TAIWANESE JOURNAL OF MATHEMATICS
Vol. 3, No. 4, pp. 539-546, December 1999

A LIMIT PROPERTY OF ARBITRARY DISCRETE INFORMATION SOURCES

Liu Wen

Abstract. In this paper the notion of stochastic conditional entropy is introduced, and the asymptotic relation between this notion and the relative entropy density is studied. A strong limit theorem which holds for arbitrary discrete information sources is obtained. In the proof an analytic technique to study the strong limit properties of discrete information sources is proposed.

1. Introduction

A question of importance in information theory is the study on the Shannon-McMillan-Breiman theorem. In previous works, conditions such as ergodicity, stationarity or asymptotic stationarity were assumed (cf. [1]-[2], [4]-[9]). In this paper we avoid these assumptions and give a strong limt theorem concerning relative entropy density and random conditional entropy, which holds for arbitrary discrete information sources. In the proof an analytic technique to study the strong limit properties of discrete information sources is proposed. The crucial part is the application of Lebesgue's theorem on differentiability of monotone functions together with the convergence theorem of infinite products.

Let $\{X_n, n \geq 1\}$ be a sequence of successive letters produced by an arbitrary information source with the alphabet $S = \{1, 2, \ldots, N\}$ and with the joint distribution

$$P(X_1 = x_1, \ldots, X_n = x_n) = p(x_1, \ldots, x_n) > 0. \tag{1}$$

Received July 31, 1997; revised December 4, 1998.
Communicated by Y.-J. Lee.
1991 *Mathematics Subject Classification*: 94A17, 60F15.
Key words and phrases: Random conditional entropy, relative entropy density, strong limit theorem, Shannon-McMillan-Breiman theorem.

Let

$$f_n(\omega) = -(1/n)\ln p(X_1, \dots, X_n), \tag{2}$$

where ω is a sample point, X_i stands for $X_i(\omega)$ for brief, and the quantity $f_n(\omega)$ is called the *relative entropy density* of $\{X_i, 1 \le i \le n\}$ (see [2]). Let

$$p_n(x_n|x_1, \dots, x_{n-1}) = P(X_n = x_1, \dots, X_{n-1} = x_{n-1}), \quad n \ge 2. \tag{3}$$

Then

$$p(x_1, \dots, x_n) = p(x_1)\prod_{k=2}^{n} p_k(x_k|x_1, \dots, x_{k-1}); \tag{4}$$

$$f_n(\omega) = -(1/n)\left[\ln p(X_1) + \sum_{k=2}^{n} \ln p_k(X_k|X_1, \dots, X_{k-1})\right]. \tag{5}$$

2. A realization of Arbitrary Information Source

Throughout this paper we shall deal with the underlying probability space $([0,1), \boldsymbol{F}, P)$, where F is the class of Borel sets in the interval $[0,1)$, and P is the Lebesgue measure. We first give a realization of an arbitrary information source with distribution (1) in the above probability space.

Split the interval $[0,1)$ into N right-semiopen intervals:

$$I_1 = [0, p(1)), \quad I_2 = [p(1), p(1) + p(2)), \dots, \quad I_N = [1 - p(N), 1).$$

These intervals will be called intervals of the first order. Proceeding inductively by splitting each nth order interval $I_{X_1 \cdots X_n}$ into N right-semiopen intervals $I_{x_1 \cdots x_n\ 1}$, $I_{x_1 \cdots x_n\ 2}, \dots$, $I_{x_1 \cdots x_n N}$ according to the ratio $p\,(x_1, \dots, x_n,\ 1) : p(x_1, \dots, x_n,\ 2) : \dots : p(x_1, \dots, x_n, N)$, the intervals of the $(n+1)$st order are created. It is easy to see that for $n \ge 1$,

$$P(I_{x_1 \cdots x_n}) = p(x_1, \dots, x_n). \tag{6}$$

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

$$X_n(\omega) = x_n, \quad \text{if } \omega \in I_{x_1 \cdots x_n}. \tag{7}$$

By (6) and (7), $\{X_n, n \ge 1\}$ has the distribution (1).

We are going to prove the following limit theorem by using the above realization.

3. Main Result

Definition 1. For $k \geq 2$, let

$$h_k(x_1, \ldots, x_{k-1}) = -\sum_{x_k=1}^{N} p_k(x_k|x_1, \ldots, x_{k-1}) \ln p_k(x_k|x_1, \ldots, x_{k-1}); \tag{8}$$

$$H_k(\omega) = h_k(X_1, \ldots, X_{k-1}). \tag{9}$$

$H_k(\omega)$ is called the *random conditional entropy.*

Theorem 1. *Let $\{X_n, n \geq 1\}$ be a sequence of succesive letters produced by an arbitrary information source with the alphabet S and the joint distribution (1), $\{a_n, n \geq 1\}$ an increasing sequence of positive real numbers such that*

$$\sum_{n=1}^{\infty} (1/a_n)^2 < \infty, \tag{10}$$

and $f_n(\omega)$ and $H_k(\omega)$ defined, respectively, by (2) and (9). Then

$$\sum_{k=2}^{\infty} (1/a_k)[\ln p_k(X_1, \ldots, X_{k-1}) + H_k(\omega)] \text{ converges a.e.}; \tag{11}$$

$$\lim_{n\to\infty} (1/a_n) \sum_{k=2}^{n} [\ln p_k(X_k|X_1, \ldots, X_{k-1}) + H_k(\omega)] = 0 \text{ a.e.} \tag{12}$$

Proof. For $k \geq 2, \lambda = 1$ or -1, let

$$\begin{aligned} Q_k(\lambda; x_1, \ldots, x_{k-1}) = \sum_{x_k=1}^{N} & p_k(x_k|x_1, \ldots, x_{k-1}) \\ & \exp\{\lambda[\ln p_k(x_k|x_1, \ldots, x_{k-1}) \\ & + h_k(x_k|x_1, \ldots, x_{k-1})]/a_k\}. \end{aligned} \tag{13}$$

Let the collection of intervals of all orders be denoted by A. Define a set function μ on $\mathbf{A}$ as follows. Let

$$\mu(I_{x_1}) = p(x_1), \tag{14}$$

and for $n \geq 2$, let

$$\mu(I_{x_1\cdots x_n}) = \frac{p(x_1,\ldots,x_n)\exp\Big\{\lambda\sum_{k=2}^{n}[\ln p_k(x_k|x_1,\ldots,x_{k-1})+h_k(x_1,\ldots,x_{k-1})]/a_k\Big\}}{\prod_{k=2}^{n}Q_k(\lambda;x_1,\ldots,x_{k-1})}. \tag{15}$$

By (4) and (13)-(15), it is easy to see that μ is an additive function on $\mathbf{A}$. Therefore there exists an increasing function f_λ defined on $[0,1]$ such that, for any $I_{x_1\cdots x_n}$,

$$\mu(I_{x_1\cdots x_n}) = f_\lambda(I^+_{x_1\cdots x_n} - f_\lambda(I^-_{x_1\cdots x_n}), \tag{16}$$

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

$$\begin{aligned} t_n(\lambda,\omega) &= \frac{f_\lambda(I^+_{x_1\cdots x_n}) - f_\lambda(I^-_{x_1\cdots x_n})}{I^+_{x_1\cdots x_n} - I^-_{x_1\cdots x_n}} \\ &= \frac{\mu(I_{x_1\cdots x_n})}{P(I_{x_1\cdots x_n})}, \quad \omega\in I_{x_1\cdots x_n}. \end{aligned} \tag{17}$$

Ler $A(\lambda)$ be the set of points of differentiability of f_λ. Then $P(A(\lambda)) = 1$ by the theorem on the existence of derivative of monotone function (cf. [3], p. 424). Let $\omega\in A(\lambda)$, and $\omega\in I_{x_1\cdots x_n}$ $(n=1,2,\ldots)$. In virtue of a property of derivative (cf. [3], p. 423), we have by (17),

$$\lim_{n\to\infty} t_n(\lambda,\omega) = \text{a finite number},\ \omega\in A(\lambda). \tag{18}$$

By (17), (14), (6) and (7),

$$\begin{aligned} t_n(\lambda,\omega) &= \frac{\mu(I_{X_1\cdots X_n})}{P(I_{X_1\cdots X_n})} \\ &= \frac{\exp\Big\{\lambda\sum_{k=2}^{n}[\ln p_k(X_k|X_1,\ldots,X_{k-1})+h_k(X_1,\ldots,X_{k-1})]/a_k\Big\}}{\prod_{k=2}^{n}Q_k(\lambda;X_1,\ldots,X_{k-1})}. \end{aligned} \tag{19}$$

For the sake of simplicity, denote $p_k(x_k|x_1,\ldots,x_{k-1})$ and $h_k(x_1,\ldots,x_{k-1})$ by p_k and h_k, respectively. We have by (8),

$$\sum_{x_k=1}^{N} p_k(\ln p_k + h_k) = 0. \tag{20}$$

By (13), (20), the inequality $0 \leq e^x - 1 - x \leq x^2 e^{|x|}$ and the entropy inequality $h_k \leq \ln N$, we have

$$
\begin{aligned}
&0 \leq Q_k(\lambda; x_1, \ldots, x_{k-1}) - 1 \\
&= \sum_{x_k=1}^{N} p_k \{\exp[\lambda(\ln p_k + h_k)/a_k] - 1 - \lambda(\ln p_k + h_k)/a_k\} \\
&\leq (1/a_k)^2 \sum_{x_k=1}^{N} p_k (\ln p_k + h_k)^2 \exp[(-\ln p_k + \ln N)/a_k].
\end{aligned} \tag{21}
$$

Since $a_k \to \infty$ (as $k \to \infty$), there exists a positive integer m such that $a_k \geq 2$ as $k \geq m$. Hence as $k \geq m$, we have by (21) and the entropy inequality,

$$
\begin{aligned}
&0 \leq Q_k(\lambda; x_1, \ldots, x_{k-1}) - 1 \\
&\leq N(1/a_k)^2 \sum_{x_k=1}^{N} p_k^{1/2} (\ln p_k + h_k)^2 \\
&< N(1/a_k)^2 \sum_{x_k=1}^{N} [p_k^{1/2} (\ln p_k)^2 - 2(\ln N) p_k^{1/2} \ln p_k + (\ln N)^2].
\end{aligned} \tag{22}
$$

Let

$$
\begin{aligned}
M_1 &= \max\{x^{1/2} (\ln x)^2, \quad 0 < x \leq 1\}; \\
M_2 &= \max\{-x^{1/2} \ln x, \quad 0 < x \leq 1\}.
\end{aligned}
$$

From (22) and (10) it follows that

$$
\begin{aligned}
&\sum_{k=m}^{\infty} [Q_k(\lambda; X_1, \ldots, X_{k-1}) - 1] \\
&< \sum_{k=m}^{\infty} (N/a_k)^2 \{M_1 + 2M_2 \ln N + (\ln N)^2\} < \infty.
\end{aligned} \tag{23}
$$

By the convergence theorem of infinite product, (23) implies that

$$
\prod_{k=2}^{\infty} [Q_k(\lambda; X_1, \ldots, X_{k-1}) \text{ converges.} \tag{24}
$$

(18), (19) and (24) imply that

$$
\begin{aligned}
&\lim_{n \to \infty} \exp\left\{\lambda \sum_{k=2}^{n} [\ln p_k(X_k | X_1, \ldots, X_{k-1}) + h_k(X_1, \ldots, X_{k-1})]/a_k\right\} \\
&= \text{a finite number a.e..}
\end{aligned} \tag{25}
$$

Letting $\lambda = 1$ and $\lambda = -1$ respectively, we have

$$\lim_{n\to\infty} \exp\left\{\sum_{k=2}^{n}[\ln p_k(X_k|X_1,\ldots,X_{k-1}) + h_k(X_1,\ldots,X_{k-1})]/a_k\right\} \tag{26}$$
$$= \text{a finite number a.e.}$$

$$\lim_{n\to\infty} \exp\left\{-\sum_{k=2}^{n}[\ln p_k(X_k|X_1,\ldots,X_{k-1}) + h_k(X_1,\ldots,X_{k-1})]/a_k\right\} \tag{27}$$
$$= \text{a finite number a.e..}$$

(26) and (27) imply that

$$\sum_{k=2}^{n}\{[\ln p_k(X_k|X_1,\ldots,X_{k-1}) + h_k(X_1,\ldots,X_{k-1})]/a_k\} \text{ converges a.e.} \tag{28}$$

i. e., (11) is true. By Kronecker's lemma, (12) follows from (28).

This completes the proof of the theorem. ■

4. Some Corollaries

Corollary 1. *Let $f_n(\omega)$ be defined by* (5)*. Then under the hypotheses of the theorem we have*

$$\lim_{n\to\infty}\left[f_n(\omega) - (1/n)\sum_{k=1}^{n} H_k(\omega)\right] = 0 \quad a.e.. \tag{29}$$

Proof. Let $a_n = 1/n$. Then (29) follows from (12) and (5) immediately. ■

Corollary 2. *Let $p > 1/2$ be a constant. Then under the hypotheses of the theorem we have*

$$\lim_{n\to\infty} n^{-1/2}(\ln n)^{-p}\sum_{k=2}^{n}[\ln p_k(X_k|X_1,\ldots,X_{k-1}) + H_k(\omega)] = 0 \quad a.e.. \tag{30}$$

Proof. Since $\sum_{n=2}^{\infty} n^{-1}(\ln n)^{-2p} < \infty$, the corollary is obtained immediately from the theorem. ■

Corollary 3. *Let $\{X_n, n \geq 1\}$ be a sequence of successive letters produced by a nonhomogeneous Markov information source with the initial distribution*

$$p(1), P(2), \cdots, p(N), \quad p(i) > 0, \quad i \in S \tag{31}$$

and the transition matrix

$$P_n = (p_n(i,j)), \quad p_n(i,j) > 0, \quad i,j \in S, \quad n \geq 1, \tag{32}$$

where $p_n(i,j) = P(X_n = j | X_{n-1} = i)$, $\{a_n, n \geq 1\}$ *is an increasing sequence of positive numbers such that* (10) *holds, and* $H(p_1, \ldots, p_n)$ *is the entropy of the distribution* $(p_1, \ldots, p_n)$. *Then*

$$\begin{aligned}\lim_{n\to\infty} (1/a_n) \sum_{k=1}^{n} \{\ln p_k(X_{k-1}, X_k) + H[p_k(X_{k-1}, \ldots, p_k(X_{k-1}, N)]\} \\ = 0 \quad a.e..\end{aligned} \tag{33}$$

Proof. By Markov property, (33) follows from (12) immediately. ■

Corollary 4. *Let* $p > 1/2$ *be a constant. Then under the hypotheses of Corollary* 3 *we have*

$$\lim_{n\to\infty} \left\{ f_n(\omega) - (1/n) \sum_{k=1}^{n} H[p_k(X_{k-1}, 1), \ldots, p_k(X_{k-1}, N)] \right\} = 0 \quad a.e., \tag{34}$$

where

$$f_n(\omega) = -(1/n) \left[\ln p(X_0) + \sum_{k=1}^{n} \ln p_k(X_{k-1}, X_k) \right] \tag{35}$$

is the relative entropy density of the Markov information source.

Proof. Let $a_n = n$. Then (34) follows from (33) and (5) directly. ■

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. P. Billingsley, *Probability and Measure*, Wiley, New York, 1986.
4. L. Breiman, The individual ergodic theorem of information theory, *Ann. Math. Statist.* **28** (1957), 809-811.

5. K. L. Chung, The ergodic theorem of information theory, *Ann. Math. Statist.* **32** (1961), 612-614.

6. J. C. Kieffer, A counterexample to Perez's generalization of the Shannon-McMillan theorem, *Ann. Probab.* **1** (1973), 362-364.

7. J. C. Kieffer, A counterexample to Perez's generalization of the Shannon-McMillan theorem, *Pacific J. Math.* **51** (1973), 203-204.

8. B. McMillan, The basic theorems of information theory, *Ann. Math. Statist.* **24** (1953), 196-216.

9. C. Shannon, A mathematial theory of communication, *Bell System Tech. J.* **27** (1948), 379-423, 623-656.

Department of Mathematics, Hebei University of Technology
Tianjin 300130, China