Density Estimation from an Individual Numerical Sequence

Andrew B. Nobel, Gusztáv Morvai, and Sanjeev Kulkarni *

July 31, 1997

Abstract

This paper considers estimation of a univariate density from an individual numerical sequence. It is assumed that (i) the limiting relative frequencies of the numerical sequence are governed by an unknown density, and (ii) there is a known upper bound for the variation of the density on an increasing sequence of intervals. A simple estimation scheme is proposed, and is shown to be L_1 consistent when (i) and (ii) apply. In addition it is shown that there is no consistent estimation scheme for the set of individual sequences satisfying only condition (i).

Appears in *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp.537-541, 1998.

Key words and phrases: Density estimation, individual sequences, bounded variation, ergodic processes.

*Andrew Nobel is with the Dept. of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260. Email: nobel@stat.unc.edu. Part of this work was completed while he was on leave as a Beckman Institute Fellow at the University of Illinois, U-C. His research was supported in part by NSF Grant DMS 9501926. Gusztáv Morvai is with Research Group for Informatics and Electronics of the Hungarian Academy of Sciences 1521 Goldmann György tér 3, Budapest, Hungary. Email: morvai@inf.bme.hu. Sanjeev Kulkarni is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544. Email: kulkarni@ee.princeton.edu.

1 Introduction

Estimation of a univariate density from a finite data set is an important problem in theoretical and applied statistics. In the most common setting, it is assumed that data are obtained from a stationary process $X_1, X_2, \ldots$ such that

$$\mathbb{P}\{X_i \in A\} = \int_A f dx \quad \text{for every Borel set} \quad A \subseteq \mathbb{R}$$

i.e. the common distribution of the X_i has density f, written $X_i \sim f$. For each $n \geq 1$ an estimate $\hat{f}_n$ of $f(\cdot)$ is produced from $X_1, \ldots, X_n$. The estimates $\{\hat{f}_n\}$ are said to be strongly L_1 consistent if $\int |\hat{f}_n - f| dx \to 0$ as $n \to \infty$ with probability one.

Common density estimation methods include histogram, kernel, nearest neighbor, orthogonal series, wavelet, spline, and likelihood based procedures. For an account of these methods, we refer the interested reader to the texts of Devroye and Györfi [4], Silverman [19], Scott [18], and Wand and Jones [20]. In establishing consistency and rates of convergence for estimation procedures like those above, many analyses assume that $X_1, X_2, \ldots$ are independent and identically distributed (i.i.d.), in which case the distribution of the process $\{X_i\}$ is completely specified by the marginal density f of X_1.

Complementing work for independent random variables, numerous results have also been obtained for stationary sequences exhibiting both short and long range dependence. Roussas [17] and Rosenblatt [16] studied the consistency and asymptotic normality of kernel density estimates from Markov processes. Similar results, under weaker conditions, were obtained by Yakowitz [21]. Györfi [5] showed that there is a simple kernel-based procedure Φ that is strongly L_2-consistent for every stationary ergodic process $\{X_i\}_{i=-\infty}^{\infty}$ 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 $E \int |h(u)|^2 du < \infty$. For additional work in this area, see also Ahmad [2], Castellana and Leadbetter [3], Györfi and Masry [7], Hall and Hart [9], and the references contained therein.

With these positive results have come examples showing that density estimation from strongly dependent processes can be problematic. In a result attributed to Shields, it was shown by Györfi, Härdle, Sarda and Vieu [8] that there are histogram density estimates, consistent for every i.i.d. process, that fail for some stationary ergodic process. Györfi and Lugosi [6] established a similar result for ordinary kernel estimates. Extending these results, Adams and Nobel [1] have recently shown that there is no density estimation procedure that is consistent for every stationary ergodic process.

With a view to considering density estimation in a more general setting, one may eliminate stochastic assumptions. Here we consider the estimation of an unknown density from an individual numerical sequence, which need not be the trajectory of a stationary stochastic process. We propose a simple estimation procedure that is applicable in a purely deterministic setting. This deterministic point of view is in line with recent work on individual sequences in information theory, statistics, and learning theory (cf. [22, 13, 12, 10]). Extending the techniques developed in this paper, Morvai, Kulkarni, and Nobel [14] consider the problem of regression estimation from individual sequences.

In many cases, results based on deterministic analyses can be applied to individual sample paths in a stochastic setting. Theorem 1 of this paper yields a positive result concerning density estimation from ergodic processes (see Corollary 1 below).

2 The Deterministic Setting

Let $f : \mathbb{R} \to \mathbb{R}$ be a univariate density function with associated probability measure $\mu_f(A) = \int_A f(x)dx$. An infinite sequence $\mathbf{x} = (x_1, x_2, \ldots)$ of numbers $x_i \in \mathbb{R}$ has *limiting density* f if

$$\hat{\mu}_n(A) \;=\; \frac{1}{n}\sum_{i=1}^{n} I\{x_i \in A\} \;\to\; \mu_f(A) \tag{1}$$

for every interval $A \subseteq \mathbb{R}$. A sequence $\mathbf{x}$ having a limiting density will be called *stationary.* Let $\Omega(f)$ be the set of stationary sequences with limiting density f.

Note that stationarity concerns the limiting behavior of relative frequencies, which need not converge to their corresponding probabilities at any particular rate. Stationarity says nothing about the mechanism by which the individual sequence $\mathbf{x}$ is produced. In particular, the limiting relative frequencies of a stationary sequence $\mathbf{x}$ are unchanged if one appends to $\mathbf{x}$ a prefix of any finite length.

The sample paths of ergodic processes provide one source of stationary sequences. The next proposition follows easily from Birkhoff's ergodic theorem.

Proposition 1 *If $X_1, X_2, \ldots$ are stationary and ergodic with $X_i \sim f$, then $\mathbf{X} = (X_1, X_2, \ldots) \in \Omega(f)$ with probability one.*

A univariate density estimation scheme is a countable collection Φ of Borel-measurable mappings $\phi_n : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$, $n \geq 1$. Thus ϕ_n associates every vector $(x_1, \ldots, x_n) \in \mathbb{R}^n$ with a function $\phi_n(\cdot : x_1, \ldots, x_n)$, which is viewed as the estimate of an unknown density associated with the sequence $x_1, \ldots, x_n$. These estimates may take negative values, and they need not integrate to one. In particular, no regularity conditions are imposed on the behavior of ϕ_n as a function of its inputs.

A scheme Φ is L_1 consistent for a a collection Ω of stationary sequences if for each $\mathbf{x} \in \Omega$,

$$\int |\phi_n(x : x_1, \ldots, x_n) - f(x)| dx \to 0\,,$$

as $n \to \infty$, where f is the limiting density of $\mathbf{x}$. A scheme Φ is universal if it is L_1 consistent for the set Ω^* of all stationary sequences. Note that, for i.i.d. data, a density estimation scheme is called universal if it is consistent for every marginal density f. The notion of universality defined above is considerable stronger, as there are no constraints apart from stationarity placed on the structure of the individual sequences. In what follows, when $\mathbf{x} = x_1, x_2, \ldots$ is fixed, $\phi(x : x_1, \ldots, x_n)$ will be denoted by $\phi_n(x)$.

Recall that the total variation of a real-valued function h defined on an interval $[a, b) \subseteq \mathbb{R}$ is given by

$$V(h : a, b) = \sup \sum_{i=1}^{n} |h(t_i) - h(t_{i-1})|\,,$$

where the supremum is taken over all finite ordered sequences $a \leq t_0 < \cdots < t_n < b$. For each nondecreasing function $\alpha : \mathbb{Z}^+ \to (0, \infty)$ let $\mathcal{F}(\alpha)$ be the set of all densities f on $\mathbb{R}$ such that $V(f : -i, i) < \alpha(i)$ for $i \geq 1$, and let

$$\Omega(\alpha) = \bigcup_{f \in \mathcal{F}(\alpha)} \Omega(f)$$

be the collection of all those stationary sequences having limiting densities in $\mathcal{F}(\alpha)$.

Given a function $\alpha(\cdot)$ as above, we propose a simple histogram based procedure that is consistent for $\Omega(\alpha)$. For each $k \geq 1$ let π_k be the partition of $\mathbb{R}$ into dyadic intervals of the form

$$A_{k,j} = \left[\frac{j}{2^k}, \frac{j+1}{2^k}\right) \quad \text{with} \quad j \in \mathbb{Z}\,,$$

and let $\pi_k[x]$ be the unique cell of π_k containing x. Let $\{b_n\}$ be any sequence of positive integers tending to infinity. For each sequence of numbers $x_1, \ldots, x_n$ and each $k \geq 1$ define histogram density estimates

$$\hat{h}_{n,k}(x) = \frac{2^k}{n} \sum_{i=1}^{n} I\{x_i \in \pi_k[x]\}\,. \tag{2}$$

Our estimate is selected from among the histograms $\hat{h}_{n,k}$ by selecting a suitable value of k. Find the partition index

$$k_n = \max\left\{1 \leq k \leq b_n : V(\hat{h}_{n,k} : -i, i) < 4\alpha(i) \quad \text{for} \quad 1 \leq i \leq k\right\} \tag{3}$$

and define

$$\phi_n^*(x : x_1, \ldots, x_n) = \hat{h}_{n,k_n}(x) \,. \tag{4}$$

If the conditions defining k_n are not satisfied for any $1 \leq k \leq b_n$, then set $\phi_n^* \equiv 0$.

Theorem 1 *Let $\alpha : \mathbb{Z}^+ \to (0, \infty)$ be a fixed, non-decreasing function. The estimation scheme $\Phi^* = \{\phi_n^*\}$ defined by (2)-(4) is L_1-consistent for $\Omega(\alpha)$. Thus for every stationary sequence $\mathbf{x}$ with limiting density $f \in \mathcal{F}(\alpha)$, $\int |\phi_n^*(x) - f(x)| dx \to 0$.*

Corollary 1 *Let $\alpha(\cdot)$ be fixed and let ϕ_n^* be defined by (2)-(4). For every stationary ergodic process $\{X_i\}$ such that $X_i \sim f$ with $f \in \mathcal{F}(\alpha)$,*

$$\int |\phi_n^*(x : X_1, \ldots, X_n) - f(x)| dx \to 0$$

as $n \to \infty$ with probability one.

Example: Fix $\gamma > 0$, and consider the class of stationary ergodic processes $\{X_i\}$ such that $X_i \sim f$ with $V(f : -\infty, \infty) < 2\gamma$. This class includes, but is not limited to, processes having uniform, exponential, and normal marginal densities with arbitrary means, under the restriction that $Var(X_i)$ is greater than $(12\gamma^2)^{-1}$, γ^{-2}, and $(2\pi\gamma^2)^{-1}$, respectively. By Corollary 1 there is a strongly consistent density estimation procedure Φ^* for this class of processes.

Remark: The variations used to define ϕ_n^* depend on the cumulative difference between the relative frequencies of adjacent cells:

$$V(\hat{h}_{n,k} : -i, i) = 2^{-k} \sum_{j=-i2^k}^{i2^k - 2} |\hat{\mu}_n(A_{k,j}) - \hat{\mu}_n(A_{k,j+1})| \,. \tag{5}$$

To find ϕ_n^*, put $x_1, \ldots, x_n$ in increasing order, and then calculate $V(\hat{h}_{n,k} : -i, i)$ for each $k = 1, \ldots, b_n$ and each $i = 1, \ldots, k$ by scanning the ordered x_i from left to right. This will require at most $O(n \log n + nb_n)$ operations.

In order to apply the procedure Φ^* described in (2)-(4), one must know before seeing $\mathbf{x}$ that the variation of its limiting density is less than a known constant on every interval of the form $[-i, i)$. The following result shows that this requirement cannot be materially weakened.

Theorem 2 *Let $\mathcal{F}$ be the collection of densities f supported on $[0,1]$ for which $V(f:0,1)$ is finite. There is no L_1 consistent density estimation scheme for*

$$\Omega = \bigcup_{f \in \mathcal{F}} \Omega(f).$$

In particular, there is no universal density estimation scheme for individual sequences.

If an upper bound on the variance of the unknown density f were known, the scheme of Theorem 1 would provide consistent estimates of f.

Given any density estimation scheme $\Phi = \{\phi_n\}$, the proof of Theorem 2 shows how one may construct a stationary sequence $\mathbf{x}$, depending on Φ, for which $\phi_n(\cdot)$ fails to converge. A related argument is used by Adams and Nobel [1] to show that there is no universal density estimation scheme for stationary ergodic processes. As a universal density estimation scheme for individual sequences would, by virtue of Proposition 1, yield a universal scheme for ergodic processes, their result also implies Theorem 2.

The proof of Theorem 1 is given in the next section after several preliminary results. The proof of Theorem 2 is given in Section 4.

3 Proof of Theorem 1

Definition: For each partition π of $\mathbb{R}$ into finite intervals and each $f \in L_1$ define

$$(f \circ \pi)(x) = \frac{1}{l(\pi[x])} \int_{\pi[x]} f(u) du \,,$$

where $l(A)$ denotes the length of an interval A. Note that $f \circ \pi$ is piecewise constant on the cells of π.

Lemma 1 *Let $\pi_1, \pi_2, \ldots$ be the partitions used to define the estimates ϕ_n^*. For each pair of integers $k, i \geq 1$,*

$$V(f \circ \pi_k : -i, i) \;\leq\; 3V(f : -i, i).$$

Moreover, if $\mathbf{x} \in \Omega(f)$ then

$$\lim_{n \to \infty} V(\hat{h}_{n,k} : -i, i) \;=\; V(f \circ \pi_k : -i, i).$$

Proof: For f non-decreasing it is immediate that $V(f \circ \pi_k : -i, i) \leq V(f : -i, i)$. If $V(f : -i, i) = C < \infty$ then $f(x) = u(x) - v(x)$ where $u(\cdot)$ and $v(\cdot)$ are non-decreasing, $V(u : -i, i) \leq C$ and $V(v : -i, i) \leq 2C$ (cf. Kolmogorov and Fomin [11]). It follows from the definition that $f \circ \pi_k = u \circ \pi_k - v \circ \pi_k$, and since u and v are non-decreasing, so are $u \circ \pi_k$ and $v \circ \pi_k$. Therefore

$$\begin{aligned} V(f \circ \pi_k : -i, i) &= V(u \circ \pi_k - v \circ \pi_k : -i, i) \\ &\leq V(u \circ \pi_k : -i, i) + V(v \circ \pi_k : -i, i) \\ &\leq V(u : -i, i) + V(v : -i, i) \\ &\leq 3C \end{aligned}$$

as the variation of the sum is less than the sum of the variations. To establish the second claim, note that as $n \to \infty$

$$\begin{aligned} V(\hat{h}_{n,k} : -i, i) &= 2^{-k} \sum_{j=-i2^k}^{i2^k-2} |\hat{\mu}_n(A_{k,j}) - \hat{\mu}_n(A_{k,j+1})| \\ &\to 2^{-k} \sum_{j=-i2^k}^{i2^k-2} |\mu_f(A_{k,j}) - \mu_f(A_{k,j+1})| \\ &= V(f \circ \pi_k : -i, i)\,. \end{aligned}$$

□

Lemma 2 *Let* $\mathbf{x} \in \Omega(\alpha)$ *with limiting density* $f \in \mathcal{F}(\alpha)$*. Then the partition index* k_n *of the density estimate* ϕ_n^* *tends to infinity with* n*.*

Proof: By Lemma 1, for arbitrary $K \geq 1$ and for all $i = 1, \ldots, K$,

$$\lim_{n\to\infty} V(\hat{h}_{n,K} : -i, i) \;=\; V(f \circ \pi_K : -i, i) \leq 3V(f : -i, i) < 3\alpha(i).$$

Thus by definition of k_n, $\liminf_{n\to\infty} k_n \geq K$. □

Proof of Theorem 1: Let $\mathbf{x} \in \Omega(\alpha)$ be a fixed stationary sequence with limiting density $f \in \mathcal{F}(\alpha)$. For each $n \geq 1$ such that $k_n \geq 1$ define the error function

$$g_n(x) = \phi_n^*(x : x_1, \ldots, x_n) - f(x) = \hat{h}_{n,k_n}(x) - f(x),$$

and note that for all $1 \leq i \leq k_n$,

$$V(g_n : -i, i) \;\leq\; V(\phi_n^* : -i, i) + V(f : -i, i) \;<\; 5\alpha(i). \tag{6}$$

Fix $\epsilon > 0$. Select an integer $L \geq 1$ such that

$$\int_{|x|\geq L} f(x)dx \leq \epsilon \tag{7}$$

and define

$$\delta = \frac{\epsilon}{L}\,. \tag{8}$$

Finally, choose an integer $K \geq 1$ so large that

$$2^{-K} < \frac{\epsilon\delta}{\alpha(L)(50\alpha(L) + 5\delta)}. \tag{9}$$

As $\mathbf{x} \in \Omega(f)$ and the partitions π_k are nested, there exists an integer $N = N(\mathbf{x}, \epsilon, f, \alpha)$ such that for $n \geq N$ one has $k_n \geq \max\{K, L\}$,

$$|\int_A g_n(x)dx| \;=\; |\hat{\mu}_n(A) - \mu_f(A)| \;<\; \frac{\delta}{2} \cdot 2^{-K} \tag{10}$$

for $A \in \pi_K$ with $A \subseteq [-L, L)$, and

$$|\hat{\mu}_n\{|x| \geq L\} - \mu\{|x| \geq L\}| \leq \epsilon\,. \tag{11}$$

For each n let

$$H_n = \{x \in \mathrm{I\!R} : |g_n(x)| > \delta\}$$

contain those points having large error, and let

$$\mathcal{H}_n = \{A \in \pi_K : A \cap H_n \neq \emptyset, A \subseteq [-L, L)\}.$$

Fix $n \geq N$ and consider a set $A \in \mathcal{H}_n$. By definition, there exists a point $x \in A$ such that $|g_n(x)| > \delta$. Assume for the moment that $g_n(x) > \delta$. It follows from (10) that there is a point $y \in A$ such that $g_n(y) < \delta/2$, and therefore

$$\sup_{x,y \in A} |g_n(x) - g_n(y)| > \delta/2\,. \tag{12}$$

As $k_n \geq L$ the variation of g_n on A is less than $5\alpha(L)$ by (6), so that for each $z \in A$,

$$g_n(z) \leq g_n(y) + 5\alpha(L) \leq \frac{\delta}{2} + 5\alpha(L)\,,$$

and

$$g_n(z) \geq g_n(x) - 5\alpha(L) \geq \frac{\delta}{2} - 5\alpha(L)\,.$$

Therefore,

$$\sup_{z \in A} |g_n(z)| \leq \frac{\delta}{2} + 5\alpha(L)\,. \tag{13}$$

A similar argument in the case $g_n(x) < -\delta$ shows that both (12) and (13) hold for each $A \in \mathcal{H}_n$. It is immediate from (12) that

$$\frac{\delta}{2}|\mathcal{H}_n| \leq V(g_n : -L, L) < 5\alpha(L),$$

and consequently

$$|\mathcal{H}_n| < \frac{10\alpha(L)}{\delta}. \tag{14}$$

For each $n \geq N$ the integrated error between ϕ_n^* and f may be decomposed as follows:

$$\begin{aligned} &\int |\phi_n^*(x) - f(x)|dx \\ \leq\ & \sum_{A \in \mathcal{H}_n} \int_A |g_n(x)|dx \ + \sum_{A \notin \mathcal{H}_n, A \subseteq [-L,L)} \int_A |g_n(x)|dx \ + \int_{|x| \geq L} |g_n(x)|dx \\ \triangleq\ & \Theta_1 + \Theta_2 + \Theta_3 \end{aligned}$$

Inequalities (13), (14) and (9) imply that

$$\Theta_1 \ \leq \ \sum_{A \in \mathcal{H}_n} \int_A (\frac{\delta}{2} + 5\alpha(L))dx \ \leq \ \left(5\alpha(L) + \frac{\delta}{2}\right) \frac{10\alpha(L)}{\delta 2^K} \ \leq \ \epsilon\,,$$

and by virtue of (8),

$$\Theta_2 \ \leq \ \int_{[-L,L)} \delta dx = \delta \cdot 2L = 2\epsilon\,.$$

Finally, it follows from (7) and (11) that

$$\Theta_3 \ \ \leq \ \ \hat{\mu}_n\{|x| \geq L\} + \mu\{|x| \geq L\} \leq 3\epsilon.$$

Combining these three bounds shows that

$$\limsup_{n \to \infty} \int |\phi_n^*(x) - f(x))|dx \leq 6\epsilon\,,$$

and as ϵ was arbitrary, the desired L_1 convergence of ϕ_n^* to f follows. □.

4 Proof of Theorem 2

The following result can be established by a straightforward extension of the Glivenko Cantelli Theorem, or by a bracketing argument (c.f. Pollard [15]).

Lemma 3 *Let $\mathcal{A}$ be the collection of all (finite and infinite) intervals in* $\mathbb{R}$. *If $\mathbf{x} \in \Omega(f)$ then*

$$\sup_{A \in \mathcal{A}} |\hat{\mu}_n(A) - \mu_f(A)| \ \to \ 0\,.$$

Proof of Theorem 2: Consider the family $\mathcal{F}_0 = \{h_1, h_2, \ldots\} \subseteq \mathcal{F}$ of Rademacher densities where

$$h_k(x) = \begin{cases} 2 & \text{if } 2j2^{-k} \leq x < (2j+1)2^{-k} \text{ for some } 0 \leq j < 2^{k-1} \\ 0 & \text{otherwise}\ . \end{cases}$$

Note that each h_j is supported on $[0,1]$ and that $\int |h_j(x)-h_k(x)|dx = 1$ whenever $j \neq k$. Let μ_k be the probability measure having density h_k, and for each finite sequence $u_1, \ldots, u_m \in [0,1]$ let

$$\Delta_k(u_1, \ldots, u_m) = \sup_{A \in \mathcal{A}} \left| \frac{1}{m} \sum_{j=1}^{m} I_A(u_j) - \mu_k(A) \right| ,$$

measure the distance between μ_k and the empirical measure of $u_1, \ldots, u_m$.

We show that if Φ is consistent for $\mathcal{F}_0$ then there is a stationary sequence $\mathbf{x}^*$ whose limiting density is identically one on $[0,1]$, but is such that $\phi(\cdot : x_1^*, \ldots, x_n^*)$ fails to have a limit in L_1. For each $k \geq 1$ select a sequence $\mathbf{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \ldots) \in \Omega(h_k)$ (e.g. a typical sample sequence from an i.i.d. process with density h_k), and define

$$m_k = \min \left\{ M : \sup_{m \geq M} \Delta_k(x_1^{(k)}, \ldots, x_m^{(k)}) \leq \frac{1}{k+1} \right\} .$$

Lemma 3 insures that m_k exists and is finite.

Fix any procedure $\Phi = \{\phi_1, \phi_2, \ldots\}$ that is consistent for $\mathcal{F}_0$ and consider the infinite sequence $\mathbf{x}^{(1)}$. As $h_1 \in \mathcal{F}_0$,

$$\int |\phi_n(x : x_1^{(1)}, \ldots, x_n^{(1)}) - h_1(x)| dx \to 0$$

as $n \to \infty$. Therefore there is an integer $n_1 \geq m_2$ and a corresponding initial segment $\mathbf{y}^{(1)} = x_1^{(1)}, \ldots, x_{n_1}^{(1)}$ of $\mathbf{x}^{(1)}$ such that

$$\int |\phi_{n_1}(x : \mathbf{y}^{(1)}) - h_1(x)| dx \leq \frac{1}{4} \quad \text{and} \quad \Delta_1(\mathbf{y}^{(1)}) \leq \frac{1}{2} .$$

Now suppose that one has constructed a sequence $\mathbf{y}^{(k)}$ of finite length n_k from initial segments of $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ such that

$$\int |\phi_{n_k}(x : \mathbf{y}^{(k)}) - h_k(x)| dx \leq 1/4 , \tag{15}$$

$$\Delta_k(\mathbf{y}^{(k)}) \leq (k+1)^{-1}, \tag{16}$$

and

$$n_k \geq k \cdot m_{k+1} . \tag{17}$$

As $\mathbf{y}^{(k)}$ is finite, the concatenation $\mathbf{y}^{(k)} \cdot \mathbf{x}^{(k+1)}$ is contained in $\Omega(h_{k+1})$. It follows from the consistency of Φ and Lemma 3 that when n is large enough each initial segment $\mathbf{y}^{(k+1)} = \mathbf{y}^{(k)} \cdot (x_1^{(k+1)}, \ldots, x_{n-n_k}^{(k+1)})$ of $\mathbf{y}^{(k)} \cdot \mathbf{x}^{k+1}$ satisfies (15) and (16) with k replaced by $k+1$. Select $n_{k+1} > n_k$ so large that the same is true of (17).

As $\mathbf{y}^{(k+1)}$ is a proper extension of $\mathbf{y}^{(k)}$, repeating the above process indefinitely yields an infinite sequence $\mathbf{x}^*$. By construction, the functions $\phi_n(\cdot) = \phi(\cdot : x_1^*, \ldots, x_n^*)$ do not converge in L_1. Indeed, it follows from (15) and the triangle inequality that $\int |\phi_{n_k} - \phi_{n_l}| dx \geq 1/2$ whenever $k \neq l$.

It remains to show that the limiting density of $\mathbf{x}^*$ is uniform on $[0,1]$. To this end, fix $k \geq 1$ and let $A \subseteq [0,1]$ be an interval of length $l(A)$. It is easily verified that

$$|\mu_k(A) - l(A)| \leq 2^{-k+1} \leq \frac{1}{k}. \tag{18}$$

Let $\hat{\mu}_n(A)$ be the empirical distribution of A under $x_1^*, \ldots, x_n^*$, and for each $1 \leq r \leq n_{k+1} - n_k$ define

$$\hat{\mu}'_{r,k}(A) = \frac{1}{r} \sum_{j=n_k+1}^{n_k+r} I_A(x_i^*)$$

It follows from the equation

$$\hat{\mu}_{n_k+r}(A) = \frac{n_k}{n_k+r} \cdot \hat{\mu}_{n_k}(A) + \frac{r}{n_k+r} \cdot \hat{\mu}'_{r,k}(A)$$

that the difference

$$\begin{aligned} |\hat{\mu}_{n_k+r}(A) - l(A)| &\leq \frac{n_k}{n_k+r} \cdot |\hat{\mu}_{n_k}(A) - l(A)| \; + \; \frac{r}{n_k+r} \cdot |\hat{\mu}'_{r,k}(A) - l(A)| \\ &\triangleq I + II. \end{aligned}$$

By virtue of (16) and (18),

$$I \;\leq\; |\hat{\mu}_{n_k}(A) - \mu_k(A)| + |l(A) - \mu_k(A)| \;\leq\; \frac{1}{k+1} + \frac{1}{k}.$$

If $n_{k+1} - n_k \geq r \geq m_{k+1}$ then

$$\Delta_{k+1}(x^*_{n_k+1}, \ldots, x^*_{n_k+r}) \;=\; \Delta_{k+1}(x_1^{(k+1)}, \ldots, x_r^{(k+1)}) \;\leq\; \frac{1}{k+2}$$

and therefore

$$II \;\leq\; |\hat{\mu}'_{r,k}(A) - \mu_{k+1}(A)| \;+\; |\mu_{k+1}(A) - l(A)| \;\leq\; \frac{1}{k+2} + \frac{1}{k+1}.$$

On the other hand, if $1 \leq r < m_{k+1}$ then (17) implies that

$$II \;\leq\; \frac{2r}{n_k+r} \;\leq\; \frac{2r}{kr+r} \;=\; \frac{2}{k+1}.$$

These bounds insure that

$$\max\{|\hat{\mu}_n(A) - l(A)| : n_k < n \leq n_{k+1}\} \leq \frac{4}{k},$$

and consequently

$$\lim_{n\to\infty} |\hat{\mu}_n(A) - l(A)| = 0.$$

As $A \in \mathcal{A}$ was arbitrary, $\mathbf{x}^*$ is stationary with limiting density $f(x) = 1$ on $[0,1]$. $\Box$

Acknowledgments

The authors wish to thank László Györfi for his helpful comments and suggestions.

References

[1] T.M. Adams and A.B. Nobel. On density estimation from ergodic processes. To appear in Ann. Probab., 1997.

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

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

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

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

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

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

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

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

[10] D. Haussler, J. Kivinen, and M. Warmuth. Tight worst-case loss bounds for predicting with expert advice. *Proc. European Conference on Computational Learning Theory*, 1994.

[11] A.N. Kolmogorov and S.V. Fomin. *Introductory Real Analysis.* Dover, Mineola, 1970.

[12] S.R. Kulkarni and S.E. Posner. Rates of convergence for nearest neighbor estimation under arbitrary sampling. *IEEE Trans. on Information Thoery*, IT-41:1028-1039, 1995.

[13] N. Merhav, M. Feder, and M. Gutman. Universal prediction of individual sequences. *IEEE Trans. on Information Thoery*, IT-38:1258-1270, 1992.

[14] G. Morvai, S. Kulkarni, and A.B. Nobel. Regression estimation from an individual sequence. In preparation. 1997.

[15] D. Pollard. *Convergence of Stochastic Processes.* Springer-Verlag, New York, 1984.

[16] M. Rosenblatt. Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference*, M. Puri editor, pages 199-213. Cambridge Univ. Press, London, 1970.

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

[18] D.W. Scott *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, 1992.

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

[20] M.P. Wand and M.C. Jones. *Kernel Smoothing.* Chapman and Hall, London, 1995.

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

[22] J. Ziv. Coding theorems for individual sequences. *IEEE Trans. on Information Thoery*, IT-24:405-412, 1978.