\min}(Y_2(\cdot)) = H(Q_{i_{\min}}) &\overset{(a)}{=} H(Q_{i_{\max}}\Pi) \overset{(b)}{=} H(Q_{i_{\max}}) \\
&= H_{\max}(Y_2(\cdot)).
\end{aligned}$$

Step (a) follows from (35). Step (b) follows from that the entropy is invariant under the permutation on the components of the probability vector $Q_{i_{\max}}$. □

In the following we show three examples of P with a symmetrical property.

Example 5: We consider the case where $M = 3$. Set $\theta_i := P_{0i}, i = 0, 1, 2$. The following three stochastic matrices $P_i, i = 1, 2, 3$ are examples of P having a symmetrical

property.

$$P_1 = \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 \\ \theta_0 & \theta_1 & \theta_2 \\ \theta_0 & \theta_1 & \theta_2 \end{bmatrix}, P_2 = \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 \\ \theta_1 & \theta_2 & \theta_0 \\ \theta_2 & \theta_0 & \theta_1 \end{bmatrix}, P_3 = \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 \\ \theta_0 & \theta_2 & \theta_1 \\ \theta_0 & \theta_1 & \theta_2 \end{bmatrix}.$$

The above three examples have some specific properties. When $P = P_1$, the source becomes the stationary memoryless source specified with $p_Y = (p_Y(0), p_Y(1), p_Y(2)) = (\theta_0, \theta_1, \theta_2)$. P_2 is a doubly stochastic matrix. When we choose $\theta_0 = \theta_1 = 0.25$, $\theta_2 = 0.50$ in P_3, $P = P_3$ coincides with the stochastic matrix in Example 2.

The efficiency of the interval algorithm is measured by the average number of coin tosses necessary to obtain the target random variable. We denote it by $\bar{L}$. According to Han and Hoshi [3], we have the following:

Lemma 3 (Han and Hoshi [3]):

$$\bar{L} H_{\min}(Y_2(\cdot)) \le H(S) \le \bar{L} H_{\max}(Y_2(\cdot)).$$

Specifically, if the stochastic matrix P of the stationary Markov information source $\{Y_t\}_{t=1,2,\cdots}$ satisfies a symmetrical property, we have $\bar{L} H(Y_2|Y_1) = H(S)$.

From this lemma we can see that an evaluation of $\bar{L}$ is reduced to an estimation of upper bound of $H(S)$. On the upper bound of this quantity, we have the following lemma.

Lemma 4:

$$H(S) \le H(X) + \log\{2M(M-1)\} + \zeta, \tag{36}$$

where $\zeta := H(S|VWX)$. For the quantity ζ, we have

$$\begin{aligned} \zeta = \sum_{x=0}^{N-1} p_Y(a^{t(x)-1}) \Bigg\{ & \sum_{a'=0}^{M-1} \sum_{a=0}^{M-2} \eta_0(a', a, x|a_{t-1}) \\ & \times H\left(p_Y^{(0)}(\cdot|a', a, x, a_{t-1})\right) \\ & + \sum_{b'=0}^{M-1} \sum_{b=1}^{M-1} \eta_1(b', b, x|a_{t-1}) \\ & \times H\left(p_Y^{(1)}(\cdot|b', b, x, a_{t-1})\right) \Bigg\}. \end{aligned} \tag{37}$$

Proof: We first prove (36). We have the following:

$$\begin{aligned} & H(S) = H(\varphi(S)\varphi_1(S)\varphi_2(S)S) = H(XVWS) \\ & = H(X) + H(VW|X) + H(S|VWX) \\ & \le H(X) + \log\{2M(M-1)\} + H(S|VWX), \end{aligned}$$

where the last inequality follows from that V is a binary random variable and that W takes values in $\mathcal{Y} \times (\mathcal{Y} - \{M-1\})$ if $V = 1$ and takes values in $\mathcal{Y} \times (\mathcal{Y} - \{0\})$ if $V = 0$. From (27), (28), (31), and (32), we have (37). □

Han and Hoshi [3] used several complicated arguments to derive the upper bound of $H(S|VWX)$. Their result is the following.

Theorem A(Han and Hoshi[3]):

$$\begin{aligned} & \frac{H(X)}{H_{\max}(Y_2(\cdot))} \le \bar{L} \le \frac{H(X)}{H_{\min}(Y_2(\cdot))} + \frac{\log\{2M(M-1)\}}{H_{\min}(Y_2(\cdot))} \\ & + \frac{h(p_{\max})}{(1-p_{\max}) H_{\min}(Y_2(\cdot))}, \end{aligned} \tag{38}$$

where

$$p_{\max} := \max_{(y_1, y_2) \in \mathcal{Y}^2} P_{y_1 y_2}$$

and $h(\cdot)$ is the binary entropy function.

Define the geometrical distribution p^* with parameter $p_{\max}$ by

$$p^* := \left({p_{\max}}^{k-1} (1 - p_{\max}) \right)_{k=1,2,\cdots}$$

Our main result on the performance analysis of the interval algorithm is the following.

Theorem 2:

$$\begin{aligned} & \frac{H(X)}{H_{\max}(Y_2(\cdot))} \le \bar{L} \le \frac{H(X)}{H_{\min}(Y_2(\cdot))} + \frac{\log\{2M(M-1)\}}{H_{\min}(Y_2(\cdot))} \\ & + \frac{h(p_{\max})}{(1-p_{\max}) H_{\min}(Y_2(\cdot))} - \frac{\Delta}{H_{\min}(Y_2(\cdot))}, \end{aligned} \tag{39}$$

where Δ is a nonnegative number defined by

$$\begin{aligned} \Delta = \sum_{x=0}^{N-1} p_Y(a^{t(x)-1}) \Bigg\{ & \sum_{a'=0}^{M-1} \sum_{a=1}^{M-1} \eta_0(a', a, x|a_{t-1}) \\ & \times D\left(p_Y^{(0)}(\cdot|a', a, x, a_{t-1}) \middle\| p^* \right) \\ & + \sum_{b'=0}^{M-1} \sum_{b=0}^{M-2} \eta_1(b', b, x|a_{t-1}) \\ & \times D\left(p_Y^{(1)}(\cdot|b', b, x, a_{t-1}) \middle\| p^* \right) \Bigg\}. \end{aligned} \tag{40}$$

Specifically, if the stochastic matrix P of the stationary Markov information source $\{Y_t\}_{t=1,2,\cdots}$ satisfies a symmetrical property, we have

$$\begin{aligned} & \frac{H(X)}{H(Y_2|Y_1)} \le \bar{L} \le \frac{H(X)}{H(Y_2|Y_1)} + \frac{\log\{2M(M-1)\}}{H(Y_2|Y_1)} \\ & + \frac{h(p_{\max})}{(1-p_{\max}) H(Y_2|Y_1)} - \frac{\Delta}{H(Y_2|Y_1)}. \end{aligned} \tag{41}$$

Proof of Theorem 2 is given in the next section. Let the upper bound of $\bar{L}$ by Han and Hoshi [3] in (38) be denoted by $\bar{L}_{\mathrm{HH}}$. Then our upper bound of $\bar{L}$ in Theorem 2 is

$$\bar{L} \le \bar{L}_{\mathrm{HH}} - \frac{\Delta}{H_{\min}(Y_2(\cdot))}. \tag{42}$$

Note that Δ is nonnegative and may almost always be positive. Hence our upper bound improves $\bar{L}_{\mathrm{HH}}$. The bound (42) is equivalent to $\bar{L}_{\mathrm{HH}} - \bar{L} \ge \Delta / H_{\min}(Y_2(\cdot))$, implying that the quantity $\Delta / H_{\min}(Y_2(\cdot))$ serves as a lower bound on the deviation of $\bar{L}_{\mathrm{HH}}$ from the true value of $\bar{L}$.

Remark: In [7], the author made a mistake in the derivation of the upper bound of $\bar{L}$. Hence the upper bound of $\bar{L}$ given

by (8) in [7] is incorrect. This upper bound should be replaced by that of $\bar{L}$ given by (39) in this paper. The formula of Δ given by (9) in [7] is also not appropriate. This formula should be replated by that of Δ given by (40) in this paper.

5. Proofs of Theorems 1 and 2

In this section we prove Theorems 1 and 2. We first prove Theorem 1. We next present Lemma 5 necessary for the proof of Theorem 2. Using Lemmas 3-5, we show Theorem 2. Proofs of Lemma 5 and Theorem 2 are quite simple and elementary.

Proof of Theorem 1: We first prove the equality (15). Using $U_X(x)$ and $L_X(x)$, the probability $p_X(x)$ of $x \in \mathcal{X}$ for the target random variable X can be expressed in the following manner:

$$\begin{aligned} p_X(x) &= U_X(x) - L_X(x) = U_X(x) + \overline{L_X(x)} - 1 \\ &= 0.b_1b_2\cdots + 0.\bar{a}_1\bar{a}_2\cdots - 1\,. \end{aligned}$$

Based on the above expression, set

$$\begin{aligned} \theta_X^{(n)}(x) := \sum_{k=1}^{n} \Big\{ & p_Y(b^{k-1})c_Y(b_k|b_{k-1}) \\ & + p_Y(a^{k-1})c_Y(\bar{a}_k|a_{k-1}) \Big\} - 1. \end{aligned}$$

The quantity $\theta_X^{(n)}(x)$ can also be written as

$$\theta_X^{(n)}(x) = 0.b_1b_2\cdots b_n + 0.\bar{a}_1\bar{a}_2\cdots\bar{a}_n \; - \; 1\,.$$

It is obvious that $\{\theta_X^{(n)}(x)\}_{n=1}^{\infty}$ is a monotone increasing sequence and satisfies

$$\lim_{n\to+\infty} \theta_X^{(n)}(x) = p_X(x)\,.$$

Since $L_X(x)$ and $U_X(x)$ first differs at the t-th place of their representations , we have $a_i = b_i$ for $i = 1, 2, \cdots, t-1$ and $b_t \geq a_t + 1$. Then we have the following:

$$\begin{aligned} \theta_X^{(t-1)}(x) &= 0.a_1a_2\cdots a_{t-1} + 0.\bar{a}_1\bar{a}_2\cdots\bar{a}_{t-1} - 1 \\ &= \{1 - p_Y(a_1a_2\cdots a_{t-1})\} - 1 = -p_Y(a^{t-1}) < 0, \\ \theta_X^{(t)}(x) &= 0.a_1a_2\cdots a_{t-1}b_t + 0.\bar{a}_1\bar{a}_2\cdots\bar{a}_{t-1}\bar{a}_t - 1 \\ &= 0.a_1a_2\cdots a_{t-1} + 0.\bar{a}_1\bar{a}_2\cdots\bar{a}_{t-1} - 1 \\ &\quad + \; 0.0_{a_1}0_{a_2}\cdots 0_{a_{t-1}}b_t + \; 0.0_{\bar{a}_1}0_{\bar{a}_2}\cdots 0_{\bar{a}_{t-1}}\bar{a}_t \\ &= p_Y(a^{t-1})\{c_Y(b_t|a_{t-1}) + c_Y(\bar{a}_t|a_{t-1})\} - p_Y(a^{t-1}) \\ &= p_Y(a^{t-1})\{c_Y(b_t|a_{t-1}) + c_Y(\bar{a}_t|a_{t-1}) - 1\} \\ &= p_Y(a^{t-1}) \sum_{a_t<a<b_t} p_Y(a|a_{t-1}) \geq 0. \end{aligned}$$

Hence we obtain

$$\begin{aligned} p_X(x) &= \theta_X^{(t)}(x) + 0.0_{\bar{a}_1}0_{\bar{a}_2}\cdots 0_{\bar{a}_{t-1}}0_{\bar{a}_t}\bar{a}_{t+1}\bar{a}_{t+2}\cdots \\ &\qquad + 0.0_{a_1}0_{a_2}\cdots 0_{a_{t-1}}0_{b_t}b_{t+1}b_{t+2}\cdots \\ &= p_Y(a^{t-1})\Bigg[\sum_{a_t<a<b_t} p_Y(a|a_{t-1}) \\ &\qquad + \sum_{k\geq t+1} p_Y(a_t^{k-1}|a_{t-1})c_Y(\bar{a}_k|a_{k-1}) \\ &\qquad + \sum_{k\geq t+1} p_Y(b_t^{k-1}|a_{t-1})c_Y(b_k|b_{k-1}) \Bigg]. \end{aligned}$$

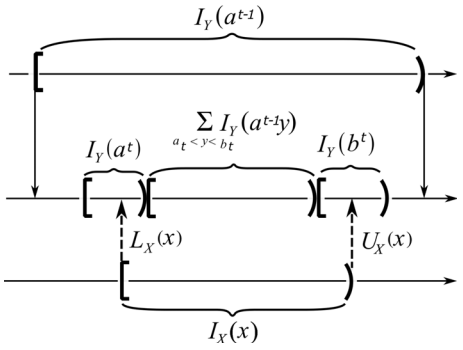


Fig. 4 Relations between $I_Y(a^{t-1})$, $I_Y(a^t)$, $I_Y(b^t)$, and $I_X(x)$.

We next prove the equality (16). We first observe that under the assumption on the p_Y-array representations of $L_X(x)$ and $U_X(x)$, we have the following:

$$I_X(x) \subseteq I_Y(a^{t-1}), \tag{43}$$

$$\left.\begin{aligned} & a^{t-1} = b^{t-1}, a_t < b_t, \\ & L_X(x) \in I_Y(a^t), U_X(x) \in I_Y(b^t). \end{aligned}\right\} \tag{44}$$

The four intervals $I_Y(a^{t-1})$, $I_Y(a^t)$, $I_Y(b^t)$, and $I_X(x)$ satisfying (43) and (44), are shown in Fig. 4. On the form of $I_X(x), x \in \mathcal{X}$ created by the interval algorithm, we have the following chain of equalities:

$$\begin{aligned} I_X(x) &\overset{(a)}{=} I_X(x) \cap I_Y(a^{t-1}) = \sum_{y\in\mathcal{Y}} I_X(x) \cap I_Y(a^{t-1}y) \\ &\overset{(b)}{=} \sum_{a_t<y<b_t} I_Y(a^{t-1}y) \\ &\quad + [L_X(x), U(a^t)) + [L(b^t), U_X(x)) \\ &\overset{(c)}{=} \sum_{a_t<y<b_t} I_Y(a^{t-1}y) \\ &\quad + \sum_{k\geq t+1} \left\{ \sum_{y>a_k} I_Y(a^{k-1}y) + \sum_{y<b_k} I_Y(b^{k-1}y) \right\}. \end{aligned} \tag{45}$$

Step (a) follows from (43). Step (b) follows from (44). Step (c) follows from Lemma 1. □

Lemma 5: For $\eta_0(a', a, x|a_{t-1})$ and $\eta_1(b', b, x|a_{t-1})$ previously defined, set

$$\xi_0 := \sum_{k\geq 1} k p_Y(a_t^{l_{k,a',a}}a|a_{t-1}),$$

$$\xi_1 := \sum_{k\geq 1} k p_Y(b_t^{\tilde{l}_{k,b',b}}b|a_{t-1}).$$

Then, we have

$$\xi_0 \leq \frac{\eta_0(a', a, x|a_{t-1})}{1 - p_{\max}}, \xi_1 \leq \frac{\eta_1(b', b, x|a_{t-1})}{1 - p_{\max}}.$$

Proof of Lemma 5: It suffices to prove the first inequality. Multiplying $p_{\max}$ to both sides of the equation of definition of ξ_0, we have

$$\xi_0 = \sum_{k\geq 1} k p_Y(a_t^{l_{k,a',a}} a|a_{t-1}), \tag{46}$$

$$\begin{aligned} p_{\max}\xi_0 &= \sum_{k\geq 1} p_{\max} k p_Y(a_t^{l_{k,a',a}} a|a_{t-1}) \\ &\overset{(a)}{=} \sum_{k\geq 1} k p_{\max} p_Y(a_t^{l_{k,a',a}}|a_{t-1}) p_Y(a|a') \\ &\overset{(b)}{\geq} \sum_{k\geq 1} k p_Y(a_t^{l_{k+1,a',a}}|a_{t-1}) p_Y(a|a') \\ &\overset{(c)}{=} \sum_{k\geq 1} k p_Y(a_t^{l_{k+1,a',a}} a|a_{t-1}) \\ &= \sum_{k\geq 2} (k-1) p_Y(a_t^{l_{k,a',a}} a|a_{t-1}). \end{aligned} \tag{47}$$

Steps (a) and (c) follow from the definition of $l_{i,a',a}, i = 1, 2, \cdots$, and the Markov property of the coin random sequences. Step (b) follows from

$$\begin{aligned} &p_{\max} p_Y(a_t^{l_{k,a',a}}|a_{t-1}) \\ &\geq p_Y(a_{l_{k,a',a}+1}^{l_{k+1,a',a}}|a_{l_{k,a',a}}) p_Y(a_t^{l_{k,a',a}}|a_{t-1}) \\ &= p_Y(a_t^{l_{k+1,a',a}}|a_{t-1}) \end{aligned}$$

for $k \geq 1$. Reducing both sides of (47) from (46), we obtain

$$(1-p_{\max})\xi_0 \leq \sum_{k\geq 1} p_Y(a_t^{l_{k,a',a}} a|a_{t-1}) = \eta_0(a', a, x|a_{t-1}),$$

completing the proof. □

Proof of Theorem 2: We first observe that

$$\begin{aligned} &H\left(p_Y^{(0)}(\cdot|a', a, x, a_{t-1})\right) \\ &= -\sum_{k\geq 1} \frac{p_Y(a_t^{l_{k,a',a}} a|a_{t-1})}{\eta_0(a', a, x|a_{t-1})} \log \frac{p_Y(a_t^{l_{k,a',a}} a|a_{t-1})}{\eta_0(a', a, x|a_{t-1})} \\ &= -\sum_{k\geq 1} \frac{p_Y(a_t^{l_{k,a',a}} a|a_{t-1})}{\eta_0(a', a, x|a_{t-1})} \\ &\quad \times \log \frac{p_Y(a_t^{l_{k,a',a}} a|a_{t-1})}{{p_{\max}}^{k-1}(1-p_{\max})\eta_0(a', a, x|a_{t-1})} \\ &\quad - \frac{1}{\eta_0(a', a, x|a_{t-1})} \sum_{k\geq 1} p_Y(a_t^{l_{k,a',a}} a|a_{t-1}) \\ &\quad \times \log\left\{{p_{\max}}^k \left(\frac{1-p_{\max}}{p_{\max}}\right)\right\} \\ &= -D\left(p_Y^{(0)}(\cdot|a', a, x, a_{t-1})\middle\| p^*\right) - \log\left(\frac{1-p_{\max}}{p_{\max}}\right) \\ &\quad - \frac{\log p_{\max}}{\eta_0(a', a, x|a_{t-1})} \sum_{k\geq 1} k p_Y(a_t^{l_{k,a',a}} a|a_{t-1}). \end{aligned}$$

Applying Lemma 5 to the last term of the above inequality, we have

$$\begin{aligned} &H\left(p_Y^{(0)}(\cdot|a', a, x, a_{t-1})\right) \\ &\leq -D\left(p_Y^{(0)}(\cdot|a', a, x, a_{t-1})\middle\| p^*\right) \\ &\quad - \log\left(\frac{1-p_{\max}}{p_{\max}}\right) - \frac{\log p_{\max}}{1-p_{\max}} \\ &= -D\left(p_Y^{(0)}(\cdot|a', a, x, a_{t-1})\middle\| p^*\right) + \frac{h(p_{\max})}{1-p_{\max}}. \end{aligned} \tag{48}$$

Similarly, we obtain

$$\begin{aligned} &H\left(p_Y^{(1)}(\cdot|b', b, x, a_{t-1})\right) \\ &\leq -D\left(p_Y^{(1)}(\cdot|b', b, x, a_{t-1})\middle\| p^*\right) + \frac{h(p_{\max})}{1-p_{\max}}. \end{aligned} \tag{49}$$

Combining Lemma 4, (48), and (49), we obtain the desired bound. □

6. Conclusion

We have given an explicit expression of interval algorithm based on number systems. We have evaluated the expected number of the Markov coin tosses in the interval algorithm and have shown that it can be characterized explicitly with the established expression of the interval algorithm.

The quantity $\Delta/H_{\min}(Y_2(\cdot))$ indicates a lower bound of the deviation of the upper bound of $\bar{L}$ obtained by Han and Hoshi [3] from the true value of $\bar{L}$. This quantity is characterized with the p_Y-ary representation and the p_Y-ary co-representation of the endpoints of the intervals corresponding to the target random numbers.

References

[1] P. Elias, "The efficient construction of an unbiased random sequences," Ann. Math. Statist., vol.43, no.3, pp.865–870, June 1976.
[2] D. Knuth and A. Yao, "The complexity of nonuniform random number generation," Algorithm and Complexity, New Directions and Results, pp.357–428, 1976.
[3] T.S. Han and M. Hoshi, "Interval algorithm for random number generation," IEEE Trans. Inform. Theory, vol.43, no.2, pp.599–611, March 1997.
[4] S. Watanabe and T.S. Han, "Interval algorithm for random number generation: Information spectrum approach," IEEE Trans. Inf. Theory, vol.66, no.3, pp.1691–1701, March 2020.
[5] T.S. Han, Information-Spectrum Methods in Information Theory, Springer, 2003.
[6] Y. Oohama, "Performance analysis of the interval algorithm for random number generation based on number systems," IEEE Trans. Inf. Theory, vol.57, no.3, pp.1177–1185, March 2011.
[7] Y. Oohama, "Performance analysis of the interval algorithm for random number generation in the case of Markov coin tossings," 2016 International Symposium on Nonlinear Theory and Its Applications (NOLTA2016), pp.245–248, 2016.
[8] F. Kanaya, "A chaos model of finite-order Markov sources and arithmetic coding," Proc. IEEE International Symposium on Information Theory, p.247, Ulm, Germany, July 1997.
[9] Y. Oohama, M. Suemitsu, and T. Kohda, "Construction of a piecewise linear one-dimensional map generating an arbitrary prescribed tree source," IEICE Trans. Fundamentals, vol.E86-A, no.9, pp.2251–2255, Sept. 2003.

Appendix: Proof of Property 1

In this appendix we prove Property 1. We first prepare a lemma necessary for the proof.

Lemma 6: For any $1 \leq k \leq n$, any $a^k \in \mathcal{Y}^k$, and any $(y, y') \in \mathcal{Y}^2$ satisfying $y < a_k < y'$, we have that if $p_Y(a^n)$, $p_Y(a^{k-1}y)$, and $p_Y(a^{k-1}y')$ are positive, then

$$U_Y(a^{k-1}y) \leq L_Y(a^k) \leq L_Y(a^n) < U_Y(a^n), \tag{A·1}$$

$$L_Y(a^n) < U_Y(a^n) \leq U_Y(a^k) \leq L_Y(a^{k-1}y'). \tag{A·2}$$

Proof: Under the condition that $p_Y(a^{k-1}y)$, $p_Y(a^{k-1}y')$, $y < a_k < y'$, and $p_Y(a^n)$ are positive, we have that $I_Y(a^{k-1}y)$, $I_Y(a^{k-1}y')$, and $I_Y(a^k)$ are not void and disjoint. We observe that $I_Y(a^{k-1}y)$ is in the left-hand side of $I_Y(a^k)$ and that $I_Y(a^{k-1}y')$ is in the right-hand side of $I_Y(a^k)$. We further observe that $I_Y(a^n) \subseteq I_Y(a^k)$. Those relations between the four intervals are shown in Fig. A·1. From the above relations between $I_Y(a^{k-1}y)$, $I_Y(a^{k-1}y')$, $I_Y(a^k)$, and $I_Y(a^n)$, for $1 \leq k \leq n$, $y < a_k < y'$, we have (A·1) and (A·2) in the lemma. □

Proof of Property 1: We first observe the following:

$$I_Y(a^{k-1}) - I_Y(a^k) = \sum_{\substack{y<a_k,\\ y>a_k}} I_Y(a^{k-1}y). \tag{A·3}$$

Taking union of both sides of (A·3) with respect to $k = m+1, \cdots, n$, we have the following:

$$I_Y(a^m) - I_Y(a^n) = J + J', \tag{A·4}$$

where we set

$$J := \sum_{k=m+1}^{n} \sum_{y<a_k} I_Y(a^{k-1}y),$$

$$J' := \sum_{k=m+1}^{n} \sum_{y>a_k} I_Y(a^{k-1}y).$$

To obtain the relation (A·4), we have used two facts. One is that

$$I_Y(a^m) \supseteq I_Y(a^{m+1}) \supseteq \cdots \supseteq I_Y(a^n).$$

The other is that the $(n-m)(M-1)$ intervals

$$I_Y(a^{k-1}y),\ (k, y) \in \bigcup_{k=m+1}^{n} \left(\{k\} \times \{y\}_{y \neq a_k}\right)$$

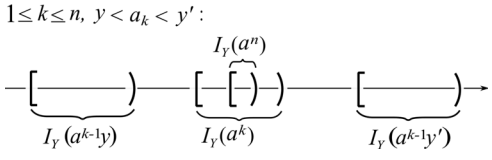


Fig. A·1 Relations between $I_Y(a^{k-1}y)$, $I_Y(a^{k-1}y')$, $I_Y(a^k)$, and $I_Y(a^n)$ for $1 \leq k \leq n$, $y < a_k < y'$.

are disjoint. On the other hand, we have

$$I_Y(a^m) - I_Y(a^n) = [L_Y(a^m), L_Y(a^n)) + [U_Y(a^n), U_Y(a^m)). \tag{A·5}$$

From (A·4) and (A·5), we have

$$[L_Y(a^m), L_Y(a^n)) + [U_Y(a^n), U_Y(a^m)) = J + J'. \tag{A·6}$$

From (A·6), we have

$$J \subseteq [L_Y(a^m), L_Y(a^n)) + [U_Y(a^n), U_Y(a^m)). \tag{A·7}$$

By (A·1) in Lemma 6, we have

$$U_Y(a^{k-1}y) < U_Y(a^n) \text{ for any } m+1 \leq k \leq n, y < a_k,$$

implying that

$$J \cap [U_Y(a^n), U_Y(a^m)) = \emptyset. \tag{A·8}$$

From (A·7) and (A·8), we have that $J \subseteq [L_Y(a^m), L_Y(a^n))$. Similary, from (A·6), we have

$$[L_Y(a^m), L_Y(a^n)) \subseteq J + J'. \tag{A·9}$$

By (A·2) in Lemma 6, we have

$$L_Y(a^n) < L_Y(a^{k-1}y) \text{ for any } m+1 \leq k \leq n, y > a_k,$$

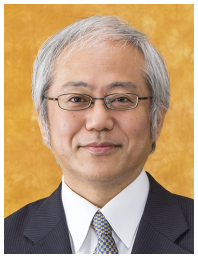
implying that

$$[L_Y(a^m), L_Y(a^n)) \cap J' = \emptyset. \tag{A·10}$$

From (A·9) and (A·10), we have that $[L_Y(a^m), L_Y(a^n)) \subseteq J$. Hence we have

$$[L_Y(a^m), L_Y(a^n)) = J, [U_Y(a^n), U_Y(a^m)) = J',$$

completing the proof. □



Yasutada Oohama was born in Tokyo, Japan, in 1963. He received the B.Eng., M.Eng., and D.Eng. degrees in mathematical engineering in 1987, 1989, and 1992, respectively, from University of Tokyo, Tokyo, Japan. From April 1992 to September 2006, he was with Kyushu University, Fukuoka, Japan. From October 2006 to April 2011, he was with University of Tokushima, Tokushima, Japan. Since May 2011, he has been with University of Electro-Communications, Tokyo, Japan. He is currently a professor at the Department of Communication Engineering and Informatics. From September 1996 to March 1997 he was a visiting scholar at the Information Systems Laboratory, Stanford University, Stanford CA. His current research interest includes basic problems in information theory and related areas.