

Large Deviations Performance of Knuth-Yao algorithm for Random Number Generation

Akisato KIMURA *
akisato@ss.titech.ac.jp

Tomohiko UYEMATSU*
uematsu@ss.titech.ac.jp

April 12, 1999

No. AK-TR-1999-02

Abstract

We investigate large deviations performance of the algorithm for random number generation proposed by Knuth and Yao.

First, we show that the number of input fair random bits per the length of output sequence approaches to the of output source almost surely, and the large deviations performance of this algorithm is equal to the one of the interval algorithm proposed by Han and Hoshi.

Next, we consider to obtain the fixed length of output sequence from the input fair random bits with fixed length. We show that the approximation error measured by the variational distance vanishes exponentially at the speed equal to the interval algorithm as the length of output sequence tends to infinity, if the number of input fair bits per output sample is above the entropy of source. Contrarily, the approximation error measured by the variational distance approaches to two exponentially at the speed equal to the interval algorithm, if the number of input fair bits per output sample is below the entropy.

*Dept. of Electrical and Electronic Eng., Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan

I. Introduction

Random number generation is a problem of simulating some prescribed target distribution by using a given source. This problem has been investigated in computer science, and has a close relation to information theory [1, 2, 3]. Some practical algorithms for random number generation have been proposed so far, i.e. [1, 3, 4, 5]. In this paper, we consider the algorithm proposed by Knuth and Yao [1].

Knuth-Yao algorithm can be regarded as a starting point of practical algorithms for random number generation, and some modified algorithm have been proposed. But the asymptotic property has not yet been made clear. On the other hand, performance of the interval algorithm proposed by Han and Hoshi [3] has already been investigated in [3, 6, 7, 8]. Especially, Uyematsu and Kanaya [6] have investigated large deviations performance of the interval algorithm, where the distribution of input source is uniform. We shall investigate large deviations performance of Knuth-Yao algorithm on the same condition as Uyematsu and Kanaya.

First, we show that the number of input fair random bits per the length of output sequence approaches to the of output source almost surely, and the large deviations performance of this algorithm is equal to the one of the interval algorithm.

Next, we consider to obtain the fixed length of output sequence from the input fair random bits with fixed length. We show that the approximation error measured by the variational distance vanishes exponentially at the speed equal to the interval algorithm as the length of output sequence tends to infinity, if the number of input fair bits per output sample is above the entropy of source. Contrarily, the approximation error measured by the variational distance approaches to two exponentially at the speed equal to the interval algorithm, if the number of input fair bits per output sample is below the entropy.

II. Basic Definitions

(a) Discrete Memoryless Source

Let \mathcal{X} be a finite set. We denote by $\mathcal{M}(\mathcal{X})$ the set of all probability distributions on \mathcal{X} . Throughout this paper, by a source X with alphabet \mathcal{X} , we mean a discrete memoryless source (DMS) of distribution $P_X \in \mathcal{M}(\mathcal{X})$. To denote a source, we will use both notations X and P_X interchangeably.

For random variable X which has a distribution P_X , we shall denote this

entropy as $H(P_X)$ and $H(X)$, interchangeably.

$$H(P_X) \triangleq - \sum_{x \in \mathcal{X}} P_X(x) \log P_X(x).$$

Further, for arbitrary distributions $P, Q \in \mathcal{M}(\mathcal{X})$, we denote by $D(P \parallel Q)$ the *information divergence*

$$D(P \parallel Q) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}.$$

Lastly, we denote by $d(P, Q)$ the *variational distance* or l_1 *distance* between two distributions P and Q on \mathcal{X}

$$d(P, Q) \triangleq \sum_{x \in \mathcal{X}} |P(x) - Q(x)|.$$

From now on, all logarithms and exponentials are to the base two.

(b) Type of Sequence

The *type* of a sequence $\mathbf{x} \in \mathcal{X}^n$ is defined as a distribution $P_{\mathbf{x}} \in \mathcal{M}(\mathcal{X})$, where $P_{\mathbf{x}}(a)$ is given by

$$P_{\mathbf{x}}(a) = \frac{1}{n} \cdot (\text{number of occurrences of } a \in \mathcal{X} \text{ in } \mathbf{x}). \quad (1)$$

We shall write \mathcal{P}_n or \mathcal{P} for the set of types of sequences in \mathcal{X}^n . We denote by T_P^n or T_P the set of sequences of type P in \mathcal{X}^n . On the contrary for a distribution $P \in \mathcal{M}(\mathcal{X})$, if $T_P \neq \emptyset$ then we denote by P the type of sequences in \mathcal{X}^n .

We introduce some well-known facts, cf. Csiszár-Körner [9]: For the set of types in \mathcal{X}^n , we have

$$|\mathcal{P}_n| \leq (n+1)^{|\mathcal{X}|} \quad (2)$$

where $|\cdot|$ denotes the cardinality of the set. For the set of sequences of type P in \mathcal{X}^n ,

$$(n+1)^{-|\mathcal{X}|} \exp(nH(P)) \leq |T_P| \leq \exp(nH(P)) \quad (3)$$

If $\mathbf{x} \in T_P$, we then have

$$Q^n(\mathbf{x}) = \exp[-n\{D(P \parallel Q) + H(P)\}]. \quad (4)$$

From (3)(4),

$$(n+1)^{-|\mathcal{X}|} \exp(-nD(P \parallel Q)) \leq Q^n(T_P) \leq \exp(-nD(P \parallel Q)) \quad (5)$$

(c) Resolvability

In this paper, we especially investigate the problem to generate a general source $\mathbf{Y} = \{Y^n\}_{n=1}^\infty$ using a uniform random number with as small size as possible. This problem is called *resolvability problem* [10]. Here, we shall introduce basic definitions for resolvability problem.

Definition 1: For an arbitrary source $\mathbf{Y} = \{Y^n\}_{n=1}^\infty$, rate R is *achievable resolvability rate* if and only if there exists a map $\varphi_n : \mathcal{U}_{M_n} \rightarrow \mathcal{Y}^n$ such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log M_n \leq R$$
$$\lim_{n \rightarrow \infty} d(Y^n, \varphi_n(U_{M_n})) = 0,$$

where $\mathcal{U}_{M_n} \triangleq \{1, 2, \dots, M_n\}$ and U_{M_n} is a uniform distribution on \mathcal{U}_{M_n} .

Definition 2 (inf achievable resolvability rate):

$$S(\mathbf{Y}) = \inf\{R \mid R \text{ is achievable resolvability rate}\}$$

As for the characterization of resolvability rate, Han and Verdú [11] proved the following fundamental theorem.

Theorem 1: For any stationary source \mathbf{Y} ,

$$S(\mathbf{Y}) = H(\mathbf{Y}) \tag{6}$$

where $H(\mathbf{Y})$ is the entropy rate of \mathbf{Y} .

III. Algorithms for Random Number Generation

In this chapter, we describe algorithms for random number generation, the interval algorithm proposed by Han and Hoshi [3], and the Knuth-Yao algorithm proposed by Knuth and Yao [1].

(a) Interval Algorithm

In this section, we describe the interval algorithm proposed by Han and Hoshi [3]. Let us consider to produce an i.i.d. random sequence $Y^n = (Y_1, Y_2, \dots, Y_n)$. Each random variable Y_i ($i = 1, 2, \dots, n$) is subject to a generic distribution $\mathbf{q} = (q_1, q_2, \dots, q_N)$. We generate this sequence by using an i.i.d. random sequence X_1, X_2, \dots , with a generic distribution $\mathbf{p} = (p_1, p_2, \dots, p_M)$. The algorithm can be described as follows.

Interval Algorithm for Generating Random Process

- 1a) Partition an unit interval $[0, 1)$ into N disjoint subinterval $J(1), J(2), \dots, J(N)$ such that

$$J(i) = [Q_{i-1}, Q_i) \quad i = 1, 2, \dots, N$$

$$Q_i = \sum_{k=1}^i q_k \quad i = 1, 2, \dots, N; \quad Q_0 = 0.$$

- 1b) Set

$$P_j = \sum_{k=1}^j p_k \quad j = 1, 2, \dots, M; \quad P_0 = 0.$$

- 2) Set $s = t = \lambda$ (null string), $\alpha_s = \gamma_t = 0$, $\beta_s = \delta_t = 1$, $I(s) = [\alpha_s, \beta_s)$, $J(t) = [\gamma_t, \delta_t)$, and $m = 1$.
- 3) Obtain an output symbol from the source X to have a value $a \in \{1, 2, \dots, M\}$, and generate the subinterval of $I(s)$

$$I(sa) = [\alpha_{sa}, \beta_{sa})$$

where

$$\alpha_{sa} = \alpha_s + (\beta_s - \alpha_s)P_{a-1}$$

$$\beta_{sa} = \alpha_s + (\beta_s - \alpha_s)P_a.$$

- 4a) If $I(sa)$ is entirely contained in some $J(ti)$ ($i = 1, 2, \dots, N$), then output i as the value of the m th random number Y_m and set $t = ti$. Otherwise, go to 5).
- 4b) If $m = n$ then stop the algorithm. Otherwise, partition the interval $J(t) \equiv [\gamma_t, \delta_t)$ into N disjoint subinterval $J(t1), J(t2), \dots, J(tN)$ such that

$$J(tj) = [\gamma_{tj}, \delta_{tj}) \quad j = 1, 2, \dots, N$$

where

$$\gamma_{tj} = \gamma_t + (\delta_t - \gamma_t)Q_{j-1}$$

$$\delta_{tj} = \gamma_t + (\delta_t - \gamma_t)Q_j$$

and set $m = m + 1$ and go to 4a).

5) Set $s = sa$ and go to 3).

Here, we denote by $T_n(\mathbf{x}, \mathbf{y})$ the length of input sequence \mathbf{x} necessary to generate the sequence $\mathbf{y} \in \mathcal{Y}^n$. Han and Hoshi [3] have showed the following bound

$$\frac{H(Y^n)}{H(X)} \leq E[T_n(X, Y^n)] \leq \frac{H(Y^n)}{H(X)} + \frac{\log(2(M-1))}{H(X)} + \frac{h(p_{\max})}{(1-p_{\max})H(X)}$$

where $p_{\max} = \max_{1 \leq j \leq M} p_j$ and $h(\cdot)$ is the binary entropy. In the special case of using fair random bits for the input sequence, we reduce this bound to the following.

$$H(Y^n) \leq E[T_n(Y^n)] \leq H(Y^n) + 3 \quad (7)$$

(b)Knuth-Yao Algorithm

In this section, we describe the Knuth-Yao algorithm proposed by Knuth and Yao [1]. Let us consider to produce an i.i.d. random sequence $Y^n = (Y_1, Y_2, \dots, Y_n)$. Each random variable Y_i ($i = 1, 2, \dots, n$) is subject to a generic distribution $\mathbf{q} = (q_1, q_2, \dots, q_N)$ on $\mathcal{Y} = \{1, 2, \dots, N\}$. We generate this random sequence by using a sequence of fair bits.

We can describe this algorithm by mapping sequences of fair bits X_1, X_2, \dots to possible outcomes Y^n by a binary tree which is called a *generating tree*. Leaves of the tree are marked by output symbols Y^n and the path to the leaf is given by a sequence of fair bits.

The generating tree must satisfy the following properties.

1. The tree should be complete, i.e. every node is either a leaf or has two descendants in the tree.
2. The probability of the leaf at depth k is $\exp(-k)$. Many leaves may be labeled with the same output symbol, but the total probability of all these leaves should be equal to the desired probability of the output symbol.
3. The expected number of fair bits required to generate Y is equal to the expected depth of this tree.

The following algorithm is to construct generating tree satisfying these properties.

Knuth-Yao Algorithm for Constructing Generating Tree

- 1a) Prepare a root of the tree, and set $i = 1$.

- 1b) For a sequence $\mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathcal{Y}^n$ from the DMS Y , the binary expansion of the probability $P_Y^n(\mathbf{y})$ be

$$P_Y^n(\mathbf{y}) = \sum_{k \geq 1} \alpha_k^{(\mathbf{y})} \exp(-k)$$

where $\alpha_k^{(\mathbf{y})} = \{0, 1\}$.

- 2) If nodes at depth $i - 1$ are not labeled with any symbols, then attach two branches and leaves to these nodes, and label one of the branches with 0, the other with 1.
- 3) For all $\mathbf{y} \in \mathcal{Y}^n$ satisfying $\alpha_i^{(\mathbf{y})} = 1$, label one of the attached leaves in step 2) with \mathbf{y} .
- 4) If all leaves are labeled, then stop the algorithm. Otherwise, set $i = i + 1$ and go to 2).

Here, we denote by $T_n(\mathbf{y})$ the length of input sequence necessary to generate the sequence $\mathbf{y} \in \mathcal{Y}^n$. Knuth and Yao [1] have showed the following bound.

$$H(Y^n) \leq E[T_n(Y^n)] \leq H(Y^n) + 2. \quad (8)$$

This bound is tighter just 1 than that of the interval algorithm (see (7)). From this, we can see that the Knuth-Yao algorithm may be better than the interval algorithm in sight of large deviations performance.

IV. Large Deviations Analysis of Length of Input Sequence

In this chapter, we investigate large deviations performance for the length of input sequence in the Knuth-Yao algorithm, and compare with the performance of the interval algorithm.

Let us consider to produce an i.i.d. random sequence $Y^n = (Y_1, Y_2, \dots, Y_n)$. Each random variable Y_i ($i = 1, 2, \dots, n$) is subject to a generic distribution P_Y on \mathcal{Y} . We generate this random sequence by using a sequence of fair bits. We denote by $T_n(\mathbf{y})$ the length of input sequence to generate $\mathbf{y} \in \mathcal{Y}^n$.

Here, we define the following function:

$$E_r(R, P_Y) = \min_{Q \in \mathcal{M}(\mathcal{Y})} \{D(Q \| P_Y) + |R - H(Q) - D(Q \| P_Y)|^+\} \quad (9)$$

$$E_{sp}(R, P_Y) = \min_{\substack{Q \in \mathcal{M}(\mathcal{Y}): \\ D(Q \| P_Y) + H(Q) \geq R}} D(Q \| P_Y) \quad (10)$$

$$F(R, P_Y) = \min_{\substack{Q \in \mathcal{M}(\mathcal{Y}): \\ D(Q \| P_Y) + H(Q) \leq R}} D(Q \| P_Y) \quad (11)$$

$$G(R, P_Y) = \min_{\substack{Q \in \mathcal{M}(\mathcal{Y}): \\ H(Q) \leq R}} D(Q \| P_Y) \quad (12)$$

where $|x|^+ = \max\{0, x\}$.

(a) Interval Algorithm

Considering to use the interval algorithm, this problem is equivalent to channel simulation (investigated by Uyematsu and Kanaya [6]) in the case that the input of channel is unique symbol. To modify the results of Uyematsu and Kanaya [6], we can easily obtain the following theorems.

Theorem 2:

$$\liminf_{n \rightarrow \infty} \left[-\frac{1}{n} \log \Pr \left\{ \frac{1}{n} T_n(Y^n) \geq R \right\} \right] \geq E_r(R, P_Y) \quad (13)$$

Also, for $R > R_{min} = -\max_{y \in \mathcal{Y}} \log P_Y(y)$

$$\lim_{n \rightarrow \infty} \left[-\frac{1}{n} \log \Pr \left\{ \frac{1}{n} T_n(Y^n) \leq R \right\} \right] = F(R, P_Y) \quad (14)$$

Further, $E_r(R, P_Y) > 0$ if and only if $R > H(Y)$ and $F(R, P_Y) > 0$ if and only if $R_{min} < R < H(Y)$.

(b) Knuth-Yao Algorithm

Considering to use the Knuth-Yao algorithm, we investigate large deviations performance in a similar manner as the interval algorithm. Then, we obtain the following theorems.

Theorem 3:

$$\liminf_{n \rightarrow \infty} \left[-\frac{1}{n} \log \Pr \left\{ \frac{1}{n} T_n(Y^n) \geq R \right\} \right] \geq E_r(R, P_Y) \quad (15)$$

where $E_r(R, P_Y)$ is given by (9).

Proof: If n is sufficiently large, without loss of generality let nR be integer. First, we modify the generating tree of the Knuth-Yao algorithm such that the depth of all leaves are nR by the following operation.

1. If depth of a leaf with a symbol $\mathbf{y} \in \mathcal{Y}^n$ is smaller than nR then construct a *sub-tree* under this leaf until the depth of all descendants are nR , and label all the descendants with \mathbf{y} .
2. If nodes at depth nR are not leaves, then cut all the branches attached under these nodes, and label these leaves with no symbol.

In this binary tree, The probability of each leaf is $\exp(-nR)$. Therefore, this operation corresponds to labeling $\left\lfloor \frac{P_Y^n(\mathbf{y})}{\exp(-nR)} \right\rfloor$ leaves with $\mathbf{y} \in \mathcal{Y}^n$ which satisfy $P_Y^n(\mathbf{y}) \geq \exp(-nR)$, where $\lfloor x \rfloor$ is the maximum integer which is not over x . The leaves with no symbol correspond to sequences of fair bits of length nR not to stop the algorithm. On the other hand, the leaves with some label correspond to sequences of length nR to stop the algorithm.

From this observation, using (2)-(5) we have

$$\begin{aligned}
& \Pr \left\{ \frac{1}{n} T_n(Y) \geq R \right\} \\
& \leq \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ P_Y^n(\mathbf{y}) \leq \exp(-nR)}} P_Y^n(\mathbf{y}) + \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ P_Y^n(\mathbf{y}) > \exp(-nR)}} \exp(-nR) \\
& \leq \sum_{\substack{Q \in \mathcal{P}_n: \\ D(Q \| P_Y) + H(Q) \geq R}} \exp(-nD(Q \| P_Y)) + \sum_{\substack{Q \in \mathcal{P}_n: \\ D(Q \| P_Y) + H(Q) < R}} \exp\{-n(R - H(Q))\} \\
& = \sum_{Q \in \mathcal{P}_n} \exp[-n\{D(Q \| P_Y) + |R - H(Q) - D(Q \| P_Y)|^+\}] \\
& \leq (n+1)^{|\mathcal{Y}|} \exp\{-nE_r(R, P_Y)\}
\end{aligned}$$

which implies (15). □

Theorem 4: For $R > R_{min} = -\max_{y \in \mathcal{Y}} \log P_Y(y)$

$$\lim_{n \rightarrow \infty} \left[-\frac{1}{n} \log \Pr \left\{ \frac{1}{n} T_n(Y^n) \leq R \right\} \right] = F(R, P_Y) \quad (16)$$

where $F(R, P_Y)$ is given by (11).

Proof: In a similar manner as the proof of Theorem 3, we obtain

$$\begin{aligned}
& \Pr \left\{ \frac{1}{n} T_n(Y) \leq R \right\} \\
&= \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ P_Y^n(\mathbf{y}) \geq \exp(-nR)}} \left[\frac{P_Y^n(\mathbf{y})}{\exp(-nR)} \right] \exp(-nR) \\
&\leq \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ P_Y^n(\mathbf{y}) \geq \exp(-nR)}} P_Y^n(\mathbf{y}) \\
&\leq \sum_{\substack{Q \in \mathcal{P}_n: \\ D(Q \| P_Y) + H(Q) \leq R}} \exp(-nD(Q \| P_Y)) \\
&\leq (n+1)^{|\mathcal{Y}|} \exp\{-nF(R, P_Y)\}
\end{aligned}$$

which implies

$$\liminf_{n \rightarrow \infty} \left[-\frac{1}{n} \log \Pr \left\{ \frac{1}{n} T_n(Y^n) \leq R \right\} \right] \geq F(R, P_Y) \quad (17)$$

for $R > R_{min}$. It should be noted that the minimum of (11) is taken over the non-empty set of Q if $R > R_{min}$.

Also, we can obtain

$$\begin{aligned}
& \Pr \left\{ \frac{1}{n} T_n(Y^n) \leq R \right\} \\
&\geq \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ P_Y^n(\mathbf{y}) \geq \exp(-nR)}} \frac{1}{2} P_Y^n(\mathbf{y}) \\
&\geq \frac{1}{2} (n+1)^{-|\mathcal{Y}|} \sum_{\substack{Q \in \mathcal{P}_n: \\ D(Q \| P_Y) + H(Q) \leq R}} \exp(-nD(Q \| P_Y)) \\
&\geq \frac{1}{2} (n+1)^{-|\mathcal{Y}|} \exp\{-n \min_{\substack{Q \in \mathcal{P}_n: \\ D(Q \| P_Y) + H(Q) \leq R}} D(Q \| P_Y)\},
\end{aligned}$$

which implies

$$\limsup_{n \rightarrow \infty} \left[-\frac{1}{n} \log \Pr \left\{ \frac{1}{n} T_n(Y^n) \leq R \right\} \right] \leq F(R, P_Y) \quad (18)$$

for $R > R_{min}$. From (17)(18), we obtain (16). \square

Theorem 5: For $R < R_{max} = -\min_{y \in \mathcal{Y}} \log P_Y(y)$

$$\limsup_{n \rightarrow \infty} \left[-\frac{1}{n} \log \Pr \left\{ \frac{1}{n} T_n(Y^n) \geq R \right\} \right] \leq E_{sp}(R, P_Y). \quad (19)$$

Further, $E_{sp}(R, P_Y) > 0$ if and only if $H(Y) < R < R_{max}$.

Proof: In a similar manner as the proof of Theorem 3, we obtain

$$\begin{aligned} & \Pr \left\{ \frac{1}{n} T_n(Y^n) \geq R \right\} \\ & \geq \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ P_Y^n(\mathbf{y}) \leq \exp(-nR)}} P_Y^n(\mathbf{y}) \\ & \geq (n+1)^{-|\mathcal{Y}|} \sum_{\substack{Q \in \mathcal{P}_n: \\ D(Q \| P_Y) + H(Q) \geq R}} \exp(-nD(Q \| P_Y)) \\ & \geq (n+1)^{-|\mathcal{Y}|} \exp\{-n \min_{\substack{Q \in \mathcal{P}_n: \\ D(Q \| P_Y) + H(Q) \geq R}} D(Q \| P_Y)\} \end{aligned}$$

which implies (19) for $R < R_{max}$. It should be noted that the minimum of (10) is taken over the non-empty set of Q if $R < R_{max}$.

$E_{sp}(R, P_Y) = 0$ if and only if $Q = P_Y$ and $H(Q) \geq R$, i.e. $R \leq H(Y)$. This implies that $E_{sp}(R, P_Y) > 0$ if and only if $H(Y) < R < R_{max}$. \square

From Theorem 3 and 4, by using the Borel-Cantelli's principle (see e.g. [12]), we immediately obtain the following corollary.

Corollary 1:

$$\lim_{n \rightarrow \infty} \frac{1}{n} T_n(Y^n) = H(Y) \quad \text{a.s.} \quad (20)$$

From Theorem 3 and 4, we can conclude that the Knuth-Yao algorithm has the same large deviations performance for the length of input sequence as the interval algorithm using a sequence of fair bits.

V. Error Exponent for Resolvability Problem

In this chapter, let us consider to produce an i.i.d. random sequence $Y^n = (Y_1, Y_2, \dots, Y_n)$ using a fixed number of fair bits. In this case, we cannot generate this random sequence exactly but approximately.

(a) Interval Algorithm

First, we modify the interval algorithm so that the algorithm outputs a specified sequence $\mathbf{y}_0 \in \mathcal{Y}^n$ whenever the algorithm does not stop with an input fair bits of length nR . The modified algorithm can be described as follows.

Modified Interval Algorithm for Generating Random Process with Fixed Number of Fair Bits

- 1) Partition an unit interval $[0, 1)$ into N disjoint subinterval $J(1), J(2), \dots, J(N)$ such that

$$J(i) = [Q_{i-1}, Q_i) \quad i = 1, 2, \dots, N$$

$$Q_i = \sum_{k=1}^i q_k \quad i = 1, 2, \dots, N; \quad Q_0 = 0.$$

- 2) Set $s = t = \lambda$ (null string), $\alpha_s = \gamma_t = 0$, $\beta_s = \delta_t = 1$, $I(s) = [\alpha_s, \beta_s)$, $J(t) = [\gamma_t, \delta_t)$, $l = 0$, and $m = 1$.

- 3) If $l = nR$ then output \mathbf{y}_0 as the output sequence Y^n , and stop the algorithm. Otherwise obtain an output symbol from the source X to have a value $a \in \{0, 1\}$, and generate the subinterval of $I(s)$

$$I(sa) = [\alpha_{sa}, \beta_{sa})$$

where

$$\alpha_{sa} = \alpha_s + \frac{1}{2}a(\beta_s - \alpha_s)$$

$$\beta_{sa} = \alpha_s + \frac{1}{2}(a+1)(\beta_s - \alpha_s),$$

and set $l = l + 1$.

- 4a) If $I(sa)$ is entirely contained in some $J(ti)$ ($i = 1, 2, \dots, N$), then set $t = ti$. Otherwise, go to 5).
- 4b) If $m = n$ then output t as the output sequence Y^n , stop the algorithm. Otherwise, partition the interval $J(t) \equiv [\gamma_t, \delta_t)$ into N disjoint subinterval $J(t1), J(t2), \dots, J(tN)$ such that

$$J(tj) = [\gamma_{tj}, \delta_{tj}) \quad j = 1, 2, \dots, N$$

where

$$\gamma_{tj} = \gamma_t + (\delta_t - \gamma_t)Q_{j-1}$$

$$\delta_{tj} = \gamma_t + (\delta_t - \gamma_t)Q_j$$

and set $m = m + 1$ and go to 4a).

5) Set $s = sa$ and go to 3).

This problem also has been investigated by Uyematsu and Kanaya [6]. They have measured the approximation error by the variational distance between the desired and approximated distribution. They have showed the following theorems for the error exponent of the interval algorithm.

Theorem 6 [6]: If the modified interval algorithm is used for random number generation of an i.i.d. source with a generic distribution P_Y , then we have

$$\liminf_{n \rightarrow \infty} \left[-\frac{1}{n} \log d(P_Y^n, \tilde{P}_Y^n) \right] \geq E_r(R, P_Y), \quad (21)$$

where \tilde{P}_Y^n denotes the output distribution of the modified interval algorithm, and $E_r(R, P_Y)$ is given by (9).

Theorem 7 [6]: If the modified interval algorithm is used for random number generation of an i.i.d. source with a generic distribution P_Y , then for $R > R_{min}$ we have

$$\lim_{n \rightarrow \infty} \left[-\frac{1}{n} \log \{2 - d(P_Y^n, \tilde{P}_Y^n)\} \right] = F(R, P_Y), \quad (22)$$

where $F(R, P_Y)$ is given by (11).

Theorem 6 implies that if the number of fair bits per the length of output sequence is above the entropy of output source, then the approximation error measured by the variational distance vanishes exponentially as the length of output sequence tends to infinity. Whereas, Theorem 7 implies that if the number of fair bits per the length of output sequence is below the entropy, the approximation error approaches to two.

Also, they have showed the following theorems for the optimum error exponent of algorithms for random number generation.

Theorem 8 [6]: Let $\tilde{P}_Y^n(\mathbf{y})$ denote a distribution on \mathcal{Y}^n using any algorithm for random number generation using fair bits of length nR . Then for $R < R_{max}$,

$$\limsup_{n \rightarrow \infty} \left[-\frac{1}{n} \log d(P_Y^n, \tilde{P}_Y^n) \right] \leq E_{sp}(R, P_Y), \quad (23)$$

where $E_{sp}(R, P_Y)$ is given by (10). Further, $E_{sp}(R, P_Y) \geq E_r(R, P_Y)$ and equality holds for $R \leq R_0$, where

$$R_0 \triangleq D(U_{|\mathcal{Y}|} \| P_Y) + \log |\mathcal{Y}|.$$

Theorem 9 [6]: Consider the optimum algorithm for random number generation using fair bits of length nR , let \widehat{P}_Y^n denote the distribution on \mathcal{Y}^n which minimizes the variational distance. Then,

$$\lim_{n \rightarrow \infty} \left[-\frac{1}{n} \log \{2 - d(P_Y^n, \widehat{P}_Y^n)\} \right] = G(R, P_Y). \quad (24)$$

Further, $G(R, P_Y) > 0$ if and only if $R < H(Y)$ and $F(R, P_Y) > G(R, P_Y)$ for $R < H(Y)$.

From Theorem 6 and 8, the algorithm proposed by Uyematsu and Kanaya is optimum for $H(Y) < R \leq R_0$, but it is still an open problem to obtain the error exponent for $R > R_0$. Also, Theorem 7 and 9 imply that the algorithm is not optimum for $R < H(Y)$.

(b) Knuth-Yao Algorithm

In a similar manner as the above section, we modify the Knuth-Yao algorithm so that the algorithm outputs a specified sequence $\mathbf{y}_0 \in \mathcal{Y}^n$ whenever the algorithm does not stop with an input fair bits of length nR .

Modified Knuth-Yao Algorithm for Constructing Generating Tree with Fixed Input Length

- 1a) Prepare a root of tree, and set $i = 1$.
- 1b) For a sequence $\mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathcal{Y}^n$ from the DMS Y , the binary expansion of the probability $P_Y^n(\mathbf{y})$ be

$$P_Y^n(\mathbf{y}) = \sum_{k \geq 1} \alpha_k^{(\mathbf{y})} \exp(-k)$$

where $\alpha_k^{(\mathbf{y})} = \{0, 1\}$.

- 2) If nodes at depth $i - 1$ are not labeled with any symbols, then attach two branches and leaves to these nodes, and label one of the branches with 0, the other with 1.
- 3) For all $\mathbf{y} \in \mathcal{Y}^n$ satisfying $\alpha_i^{(\mathbf{y})} = 1$, label one of the attached leaves in step 2) with \mathbf{y} .
- 4a) If $i = nR$ then label all the leaves labeled with no label with \mathbf{y}_0 , and stop the algorithm.
- 4b) If all leaves are labeled, then stop the algorithm. Otherwise, set $i = i + 1$ and go to 2).

We measure the approximation error by the variational distance between the desired and approximated distribution. Then, we obtain the following theorems.

Theorem 10: If the modified Knuth-Yao algorithm is used for random number generation of an i.i.d. source with a generic distribution P_Y , then we have

$$\liminf_{n \rightarrow \infty} \left[-\frac{1}{n} \log d(P_Y^n, \tilde{P}_Y^n) \right] \geq E_r(R, P_Y), \quad (25)$$

where \tilde{P}_Y^n denotes the output distribution of the modified Knuth-Yao algorithm, and $E_r(R, P_Y)$ is given by (9).

Proof:

$$\begin{aligned} d(P_Y^n, \tilde{P}_Y^n) &= \sum_{\mathbf{y} \in \mathcal{Y}^n} |P_Y^n(\mathbf{y}) - \tilde{P}_Y^n(\mathbf{y})| \\ &= \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ \mathbf{y} \neq \mathbf{y}_0}} |P_Y^n(\mathbf{y}) - \tilde{P}_Y^n(\mathbf{y})| + \left| \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ \mathbf{y} \neq \mathbf{y}_0}} (P_Y^n(\mathbf{y}) - \tilde{P}_Y^n(\mathbf{y})) \right| \\ &= 2 \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ \mathbf{y} \neq \mathbf{y}_0}} |P_Y^n(\mathbf{y}) - \tilde{P}_Y^n(\mathbf{y})| \\ &\leq 2 \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ P_Y^n(\mathbf{y}) \geq \exp(-nR)}} \exp(-nR) + 2 \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ P_Y^n(\mathbf{y}) \leq \exp(-nR)}} P_Y^n(\mathbf{y}) \\ &\leq 2(n+1)^{|\mathcal{Y}|} \exp\{-n \min_{Q \in \mathcal{P}_n} (D(Q \| P_Y) + |R - H(Q) - D(Q \| P_Y)|^+)\}, \end{aligned}$$

which implies (25). \square

Theorem 11: If the modified Knuth-Yao algorithm is used for random number generation of an i.i.d. source with a generic distribution P_Y , then for $R > R_{min}$ we have

$$\lim_{n \rightarrow \infty} \left[-\frac{1}{n} \log \{2 - d(P_Y^n, \tilde{P}_Y^n)\} \right] = F(R, P_Y), \quad (26)$$

where $F(R, P_Y)$ is given by (11).

Proof: From the equality $a + b = |a - b| + 2 \min(a, b)$, we obtain

$$\begin{aligned}
2 - d(P_Y^n, \tilde{P}_Y^n) &= 2 \sum_{\mathbf{y} \in \mathcal{Y}^n} \min(P_Y^n(\mathbf{y}), \tilde{P}_Y^n(\mathbf{y})) \\
&= 2 \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ \mathbf{y} \neq \mathbf{y}_0}} \tilde{P}_Y^n(\mathbf{y}) + 2P_Y^n(\mathbf{y}_0) \\
&\leq 2 \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ P_Y^n(\mathbf{y}) \geq \exp(-nR)}} P_Y^n(\mathbf{y}) + 2 \exp(-nR) \\
&\leq 2(n+1)^{|\mathcal{Y}|} \exp\{-n \min_{\substack{Q \in \mathcal{P}_n: \\ D(Q \| P_Y) + H(Q) \leq R}} D(Q \| P_Y)\} + 2 \exp(-nR),
\end{aligned}$$

which implies

$$\liminf_{n \rightarrow \infty} \left[-\frac{1}{n} \log d(P_Y^n, \tilde{P}_Y^n) \right] \geq F(R, P_Y)$$

for $R > R_{min}$. We can obtain the reverse inequality as follows.

$$\begin{aligned}
2 - d(P_Y^n, \tilde{P}_Y^n) &= 2 \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ \mathbf{y} \neq \mathbf{y}_0}} \tilde{P}_Y^n(\mathbf{y}) + 2P_Y^n(\mathbf{y}_0) \\
&\geq 2 \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ P_Y^n(\mathbf{y}) \geq \exp(-nR)}} \frac{1}{2} P_Y^n(\mathbf{y}) \\
&\geq (n+1)^{-|\mathcal{Y}|} \exp\{-n \min_{\substack{Q \in \mathcal{P}_n: \\ D(Q \| P_Y) + H(Q) \leq R}} D(Q \| P_Y)\}.
\end{aligned}$$

Then, we obtain (26). □

From these theorems Theorem 6 and 7, we can conclude that the Knuth-Yao algorithm has the same error exponent measured by the variational distance as the interval algorithm using a sequence of fair bits.

VI. Conclusion

We have investigated large deviations performance of the Knuth-Yao algorithm. We have clarified some asymptotic properties, and we have showed that the performance is the same as the interval algorithm using a sequence of fair bits. As future reserches, we are going to generalize our results to more complex sources.

References

- [1] D. Knuth and A. Yao, "The complexity of nonuniform random number generation," *Algorithm and Complexity, New Directions and Results*, pp.357-428, ed. by J. F. Traub, Academic Press, New York, 1976.
- [2] S. Vembu and S. Verdú, "Generating random bits from an arbitrary source: Fundamental limits," *IEEE Trans. on Inform. Theory*, vol.IT-41, pp.1322-1332, 1995.
- [3] T. S. Han and M. Hoshi, "Interval algorithm for random number generation," *IEEE Trans. on Inform. Theory*, vol.43, pp.599-611, Mar. 1997.
- [4] F. Kanaya, "An asymptotically optimal algorithm for generating Markov random sequences," *Proc. of SITA '97*, pp.77-80, Matsuyama, Japan, Dec., 1997 (in Japanese).
- [5] Y. Ohama, "Fixed to fixed length random number generation using one dimensional piecewise linear maps," *Proc. of SITA '98*, pp.57-60, Gifu, Japan, Dec., 1998 (in Japanese).
- [6] T. Uyematsu and F. Kanaya, "Methods of channel simulation achieving conditional resolvability by statistically stable transformation," *submitted to IEEE Trans. on Inform. Theory*.
- [7] O. Uchida and T. S. Han, "Performance analysis of interval algorithm for generating Markov processes," *Proc. of SITA '98*, pp.65-68, Gifu, Japan, Dec., 1998.
- [8] A. Kimura and T. Uyematsu, "Large deviations performance of interval algorithm for random number generation," *Proc. of Memorial Workshop for 50th Anniversary of Shannon Theory*, pp.1-4, Yamanashi, Japan, Jan. 1999.
- [9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
- [10] T. S. Han: *Information-Spectrum Methods in Information Theory*, Baifukan, Tokyo, 1998 (in Japanese).
- [11] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. on Inform. Theory*, vol.IT-39, pp.752-772, May. 1993.

- [12] P. C. Shields: *The ergodic theory of discrete sample paths*, Graduate Studies in Math. vol.13, American Math. Soc., 1996.