Some limits to nonparametric estimation for ergodic processes

Hayato Takahashi
The Institute of Statistical Mathematics,
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan.
e-mail: hayato.takahashi@ieee.org.

Abstract—A new negative result for nonparametric distribution estimation of binary ergodic processes is shown. The problem of estimation of distribution with any degree of accuracy is studied. Then it is shown that for any countable class of estimators there is a zero-entropy binary ergodic process that is inconsistent with the class of estimators. Our result is different from other negative results for universal forecasting scheme of ergodic processes. We also introduce a related result by B. Weiss.

Index Terms—ergodic process, cutting and stacking, nonparametric estimation, computable function.

I. INTRODUCTION

Let $X_1, X_2, \ldots$ be a binary-valued ergodic process and $P$ be its distribution. In this paper we study nonparametric estimation of binary-valued ergodic processes with any degree of accuracy. Let $S$ and $\Omega$ be the set of finite binary strings and the set of infinite binary sequences, respectively. Let $\Delta(x) := \{x_\omega | \omega \in \Omega\}$, where $x_\omega$ is the concatenation of $x \in S$ and $\omega$, and write $P(x) = P(\Delta(x))$. For $x \in S$, $|x|$ is the length of $x$. Let $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ be the set of natural numbers, the set of integers, and the set of rational numbers, respectively. From ergodic theorem, there is a function $r$ such that for $x \in S, n, k \in \mathbb{N}$,

$$P(\{\Delta(y) | |P(x) - 1/|y| \sum_{i=1}^{|y|} I_{y_i \neq x_i}] \geq 1/k, |y| = n\}) < r(n, k, x),$$

$$\forall x, k \lim_{n} r(n, k, x) = 0,$$

where $I$ is the indicator function and $y_i = y_i \cdot y_{i+1} \cdot y_j$ for $y = y_1 \cdot y_n, i \leq j \leq n$. $r$ is called convergence rate. If $r$ is given, we know how much sample size is necessary to estimate the distribution with prescribed accuracy. However it is known that there is no universal convergence rate for ergodic theorem. If $r$ is not known, ergodic theorem does not help to estimate the distribution with prescribed accuracy. Here a natural question arise: for any binary-valued ergodic process, is it always possible to estimate the distribution with any degree of accuracy with positive probability? We show that this problem has a negative answer, i.e., for any countable class of estimators there is a zero-entropy binary ergodic process that is not estimated from this class of estimators with positive probability. In particular, since the set of computable functions is countable, we see that there is a zero-entropy binary ergodic process that is inconsistent with computable estimators. Our result is not derived from other negative results for universal forecasting scheme of ergodic processes, see Remark 2.

Let $x \subseteq y$ if $x$ is a prefix of $y$. $f$ is called estimator if $f(x, k, y) \in \mathbb{Q}$ is defined for $(x, k, y) \in S \times \mathbb{N} \times S$.

$$\forall z \exists y f(x, k, z) = f(x, k, y).$$

For $\omega \in \Omega$, let $f(x, k, \omega) := f(x, k, y)$ if $f(x, k, y)$ is defined and $y \subseteq \omega$. We say that $f$ estimates $P$ if

$$P(\omega | \forall x, k f(x, k, \omega)) \text{ is defined and } |P(x) - f(x, k, \omega)| < \frac{1}{k} > 0.$$}

Here $\omega$ is a sample sequence and the minimum length of $y \subseteq \omega$ for which $f(x, k, y)$ is defined is a stopping time.

In this paper, we construct an ergodic process that is not estimated from any given countable set of estimators:

Theorem 1.

$$\forall F : \text{countable set of estimators}$$

$$\exists P \text{ ergodic and zero entropy } \forall f \in F$$

$$P(\omega | \forall x, k f(x, k, \omega)) \text{ is defined and } |P(x) - f(x, k, \omega)| < \frac{1}{k} = 0.$$}

We say that $P$ is effectively estimated if there is a partial computable $f$ that satisfies (2) and (3). Since the set of partial computable estimators is countable, we have

Corollary 1. There is a zero entropy ergodic process that is not effectively estimated.

If $r$ in (1) is computable then it is easy to see that $P$ is effectively estimated. For example, i.i.d. processes of finite alphabet are effectively estimated, see Leeuw et al. 3.

As stated above, a difficulty of effective estimation of ergodic processes comes from that there is no universal convergence rate for ergodic theorem. In Shields pp.171 7, it is shown that for any given decreasing function $r$, there is an ergodic process that satisfies

$$\exists N \forall n \geq N P(|P(1) - \sum_{i=1}^{n} I_{X_i = 1/n}| \geq 1/2) > r(n).$$

In particular if $r$ is chosen such that $r$ decreases to 0 asymptotically slower than any computable function then $r$ is not
computable. In V’yugin [9], a binary-valued computable stationary process with incomputable convergence rate is shown.

It is possible that an ergodic process is effectively estimated even if the convergence rate is not computable.

**Theorem 2.** For any decreasing \( r \), there is a zero entropy ergodic process that is effectively estimated and satisfies (4).

For proofs of Theorem 1 and 2 see [8].

**Remark 1.** (i) \( P \) is computable \( \Rightarrow \) (ii) convergence rate \( r \) in (1) is upper semi-computable (effectively approximated from above) \( \Rightarrow \) (iii) \( P \) is effectively estimated. None of the converse is true.

**Remark 2.** In Cover [2], two problems about prediction of ergodic processes are posed. Problem 1: Is there a universal scheme \( f \) such that \( \lim_{n \to \infty} |f(X_{n}^{i-1}) - P(X_n|X_{0}^{n-1})| \to 0 \), a.s. for all binary-valued ergodic \( P \)?

Problem 2: Is there a universal scheme \( f \) such that \( \lim_{n \to \infty} |f(X_{n}^{i-1}) - P(X_0|X_{0}^{n-1})| \to 0 \), a.s. for all binary-valued ergodic \( P \)?

Problem 2 was affirmatively solved by Ornstein [5], [10].

It is possible that an ergodic process is effectively estimated but is much stronger statement than the fact that \( P \) is not estimated in the sense of Theorem 1.

The difference between this result and Theorem 1 is that (6) requires the universality of \( f \) but is much stronger statement than the fact that \( P \) is not estimated in the sense of Theorem 1.

**Acknowledgement**

The author thanks Prof. Teturo Kamae (Matsuyama Univ.) and Prof. Benjamin Weiss (Hebrew Univ.) for helpful discussions and valuable comments.

**References**

[1] D. H. Bailey. Sequential schemes for classifying and predicting ergodic processes. PhD thesis, Stanford Univ., 1976.

[2] T. M. Cover. Open problems in information theory. In 1975 IEEE Joint Workshop on Information Theory, pages 35–36, 1975.

[3] K. de Leeuw, E. F. Moore, C. E. Shannon, and N. Shapiro. Computability by probabilistic machines. In C. E. Shannon and J. McCarthy, editors, Automata Studies, pages 183–212. Princeton Univ. Press, 1956.

[4] L. Györﬁ, G. Morvai, and S. J. Yakowitz. Limits to consistent online forecasting for ergodic time series. IEEE Trans. Inform. Theory, 44(2):886–892, 1988.

[5] D. S. Ornstein. Guessing the next output of a stationary process. Israel J. Math., 30(3):292–296, 1978.

[6] B. Ya. Ryabko. Prediction of random sequences and universal coding. Probl. Inf. Transm., 24:87–96, 1988.

[7] P. Shields. The ergodic theory of discrete sample paths. Amer. Math. Soc., 1996.

[8] H. Takahashi. Computational limits to nonparametric estimation for ergodic processes. http://arXiv.org/abs/1002.1559, submitted to IEEE Trans IT.

[9] V. V. V’yugin. Ergodic theorems for individual random sequences. Theor. Comp. Sci., 207:343–361, 1998.

[10] B. Weiss. Single Orbit Dynamics. Amer. Math. Soc., 2000.