DENSITY ESTIMATION USING SPLINE PROJECTION KERNELS

by

Su-Yun Huang
Department of Mathematics
Wayne State University
Detroit, MI

and

W.J. Studden*
Department of Statistics
Purdue University
West Lafayette, IN

Technical Report # 90-52

Department of Statistics
Purdue University

September, 1990

*Research supported by NSF under Grant No. DMS–8802535.

△

DENSITY ESTIMATION USING SPLINE PROJECTION KERNELS

by

Su-Yun Huang and W.J. Studden

ABSTRACT

Density estimation is considered using a kernel derived from the L_2 projection onto certain spline spaces. The projection can be written as $f_p(x) = \int K(x,y)f(y)dy$. An estimate based on a sample $X_1, \ldots, X_n$ from f is given by $\hat{f}_p(x) = n^{-1}\Sigma_i K(x, X_i)$. The asymptotic bias and variance at a point are derived. It is found the quadratic spline projection kernel estimator has the same asymptotic bias as histosplines and has improved asymptotic variance.

AMS 1970 subject classification, 62G05

Key words and phrases: Density estimation, Spline projection kernel, histospline.

1. Introduction

The main purpose of this paper is to study the asymptotic behavior of the kernel involved in the $L_2(\mu)$ spline projection of functions onto certain linear spaces of splines on the real line. This $L_2(\mu)$ projection of f, for Lebesgue measure, can generally be written as

$$f_p(x) = \int K(x,y)f(y)dy. \tag{1.1}$$

If $N'(x) = (\ldots, N_{-1}(x), N_0(x), N_1(x), \ldots)$ forms a basis for the linear space then $K(x,y)$ can usually be written as

$$K(x,y) = N'(x)M^{-1}N(y) \tag{1.2}$$

where $M_{ij} = \int N_i(x)N_j(x)dx$. We will be studying here spaces of linear, quadratic or cubic splines with a knot sequence becoming dense over the interval of interest. In such cases it is clear that for fixed $x, K(x,y)$ integrates to one since the function $f \equiv 1$ is in the linear space, and $K(x,y)$ should approach a "delta function" at x as the knot sequence becomes dense. The kernel $K(x,y)$ here is possible negative. If $X_1, X_2, \ldots, X_n$ is a random sample from a density f, it might not be unreasonable to estimate f by

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

Our study is motivated mainly from two sources. These are the papers by Boneva, Kendall and Stefanov (1971) (BKS) and Silverman (1985). In the first paper histosplines were introduced. These were quadratic splines which smoothed histograms and matched areas over each bin. To be somewhat more precise, they solved the following minimization problem: let $W_2'(-\infty,\infty) = \{f : f, f' \in L_2(-\infty,\infty)\}$ and minimize $\int (f'(x))^2 dx$ for $f \in W_2'(-\infty,\infty)$ subject to $\int_{jh}^{(j+1)h} f(x)dx = h_j, -\infty < j < \infty$, where h_j is the fraction of observations falling in $[jh, (j+1)h)$. The resulting histospline $\hat{f}_{BKS}(x)$ can be shown to be approximately

$$\hat{f}_{BKS}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right). \tag{1.4}$$

The kernel K in (1.4) is the delta spline obtained by solving the above minimization problem with bin size $h = 1$ and $h_0 = 1, h_j = 0$ for $j \neq 0$. We will show that the estimate $\hat{f}_p$ in (1.3) behaves very much like $\hat{f}_{BKS}$. The kernel $K(x,y)$ from (1.3) is not however of a translation type as in (1.4). For the spline of order d (degree $d-1$) the kernel behaves, for fixed x, like a convex combination of d translation kernels. The weights depend on x and are certain nonnegative splines adding to one.

The paper by Silverman (1985) is concerned mainly with the regression smoothing spline. The smoothing spline estimates f from data $Y_i = f(x_i) + \epsilon_i$ by minimizing

$$S(f) = \sum_{i=1}^{n} (Y_i - f(x_i))^2 + \alpha \int (f(x))^2 dx. \tag{1.5}$$

The function minimizing (1.5) is a natural cubic spline $\hat{f}_s$ with knots at each observation point x_i. In Silverman (1985) it is shown that

$$\hat{f}_s(x) = \frac{1}{n} \sum_{i=1}^{n} G(x, x_i) Y_i \tag{1.6}$$

where $G(x,y) = \frac{1}{p(y)} \frac{1}{h(y)} K\left(\frac{x-y}{h(y)}\right)$. The function p is the local density of the x_i, $h(x) = \alpha^{1/4} n^{-1/4} p(x)^{-1/4}$ and K is a certain kernel. In the discussion to the paper Drs. R.L. Parker and J.A. Rice inquire and comment on their use of a regression spline and a penalized version of it. Such splines would use knots other than the n data points x_i and usually many fewer knots. The regression spline involves the usual regression of the data onto a set of splines and can be written as a projection

$$\hat{f}_{PR}(x) = \frac{1}{n} \sum_{i=1}^{n} N'(x) M^{-1}(n) N(x_i) Y_i \tag{1.7}$$

Here $N'(x) = (\ldots, N_{-1}(x), N_0(x), N_1(x), \ldots)$ denote a basis for the splines and $M(n)$ depends on the data points x_i i.e.

$$M_{ij}(n) = \frac{1}{n} \sum_{k=1}^{n} N_i(x_k) N_j(x_k). \tag{1.8}$$

If the points x_i for larger n again have limiting density $p(x)$ then

$$M_{ij}(n) \simeq \int N_i(x) N_j(x) p(x) dx \tag{1.9}$$

which would give (1.2) if $p(x) \equiv 1$. Parker and Rice comment that the kernel

$$K_n(x,y) = N'(x) M^{-1}(n) N(y)$$

as a function of y has more pronounced "side lobes" than the corresponding kernel for the smoothing spline.

In this paper we will concentrate on the problem of L_2 (Lebesgue) spline projection and apply it to density estimation. However, the spline projection can extend to $L_2(\mu)$ case for fairly general μ and can apply to the regression problem where the design points are assumed to have a limiting distribution μ. This paper is organized as follows. In Section 2 we formulate the linear, quadratic and cubic spline projection kernels with equi-spaced knots. Some properties of the spline projection kernels are also stated in Section 2. In Section 3 we give the asymptotic behavior for the estimator in (1.3). In Section 4 we compare the asymptotic behavior of the quadratic estimator in (1.3) with the BKS histospline. It is found that both estimators have the same asymptotic bias in estimating $f(x)$ and $f'(x)$. However, the quadratic estimator in (1.3) has smaller variance. In Section 5 we briefly describe the utilization of variable knots. Proofs are in an appendix.

2. The spline projection kernels

For simplicity of notations and formulas we may choose the knot sequence $\xi = \{i\}_{i=-\infty}^{\infty}$ for even order splines, or $\xi = \{i - \frac{1}{2}\}_{i=-\infty}^{\infty}$ for odd order splines. Given the order d and knot sequence ξ, let $N'(x) = (\ldots, N_{-1}(x), N_0(x), N_1(x), \ldots)$ be the normalized B-spline basis indexed in such a way that each $N_i(x)$ is centered at i. Let

$$S^2_{d,\xi} = \{s(x) : s(x) = N'(x)\theta \text{ with } \theta \in \ell_2\}.$$

The kernels associated with the L_2 projection onto $S^2_{d,\xi}$ are given below. For the space of linear, quadratic or cubic splines ($d = 2, 3$ or 4), the (i,j)-th entry of M^{-1} is given by

$$(M^{-1})_{ij} = \sum_{\ell=1}^{d-1} C_\ell \gamma_\ell^{|i-j|},$$

where C_ℓ's and γ_ℓ's depend on the order d. For $d = 2$,

$$\gamma_1 = -2 + \sqrt{3} \text{ and } C_1 = \sqrt{3}.$$

For $d = 3, \gamma_1$ and γ_2 are roots, with magnitude less than one, of $\gamma^4 + 26\gamma^3 + 66\gamma^2 + 26\gamma + 1 = 0$. The constants C_1 and C_2 can be solved from the system of equations.

$$C_1 S(\gamma_1) + C_2 S(\gamma_2) = 0$$

$$C_1 T(\gamma_1) + C_2 T(\gamma_2) = 120,$$

where $S(\gamma) = \gamma + 26 + 66\gamma + 26\gamma^2 + \gamma^3$ and $T(\gamma) = \gamma^2 + 26\gamma + 66 + 26\gamma + \gamma^2$. The results are

$$\gamma_1 = \frac{-13 + \sqrt{105} + \sqrt{270 - 26\sqrt{105}}}{2} \simeq -.4305\ 75344$$

$$\gamma_2 = \frac{-13 - \sqrt{105} + \sqrt{270 + 26\sqrt{105}}}{2} \simeq -.04330\ 96288$$

$$C_1 \simeq 3.0949\ 86517$$

$$C_2 \simeq -.2528\ 15597.$$

For $d = 4$, γ_1, γ_2 and γ_3 are roots, with magnitude less than one, for $\gamma^6 + 120\gamma^5 + 1191\gamma^4 + 2416\gamma^3 + 1191\gamma^2 + 120\gamma + 1 = 0$; and where C_1, C_2 and C_3 can be solved from the following systems of equations

$$C_1 S(\gamma_1) + C_2 S(\gamma_2) + C_3 S(\gamma_3) = 0$$

$$C_1 T(\gamma_1) + C_2 T(\gamma_2) + C_3 T(\gamma_3) = 0$$

$$C_1 Z(\gamma_1) + C_2 Z(\gamma_2) + C_3 Z(\gamma_3) = 5040$$

where

$$S(\gamma) = \gamma + 120 + 1191\gamma + 2416\gamma^2 + 1191\gamma^3 + 120\gamma^4 + \gamma^5$$

$$T(\gamma) = \gamma^2 + 120\gamma + 1191 + 2416\gamma + 1191\gamma^2 + 120\gamma^3 + \gamma^4$$

$$Z(\gamma) = \gamma^3 + 120\gamma^2 + 1191\gamma + 2416 + 1191\gamma + 120\gamma^2 + \gamma^3.$$

The results are

$$\gamma_1 \simeq -.5352\ 80431$$

$$\gamma_2 \simeq -.1225\ 54615$$

$$\gamma_3 \simeq -.0091\ 48695$$

and

$$C_1 \simeq 6.0162\ 84002$$

$$C_2 \simeq -1.0558\ 20294$$

$$C_3 \simeq 0.0042\ 69179.$$

From the above discussion, the kernel can be written as

$$K(x, y) = \sum_{i,j} N_i(x) N_j(y) \left(\sum_{\ell=1}^{d-1} C_\ell \gamma_\ell^{|i-j|} \right).$$

Define

$$H(x) = \sum_i N_i(x) \left(\sum_{\ell=1}^{d-1} C_\ell \gamma_\ell^{|i-j|} \right).$$

It is straightforward to check that

$$K(x,y) = \sum_{\ell=-[\frac{d-1}{2}]}^{[\frac{d}{2}]} N_\ell(w) H(x-i-\ell),$$

if $y \in [\xi_i, \xi_{i+1})$ and $w = y - \xi_i$. Note that $\sum_{\ell=-[\frac{d-1}{2}]}^{[\frac{d}{2}]} N_\ell(w) \equiv 1$. The spline projection kernel, for fixed y, can be written as a convex combination of d translations of H. The weights depend on the distance between y and the nearest knot and are certain nonnegative splines adding to one. There are certain similarities between $H(x)$ and the delta spline in BKS (1971). The delta spline dies out at rate $\gamma = -2+\sqrt{3}$ and, as one moves one cell further out, the present cell can be obtained merely by multiplying the past cell with the factor $\gamma = -2+\sqrt{3}$. For $d=2, H(x)$ has the same properties with the same factor $\gamma = -2+\sqrt{3}$. The difference is that the delta spline is parabolic while $H(x)$ is linear. For $d=3, H(x)$ is the sum of two parabolas in each cell and each parabola dies out at its own rate. As one moves one cell further out, each parabola in the present cell can be obtained by multiplying the parabola in the past cell with its own factor. For $d=4, H(x)$ is the sum of three cubic curves in each cell, which have analogous properties as described above. Pictures of $H(x)$ and $K(x,y)$ with various values of y are given in Figure 1 and Figure 2.

Following is a list of properties of the L_2 spline projection.

Property 1. The spline space $S^2_{d,\xi}$ is a reproducing kernel Hilbert space with the reproducing kernel $K(x,y)$.

Property 2. The integral transform in (2.1) is well-defined for $f \in L_1$. This desirable property enables us to apply the "projection" to the general density function. Actually, the integral transform in (2.1) can extend to functions satisfying the conditions

$$f \in L_1^{local} = \{f : \int_A |f(x)|dx < \infty \text{ for all measurable } A \text{ with finite measures}\} \qquad (2.2)$$

and

$$f(x) = O(x^\alpha) \text{ for some } \alpha \text{ as } x \to \infty. \qquad (2.3)$$

Property 3. Define another spline space

$$S_{d,\xi} = \{x(x) : s(x) = N'(x)\theta \text{ with } \theta \in \ell_\infty\}.$$

The kernel $K(x,y)$ reproduces $S_{d,\xi}$. That is,

$$\int_{-\infty}^{\infty} K(x,y)s(y)dy = s(x) \text{ for } s \in S_{d,\xi}. \tag{2.4}$$

Property 4. Let $\mathcal{P}$ be the space of polynomials of order d (degree $\leq d-1$). The kernel reproduces $\mathcal{P}$. That is,

$$\int_{-\infty}^{\infty} K(x,y)p(y)dy = p(x) \text{ for } p \in \mathcal{P}. \tag{2.5}$$

Especially,

$$\int_{-\infty}^{\infty} K(x,y)dy \equiv 1.$$

Property 5.

$$\int_{-\infty}^{\infty} H(x)dx = 1.$$

Property 6. The kernel decays to zero at an exponential rate as $|x-y| \to 0$. To be more precise, we have

$$|K(x,y)| \leq C\gamma^{n(x,y)},$$

for some constants $C > 0$ and $0 < \gamma < 1$, where $n(x,y)$ is the number of knots between x and y.

Property 7.

$$\int_{-\infty}^{\infty} K(x,y)y^d dy - x^d = -B_d(w)$$

where $w = x - \xi_i$ if $x \in [\xi_i, \xi_{i+1})$ and $B_d(w)$ is the d-th Bernoulli polynomial. The above equality says the bias of x^d, if approximated by its integral transform in (1.2), behaves like the d-th Bernoulli polynomial in each cell.

All of the above materials are discussed under the case of unit knot distance. It is just straightforward to extend the results to the case where the knot distance is h. For examples, the kernel scaled by h becomes $K\left(\frac{x}{h}, \frac{b}{h}\right)/h$ and the bias in Property 7 is $-B_d(w)h^d$ with $w = \frac{x-\xi_i}{h}$.

3. Asymptotics

This section is mainly devoted to the asymptotic behavior of the estimator in (1.3). We assume the knots are equi-spaced with distance h.

Theorem 1. Suppose $f^{(d)}(x)$ is uniformly continuous. For a fixed x, we have

$$E\hat{f}_p^{(k)}(x) - f^{(k)}(x) = -\frac{f^{(d)}(x)}{(d-k)!}B_{d-k}(w)h^{d-k} + O(h^{d-k+1}), k = 0, 1, \ldots, d-1,$$

where $w = \frac{x-\xi_i}{h}$ if $x \in [\xi_i, \xi_{i+1})$.

Theorem 2. Suppose $f(x)$ is uniformly continuous and that $nh \to \infty$ and $h \to 0$ as $n \to \infty$. For a fixed x, we have

$$\begin{aligned}\text{Var } \hat{f}_p(x) &= \frac{1}{n}f(x)K(x,x) + O\left(\frac{1}{n}\right)\\ &= \frac{1}{nh}f(x)K_0(w,w) + O\left(\frac{1}{n}\right)\end{aligned}$$

where K_0 is the spline projection kernel of order d and integer knots, and $w = \frac{x-\xi_i}{h}$ if $x \in [\xi_i, \xi_{i+1})$.

Pictures of $B_d(w)$ and $K(w,w)$ for $d = 2, 3$ or 4 are in Figure 3 and Figure 4. From Figure 3, gives an idea of how the spline projection approximates the true functions and how the positioning of knots affect the approximation. From Figure 4, we can see as one moves away from knots, the estimator in (1.3) tends to have smaller variance if the order is even, or larger variance if the order is odd.

4. Comparison with histospline

Lii and Rosenblatt (1975) derived the asymptotics for histosplines on finite interval [0, 1]. What they did is as follows. Place knots at $\xi_i = \frac{i}{\ell+1}, i = 0, 1, \ldots, \ell+1$. Let $s(x)$ be the cubic spline interpolator of the sample c.d.f. $F_n(x)$ with knots $\xi = \{\xi_i\}_{i=0}^{\ell+1}$ and with the same known boundary conditions $s'(0) = f(0)$ and $s'(1) = f(1)$ (or some other known boundary conditions, e.g., $s''(0) = f'(0) = s''(1) = f'(1) = 0$ or $f(x)$ is known to be periodic and the periodic boundary conditions $s'(0) = s'(1)$ and $s''(0) = s''(1)$ are imposed.). If the boundary behavior of f is unknown, then the boundary conditions $s''(0) = s''(1) = 0$ will be imposed, and the resulting curve is the smoothest one in the sense that $\int_0^1 (s''(x))^2 dx$ is smallest. The estimator $s'(x)$ is used to estimate $f(x)$.

The following theorems concerning the asymptotics of $s'(x)$ away from the boundary.

Theorem 1. Let $f \in C^3[0,1]$ and $w = \frac{x-\xi_i}{h}$, where $h = \frac{1}{\ell+1}$ is the distance between two knots. Then we have

$$\begin{aligned} \text{bias } &= ES'(x) - f(x) = -\frac{f^{(3)}(x)}{3!} B_3(w) h^3 + o(h^3) \\ ES''(x) - f'(x) &= -\frac{f^{(3)}(x)}{2!} B_2(w) h^2 + o(h^2) \end{aligned}$$

if $0 < x < 1$ is fixed and $x \in [\xi_i, \xi_{i+1})$ as $\ell \to \infty$. (See Lii and Rosenblatt, 1975 and Rosenblatt, 1976.)

Theorem 2. Let f be continuous on [0, 1] and let $\sigma = \sqrt{3} - 2$. Then the variance of $s'(x)$ is

$$\text{var } s'(x) = \frac{f(x)}{nh} A(w) + o\left(\frac{h}{n}\right)$$

if $0 < x < 1$ is fixed, w is the same as the above theorem and $nh \to \infty$ and $h \to 0$ as $n \to \infty$. Here

$$\begin{aligned} A(w) = 1 &- \frac{3(1-\sigma)}{2+\sigma}\left(2w^2 - 2w + \frac{1}{3}\right) + \frac{9}{4}\left(\frac{1-\sigma}{2+\sigma}\right)^2 \\ &\left\{\left(2w^2 - 2w + \frac{1}{3}\right)^2 + \left[\left(w^2 - \frac{1}{3}\right) + \sigma\left(\frac{1}{3} - (1-w)^2\right)\right]^2 \frac{1}{1-\sigma^2}\right. \\ &\left. + \left[\left(w^2 - \frac{1}{3}\right) + \frac{1}{\sigma}\left(\frac{1}{3} - (1-w)^2\right)\right]^2 \frac{\sigma^2}{1-\sigma^2}\right\}. \end{aligned}$$

Though the above theorems are done for estimators in finite interval, the authors believe, for histosplines in the real line, the bias and variance cannot be better than those given in the above theorems and may achieve the same results if further conditions, $f^{(3)}$ is uniformly continuous on the real line in Theorem 1, or f is uniformly continuous on the real line in Theorem 2, are met. Compare Theorem 1 and Theorem 2 in Section 3 and Section 4. The biases for estimating $f(x)$ and $f'(x)$ are asymptotically the same for both estimators. The variances are of the same order of magnitude $O\left(\frac{1}{nh}\right)$, however, the constant term in the spline projection kernel estimator is smaller than that in the histospline. A plot comparing $A(w)$ and $K_0(w,w)$ is in Figure 5. Figure 6 is 10 independent estimates using quadratic spline projection kernel with h=1. Each estimate is based on 100 random numbers drawn from the Gumbel distribution with density function

$$f(x) = e^{-x-e^{-x}}.$$

Compare Figure 6 with Figure 10b in BKS (1971). Figure 6 shows less sampling fluctuation.

5. Variable knots

Instead of equi-spaced knots, we may consider variable knots. Suppose the knot sequences have a limiting density $p(x)$. To adapt the amount of smoothing to the local density $p(x)$, the estimate may be constructed out of $K\left(\frac{x}{h(x)}, \frac{y}{h(y)}\right)/\sqrt{h(x)h(y)}$, where K is the spline projection kernel with unit knot distance and certain order d and $h(x) = \frac{h}{p(x)}$, where h controls the overall smoothness. The estimate is given by

$$\hat{f}_p(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x}{h(x)}, \frac{X_i}{h(X_i)}\right)/\sqrt{h(x)h(X_i)}.$$

Those X_i's which are not close to x have little contribution to the estimate. Therefore it may not be unreasonable to replace $h(X_i)$ by $h(x)$ to avoid the local smoothing parameters from depending on the data.

An estimate for $f(x)$ is then

$$\hat{f}_p(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h(x)} K\left(\frac{x}{h(x)}, \frac{X_i}{h(x)}\right).$$

Studies (Huang, 1990) have suggested to choose $p(x)$ proportional to

$$\left[\frac{(f^{(d)}(x))^2}{f(x)}\right]^{\frac{1}{2d+1}}.$$

That is, the local smoothing parameter should be chosen proportional to

$$\left[\frac{f(x)}{(f^{(d)}(x))^2}\right]^{\frac{1}{2d+1}}.$$

This is similar to the results in Parzen (1962), he has the optimum value of h (in the sense of minimizing mean square error at x) as

$$h_{opt}(x)\alpha\left[\frac{f(x)}{(f^{(d)}(x))^2}\right]^{\frac{1}{2d+1}},$$

where d is the characteristic exponent for the Fourier transform of the kernel K.

Appendix

Proof of Property 1: $S^2_{d,\xi}$ is a Hilbert space which admits $K(x,y)$. By the uniqueness of the reproducing kernel, $S^2_{d,\xi}$ is a reproducing kernel Hilbert space with the reproducing kernel $K(x,y)$. (See Saburon Saitoh, 1989)

Proof of Property 2: This can be easily seen by the fact that the (i,j)-th entry of M^{-1} goes to zero at an exponential rate as $|i-j| \to \infty$.

Proof of Property 3: By Property 2, the left part of (2.4) is well-defined. Then

$$\begin{aligned}
& \int_{-\infty}^{\infty} K(x,y)s(y)dy \\
= & \int_{-\infty}^{\infty} N'(x)M^{-1}N(y)N'(y)\theta dy \\
= & \; N'(x)\theta \\
= & \; s(x).
\end{aligned}$$

Proof of Property 4: Given $p(x) \in \mathcal{P}, p(x)$ can be written as

$$p(x) = N'(x)\theta$$

where $\theta = (\ldots, \theta_{-1}, \theta_0, \theta_1, \ldots)$ satisfies $\theta_n = O(n^{d-1})$. By Property 2, the left side of (2.5) is well-defined. Then

$$\begin{aligned}
& \int_{-\infty}^{\infty} K(x,y)p(y)dy \\
= & \int_{-\infty}^{\infty} N'(x)M^{-1}N(y)N'(y)\theta dy \\
= & \; N'(x)\theta = p(x).
\end{aligned}$$

Proof of Property 5: Recall that $K(x,y)$ can be written as

$$K(x,y) = \sum_{\ell=-[\frac{d-1}{2}]}^{[\frac{d}{2}]} N_\ell(w)H(x-i-\ell), \quad w = y - \xi_i.$$

Then

$$1 \equiv \int_{-\infty}^{\infty} K(x,y)dx$$

$$= \left(\sum_{\ell=-[\frac{d-1}{2}]}^{[\frac{d}{2}]} N_\ell(w) \right) \int_{-\infty}^{\infty} H(x)dx$$

$$= \int_{-\infty}^{\infty} H(x)dx.$$

Proof of Property 6: This is clear from the fact that the (i,j)-th entry in M^{-1} is of form

$$(M^{-1})_{ij} = \sum_{\ell=1}^{d-1} C_\ell \gamma_\ell^{|i-j|}.$$

Proof of Property 7: Without loss of generality, we may take integer knots. It suffices to check

i) $x^d - B_d(w)$ can be written as $N'(x)\theta$, and

ii) $\int_{-\infty}^{\infty} B_d(w) N_i(x)dx = 0$ for all i.

Check i): $x^d - B_d(w) = x^d - B_d(x-[x]) = x^d - (x-[x])^d + \sum_{k=0}^{d-1} a_k(x-[x])^k$ for some suitable a_k's. Therefore i) holds.

Check ii): Note that $N_i(x)$, having support on $[i, i+d]$, is the density for the sum of d independent uniform $(i, i+1)$ random variables. Let $X = U_1 + \ldots + U_d$, where $U_k \overset{i.i.d.}{\sim}$ Uniform $(i, i+1)$. It can be shown that $X - [X] \sim$ uniform (0, 1). Therefore

$$\int_{-\infty}^{\infty} B_d(w) N_i(x)dx = EB_d(X-[X])$$

$$= \int_0^1 B_d(t)dt = 0.$$

Proof of Theorem 1 in Section 3: We will show the case $k = 0$ first

$$E\hat{f}_p(x) - f(x)$$

$$= \int_{-\infty}^{\infty} K(x,y)(f(y) - f(x))dy$$

$$= \int_{-\infty}^{\infty} K(x,y)\Big(f'(x)(y-x) + \ldots + \frac{f^{(d-1)}(x)}{(d-1)!}(y-x)^{d-1} + \frac{f^{(d)}(\xi_{x,y})}{d!}(y-x)^d \Big) dy \qquad (A.1)$$

where $\xi_{x,y}$ is some number between x and y. Since $K(x,y)$ reproduces polynomials of order d,

$$(A.1) = \int_{-\infty}^{\infty} K(x,y)\frac{f^{(d)}(\xi_{x,y})}{d!}(y-x)^d dy$$

$$= \int_{-\infty}^{\infty} K(x,y)\frac{f^{(d)}(x)}{d!}(y-x)^d dy + \int_{-\infty}^{\infty} K(x,y)\frac{f^{(d)}(\xi_{x,y}) - f^{(d)}(x)}{d!}(y-x)^d dy$$

$$= -\frac{f^{(d)}(x)}{d!}B_d(w)h^d + \int_{-\infty}^{\infty} K(x,y)\frac{f^{(d)}(\xi_{x,y}) - f^{(d)}(x)}{d!}(y-x)^d dy \qquad (A.2)$$

The proof can be completed by showing the second term in (A.2) is $O(h^{d+1})$.

$$\Big|\int_{-\infty}^{\infty} K(x,y)\frac{f^{(d)}(\xi_{x,y}) - f^{(d)}(x)}{d!}(y-x)^d dy\Big|$$

$$= \Big|\sum_{m=-\infty}^{\infty} \int_{\xi_i+(m-1)h}^{\xi_i+mh} K(x,y)\frac{f^{(d)}(\xi_{x,y}) - f^{(d)}(x)}{d!}(y-x)^d dy\Big|$$

$$\leq \sum_{m=-\infty}^{\infty} \frac{C\gamma^{|m-1|}}{hd!} w(f^{(d)}, mh)(mh)^d h \qquad (A.3)$$

where C and γ are constants given in Property 6 of section 2 and $w(f,h)$ is the first modulus of smoothness of f defined as $w(f,h) = \max_x |f(x) - f(x+h)|$. Since $f^{(d)}(x)$ is uniformly continuous, $w(f^{(d)}, mh) = O(mh)$. Thus $(A.3) = O(h^{d+1})$. Therefore, we have

$$E\hat{f}_p(x) - f(x) = -\frac{f^{(d)}(x)}{d!}B_d(w)h^d + O(h^{d+1}). \qquad (A.4)$$

For $k > 0$,

$$E\hat{f}_p^{(k)}(x) - f^{(k)}(x) = \int_{-\infty}^{\infty} \frac{d^k}{dx^k}K(x,y)f(y)dy - f^{(k)}(x). \qquad (A.5)$$

We can exchange the order of integration and derivative because $K(x, y)$ can be written as $N'(x)M^{-1}N(y)$. Then

$$(A.5) = \frac{d^k}{dx^k} \int_{-\infty}^{\infty} K(x,y)f(y)dy - f^{(k)}(x). \qquad (A.6)$$

By (A.4) and the result $\frac{d^k}{dx^k} B_d(w) = \frac{d!}{(d-k)!} B_{d-k}(w)/h^k$, we have

$$(A.6) = -\frac{f^{(d)}(x)}{(d-k)!} B_{d-k}(w)h^{d-k} + O(h^{d-k+1}).$$

Proof of Theorem 2 in Section 3:

$$\begin{aligned}
\text{Var } \hat{f}_p(x) &= \frac{1}{n}\int_{-\infty}^{\infty} K^2(x,y)f(y)dy - \frac{1}{n}\left(\int_{-\infty}^{\infty} K(x,y)f(y)dy\right)^2 \\
&= \frac{1}{n}\int_{-\infty}^{\infty} K^2(x,y)(f(x) + f(y) - f(x))dy \\
&\quad - \frac{1}{n}\left(\int_{-\infty}^{\infty} K(x,y)f(y)dy\right)^2 \\
&= \frac{1}{n}K(x,x)f(x) + \frac{1}{n}\int_{-\infty}^{\infty} K^2(x,y)(f(y) - f(x))dy \\
&\quad - \frac{1}{n}\left(\int_{-\infty}^{\infty} K(x,y)f(y)dy\right)^2
\end{aligned}$$

It is clear $K(x, x) = \frac{1}{h}K_0(w, w)$. To complete the theorem we need to show

$$\int_{-\infty}^{\infty} K^2(x,y)(f(y) - f(x))dy = O(1) \text{ and}$$

$$\int_{-\infty}^{\infty} K(x,y)f(y)dy = O(1)$$

We have

$$\Big|\int_{-\infty}^{\infty} K^2(x,y)(f(y)-f(x))dy\Big|$$

$$= \Big|\sum_{m=-\infty}^{\infty} \int_{\xi_i+(m-1)h}^{\xi_i+mh} K^2(x,y)(f(y)-f(x))dy\Big|$$

$$\leq \sum_{m=-\infty}^{\infty} \frac{C^2\gamma^{2|m-1|}}{h^2} w(f,mh)h = O(1)$$

and

$$\Big|\int_{-\infty}^{\infty} K(x,y)f(y)dy\Big|$$

$$= \Big|\sum_{m=-\infty}^{\infty} \int_{\xi_i+(m-1)h}^{\xi_i+mh} K(x,y)f(y)dy\Big|$$

$$\leq \sum_{m=-\infty}^{\infty} \frac{C\gamma^{|m-1|}}{h} M_f h = O(1)$$

where $M_f = \max_x |f(x)| < \infty$.

References

1. Boneva, L.I., Kendall, D. and Stefanov, I. (1971). Spline transformations: Three new diagnostic aids for the statistical data–analyst. *J. Roy. Statist. Soc. B* **33**, 1–70.
2. Huang, S.Y. (1990). Nonparametric density estimation by spline projection kernels. Ph.D. Thesis, Purdue University.
3. Lii, K.S. and Rosenblatt, M. (1975). Asymptotic behavior of a spline estimate of a density function. *Comput. Math. Appl.* **1**, 223–235.
4. Parzen, E. (1962). On the estimation of the probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076.
5. Rosenblatt, M. (1976). Asymptotic and representation of cubic splines. *J. Approx. Theory* **17**, 332–343.
6. Saitoh, S. (1989). Theory of reproducing kernels and its applications. John Wiley & Sons.
7. Silverman, B.W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. B* **46**, 1–52.

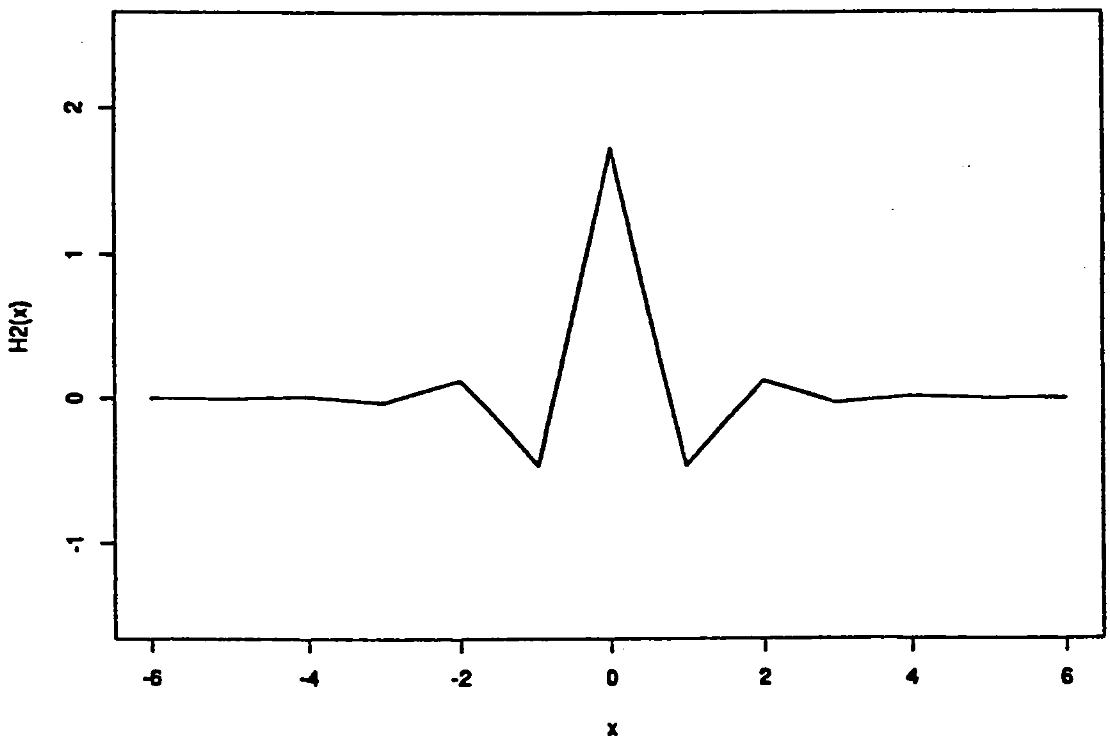

Figure 1a

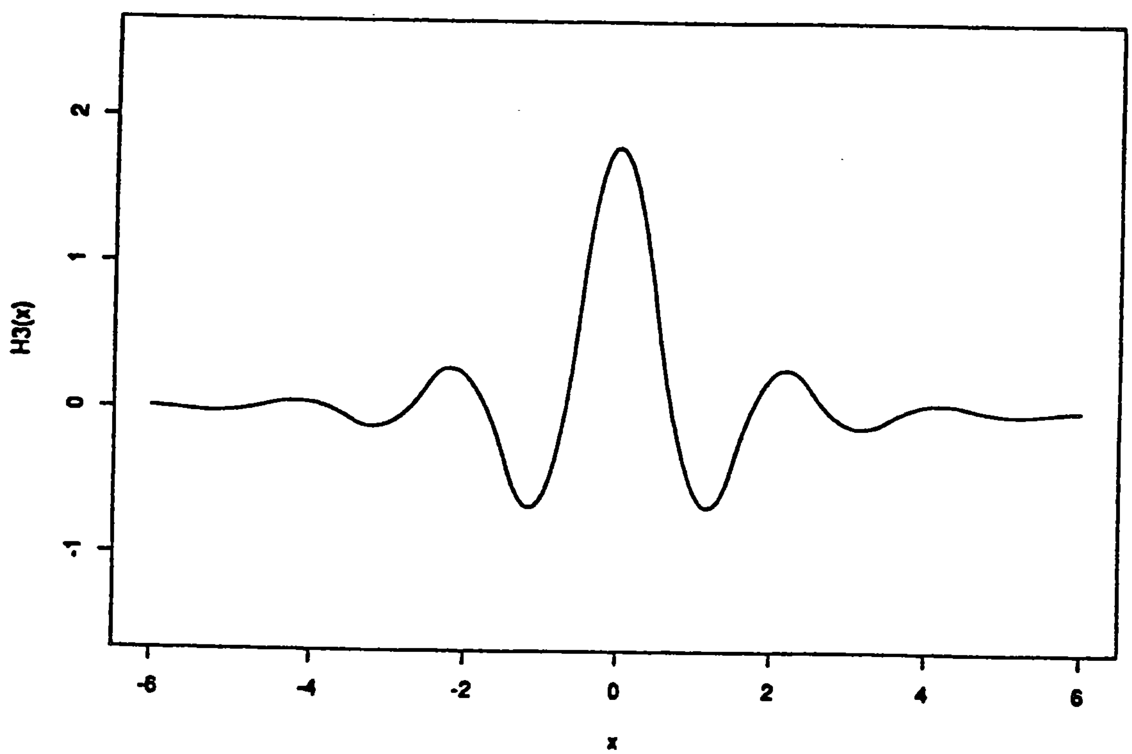

Figure 1b

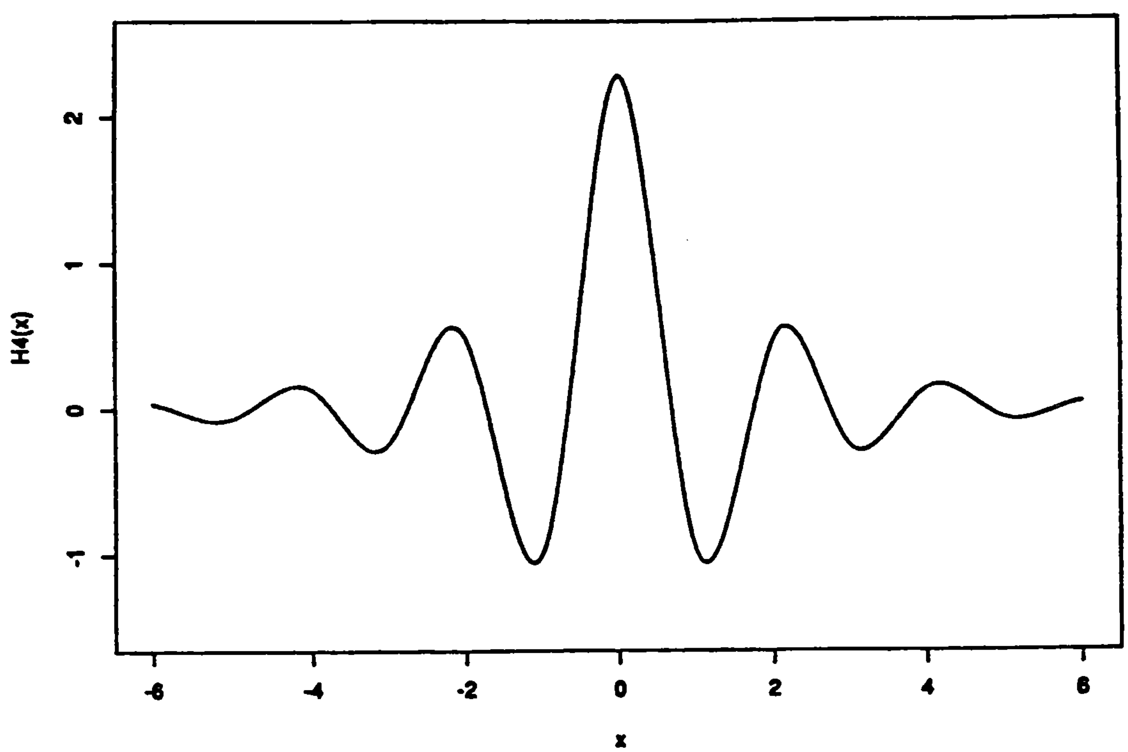

Figure 1c

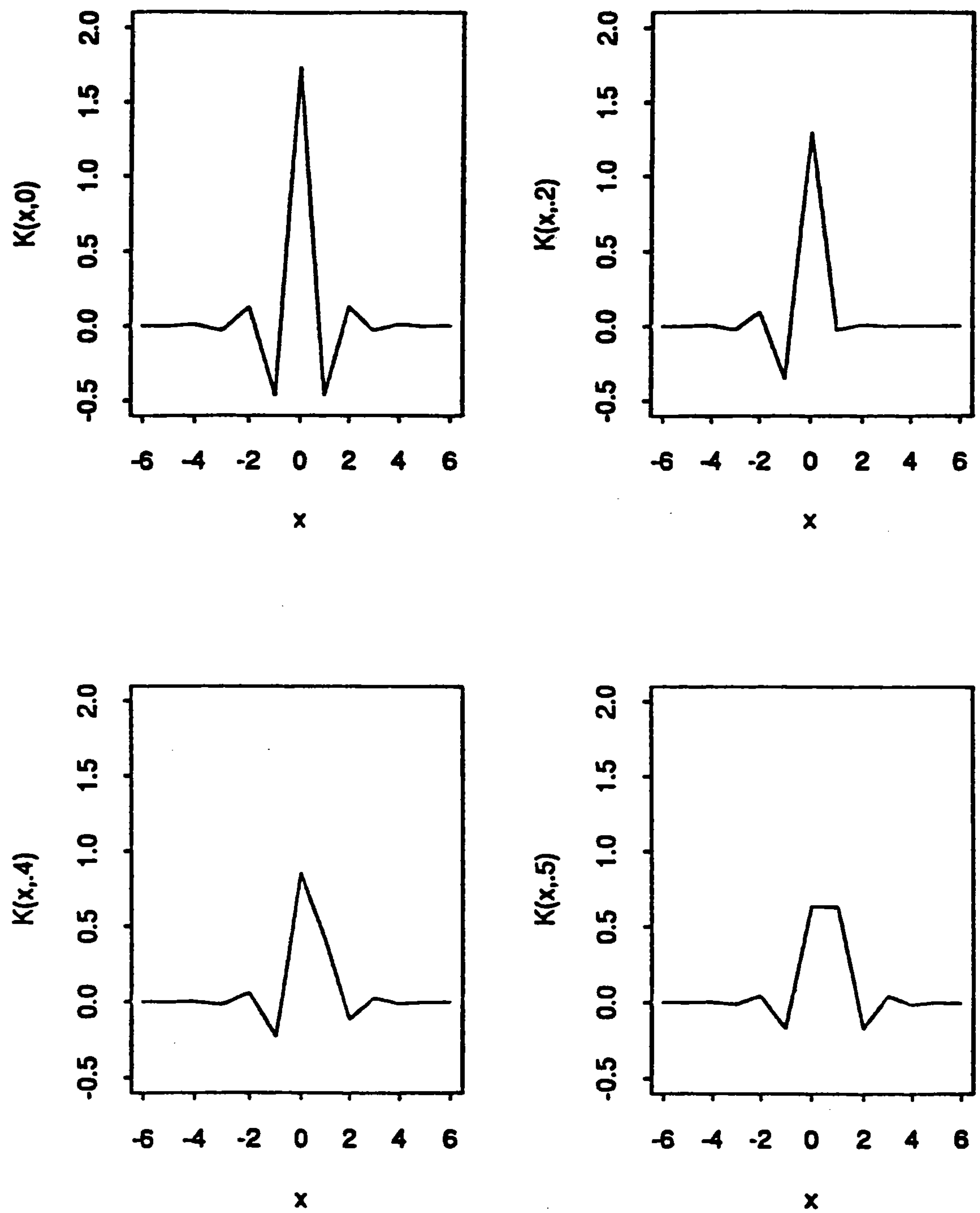

Figure 2a

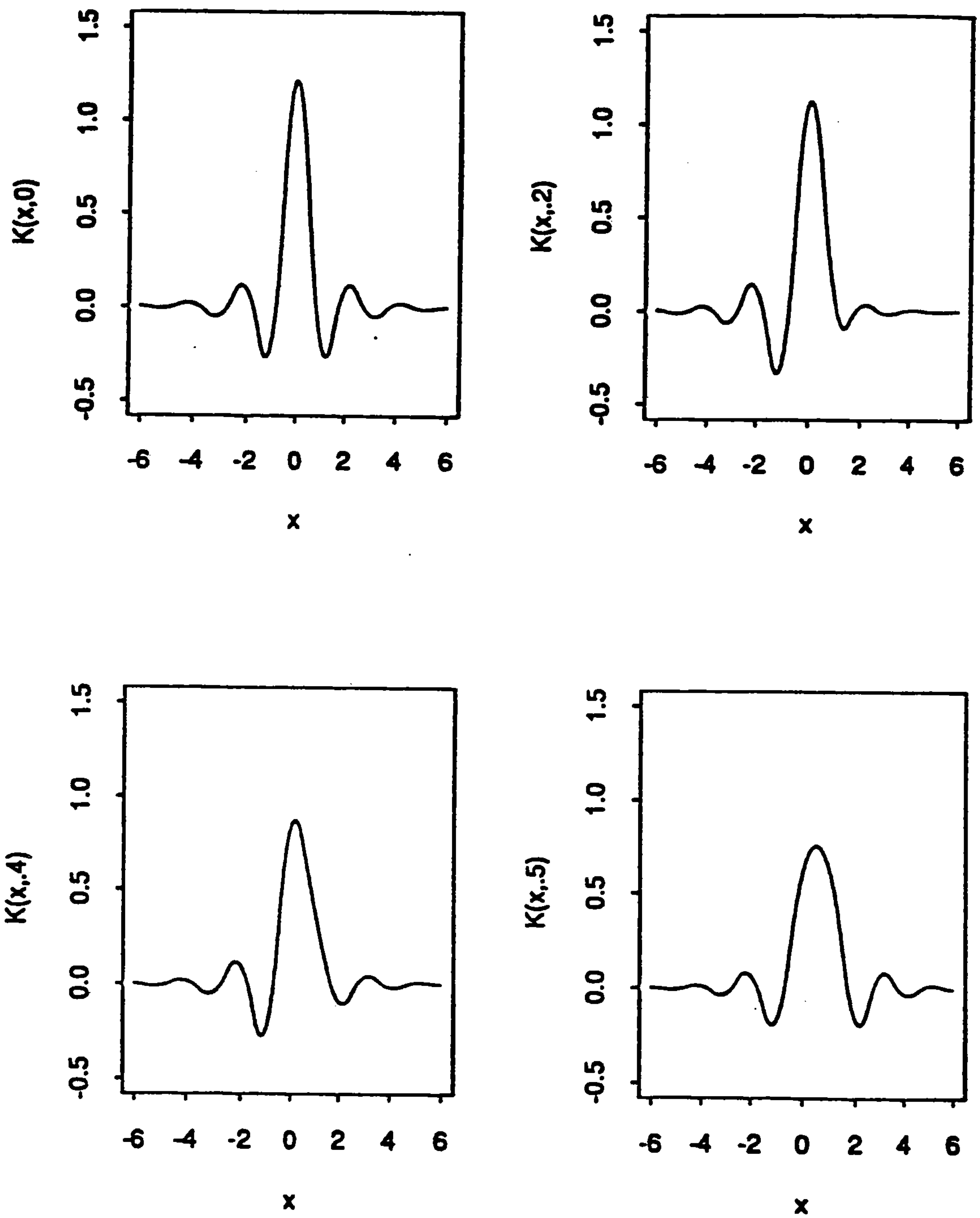

Figure 2b

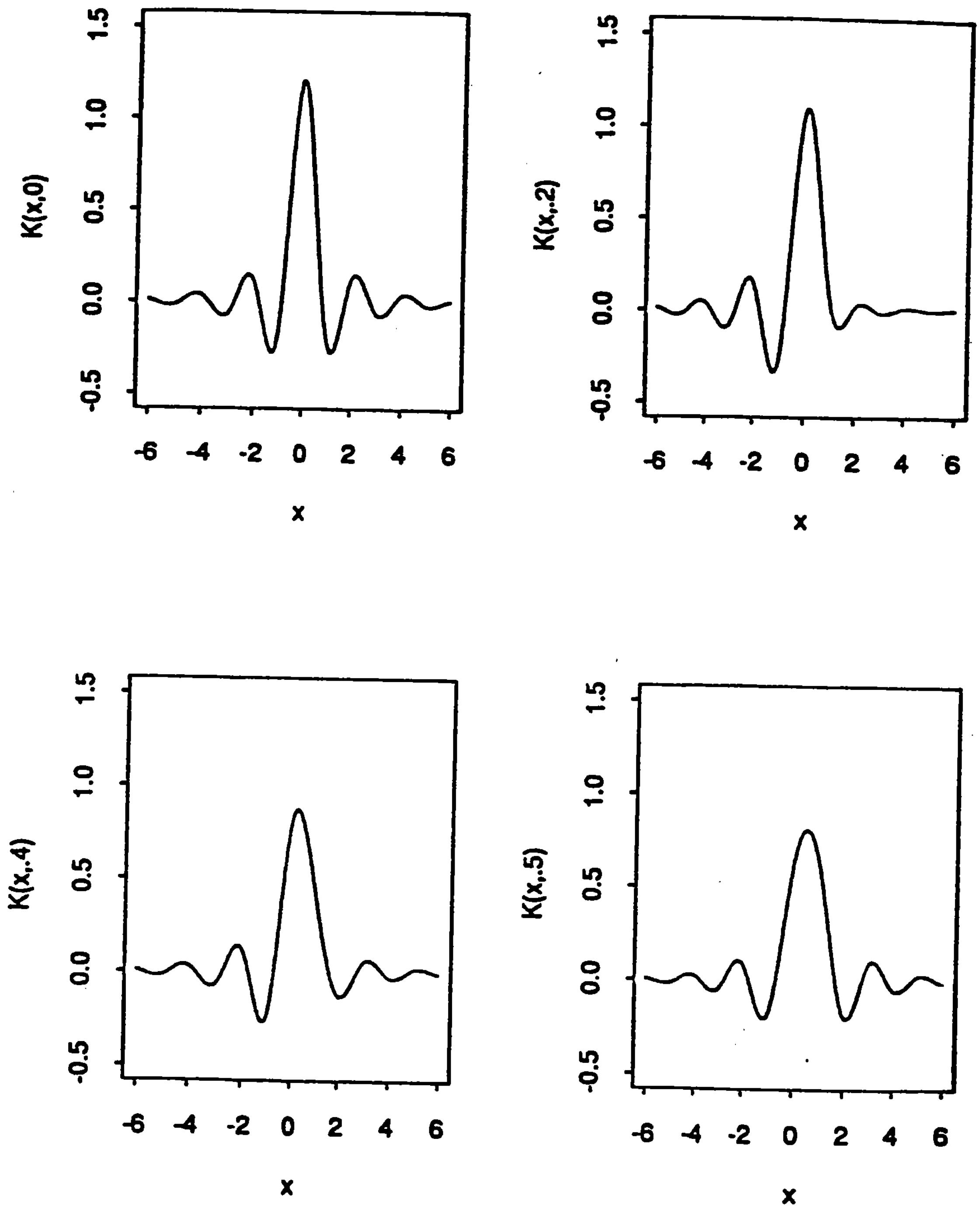

Figure 2c

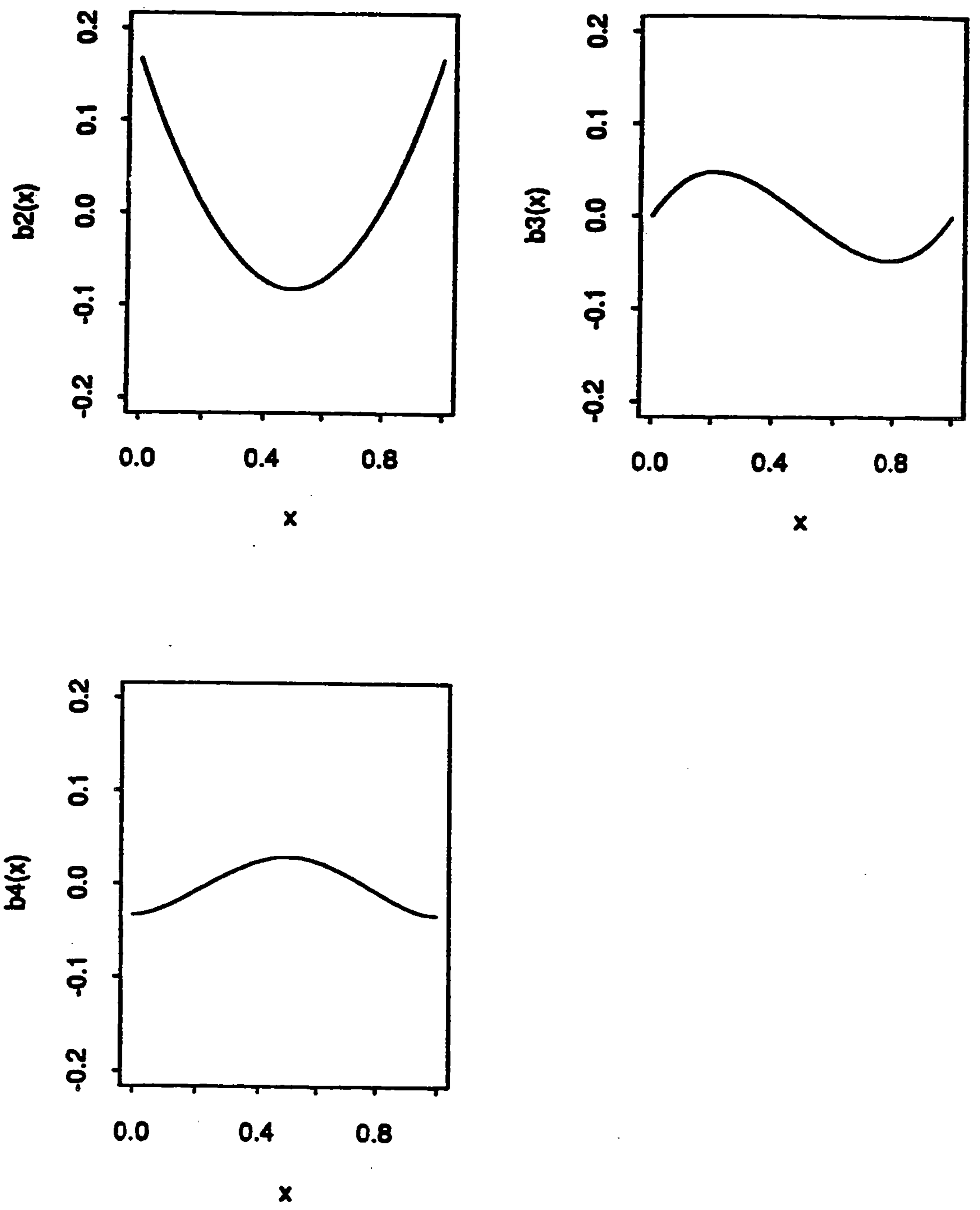
b2(x)
-0.2
-0.1
0.0
0.1
0.2
0.0
0.4
0.8
x
b3(x)
-0.2
-0.1
0.0
0.1
0.2
0.0
0.4
0.8
x
b4(x)
-0.2
-0.1
0.0
0.1
0.2
0.0
0.4
0.8
x

Figure 3

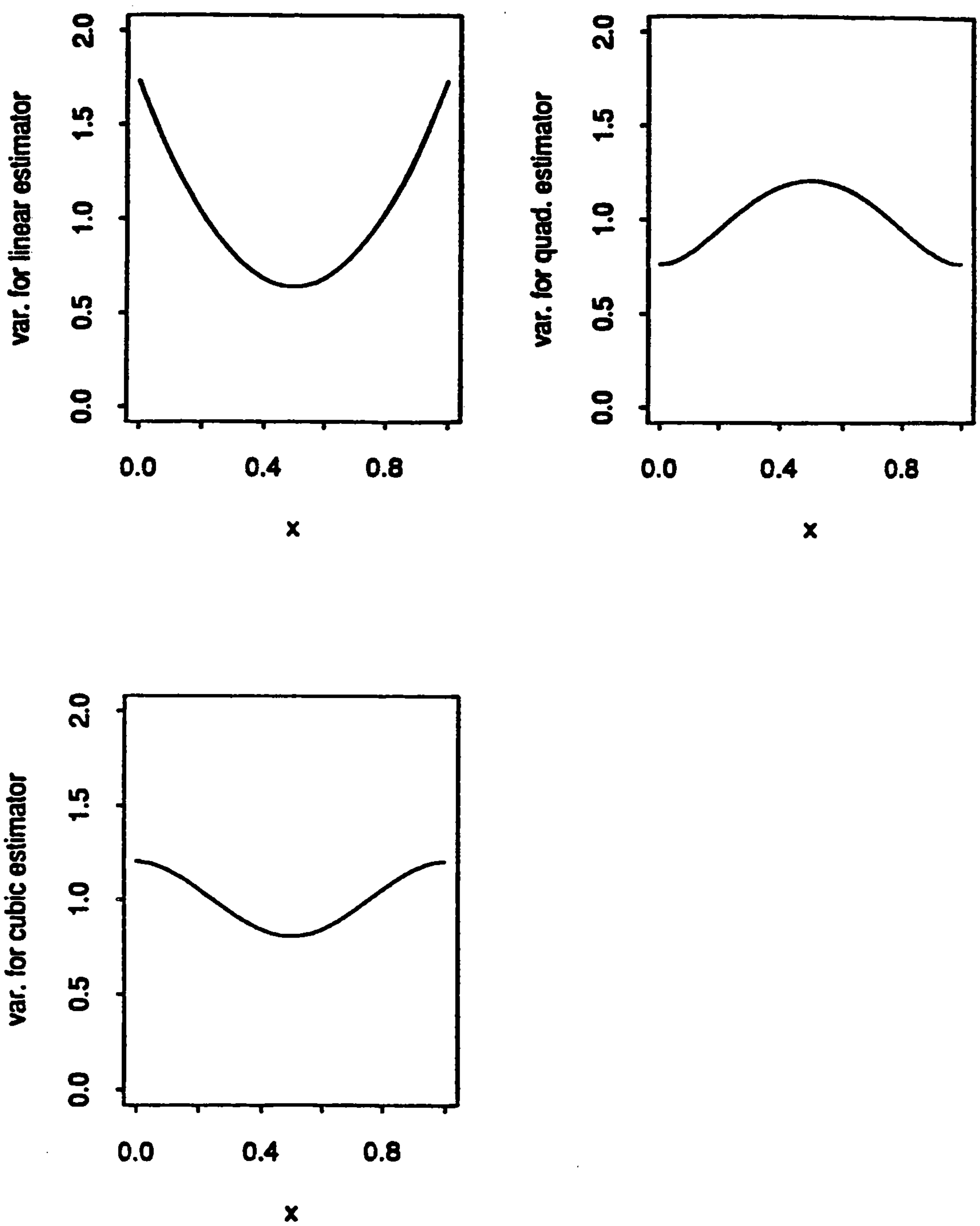
var. for linear estimator
var. for quad. estimator
var. for cubic estimator
0.0
0.5
1.0
1.5
2.0
0.0
0.4
0.8
x

Figure 4

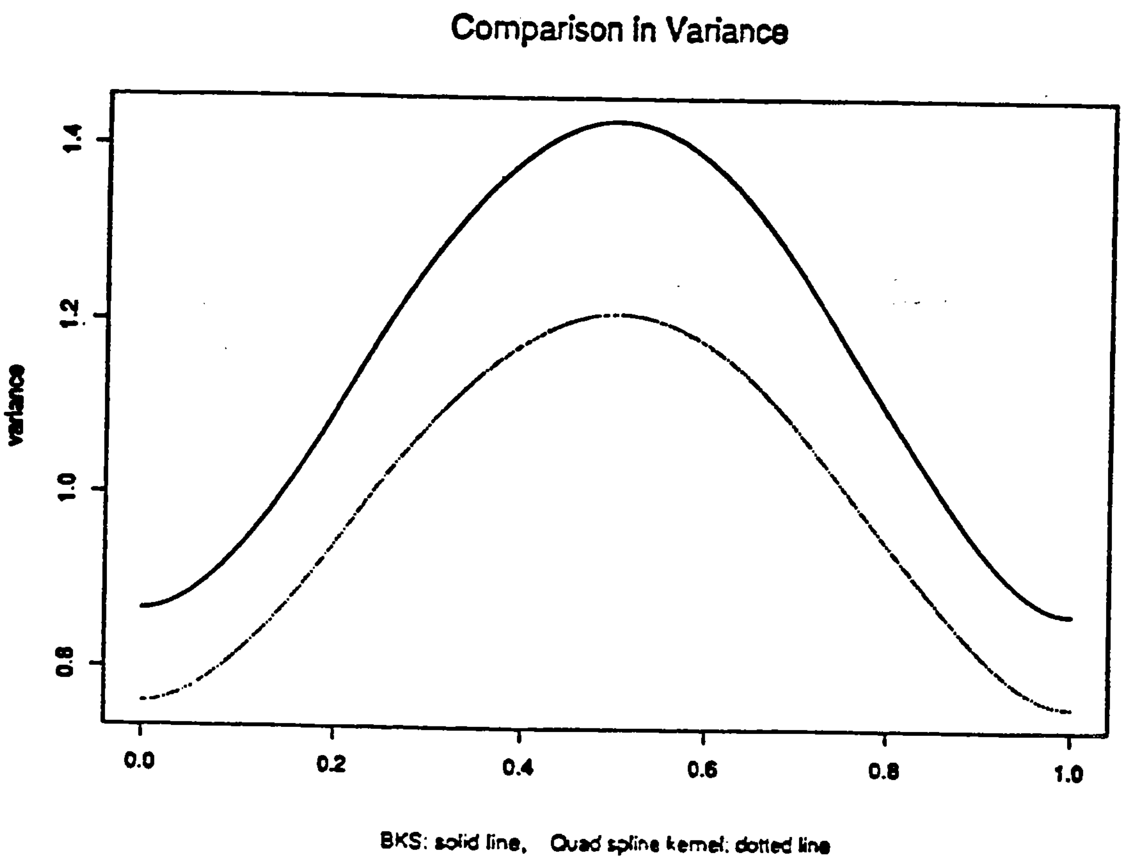
Comparison in Variance
variance
1.4
1.2
1.0
0.8
0.0
0.2
0.4
0.6
0.8
1.0
BKS: solid line, Quad spline kernel: dotted line

Figure 5

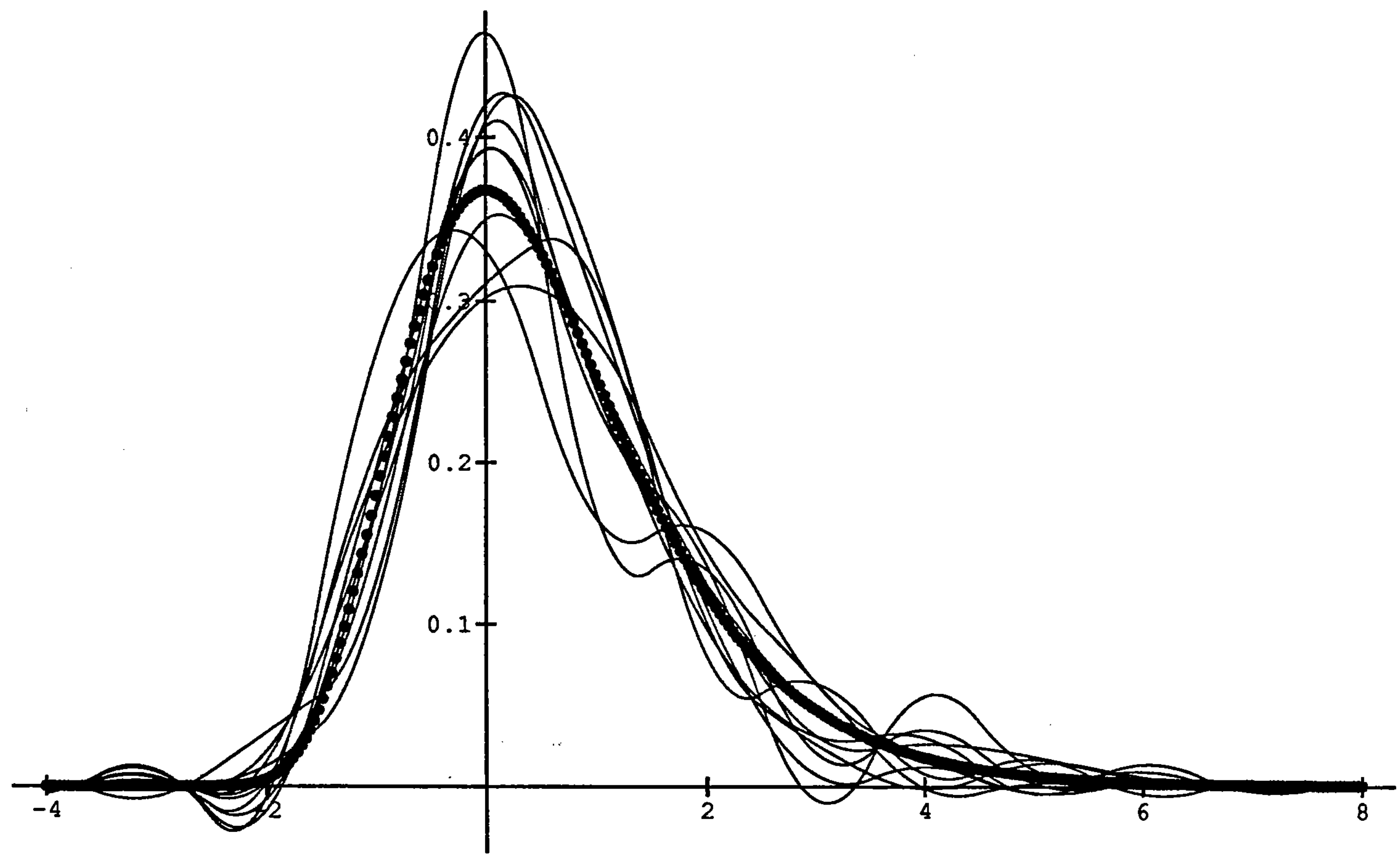

Figure 6