The Annals of Probability
1990, Vol. 18, No. 2, 829–839

RELATIVE ENTROPY DENSITIES AND A CLASS OF LIMIT THEOREMS OF THE SEQUENCE OF m-VALUED RANDOM VARIABLES

BY LIU WEN

Hebei Institute of Technology

Let $\{X_n,\ n \geq 1\}$ be a sequence of random variables taking values in $S = \{1, 2, \dots, m\}$ with distribution $p(x_1, \dots, x_n)$, $(p_{i1}, p_{i2}, \dots, p_{im})$, $i = 1, 2, \dots,$ a sequence of probability distributions on S, and $\varphi_n = (1/n)\log p(X_1, \dots, X_n) - (1/n)\sum_{i=1}^n \log p_{iX_i}$ the entropy density deviation, relative to the distribution $\prod_{i=1}^n p_{ix_i}$, of $\{X_i,\ 1 \leq i \leq n\}$. In this paper the relation between the limit property of φ_n and the frequency of given values in $\{X_n\}$ is studied.

Let $\{X_n,\ n \geq 1\}$ be a sequence of random variables taking values in $S = \{1, 2, \dots, m\}$ with the joint distribution

$$(1)\quad P(X_1 = x_1, \dots, X_n = x_n) = p(x_1, \dots, x_n) > 0, \qquad x_i \in S, 1 \leq i \leq n.$$

Let

$$(2)\qquad f_n(\omega) = -(1/n)\log p(X_1, \dots, X_n),$$

where log is the natural logarithm. $f_n(\omega)$ is called the relative entropy density of $\{X_k,\ 1 \leq k \leq n\}$ (see [1]). A question of importance in information theory is the nature and existence in some sense, of $f_n(\omega)$ (cf. [1]–[4] and [6]–[8]). The purpose of this paper is to study the relation between the relative entropy density and a class of limit theorems.

Let $\{X_n,\ n \geq 1\}$ be a sequence of random variables with distribution (1). Then $\{X_n,\ n \geq 1\}$ are independent if and only if there exists a sequence of probability distributions on S,

$$(3)\qquad (p_{i1}, p_{i2}, \dots, p_{im}), \qquad p_{ik} > 0, k = 1, 2, \dots,$$

such that

$$(4)\qquad p(x_1, \dots, x_n) = \prod_{i=1}^n p_{ix_i}.$$

In this case, we have

$$(5)\qquad p(X_n = i) = p_{ni},$$

$$(6)\qquad f_n(\omega) = -(1/n)\sum_{i=1}^n \log p_{iX_i}, \qquad n = 1, 2, \dots.$$

Received November 1988; revised March 1989.
AMS 1980 *subject classifications*. Primary 60F15; secondary 94A17.
Key words and phrases. Limit theorem, entropy, relative entropy density, almost stationary sequences.

DEFINITION. Let $\{X_n,\ n \geq 1\}$ be a sequence of random variables with distribution (1), and let (3) be a sequence of probability distribution on S. The difference

$$\varphi_n(\omega) = (1/n)\log p(X_1, \dots, X_n) - (1/n)\sum_{i=1}^{n} \log p_{iX_i} \tag{7}$$

will be called the entropy density deviation, relative to the distribution (4) of independent type, of $\{X_i,\ 1 \leq i \leq n\}$.

LEMMA. *Assume $c \geq 0$ is a constant, and let*

$$g(c, \lambda) = \frac{1}{\log \lambda}(\lambda - 1 + c) - 1, \qquad \lambda > 0, \lambda \neq 1. \tag{8}$$

Then if $c > 0$, $g(c, \lambda)$ (as a function of λ) attains its smallest value on the interval $(1, \infty)$ at $\lambda = \beta(c) \in (1, \infty)$, where $\beta(c)$ is the unique solution of the equation

$$\lambda(\log \lambda - 1) + 1 = c \tag{9}$$

on the interval $(1, \infty)$; if $0 < c < 1$, $g(c, \lambda)$ attains its largest value on the interval $(0, 1)$ at $\alpha(c) \in (0, 1)$, where $\alpha(c)$ is the unique solution of (9) on the interval $(0, 1)$. Moreover

$$g(c, \alpha(c)) = \alpha(c) - 1, \tag{10}$$

$$g(c, \beta(c)) = \beta(c) - 1, \tag{11}$$

$$\lim_{c \to 0^+} \alpha(c) = 1, \tag{12}$$

$$\lim_{c \to 0^+} \beta(c) = 1. \tag{13}$$

PROOF. Differentiating $g(c, \lambda)$ with respect to λ, we obtain

$$g'(c, \lambda) = [\lambda(\log \lambda - 1) + 1 - c]/\lambda(\log \lambda)^2. \tag{14}$$

Letting $g'(c, \lambda) = 0$, (9) follows. Let

$$\psi(\lambda) = \lambda(\log \lambda - 1) + 1, \qquad \lambda > 0. \tag{15}$$

Since $\psi(\lambda)$ is increasing on the interval $(1, \infty)$, and $\psi(1) = 0$, $\lim_{\lambda \to \infty} \psi(\lambda) = \infty$, so (9) has unique solution $\beta(c)$ on $(1, \infty)$. It is obvious that g attains its smallest value on the interval $(1, \infty)$ at $\lambda = \beta(c)$. Moreover, by (9),

$$\log \beta(c) = 1 + (c - 1)/\beta(c). \tag{16}$$

By (16) and (8), (11) follows. Since $\psi(\lambda)$ is decreasing on the interval $(0, 1)$, and $\psi(1) = 0$, $\lim_{\lambda \to 0^+} \psi(\lambda) = 1$, so (9) has unique solution $\alpha(c)$ on $(0, 1)$, and

g attains its largest value on $(0, 1)$ at $\lambda = \alpha(c)$. Similarly, by (9) and (8), (10) follows. Since $\psi(1) = 1$ by the continuity of ψ, (12) and (13) follow. □

THEOREM 1. *Let $\{X_n, n \geq 1\}$ be a sequence of random variables with distribution* (1), $k \in S$, *$S_n(k, \omega)$ be the number of k in the sequence $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$, $\varphi_n(\omega)$ be defined by* (7), *and c be a nonnegative constant. Let*

$$b_k = \limsup_n (1/n) \sum_{i=1}^{n} p_{ik}, \tag{17}$$

$$D(c) = \left\{\omega: \limsup_n \varphi_n(\omega) \leq c\right\}. \tag{18}$$

Then:

(a) *If $c \geq 0$ and $b_k > 0$, then*

$$\limsup_n (1/n)\left[S_n(k, \omega) - \sum_{i=1}^{n} p_{ik}\right] \leq b_k[\beta(c/b_k) - 1] \quad a.e., \ \omega \in D(c), \tag{19}$$

where $\beta(0) = 1$, and $\beta(c/b_k)$ is defined as in the lemma if $c > 0$.

(b) *If $0 \leq c < b_k$ and $b_k > 0$, then*

$$\liminf_n (1/n)\left[S_n(k, \omega) - \sum_{i=1}^{n} p_{ik}\right] \geq b_k[\alpha(c/b_k) - 1] \quad a.e., \ \omega \in D(c), \tag{20}$$

where $\alpha(0) = 1$, and $\alpha(c/b_k)$ is defined as in the lemma if $c > 0$.

(c) *If $c \geq 0$, then*

$$\liminf_n (1/n)\left[S_n(k, \omega) - \sum_{i=1}^{n} p_{ik}\right] \geq -b_k \quad a.e., \ \omega \in D(c). \tag{21}$$

(d) *If $c \geq 0$ and $b_k = 0$, then*

$$\lim_n (1/n)\left[S_n(k, \omega) - \sum_{i=1}^{n} p_{ik}\right] = 0 \quad a.e., \ \omega \in D(c). \tag{22}$$

PROOF. Throughout this paper we shall deal with the underlying probability space $([0, 1), \mathscr{F}, P)$, where $\mathscr{F}$ is the class of Lebesgue measurable sets in the interval $[0, 1)$, and P is the Lebesgue measure. We first give, in the above probability space, a realization of the sequence of random variables with distribution (1).

Divide the interval $[0, 1)$ into a m right-semiopen intervals:

$$\delta_1 = [0, p(1)), \delta_2 = [p(1), p(1) + p(2)), \ldots, \delta_m = [1 - p(m), 1).$$

These intervals will be called intervals of the first order. Proceeding inductively, suppose the m^n nth order intervals $\{\delta_{x_1 \cdots x_n}, x_i = 1, 2, \ldots, m, 1 \leq i \leq n\}$ have been defined. Then dividing the right-semiopen interval $\delta_{x_1 \cdots x_n}$ into m

right-semiopen intervals $\delta_{x_1 \cdots x_n 1}, \ldots, \delta_{x_1 \cdots x_n 2}, \ldots, \delta_{x_1 \cdots x_n m}$ according to the ratio

$$p(x_1, \ldots, x_n, 1) : p(x_1, \ldots, x_n, 2) : \cdots : p(x_1, \ldots, x_n, m),$$

the intervals of the $(n+1)$st order are created. It is easy to see that for $n \geq 1$,

$$P(\delta_{x_1 \cdots x_n}) = p(x_1, \ldots, x_n). \tag{23}$$

Define, for $n \geq 1$, a random variable X_n: $[0, 1) \to S$ as follows:

$$X_n(\omega) = x_n, \quad \text{if } \omega \in \delta_{x_1 \cdots x_n}. \tag{24}$$

By (23) and (24),

$$\{\omega\colon X_1 = x_1, \ldots, X_n = x_n\} = \delta_{x_1 \cdots x_n},$$

$$P(X_1 = x_1, \ldots, X_n = x_n) = p(x_1, \ldots, x_n);$$

hence $\{X_n, n \geq 1\}$ has distribution (1).

Let the collection of intervals of all orders [including the zeroth order interval $[0, 1)$] be denoted by $\mathscr{A}$, and $\lambda > 0$ be a constant. Define a set function μ on $\mathscr{A}$ as follows: Let $\mu([0, 1)) = 1$. Assume $\delta_{x_1 \cdots x_n}$ is an interval of nth order, $k \in S$ and $s_n(k)$ is the number of k in $x_1, x_2, \ldots, x_n$. Then let

$$\mu(\delta_{x_1 \cdots x_n}) = \lambda^{s_n(k)} \prod_{i=1}^{n} \left[p_{i x_i} / (1 + (\lambda - 1) p_{ik}) \right]. \tag{25}$$

Define, for $1 \leq k \leq m$, m functions I_k: $S \to \{0, 1\}$ as follows:

$$I_k(s) = \begin{cases} 1, & \text{as } s = k, \\ 0, & \text{as } s \neq k. \end{cases} \tag{26}$$

It is easy to see that

$$I_j(x_i) = 1, \qquad \sum_{i=1}^{m} I_j(i) = 1, \tag{27}$$

$$s_n(k) = \sum_{i=1}^{n} I_k(X_i), \tag{28}$$

$$s_n(k) = s_{n-1}(k) + I_k(x_n) \quad [\text{let } s_0(k) = 0]. \tag{29}$$

By (25) and (29), we have for $n > 1$,

$$\begin{aligned} \sum_{x_n=1}^{m} \mu(\delta_{x_1 \cdots x_n}) &= \left(\lambda^{s_{n-1}(k)} \prod_{i=1}^{n-1} \frac{p_{i x_i}}{1 + (\lambda - 1) p_{ik}} \right) \left(\sum_{x_n=1}^{m} \lambda^{I_k(x_n)} \frac{p_{n x_n}}{1 + (\lambda - 1) p_{nk}} \right) \\ &= \mu(\delta_{x_1 \cdots x_{n-1}}) \left[\sum_{x_n \neq k} \frac{p_{n x_n}}{1 + (\lambda - 1) p_{nk}} + \frac{p_{nk}}{1 + (\lambda - 1) p_{nk}} \right] \\ &= \mu(\delta_{x_1 \cdots x_{n-1}}). \end{aligned} \tag{30}$$

Similarly,

$$\sum_{x_1=1}^{m} \mu(\delta_{x_1}) = 1 = \mu([0,1)). \tag{31}$$

By (30) and (31), it is easy to see that μ is an additive set function on $\mathscr{A}$. Hence there exists an increasing function f_λ defined on $[0,1)$ such that, for any $\delta_{x_1 \cdots x_n}$,

$$\mu(\delta_{x_1 \cdots x_n}) = f_\lambda(\delta^+_{x_1 \cdots x_n}) - f_\lambda(\delta^-_{x_1 \cdots x_n}), \tag{32}$$

where $\delta^-_{x_1 \cdots x_n}$ and $\delta^+_{x_1 \cdots x_n}$ denote, respectively, the left and right endpoints of $\delta_{x_1 \cdots x_n}$. Let

$$t_n(\lambda, \omega) = \frac{\mu(\delta_{x_1 \cdots x_n})}{P(\delta_{x_1 \cdots x_n})} = \frac{f_\lambda(\delta^+_{x_1 \cdots x_n}) - f_\lambda(\delta^-_{x_1 \cdots x_n})}{\delta^+_{x_1 \cdots x_n} - \delta^-_{x_1 \cdots x_n}}, \qquad \omega \in \delta_{x_1 \cdots x_n}. \tag{33}$$

Let $A_k(\lambda)$ be the set of points of differentiability of f_λ. Then (cf. [5], page 345)

$$\lim_n t_n(\lambda, \omega) = \text{finite number}, \qquad \omega \in A_k(\lambda), \tag{34}$$

and $P(A_k(\lambda)) = 1$ by the theorem on the existence of the derivative of a monotone function. By (34),

$$\limsup_n (1/n) \log t_n(\lambda, \omega) \le , \qquad \omega \in A_k(\lambda). \tag{35}$$

By (23), (25) and (33),

$$t_n(\lambda, \omega) = \left[\lambda^{S_n(k,\omega)} \prod_{i=1}^{n} \frac{p_{iX_i}}{1 + (\lambda - 1)p_{ik}} \right] \Big/ p(X_1, \dots, X_n), \qquad \omega \in [0,1) \tag{36}$$

We have by (36) and (7),

$$\frac{1}{n} \log t_n(\lambda, \omega) = \frac{S_n(k,\omega)}{n} \log \lambda - \frac{1}{n} \sum_{i=1}^{n} \log(1 + (\lambda - 1)p_{ik}) - \varphi_n(\omega), \qquad \omega \in [0,1). \tag{37}$$

By (37) and (35),

$$\limsup_n \left[\frac{S_n(k,\omega)}{n} \log \lambda - \frac{1}{n} \sum_{i=1}^{n} \log(1 + (\lambda - 1)p_{ik}) - \varphi_n(\omega) \right] \le 0, \qquad \omega \in A_k(\lambda). \tag{38}$$

(a) Letting $\lambda > 1$, and dividing the two sides of (38) by $\log \lambda$, we have

$$\limsup_n \left[-\frac{1}{n} S_n(k,\omega) - \frac{1}{n} \sum_{i=1}^n \frac{\log(1+(\lambda-1)p_{ik})}{\log \lambda} - \frac{\varphi_n(\omega)}{\log \lambda} \right]$$

$$(39)\qquad \begin{aligned} &= \limsup_n \left\{ \frac{1}{n} \left[S_n(k,\omega) - \sum_{i=1}^n p_{ik} \right] \right\} \\ &\quad - \frac{1}{n} \sum_{i=1}^n \frac{\log(1+(\lambda-1)p_{ik})}{\log \lambda} - p_{ik} - \frac{\varphi_n(\omega)}{\log \lambda} \\ &\le 0, \qquad \omega \in A_k(\lambda). \end{aligned}$$

By (39), (17), (18) and the inequality $\log(1+x) \le x$, $x \ge 0$, we have

$$(40)\qquad \begin{aligned} &\limsup_n \frac{1}{n} \left[S_n(k,\omega) - \sum_{i=1}^n p_{ik} \right] \\ &\quad \le \limsup_n \frac{1}{n} \sum_{i=1}^n \left[\frac{(\lambda-1)p_{ik}}{\log \lambda} - p_{ik} \right] + \frac{c}{\log \lambda} \\ &\quad = \left(\frac{\lambda-1}{\log \lambda} - 1 \right) \limsup_n \frac{1}{n} \sum_{i=1}^n p_{ik} + \frac{c}{\log \lambda} \\ &\quad = b_k \left(\frac{\lambda-1}{\log \lambda} - 1 \right) + \frac{c}{\log \lambda}, \qquad \omega \in A_k(\lambda) \cap D(c). \end{aligned}$$

If $b_k > 0$, then we have by (40) and (8),

$$(41)\qquad \limsup_n (1/n) \left[S_n(k,\omega) - \sum_{i=1}^n p_{ik} \right] \le b_k g(c/b_k, \lambda),$$
$$\omega \in A_k(\lambda) \cap D(c).$$

Letting, in the case $c > 0$, $\lambda = \beta(c/b_k)$, we have by use of (11),

$$(42)\qquad \limsup_n (1/n) \left[S_n(k,\omega) - \sum_{i=1}^n p_{ik} \right] \le b_k[\beta(c/b_k) - 1],$$
$$\omega \in A_k(\beta(c/b_k)) \cap D(c).$$

Since $P(A_k(\lambda)) = 1$, (19) follows from (42). In the case $c = 0$, choose $\lambda_i > 1$, $i = 1, 2, \ldots,$ such that $\lambda_i \to 1$ (as $i \to \infty$), and let

$$H^*(k) = \bigcap_{i=1}^\infty (A_k(\lambda_i) \cap D(0)):$$

Then for all $i \ge 1$, we have by (41),

$$(43)\quad \limsup_n (1/n) \left[S_n(k,\omega) - \sum_{i=1}^n p_{ik} \right] \le b_k g(0, \lambda_i), \qquad \omega \in H^*(k).$$

Since $\lim_{i\to\infty} g(0, \lambda_i) = 0$, we have by (43),

$$\limsup_n (1/n)\left[S_n(k,\omega) - \sum_{i=1}^n p_{ik}\right] \le 0, \qquad \omega \in H^*(k). \tag{44}$$

Since $H^*(k) \subset D(0)$ and $P(H^*(k)) = P(D(0))$, (19) follows from (44) as $c = 0$.

(b) Letting $0 < \lambda < 1$, and dividing the two sides of (38) by $\log \lambda$, we have

$$\begin{aligned}
&\liminf_n \left\{\frac{1}{n}\left[S_n(k,\omega) - \sum_{i=1}^n \frac{\log(1 + (\lambda - 1)p_{ik})}{\log \lambda}\right] - \frac{\varphi_n(\omega)}{\log \lambda}\right\} \\
&\quad = \liminf_n \left\{\frac{1}{n}\left[S_n(k,\omega) - \sum_{i=1}^n p_{ik}\right]\right. \\
&\qquad \left. - \frac{1}{n}\sum_{i=1}^n \left[\frac{\log(1 + (\lambda - 1)p_{ik})}{\log \lambda} - p_{ik}\right] - \frac{\varphi_n(\omega)}{\log \lambda}\right\} \\
&\quad \ge 0, \qquad \omega \in A_k(\lambda).
\end{aligned} \tag{45}$$

By (45), (18), (17) and the inequalities $\log(1 + x) \le x$, $-1 < x \le 0$, and $0 < (\lambda - 1)/\log \lambda < 1$, $0 < \lambda < 1$, we have

$$\begin{aligned}
&\liminf_n \frac{1}{n}\left[S_n(k,\omega) - \sum_{i=1}^n p_{ik}\right] \\
&\quad \ge \liminf_n \frac{1}{n}\sum_{i=1}^n \left[\frac{\log(1 + (\lambda - 1)p_{ik})}{\log \lambda} - p_{ik}\right] + \liminf_n \frac{\varphi_n(\omega)}{\log \lambda} \\
&\quad \ge \liminf_n \frac{1}{n}\sum_{i=1}^n \left[\frac{\log(1 + (\lambda - 1)p_{ik})}{\log \lambda} - p_{ik}\right] + \liminf_n \frac{\varphi_n(\omega)}{\log \lambda} \\
&\quad \ge \liminf_n \left(\frac{\lambda - 1}{\log \lambda} - 1\right)\frac{1}{n}\sum_{i=1}^n p_{ik} + \frac{c}{\log \lambda} \\
&\quad = \left(\frac{\lambda - 1}{\log \lambda} - 1\right)b_k + \frac{c}{\log \lambda}, \qquad \omega \in A_k(\lambda) \cap D(c).
\end{aligned} \tag{46}$$

If $b_k > 0$, then we have by (46) and (8),

$$\liminf_n (1/n)\left[S_n(k,\omega) - \sum_{i=1}^n p_{ik}\right] \ge b_k g(c/b_k, \lambda), \tag{47}$$

$$\omega \in A_k(\lambda) \cap D(c).$$

Letting, in the case $0 < c < b_k$, $\lambda = \alpha(c/b_k)$, we have by (10),

$$\liminf_n (1/n)\left[S_n(k,\omega) - \sum_{i=1}^n p_{ik}\right] \ge b_k[\alpha(c/b_k) - 1], \tag{48}$$

$$\omega \in A_k(\alpha(c/b_k)) \cap D(c).$$

Since $P(A(\lambda)) = 1$, (20) follows from (48). In the case $c = 0$, choose $\tau_i \in (0, 1)$, $i = 1, 2, \ldots,$ such that $\tau_i \to 1$ (as $i \to \infty$), and let

$$H^*(k) = \bigcap_{i=1}^{\infty} (A_k(\tau_i) \cap D(0)).$$

Then for all $i \geq 1$, we have by (47),

$$\text{(49)} \quad \liminf_n (1/n)\left[S_n(k, \omega) - \sum_{i=1}^{n} p_{ik}\right] \geq b_k g(0, \tau_i), \qquad \omega \in H_*(k).$$

Since $\lim_i g(0, \tau_i) = 0$, we have by (49),

$$\text{(50)} \qquad \liminf_n (1/n)\left[S_n(k, \omega) - \sum_{i=1}^{n} p_{ik}\right] \geq 0, \qquad \omega \in H_*(k).$$

Since $H_*(k) \subset D(0)$ and $P(H_*(k)) = P(D(0))$, (20) follows from (50) as $c = 0$.

(c) For arbitrary $c \geq 0$, choose $\lambda_i \in (0, 1)$, $i = 1, 2, \ldots,$ such that $\lambda_i \to 0$ (as $i \to \infty$), and let $A = \bigcap_{i=1}^{\infty}(A(\lambda_i) \cap D(c))$. Then for all $i \geq 1$, we have by (46),

$$\text{(51)} \qquad \liminf_n \frac{1}{n}\left[S_n(k, \omega) - \sum_{i=1}^{n} p_{ik}\right] \geq \left(\frac{\lambda_i - 1}{\log \lambda_i - 1}\right) b_k + \frac{c}{\log \lambda_i}, \qquad \omega \in A.$$

Since

$$\lim_i \left[\left(\frac{\lambda_i - 1}{\log \lambda_i} - 1\right) b_k + \frac{c}{\log \lambda_i}\right] = -b_k,$$

it follows from (51) that

$$\text{(52)} \qquad \liminf_n (1/n)\left[S_n(k, \omega) - \sum_{i=1}^{n} p_{ik}\right] \geq -b_k, \qquad \omega \in A.$$

Since $A \subset D(c)$ and $P(A) = P(D(c))$, (21) follows from (52).

(d) In the case $b_k = 0$, choose $\lambda_i \in (0, 1)$, $\tau_i \in (1, \infty)$ such that $\lambda_i \to 0$, $\tau_i \to \infty$ (as $i \to \infty$), and let $B = \bigcap_{i=1}^{\infty}[A(\lambda_i) \cap A(\tau_i) \cap D(c)]$. Then for all $i \geq 1$ we have by (40) and (46),

$$\text{(53)} \qquad \limsup_n \frac{1}{n}\left[S_n(k, \omega) - \sum_{i=1}^{n} p_{ik}\right] \leq \frac{c}{\log \tau_i}, \qquad \omega \in B,$$

$$\text{(54)} \qquad \liminf_n \frac{1}{n}\left[S_n(k, \omega) - \sum_{i=1}^{n} p_{ik}\right] \geq \frac{c}{\log \lambda_i}, \qquad \omega \in B.$$

Since

$$\lim_i (c/\log \tau_i) = \lim_i (c/\log \lambda_i) = 0,$$

we have by (53) and (54),

$$\lim_n (1/n)\left[S_n(k,\omega) - \sum_{i=1}^n p_{ik}\right] = 0, \qquad \omega \in B. \tag{55}$$

Since $B \subset D(c)$ and $P(B) = P(D(c))$, (22) follows from (55). □

COROLLARY. *Under the assumption of Theorem* 1,

$$\lim_n (1/n)\left[S_n(k,\omega) - \sum_{i=1}^n p_{ik}\right] = 0 \quad a.e., \omega \in D(0). \tag{56}$$

PROOF. Letting $c = 0$ in the Theorem 1, (56) follows from (19) and (20). □

In case of a stationary sequence $-(1/n)\log p(X_1, \dots, X_n)$ tends to the entropy of the process, and this implies that $P(D(0)) = 1$, i.e., Theorem 1 is applicable with $c = 0$. Similar behavior can be expected in the cases of almost stationary sequences.

It will be shown in the following remarks that the conditions of Theorem 1, i.e., $P(D(c)) > 0$ can be satisfied in some nontrivial cases.

REMARK 1. Assume $m = 2$, $p_{ik} = \frac{1}{2}$, $k = 1,2$; $i = 1,2,\dots,$ and let

$$d_n = \max p(x_1, \dots, x_n), \qquad x_i = 1,2; i = 1,2,\dots,n.$$

If

$$\limsup_n d_n^{1/n} \le \tfrac{1}{2}e^c, \tag{57}$$

then $D(c) = [0,1)$.

PROOF. Let

$$r_n(\omega) = p(X_1, \dots, X_n) \Big/ \prod_{i=1}^n p_{iX_i}, \qquad \omega \in [0,1). \tag{58}$$

Then

$$D(c) = \left\{\omega, \limsup_n \left[r_n(\omega)\right]^{1/n} \le e^c\right\}. \tag{59}$$

By the assumptions and (58),

$$r_n(\omega) = 2^n p(X_1, \dots, X_n) \le 2^n d_n \tag{60}$$

and $D(c) = [0, 1)$ follows by (60), (57) and (59). □

REMARK 2. Assume $m = 2$, $p_{ik} = \frac{1}{2}$, $k = 1,2$; $i = 1,2,\dots$. If

$$\limsup_n p(X_1, \dots, X_{n+1})/p(X_1, \dots, X_n) \le \tfrac{1}{2}e^c \quad \text{a.e.}, \tag{61}$$

then $P(D(c)) = 1$.

PROOF. Since $\limsup a_n^{1/n} \le \limsup a_{n+1}/a_n$ for positive real numbers a_n, $n \ge 1$, we have by (58) and (61),

$$\begin{aligned}\limsup_n [r_n(\omega)]^{1/n} &= 2 \limsup [p(X_1, \dots, X_n)]^{1/n} \\ &\le 2 \limsup_n p(X_1, \dots, X_{n+1})/p(X_1, \dots, X_n) \le e^c \quad \text{a.e.,}\end{aligned} \tag{62}$$

and $P(D(c)) = 1$ follows by (62) and (59). □

THEOREM 2. *Under the assumptions of Theorem* 1, *if* $c \ge 0$, $b_k > 0$, *then*

$$\limsup_n (1/n)\left[S_n(k, \omega) - \sum_{i=1}^n p_{ik}\right] \le 2\sqrt{b_k c} + c \quad a.e., \omega \in D(c); \tag{63}$$

if $0 \le c < b_k$, $b_k > 0$, *then*

$$\limsup_n (1/n)\left[S_n(k, \omega) - \sum_{i=1}^n p_{ik}\right] \ge -2\sqrt{b_k c} \quad a.e., \omega \in D(c). \tag{64}$$

PROOF. Letting $\lambda > 1$, by use of the inequality $\log \lambda > 1 - 1/\lambda$, $\lambda > 1$, it follows from (40) that

$$\begin{aligned}\limsup_n \frac{1}{n}\left[S_n(k, \omega) - \sum_{i=1}^n p_{ik}\right] &\le b_k\left(\frac{\lambda - 1}{1 - 1/\lambda} - 1\right) + \frac{c}{1 - 1/\lambda} \\ &= b_k(\lambda - 1) + \frac{c}{\lambda - 1}, \qquad \omega \in A_k(\lambda) \cap D(c).\end{aligned} \tag{65}$$

It is easy to see that if $c > 0$ and $b_k > 0$, then the function $\psi(\lambda) = b_k(\lambda - 1) + c\lambda/(\lambda - 1)$, $\lambda > 1$, attains, at $\lambda = 1 + \sqrt{c/b_k}$, its smallest value $\psi(1 + c/b_k) = 2\sqrt{b_k c} + c$ on the interval $(1, \infty)$. Letting $\lambda = 1 + c/b_k$ in (65), it follows that

$$\begin{aligned}\limsup_n (1/n)\left[S_n(k, \omega) - \sum_{i=1}^n p_{ik}\right] &\le 2\sqrt{b_k c} + c, \\ &\omega \in A_k\left(1 + \sqrt{c/b_k}\right) \cap D(c).\end{aligned} \tag{66}$$

Since $P(A_k(\lambda)) = 1$, (63) follows from (66).

Letting $0 < \lambda < 1$, by use of the inequalities $1 - 1/\lambda < \log \lambda < 0$ and $\log \lambda < \lambda - 1 < 0$, $0 < \lambda < 1$, it follows from (46) that

$$\begin{aligned}\limsup_n \frac{1}{n}\left[S_n(k, \omega) - \sum_{i=1}^n p_{ik}\right] &\ge \left(\frac{\lambda - 1}{1 - 1/\lambda} - 1\right) b_k + \frac{c}{\lambda - 1} \\ &= b_k(\lambda - 1) + \frac{c}{\lambda - 1}, \\ &\omega \in A(\lambda) \cap D(c).\end{aligned} \tag{67}$$

It is easy to see that if $0 < c < b_k$ and $b_k > 0$, then the function $h(\lambda) = b_k(\lambda - 1) + c/(\lambda - 1)$, $0 < \lambda < 1$, attains, at $\lambda = 1 - \sqrt{c/b_k}$, its largest value $h(1 - \sqrt{c/b_k}) = -2\sqrt{b_k c}$ on the interval (0, 1). Letting $\lambda = 1 - \sqrt{c/b_k}$ in (67), it follows that

$$\liminf_n (1/n)\left[S_n(k,\omega) - \sum_{i=1}^{n} p_{ik}\right] \geq -2\sqrt{b_k c}, \tag{68}$$

$$\omega \in A_k\left(1 - \sqrt{c/b_k}\right) \cap D(c).$$

Since $P(A_k(\lambda)) = 1$, (64) follows from (68).

By (56) and (22), (63) and (64) are also true as $c = 0$ or $b_k = 0$. □

REFERENCES

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

[2] Breiman, L. (1957). The individual ergodic theorem of information theory. *Ann. Math. Statist.* **28** 809–811.

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

[4] Feinstein, A. (1954). A new basic theory of information theory. *IEEE Trans. P.G.I.T.* 2–22.

[5] Hildebrandt, T. H. (1963). *Introduction to the Theory of Integration*. Academic, New York.

[6] Kiefer, J. C. (1974). A simple proof of the Mog–Perez generalization of the Shannon–McMillan theorem. *Pacific J. Math.* **51** 203–204.

[7] McMillan, B. (1953). The basic theorems of information theory. *Ann. Math. Statist.* **24** 169–219.

[8] Shannon, C. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27** 379–423, 623–656.

Department of Mathematics
Hebei Institute of Technology
Tianjin, 300130
People's Republic of China