Channel Simulation by Interval Algorithm: A Performance Analysis of Interval Algorithm [*]

Tomohiko UYEMATSU[†] and Fumio KANAYA[††]

uematsu@ss.titech.ac.jp fkanaya@info.shonan-it.ac.jp

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

[††] Dept. of Information Science,
Shonan Institute of Technology,
Fujisawa-shi, Kanagawa 251-0046, Japan

March 24, 1999

[*] The material in this correspondence was presented in part at 1999 IEEE Information Theory Workshop, Kruger National Park, South Africa, June 20-25, 1999.

Abstract

This paper deals with the problem of simulating a discrete memoryless channel and proposes two algorithms for channel simulation by using the interval algorithm. The first algorithm provides exact channel simulation and the number of fair random bits per input sample approaches the conditional resolvability of the channel with probability one. The second algorithm provides approximate channel simulation and the approximation error measured by the variational distance vanishes exponentially as the block length tends to infinity, when the number of fair random bits per input sample is above the conditional resolvability. Further, some asymptotic properties of these algorithms as well as the original interval algorithm for random number generation are clarified.

Keywords: channel simulation, interval algorithm, conditional resolvability, conditional entropy, random number generation

I. Introduction

The minimum randomness necessary to simulate an arbitrary given channel was first investigated by Steinberg and Verdú [1], as a complementary problem to that of approximating output statistics introduced by Han and Verdú [2]. In that problem, they measured the complexity of the simulator by the number of fair bits per input sample required to generate every realization of the simulated process and they adopted the variational distance as a measure of similarity between probability distributions. They considered the channel simulation such that the variational distance goes to zero as the block length tends to infinity and showed that the *conditional resolvability* defined by the minimum number of fair bits per input sample is equal to both the conditional sup-entropy and the minimum achievable fixed-length source coding rate with side information. However, they did not consider practical methods to simulate the channel nor the problem of simulating the channel without any approximation.

This paper deals with the problem of simulating a discrete memoryless channel (DMC) and proposes two algorithms for channel simulation that achieve the conditional resolvability. Both algorithms can be regarded as special cases of the interval algorithm proposed by Han and Hoshi for random number generation [3]. The first algorithm provides exact channel simulation. By investigating large deviations performance of the empirical number of fair bits, we show that the number of fair random bits per input sample approaches the conditional resolvability of the channel with probability one. The second algorithm provides approximate channel simulation. We demonstrate that the approximation error measured by the variational distance between the desired and approximate distributions, vanishes exponentially as the block length tends to infinity, when the number of fair random bits per input sample is above the conditional resolvability. On the contrary, the approximation error approaches the value of two exponentially, when the number of fair random bits per input sample is below the conditional resolvability. Further, we show that the second algorithm can achieve the optimum error exponent, when the number of fair random bits per input sample is within some range above the conditional resolvability.

Finally, since random number generation is a special case of channel simulation, all asymptotic properties obtained for channel simulation are directly applied to the interval algorithm for random number generation. In section V, we briefly describe some asymptotic properties of the interval algorithm, when it is specialized in random number generation.

II. Basic Definitions

(a) Discrete memoryless sources and channels

Let $\mathcal{X}$, $\mathcal{Y}$ be finite sets. We denote by $\mathcal{M}(\mathcal{X})$ [resp. $\mathcal{M}(\mathcal{Y})$] the set of all probability distributions on $\mathcal{X}$ (resp. $\mathcal{Y}$). Similarly, we denote by $\mathcal{M}(\mathcal{Y}|\mathcal{X})$ the set of all conditional distributions $W(\cdot|\cdot)$ such that $W(\cdot|a) \in \mathcal{M}(\mathcal{Y})$ for every $a \in \mathcal{X}$. 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. Similarly, a discrete memoryless channel (DMC) $W : \mathcal{X} \to \mathcal{Y}$ is given by a conditional distribution $W \in \mathcal{M}(\mathcal{Y}|\mathcal{X})$. A joint input-output process with a distribution $P_{XY} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ will be denoted by XY and P_{XY}.

For random variables (RV's) X and Y such that X has a distribution P_X and Y is connected with X by a DMC $W : \mathcal{X} \to \mathcal{Y}$, we shall denote their conditional entropy as $H(W|P_X)$ and $H(Y|X)$ interchangeably. Further, for arbitrary distributions $P, Q \in \mathcal{M}(\mathcal{X})$ and conditional distributions $V, W \in \mathcal{M}(\mathcal{Y}|\mathcal{X})$, we denote by $D(P \parallel Q)$ and $D(V \parallel W|P)$, the information divergence

$$D(P \parallel Q) \stackrel{\triangle}{=} \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)},$$

and the conditional information divergence

$$D(V \parallel W|P) \stackrel{\triangle}{=} \sum_{x \in \mathcal{X}} P(x) D(V(\cdot|x) \parallel W(\cdot|x)),$$

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

(b) Conditional resolvability

Here, we shall give some necessary definitions for conditional resolvability introduced by Steinberg and Verdú [1].

Definition 1: The variational distance or ℓ_1 distance between two distributions

P and Q on A is

$$d(P, Q) = \sum_{a \in A} |P(a) - Q(a)|.$$

Definition 2 [2]: The resolution $R(P)$ of a distribution $P \in \mathcal{M}(\mathcal{X})$ is the minimum $\log m$ such that P is an m-type (i.e., its masses are integer multiples of $1/m$). If such an m does not exist, $R(P) = \infty$.

Definition 3: The resolution $R_c(W)$ of a DMC $W : \mathcal{X} \to \mathcal{Y}$ is defined as

$$R_c(W) = \max_{a \in \mathcal{X}} R(W(\cdot|a)).$$

Definition 4: For a source P_X and a DMC $W : \mathcal{X} \to \mathcal{Y}$, let $P_X \cdot W$ be the joint distribution of input-output process of the DMC W. Then, R is ϵ-achievable resolution rate of XY for given X, if for every $\gamma > 0$ there exists a channel $\tilde{W}^n : \mathcal{X}^n \to \mathcal{Y}^n$ satisfying

$$\frac{1}{n} R_c(\tilde{W}^n) < R + \gamma$$

and

$$d(P_X^n \cdot W^n, P_X^n \cdot \tilde{W}^n) < \epsilon$$

for all sufficiently large n, where P_X^n (resp. W^n) denotes the n-th memoryless extension of P_X (resp. W).

Definition 5: $\sigma_\epsilon(XY|X)$ is the minimum ϵ-achievable resolution rate of XY given X. Especially, $\sigma(XY|X)$ given by

$$\sigma(XY|X) = \lim_{\epsilon \to 0} \sigma_\epsilon(XY|X) = \sup_{\epsilon > 0} \sigma_\epsilon(XY|X)$$

is the conditional resolvability of XY given X.

The next theorem indicates the relation between the conditional resolvability and the conditional entropy.

Theorem 1 [1]: For random variables X and Y such that X has a distribution P_X and Y is connected with X by a DMC $W : \mathcal{X} \to \mathcal{Y}$,

$$\sigma(XY|X) = H(Y|X).$$

III. Channel simulation by interval algorithm

In this section, we propose two algorithms for channel simulation by the interval algorithm [3]. Especially, we consider the simulation of DMC by using an unbiased coin and we investigate the required number of coin tosses.

(a) Required number of coin tosses for exact channel simulation

First, we propose an algorithm for channel simulation by using a version of the interval algorithm such that the partition of the unit interval depends on the input sequence. For the sake of simplicity and without any loss of generality, we assume that $\mathcal{Y} = \{1, 2, \cdots, M\}$.

Interval algorithm for channel simulation:

1) Let $x_1x_2\cdots x_n \in \mathcal{X}^n$ be a given input sequence. Set $m = 1, s = t = \lambda$ (null string), $\alpha_s = \gamma_t = 0, \beta_s = \delta_t = 1, I(s) = [\alpha_s, \beta_s)$.

2) Partition the interval $J(t) \overset{\triangle}{=} [\gamma_t, \delta_t)$ into M disjoint subintervals $J(t1), J(t2), \cdots, J(tM)$ such that

$$J(tj) \;=\; [\gamma_{tj}, \delta_{tj}) \quad (j = 1, 2, \cdots, M)$$

where

$$\begin{aligned} \gamma_{tj} &= \gamma_t + (\delta_t - \gamma_t) Q_{j-1}(x_1) \\ \delta_{tj} &= \gamma_t + (\delta_t - \gamma_t) Q_j(x_1) \\ Q_j(x_1) &= \sum_{k=1}^{j} W(k|x_1) \quad (j = 1, 2, \cdots, M;\; Q_0(x_1) = 0). \end{aligned}$$

3) Toss the unbiased coin Z such that it obtains a value $a \in \{0, 1\}$, and generate the subinterval of $I(s)$

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

where

$$\begin{aligned} \alpha_{sa} &= \alpha_s + (\beta_s - \alpha_s) a/2 \\ \beta_{sa} &= \alpha_s + (\beta_s - \alpha_s)(a + 1)/2. \end{aligned}$$

4a) If $I(sa)$ is entirely contained in some $J(ti) \quad (i = 1, 2, \cdots, M)$, generate the output i as the value of the mth output symbol Y_m and set $t = ti$. Otherwise go to 5).

4b) If $m = n$, stop the algorithm. Otherwise, partition the interval $J(t) \overset{\triangle}{=} [\gamma_t, \delta_t)$ into M disjoint subintervals $J(t1), J(t2), \cdots, J(tM)$ such that

$$J(tj) \;=\; [\gamma_{tj}, \delta_{tj}) \quad (j = 1, 2, \cdots, M)$$

where

$$\begin{aligned}
\gamma_{tj} &= \gamma_t + (\delta_t - \gamma_t) Q_{j-1}(x_{m+1}) \\
\delta_{tj} &= \gamma_t + (\delta_t - \gamma_t) Q_j(x_{m+1}) \\
Q_j(x_{m+1}) &= \sum_{k=1}^{j} W(k|x_{m+1}) \quad (j = 1, 2, \cdots, M;\ Q_0(x_{m+1}) = 0),
\end{aligned}$$

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

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

Though we can exactly simulate a DMC by the interval algorithm for channel simulation, the number of coin tosses to generate an output sequence cannot be specified in advance. Hence, for a given sequence $\boldsymbol{x} \in \mathcal{X}^n$, let $T_n(\boldsymbol{x})$ be the random variable indicating the number of coin tosses to generate a sequence in $\mathcal{Y}^n$. The next theorem gives a large deviations performance of $T_n(\boldsymbol{x})$.

Theorem 2: If the input process $X^n = X_1 X_2 \cdots X_n$ is an i.i.d. with a generic distribution P_X and the output process Y^n is connected to X^n by a DMC W, then the number of coin tosses $T_n(X^n)$ necessary to generate a sequence in $\mathcal{Y}^n$ satisfies

$$\begin{aligned}
&\liminf_{n\to\infty} \left[-\frac{1}{n} \log \Pr\{n^{-1} T_n(X^n) \geq R\} \right] \\
&\quad = \liminf_{n\to\infty} \left[-\frac{1}{n} \log \left(\sum_{\boldsymbol{x} \in \mathcal{X}^n} P_X^n(\boldsymbol{x}) \Pr\{n^{-1} T_n(\boldsymbol{x}) \geq R\} \right) \right] \\
&\quad \geq E_r(R, P_X, W), \qquad (1)
\end{aligned}$$

where

$$\begin{aligned}
E_r(R, P_X, W) \overset{\triangle}{=} & \min_{Q \in \mathcal{M}(\mathcal{X}), V \in \mathcal{M}(\mathcal{Y}|\mathcal{X})} [D(Q \cdot V \parallel P_X \cdot W) \\
& + |R - H(V|Q) - D(V \parallel W|Q)|^+] , \qquad (2)
\end{aligned}$$

and $|x|^+ = \max\{0, x\}$. Further, $E_r(R, P_X, W) > 0$ if and only if $R > H(Y|X)$. On the other hand,

$$\begin{aligned}
&\lim_{n\to\infty}\left[-\frac{1}{n}\log\Pr\{n^{-1}T_n(X^n)\le R\}\right] \\
&= \lim_{n\to\infty}\left[-\frac{1}{n}\log\left(\sum_{\boldsymbol{x}\in\mathcal{X}^n} P_X^n(\boldsymbol{x})\Pr\{n^{-1}T_n(\boldsymbol{x})\le R\}\right)\right] \\
&= F(R, P_X, W), \qquad (3)
\end{aligned}$$

where

$$F(R, P_X, W) \overset{\triangle}{=} \min_{\substack{Q\in\mathcal{M}(\mathcal{X}), V\in\mathcal{M}(\mathcal{Y}|\mathcal{X}):\\ D(V\|W|Q)+H(V|Q)\le R}} D(Q\cdot V \parallel P_X\cdot W). \qquad (4)$$

It should be noted that $F(R, P_X, W)$ may be $+\infty$. Further, $F(R, P_X, W) > 0$ if and only if $R < H(Y|X)$.

Combining Theorem 2 and Borel-Cantelli's lemma (see e.g. [4]), we immediately obtain the following corollary.

Corollary 1: If the input process $X^n = X_1X_2\cdots X_n$ is an i.i.d. with a generic distribution P_X, and the output process Y^n is connected to X^n by a DMC W, then the number of coin tosses $T_n(X^n)$ necessary to generate Y^n satisfies

$$\lim_{n\to\infty}\frac{1}{n}T_n(X^n) = H(Y|X) \quad a.s. \qquad (5)$$

We conclude that the interval algorithm for channel simulation is asymptotically optimum from the viewpoint of the necessary random bits per input sample.

(b) Channel simulation with fixed number of coin tosses

Next, we consider another problem of channel simulation where the number of coin tosses is specified. In this case, we cannot simulate the channel exactly but approximately. First, we modify the interval algorithm for channel simulation such that the algorithm outputs a *dummy sequence* $11\cdots1$, whenever the algorithm does not stop within N coin tosses.

Interval algorithm for channel simulation with n coin tosses:

1) Let $x_1x_2\cdots x_n \in \mathcal{X}^n$ be a given input sequence. Set $m=1, l=0, s=t=\lambda$ (null string), $\alpha_s=\gamma_t=0, \beta_s=\delta_t=1, I(s)=[\alpha_s,\beta_s)$.

2) Partition the interval $J(t) \overset{\triangle}{=} [\gamma_t,\delta_t)$ into M disjoint subintervals $J(t1), J(t2), \cdots, J(tM)$ such that

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

where

$$\begin{aligned}\gamma_{tj} &= \gamma_t+(\delta_t-\gamma_t)Q_{j-1}(x_1)\\ \delta_{tj} &= \gamma_t+(\delta_t-\gamma_t)Q_j(x_1)\\ Q_j(x_1) &= \sum_{k=1}^{j} W(k|x_1) \quad (j=1,2,\cdots,M;\ Q_0(x_1)=0).\end{aligned}$$

3) If $l=n$, generate the output $11\cdots1$ as the output sequence Y^n and stop the algorithm. Otherwise, toss the unbiased coin Z such that it obtains a value $a\in\{0,1\}$ and generate the subinterval of $I(s)$

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

where

$$
\begin{aligned}
\alpha_{sa} &= \alpha_s + (\beta_s - \alpha_s)a/2 \\
\beta_{sa} &= \alpha_s + (\beta_s - \alpha_s)(a+1)/2.
\end{aligned}
$$

Set $l = l + 1$.

4a) If $I(sa)$ is entirely contained in some $J(ti)$ $(i = 1, 2, \cdots, M)$, set $t = ti$. Otherwise go to 5).

4b) If $m = N$, generate the output t as the output sequence Y^n and stop the algorithm. Otherwise, partition the interval $J(t) \stackrel{\triangle}{=} [\gamma_t, \delta_t)$ into M disjoint subintervals $J(t1), J(t2), \cdots, J(tM)$ such that

$$
J(tj) \;=\; [\gamma_{tj}, \delta_{tj}) \quad (j = 1, 2, \cdots, M)
$$

where

$$
\begin{aligned}
\gamma_{tj} &= \gamma_t + (\delta_t - \gamma_t)Q_{j-1}(x_{m+1}) \\
\delta_{tj} &= \gamma_t + (\delta_t - \gamma_t)Q_j(x_{m+1}) \\
Q_j(x_{m+1}) &= \sum_{k=1}^{j} W(k|x_{m+1}) \quad (j = 1, 2, \cdots, M;\ Q_0(x_{m+1}) = 0),
\end{aligned}
$$

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

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

The next theorem shows that the approximation error measured by the variational distance between the desired and approximate output distributions vanishes exponentially, when the number of fair random bits per input sample is above the conditional resolvability $H(Y|X)$.

Theorem 3: For an input sequence $\boldsymbol{x} \in \mathcal{X}^n$, let $\tilde{W}^n(\boldsymbol{y}|\boldsymbol{x})$ denote the conditional probability of an output sequence $\boldsymbol{y} \in \mathcal{Y}^n$ generated by the interval algorithm with nR coin tosses. Then, we have

$$\liminf_{n\to\infty} \left[-\frac{1}{n} \log d(P_X^n \cdot W^n, P_X^n \cdot \tilde{W}^n) \right] \geq E_r(R, P_X, W), \tag{6}$$

where $E_r(R, P_X, W)$ is given by (2).

The following theorem is the converse of Theorem 3.

Theorem 4: For any $\hat{W}^n \in \mathcal{M}(\mathcal{Y}^n|\mathcal{X}^n)$ with its resolution $R_c(\hat{W}^n) = nR$, we have

$$\limsup_{n\to\infty} \left[-\frac{1}{n} \log d(P_X^n \cdot W^n, P_X^n \cdot \hat{W}^n) \right] \leq E_{sp}(R, P_X, W), \tag{7}$$

where

$$E_{sp}(R, P_X, W) \overset{\triangle}{=} \min_{\substack{Q \in \mathcal{M}(\mathcal{X}), V : \in \mathcal{M}(\mathcal{Y}|\mathcal{X}): \\ D(V\|W|Q) + H(V|Q) \geq R}} D(Q \cdot V \parallel P_X \cdot W). \tag{8}$$

It should be noted that $E_{sp}(R, P_X, W)$ may be $+\infty$. Further, $E_{sp}(R, P_X, W) \geq E_r(R, P_X, W)$ and equality holds for $R \leq R_o$, where

$$R_o \overset{\triangle}{=} D(V_o \parallel W|P_X) + \log M, \tag{9}$$

and $V_o(b|a) \overset{\triangle}{=} 1/M$ for every $a \in \mathcal{X}$ and $b \in \mathcal{Y}$.

According to Theorem 4, we can conclude that the bound obtained in Theorem 3 is tight whenever $R \leq R_o$, because $\tilde{W}^n$ in (6) obviously satisfies $R_c(\tilde{W}^n) = nR$. However, it is still an open problem to determine the error exponent for $R > R_o$.

The following theorems show that the approximation error approaches the value of two exponentially, when the number of fair random bits per input sample is below the conditional resolvability.

Theorem 5: For an input sequence $\boldsymbol{x} \in \mathcal{X}^n$, let $\tilde{W}(\boldsymbol{y}|\boldsymbol{x})$ denote the conditional probability of an output sequence $\boldsymbol{y} \in \mathcal{Y}^n$ generated by the interval algorithm with nR coin tosses. Then,

$$\lim_{n\to\infty}\left[-\frac{1}{n}\log\{2-d(P_X^n \cdot W^n, P_X^n \cdot \tilde{W}^n)\}\right] = F(R, P_X, W), \tag{10}$$

where $F(R, P_X, W)$ is given by (4).

Theorem 6: Let $\overline{W}^n \in \mathcal{M}(\mathcal{Y}^n|\mathcal{X}^n)$ denote the conditional probability which minimizes the variational distance $d(P_X^n \cdot W^n, P_X^n \cdot \overline{W}^n)$ under the condition $R_c(\overline{W}^n) = nR$. Then,

$$\lim_{n\to\infty}\left[-\frac{1}{n}\log\{2-d(P_X^n \cdot W^n, P_X^n \cdot \overline{W}^n)\}\right] = G(R, P_X, W), \tag{11}$$

where

$$G(R, P_X, W) \triangleq \min_{\substack{Q\in\mathcal{M}(\mathcal{X}), V:\in\mathcal{M}(\mathcal{Y}|\mathcal{X}):\\ H(V|Q)\le R}} D(Q \cdot V \parallel P_X \cdot W). \tag{12}$$

Obviously $G(R, P_X, W) < F(R, P_X, W)$ for $R < H(Y|X)$. Therefore, the interval algorithm with nR coin tosses cannot achieve the optimum exponent, whenever $R < H(Y|X)$.

IV. Proofs of Theorems

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

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

We shall write $\mathcal{P}_n$ for the set of types of sequences in $\mathcal{X}^n$. The *joint type* $P_{\boldsymbol{x},\boldsymbol{y}}$ of two sequences $\boldsymbol{x} \in \mathcal{X}^n$ and $\boldsymbol{y} \in \mathcal{Y}^n$ is the distribution on $\mathcal{X} \times \mathcal{Y}$, defined similarly. The set of sequences of type P in $\mathcal{X}^n$ is denoted by T_P^n or T_P. Further, for every $\boldsymbol{x} \in \mathcal{X}^n$ and $\boldsymbol{y} \in \mathcal{Y}^n$, if $\boldsymbol{x}$ and $\boldsymbol{y}$ have the joint type $P_{\boldsymbol{x},\boldsymbol{y}}(a,b) = P_{\boldsymbol{x}}(a)V(b|a)$, then we shall say that $\boldsymbol{y}$ has the *conditional type* V given $\boldsymbol{x}$. The set of such $\boldsymbol{y}$ will be denoted by $T_V(\boldsymbol{x})$. We shall denote by $\mathcal{V}(P)$ or $\mathcal{V}_n(P)$ the set of stochastic matrices $V : \mathcal{X} \to \mathcal{Y}$ such that $T_V(\boldsymbol{x}) \neq \emptyset$ for a sequence $\boldsymbol{x}$ of type P.

We introduce some well-known facts, cf. Csiszár-Körner [5]: For the set of types and the set of stochastic matrices, we have

$$\begin{aligned} |\mathcal{P}_n| &\leq (n+1)^{|\mathcal{X}|}, && (13)\\ |\mathcal{V}_n(P)| &\leq (n+1)^{|\mathcal{X}|\cdot|\mathcal{Y}|}, && (14) \end{aligned}$$

where $|\cdot|$ denotes the cardinality of the set. If $P \in \mathcal{P}_n$ then

$$(n+1)^{-|\mathcal{X}|} \exp\{-nD(P \parallel Q)\} \leq Q^n(T_P) \leq \exp\{-nD(P \parallel Q)\}. \tag{15}$$

If V is any conditional type of sequences in $\mathcal{Y}^n$ given $\boldsymbol{x} \in T_P$ then

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|} \exp\{nH(V|P)\} \leq |T_V(\boldsymbol{x})| \leq \exp\{nH(V|P)\}. \tag{16}$$

Further, if $\boldsymbol{x} \in T_P$ and $\boldsymbol{y} \in T_V(\boldsymbol{x})$, we then have

$$W^n(\boldsymbol{y}|\boldsymbol{x}) = \exp\{-n[D(V \parallel W|P) + H(V|P)]\}. \tag{17}$$

Proof of Theorem 2: Due to the nature of the interval algorithm, we can correspond each $\boldsymbol{y} \in \mathcal{Y}^n$ to several distinct subintervals $J(\boldsymbol{y})$ of $[0,1)$ with width $W^n(\boldsymbol{y}|\boldsymbol{x})$. On the other hand, partition a unit interval $[0,1)$ into $\exp(nR)$ subintervals

$$I_i \overset{\triangle}{=} [(i-1)\exp(-nR), i\exp(-nR)) \quad i = 1, 2, \cdots, \exp(nR).$$

Then, the interval I_i corresponds to each outcome of nR coin tosses. If the subinterval I_i is completely included in some $J(\boldsymbol{y})$, then the sequence of coin tosses corresponding to I_i can terminate the algorithm. By using this observation, (13)-(17) and the relation

$$D(Q \cdot V \parallel P_X \cdot W) = D(Q \parallel P_X) + D(V \parallel W|Q), \tag{18}$$

we have

$$\begin{aligned}
\Pr\{T_n(X^n) \geq nR\} &= \sum_{\boldsymbol{x} \in \mathcal{X}^n} P_X^n(\boldsymbol{x}) \Pr\{T_n(\boldsymbol{x}) \geq nR\} \\
&\leq \sum_{\boldsymbol{x} \in \mathcal{X}^n} P_X^n(\boldsymbol{x}) \Bigg(\sum_{\substack{\boldsymbol{y} \in \mathcal{Y}^n: \\ W^n(\boldsymbol{y}|\boldsymbol{x}) \geq \exp(-nR)}} 2\exp(-nR) + \sum_{\substack{\boldsymbol{y} \in \mathcal{Y}^n: \\ W^n(\boldsymbol{y}|\boldsymbol{x}) < \exp(-nR)}} W^n(\boldsymbol{y}|\boldsymbol{x}) \Bigg) \\
&\leq \sum_{Q \in \mathcal{P}_n} \exp\{-nD(Q \parallel P_X)\} \\
&\quad \times \Bigg(\sum_{\substack{V \in \mathcal{V}_n(Q): \\ D(V \| W|Q) + H(V|Q) \leq R}} 2\exp\{-n(R - H(V|Q))\} \\
&\quad + \sum_{\substack{V \in \mathcal{V}_n(Q): \\ D(V \| W|Q) + H(V|Q) > R}} \exp\{-nD(V \parallel W|Q)\} \Bigg) \\
&\leq 2 \sum_{Q \in \mathcal{P}_n, V \in \mathcal{V}_n(Q)} \exp\{-n(D(Q \cdot V \parallel P_X \cdot W) \\
&\quad + |R - H(V|Q) - D(V \parallel W|Q)|^+)\} \\
&\leq 2(n+1)^{|\mathcal{X}||\mathcal{Y}| + |\mathcal{X}|} \exp\{-nE_r(R, P_X, W)\}
\end{aligned}$$

which implies (1).

On the other hand, $E_r(R, P_X, W) = 0$ if and only if $Q \in \mathcal{M}(\mathcal{X})$ and $V \in \mathcal{M}(\mathcal{Y}|\mathcal{X})$ satisfy both $Q \cdot V = P_X \cdot W$ and $D(V \parallel W|Q) + H(V|Q) \geq R$. However, if Q and V satisfy $Q \cdot V = P_X \cdot W$, then we have

$$D(V \parallel W|Q) + H(V|Q) = H(W|P_X) = H(Y|X).$$

This implies $E_r(R, P, W) > 0$ if and only if $R > H(Y|X)$.

Next, we show (3). In a similar manner as in the proof of (1), by using (13)-(18), we obtain

$$\begin{aligned}
&\Pr\{T_n(X^n) \leq nR\} \\
&= \sum_{\boldsymbol{x} \in \mathcal{X}^n} P_X^n(\boldsymbol{x}) \Pr\{T_n(\boldsymbol{x}) \leq nR\} \\
&\leq \sum_{\boldsymbol{x} \in \mathcal{X}^n} P_X^n(\boldsymbol{x}) \sum_{\substack{\boldsymbol{y} \in \mathcal{Y}^n: \\ W^n(\boldsymbol{y}|\boldsymbol{x}) \geq \exp(-nR)}} W^n(\boldsymbol{y}|\boldsymbol{x}) \\
&\leq \sum_{\substack{Q \in \mathcal{P}_n, V \in \mathcal{V}_n(Q): \\ D(V \parallel W|Q) + H(V|Q) \leq R}} \exp\{-n(D(Q \parallel P_X) + D(V \parallel W|Q))\} \\
&\leq (n+1)^{|\mathcal{X}| + |\mathcal{X}||\mathcal{Y}|} \exp\{-nF(R, P_X, W)\},
\end{aligned}$$

which implies

$$\liminf_{n \to \infty} \left[-\frac{1}{n} \log \Pr\{T_n(X^n) \leq nR\} \right] \geq F(R, P_X, W).$$

The reverse inequality can be obtained similarly by using the relation

$$\begin{aligned}
&\Pr\{T_n(X^n) \leq nR\} \\
&\geq \sum_{\boldsymbol{x} \in \mathcal{X}^n} P_X^n(\boldsymbol{x}) \sum_{\substack{\boldsymbol{y} \in \mathcal{Y}^n: \\ W^n(\boldsymbol{y}|\boldsymbol{x}) \geq 2\exp(-nR)}} \frac{1}{3} W^n(\boldsymbol{y}|\boldsymbol{x})
\end{aligned}$$

$$\geq \frac{1}{3}\exp\{-n \min_{\substack{Q\in\mathcal{P}_n, V\in\mathcal{V}_n(Q):\\ D(V\|W|Q)+H(V|Q)\leq R-1/n}} D(Q\cdot V \parallel P_X\cdot W)\}.$$

Lastly, $F(R, P_X, W) = 0$ if and only if Q and V satisfy both $Q\cdot V = P_X\cdot W$ and $D(V \parallel W|Q) + H(V|Q) \leq R$. In such a case, $H(W|P_X) = H(Y|X) \leq R$ must hold. Hence, we have $F(R, P_X, W) > 0$ if and only if $R < H(Y|X)$. □

Proof of Theorem 3: Let $\tilde{W}^n(\boldsymbol{y}|\boldsymbol{x})$ be the probability of obtaining the output $\boldsymbol{y}\in\mathcal{Y}^n$ for a given input $\boldsymbol{x}\in\mathcal{X}^n$ by the interval algorithm with nR coin tosses. Then, for $\boldsymbol{y}\neq 11\cdots 1$,

$$0 \leq W^n(\boldsymbol{y}|\boldsymbol{x}) - \tilde{W}^n(\boldsymbol{y}|\boldsymbol{x}) \leq \begin{cases} 2\exp(-nR) & \text{if } W^n(\boldsymbol{y}|\boldsymbol{x}) \geq \exp(-nR), \\ W^n(\boldsymbol{y}|\boldsymbol{x}) & \text{otherwise .}\end{cases} \tag{19}$$

According to (19), for every sequence $\boldsymbol{x}\in\mathcal{X}^n$, we have

$$\begin{aligned}
&\sum_{\boldsymbol{y}\in\mathcal{Y}^n} |\tilde{W}^n(\boldsymbol{y}|\boldsymbol{x}) - W^n(\boldsymbol{y}|\boldsymbol{x})| \\
&= \sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ \boldsymbol{y}\neq 11\cdots 1}} |\tilde{W}^n(\boldsymbol{y}|\boldsymbol{x}) - W^n(\boldsymbol{y}|\boldsymbol{x})| + \left|\sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ \boldsymbol{y}\neq 11\cdots 1}} (W^n(\boldsymbol{y}|\boldsymbol{x}) - \tilde{W}^n(\boldsymbol{y}|\boldsymbol{x}))\right| \\
&\leq 2\sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ \boldsymbol{y}\neq 11\cdots 1}} |\tilde{W}^n(\boldsymbol{y}|\boldsymbol{x}) - W^n(\boldsymbol{y}|\boldsymbol{x})| \\
&\leq 4\sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ W^n(\boldsymbol{y}|\boldsymbol{x})\geq\exp(-nR)}} \exp(-nR) + 2\sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ W^n(\boldsymbol{y}|\boldsymbol{x})<\exp(-nR)}} W^n(\boldsymbol{y}|\boldsymbol{x}).
\end{aligned}$$

In a similar manner as in the proof of (1) in Theorem 2, we have

$$d(P_X^n\cdot W^n, P_X^n\cdot\tilde{W}^n) = \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}^n\times\mathcal{Y}^n} P_X^n(\boldsymbol{x})|\tilde{W}^n(\boldsymbol{y}|\boldsymbol{x}) - W^n(\boldsymbol{y}|\boldsymbol{x})|$$

$$\begin{aligned} &\le\; 4 \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_X^n(\boldsymbol{x}) \left(\sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ W^n(\boldsymbol{y}|\boldsymbol{x})\ge\exp(-nR)}} \exp(-nR) + \sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ W^n(\boldsymbol{y}|\boldsymbol{x})<\exp(-nR)}} W^n(\boldsymbol{y}|\boldsymbol{x}) \right) \\ &\le\; 4(n+1)^{|\mathcal{X}||\mathcal{Y}|+|\mathcal{X}|}\exp\{-nE_r(R,P_X,W)\}, \end{aligned}$$

which implies (6). □

Proof of Theorem 4: Since $R_c(\hat{W}^n) = nR$, we have

$$|\hat{W}^n(\boldsymbol{y}|\boldsymbol{x}) - W^n(\boldsymbol{y}|\boldsymbol{x})| \ge W^n(\boldsymbol{y}|\boldsymbol{x}) \quad \text{if } W^n(\boldsymbol{y}|\boldsymbol{x}) \le \exp(-nR)/2.$$

Hence, for every sequence $\boldsymbol{x} \in T_Q$,

$$\begin{aligned} &\sum_{\boldsymbol{y}\in\mathcal{Y}^n} |\hat{W}^n(\boldsymbol{y}|\boldsymbol{x}) - W^n(\boldsymbol{y}|\boldsymbol{x})| \\ &\quad\ge \sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ W^n(\boldsymbol{y}|\boldsymbol{x})\le\exp(-nR)/2}} W^n(\boldsymbol{y}|\boldsymbol{x}) \\ &\quad\ge (n+1)^{-|\mathcal{X}||\mathcal{Y}|}\exp\{-n \min_{\substack{V\in\mathcal{V}_n(Q):\\ D(V\|W|Q)+H(V|Q)\ge R+1/n}} D(V \parallel W|Q)\}. \end{aligned}$$

This implies that

$$\begin{aligned} d(P_X^n\cdot W^n, P_X^n\cdot\hat{W}^n) &= \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}^n\times\mathcal{Y}^n} P_X^n(\boldsymbol{x})|\hat{W}^n(\boldsymbol{y}|\boldsymbol{x}) - W^n(\boldsymbol{y}|\boldsymbol{x})| \\ \ge\; (n+1)^{-|\mathcal{X}||\mathcal{Y}|} &\sum_{Q\in\mathcal{P}_n} P_X^n(T_Q)\exp\{-n \min_{\substack{V\in\mathcal{V}_n(Q):\\ D(V\|W|Q)+H(V|Q)\ge R+1/n}} D(V \parallel W|Q)\} \\ \ge\; (n+1)^{-|\mathcal{X}||\mathcal{Y}|-|\mathcal{X}|}&\exp\{-nE_{sp}(R+1/n,P_X,W)\}. \end{aligned}$$

By using the continuity of divergence and entropy, we can obtain (7).

From (2) and (8), it is easy to see $E_{sp}(R,P_X,W) \ge E_r(R,P_X,W)$. In what follows, we investigate when the equality holds. First, rewrite $E_r(R,P_X,W)$ as follows.

$$E_r(R,P_X,W) = \min[E_{sp}(R,P_X,W), E_1(R,P_X,W)], \tag{20}$$

where

$$E_1(R, P_X, W) \;\overset{\triangle}{=}\; \min_{\substack{Q\in\mathcal{M}(\mathcal{X}), V\in\mathcal{M}(\mathcal{Y}|\mathcal{X}):\\ D(V\|W|Q)+H(V|Q)\le R}} [R - H(V|Q) + D(Q \parallel P_X)]. \tag{21}$$

Obviously, we can see that

$$E_1(R, P_X, W) \ge R - \log M, \tag{22}$$

where the equality holds if and only if $H(V|Q) = \log M$ and $Q = P_X$. However for every $V \in \mathcal{M}(\mathcal{Y}|\mathcal{X})$ satisfying $H(V|P_X) = \log M$, we have

$$D(V \parallel W|P_X) + H(V|P_X) = D(V_0 \parallel W|P_X) + H(V_0|P_X) = R_0.$$

Therefore, if $R \ge R_0$,

$$E_r(R, P_X, W) = R - \log M.$$

On the other hand, note that $-H(V|Q)+D(Q \parallel P_X)$ [resp. $D(V \parallel W|Q)+H(V|Q)$] is a convex (resp. linear) function of V for an arbitrary fixed Q. Then, for $R \le R_0$, the minimum of (21) can be attained at its boundary, that is,

$$\begin{aligned} E_1(R, P_X, W) &= \min_{\substack{Q\in\mathcal{M}(\mathcal{X}), V\in\mathcal{M}(\mathcal{Y}|\mathcal{X}):\\ D(V\|W|Q)+H(V|Q)=R}} [R - H(V|Q) + D(Q \parallel P_X)] \\ &= \min_{\substack{Q\in\mathcal{M}(\mathcal{X}), V\in\mathcal{M}(\mathcal{Y}|\mathcal{X}):\\ D(V\|W|Q)+H(V|Q)=R}} D(Q\cdot V \parallel P_X \cdot W). \end{aligned}$$

The above equation and (8) implies $E_1(R, P_X, W) \ge E_{sp}(R, P_X, W)$ for $R \le R_0$. Hence, $E_r(R, P_X, W) = E_{sp}(R, P_X, W)$ whenever $R \le R_0$. □

Proof of Theorem 5: For a given $\boldsymbol{x} \in \mathcal{X}^n$, according to the identity $a + b = |a - b| + 2\min(a, b)$ and $\tilde{W}^n(\boldsymbol{y}|\boldsymbol{x}) \le W^n(\boldsymbol{y}|\boldsymbol{x})$ for $\boldsymbol{y} \ne 11\cdots 1$, we have

$$\begin{aligned}
&2 - \sum_{\boldsymbol{y}\in\mathcal{Y}^n} |\tilde{W}^n(\boldsymbol{y}|\boldsymbol{x}) - W^n(\boldsymbol{y}|\boldsymbol{x})| \\
&\quad = 2 \sum_{\boldsymbol{y}\in\mathcal{Y}^n} \min(\tilde{W}^n(\boldsymbol{y}|\boldsymbol{x}), W^n(\boldsymbol{y}|\boldsymbol{x})) \\
&\quad \le 2 \sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\boldsymbol{y}\ne 11\cdots 1\\ W^n(\boldsymbol{y}|\boldsymbol{x})\ge\exp(-nR)}} W^n(\boldsymbol{y}|\boldsymbol{x}) + 2W^n(11\cdots 1|\boldsymbol{x}) \\
&\quad \le 2 \sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ W^n(\boldsymbol{y}|\boldsymbol{x})\ge\exp(-nR)}} W^n(\boldsymbol{y}|\boldsymbol{x}) + 2\exp(-nR).
\end{aligned}$$

Here, the last inequality holds either for the case $W^n(11\cdots 1|\boldsymbol{x}) \ge \exp(-nR)$ or $W^n(11\cdots 1|\boldsymbol{x}) < \exp(-nR)$. Therefore,

$$\begin{aligned}
&2 - d(P_X^n \cdot \tilde{W}^n, P_X^n \cdot W^n) \\
&\quad \le 2 \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_X^n(\boldsymbol{x}) \left(\sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ W^n(\boldsymbol{y}|\boldsymbol{x})\ge\exp(-nR)}} W^n(\boldsymbol{y}|\boldsymbol{x}) + \exp(-nR) \right),
\end{aligned}$$

and in a similar manner as in the proof of (3) in Theorem 2, we can obtain

$$\liminf_{n\to\infty} \left[-\frac{1}{n} \log\{2 - d(P_X^n \cdot \tilde{W}^n, P_X^n \cdot W^n)\} \right] \ge F(R, P_X, W).$$

The reverse inequality comes from the relation

$$\begin{aligned}
2 - \sum_{\boldsymbol{y}\in\mathcal{Y}^n} |\tilde{W}^n(\boldsymbol{y}|\boldsymbol{x}) - W^n(\boldsymbol{y}|\boldsymbol{x})| &\ge 2 \sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ W^n(\boldsymbol{y}|\boldsymbol{x})\ge 2\exp(-nR)}} \min(\tilde{W}^n(\boldsymbol{y}|\boldsymbol{x}), W^n(\boldsymbol{y}|\boldsymbol{x})) \\
&\ge 2 \sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ W^n(\boldsymbol{y}|\boldsymbol{x})\ge 2\exp(-nR)}} \frac{1}{3} W^n(\boldsymbol{y}|\boldsymbol{x}).
\end{aligned}$$

□

Proof of Theorem 6: First, we shall show the converse part. It is easy to see that

$$\begin{aligned}
2-\sum_{\boldsymbol{y}\in\mathcal{Y}^n}|\overline{W}^n(\boldsymbol{y}|\boldsymbol{x})-W^n(\boldsymbol{y}|\boldsymbol{x})| &= 2\sum_{\boldsymbol{y}\in\mathcal{Y}^n}\min(\overline{W}^n(\boldsymbol{y}|\boldsymbol{x}),W^n(\boldsymbol{y}|\boldsymbol{x}))\\
&\le 2\sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ \overline{W}^n(\boldsymbol{y}|\boldsymbol{x})>0}} W^n(\boldsymbol{y}|\boldsymbol{x}).
\end{aligned}$$

Since $\overline{W}^n$ has a resolution nR, $\overline{W}^n(\boldsymbol{y}|\boldsymbol{x})$ must be positive for at most $\exp(nR)$ sequences in $\mathcal{Y}^n$ for a given $\boldsymbol{x}\in\mathcal{X}^n$. Hence, we have

$$\begin{aligned}
&2-d(P_X^n\cdot\overline{W}^n,P_X^n\cdot W^n)\\
&\le 2\sum_{Q\in\mathcal{P}_n}P_X^n(\boldsymbol{x})\sum_{\substack{\boldsymbol{y}\in\mathcal{Y}^n:\\ \overline{W}^n(\boldsymbol{y}|\boldsymbol{x})>0}} W^n(\boldsymbol{y}|\boldsymbol{x})\\
&\le 2\sum_{Q\in\mathcal{P}_n}\exp\{-nD(Q\parallel P_X)\}\sum_{V\in\mathcal{V}_n(Q)}\exp\{-nD(V\parallel W|Q)\}\\
&\quad\times\min\left\{1,\frac{\exp(nR)}{\exp\{nH(V|Q)\}}\right\}\\
&\le 2(n+1)^{|\mathcal{X}|+|\mathcal{X}||\mathcal{Y}|}\exp\{-n\min_{Q\in\mathcal{M}(\mathcal{X}),V\in\mathcal{M}(\mathcal{Y}|\mathcal{X})}[D(Q\cdot V\parallel P_X\cdot W)\\
&\quad+|H(V|Q)-R|^+]\}. \qquad (23)
\end{aligned}$$

For a fixed Q, $S(Q)\overset{\triangle}{=}\{V\in\mathcal{M}(\mathcal{Y}|\mathcal{X}):H(V|Q)\ge R\}$ is a convex set. If we restrict the region of V to $S(Q)$, $D(Q\cdot V\parallel P_X\cdot W)+H(V|Q)-R$ is a linear function of V. Then, we have

$$\min_{V\in S(Q)}[D(Q\cdot V\parallel P_X\cdot W)+|H(V|Q)-R|^+]=\min_{\substack{V\in\mathcal{M}(\mathcal{Y}|\mathcal{X}):\\ H(V|Q)=R}}D(Q\cdot V\parallel P_X\cdot W).$$

Therefore the region of taking the minimum for V in (23) may be restricted to $H(V|Q)\le R$, and we obtain

$$\liminf_{n\to\infty}\left[-\frac{1}{n}\log\{2-d(P_X^n\cdot\tilde{W}^n,P_X^n\cdot W^n)\}\right]\ge G(R,P_X,W).$$

Next we shall show the achievability part. For any $\boldsymbol{x} \in T_Q$, we can select a conditional type $V' \in \mathcal{V}(Q)$ such that $\exp\{nH(V'|Q)\} \leq \exp(nR)$ and V' minimizes $D(V' \parallel W|Q)$. Then, we can choose a set $\overline{T}(\boldsymbol{x})$ satisfying both $\overline{T}(\boldsymbol{x}) \subset T_{V'}(\boldsymbol{x})$ and $|T_{V'}(\boldsymbol{x})|/3 \leq |\overline{T}(\boldsymbol{x})| < |T_{V'}(\boldsymbol{x})|/2$. By using the set $\overline{T}(\boldsymbol{x})$, we assign the conditional probability $\overline{W}^n(\boldsymbol{y}|\boldsymbol{x})$ as follows. For $\boldsymbol{y} \in \overline{T}(\boldsymbol{x})$, we assign

$$\overline{W}^n(\boldsymbol{y}|\boldsymbol{x}) = \left\lfloor \frac{\exp(nR)}{|\overline{T}(\boldsymbol{x})|} \right\rfloor \exp(-nR) \quad \text{or} \quad \left(\left\lfloor \frac{\exp(nR)}{|\overline{T}(\boldsymbol{x})|} \right\rfloor + 1 \right) \exp(-nR)$$

such that

$$\sum_{\boldsymbol{y} \in \overline{T}(\boldsymbol{x})} \overline{W}^n(\boldsymbol{y}|\boldsymbol{x}) = 1,$$

where $\lfloor z \rfloor$ denotes the maximum integer less than or equal to z. On the other hand, we assign $\overline{W}^n(\boldsymbol{y}|\boldsymbol{x}) = 0$ for $\boldsymbol{y} \notin \overline{T}(\boldsymbol{x})$. Obviously, $R_c(\overline{W}^n) = nR$. Since $2|\overline{T}(\boldsymbol{x})| < |T_{V'}(\boldsymbol{x})| \leq \exp\{nH(V'|Q)\} \leq \exp(nR)$, we have

$$\overline{W}^n(\boldsymbol{y}|\boldsymbol{x}) \geq \left(\frac{2\exp(nR)}{|T_V(\boldsymbol{x})|} - 1 \right) \exp(-nR) \geq \frac{1}{|T_{V'}(\boldsymbol{x})|} \geq W^n(\boldsymbol{y}|\boldsymbol{x}),$$

for every $\boldsymbol{y} \in \overline{T}(\boldsymbol{x})$. Hence, by using the above inequality and the identity $a + b = |a - b| + 2\min(a, b)$, we have

$$\begin{aligned}
& 2 - \sum_{\boldsymbol{y} \in \mathcal{Y}^n} |\overline{W}^n(\boldsymbol{y}|\boldsymbol{x}) - W^n(\boldsymbol{y}|\boldsymbol{x})| \\
&= 2 - \sum_{\boldsymbol{y} \in \mathcal{Y}^n} (\overline{W}^n(\boldsymbol{y}|\boldsymbol{x}) + W^n(\boldsymbol{y}|\boldsymbol{x})) + 2 \sum_{\boldsymbol{y} \in \mathcal{Y}^n} \min(\overline{W}^n(\boldsymbol{y}|\boldsymbol{x}), W^n(\boldsymbol{y}|\boldsymbol{x})) \\
&= 2 \sum_{\boldsymbol{y} \in \mathcal{Y}^n} \min(\overline{W}^n(\boldsymbol{y}|\boldsymbol{x}), W^n(\boldsymbol{y}|\boldsymbol{x})) \\
&= 2 \sum_{\boldsymbol{y} \in \overline{T}(\boldsymbol{x})} W^n(\boldsymbol{y}|\boldsymbol{x}) \\
&\geq \frac{2}{3} W^n(T_{V'}(\boldsymbol{x})|\boldsymbol{x})
\end{aligned}$$

$$\geq \frac{2}{3}(n+1)^{-|\mathcal{X}||\mathcal{Y}|}\exp\{-n \min_{\substack{V\in\mathcal{V}_n(Q):\\ H(V|Q)\leq R}} D(V \parallel W|Q)\},$$

where the last inequality follows from the choice of V'. This implies that

$$\begin{aligned} &2 - d(P_X^n \cdot \overline{W}^n, P_X^n \cdot W^n) \\ &\quad \geq \frac{2}{3}(n+1)^{-|\mathcal{X}|-|\mathcal{X}||\mathcal{Y}|}\exp\{-n \min_{Q\in\mathcal{P}_n} D(Q \parallel P_X)\} \\ &\quad\quad \times \exp\{-n \min_{\substack{V\in\mathcal{V}_n(Q):\\ H(V|Q)\leq R}} D(V \parallel W|Q)\}. \end{aligned}$$

Therefore, we have

$$\limsup_{n\to\infty}\left[-\frac{1}{n}\log\{2 - d(P_X^n \cdot \overline{W}^n, P_X^n \cdot W^n)\}\right] \leq G(R, P_X, W).$$

□

V. Performance analysis of random number generation by interval algorithm

If the input sequence $\boldsymbol{x} \in \mathcal{X}^n$ consists of one symbol, i.e. $\boldsymbol{x} = 11\cdots 1$, the proposed algorithms generate sequences of an i.i.d. source with a distribution $W(\cdot|1)$. In such a case, the first algorithm can be reduced to the original interval algorithm for random number generation [3], and our analysis can be directly applied to the original interval algorithm.

Consider the case where the interval algorithm is used as random number generation of an i.i.d. source with a generic distribution P_Y. Then, Theorem 2 indicates a large deviations performance of the interval algorithm for random number generation. Especially, the number of coin tosses T_n necessary to generate a sequence in

$\mathcal{Y}^n$ satisfies

$$\lim_{n\to\infty} \frac{1}{n} T_n = H(Y) \quad a.s. \tag{24}$$

Remark 1: As for the random number generation, Han and Hoshi have already shown that the average number of coin tosses per output sample approaches $H(Y)$ [3]. On the other hand, Theorem 2 clarifies that the number of coin tosses per output sample approaches $H(Y)$ with probability one, which is not implied in [3].

The second algorithm can be used as random number generation with specified number of coin tosses. In this situation, the random number is generated not exactly but approximately within a nonzero but arbitrary small tolerance in terms of variational distance. Such random number generation has been studied by Han and Verdú [2], but the rate of convergence in terms of variational distance has not yet been investigated even for an i.i.d. source. On the other hand, the performance analysis for the second algorithm gives some fundamental solutions for the rate of convergence. The next three corollaries are direct applications of Theorems 3-5.

Corollary 2: If the interval algorithm with nR coin tosses is used as random number generation of an i.i.d. source with a generic distribution P_Y, then we have

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

where $\tilde{P}_Y^n$ denotes the output distribution of the interval algorithm with nR coin tosses and

$$E_r(R, P_Y) \stackrel{\triangle}{=} \min_{Q \in \mathcal{M}(\mathcal{Y})} [D(Q \parallel P_Y) + |R - H(Q) - D(Q \parallel P_Y)|^+]. \tag{26}$$

Further, $E(R, P_Y) > 0$ if and only if $R > H(Y)$.

Corollary 3: Let $\hat{P}_Y^n$ denote any probability distribution with its resolution $R(\hat{P}_Y^n) = nR$. Then, for a given distribution $P_Y \in \mathcal{M}(\mathcal{Y})$,

$$\limsup_{n\to\infty}\left[-\frac{1}{n}\log d(P_Y^n, \hat{P}_Y^n)\right] \le E_{sp}(R, P_Y), \tag{27}$$

where

$$E_{sp}(R, P_Y) \stackrel{\triangle}{=} \min_{\substack{Q\in\mathcal{M}(\mathcal{Y}):\\ D(Q\|P_Y)+H(Q)\ge R}} D(Q \parallel P_Y). \tag{28}$$

Further, $E_{sp}(R, P_Y) \ge E_r(R, P_Y)$ and equality holds for $R \le R_o$, where

$$R_o \stackrel{\triangle}{=} D(Q_o \parallel P_Y) + \log M, \tag{29}$$

and $Q_o(b) \stackrel{\triangle}{=} 1/M$ for every $b \in \mathcal{Y}$.

Corollary 4: If the interval algorithm with nR coin tosses is used as random number generation of an i.i.d. source with a generic distribution P_Y, then we have

$$\lim_{n\to\infty}\left[-\frac{1}{n}\log\{2 - d(P_Y^n, \tilde{P}_Y^n)\}\right] = F(R, P_Y), \tag{30}$$

where $\tilde{P}_Y^n$ denotes the output distribution of the interval algorithm with nR coin tosses and

$$F(R, P_Y) \stackrel{\triangle}{=} \min_{\substack{Q\in\mathcal{M}(\mathcal{Y}):\\ D(Q\|P_Y)+H(Q)\le R}} D(Q \parallel P_Y). \tag{31}$$

Further, $F(R, P_Y) > 0$ whenever $R < H(Y)$.

These corollaries indicate the following properties of the interval algorithm for random number generation: (1) If the number of coin tosses per input sample is above resolvability $H(Y)$, the approximation error measured by the variational distance vanishes exponentially as the block length tends to infinity. Further, the interval algorithm achieves the optimum error exponent, when the number of coin tosses per input sample is less than R_o. (2) If the number of coin tosses per input sample is below the resolvability, the approximation error approaches the value of two exponentially.

The last corollary is a direct consequence of Theorem 6 and clarifies the optimum exponent below the resolvability.

Corollary 5: Let $\overline{P}_Y^n$ denote the distribution which minimizes the variational distance $d(P_Y^n, \overline{P}_Y^n)$ under the condition $R_c(\overline{P}_Y^n) = nR$. Then,

$$\lim_{n\to\infty}\left[-\frac{1}{n}\log\{2 - d(P_Y^n, \overline{P}_Y^n)\}\right] = G(R, P_Y), \tag{32}$$

where

$$G(R, P_Y) \stackrel{\triangle}{=} \min_{\substack{Q\in\mathcal{M}(\mathcal{Y}):\\ H(Q)\le R}} D(Q \parallel P_Y). \tag{33}$$

From Corollaries 4 and 5, we can conclude that the interval algorithm with nR coin tosses is not optimum for $R < H(Y)$.

VI. Conclusion

We have proposed two algorithms for channel simulation based on the interval algorithm. The first algorithm provides an exact channel simulation, while the second one provides an approximate channel simulation. We have clarified some asymptotic properties of these algorithms as well as the random number generation by the interval algorithm. Regarding future research, we shall generalize our results to more complex channels, such as channels with memory.

Acknowledgment

The authors wish to thank Prof. Te Sun Han of University of Electro-Communications for his valuable discussions and comments, which led to an improvement of the original manuscript. They would also like to thank the anonymous referees for their detailed and very helpful suggestions and corrections that much improved this correspondence.

References

[1] Y. Steinberg and S. Verdú, "Channel simulation and coding with side information," *IEEE Trans. on Inform. Theory*, vol.40, pp.634-646, 1994.

[2] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. on Inform. Theory*, vol.IT-39, pp.752-772, 1993.

[3] T. S. Han and M. Hoshi, "Interval algorithm for random number generation," *IEEE Trans. on Inform. Theory*, vol.43, pp.599-611, 1997.

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

[5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.