Journal of Mathematical Inequalities
Volume 16, Number 3 (2022), 1151–1167
doi:10.7153/jmi-2022-16-77

STRONG DEVIATION THEOREMS FOR GENERAL INFORMATION SOURCES

SHU CHEN, WEI-CAI PENG* AND ZHONG-ZHI WANG

(*Communicated by Z. S. Szewczak*)

Abstract. In this paper, we first introduce some new concepts of generalized likelihood ratio, upper/lower generalized divergence rate and upper/lower generalized relative entropy, as a measure of randomness to characterize the deviation between generalized information sources and memoryless (i.e., independent) sources. Then, by adopting pure analysis method on studying probability limit theory, a class of strong limit theorems and strong deviation theorems for generalized information sources and generalized information source entropy density are established. The outcomes extend some existing results of [10] and [21].

1. Introduction

Since [17] established the pioneering work in information theory field, many scholars have achieved a series of in-depth and fruitful research on the theoretical and application basis of information theory, and achieved rich results. Many studies, such as [7], [9], devoted to studying the more general and abstract mathematical model of the axiomatic system of information theory, and obtained more general results to the basic theory of information. In the United States, a group of excellent engineers and technicians are committed to the realization of effective information processing and reliable transmission, and have made outstanding contributions to the transformation of information theory into information technology.

However, most of the above literature are based on the assumption that the source is stable (or ergodic). [6] successfully introduced the concept of upper/lower probability limit, which provided a new idea for the study of general information sources and called this method spectral information method. Han and his collaborators discussed the coding theorem, information distortion rate and hypothesis test of general (generalized) sources and achieved a lot of meaningful results. His systematic achievements are summarized in [6].

Mathematics subject classification (2020): 60F15.

Keywords and phrases: General information source, upper/lower generalized divergence-density rate, general relative entropy.

Foundation item: The Anhui Province Social Science Innovation Development Research Project (2021CX077), the National Natural Science Foundation of China (No. 11571142), the National Social Science Fundation of China (No. 21BJY213), University Key Project of Natural Science Foundation of Anhui Province (No. KJ2021A1031, KJ2021A1034, KJ2019A0683), the University Outstanding Young Talents Project of Anhui Province (No. gxyq2021018, gxyq2019082) and Key Project of Natural Science Foundation of Chaohu Univesity (No. XLZ-202105, XLZ-201801). Authors are alphabetized by last names.

* Corresponding author.

An important problem in information theory is the limit property of sample entropy. In the literature [17], discussed the asymptotic equipartition property (AEP) of independent identically distributed processes and discussed the case of stationary ergodic processes. [12] and [3] proved the AEP of ergodic sources on a finite alphabet set, this is the famous Shannon-McMillan-Breiman theorem. [4] made the extension to the case of countable set. [13], [16] and [8] obtained the $\mathscr{L}_1$ convergence of continuous ergodic sources. [2] and [14] obtained almost everywhere convergence of real valued ergodic processes. [1] skillfully adopted the sandwich method to give a general asymptotic equipartition property. [11], [10], [5], [15], [19], [18] and [20], by using analysis method, discussed AEP and strong deviation theorem for homogeneous Markov chains, non-homogeneous hidden Markov chains, tree-indexed Markov chains and Markov chains in random environment.

As mentioned in the introduction of [6], it is difficult to obtain significant results without appropriate restrictions on the sources. Motivated by the work of [6], [11], this paper establishes some new deviation theorems for general information sources, which generalizes the results of [10] and [21]. It is worth pointing out that we have no restrictions on the information sources.

The rest of this paper is organized as follows. Section 2 gives some basic concepts and definitions. In Section 3, a kind of limit properties and strong deviation limit theorems for generalized sources are established.

2. Preliminaries

In this section, we give some definitions and notations which will be used henceforward. We start by introducing the notations. Throughout this paper, all random variables are defined on a fixed probability space $(\Omega, \mathscr{F}, \mu)$. Let $\Xi = \{\xi^n = (\xi_1^{(n)}, \cdots, \xi_n^{(n)})\}_{n\in\mathbb{N}^+}$ be a general information source, where $\{\xi_i^{(n)}, 1 \leqslant i \leqslant n\}$ is a random variable over the n-th Cartesian product $\mathscr{X}^n$ of an arbitrary discrete source alphabet $\mathscr{X} = \{a_1, a_2, \cdots\}$ and $\mathbb{N}^+$ is the set of all positive integers.

Assume that the joint distribution of ξ^n are

$$\begin{aligned}&p_n(x_1^{(n)}, \cdots, x_n^{(n)})\\ =&\mu(\xi_1^{(n)} = x_1^{(n)}, \cdots, \xi_n^{(n)} = x_n^{(n)}) > 0,\ \ x_i^{(n)} \in \mathscr{X},\ \ i = 1, 2, \cdots, n.\end{aligned} \tag{1}$$

In general, the information sources satisfy the consistency condition

$$\xi_i^{(n)} = \xi_i^{(m)}, \quad \forall\ i = 1, 2, \ldots, m,$$

for arbitrary $m < n$ and $m, n \in \mathbb{N}^+$ and are usually called the stochastic processes. But, the general sources considered here are not required to satisfy the consistency condition, they contain various sources, it contains all of non-stationary and/or non-ergodic sources.

Further, suppose that the marginal distribution of $\xi_i^{(n)}$ is

$$p_{ni}(x_i^{(n)}) = \mu(\xi_i^{(n)} = x_i^{(n)}),\ \ i = 1, 2, \cdots, n.$$

and denote

$$q_n(x_1^{(n)},\cdots,x_n^{(n)}) = \prod_{i=1}^{n} p_{ni}(x_i^{(n)}), x_i^{(n)} \in \mathscr{X}, \ \ n \in \mathbb{N}^+.$$

By the Kolomogrov measure extension theorem, there exists a probability measure (denoted by $\widetilde{\mu}$) on $(\Omega,\mathscr{F})$ such that

$$\widetilde{\mu}(\xi_1^{(n)} = x_1^{(n)},\cdots,\xi_n^{(n)} = x_n^{(n)}) = q_n(x_1^{(n)},\cdots,x_n^{(n)}), \tag{2}$$

i.e. $\Xi = \{\xi^n\}_{n\in\mathbb{N}^+}$, for fixed n, are independent under probability measure $\widetilde{\mu}$.

DEFINITION 1. Let μ and $\widetilde{\mu}$ be defined as in (1) and (2), respectively. Let $\xi^n = (\xi_1^{(n)},\cdots,\xi_n^{(n)})_{n\in\mathbb{N}^+}$ be a random vector, $q_n(\xi^n)$ and $p_n(\xi^n)$ be distribution functions of ξ^n, define

$$\overline{h}_{\mu}^{\widetilde{\mu}}(\omega) := \limsup_n \frac{1}{n} \log \frac{q_n(\xi^n)}{p_n(\xi^n)} \tag{3}$$

and

$$\underline{h}_{\mu}^{\widetilde{\mu}}(\omega) := \liminf_n \frac{1}{n} \log \frac{q_n(\xi^n)}{p_n(\xi^n)} \tag{4}$$

$\overline{h}_{\mu}^{\widetilde{\mu}}(\omega)$ and $\underline{h}_{\mu}^{\widetilde{\mu}}(\omega)$ are called the upper and lower generalized divergence-density rate of probability measure μ relative to $\widetilde{\mu}$, respectively.

DEFINITION 2. Let $D(p_n||q_n) = \mathbb{E}\log\left[\frac{p_n(\xi^n)}{q_n(\xi^n)}\right]$, define

$$\underline{D}(\mu||\widetilde{\mu}) := \liminf_n \frac{1}{n} D(p_n||q_n), \tag{5}$$

and

$$\overline{D}(\mu||\widetilde{\mu}) := \limsup_n \frac{1}{n} D(p_n||q_n). \tag{6}$$

$\overline{D}(\mu\|\widetilde{\mu})$ and $\underline{D}(\mu\|\widetilde{\mu})$ are called the sup-divergence rate and the inf-divergence rate of μ respect to $\widetilde{\mu}$, respectively.

Let $\Xi = \{\xi^n = (\xi_1^{(n)},\cdots\xi_n^{(n)})\}_{n\in\mathbb{N}^+}$ be a general information source. Denote $\widetilde{\xi}_i^{(n)} = \log\frac{1}{p_{ni}(\xi_i^{(n)})}$, $1 \leqslant i \leqslant n$, $n \in \mathbb{N}^+$. For every n,i $(1 \leqslant i \leqslant n)$, defining,

$$\mathscr{L}_n(\omega) := \frac{p_n(\xi^n)}{\prod_{i=1}^{n} p_{ni}(\xi_i^{(n)})},$$

$$G_i^{(n)}(r) := \mathbb{E}e^{r\widetilde{\xi}_i^{(n)}} = \sum_{x_i^{(n)}\in\mathscr{X}} \left[p_{ni}(x_i^{(n)})\right]^{1-r} \quad (r > 0)$$

be the likelihood ratio and the generating function of r.v. $\widetilde{\xi}_i^{(n)}$, respectively.

Hereafter in this context, log means the natural logarithm unless stated otherwise. We will use the convention that $0\log 0 = 0$, which is easily justified by continuity since $x\log x \to 0$ as $x \to 0$.

LEMMA 1. ([1]) *Let* $\{\xi_n\}_{n\in\mathbb{N}^+}$ *be a sequence of nonegative r.v.'s with* $\mathbb{E}\xi_n \leqslant 1$, *then*

$$\limsup_n \frac{1}{n}\log \xi_n \leqslant 0 \qquad a.s.$$

3. Main results and proofs

With the preliminaries accounted for, the main results may now be established.

THEOREM 1. *Let* $\Xi = \{\xi^n = (\xi_1^{(n)}, \cdots \xi_n^{(n)})\}_{n\in\mathbb{N}^+}$, $\underline{h}_\mu^{\tilde{\mu}}(\omega)$, $\mathscr{L}_n(\omega)$, $G_i^{(n)}(r)$ *be given as above. Let*

$$\mathscr{D} := \left\{\omega : \underline{h}_\mu^{\tilde{\mu}}(\omega) > -\infty\right\}. \tag{7}$$

If there exists a positive constant r_0, *such that* $G_i^{(n)}(r_0) < \infty$, $1 \leqslant i \leqslant n$, $n \in \mathbb{N}^+$ *and*

$$\limsup_n \frac{1}{n}\sum_{i=1}^n G_i^{(n)}(r_0) = G(r_0) < \infty, \tag{8}$$

then

$$\liminf_n \frac{1}{n}\sum_{i=1}^n \left[\widetilde{\xi}_i^{(n)} - H(\xi_i^{(n)})\right] \geqslant 0 \quad a.s. \quad \omega \in \mathscr{D} \tag{9}$$

and

$$\limsup_n \frac{1}{n}\sum_{i=1}^n \left[\widetilde{\xi}_i^{(n)} - H(\xi_i^{(n)})\right] \leqslant \alpha(\underline{h}_\mu^{\tilde{\mu}}(\omega)) \quad a.s. \quad \omega \in \mathscr{D} \tag{10}$$

where $H(\xi_i^{(n)}) = \mathbb{E}\widetilde{\xi}_i^{(n)}$ *is the entropy of r.v.* $\xi_i^{(n)}$, *and*

$$\alpha(x) = \inf_r\{g(r,x), 0 < r < r_0\}, \quad x \leqslant 0 \tag{11}$$

$$g(r,x) = \frac{2re^{-2}G(r_0)}{(r_0-r)^2} - \frac{x}{r}, \quad x \leqslant 0 \tag{12}$$

and

$$\alpha(0) = 0 \leqslant \alpha(x) \leqslant g\left(\frac{r_0\sqrt{-x}}{1-\sqrt{-x}}, 0\right) = \left[\frac{2e^{-2}G(r_0)+1}{r_0}\right]\sqrt{-x}(1+\sqrt{-x}).$$

Proof. For $r \in (-\infty, r_0]$, define

$$G_i^{(n)}(r,\xi_i^{(n)}) := \frac{1}{G_i^{(n)}(r)}\left[p_{ni}(\xi_i^{(n)})\right]^{1-r}.$$

It can be checked that $G_i(r,x)$ is a probability mass function on $\mathscr{X}$. Put

$$r_n(x_1^{(n)},\cdots,x_n^{(n)}) := \prod_{i=1}^{n} G_i^{(n)}(r,x_i^{(n)}) = \prod_{i=i}^{n}\frac{1}{G_i^{(n)}(r)}\left[p_{ni}(x_i^{(n)})\right]^{1-r},$$

and

$$\Lambda_n^{(1)}(r,\omega) := \frac{r_n(\xi_1^{(n)},\cdots,\xi_n^{(n)})}{p_n(\xi_1^{(n)},\cdots,\xi_n^{(n)})}, \quad n \in \mathbb{N}^+.$$

Note that

$$\begin{aligned}\mathbb{E}_\mu \Lambda_n^{(1)}(r,\omega) =&\Sigma_{x_1^{(n)},\cdots,x_n^{(n)}} \frac{r_n(x_1^{(n)},\cdots,x_n^{(n)})}{p_n(x_1^{(n)},\cdots,x_n^{(n)})}\cdot p_n(x_1^{(n)},\cdots,x_n^{(n)})\\ =&\Sigma_{x_1^{(n)},\cdots,x_n^{(n)}} r_n(x_1^{(n)},\cdots,x_n^{(n)})\\ \leqslant& 1.\end{aligned}$$

Obviously, Lemma 1 implies that

$$\limsup_n \frac{1}{n}\log \Lambda_n^{(1)}(r,\omega) \leqslant 0 \quad a.s.$$

Noting

$$\log \Lambda_n^{(1)}(r,\omega) = \sum_{i=1}^{n} r\widetilde{\xi}_i^{(n)} - \sum_{i=1}^{n}\log G_i^{(n)}(r) - \log \mathscr{L}_n(\omega),$$

thus,

$$\limsup_n \frac{1}{n}\left[\sum_{i=1}^{n} r\widetilde{\xi}_i^{(n)} - \sum_{i=1}^{n}\log G_i^{(n)}(r) - \log \mathscr{L}_n(\omega)\right] \leqslant 0 \quad a.s. \tag{13}$$

Letting $r = 0$ in (13), we have

$$h_*(\omega) := \liminf_n \frac{1}{n}\log \mathscr{L}_n(\omega) \geqslant 0 \quad a.s.$$

Hence

$$\overline{h}_\mu^{\widetilde{\mu}}(\omega) = \limsup_n \frac{1}{n}\log\frac{1}{\mathscr{L}_n(\omega)} \leqslant 0 \quad a.s.$$

By formulas (4), (7) and (13), the properties of superior limit and the inequality $1-\frac{1}{x}\leqslant \log x\leqslant x-1\ \ (x>0)$, we have

$$
\begin{aligned}
&\limsup_n \frac{r}{n}\sum_{i=1}^n\left[\widetilde{\xi}_i^{(n)}-H(\xi_i^{(n)})\right]\\
\leqslant&\limsup_n \frac{1}{n}\sum_{i=1}^n\left[\log G_i^{(n)}(r)-rH(\xi_i^{(n)})\right]-\underline{h}_\mu^{\tilde{\mu}}(\omega)\\
\leqslant&\limsup_n \frac{1}{n}\sum_{i=1}^n\left[G_i^{(n)}(r)-1-rH(\xi_i^{(n)})\right]-\underline{h}_\mu^{\tilde{\mu}}(\omega)\\
=&\limsup_n \frac{1}{n}\sum_{i=1}^n\mathbb{E}(e^{r\widetilde{\xi}_i^{(n)}}-1-r\widetilde{\xi}_i^{(n)})-\underline{h}_\mu^{\tilde{\mu}}(\omega)\quad a.s.\ \omega\in\mathscr{D}.
\end{aligned}\tag{14}
$$

Using the inequality $0\leqslant e^x-1-x\leqslant\frac{1}{2}(x)^2e^{|x|}$, we have by (14) that

$$
\begin{aligned}
&\limsup_n \frac{r}{n}\sum_{i=1}^n\left[\widetilde{\xi}_i^{(n)}-H(\xi_i^{(n)})\right]\\
\leqslant&\frac{r^2}{2}\limsup_n \frac{1}{n}\sum_{i=1}^n\mathbb{E}\left[(\widetilde{\xi}_i^{(n)})^2e^{|r\widetilde{\xi}_i^{(n)}|}\right]-\underline{h}_\mu^{\tilde{\mu}}(\omega)\quad a.s.\ \omega\in\mathscr{D}.
\end{aligned}\tag{15}
$$

Since

$$
\sup\left\{x^{(r_0+r)}(\log x)^2,0\leqslant x\leqslant 1\right\}\leqslant\frac{4e^{-2}}{(r_0+r)^2},\quad r<0\ \text{ and }\ r\neq -r_0\tag{16}
$$

and

$$
\sup\left\{x^{(r_0-r)}(\log x)^2,0\leqslant x\leqslant 1\right\}\leqslant\frac{4e^{-2}}{(r_0-r)^2},\quad 0<r<r_0.\tag{17}
$$

Let $r<0$ and $r\neq -r_0$ in (15), we have by (16) and (17)

$$
\begin{aligned}
&r\liminf_n \frac{1}{n}\sum_{i=1}^n\left[\widetilde{\xi}_i^{(n)}-H(\xi_i^{(n)})\right]\\
\leqslant&\frac{r^2}{2}\limsup_n \frac{1}{n}\sum_{i=1}^n\sum_{x_i^{(n)}\in\mathscr{X}}p_{ni}(x_i^{(n)})e^{r\log p_{ni}(x_i^{(n)})}\left[\log p_{ni}(x_i^{(n)})\right]^2-\underline{h}_\mu^{\tilde{\mu}}(\omega)\\
=&\frac{r^2}{2}\limsup_n \frac{1}{n}\sum_{i=1}^n\sum_{x_i^{(n)}\in\mathscr{X}}\left[p_{ni}(x_i^{(n)})\right]^{1-r_0}\left[p_{ni}(x_i^{(n)})\right]^{r_0+r}\left[\log p_{ni}(x_i^{(n)})\right]^2-\underline{h}_\mu^{\tilde{\mu}}(\omega)\\
\leqslant&\frac{2r^2e^{-2}G(r_0)}{(r+r_0)^2}-\underline{h}_\mu^{\tilde{\mu}}(\omega)\quad a.s.\quad \omega\in\mathscr{D}.
\end{aligned}\tag{18}
$$

By formulas (4), (8) and (18), we have

$$
\liminf_n \frac{1}{n}\sum_{i=1}^n\left[\widetilde{\xi}_i^{(n)}-H(\xi_i^{(n)})\right]\geqslant\frac{2re^{-2}G(r_0)}{(r+r_0)^2}-\frac{\underline{h}_\mu^{\tilde{\mu}}(\omega)}{r}\ \ a.s.\ \omega\in\mathscr{D}.\tag{19}
$$

Taking $r \to -\infty$ in (19), we have $\frac{2re^{-2}G(r_0)}{(r+r_0)^2} - \frac{\underline{h}_\mu^{\tilde{\mu}}(\omega)}{r} \to 0$, thus (9) follows

Similarly, put $0 < r < r_0$ in (15), we have by (10) and (17)

$$
\begin{aligned}
&\limsup_n \frac{1}{n}\sum_{i=1}^n \left[\widetilde{\xi}_i^{(n)} - H(\xi_i^{(n)})\right] \\
\leqslant &\frac{r}{2}\limsup_n \frac{1}{n}\sum_{i=1}^n \sum_{x_i^{(n)}\in\mathscr{X}} p_{ni}(x_i^{(n)})e^{-r\log p_{ni}(x_i^{(n)})}\left[\log p_{ni}(x_i^{(n)})\right]^2 - \frac{\underline{h}_\mu^{\tilde{\mu}}(\omega)}{r} \\
= &\frac{r}{2}\limsup_n \frac{1}{n}\sum_{i=1}^n \sum_{x_i^{(n)}\in\mathscr{X}} \left[p_{ni}(x_i^{(n)})\right]^{1-r_0}\left[p_{ni}(x_i^{(n)})\right]^{r_0-r}\left[\log p_{ni}(x_i^{(n)})\right]^2 - \frac{\underline{h}_\mu^{\tilde{\mu}}(\omega)}{r} \\
\leqslant &\frac{2re^{-2}G(r_0)}{(r-r_0)^2} - \frac{\underline{h}_\mu^{\tilde{\mu}}(\omega)}{r} \\
= &g(r, \underline{h}_\mu^{\tilde{\mu}}(\omega)) \\
\leqslant &\alpha(\underline{h}_\mu^{\tilde{\mu}}(\omega)), \quad a.s. \quad \omega \in \mathscr{D}.
\end{aligned}
$$

From (11) and (12), we have for every $x \leqslant 0$,

$$0 \leqslant \alpha(x) \leqslant g\left(\frac{r_0\sqrt{-x}}{1+\sqrt{-x}}, x\right) = \left[\frac{2e^{-2}G(r_0)+1}{r_0}\right]\sqrt{-x}(1+\sqrt{-x}).$$

and

$$\alpha(0) \leqslant \lim_{x\to 0} g\left(\frac{r_0\sqrt{-x}}{1+\sqrt{-x}}, x\right) = 0.$$

hence

$$\alpha(0) = 0.$$

The proof is completed. □

Consider an independent and identically distributed information source $\xi = \{\xi_i\}_{i=1}^\infty$ with source alphabet $\mathscr{X}$ and Shannon entropy $H(\xi)$. The asymptotic equipartition property (AEP) is the assertion that

$$-\frac{1}{n}\log p(\xi_1, \xi_2, \cdots, \xi_n) \to H(\xi)$$

either in a sense of $\mathscr{L}_1$ convergence, convergence in probability or with probability one as n approaches to ∞. The AEP is fundamental to information theory and is called the Shannon-McMillan theorem in information theory.

Let $\Xi = \{\xi^n = (\xi_1^{(n)}, \cdots, \xi_n^{(n)})\}_{n\in\mathbb{N}^+}$ be a general information sources. Denote $p_n(x_1^{(n)}, \cdots, x_n^{(n)}) = \mu(\xi_1^{(n)} = x_1^{(n)}, \cdots, \xi_n^{(n)} = x_n^{(n)})$. Let

$$f_n(\omega) := -\frac{1}{n}\log p_n(\xi^n), \tag{20}$$

which is called the generalized information source entropy density of $p_n(\xi_1^{(n)}, \cdots, \xi_n^{(n)})$.

THEOREM 2. *Under the conditions of Theorem 1, we have*

$$\liminf_n \left[f_n(\omega) - \frac{1}{n}H(\xi^n)\right] \geqslant \underline{h}_\mu^{\tilde{\mu}}(\omega) + \underline{D}(\mu||\widetilde{\mu}) \quad a.s. \quad \omega \in \mathscr{D}.$$

and

$$\limsup_n \left[f_n(\omega) - \frac{1}{n}H(\xi^n)\right] \leqslant \overline{h}_\mu^{\tilde{\mu}}(\omega) + \alpha(\underline{h}_\mu^{\tilde{\mu}}(\omega)) + \overline{D}(\mu||\widetilde{\mu}), \quad a.s. \quad \omega \in \mathscr{D}.$$

where $H(\xi^n) = \mathbb{E}\left[-\log p_n(\xi_1^{(n)}, \cdots, \xi_n^{(n)})\right]$ *is the joint entropy of random vector* ξ^n.

Proof. From (5), (9) and (20), we have

$$\begin{aligned}
&\liminf_n \left[f_n(\omega) - \frac{1}{n}H(\xi^n)\right] \\
\geqslant &\liminf_n \frac{1}{n}\log\frac{1}{\mathscr{L}_n(\omega)} + \liminf_n \frac{1}{n}\sum_{i=1}^n \left[\widetilde{\xi}_i^{(n)} - H(\xi_i^{(n)})\right] \\
&+ \liminf_n \frac{1}{n}\left[\sum_{i=1}^n H(\xi_i^{(n)}) - H(\xi^n)\right] \\
\geqslant &\underline{h}_\mu^{\tilde{\mu}}(\omega) + \underline{D}(\mu||\widetilde{\mu}) \quad a.s. \quad \omega \in \mathscr{D}.
\end{aligned}$$

Analogously, we have by formulas (6), (10) and (20)

$$\begin{aligned}
&\limsup_n \left[f_n(\omega) - \frac{1}{n}H(\xi^n)\right] \\
\leqslant &\limsup_n \frac{1}{n}\log\frac{1}{\mathscr{L}_n(\omega)} + \limsup_n \frac{1}{n}\sum_{i=1}^n \left[\widetilde{\xi}_i^{(n)} - H(\xi_i^{(n)})\right] \\
&+ \limsup_n \frac{1}{n}\left[\sum_{i=1}^n H(\xi_i^{(n)}) - H(\xi^n)\right] \\
\leqslant &\overline{h}_\mu^{\widetilde{\mu}}(\omega) + \alpha(\underline{h}_\mu^{\tilde{\mu}}(\omega)) + \overline{D}(\mu||\widetilde{\mu}) \quad a.s. \quad \omega \in \mathscr{D},
\end{aligned}$$

which implies Theorem 2 holds. □

COROLLARY 1. *Put* $\mu = \widetilde{\mu}$, *i.e.* $\{\xi_i^{(n)},\ 1 \leqslant i \leqslant n\}_{n\in\mathbb{N}^+}$ *are independent for every fixed* n. *If there exists a positive constant* r_0 *such that* (10) *holds, then*

$$\limsup_n \left[f_n(\omega) - \frac{1}{n}H(\xi^n)\right] = 0 \quad a.s. \quad \omega \in \mathscr{D}.$$

Proof. Noticing that in this case, we have $\overline{D}(\mu||\widetilde{\mu}) = \underline{D}(\mu||\widetilde{\mu}) = 0$, $\underline{h}_\mu^{\tilde{\mu}}(\omega) = \overline{h}_\mu^{\tilde{\mu}}(\omega) \equiv 0$, $\mathscr{D} = \Omega$ and $\alpha(0) = 0$. Corollary 1 follows immediately. □

In the remainder of this section we make the restriction that the source alphabet $\mathscr{X}$ is finite, i.e. $\mathscr{X} = \{a_1, a_2, \cdots, a_{|\mathscr{X}|}\}$, where $|\cdot|$ denotes cardinality operator.

COROLLARY 2. *Under the conditions of Theorem 2, we have*

$$\liminf_n \frac{1}{n}\sum_{i=1}^{n}\left[\widetilde{\xi}_i^{(n)} - H(\xi_i^{(n)})\right] \geqslant 0 \quad a.s. \quad \omega \in \mathscr{D}$$

and

$$\limsup_n \frac{1}{n}\sum_{i=1}^{n}\left[\widetilde{\xi}_i^{(n)} - H(\xi_i^{(n)})\right] \leqslant \beta(\underline{h}_\mu^{\tilde{\mu}}(\omega)) \quad a.s. \quad \omega \in \mathscr{D}$$

$$\liminf_n \left[f_n(\omega) - \frac{1}{n}H(\xi^n)\right] \geqslant \underline{h}_\mu^{\tilde{\mu}}(\omega) + \underline{D}(\mu||\widetilde{\mu}) \quad a.s. \quad \omega \in \mathscr{D} \tag{21}$$

and

$$\limsup_n \left[f_n(\omega) - \frac{1}{n}H(\xi^n)\right] \leqslant \overline{h}_\mu^{\tilde{\mu}}(\omega) + \beta(\underline{h}_\mu^{\tilde{\mu}}(\omega)) + \overline{D}(\mu||\widetilde{\mu}) \quad a.s. \quad \omega \in \mathscr{D} \tag{22}$$

where

$$\beta(x) = \inf_r \{h(r,x), 0 < r < 1\}, \quad x \leqslant 0$$

$$h(r,x) = \frac{2e^{-2}|\mathscr{X}|r}{(1-r)^2} - \frac{x}{r} \geqslant 2\sqrt{-\frac{2e^{-2}|\mathscr{X}|x}{(1-r)^2}}, \quad x \leqslant 0$$

and

$$0 = \beta(0) \leqslant \beta(x) \leqslant h\left(\frac{\sqrt{-x}}{1+\sqrt{-x}}, x\right) = (2e^{-2}|\mathscr{X}|+1)\sqrt{-x}(1+\sqrt{-x}). \tag{23}$$

Proof. Let $r_0 = 1$, note that $G(r_0) = G(1) \leqslant |\mathscr{X}|$. The assertion follows directly from Theorem 1. □

LEMMA 2. *Let Ξ be a generalized information source, then $\left\{-\frac{1}{n}\log p_n(\xi^n)\right\}_{n\in\mathbb{N}^+}$ are uniformly integrable.*

Proof. For each nonnegative integer k define the sets

$$B_k(1,n) := \left\{\omega : -\frac{1}{n}\log p_n(\xi^n) \in [k, k+1)\right\}$$

and hence if $\omega \in B_k(1,n)$ we have

$$k \leqslant -\frac{1}{n}\log p_n(\xi^n) < k+1$$

or

$$e^{-n(k+1)} < p_n(\xi^n) \leqslant e^{-nk}.$$

Thus for any k, we have

$$\begin{aligned}\sum_{\omega\in B_k(1,n)}\Big[-\frac{1}{n}\log p_n(\xi^n)\Big] <&(k+1)\mu(B_k(1,n))\\ =&(k+1)\sum_{\omega\in B_k(1,n)} e^{-nk}\\ \leqslant&(k+1)e^{-nk}|\mathscr{X}|^n.\end{aligned}$$

There are at most $|\mathscr{X}|^n$ possible n-tuples corresponding to thin cylinders in $B_k(1,n)$ and each probability less than e^{-nk}.

To prove uniform integrability we must show uniform convergence to 0 as $k\to\infty$ of the integral

$$\begin{aligned}\mathbb{E}A_k(1,n) :=& \sum_{\omega\in\left\{-\frac{1}{n}p_n(x_1^{(n)},\cdots,x_n^{(n)})\geqslant k\right\}} -\frac{1}{n}p_n(x_1^{(n)},\cdots,x_n^{(n)})\cdot p_n(x_1^{(n)},\cdots,x_n^{(n)})\\ =&\sum_{i=0}^{\infty}\sum_{\omega\in B_{k+i}(1,n)} -\frac{1}{n}p_n(x_1^{(n)},\cdots,x_n^{(n)})\cdot\log p_n(x_1^{(n)},\cdots,x_n^{(n)})\\ \leqslant&\sum_{i=0}^{\infty}(k+i+1)|\mathscr{X}|^n e^{-n(k+i)}e^{-n(k+i)}\\ \leqslant&\sum_{i=0}^{\infty}(k+i+1)e^{-n(2k+2i-\log|\mathscr{X}|)}.\end{aligned}$$

Notice that, taking k large enough so that $2k>\log|\mathscr{X}|$, then the exponential term is bound above by the special case $n=1$ and we have the bound

$$\mathbb{E}A_k(1,n)\leqslant\sum_{i=0}^{\infty}(k+i+1)e^{-(2k+2i-\log|\mathscr{X}|)}$$

a bound which is finite and independent of i and n. Taking $k\to\infty$, the sum can easily be shown to go to zero so that prove $\left\{-\frac{1}{n}\log p_n(\xi^{(n)})\right\}_{n\in\mathbb{N}^+}$ are uniformly integrable. $\square$

COROLLARY 3. *Let $\underline{h}_\mu^{\tilde\mu}(\omega)$ and $h_*(\omega)$ be defined as above, then*

$$\liminf_n\left[f_n(\omega)-\frac{1}{n}H(\xi^n)\right]\geqslant\underline{h}_\mu^{\tilde\mu}(\omega)+\mathbb{E}(h_*(\omega))\quad a.s.\quad \omega\in\mathscr{D}. \tag{24}$$

and

$$\begin{aligned}\limsup_n\left[f_n(\omega)-\frac{1}{n}H(\xi^n)\right]\leqslant&(2e^{-2}|\mathscr{X}|+1)\sqrt{-\underline{h}_\mu^{\tilde\mu}(\omega)}\left[1+\sqrt{-\underline{h}_\mu^{\tilde\mu}(\omega)}\right]\\ &+\mathbb{E}(-\underline{h}_\mu^{\tilde\mu}(\omega))\quad a.s.\quad\omega\in\mathscr{D}.\end{aligned}\tag{25}$$

Proof. Since $\left\{\frac{\log \mathscr{L}_n(\omega)}{n}, n \geqslant 1\right\}$ are uniformly integrable, by the Fatou Lemma, we have

$$\underline{D}(p_n \parallel q_n) = \liminf_{n} \mathbb{E}(\log \mathscr{L}_n(\omega)) \geqslant \mathbb{E}(h_*(\omega)), \tag{26}$$

$$\overline{D}(p_n \parallel q_n) = \limsup_{n} \mathbb{E}(\log \mathscr{L}_n(\omega)) \leqslant \mathbb{E}(\underline{h}_\mu^{\tilde{\mu}}(\omega)), \tag{27}$$

then (24) follows from (21) , (23) and (26), and (25) follows from (22), (23) and (27). □

COROLLARY 4. *If*

$$\underline{h}_\mu^{\tilde{\mu}}(\omega) = 0 \quad a.s.,$$

then

$$\lim_n \left[f_n(\omega) - \frac{1}{n} H(\xi^n) \right] = 0 \quad a.s. \tag{28}$$

Proof. In this case, $\mu(\mathscr{D}) = 1$, $\mathbb{E}(-\underline{h}_\mu^{\tilde{\mu}}(\omega)) = 0$, and $h_*(\omega) \geqslant 0 \quad a.s.$, (28) follows from (24) and (25). □

Next, we adopt the method presented above to get some upper bounds on the generalized divergence-density rate.

THEOREM 3. *Let $a_k \in \mathscr{X}$, $S_n(a_k, \omega)$ be the number of occurrences of a_k in sequence $(\xi_1^{(n)}, \cdots, \xi_n^{(n)})$, c be a constant with $(0 \leqslant c \leqslant 1)$. Let*

$$S_*(a_k, \omega) := \left\{ \omega : \liminf_{n \to \infty} \frac{1}{n} \left[S_n(a_k, \omega) - \sum_{i=1}^{n} p_{ni}(a_k) \right] \geqslant c \right\}$$

and

$$S_*(c) := \bigcup_{k=1}^{|\mathscr{X}|} S_*(a_k, \omega),$$

then

$$\overline{h}_\mu^{\tilde{\mu}}(\omega) \leqslant c - (1+c)\log(1+c) \quad a.s. \quad \omega \in S_*(c).$$

Proof. Let t be a positive constant, define

$$\Lambda_n^{(2)}(t, \omega) := \frac{t^{S_n(a_k,\omega)} \prod_{i=1}^{n} \frac{p_{ni}(a_k)}{1+(t-1)p_{ni}(a_k)}}{p_n(x_1^{(n)}, \cdots, x_n^{(n)})}. \tag{29}$$

Note that

$$
\begin{aligned}
&\mathbb{E}\Lambda_n^{(2)}(t,\omega)\\
=&\Sigma_{x_1^{(n)},\cdots,x_n^{(n)}} \frac{t^{S_n(a_k,\omega)}\prod_{i=1}^n \frac{p_{ni}(a_k)}{1+(t-1)p_{ni}(a_k)}}{p_n(x_1^{(n)},\cdots,x_n^{(n)})}\cdot p_n(x_1^{(n)},\cdots,x_n^{(n)})\\
=&\Sigma_{x_1^{(n)},\cdots,x_n^{(n)}} t^{S_n(a_k,\omega)}\prod_{i=1}^n \frac{p_{ni}(a_k)}{1+(t-1)p_{ni}(a_k)}\\
=&\Sigma_{x_1^{(n)},\cdots,x_n^{(n)}} \prod_{i=1}^n \frac{p_{ni}(a_k)\cdot t^{\mathbf{1}_{\{a_k\}}(x_i^{(n)})}}{1+(t-1)p_{ni}(a_k)}
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{x_i^{(n)}} p_{ni}(x_i^{(n)})\cdot t^{\mathbf{1}_{\{a_k\}}(x_i^{(n)})}\\
=&\sum_{x_i^{(n)}=\{a_k\}} p_{ni}(x_i^{(n)})\cdot t^{\mathbf{1}_{\{a_k\}}(x_i^{(n)})}+\sum_{x_i^{(n)}\neq\{a_k\}} p_{ni}(x_i^{(n)})\cdot t^{\mathbf{1}_{\{a_k\}}(x_i^{(n)})}\\
=&\,p_{ni}(a_k)\cdot t+(1-p_{ni}(a_k))\\
=&\,1+(t-1)p_{ni}(a_k).
\end{aligned}
$$

Hence $\mathbb{E}\Lambda_n^{(2)}(t,\omega)\leqslant 1$, then by Lemma 1, we have

$$
\limsup_n \frac{1}{n}\log\Lambda_n^{(2)}(t,\omega)\leqslant 0 \quad a.s. \tag{30}
$$

From (29) and (30), we have

$$
\liminf_{n\to\infty}\left\{\frac{1}{n}\log\mathscr{L}_n(\omega)+\frac{1}{n}\sum_{i=1}^n\log\left[1+(t-1)p_{ni}(a_k)\right]-\frac{1}{n}S_n(a_k,\omega)\log t\right\}\geqslant 0,
$$

thus

$$
\begin{aligned}
-\overline{h}_\mu^{\widetilde{\mu}}(\omega)&=\liminf_{n\to\infty}\frac{1}{n}\log\mathscr{L}_n(\omega)\\
&\geqslant\liminf_{n\to\infty}\left\{\frac{1}{n}S_n(a_k,\omega)\log t-\frac{1}{n}\sum_{i=1}^n\log\left[1+(t-1)p_{ni}(a_k)\right]\right\}.
\end{aligned} \tag{31}
$$

Let $t > 1$, (31) and the inequality $\log(1+x) \geqslant x (x \geqslant 0)$ imply that

$$\begin{aligned} -\overline{h}_{\mu}^{\widetilde{\mu}}(\omega) &\geqslant \log t \cdot \liminf_{n\to\infty} \frac{1}{n}\left[S_n(a_k,\omega) - \sum_{i=1}^{n} p_{ni}(x_i^{(n)})\right] \\ &\quad - \log t \cdot \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\log[1+(t-1)p_{ni}(a_k)]}{\log t} - p_{ni}(a_k)\right\} \\ &\geqslant \log t \cdot \liminf_{n\to\infty} \frac{1}{n}\left[S_n(a_k,\omega)\log t - \sum_{i=1}^{n} p_{ni}(a_k)\right] \\ &\quad - \log t \cdot \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\log[1+(t-1)p_{ni}(a_k)]}{\log t} - p_{ni}(a_k)\right\} \\ &\geqslant \log t\left\{c - \limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(t-1)p_{ni}(a_k)}{\log t} - p_{ni}(a_k)\right]\right\} \\ &\geqslant (1+c)\log t + 1 - t \quad a.s. \quad \omega \in S_*(c), \end{aligned}$$

thus

$$\overline{h}_{\mu}^{\widetilde{\mu}}(\omega) \leqslant t - 1 - (1+c)\log t \quad a.s. \quad \omega \in S_*(c).$$

We can find that, for any $c > 0$, $g(t) = (1+c)\log t + 1 - t$ attains the maximum value $g(1+c) = (1+c)\log(1+c) - c$ at $t = 1+c$. Therefore, we have for $c > 0$

$$\overline{h}_{\mu}^{\widetilde{\mu}}(\omega) \leqslant c - (1+c)\log(1+c) \quad a.s. \quad \omega \in S_*(c) \quad \square$$

THEOREM 4. *Under of the conditions of Theorem 3, let*

$$S^*(a_k,c) = \left\{\omega : \limsup_{n} \frac{1}{n}\left[S_n(a_k,\omega) - \sum_{i=1}^{n} p_{ni}(a_k)\right] \leqslant -c\right\} \tag{32}$$

and

$$S^*(c) = \bigcup_{k=1}^{|\mathscr{X}|} S^*(a_k,c),$$

then, if $0 \leqslant c < 1$ *we have*

$$\overline{h}_{\mu}^{\widetilde{\mu}}(\omega) \leqslant (c-1)\log(1-c) - c \quad a.s. \quad \omega \in S^*(c).$$

If $c = 1$, *we have*

$$\overline{h}_{\mu}^{\widetilde{\mu}}(\omega) \leqslant -1 \quad a.s. \quad \omega \in S^*(1).$$

Proof. Put $0<t<1$, by (31), (32) and the inequality $\log(1+x)\leqslant x$ $(-1<x\leqslant 0)$, we have

$$\begin{aligned}-\frac{1}{\log t}\cdot\overline{h}_{\mu}^{\widetilde{\mu}}(\omega) &\leqslant \limsup_{n}\frac{1}{n}\left[S_n(a_k,\omega)-\sum_{i=1}^{n}p_{ni}(x_i^{(n)})\right]\\ &\quad-\limsup_{n}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\log\left[1+(t-1)p_{ni}(x_i^{(n)})\right]}{\log t}-p_{ni}(x_i^{(n)})\right\}\\ &\leqslant \limsup_{n}\frac{1}{n}\left[S_n(a_k,\omega)-\sum_{i=1}^{n}p_{ni}(x_i^{(n)})\right]\\ &\quad-\liminf_{n}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(t-1)p_{ni}(x_i^{(n)})}{\log t}-p_{ni}(x_i^{(n)})\right]\\ &\leqslant -c+1-\frac{t-1}{\log t}\quad a.s.\quad \omega\in S^*(c).\end{aligned}$$

Thus

$$\overline{h}_{\mu}^{\widetilde{\mu}}(\omega)\leqslant (c-1)\log t+t-1\quad a.s.\quad \omega\in S^*(c).$$

It is easy to see that $g(t)=(1-c)\log t+t-1$ $(0<t\leqslant 1)$ attains the maximum $g(1-c)=(c-1)\log(1-c)-c$ on $(0,1]$. If $0<c<1$, we have

$$\overline{h}_{\mu}^{\widetilde{\mu}}(\omega)\leqslant (c-1)\log(1-c)-c\quad a.s.\quad \omega\in S^*(c).$$

Similarly, if $c=1$, we have

$$\overline{h}_{\mu}^{\widetilde{\mu}}(\omega)\leqslant -1\quad a.s.\quad \omega\in S^*(1).$$

The proof is completed. □

THEOREM 5. *Assume that $(p_1,\cdots,p_{|\mathscr{X}|})$ with $p_i>0$ and $\sum_{i=1}^{|\mathscr{X}|}p_i=1$ is a distribution on $\mathscr{X}$. Denote $p_{\max}=\max\{p_1,\cdots,p_{|\mathscr{X}|}\}$ and $p_{\min}=\min\{p_1,\cdots,p_{|\mathscr{X}|}\}$. Let $\mathscr{X}_0=\{a_i:p_i=p_{\min},a_i\in\mathscr{X}\}$, $\mathscr{X}_1=\{a_i:p_i=p_{\max},a_i\in\mathscr{X}\}$. Let $a_i\in\mathscr{X}$, $S_n(a_i,\omega)$ be the number of occurrences of a_i in segement $(\xi_1^{(n)},\cdots,\xi_n^{(n)})$. If*

$$\limsup_{n\to\infty}\frac{S_n(a_i,\omega)}{n}\geqslant p_i,\ for\ all\ a_i\in\mathscr{X}-\mathscr{X}_0\quad a.s.$$

or

$$\limsup_{n\to\infty}\frac{S_n(a_i,\omega)}{n}\leqslant p_i,\ for\ all\ a_i\in\mathscr{X}-\mathscr{X}_1\quad a.s.$$

then

$$\limsup_{n\to\infty}f_n(\omega)\leqslant H(p_1,\cdots,p_{|\mathscr{X}|})\qquad a.s.$$

Proof. Set

$$\Lambda_n^{(3)}(\omega)=\frac{\prod_i^n\sum_{k=1}^{|\mathscr{X}|}p_k\mathbf{1}_{\{\xi_i^{(n)}=a_k\}}}{p_n(\xi^n)}.$$

It is not hard to verify that $\mathbb{E}\Lambda_n^{(3)}=1$, so we have $\limsup_n \frac{1}{n}\log\Lambda_n^{(3)}(\omega)\leqslant 0 \quad a.s.$

Note that

$$\frac{1}{n}\log\Lambda^{(3)}(\omega)=\sum_{i:a_i\in\mathscr{X}-\mathscr{X}_0}\frac{S_n(a_i,\omega)}{n}\log p_i+\sum_{i:a_i\in\mathscr{X}-\mathscr{X}_0}\left(1-\frac{S_n(a_i,\omega)}{n}\right)\log p_{\min}$$
$$-\frac{1}{n}\log p_n(\xi^n),$$

thus

$$\limsup_n f_n(\omega)\leqslant\limsup_n\sum_{i:a_i\in\mathscr{X}-\mathscr{X}_0}\frac{S_n(a_i,\omega)}{n}\log\frac{p_{\min}}{p_i}-\log p_{\min}\ \ a.s.$$

$$\limsup_n f_n(\omega)\leqslant\sum_{i:a_i\in\mathscr{X}-\mathscr{X}_0}\log\frac{p_{\min}}{p_i}\limsup_n\frac{S_n(a_i,\omega)}{n}-\log p_{\min}\ \ a.s.$$

$$\limsup_n f_n(\omega)\leqslant\sum_{i:a_i\in\mathscr{X}-\mathscr{X}_0}p_i\log\frac{p_{\min}}{p_i}-\log p_{\min}\ \ a.s.$$

$$\limsup_n f_n(\omega)\leqslant-\sum_{i=1}^{|\mathscr{X}|}p_i\log p_i\ \ a.s.$$

i.e.

$$\limsup_n f_n(\omega)\leqslant H(p_1,\cdots,p_{|\mathscr{X}|})\ \ a.s.$$

Notice that

$$\frac{1}{n}\log\Lambda^{(3)}(\omega)=\sum_{i:a_i\in\mathscr{X}-\mathscr{X}_1}\frac{S_n(a_i,\omega)}{n}\log p_i+\sum_{i:a_i\in\mathscr{X}-\mathscr{X}_1}\left(1-\frac{S_n(a_i,\omega)}{n}\right)\log p_{\max}$$
$$-\frac{1}{n}\log p_n(\xi^n).$$

Similarly, we have

$$\limsup_n f_n(\omega)\leqslant\limsup_n\sum_{i:a_i\in\mathscr{X}-\mathscr{X}_1}\frac{S_n(a_i,\omega)}{n}\log\frac{p_{\max}}{p_i}-\log p_{\max}\ \ a.s.$$

$$\limsup_n f_n(\omega)\leqslant\sum_{i:a_i\in\mathscr{X}-\mathscr{X}_1}\log\frac{p_{\max}}{p_i}\limsup_n\frac{S_n(a_i,\omega)}{n}-\log p_{\max}\ \ a.s.$$

$$\limsup_n f_n(\omega)\leqslant\sum_{i:a_i\in\mathscr{X}-\mathscr{X}_1}p_i\log\frac{p_{\max}}{p_i}-\log p_{\max}\ \ a.s.$$

thus

$$\limsup_n f_n(\omega) \leqslant -\sum_{i=1}^{|\mathscr{X}|} p_i \log p_i \ \ a.s.$$

Therefore, Theorem 5 follows, i.e.

$$\limsup_n f_n(\omega) \leqslant H(p_1, \cdots, p_{|\mathscr{X}|}) \quad a.s. \quad \square$$

COROLLARY 5. $\limsup_n f_n \leqslant \log|\mathscr{X}| \ \ a.s.$

Proof. Let

$$\Lambda^{(4)}(\omega) := \frac{\frac{1}{|\mathscr{X}|^n}}{p_n(\xi^n)}.$$

Obviously,

$$\begin{aligned}
\mathbb{E}\Lambda^{(4)}(\omega) &= \sum_{x_1^{(n)} \in \mathscr{X}} \cdots \sum_{x_n^{(n)} \in \mathscr{X}} \frac{\frac{1}{|\mathscr{X}|^n}}{p_n(x_1^{(n)}, \cdots, x_n^{(n)})} \cdot p_n(x_1^{(n)}, \cdots, x_n^{(n)}) \\
&= \frac{1}{|\mathscr{X}|^n} \cdot |\mathscr{X}|^n \\
&= 1,
\end{aligned}$$

hence

$$\limsup_n \frac{1}{n} \Lambda^{(4)}(\omega) \leqslant 0 \ \ a.s.$$

which indicates that

$$\limsup_n \frac{1}{n} [-\log|\mathscr{X}|^n - \log p_n(\xi^n)] \leqslant 0 \ \ a.s.$$

i.e.

$$\limsup_n f_n \leqslant \log|\mathscr{X}| \ \ a.s. \quad \square$$

REFERENCES

[1] P. H. ALGOET, T. M. COVER, *A sandwich proof of the Shannon-McMillan-Breiman theorem*, The annals of probability, 899–909, 1988.

[2] A. R. BARRON, *The strong ergodic theorem for densities: generalized Shannon-McMillan-Breiman theorem*, The annals of Probability **13** (4), 1292–1303, 1985.

[3] L. BREIMAN, *The individual ergodic theorem of information theory*, The Annals of Mathematical Statistics **28** (3), 809–811, 1957.

[4] K. L. CHUNG, *A note on the ergodic theorem of information theory*, Mathematical Division, Office of Scientific Research, US Air Force, 1960.

[5] H. DANG, W. YANG, Z. SHI, *The strong law of large numbers and the entropy ergodic theorem for nonhomogeneous bifurcating Markov chains indexed by a binary tree*, IEEE Transactions on Information Theory **61** (4), 1640–1648, 2015.
[6] T. S. HAN, *Information-spectrum methods in information theory*, Springer, New York, 2003.
[7] A. Y. KHINCHIN, *Mathematical foundations of information theory*, Courier Corporation, 2013.
[8] J. C. KIEFFER, et al., *A simple proof of the Moy-Perez generalization of the Shannon-McMillan theorem*, Pacific Journal of Mathematics **51** (1), 203–206, 1974.
[9] A. N. KOLMOGOROV, A. T. BHARUCHA-REID, *Foundations of the theory of probability: Second English Edition*, Courier Dover Publications, 2018.
[10] W. LIU, *The Strong Deviation Theorms and Analytical Methods* (in Chinese), Beijing, Science Press, 2003.
[11] W. LIU, W. YANG, *The Markov approximation of the sequences of N-valued random variables and a class of small deviation theorems*, Stochastic Processes and their Applications **89** (1), 117–130, 2000.
[12] B. MCMILLAN, *The basic theorems of information theory*, The Annals of Mathematical Statistics, 196–219, 1953.
[13] S.-T. C. MOY, *Generalizations of Shannon-McMillan theorem*, Pacific J. Math **11** (2), 705–714, 1961.
[14] S. OREY, *On the shannon-perez-moy theorem*, Contemp. Math **41**, 319–327, 1985.
[15] W. PENG, W. YANG, Z. SHI, *Strong law of large numbers for Markov chains indexed by spherically symmetric trees*, Probability in the Engineering and Informational Sciences **29** (3), 473–481.
[16] J. PIERCE, *The early days of information theory*, IEEE Transactions on Information Theory **19**, (1), 3–8, 1973.
[17] C. E. SHANNON, *A mathematical theorem of communication*, Bell System Technical Journal **27**, (3), 379–423, 1948.
[18] Z. WANG, *A kind of asymptotic properties of moving averages for Markov chains in Markovian environments*, Communications in Statistics-Theory and Methods **46** (22), 10926–10940, 2017.
[19] Z. WANG, W. YANG, *The generalized entropy ergodic theorem for nonhomogeneous Markov chains*, Journal of Theoretical Probability **29** (3), 761–775, 2016.
[20] Z. WANG, W. YANG, *Markov approximation and the generalized entropy ergodic theorem for non-null stationary process*, Proceedings-Mathematical Sciences **130** (1), 1–19, 2020.
[21] Z.-Z. WANG, S.-S. YANG, H.-F. JIANG, F.-Q. DING, *On strong limit theorems for general information sources with an application to AEP*, Communications in Statistics-Theory and Methods **50** (6), 1387–1399, 2021.

(Received January 7, 2022)

Shu Chen
School of Mathematics & Physics
AnHui University of Technology
Ma'anshan, 243002, China
e-mail: 13855217135@163.com

Wei-cai Peng
Department of Mathematics and Statistics
Chaohu University
Hefei, 238000, China
e-mail: weicaipeng@126.com

Zhong-zhi Wang
School of Mathematics & Physics
AnHui University of Technology
Ma'anshan, 243002, China
e-mail: zhongzhiw@126.com