Ann. Inst. Statist. Math.
Vol. 52, No. 3, 471–480 (2000)

PROBABILITY DENSITY FUNCTION ESTIMATION USING GAMMA KERNELS

SONG XI CHEN*

Department of Statistical Science, La Trobe University, Victoria 3083, Australia

(Received May 25, 1998; revised April 12, 1999)

Abstract. We consider estimating density functions which have support on $[0, \infty)$ using some gamma probability densities as kernels to replace the fixed and symmetric kernel used in the standard kernel density estimator. The gamma kernels are non-negative and have naturally varying shape. The gamma kernel estimators are free of boundary bias, non-negative and achieve the optimal rate of convergence for the mean integrated squared error. The variance of the gamma kernel estimators at a distance x away from the origin is $O(n^{-4/5}x^{-1/2})$ indicating a smaller variance as x increases. Finite sample comparisons with other boundary bias free kernel estimators are made via simulation to evaluate the performance of the gamma kernel estimators.

Key words and phrases: Boundary bias, gamma kernels, local linear estimators, variable kernels.

1. Introduction

This paper considers estimation of a probability density function that has bounded support on $[0, \infty)$. It is well known in nonparametric kernel density estimation that the bias of the standard kernel density estimator is of a larger order near the boundary than that in the interior. This bias phenomena is called boundary bias or edge effect. There are methods available for removing boundary bias such as data reflection (Schuster (1985)), boundary kernels (Müller (1991, 1993) and Müller and Wang (1994)), generating pseudodata (Cowling and Hall (1996)), hybrid method (Hall and Wehrly, (1991)), empirical transformation (Marron and Ruppert (1994)), the local linear estimator (Lejeune and Sarda (1992) and Jones (1993)), data binning and a local polynomial fitting on the bin counts (Cheng *et al.* (1997)) and others.

Recently, Brown and Chen (1999) and Chen (1999, 2000) have proposed using some beta density functions as kernels to estimate curves whose supports are compact intervals. In this paper the idea of beta kernel smoothing is extended to estimating densities whose supports are bounded from one end only. Two classes of gamma density functions are considered as kernels to formulate two density estimators. The gamma kernel estimators are free of boundary bias, always non-negative and achieve the optimal rate of convergence in the mean integrated square error within the class of non-negative kernel density estimators. The gamma kernel estimators have a unique feature not shared by the beta kernel estimators in that the variance reduces as the position where the smoothing is made moves away from the boundary. This feature is attractive in estimation of densities that have sparse areas.

* Now at Department of Statistics and Applied Probability, National University of Singapore, Singapore 117543, Singapore.

The paper is structured as follows. Two gamma kernel estimators are introduced and their bias is assessed in Section 2. The variance properties are studied in Section 3. In Section 4, the mean integrated squared errors and the optimal smoothing bandwidths are derived. A data set is analyzed as an example in Section 5. Section 6 presents results from a simulation study.

2. *Gamma kernel estimators and their bias*

Let $X_1, \ldots, X_n$ be a random sample from a distribution with an unknown probability density function f which is defined on $[0, \infty)$. We assume throughout the paper that f has a continuous second derivative, and that both $\int_0^\infty f'^2(x)dx$ and $\int_0^\infty \{xf''(x)\}^2 dx$ are finite. The standard kernel density estimator for f is

$$\hat{f}(x) = (nh)^{-1} \sum K\{h^{-1}(x - X_i)\} \tag{2.1}$$

where K and h are the kernel function and the smoothing bandwidth respectively. Comprehensive reviews of the kernel smoothing method are available in Silverman (1986) and Wand and Jones (1995). The kernel function K is usually symmetric and is regarded as less important than the smoothing bandwidth. However, a fixed symmetric kernel is not appropriate for fitting densities with bounded supports as it causes boundary bias. For the situation we are interested in, where the support of f is $[0, \infty)$, the expected value of $\hat{f}(x)$ at $x = 0$ is only half of the real density value.

The cause of boundary bias is due to that the fixed symmetric kernel which allocates weight outside the density support when smoothing is made near the boundary. A remedy is to use kernels that never assign weight outside the support. In the context of nonparametric regression, Chen (2000) proposes using the density of Beta$\{x/b + 1, (1 - x)/b + 1\}$ distribution as the kernels, where b is a smoothing bandwidth, to replace the fixed kernel in the Gasser-Müller estimator. In the case of $X_1, \ldots, X_n$ being confined within an interval, Chen (1999) studies the properties of density estimators using the beta kernels. The beta kernels provide a flexible family of kernel functions with varying shape and the amount of smoothing within the interval.

When the underlying density f is defined in $[0, \infty)$, the beta kernels should be replaced by gamma kernels as the gamma density functions have flexible shapes and locations within $[0, \infty)$. Let $K_{p,q}$ be the density function of a Gamma(p, q) random variable. The first class of gamma kernels considered are

$$K_{x/b+1,b}(t) = \frac{t^{x/b} e^{-t/b}}{b^{x/b+1}\Gamma(x/b + 1)}$$

where b is a smoothing parameter satisfying the condition that $b \to 0$ and $nb \to \infty$ as $n \to \infty$. The first gamma kernel estimator is

$$\hat{f}_1(x) = n^{-1} \sum_{i=1}^{n} K_{x/b+1,b}(X_i).$$

It is similar to the standard kernel estimator (2.1), only replaces the fixed symmetric kernel K with the gamma kernels. Notice that

$$E\{\hat{f}_1(x)\} = \int_0^\infty K_{x/b+1,b}(y) f(y) dy = E\{f(\xi_x)\}$$

where ξ_x is the gamma$(x/b+1, b)$ random variable. From the standard theory on gamma distribution, $\mu_x = E(\xi_x) = x + b$ and $\text{Var}(\xi_x) = xb + b^2$. By Taylor expansion,

$$E\{f(\xi_x)\} = f(\mu_x) + \frac{1}{2}f^{''}(x)\,\text{Var}(\xi_x) + o(b) \tag{2.2}$$
$$= f(x) + b\left\{f^{'}(x) + \frac{1}{2}xf^{''}(x)\right\} + o(b).$$

As the bias is $O(b)$ near the origin and in the interior, the density estimator is not subject to boundary bias.

The involvement of $f^{'}$ in the bias is less desirable, which is due to the fact that x is not the mean of the gamma kernel $K_{x/b+1,b}$; rather, it is the mode. It should be noted that x is the mean of $K_{x/b,b}$, but $K_{x/b,b}$ is unbounded near $x = 0$. A compromise would be to use $K_{x/b,b}$ for x in the interior and $K_{x/b+1,b}$ at $x = 0$. It needs a bridge to connect them smoothly. A simple choice is to use the kernels $K_{\rho_b(x),b}$ where

$$\rho_b(x) = \begin{cases} x/b & \text{if} \quad x \geq 2b; \\ \frac{1}{4}(x/b)^2 + 1 & \text{if} \quad x \in [0, 2b). \end{cases}$$

This leads to the second gamma kernel estimator:

$$\hat{f}_2(x) = n^{-1} \sum_{I=1}^{n} K_{\rho_b(x),b}(X_i).$$

Using the same method as in (2.2), it may be shown that

$$Bias\{\hat{f}_2(x)\} = \begin{cases} \frac{1}{2}xf''(x)b + o(b) & \text{if} \quad x \geq 2b; \\ \xi_b(x)bf'(x) + o(b) & \text{if} \quad x \in [0, 2b); \end{cases} \tag{2.3}$$

where $\xi_b(x) = (1-x)\{\rho_b(x) - x/b\}/\{1 + b\rho_b(x) - x\}$. Now $f^{'}$ is removed from the bias in the interior, and is only present in a small area near the origin but is compensated by the disappearance of $f^{''}$. As $\int_0^\infty \{xf^{''}(x)\}^2 dx < \infty$, $xf^{''}(x)$ converges to zero as $x \to \infty$. So, the bias will be smaller as x increases. Relative to the local linear and the boundary kernel estimators, the bias of the gamma kernel estimators may be larger as x is large; however this is compensated by a reduced variance as shown in the next section. The integrated squared bias is

$$IB^2\{\hat{f}_2(x)\} = \int_0^\infty Bias^2\{\hat{f}_2(x)\}dx = \frac{1}{4}b^2 \int_0^\infty \{xf''(x)\}^2 dx + o(b^2) \tag{2.4}$$

which does not involve $f^{'}$ in the leading term.

Figure 1 displays the kernels $K_{x/b+1,b}$ and $K_{\rho_b(x),b}$ for some selected x-values. There are two important features about the gamma kernels. One is that all the gamma kernels are non-negative which implies that the gamma kernel estimators are non-negative. The other is that the shape of the kernels changes according to the value of x. This varying kernel shape changes the amount of smoothing applied by the gamma kernel estimators as the variance of $K_{x/b+1,b}$ is $xb + b^2$, and that of $K_{\rho_b(x),b}$ has a similar form.

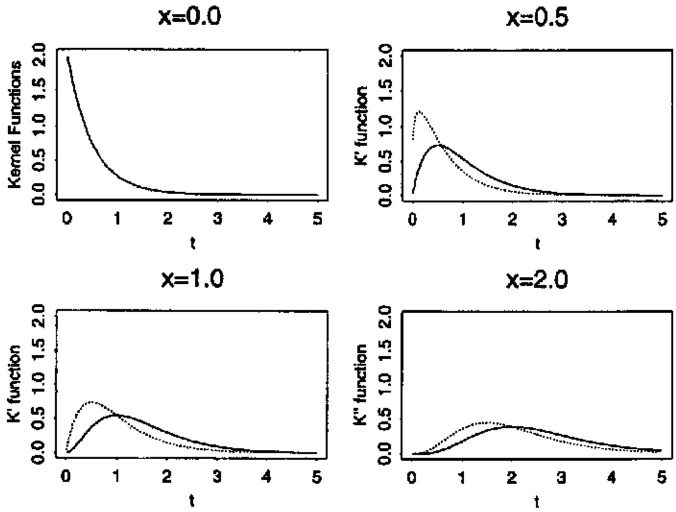

Fig. 1. Gamma kernels $K_{x/b+1,b}(t)$ (in solid lines) and $K_{\rho_b(x),b}(t)$ (in dotted lines) for $b = 0.2$.

3. Variance

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

(3.1) $$\mathrm{Var}\{\hat{f}_1(x)\} = n^{-1}\,\mathrm{Var}\{K_{x/b+1,b}(X_i)\} = n^{-1}E\{K_{x/b+1,b}(X_i)\}^2 + O(n^{-1}).$$

Let η_x be a gamma$(2x/b+1, b)$ random variable. Then,

$$E\{K_{x/b+1,b}(X_i)\}^2 = B_b(x)E\{f(\eta_x)\}$$

where

(3.2) $$B_b(x) = \frac{b^{-1}\Gamma(2x/b+1)}{2^{2x/b+1}\Gamma^2(x/b+1)}.$$

Let $R(z) = \sqrt{2\pi}e^{-z}z^{z+1/2}/\Gamma(z+1)$ for $z \geq 0$. Expressing the gamma function appeared in (3.2) in terms of R, we have

(3.3) $$B_b(x) = \frac{b^{1/2}x^{-1/2}R^2(x/b)}{2\sqrt{\pi}R(2x/b)}.$$

According to Lemma 3 of Brown and Chen (1999), $R(z)$ is a monotonic increasing function which converges to 1 as $z \to \infty$ and $R(z) < 1$ for any $z > 0$. Thus, $R^2(x/b)/R(2x/b) < 1$ and

(3.4) $$B_b(x) \leq \frac{b^{1/2}x^{-1/2}}{2\sqrt{\pi}}$$

indicating that B_b is bounded from above by the term on the right hand side. Also implied from (3.3) is that for b small enough

$$B_b(x) \sim \begin{cases} \dfrac{1}{2\sqrt{\pi}}b^{-1/2}x^{-1/2} & \text{if } \ x/b \to \infty; \\ \dfrac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}b^{-1} & \text{if } \ x/b \to \kappa \end{cases}$$

for a non-negative constant κ. This gives

$$\mathrm{Var}\{\hat{f}_1(x)\} \sim \begin{cases} \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} x^{-1/2} f(x) & \text{if } x/b \to \infty; \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)} n^{-1} b^{-1} f(x) & \text{if } x/b \to \kappa. \end{cases}$$

The variance of $\hat{f}_2$ is similar, except that the multiplier in front of $n^{-1}b^{-1}$ in the case of $x/b \to \kappa$ has a slightly different form.

Later derivation shows that the optimal $b = O(n^{-2/5}) = O(h^2)$ where h is the optimal bandwidth used by the local linear and the boundary kernel estimators. This means that the asymptotic variance of $\hat{f}_1(x)$ is $O\{(nh^2)^{-1}\}$ in the boundary area and is a larger order than that of the other kernel estimators which have increased variance coefficients instead. However, it can be a entirely different story in finite sample situations. Let the variance coefficient function be the multiplier of $(nh)^{-1}f(x)$ or $(nb^{1/2})^{-1}f(x)$ in the leading term of the variance expansion. So, that for $\hat{f}_1$ denoted as $V_{gam1}(x;b)$ is just $\sqrt{b}B_b(x)$; and that for the local linear smoother using the Biweight kernel, denoted as $V_{lls}(x;h)$, is given in Jones (1993). To make the amount of smoothing in the same scale, we let $h = \sqrt{b}$. It can be shown by ploting the two variance coefficient functions, that when $b > 0.01$ $V_{gam1}(x;b) < V_{lls}(x;\sqrt{b})$ for all $x \in [0,\infty)$. It is only when $b \le 0.01$ $V_{gam1}(x;b)$ starts to be larger near zero, reflecting a larger asymptotic variance near the boundary, but smaller as x moves away from the boundary area. The exact sample size corresponding to $b = 0.01$ depends on the underlying density f. The above observations on the variance coefficients have been confirmed by the simulation reported in Section 6.

A unique feature for the gamma estimators is the variance coefficient decreases as x increases as $V_{gam1}(x;b) \sim \frac{1}{2\sqrt{\pi}}x^{-1/2}$. This is in contrast to other kernel estimators whose variance coefficients remind constant outside the boundary area; and is because the gamma kernels has a larger support $[0,\infty)$ than that of a compact kernel and consequently has a larger effective sample size. This larger effective sample size is desirable for estimating densities having sparse areas as more data points can be pooled to smooth in areas with fewer observations. The reduced variance when x is large is balanced by an increased bias as shown in the previous section. The direct involvement of the factor x disappears in the optimal mean square errors. To appreciate this, note that the mean square error of $\hat{f}_2(x)$ for $x/b \to \infty$ is

$$MSE\{\hat{f}_2(x)\} = \frac{1}{4}\{xf''(x)\}^2 b^2 + \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} x^{-1/2} f(x) + o\{b^2 + (nb)^{-1/2}\}.$$

It can be shown that the optimal mean square error is

$$\frac{5}{4^{4/5}} \left\{ \frac{1}{2\sqrt{\pi}} f(x) \right\}^{4/5} \{f''(x)\}^{2/5} n^{-4/5}, \tag{3.5}$$

which is the same as that of the standard kernel density estimator using the Gaussian kernel. So, the gamma kernels in $[0,\infty)$ are in a sense equivalent to the Gaussian kernel in $(-\infty,\infty)$.

4. Global Properties

The increase in the variance near the boundary has negligible impact on the integrated variance. To appreciate this, define $\delta = b^{1-\epsilon}$ where $0 < \epsilon < 1$ and notice that for $i = 1$ or 2,

$$\begin{aligned}(4.1)\qquad \int_0^\infty \mathrm{Var}\{\hat{f}_i(x)\}dx &= \int_0^\delta + \int_\delta^\infty \mathrm{Var}\{\hat{f}_i(x)\}dx \\ &= \int_\delta^\infty \frac{1}{2\sqrt{\pi}}\, x^{-1/2}n^{-1}b^{-1/2}f(x)dx + O(n^{-1}b^{-\epsilon}) \\ &= \frac{1}{2\sqrt{\pi}}n^{-1}b^{-1/2}\int_0^\infty x^{-1/2}f(x)dx + o(n^{-1}b^{-1/2})\end{aligned}$$

by choosing ϵ properly and noting that $\int_0^\infty x^{-1/2}f(x)dx$ is finite.

Combining (2.2), (2.4) and (4.1),

$$\begin{aligned}(4.2)\qquad MISE(\hat{f}_1) = b^2 \int_0^\infty &\left\{xf'(x) + \frac{1}{2}xf''(x)\right\}^2 dx \\ &+ \frac{1}{2\sqrt{\pi}}n^{-1}b^{-1/2}\int_0^\infty x^{-1/2}f(x)dx + o(n^{-1}b^{-1/2} + b^2)\end{aligned}$$

and

$$\begin{aligned}(4.3)\qquad MISE(\hat{f}_2) = \frac{1}{4}b^2 &\int_0^\infty \{xf''(x)\}^2 dx + \frac{1}{2\sqrt{\pi}}n^{-1}b^{-1/2}\int_0^\infty x^{-1/2}f(x)dx \\ &+ o(n^{-1}b^{-1/2} + b^2).\end{aligned}$$

The optimal bandwidths which minimize the leading terms in (4.2) and (4.3) are

$$b_1^* = \frac{\left[\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}f(x)dx\right]^{2/5}}{4^{2/5}\left[\int_0^\infty \left\{xf'(x) + \frac{1}{2}xf''(x)\right\}^2 dx\right]^{2/5}}n^{-2/5} \quad \text{and}$$

$$b_2^* = \frac{\left[\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}f(x)dx\right]^{2/5}}{\left[\int_0^\infty \{xf''(x)\}^2 dx\right]^{2/5}}n^{-2/5}.$$

So, the optimal bandwidths are $O(n^{-2/5})$ as compared $O(n^{-1/5})$ for the other kernel estimators. Substituting the above optimal bandwidths, the optimal mean integrated squared errors are

$$\begin{aligned}MISE^*(\hat{f}_1) &= \frac{5}{4^{4/5}}\left[\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}f(x)dx\right]^{4/5} \\ &\quad \cdot \left[\int_0^\infty \left\{xf'(x) + \frac{1}{2}xf''(x)\right\}^2 dx\right]^{1/5} n^{-4/5} \quad \text{and}\end{aligned}$$

$$MISE^*(\hat{f}_2) = \frac{5}{4^{4/5}}\left[\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}f(x)dx\right]^{4/5}\left[\int_0^\infty \{xf''(x)\}^2 dx\right]^{1/5} n^{-4/5}.$$

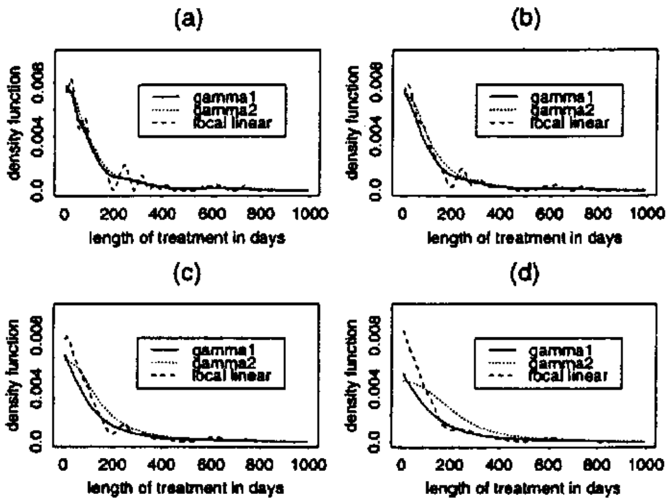

Fig. 2. Estimated density curves using the gamma, the local linear and the boundary kernel estimator for the suicide data. The bandwidths used in panel (b) are given by LSCV, and those used in (a) are half of those used in (b).

Thus, both $\hat{f}_1$ and $\hat{f}_2$ achieve the optimal rate of convergence for the mean integrated squared error for using non-negative kernels.

It may be shown that for any density function f, if both $\int_0^\infty \{f'(x)\}^2 dx$ and $\int_0^\infty \{f''(x)\}^2 dx$ are finite, then

$$\int_0^\infty \left\{f'(x) + \frac{1}{2}xf''(x)\right\}^2 dx \geq \int_0^\infty \left\{\frac{1}{2}xf''(x)\right\}^2 dx.$$

This means that $MISE^\star(\hat{f}_1) \geq MISE^\star(\hat{f}_2)$ and $b_1^\star \geq b_2^\star$. Therefore, $\hat{f}_2$ should have a better global performance and use a smaller bandwidth than $\hat{f}_1$.

5. An example

We apply the gamma kernel estimators on the suicide data given in Silverman ((1986), p. 8.), which consists of 86 observations on the treatment lengths (in days) of patients in a suicide study. Müller and Zhou (1991) employed variable bandwidth boundary kernel estimation to the data by applying different amount of smoothing at different location. Clearly the underlying density function has support on $[0, \infty)$. The standard kernel estimator will suffer from boundary bias as there are relatively large number observations near $x = 0$. We apply the gamma kernel estimators on the data set along with the local linear estimator and the boundary kernel estimator using kernels proposed in Müller and Wang (1994). We do not consider the Jones-Foster estimator as the local linear estimates are well above zero near the boundary $x = 0$ due to many observations there.

The least squares cross-validation (LSCV) was used to choose the bandwidths. All the data points with equal weights were used in computing the LSCV score functions for a grid of bandwidth values. The LSCV score functions for the two gamma estimators had distinct minimums at $b_{1cv} = 19.4$ and $b_{2cv} = 27.7$ respectively. However, those of the local linear and the boundary kernel estimators had a very wide flat valley for

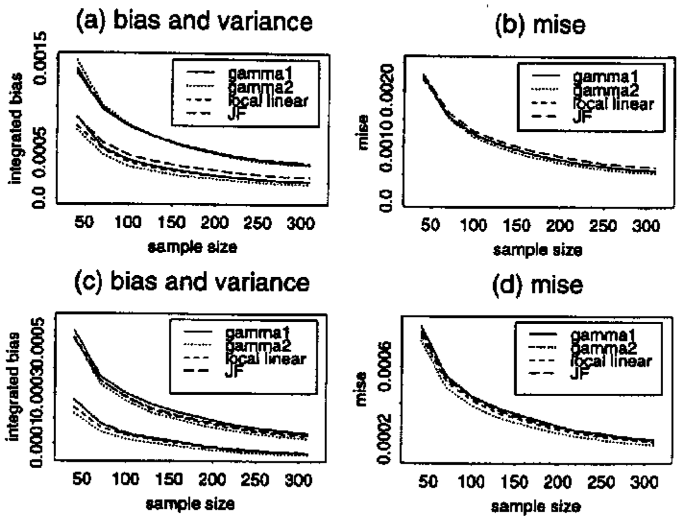

Fig. 3. Integrated square bias and variance, and the mean integrated square errors for gamma(2, 1) in (a) and (b) and gamma(3, 1) in (c) and (d) densities. In (a) and (c) the integrated square bias are below the integrated variance.

$h \in [70, 250]$ and were minimized at $h_{1cv} = 195$ and $h_{2cv} = 171$ respectively which seemed to be large. This certainly confirmed the existed concern for the efficiency of LSCV. It was based on this concern, in additional to the bandwidths prescribed by the LSCV, we also used bandwidths values half of those prescribed by the LSCV. We see in Fig. 2 that the estimates using LSCV bandwidths in panel (b) seem to oversmooth whereas that given in panel (a) seem to be quite reasonable. The bumps near $x = 300$ and 600 were picked up by the density estimates in (a). There were no much difference among the estimates except near the origins where the second gamma estimator had a slight shoulder near the origin, and the two non-gamma estimators were slightly more responsive to the two bumps in panel (a) as they used compact kernels.

6. Simulation results

Five estimators were considered in the simulation study: the two gamma kernel estimators, the local linear estimator of Jones (1993), the non-negative estimator of Jones and Foster (1996) and the boundary kernel estimators of Müller and Wang (1994). Random samples were generated from gamma(2, 1) and gamma(3, 1) distributions using the routine given in Press *et al.* (1992). The size of the random sample ranged from 40 to 310, and the number of simulation was 1000.

For each simulated sample and each estimator considered, the smoothing bandwidth was chosen by directly minimizing the integrated squared error

$$ISE(b) = \int_0^9 \{\hat{f}(x) - f(x)\}^2 \, dx$$

as the densities virtually having zero values outside $[0, 9]$. The minimization of the integrated squared error with respect to the smoothing bandwidth was carried out by the golden search algorithm given in Press *et al.* (1992). By substituting the optimal bandwidths, the average integrated squared bias, the integrated variance and the integrated squared errors were calculated as measures of performance for each of the estimators.

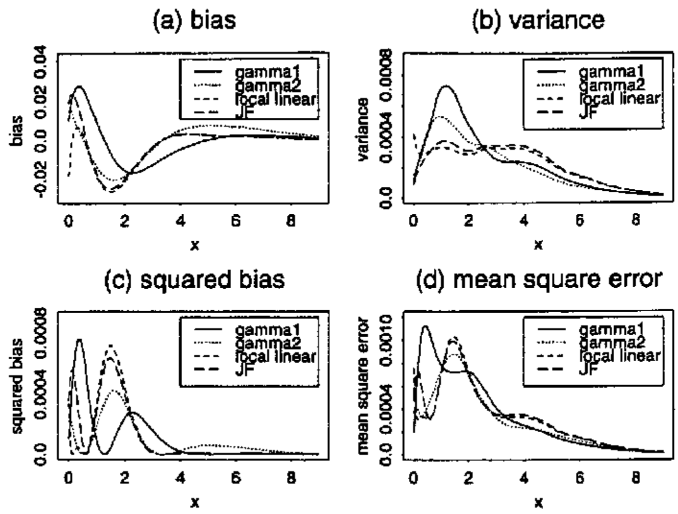

Fig. 4. Point-wise bias and variance for estimating the density of Gamma(3, 1) for $n = 160$.

Figure 3 presents the average integrated square bias and variance, and the average integrated square errors for the estimators considered. To make the figure less crowded and considering that $\hat{f}_2$ has smaller mean integrated square error than $\hat{f}_1$ as revealed in Section 4, we will not report the results of the first gamma kernel estimator $\hat{f}_1$. However, the simulation showed that $\hat{f}_1$ performed similar to the Jones-Foster estimator. We observe in Fig. 3 that the bias, the variance and the integrated squared errors for the four estimators were all decreasing and seem to converge as the sample size increased. Even though the overall differences in the three measures of performance among the estimators were not large, there were some noticeable differences. Firstly, $\hat{f}_2$ had the best performance for almost all the cases considered except for gamma(2, 1) when $n = 40$. The second best performer was the local linear estimator followed by the Jones-Foster estimator. The local linear had much smaller bias and slightly smaller variance than its non-negative modification. It was surprising to see the boundary kernel estimator had the largest average integrated bias, variance and square errors for almost all the cases considered. The average bandwidth used by the estimator was larger than the local linear and the Jones-Foster estimators in all the case considered.

In Fig. 4 we present the point-wise bias, variance and mean square errors of the four estimators for $x \in [0, 9]$ for the gamma(3, 1) density and $n = 160$. We found the gamma kernel estimator had smaller variance than the other estimators when $x > 3$, confirming the theory that the variance is $O(n^{-1}b^{-1/2}x^{-1/2})$. This smaller variance was earned at the expense of the bias, as revealed in (a) and (c) of Fig. 4 where the corresponding biases were larger. However, the price was worthwhile as the gamma kernel estimator had smaller mean square errors when $x > 3$. When $x < 3$ there was no clear winner among the estimators. The mean square error of $\hat{f}_2$ was very reasonable compared with the other estimators. The negative bias of the local linear and the boundary kernel estimators near $x = 0$ indicates it takes negative values. It is interesting to see the gamma estimator did not have significantly larger variances at $x = 0$. In fact the variance was much smaller than the non-negative local linear and the boundary kernel estimators near $x = 0$. This reassures the findings in Section 3.

Acknowledgements

The author thanks two referees for constructive comments that improve the presentation of the paper.

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). Beta kernel estimators for density functions, *Comput. Statist. Data Anal.*, **31**, 131–145.

Chen, S. X. (2000). Beta kernel smoothers for regression curves, *Statist. Sinica*, **10**, 73–91.

Cheng, M. Y., Fan, J. and Marron, J. S. (1997). On automatic boundary corrections, *Ann. Statist.*, **25**, 1691–1708.

Cowling, A. and Hall, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation, *J. R. Statist. Soc. B*, **58**, 551–563.

Hall, P. and Wehrly, T. E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators, *J. Amer. Statist. Assoc.*, **86**, 665–672.

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

Jones, M. C. and Foster, P. J. (1996). A simple nonnegative boundary correction method for kernel density estimation, *Statist. Sinica.*, **6**, 1005–1013.

Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions, *Comput. Statist. Data Anal.*, **14**, 457–471.

Marron, J. S. and Ruppert, D. (1994). Transformation to reduce boundary bias in kernel density estimation, *J. R. Statist. Soc. B*, **56**, 653–671.

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

Müller, H.-G. (1993). On the boundary kernel method for non-parametric curve estimation near endpoints. *Scand. J. Statist.*, **20**, 313–328.

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

Müller, H.-G. and Zhou, H. (1991). A discussion to "Transformations in density estimation" by Wand, M., Marron, J. S. and Ruppert, D., *J. Amer. Statist. Assoc.*, **86**, 356–358.

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

Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities, *Commun. Statist. Theory Methods*, **14**, 1123–1136.

Silverman, B. W. (1986). *Density Estimation*, Chapman and Hall, London.

Wand, M. and Jones, M. C. (1995). *Kernel Smoothing*, Chapman and Hall, London.