Limits to nonparametric estimation for ergodic processes

高橋　勇人　　データ科学研究系　特任研究員

In [3], problems about universal prediction of ergodic processes are posed. There are positive and negative results for these problems, see [2, 5, 6, 4, 9]. Similarly in [1], it is shown that that there is no universal consistent density estimation procedure for all ergodic processes, i.e., for any density estimation procedure there is an ergodic P that is inconsistent with the procedure. In this paper we study nonparametric estimation of ergodic processes. In particular we study estimation of distribution of binary valued ergodic processes with any given accuracy. Unlike with negative results for universal forecasting and density estimation (i.e., they constructed an ergodic process that is inconsistent with a given estimator), we show that for any given countable class of estimators, there is an ergodic process that is inconsistent with that class of estimators.

Let S, $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ be the set of finite binary strings, the set of natural numbers, the set of integers, and the set of rational numbers, respectively. Let P be an ergodic process on $\Omega := \{0,1\}^\infty$. Let $\Delta(x) := \{x\omega | \omega \in \Omega\}$, where $x\omega$ is the concatenation of $x \in S$ and ω, and write $P(x) = P(\Delta(x))$. For $x \in S$, $|x|$ is the length of x. Let $x \sqsubseteq y$ if x is a prefix of y. f is called estimator if $\exists D_f \subseteq S \times \mathbb{N} \times S,\ f : D_f \to \mathbb{Q}$ and

$$f(x,k,y) \text{ is defined, i.e., } (x,k,y) \in D_f \Rightarrow \forall z \sqsupseteq y, f(x,k,z) = f(x,k,y). \tag{1}$$

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

$$P(\omega | \forall x,k,\ f(x,k,\omega) \text{ is defined and } |P(x) - f(x,k,\omega)| < \frac{1}{k}) > 0. \tag{2}$$

Here ω is a sample sequence and the minimum length of $y \sqsubset \omega$ 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

$\forall F$: *countable set of estimators* $\exists P$ *ergodic and zero entropy* $\forall f \in F$

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

Note that the countability condition is necessary in the above theorem. For example if F is the whole class of estimators (which is not countable) and if f is a rational approximation of P then f $(\in F)$ estimates P.

We say that P is *effectively estimated* if there is a partial computable f that satisfies (1) and (2). Let F be the partial computable estimators in Theorem 1, then we have

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

Let r be a function such that for $x \in S$, $n, k \in \mathbb{N}$,

$$P(\cup\{\Delta(y) | \ |P(x) - \sum_{i=1}^{|y|-|x|+1} I_{y_i^{i+|x|-1}=x}/|y|| > 1/k, |y| = n\}) < r(n,k,x),$$

where I is the indicator function and $y_i^j = y_i y_{i+1} \cdots y_j$ for $y = y_1 \cdots y_n, i \le j \le n$. From ergodic theorem, we have $\forall x,k,\ \lim_n r(n,k,x) = 0$. If r 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.

The difficulty of effective estimation of ergodic processes comes from that there is no universal convergence rate for ergodic theorem. In pp.171 [7], it is shown that for any given decreasing function r, there is an ergodic process that satisfies

$$\exists N \forall n \ge N,\ P(|P(1) - \sum_{i=1}^{n} I_{X_i=1}/n| \ge 1/2) > r(n). \tag{3}$$

The above inequality shows that there is no universal convergence rate. In particular if r is chosen such that r decreases to 0 asymptotically slower than any computable function then r is not computable. In [8], an incomputable convergence rate is shown in another way.

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 (3).*

Remark 1 *P is called computable if there is a computable $p : S \times \mathbb{N} \to \mathbb{Q}$ such that $\forall x,k,\ |P(x) - p(x,k)| < \frac{1}{k}$. From definition, if P is computable, it is effectively estimated. Moreover we have i) P is computable $\Rightarrow$ ii) convergence rate of P is computable $\Rightarrow$ iii) P is effectively estimated. None of the converse is true.*

References

[1] T. M. Adams and A. B. Nobel. On density estimation from ergodic processes. *Ann. Probab.*, 26(2):794–804, 1998.

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

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

[4] L. Györfi, G. Morvai, and S. J. Yakowitz. Limits to consistent on-line forcasting 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] V. V. V'yugin. Ergodic theorems for individual random sequences. *Theor. Comp. Sci.*, 207:343–361, 1998.

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