Statistica Sinica **9**(1999), 137-151

DENSITY ESTIMATION BY WAVELET-BASED REPRODUCING KERNELS

Su-Yun Huang

Academia Sinica

Abstract: Density estimation by wavelet-based reproducing kernels is studied. Asymptotic bias and variance are derived. Estimators using spline-wavelets and Daubechies wavelets are presented as examples. Kernel order and kernel efficiency are also discussed.

By an integral property of the bias and an idea from Scott's averaged shifted histograms, a bias reduction technique based on a grid point average is proposed. This bias reduction technique is shown to be variance stable.

Key words and phrases: Asymptotics, Bernoulli numbers, Bernoulli polynomials, density estimation, efficiency, multiresolution approximation, projection kernel, reproducing kernel, reproducing kernel Hilbert space, wavelets.

1. Introduction

This article studies the density estimation problem based on i.i.d. observations $X_1, \ldots, X_n$ from a distribution with p.d.f. $f(x)$. We consider estimators of the form

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K(x, X_i), \tag{1.1}$$

where $K(x, y)$ is positive definite and satisfies the condition $K(x, y) = K(y, x)$. Such a positive definite kernel arises naturally from a reproducing kernel Hilbert space (RKHS in short). (See Wahba (1990) for the applications of reproducing kernels in spline models and see also Saitoh (1989) for the general theory of RKHS.)

Our main interests are restricted to projection kernels derived from an L_2 multiresolution approximation. Such kernels are reproducing kernels (Meyer (1990, 1992)). The estimator (1.1) is the so called linear wavelet estimator. There is a natural duality between RKHS's and stochastic processes (Parzen (1961)). Via this duality, Bayesian interpretation and inferences (for a regression estimator) can be established (Kimeldorf and Wahba (1970); Wahba (1978, 1990)). In this article, our focus is on the study of asymptotic behavior and bias reduction. The Bayesian aspects of a wavelet estimator is studied elsewhere.

We now introduce some notation. Let the nested sequence of closed subspaces $\cdots V_{j-1} \subset V_j \subset V_{j+1} \cdots, j \in Z$, be a multiresolution approximation to $L_2(R)$. A function $\phi(x)$ is called a wavelet scaling function (or a father wavelet) if $\int_{-\infty}^{\infty} \phi(x)dx = 1$ and $\{\phi(x-k)\}_{k\in Z}$ forms an orthonormal basis for V_0. For every $j \in Z$, define W_j to be the orthogonal complement of V_j in V_{j+1}. A function $\psi(x)$ is called a (mother) wavelet if $\{\psi(x-k)\}_{k\in Z}$ forms an orthonormal basis for W_0. (See Daubechies (1992), Mallat (1989) and Meyer (1990, 1992) for some general theory on wavelets and multiresolution analysis.)

The projection of a function $f(x)$ in $L_2(R)$ onto the space V_j is given by $\mathcal{P}_h f(x) = \int_{-\infty}^{\infty} K_h(x,y)f(y)dy$, where

$$K_h(x,y) = \frac{1}{h} K\left(\frac{x}{h}, \frac{y}{h}\right) \tag{1.2}$$

with $K(x,y) = \sum_{k\in Z} \phi(x-k)\phi(y-k)$ and $h = 2^{-j},\ j \in Z$. A density estimator based on projection kernel

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x, X_i) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x}{h}, \frac{X_i}{h}) \tag{1.3}$$

was proposed by Kerkyacharian and Picard (1992), Huang (1990), Huang and Studden (1993a). The effective bandwidth was $h \in R^+$ in Huang (1990) and Huang and Studden (1993a) and $h = 2^{-j}, j \in Z$ in Kerkyacharian and Picard (1992).

The rest of the article is organized as follows. In Section 2, we give some preliminaries. Sections 3 and 4 are on asymptotic bias and variance. In Section 5, a bias reduction method based on grid point averaging is proposed and shown to be variance stable. Some efficiency discussion is in Section 6.

2. Preliminaries

Some definitions and properties, needed later for the study of local asymptotics, are given below.

Definition 2.1. A kernel $K(x,y)$ is said to be of order m if and only if it satisfies the moment conditions:

$$\int_{-\infty}^{\infty} K(x,y)\ y^{\ell} dy = \begin{cases} 1, & \ell = 0, \\ x^{\ell}, & \ell = 1, \ldots, m-1, \\ \alpha(x) \neq x^m, & \ell = m. \end{cases}$$

(This definition is compatible with that of a convolution kernel.)

We say that the wavelet $\psi(x)$ has vanishing moments of order m if and only if the following moment conditions are met.

$$\int_{-\infty}^{\infty} \psi(y) y^{\ell} dy = 0,\ \ell = 0, 1, \ldots, m-1; \ \text{ and } \int_{-\infty}^{\infty} \psi(y) y^m dy \neq 0.$$

Proposition 2.2. *Given a multiresolution approximation, a wavelet $\psi(x)$ has vanishing moments of order m if and only if the associated projection kernel $K(x, y)$ in* (1.2) *is of order m.*

Proof. Every polynomial of order $m + 1$ (i.e. degree m or less) has a unique representation as

$$p(x) = \sum_{k \in Z} \langle p, \phi(\cdot - k) \rangle \phi(x - k) + \sum_{j=0}^{\infty} \sum_{k \in Z} b_{j,k} \sqrt{2^j} \psi(2^j x - k),$$

where $b_{j,k} = \langle p, \sqrt{2^j} \psi(2^j \cdot -k) \rangle$. Suppose that $\psi(x)$ has vanishing moments of order m. Then, for $\ell = 0, \dots, m - 1$,

$$x^\ell = \sum_{k \in Z} \langle y^\ell, \phi(y - k) \rangle \phi(x - k) = \int_{-\infty}^{\infty} K(x, y) y^\ell dy. \tag{2.1}$$

However, for $\ell = m$, some $b_{j,k}$'s are non-zero and hence the representation (2.1) does not hold.

Now suppose that $K(x, y)$ is of order m. For $\ell = 0, \dots, m-1$, the 'projection' (integral transform) of x^ℓ onto W_0 is given by

$$(\mathcal{P}_{1/2} - \mathcal{P}_1) x^\ell = \int_{-\infty}^{\infty} \Big[2K(2x, 2y) - K(x, y) \Big] y^\ell dy = x^\ell - x^\ell = 0$$

by the reproducing property of $K(x, y)$. Therefore $\int_{-\infty}^{\infty} \psi(y) y^\ell dy = 0$.

The wavelet subspace V_j is a RKHS with the unique reproducing kernel $K_{2^{-j}}(x, y)$. Proposition 2.2 says that the order of vanishing moments or equivalently the order of reproducing ability is intrinsic for a multiresolution approximation and is independent of the choices of wavelet bases.

Definition 2.3. We say that a sequence of multiresolution approximation, $\dots, V_{j-1}, V_j, V_{j+1}, \dots$, is symmetric if and only if the projection kernel onto V_0 satisfies the condition $K(-x, y) = K(x, -y)$.

Note that if a multiresolution approximation sequence is symmetric under Definition 2.3, then the projection operator, onto any V_j, maps an even function to an even function. The usual definition for symmetry is based on whether the scaling function $\phi(x)$ is symmetric or not. Definition 2.3 is a bit more general. It is defined on a multiresolution approximation and is independent of choices of wavelet bases.

Also note that $K(-x, y)$ is the time-reversed kernel corresponding to the time-reversed data in a regression setting. (See Antoniadis, Gregoire and McKeague (1994) for more discussion on symmetrization.)

3. Asymptotic Bias

In this and the next sections, the asymptotic bias and variance of the estimator (1.3) are studied. Spline-wavelets and Daubechies wavelets are presented as examples. Define $b_m(x) = x^m - \int_{-\infty}^{\infty} K(x,y)y^m dy$. It is easy to establish the following results.

Proposition 3.1. *Suppose that $K(x,y)$ is of order m.*
(a) *Then the function $b_m(x)$ is periodic with period one.*
(b) *For $h > 0$, we have $x^m - \int_{-\infty}^{\infty} K_h(x,y)y^m dy = h^m b_m\left(\frac{x}{h}\right)$.*
(c) *If the approximation is symmetric, then $b_m(-x) = (-1)^m b_m(x)$.*

When m is even, $b_m(x)$ is symmetric about zero. Since $b_m(x)$ is periodic with period one, $b_m(x)$ is also symmetric about all the points $x = k/2, k \in Z$. When m is odd $b_m(x)$ is anti-symmetric about zero and hence anti-symmetric about all the points $x = k/2, k \in Z$. A plot of $b_m(x)$ can provide a visual idea of the degree of symmetry of the multiresolution approximation. (See Examples 1 and 2 as well as Figures 3.1 and 3.2 below for spline wavelets and Daubechies wavelets.)

Theorem 3.2. *Assume that $f(x)$ belongs to the function space below:*

$$Lip^{m,\alpha}(R) = \{f \in C^m(R) : |f^{(m)}(x) - f^{(m)}(y)| \le A|x-y|^\alpha,\ x,y \in R\},$$

for some $\alpha > 0$ and $A > 0$. Assume that the kernel has the following localization property: $\int_{-\infty}^{\infty} |K(x,y)(y-x)^{m+\alpha}|\, dy \le C,$ for some $C > 0$. Also assume that $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Then for a fixed x, we have

$$E\hat{f}(x) - f(x) = \frac{-1}{m!} f^{(m)}(x) b_m\left(\frac{x}{h}\right) h^m + O(h^{m+\alpha}). \tag{3.1}$$

Moreover, if $f^{(m)}$ is in $L_2(R)$, then the integrated squared bias is

$$||E\hat{f} - f||_2^2 = \frac{b_{2m}}{(2m)!} ||f^{(m)}||_2^2 h^{2m} + O(h^{2(m+\alpha)}),$$

where $b_{2m} = (2m)!(m!)^{-2} \int_0^1 b_m^2(x)dx$.

Proof. We first observe that

$$\mathrm{E}\hat{f}(x) - f(x) = \int_{-\infty}^{\infty} K_h(x,y)(f(y) - f(x))dy$$

$$= \int_{-\infty}^{\infty} K_h(x,y)\Big(\sum_{i=1}^{m-1} \frac{f^{(i)}(x)}{i!}(y-x)^i + \frac{f^{(m)}(\xi_{x,y})}{m!}(y-x)^m\Big)dy$$

$$= \int_{-\infty}^{\infty} K_h(x,y) \frac{f^{(m)}(\xi_{x,y})}{m!}(y-x)^m dy$$

$$= \frac{-1}{m!} f^{(m)}(x) b_m\left(\frac{x}{h}\right) h^m + \int_{-\infty}^{\infty} K_h(x,y)\Big(\frac{f^{(m)}(\xi_{x,y}) - f^{(m)}(x)}{m!}\Big)(y-x)^m dy,$$

where $\xi_{x,y}$ is some number lying between x and y. It will be shown below that the second term in the last equality is $O(h^{m+\alpha})$. By the localization assumption of $K(x,y)$, we have

$$\begin{aligned}&\left|\int_{-\infty}^{\infty} K_h(x,y)\Big(\frac{f^{(m)}(\xi_{x,y})-f^{(m)}(x)}{m!}\Big)(y-x)^m dy\right|\\ &\le \frac{A}{m!}\int_{-\infty}^{\infty}|K_h(x,y)(y-x)^{m+\alpha}|\,dy\\ &= \frac{Ah^{m+\alpha}}{m!}\int_{-\infty}^{\infty}|K(x/h,y)(y-x/h)^{m+\alpha}|\,dy = O(h^{m+\alpha}).\end{aligned}$$

The second assertion concerning the integrated squared bias follows easily from the first assertion and the square integrability of $f^{(m)}$.

When $\alpha = 0$, expression (3.1) in Theorem 3.2 can be improved upon with the remainder term of order $o(h^m)$. Assume that $f^{(m)}$ is bounded and continuous. Improvement can be done by bounding

$$\begin{aligned}&h^{-m}\int_{-\infty}^{\infty}\Big(f^{(m)}(\xi_{x,y})-f^{(m)}(x)\Big)\,|K_h(x,y)(y-x)^m|dy\\ &=\int_{-\infty}^{\infty}\Big(f^{(m)}(\xi'_y\ h)-f^{(m)}(x)\Big)\,|K(x',y)(y-x')^m|dy\\ &=\int_{-\infty}^{\infty}\Big(f^{(m)}(\xi'_{y+[x']}\ h)-f^{(m)}(x)\Big)\,|K(w,y)(y-w)^m|dy, \qquad (3.2)\end{aligned}$$

where $x' = x/h$, ξ'_z denotes some number between x' and z, $[\ \cdot\ ]$ denotes the greates integer function and $0 \le w = x' - [x'] < 1$. The measure induced by $\{\sup_{0\le w<1}|K(w,y)(y-w)^m|\}dy$ on the real line is finite. Also note that $\xi'_{y+[x']}\ h \to x$, as $h \to 0$. By the Bounded Convergence Theorem, (3.2) is $o(1)$ as $h \to 0$.

Note that, under the multiresolution approximation framework, $h = 2^{-j}, j \in Z$. However, under the projection kernel (or reproducing kernel) framework, h can also be generalized and treated as a continuous smoothing parameter. The nice idea that h can be taken as a continuous smoothing parameter can also be found in Huang (1990), Huang and Studden (1993a, 1993b), and Hall and Patil (1995a).

Below we present two examples: spline-wavelets and Daubechies wavelets. Spline-wavelets have exponential localization (i.e., $K(x,y) \to 0$ exponentially fast, as $|x-y| \to \infty$) and Daubechies wavelets have compact support. Therefore the localization property in Theorem 3.2 is met for both cases.

Example 1. (Battle-Lemarié spline-wavelets.) Consider density estimator using spline-wavelets. From results obtained in Huang and Studden (1993a), we have

$$b_m(x) = B_m(x) \text{ for } x \in (0,1) \text{ and } b_{2m} = |B_{2m}|,$$

where $B_m(x)$ is the mth Bernoulli polynomial and B_{2m} is the $2m$th Bernoulli number. Plots of Bernoulli polynomials are presented in Figure 3.1.

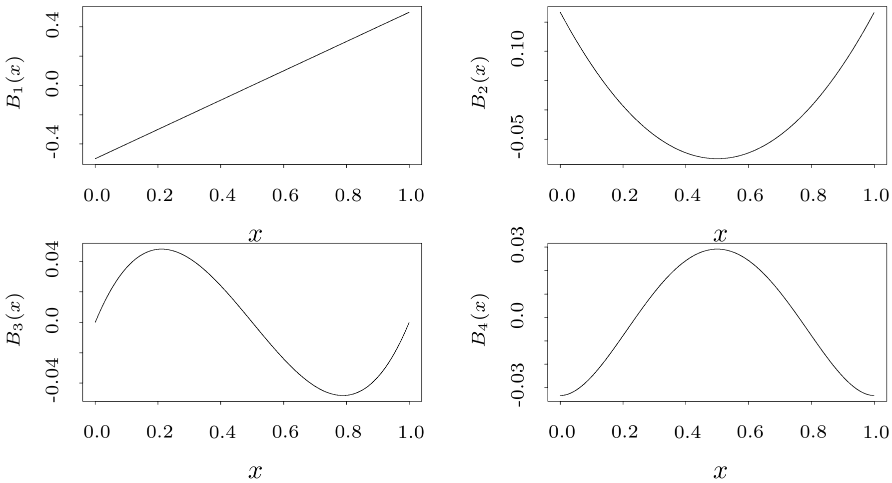

Figure 3.1. Bias plots using spline-wavelets

Example 2. (Wavelets by Daubechies (1988).)

Let ${}_N\phi(x)$ denote the Daubechies' scaling function supported on $[0, 2N-1]$. The multiresolution approximation associated with Daubechies' ${}_N\phi$ system has order $m = N$. An expression for $b_N(x)$ is given below.

$$b_N(x) = x^N - \sum_{\ell=0}^{N} C(N,\ell) a_N^\ell \Phi_N^{N-\ell}(x),$$

where $C(N,\ell) = N!/\ell!(N-\ell)!$ and

$$a_N^\ell = \int_0^{2N-1} {}_N\phi(x) x^\ell dx, \quad \Phi_N^{N-\ell}(x) = \sum_{k=-2N+2}^{0} k^{N-\ell}\, {}_N\phi(x-k).$$

Daubechies' wavelet for $N = 1$ corresponds to the Haar wavelet. The associated $b_1(x)$ is simply the first Bernoulli polynomial $B_1(x)$. Therefore it is omitted from Figure 3.2. Plots of $b_m(x)$ are presented in Figure 3.2 for $m = 2, 3, 4$ and 5. Values of b_{2m} are listed in Table 6.1, Section 6. The value of b_{2m}, small or large, plays an influential role in comparing the relative efficiency of wavelet-based reproducing kernels. The numbers b_{2m} in this example are much larger than their competitive Bernoulli numbers.

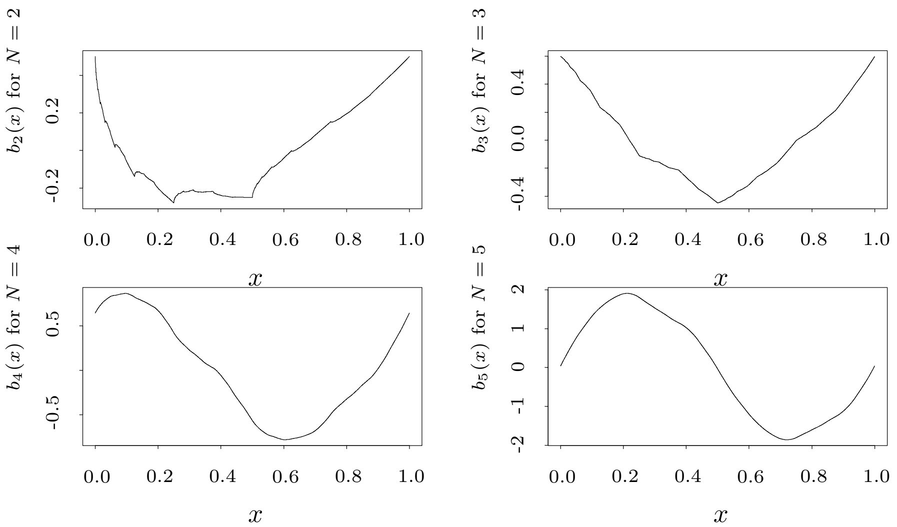

Figure 3.2. Bias plots using Daubechies wavelets

4. Asymptotic Variance

In this section, we derive the expression of aymptotic variance for estimator (1.3). Antoniadis, Gregoire and McKeague (1994) also have studied the same problem and have given the order of magnitude at dyadic points.

Theorem 4.1. *Suppose that $f(x) \in C^1(R)$ and that $f(x)$ and $f'(x)$ are uniformly bounded. For a fixed x, we have*

$$\operatorname{Var} \hat{f}(x) = \frac{1}{nh} f(x) V(\frac{x}{h}) + O(\frac{1}{n}),$$

where $V(x) = \int_{-\infty}^{\infty} K^2(x,y)dy = K(x,x)$. Moreover the integrated variance is given by $\int_{-\infty}^{\infty} \operatorname{Var} \hat{f}(x)dx = v/(nh) + O(n^{-1})$, whre $v = \int_0^1 V(x)dx$.

Proof. Start with

$$\begin{aligned}
\operatorname{Var} \hat{f}(x) &= \frac{1}{n}\int_{-\infty}^{\infty} K_h^2(x,y)f(y)dy - \frac{1}{n}\Big(\int_{-\infty}^{\infty} K_h(x,y)f(y)dy\Big)^2 \\
&= \frac{1}{n} f(x)\int_{-\infty}^{\infty} K_h^2(x,y)dy + \frac{1}{n}\int_{-\infty}^{\infty} K_h^2(x,y)\left(f(y)-f(x)\right)dy \\
&\quad -\frac{1}{n}\Big(\int_{-\infty}^{\infty} K_h(x,y)f(y)dy\Big)^2 \\
&= \frac{1}{nh} f(x) V(\frac{x}{h}) + \frac{1}{n}\int_{-\infty}^{\infty} K_h^2(x,y)(f(y)-f(x))dy \\
&\quad -\frac{1}{n}\Big(\int_{-\infty}^{\infty} K_h(x,y)f(y)dy\Big)^2.
\end{aligned}$$

Below we show that the second and the third terms in the last equality are of order $O(n^{-1})$.

$$\left|\frac{1}{n}\int_{-\infty}^{\infty} K_h^2(x,y)\,(f(y)-f(x))\,dy\right| \le \frac{1}{n}\sup_{x\in R}|f'(x)|\frac{1}{h^2}\int_{-\infty}^{\infty} K^2\left(\frac{x}{h},\frac{y}{h}\right)|y-x|dy$$
$$\le \frac{1}{n}\sup_{x\in R}|f'(x)| \sup_{s,t\in R}|K(s,t)|\int_{-\infty}^{\infty}\left|K\left(\frac{x}{h},t\right)\left(t-\frac{x}{h}\right)\right|dt = O(\frac{1}{n}).$$

By the uniform boundedness of $f(x)$, it is easy to see that

$$\frac{1}{n}\left(\int_{-\infty}^{\infty} K_h(x,y)f(y)dy\right)^2 = O(\frac{1}{n}).$$

Therefore, we have the first assertion of the theorem.

The second assertion can be easily obtained by bounding the following two integrals. One is $n^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_h^2(x,y)\,(f(y)-f(x))\,dy\ dx = 0$. The other is $n^{-1}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K_h(x,y)f(y)dy\right)^2 dx = n^{-1}||\mathcal{P}_h f||_2^2 = O(n^{-1})$.

Note that, for any multiresolution approximation, we have

$$v \equiv \int_0^1 K(x,x)dx = \int_0^1 \sum_{k=-\infty}^{\infty}\phi(x-k)\phi(x-k)dx = 1.$$

That is, the value of v is always independent of choices of systems of multiresolution approximation. Also note that, when a multiresolution approximation is symmetric, the function $V(x)$ is symmetric about $x = k/2, k \in Z$.

Example 1. (continued) Plots of $V(x)$ are presented in Figure 4.1. Note that they are symmetric.

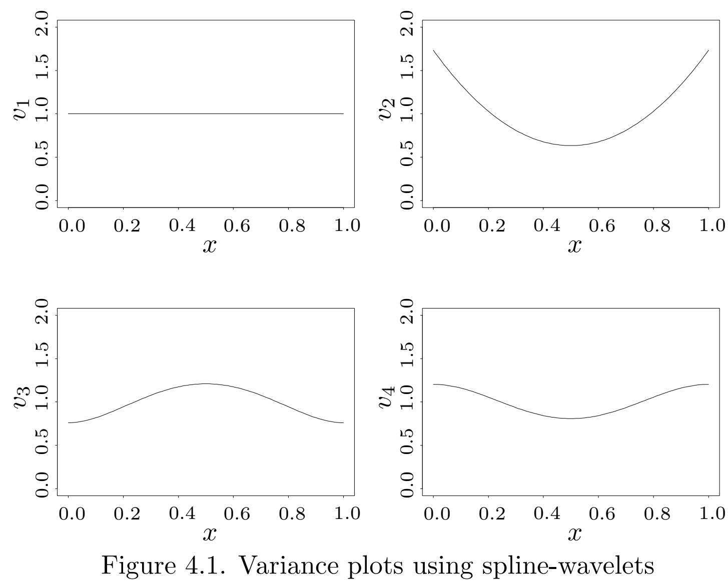

Figure 4.1. Variance plots using spline-wavelets

Example 2. (continued) Plots of $V(x)$ are presented in Figure 4.2. Since the case $N = 1$ coincides with the spline of order 1, it is omitted from Figure 4.2. These $V(x)$ are not symmetric. However, the approximation induced by the system ${}_5\phi$ is quite close to symmetry. Also note that the variance of spline-wavelet estimator is more stable than that of Daubechies wavelets.

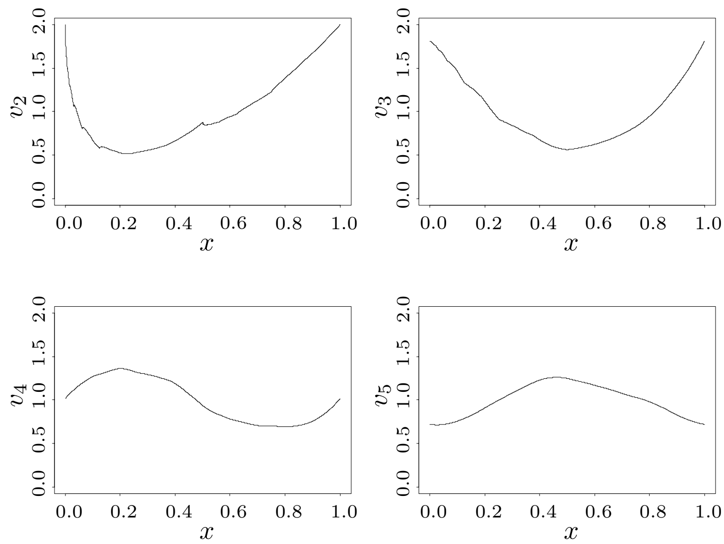

Figure 4.2. Variance plots using Daubechies wavelets

5. Averaged Shifted Kernel Estimator: A Bias Reduction Method

In view of Theorems 3.2 and 4.1 and Examples 1 and 2, the pointwise asymptotic bias and variance show an intriguing dependence of the grid point positions (i.e. the knot positions in the language of spline theory, or the dyadic points in the language of multiresolution analysis). This oscillatory effect can also be found in Huang and Studden (1993a) and Hall and Patil (1995b). To remove or at least to lessen the dependence effect of grid point positions, we propose an averaged shifted kernel method. The shift-and-average technique can also be found in Scott (1985) on density estimation by histogram, Huang (1992) on density estimation, Coifman and Donoho (1995) on nonparametric regression and Nason and Silverman (1995) on stationary wavelet transform. The effect of bias reduction by Coifman and Donoho (1995) is mainly based on simulation. So far there are not many theoretical results on such procedure either in regression or density estimation. In Proposition 5.1, we present a key observation which allows us to remove the grid point dependence and provide bias reduction as well. We also show that the shift-and-average procedure is variance stable.

Proposition 5.1. *We have* $\int_0^1 b_m(x)dx = 0$.

Proof. Since $K_0(x,y)$ reproduces the polynomial $p(x) = 1$, we have $1 = \int_{-\infty}^{\infty} K_0(x,y)dy = \sum_{k\in Z}\phi(x-k)$. Then

$$\begin{aligned}\int_0^1 b_m(x)dx &= \int_0^1 b_m(x)\sum_{k\in Z}\phi(x-k)dx = \sum_{k\in Z}\int_0^1 b_m(x)\phi(x-k)dx\\ &= \sum_{k\in Z}\int_{-k}^{-k+1} b_m(x)\phi(x)dx = \int_{-\infty}^{\infty} b_m(x)\phi(x)dx\\ &= \int_{-\infty}^{\infty}\phi(x)\Big(x^m - \int_{-\infty}^{\infty} K(x,y)y^m dy\Big)dx\\ &= \int_{-\infty}^{\infty} x^m\phi(x)dx - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(x)K(x,y)y^m dxdy\\ &= \int_{-\infty}^{\infty} x^m\phi(x)dx - \int_{-\infty}^{\infty} y^m\phi(y)dy = 0.\end{aligned}$$

Note that $b_m(x)$ appears as a key factor in the first order term of a bias in equation (3.1). Consider an averaged shifted kernel

$$K_h^{(q)}(x,y) = \frac{1}{q}\sum_{\ell=0}^{q-1} K_h\Big(x+\frac{\ell h}{q}, y+\frac{\ell h}{q}\Big). \tag{5.1}$$

A density estimator based on the averaged shifted kernel

$$\hat{f}_{ASKE,q}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h^{(q)}(x, X_i)$$

is proposed. The idea of shifting grid points and then taking the average was originated by Scott's (1985) averaged shifted histograms. As $q\to\infty$, the wavelet estimator is an estimator based on the kernel

$$K_h^{\infty}(x,y) = \int_0^1 K_h(x+th, y+th)dt.$$

Proposition 5.2. *For each shifted kernel $K_h(x+\ell h/q, y+\ell h/q)$, the reproducing ability is of order m. Therefore, $K_h^{(q)}(x,y)$ is still of order m.*

Proposition 5.3. *We have $x^m - \int_{-\infty}^{\infty} K(x+\frac{\ell}{q}, y+\frac{\ell}{q})y^m dy = b_m(x+\frac{\ell}{q})$.*

Proofs for Proposition 5.2 and Proposition 5.3 are straightforward and are omitted here.

Theorem 5.4. *Assume conditions in Theorem 3.2. We have, for a fixed x,*

$$E\,\hat{f}_{ASKE,q}(x) - f(x) = \frac{-f^{(m)}(x)}{m!}\,\frac{1}{q}\Big(\sum_{\ell=0}^{q-1} b_m\Big(\frac{x}{h}+\frac{\ell}{q}\Big)\Big)h^m + O(h^{m+\alpha}),$$

where $O(h^{m+\alpha})$ is uniform in q.

Proof.

$$\mathrm{E}\ \hat{f}_{ASKE,q}(x) - f(x) = \int_{-\infty}^{\infty} K_h^{(q)}(x,y)\,(f(y)-f(x))\,dy$$

$$= \frac{1}{q}\sum_{\ell=0}^{q-1}\int_{-\infty}^{\infty} K_h\Big(x+\frac{\ell h}{q}, y+\frac{\ell h}{q}\Big)(f(y)-f(x))dy$$

$$= \frac{1}{q}\sum_{\ell=0}^{q-1}\int_{-\infty}^{\infty} K_h\Big(x+\frac{\ell h}{q}, y+\frac{\ell h}{q}\Big)\frac{f^{(m)}(\xi_{x,y})}{m!}(y-x)^m dy$$

$$= \frac{-f^{(m)}(x)}{m!}\frac{1}{q}\Big(\sum_{\ell=0}^{q-1} b_m\Big(\frac{x}{h}+\frac{\ell}{q}\Big)\Big)h^m$$

$$+\frac{1}{q}\sum_{\ell=0}^{q-1}\int_{-\infty}^{\infty} K_h\Big(x+\frac{\ell h}{q}, y+\frac{\ell h}{q}\Big)\frac{f^{(m)}(\xi_{x,y})-f^{(m)}(x)}{m!}(y-x)^m dy,$$

where $\xi_{x,y}$ is some number between x and y.

For each $\ell = 0, \dots, q-1$,

$$\Big|\int_{-\infty}^{\infty} K_h\Big(x+\frac{\ell h}{q}, y+\frac{\ell h}{q}\Big)\frac{f^{(m)}(\xi_{x,y})-f^{(m)}(x)}{m!}(y-x)^m dy\Big|$$

$$\le \frac{Ah^{m+\alpha}}{m!}\int_{-\infty}^{\infty}\Big|\frac{1}{h}K\left(\frac{x}{h}+\frac{\ell}{q}, \frac{y}{h}+\frac{\ell}{q}\Big)\Big(\frac{y-x}{h}\right)^{m+\alpha} dy\Big| \le \frac{ACh^{m+\alpha}}{m!},$$

where A and C are constants defined in Theorem 3.4. Therefore

$$\frac{1}{q}\sum_{\ell=0}^{q-1}\int_{-\infty}^{\infty} K_h\Big(x+\frac{\ell h}{q}, y+\frac{\ell h}{q}\Big)\frac{f^{(m)}(\xi_{x,y})-f^{(m)}(x)}{m!}(y-x)^m dy = O(h^{m+\alpha})$$

uniformly in q.

Theorem 5.5. *Assume conditions in Theorem* 4.1. *We have, for a fixed x,*

$$\mathrm{Var}\,\hat{f}_{ASKE,q}(x) \le \frac{1}{nh}f(x)K_0^{(q)}\Big(\frac{x}{h},\frac{x}{h}\Big)+O\big(\frac{1}{n}\big),$$

where $O(n^{-1})$ is uniform in q.

Proof. Begin with

$$\mathrm{Var}\ \hat{f}_{ASKE,q}(x) = \frac{1}{n}\mathrm{Var}\ K_h^{(q)}(x,X)$$

$$= \frac{1}{n}\int_{-\infty}^{\infty} K_h^{(q)}(x,y)^2 f(y)dy - \frac{1}{n}\Big(\int_{-\infty}^{\infty} K_h^{(q)}(x,y)f(y)dy\Big)^2. \qquad (5.2)$$

Below we deal with the first term in (5.2) and then show that the second term in (5.2) is of order $O(n^{-1})$ uniformly in q. The first term is

$$\begin{aligned}
&\frac{1}{n}\int_{-\infty}^{\infty} K_h^{(q)}(x,y)^2 f(y)dy\\
&=\frac{1}{n}\int_{-\infty}^{\infty}\frac{1}{q^2}\Big(\sum_{\ell=0}^{q-1}K_h(x+\frac{\ell h}{q},y+\frac{\ell h}{q})\Big)^2 f(y)dy\\
&\le\frac{1}{nq}\int_{-\infty}^{\infty}\sum_{\ell=0}^{q-1}K_h^2\Big(x+\frac{\ell h}{q},y+\frac{\ell h}{q}\Big)f(y)dy\\
&=\frac{1}{nq}\sum_{\ell=0}^{q-1}\int_{-\infty}^{\infty}K_h^2\Big(x+\frac{\ell h}{q},y+\frac{\ell h}{q}\Big)f(x)dy\\
&\quad+\frac{1}{nq}\sum_{\ell=0}^{q-1}\int_{-\infty}^{\infty}K_h^2\Big(x+\frac{\ell h}{q},y+\frac{\ell h}{q}\Big)(f(y)-f(x))dy\\
&=\frac{1}{nh}f(x)K^{(q)}\Big(\frac{x}{h},\frac{x}{h}\Big)+\frac{1}{nq}\sum_{\ell=0}^{q-1}\int_{-\infty}^{\infty}K_h^2\Big(x+\frac{\ell h}{q},y+\frac{\ell h}{q}\Big)(f(y)-f(x))dy.
\end{aligned}$$

Note that

$$\begin{aligned}
&\Big|\frac{1}{nq}\sum_{\ell=0}^{q-1}\int_{-\infty}^{\infty}K_h^2\Big(x+\frac{\ell h}{q},y+\frac{\ell h}{q}\Big)\Big(f(y)-f(x)\Big)dy\Big|\\
&\le\frac{1}{nq}\sup_{x\in R}|f'(x)|\sum_{\ell=0}^{q-1}\int_{-\infty}^{\infty}K_h^2\Big(x+\frac{\ell h}{q},y+\frac{\ell h}{q}\Big)|y-x|dy\\
&\le\frac{1}{nq}\sup_{x\in R}|f'(x)|\sup_{s,t\in R}|K(s,t)|\sum_{\ell=0}^{q-1}\int_{-\infty}^{\infty}\Big|K\Big(\frac{x}{h}+\frac{\ell}{q},t\Big)\Big(t-\frac{x}{h}-\frac{\ell}{q}\Big)\Big|dt\\
&\le\frac{1}{n}\sup_{x\in R}|f'(x)|\sup_{s,t\in R}|K(s,t)|\sup_{s\in R}\int_{-\infty}^{\infty}\Big|K(s,t)(t-s)\Big|dt=O(\frac{1}{n})
\end{aligned}$$

independently of q. The second term in (5.2) is of order $O(n^{-1})$, since

$$\frac{1}{n}\Big(\int_{-\infty}^{\infty}K_h^{(q)}(x,y)f(y)dy\Big)^2\le\frac{1}{n}\Big(\sup_{x\in R}|f(x)|\sup_{x\in R}\int_{-\infty}^{\infty}\Big|K(x,y)\Big|dy\Big)^2=O(\frac{1}{n}).$$

Note that the number of shifts q can be chosen independent of n and h. By letting $q\to\infty$, we have

$$\frac{1}{q}\Big(\sum_{\ell=0}^{q-1}b_m\Big(\frac{x}{h}+\frac{\ell}{q}\Big)\Big)\to\int_0^1 b_m\Big(\frac{x}{h}+t\Big)dt=0,$$

and

$$K^{(q)}\Big(\frac{x}{h},\frac{x}{h}\Big) \to \int_0^1 K(x,x)dx = 1.$$

Actually the variance is getting more and more stable in the sense that

$$\sup_{x\in R}\Big|K^{(q)}(x,x)-1\Big| \to 0, \ \ \text{as } q\to\infty.$$

The action of shift-and-average is a smoothing operation on the grid points rather than on the data points. It is well known that when a smoothing operation is applied to the data points, there is a trade-off between bias and variance. However it is not the case when the smoothing operation is applied to the grid points.

6. Efficiency

In a wavelet method the bandwidth selection problem is often considered in a discrete manner. Bandwidths of the form $h = 2^{-j}$ are considered. However, from the kernel point of view, bandwidths can be chosen in a continuous manner. Any automatic bandwidth selector for a convolution kernel estimator can be tried out for estimator (1.3). Theoretical optimal bandwidth can be obtained by minimizing the asymptotic IMSE given below.

$$\text{IMSE} \simeq \frac{b_{2m}}{(2m)!}||f^{(m)}||_2^2 h^{2m} + \frac{v}{nh}. \tag{5.3}$$

The optimal bandwidth is

$$h_{\text{opt}} = \Big(||f^{(m)}||_2^2\Big)^{-1/(2m+1)}\Big(\frac{(2m-1)!v}{b_{2m}}\Big)^{1/(2m+1)} n^{-1/(2m+1)}. \tag{5.4}$$

The above formulae (5.3) and (5.4) are also valid for a convolution type kernel with $v = \int_{-\infty}^{\infty} k^2(t)dt$ and $b_{2m} = (2m)!(m!)^{-2}\Big(\int_{-\infty}^{\infty} K(t)t^m dt\Big)^2$. Plugging h_{opt} into (5.3), we have

$$\text{IMSE}_{\text{opt}} \simeq \frac{2m+1}{2m}\Big(\frac{b_{2m}}{(2m-1)!}\Big)^{1/2m+1}||f^{(m)}||_2^{2/2m+1}\Big(\frac{v}{n}\Big)^{2m/2m+1}. \tag{5.5}$$

Let $C_m(K) = b_{2m}^{1/2m+1}v^{2m/2m+1}$. We define the relative efficiency of $K^\star$ to K as

$$\text{rel eff} = \{C_m(K)/C_m(K^\star)\}^{(2m+1)/2m}. \tag{5.6}$$

The above definitions of $C_m(K)$ and relative efficiency are compatible with those for convolution kernels in Silverman (1986), Sections 3.3 and 3.6. The $C_m(K)$ here differs from that in Silverman only by a constant factor which will be cancelled off in the calculation of relative efficiency.

Below we compare the asymptotic relative effeciency of several kernels (indicated by $\star$) to the spline projection kernels.

Table 6.1. Asymptotic relative efficiency

$m = 2$	$b^{\star}_{2m}$	$v^{\star}$	efficiency
Epanechnikov kernel*	6/25	3/5	1.0175
Gaussian*	6	$1/(2\sqrt{\pi}\ )$	0.9678
minimum variance kernel*	2/3	1/2	0.9457
Daubechies*	.3000532	1	0.5773
$m = 3$			
Daubechies*	1.785819	1	0.4870
$m = 4$			
Daubechies*	21.61719	1	0.4452
optimal kernel*	10/63	5/4	0.6582
minimum variance kernel*	18/35	9/8	0.6314
$m = 5$			
Daubechies*	436.472568	1	0.4207
$m = 6$			
Daubechies*	13178.74404	1	0.4045
optimal kernel*	700/5577	1575/832	0.5601
minimum variance kernel*	100/231	225/128	0.5440

The efficiency discussion above is intended for comparison between kernels when the underlined function is smooth. Table 6.1 should be read with that in mind. The advantages of using wavelets mostly appear where functions are not everywhere smooth but only smooth in a global sense. (See Kerkyacharian and Picard (1992, 1993) for important and fundamental results on minimax optimality in Besov spaces and also on saturation spaces for the minimax optimality.) One interesting thing to note: the spline wavelets (Battle-Lemarié) have similar smoothing effect as classical convolution kernels meanwhile retaining the nice properties of wavelets.

References

Agarwal, G. G. and Studden, W. J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing splines. *Ann. Statist.* **8**, 1307–1325.

Antoniadis, A., Gregoire, G. and McKeague, I. W. (1994). Wavelet methods for curve estimation. *J. Amer. Statist. Assoc.* **89**, 1340-1353 .

Coifman, R. R. and Donoho, D. L. (1995). Translation-invariant denoising. Wavelets and Statistics. In *Lecture Notes in Statistics* **103** (Edited by Antoniadis and Oppenheim), 125-150. Springer-Verlag.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **41**, 909-996.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Academic Press .

Donoho, D. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425-455.

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. Smoothing Techniques for Curve Estimation In *Lecture notes in Mathematics* **757** (Edited by Gasser and Rosenblatt), 23-68. Springer-Verlag.

Hall, P. and Patil, P. (1995a). Formulae for mean integrated squared error of nonlinear wavelet-based estimators. *Ann. Statist.* **23**, 905-928.

Hall, P. and Patil, P. (1995b). On wavelet methods for estimating smooth functions. *Bernoulli* **1**, 41-58 .

Huang, S. Y. (1990). Nonparametric density estimation by spline projection kernels. Ph.D. Thesis, Department of Statistics, Purdue University.

Huang, S. Y. (1992). Averaged shifted spline projection kernel estimator: an effective bias reducing nonparametric density estimator. Unpublished manuscript, Department of Mathematics, Wayne State University. Abstract in *Inst. Math. Statist. Bulletin* **21**, 52, 1992.

Huang, S. Y. and Studden, W. J. (1993a). Density estimation using spline projection kernels. *Comm. Statist.-Theory Meth.* **22**, 3263-3285.

Huang, S. Y. and Studden, W. J. (1993b). An equivalent kernel method for least squares spline regression. *Statist. Decisions* supp. **3**, 179-201.

Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov spaces. *Statist. Probab. Letters.* **13**, 15-24.

Kerkyacharian, G. and Picard, D. (1993). Density estimation by kernel and wavelets methods: Optimality of Besov spaces. *Statist. Probab. Letters.* **18**, 327-336.

Kimeldorf, G. and Wahba, G. (1970). A correspondence between Bayesian estimation in stochastic processes and smoothing by splines. *Ann. Math. Statist.* **41**, 495-562.

Mallat, S. (1989). Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Trans. Amer. Math. Soc.* **315**, 69-87.

Meyer, Y. (1990). *Ondelettes et Opérateurs.* volume 1. Hermann, Paris.

Meyer, Y. (1992). *Wavelets and Operators.* Cambridge University Press.

Nason, G. P. and Silverman, B. W. (1995). The stationary wavelet transform and some statistical applications. Wavelets and Statistics. In *Lecture Notes in Statistics* **103** (Edited by Antoniadis and oppenheim), 281-300. Springer-Verlag.

Parzen, E. (1961). An approach to time series analysis. *Ann. Math. Statist.* **32**, 951-989.

Saitoh, S. (1989). *Theory for Reproducing Kernels and its Applications.* Wiley, New York.

Scott, D. (1985). Averaged shifted histograms: effective nonparametric density estimators in several dimensions. *Ann. Statist.* **13**, 1024-1040.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. B*, **40**, 364-372.

Wahba, G. (1990). *Spline Models for Observational Data.* Society for Industrial and Applied Mathematics.

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.

E-mail: syhuang@stat.sinica.edu.tw

(Received April 1996; accepted June 1997)