Sankhyā: The Indian Journal of Statistics
1995, Volume 57, Series A, Pt. 1, pp. 56–67

KERNEL TYPE DENSITY ESTIMATES FOR POSITIVE VALUED RANDOM VARIABLES

By ISHA BAGAI and B.L.S. PRAKASA RAO
Indian Statistical Institute

SUMMARY. We propose kernel type estimators for the density function of non negative random variables, where the kernel function is a probability density function on $(0,\infty)$. Properties of these estimators are discussed. A kernel, that minimizes the integrated mean square error is obtained. It is shown, however, that any reasonable kernel gives almost the same mean square error. On the basis of simulation studies the use of exponential kernels is recommended.

1. Introduction

Silverman (1986) used two data sets to illustrate various techniques of density estimation where the random variables take only positive values. Since the random variable of interest is positive its density function has support on $(0,\infty)$. Hence, an estimate of the density should have support on $(0,\infty)$.

The most commonly used density estimator is the kernel estimator, which has been extensively studied (see, for example, Rosenblatt (1956), Parzen (1962), Prakasa Rao (1983), Silverman (1986)). The kernel estimator based on a random sample $X_1, X_2, \ldots, X_n$ from a population with density function f is defined by

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K^*\left(\frac{x-X_i}{h_n}\right), \forall\ x \in \mathbb{R} \qquad \ldots (1.1)$$

where h_n is called the bandwidth and K^* is the kernel function. In practice, $\{h_n\}$ is chosen so that $h_n \to 0$ as $n \to \infty$ and the kernel function $K^*(\cdot)$ itself is a symmetric probability density function on the entire real line. If K^* is continuous and differentiable then $\hat{f}_n(x)$ will also be continuous and differentiable. But $\hat{f}_n(x)$ might take positive values even for $x \in (-\infty, 0]$, which is not desirable when the random variable is positive. Silverman (1986) mentions some adaptations of the existing methods when the support of the density to be estimated is not the whole real line.

Paper received. December 1990; revised July 1994.

AMS (1991) *subject classification.* Primary 62G05; secondary 62 N05.

Key words and phrases. Density estimation, Kernel-type estimators, reliability, standard exponential kernels.

Here, we propose and discuss the properties of kernel estimators for densities with support on the positive half of the real line. We use them to estimate the density function for a data set considered by Silverman (1986). Finally, some simulation studies are performed to compare the density estimates with the original density functions.

2. Kernel estimation

2.1 *Preliminaries.* Let X_1, X_2, ..., X_n be a random sample of size n from a population with density function f with support on $(0, \infty)$. Further, let $K(\cdot)$ be a bounded density function with support on $(0, \infty)$, satisfying the condition

$$\int_0^\infty x^2 K(x) dx < \infty. \qquad \dots (2.1)$$

Here, we propose a kernel type estimator for $f(x)$ given by

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), x \geq 0 \qquad \dots (2.2)$$

where $\{h_n\}$ is a sequence chosen such that $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.

It can be easily seen that $f_n(x)$ is a probability density function with support on $(0, \infty)$. The only difference between $\hat{f}_n(x)$ and $f_n(x)$ is that $\hat{f}_n(x)$ is based on a kernel possibly with support extending beyond $(0, \infty)$.

The estimator $f_n(x)$ can also be written in the form

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_{(i)}}{h_n}\right)$$

where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are the ordered observations. Thus, it is easily seen that for x such that $X_{(r)} < x \leq X_{(r+1)}$, only the first r order statistics contribute to the value of the estimator $f_n(x)$ of the density function $f(x)$.

Assume that f is twice continuously differentiable on $(0, \infty)$. Then, following Parzen (1962),

$$E[f_n(x)] \simeq f(x) - h_n f'(x)\gamma_1 + \frac{h_n^2}{2} f''(x)\gamma_2, \qquad \dots (2.3)$$

where $\gamma_j = \int_0^\infty x^j K(x) dx, \quad j = 1, 2.$

Hence, the bias is

$$B_n(x) \simeq -h_n f'(x)\gamma_1 + \frac{h_n^2}{2} f''(x)\gamma_2. \qquad \dots (2.4)$$

Furthermore, the variance of $f_n(x)$ is

$$\begin{aligned} V_n(x) &\simeq \frac{1}{nh_n}[f(x)\beta_0 - h_n f'(x)\beta_1 + \frac{h_n^2}{2} f''(x)\beta_2] \\ &- \frac{1}{n}[B_n(x) + f(x)]^2 \end{aligned} \qquad \dots (2.5)$$

where, $\beta_j = \int_0^\infty x^j K^2(x)dx, \quad j = 0, 1, 2.$

Therefore,

$$V_n(x) \simeq \frac{1}{nh_n} f(x)\beta_0 + 0\left(\frac{1}{n}\right). \qquad \ldots (2.6)$$

From (2.3) and (2.6) it follows that $E[f_n(x)] \to f(x)$ and $V_n(x) \to 0$ when $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. Thus $f_n(x)$, like the usual kernel estimator $\hat{f}_n(x)$, is asymptotically unbiased and weakly consistent for $f(x)$. It is easy to see that $[f_n(x) - E(f_n(x))]/[V_n(x)]^{\frac{1}{2}}$ is asymptotically normally distributed, (see for example, Prakasa Rao, 1983, p. 61).

The mean square error (MSE) of $f_n(x)$, denoted by $M_n(x)$, is given by

$$M_n(x) \simeq \frac{1}{nh_n} f(x)\beta_0 + h_n^2 (f'(x))^2 \gamma_1^2, \qquad \ldots (2.7)$$

and the integrated mean square error (IMSE) of $f_n(x)$ is

$$\begin{aligned} Q_n(x) &= \int_0^\infty M_n(x)dx \\ &\simeq \frac{1}{nh_n}\beta_0 + h_n^2\gamma_1^2 \int_0^\infty (f'(x))^2 dx. \end{aligned} \qquad \ldots (2.8)$$

2.2 *Optimal bandwidth and optimal kernel.* Here we consider the optimal choice of the sequence h_n and the kernel $K(\cdot)$, the one which minimizes the IMSE. Let $h_n = kn^{-\alpha}, \alpha > 0, k > 0$. Then, following Prakasa Rao (1983, 34-35), we get that IMSE is minimized when $\alpha = 1/3$ and the optimal choice of k is

$$k^* = \left(\frac{\beta_0}{2\gamma_1^2 \int_0^\infty (f'(x))^2 dx}\right)^{1/3}. \qquad \ldots (2.9)$$

With the above choice of k^* and α, the optimal value of h_n is

$$h_n(opt.) = n^{-1/3}\left(\frac{\beta_0}{2\gamma_1^2 \int_0^\infty (f'(x))^2 dx}\right)^{1/3}, \qquad \ldots (2.10)$$

and the minimum IMSE is given by

$$2^{-4/3} 3 n^{-2/3} \beta_0^{2/3} \gamma_1^{2/3} [\int_0^\infty (f'(x))^2 dx]. \qquad \ldots (2.11)$$

It is interesting to note that h_n(opt.) and hence minimum IMSE depend on the unknown density function, which is also true for h_n^* (opt.) for $\hat{f}_n(x)$. The usual approach is to choose h_n^* with reference to some standard density. The normal density is often chosen leading to the kernel $K^*(x)$ in the cases discussed by Silverman (1986). One could similarly choose the exponential density when the data is restricted to the positive half of the real line for determining h_n. Another approach is to make a data driven choice for h_n as in the classical density estimation case discussed by Duin (1976) and Stone (1984) among others. Or, as a referee has suggested, one could first estimate the terms in h_n(opt.) consistently and then by weak convergence arguments get an efficient estimator. These approaches need to be investigated.

The optimal choice of the kernel $K(x)$ can be found as in Epanenchnikov (1969). Minimize

$$\begin{aligned} C(K) &= (\gamma_1\beta_0)^{2/3} \\ &= [\int_0^\infty zK(z)dz]^{2/3}[\int_0^\infty K^2(z)dz]^{2/3} \end{aligned} \qquad \ldots (2.12)$$

subject to $\gamma_1 = a$ (prefixed) where K satisfies (2.1). Then, by calculus of variations, the optimal kernel is

$$K_{\text{opt}}(x) = \begin{cases} \dfrac{2}{3a} - \dfrac{2x}{9a^2} &, \text{ if } 0 < x \le 3a, \\ 0 &, \text{ otherwise.} \end{cases} \qquad \ldots (2.13)$$

Consider the following kernels

$$K_E(x) = \begin{cases} \dfrac{1}{a} e^{-x/a} &, x \ge 0, \\ 0 &, \text{otherwise,} \end{cases}$$

and

$$K_U(x) = \begin{cases} \dfrac{1}{2a} &, 0 \le x \le 2a, \\ 0 &, \text{otherwise.} \end{cases}$$

Then $\int_0^\infty K_{\text{opt}}^2(x)dx = \frac{4}{9a}$ and $\int_0^\infty K_E^2(x)dx = \int_0^\infty K_U^2(x)dx = \frac{1}{2a}$.

Therefore, the relative efficiency is

$$r = \frac{\int_0^\infty K_E^2(x)dx}{\int_0^\infty K_{\text{opt}}^2(x)dx} = \frac{\int_0^\infty K_U^2(x)dx}{\int_0^\infty K_{\text{opt}}^2(x)dx} = 1.125.$$

Hence, $K_{\text{opt}}(x)$ is optimal in the above sense but $K_E(x)$ and $K_U(x)$ are almost as good.

2.3 *Comparison between $f_n(x)$ and $\hat{f}_n(x)$.* It would be interesting to compare h_n(opt.) corresponding to $f_n(x)$ and h_n^*(opt.) for $\hat{f}_n(x)$. Observe that h_n(opt.) given by (2.10) is of order $n^{-1/3}$ and h_n^*(opt.) for $\hat{f}_n(x)$ is of order $n^{-1/5}$ (see, Prakasa Rao, 1983, page 67). If a kernel $K(x)$ with support contained in $(0, \infty)$ is used, then it ensures that $f_n(x)$ has its support on $(0, \infty)$ and is also a probabilty density. This is not the case if one uses $K^*(x)$ whose support extends possibly beyond $(0,\infty)$. It should be noted that the rate of convergence can always be improved if one does not insist on non-negative kernels (Prakasa Rao (1983)).

Further, it is well known that (see, Prakasa Rao (1983)) $\text{var}(\hat{f}_n(x)) \simeq \frac{1}{nh_n} f(x) \int_{-\infty}^{\infty} K^{*^2}(y)dy$. Then the ratio of $M_n(x)$, the MSE $(f_n(x))$ to $\text{var}(\hat{f}_n(x))$ is given by

$$\frac{M_n(x)}{\text{var}(\hat{f}_n(x))} = \frac{\beta_0}{a_0} + \frac{nh_n^3(f'(x))^2\gamma_1^2}{f(x)a_0}, \qquad \ldots (2.14)$$

where $f_n(x)$ is computed using kernel K and $\hat{f}_n(x)$ is computed using kernel K^* and the same bandwidth $\{h_n\}$ is used. Here,

$$a_0 = \int_{-\infty}^{\infty} K^{*^2}(y)dy.$$

For the optimal choice of $K(x)$ given in (2.13), that of $K^*(x)$ given by Epanechnikov (1969) and h_n given by (2.10), we can see that

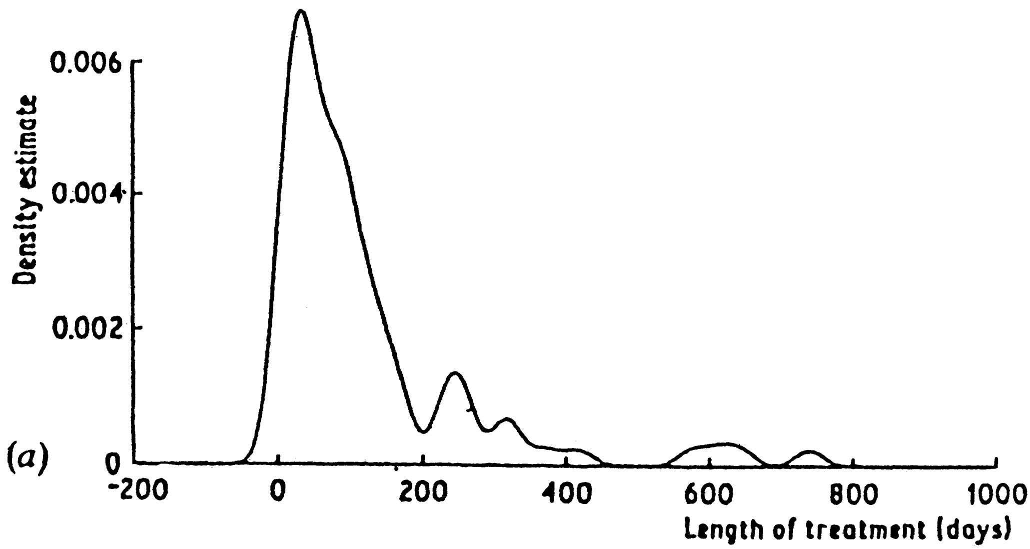
$$\frac{M_n(x)}{\text{var}(\hat{f}_n(x))} = \frac{20\sqrt{5}}{27a}\left\{1 + \frac{(f'(x))^2}{2f(x)\int_0^{\infty}(f'(x))^2)dx}\right\} \qquad \ldots (2.15)$$

3. An example

Here we consider a data set used by Silverman ((1986), page 8) to exhibit the performance of the optimum kernel function $K_{\text{opt}}(x)$ and the standard exponential kernel function $K_{SE}(x)$. The data set consists of lengths of 86 spells of psychiatric treatment undergone by patients used as controls in a study of the relationship between suicide risk and time under treatment. A spell of treatment refers to a period of continous contact with the psychiatric services, including times at which the patient is in the community and attending out-patient appointments. Many of the patients had more than one spell of treatment. The data is given below (see Silverman (1986)).

Table 3.1. Length of treatment spells (in days) of control patients in suicide study

1	1	1	5	7	8	8	13	14	14	17
18	21	21	22	25	27	27	30	30	31	31
32	34	35	36	37	38	39	39	40	49	49
54	56	56	62	63	65	65	67	75	76	79
82	83	84	84	84	90	91	92	93	93	103
103	111	112	119	122	123	126	129	134	144	147
153	163	167	175	228	231	235	242	256	256	257
311	314	322	369	415	573	608	640	737		



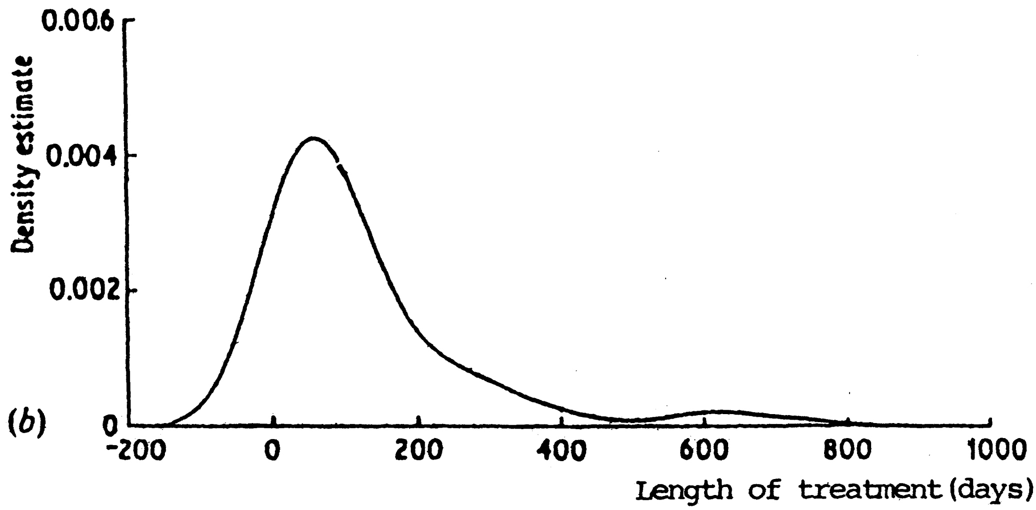


Fig.3.1 Kernel estimates for suicide study data. Bandwidth :
(a) 20 (b) 60

(Both the figures have been taken from Silverman (1986))

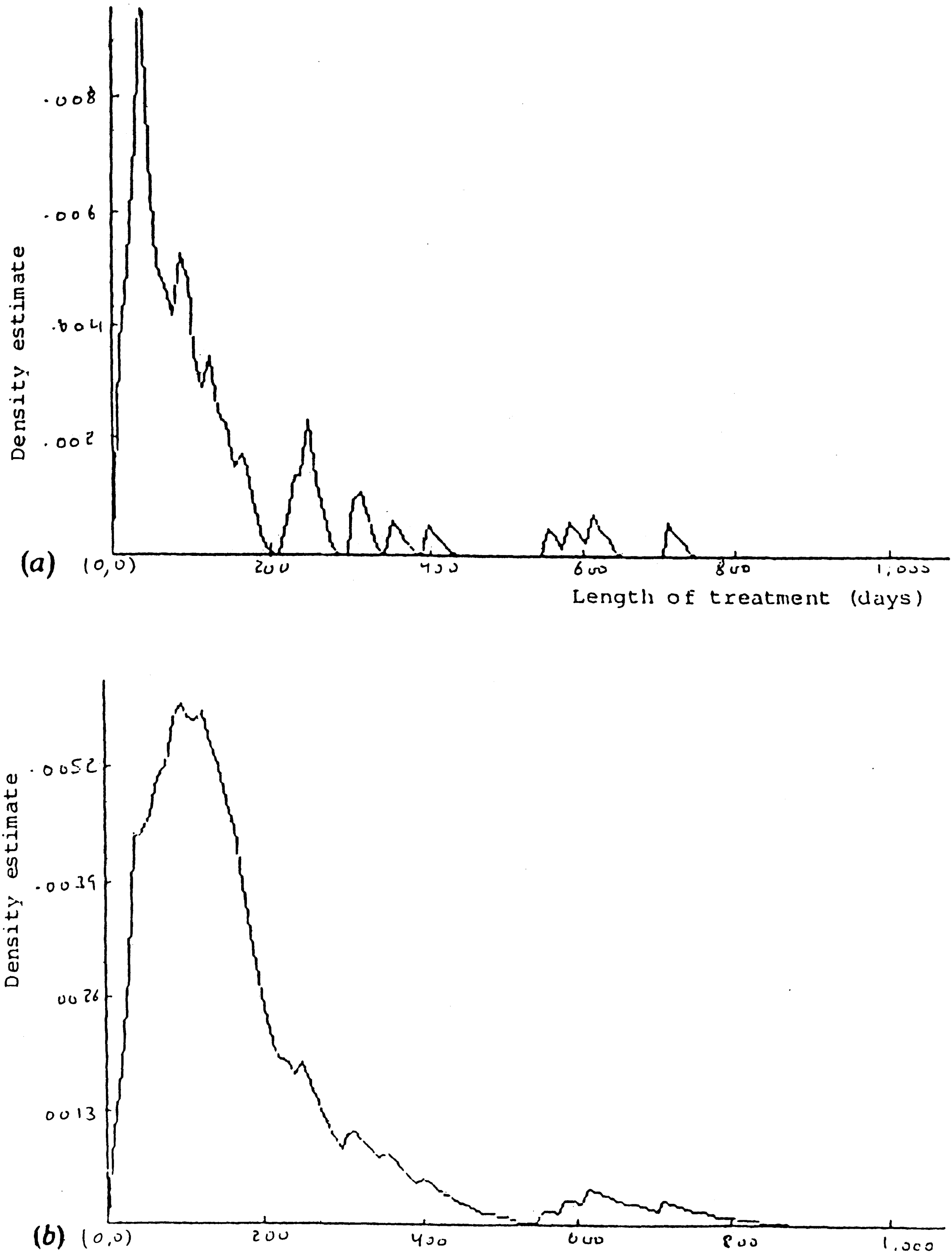


Fig.3.2 Kernel estimates for suicide study data with optimal kernel function. Bandwidth (a) 20 (b) 60

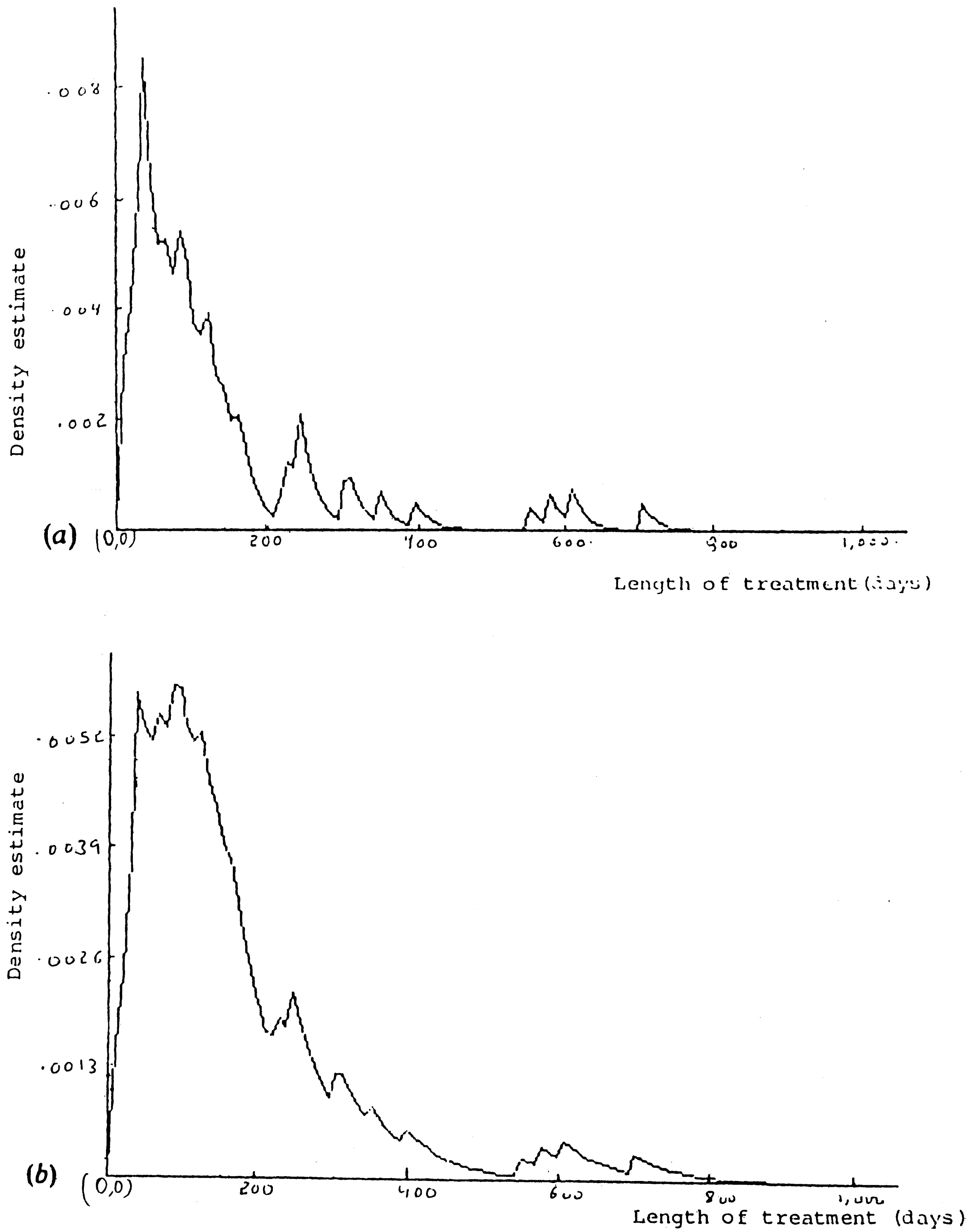


Fig. 3.3 Kernel estimates for suicide study data with standard exponential kernel function. Bandwidth (a) 20; (b) 60.

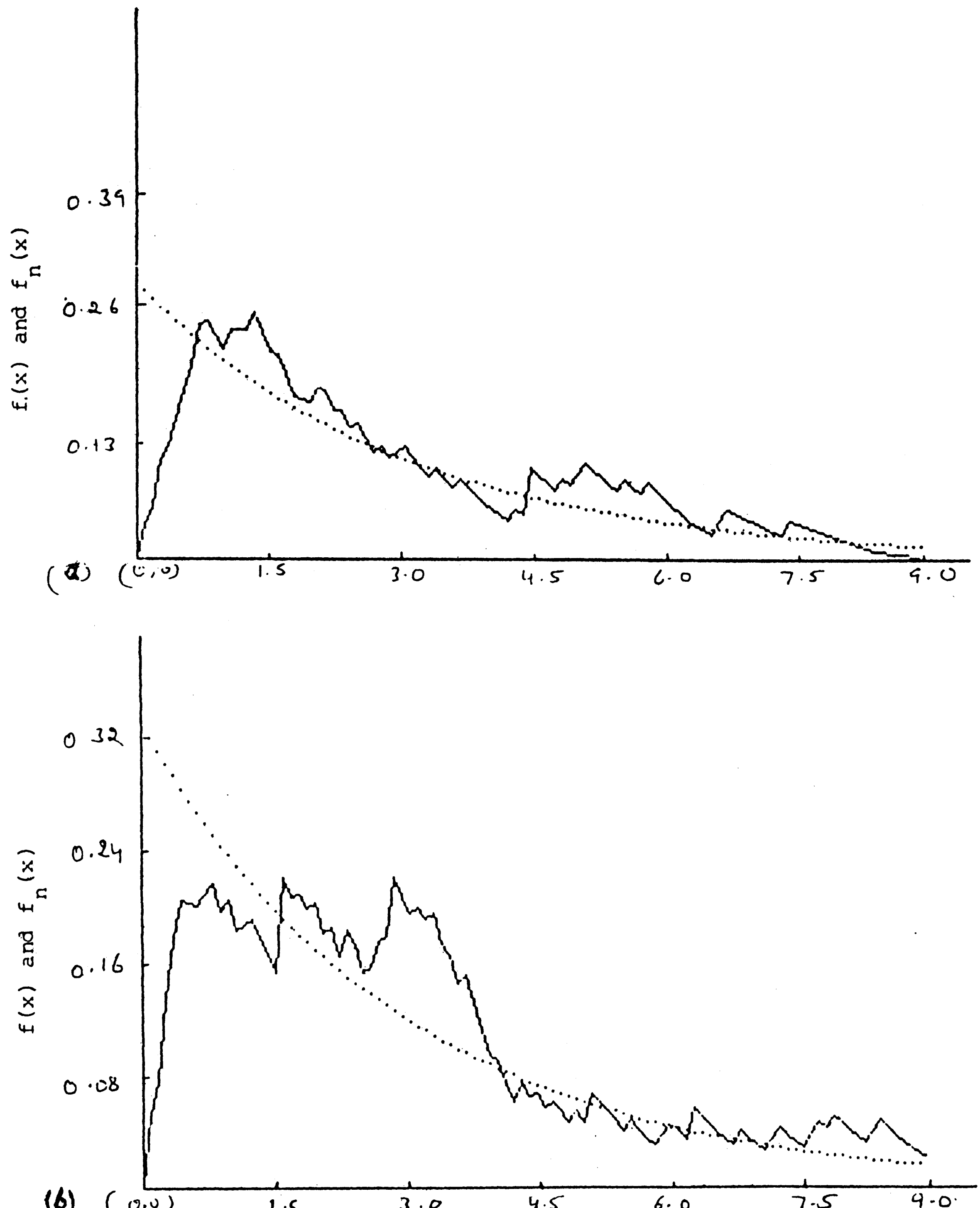


Fig. 4.1 Exponential density function (f(x)) with mean a (...) and its kernel estimat $f_n(x)$ with optimal kernel function (—

Sample Size (a) n = 50, (b) n = 100.

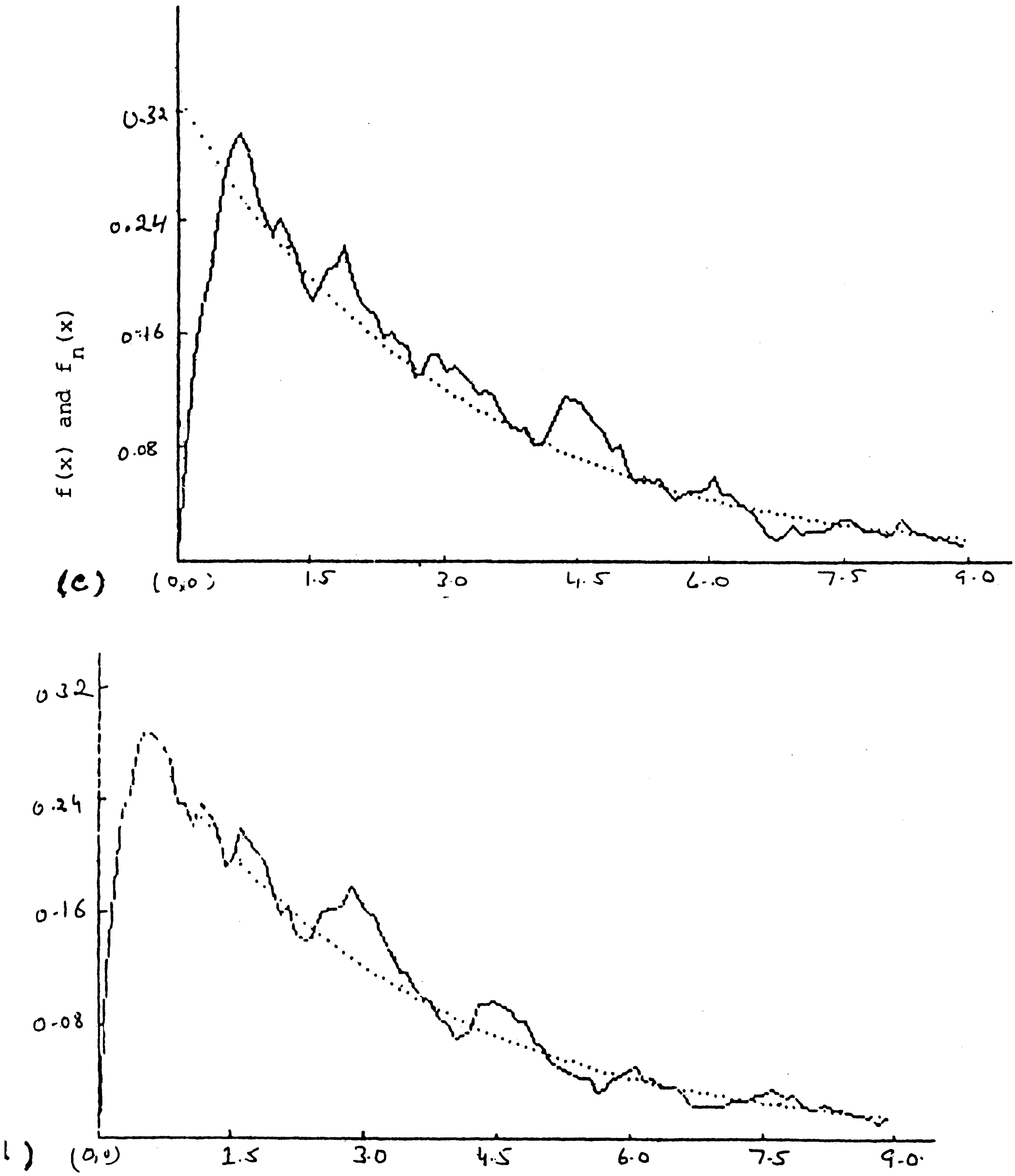


Fig. 4.2 Exponential density function (f(x)) with mean a (...) and its kernel estimate $f_n(x)$ with optimal kernel function (——). Sample Size (c) n = 500, (d) n = 1000.

Figure 3.1 (taken from Silverman (1986), page 18) shows the density estimator $\hat{f}_n(x)$. Figures 3.2 and 3.3 give $f_n(x)$ for the suicide data with the bandwidth h_n as 20 and 60 and kernel function $K_{\text{opt}}(x)$ and $K_{SE}(x)$ ($K_E(x)$ with $a = 1$, i.e. the standard exponential kernel).Comparing $\hat{f}_n(x)$ with $f_n(x)$ we see that the shape of $f_n(x)$ is essentially like that of $\hat{f}_n(x)$. If h_n is too small, then some additional peaks are observed, where as when h_n is too large then the density curve is smooth. Observe that the optimum kernel $K_{\text{opt}}(x)$ involves the parameter a which is also the mean of the distribution with density $K_{\text{opt}}(x)$. The value of a can be chosen by estimating the mean from the data. Note that there are quite a few observations close to the origin which result in some under smoothing near the origin. However, the essential difference between $f_n(x)$ and $\hat{f}_n(x)$ is that there is no loss of area when $f_n(x)$ is used, whereas some area under the curve $\hat{f}_n(x)$ lies to the left of the origin. This is not desirable as density estimators for positive valued random variables should have support $(0, \infty)$. But, if one ignores the area to the left of the origin, then $\hat{f}_n(x)$ would not integrate out to 1.

4. Simulation results

It would be interesting to compare $f_n(x)$ using the optimal kernel function with the actual density function $f(x)$. Random samples of size n were generated from

$$f(x) = \frac{1}{a} \exp\left[\frac{-x}{a}\right], x \geq 0, a \geq 0, \qquad \ldots (4.1)$$

for various choices of mean a. The density $f_n(x)$ was computed using the optimal kernel function with mean a and $h_n = o(n^{-1/3})$. Both the curves $f(x)$ and $f_n(x)$ were plotted on the same graph. The computer program is available from the authors.

The graphs with $a = 3$ and $n = 50, 100, 500$ and 1,000 are given below. See Fig. 4.1 and 4.2. The kernel type density estimator $f_n(x)$ looks like the original density function $f(x)$, and it gets closer and closer to $f(x)$ as the sample size increases.

5. Concluding section

Thus, when one is restricted to lifetimes or related data in reliability and survival analysis, one would recommend the use of the optimal kernel given by (2.13) or the standard exponential kernel function for estimating the density.

However, there are several other ways of handling positive data. For instance, one could use the usual kernel estimate, truncate at zero and rescale the density for the positive part; or one could transform the original data, say, by a logarithmic transformation, and then retransform the estimated density back. Relative merits and demerits of these techniques with the one suggested by us need to be investigated.

Acknowledgements. The authors are grateful to Dr. K. Balasubramaniam for his help in the computation of various density estimates and the referee for his suggestions.

REFERENCES

DUIN, R.P.W. (1976). 'On the choice of smoothing parameters for Parzen estimators of probability density functions', *IEE-Trans.Comput.*, **C-25**, 1175-1179.

EPANECHNIKOV, V.A. (1969). 'Nonparametric estimation of a multidimensional probability density', *Theor. Probab. Appl.*, **14**, 153-158.

PARZEN, E. (1962). 'On estimation of a probability density function and mode', *Ann. Math. Statist.*, **33**, 1065-1067.

PRAKASA RAO, B.L.S. (1983). *Nonparametric Functional Estimation.* New York Academic Press.

ROSENBLATT, M. (1956). 'Remarks on some nonparametric estimates of a density function', *Ann. Math. Statist.*, **27**, 832-837.

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

STONE, C.J. (1984). 'An asymptotically optimal window selection rule for kernel density estimates', *Ann. Statist.*, **12**, 1285-1297.

INDIAN STATISTICAL INSTITUTE
DELHI CENTRE
7 S.J.S. SANSALWAL MARG
NEW DELHI
INDIA