Chen, Song Xi

Working Paper

Local linear smoothers using asymmetric kernels

SFB 373 Discussion Paper, No. 1999,100

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Chen, Song Xi (1999) : Local linear smoothers using asymmetric kernels, SFB 373 Discussion Paper, No. 1999,100, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10046875

This Version is available at:
https://hdl.handle.net/10419/61788

Local Linear Smoothers Using Asymmetric Kernels

SONG XI CHEN
Department of Statistical Science
La Trobe University, VIC 3083, Australia
Song.Chen@latrobe.edu.au

ABSTRACT. This paper considers using asymmetric kernels in local linear smoothing to estimate a regression curve with bounded support. The asymmetric kernels are either beta kernels if the curve has a compact support or gamma kernels if the curve is bounded from one end only. While possessing the standard benefits of local linear smoothing, the local linear smoother using the beta or gamma kernels offers some extra advantages in aspects of having finite variance and resistance to sparse design. These are due to their flexible kernel shape and the support of the kernel matching the support of the regression curve.

Key Words: Beta kernels; Gamma kernels; Local linear smoother; Nonparametric regression; Sparse region.

Runing Head: Smoothing Using Asymetric Kernels.

Contact Details: Song Xi Chen, Department of Statistical Science, La Trobe University, VIC 3083, Australia. Telephone: 61-3-9479 3665; Fax: 61-3-9479 2466; Email: Song.Chen@latrobe.edu.au.

The author wish to submit the final manuscript electronically in LATEX.

1. INTRODUCTION

In recent years, local polynomial smoothing (Stone, 1977; Cleveland, 1979; Cleveland and Delvin, 1988) has been shown by Fan (1993), Fan and Gijbels (1992), Hastie and Loader (1993), Ruppert and Wand (1994) and others to be an effective smoothing method in nonparametric regression. It has the advantages of achieving full asymptotic minimax efficiency and automatically correcting for boundary bias. An excellent review of local polynomial smoothing is given in Fan and Gijbels (1996). The standard application of the local polynomial smoothing has been focused on employing symmetric kernels, among them the compact kernels are popular choices. However, local polynomial smoothers using a compact kernel have a problem as the variance is unbounded in finite sample as revealed in Seifert and Gasser (1996). Seifert and Gasser also find that using the Gaussian kernel has attractive variance behaviour. This is because the Gaussian kernel has unbounded support that leads to a key matrix in local polynomial smoothing being non-singular, thus making the finite sample variance bounded.

When the curve under consideration has a bounded support, kernels whose support matches the support of the curve should have the same attractive variance properties as the Gaussian kernel, and are more appealing as no weights are allocated outside the support of the curve. Recently, beta kernels, which are densities of some beta distributions, have been proposed to smooth the Bernstein polynomials in Brown and Chen (1999), to construct Gasser-Müller estimator in Chen (1999a) and density estimators in Chen (1999b). Related gamma kernels are proposed for density estimation in Chen (1998). These beta and gamma smoothers are free of boundary bias, achieve $n^{-4/5}$-order convergence for their mean integrated square errors and have attractive finite sample properties. It is also found, however, that their asymptotic mean square errors are of a larger order near the boundary. Even though this happens only in a small boundary area so small that it registers no effect on the mean integrated square error, it does pose an "asymptotic" hiccup for this nice idea of smoothing.

In this paper we consider local linear smoothing using the beta and gamma kernels. It turns out that local linear beta or gamma kernel smoothers remove the problem

of increased mean square error near the boundary, while maintaining the usual good properties of standard local linear smoothing with a fixed symmetric kernel. Using the beta or gamma kernels offers some extra benefits. Firstly, it is a kind of adaptive smoothing as both the beta and the gamma kernels have varying shapes and varying amounts of smoothness. Secondly, it has finite sample variance as long as there are two different design points not on the boundary of the support. The third is that the effective sample size is increased and thus the finite sample variance of the curve estimates can be reduced. These features are due to the fact that the support of the kernels matches the support of the regression curve. And the latter can make the local linear smoother having smaller variance when the curve has sparse areas. It has been shown that a local linear smoother using a symmetric kernel implicitly uses asymmetric kernels as well. However, the support of the kernels does not match that of the curve. This can produce problem in the variance when the design is sparse.

The paper is structured as follows. Section 2 introduces the local linear smoother using either beta or gamma kernels. The general properties using beta kernels are studied in Section 3, whereas those using gamma kernels are given in Section 4. Section 5 considers finite sample variance properties. A data set from a line transect survey is analyzed in Section 6. Section 7 presents some simulation results. Some derivations are given in the Appendix.

2. BETA OR GAMMA KERNEL BASED LOCAL LINEAR SMOOTHER

Let $Y_1, \ldots, Y_n$ be the responses of n design points $X_1, \ldots, X_n$ from a regression model

$$Y_i = m(X_i) + \epsilon_i \quad i = 1, \ldots, n, \tag{1}$$

where $m(\cdot)$ is an unknown regression function with bounded support $\mathcal{S}$ and the residual ϵ_i are uncorrelated random variables with zero mean and variance $\sigma^2(X_i)$. We consider in this paper that $\mathcal{S}$ is either $[0,1]$ or $[0,\infty)$ which are two typical forms of bounded supports.

The kernel we use to smooth at x is either

$$K_{x,b}(t) = \frac{t^{x/b}(1-t)^{(1-x)/b}}{B\{x/b+1, (1-x)/b+1\}}\, I(t \in \mathcal{S}) \quad \textit{if } \mathcal{S} = [0,1]$$

or

$$K_{x,b}(t) = \Gamma^{-1}(x/b+1)b^{-x/b}t^{x/b}e^{-t/b}\ I(t \in \mathcal{S}) \quad \textit{if } \mathcal{S} = [0,\infty)$$

where B and Γ are the beta and gamma functions respectively, and b is a smoothing parameter. Clearly, the kernel is the density of Beta$\{x/b+1, (1-x)/b+1\}$ or Gamma$\{x/b+1, b\}$ distribution. Both beta and gamma kernels have varying kernel shapes and become more asymmetric as x moves towards the boundary. Both beta and gamma kernels have $\mathcal{S}$ as the support, matching that of the regression curve.

The local linear smoother for m using either the beta or the gamma kernel $K_{x,b}$ is obtained by finding a and b that minimize

$$\sum_{j=1}^{n}\{Y_j - a - b(x-X_j)\}^2 K_{x,b}(X_j).$$

Let $S_l(x) = n^{-1}\sum_{j=1}^{n}(x-X_j)K_{x,b}(X_j)$ for $l = 0,1,2$. The local linear smoother $\hat{m}(x) = \hat{a}$, the solution of a to the above optimization, and has a detailed form

$$\hat{m}(x) = \sum_{j=1}^{n} w_j(x)Y_j / \sum_{j=1}^{n} w_j(x) \tag{2}$$

where the local linear weight

$$w_j(x) = n^{-1}\{S_2(x) - S_1(x)(x-X_j)\}K_{x,b}(X_j).$$

Let f be the density function of the design points. We assume throughout the paper that, for some positive constants f_c and σ_c^2,

$$\begin{aligned}
&(i)\ m^{(2)} \in C(\mathcal{S}),\ f(\cdot) \text{ and } \sigma^2(\cdot) \text{ obey a first order Lipschitz condition in } \mathcal{S};\\
&(ii)\ f(x) \geq f_c > 0 \text{ and } \sigma^2(x) \leq \sigma_c^2 \text{ for all } x \in \mathcal{S};\\
&(iii)\ b \to 0 \text{ and } nb^2 \to \infty \text{ as } n \to \infty;
\end{aligned} \tag{3}$$

Let ξ_x be the beta or gamma random variable with $K_{x,b}$ as its density, $p_l(x) = E(\xi_x - x)^l$ for $l = 0,1,2\ldots$ and

$$A_b(x) = \begin{cases} \frac{B\{2x/b+1, 2(1-x)/b+1\}}{B^2\{x/b+1,(1-x)/b+1\}} & \textit{if } \mathcal{S} = [0,1]; \\ \frac{\Gamma(2x/b+1)}{b2^{2x/b+1}\Gamma^2(x/b+1)} & \textit{if } \mathcal{S} = [0,\infty). \end{cases}$$

We first present the following general theorem, whose proof is given in the appendix.

Theorem 1. Assume the conditions given in (3). Then for any $x \in \mathcal{S}$

$$Bias\{\hat{m}(x)\} = \tfrac{1}{2}m^{(2)}(x)p_2(x) + o\{p_2(x)\} + O\{n^{-1}A_b(x)\} \quad \text{and} \tag{4}$$

$$Var\{\hat{m}(x)\} = n^{-1}A_b(x)\sigma^2(x)f^{-1}(x) + o\{n^{-1}A_b(x)\}. \tag{5}$$

Remark. The magic power of the local linear (polynomial) smoothing is to make the leading bias term free of the first derivatives of m and f. This power is maintained in beta or gamma kernels. It is this missing first derivatives in the bias that removes the increased mean square error near the boundary which was associated with the earlier beta/gamma kernel smoothers.

3. ASYMPTOTIC PROPERTIES USING BETA KERNELS

We study the asymptotic properties of the local linear beta smoother $\hat{m}(x)$ assuming $S = [0,1]$ in this section. To simplify notations, we define $x \in \mathcal{S}$ to be a

$$\text{“}interior\ x\text{”}\ \textit{if}\ \text{“}x/b\ \textit{and}\ (1-x)/b \to \infty\text{”}\ \textit{and}$$

$$\text{“}boundary\ x\text{”}\ \textit{if}\ \text{“}x/b\ \textit{or}\ (1-x)/b \to \kappa\text{”}$$

where κ is a nonnegative constant. Clearly any fixed $x \in [0,1]$ which is free of n is an interior point.

From the basic properties of beta random variables, $E(\xi_x) = (x+b)/(1+2b)$ and $Var(\xi_x) = b(x+1)(1-x+1)/\{(1+2b)^2(1+3b)\}$. Thus,

$$\begin{aligned} p_2(x) &= \frac{bx(1-x) + b^2\{2-4x(1-x)\} + b^3\{4-12x(1-x)\}}{(1+2b)^2(1+3b)} \\ &= \begin{cases} bx(1-x) + O(b^2) & \textit{for interior } x; \\ (2+\kappa)b^2 + O(b^3) & \textit{for boundary } x. \end{cases} \end{aligned} \tag{6}$$

Chen (1999a) shows that for small b

$$A_b(x) = \begin{cases} b^{-1/2}\frac{1}{\sqrt{4\pi}\sqrt{x(1-x)}} + o(b^{-1/2}) & \textit{for interior } x; \\ b^{-1}\frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} + o(b^{-1}) & \textit{for boundary } x. \end{cases} \tag{7}$$

From the Theorem 1, (6) and (7), we have

$$Bias\{\hat{m}(x)\} = \begin{cases} \frac{1}{2}x(1-x)m^{(2)}(x)b + O(b^2) & \textit{for interior } x; \\ \\ \frac{1}{2}(2+\kappa)m^{(2)}(x)b^2 + O(b^3) & \textit{for boundary } x \end{cases} \tag{8}$$

and

$$Var\{\hat{m}(x)\} = \begin{cases} n^{-1}b^{-1/2}\frac{\sigma^2(x)}{\sqrt{4\pi}\sqrt{x(1-x)}f(x)} + o(n^{-1}b^{-1/2}) & \textit{for interior } x; \\ n^{-1}b^{-1}\frac{\sigma^2(x)\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)f(x)} + o(n^{-1}b^{-1}) & \textit{for boundary } x. \end{cases} \tag{9}$$

In the boundary areas, in terms of the order of magnitude of b the bias is of a smaller order whereas the variance is of larger order than those in the interior. However, b does not represent the total amount of smoothing used; rather, $p_2(x)$ is the real amount of smoothing used at x. The trade-off between the bias and the variance is directly due to $p_2(x)$ having different orders between the boundary and the interior as shown in (6).

By adjusting b so that $p_2(x)$ is of the same order within $[0,1]$, the mean square error can be made of order $n^{-4/5}$ everywhere within $[0,1]$. To appreciate this, notice that

$$MSE\{\hat{m}(x)\} = \begin{cases} n^{-1}b^{-1/2}\frac{\sigma^2(x)}{\sqrt{4\pi}\sqrt{x(1-x)}f(x)} + \frac{1}{4}x^2(1-x)^2\{m^{(2)}(x)\}^2b^2 \textit{ or} \\ n^{-1}b^{-1}\frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} + \frac{1}{4}(2+\kappa)^2\{m^{(2)}(x)\}^2b^4 \end{cases}$$

respectively in the interior with error terms of $o(n^{-1}b^{-1/2}+b^2)$ or in the boundary with error terms of $o(n^{-1}b^{-1}+b^4)$. The optimal bandwidth is

$$b^\star(x) = \begin{cases} \left[\frac{\sigma^2(x)}{\sqrt{4\pi}x^{5/2}(1-x)^{5/2}f(x)\{m^{(2)}(x)\}^2}\right]^{2/5} n^{-2/5} & \textit{for interior } x; \\ \left[\frac{\Gamma(2\kappa+1)\sigma^2(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)f(x)(2+\kappa)\{m^{(2)}(x)\}^2}\right]^{1/5} n^{-1/5} & \textit{for boundary } x. \end{cases} \tag{10}$$

The optimal mean square error, with an error term of $o(n^{-4/5})$, is

$$MSE^\star\{\hat{m}(x)\} = \begin{cases} \frac{5}{4}n^{-4/5}\left\{\frac{\sigma^2(x)m^{(2)}(x)}{\sqrt{4\pi}f(x)}\right\}^{4/5} & \textit{for interior } x; \\ \frac{5}{4}n^{-4/5}(2+\kappa)^{1/5}\left\{\frac{\Gamma(2\kappa+1)\sigma^2(x)m^{(2)}(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)f(x)}\right\}^{4/5} & \textit{for boundary } x. \end{cases} \tag{11}$$

Hence, the optimal mean square error is of order $n^{-4/5}$ throughout $[0,1]$. This improves the beta kernel estimator considered in Chen (1999a) whose mean square error

is of order $n^{-2/3}$ in the boundary areas and is of order $n^{-4/5}$ in the interior. The optimal bandwidth given in (10) prescribes a larger b value in the boundary to offset a reduced value of $bx(1-x)$ as x approaches the boundaries. By doing so, the total amount of smoothness $p_2(x) = O(n^{-2/5})$ throughout $[0,1]$.

It may be shown in a manner similar to that given in Chen (1999a) that the bias and the variance in the boundary areas have negligible contribution to the mean integrated square error, that is with an error term of $o(n^{-1}b^{-1/2} + b^2)$

$$MISE(\hat{m}) = n^{-1}b^{-1/2}\int_0^1 \frac{\sigma^2(x)}{\sqrt{4\pi}\sqrt{x(1-x)}f(x)}dx + b^2\tfrac{1}{4}\int_0^1 x^2(1-x)^2\{m^{(2)}(x)\}^2dx.$$

The optimal global bandwidth

$$b^\star = \left\{\tfrac{1}{\sqrt{4\pi}}\int_0^1 \frac{\sigma^2(x)}{\sqrt{x(1-x)}f(x)}dx\right\}^{2/5}\left\{\int_0^1\{x(1-x)m^{(2)}(x)\}^2dx\right\}^{-2/5} n^{-2/5}, \tag{12}$$

and the optimal mean integrated square error is

$$MISE^\star(\hat{m}) = \tfrac{4}{5}\left\{\tfrac{1}{\sqrt{4\pi}}\int_0^1 \frac{\sigma^2(x)}{\sqrt{x(1-x)}f(x)}dx\right\}^{4/5}\left\{\int_0^1\{x(1-x)m^{(2)}(x)\}^2dx\right\}^{-4/5} n^{-4/5}. \tag{13}$$

It is interesting to see that the optimal mean integrated square error and the optimal mean square error for interior x coincide respectively with those of the local linear smoother using the Gaussian kernel. So, the beta kernels are asymptotically equivalent to the Gaussian kernel in a global sense within $[0,1]$, and in a point-wise sense in the interior of $[0,1]$. However, they are not asymptotically equivalent in the boundary areas and their finite sample handling of the smoothing may be quite different.

4. ASYMPTOTIC PROPERTIES USING GAMMA KERNELS

In this section, we investigate the asymptotic properties of the local linear gamma smoother assuming $I = [0,\infty)$. We define $x \in S$ to be a

$$\textit{"interior } x\textit{"} \text{ if } \textit{"}x/b \to \infty\textit{"} \text{ or } \textit{"boundary } x\textit{"} \text{ if } \textit{"}x/b \to \kappa\textit{"}. \tag{14}$$

As ξ_x is the Gamma$(x/b+1, b)$ random variable, $E(\xi_x) = x+b$ and $Var(\xi_x) = bx+b^2$. Thus,

$$p_2(x) = bx + 2b^2. \tag{15}$$

According to Chen (1998) for small b, when $S = [0, \infty)$,

$$A_b(x) = \begin{cases} b^{-1/2}/\{\sqrt{4\pi}\sqrt{x}\} + o(b^{-1/2}) & \textit{for interior } x; \\ b^{-1}\Gamma(2\kappa+1)/\{2^{2\kappa+1}\Gamma^2(\kappa+1)\} + o(b^{-1}) & \textit{for boundary } x. \end{cases} \tag{16}$$

This is almost the same as that of the beta kernel given in (7) except that $\sqrt{1-x}$ does not appear when x is in the interior as $x = 1$ is no longer a boundary point.

Substituting (15) and (16) into Theorem 1, we will reproduce the formulae from (8) to (13) except that $1 - x$ does not appear in these formulae when x is in the interior and the integration interval is $S = [0, \infty)$.

However, using the gamma kernels produces some unique and interesting properties. Notice that

$$Bias\{\hat{m}(x)\} = \begin{cases} \frac{1}{2}xm^{(2)}(x)b + O(b^2) & \textit{interior } x \\ \frac{1}{2}(2+\kappa)m^{(2)}(x)b^2 + O(b^3) & x \textit{ in boundary} \end{cases} \tag{17}$$

and

$$Var\{\hat{m}(x)\} = \begin{cases} n^{-1}b^{-1/2}\frac{\sigma^2(x)}{\sqrt{4\pi}\sqrt{x}f(x)} + o(n^{-1}b^{-1/2}) \ \textit{interior } x \\ \\ n^{-1}b^{-1}\frac{\sigma^2(x)\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)f(x)} + o(n^{-1}b^{-1}) \ \textit{boundary } x. \end{cases} \tag{18}$$

So, the variance decreases when x increases as $x^{-1/2}$ appeared in the leading term of the asymptotic variance. This property is highly desirable when estimating curves with sparse regions in the upper tail of the design density f. This is a property not shared by using other fixed symmetric kernels including the Gaussian kernel. The reduced variance for large x is gained at the price of increasing bias. This is equivalent to the strategy of using larger bandwidth values in areas where the design is sparse to increase the bias in return for reduced variance. The gamma kernel carries this strategy automatically in a natural manner.

5. FINITE SAMPLE VARIANCE

One problem with a compact kernel based local polynomial smoother, as revealed in Serfeit and Gasser (1996), is that its finite sample variance can be infinite. This is because, in the case of the local linear smoother $\sum w_j(x) = S_2(x)S_0(x) - S_1^2(x)$, the denominator of (2) has a positive probability of being zero. To avoid a zero denominator, Fan (1993) added n^{-2} in the denominator. It is shown in the following

that using the beta or gamma kernels can eliminate the problem. The results are valid for local polynomial smoothers rather than just local linear smoothers.

Let

$$X_p = \begin{pmatrix} 1\,(x - X_1) \cdots (x - X_1)^p \\ \cdots\cdots\cdots \\ 1\,(x - X_n) \cdots (x - X_n)^p \end{pmatrix} \quad \text{and} \quad W = diag\,(K_{x,b}(X_j))\,.$$

Lemma. If there are at least $p + 1 \leq n$ different design points not being on the boundary of $\mathcal{S} = [0, 1]$, then $X^T W X$ is non-singular.

Proof: The condition of the lemma implies that $rank(W) \geq p+1$ and $rank(X_p) = p + 1$. So, $rank(X_p^T W X_p) = p + 1$. Therefore, $X_p^T W X_p$ is non-singular.

In the case of local linear ($p = 1$), it may be shown by some matrix algebra that

$$\sum w_j(x) = |X_p^T W X_p| = \tfrac{1}{2} \textstyle\sum_{i \neq j} (X_i - X_j)^2 K_{x,b}(X_i) K_{x,b}(X_j), \tag{19}$$

which is a weighted measure of the spread of the design points. So, if there are two different design points not being on the boundary of $\mathcal{S}$ $\sum w_j(x) \neq 0$ as maintained in the Lemma. Note in passing that (19), which is valid for symmetric kernels and appears has not been noticed before, is useful in computation as recursive formula can be developed.

Based on the lemma, it is readily true that the probability of $X_p^T W X_p$ being singular is zero if the design variable is either continuous or fixed but there are at least $p + 1$ different points not being on the boundary of $\mathcal{S}$. The later case includes any equally spaced design as long as $n \geq p+3$. The condition on the design variable is very weak and is satisfied in almost all the practical cases. Therefore, in finite samples the local polynomial beta or gamma smoother has finite variance with probability 1. This revises a corollary in Serfeit and Gasser (1996) which maintained that local linear fits have finite variance if and only if the kernel function has noncompact support.

6. AN EXAMPLE

In this section we analyze Laake's (1978) stake data, a well known data set in line transect survey. Line transect survey (Seber,1982; and Buckland et al., 1993) is a popular methodology for estimating the abundance of biological populations. It

has been an important tool for wildlife management and conservation. To estimate size of a population, an observer traverses a distance L along randomly allocated non-overlapping transect lines within the survey area. Each object sighted from the transect lines is counted and its perpendicular distance x from the line is measured. In 1978 when the theory of line transect survey was still in its maturing stage, Jeff Laake placed 150 wooden stakes randomly in a rectangular area. Eleven observers were asked to traverse a transect line of 1000 meters long, cut through the middle of the rectangle, to detect the stakes. The purpose of the experiment was to study the performance of some line transect estimators as the real population size in almost all the applications of line transect surveys is unknown in practice.

Let $Y_i = 1$ if the i-th stake which is distance X_i away from the transect line; and $Y_i = 0$ otherwise. Let $m(x)$ be the conditional probability of detecting an object given that the object is at distance x from the transect line. Then, Y_i is binomial$\{1, m(X_i)\}$ distributed, and thus $m(x)$ is the mean function and $\sigma^2(x) = m(x)\{1 - m(x)\}$.

An essential assumption in conventional line transect surveys is that $m(0) = 1$; that is, certain detection can be achieved for objects situated right on the transect line (the walking path). Other assumptions are that g is monotone decreasing and $m(x)$ exhibits " a shoulder" near $x = 0$; that is, $m'(0) = 0$. The interest of the current analysis is on estimating the detection function $m(x)$ for two of the eleven observers to see if m obeys the above assumptions. knowledge on the detection pattern is also vital in providing possible models for the line transect data, and is useful in training observers to detect properly.

Local linear estimates of the detection functions using the beta kernels are given in Figure 1 (panels (b) and (d)) together with the data sets (panels (a) and (c)). To apply the beta kernel based smoother, the x_i were first linearly transformed from $[0, 20]$ to $[0, 1]$. A smoothing bandwidth was obtained by the so-called plug-in method based on the optimal bandwidth given in (12). As we know the stakes were placed randomly, $f(x) = 1$ for $x \in [0, 1]$. Estimates of $m(x)$ were obtained via fitting a quadratic polynomial to each of the two data sets, similarly to the rule of thumb bandwidth selection method outlined in Fan and Gijbles (1996,p111f). The estimate of $\sigma^2(x)$

was obtained by substituting the estimate for $m(x)$. From (12), we had $b_1 = 0.143$ for observer 1 and $b = 0.119$ for observer 2. The curve estimates employing these bandwidth values (in solid lines) were in fact very similar to those of the quadratic polynomial fits (in dotted lines). While this was quite remarkable, it did indicated that the above bandwidths were perhaps a little bit too large. The curves in dashed lines were estimates using only a quarter of the above bandwidths, which picked up more local features such as a bump at $x = 4$ for the second observer.

The estimates reveal quite different detection patterns between the two. They show that the first observer had better ability of detection than the second as $\hat{m}_1(0) \approx 1$ and $\hat{m}_1$ maintained a higher level than $\hat{m}_2$ throughout. That $\hat{m}_1(x)$ was much larger than 0 even when x was towards the right boundary was the most impressive.

The two beta kernel estimates of $m_2(0)$ were all much less than 1, indicating that there is some evidence that $m_2(0) < 1$. This was hardly believable as the survey was run on land under good weather condition and the observer still had poor detection for objects which were very close to him. Notice that the polynomial fit did not reveal this problem at all as it was a global fit and tended to miss some local features. Of course, to vigorously prove that $m_2(0) < 1$, we have to construct an one sided confidence interval for $m_2(0)$, which will not be discussed here. Also the estimates of m_2 did not exibit any shoulders near $x = 0$ at all. The above analysis indicates that we cannot take the assumptions of line transect surveys for granted, and analyses are needed for each data set to check their validity before they are incorporated in various models.

7. SIMULATION RESULTS

A simulation study was carried out designed to investigate the performance of the beta or gamma kernel based local linear smoother. For comparison purposes, the local linear smoother using the Epanechnikov and the Gaussian kernels are also considered.

The data were generated according to two regression models based on:

$$Y_i = m(X_i) + \epsilon_i$$

where ϵ_i are independent $N(0, 0.05^2)$ random variables. In the first model, $m(x) = (x - 0.5)^2 I(0 \leq x \leq 1)$ and X_i were independent $|N(0, 0.3^2)|$ random variables truncated on $[-1, 1]$. In the second model, $m(x) = \exp(-x) + \exp\{-4(x-1)^2\}$ for $x > 0$ and X_i were independent Gamma(2) random variables. So, in both models the design density was relatively sparse towards the right end of the support. As the regression curve is compactly supported in the first case, the gamma kernel based local linear smoother is not considered here. And similarly, the beta kernel based smoother was not considered in the second case. The simulation results were all based on 1000 simulations with the random variables generated using the algorithm given in Press *et. al.* (1992). The sample sizes used in each model were $n = 100, 200$ and 400.

The performances of the smoothers were evaluated over a gird of equally spaced points within $[0, 1]$ in the first and $[0, 5.5]$ in the second model. In each simulation for each smoother at each point, the bandwidth that minimizes the average squared error was chosen by a golden search algorithm given in Press *el. al.* (1992), and was then used to compute the curve estimate. After 1000 simulations, the average squared bias and variance were obtained as measures of performance.

The simulation results are summarized in Figure 2 for the first model and Figure 3 for the second model. Both figures show that local linear smoothers using the beta or gamma kernels had the best variance for almost all the sample sizes considered across the entire range. They had significantly smaller variance when $x > 0.8$ in the first model and when $x > 3$ in the second where the design density was sparse. The Gaussian kernel based smoother performed better than that based on the Epanechnikov kernel, which confirmed the results of Seifert and Gasser (1996). The variance of the compact kernel based smoother towards the boundary of the support was a serious concern in the second case.

The increase in both bias and variance as x increases in Figure 2 was due to (i) $m''(x)$ being constant and (ii) the design density is monotonic decreasing within $[0, 1]$. The U-shape of the variance in Figure 3 was due to the Gamma(2) design distribution used which has a mode at $x = 1$. The increase in bias in Figure 3 when $x \in [1, 2]$ was due to the fact that $\{m''(x)\}^2$ peaks at $x = 1.5$. The rise in bias at $x = 0$ was due

to the combination of the boundary effect and the relatively lower level design. It is quite interesting to see the gamma kernel based smoother had both their bias and variance drop at $x = 0$. As expected, the bias of the gamma kernel based smoother was the largest in the sparse area, but was a good price paid in return for a much larger reduction in the variance. It was a little surprising to see the good performance of the gamma kernel based estimator at $x = 0$.

APPENDIX: PROOF OF THE THEOREM

We only prove the theorem for the case of beta kernels as that for the gamma kernels can be derived in a similar way.

Let $\mu_l(x) = E\{S_l(x)\}$, $T_l(x) = n^{-1}\sum_{j=1}^{n}(x-X_j)^l m(X_j)K_{x,b}(X_j)$ and $\nu_l = E\{T_l(x)\}$ for $l = 1, 2, 3$, and $r(x) = m(x)f(x)$. It can be shown that

$$\mu_l(x) = (-1)^l \sum_{j=0}^{2-l} f^{(j)}(x)p_j(x)/j! + o\{p_2(x)\} \quad \text{and} \tag{A.1}$$

$$\nu_l(x) = (-1)^l \sum_{j=0}^{2-l} r^{(j)}(x)p_j(x)/j! + o\{p_2(x)\}. \tag{A.2}$$

Notice that

$$\begin{aligned} Cov\{S_{l_1}(x), S_{l_2}(x)\} &= n^{-1}\int K_{x,b}^2(t)(x-y)^{l_1+l_2}f(y)dy - n^{-1}\mu_{l_1}\mu_{l_2} \\ &= n^{-1}A_b(x)E\{(x-\gamma_x)^{l_1+l_2}f(\gamma_x)\} - n^{-1}\mu_{l_1}\mu_{l_2} \end{aligned} \tag{A.3}$$

where γ_x is the Beta$\{2x/b+1, 2(1-x)/b+1\}$ random variable. From Chen (1999a), $A_b(x) = O\{b^{-1/2}x^{-1/2}(1-x)^{-1/2}\}$. Also $E\{(x-\gamma_x)^{l_1+l_2}f(\gamma_x)\} = O\{\mu_{l_1+l_2}(x)\}$. Thus, it may be shown that the second term on the right hand side of (A.3) is of a smaller order than the first term. Therefore,

$$Cov\{S_{l_1}(x), S_{l_2}(x)\} = \begin{cases} O\{n^{-1}A_b(x)\} & \text{if } l_1 + l_2 = 0; \\ o\{n^{-1}A_b(x)\} & \text{if } l_1 + l_2 \geq 1. \end{cases}$$

In general, we have

$$Cov\{H_{l_1}(x), H_{l_2}(x)\} = \begin{cases} O\{n^{-1}A_b(x)\} & \text{if } l_1 + l_2 = 0; \\ o\{n^{-1}A_b(x)\} & \text{if } l_1 + l_2 \geq 1 \end{cases} \tag{A.4}$$

where H_l can be either S_l or T_l.

Let $\hat{r}(x) = S_2(x)T_0(x) - S_1(x)T_0(x)$ and $\hat{q}(x) = S_2(x)S_0(x) - S_1^2(x)$. Then $\hat{m}(x) = \hat{r}(x)/\hat{q}(x)$ and based on (A.4) we have

$$E\{\hat{m}(x)\} = \frac{\mu_2(x)\nu_0(x) - \mu_1(x)\nu_1(x)}{\mu_2(x)\mu_0(x) - \mu_1^2(x)} + O\{n^{-1}A_b(x)\}.$$

Substituting (A.1) and (A.2), and using the standard derivation for the bias of local linear smoothers, we may derive (4) of the theorem.

Note that

$$\begin{aligned} Var\{\hat{m}(x)\} &= Var[E\{\frac{\sum w_j(x)Y_j}{\sum w_j(x)}|X_1,\cdots,X_n\}] + E[Var\{\frac{\sum w_j(x)Y_j}{\sum w_j(x)}|X_1,\cdots,X_n\}] \\ &= Var\{\hat{r}(x)/\hat{q}(x)\} + E\Big[\frac{\sum w_j^2(x)\sigma^2(X_j)}{\{\sum w_j(x)\}^2}\Big] \end{aligned} \tag{A.5}$$

as $\sum w_j(x)m(X_j) = \hat{r}(x)$ and $\sum w_j(x) = \hat{q}(x)$.

Let $\eta(x) = \mu_2(x)\mu_0(x) - \mu_1^2(x)$ and $\beta(x) = \mu_2(x)\nu_0(x) - \mu_1(x)\nu_1(x)$. Then,

$$\begin{aligned} Var\{\hat{r}(x)/\hat{q}(x)\} &= \eta^{-2}(x)Var\{\hat{r}(x)\} - 2\eta^{-3}(x)\beta(x)Cov\{\hat{r}(x),\hat{q}(x)\} \\ &+ \eta^{-4}(x)\beta^2(x)Var\{\hat{q}(x)\} + o\{n^{-1}A_b(x)\} \\ &= \eta^{-2}(x)\mu_2^2(x)\Big[Var\{T_0(x)\} - 2\eta^{-1}(x)\beta(x)Cov\{T_0(x),S_0(x)\} \\ &+ \eta^{-2}(x)\beta^2(x)Var\{S_0(x)\}\Big] + o\{n^{-1}A_b(x)\}. \end{aligned}$$

We use (A.4) in the derivation of the last equation. It may be shown in a similar fashion to that for deriving (A.4) that, with error terms of $o\{n^{-1}A_b(x)\}$,

$$\begin{aligned} Var\{T_0(x)\} &= n^{-1}A_b(x)f(x)m^2(x),\ Cov\{T_0(x),S_0(x)\} = n^{-1}A_b(x)f(x)m(x) \text{ and} \\ Var\{S_0(x)\} &= n^{-1}A_b(x)f(x). \end{aligned}$$

These and the fact that $\eta^{-1}(x)\beta(x) = m(x) + o(1)$ lead to

$$Var\{\hat{r}(x)/\hat{q}(x)\} = o\{n^{-1}A_b(x)\}. \tag{A.6}$$

To work out the second term on the right of (A.5), we define

$$W_l(x) = n^{-1}\sum_{j=1}^{n}(x - X_j)^l K^2_{x/b+1,(1-x)/b+1}(X_j)\sigma^2(X_j)$$

for non-negative integer l. It may be shown that

$$\omega_l(x) = E\{W_l(x)\} = \begin{cases} A_b(x)\sigma^2(x)f(x) + o\{A_b(x)\} & \text{if } l = 0; \\ o\{A_b(x)\} & \text{if } l \geq 1. \end{cases} \tag{A.7}$$

Then

$$\begin{aligned} & E\Big[\frac{\sum w_j^2(x)\sigma^2(X_j)}{\{\sum w_j(x)\}^2}\Big] \\ = \; & n^{-1}E\Big[\frac{S_2^2(x)W_0(x) - 2S_1(x)S_2(x)W_1(x) + S_1^2(x)W_2(x)}{\{S_2(x)S_0(x) - S_1^2(x)\}^2}\Big] \\ = \; & n^{-1}\left(\frac{\mu_2^2(x)\omega_0(x) - 2\mu_1(x)\mu_2(x)\omega_1(x) + \mu_1^2(x)\omega_2(x)}{\{\mu_2(x)\mu_0(x) - \mu_1^2(x)\}^2}\right)\{1 + O(1)\} \\ = \; & n^{-1}Z_0(x)/\mu_0^2(x) + o\{n^{-1}A_b(x)\} \\ = \; & n^{-1}A_b(x)\sigma^2(x)/f(x) + o\{n^{-1}A_b(x)\} \end{aligned} \tag{A.8}$$

Combining (A.6) and (A.8) we prove (5) in the lemma, and thus finish the proof of the theorem.

REFERENCES

Brown, B. M. and Chen, S. X, "Beta-Bernstein smoothing for regression curves with compact supports," *Scandinavian Journal of Statistics* **26** (1999), 47-59.

Buckland, S.T., Anderson, D. R., Burnham, K.P. and Laake, J. L, *Distance Sampling*. London: Chapman and Hall, 1993.

Chen, S.X, "Beta Kernel Smoother for Regression Curves", *Statistics Sinica* to appear (2000).

Chen, S.X, "Beta Kernel Estimators for Density Functions", *Computational Statistics and Data Analysis*. 31 (1999), 131-145.

Chen, S.X, "Probability Density Function Estimation Using Gamma Kernels," *Annals of Institute of Statistical Mathematics*, to appear (2000).

Cleveland, W. S, Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Association*. 74 (1979), 829-836.

Cleveland, W. S. and Devlin, S, Locally weighted regression: an approach to regression analysis by local fitting, *Journal of the American Statistical Association*. 83 (1988), 596-610.

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

Fan, J. Q. and Gijbels, I, Variable bandwidth and local linear regression smoothers. *The Annals of Statistics.* **20** (1992), 2008-2036.

Fan, J. Q. and Gijbels, I, "Local Polynomial Modelling and Its Applications" London, U.K.: Chapman and Hall, 1996.

Hastie, T., and Loader, C, Local Regression: Automatic Kernel Carpentry, *Statistical Science.* 8 (1993), 120-143.

Laake, J. L, Line transect estimators robust to animal movement, MS thesis, Utah State University, 1978.

Press, W. H., Flannery, B. F., Teukolsky, S. A. & Vetterling, W.T, "Numerical Recipes: the Art of Scientific Computing", Cambridge, U.K.: Cambridge University Press, 1992.

Ruppert, D., and Wnad, M. P, Multivariate Locally Weighted Least Sqaures Regression, *The Annals of Statistics.* 22 (1994), 1346-1370.

Seber, G.A.F, "The Estimation of Animal Abundance", 2nd edition. London: Griffin, 1982.

Seifert, B. and Gasser, Th., Finite sample variance of local polynomials: analysis and solutions. *Journal of the American Statistical Association.* **91** (1996), 267-275.

Stone, C. J, Consistent Nonparametric Regression, *The Annals of Statistics*, 5 (1977), 595-645.

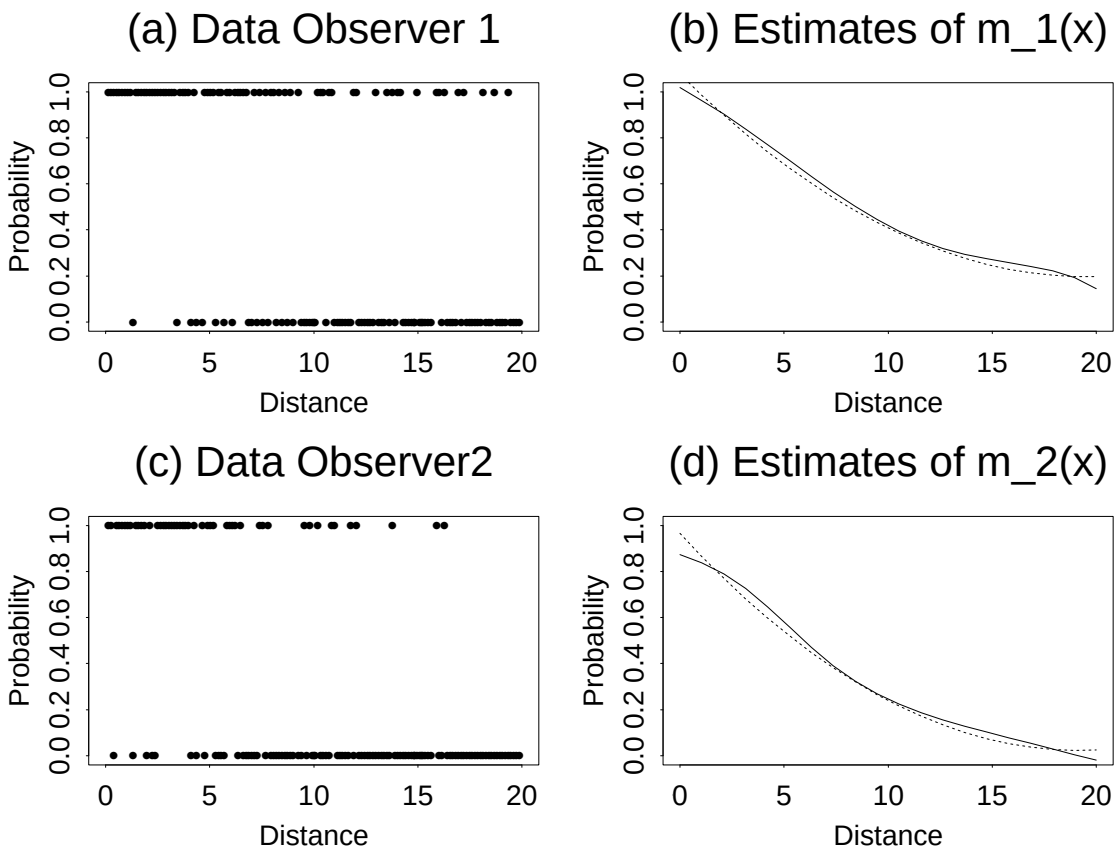

Figure 1. Local linear estimates of the detection functions for the Stake Data using the beta kernels: solid lines with $b = b^\star$ and dotted lines with $b = b^\star/4$, and quadratic polynomial fits: dotted lines.

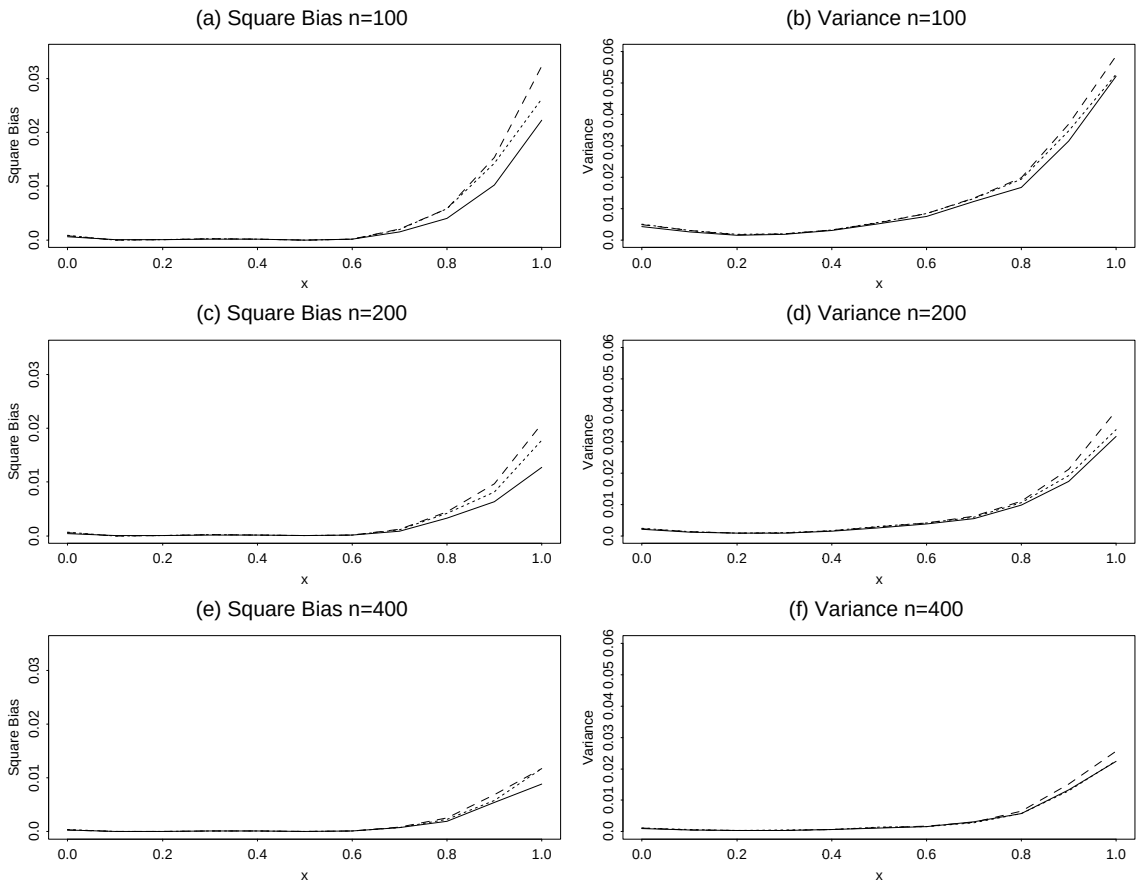

Figure 2. Squared Bias and Variance of the local linear smoother using the beta kernels (solid lines), the Epanechnikov kernel (dotted lines) and the Gaussian kernel (dashed lines); $m(x) = (x - 0.5)^2 I(0 \leq x \leq 1)$ *and the design points* x_i *are truncated* $N(0, 0.3^2)$.

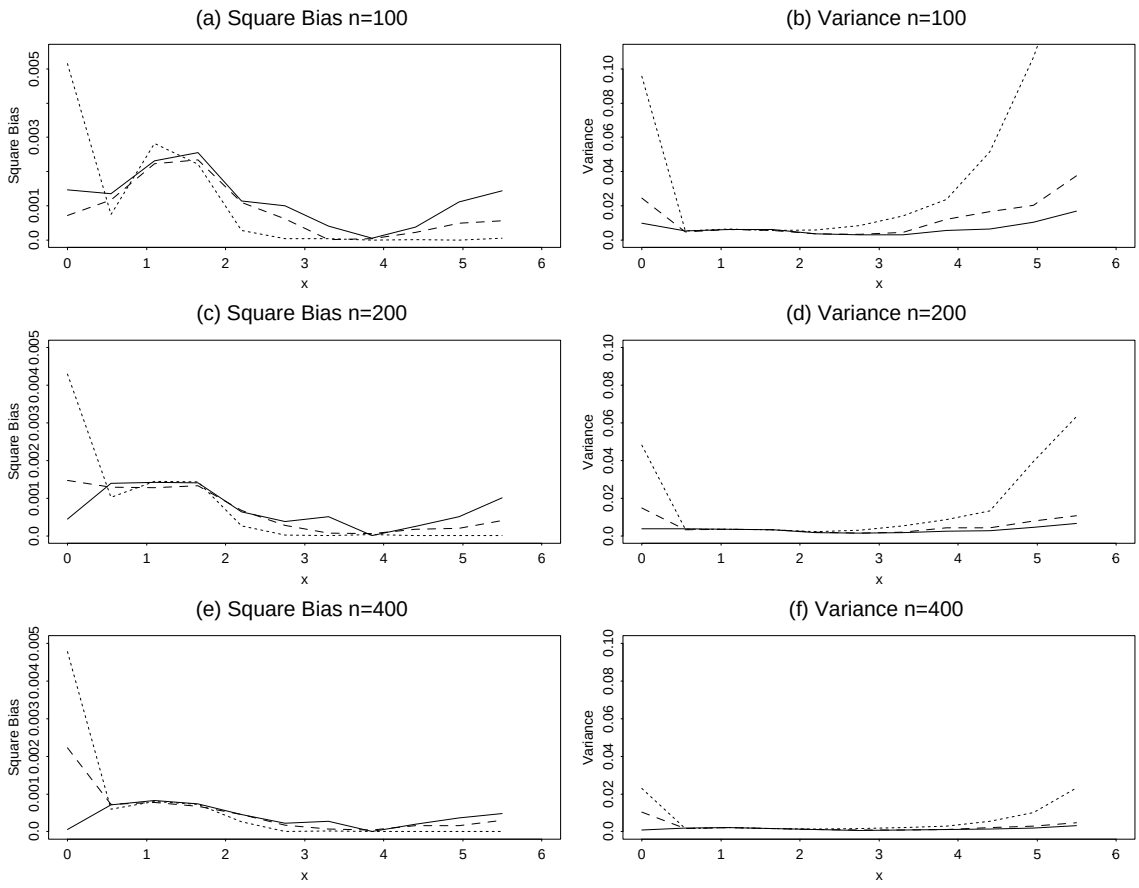

Figure 3. Squared Bias and Variance of the local linear smoother using the gamma kernels (solid lines), the Epanechnikov kernel (dotted lines) and the Gaussian kernel (dashed lines); $m(x) = \exp(-x) + \exp\{-4(x-1)^2\}I(x > 0)$ *and the design points* x_i *are Gamma*(2).