Dec. 2006 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Vol. 23 No. 6

Article ID:1005-3085(2006)06-1068-07

Finite Element Density Estimator with Multi-stage Adaptive Filters*

DAI Wan-yang

(Department of Mathematics, Nanjing University, Nanjing 210093)

Abstract: Based on sample data which may be correlated and have certain ergodic property, a finite element estimator with multi-stage adaptive filters is designed to estimate the associated (stationary) density function. Numerical examples are presented to show the effectiveness of the estimator. Meanwhile, the stability property of the estimator is also discussed.

Keywords: density estimation; finite element interpolation; adaptive filter; weak convergence

Classification: AMS(2000) 62G07 **CLC number:** O21 **Document code:** A

1 Introduction

In this paper, we design a finite element estimator with adaptive filters to estimate the (stationary) probability density from a sequence of observed sample data, and at the same time, we present numerical implementation experiences to show the effectiveness of the estimator. Compared with the well studied kernel methods such as those summarized in [1], the designed estimator has some flexibility and advantages in estimating densities with certain non-smooth corners and/or discontinuous jump points, and in handling boundary bias. Some related works, which use finite element and wavelet interpolations to estimate (stationary) density functions, can be found, for examples, in [2,3,4].

In our numerical implementations and stability study for the estimator, we allow the sequence of observations to be correlated in certain way. Specifically, it is assumed that the observed sample data denoted by $\{X_1, X_2, \cdots, X_n\}$ have the following ergodic property (strong law of large numbers),

$$\frac{f(X_1)+f(X_2)+\cdots+f(X_n)}{n} \to E_\pi f(X) = \int_{x\in S} f(x)\mathrm{d}\pi(x), \quad \text{as } n\to\infty, \text{ a.s.} \tag{1}$$

where $\pi(\cdot)$ is some unknown probability measure with density function $p(x)$, X is a random variable associated with $\pi(\cdot)$, $E_\pi f(X)$ denotes the expectation of $f(X)$ under π and f is a continuous (or more generally, Borel integrable) function. There are a number of applications which have the average convergence property as described in (1). For examples, independent and identically distributed (i.i.d.) observations drawn from an unknown probability density $p(x)$ (to be estimated); $\{X_1, X_2, \cdots, X_n\}$ are sample data or are generated from a stationary

Received date: 2004-09-06. **Biography**: Dai Wanyang (Born in 1963), Male, Ph.D., Professor, Research fields: Probability and Statistics, Stochastic Processes and Stochastic Networks.

*__Foundation item:__ The National Natural Science Foundation of China (10371053).

process (sequence), in this case, π will be the stationary distribution of the process or sequence (see corresponding Ergodic Theorems in [5] for details); $\{X_1, X_2, \cdots, X_n\}$ are sample data generated from some Harris recurrent Markovian process, namely, the probability of an infinite number of returns to an arbitrary set $A \subset S$ is one, in this case, π is the stationary distribution for the process (the associated Ergodic Theorem can be found in [6]).

The rest of the paper is organized as follows. In Section 2, we present the estimator and discuss its related stability. In Section 3, several numerical examples are presented to illustrate the usage and effectiveness of the estimator. Finally, in Section 4, we prove the stability theorem.

2 Density Estimator with Multi-stage Adaptive Filters

In this section, we design a finite element algorithm to estimate $p(x)$ and present the associated weak convergence theorem. To define our finite element global basis, we introduce the following shape function

$$\phi(x) = 1 - |x| \quad \text{for} \quad x \in [-1, 1]. \tag{2}$$

For $m \in \{1, 2, \cdots\}$, let $b_m = m$ and $h = b_m/m^3$, or $b_m/(\alpha m^2)$ with α being some suitable positive constant, moreover, subdivide the interval $[-b_m, b_m]$ into a union of elements by the uniform mesh $-b_m = x_{-m^3} < \cdots < x_{-1} < x_0 = 0 < x_1 \cdots < x_{m^3} = b_m$, then we define the following chapeau or hat functions (see, for example, [7])

$$\phi_i(x) = \begin{cases} \phi\left(\frac{x-ih}{h}\right) & \text{if } x \in [(i-1)h, (i+1)h] \cap [-b_m, b_m], \\ 0 & \text{otherwise}, \end{cases} \tag{3}$$

where the index $i \in \{-m^3, -(m^3-1), \cdots, -1, 0, 1, \cdots, (m^3-1), m^3\}$. Thus, the probability density $p(x)$ can be approximated by the below finite element interpolation

$$p^m(x) = \sum_{i=-m^3}^{m^3} c_i \phi_i(x), \tag{4}$$

where the coefficient c_i for each i is constant. Hence, from the sample data $\{X_1, \cdots, X_n\}$ and each integer $d \geq 1$, we can design the following density estimator

$$p_{m,n,d}(x) = p_{m,n,d}(x; X_1, \cdots, X_n) = \sum_{i=-m^3}^{m^3} \hat{c}_i^d \phi_i(x), \tag{5}$$

where $\hat{c}_i^d$ is the dth $(d \geq 1)$ estimated coefficient for each $i \in \{-m^3, \cdots, m^3\}$ via the following adaptive smoothing filter,

$$\hat{c}_i^{d+1} = \sum_{j=i-\beta_i}^{i+\beta_i} \alpha_j \hat{c}_j^d \quad \text{with} \quad \hat{c}_i = \frac{1}{nh} \sum_{j=1}^{n} \phi_i(X_j), \tag{6}$$

where $(2\beta_i+1)$ is the moving average window size of the filter in (6), the integer-valued constant β_i can be chosen dynamically with the variation of the index i, moreover, the coefficients $\{\alpha_j\}$

are the associated moving average weights satisfying

$$\sum_{j=i-\beta_i}^{i+\beta_i} \alpha_j = 1 \quad \text{and} \quad \alpha_j \geq 0,\ \alpha_{i-j} = \alpha_{i+j} \ \text{for all} \ j \in \{0, 1, \cdots, \beta_i\}. \tag{7}$$

From our numerical examples presented in the next section, one can see that $p_{m,n,d}(x)$ is indeed a good approximation to $p(x)$ for suitable chosen integer d. Moreover, the estimator (5) has some flexibility and advantages in estimating densities with non-smooth corners and/or discontinuous jump points, and in handling boundary bias. As for the stability of the estimator, we have the following limit theorem.

Theorem 1 If $E|X| < \infty$, then for each m and $d \geq 1$, $p_{m,n,d}(\cdot)$ in (5) converges to a unique Borel measure $\mu_m(\cdot)$ weakly in the sense that, for each $f \in C_c(R)$ (the space of continuous real-valued functions with compact support on $R = (-\infty, \infty)$),

$$\int_R f(x) p_{m,n,d}(x) \mathrm{d}x \to \int_R f(x) \mathrm{d}\mu_m(x) \quad \text{as} \ n \to \infty. \tag{8}$$

Remark We conjecture that, for suitable chosen d, the Radon Nikodym deriveative of μ_m in terms of the Lebesgue measure $\mathrm{d}x$ equals $p(x)$ almost everywhere as m tends to infinite. Here we leave this conjecture as an open problem for future study.

3 Numerical Examples

In this section, we present several examples as given in Figure 1 and Figure 2 to compare the estimating performance among the finite element density estimator (5), exact density and the widely discussed kernel method (see, for example, [1]). Here, we also point out that, from our simulation experiences, we believe that estimation errors appeared in our examples are due to generated pseudo random numbers which are treated as treated sample data. In fact, how to generate accurate random numbers is also an active research area.

To be convenient for readers, we recall here the kernel estimation formula

$$\hat{p}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right). \tag{9}$$

In (9), the kernel $K(\cdot)$ has various choices and the standard normal density is chosen in our implementation. In addition, h_n should also be carefully selected such that it satisfies $h_n \to 0$ and $nh_n \to \infty$. From existed numerical results and our own simulation experience, we find that $h_n = 1/(1.05 * n^{0.2})$ is a good choice.

Example 1 In this example, the following standard normal density is discussed

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad -\infty < x < \infty. \tag{10}$$

Based on two uniform random number sequences r_n^1 and r_n^2 derived from the Kiss generator (see, for example, [6]), we use the transformation method to generate associated normal sample data $\{x_n\}$.

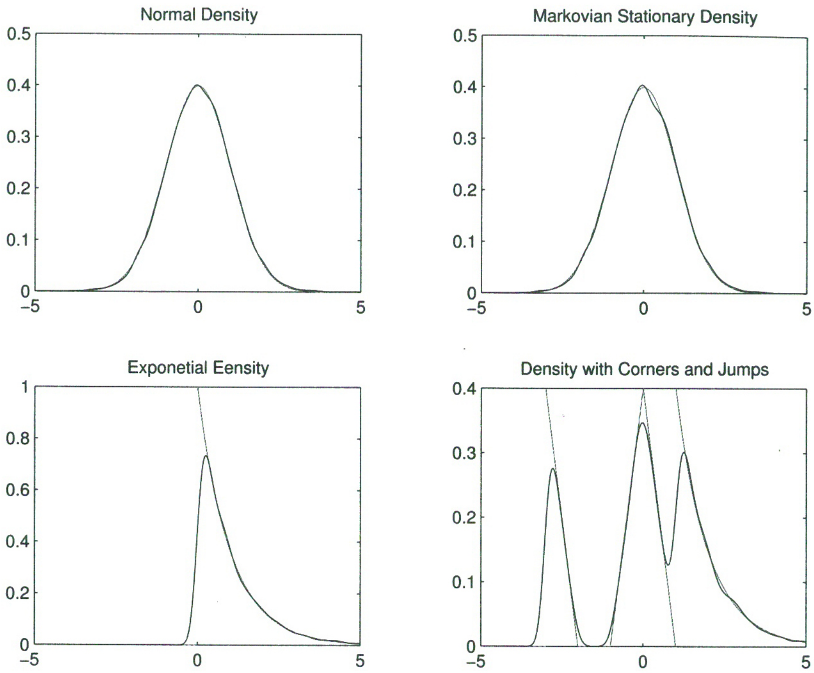

Figure 1: Estimated densities by Kernel method with $h_n = 1/(1.05 * n^{0.2})$.

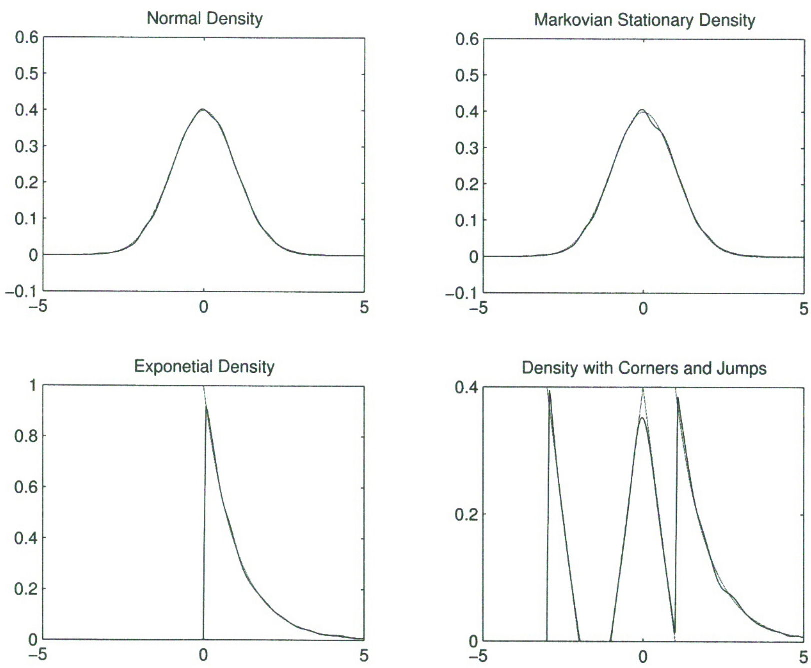

Figure 2: Estimated densities by finite element method with $m = 5$ *and* $d = 4$.

Example 2 In this example, we consider an ergodic Markovian process whose stationary density is the standard normal one given by (10). The sample data for the process can be generated from a Markov Chain Monte Carlo (MCMC) method. In our simulation, we employed the following independent Metropolis-Hastings algorithm (see, for example, [6]). Concretely, for the given standard normal sample data $\{x_n\}$ in Example 1, we used the following procedure to generate Markovian sample data $\{z_n\}$.

1) Given x_n, generate a random number y_n from $u(y)$ which is the uniform density function over $[0, 1]$.

2) Take

$$z_{n+1} = \begin{cases} y_n & \text{with probability } \min\left\{\frac{p(y_n)u(x_n)}{p(x_n)u(y_n)}, 1\right\}, \\ x_n & \text{otherwise.} \end{cases}$$

Example 3 In this example, the following exponential density is considered,

$$p(x) = \begin{cases} e^{-x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Example 4 In this example, we deal with the following density with non-smooth corners and discontinuous jump points,

$$p(x) = \begin{cases} \frac{1}{5}(-4-2x) & \text{if } -3 \leq x \leq -2, \\ \frac{1}{5}(2+2x) & \text{if } -1 \leq x \leq 0, \\ \frac{1}{5}(2-2x) & \text{if } 0 \leq x \leq 1, \\ \frac{2}{5}e^{-(x-1)} & \text{if } x \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

4 Proof of Theorem 1

For any given function $f \in C_c(R)$, let K be the compact set

$$K \equiv \text{the closure of the set } \{x : f(x) \neq 0\}. \tag{11}$$

Thus, there exists a sufficiently large positive integer M such that $K \subset [-M, M]$. Moreover, for a given sample sequence $\{X_1, X_2, \cdots\}$ satisfying (1) and a given $\epsilon > 0$, there exists a sufficiently large N such that

$$\left|\frac{\phi_i(X_1) + \phi_i(X_2) + \cdots + \phi_i(X_n)}{n} - E_\pi \phi_i(X)\right| < \epsilon \tag{12}$$

for all $n \geq N$ and all $i \in \{-m^3, -(m^3-1), \cdots, -1, 0, 1, \cdots, (m^3-1), m^3\}$. It follows from (3), (5)-(7) and (12) that

$$\begin{aligned}
&\int_{-\infty}^{\infty} f(x)p_{m,n,d}(x)\mathrm{d}x \qquad (13)\\
= &\sum_{i=-m^3}^{m^3} \hat{c}_i^d \int_{-M\vee b_m}^{M\wedge b_m} f(x)\phi_i(x)\mathrm{d}x\\
= &\sum_{i=-m^2(M\vee b_m)}^{m^2(M\wedge b_m)} \sum_{j_{d-1}=i-\beta_i}^{i+\beta_i} \alpha_{j_{d-1}} \cdots \sum_{j_1=j_2-\beta_{j_2}}^{j_2+\beta_{j_2}} \alpha_{j_1}\hat{c}_{j_1} \int_{-M\vee b_m}^{M\wedge b_m} f(x)\phi_i(x)\mathrm{d}x\\
= &\sum_{i=-m^2(M\vee b_m)}^{m^2(M\wedge b_m)} \sum_{j_{d-1}=i-\beta_i}^{i+\beta_i} \alpha_{j_{d-1}} \cdots \sum_{j_1=j_2-\beta_{j_2}}^{j_2+\beta_{j_2}} \Big(\frac{\alpha_{j_1}}{h} E\phi_{j_1}(X)+\\
&\frac{1}{h^2}O_n(\epsilon) + \frac{1}{h^2}o_n(\epsilon)\Big) \int_{-M\vee b_m}^{M\wedge b_m} f(x)\phi_i(x)\mathrm{d}x,
\end{aligned}$$

where for each $n > N$, $O_n(\epsilon)$ denotes the quantity of the same order as the value ϵ, i.e., $O_n(\epsilon) = c_n\epsilon$ for some constant c_n independent of ϵ, and $o_n(\epsilon)$ is infinitesimal in terms of ϵ, and in (13), we used the following convention that, if $d = 1$,

$$\sum_{j_{d-1}=i-\beta_i}^{i+\beta_i} \alpha_j \cdots \sum_{j_1=j_2-\beta_{j_2}}^{j_1+\beta_{j_2}} \alpha_{j_1}\hat{c}_{j_1} = \hat{c}_i.$$

In (13), let $\epsilon \to 0$, we know that the following claim holds for each given d and m,

$$-\infty < I(f) \equiv \lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)p_{m,n,d}(x)\mathrm{d}x < \infty. \qquad (14)$$

Since $0 \leq \phi_i(x) \leq 1$, we have that $I(f) \geq 0$ if $f \geq 0$. Moreover, one can easily show that $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ for all α, β and all functions $f, g \in C_c(S)$, that is, $I(\cdot)$ is a positive linear functional on $C_c(S)$. Notice that the space $S = R = (-\infty, \infty)$ is a locally compact Hausdorff space, then by Riesz-Markov representation theorem (see, for example, [8]), there exists a Borel measure μ_m for each m on the Borel measurable space $(S, \mathcal{B}(S))$ with $\mathcal{B}(S)$ denoting the Borel σ-field of S, such that,

$$I(f) = \int_S f\mathrm{d}\mu_m, \qquad (15)$$

for each $f \in C_c(S)$. Furthermore, since S is both locally compact and σ-compact separable metric space, every Borel set of S is a Baire set and the converse is also true by the argument on page 332 in [8], and thus, all Borel measures on S are regular (and hence inner regular) by Corollary 12 in [8]. Therefore, the Borel measure μ_m appeared in (15) is unique by the Riese-Markov representation theorem again. Hence, for each m, we have

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)p_{m,n,d}(x)\mathrm{d}x = I(f) = \int_S f\mathrm{d}\mu_m \qquad (16)$$

for each $f \in C_c(S)$. Hence we complete the proof of the theorem.

References:

[1] Simonoff J S. Smoothing Methods in Statistics[M]. Springer Series in Statistics, Spronger-Verlag, 1996
[2] Dai W. Brownian approximations for queueing networks with finite buffers: modeling, heavy traffic analysis and numerical implementations[D]. Ph.D. Thesis, School of Mathematics, Georgia Institute of Technology, 1996
[3] Shen X, Chen H, Dai J G, et al. The finite element method for computing the stationary distribution of an SRBM in a hypercube with applications to finite buffer queueing networks[J]. Queueing Systems, 2002,42:33-62
[4] Donoho D L, Johnstone I M, Kerkyacharian G, et al. Density estimation by wavelet thresholding[J]. Annals of Statistics, 1995,24:508-539
[5] Kallenberg O. Foundations of Modern Probability[M]. Berlin: Springer, 1997
[6] Robert C P, Casella G. Monte Carlo Statistical Methods[M]. Berlin: Springer, 1999
[7] Hall C A, Porsching T A. Numerical Analysis of Partial Differential Equations[M]. New Jersey: Prentice Hall, Englewood Cliffs, 1990
[8] Royden H L. Real analysis[M]. Third Edition, New York: Macmillan Publishing Company, 1988

具有多级自适应滤波器的有限元密度估计器

戴万阳

(南京大学数学系，南京 210093)

摘 要: 基于可能相关并具有某种遍历性的样本数据，该文设计了一个具有多级自适应滤波器的有限元估计器去估计相应的（平稳）密度函数。数值例子表明了估计器的有效性。同时也讨论了一个有关估计器稳定性的性质。

关键词: 密度估计；有限元插值；自适应滤波；弱收敛