DEPARTMENT OF MATHEMATICS
AND
COMPUTER SCIENCE

TECHNICAL REPORT SERIES

Inequalities and Bounds for Kernel Length-Biased Density Estimation

by

Broderick O. Oluyede
Department of Mathematics and Computer Science
Georgia Southern University, Statesboro, GA 30460-8093

Number 2002-008
Submitted: March 28, 2002

INEQUALITIES AND BOUNDS FOR KERNEL LENGTH-BIASED DENSITY ESTIMATION

Broderick O. Oluyede
Department of Mathematics and Computer Science
Georgia Southern University, Statesboro, GA 30460
boluyede@gasou.edu

Abstract. In this note non-parametric estimates of the length-biased probability density function and related reliability measures are presented. Non-parametric estimates are also presented under random censoring. Inequalities and bounds for the error of kernel estimators used for the estimation of length-biased probability densities are obtained. Non-asymptotic bounds and stochastic convergence results are established. Inference for length-biased energy functions is developed and implemented.

AMS (MOS) Subject Classifications: 62G07, 62B10.

Keywords: Stochastic inequalities, L_1 density estimation, Random censoring, Stochastic convergence.

1. Introduction

Weighted distributions occur naturally in a wide variety of sampling plans. If items' length are distributed according to the cumulative distribution function(cdf) F, and if the probability of selecting any particular item is proportional to its length, then the lengths of the sampled items have the length-biased distribution. A survey of applications of weighted distributions in general and the length-biased distributions in particular are given by Patil and Rao (1977, 1978). Vardi (1982) derived a non-parametric maximum likelihood estimate of a lifetime distribution F on the basis of two independent samples, one sample of size m from F and the other a sample of size n from the length-biased distribution of F, and studied its distributional properties.

In this paper, we consider the problem estimating the length-biased probability density function and related reliability measures using kernels under random censoring and non-censoring and obtain inequalities and non-asymptotic bounds for the errors of the kernel density estimates. Some stochastic convergence results are also presented. In a wide variety of situations including complex components and test environments, it might be appropriate to assume that the hazard rate increases as the components age, so that the reliability function $\overline{F}$ can be restricted to the class of functions with increasing hazard rate (IHR), where $\overline{F}(x) = \int_x^\infty f(u)du > 0$ for all $x > 0$ and $f(u)$ is the probability density function (pdf). In this case, $\overline{F}$ is absolutely continuous and has IHR if $\lambda_F(x)$ is non-decreasing in x, where $\lambda_F(x) = f(x)/\overline{F}(x)$ is the hazard rate. Similarly, $\overline{F}$ is DHR if $\lambda_F(x)$ is non-increasing in x.

This paper is organized as follows. Section 2 contains some basic results and comparisons. In section 3, we present non-parametric estimates of the length-biased pdf and related reliability measures as well as some inequalities and bounds for the kernel density estimates. Estimates of the pdf and related reliability measures are also presented under random censoring. These results are used to construct test for equality of length-biased informational energies. Section 4 is concerned with results on inference for length-biased reliability and energy functions as well as stochastic convergence. This paper concludes with a discussion in section 5.

2. Some Basic Results and Comparisons

In this section, we present some definitions, basic results and comparisons. Let $\mathcal{F}$ be the class of non-increasing absolutely continuous probability density functions on $[0, \infty)$ satisfying

$$H(0) = 0, \lim_{x \to \infty} H(x) = 1, Sup\{x : H(x) < 1\} = \infty. \tag{1}$$

Note that if the mean of a random variable in $\mathcal{F}$ is finite, it is positive. Assume that the pdf $f \in \mathcal{F}$ is such that $f(0) \leq \Delta$, $\Delta > 0$. Note that $g(0) \leq \Delta$, where $g(x) = xf(x)/\mu_F$ is the length-biased pdf and assumed to be bounded. The length-biased cdf G corresponding to the distribution function F is given by

$$G(y) = \mu^{-1} \int_0^y x dF(x), \tag{2}$$

$y \geq 0$, where $\mu_F = \int_0^\infty \overline{F}(x)dx < \infty$.

The corresponding hazard rate is given by

$$\lambda_G(x) = x\lambda_F(x)/V_F(x), \tag{3}$$

where $V_F(x) = M_F(x) + x$ is called the vitality function. The quantity $M_F(x) = \{\overline{F}(x)\}^{-1} \int_x^\infty \overline{F}(y)dy$, $\overline{F}(x) > 0$ is the mean residual life function. The length-biased reliability function can be written as

$$\overline{G}(x) = V_F(x)\overline{F}(x)/\mu_F. \tag{4}$$

If $V_F(x)$ is increasing in $x \geq 0$, then $\overline{G}(x)/\overline{F}(x)$ is increasing in $x \geq 0$. Note that

$$\overline{G}(x) = \mu^{-1} \int_x^\infty y dF(y) \geq \alpha^{-1}\overline{F}(x) \geq \overline{F}(x), \tag{5}$$

where $\alpha^{-1} = \overline{G}(0)/\overline{F}(0)$. Clearly $V_G(x) \geq V_F(x)$ for all $x \geq 0$ is equivalent to $M_G(x) \geq M_F(x)$ for all $x \geq 0$, since $V_F(x) = E[X - x|X > x] + x$, $x \geq 0$.

Proposition 1 . *Let $\lambda_F^-(x) = f(x)/F(x)$, $F(x) > 0$. Define $\lambda_G^-(x)$ in a similar manner. Then G has heavier tails than F if and only if*

$$\lambda_G(x)/\lambda_G^-(x) \geq \lambda_F(x)/\lambda_F^-(x) \tag{6}$$

for all $x \geq 0$. That is, above a certain value an upper tail of G always has more mass than the corresponding upper tail of F. □

Definition 1 *Let $f(x, \theta)$ be a pdf. The energy function associated with $f(x, \theta)$ is given by*

$$E(F(x, \theta)) = \int f^2(x, \theta)dx, \tag{7}$$

where $\theta \in \Theta$, and Θ is the parameter space.

Definition 2 . *Let h and l be two nonnegative bounded real functions on $\Re$. We say h is exponentially dominated by l if for each $\epsilon \in (0, 1)$, there exist $A(\epsilon) < \infty$ such that*

$$h(x) \leq A(\epsilon)l(x)^{1-\epsilon} \tag{8}$$

for all x. If h and l are exponentially dominated by each other, they are said to be exponentially equivalent.

The usefulness of the above definition is in the comparisons of small values of bounded nonnegative functions h and l respectively. □

Theorem 1 . *Suppose the original pdf f is exponentially dominated by the length-biased pdf g, then the energy function $E(F) = \int f^2(x)d(x)$ is exponentially dominated by $E(G)$.*

P r o o f: Let $h(x) = f^2(x)$ and $l(x) = g^2(x)$, and apply Jensen's inequality to the concave function $y \mapsto y^{1-\epsilon}$ to obtain

$$\begin{aligned} E(f) &= \int h(x)dx \\ &\leq \int C(\epsilon)(l(x))^{1-\epsilon}dx \\ &\leq C(\epsilon)(\int(l(x)dx)^{1-\epsilon}dx \\ &= C(\epsilon)(E(g))^{1-\epsilon}, \end{aligned} \tag{9}$$

where $C(\epsilon) = (A(\epsilon))^2$. □

3. Inequalities and Bounds for Length-Biased Estimation Using Kernels

In this section, we have independent and identically distributed (iid) observations from f and a kernel K, with $K \geq 0$ and $\int K(x)dx = 1$. The well known density estimate of f is

$$f_n(x) = (nh_n)^{-1} \sum_{i=1}^{n} K((x - X_i)/h_n), \tag{10}$$

where $\{h_n : n \geq 1\}$ is a sequence of positive constant tending to zero. A natural estimate of the length-biased pdf g based on the complete set of observations $X_1, X_2, ..., X_n$ and K is given by

$$g_n(x) = xf_n(x)/\mu_n, \tag{11}$$

where μ_n is an estimate of μ_F. Using the estimate $\mu_n = n/\sum_{i=1}^{n} X_i^{-1}$, we have

$$g_n(x) = x(n^2 h_n)^{-1} \sum_{i=1}^{n} K((x - X_i)/h_n)/\sum_{i=1}^{n} X_i^{-1}. \tag{12}$$

The length-biased pdf can also be estimated by using the fact that $\mu_n \leq \mu_n^* = \sum_{i=1}^{n} X_i/n$, to get

$$g_n^*(x) = x(nh_n\mu_n^*)^{-1} \sum_{i=1}^{n} K((x - X_i)/h_n). \tag{13}$$

Next we present estimates of the length-biased pdf and related reliability measures under random censoring. Let the random variable X_i be censored on the right by the random variable D_i, leading to the observation of $T_i = X_i \bigwedge D_i$ and $\delta_i = I(X_i \leq D_i)$, where $\bigwedge$ denote minimum and $I(.)$ the indicator random variable of the event in parenthesis. We assume that the censoring times D_i are independent and identically distributed (iid) and independent of the X_i, $i = 1, 2, ..., n$. The estimate of the length-biased pdf based on the censored data $\{(T_i, \delta_i), i = 1, 2, ..., n\}$ and K is given by

$$g_n^{**}(x) = x(nh_n\mu_n^{**})^{-1} \sum_{i=1}^{n} K((x - X_i)/h_n)q_i, \tag{14}$$

where $q_i = F_n^*(T_{(i)}) - F_n^*(T_{(i-1)})$, $i = 2, 3, ..., n$, $q_1 = F_n^*(T_{(1)})$, and

$$\begin{aligned} \overline{F}_n^*(x) &= 1 - F_n^*(x) \\ &= \prod_{i:T_{(i)} \leq x < T_{(n)}} ((n-i)/(n-i+1))^{\delta_i}, \end{aligned} \tag{15}$$

$\mu_n^{**} = n/\sum_{i=1}^n T_i^{-1}$ and $T_{(i)}$ the order statistics corresponding to T_i. The quantity $\overline{F}_n^*$ is the Kaplan-Meier (K-M) estimate of the reliability function.

The estimate of the length-biased distribution function and hazard rate are given for values of x for which $\overline{G}_n^{**}(x) > 0$, by

$$G_n^{**}(x) = \int_0^x g_n^{**}(y)dy, \tag{16}$$

where $g_n^{**}(x)$ is given by (14), and

$$\tilde{\lambda}_{F_n}(x) = g_n^{**}(x)/\overline{G}_n^{**}(x). \tag{17}$$

On the other hand, if the censored data $\{(T_i, \delta_i), i = 1, 2, ..., n\}$ is available from the length-biased distribution, then an estimate of the parent (original) pdf f is given by

$$f_n^*(x) = \tilde{\mu_n}\tilde{g}_n^*(x)/x, \tag{18}$$

for $x > 0$, where

$$\tilde{g}_n^*(x) = (nh_n)^{-1}\sum_{i=1}^n K((x - X_i)/h_n)p_i, \tag{19}$$

$p_i = \hat{G}_n^*(T_{(i)}) - \hat{G}_n^*(T_{(i-1)}),\ i = 2, 3, ..., n,\ p_1 = \hat{G}_n^*(T_{(1)}),\ \tilde{\mu}_n = n/\sum_{i=1}^n T_i^{-1}$ or any consistent estimate of μ_F and $\overline{\tilde{G}}_n^*$ is the Kaplan-Meier estimate of the reliability function.

Next we present some inequalities and non-asymptotic bounds for the length-biased estimates given above.

Theorem 2 . *Let $g_n(x) = xf_n(x)/\mu_n$ where f_n and μ_n are given by (12). If the kernel K is left sided then for all $g \in \mathcal{F}$,*

$$E\int_0^\infty |g_n(x) - g(x)|dx \le (\int_0^\infty (xf(x))^{1/2}dx)((C^*/nh_n\mu)^{1/2}) + g(0)h_nC^{**}, \tag{20}$$

where $C^* = \int K^2(y)dy$, $C^{**} = \int |y|K(y)dy$, and $\mu = \int \overline{G}(y)dy$.

P r o o f: We have

$$|g_n(x) - g(x)| \le |g_n(x) - E(g_n(x))| + |E(g_n(x)) - g(x)|, \tag{21}$$

so that for fixed x,

$$E\int_0^\infty |g_n(x) - g(x)|dx \le E\int_0^\infty |g_n(x) - E(g_n(x))|dx + E\int_0^\infty |E(g_n(x)) - g(x)|dx.$$

Note that for fixed x, (Devroye(1987)),

$$Var(g_n(x)) \leq (nh_n)^{-1} g(x) \int K^2(y)dy. \tag{22}$$

Applying Fubini theorem and Cauchy-Schwartz inequality we have

$$\begin{aligned} E\int_0^\infty |g_n(x) - E(g_n(x))|dx &\leq \int_0^\infty (Var(g_n(x))^{1/2}dx \\ &\leq \int_0^\infty ((nh_n)^{-1} g(x) \int K^2(y)dy)^{1/2}dx \\ &= \int_0^\infty ((nh_n\mu)^{-1} C^* x f(x))^{-1/2}dx \\ &\leq (\int_0^\infty (xf(x))^{1/2}dx)((C^*/nh_n\mu)^{1/2}). \end{aligned} \tag{23}$$

Also, on applying Lemma 1 given by Datta (1992) we get

$$E\int_0^\infty |E(g_n(x)) - g(x)|dx \leq g(0)h_n C^{**}. \tag{24}$$

Consequently,

$$E\int_0^\infty |g_n(x) - g(x)|dx \leq (\int_0^\infty (xf(x))^{1/2}dx)((C^*/nh_n\mu)^{1/2}) + g(0)h_n C^{**}. \tag{25}$$

□

Corollary 1 . *For all n,*

$$sup_{g\in\mathcal{F}} E\int_0^1 |g_n(x) - g(x)|dx \leq ((C^*/nh_n\mu)^{1/2}), \tag{26}$$

for all $h_n > 0$.

P r o o f: This follows from the fact that $\int_0^\infty (xf(x))^{1/2}dx \leq 1$ and $g(0) < \Delta$. □

Corollary 2 . *For the length-biased kernel density estimate given in (12)*

$$E\int_0^1 |g_n(x) - Eg_n(x)|dx \leq ((C^*/nh_n\mu)^{1/2}). \tag{27}$$

P r o o f: Note that for fixed x, (Devroye(1987)),

$$Var(g_n(x)) \leq (nh_n)^{-1} g(x) \int K^2(y)dy, \tag{28}$$

so that

$$\int_0^\infty Var(g_n(x))dx \leq (nh_n)^{-1} \int_0^\infty g(x)dx \int K^2(y)dy. \tag{29}$$

That is,

$$\int_0^\infty Var(g_n(x))dx \leq (nh_n)^{-1} \int K^2(y)dy.$$

Consequently,

$$\begin{aligned} E\int_0^1 |g_n(x) - E(g_n(x))|dx &\leq \int_0^1 (Var(g_n(x))^{1/2}dx \\ &\leq (\int_0^\infty (Var(g_n(x))dx)^{1/2} \\ &\leq ((nh_n)^{-1}C^*)^{1/2}, \end{aligned} \tag{30}$$

where $C^* = \int K^2(y)dy$. The inequalities in (30) follow directly from the Cauchy-Schwartz inequality. □

4. Inference for Length-Biased Energy Measure

In this section, statistical inference for length-biased energy function is presented. Some convergence results are given.

Theorem 3 . *Under appropriate conditions (Parzen (1962)), if* $(nh_n)^{1/2}(\mu_n - \mu) \xrightarrow{P} 0$, *then*

$$(nh_n)^{1/2}(g_n(x) - E(g_n(x)) \xrightarrow{D} N(0, \sigma^2(x)), \tag{31}$$

for $x \in [0, a], a < \infty$, *where*

$$\sigma^2(x) = g(x) \int_{\Re} K^2(y)dy, \tag{32}$$

and $\xrightarrow{D}$ *denote convergence in distribution.*

P r o o f: The proof follows from the result of Parzen (1962) page 1073 and Roussas (1990) Theorem 3.1. □

Theorem 4 .

1. *If* T_i^{-1} *has finite second moment, then*

$$(nh_n)^{1/2}(\mu_n^* - \mu) \xrightarrow{P} 0.$$

2. *If* $E(T_i^{2r}) < \infty$, *where* r *is a natural number, then*

$$(nh_n)^{1/2}(f_n^*(x) - E(f_n^*(x)) \longrightarrow 0 \tag{33}$$

as $n \longrightarrow \infty$, *where* f_n^* *is given by* (18).

P r o o f: The proof follows from the result of Van Ryzin (1969), Theorem 1. □

Theorem 5 . *Let* $\hat{\theta}$ *be the maximum likelihood estimate of* $\theta \in \Theta$. *If* $B = (b_1, b_2, ..., b_k)^T$ *and* $\theta = (\theta_1, \theta_2, ..., \theta_k)^T$ *where* $b_i = \partial E(\theta)/\partial\theta_i$ *and* $\sigma^2(\theta) = B^T I_G^{-1}(\theta)B > 0$, *then*

$$\sqrt{n}(E(\hat{\theta}) - E(\theta)) \xrightarrow{L} N(0, B^T I_G^{-1} B), \tag{34}$$

as $n \to \infty$, *where* $I_G(\theta)$ *is the Fisher information matrix.*

P r o o f: The asymptotic normality of $\sqrt{n}(\hat{\theta} - \theta)$ and a Taylor's expansion of $E(G_\theta) = E(\theta)$ around θ, gives the desired result. □

The results above can be used for the purpose of statistical inference. Now consider the hypothesis, $H_0 : E(G_\theta) = E(G_{\theta_0})$ against $H_0 : E(G_\theta) > E(G_{\theta_0})$, where $E(G_{\theta_0}) = E(\theta_0)$ is a specified value of the length-biased population informational energy. An appropriate test statistics for testing the hypothesis is given by

$$T^* = T^2 I_{(0,\infty)}(T), \tag{35}$$

where

$$T = \frac{\sqrt{n}(E(\hat{\theta}) - E(\theta_0))}{\sigma^2(\hat{\theta})}. \tag{36}$$

The statistic T has in the limit the standard normal distribution so that T^2 has a chi-square distribution with one degree of freedom.

A size $\alpha-$test will reject H_0 if $T^* > \chi^2_{1,2\alpha}$. This follows from the fact that

$$\begin{aligned} \lim_{n\to\infty} P_{H_0}(T^* > C) &= \lim_{n\to\infty} P_{H_0}(T^2 I_{(0,\infty)}(T) > C) \\ &= \frac{1}{2} P_{H_0}(\chi^2_{(1)} > C) \end{aligned} \tag{37}$$

if $C > 0$, and 1 if $C \le 0$, where $\chi^2_{(1)}$ denote a random variable having a chi-square distribution with one degree of freedom.

A test of equality of several informational energies, that is

$$H_0 : E(\theta_1) = E(\theta_2) = ... = E(\theta_k), \tag{38}$$

rejects H_0 if $T_1 > C$, where

$$T_1 = \sum_{i=1}^{k} \{(E(\hat{\theta}_i) - \delta(\hat{\theta}_i))/\sqrt{\sigma^2(\hat{\theta}_i)/n_i}\}^2, \tag{39}$$

$\delta(\hat{\theta}_i) = \frac{\sum_{i=1}^{k} E(\hat{\theta}_i)/[\sigma^2(\hat{\theta}_i)/n_i]}{\sum_{i=1}^{k}[\sigma^2(\hat{\theta}_i)/n_i]}$, and $\sigma^2(\theta) = B^T I_G^{-1}(\theta)B > 0$, and C is chosen such that

$$P_{H_0}(T_1 > C) = \alpha. \tag{40}$$

The statistic T_1 has in the limit as $n = \sum_{i=1}^{k} n_i$ goes to infinity the chi-square distribution with $k-1$ degrees of freedom. Consequently, the null hypothesis is rejected at level α if $T > \chi^2_{k-1;\alpha}$.

Let $X_{11}, X_{12}, ..., X_{1n_i}$, be independent random samples with length-biased distribution functions G_i, $i = 1, 2$, respectively. An estimate of the energy function $E(G_j) = \int g_j^2(x)dx$ where $g(x) = xf(x)/\mu_F$ proposed by Bhattacharyya and Roussas (1969) is given by

$$\tilde{E}(G_j) = \int \hat{g}_j^2(x)dx, \tag{41}$$

where

$$\hat{f}_j(x) = [n_j h_j]^{-1} \sum_{i=1}^{n_j} K((x - X_{ji})/h_j), \tag{42}$$

$\hat{g}_j(x) = x\hat{f}_j(x)/\mu_n$ for fixed x, μ_n a consistent estimate of μ_F, h_j is a bandwidth, and K is a known symmetric and bounded function probability density function such that $\lim_{y\to\infty} yK(y) = 0$. See equations (12) and (13). Ahmad (1976) proposed the estimate

$$\hat{E}(F_j) = [n_j^2 h_j]^{-1} \sum_{r=1}^{n_j}\sum_{s=1}^{n_j} K((X_{jr} - X_{js})/h_j). \tag{43}$$

Bhattacharyya and Roussas (1969) estimate above is a special case of Ahmad (1976) estimate, since

$$\tilde{E}(F_j) = [n_j^2 h_j]^{-2} \sum_{r=1}^{n_j}\sum_{s=1}^{n_j} K^{(2)}((X_{jr} - X_{js})/h_j), \tag{44}$$

where $K^{(2)}(y)$ is the convolution of $K(y)$ with itself. See Ahmad and Kochar (1988) for details.

A test statistics for testing $H_0 : E(G_1) = E(G_2)$ is given by

$$T(G_1, G_2) = \int (\hat{E}(G_1) - \hat{E}(G_2))^2 dx. \tag{45}$$

In the case $h_1 = h_2 = h$, $g_1 = g$ is fixed probability density function and h is a function such that $g_2 = g + \gamma h$ is a probability density function for sufficiently small $|\gamma|$, the α-level test rejects H_0 if $T(G) = T(g) > t_g$, where

$$P_{H_0}(T(g) > t_g) = \alpha, \tag{46}$$

and t_g is the α-level critical point of the distribution of $T(G)$ under the null hypothesis $H_0 : \gamma = 0$, that is $\mathrm{E}(\mathrm{G}_1) = E(G_2)$.

Let $H_1 = H_1(\gamma)$ denote the alternative hypothesis that $\gamma = \delta/\sqrt{nh}, \delta \neq 0$, then

$$\pi(\delta) = \lim_{n\to\infty} P_{H_1}(T(g) > t_g) \to 1, \tag{47}$$

as $|\delta| \to 0$ provided $0 < |\delta| < \infty$. Also, $\alpha < \pi(\delta) < 1$ for $0 < |\delta| < \infty$. $\square$

5. Discussion

In this paper, estimates of length-biased pdf and related reliability functions using kernels are presented. Inequalities, bounds and convergence results are also presented for the length-biased pdf. The use of the length-biased energy function for statistical comparisons and inferences in terms of parameter sets is developed. An intuitive grasp of notions involving informational energy functions follows by noting that the scale parameters for the distributions are ordered. Non-parametric estimates under random censoring are also presented. See references therein. Procedures for testing for homogeneity of length-biased energy functions are obtained and implemented.

Acknowledgement. The author wishes to thank the referees for valuable comments and suggestions.

References

[1] Ahmad, I.A., *On Asymptotic Properties of an Estimate of a Functional of a Probability Density Function*, Scan. Actuarial J. 3, 176-181,(1976)

[2] Ahmad, I.A. and Kochar, S.C., *Testing for Dispersive Order*, Statist. and Probab. Letters 7, 179-185, (1988)

[3] Bhattacaryya, G.K. and Roussas, G.G., *Estimation of a Certain Functional of a Probability Density Function*, Skand. Aktuar. Tidskr. 52, 201-206 (1969)

[4] Datta, S., *Some Nonasymptotic Bounds for L_1 Density Estimation Using Kernels*, Ann. Statist. Vol. 20, No. 3, 1658-1667, (1992)

[4] Devroye, L., *A Course in Density Estimation*, Birkhauser, Boston, (1987)

[5] Parzen, E.,*On Estimation of a Probability Density and Mode*, Ann. Math. Statist. 33, 1065-1076, (1962)

[6] Patil, G.P., Rao, C.R., *The Weighted Distributions: A survey and their applications. Applications in Statistics*, ed. P.R. Krishnaiah, North Holland Publishing Company, 3 83-405, (1977)

[7] Patil, G.P., Rao, C.R., *Weighted Distributions and Size-Biased Sampling with Applications to Wildlife and Human Families*, Biometrics, 34, 179-189, (1978)

[8] Roussas, G.G., *Asymptotic Normality of the Kernel Estimate under Dependence Conditions: Application to Hazard Rate*, J. Statist. Plann. Inference, Vol. 25, 181-194, (1990)

[9] Shannon C.E., *A Mathematical Theory of Communication*, Bell System Technical Journal 27, 379-423, (1948)

[10] Vardi, Y., *Nonparametric Estimation in the Presence of Bias*, Ann. Statist., Vol. 10, No. 2, 616-620, (1982)

[11] Van Ryzin, J., *On Strong Consistency of Density Estimates*, Ann. Statist., Vol. 40, No. 5, 1765-1772, (1969)