Nonparametric Statistics, Vol. 1, pp. 127–141
Reprints available directly from the publisher
Photocopying permitted by license only

LENGTH BIASED DENSITY ESTIMATION OF FIBRES

G. D. RICHARDSON,[1] M. K. KAZEMPOUR,[2] and B. B. BHATTACHARYYA[3]

[1]Department of Mathematics, University of Central Florida, Orlando, FL 32816, USA and [2]Department of Statistics, University of Central Florida, Orlando, FL 32816, USA and [3]Department of Statistics, North Carolina State University, Raleigh, NC 27607, USA

Cox (1969) discussed several procedures used in sampling of textile fibres. One such procedure is called "length biased" or weighted sampling and occurs when the chance of selection is proportional to fibre length. Cox considered the problem of estimating the unweighted distribution function F at a fixed $x>0$ and compared the asymptotic variance of estimators based on length biased samples with those based on unweighted samples. Consideration here is devoted to estimating the probability density function f at a fixed $x>0$ based on length biased samples.

It is shown, under suitable regularity conditions, that the square of the bias of the weighted estimator is less (greater) than the square of the bias of the Parzen (1962)–Rosenblatt (1956) kernel estimator of $f(x)$ based on unweighted observations when $(f'(x)/x)(f^{(2)}(x)+f'(x))<0(>0)$ and n is sufficiently large. Moreover, the variance of the length biased estimator is less (greater) than that of the unweighted estimator when $x>\mu(x<\mu)$ for all n sufficiently large, where μ denotes the mean with respect to f. An optimal window width $h_n(x)$ is given which makes the asymptotic mean square error of the length biased estimator a minimum. Under regularity assumptions, it is shown that the optimal asymptotic mean square error of the length biased estimator at x is less than that for the unweighted estimator exactly when $(\mu/x)^3\,|g^{(2)}(x)/f^{(2)}(x)|<1$. Moreover, simulations are undertaken to compare the two estimators for several sample sizes.

KEYWORDS: Length biased density, kernel estimator, order of bias and variance, optimal asymptotic mean square error.

1. INTRODUCTION

Let the random variable X denote fibre length and have probability density function f relative to Lebesgue measure and assume that the mean μ with respect to f is a real number. A random sample having common probability density function f is referred to as an *unweighted* sample whereas one based on

$$g(x)=\begin{cases} xf(x)/\mu, & x\geq 0 \\ 0, & x<0 \end{cases}$$

is called *length biased* or *weighted* sample. For a fixed $x>0$, $\tilde{f}_n(x)$ and $\hat{f}_n(x)$, respectively, denote the unweighted and weighted estimators of $f(x)$. Cox (1969) discusses estimators for both μ and the distribution function $F(x)$ of $f(x)$, for x fixed and positive, based on length biased and unweighted samples. Asymptotic variances of these estimators are compared. Recently Gill, Vardi, and Wellner

(1988) investigated asymptotic properties of estimators of F based on multiple length biased samples. They consider a general weight function which includes the length biased case.

Bhattacharyya, Franklin, and Richardson (1988) discussed an estimator $\hat{f}_n(x)$ of $f(x)$ based on length biased sampling. Under suitable regularity conditions, it was shown that $\hat{f}_n(x)$ is asymptotically normal and, moreover, if X has a finite fourth moment relative to g, then the mean square error of $\hat{f}_n(x)$ converges to zero. This work represents a continuation of the above mentioned. It is shown here that, under appropriate regularity conditions, the mean square error of $\hat{f}_n(x)$ converges to zero when X has a finite second moment with respect to g. Furthermore, the order of the bias and variance of $\hat{f}_n(x)$ are determined and large sample comparisons are made with a natural unweighted estimator $\tilde{f}_n(x)$ of $f(x)$. Simulation results are obtained for selected sample sizes when f is a gamma density. The estimator $\hat{f}_n(x)$ has a smaller mean square error than that for $\tilde{f}_n(x)$ when x is sufficiently large whereas the reverse order is true for small x-values.

It should be emphasized that our estimator of $f(x)$ is for fixed positive values of x. Moreover, the condition $E_g(X_1^{-2}) < \infty$ is needed in order to verify the assumptions of section 3 used in the development of the bias and variance for $\hat{f}_n(x)$. This leads us to make the assumption that $f(0) = 0$, which makes f continuous at zero, and hence there is no need to use a modified kernel estimator studied by Schuster (1985) and others.

2 MEAN SQUARE ERROR

The notation $\{X_i\}$ will always denote a sequence of independent and identically distributed random variables each having length biased density g. Let $\phi_n(\mathbf{X}_n)$ denote any estimator of μ such that $E(\phi_n^2(\mathbf{X}_n))$ is finite for each positive integer n. Since $f(x) = \mu g(x)/x$, it seems reasonable to define our estimator of $f(x)$ by

$$\hat{f}_n(x) = \phi_n(\mathbf{X}_n)\hat{g}_n(x)/x, \tag{2.1}$$

where $\hat{g}_n(x) = (1/nh_n)\sum_1^n K((x - X_i)/h_n)$ denotes the kernel estimator due to Parzen (1962) and Rosenblatt (1956). It was shown by Bhattacharyya, Franklin, and Richardson (1988, Theorem 2.3) that, under regularity conditions, the mean square error of $\hat{f}_n(x)$ converges to zero when $\phi_n(\mathbf{X}_n)$ is the harmonic mean and $E(X_1^4)$ is finite. It is shown here that the result is still valid under reasonable regularity assumptions when $\phi_n(\mathbf{X}_n)$ is a general estimator of μ and $E(X_1^2)$ is finite. The following notation is used throughout the paper.

NOTATION

$\psi_n(s) = E[\phi_n(s, X_2, \ldots, X_n)]$, $\delta_n(s) = E[\phi_n^2(s, X_2, \ldots, X_n)]$, $\gamma_n(s, t) = E[\phi_n^2(s, t, X_3, \ldots, X_n)]$; $\Psi_n = \psi_n \cdot g$, $\nabla_n = \delta_n \cdot g$, and $\Gamma_n(s, t) = \gamma_n(s, t)g(s)g(t)$.

The assumptions listed below are used to show that the mean square error of $\hat{f}_n(x)$ converges to zero. Conditions C and D are needed in order to apply the dominated convergence theorem. Note that the requirement $\lim_{y\to\infty} |yK(y)| = 0$

is not assumed here. The notation $h(x, y) \in L(X \times Y)$ means that $\int_{X\times Y} h(x, y)\, dx\, dy$ is finite.

Assumptions

A: K is a bounded, even, density function and $\phi_n(\mathbf{x}_n)$ is exchangeable, that is, $\phi_n(x_1, \ldots, x_n) = \phi_n(x_{i_1}, \ldots, x_{i_n})$ for each permutation $\{i_1, \ldots, i_n\}$ of $\{1, \ldots, n\}$.

B: (1) $h_n \to 0$ (2) $nh_n \to \infty$.

C: For each fixed real number x,
(1) $\psi_n(s_n) \to \mu$ when $s_n \to x$
(2) $\delta_n(s_n) \to \mu^2$ when $s_n \to x$
(3) $\gamma_n(s_n, t_n) \to \mu^2$ when $s_n \to x$, $t_n \to x$.

D: (1) $|\Psi_n(x - h_n y)| \le l(y)$ where $l(y)K(y) \in L(R)$
(2) $|\nabla_n(x - h_n y)| \le m(y)$ where $m(y)K^2(y) \in L(R)$
(3) $|\Gamma_n(x - h_n y, x - h_n z| \le n(y, z)$ where $n(y, z)K(y)K(z) \in L(R \times R)$.

REMARK 2.1. It is shown in the Appendix that if g is bounded, continuous at x, and $E(X_1^2)$ is finite, then conditions C and D are satisfied when

$$\phi_n(\mathbf{x}_n) = \begin{cases} n \Big/ \sum_1^n x_i^{-1}, & \text{all } x_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

is the harmonic mean. Alternately, if K has compact support, f is continuous at x, ϕ_n is the harmonic mean, and $E(X_1^2)$ is finite, then these conditions also hold. The proof is supplied for conditions C(3) and D(3) while the remaining parts follow in a similar manner.

THEOREM 2.1. Assume that conditions A–D are satisfied and that f is continuous at $x > 0$. Then MSE $\hat{f}_n(x)$ converges to zero.

Proof. Let R_n denote n-dimensional Euclidean space. Since $\phi_n(\mathbf{x}_n)$ is exchangeable, $E(\hat{f}_n(x)) = (1/h_n x)E[K((x - X_1)/h_n)\phi_n(\mathbf{X}_n)] = (1/h_n x) \times \int_{R_n} K((x - x_1)/h_n)\phi_n(\mathbf{x}_n)g(\mathbf{x}_n)\, d\mathbf{x}_n$. Denote $y = (x - x_1)/h_n$; then $E(\hat{f}_n(x)) = (1/x)\int_{R_n} K(y)\phi_n(x - h_n y, x_2, \ldots, x_n)g(x - h_n y) \prod_2^n g(x_i)\, dy \prod_2^n dx_i =$ (by Fubini's Theorem) $(1/x)\int_R K(y)E[\phi_n(x - h_n y, X_2, \ldots, X_n)]g(x - h_n y)\, dy = (1/x) \times \int_R K(y)\Psi_n(x - h_n y)\, dy \to (1/x)\mu g(x) = f(x)$ by assumptions A, B(1), C(1), D(1) and the fact that g is continuous at x. Hence the bias of $\hat{f}_n(x)$ converges to zero.

Moreover, $\operatorname{var} \hat{f}_n(x) = (1/(nh_n x)^2)n \operatorname{var}[K((x - X_1)/h_n)\phi_n(\mathbf{X}_n)] + (n(n-1)/(nh_n x)^2) \operatorname{cov}[K((x - X_1)/h_n)\phi_n(\mathbf{X}_n), K((x - X_2)/h_n)\phi_n(\mathbf{X}_n)]$. Note that

$$(1/h_n) \operatorname{var}[K((x - x_1/h_n)\phi_n(\mathbf{X}_n)] = (1/h_n)\int_{R_n} K^2((x - x_1)/h_n) \times \phi_n^2(\mathbf{x}_n)g(\mathbf{x}_n)\, d\mathbf{x}_n - (1/h_n)(E[K((x - X_1)/h_n)\phi_n(\mathbf{X}_n)])^2$$

$$= \text{(by Fubini's Theorem)}$$

$$\int_R K^2(y)E[\phi_n^2(x - h_n y, X_2, \ldots, X_n)]g(x - h_n y)\, dy - h_n x^2(E(\hat{f}_n(x)))^2$$

$$= \int_R K^2(y)\nabla_n(x - h_n y)\, dy - h_n x^2(E(\hat{f}_n(x)))^2 \to \mu^2 g(x)\int_R K^2(y)\, dy$$

by B(1), C(2), D(2), continuity of g at x and the fact that $E(\hat{f}_n(x)) \to f(x)$. Hence it follows from B(2) that $(1/(nh_n x)^2)n \operatorname{var}[K((x - X_1)/h_n)\phi_n(\mathbf{X}_n)]$ converges to zero.

Finally,

$$
\begin{aligned}
&(1/h_n^2)\operatorname{cov}[K((x - X_1)/h_n)\phi_n(\mathbf{X}_n),\ K((x - X_2)/h_n)\phi_n(\mathbf{X}_n)] \\
&= \int_{R_2} K(y)K(z)E[\phi_n^2(x - h_n y, x - h_n z, X_3, \ldots, X_n)]g(x - h_n y)g(x - h_n z)\,dy\,dz \\
&\quad - (1/h_n^2)(E[K((x - X_1)/h_n)\phi_n(\mathbf{X}_n)])^2 \\
&= \int_{R_2} K(y)K(z)\Gamma_n(x - h_n y, x - h_n z)\,dy\,dz \\
&\quad - (E[(1/h_n)K((x - X_1)/h_n)\phi_n(\mathbf{X}_n)])^2 \to \mu^2(g(x))^2 - (\mu g(x))^2 = 0
\end{aligned}
$$

by our assumptions. It follows that the mean square error of $\hat{f}_n(x)$ converges to zero. □

3. ORDER OF CONVERGENCE

Let us first find the order of the bias of $\hat{f}_n(x)$. The technique is to use Taylor's formula to expand $E(\hat{f}_n(x))$ in terms of h_n and then apply the dominated convergence theorem to the remainder term. The notation used here is consistent with that defined in the preceding section. It is assumed that $\psi_n(s)$ possesses a continuous second derivative for all $s \in R$. Additional assumptions needed are listed below. Let $I = [0, 1]$.

Assumptions

E: (1) $(\psi_n(x) - \mu)/h_n^2 \to 0$
(2) $\Psi_n^{(2)}(s_n) \to \mu g^{(2)}(x)$ when $s_n \to x$
(3) $|\Psi_n^{(2)}(x - h_n y s)| \le l(s, y)$ where $y^2 K(y) l(s, y) \in L(I \times R)$.

REMARK 3.1. Suppose that $E(X_1^{-2})$ and $E(X_1^3)$ are each finite, g has a continuous second derivative in a neighborhood of x, ϕ_n is the harmonic mean, K satisfies assumption A and has compact support, $h_n \to 0$ and $nh_n^2 \to \infty$. Then the conclusions of Theorem 3.1 and Corollary 3.1 below are valid. In addition, if $E(X_1^4)$ is finite and $f(x) > 0$, then the results of Theorem 3.2, Corollary 3.2 and Theorem 3.3 are valid. An outline of a proof of these facts is given in the Appendix.

THEOREM 3.1. Suppose that conditions A, B(1), E are satisfied and that $g(s)$ has a continuous second derivative for all $s \in R$. Then for $x > 0$, $\operatorname{bias} \hat{f}_n(x)/h_n^2 \to (\mu g^{(2)}(x)/2x) \int_R y^2 K(y)\,dy$.

Proof. It was shown in the proof of Theorem 2.1 that $E(\hat{f}_n(x)) = (1/x) \int_R K(y)\Psi_n(x - h_n y)\,dy$. An application of Taylor's formula gives that $\Psi_n(x - h_n y) = \Psi_n(x) - h_n y \Psi_n'(x) + h_n^2 y^2 \int_I (1 - s)\Psi_n^{(2)}(x - h_n y s)\,ds$. Since K is an even density function, $E(\hat{f}_n(x)) = \Psi_n(x)/x + (h_n^2/x) \int_{I \times R} y^2 K(y)(1 - s)\Psi_n^{(2)}(x - h_n y s)\,ds\,dy$ and thus $\operatorname{bias} \hat{f}_n(x)/h_n^2 = (g(x)/x)[(\psi_n(x) - \mu)/h_n^2] + (1/x) \int_{I \times R} y^2 K(y)(1 - s)\Psi_n^{(2)}(x - h_n y s)\,ds\,dy$. The second term converges to $(\mu g^{(2)}(x)/x) \int_{I \times R} y^2 K(y)(1 - s)\,ds\,dy = (\mu g^{(2)}(x)/2x) \int_R y^2 K(y)\,dy$ by the domin-

ated convergence theorem and the desired conclusion follows from assumption $E(1)$. □

Let $\{X_i'\}$ denote a sequence of independent and identically distributed random variables each having density f. Define

$$\tilde{f}_n(x) = (1/nh_n) \sum_1^n K((x - X_i')/h_n) \tag{3.1}$$

to be the Parzen (1962)-Rosenblatt (1956) kernel estimator of $f(x)$. It is known that if in addition to our assumptions on K and h_n, $y^2K(y) \in L(R)$, f is bounded and $f^{(2)}(x)$ exists, then bias $\tilde{f}_n(x)/h_n^2 \to (f^{(2)}(x)/2) \int_R y^2K(y)\, dy$; for example, see Shapiro (1969, Theorem 6). However, this proof is also valid when the assumptions that $y^2K(y) \in L(R)$ and f is bounded are replaced by $|f(x - h_n(y)| \leq r(y)$ where $y^2K(y)r(y) \in L(R)$. Note that the latter implies that $|f(x)| \leq r(y)$ for each $y \in R$ since f is continuous at x and thus $y^2K(y)r(y) \in L(R)$ implies that $y^2K(y) \in L(R)$.

The next result follows from a simple calculation comparing the two squared biases.

COROLLARY 3.1. Suppose that the assumptions made in the preceding theorem are satisfied and, moreover, $|f(x - h_n y)| \leq r(y)$ where $y^2K(y)r(y) \in L(R)$. Then the square of the bias for $\hat{f}_n(x)$ is less (greater) than that for $\tilde{f}_n(x)$ when $(f'(x)/x)[f^{(2)}(x) + f'(x)/x] < 0$ (>0) and n is sufficiently large.

Next, let us consider the order of the variance of $\hat{f}_n(x)$. Additional assumptions needed in order to apply the dominated convergence theorem are listed below.

Assumptions

F: (1) $L_n(y, z) = nh_n(\Gamma_n(x - h_n y, x - h_n z) - \Psi_n(x - h_n y)\Psi_n(x - h_n z)) \to 0$
(2) $|L_n(y, z)| \leq \nu(y, z)$ where $K(y)K(z)\nu(y, z) \in L(R \times R)$.

THEOREM 3.2. Suppose that assumptions A, B(1), C(1), C(2), D(1), D(2), F are satisfied and f is continuous at $x > 0$. Then $nh_n \operatorname{var} \hat{f}_n(x) \to (\mu f(x)/x) \int_R K^2(y)\, dy$.

Proof. Since ϕ_n is exchangeable, $nh_n \operatorname{var} \hat{f}_n(x) = (1/h_n x^2) \operatorname{var} \phi_n(\mathbf{X}_n)K((x - X_1)/h_n) + ((n-1)/h_n x^2) \operatorname{cov}[\phi_n(\mathbf{X}_n)K((x - X_1)/h_n), \phi_n(\mathbf{X}_n)K((x - X_2)/h_n)]$. It was shown in Theorem 2.1 that $(1/h_n x^2) \operatorname{var} \phi_n(\mathbf{X}_n)K((x - X_1)/h_n) \to (\mu f(x)/x) \int_R K^2(y)\, dy$. Moreover, it is straightforward to show that $J_n = (n/h_n) \operatorname{cov}[\phi_n(\mathbf{X}_n) K((x - X_1)/h_n), \phi_n(\mathbf{X}_n)K((x - X_2)/h)] = nh_n \int_{R_2} K(y)K(z) \times [\Gamma_n(x - h_n y, x - h_n z) - \Psi_n(x - h_n y)\Psi_n(x - h_n z)]\, dy\, dz$ and thus assumption F and the dominated convergence theorem give that $J_n \to 0$. □

COROLLARY 3.2. Assume that the conditions of the preceding theorem are satisfied and that $|f(x - h_n y)| \leq r(y)$ and $K(y)r(y) \in L(R)$. If $f(x) > 0$, then the variance of $\hat{f}_n(x)$ is less (greater) than that for $\tilde{f}_n(x)$ when $x > \mu (x < \mu)$ and n is sufficiently large.

Proof. A simplified version of the proof of Theorem 3.2 gives the well-known result that $nh_n \operatorname{var} \tilde{f}_n(x) \to f(x) \int_R K^2(y)\, dy$ and thus the desired conclusion follows. □

Let us use Theorems 3.1 and 3.2 to find the optimal asymptotic mean square error for $\hat{f}_n(x)$ and compare it with that of $\tilde{f}_n(x)$.

THEOREM 3.3. Suppose that assumptions A, B(1), C(1), C(2), D(1), D(2), E, F are satisfied and that $g(s)$ has a continuous second derivative for each s. Moreover, assume that $|f(x-h_n y)| \le r(y)$, $y^2K(y)r(y) \in L(R)$, and $f(x) > 0$. Then

(i) the optimal $h_n(x)$ making $\mathrm{MSE}\,\hat{f}_n(x)$ a minimum is $[xf(x)\int_R K^2(y)\,dy/n\mu(g^{(2)}(x)\int_R y^2K(y)\,dy)^2]^{1/5}$

(ii) the optimal asymptotic $\mathrm{MSE}\,\hat{f}_n(x)$ is $(5/4)[\mu^6(f(x))^4(g^{(2)}(x))^2(\int_R K^2(y)\,dy)^4 \times (\int_R y^2K(y)\,dy)^2/n^4x^6]^{1/5}$.

(iii) the optimal asymptotic $\mathrm{MSE}\,\hat{f}_n(x)$ is less than the optimal asymptotic $\mathrm{MSE}\,\tilde{f}_n(x)$ *iff* $(\mu/x)^3\,|g^{(2)}(x)/f^{(2)}(x)| < 1$.

Proof. Recall that

$$\mathrm{MSE}\,\hat{f}_n(x) \sim \mu f(x)\int_R K^2(y)\,dy/nh_nx + h_n^4(\mu g^{(2)}(x)\int_R y^2K(y)\,dy)^2/4x^2$$

and thus the optimal $h_n(x)$ is obtained by differentiation and, moreover, part (ii) is obtained by substitution of part (i). The optimal asymptotic $\mathrm{MSE}\,\tilde{f}_n(x)$ is

$$(5/4)[(f(x))^4(f^{(2)}(x))^2(\int_R K^2(y)\,dy)^4(\int_R y^2K(y)\,dy)^2/n^4]^{1/5}.$$

A comparison of this result with part (ii) gives the desired conclusion in part (iii). □

4 SIMULATION

Asymptotic properties of the two estimators $\hat{f}_n(x)$ and $\tilde{f}_n(x)$ defined in (2.1) and (3.1), respectively, have been compared in the preceding section. The purpose of this section is to compare the two estimators through simulations for various sample sizes. The family of gamma densities

$$f(x) = x^{\alpha-1}e^{-x}/\Gamma(\alpha), \qquad x \ge 0 \tag{4.1}$$

is selected and simulation results are obtained for the cases of $\alpha = 2, 4$. The values of $f(x)$ are quite small for large values of α. Recall that

$$g(x) = x^{\alpha}e^{-x}/\Gamma(\alpha+1), \qquad x \ge 0. \tag{4.2}$$

The Epanechnikov kernel

$$K(y) = \begin{cases} (3/4\sqrt{5})(1-y^2/5), & |y| < \sqrt{5} \\ 0, & \text{otherwise} \end{cases}$$

is chosen and the estimator of μ selected is the harmonic mean.

The optimal values of $h_n(x)(h_n'(x))$ making $\mathrm{MSE}\,\hat{f}_n(x)$ ($\mathrm{MSE}\,\tilde{f}_n(x)$) a minimum are used. This extends the simulation study of Bhattacharyya, Franklin, and Richardson (1988) where the optimal $h_n(x)$ was not used in $\hat{f}_n(x)$. For the case mentioned above, it follows from Theorem 3.3 that

$$h_n(x) = [3\sqrt{5}xf(x)/25n\alpha(g^{(2)}(x))^2]^{1/5}$$

and

$$h_n'(x) = [3\sqrt{5}f(x)/25n(f^{(2)}(x))^2]^{1/5}.$$

These formulas are valid except when $g^{(2)}(x)$ or $f^{(2)}(x)$ are sufficiently close to zero. In the latter case, modified values of $h_n(x)(h_n'(x))$ are obtained in a neighborhood of the singularity by expanding $g(f)$ in a Taylor series about x through the fourth derivative and equating the second derivative to zero throughout the neighborhood. The modified optimal formulas for $h_n(x)$ and $h_n'(x)$ are

$$h_n^*(x) = [1176\sqrt{5}xf(x)/625n\alpha(g^{(4)}(x))^2]^{1/9}$$

and

$$h_n'^*(x) = [1176\sqrt{5}f(x)/625n(f^{(4)}(x))^2]^{1/9}.$$

The new value of $h_n(x)(h_n'(x))$ is used in formula 2.1 (3.1) throughout a neighborhood of the singularity.

The simulation began by first generating n random numbers from the density $f(x)$ in (4.1) by using the IMSL subroutine RANGAM. The values generated were used to calculate $\tilde{f}_n(x)$ for various values of x. A simulation size of 600 was used and the average of $\tilde{f}_n(x)$ was recorded for various values of x. Similarly, n random numbers were generated from $g(x)$ in (4.2) and the average of 600 values of $\hat{f}_n(x)$ was recorded for corresponding x-values. The simulation size of 600 was chosen because changes in the means and variances of the estimators were less than 5% when the simulation size increased from 400 to 600. The effect of the sample size was illustrated for values of $n = 50$ and 200 when $\alpha = 2$ and $\alpha = 4$. Table 1 list values of $f(x)$, $\text{AVE}\,\hat{f}_n(x)$, $\text{AVE}\,\tilde{f}_n(x)$ along with the square of the bias and mean square error for each estimator when $\alpha = 2$ and $n = 50$. Other combinations of α and n are listed in the remaining tables and graphical presentations are given in Figures 1–8.

The following observations are made. Note that the mean square error for $\tilde{f}_n(x)$ is smaller than that for $\hat{f}_n(x)$ in the initial portion of the domain whereas the reverse ordering is true in the upper tail. Moreover, the square of the bias for $\hat{f}_n(x)$ is smaller than that for $\tilde{f}_n(x)$ in a neighborhood of the mean of f whereas the variance has the reverse effect and, consequently, the difference in the mean square errors of the two estimators is small in this region. However, the values of $\hat{f}_n(x)$ are closer to $f(x)$ in this region. It should also be observed that the mean square error of each estimator decreased as n increased. In conclusion, at least

Table 1. $\phi(\mathbf{X}_n) = n/\sum_{i=1}^n (1/X_i)$

x	$f(x)$	$\text{AVE}\,\hat{f}_n(x)$	$\text{AVE}\,\tilde{f}_n(x)$	$\text{bias}^2\,\hat{f}_n(x)$	$\text{bias}^2\,\tilde{f}_n(x)$	$\text{MSE}\,\hat{f}_n(x)$	$\text{MSE}\,\hat{f}_n(x)$
0.5	.3033	.2228	.2665	.006478	.001354	.010822	.005261
1.0	.3679	.3190	.3309	.002392	.001364	.005462	.003663
1.5	.3347	.3220	.2980	.000162	.001346	.002575	.002091
2.0	.2707	.2576	.2546	.000171	.000258	.001903	.000569
2.5	.2052	.1932	.2137	.000144	.000072	.001199	.000398
3.0	.1494	.1215	.1636	.000774	.000203	.001199	.000775
3.5	.1057	.0520	.1187	.002878	.000168	.002959	.000728
4.0	.0733	.0592	.0840	.000197	.000114	.000382	.000554
4.5	.0500	.0462	.0583	.000014	.000069	.000166	.000387
5.0	.0337	.0345	.0403	.000001	.000043	.000106	.000260

$f(x)$ is a gamma density function with mean = 2. Estimators from sample of $n = 50$ observations.

Table 2. $\phi(\mathbf{X}_n) = n/\sum_{i=1}^n (1/X_i)$

x	$f(x)$	AVE $\hat{f}_n(x)$	AVE $\tilde{f}_n(x)$	bias^2 $\hat{f}_n(x)$	bias^2 $\tilde{f}_n(x)$	MSE $\hat{f}_n(x)$	MSE $\tilde{f}_n(x)$
0.5	.3033	.2030	.2808	.010061	.000503	.011563	.001813
1.0	.3679	.3253	.3463	.001814	.000464	.003220	.001328
1.5	.3347	.3441	.3158	.000088	.000356	.001299	.000703
2.0	.2707	.2775	.2619	.000046	.000077	.000778	.000205
2.5	.2052	.2055	.2124	.000000	.000051	.000387	.000223
3.0	.1494	.1186	.1580	.000948	.000074	.001094	.000295
3.5	.1057	.0512	.1136	.002973	.000062	.003001	.000266
4.0	.0733	.0549	.0798	.000337	.000043	.000391	.000196
4.5	.0500	.0431	.0554	.000047	.000029	.000089	.000135
5.0	.0337	.0322	.0380	.000002	.000019	.000030	.000092

$f(x)$ is a gamma density function with mean = 2. Estimators from sample of $n = 200$ observations.

for the class of gamma densities, the estimator $\hat{f}_n(x)$ performed better than $\tilde{f}_n(x)$ when x is in a neighborhood of the mean of f and when x is in the upper tail. The superiority of $\hat{f}_n(x)$ for sufficiently large values of x is intuitively clear since due to the sampling bias one obtains more large observations than would be the case if there were no bias. Moreover, it should be emphasized that the estimator $\tilde{f}_n(x)$ cannot actually be found since observations from f are unavailable.

APPENDIX: PROOF OF REMARKS 2.1, 3.1

Proof of Remark 2.1. Assume that g is bounded, continuous at x, $E(X_1^2)$ is finite, and let $\phi_n(\mathbf{x}_n)$ denote the harmonic mean. Recall that

$$\phi_n^2(s, t, X_3, \ldots, X_n) = n^2/(s^{-1} + t^{-1} + \textstyle\sum_3^n X_i^{-1})^2 \le (n/\sum_3^n X_i^{-1})^2$$
$$= (n/(n-2))^2((n-2)/\textstyle\sum_3^n X_i^{-1})^2 \le (n/(n-2))^2(\sum_3^n X_i^2/(n-2))$$

Table 3. $\phi(\mathbf{X}_n) = n/\sum_{i=1}^n (1/X_i)$

x	$f(x)$	AVE $\hat{f}_n(x)$	AVE $\tilde{f}_n(x)$	bias^2 $\hat{f}_n(x)$	bias^2 $\tilde{f}_n(x)$	MSE $\hat{f}_n(x)$	MSE $\tilde{f}_n(x)$
0.5	.0126	.0329	.0192	.000126	.000043	.002775	.000352
1.0	.0613	.1382	.0718	.005913	.000110	.010788	.000623
1.5	.1255	.1767	.1189	.002620	.000044	.004679	.000561
2.0	.1804	.1395	.1654	.001675	.000227	.002191	.001123
2.5	.2138	.1612	.1970	.002764	.000280	.003253	.001218
3.0	.2240	.2010	.2065	.000530	.000307	.001403	.000996
3.5	.2158	.2022	.1978	.000185	.000322	.000988	.000738
4.0	.1954	.1874	.1779	.000063	.000306	.000704	.000524
4.5	.1687	.1612	.1555	.000057	.000175	.000519	.000300
5.0	.1404	.1334	.1387	.000049	.000003	.000352	.000149
5.5	.1133	.0977	.1193	.000244	.000035	.000394	.000250
6.0	.0892	.0763	.0968	.000167	.000058	.000301	.000339
6.5	.0688	.0574	.0766	.000131	.000060	.000237	.000318
7.0	.0521	.0423	.0593	.000096	.000052	.000169	.000268
7.5	.0389	.0381	.0451	.000001	.000039	.000086	.000215
8.0	.0286	.0298	.0339	.000001	.000028	.000067	.000163
8.5	.0208	.0230	.0254	.000005	.000021	.000051	.000124
9.0	.0150	.0175	.0190	.000005	.000016	.000039	.000095

$f(x)$ is a gamma density function with mean = 4. Estimators from sample of $n = 50$ observations.

Table 4. $\phi(\mathbf{X}_n) = n/\sum_{i=1}^{n}(1/X_i)$

x	$f(x)$	AVE $\hat{f}_n(x)$	AVE $\tilde{f}_n(x)$	bias^2 $\hat{f}_n(x)$	bias^2 $\tilde{f}_n(x)$	MSE $\hat{f}_n(x)$	MSe $\tilde{f}_n(x)$
0.5	.0126	.0213	.0165	.000076	.000015	.000818	.000106
1.0	.0613	.1219	.0677	.003674	.000041	.005077	.000209
1.5	.1255	.1719	.1203	.002156	.000027	.002975	.000218
2.0	.1804	.1247	.1711	.003111	.000088	.003284	.000396
2.5	.2138	.1500	.2033	.004069	.000110	.004238	.000418
3.0	.2240	.2074	.2132	.000277	.000118	.000604	.000389
3.5	.2158	.2127	.2048	.000004	.000121	.000320	.000320
4.0	.1954	.2020	.1847	.000044	.000114	.000330	.000239
4.5	.1687	.1754	.1595	.000044	.000084	.000269	.000142
5.0	.1404	.1442	.1401	.000014	.000000	.000169	.000061
5.5	.1133	.0924	.1177	.000438	.000019	.000494	.000110
6.0	.0892	.0685	.0945	.000432	.000028	.000476	.000121
6.5	.0688	.0499	.0742	.000357	.000029	.000386	.000110
7.0	.0521	.0358	.0570	.000266	.000024	.000284	.000093
7.5	.0389	.0355	.0430	.000011	.000017	.000037	.000076
8.0	.0286	.0278	.0321	.000001	.000012	.000021	.000057
8.5	.0208	.0214	.0239	.000000	.000009	.000015	.000043
9.0	.0150	.0162	.0177	.000001	.000007	.000012	.000032

$f(x)$ is a gamma density function with mean = 4. Estimators from sample of $n = 200$ observations.

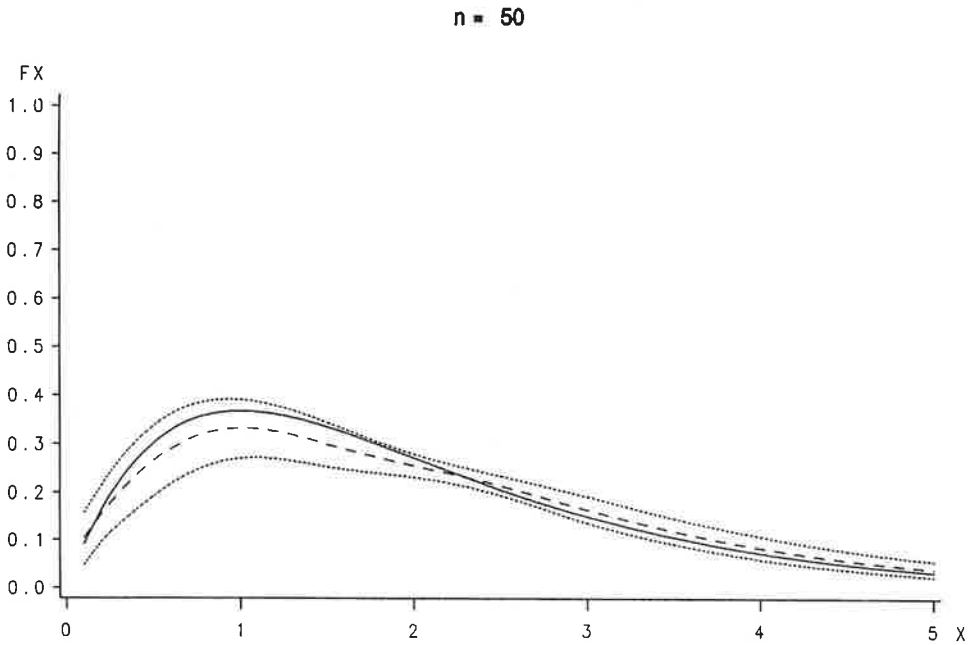

Figure 1. $\tilde{f}_n(x)$. (1) $f(x) = xe^{-x}$ is plotted as the solid line. (2) Averages of $\tilde{f}_n(x)$ are plotted as the dashed line. (3) Averages of $\tilde{f}_n(x) \pm \sqrt{\text{MSE}\,\tilde{f}_n(x)}$ are shown as dotted lines.

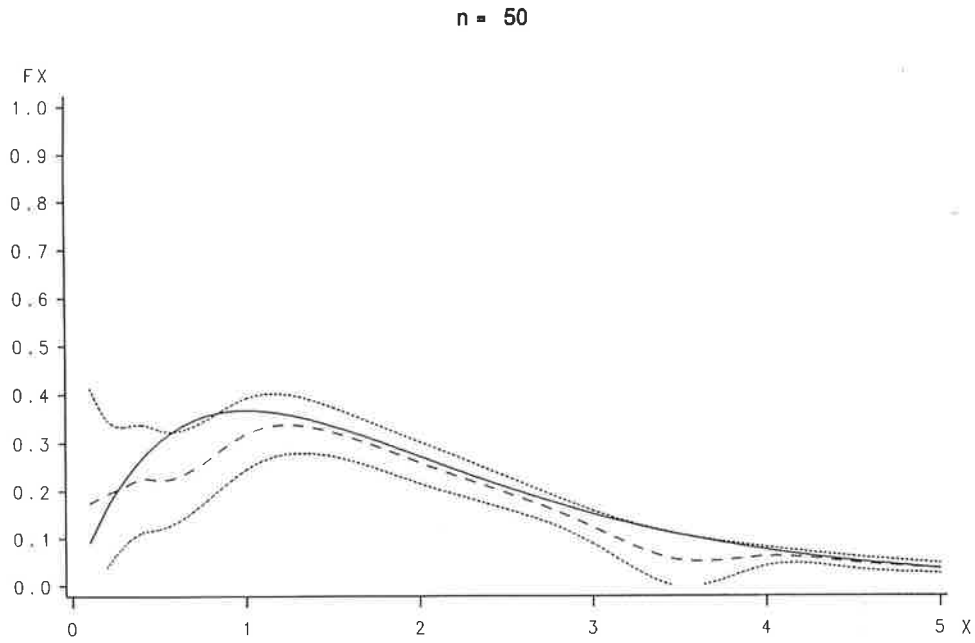

Figure 2. $\hat{f}_n(x)$. (1) $f(x) = xe^{-x}$ is plotted as the solid line. (2) Averages of $\hat{f}_n(x)$ are plotted as the dashed line. (3) Averages of $\hat{f}_n(x) \pm \sqrt{\text{MSE}\,\hat{f}_n(x)}$ are shown as dotted lines.

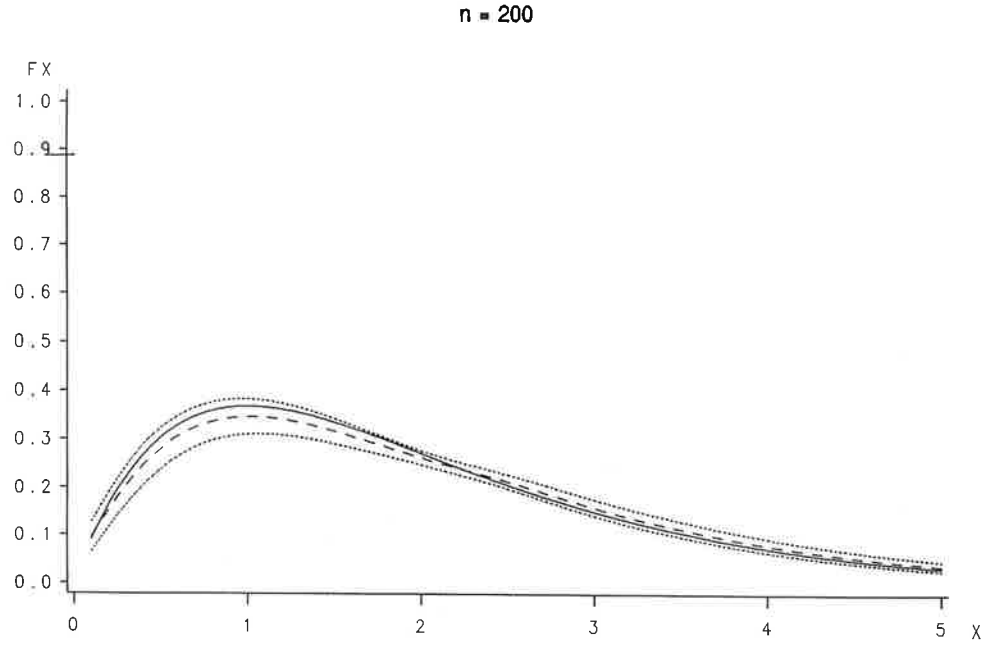

Figure 3. $\tilde{f}_n(x)$. (1) $f(x) = xe^{-x}$ is plotted as the solid line. (2) Averages of $\tilde{f}_n(x)$ are plotted as the dashed line. (3) Averages of $\tilde{f}_n(x) \pm \sqrt{\text{MSE}\,\tilde{f}_n(x)}$ are shown as dotted lines.

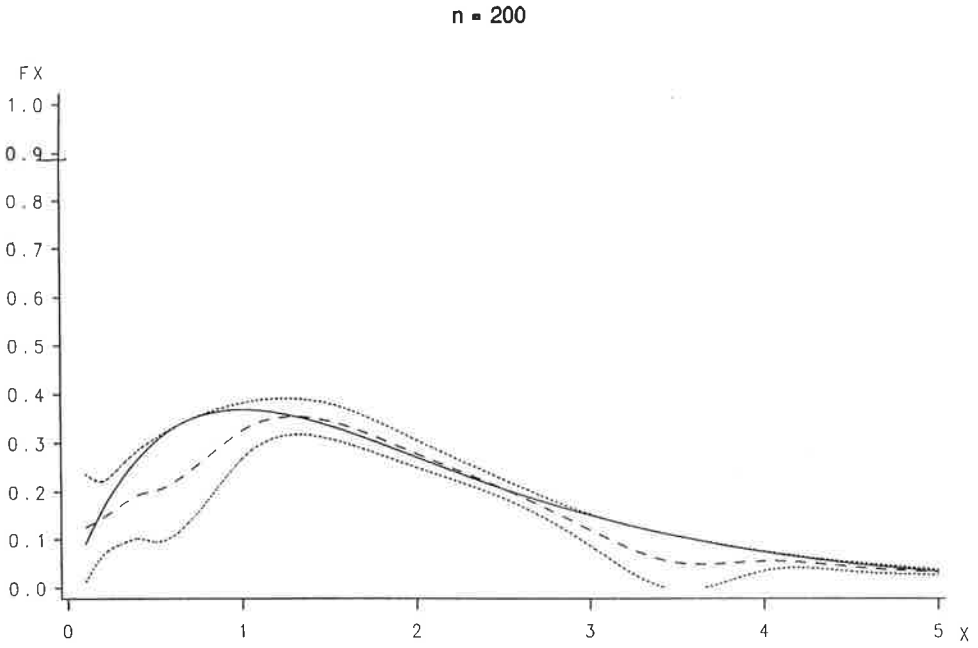

Figure 4. $\hat{f}_n(x)$. (1) $f(x) = xe^{-x}$ is plotted as the solid line. (2) Averages of $\hat{f}_n(x)$ are plotted as the dashed line. (3) Averages of $\hat{f}_n(x) \pm \sqrt{\text{MSE}\,\hat{f}_n(x)}$ are shown as dotted lines.

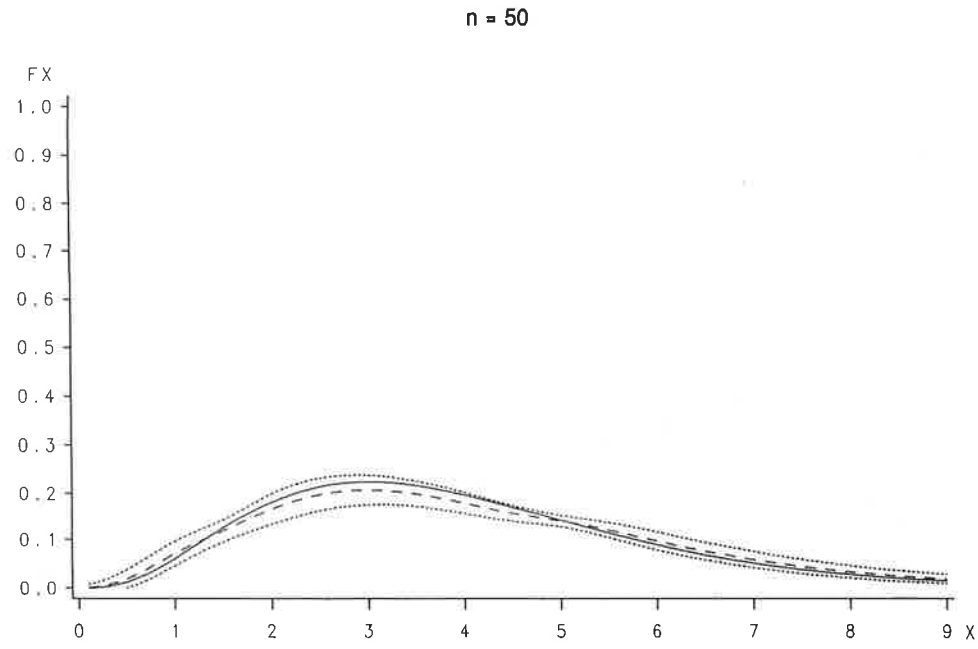

Figure 5. $\tilde{f}_n(x)$. (1) $f(x) = \frac{1}{6}x^3e^{-x}$ is plotted as the solid line. (2) Averages of $\tilde{f}_n(x)$ are plotted as the dashed line. (3) Averages of $\tilde{f}_n(x) \pm \sqrt{\text{MSE}\,\tilde{f}_n(x)}$ are shown as dotted lines.

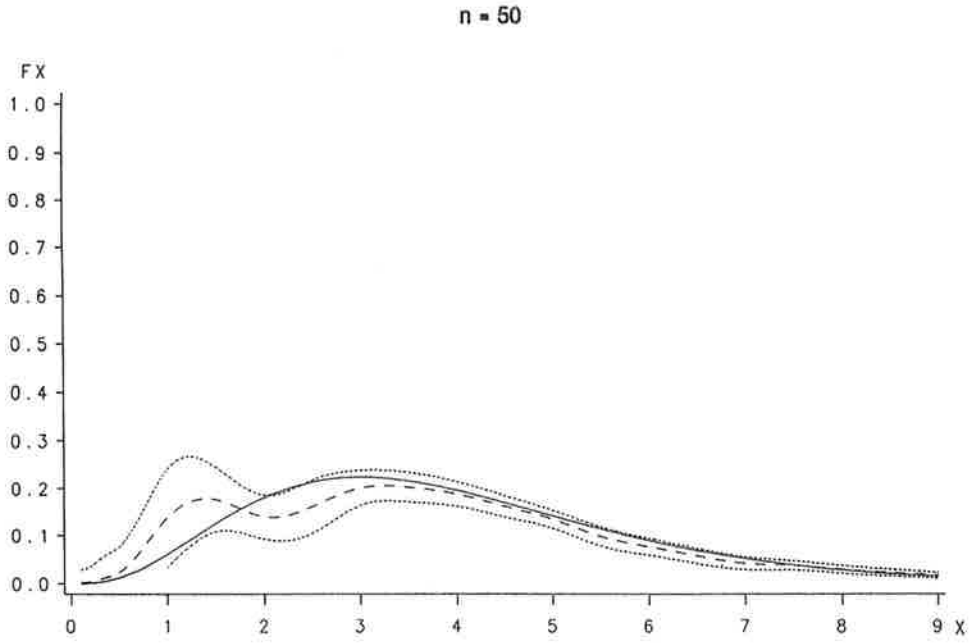

Figure 6. $\hat{f}_n(x)$. (1) $f(x) = \frac{1}{6}x^3e^{-x}$ is plotted as the solid line. (2) Averages of $\hat{f}_n(x)$ are plotted as the dashed line. (3) Averages of $\hat{f}_n(x) \pm \sqrt{\text{MSE}}\hat{f}_n(x)$ are shown as dotted lines.

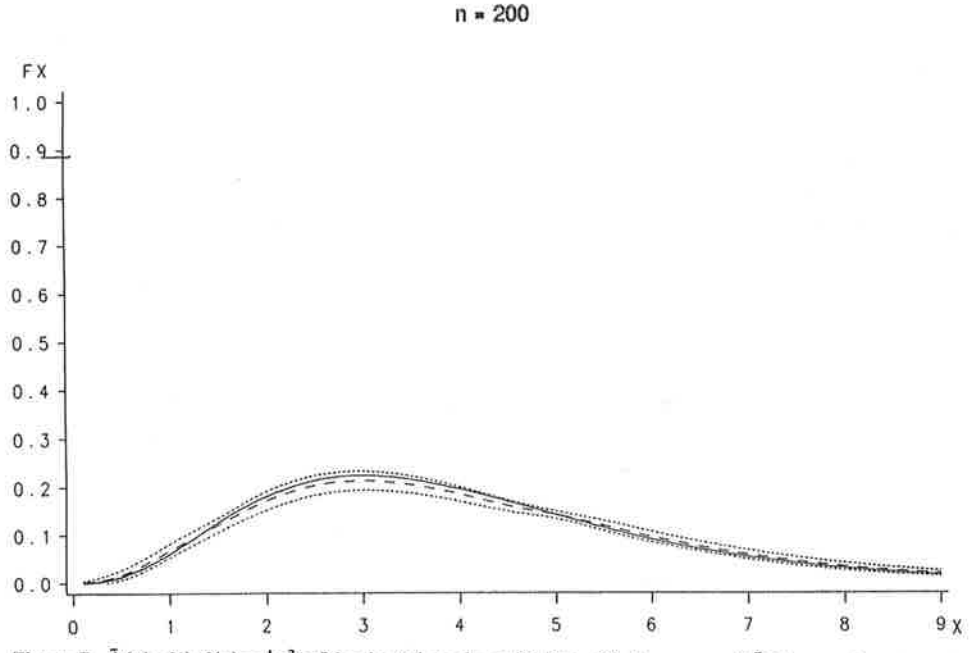

Figure 7. $\bar{f}_n(x)$. (1) $f(x) = \frac{1}{6}x^3e^{-x}$ is plotted as the solid line. (2) Averages of $\bar{f}_n(x)$ are plotted as the dashed line. (3) Averages of $\bar{f}_n(x) \pm \sqrt{\text{MSE}}\bar{f}_n(x)$ are shown as dotted lines.

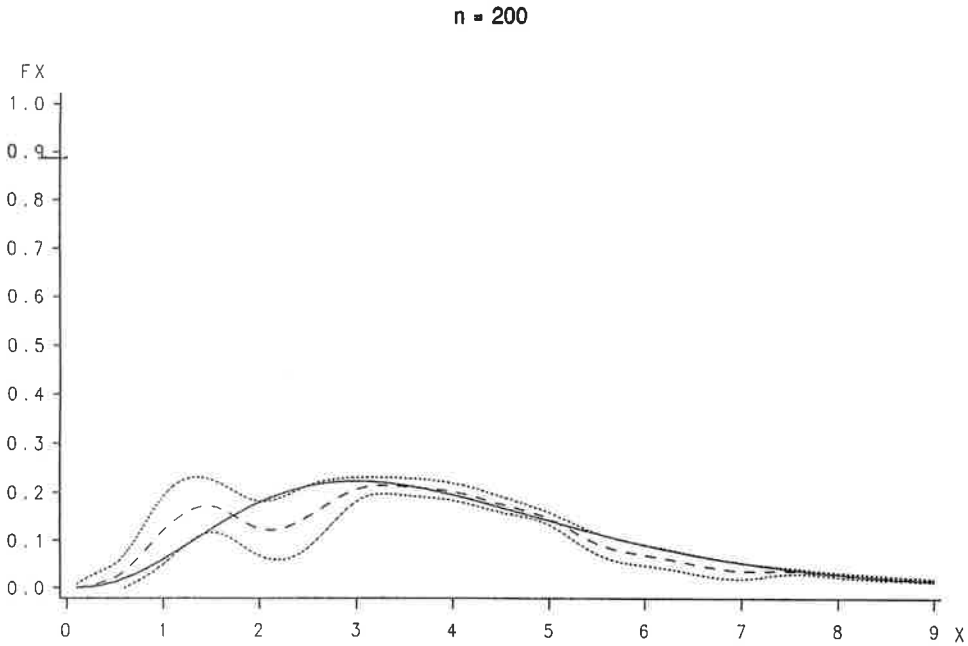

Figure 8. $\hat{f}_n(x)$. (1) $f(x) = \frac{1}{6}x^3e^{-x}$ is plotted as the solid line. (2) Averages of $\hat{f}_n(x)$ are plotted as the dashed line. (3) Averages of $\hat{f}_n(x) \pm \sqrt{\text{MSE}\,\hat{f}_n(x)}$ are shown as dotted lines.

and thus $\{\phi_n^2(s, t, X_3, \ldots, X_n) \mid s, t, \varepsilon R, n \geq 1\}$ is uniformly integrable. It follows that if $s_n \to x$ and $t_n \to x$, then $\gamma_n(s_n, t_n) = E[\phi_n^2(s_n, t_n, X_3, \ldots, X_n)] \to \mu^2$ and thus condition C(3) is satisfied. Furthermore, $\gamma_n(s, t) = E[\phi_n^2(s, t, X_3, \ldots, X_n)] \leq ME(X_1^2)$ and since g is bounded, condition D(3) holds. A similar proof can be given to show that the remaining conditions of C and D are satisfied.

Next, suppose that boundedness of g is replaced by the assumption that K has support contained in $[-k, k]$, $k > 0$. The proof given above shows that C(3) is still valid. In this case, y and z may be restricted to the interval $[-k, k]$ and thus $g(x - h_n y)$ is a bounded function of $y \in [-k, k]$ uniformly in n, for all n sufficiently large, and thus D(3) is satisfied for n sufficiently large. The remaining parts in C and D are shown in a like manner. □

LEMMA A. Assume that $\{X_i\}$ is a sequence of independent and idencically distributed random variables each having length biased density g such that $E(X_1^{-2})$ and $E(X_1^3)(E(X_1^4))$ are each finite. If $\phi_n(\mathbf{x}_n)$ denotes the harmonic mean, then $E(\phi_n(\mathbf{X}_n)) - \mu = 0(n^{-1})(E(\phi_n^2(\mathbf{X}_n)) - \mu^2 = 0(n^{-1})$.

Proof. Define $g(y) = y^{-1}$ for each $y > 0$ and let $\theta = \mu^{-1}$ and thus $g(y) = g(\theta) + g'(\theta)(y - \theta) + (g^{(2)}(\xi)/2)(y - \theta)^2$, where ξ lies between y and θ, $y > 0$. Denote by $Y_n = \sum_1^n X_i^{-1}/n$ and thus $g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + (g^{(2)}(\xi)/2)(Y_n - \theta)^2$, or $Y_n^{-1} = \theta^{-1} - \theta^{-2}(Y_n - \theta) + \xi_n^{-3}(Y_n + \theta q^2$, where ξ_n is a random variable assuming values between Y_n and θ. Recall that $E(Y_1) = \mu^{-1} = \theta$ and thus $E(Y_n) = \theta$. Hence $E[n(\phi_n(\mathbf{X}_n) - \mu)] = E[n(Y_n^{-1} - \mu)] = nE[\xi_n^{-3}(Y_n - \theta)^2] = nE[\xi_n^{-3}(Y_n -$

$\theta)^2 \cdot 1_{S_n}(Y_n)] + nE[\xi_n^{-3}(Y_n - \theta)^2 \cdot 1_{T_n}(Y_n)]$, where $S_n = \{Y_n \geq \theta\}$ with complement T_n. Note that the first term is bounded above by $n\theta^{-3}$ var $Y_n = \theta^{-3}$ var X_1^{-1} which is finite for each $n \geq 1$.

The second term is bounded above by $nE[Y_n^{-3}(Y_n - \theta)^2] \leq nE[(\bar{X}_n)^3(Y_n - \theta)^2] \leq E[\sum_1^n X_i^3(Y_n - \theta)^2]$. Note that $X_i^3(Y_n - \theta)^2 = \sum_{j=1}^n X_i^3(X_j^{-1} - \theta)^2/n^2 + \sum_{j \neq k} X_i^3(X_j^{-1} - \theta)\ (X_k^{-1} - \theta)/n^2$ and thus $E[X_i^3(Y_n - \theta)^2] \leq cn^{-1}$ for all $n \geq 1$ since the last term has zero expectation. Hence $nE[\xi_n^{-3}(Y_n - \theta)^2 \cdot 1_{T_n}(Y_n)] \leq c$ for all $n \geq 1$ and consequently $E(\phi_n(\mathbf{X}_n)) - \mu = 0(n^{-1})$. The other part of the lemma is proved in a similar manner by letting $g(y) = y^{-2}$ for $y > 0$. □

Proof of Remark 3.1. (Outline) The verification of conditions E(2) and E(3) is similar to that given in the proof of Remark 2.1 above and is omitted. Let us prove that E(1) is satisfied.

Observe that

$$
\begin{aligned}
n\,|\phi_n(x, X_2, \ldots, X_n) - \phi_n(\mathbf{X}_n)| &= n^2\,|1/(x^{-1} + \textstyle\sum_2^n X_i^{-1}) - 1/\sum_1^n X_i^{-1}| \\
&= n^2\,|X_1^{-1} - x^{-1}|/(x^{-1} + \textstyle\sum_2^n X_i^{-1})(\sum_1^n X_i^{-1}| \\
&\leq n^2\,|X_1^{-1} - x^{-1}|\,/(\textstyle\sum_2^n X_i^{-1})^2 \leq |X_1^{-1} - x^{-1}|(\sum_2^n X_i/(n-1))^2(n/(n-1))^2 \\
&\leq |X_1^{-1} - x^{-1}|\,(\textstyle\sum_2^n X_i^2/(n-1))(n/(n-1))^2.
\end{aligned}
$$

Hence $\psi_n(x) - E(\phi_n(\mathbf{X}_n)) = 0(n^{-1})$ and thus by Lemma A, $\psi_n(x) - \mu = (\psi_n(x) - E(\phi_n(\mathbf{X}_n))) + (E(\phi_n(\mathbf{X}_n)) - \mu) = 0(n^{-1})$. Then $(\psi_n(x) - \mu)/h_n^2 = n(\psi_n(x) - \mu)/nh_n^2 \to 0$ since $nh_n^2 \to \infty$ and thus Theorem 3.1 and Corollary 3.1 are valid in this case.

Next, let us show the assumption F holds. Since g is continuous at x, $h_n \to 0$, K has support contained in $[-k, k]$, F(1) and F(2) will follow by showing that $M_n(y, z) = \gamma_n(x - h_n y, x - h_n z) - \psi_n(x - h_n y)\psi_n(x - h_n z) = 0(n^{-1})$. It is convenient to write $nM_n(x, y)$ as the sum of the following terms:

$$
\begin{aligned}
N_1 &= n[\gamma_n(x - h_n y, x - h_n z) - E(\phi_n^2(\mathbf{X}_n))], \qquad N_2 = n[E(\phi_n^2(\mathbf{X}_n)) - \mu^2], \\
N_3 &= n[\mu^2 - (E(\phi_n(\mathbf{X}_n)))^2], \qquad N_4 = n[(E(\phi_n(\mathbf{X}_n)))^2 - (\psi_n(x - h_n y))^2],
\end{aligned}
$$

and $N_5 = n[(\psi_n(x - h_n y))^2 - \psi_n(x - h_n y)\psi_n(x - h_n z)]$.

It follows by Lemma A that N_2 is 0(1) and also $N_3 = n(\mu - E(\phi_n(\mathbf{X}_n)))(\mu + E(\phi_n(\mathbf{X}_n))$ is 0(1) by Lemma A and the fact that $E(\phi_n(\mathbf{X}_n)) \leq E(X_1)$. Next, let us verify that $N_1 = 0(1)$. Denote by $s_n = x - h_n y$ and $t_n = x - h_n z$ and write

$$
\begin{aligned}
U_n &= n\,|\phi_n^2(s_n, t_n, X_3, \ldots, X_n) - \phi_n^2(\mathbf{X}_n)| \\
&= n\,|\phi_n(s_n, t_n, X_3, \ldots, X_n) - \phi_n(\mathbf{X}_n)|\,|\phi_n(s_n, t_n, X_3, \ldots, X_n) \\
&\quad + \phi_n(\mathbf{X}_n)|.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&n\,|\phi_n(s_n, t_n, X_3, \ldots, X_n) - \phi_n(\mathbf{X}_n)| \\
&\quad = n^2\,|1/(s_n^{-1} + t_n^{-1} + \textstyle\sum_3^n X_i^{-1}) - 1/\sum_1^n X_i^{-1}| \\
&\quad \leq n^2(X_1^{-1} + X_2^{-1} + s_n^{-1} + t_n^{-1})/(\textstyle\sum_3^n X_i^{-1})^2.
\end{aligned}
$$

Also observe that $|\phi_n(s_n, t_n, X_3, \ldots, X_n) + \phi_n(\mathbf{X}_n)| \leq 2n/\sum_3^n X_i^{-1}$ and thus $U_n \leq C(X_1^{-1} + X_2^{-1} + s_n^{-1} + t_n^{-1})(\sum_3^n X_i^3/n)$. Since K has compact support, $s_n^{-1}(t_n^{-1})$ is uniformly bounded in $y(z)$ and n, when n is sufficiently large. Hence $|N_1| \leq E(U_n)$ is uniformly bounded in y, z, and n, when n is sufficiently large, and thus $N_1 = 0(1)$. It is straightforward to verify that N_4 and N_5 are also $0(1)$ and, moreover, each N_i is dominated by a constant uniformly in y, $z \in [-k, k]$ and n, when n is sufficiently large. Hence both F(1) and F(2) are satisfied and thus Remark 3.1 is valid. □

References

Bhattacharyya, B. B., Franklin, L. A., and Richardson, G. D. (1988), "A Comparison of Nonparametric Unweighted and Length-Biased Density Estimation of Fibres," *Communications in Statistics – Theory and Methods*, **17**, 3629–3644.

Cox, D. R. (1969), "Some Sampling Problems in Technology," pp. 506–527, Symposium on the Foundations of Survey Sampling held at University of North Carolina at Chapel Hill, April 22–26, 1968, edited by N. L. Johnson and H. Smith, Jr., New York: John Wiley.

Gill, R. D., Vardi, Y., and Wellner, J. A. (1988), "Large Sample Theory of Empirical Distributions in Biased Sampling Models," *The Annals of Statistics*, **16**, 1069–1112.

Parzen, E. (1962), "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, **33**, 1065–1076.

Rosenblatt, M. (1956), "Remarks on Some Non-Parametric Estimates of a Density Function, *The Annals of Mathematical Statistics*, **27**, 832–837.

Schuster, E. F. (1985), "Incorporating Support Constraints into Nonparametric Estimators of Densities," *Communications in Statistics – Theory and Methods*, **14**, 1123–1136.

Shapiro, H. S. (1969) *Smoothing and Approximation of Functions*, New York: Van Nostrand.