tends to unity as $n \to \infty$ (while N is kept fixed), we have constructed a sequence of fixed-rate block codes that satisfies (8).

To estimate the number of codewords (and hence the rate) at the first level code, we apply Lemma 4 by setting $R = N(R_N + \delta)$, $u_i = G_N^1(\boldsymbol{x}_i)$, and $J = |\mathcal{Y}|^N$, where the latter assignment expresses the fact that in the finite reproduction alphabet case, the guessing list size need not exceed the total number of possible reproduction vectors. Thus we can upper-bound the number of codewords in the first level by

$$M_1 \le (n+1)^{|\mathcal{Y}|^N} \exp\{n[N(R_N+\delta) + \ln(2\ln|\mathcal{Y}|^N + 2)]\} = \exp\left\{Nn\left[R_N + \delta + \frac{|\mathcal{Y}|^N \ln(n+1)}{Nn} + \frac{\ln(2N\ln|\mathcal{Y}|+2)}{N}\right]\right\}. \tag{A.7}$$

Letting $n \to \infty$ for fixed N, we see that the exponent of this expression tends to $R_N + \delta + \ln(2N\ln|\mathcal{Y}|+2)/N$. In the same manner, one can verify that the total number of codewords at the second level satisfies

$$\limsup_{n\to\infty} \frac{1}{nN} \ln M_2 \le R_N + \Delta_N + 2\delta + \frac{1}{N}[\ln(2N\ln|\mathcal{Y}|+2) + \ln(2N\ln|\mathcal{Z}|+2)].$$

Clearly, there exists a constant c (that depends solely on $|\mathcal{Y}|$ and $|\mathcal{Z}|$) such that $c\ln(N+1)/N$ upper-bounds the $O(\log N/N)$ terms in the exponents of both M_1 and M_2, for all N. Finally, since δ is arbitrarily small, this implies that

$$(R_N + c\ln(N+1)/N,\ R_N + \Delta_N + c\ln(N+1)/N,\ D_1,\ D_2)$$

is an achievable quadruple w.r.t. P' by definition.

References

[1] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1041–1056, May 1998; see also, EE-1015, Dept. Elec. Eng., Tech. Rep. CC-141, Technion, Haifa, Israel, Feb. 1996.

[2] V. Koshélev, "Hierarchical coding of discrete sources," *Probl. Pered. Inform.*, vol. 16, no. 3, pp. 31–49, 1980.

[3] ——, "Estimation of mean error for a discrete successive approximation scheme," *Probl. Pered. Inform.*, vol. 17, no. 3, pp. 20–33, July–Sept. 1981.

[4] H. Yamamoto, "Source coding theory for cascade and branching communication systems," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 299–308, May 1981.

[5] W. H. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–275, Mar. 1991.

[6] B. Rimoldi, "Successive refinement of information: Characterization of achievable rates," *IEEE Trans. Inform. Theory*, vol. 40, pp. 253–259, Jan. 1994.

[7] A. Kanlis and P. Narayan, "Error exponents for successive refinement by partitioning," *IEEE Trans. Inform. Theory*, vol. 42, pp. 275–282, Jan. 1996.

[8] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.

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

Almost-Sure Variable-Length Source Coding Theorems for General Sources

Jun Muramatsu and Fumio Kanaya

***Abstract*—Source coding theorems for general sources are presented. For a source μ, which is assumed to be a probability measure on all strings of infinite-length sequence with a finite alphabet, the notion of almost-sure sup entropy rate is defined; it is an extension of the Shannon entropy rate. When both an encoder and a decoder know that a sequence is generated by μ, the following two theorems can be proved: 1) in the almost-sure sense, there is no variable-rate source coding scheme whose coding rate is less than the almost-sure sup entropy rate of μ. and 2) in the almost-sure sense, there exists a variable-rate source coding scheme whose coding rate achieves the almost-sure sup entropy rate of μ.**

***Index Terms*—Almost-sure sup entropy rate, general sources, source coding theorems.**

I. Introduction

Throughout this correspondence, let $\hat{A}$ be a finite set and $(\hat{\mathcal{A}}^\infty, \mathcal{F})$ a measurable space, where $\hat{\mathcal{A}}^\infty$ is the set of all strings of infinite length that can be formed from the symbols in $\hat{\mathcal{A}}$, and $\mathcal{F}$ is a σ-field of subsets of $\hat{\mathcal{A}}^\infty$.

Let μ be a probability measure defined on $(\hat{\mathcal{A}}, \mathcal{F})$. Then, we call $(\hat{\mathcal{A}}, \mathcal{F}, \mu)$ a probability space. We call μ a *general source* or simply a *source*. It should be noted that μ satisfies consistency restrictions. Traditionally, a source is defined as a sequence of random variables $\hat{X} \equiv \{\hat{X}n\}_{n=1}^\infty$, but if $\hat{X}$ satisfies consistency restrictions

$$\sum_{\hat{x}^{n+1} \in \hat{A}} \mathrm{Prob}\,(\hat{X}^{n+1} = \hat{x}^{n+1}) = \mathrm{Prob}\,(\hat{X}^n = \hat{x}^n), \qquad \forall \hat{x}^n \in \hat{A}^n, \quad \forall n \in \mathbb{N}$$

we can construct the probability measure $\mu_{\hat{X}}$ satisfying

$$\mu_{\hat{X}}^n(\hat{x}^n) \equiv \mathrm{Prob}\,(\hat{X}^n = \hat{x}^n)$$

where $\mu_{\hat{X}}^n$ is a probability distribution on $\hat{\mathcal{A}}^n$ induced by $\mu_{\hat{X}}$. Then, $\mu_{\hat{X}}$ can be considered as a general source.

We will prove almost-sure source coding theorems for general sources, placing no assumption on sources except consistency restrictions. To this end, we define the almost-sure sup entropy rate of a general source μ. Assuming that an encoder and a decoder know that a string is produced by μ, we can make the following two statements:

1) There is no variable-length code such that the coding rate of this code is less than the almost-sure sup entropy rate of the source with probability 1.
2) There exists a variable-length code such that the coding rate of this code is equal to the almost-sure sup entropy rate of the source with probability 1.

Manuscript received July 17, 1997; revised April 26, 1998. The material in this correspondence was presented in part at the 19th Symposium on Information Theory and Its Applications, Hakone, Japan, December 3–6, 1996 (in Japanese).

J. Muramatsu is with the NTT Communication Science Laboratories, Kyoto 619-0237, Japan.

F. Kanaya is with the Department of Information Science, Shonan Institute of Technology, Kanagawa 251-0046, Japan.

Communicated by M. Feder, Associate Editor for Source Coding.

Publisher Item Identifier S 0018-9448(99)00625-2.

Historically, various notions of entropy have been defined under different assumptions on the source; by using these notions, coding theorems have been proved for various types of codes and for different definitions of a coding rate. We survey some existing notions of entropy and discuss them vis-à-vis our newly defined notion of the almost-sure sup entropy rate.

We begin by noting that for noiseless coding of sources, various situations occur depending on the definition of the coding rate as well as the type of code.

Fixed-Length Code: For any $\varepsilon > 0, \gamma > 0$ and sufficiently large n, there exists a set $\{\hat{x}_1^n, \hat{x}_2^n, \cdots, \hat{x}_M^n\} \subset \hat{\mathcal{A}}^n$ such that

$$\frac{1}{n} \log_2 M < R + \gamma$$

$$\mu(\{\hat{x} \in \hat{\mathcal{A}}^\infty; \quad \hat{x}^n \notin \{\hat{x}_1^n, \hat{x}_2^n, \cdots, \hat{x}_M^n\}\}) \leq \varepsilon.$$

Variable-Length Code with Average Coding Rate Defined: There exists a variable-length code $\{(\varphi_n, \varphi_n^{-1})\}_{n=1}^\infty$ such that

$$\limsup_{n\to\infty} E_{\mu^n}\left[\frac{1}{n}\,\ell(\varphi_n(\hat{x}^n))\right] \leq R.$$

Variable-Length Code with Coding Rate in Probability Defined: There exists a variable-length code $\{(\varphi_n, \varphi_n^{-1})\}_{n=1}^\infty$ such that for any $\varepsilon > 0$

$$\lim_{n\to\infty} \mu^n\left(\left\{\hat{x} \in \hat{\mathcal{A}}^n; \frac{1}{n}\,\ell(\varphi_n(\hat{x}^n)) < R + \varepsilon\right\}\right) = 1.$$

Variable-Length Code with Almost-Sure Coding Rate Defined: There exists a variable-length code $\{(\varphi_n, \varphi_n^{-1})\}_{n=1}^\infty$ such that

$$\limsup_{n\to\infty} \frac{1}{n}\,\ell(\varphi_n(\hat{x}^n)) \leq R, \quad \mu\text{-a.s.}$$

that is,

$$\mu\left(\left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \limsup_{n\to\infty} \frac{1}{n}\,\ell(\varphi_n(\hat{x}^n)) \leq R\right\}\right) = 1.$$

It should be noted that from a fixed-length code of rate level R, we can always construct a variable-length code whose rate converges to R in probability.

Coding theorems for a stationary ergodic source are proved via the asymptotic equipartition property (AEP), which assures the convergence

$$\frac{1}{n} \log_2 \frac{1}{\mu^n(\hat{x}^n)} \to H_\mu^*.$$

Here, H_μ^* denotes entropy rate, which has various definitions as follows. In addition, it should be noted that senses of convergence are also different.

First of all, Shannon [11] defined the entropy of an independent and identically distributed (i.i.d.) source by

$$H_\mu^{\mathrm{S}} \equiv \sum_{\hat{x}^1 \in \hat{\mathcal{A}}^1} \mu^1(\hat{x}^1) \log_2 \frac{1}{\mu^1(\hat{x}^1)}.$$

He proved the AEP in the sense of convergence in probability and showed that H_μ^{S} is the minimum coding rate for any i.i.d. source μ.

For stationary ergodic sources μ, McMillan [9] defined the entropy rate by

$$H_\mu^{\mathrm{SMB}} \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{\hat{x}^n \in \hat{\mathcal{A}}^n} \mu^n(\hat{x}^n) \log_2 \frac{1}{\mu^n(\hat{x}^n)}.$$

He showed the AEP in mean convergence. Breiman [2] extended the above result to show that μ satisfies the AEP in almost-sure convergence to H_μ^{SMB}.

For general sources, Verdú and Han [12] proved that AEP is equivalent to fixed-length source coding theorems for nonzero-entropy finite-alphabet sources.

Since almost-sure convergence implies convergence in probability and also convergence of the average under a certain condition of boundedness, each coding rate in the above-mentioned coding situations is equal to H_μ^{SMB} for any stationary ergodic source. However, when a source does not satisfy ergodicity, coding rates are not necessarily the same. Furthermore, coding rates do not always converge. To deal with these situations, different notions of entropy are studied and used to prove the coding theorems.

Kieffer [8] considered a variable-length code with the average coding rate defined. He defined the entropy rate H_μ^{K} of a general source μ by

$$\overline{H}_\mu^{\mathrm{K}} \equiv \limsup_{n\to\infty} \frac{1}{n} \sum_{\hat{x}^n \in \hat{\mathcal{A}}^n} \mu^n(\hat{x}^n) \log_2 \frac{1}{\mu^n(\hat{x}^n)}$$

and showed that $\overline{H}_\mu^{\mathrm{K}}$ is the optimum average coding rate. Clearly, $\overline{H}_\mu^{\mathrm{K}} = H_\mu^{\mathrm{SMB}}$ if μ is stationary and ergodic.

Han and Verdú [7] considered fixed-length coding of general sources and defined the sup entropy rate $\overline{H}_\mu^{\mathrm{HV}}$ of a general source μ by

$$\overline{H}_\mu^{\mathrm{HV}} \equiv \inf\left\{h; \lim_{n\to\infty} \mu^n\left(\left\{\hat{x}^n \in \hat{\mathcal{A}}^n; \frac{1}{n} \log_2 \frac{1}{\mu^n(\hat{x}^n)} > h\right\}\right) = 0\right\}$$

where the random variable

$$\frac{1}{n} \log_2 \frac{1}{\mu_{\hat{X}}^n(\hat{X}n)}$$

is called the *entropy density rate* and its distribution is called the *information spectrum of the source* X (cf. [6]). They proved that $\overline{H}_\mu^{\mathrm{HV}}$ is the optimal coding rate of the source μ. If μ is stationary ergodic, then $\overline{H}_\mu^{\mathrm{HV}} = H_\mu^{\mathrm{SMB}}$.

Remark 1: Since a source was considered to be a general sequence of random variables, Han and Verdú did not necessarily assume that a source satisfies the consistency restriction. On the other hand, to define the almost-sure sup entropy rate we assume *a priori* knowledge for the probability distribution on $\hat{\mathcal{A}}^\infty$, so our source satisfies the consistency restriction. In addition, it should be noticed that we assume a finite source alphabet, which was also not necessarily assumed by them.

Our new definition of the almost-sure sup entropy rate is defined to deal with the *worst case* optimal coding rate of the variable-length code with probability 1. We prove that it satisfies source coding theorems in the almost-sure sense. This provides yet another extension of the Shannon–McMillan–Breiman entropy rate, but it should be remarked that our definition is strictly different from both the Kieffer entropy rate and the Han–Verdú sup entropy rate. We discuss the matter in Section V.

II. Source Coding Theorems

Let μ be a general source. We define the sup μ-complexity rate of an element of $\hat{\mathcal{A}}^\infty$.

Definition 1: A function $\overline{h}_\mu\colon \hat{\mathcal{A}}^\infty \to [0, \infty)$ defined by

$$\overline{h}_\mu(\hat{x}) \equiv \limsup_{n\to\infty} \frac{1}{n} \log_2 \frac{1}{\mu^n(\hat{x}^n)}, \qquad \hat{x} \in \hat{\mathcal{A}}^\infty$$

is called the *sup* μ*-complexity rate function*. We call $\overline{h}_\mu(\hat{x})$ the *sup* μ-complexity rate of $\hat{x} \in \hat{\mathcal{A}}^\infty$. It should be noted that $\overline{h}_\mu$ is a measurable function on $(\hat{\mathcal{A}}^n, \mathcal{F}, \mu)$. Intuitively, the sup μ-complexity rate of a sequence with infinite length is the amount of information per source symbol that we need to faithfully describe the sequence; it is assumed that we know *a priori* the probability distribution μ of the space for all sequences with infinite length.

Next is the definition of the essential sup, which is familiar in measure theory.

Definition 2: Let f be a measurable function on a measurable space $(\hat{\mathcal{A}}^n, \mathcal{F}, \mu)$. Then, we define μ-ess. sup f by

$$\mu\text{-ess. sup } f \equiv \inf\{\alpha; f(x) \le \alpha, \mu\text{-a.s.}\}$$

and call it the *essential sup* of f under μ.

We will now define the almost-sure sup entropy rate of μ.

Definition 3: The *almost-sure sup entropy rate* $\overline{H}_\mu$ of μ is defined by

$$\overline{H}_\mu \equiv \mu\text{-ess. sup } \overline{h}_\mu.$$

We define the variable-length noiseless code and its coding rate. Let $\mathcal{B} \equiv \{0, 1\}$ and $\mathcal{B}^* \equiv \cup_{n=1}^\infty \mathcal{B}^n$. Let ℓ: $\mathcal{B}^* \to \mathbb{N}$ be the length function of finite binary strings. Hereafter, we use notation $a * b$ to denote concatenation of strings a and b.

Definition 4: Given two strings $b_1, b_2 \in \mathcal{B}^*$, we say b_1 is a *prefix* of b_2 if there is a string $b_3 \in \mathcal{B}^*$ such that $b_2 = b_1 * b_3$. A subset $\mathcal{B}_{\mathrm{p}}^*$ of $\mathcal{B}^*$ is said to be a *prefix set* if $b_1 = b_2$ whenever b_1, b_2 are members of $\mathcal{B}_{\mathrm{p}}^*$ and b_1 is a prefix of b_2.

Definition 5: A sequence $\{(\varphi_n, \varphi_n^{-1})\}_{n=1}^\infty$ of pairs of functions is called a *variable-length noiseless code* if φ_n: $\hat{\mathcal{A}}^n \to \mathcal{B}^*$ is injective, $\varphi_n(\hat{\mathcal{A}}^n)$ is a prefix set, and φ_n^{-1}: $\varphi_n(\hat{\mathcal{A}}^n) \to \hat{\mathcal{A}}^n$ is defined to satisfy

$$\varphi_n^{-1}(\varphi_n(\hat{x}^n)) = \hat{x}^n, \qquad \forall \hat{x}^n \in \hat{\mathcal{A}}^n$$

for each $n \in \mathbb{N}$.

Let $\varphi \equiv \{(\varphi_n, \varphi_n^{-1})\}_{n=1}^\infty$ be a variable-length noiseless code. We define the almost-sure coding rate of the source μ by the code φ as follows.

Definition 6: We define $\overline{r}_\varphi(\hat{x})$ by

$$\overline{r}_\varphi(\hat{x}) \equiv \limsup_{n\to\infty} \frac{1}{n} \ell(\varphi_n(\hat{x}^n))$$

and call it the *coding rate of* $\hat{x} \in \hat{\mathcal{A}}^\infty$ by φ. We define $\overline{R}_\varphi$ by

$$\overline{R}_{\varphi,\mu} \equiv \mu\text{-ess. sup } \overline{r}_\varphi$$

and call it the *almost-sure coding rate of* μ by φ.

The following theorems tell us that the almost-sure sup entropy rate is the optimal coding rate of a variable-length noiseless code.

Theorem 1: (Converse) If φ is a variable-length noiseless code, then

$$\overline{r}_{\varphi,\mu}(\hat{x}) \ge \overline{h}_\mu(\hat{x}), \quad \mu\text{-a.s.} \tag{1}$$

and

$$\overline{R}_{\varphi,\mu} \ge \overline{H}_\mu.$$

Theorem 2: (Achievability) There exists a variable-length noiseless code φ_μ such that

$$\overline{r}_{\varphi,\mu}(\hat{x}) = \overline{h}_\mu(\hat{x}), \quad \mu\text{-a.s.} \tag{2}$$

and

$$\overline{R}_{\varphi_\mu,\mu} = \overline{H}_\mu.$$

Their proofs will follow.

III. Proof of the Converse Theorem

To prove Theorem 1, we prepare the following lemmas.

Lemma 3 ([3, Theorem 5.2.1]): If $\mathcal{B}_{\mathrm{p}}^*$ is a prefix set, then the Kraft inequality

$$\sum_{b^* \in \mathcal{B}_{\mathrm{p}}^*} 2^{-\ell(b^*)} \le 1$$

is satisfied.

Lemma 4 ([1, Theorem 3.1]): Let μ be a general source and $\{(\varphi_n, \varphi_n^{-1})\}_{n=1}^\infty$ a variable-length noiseless code. Then

$$\ell(\varphi_n(\hat{x}^n)) \ge \log_2 \frac{1}{\mu^n(\hat{x}^n)} - \log_2 n - 2 \log_2 \log_2 n, \quad \mu\text{-a.s.}$$

Lemma 5: Let f and g be measurable functions on the same probability space $(\hat{\mathcal{A}}^\infty, \mathcal{F}, \mu)$ such that

$$f(x) \ge g(x), \quad \mu\text{-a.s.}$$

Then

$$\mu\text{-ess. sup } f \ge \mu\text{-ess. sup } g.$$

Proof: Assume that

$$f(x) \le \alpha, \quad \mu\text{-a.s.} \tag{3}$$

Since

$$\begin{aligned} g(x) &\le f(x) \\ &\le \alpha, \quad \mu\text{-a.s.} \end{aligned}$$

we have

$$\mu\text{-ess. sup } g \le \alpha$$

by the definition of μ-ess. sup g. We can take an arbitrary α under condition (3). Therefore, we have

$$\mu\text{-ess. sup } g \le \mu\text{-ess. sup } f$$

by the definition of μ-ess. sup f. □

We now prove the converse theorem.

Proof: (Proof of Theorem 1) By Lemmas 3 and 4

$$\begin{aligned} \overline{r}_\varphi(\hat{x}) &= \limsup_{n\to\infty} \frac{1}{n} \ell(\varphi_n(\hat{x}^n)) \\ &\ge \limsup_{n\to\infty} \frac{1}{n} \log_2 \frac{1}{\mu^n(\hat{x}^n)} \\ &= \overline{h}_\mu(\hat{x}), \quad \mu\text{-a.s.} \end{aligned}$$

Therefore, we have (1) and

$$\begin{aligned} \overline{R}_{\varphi,\mu} &= \mu\text{-ess. sup}\overline{r}_\varphi \\ &\ge \mu\text{-ess. sup}\overline{h}_\mu \\ &= \overline{H}_\mu \end{aligned}$$

by Lemma 5. □

IV. Proof of the Achievability Theorem

To prove Theorem 2, we prepare the following lemma, which assures a nice coding property of the prefix set.

Lemma 6 ([3, Theorem 5.2.1]): Let $\hat{\mathcal{A}}_{\rm p}^n \subset \hat{\mathcal{A}}^n$ and assume that L: $\hat{\mathcal{A}}_{\rm p}^n \to \mathbb{N}$ satisfies the Kraft inequality

$$\sum_{\hat{x}^n \in \hat{\mathcal{A}}_{\rm p}^n} 2^{-L(\hat{x}^n)} \le 1. \tag{4}$$

Then, there exists a prefix set $\mathcal{B}_{\rm p}^*$ and a bijection $\varphi_n:\hat{\mathcal{A}}_{\rm p}^n \to \mathcal{B}_{\rm p}^*$ such that $\ell(\varphi_n(\hat{x}^n)) = L(\hat{x}^n)$ for any $\hat{x}^n \in \hat{\mathcal{A}}_{\rm p}^n$.

We now prove the existence theorem.

Proof: (Proof of Theorem 2) Let $L:\hat{\mathcal{A}}^n \to \mathbb{N}$ be defined by

$$L_\mu(\hat{x}^n) \equiv \begin{cases} 1 + \left\lceil \log_2 \dfrac{1}{\mu^n(\hat{x}^n)} \right\rceil, & \text{if } \mu^n(\hat{x}^n) > 0 \\ 1 + \lceil \log_2 |\hat{\mathcal{A}}^n| \rceil, & \text{if } \mu^n(\hat{x}^n) = 0 \end{cases}$$

for $\hat{x}^n \in \hat{\mathcal{A}}^n$. Since L satisfies (4), it follows from Lemma 6 that we can construct a variable-length noiseless code

$$\varphi_\mu = \{(\varphi_{\mu,n}, \varphi_{\mu,n}^{-1})\}_{n=1}^\infty$$

such that $\ell(\varphi_{\mu,n}(\hat{x}^n)) = L_\mu(\hat{x}^n)$ for any $\hat{x}^n \in \hat{\mathcal{A}}^n$.

On the other hand, since μ satisfies the consistency restriction

$$\begin{aligned} &\mu(\{\hat{x} \in \hat{\mathcal{A}}^\infty; \mu^n(\hat{x}^n) > 0, \quad \forall n \in \mathbb{N}\}) \\ &\quad = 1 - \mu(\{\hat{x} \in \hat{\mathcal{A}}^\infty; \exists n \in \mathbb{N} \text{ s.t. } \mu^n(\hat{x}^n) = 0\}) \\ &\quad = 1 - \mu\left(\bigcup_{n=1}^\infty \{\hat{x} \in \hat{\mathcal{A}}^\infty; \mu^n(\hat{x}^n) = 0\}\right) \\ &\quad \ge 1 - \sum_{n=1}^\infty \mu(\{\hat{x} \in \hat{\mathcal{A}}^\infty; \mu^n(\hat{x}^n) = 0\}) \\ &\quad = 1 - \sum_{n=1}^\infty \mu^n(\{\hat{x}^n \in \hat{\mathcal{A}}^n; \mu^n(\hat{x}^n) = 0\}) \\ &\quad = 1 \end{aligned}$$

and we have

$$L_\mu(\hat{x}^n) = 1 + \left\lceil \log_2 \frac{1}{\mu^n(\hat{x}^n)} \right\rceil, \quad \mu\text{-a.s.}$$

Hence

$$\begin{aligned} \overline{r}_{\varphi_\mu}(\hat{x}) &= \limsup_{n\to\infty} \frac{1}{n} \ell(\varphi_{\mu,n}(\hat{x}^n)) \\ &= \limsup_{n\to\infty} \frac{1}{n} \left[1 + \left\lceil \log_2 \frac{1}{\mu^n(\hat{x}^n)} \right\rceil\right] \\ &= \limsup_{n\to\infty} \frac{1}{n} \log_2 \frac{1}{\mu^n(\hat{x}^n)} \\ &= \overline{h}_\mu(\hat{x}), \quad \mu\text{-a.s.} \end{aligned}$$

and we have (2). This gives us

$$\begin{aligned} \overline{R}_{\varphi_\mu,\mu} &= \mu\text{-ess. sup } \overline{r}_{\varphi_\mu} \\ &= \mu\text{-ess. sup } \overline{h}_\mu \\ &= \overline{H}_\mu. \end{aligned}$$

□

Remark 2: The code constructed in the proof of the theorem is based on a probability distribution. In the event that

$$\limsup_{n\to\infty} \frac{1}{n} L_\mu(\hat{x}^n) > \log_2 |\hat{\mathcal{A}}|$$

has positive probability, then the code might be worse than the code formed by a simple binary representation of the alphabet. However, the following theorem assures us that the constructed code is no worse than a simple binary encoding of the alphabet.

Theorem 7: Let μ be a general source. Then

$$0 \le \overline{H}_\mu \le \log_2 |\hat{\mathcal{A}}|.$$

Proof: First, we prove $0 \le \overline{H}_\mu$ by contradiction. Assume that $\overline{H}_\mu < 0$. Then,

$$\overline{h}_\mu(\hat{x}) < 0, \quad \mu\text{-a.s.} \tag{5}$$

On the other hand, for all $\hat{x} \in \hat{\mathcal{A}}^\infty$, we have $\mu(\hat{x}^n) \le 1$ for every $n \ge 1$, so $\overline{h}_\mu(\hat{x}) \ge 0$. Since this contradicts (5), we must have $0 \le \overline{H}_\mu$.

Next, we prove $\overline{H}_\mu \le \log_2 |\hat{\mathcal{A}}|$. Let $\varepsilon > 0$. Then for every $n \ge 1$

$$\begin{aligned} &\mu\left(\left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{n} \log_2 \frac{1}{\mu^n(\hat{x}^n)} \ge \log_2 |\hat{\mathcal{A}}| + \varepsilon\right\}\right) \\ &\quad = \mu^n\left(\left\{\hat{x}^n \in \hat{\mathcal{A}}^n; \mu^n(\hat{x}^n) \le 2^{-n[\log_2 |\hat{\mathcal{A}}|+\varepsilon]}\right\}\right) \\ &\quad \le 2^{-n[\log_2 |\hat{\mathcal{A}}|+\varepsilon]} \cdot |\hat{\mathcal{A}}|^n \\ &\quad = 2^{-n\varepsilon}. \end{aligned}$$

Hence

$$\sum_{n=1}^\infty \mu\left(\left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{n} \log_2 \frac{1}{\mu^n(\hat{x}^n)} \ge \log_2 |\hat{\mathcal{A}}| + \varepsilon\right\}\right) < \infty.$$

By the Borel-Cantelli Lemma (cf. [4, Theorem 1.2]) we have

$$\begin{aligned} &\mu\left(\bigcup_{n=1}^\infty \bigcap_{m=n}^\infty \left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{m} \log_2 \frac{1}{\mu^m(\hat{x}^n)} < \log_2 |\hat{\mathcal{A}}| + \varepsilon\right\}\right) \\ &\quad = 1 - \mu\left(\bigcap_{n=1}^\infty \bigcup_{m=n}^\infty \left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{m} \log_2 \frac{1}{\mu^m(\hat{x}^n)} \ge \log_2 |\hat{\mathcal{A}}| + \varepsilon\right\}\right) \\ &\quad = 1. \end{aligned}$$

It follows that there exists $N_{\hat{x},\varepsilon} \in \mathbb{N}$ for μ-a.s.$\hat{x}$ such that if $n > N_{\hat{x},\varepsilon}$

$$\frac{1}{n} \log_2 \frac{1}{\mu^n(\hat{x}^n)} < \log_2 |\hat{\mathcal{A}}| + \varepsilon$$

which implies

$$\limsup_{n\to\infty} \frac{1}{n} \log_2 \frac{1}{\mu^n(\hat{x}^n)} \le \log_2 |\hat{\mathcal{A}}| + \varepsilon, \quad \mu\text{-a.s.}$$

Consequently, $\overline{H}_\mu \le \log_2 |\hat{\mathcal{A}}| + \varepsilon$ by the definition of $\overline{H}_\mu$. Since $\varepsilon > 0$ is arbitrary, it is proved that $\overline{H}_\mu \le \log_2 |\hat{\mathcal{A}}|$. □

V. Relations with Other Notions of Entropy

In this section, we discuss the relations between our new definition of the almost-sure sup entropy rate and other notions of entropy. We denote $\overline{H}_\mu^{\rm MK}$ as the almost-sure sup entropy rate defined in Section II.

The next theorem immediately follows from the AEP for a stationary ergodic source. As we have already seen a similar relation holds for $\overline{H}_\mu^{\rm K}$ and $\overline{H}_\mu^{\rm HV}$.

Theorem 8: Let μ be a stationary ergodic source. Then

$$\overline{H}_\mu^{\rm MK} = H_\mu^{\rm SMB}.$$

We now compare $\overline{H}_\mu^{\rm MK}$ with $\overline{H}_\mu^{\rm K}$. Kieffer considered the average coding rate, whereas we consider the worst case coding rate. The following theorem shows the relation among $\overline{H}_\mu^{\rm MK}, \overline{H}_\mu^{\rm K}$ and the average of the μ-complexity rate $\overline{h}_\mu$.

Theorem 9: Let μ be a general source. Then

$$\overline{H}_\mu^{\rm K} \le E_\mu[\overline{h}_\mu(\hat{x})] \le \overline{H}_\mu^{\rm MK}.$$

Proof: The left-hand side inequality follows from Fatou's lemma. The right-hand side inequality follows from the definition of $\overline{H}_\mu^{\rm MK}$. $\square$

Next, we compare $\overline{H}_\mu^{\rm MK}$ with $\overline{H}_\mu^{\rm HV}$, which both deal with the worst case coding rate. However, the former considers variable-rate coding in an "almost-sure" sense, while the latter considers fixed-rate coding in an "in-probability" sense. From the definitions, $\overline{H}_\mu^{\rm MK}$ and $\overline{H}_\mu^{\rm HV}$ can be transformed to

$$\overline{H}_\mu^{\rm HV} = \inf\left\{h; \lim_{n\to\infty} \mu\left(\left\{\hat{x}^n \in \hat{\mathcal{A}}^\infty; \frac{1}{n}\log_2 \frac{1}{\mu^n(\hat{x}^n)} \le h\right\}\right) = 1\right\}$$

and

$$\overline{H}_\mu^{\rm MK} = \inf\left\{h; \mu\left(\left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \lim_{n\to\infty}\frac{1}{n}\log_2 \frac{1}{\mu^n(\hat{x}^n)} \le h\right\}\right) = 1\right\}$$

respectively. This is the definite difference between $\overline{H}_\mu^{\rm MK}$ and $\overline{H}_\mu^{\rm HV}$. Our definition of $\overline{H}_\mu^{\rm MK}$ is based on the sup μ-complexity rate, which is the amount of information needed per symbol to faithfully describe any specific source string when we know beforehand the source μ. The definition of $\overline{H}_\mu^{\rm HV}$ is based on the notion of the entropy density rate, which is a random variable. At this point, it is worthwhile to notice the following difference between coding in an "in-probability" sense and coding in an "almost-sure" sense. Let $\{\mathcal{C}_n\}_{n=1}^\infty$ be a fixed-length code of coding rate $R \ge \overline{H}_\mu^{\rm HV}$ for a general source μ. Then, whether or not any given output string $\hat{x}$ from the source is faithfully coded by the code $\{\mathcal{C}_n\}_{n=1}^\infty$ remains unknown until the coding of $\hat{x}$ is completed. On the other hand, let $\{(\varphi_n, \varphi_n^{-1})\}_{n=1}^\infty$ be a variable-length code of almost-sure coding rate $R \ge \overline{H}_\mu^{\rm MK}$ for the source μ. For any given output string $\hat{x}$ from the source, one is guaranteed beforehand that it will faithfully be encoded by the code $\{(\varphi_n, \varphi_n^{-1})\}_{n=1}^\infty$ in the long run.

Next, we show the relation between $\overline{H}_\mu^{\rm HV}$ and $\overline{H}_\mu^{\rm MK}$.

Theorem 10: Let μ be a general source. Then

$$\overline{H}_\mu^{\rm HV} \le \overline{H}_\mu^{\rm MK}.$$

Proof: Let $\varepsilon > 0$. From the definition of $\overline{H}_\mu^{\rm MK}$

$$\limsup_{n\to\infty} \frac{1}{n}\log_2 \frac{1}{\mu^n(\hat{x})} \le \overline{H}_\mu^{\rm MK} + \varepsilon, \quad \mu\text{-a.s.}$$

so we have

$$\mu\left(\bigcap_{k=1}^\infty \bigcup_{n=1}^\infty \bigcap_{m=n}^\infty \left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{m}\log_2 \frac{1}{\mu^m(\hat{x})} \le \overline{H}_\mu^{\rm MK} + \varepsilon + \frac{1}{k}\right\}\right) = 1. \tag{6}$$

On the other hand, since

$$\bigcup_{n=1}^\infty \bigcap_{m=n}^\infty \left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{m}\log_2 \frac{1}{\mu^m(\hat{x})} \le \overline{H}_\mu^{\rm MK} + 2\varepsilon\right\} \supset \bigcap_{k=1}^\infty \bigcup_{n=1}^\infty \bigcap_{m=n}^\infty \left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{m}\log_2 \frac{1}{\mu^m(\hat{x})} \le \overline{H}_\mu^{\rm MK} + \varepsilon + \frac{1}{k}\right\}$$

we have from (6)

$$\mu\left(\bigcup_{n=1}^\infty \bigcap_{m=n}^\infty \left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{m}\log_2 \frac{1}{\mu^m(\hat{x})} \le \overline{H}_\mu^{\rm MK} + 2\varepsilon\right\}\right) = 1$$

which is equivalent to

$$\mu\left(\bigcap_{n=1}^\infty \bigcup_{m=n}^\infty \left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{m}\log_2 \frac{1}{\mu^m(\hat{x})} > \overline{H}_\mu^{\rm MK} + 2\varepsilon\right\}\right) = 0. \tag{7}$$

Since for any $n \in \mathbb{N}$

$$\left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{n}\log_2 \frac{1}{\mu^n(\hat{x})} > \overline{H}_\mu^{\rm MK} + 2\varepsilon\right\} \subset \bigcup_{m=n}^\infty \left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{m}\log_2 \frac{1}{\mu^m(\hat{x})} > \overline{H}_\mu^{\rm MK} + 2\varepsilon\right\}$$

we have from (7) and [5, Lemma 4.6.3]

$$\begin{aligned}\lim_{n\to\infty} \mu\left(\left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{n}\log_2 \frac{1}{\mu^n(\hat{x})} > \overline{H}_\mu^{\rm MK} + 2\varepsilon\right\}\right) &\le \lim_{n\to\infty} \mu\left(\bigcup_{m=n}^\infty \left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{m}\log_2 \frac{1}{\mu^m(\hat{x})} > \overline{H}_\mu^{\rm MK} + 2\varepsilon\right\}\right)\\ &= 0\end{aligned}$$

implying that

$$\lim_{n\to\infty} \mu\left(\left\{\hat{x} \in \hat{\mathcal{A}}^\infty; \frac{1}{n}\log_2 \frac{1}{\mu^n(\hat{x})} > \overline{H}_\mu^{\rm MK} + 2\varepsilon\right\}\right) = 0.$$

Thus it follows from the definition of $\overline{H}_\mu^{\rm HV}$ that

$$\overline{H}_\mu^{\rm HV} \le \overline{H}_\mu^{\rm MK} + 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we have

$$\overline{H}_\mu^{\rm HV} \le \overline{H}_\mu^{\rm MK}.$$ $\square$

Then, combining Theorems 7 and 10 with the known fact that $0 \le \overline{H}_\mu^{\rm K} \le \overline{H}_\mu^{\rm HV}$ (cf. [6]), we have the following theorem.

Theorem 11: Let μ be a general source. Then

$$0 \le \overline{H}_\mu^{\rm K} \le \overline{H}_\mu^{\rm HV} \le \overline{H}_\mu^{\rm MK} \le \log_2 |\hat{\mathcal{A}}|.$$

Han [6] presented a source such that $\overline{H}_\mu^{\rm K} < \overline{H}_\mu^{\rm HV}$, and this fact shows that there is a rigorous difference between $\overline{H}_\mu^{\rm K}$ and $\overline{H}_\mu^{\rm HV}$. At this point, one may also be interested in the problem of whether there exists a strict difference between $\overline{H}_\mu^{\rm HV}$ and $\overline{H}_\mu^{\rm MK}$. The next theorem gives the answer.

Theorem 12: There exists a source μ such that $\overline{H}_\mu^{\rm HV} < \overline{H}_\mu^{\rm MK}$.

Proof: The following proof, in which we construct the source μ satisfying $\overline{H}_\mu^{\rm HV} < \overline{H}_\mu^{\rm MK}$, is due to [10].

Let $\{s_k\}_{k=1}^\infty$ be the strictly increasing sequence of positive integers and $\{p_k\}_{k=1}^\infty$ the sequence such that $0 < p_k < 1$ for each $k \in \mathbb{N}$. We assume that $\{p_k\}_{k=1}^\infty$ and $\{s_k\}_{k=1}^\infty$ have the following limits:

$$\lim_{k\to\infty} \frac{s_{k-1}}{s_k} = 0 \tag{8}$$

$$\lim_{k\to\infty} p_k = 0 \tag{9}$$

$$\sum_{k=1}^\infty \frac{p_k[|\hat{\mathcal{A}}|^{t_k} - 1]}{|\hat{\mathcal{A}}|^{t_k}} = \infty \tag{10}$$

where $s_0 \equiv 0$ and $t_k \equiv s_k - s_{k-1}$. For example,

$$s_k \equiv k!$$

$$p_k \equiv \frac{|\hat{\mathcal{A}}|^{t_k}}{4k[|\hat{\mathcal{A}}|^{t_k} - 1]}$$

satisfy the above conditions.

$$\mu^n(\hat{x}^n) \equiv \left[\prod_{i=1}^{k(n)} \nu_i(\hat{x}_{s_{i-1}+1}^{s_i})\right] \cdot \left[\sum_{\hat{y}^{s_{k(n)+1}-n} \in \hat{\mathcal{A}}^{s_{k(n)+1}-n}} \nu_{k(n)+1}(\hat{x}_{s_{k(n)}+1}^{n} * \hat{y}^{s_{k(n)+1}-n})\right]$$

Let ν_k be the probability distribution on $\hat{\mathcal{A}}^{t_k}$ defined by

$$\nu_k(\hat{x}t_k) \equiv \begin{cases} 1 - p_k + \dfrac{p_k}{|\hat{\mathcal{A}}|^{t_k}}, & \text{if } \hat{x}^{t^k} = \hat{a}^{t_k} \\ \dfrac{p_k}{|\hat{\mathcal{A}}|^{t_k}}, & \text{if } \hat{x}^{t^k} \neq \hat{a}^{t_k} \end{cases}$$

where the symbol $\hat{a} \in \hat{\mathcal{A}}$ is given in advance. We define μ^n by the expression at the bottom of this page for each $\hat{x}^n \in \hat{\mathcal{A}}^n$, where $k(n) \equiv \max\{k; s_k \leq n\}$ and$\hat{x}_i^j \equiv (\hat{x}_i, \cdots, \hat{x}_j)$. It is easily verified that $\{\mu^n\}_{n=1}^{\infty}$ satisfies the consistency restrictions. Now we can construct the general source from $\{\mu^n\}_{n=1}^{\infty}$.

First, we prove that $\overline{H}_{\mu}^{\mathrm{HV}} = 0$. Toward it let T_n be the subset of $\hat{\mathcal{A}}^n$ defined by

$$T_n \equiv \{\hat{x}^n; \hat{x}_{s_{k(n)-1}+1}^{n} = \hat{a}^{n-s_{k(n)-1}}\}.$$

Then, the probability of T_n is

$$\begin{aligned} \mu^n(T_n) &= \left[1 - p_{k(n)} + \frac{p_{k(n)}}{|\hat{\mathcal{A}}|^{t_{k(n)}}}\right] \\ &\quad \cdot \left[1 - p_{k(n)+1} + \frac{p_{k(n)+1}}{|\hat{\mathcal{A}}|^{t_{k(n)+1}}} \right. \\ &\quad \left. + [|\hat{\mathcal{A}}|^{s_{k(n)+1}-n} - 1] \cdot \frac{p_{k(n)+1}}{|\hat{\mathcal{A}}|^{t_{k(n)+1}}}\right] \\ &= \left[1 - p_{k(n)} + \frac{p_{k(n)}}{|\hat{\mathcal{A}}|^{t_{k(n)}}}\right] \\ &\quad \cdot \left[1 - p_{k(n)+1} + \frac{p_{k(n)+1}}{|\hat{\mathcal{A}}|^{n-s_{k(n)}}}\right]. \end{aligned}$$

We can regard $\{T_n\}_{n=1}^{\infty}$ as the fixed-rate code. Error probability is evaluated by

$$\begin{aligned} 1 - \mu^n(T_n) &\leq p_{k(n)} - \frac{p_{k(n)}}{|\hat{\mathcal{A}}|^{t_{k(n)}}} + p_{k(n)+1} \frac{p_{k(n)+1}}{|\hat{\mathcal{A}}|^{n-s_{k(n)}}} \\ &\leq p_{k(n)} + p_{k(n)+1}. \end{aligned}$$

To prove the first inequality, we use the fact that $1 - \alpha\beta \leq [1 - \alpha] + [1 - \beta]$ if $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$. By (9), the error probability tends to zero as n goes to infinity. From the converse theorem for the fixed-rate coding of general sources (cf. [7]) and by using (8), we have

$$\begin{aligned} \overline{H}_{\mu}^{\mathrm{HV}} &\leq \limsup_{n\to\infty} \frac{1}{n} \log_2 |T_n| \\ &= \limsup_{n\to\infty} \frac{1}{n} \log_2 |\hat{\mathcal{A}}|^{s_{k(n)-1}} \\ &\leq \limsup_{n\to\infty} \frac{s_{k(n)-1}}{s_{k(n)}} \log_2 |\hat{\mathcal{A}}| \\ &= 0 \end{aligned}$$

which implies that $\overline{H}_{\mu}^{\mathrm{HV}} = 0$.

Next, we prove that $\overline{H}_{\mu}^{\mathrm{MK}} = \log_2 |\hat{\mathcal{A}}|$. Toward it let

$$A_k \equiv \{\hat{x} \in \hat{\mathcal{A}}^{\infty}; \hat{x}_{s_{k-1}+1}^{s_k} \neq \hat{a}^{t_k}\}.$$

It is evident from the definition of μ that $\{A_k\}_{k=1}^{\infty}$ is a set of independent events. By (10), we have

$$\sum_{k=1}^{\infty} \mu(A_k) = \sum_{k=1}^{\infty} p_k[|\hat{\mathcal{A}}|^{t_k} - 1]/|\hat{\mathcal{A}}|^{t_k} = \infty$$

and the Borel-Cantelli lemma (cf. [4, Theorem 1.2]) gives us

$$\mu\left(\bigcap_{i=1}^{\infty} \bigcup_{k=i}^{\infty} A_k\right) = 1.$$

This implies that for μ-a.s. $\hat{x}$ and for any n there exists $k_{\hat{x}} > k(n)$ such that $\hat{x} \in A_{k_x}$. Therefore, letting $n_{\hat{x}} \equiv s_{k_{\hat{x}}} \geq s_{k(n)+1} > n$, we have for μ-a.s. $\hat{x}$ and for infinitely many $n_{\hat{x}} > n$

$$\begin{aligned} \frac{1}{n_{\hat{x}}} \log_2 \frac{1}{\mu^{n_{\hat{x}}}(\hat{x}^{n_{\hat{x}}})} &> \frac{1}{n_{\hat{x}}} \log_2 \frac{1}{\mu(\hat{x}^{t_{k_{\hat{x}}}})} \\ &= \frac{1}{n_{\hat{x}}} \log_2 \frac{|\hat{\mathcal{A}}|^{t_{k_{\hat{x}}}}}{p_{k_{\hat{x}}}} \\ &> \frac{1}{n_{\hat{x}}} \log_2 |\hat{\mathcal{A}}|^{t_{k_{\hat{x}}}} \\ &= \frac{t_{k_{\hat{x}}}}{n_{\hat{x}}} \log_2 |\hat{\mathcal{A}}| \\ &> \frac{s_{k_{\hat{x}}} - s_{k_{\hat{x}}} - 1}{s_{k_{\hat{x}}}} \log_2 |\hat{\mathcal{A}}| \\ &= \left[1 - \frac{s_{k_{\hat{x}}} - 1}{s_{k_{\hat{x}}}}\right] \log_2 |\hat{\mathcal{A}}|. \end{aligned}$$

By using (8) we have

$$\limsup_{n\to\infty} \frac{1}{n} \log_2 \frac{1}{\mu^n(\hat{x}^n)} = \log_2 |\hat{\mathcal{A}}|, \quad \mu\text{-a.s.}$$

This implies that $\overline{H}_{\mu}^{\mathrm{MK}} = \log_2 |\hat{\mathcal{A}}|$. Therefore, it is concluded that μ satisfies $\overline{H}_{\mu}^{\mathrm{HV}} < \overline{H}_{\mu}^{\mathrm{MK}}$.

Acknowledgment

The authors wish to thank Prof. T. S. Han and Mr. S. Miyake for their helpful discussions and useful comments on this paper. Constructive comments and suggestions by anonymous reviewers and Associate Editor M. Feder have significantly improved the presentation of our results.

References

[1] A. R. Barron, "Logically smooth density estimation," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1985.
[2] L. Breiman, "The individual ergodic theorems of information theory," *Ann. Math. Stat.*, vol. 28, pp. 809–811, 1957.
[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[4] J. L. Doob, *Stochastic Processes*. New York: Wiley, 1953.
[5] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*. New York: Springer-Verlag, 1988.
[6] T. S. Han, *Information-Spectrum Methods in Information Theory* (Lecture Notes, Grad. School Inform. Syst., Univ. Electro-Commun., Japan), Mar. 1997
[7] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, pp. 752–772, May 1993.
[8] J. C. Kieffer, "Finite-state adaptive block to variable-length noiseless coding of a nonstationary information source," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1259–1263, Nov. 1989.
[9] B. McMillan, "The basic theorems of information theory," *Ann. Math. Stat.*, vol. 24, pp. 196–219, 1953.
[10] S. Miyake, private communication.
[11] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
[12] S. Verdú and T. S. Han, "The role of the asymptotic equipartition property in noiseless source coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 847–857, May 1997.