Applied Mathematics & Information Sciences **2**(3) (2008), 333–343
– An International Journal

Almost Sure Limit Behavior of Discrete Random Sequence with Application to Arbitrary Information Sources

Zhong-Zhi Wang[1] and Keyue Ding[2]

[1]Faculty of Mathematics & Physics, AnHui University of Technology
Ma'anshan, 243002, China
Email Address: wzz30@ahut.edu.cn

[2]National Cancer Institute of Canada Clinical Trials Group, Queen's University
Kingston, Ontario, K7L 3N6, Canada
Email Address: KDing@ctg.queensu.ca

Received February 11, 2008; Revised June 3, 2008

In this paper, the notion of the limit of logarithmic likelihood ratio of random sequences, as a measure of $dissimilarity$ between two probability measures, is introduced. After establishing a ratio of two measures by means of constructing a new probability measure, we obtained the strong random deviation theorems for partial sums of functions of arbitrary discrete random variables under suitable restrictive conditions. As a direct application, we used our results to derive some limit properties of discrete information source.

Keywords: Random sequence, random limit logarithmic likelihood ratio, strong deviation theorem, conditional entropy.

1 Introduction

In recent years, important progresses have been made in the field of deviation of the average of random variables from the expectations of their marginals or the reference measure. The main problem of the research area, tracing back to Liu [5] and [6], Liu and Yang [7] and Wang [9], is to determine a relationship between the true probability distribution and its marginals. The present paper focuses on the the strong deviation theorems for partial sums of arbitrary discrete random variables in more general settings.

The first author's research is supported by he National Natural Science Foundation of China (10571076) and Science Foundation of the Education Committee of AnHui Province (2006Kj 246B). K. Ding's research is supported by a grant from Natural Sciences and Engineering Research Council of Canada

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $\{\mathcal{F}_n, n \geq 0\}$ be an increasing sequence of $\sigma - fields$ with $\mathcal{F}_n \subset \mathcal{F}$ for $n \geq 0$, and suppose that $\{X_n, \mathcal{F}_n, n \geq 0\}$ be an adapted random variables taking values in $S = \{t_0, t_1, \ldots\}$ with a joint distribution of

$$p_n(x_1, \ldots, x_n),\ x_i \in S,\ n = 1, 2, \ldots \tag{1.1}$$

Without loss of generality, we may assume $p_n(x_0, \ldots, x_n) > 0$. Set

$$p_n(x_n|x_0, \ldots, x_{n-1}) = \mathbb{P}(X_n = x_n|X_0 = x_0, \ldots, X_{n-1} = x_{n-1}),\ \ n \geq 1, \tag{1.2}$$

and Π be another measure on Ω defined as follows.

$$\Pi(X_0 = x_0, \ldots, X_n = x_n) = \pi_n(x_0, \ldots, x_n) = \pi_0(x_0) \prod_{i=1}^{n} \pi_i(x_i|x_0, \ldots, x_{i-1}). \tag{1.3}$$

For $n \geq 0$, let $f_n(x_0, \cdots, x_n)$ denote a real-valued measurable function defined on S^{n+1}.

Definition 1.1. Let $\{X_n, n \geq 0\}$ be a sequence of random variables, and $\mathbb{P}, \Pi$ be two probability measures defined as above, $\{\sigma_n, n \geq 1\}$ be a integer valued sequence with $\sigma_n \uparrow \infty$. Let

$$L_n(\omega) = \frac{\Pi_n(X_0, \ldots, X_n)}{\mathbb{P}_n(X_0, \ldots, X_n)}, \tag{1.4}$$

$$\Lambda_n(\omega) = \log L_n(\omega). \tag{1.5}$$

The random variable

$$\gamma(\omega) = -\liminf_{n} \frac{\Lambda_n(\omega)}{\sigma_n}, \tag{1.6}$$

is called the limit of the random logarithmic likelihood ratio, relative to the measure Π, of $X_n, n \geq 1$, where $\log$ is the natural logarithm, ω is the sample point and X_n stands for $X_n(\omega)$.

2 Some General Strong Deviation Theorems

In this section, we shall first examine the connection between the Strong Deviation Theorems for arbitrary random variables and the Chung's type conditions [4, Theorem 1]. Let $\{a_n\}$ be a sequence of positive real numbers such that $a_{n+1} > a_n$ and $\lim_{n\to\infty} a_n = \infty$. Let $\{\Psi_n(t)\}$ be a sequence of positive, even, continuous functions such that, for $t_1, t_2 \in \mathbb{R}_+, t_1 \leq t_2$, there are constants $0 < \alpha_n \leq 2, K_n > 0 (n \geq 1)$, satisfy,

$$\frac{t_1^{\alpha_n}}{\Psi_n(t_1)} \leq K_n \frac{t_2^{\alpha_n}}{\Psi_n(t_2)}. \tag{2.1}$$

For simplicity, we denote

$$Z_n = f_n(X_0, \ldots, X_n),\ \ n \geq 0, \tag{2.2}$$

and

$$W_n = Z_n \cdot I_{[|Z_n| \leq a_n]}, \tag{2.3}$$

where $I_{[\cdot]}$ is the indicator function.

Theorem 2.1. *Let $\{X_n, n \geq 0\}, \{f_n, n \geq 0\}, \{\sigma_n, n \geq 1\}$ and $\gamma(\omega)$ be defined as in Section 1 with $\sigma_n \uparrow \infty$. Let*

$$\mathcal{J} = \{\omega : \gamma(\omega) < \infty\}, \tag{2.4}$$

then

$$\limsup_n \frac{1}{\sigma_n} \sum_{i=1}^{\sigma_n} K_i \mathbb{E}_\Pi[\frac{\Psi_i(W_n)}{\Psi_i(a_i)}|X_0, \dots, X_{n-1}] = c(\omega) < \infty, \ \mathbb{P} - a.s. \tag{2.5}$$

implies that

$$\liminf_n \frac{1}{\sigma_n} \sum_{i=1}^{\sigma_n} \frac{W_i - \mathbb{E}_\Pi[W_i|X_0, \dots, X_{i-1}]}{a_i} \geq \alpha(\gamma(\omega), c(\omega)), \ \ \mathbb{P} - a.s. \, \omega \in \mathcal{J}. \tag{2.6}$$

$$\limsup_n \frac{1}{\sigma_n} \sum_{i=1}^{\sigma_n} \frac{W_i - \mathbb{E}_\Pi[W_i|X_0, \dots, X_{i-1}]}{a_i} \leq \beta(\gamma(\omega), c(\omega)), \ \ \mathbb{P} - a.s. \, \omega \in \mathcal{J}. \tag{2.7}$$

where $\mathbb{E}_\Pi$ denotes expectation under Π,

$$\alpha(x, y) = \sup\{\varphi(s, x, y), \ s < 0\}, \qquad 0 \leq x, y < +\infty, \tag{2.8}$$

$$\beta(x, y) = \inf\{\varphi(s, x, y), \ s > 0\}, \qquad 0 \leq x, y < +\infty, \tag{2.9}$$

$$\varphi(s, x, y) = \frac{x}{s} + \frac{1}{2} s e^{2|s|} y, \quad 0 \leq x, y < \infty. \tag{2.10}$$

and

$$\alpha(x, y) \leq 0, \ \beta(x, y) \geq 0, \ 0 \leq x, y < \infty. \tag{2.11}$$

$$\alpha(0, y) = \alpha(x, 0) = \beta(0, y) = \beta(x, 0) = 0, \ \ 0 \leq x, y < \infty. \tag{2.12}$$

$$\lim_{x \to 0+} \alpha(x, y) = \lim_{x \to 0+} \beta(x, y) = 0. \tag{2.13}$$

Proof. Define

$$D_{x_0, \dots, x_n} = \{\omega : X_i = x_i, 0 \leq i \leq n\}, \ \ x_i \in S,$$

then

$$\mathbb{P}(D_{x_0}) = p_0(x_0), \tag{2.14}$$

and

$$\mathbb{P}(D_{x_0, \dots, x_n}) = p_n(x_0, \dots, x_n) = p_0(x_0) \prod_{i=1}^{n} p_i(x_i|x_0, \dots, x_{i-1}), \tag{2.15}$$

$D_{x_0,\ldots,x_n}$ is called an nth-order elementary cylinder. Let $\mathcal{N}_n$ be the collection of nth-order elementary cylinders, $\mathcal{N}$ be the collection of ϕ and Ω and all cylinder sets. Let

$$Y_i := \frac{W_i - \mathbb{E}_\Pi[W_i|X_0,\ldots,X_{i-1}]}{a_i}, \quad y_i := \frac{w_i - \mathbb{E}_\Pi[W_i|X_0 = x_0,\ldots,X_{i-1} = x_{i-1}]}{a_i},$$

$$Q_0(s) := \pi_0(x_0).$$

For a real number of s, let

$$Q_i(s;x_0,\ldots,x_{i-1}) := \mathbb{E}_\Pi \exp[sY_i|X_0 = x_0,\ldots,X_{i-1} = x_{i-1}], \tag{2.16}$$

$$\pi_i(s;x_0,\ldots,x_i) := \frac{\pi_i(x_i|x_0,\ldots,x_{i-1})\exp(sy_i)}{Q_i(s;x_0,\ldots,x_{i-1})}. \tag{2.17}$$

Define a set function μ_s on $\mathcal{N}$ as follows:

$$\mu_s(\phi) = 0, \ \ \mu_s(\Omega) = 1, \ \ \mu_s(D_{x_0}) = \pi_0(x_0), \tag{2.18}$$

$$\begin{aligned}\mu_s(D_{x_0,\ldots,x_n}) &= \mu_s(D_{x_0,\ldots,x_{n-1}})\pi_n(s;x_0,\ldots,x_n) \\ &= \pi_0(x_0)\prod_{i=1}^n \pi_i(s;x_0,\ldots,x_i).\end{aligned} \tag{2.19}$$

It follows from (2.16)-(2.19) that μ_s is a measure on $\mathcal{N}$. Since $\mathcal{N}$ is a semi-algebra, μ_s has a unique extension to the $\sigma-$field $\sigma(\mathcal{N})$. Let

$$L_n(s;\omega) := \frac{\mu_s(D_{X_0,\ldots,X_n})}{\mathbb{P}(D_{X_0,\ldots,X_n})}. \tag{2.20}$$

It is easy to see that $\{\mathcal{N}_n, n \geq 1\}$ is a net relative to $(\Omega, \mathcal{F}, \mathbb{P})$. By Stromberg and Hewitt [8], there exists $A(s) \in \sigma(\mathcal{N}), \mathbb{P}(A(s)) = 1$, such that

$$\lim_n L_n(s;\omega) = \text{a finite number depending on } \omega, \omega \in A(s). \tag{2.21}$$

It implies that

$$\limsup_n \sigma_n^{-1} \log L_{\sigma_n}(s;\omega) \leq 0, \quad \omega \in A(s). \tag{2.22}$$

By (2.16)-(2.19) and (1.5), we have

$$\log L_{\sigma_n}(s;\omega) = s\sum_{i=1}^{\sigma_n} Y_i - \sum_{i=1}^{\sigma_n} \log Q_i(s) + \Lambda_{\sigma_n}. \tag{2.23}$$

Combing (2.22) and (2.23), we find that

$$\limsup_n \frac{1}{\sigma_n}\Big[\sum_{i=1}^{\sigma_n} sY_i + \Lambda_{\sigma_n} - \sum_{i=1}^{\sigma_n} \log Q_i(s)\Big] \leq 0, \ \omega \in A(s). \tag{2.24}$$

By (1.6) and (2.24), we have

$$\limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n} sY_i \le \gamma(\omega) + \limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n} \log Q_i(s),\ \omega \in A(s). \tag{2.25}$$

By (2.16) and the inequality $0 \le e^x - 1 - x \le x^2 e^{|x|}/2$, for all $x \in \mathbb{R}$, and the fact that $|Y_i| \le 2$, we have

$$\begin{aligned} 0 \le Q_i(s) - 1 &= \mathbb{E}_\Pi[\exp(sY_i) - 1 - sY_i] \\ &\le \frac{1}{2}s^2 e^{2|s|}\mathbb{E}_\Pi Y_i^2 \le \frac{1}{2}s^2 e^{2|s|}\mathbb{E}_\Pi[(\frac{W_i}{a_i})^2 | X_0, \dots, X_{i-1}]. \end{aligned} \tag{2.26}$$

Now with (2.1), we have

$$\Big(\frac{|W_i|}{a_i}\Big)^2 \le \Big(\frac{|W_i|}{a_i}\Big)^{\alpha_i} \le K_i \frac{\Psi_i(|W_i|)}{\Psi_i(a_i)}, \tag{2.27}$$

which implies

$$\mathbb{E}_\Pi[(\frac{|W_i|}{a_i})^2 | X_0, \dots, X_{i-1}] \le K_i \mathbb{E}_\Pi[\frac{\Psi_i(W_i)}{\Psi_i(a_i)} | X_0, \dots, X_{i-1}]. \tag{2.28}$$

Hence

$$0 \le Q_i(s) - 1 \le \frac{1}{2}s^2 e^{2|s|} K_i \mathbb{E}_\Pi[\frac{\Psi_i(W_i)}{\Psi_i(a_i)} | X_0, \dots, X_{i-1}]. \tag{2.29}$$

We set

$$\mathcal{H} = \{\omega : \limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n} K_i \mathbb{E}_\Pi[\frac{\Psi_i(W_i)}{\Psi_i(a_i)} | X_0, \dots, X_{i-1}] = c(\omega) < \infty\}. \tag{2.30}$$

From (2.26), (2.29) and (2.30), we have

$$0 \le \limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}[Q_i(s) - 1] \le \frac{1}{2}s^2 e^{2|s|} c(\omega), \quad \omega \in \mathcal{H} \cap A(s). \tag{2.31}$$

By the inequality $0 \le \log x \le x - 1 \ \ (x \ge 1)$ and (2.31), we have

$$0 \le \limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n} \log Q_i(s) \le \frac{1}{2}s^2 e^{2|s|} c(\omega),\ \ \omega \in \mathcal{H} \cap A(s). \tag{2.32}$$

By (1.6), (2.25), and (2.32), we have

$$\limsup_n \frac{s}{\sigma_n}\sum_{i=1}^{\sigma_n} Y_i \le \gamma(\omega) + \frac{1}{2}s^2 e^{2|s|} c(\omega),\ \omega \in \mathcal{H} \cap \mathcal{J} \cap A(s). \tag{2.33}$$

For $s < 0$, dividing both sides of (2.33) by s, we find that

$$\liminf_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n} Y_i \ge \frac{\gamma(\omega)}{s} + \frac{1}{2}s e^{2|s|} c(\omega),\ \ \omega \in \mathcal{H} \cap \mathcal{J} \cap A(s). \tag{2.34}$$

Note that $\mathbb{P}(A(s)\cap\mathcal{H})=1$, taking superimum over $s<0$, we get (2.6).

For $s>0$, by (2.33), we have

$$\limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n} Y_i \le \frac{\gamma(\omega)}{s}+\frac{1}{2}se^{2|s|}c(\omega),\ \ \omega\in\mathcal{J}, \tag{2.35}$$

which proves (2.7). □

Corollary 2.1. *Under the conditions of Theorem 2.1, if $c(\omega)=0$ or $\gamma(\omega)=0, a.s.$ then*

$$\lim_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\frac{W_i-\mathbb{E}_\Pi[W_i|X_0,\ldots,X_{i-1}]}{a_i}=0,\ \ \mathbb{P}-a.s. \tag{2.36}$$

Our next result is to estimate the strong deviation of $\{f_i(X_0,\ldots,x_i)\}$ from its conditional means, which will require the method of conditional moment generating functions.

Theorem 2.2. *Under the above set up, if there is $\theta(\omega)>0\ a.s., 0<a<b$, such that*

$$\limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\mathbb{E}_\Pi\{\exp[|bZ_i|]|X_0,\ldots,X_{i-1}\}\le\theta(\omega).\ \ \mathbb{P}-a.s. \tag{2.37}$$

then, if $\rho(\gamma(\omega))<a$, we have

$$\limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\{Z_i-\mathbb{E}_\Pi[Z_i|X_0,\ldots,X_i]\}\le\frac{2\gamma(\omega)}{\rho(\gamma(\omega))},\ \ \mathbb{P}-a.s.,\omega\in\mathcal{J} \tag{2.38}$$

$$\liminf_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\{Z_i-\mathbb{E}_\Pi[Z_i|X_0,\ldots,X_i]\}\ge-\frac{2\gamma(\omega)}{\rho(\gamma(\omega))},\ \ \mathbb{P}-a.s.,\omega\in\mathcal{J} \tag{2.39}$$

Before proceeding with the proof, we need the following lemma.

Lemma 2.1. *Let a,b,c be positive constants and*

$$g(t,c)=\frac{2\theta t}{e^2(a-b)^2}+\frac{c}{t}\ , \tag{2.40}$$

where $\theta>0$. Then $g(t,c)$ attains its maximum value at $t=\rho(c)$, where $t=\rho(c)(0<|t|<a<b)$ is the unique solution of equation

$$2\theta t^2=e^2c(a-b)^2, \tag{2.41}$$

and

$$g(\rho(c),c)=\frac{2c}{\rho(c)}, \tag{2.42}$$

$$g(-\rho(c),c)=-\frac{2c}{\rho(c)}. \tag{2.43}$$

Proof of Theorem 2.2: Let t ($|t| < a$) be a real number, define

$$M_0(t) := \mathbb{E}_\Pi[\exp(tZ_0)]. \tag{2.44}$$

For $i \geq 1$, define

$$\begin{aligned} M_i(t; x_0, \ldots, x_i) &:= \mathbb{E}_\Pi[\exp\{tZ_i(\omega)\}|X_0 = x_0, \ldots, X_{i-1} = x_{i-1}] \\ &= \sum_{x_i \in S} \exp\{tf_i(x_0, \ldots, x_i)\}\pi_i(x_i|x_0, \ldots, x_{i-1}). \end{aligned} \tag{2.45}$$

$M_i(t; x_0, \ldots, x_i)$ is called the conditional moment generating function (CMGF) of $Z_i(\omega)$ given $X_0 = x_0, \ldots, X_{i-1} = x_{i-1}$. For each real number $t(|t| < a)$ and nonnegative integer i, we set

$$m_i(t; x_0, \ldots, x_i) := \frac{\exp\{tf_i(x_0, \ldots, x_i)\}\pi_i(x_i|x_0, \ldots, x_{i-1})}{M_i(t; x_0, \ldots, x_i)}. \tag{2.46}$$

As in Theorem 2.1, we define a set function ν_t on $\mathcal{N}$ as follows:

$$\nu_t(\phi) = 0, \ \ \nu_t(\Omega) = 1, \ \ \nu_t(D_{x_0}) = \pi_0(x_0), \tag{2.47}$$

$$\begin{aligned} \nu_t(D_{x_0,\ldots,x_n}) &= \nu_t(D_{x_0,\ldots,x_{n-1}})m_n(t; x_0, \ldots, x_n) \\ &= \pi_0(x_0)\prod_{i=1}^{n} m_i(t; x_0, \ldots, x_i). \end{aligned} \tag{2.48}$$

Let

$$L_n(t; \omega) := \frac{\nu_t(D_{X_0,\ldots,X_n})}{\mathbb{P}(D_{X_0,\ldots,X_n})}. \tag{2.49}$$

As in the proof of Theorem 2.1, we know that there exists a set $B(t) \in \sigma(\mathcal{N})$ with $\mathbb{P}(B(t)) = 1$, such that

$$\lim_n L_{\sigma_n}(t; \omega) = a\ finite\ number\ depend\ on\ \omega, \omega \in B(t), \tag{2.50}$$

so

$$\limsup_n \sigma_n^{-1} \log L_{\sigma_n}(t; \omega) \leq 0, \quad \omega \in B(t). \tag{2.51}$$

From (2.46), (2.47), and (2.49), we have

$$\begin{aligned} \frac{1}{\sigma_n} \log L_{\sigma_n}(t; \omega) &= \frac{t}{\sigma_n}\sum_{i=1}^{\sigma_n} Z_i - \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n} \log \mathbb{E}_\Pi[\exp\{tZ_i\}|X_0, \ldots, X_{i-1}] \\ &\quad - \frac{1}{\sigma_n}\log(p_n(X_0, \ldots, X_i)/[\pi_0(X_0)\prod_{i=1}^{\sigma_n}\pi_i(X_i|X_0, \ldots, X_{i-1})]). \end{aligned} \tag{2.52}$$

Combining $(2.51), (2.52)$ and (1.6), we have

$$\limsup\{\frac{t}{\sigma_n}\sum_{i=1}^{\sigma_n} Z_i - \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\log \mathbb{E}_\Pi[\exp\{tZ_i|X_0,\ldots,X_{i-1}]\} \leq \gamma(\omega), \ \ \omega\in B(t). \tag{2.53}$$

and

$$\begin{aligned}\limsup_n \frac{t}{\sigma_n}\sum_{i=1}^{\sigma_n}\{Z_i - \mathbb{E}_\Pi[Z_i|X_0,\ldots,X_{i-1}]\} \\ \leq \limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\{\log \mathbb{E}_\Pi[\exp tZ_i|X_0,\ldots,X_{i-1}] - \mathbb{E}_\Pi[tZ_i|X_0,\ldots,X_{i-1}]\} \\ + \gamma(\omega), \ \ \omega\in B(t). \end{aligned} \tag{2.54}$$

Note that

$$\limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\{\log \mathbb{E}_\Pi[e^{tZ_i}|X_0,\ldots,X_{i-1}] - \mathbb{E}_\Pi[tZ_i|X_0,\ldots,X_{i-1}]\}$$

$$\leq \limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\{[\mathbb{E}_\Pi[e^{tZ_i}|X_0,\ldots,X_{i-1}] - 1 - \mathbb{E}_\Pi[tZ_i|X_0,\ldots,X_{i-1}]\} \tag{2.55}$$

$$\leq \limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\mathbb{E}_\Pi[(e^{tZ_i} - 1 - tZ_i)|X_0,\ldots,X_{i-1}]$$

$$\leq \limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\mathbb{E}_\Pi[\frac{t^2}{2}Z_i^2 e^{|tZ_i|}|X_0,\ldots,X_{i-1}] \tag{2.56}$$

$$= \frac{t^2}{2}\limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\mathbb{E}_\Pi[e^{b|Z_i|}Z_i^2 e^{(|t|-b)|Z_i|}|X_0,\ldots,X_{i-1}]$$

$$\leq \frac{t^2}{2}\limsup_n \frac{1}{\sigma_n}\sum_{i=1}^{\sigma_n}\mathbb{E}_\Pi[e^{b|Z_i|}\frac{4e^{-2}}{(|t|-b)^2}|X_0,\ldots,X_{i-1}] \tag{2.57}$$

$$\leq \frac{2\theta(\omega)t^2}{e^2(a-b)^2}, \quad \mathbb{P}-a.s. \tag{2.58}$$

Here we have used the inequality $\log x \leq x-1$ (for $x>0$) in (2.55) , and $0 \leq e^x - 1 - x \leq x^2 e^{|x|}/2$ for all $x\in\mathbb{R}$ in (2.56). While (2.58) follows from the fact that $g(x) = x^2 e^{-hx}(x>0, h>0)$ attains its maximum value at $x = 2/h$.

By (2.55)-(2.58), we have

$$\limsup_n \frac{t}{\sigma_n}\sum_{i=1}^{\sigma_n}\{Z_i - \mathbb{E}_\Pi[Z_i|X_0,\ldots,X_{i-1}]\} \leq \frac{2\theta(\omega)t^2}{e^2(a-b)^2} + \gamma(\omega), \mathbb{P}-a.s.\, \omega\in\mathcal{J} \tag{2.59}$$

For $0 < t < a$, from Lemma 2.4 and inequality (2.59), we have

$$\limsup_n \frac{1}{\sigma_n} \sum_{i=1}^{\sigma_n} \{Z_i - \mathbb{E}_\Pi[Z_i|X_0, \ldots, X_{i-1}]\}$$
$$\leq \frac{2\theta(\omega)t}{e^2(a-b)^2} + \frac{\gamma(\omega)}{t} = g(t, \gamma(\omega)) \leq \frac{2\gamma(\omega)}{\rho(\gamma(\omega))}, \mathbb{P} - a.s.\ \omega \in \mathcal{J} \tag{2.60}$$

which proves (2.38).

For $-a < t < 0$, (2.39) can be proven in the same way. □

Corollary 2.2. *Under the conditions of Theorem 2.2, if there is a $\theta(\omega) > 0$ and $b > 0$, such that*

$$\limsup_n \frac{1}{\sigma_n} \sum_{i=1}^{\sigma_n} \mathbb{E}\{\exp[|bZ_i|]|X_0, \ldots, X_{i-1}\} \leq \theta(\omega).\ \ \mathbb{P} - a.s. \tag{2.61}$$

then

$$\lim_n \frac{1}{\sigma_n} \sum_{i=1}^{\sigma_n} \{Z_i - \mathbb{E}[Z_i|X_0, \ldots, X_i]\} = 0,\ \ \mathbb{P} - a.s. \tag{2.62}$$

Proof. In the proof of Theorem 2.2, if we take $\Pi = \mathbb{P}$, then (2.62) follows directly. □

In many statistical models, when assumptions cannot be safely made about the dependence structure of a model, a natural approach is to regard the data as coming from some specified distribution $\mathbb{P}$. But $\mathbb{P}$ may be difficult to work with, and practitioners may be led to use the production of marginals Π, which is an approximation. Suppose that $\tilde{x}^n = (x_0, \ldots, x_n)$ were generated from a distribution family denoted $\mathbb{P}(\tilde{x}^n)$, but that $\mathbb{P}(\tilde{x}^n)$ is not known in detail. One may try to model the data by using a different conditional distribution $\Pi(\tilde{x}^n) = \prod_{i=0}^n p_i(x_i)$, which assumes independence even when this is not valid. We take the $\pi's$ to be the marginal from $\mathbb{P}$, a nature starting choice. Here we examine the deviation, based on likelihood ratio, of the average $(1/n)\sum_{i=0}^n X_i$ from $(1/n)\sum_{i=0}^n E_\Pi X_i$.

Let $\{X_n, n \geq 0\}$ be a stochastic sequence on the probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$, taking values in S and with the joint distribution (1.1) and their marginal distributions are

$$p_i(x_i) = \mathbb{P}(X_i = x_i),\ \ x_i \in S. \tag{2.63}$$

Let Π be the measure of the product of the marginal distributions

$$\Pi(X_0 = x_0, \ldots, X_n = x_n) = \pi(x_0, \ldots, x_n) = \prod_{i=0}^n p_i(x_i). \tag{2.64}$$

Hence

$$\mathbb{E}_\Pi X_i = \sum_{x_i \in S} x_i p_i(x_i). \tag{2.65}$$

Corollary 2.3. *Under the assumptions of Theorem 2.2, let* Π *be as defined in (2.64), such that*

$$\limsup_{n} \frac{1}{\sigma_n} \sum_{i=1}^{\sigma_n} \mathbb{E}_\Pi\{\exp[|bX_i|]\} \leq \theta(\omega). \tag{2.66}$$

Then, for $\rho(\gamma(\omega)) < a$*, we have*

$$\limsup_{n} \frac{1}{\sigma_n} \sum_{i=1}^{\sigma_n} \{X_i - \mathbb{E}_\Pi X_i\} \leq \frac{2\gamma(\omega)}{\rho(\gamma(\omega))}, \ \ \mathbb{P} - a.s., \omega \in \mathcal{J} \tag{2.67}$$

$$\liminf_{n} \frac{1}{\sigma_n} \sum_{i=1}^{\sigma_n} \{X_i - \mathbb{E}_\Pi X_i\} \geq -\frac{2\gamma(\omega)}{\rho(\gamma(\omega))}, \ \ \mathbb{P} - a.s., \omega \in \mathcal{J}. \tag{2.68}$$

3 Limit Properties of Arbitrary Discrete Information Sources

A question of importance in information theory is the study on Shannon-mcmillan-Breiman theorem. In previous works, conditions such as ergodic,stationary or asymptotic stationary were assumed, see e.g. [1–3]. In this section we avoid these assumptions by the technique of CMGF and give a strong deviation theorem regarding the relative discrete information density and random conditional entropy, which holds for arbitrary discrete information sources. Before providing our next result, we review some basic concepts of information.

Let $\{X_n, n \geq 0\}$ be a sequence of successive letters produced by an arbitrary information source with the alphabet $S = \{t_0, t_1, \ldots, \}$ and with the joint distribution of (1.1), let

$$f_n(\omega) = -\frac{1}{n} \log p(X_1, \ldots, X_n), \tag{3.1}$$

where ω is a sample point, and the quantity $f_n(\omega)$ is called the $relative\ entropy\ density$ of $\{X_i, 1 \leq i \leq n\}$. Also let Π be another information source with the joint distribution (1.3)

Definition 3.1. For $i \geq 1$, let

$$h_i(x_1, \ldots, x_{i-1}) = -\sum_{x_i \in S} \pi(x_i|x_0, \ldots, x_{i-1}) \log \pi(x_i|x_0, \ldots, x_{i-1}), \tag{3.2}$$

$$H_i(\omega) = h_i(x_1, \ldots, x_{i-1}). \tag{3.3}$$

$H_i(\omega)$ is called the random conditional entropy of $\{X_i, 0 \leq i \leq n\}$.

Theorem 3.1. *Let* $\{X_n, n \geq 0\}$ *be a sequence of succesive letters produced by an arbitrary information source with the alphabet* S *and the joint distribution (1.1),* $f_n(\omega), \gamma(\omega)$ *and*

$H_i(\omega)$ be defined in (3.1), (1.6) and (3.3). If there is $b > a > 0$ such that

$$\limsup_n \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\Pi}\{[p(X_i|X_0,\ldots,X_{i-1})]^b|X_0,\ldots,X_{i-1}\} \le \theta(\omega).\ \mathbb{P}-a.s. \tag{3.4}$$

then, for $\rho(\gamma(\omega)) < a$, we have

$$\limsup_n\{f_n(\omega) - \frac{1}{n}\sum_{i=1}^{n} H_i(\omega)\} \le \frac{2\gamma(\omega)}{\rho(\gamma(\omega))},\ \ \mathbb{P}-a.s., \omega \in \mathcal{J}, \tag{3.5}$$

$$\liminf_n\{f_n(\omega) - \frac{1}{n}\sum_{i=1}^{n} H_i(\omega)\} \ge -\frac{2\gamma(\omega)}{\rho(\gamma(\omega))},\ \ \mathbb{P}-a.s., \omega \in \mathcal{J}. \tag{3.6}$$

Proof. In Theorem 2.2, let $f_i(x_0,\ldots,x_n) = \log p_i(x_i|x_0,\ldots,x_{i-1})$, (3.5) and (3.6) follow immediately from (2.38) and (2.39). □

References

[1] P. H. Algoet and T. M. Cover, A sandwich proof of the Shannon-McMillan-Breiman theorem, *Ann. Probab.* **16** (1988), 899–909.

[2] A. R. Barron, The strong ergodic theorem for densities: generalized Shannon-McMillan-Breiman theorem, *Ann. Probab.* **13** (1985), 1292–1303.

[3] K. L. Chung, The ergodic theorem of information thoery, *Ann. Math. Statist.* **32** (1961), 612–614.

[4] C. Jardas, J. Pecaric and N. Sarapa, A note on Chung's law of large numbers, *J. Math. Analysis Appl.* **217** (1998), 328–334.

[5] W. Liu, A kind of strong deviation theorems for the sequence of nonnegative integer-valued random variables, *Stat. Probab. Letts.* **32** (1997), 269–276.

[6] W. Liu, Some limit properties of multivariate function sequence of discrete random variables, *Stat. Probab. Letts.* **61** (2003), 41–50.

[7] W. Liu and W. G. Yang, The Markov approximation of the sequences of N-valued random variables and a class of small deviation theorems, *Stochastic Process. Appl.* **89** (2000), 117–130.

[8] K. R. Stromberg and E. Hewitt, *Real and Abstract Analysis – A Mordern Treament of the Theory of Functions of Real Variable*, Springer, New York, 1994.

[9] Z. Z. Wang, A class of random deviation theorems for sums of nonnegative stochastic sequence and strong law of large numbers, *Stat. Probab. Letts.* **76** (2006), 2017–2026.