Statistica Sinica **10**(2000), 73-91

BETA KERNEL SMOOTHERS FOR REGRESSION CURVES

Song Xi Chen

La Trobe University

Abstract: This paper proposes beta kernel smoothers for estimating curves with compact support by employing a beta family of densities as kernels. These beta kernel smoothers are free of boundary bias, achieve the optimal convergence rate of $n^{-4/5}$ for mean integrated squared error and always allocate non-negative weights. In the context of regression, a comparison is made between one of the beta smoothers and the local linear smoother. Its mean integrated squared error is comparable with that of the local linear smoother. Situations where the beta kernel smoother has a smaller mean integrated squared error are given. Extensions to probability density estimation are discussed.

Key words and phrases: Beta kernels, boundary bias, local linear regression, mean integrated square error, nonparametric regression.

1. Introduction

In the traditional kernel methods for curve estimation, it has been widely regarded that the performance of the kernel methods depends largely on the smoothing bandwidth, and depends very little on the form of the kernel. Most kernels used are symmetric kernels and, once chosen, are fixed. This may be efficient for estimating curves with unbounded supports, but not for curves which have compact support and are discontinuous at boundary points. For curves of this type, a fixed form of kernel leads to boundary bias.

Boundary bias is a well known problem and many authors have suggested ways for removing it. In the context of nonparametric regression, Gasser and Müller (1979), Müller (1991) and Müller and Wang (1994) proposed the use of boundary kernels, while Rice (1984) used Richardson's extrapolation to combine two kernel estimates with different bandwidths. In density estimation, Schuster (1985) proposed data reflection, Marron and Ruppert (1994) considered using empirical transformations, and Jones (1993) proposed a framework of jackknife methods for correcting boundary bias. In recent years, it has been shown by Fan and Gijbels (1992) and Fan (1993), that in nonparametric regression the local linear smoother is free of boundary bias and achieves the optimal rate of convergence for mean integrated squared error. It is interesting to note that even a local linear smoother uses a fixed kernel in its initial form, the local least-squares regression implicitly employs different kernels at different places.

This paper investigates kernel smoothers for curves with compact support based on certain beta density functions. There are two unique features about the beta kernels. One is that both the shape of the beta kernels and the amount of smoothing vary according to the position where the curve estimation is made; the other is that the beta kernels assign no weight outside the data support. The beta kernel estimators respect the range of the curves as only nonnegative weights are assigned and are free of boundary bias; that is, the order of magnitude of the bias is not increased near the boundaries.

The idea of beta kernel smoothing was first considered in Brown and Chen (1999). This was motivated by the Bernstein theorem in mathematical function analysis, which says that the Bernstein polynomials associated with a continuous function on $[0, 1]$ converge uniformly to that function. The same rate of convergence is achieved by the Bernstein polynomials throughout the entire domain and thus there is no boundary bias. The Bernstein polynomials use a special type of beta kernel — the binomial probability function. However, they undersmooth as a bandwidth of order $n^{-1/2}$ is implicitly used. To overcome the undersmoothing problem, Brown and Chen (1999) proposed a kernel smoother which smooths the Bernstein polynomials by a family of beta densities. It turns out that the estimator does not have an enlarged bias near the boundaries and achieves the optimal rate of convergence for mean integrated squared error. However, the estimator applies only for regression curves with equally spaced design points. The present paper generalizes it to arbitrary designs.

The paper is structured as follows. Section 2 introduces a beta kernel estimator for regression curves. Its bias and variance are studied in Section 3, and its mean integrated squared error and the optimal bandwidth value are considered in Section 4. In Section 5 a modified smoother is introduced and a comparison with the local linear smoother is made in Section 6. Some simulation results are reported in Section 7. Extensions to probability density estimation are given in Section 8.

2. A Beta Kernel Estimator

This paper concentrates on the properties of beta kernel smoothers of Gasser-Müller type for regression curves. Extensions to density estimation are presented briefly in Section 8, and more details are available in Chen (1999).

Suppose n observations $y_1, \ldots, y_n$, are the responses at design points $x_1, \ldots,$ x_n, within $[0, 1]$ from the model

$$y_i = m(x_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where $m(x)$ is an unknown function defined in $x \in [0, 1]$ and the residuals ϵ_i are uncorrelated random variables with zero mean and variance $\sigma^2(x_i)$. The design points are generated by a probability density function f and are ordered such that $0 \leq x_1 \leq \cdots \leq x_n \leq 1$. See Müller (1984) for details on f.

Let $K_{\alpha,\beta}$ denote the density function of a Beta(α,β) random variable. We propose using

$$K_{x/b+1,(1-x)/b+1}(t) = \frac{t^{x/b}(1-t)^{(1-x)/b}\ I(0 \leq t \leq 1)}{B\{x/b+1,(1-x)/b+1\}}$$

as the kernel to smooth at x, where B is the beta function and b is a smoothing parameter satisfying $b \to 0$ as $n \to \infty$. As shown in Figure 1, the shape of kernel $K_{x/b+1,(1-x)/b+1}$ varies according to the value of x. It is symmetric at $x = 0.5$, and becomes asymmetric as x moves towards the boundary points 0 or 1. One important feature of the beta kernels is that they assign no weight outside the interval $[0,1]$.

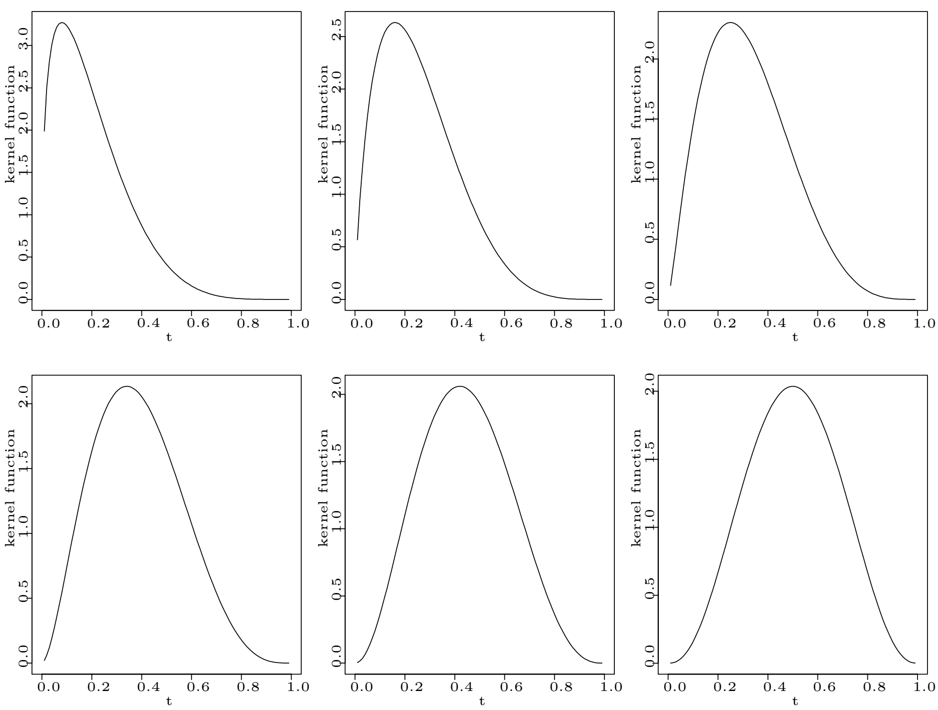

Figure 1. Beta kernels $K_{x/b+1,(1-x)/b+1}(t)$ with $b = 0.2$.

The first beta kernel estimator for m at $x \in [0,1]$ is

$$\hat{m}_1(x) = \sum_{i=1}^{n} y_i \int_{s_{i-1}}^{s_i} K_{x/b+1,(1-x)/b+1}(t)dt, \tag{2.2}$$

where $s_i = \frac{1}{2}(x_i + x_{i+1})$ for $i = 1, \ldots, n-1$, $s_0 = 0$ and $s_n = 1$. Estimator (2.2) is a Gasser-Müller type estimator (Gasser and Müller (1979)) but replaces a fixed symmetric kernel by a beta kernel.

Even though the beta smoother is defined for $x_i \in [0,1]$, it can be extended to situations where the x_i are confined in an interval $[w_1, w_2]$ by linearly transforming the design points to $[0,1]$.

We have been calling b the smoothing bandwidth. It should be pointed out that the real smoothing bandwidth used at x is approximately $x\,(1-x)\,b$, as that is approximately the variance of a Beta$\{x/b+1,(1-x)/b+1\}$ distribution. So the beta kernel smoothing changes not only the shape of the kernel but also implicitly changes the amount of smoothing according to the position where the smoothing is made.

To convey results on the beta smoother, we assume throughout the paper that

(i) $m^{(2)} \in C[0,1]$, $f(\cdot)$ and $\sigma^2(\cdot)$ obey a first order Lipschitz condition in [0, 1];
(ii) $f(x) \geq f_c > 0$ and $\sigma^2(x) \leq \sigma_c^2$ for all $x \in [0,1]$;
(iii) $b \to 0$ and $nb^2 \to \infty$ as $n \to \infty$;
(iv) the design points are fixed. (2.3)

3. Local Properties

Here we study the bias and variance of $\hat{m}_1(x)$ at any fixed $x \in [0,1]$. The bias is

$$Bias\{\hat{m}_1(x)\} = E\{\hat{m}_1(x)\} - m(x) = I_{n1}(x) + I_{n2}(x),$$

where

$$\begin{aligned} I_{n1}(x) &= E\{\hat{m}_1(x)\} - \int_0^1 m(t) K_{x/b+1,(1-x)/b+1}(t)dt \\ &= \sum_{i=1}^n \int_{s_{i-1}}^{s_i} \{m(x_i) - m(t)\} K_{x/b+1,(1-x)/b+1}(t)dt, \\ I_{n2}(x) &= \int_0^1 m(t) K_{x/b+1,(1-x)/b+1}(t)dt - m(x) = E\{m(\xi_x)\} - m(x), \end{aligned}$$

and ξ_x is a Beta$\{x/b+1,(1-x)/b+1\}$ random variable. It is easy to show that $I_{n1}(x) = O(n^{-1})$. A derivation, deferred until Appendix 1, shows that under the assumptions in (2.3),

$$I_{n2}(x) = b(1-2x)m^{(1)}(x) + \tfrac{1}{2}bx(1-x)m^{(2)}(x) + o(b) \tag{3.1}$$

uniformly for all $x \in [0,1]$. Then

$$Bias\{\hat{m}_1(x)\} = b(1-2x)m^{(1)}(x) + \tfrac{1}{2}bx(1-x)m^{(2)}(x) + o(b) + O(n^{-1}) \tag{3.2}$$

uniformly for all $x \in [0,1]$. Thus, the bias is of the same order of magnitude for all $x \in [0,1]$. Notice that $b \approx h^2$, i.e., b and h^2 are of the same order, where h is a smoothing bandwidth used by other kernel smoothers.

The $b(1-2x)m^{(1)}(x)$ term will disappear from (3.2) if $K_{x/b,(1-x)/b}$, rather than $K_{x/b+1,(1-x)/b+1}$, is used as the kernel. The reason for using $K_{x/b+1,(1-x)/b+1}$ is that it is bounded in $[0,1]$. We shall return to this point when we propose a modified beta kernel smoother in Section 5.

The variance of $\hat{m}_1(x)$ is

$$\mathrm{Var}\,\{\hat{m}_1(x)\} = \sum_{i=1}^{n} \sigma^2(x_i)\{\int_{s_{i-1}}^{s_i} K_{x/b+1,(1-x)/b+1}(t)dt\}^2. \tag{3.3}$$

A derivation, deferred until Appendix 2, shows that under assumptions in (2.3),

$$\sum_{i=1}^{n} \sigma^2(x_i)\{\int_{s_{i-1}}^{s_i} K_{x/b+1,(1-x)/b+1}(t)dt\}^2 = n^{-1}A_b(x)\{\sigma^2(x)f^{-1}(x) + O(b+n^{-1})\} \tag{3.4}$$

uniformly for $x \in [0,1]$, where

$$A_b(x) = \frac{B\{2x/b+1, 2(1-x)/b+1\}}{B^2\{x/b+1, (1-x)/b+1\}}.$$

A derivation, deferred until Appendix 3, shows that for small b,

$$A_b(x) \le \frac{b\,(b^{-1}+1)^{3/2}}{2\sqrt{\pi}\sqrt{x(1-x)}} \quad \text{for any} \quad x \in [0,1] \tag{3.5}$$

and

$$A_b(x) \sim \begin{cases} \frac{b^{-1/2}}{2\sqrt{\pi}\sqrt{x(1-x)}} & \text{if } x/b \text{ and } (1-x)/b \to \infty \\ \frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} b^{-1} & \text{if } x/b \text{ or } (1-x)/b \to \kappa, \end{cases}$$

where κ is a nonnegative constant. Thus,

$$\mathrm{Var}\,\{\hat{m}_1(x)\} = \begin{cases} \frac{\sigma^2(x)n^{-1}\ b^{-1/2}}{2\sqrt{\pi}\ \sqrt{x(1-x)}\ f(x)}\ \{1+O(n^{-1})\} & \text{if } x/b \text{ and } (1-x)/b \to \infty \\ \frac{\sigma^2(x)\Gamma(2\kappa+1)n^{-1}b^{-1}}{2^{2\kappa+1}\Gamma^2(\kappa+1)f(x)}\{1+O(n^{-1})\} & \text{if } x/b \text{ or } (1-x)/b \to \kappa. \end{cases} \tag{3.6}$$

In particular,

$$\mathrm{Var}\,\{\hat{m}_1(0)\} = \frac{\sigma^2(0)n^{-1}b^{-1}}{2f(0)}\{1+O(n^{-1})\}.$$

The results in (3.6) show that the variance is of $O(n^{-1}b^{-1/2})$ within the interior of $[0,1]$, but is of a larger order, $O(n^{-1}b^{-1})$, in areas near the boundaries.

However, since $b \approx h^2$, the size of the areas is of order b and is one order of magnitude smaller than h, the size of boundary bias areas. This makes the contribution of the variance in the boundary areas to the mean integrated squared error negligible, as shown in the next section.

4. Global Properties

From (3.2) the integrated squared bias of $\hat{m}_1$ is

$$b^2 \int_0^1 \{(1-2x)m^{(1)}(x) + \tfrac{1}{2}x(1-x)m^{(2)}(x)\}^2 dx + o(b^2 + n^{-1}).$$

Although Var $\{\hat{m}_1(x)\}$ is $O(n^{-1}b^{-1})$ in areas near the boundaries, the integrated variance is still $O(n^{-1}b^{-1/2})$. To appreciate why, we note that (3.4) and the upper bound in (3.5) mean that

$$\begin{aligned}\int_0^1 \text{Var}\,\{\hat{m}_1(x)\}dx &\le \tfrac{1}{2\sqrt{\pi}} n^{-1} b(b^{-1}+1)^{3/2} \textstyle\int_0^1 \sigma^2(x) f^{-1}(x)\{x(1-x)\}^{-1/2} dx \\ &\le \tfrac{\sqrt{\pi}}{2}\sigma_c^2 f_c^{-1} n^{-1} b(b^{-1}+1)^{3/2},\end{aligned}$$

where σ_c^2 and f_c are the upper and lower bound of $\sigma^2(x)$ and $f(x)$ respectively, as assumed in (2.3).

In fact (3.6) means that for any $\delta = b^{1-\epsilon}$ where $0 < \epsilon < 1/2$,

$$\begin{aligned}\int_0^1 \text{Var}\,\{\hat{m}_1(x)\}dx &= \int_0^\delta + \int_\delta^{1-\delta} + \int_{1-\delta}^1 \text{Var}\,\{\hat{m}_1(x)\}dx \\ &= \int_\delta^{1-\delta} \text{Var}\,\{\hat{m}_1(x)\}dx + O(n^{-1}b^{-\epsilon}) \\ &= \tfrac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} \textstyle\int_0^1 \frac{\sigma^2(x)}{f(x)\sqrt{x(1-x)}} dx + o(n^{-1}b^{-1/2}).\end{aligned}$$

So the mean integrated squared error is

$$\begin{aligned}MISE\{\hat{m}_1(x)\} &= b^2 \int_0^1 \{(1-2x)m^{(1)}(x) + \tfrac{1}{2}x(1-x)m^{(2)}(x)\}^2 dx \\ &\quad + \tfrac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} \textstyle\int_0^1 \frac{\sigma^2(x)}{f(x)\sqrt{x(1-x)}} dx + o(b^2 + n^{-1}b^{-1/2}). \quad (4.1)\end{aligned}$$

The bandwidth which minimises the leading terms in (4.1) is

$$\begin{aligned}b_1^\star = \Big[\int_0^1 \frac{2^{-3}\pi^{-1/2}\sigma^2(x)}{f(x)\sqrt{x(1-x)}} dx\Big]^{2/5} \Big[\int_0^1 \{(1-2x)m^{(1)}(x) \\ + \tfrac{1}{2}x(1-x)m^{(2)}(x)\}^2 dx\Big]^{-2/5} n^{-2/5}, \quad (4.2)\end{aligned}$$

which is of order $n^{-2/5}$. Thus, $b_1^\star \approx h^2$ where h is the optimal bandwidth used by other kernel smoothers.

When the variance is a constant function and the design points are uniformly distributed (that is, $\sigma^2(x) = \sigma^2$ and $f(x) = 1$),

$$MISE\{\hat{m}_1(x)\} = b^2 \int_0^1 \{(1-2x)m^{(1)}(x) + \tfrac{1}{2}x(1-x)m^{(2)}(x)\}^2 dx \\ + \tfrac{\sqrt{\pi}}{2} n^{-1}b^{-1/2} + o(b^2 + n^{-1}b^{-1/2})$$

and

$$b_1^\star = 2^{-6/5}\pi^{1/5}\,\sigma^{4/5}\left[\int_0^1 \{(1-2x)m^{(1)}(x) + \tfrac{1}{2}x(1-x)m^{(2)}(x)\}^2 dx\right]^{-2/5} n^{-2/5}.$$

5. A Modified Estimator

Equations (4.1) and (4.2) imply that the beta smoother $\hat{m}_1$ achieves the $O(n^{-4/5})$ optimal rate of convergence for mean integrated squared error. It also has a simple form. However, we see from (3.2) and (4.1) that $m^{(1)}$ appears in the bias and ultimately in the mean integrated squared error. This makes the mean integrated squared error complicated and a comparison with other kernel smoothers, especially the local linear smoother, difficult.

One way to remove $m^{(1)}$ is to modify the beta kernels. The appearance of $m^{(1)}$ in the bias of $\hat{m}_1$ arises because x is the mode and not the mean of the beta kernel $K_{x/b+1,(1-x)/b+1}$. However, x is the mean of $K_{x/b,(1-x)/b}$. This leads us to propose the following modified beta kernels

$$K^\star_{x,b}(t) = \begin{cases} K_{x/b,(1-x)/b}, & \text{if } x \in (2b, 1-2b)\ , \\ K_{\rho(x,b),(1-x)/b}(t), & \text{if } x \in [0, 2b], \\ K_{x/b,\rho(1-x,b)}(t), & \text{if } x \in [1-2b, 1], \end{cases} \tag{5.1}$$

where $\rho(x,b) = 2b^2 + 2.5 - \sqrt{4b^4 + 6b^2 + 2.25 - x^2 - x/b}$. Note that $\rho(x,b)$ is monotonic increasing in $[0,2b]$, $\rho(0,b) = 1$, and $y = x/b$ serves as a tangent line at $x = 2b$. The arrangement for x in the boundary areas is to keep the kernels bounded.

The modified beta kernel estimator is

$$\hat{m}_2(x) = \sum_{i=1}^n y_i \int_{s_{i-1}}^{s_i} K^\star_{x,b}(t)dt.$$

By slightly modifying the derivation in Appendix 1, it can be shown that

$$Bias\{\hat{m}_2(x)\} = \begin{cases} \tfrac{1}{2}x(1-x)m^{(2)}(x)b + O(b^2), & \text{if } x \in (b, 1-b)\ , \\ \xi(x,b)m^{(1)}(x)b + O(b^2), & \text{if } x \in [0, 2b], \\ -\xi(1-x,b)m^{(1)}(x)b + O(b^2), & \text{if } x \in [1-2b, 1], \end{cases}$$

where $\xi(x,b)=\{(1-x)\{\rho(x,b)-x/b\}/\{1+b\rho(x,b)-x\}$ is a bounded function for $x\in[0,2b]$. The integrated square bias is

$$\tfrac{1}{4}b^2\int_0^1\{x(1-x)m^{(2)}(x)\}^2dx+o(b^2+n^{-1}).$$

As promised, $m^{(1)}$ has disappeared. As x is still not the mean of the boundary beta kernels, $m^{(1)}$ still appears but $m^{(2)}$ disappears in the point-wise bias when x is close to 0 or 1.

A derivation like that at (3.6) shows that, for any finite κ, when $b\to 0$,

$$\mathrm{Var}\,\{\hat m_2(x)\}=\begin{cases}\frac{\sigma(x)^2n^{-1}\,b^{-1/2}}{2\sqrt{\pi}\,\sqrt{x(1-x)}\,f(x)}\{1+O(n^{-1})\}, & \text{if } x/b\text{and } (1-x)/b\to\infty,\\ \frac{1}{2}\sigma^2(x)(1-x)n^{-1}b^{-1}f^{-1}(x)+O(n^{-1}), & \text{if } x/b\to\kappa\ ,\\ \frac{1}{2}\sigma^2(x)xn^{-1}b^{-1}f^{-1}(x)+O(n^{-1}), & \text{if } (1-x)/b\to\kappa.\end{cases}$$

The mean integrated squared error is

$$MISE\{\hat m_2(x)\}=b^2\int_0^1\{\tfrac{1}{2}x(1-x)m^{(2)}(x)\}^2dx+\tfrac{1}{2\sqrt{\pi}}n^{-1}b^{-1/2}\int_0^1\frac{\sigma^2(x)}{f(x)\sqrt{x(1-x)}}dx$$
$$+o(b^2+b^{-1/2}n^{-1}). \qquad (5.2)$$

The optimal bandwidth is

$$b_2^\star=2^{-2/5}\,\pi^{-1/5}\left[\int_0^1\frac{\sigma^2(x)}{f(x)\sqrt{x(1-x)}}dx\right]^{2/5}\left[\int_0^1\{x(1-x)m^{(2)}(x)\}^2dx\right]^{-2/5}n^{-2/5}. \qquad (5.3)$$

Substituting the optimal bandwidth in (5.3) into (5.2), we have the optimal mean integrated squared error

$$MISE^\star=\tfrac{5}{4}\left(\tfrac{1}{2\sqrt{\pi}}\right)^{4/5}\left[\int_0^1\frac{\sigma^2(x)}{f(x)\sqrt{x(1-x)}}dx\right]^{4/5}\left[\int_0^1x^2(1-x)^2\{m^{(2)}(x)\}^2dx\right]^{1/5}n^{-4/5}$$
$$+o(n^{-4/5}). \qquad (5.4)$$

We cannot use (5.3) to choose b in practice because of the unknown derivative of m. Cross validation, the details of which are available in Müller (1988) and Härdle (1990), can be used to choose the bandwidth.

Comparing the two beta estimators, we find $\hat m_2$ has simpler forms for its bias and mean integrated squared error while the definition of $\hat m_1$ is simpler. Both estimators have the same optimal mean integrated squared error of $O(n^{-4/5})$. However, for any function m, if both $\int_0^1\{m^{(1)}(x)\}^2dx$ and $\int_0^1\{m^{(2)}(x)\}^2dx$ are finite it may be shown, using integration by parts, that

$$\int_0^1\left\{(1-2x)m^{(1)}(x)+\tfrac{1}{2}x(1-x)m^{(2)}(x)\right\}^2dx\ge\int_0^1\{\tfrac{1}{2}x(1-x)m^{(2)}(x)\}^2dx. \qquad (5.5)$$

This means that $\hat{m}_2$ has smaller integrated bias and mean integrated error than $\hat{m}_1$, and is thus recommeded.

6. Comparison with the Local Linear Smoother

In this section we compare the performance of the beta kernel smoother $\hat{m}_2$ with the local linear smoother in terms of mean squared error and mean integrated squared error respectively. The local linear smoother denoted as $\hat{m}_{ll}$ is the one considered in Fan (1993), with a kernel K which is a density function itself. To make the two estimators comparable we transform b to h, the bandwidth of the local linear smoother, using $b = h^2$.

We first compare at x in the interior of $[0, 1]$, where both x/b and $(1-x)/b \to \infty$. In this case the optimal mean squared errors are, respectively,

$$MSE^\star\{\hat{m}_2(x)\} = \tfrac{5}{4}\{\tfrac{\sigma^2(x)}{\sqrt{4\pi}f(x)}\}^{4/5}\{m^{(2)}(x)\}^{2/5}n^{-4/5} \quad \text{and}$$
$$MSE^\star\{\hat{m}_{ll}(x)\} = \tfrac{5}{4}\{\tfrac{\sigma^2(x)R(K)}{f(x)}\}^{4/5}\{\sigma_K^2 m^{(2)}(x)\}^{2/5}n^{-4/5},$$

where $\sigma_K^2 = \int u^2 K(u)du$ and $R(K) = \int K^2(u)du$.

The relative efficiency of $\hat{m}_2$ to $\hat{m}_{ll}$ is

$$\frac{MSE^\star\{\hat{m}_{ll}(x)\}}{MSE^\star\{\hat{m}_2(x)\}} = \{\sqrt{4\pi}R(K)\sigma_K\}^{4/5}.$$

This ratio is 1 if K is the Gaussian kernel and 0.951 if K is the optimal Epanechnikov kernel. Thus, both smoothers have almost the same performance in the interior.

In the boundary area, the variance of $\hat{m}_2(x)$ is of order $n^{-1}h^{-2}$ whereas the variance of $\hat{m}_{ll}$ has an increased coefficient while maintaining its order of $n^{-1}h^{-1}$, as reported by Fan and Gijbels (1992). We plot in Figure 2 the variance coefficient functions for both the beta and the local linear smoothers for four levels of bandwidths. The variance coefficient function for the beta smoother is $V_b(x) = \sqrt{b}A_b(x)$, whereas that for the local linear smoother is $V_{ll}(x) = v(x)$, defined in Fan and Gijbels (1992, p.2015) using the Epanechnikov kernel. The reason for using $V_b(x)$ and $V_{ll}(x)$ is that they are the different coefficients of the dominant variance terms, apart from common factors $\sigma^2(x)f^{-1}(x)n^{-1}$. To make the amount of smoothing used by the two smoothers of the same scale we let $h = b^{1/2}$. We see that, when the bandwidth is at a higher level in plots (1), (2) and (3), the beta smoother has smaller variance coefficients. It is only when b is less than 0.00613 that the beta smoother starts to have larger variance at the boundaries. The sample size corresponding to a b of 0.00613 depends on $m(\cdot)$, $\sigma^2(\cdot)$ and $f(\cdot)$. If $m(x) = \exp(-ax^2)$, $\sigma(x) \equiv 0.05$ and $f(x) = 1$, the sample size is $n = 140$ for $a = 4$, and $n = 400$ for $a = 1$. If the level of noise is increased from $\sigma = 0.05$ to $\sigma = 0.1$, the sample sizes will be approximately 1140 and 1600 respectively. Therefore, in general terms, the beta smoother has a worse variance near the boundaries only when the sample size is very large.

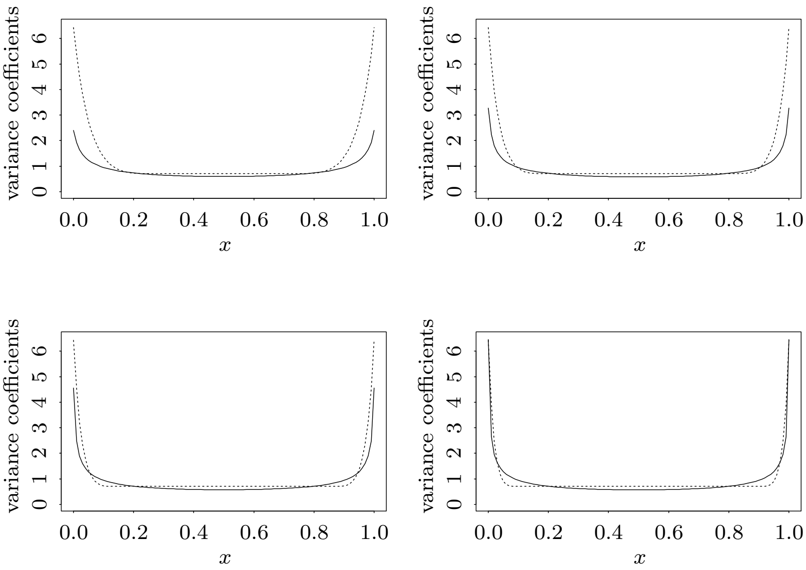

Figure 2. Variance coefficient functions for the beta kernel (solid lines) and the local linear (dashed lines) smoothers.

Next we compare mean integrated squared errors of the two smoothers. From (5.4) and the similar expression for $\hat{m}_{ll}$ in Fan (1993), we have

$$r(\hat{m}_2, \hat{m}_{ll}) = \frac{MISE^{\star}\{\hat{m}_2(x)\}}{MISE^{\star}\{\hat{m}_{ll}(x)\}}$$
$$= \Big[\frac{(\sqrt{4\pi})^{-1}\int_0^1 \frac{\sigma^2(x)}{f(x)\sqrt{x(1-x)}}dx}{R(K)\sigma_K\int_0^1 \frac{\sigma^2(x)}{f(x)}dx}\Big]^{4/5}\Big[\frac{\int_0^1 x^2(1-x)^2\{m^{(2)}(x)\}^2dx}{\int_0^1\{m^{(2)}(x)\}^2dx}\Big]^{1/5}.$$

The value of this ratio is quite difficult to assess. It is easier to assume $f(t) = 1$ and $\sigma^2(x)$ is a constant function, which means equally spaced design points. In this case,

$$r(\hat{m}_2, \hat{m}_{ll}) = \{\frac{\sqrt{\pi}}{2R(K)\sigma_K}\}^{4/5}\Big[\frac{\int_0^1 x^2(1-x)^2\{m^{(2)}(x)\}^2dx}{\int_0^1\{m^{(2)}(x)\}^2dx}\Big]^{1/5}. \qquad (6.1)$$

The first factor on the right is $\pi^{4/5}$ for the the Epanechnikov kernel and is close to $\pi^{4/5}$ for other standard kernels, including the biweight and the Gaussian kernels.

The second factor is less than 1 depending on $m^{(2)}$. There are situations when $r(\hat{m}_2, \hat{m}_{ll}) < 1$, which means that $\hat{m}_2$ has a smaller mean integrated squared error than $\hat{m}_{ll}$. Convenient examples are $m(x) = Cx^p$ or $m(x) = C(1-x)^p$ for $p \geq 7$, where C is any constant.

7. Empirical Results

The beta kernel smoothers use only non-negative weights. A good feature of using non-negative weights is that the curve estimates respect the range of the data. Let $m(x)$ be an unknown probability function and Y_i be binary 0 and 1 response variables. The beta kernel estimates for $m(x)$ then always lie in $[0,1]$ whereas the local linear smoother may give values outside the range, so that the logistic transform is needed to make it range respecting. In Figure 3, we display five estimated curves by applying $\hat{m}_1$ and local linear smoothers to five simulated binary samples of size 40 generated from $\text{Bin}\{1, m(x_i)\}$ for equally spaced design points in [0,1] and $m(x) = \exp(-4x^2)$. The bandwidths used were $b = 0.089$ and $h = 0.254$; both were global optimum. We see two of the local linear estimates were outside $[0,1]$ near $x = 0$ and $x = 1$ respectively. The beta estimates were all within $[0,1]$. Apart from that, the two sets of estimates were quite similar. We notice that the variation among the estimates was small near the boundaries and large in the middle.

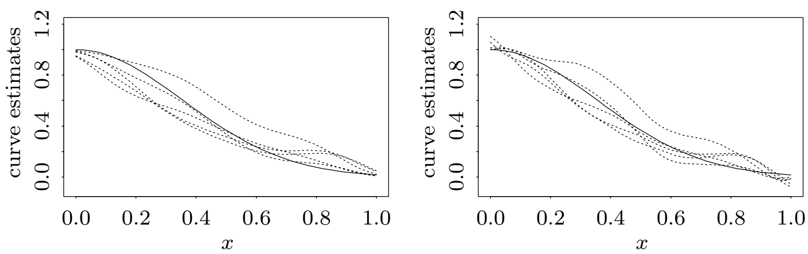

Figure 3. Five estimated curves for simulated $Y_i \sim Bin\{1, m(x_i)\}$ where $m(x) = \exp(-4x^2)$.

Next we present results from simulation studies designed to investigate the performance of the proposed beta kernel estimators and compare them with the local linear smoother. The regression curves considered are

$$m(x) = x^p, \tag{7.1}$$

where $p = 4$ and $p = 10$ respectively. It can easily be shown from the comparison made in Section 7 that the local linear smoother has a smaller optimal mean integrated squared error than $\hat{m}_2$ when $p = 4$, while the reverse holds when $p = 10$. The aim of the simulation study is to verify these things. We would also like to observe the performance of $\hat{m}_1$.

The fixed design points, x_i, are taken at equally spaced points in $[0, 1]$ and the ϵ_i are uncorrelated normal random variables with zero mean and standard deviation $\sigma = 0.05$. The biweight kernel

$$K(u) = \frac{15}{16}(1-u^2)^2 I(|u| < 1)$$

is used by the local linear smoother.

Table 1 contains the optimal average integrated squared errors and their standard errors for the estimators based on 1000 simulations. For each simulation, optimal smoothing bandwidths, which minimize the integrated squared errors by the golden section search algorithm described in Press, Flannery, Teukolsky and Vetterling (1992), are used. The optimal integrated squared errors are determined by using the optimal bandwidths. To confirm the theoretical expansions for the mean integrated squared errors, theoretical optimal mean integrated squared errors derived from (4.1) and (4.2), (5.4) and from Fan (1993) are also given for the three smoothers.

Table 1. Simulated optimal average integrated squared errors and their standard errors for beta and local linear smoothers with $\sigma = 0.05$. The columns headed "predic." and "real" give 10^3 times the optimal mean square errors based on the theoretical expansion and the direct mean square error calculation respectively. The standard errors, multiplied by 10^3, are given in parentheses.

(a) $m(x) = x^{10}$

n	$\hat{m}_1$		$\hat{m}_2$		$\hat{m}_{ll}$	
	predic.	real	predic.	real	predic.	real
20	1.692	0.953 (0.45)	0.986	0.934 (0.45)	1.166	1.027 (0.48)
40	0.972	0.668 (0.29)	0.566	0.564 (0.24)	0.670	0.603 (0.26)
60	0.703	0.500 (0.19)	0.410	0.414 (0.17)	0.484	0.435 (0.18)
100	0.467	0.343 (0.13)	0.272	0.278 (0.11)	0.322	0.292(0.17)
140	0.357	0.261 (0.09)	0.208	0.209 (0.08)	0.246	0.222 (0.09)
180	0.292	0.222 (0.08)	0.170	0.175 (0.06)	0.201	0.185 (0.07)

(b) $m(x) = x^4$

n	$\hat{m}_1$		$\hat{m}_2$		$\hat{m}_{ll}$	
	predic.	real	predic.	real	predic.	real
20	1.143	0.904 (0.46)	0.912	0.725 (0.39)	0.737	0.665 (0.39)
40	0.657	0.539 (0.25)	0.524	0.452 (0.20)	0.423	0.373 (0.21)
60	0.475	0.409 (0.18)	0.379	0.329 (0.15)	0.306	0.281 (0.15)
100	0.316	0.287 (0.11)	0.252	0.219 (0.10)	0.203	0.187 (0.09)
140	0.241	0.218 (0.09)	0.192	0.169 (0.07)	0.155	0.143 (0.07)
180	0.197	0.165 (0.06)	0.157	0.136 (0.06)	0.127	0.116 (0.06)

The results in Table 1 show that $\hat{m}_2$ had smaller simulated average squared errors than $\hat{m}_{ll}$ when $m(x) = x^{10}$ for all the sample sizes considered, while the reverse holds when $m(x) = x^4$. This confirms the theoretical comparison given in Section 6. The simulated average squared errors of the three smoothers appeared to converge to the theoretical mean integrated squared errors as n increased, which confirmed the expansions developed in Sections 4 and 5. The first beta kernel smoother $\hat{m}_1$ had larger simulated squared errors than $\hat{m}_2$ in both tables. However, this was not surprising in view of (5.5).

8. Extensions to Density Estimation

Estimators for density functions using the beta kernels are given in this section. Let $X_1, \ldots, X_n$ be a random sample from a distribution with an unknown probability density function f having compact support $[0, 1]$. We assume $f^{(2)} \in C[0, 1]$ and the conditions (ii) and (iii) in (2.3).

The analogues of $\hat{m}_1$ and $\hat{m}_2$ in density estimation are

$$\hat{f}_1(x) = n^{-1} \sum_{i=1}^{n} K_{x/b+1,(1-x)/b+1}(X_i) \quad \text{and} \quad \hat{f}_2(x) = n^{-1} \sum_{I=1}^{n} K^{\star}_{x,b}(X_i).$$

Similar to (3.2) and the bias of $\hat{m}_2$ given in Section 5, we have

$$Bias\{\hat{f}_1(x)\} = \{(1 - 2x) f^{(1)}(x) + \tfrac{1}{2} x(1 - x) f^{(2)}(x)\} b + O(b^2) \quad \text{and}$$

$$Bias\{\hat{f}_2(x)\} = \begin{cases} \frac{1}{2} x(1 - x) f^{(2)}(x) b + O(b^2), & \text{if } x \in [b, 1 - 2b], \\ \xi(x) b f^{(1)}(x) + O(b^2), & \text{if } x \in [0, 2b), \\ -\xi(1 - x) b f^{(1)}(x) + O(b^2), & \text{if } x \in (1 - 2b, 1], \end{cases}$$

where the remainder terms are uniformly $O(b^2)$ for $x \in [0, 1]$. The biases are of $O(b)$ throughout $[0, 1]$, indicating that the estimators are free of boundary bias.

Using a method similar to, but slightly easier than, that for deriving (3.6) we have

$$\text{Var}\,\{\hat{f}_1(x)\} = \begin{cases} \frac{1}{2\sqrt{\pi}} \frac{n^{-1} b^{-1/2}}{\{x(1-x)\}^{1/2}} \{f(x) + O(n^{-1})\} & \text{if } x/b \text{ and } (1 - x)/b \to \infty; \\ \frac{\Gamma(2\kappa+1)}{2^{2\kappa+1} \Gamma^2(\kappa+1)} n^{-1} b^{-1} \{f(x) + O(n^{-1})\} & \text{if } x/b \text{ or } (1 - x)/b \to \kappa. \end{cases} \tag{8.1}$$

The variance of $\hat{f}_2$ can be worked out in a similar fashion. The only difference is that the multiplier in front of $n^{-1} b^{-1}$ in the case that x/b or $(1 - x)/b \to \kappa$ has a slightly different form. The mean integrated square errors and the optimal bandwidths for both estimators can be obtained in a similar fashion to those of the regression estimators. Simulation studies which compare the performance of the beta density estimators with some of the local linear density estimators are reported in Chen (1999).

Acknowledgements

The author thanks Professor Bruce M. Brown for stimulating discussions, and Dr. Ming Yen Cheng and two referees for beneficial comments.

Appendix

Derivation of (3.1).

Let μ_{ξ_x} and $\sigma^2_{\xi_x}$ be the mean and the variance of ξ_x respectively. From Johnson, Kotz and Balakrishnan (1994) it may be shown that there exists a constant M such that

$$\mu_{\xi_x} = x + b(1-2x) + \Delta_1(x) \quad \text{and} \tag{A.1}$$
$$\sigma^2_{\xi_x} = bx(1-x) + \Delta_2(x), \tag{A.2}$$

where $\Delta_j(x) \le Mb^2$ for $j = 1$ and 2. Hence, the remainder terms in the above expansions for μ_{ξ_x} and $\sigma^2_{\xi_x}$ are uniformly $O(b^2)$.

A Taylor expansion gives

$$m(\xi_x) = m(x) + m^{(1)}(x)(\xi_x - x) + \tfrac{1}{2}m^{(2)}(x)(\xi_x - x)^2 + r(\xi_x - x), \tag{A.3}$$

where the remainder term

$$r(\xi_x - x) = \int_0^{\xi_x - x} (\xi_x - x - t)\{m^{(2)}(x+t) - m^{(2)}(x)\}dt.$$

Let g be the density of $(\xi_x - x)/\sqrt{b}$. Then

$$\begin{aligned} E\{r(\xi_x - x)\} &= \int g(y)dy \int_0^{\sqrt{b}y} (\sqrt{b}y - t)\{m^{(2)}(x+t) - m^{(2)}(x)\}dt \\ &= b \int g(y)dy \int_0^{y} (y - t)\{m^{(2)}(x+\sqrt{b}t) - m^{(2)}(x)\}dt. \end{aligned}$$

As $m^{(2)}$ is uniformly continuous in $[0,1]$, by the Dominated Convergence Theorem the integral on the right hand side converges uniformly to zero.

Then taking the expectation of both sides of (A.3), and from (A.1) and (A.2), we have

$$\sup_{x\in[0,1]} \left| E\{m(\xi_x)\} - b(1-2x)m^{(1)}(x) - \tfrac{1}{2}bx(1-x)m^{(2)}(x) \right| \le \epsilon_1 b,$$

where $\epsilon_1 \to 0$ uniformly for all $x \in [0,1]$ as $n \to \infty$. Thus, we obtain (3.1).

Derivation of (3.5).

Let $R(z) = \sqrt{2\pi}e^{-z}z^{z+1/2}/\Gamma(z+1)$ for $z \ge 0$. According to Lemma 3 of Brown and Chen (1998), $R(z)$ converges to 1 as $z \to \infty$ and $R(z) < 1$ for any

$z > 0$. Taking the derivative on $R(z)$ and using a well known expansion of $\Gamma^{(1)}(z+1)/\Gamma(z+1)$, we can prove that $R(z)$ is monotonic increasing.

Write $A_b(x)$ in terms of Gamma functions,

$$A_b(x) = \frac{\Gamma(2x/b+1)\Gamma\{2(1-x)/b+1\}\Gamma^2(1/b+2)}{\Gamma^2(x/b+1)\Gamma^2\{(1-x)/b+1\}\Gamma(2/b+2)}. \tag{A.4}$$

Expressing the gamma functions in terms of their Stirling's formulae and R-functions, we have

$$A_b(x) = \frac{e^{-1}}{\sqrt{4\pi}} \frac{b(b^{-1}+1)^{3/2}\{1+(2b^{-1}+5/2)^{-1}\}^{2/b+3}}{\sqrt{x(1-x)}} S(b,x), \tag{A.5}$$

where

$$S(b,x) = \frac{R^2(x/b)R^2\{(1-x)/b\}R(2/b+1)}{R(2x/b)R\{2(1-x)/b\}R^2(1/b+1)}. \tag{A.6}$$

As R is monotonic increasing and $R(z) < 1$ for all $z \geq 0$,

$$S(b,x) \leq R(2/b+1)/R^2(1/b+1) \to 1 \quad \text{as} \quad b \to 0. \tag{A.7}$$

The upper bound for $A_b(x)$ can be obtained from (A.5), (A.7) and the fact that $\{1+(2b^{-1}+5/2)^{-1}\}^{2/b+3} \to e$ as $b \to 0$.

Next we establish the convergence results. If x/b and $(1-x)/b \to \infty$, $S(b,x) \to 1$ because each R-function in (A.6) converges to 1 as $b \to 0$. Thus, $A_b(x)$ converges to its upper bound.

To prove the case when $x/b \to \kappa$, where κ is a nonnegative constant, we notice from (A.4) that

$$\begin{aligned} A_b(x) &\sim \frac{\Gamma(2\kappa+1)R^2\{(1-x)/b\}\,R(2/b+1)b^{-1/2}(b^{-1}+1)^{1/2}(1+b)^{2/b+3/2}}{\Gamma^2(\kappa+1)R\{2(1-x)/b\}R^2(1/b+1)2^{1+2\kappa}e\,(1-x)^{1/2}(1+b/2)^{2/b+3/2}} \\ &\sim \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)} b^{-1} \end{aligned}$$

as $x \to 0$ when $x/b \to \kappa$. The proof for the case $(1-x)/b \to 0$ is similar.

Derivation of (3.4).

Define $J_n = \sum_{i=1}^n \sigma^2(x_i) w_i^2(x)$ where $w_i(x) = \int_{s_{i-1}}^{s_i} K_{x/b+1,(1-x)/b+1}(t)dt$. According to the Mean Value Theorem, there exist t_i and θ_i in $[s_{i-1}, s_i]$ such that

$$J_n = \sum_{i=1}^n \sigma^2(x_i)(s_i - s_{i-1})^2 K^2_{x/b+1,(1-x)/b+1}(t_i).$$

From Müller (1988, p.27f) or Jennen-Steinmetz and Gasser (1988), $s_i - s_{i-1} = \{nf(\xi_i)\}^{-1} + O(n^{-2})$ for some $\xi_i \in [x_{i-1}, x_i]$ and $\sigma^2(x_i) = \sigma^2(t_i) + O(n^{-1})$ under the assumptions (i) and (ii) in (2.3). As f obeys the Lipschitz condition and $s_i - s_{i-1} = O(n^{-1})$, we have $f^{-1}(\xi_i) = f^{-1}(t_i) + O(n^{-1})$. Thus,

$$J_n = \sum_{i=1}^{n}[\sigma^2(t_i)\{nf(t_i)\}^{-1} + O(n^{-2})](s_i - s_{i-1})K^2_{x/b+1,(1-x)/b+1}(t_i)$$
$$= n^{-1}P_{n1} + O(n^{-2}P_{n2}), \tag{A.8}$$

where $P_{n1} = \sum_{i=1}^{n}(s_i - s_{i-1})\sigma^2(t_i)f^{-1}(t_i)K^2_{x/b+1,(1-x)/b+1}(t_i)$ and $P_{n2} = \sum_{i=1}^{n}(s_i - s_{i-1})K^2_{x/b+1,(1-x)/b+1}(t_i)$ are the Riemann sums of $\int_0^1 \sigma^2(t)f^{-1}(t)K^2_{x/b+1,(1-x)/b+1}(t)dt$ and $\int_0^1 f^{-1}(t)K^2_{x/b+1,(1-x)/b+1}(t)dt = A_b(x)$, respectively.

We want to prove that uniformly, as $n \to \infty$,

$$n^{-1}P_{n1} = n^{-1}\int_0^1 \sigma^2(t)f^{-1}(t)K^2_{x/b+1,(1-x)/b+1}(t)dt\{1 + o(1)\} \quad \text{and} \tag{A.9}$$
$$n^{-1}P_{n2} = n^{-1}A_b(x)\{1 + o(1)\}. \tag{A.10}$$

We give a proof only for (A.10) as that for (A.9) is similar. Note that $A_b(x) = n^{-1}\int_0^1 K^2_{x/b+1,(1-x)/b+1}(t)dt$. From the Mean Value Theorem, there exist $\theta_i \in [s_{i-1}, s_i]$ such that

$$|n^{-1}P_{n2} - n^{-1}\int_0^1 K^2_{x/b+1,(1-x)/b+1}(t)dt| = n^{-1}|\sum_{i=1}^{n}(s_i - s_{i-1})\Delta_i(x,b)|,$$

where $\Delta_i(x,b) = K^2_{x/b+1,(1-x)/b+1}(t_i) - K^2_{x/b+1,(1-x)/b+1}(\theta_i)$. Only the proof of (A.10) for $x \le 1/2$ is given as $x > 1/2$ is covered due to the symmetry of the beta kernels with respect to x.

Let $w(x) = K_{x/b+1,(1-x)/b+1}(x)$. As x is the mode of the beta kernel,

$$K_{x/b+1,(1-x)/b+1}(t) \le w(x) \quad \text{for any } t \in [0,1].$$

As $w(x)$ is symmetric about 1/2 and is monotonic decreasing for $x \in [0, 1/2]$,

$$K_{x/b+1,(1-x)/b+1}(t) \le w(0) = b^{-1}(1+b) \quad \text{for any } t, x \in [0,1]. \tag{A.11}$$

A tighter upper bound is

$$K_{x/b+1,(1-x)/b+1}(t) \le C_1 b^{-1/2}\{x(1-x)\}^{-1/2} \tag{A.12}$$

which can be derived using Stirling's formula and properties of the R-function. Throughout the proof C_j denotes some positive constant.

The first case considered is $x \le b^{1-\epsilon}$ for some $0 < \epsilon < 1/2$. Notice that $T_{x,b} = n^{-1}\sum_{i=1}^{n}(s_i - s_{i-1})\Delta_i(x,b)$ can be partitioned into three sums, say $I_{1,x,b}, I_{2,x,b}$

and $I_{3,x,b}$, where for $\delta = 1/2 + \epsilon$,

$$I_{1,x,b} = n^{-1} \sum_{s_i < b^\delta} (s_i - s_{i-1})\Delta_i(x, b),$$
$$I_{2,x,b} = n^{-1} \sum_{b^\delta \le s_{i-1} s_i \le 1-b^\delta} (s_i - s_{i-1})\Delta_i(x, b) \quad \text{and}$$
$$I_{3,x,b} = n^{-1} \sum_{s_{i-1} > 1-b^\delta} (s_i - s_{i-1})\Delta_i(x, b).$$

From (A.11) and the upper bound of $A_b(x)$ given in (3.5), we have for $j = 1$ and $j = 3$,

$$|I_{j,x,b}| \le O(1)n^{-1}b^{\delta-2} \le O(1)n^{-1}b^{-1/2}\{x(1-x)\}^{-1/2}b^{\delta-1/2-\epsilon/2} = o\{n^{-1}A_b(x)\} \tag{A.13}$$

uniformly for any $x \le b^{1-\epsilon}$ as $\delta = 1/2+\epsilon$. Let $v_x(t) = (t/x)^x\{(1-t)/(1-x)\}^{1-x}$. Using a derivation similar to that in (A.12) and the fact that

$$|v_x^{1/b}(t_i) - v_x^{1/b}(\theta_i)| \le C_2 b^{-1}|v_x(t_i) - v_x(\theta_i)|$$

as $v_x(t) \le 1$, we have

$$|K_{x/b+1,(1-x)/b+1}(t_i) - K_{x/b+1,(1-x)/b+1}(\theta_i)| \le C_3 b^{-3/2}\{x(1-x)\}^{-1/2}|v_x(t_i) - v_x(\theta_i)|.$$

The derivative of $v_x(t)$ with respect to t is

$$v_x^{(1)}(t) = \{t/(1-t)\}^x\{(1-x)/x\}^x(x-t)/\{t(1-x)\}.$$

Note that $\{(1-x)/x\}^x \le 1.5$ and $t/(1-t)$ is monotonic increasing in $[0, 1]$. Thus, for $t \in [b^\delta, 1-b^\delta]$ and $x \le b^{1-\epsilon}$,

$$|v_x^{(1)}(t)| \le 3b^{-\delta x}(x/t + 1) \le 3b^{-\delta x+1-\epsilon-\delta} + 3 \le 4$$

as $\lim_{b\to 0} b^{-\delta b^{1-\epsilon}+1/2} = 0$. This means that $v_x(t)$ has bounded first derivative for $t \in [b^\delta, 1-b^\delta]$. Hence,

$$|K_{x/b+1,(1-x)/b+1}(t_i) - K_{x/b+1,(1-x)/b+1}(\theta_i)| \le C_4 b^{-3/2}\{x(1-x)\}^{-1/2}|s_i - s_{i-1}|. \tag{A.14}$$

Using (A.11) and (A.14),

$$|I_{2,x,b}| \le C_5 n^{-2}b^{-5/2}\{x(1-x)\}^{-1/2} = o\{n^{-1}A_b(x)\}$$

uniformly since $nb^2 \to \infty$, as assumed in (2.3). This completes the proof for the case $x \le b^{1-\epsilon}$.

The second case is $b^{1-\epsilon} < x \leq 1/2$. The derivative of the beta kernel with respect to t is

$$K^{(1)}_{x/b+1,(1-x)/b+1}(t) = b^{-1}(1+b)\{K_{x/b,(1-x)/b+1}(t) - K_{x/b+1,(1-x)/b}(t)\}.$$

As $x/b \to \infty$ and $(1-x)/b \to \infty$ in this case, we have a situation similar to (A.11):

$$|K^{(1)}_{x/b+1,(1-x)/b+1}(t)| \leq C_6 b^{-2}. \tag{A.15}$$

From the Mean Value Theorem,

$$|T_{x,b}| = 2n^{-1}|\sum(s_i - s_{i-1})K_{x/b+1,(1-x)/b+1}(t_i^{'})K^{(1)}_{x/b+1,(1-x)/b+1}(t_i^{'})(\theta_i - t_i)|$$

where $t_i^{'} \in (t_i, \theta_i)$. From (A.12) and (A.15),

$$|T_{x,b}| \leq C_7 n^{-2} b^{-5/2}\{x(1-x)\}^{-1/2} = o\{n^{-1}A_b(x)\}$$

uniformly for $b^{1-\epsilon} < x \leq 1/2$. Thus, the proof of (A.10) is complete.

Some algebra shows that

$$\int_0^1 \sigma^2(t)f^{-1}(t)K^2_{x/b+1,(1-x)/b+1}(t)dt = A_b(x)E\{\sigma^2(\gamma_x)f^{-1}(\gamma_x)\},$$

where γ_x is the Beta random variable up to the constant $A_b(x)$. Using the same method as in Section 9.1, it may be shown that

$$\int_0^1 f^{-1}(t)K^2_{x/b+1,(1-x)/b+1}(t)dt = A_b(x)\{\sigma^2(x)f^{-1}(x) + O(b)\}. \tag{A.16}$$

Then, (3.4) can be established by combining (A.9), (A.10) and (A.16).

References

Brown, B. M. and Chen, S. X. (1999). Beta-Bernstein smoothing for regression curves with compact supports. *Scand. J. Statist.* **26**, 47-59.

Chen, S. X. (1999). A beta kernel estimator for density functions with compact supports. *Comput. Statist. Data Anal.* **31**, 131-145.

Fan, J. Q. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.* **20**, 2008-2036.

Fan, J. Q. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21**, 196-216.

Gasser, Th. and Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation* (Edited by Th. Gasser and M. Rosenblatt), 23-68. Springer, Heidelberg.

Gasser, Th. and Müller, H.-G. (1984). Estimating regression functions and their derivatives by kernel methods. *Scand. J. Statist.* **11**, 171-185.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

Jennen-Steinmetz, C. and Gasser, Th. (1988). A unifying approach to nonparametric regression estimation. *J. Amer. Statist. Assoc.* **83**, 1084-1089.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*. Wiley, New York.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statist. Comput.* **3**, 135-146.

Marron, J. S. and Ruppert, D. (1994). Transformation to reduce boundary bias in kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **56**, 653-672.

Müller, H.-G. (1984). Optimal designs for nonparametric kernel regression. *Statist. Probab. Lett.* **2**, 285-290.

Müller, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*. Springer-Verlag, Berlin.

Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* **78**, 521-530.

Müller, H.-G. and Wang, J. L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50**, 61-76.

Press, W. H., Flannery, B. F., Teukolsky, S. A. and Vetterling, W. T. (1992). *Numerical Recipes: the Art of Scientific Computing*. Cambridge University Press, Cambridge.

Rice, J. A. (1984). Boundary modification for kernel regression. *Comm. Statist. Theory Methods* **13**, 893-900.

Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Comm. Statist. Ser. A* **14**, 1123-1136.

Department of Statistical Science, La Trobe University, VIC 3083, Australia.

E-mail: song.chen@latrobe.edu.au

(Received October 1997; accepted March 1999)