On Density Estimation from Ergodic Processes

Terrence M. Adams and Andrew B. Nobel *

June 14, 1997

Abstract

We consider the problem of L_p-consistent density estimation from the initial segments of strongly dependent processes. It is shown that no procedure can consistently estimate the one-dimensional marginal density of every stationary ergodic process for which such a density exists. A similar result is established for the problem of estimating the support of the marginal distribution of an ergodic process.

Appears in: Annals of Probability, Vol.26, pp.794-804

AMS 1991 subject classifications: 60G10, 62G07, 60G17.

Key words and phrases: Density estimation, ergodic processes, cutting and stacking, counterexample.

*Work supported in part by NSF Grant DMS-9501926.

1 Introduction

Let μ be a probability measure on the Borel subsets $\mathcal{B}$ of the half-open unit interval $[0, 1)$. A measurable transformation $T : [0, 1) \to [0, 1)$ is μ-preserving if $\mu(T^{-1}B) = \mu(B)$ for each $B \in \mathcal{B}$, and is ergodic if for each B such that $T^{-1}B = B$ either $\mu(B) = 0$ or $\mu(B) = 1$. A sequence $\mathbf{X} = \{X_1, X_2, \ldots\}$ of random variables defined on $([0, 1), \mathcal{B}, \mu)$ is said to be stationary and ergodic if there exists an ergodic μ-preserving transformation T and a Borel measurable function $g : [0, 1) \to \mathrm{I\!R}$ such that

$$X_i(\omega) = g(T^{i-1}\omega) \quad i = 1, 2, \ldots$$

for μ-a.e. $\omega \in [0, 1)$. If the distribution $\nu = \mu \circ g^{-1}$ of each random variable X_i is absolutely continuous with respect to Lebesgue measure λ, then X_i is distributed according to the the probability density $f = d\nu/d\lambda$, written $X_i \sim f$. Estimation of f from finitely many observations of the process $\mathbf{X}$ is an important and well studied problem in applied and theoretical statistics. The density estimation problem and its potential solutions can be formalized as follows.

Problem: Given an ergodic process $\mathbf{X} = X_1, X_2, \ldots \in \mathrm{I\!R}$ with $X_i \sim f$, select integrable functions $\hat{f}_1, \hat{f}_2, \ldots$ such that (i) $\hat{f}_n$ depends only on $X_1, \ldots, X_n$, and (ii) $\int |\hat{f}_n - f| dx \to 0$ in probability as $n \to \infty$.

Definition: Let $\mathrm{I\!R}^* = \cup_{n=1}^{\infty} \mathrm{I\!R}^n$ contain all finite sequences of real numbers. A density estimation procedure is a mapping $\Phi : \mathrm{I\!R}^* \to L_1(\mathrm{I\!R})$ that associates every finite sequence $x_1, \ldots, x_n \in \mathrm{I\!R}$ with an integrable function $\Phi(\cdot; x_1, \ldots, x_n)$.

Remark: In what follows we restrict our attention to procedures that are measurable, in the sense that $(x_1, \ldots, x_n) \mapsto \int |\Phi(u; x_1, \ldots, x_n) - h(u)| du$ is a Borel measurable map from $\mathrm{I\!R}^n$ to $\mathrm{I\!R}$ for each $n \geq 1$, and each $h \in L_1(\mathrm{I\!R})$. Aside from measurability, no regularity conditions are imposed on the behavior of Φ as a function of its input. The estimates $\Phi(\cdot; x_1, \ldots, x_n)$ may take negative values, and need not integrate to one.

Definition: A density estimation procedure Φ is weakly L_1-consistent (or simply consistent) for an ergodic process $\mathbf{X}$ with $X_i \sim f$ if as n tends to infinity

$$\int |\Phi(u; X_1, \ldots, X_n) - f(u)| du \to 0 \text{ in probability.}$$

The procedure Φ is consistent for a family $\mathcal{P}$ of ergodic processes if it is consistent for each $\mathbf{X} \in \mathcal{P}$, and is said to be universal if it is consistent for every ergodic process $\mathbf{X}$ such that the distribution of X_1 is absolutely continuous. Strong consistency is defined as above, with almost sure convergence replacing convergence in probability.

Common density estimation methods include histogram, kernel, nearest neighbor, orthogonal series, and likelihood based procedures. For a general account of these and other methods, see Devroye and Györfi (1985) and Silverman (1986). In establishing consistency, rates of convergence, and central limit theorems for a specific procedure, most analyses assume that $X_1, X_2, \ldots$ are independent and identically distributed (i.i.d.). It is known for instance that suitable versions of the methods above are consistent for all, or almost all, i.i.d. processes.

Numerous results have also been obtained for dependent random variables under various mixing conditions. Roussas (1967) and Rosenblatt (1970) studied the consistency and asymptotic normality of kernel density estimates from Markov processes. Similar results, under substantially weaker conditions, were obtained by Yakowitz (1989). Ahmad (1979) established the strong consistency of orthogonal series estimates under α mixing conditions. Györfi (1981) showed that there is a simple kernel-based procedure Φ that is strongly L_2-consistent for every stationary ergodic process $\mathbf{X} = \{X_i\}_{i=-\infty}^{\infty}$ such that (i) the conditional distribution of X_1 given $\{X_i : i \leq 0\}$ is absolutely continuous with probability one, and (ii) the corresponding conditional density h satisfies $E \int |h(u)|^2 du < \infty$.

Castellana and Leadbetter (1986) studied pointwise consistency and central limit theorems for kernel density estimates using a dependence index based on the difference between joint bivariate densities and the product of their marginals. Györfi and Masry (1990) established the strong L_1 consistency of multivariate recursive kernel density estimates for both ρ and α mixing processes under weak conditions on the mixing coefficients. Hall and Hart (1990) established convergence rates for kernel density estimates from infinite order moving average processes. The cited papers contain many additional references to research in this area.

In spite of these positive results, difficulties can arise in estimating densities from strongly dependent processes. These difficulties are most clearly seen in the case of histogram estimates based on regular partitions, in which the error that results from estimating the probability of a cell by its relative frequency is magnified by the inverse of the cell width. Shrinking cell widths ensuring consistent density estimates for any i.i.d. process can be fixed in advance of the data. To obtain consistent estimates for families of strongly dependent processes, the cell widths must shrink in a *data-dependent* fashion. In a result attributed to Shields, it was shown by Györfi, Härdle, Sarda and Vieu (1989) that there exists cell widths, suitable for every i.i.d. process, such that the associated histograms fail to produce consistent density estimates from a suitably constructed ergodic process. Györfi and Lugosi (1992) exhibited an ergodic process $\mathbf{X}$ for which a standard kernel density estimate with bandwidths $h_n \to 0$ and $nh_n \to \infty$ fails to be consistent.

Taken together, these positive and negative results lead to the following question, which was asked by Györfi (1981) and Györfi and Lugosi (1992):

> Is there some (measurable) procedure Φ that is weakly L_p consistent for every stationary ergodic $\mathbf{X}$ having a one-dimensional marginal density?

Our principal result, given in Theorem 1 below, shows that the answer to this question is 'No'. In preliminary work, based on different methods than those developed here, Yakowitz and Heyde (1997) have announced a similar negative result when $p = 2$.

In light of Theorem 1, one may ask about the existence of solutions to the more general problem of estimating the marginal distribution of an ergodic process. The answers depend on the type of estimation. The classical Glivenko-Cantelli theorem, which relies only on the convergence of relative frequencies to probabilities, extends easily to the ergodic case. Devroye and Gyorfi (1990) have shown that there is no universal procedure for estimating the marginal distribution of an i.i.d. process in the total variation norm. On the other hand, for the family of i.i.d. processes whose marginal distributions are absolutely continuous, there is an L_1-consistent density estimation scheme, which provides distribution estimates consistent in total variation. The discussion following Theorem 1 shows that, for the family of ergodic processes with absolutely continuous marginal distributions, there is no distribution estimation scheme consistent in total variation. In fact, it is shown in Corollary 4 that, for such processes, one cannot even estimate the Lebesgue measure of the support of their marginal distribution.

The next section contains two positive results intended to highlight the distinction between collections of densities and families of dependent processes. Section 3 is devoted to the statement and proof of Theorem 1. Section 4 presents several corollaries, including two results on estimating the support of the marginal distribution of an ergodic process.

2 Two Positive Results

For i.i.d. processes there is a one to one correspondence between collections of densities and collections of processes. No such relation exists in the dependent case, where the investigation of consistency involves constraints on the candidate densities and constraints on the structure of the underlying processes. For suitably constrained families of candidate densities, no assumptions concerning the dependence of the process are necessary. Let $\alpha : \{1, 2, \ldots\} \to [0, \infty)$ be non-decreasing, and let $\mathcal{F}(\alpha)$ contain all those densities $f : \mathbb{R} \to \mathbb{R}$ such that, for each $i \geq 1$, the variation of f on $[-i, i]$ is at most $\alpha(i)$. Nobel, Morvai and Kulkarni (1996) established the following result.

Theorem A *Let $\alpha(\cdot)$ be known. Then there is a strongly consistent procedure Φ for the family $\mathcal{P}$ of all ergodic processes $\mathbf{X}$ such that $X \sim f$ with $f \in \mathcal{F}(\alpha)$.*

This result suggests that the existence of a consistent estimation procedure for a family of strongly dependent processes might require a compactness type condition on the set of candidate densities. The next result shows that this is not the case. In particular, one can almost surely distinguish between the members of any countable family of ergodic processes.

Let $\mathcal{P} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots\}$ be a countable family of stationary ergodic processes indexed by $\{1, 2, \ldots\}$, each defined on the same probability space $(\Omega, \mathcal{F}, \mu)$. Assume that the elements of $\mathcal{P}$ are distinct in the sense that no two processes have the same k-dimensional marginal distributions for each $k \geq 1$. The following result, which shows that one can distinguish between the processes in $\mathcal{P}$, is due to Barron (1985), who gave two proofs using martingale theory. The proof below relies on the ergodic theorem.

Lemma 1 *There exists a procedure $\Psi : \mathbb{R}^* \to \{1, 2, \ldots\}$ such that for each process $\mathbf{X}^{(l)} \in \mathcal{P}$ the cardinality of $\{n : \Psi(X_1^{(l)}, \ldots, X_n^{(l)}) \neq l\}$ is finite with μ-probability one.*

Note: It follows from Lemma 1 that there is a consistent density estimation procedure Φ for a countable family of ergodic processes whose marginal densities form a dense subset of the collection of all densities on $\mathbb{R}$.

Proof: Fix $i \neq j$. By assumption there exists $k \geq 1$ and a k-dimensional Borel set $A_{i,j}$ such that

$$\epsilon_{i,j} = |\mu\{(X_1^{(i)}, \ldots, X_k^{(i)}) \in A_{i,j}\} \,-\, \mu\{(X_1^{(j)}, \ldots, X_k^{(j)}) \in A_{i,j}\}| \,>\, 0\,.$$

For each $n \geq k$ let $D_n(i,j)$ contain all those vectors $(x_1, \ldots, x_n) \in \mathbb{R}^n$ such that

$$\left| \frac{1}{n-k+1} \sum_{l=1}^{n-k+1} I\{(x_l, \ldots, x_{l+k-1}) \in A_{i,j}\} \,-\, \mu\{(X_1^{(i)}, \ldots, X_k^{(i)}) \in A_{i,j}\} \right| \,<\, \frac{\epsilon_{i,j}}{2}$$

Fix $m \geq 2$ for the moment. By the ergodic theorem there exists an integer n_m such that for each $i \neq j$ with $1 \leq i, j \leq m$,

$$\mu\{(X_1^{(i)}, \ldots, X_{n_m}^{(i)}) \in D_{n_m}^c(i,j)\} \,\leq\, \frac{1}{m^3} \tag{1}$$

and

$$\mu\{(X_1^{(j)},\ldots,X_{n_m}^{(j)}) \in D_{n_m}(i,j)\} \;\leq\; \frac{1}{m^3}\,. \tag{2}$$

Restricting i,j to $1,\ldots,m$, let $B_m(i) = \bigcap_{j\neq i} D_{n_m}(i,j)$ and define sets

$$F_m(i) = \bigcap_{j\neq i} B_m(j)^c \cap B_m(i), \quad i = 1,\ldots,m-1; \quad F_m(m) = \left(\bigcup_{i=1}^{m-1} F_m(i)\right)^c,$$

so that $F_m(1),\ldots,F_m(m)$ form a partition of $\mathbb{R}^{n_m}$. Let $n_1 = 1$, $F_1(1) = \mathbb{R}$, and assume without loss of generality that $n_m > n_{m-1}$.

Given a sequence of numbers $x_1,\ldots,x_n$, find the largest integer m such that $n \geq n_m$. If $(x_1,\ldots,x_{n_m}) \in F_m(l)$ then set $\Psi(x_1,\ldots,x_n) = l$. For $n_m \leq n < n_{m+1}$ the procedure attempts to distinguish between $\mathbf{X}_1,\ldots,\mathbf{X}_m$. If $\mathbf{X}^{(l)} \in \mathcal{P}$, then for each $m > l$,

$$\begin{aligned}
&\mu\{(X_1^{(l)},\ldots,X_{n_m}^{(l)}) \notin F_m(l)\} \\
&\quad\leq\; \mu\left\{(X_1^{(l)},\ldots,X_{n_m}^{(l)}) \in B_m^c(l)\right\} \;+\; \sum_{k\neq l} \mu\left\{(X_1^{(l)},\ldots,X_{n_m}^{(l)}) \in B_m(k)\right\}.
\end{aligned}$$

By virtue of (1) and (2), each term on the right hand side of the inequality is less than m^{-2}, and consequently

$$\sum_{m=1}^{\infty} \mu\{\Psi(X_1^{(l)},\ldots,X_{n_m}^{(l)}) \neq l\} \;<\; \infty\,.$$

Therefore $\Psi(X_1^{(l)},\ldots,X_{n_m}^{(l)}) \neq l$ for finitely many m with μ-probability one, and the result follows. □

Ornstein and Weiss (1990) described a universal procedure that will estimate a Bernoulli process (in the $\bar{d}$ sense) from finite initial segments of almost every sample path. They also gave several counterexamples showing there is no procedure that gives $\bar{d}$-consistent estimates of every finite alphabet stationary ergodic process.

3 A Counterexample

It is shown below that there is no universal density estimation procedure for ergodic processes. We restrict ourselves throughout to processes $\mathbf{X} = \{X_1, X_2, \ldots\}$, defined on $([0,1),\mathcal{B})$, and of the form $X_i(\omega) = T^i\omega$. For our purposes, this restriction implies no loss of generality.

Given a procedure Φ that is assumed to be universal, we exhibit an ergodic process $\{X_i\}$ having uniform marginal distribution, such that $\Phi(\cdot; X_1,\ldots,X_n)$ fails to converge. The process $\{X_i\}$ fools the procedure into believing that it is seeing a process with an oscillatory marginal distribution, and it does this infinitely often. In fact, the assumed universality of Φ is the key to its failure. The process $\{X_i\}$ is constructed using the method of cutting and stacking, an introduction to which can be found in Shields (1991) and Friedman (1970).

3.1 Pairwise Cutting and Stacking

A *column* of height m and width b is an ordered collection $C = \{I_j : 1 \leq j \leq m\}$ of m disjoint intervals $I_j = [a_j, a_j + b) \subseteq [0, 1)$ each having length b. One views the intervals I_j as being stacked on top of one another, with I_1 on the bottom, and I_j placed directly above I_{j-1}. Associated with every column C is a transformation $T_C : \bigcup_{j=1}^{m-1} I_j \to \bigcup_{j=2}^{m} I_j$ that maps each point in the first $m-1$ levels of C to the point directly above it:

$$T_C(a_j + s) = a_{j+1} + s\,.$$

for each $s \in [0, b)$ and each $1 \leq j < m$.

A column C' is said to be a *2-cut* of C if it is obtained by cutting C in half vertically, and then stacking the intervals to the right of the cut directly on top of those to the left. Thus C' is twice as high and half as wide as C, and is of the form $C' = \{I'_1, \ldots, I'_{2m}\}$, where

$$I'_j = \begin{cases} [a_j, a_j + b/2) & \text{if } 1 \leq j \leq m \\ [a_{j-m} + b/2, a_{j-m} + b) & \text{if } m + 1 \leq j \leq 2m\,. \end{cases}$$

Note that the associated transformation $T_{C'}$ is an extension of T_C, in that $T_{C'}(\omega) = T_C(\omega)$ for $\omega \in \bigcup_{j=1}^{m-1} I_j$. In particular, $T_{C'}$ maps the left half of I_m to the right half of I_1.

Let $C = \{I_1, \ldots, I_m\}$ be an initial column with support $U = \bigcup_{j=1}^{m} I_j$, and let $C_1, C_2, \ldots$ be successive 2-cuts of C. Then the mappings $T_{C_1}, T_{C_2}, \ldots$ form a chain whose limit

$$T^*_C(\omega) = \lim_{n\to\infty} T_{C_n}(\omega)\,.$$

is defined for each $\omega \in U$. Each map T_{C_n} is measurable, and such that $\lambda(T_{C_n}^{-1} A) = \lambda(A)$ for each Borel subset A of $[0, 1]$. As the range and domain of T_{C_n} increase to U, the limit T^*_C is measurable and preserves the normalized restriction of Lebesgue measure to U. In addition, it follows from Theorem 6.2 of Friedman (1970) that T^*_C is ergodic. If the initial column $C = \{[0, 1)\}$ then T^*_C is the von Neumann-Kakutani adding machine (c.f. von Neumann (1932)).

For each $k \geq 1$ let π_k be the partition of $[0, 1)$ into k left closed, right open subintervals of length $1/k$. Let $\mathcal{L}_k$ be the collection of bijections $\phi : \{1, 2, \ldots, k\} \times \{1, 2\} \to \pi_{2k}$. Each element $\phi \in \mathcal{L}_k$ uniquely describes two disjoint columns

$$C_1 = \{\phi(i, 1) : 1 \leq i \leq k\} \quad \text{and} \quad C_2 = \{\phi(i, 2) : 1 \leq i \leq k\}$$

of height k composed of intervals from π_{2k}. Conversely, any two such columns can be described by an element of $\mathcal{L}_k$. Let $U_j = \bigcup_{i=1}^k \phi(i,j)$ be the support of C_j, and define probability measures

$$\mu_\phi^j(A) = 2\lambda(A \cap U_j) \quad j = 1,2$$

with corresponding densities $f_\phi^j(x) = 2I_{U_j}(x)$. Define the transformation

$$T_\phi(\omega) = \begin{cases} T^*_{C_1}(\omega) & \text{if } \omega \in U_1 \\ T^*_{C_2}(\omega) & \text{if } \omega \in U_2. \end{cases}$$

It follows from the remarks above that T_ϕ is ergodic and measure preserving on $([0,1), \mathcal{B}, \mu_\phi^j)$ for $j = 1,2$. Let $\mathcal{L} = \bigcup_{k=1}^\infty \mathcal{L}_k$, and let $\mathcal{P}_1$ contain all those stationary ergodic processes on $([0,1), \mathcal{B}, \mu_\phi^j)$ of the form $X_i(\omega) = T_\phi^{i-1}\omega$ with $\phi \in \mathcal{L}$ and $j = 1,2$.

3.2 Principal Result

Theorem 1 *No density estimation procedure Φ is weakly L_1 consistent for every ergodic process having an absolutely continuous marginal distribution.*

Proof: Assume that Φ is weakly consistent for the family $\mathcal{P}_1$, and let $\epsilon_1, \epsilon_2, \ldots > 0$ be such that $\sum_{n=1}^\infty \epsilon_n < \infty$. We construct an ergodic measure preserving transformation T on $([0,1), \mathcal{B}, \lambda)$ such that

$$\limsup_{n\to\infty} \int_0^1 |\Phi(u; T^0\omega, \ldots, T^{n-1}\omega) - 1| du \geq \frac{1}{2} \tag{3}$$

for λ-almost every $\omega \in [0,1)$. The transformation T is defined by a sequence of pairs of columns $\{C_k^{(1)}, C_k^{(2)} : k \geq 1\}$.

Let $C_1^{(1)} = \{[0,1/2)\}$ and $C_1^{(2)} = \{[1/2,1)\}$ be initial columns, having a corresponding bijection $\phi_1 \in \mathcal{L}$. By assumption, there is an integer l_1 such that each of the sets

$$A_{1,j} = \left\{\omega : \int_0^1 |\Phi(u; T_{\phi_1}^0\omega, \ldots, T_{\phi_1}^{l_1-1}\omega) - f_{\phi_1}^j(u)| du \geq \frac{1}{2}\right\} \quad j = 1,2$$

has measure $\mu_{\phi_1}^j(A_{1,j}) < \epsilon_1$. Setting $A_1 = A_{1,1} \cup A_{1,2}$, it follows that $\lambda(A_1) < \epsilon_1$. Choose $m_1 \geq 1$ such that $l_1/2^{m_1} < \epsilon_1$, and let $C_2^{(j)}, \ldots, C_{m_1}^{(j)}$ be successive 2-cuts of $C_1^{(j)}$ for $j = 1,2$.

Suppose now that for some $n \geq 2$ we have selected integers $m_1, m_2, \ldots, m_{n-1}$ and formed two columns $C_{s(n-1)}^{(1)}$ and $C_{s(n-1)}^{(2)}$ where $s(n-1) = n - 2 + \sum_{i=1}^{n-1} m_i$. Mingle the two columns by stacking $C_{s(n-1)}^{(2)}$ on top of $C_{s(n-1)}^{(1)}$, and then cutting this new column in half. Let $C_{s(n-1)+1}^{(1)}$ be those intervals to the left of the cut and $C_{s(n-1)+1}^{(2)}$ those to the right, and let $\phi_n \in \mathcal{L}$ be the bijection corresponding to this new pair of columns. As Φ is weakly consistent for $\mathcal{P}_1$, there is an integer l_n for which $A_n = A_{n,1} \cup A_{n,2}$ with

$$A_{n,j} = \left\{ \omega : \int_0^1 |\Phi(u; T^0_{\phi_n}\omega, \ldots, T^{l_n-1}_{\phi_n}\omega) - f^j_{\phi_n}(u)| du \, \geq \, \frac{1}{2} \right\} ,$$

has Lebesgue measure $\lambda(A_n) < \epsilon_n$. Choose $m_n \geq 1$ so that $s(n) = s(n-1)+1+m_n$ satisfies

$$\frac{l_n}{2^{s(n)}} < \epsilon_n \, ,$$

and let $C^{(j)}_{s(n-1)+2}, \ldots, C^{(j)}_{s(n)}$ be successive 2-cuts of $C^{(j)}_{s(n-1)+1}$ for $j = 1, 2$. Continuing in this fashion we obtain a sequence $\{C^{(1)}_k, C^{(2)}_k : k \geq 1\}$ of pairs of columns, each having an associated transformations T_k. By construction, T_{k+1} extends T_k, and the domain of T_k increases to $[0, 1)$. The limiting transformation $T : [0, 1) \to [0, 1)$ is invertible and preserves Lebesgue measure. At each mingling operation the behavior of the limiting transformation T is described by a single column. As the width of these columns tends to zero, it follows from Theorem 6.2 of Friedman (1970) that T is ergodic. Thus the process $X_i(\omega) = T^{i-1}\omega$, $i \geq 1$, is stationary and ergodic, with marginal density $f \equiv 1$ on $[0, 1)$.

At each stage k the columns $C^{(1)}_k$ and $C^{(2)}_k$ are of height 2^{k-1} and have width 2^{-k}. Let $R(n, l)$ be the union of the top l intervals of $C^{(1)}_{s(n)}$ and $C^{(2)}_{s(n)}$. The choice of m_n insures that $\lambda(R(n, l_n)) \leq 2\epsilon_n$. By design, if $\omega \in R(n, l_n)^c \cap A^c_n$ then

$$T^0\omega, T\omega, \ldots, T^{l_n-1}\omega \; = \; T^0_{\phi_n}\omega, T_{\phi_n}\omega, \ldots, T^{l_n-1}_{\phi_n}\omega$$

and

$$\max_{j=1,2} \int_0^1 |\Phi(u; T^0_{\phi_n}\omega, \cdots, T^{l_n-1}_{\phi_n}\omega) - f^j_{\phi_n}(u)| du \; < \; \frac{1}{2} \, .$$

Since $\int_0^1 |1 - f^j_{\phi_n}| du = 1$, for each such ω,

$$\int_0^1 |\Phi(u; T^0\omega, \cdots, T^{l_n-1}\omega) - 1| du \; \geq \; \frac{1}{2} .$$

Thus if $B \triangleq \bigcap_{r=1}^\infty \bigcup_{n=r}^\infty (R(n, l_n) \cup A_n)$ then for each $\omega \in B^c$,

$$\limsup_{k\to\infty} \int_0^1 |\Phi(u; T^0\omega, \cdots, T^k\omega) - 1| du \; \geq \; \frac{1}{2} \, .$$

Summability of the ϵ_n insures that $\lambda(B) = 0$. This completes the proof. $\Box$.

Remark: A careful examination of the proof shows that, subsequent to each mingling operation, each column consists of alternating dyadic intervals in $[0, 1)$. More precisely, $C^{(j)}_{s(n)+1}$ contains the intervals

$$\left[\frac{2r + j - 3}{2^{s(n)+1}}, \frac{2r + j - 2}{2^{s(n)+1}} \right) \quad 1 \leq r \leq 2^{s(n)},$$

though they do not appear in their natural increasing order. It is therefore enough to let $\mathcal{P}_1$ contain only those processes whose associated bijections separately re-order the even and odd intervals of some dyadic partition. The processes in this restricted class have densities supported on alternating cells of the k'th dyadic partition of $[0,1)$, for some $k \geq 1$.

4 Some Corollaries

The transformation T constructed in the proof of Theorem 1 is uniquely specified by the integer sequence $\mathbf{m} = (m_1, m_2, \ldots)$, which describes the number of 2-cuts performed between successive mingling of the columns. Thus we may write $T = T_{\mathbf{m}}$. Let $\mathcal{P}_2$ be the family of processes defined by $X_i(\omega) = T_{\mathbf{m}}^i \omega$ for $\omega \in [0,1)$, where $\mathbf{m}$ ranges over all sequences of non-negative integers. Each element of $\mathcal{P}_2$ is stationary and ergodic with uniform marginal distribution on $[0,1)$. By virtue of Lemma 1, $\mathcal{P}_2$ is necessarily uncountable. The following corollaries are immediate from the proof of Theorem 1.

Corollary 1 *There is no consistent density estimation procedure for $\mathcal{P}^* = \mathcal{P}_1 \cup \mathcal{P}_2$.*

Remark: By modifying the construction of Theorem 1, one can establish an analogous result for a family of processes $\mathcal{P}^{**}$, each element of which is generated by a mixing transformation T.

Corollary 2 *For each $p \geq 1$, no density estimation procedure Φ is weakly L_p consistent for every ergodic process having a marginal density in L_p.*

As $\mathcal{P}_1$ is countable and every process in $\mathcal{P}_2$ has a uniform density, there exist consistent density estimation procedures for each family individually. Let Φ_1 be consistent for $\mathcal{P}_1$, and Φ_2 be consistent for $\mathcal{P}_2$. Given these procedures, the density estimation problem may, in principle, be solved by identifying the family to which the observed process $\mathbf{X} \in \mathcal{P}^*$ belongs. This observation and Corollary 1 lead to counterexamples for other problems.

For a given ergodic process $\mathbf{X} = \{X_i\}$, define $X_1^n = X_1, \ldots, X_n$, and let $S_X \subseteq \mathbb{R}$ be the support of the distribution of X_i. The following results show there is no universal procedure that will estimate S_X, or even $\lambda(S_X)$, from the finite initial segments of $\mathbf{X}$.

Corollary 3 *There is no procedure $\Psi : \mathbb{R}^* \to \mathcal{B}$ such that $\lambda\left(\Psi(X_1^n) \Delta S_X\right) \to 0$ in probability for every process $\mathbf{X} = \{X_i\} \in \mathcal{P}^*$.*

Corollary 4 *There is no procedure* $\Theta : \mathbb{R}^* \to \mathcal{B}$ *such that* $\Theta(X_1^n) \to \lambda(S_X)$ *in probability for every process* $\mathbf{X} = \{X_i\} \in \mathcal{P}^*$.

Proofs: It suffices to prove Corollary 4. If such a mapping Θ existed, then the compound procedure Φ defined by

$$\Phi(u : X_1^n) = \begin{cases} \Phi_1(u; X_1^n) & \text{if } \Theta(X_1^n) \leq 3/4 \\ \Phi_2(u; X_1^n) & \text{if } \Theta(X_1^n) > 3/4. \end{cases}$$

would be consistent for $\mathcal{P}^*$, which contradicts Corollary 1. Both Corollaries may also be proved directly by arguments similar to that given in Theorem 1.

Definition: A density estimation procedure $\Phi(\cdot)$ is invariant if for every sequence $x_1, x_2, \ldots \in \mathbb{R}$ and every density f, $\int |\Phi(u; x_1, \ldots, x_n) - f(u)| du \to 0$ if and only if $\int |\Phi(u; x_2, \ldots, x_{n+1}) - f(u)| du \to 0$.

Corollary 5 *For every invariant procedure* Φ *there is a process* $\mathbf{X} \in \mathcal{P}^*$ *with* $X_i \sim f$ *such that*

$$\limsup_{n \to \infty} \int_0^1 |\Phi(u; X_1^n) - f(u)| du \; > \; 0 \quad a.e. \tag{4}$$

Proof: If Φ is invariant then for each $j = 1, 2$ and every function $\phi \in \mathcal{L}$ the set

$$\left\{ \omega : \int_0^1 |\Phi(u; T_\phi^0 \omega, \ldots, T_\phi^{n-1} \omega) - f_\phi^j(u)| du \to 0 \right\}$$

is invariant under T_ϕ, and therefore its measure under μ_ϕ^j is either zero or one. As a consequence, if (4) fails to hold for each $\mathbf{X} \in \mathcal{P}_1$ then Φ is consistent for $\mathcal{P}_1$. In this case the proof of Theorem 1 shows that (4) holds for some $\mathbf{X} \in \mathcal{P}_2$. $\Box$

References

[1] I.A. Ahmad (1979). Strong consistency of density estimation by orthogonal series methods for dependent variables with applications. *Ann. Inst. Statist. Math.*, 31 279-288.

[2] A.R. Barron (1985). Logically smooth density estimation. Technical Report TR 56, Department of Statistics, Stanford University.

[3] J.V. Castellana and M.R. Leadbetter (1986). On smoothed probability density estimation for stationary processes. *Stoch. Proc. Appl.*, 21 179-193.

[4] L. Devroye and L. Györfi (1985). *Nonparametric Density Estimation: the L_1-view.* John Wiley, New York.

[5] L. Devroye and L. Györfi (1990). No empirical probability measure can converge in the total variation sense for all distributions. *Annals of Statistics*, 18 1496-1499.

[6] N.A. Friedman (1970). *Introduction to Ergodic Theory.* Van Nostrand Reinhold.

[7] L. Györfi (1981). Strongly consistent density estimate from ergodic sample. *J. Multivariate Analysis*, 11 81-84.

[8] L. Györfi and G. Lugosi (1992). Kernel density estimation from ergodic sample is not universally consistent. *Comput. Stat. Data Anal.*, 14 437-442.

[9] L. Györfi and E. Masry (1990). The L_1 and L_2 strong consistency of recursive kernel density estimation from dependent samples. *IEEE Trans. Inform. Theory*, 36 531-539.

[10] L. Györfi, W. Härdle, P. Sarda, and P. Vieu (1989). *Nonparametric Curve Estimation from Time Series.* Springer-Verlag, Berlin.

[11] P. Hall and J.D. Hart (1990). Convergence rates in density estimation for data from infinite-order moving average processes. *Probab. Th. Rel. Fields*, 87 253-274.

[12] J. von Neumann (1932). Zur Operatorenmethode in der klassichen Mechanik. *Ann. Math.*, 33 587-642.

[13] A.B. Nobel, G. Morvai, and S. Kulkarni (1996). Density estimation from an individual numerical sequence. Submitted to *IEEE Trans. Info. Theory.*

[14] D. Ornstein and B. Weiss (1990). How sampling reveals a process. *Ann. Prob.*, 18 905-930.

[15] M. Rosenblatt (1970). Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference*, M. Puri editor, Cambridge Univ. Press, London.

[16] G. Roussas (1967). Nonparametric estimation in Markov processes. *Ann. Inst. Statist. Math.*, 21 73-87.

[17] P.C. Shields (1991). Cutting and stacking: a method for constructing stationary processes. *IEEE Trans. Inform. Theory*, 37 1605-1617.

[18] B.W. Silverman (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

[19] S. Yakowitz (1989). Nonparametric density and regression estimation for Markov sequences without mixing assumptions. *J. Multivar. Analysis*, 30 124-136.

[20] S. Yakowitz and C. Heyde (1997). Private communication.

Terrence M. Adams
Department of Mathematics
The Ohio State University
Columbus, OH 43210
Email: tadams@math.ohio-state.edu

Andrew B. Nobel
Department of Statistics
University of North Carolina
Chapel Hill, NC 27599-3260
Email: nobel@stat.unc.edu