Technical Report No. 01/07, January 2007

A NEW SMOOTH DENSITY ESTIMATOR FOR NON-NEGATIVE RANDOM VARIABLES

Yogendra P. Chaubey, Arusharka Sen and Pranab K. Sen

A New Smooth Density Estimator for Non-negative Random Variables

Yogendra P. Chaubey
Department of Mathematics and Statistics
Concordia University, Montreal, Canada H4B 1R6
E-mail: chaubey@alcor.concordia.ca
Arusharka Sen
Department of Mathematics and Statistics
Concordia University, Montreal, Canada H4B 1R6
Pranab K. Sen
Department of Statistics and Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA NC 27599-7400

December 13, 2006

ABSTRACT

Commonly used kernel density estimators may not provide admissible values of the density or its functionals at the boundaries for densities with restricted support. For smoothing the empirical distribution a generalization of the Hille's lemma, considered here, alleviates some of the problems of kernel density estimator near the boundaries. For nonnegative random variables which crop up in reliability and survival analysis, the proposed procedure is thoroughly explored; its consistency and asymptotic distributional results are established under appropriate regularity assumptions. Methods of obtaining smoothing parameters through cross-validation are given, and graphical illustrations of the estimator for continuous (at zero) as well as discontinuous densities are provided.

KEY WORDS: Asymptotics; boundary correction; cross-validation; empirical distribution; hazard function; Hille's lemma; kernel density estimator; survival function.

1 Introduction

In reliability and survival analysis, typically, a non-negative random variable (r.v.) X, admitting a continuous probability density function (pdf) $f(x)$, is conceived. The related objects of interest are: the *cumulative distribution function* (cdf) F, the *survival function* (sf) S, defined by

$$S(x) = 1 - F(x) = \int_x^\infty f(y)dy \quad x \geq 0, \tag{1.1}$$

the hazard function $r(x) := f(x)/S(x)$ and so on.

Based on a random sample $(X_1, X_2, ..., X_n)$, the *empirical distribution function* (edf) F_n is defined as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), x \geq 0, \tag{1.2}$$

and $S_n(x) = 1 - F_n(x)$, the empirical survival function is $n^{-1}\sum_{i=1}^n I(X_i > x), x \geq 0$. In a broad sense, they are optimal nonparametric estimators (of F and S respectively). However,they are step functions, and hence, do not directly amend to estimation of $f(.)$ (and as a result, for instance, to that of $r(.)$), which require the estimation of the derivative of F (or S). Kernel smoothing, histogram methods, splines, and orthogonal functions (among others) have therefore been explored for smooth estimation of $f(.)$ and its functionals; we refer to Eubank (1988), Devroye(1989) and Wand and Jones (1995) where other references have been cited.

The smooth *kernel estimator* [Rosenblatt (1956), Parzen (1962)] of $f(.)$ is of the form

$$\hat{f}_n(x) = (nh_n)^{-1}\sum_{i=1}^{n} k((X_i - x)/h_n), \tag{1.3}$$

where $h_n(> 0)$, known as the *band-width* is so chosen that $h_n \to 0$ but $nh_n \to \infty$, as $n \to \infty$; $k(.)$ is termed the *kernel function*, and it is typically assumed to be a symmetric pdf with zero mean and unit variance. This estimator suffers from the following two kinds of *boundary bias*:

B1) *Positive mass outside support.* Silverman (1986) noted the inadequacy of the kernel estimator in assigning positive mass to some $x \in (-\infty, 0)$, while illustrating the method for random variables taking only positive values, as in reliability and survival analysis. Even otherwise, as remarked in Wand, Marron and Rupert (1991), the estimator in (1.3) “works well for densities that are not far from Gaussian in shape,” however, “it can perform very poorly when the shape is far from Gaussian,” especially near the boundaries.

B2) *Failure to estimate discontinuity at boundary.* For densities on $[0, \infty)$ with $f(0) > 0$, such as the Exponential density, $f_n(0)$ often does not consistently estimate $f(0)$.

Among different approaches suggested to deal with data on $[0, \infty)$ is the transformation method. For example by taking logarithmic transformation of data, standard arguments lead to the estimator of the untransformed data given by $\hat{f}_n(x) = (1/x)\hat{g}(\log x)$, where $\hat{g}(.)$ denotes the kernel density estimator using the transformed data. The presence of the multiplier $1/x$, usually gives rise to a spike in the density graph, which may not be an attractive feature of this estimator. General transformation methods were studied by Wand, Marron and Rupert (1991) and Rupert and Wand (1992) amongst others. However, these methods could not be fully satisfactory for reducing the boundary bias. Marron and Rupert (1994) in a later paper, address this problem in detail by proposing a three step computer intensive transformation method. Simplicity of the transformation method still makes this an attractive choice for smoothing the histogram, but the interest still persists in finding a method similar to kernel smoothing without the data transformation. This desire led Bagai and Prakasa Rao (1996) to propose replacing the kernel k by a non-negative density function k^*, such that $\int_0^\infty x^2 k^*(x)dx < \infty$. They show that the resulting estimator has similar asymptotic properties as the usual kernel estimator under some regularity conditions. This certainly alleviates the problem of positive probability in the negative region, however, as noted in Bagai and Prakasa Rao (1996), for estimating $f(x)$, only the first r order statistics contribute to the value of the modified estimator, where $X_{(r)} < x \leq X_{(r+1)}$, $X_{(i)}$ denoting the i^{th} order statistic. This may affect the behavior of the smooth estimator at the boundary.

Chaubey and Sen (1996) proposed a density estimator as the derivative of a smooth version of the *edf* by adapting the so called Hille's (1948) smoothing lemma, *albeit* in a stochastic setup which, in contrast to the proposal of Bagai and Prakasa Rao (1996), uses the whole data.

An interesting class of estimators was proposed by Chen (1999), using Gamma kernels, as well as Scaillet (2004), using inverse Gaussian and reciprocal inverse Gaussian kernels. These estimators do not suffer from boundary bias. However, their variances blow up at $x = 0$, to circumvent which the authors give two kinds of variance formulae, one for $x/b \to \infty$ and the other for $x/b \to \kappa > 0$, where b is the bandwidth. This appears somewhat arbitrary to us as it does not give a clear picture of what happens at or near $x = 0$. Secondly, there is no *graphical* illustration of the method for densities such as the Exponential, so it is not clear how it works regarding discontinuity at the boundary (see B2 above).

In this paper, we propose an estimator based on a generalization of Hille's smoothing lemma coupled with a perturbation idea to take care of the boundary bias. In Section 2, we present the generalized lemma and derive our estimator from it. The estimator is simple, lends itself easily to estimation of functionals of density such as its derivatives and to smoothing parameter choice by cross-validation via an explicit variance formula. In this section we also show that the estimators mentioned above, namely the kernel, the log-transformation, those of Chaubey and Sen (1996), Chen (2000) and Scaillet (2004), are in fact *all* motivated by this general lemma. Moreover, we point out that our perturbation idea is a general technique and can be successfully applied to, for instance, the Chen (2000) and Scaillet (2004) estimators to prevent their variances from blowing up at $x = 0$.

Section 3 gives the asymptotic properties of the proposed estimator, namely *uniform consistency* and *asymptotic normality.* In Section 4, we present methods of smoothing parameter choice and a simulation study. To demonstrate effectiveness of our method in handling boundary bias, we estimate both *Weibull* ($f(0) = 0$, f *continuous* at 0) and *Exponential* ($f(0) > 0$, f *discontinuous* at 0) densities, and the results are very satisfactory. The proofs of results in Section 3 are deferred to the Appendix.

2 A General Smooth Estimator of the Density Function

The following discussion gives a general approach to density estimation which is specialized to the case of *non-negative* data. The key to the proposal is the following generalization of the Hille's lemma, which is a slight variation of Lemma 1 given in Feller (1965, §VII.1).

Lemma 1: *Let u be any bounded and continuous function. Let $G_{x,n}, n = 1, 2, ...$ be a family of distributions with mean $\mu_n(x)$ and variance $h_n^2(x)$ then we have as $\mu_n(x) \to x$ and $h_n(x) \to 0$*

$$\tilde{u}(x) = \int_{-\infty}^{\infty} u(t) dG_{x,n}(t) \to u(x). \tag{2.1}$$

The convergence is uniform in every subinterval in which $h_n(x) \to 0$ and u is uniformly continuous.

This generalization may be adapted for smooth estimation of the distribution function by replacing $u(x)$ by the empirical distribution function $F_n(x)$ as given below ;

$$\tilde{F}_n(x) = \int_{-\infty}^{\infty} F_n(t) dG_{x,n}(t) \tag{2.2}$$

Strong convergence of $\tilde{F}_n(x)$ parallels to that of the strong convergence of the empirical distribution function as stated in the following theorem.

Theorem 1: *If $h \equiv h_n(x) \to 0$ for every fixed x as $n \to \infty$ we have*

$$\sup_x |\tilde{F}_n(x) - F(x)| \overset{a.s.}{\to} 0 \tag{2.3}$$

as $n \to \infty$.

Technically, $G_{x,n}$ can have any support but it may be prudent to choose it so that it has the same support as the random variable under consideration; because this will get rid of the problem of the estimator assigning positive mass to undesired region.

For $\tilde{F}_n(x)$ to be a proper distribution function, $G_{x,n}(t)$ must be a decreasing function of x, which can be shown using an alternative form of $\tilde{F}_n(x)$:

$$\tilde{F}_n(x) = 1 - \frac{1}{n} \sum_{i=1}^{n} G_{x,n}(X_i). \tag{2.4}$$

This leads us to propose a smooth estimator of the density given by

$$\tilde{f}_n(x) = \frac{d\tilde{F}_n(x)}{dx} = -\frac{1}{n} \sum_{i=1}^{n} \frac{d}{dx} G_{x,n}(X_i). \tag{2.5}$$

Densities with Non-Negative Support

Using the representation (2.4), we now propose the following estimators of the distribution and density functions with support $[0, \infty)$, which generalizes the estimator in Chaubey and Sen (1996). Let $Q_v(x)$ represent a distribution on $[0, \infty)$ with mean 1 and variance v^2, then an estimator of $F(x)$ is given by

$$F_n^+(x) = 1 - \frac{1}{n} \sum_{i=1}^{n} Q_{v_n}\left(\frac{X_i}{x}\right), \tag{2.6}$$

where $v_n \to 0$ as $n \to \infty$. Obviously, this choice uses $G_{(x,n)}(t) = Q_{v_n}(t/x)$ which is a decreasing function of x.

This leads to the following density estimator

$$\frac{d}{dx}(F_n^+(x)) = \frac{1}{nx^2}\sum_{i=1}^{n} X_i \; q_{v_n}\left(\frac{X_i}{x}\right),$$

where $q_v(.)$ denotes the density corresponding to the distribution function $Q_v(.)$.

However, *the above estimator may not be defined at* $x = 0$, except in cases where $\lim_{x\to 0} \frac{d}{dx}(F_n^+(x))$ exists. Moreover, this limit is typically zero, which is acceptable only when we are estimating a density f with $f(0) = 0$.

Hence in view of the more general case where $0 \leq f(0) < \infty$, we considered the following *perturbed* version of the above density estimator:

$$f_n^+(x) = \frac{1}{n(x+\epsilon_n)^2}\sum_{i=1}^{n} X_i \; q_{v_n}\left(\frac{X_i}{x+\epsilon_n}\right), \; x \geq 0 \tag{2.7}$$

where $\epsilon_n \downarrow 0$ at an appropriate (sufficiently slow) rate as $n \to \infty$. In the sequel, we illustrate our method by taking $Q_v(.)$ to be the Gamma ($\alpha = 1/v^2, \beta = v^2$) distribution function.

Next we present a comparison of our approach with some existing estimators.

Kernel Estimator. The usual kernel estimator is a special case of the representation given by Eq. (2.5), by taking $G_{x,n}(.)$ as

$$G_{x,n}(t) = K\left(\frac{t-x}{h}\right), \tag{2.8}$$

where $K(.)$ is a distribution function with mean zero and variance 1.

Transformation Estimator of Wand *et al.* The well known logarithmic transformation approach of Wand, Marron and Rupert (1991) leads to the following density estimator:

$$\tilde{f}_n^{(L)}(x) = \frac{1}{nh_n x}\sum_{i=1}^{n} k(\frac{1}{h_n}\log(X_i/x)),$$

where $k(.)$ is a density function (kernel) with mean zero and variance 1. This is easily seen to be a special case of Eq. (2.5), taking $G_{x,n}$ again as in Eq. (2.8) but applied to $\log x$. This approach, however, creates problem at the boundary which led Rupert and Marron (1994) to propose modifications that are computationally intensive.

Estimators of Chen and Scaillet. Chen's (2000) estimator is of the form

$$\hat{f}_C(x) = \frac{1}{n} \sum_{i=1}^{n} g_{x,n}(X_i),$$

where $g_{x,n}(.)$ is the Gamma$(\alpha = a(x,b), \beta = b)$ density with $b \to 0$ and $ba(x,b) \to x$. This also can be motivated from Eq. (2.1) as follows: take $u(t) = f(t)$ and note that the integral $\int f(t) g_{x,n}(t) dt$ can be estimated by $n^{-1} \sum_{i=1}^{n} g_{x,n}(X_i)$. This approach controls the boundary bias at $x = 0$; however, the variance blows up at $x = 0$, and computation of mean integrated squared error (MISE) is not tractable. Moreover, estimators of derivatives of the density are not easily obtainable because of the appearance of x as argument of the Gamma function.

Scaillet's (2004) estimators replace the Gamma kernel by inverse Gaussian (IG) and reciprocal inverse Gaussian (RIG) kernels. These estimators are more tractable than Chen's; however, the IG-kernel estimator assumes value zero at $x = 0$, which is not desirable when $f(0) > 0$, and the variances of the IG as well as the RIG estimators blow up at $x = 0$. Bouezmarni and Scaillet (2005), however, demonstrate good finite-sample performance of these estimators.

It is interesting to note that one can immediately define a *Chen-Scaillet version* of our estimator, namely,

$$f_{n,C}^{+}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{x} \, q_{v_n} \left(\frac{X_i}{x} \right);$$

on the other hand, *our version* (i.e., perturbed version) of the Chen-Scaillet estimator would be

$$\hat{f}_C^{+}(x) = \frac{1}{n} \sum_{i=1}^{n} g_{x+\epsilon_n,n}(X_i);$$

we strongly believe that this version will not have the problem of variance blowing up at $x = 0$.

The above comments show that the estimator of this paper is based on two very general ideas: a) *Hille's Lemma*, which is seen to be the motivation behind *all* density estimators; b) the *perturbation* idea, which can handle the boundary problems (not just bias but variance too) of most estimators.

Our estimator, besides being straightforward and free from boundary problems, yields itself to bandwidth selection by cross validation. Furthermore it is *uniformly consistent* on $[0, \infty)$ as well as asymptotically normal (see Theorems 3 and 4 in the next section).

3 Asymptotic Properties of Estimators

3.1 Asymptotic Properties of $F_n^+(x)$

The strong consistency holds in general for the estimator $\tilde{F}_n(x)$ as is clear from Theorem 2 and hence it naturally extends to the estimator $F_n^+(x)$. The representation given by Eq. (2.6) allows us to study further its asymptotic distribution and other properties. First, we show that $F_n^+(x)$ is asymptotically unbiased. Note that from Eq. (2.2) we can write, by taking the expectation inside the integral sign, as justified by Fubini's theorem:

$$E[F_n^+(x))] = \int_0^\infty F(ux) q_{v_n}(u) du.$$

Assuming the boundedness of the 2nd derivative of the density, we can write,

$$F(ux) = F(x) + x(u-1)f(x) + (1/2)x^2(u-1)^2 f''(x) + o((x-1)^2).$$

Substituting this in the previous equation, we have

$$E[F_n^+(x)] \approx F(x) + (1/2)t^2 f''(x) v_n^2 \tag{3.1}$$

Therefore, assuming that $v_n \to 0$ as $n \to \infty$, we find the smooth estimator to be asymptotically unbiased. Moreover, we can show that for large n, the smooth estimator can be arbitrarily close to the *edf* by proper choice of v_n, as given in the following theorem.

Theorem 2: *Assuming that f has a bounded derivative, let $nv_n^2 = o(n^{-1})$, then for some $\delta > 0$, we have, with probability one,*

$$\sup_{x \geq 0} |F_n^+(x) - F_n(x)| = O\left(n^{-3/4}(\log n)^{1+\delta}\right). \tag{3.2}$$

For the case of Poisson weights, Chaubey and Sen (1996) obtained the same rate using the properties of tail sum of Poisson probabilities. This shows that the asymptotic distribution of $F_n^+(x)$ can be obtained through that of $F_n(t)$, namely,

$$\sqrt{n}(F_n^+(x) - F(x)) \sim AN(0, F(x)(1 - F(x)).$$

In the next section, we study the properties of the derived density estimator given in Eq. (2.7).

3.2 Asymptotic Properties of $f_n^+(x)$

The formula given in Eq. (2.7) for the density estimator is useful for computational purpose, however, the following integral representation,

$$f_n^+(x) = \int_0^\infty F_n(t)\frac{d}{dx}[g_{x+\epsilon_n,n}(t)]dt, \tag{3.3}$$

where $g_{x,n}(t) = \frac{d}{dt}G_{x,n}(t)$, will be used for studying the asymptotic properties.

First we establish uniform strong consistency of the density estimator $f_n^+(x)$ as given in the following theorem. As can be seen in this theorem, the convergence of $v_n \to 0$ is coupled with an added condition on the derivatives of the densities $q_{v_n}(.)$. Here onwards we will omit the subscript n from v_n and assume $v \equiv v_n$.

Theorem 3: *If*

A. $v_n \to 0$, $\epsilon_n \to 0$ *as* $n \to \infty$ *and*

B. $\sup_{x\geq 0}\int_0^\infty |\frac{d}{dx}[g_{x+\epsilon_n,n}(t)]|dt = o\left((\frac{\log\log n}{n^{1/2}})^{-1}\right)$,

C. $\sup_{u>0,v>0} uq_v(u) < \infty$,

D. $f(\cdot)$ *is Lipschitz continuous on* $[0,\infty)$,

then we have

$$\sup_{x\geq 0}|f_n^+(x) - f(x)| \xrightarrow{a.s.} 0 \tag{3.4}$$

as $n \to \infty$.

Remark 1: Condition B of Theorem 3 holds, for example, if we take $g_{x,n}(t) = \frac{1}{x}q_v(t/x)$, where $q_v(.)$ is the the gamma density given by,

$$q_v(t) = \frac{1}{\beta^\alpha\Gamma(\alpha)}t^{\alpha-1}\mathrm{e}^{-\frac{t}{\beta}},\ t>0,$$

where

$$\alpha = 1/v^2 \text{ and } \beta = 1/\alpha. \tag{3.5}$$

In this case

$$\begin{aligned}\int_0^\infty |\frac{d}{dx}g_{x+\epsilon_n,n}(t)|dt &= \frac{1}{(x+\epsilon_n)^2v^2}\int_0^\infty |t-(x+\epsilon_n)|g_{x+\epsilon_n,n}(t)dt \\ &= O\left(\frac{1}{(x+\epsilon_n)v}\right),\end{aligned}$$

so that $\sup_{x\geq 0}\int_0^\infty |\frac{d}{dx}g_{x+\epsilon_n,n}(t)|dt = O((v\epsilon_n)^{-1})$.

Choosing $v\epsilon_n = O(n^{-\frac{1}{2}+\delta})$ for some $0<\delta<1/2$, will satisfy Condition B of the theorem.

Equation (2.7) shows that the density estimator is the mean of *i.i.d.* random variables, $Y_{in} = (X_i/(x+\epsilon_n)^2)\; q_v(X_i/(x+\epsilon_n)), i = 1,2,...,n$. The following theorem gives conditions, on q_v and f, under which it is asymptotically normal and gives the form of its asymptotic variance.

Theorem 4: *Assume the following conditions:*

(F) $f(\cdot)$ *is Lipschitz continuous on* $[0,\infty)$*;*

(G1) $\int_0^\infty (q_v(t))^m dt = O(v^{-(m-1)})$ *as* $v\to 0$*, for* $1\leq m\leq 3$*, and* $I_2(q) := \lim_{v\to 0} v\int_0^\infty (q_v(t))^2 dt$ *exists;*

(G2) with $q^*_{m,v}(t) := (q_v(t))^m / \int_0^\infty (q_v(w))^m dw$*,* $1\leq m\leq 3$*, and as* $v\to 0$*,*

(i) $\mu_{m,v} := \int_0^\infty t q^*_{m,v}(t)dt = 1 + O(v)$*,*

(ii) $\sigma^2_{m,v} := \int_0^\infty (t-\mu_{m,v})^2 q^*_{m,v}(t)dt = O(v^2)$*,*

(iii) $\sup_{0<v<\epsilon}\int_0^\infty t^{4+\delta} q^*_{m,v}(t)dt < \infty$*, for some* $\delta>0,\ \epsilon>0$*,*

(H) $nv\to\infty$*,* $nv\epsilon_n\to\infty$*,* $nv^3\to 0$*,* $nv\epsilon_n^2\to 0$ *as* $n\to\infty$*;*

then as $n\to\infty$*,*

(a) $\sqrt{nv}(f_n^+(x) - f(x)) \to$ *Normal* $\left(0, I_2(q)\frac{f(x)}{x}\right)$*, for* $x>0$*,*

(b) $\sqrt{nv\epsilon_n}(f_n^+(0) - f(0)) \to$ *Normal* $(0, I_2(q)f(0))$

in distribution.

Remark 2: It is easily seen that the conditions G2, (i) – (iii), mean the following: let $T^*_{m,v}$ be a random variable with density $q^*_{m,v}$; then $T^*_{m,v}\to 1$ in L_p for $1\leq p\leq 4$ as $v\to 0$, for $m = 1,2,3$.

Remark 3: We illustrate the conditions (G1) and (G2) with $q_v(t)$ as the Gamma density as given in Eq. (3.5). For $m\geq 1$, $(q_v(t))^m = (\int_0^\infty (q_v(w))^m dw)\; q^*_{m,v}(t)$, where $q^*_{m,v}(t)$ is a Gamma density with $\alpha = \frac{m}{v^2} - m + 1, \beta = \frac{v^2}{m}$ and

$$
\begin{aligned}
\int_0^\infty (q_v(t))^m dt &= \frac{(\frac{1}{v^2})^{\frac{m}{v^2}}}{(\frac{m}{v^2})^{\frac{m}{v^2}-m+1}}.\frac{\Gamma(\frac{m}{v^2}-m+1)}{(\Gamma(\frac{1}{v^2}))^m} \\
&\approx \frac{1}{\sqrt{m(2\pi)^{m-1}}}.\frac{1}{v^{m-1}\sqrt{1-v^2}}, \text{ as } v \to 0,
\end{aligned}
$$

using *Stirling's approximation* for the Gamma function. Thus we verify that for any $m \geq 1$,

(G1) $I_2(q) = \lim_{v\to 0} v \int_0^\infty (q_v(t))^2 dt = 1/\sqrt{4\pi}$ exists;

(G2)(i) $\mu_{m,v} = \int_0^\infty t q^*_{m,v}(t)dt = 1 - ((m-1)/m)v^2$;

(G2)(ii) $\sigma^2_{m,v} = \int_0^\infty (t-\mu_{m,v})^2 q^*_{m,v}(t)dt = (1-((m-1)/m)v^2)v^2/m$;

(G2)(iii) for any $k \geq 1$, and any $\epsilon > 0$,

$$
\sup_{0<v<\epsilon} \int_0^\infty t^k q^*_{m,v}(t)dt = \sup_{0<v<\epsilon} \frac{\Gamma(k+\frac{m}{v^2}-m+1)}{\frac{m}{v^2}\Gamma(\frac{m}{v^2}-m+1)} = O(1+(\frac{k-m+1}{m})v^2) < \infty.
$$

4 Cross-Validation and Numerical Results

The leading terms in the bias and variance of $f_n^+(x)$ may be shown to be as given in the following equation:

$$
\begin{aligned}
Bias[f_n^+(x)] &= x\, f'(x)v_n^2 + \epsilon_n\, f'(x)(1+v_n^2) + o(v_n^2+\epsilon_n) \\
&= (x\, v_n^2 + \epsilon_n)f'(x) + o(v_n^2+\epsilon_n), \quad v_n^2 \to 0, and\ \epsilon_n \to 0. \qquad (4.1) \\
Var[f_n^+(x)] &\approx \frac{1}{n(x+\epsilon_n)}\frac{I_2(q)}{v_n}\,[f(x) + (x+\epsilon_n)xf'(x)O(v_n) + (x+\epsilon_n)\epsilon_n f'(x) \\
&\quad + o((x+\epsilon_n)xf'(x)O(v_n) + (x+\epsilon_n)\epsilon_n f'(x))] \\
&= \frac{I_2(q)f(x)}{nv_n(x+\epsilon_n)} + o((nv_n)^{-1}), \ \ v_n \to 0\ \epsilon_n \to 0,\ \ nv_n \to \infty. \qquad (4.2)
\end{aligned}
$$

Therefore, by combining the above formulas we obtain the *mean squared error* as

$$
\begin{aligned}
MSE[f_n^+(x)] &\approx \frac{I_2(q)f(x)}{nv_n(x+\epsilon_n)} + [(x\, v_n^2+\epsilon_n)f'(x)]^2 \\
&\quad + o(v_n^2+\epsilon_n) + o((nv_n)^{-1}), \qquad (4.3)
\end{aligned}
$$

and the *mean integrated squared error*:

$$\begin{aligned} MISE(f_n^+) &= \int_0^\infty MSE[f_n^+(x)]\,dx \\ &\approx \frac{I_2(q)}{nv_n}\int_0^\infty \frac{f(x)}{x+\epsilon_n}\,dx + \int_0^\infty [(x\,v_n^2+\epsilon_n)f'(x)]^2\,dx \\ &\quad + o(v_n^2+\epsilon_n) + o((nv_n)^{-1}) \end{aligned} \tag{4.4}$$

The leading term of $MISE$ is defined as the *asymptotic MISE*:

$$AMISE[f_n^+] = \frac{I_2(q)}{nv_n}\int_0^\infty \frac{f(x)}{x+\epsilon_n}\,dx + \int_0^\infty [(x\,v_n^2+\epsilon_n)f'(x)]^2\,dx \tag{4.5}$$

To derive the optimal v_n and ϵ_n for estimating $f(x)$ by minimizing $AMISE[f_n^+]$, let us rewrite it as

$$A(v,\epsilon) := AMISE[f_n^+] = \frac{C_0}{nv}\int_0^\infty \frac{f(x)}{x+\epsilon}\,dx + C_1^2v^4 + 2C_2v^2\epsilon + C_3\epsilon^2 \tag{4.6}$$

Existence of unique minimizers. First, it is easy to see from Eq.(4.6) that $A(v,\epsilon)$ is a strictly convex function in $v>0$, $\epsilon>0$, as follows: $h_x(\epsilon)=1/(x+\epsilon)$ is convex for each $x>0$; hence $\int_0^\infty \frac{f(x)}{x+\epsilon}dx = \int_0^\infty h_x(\epsilon)f(x)dx$ is also obviously a convex, and decreasing, function. Since $1/v$ is also a decreasing convex function, it follows that $\frac{C_0}{nv}\int_0^\infty \frac{f(x)}{x+\epsilon}\,dx$ is convex. Further, since $(C_1^2v^4+2C_2v^2\epsilon+C_3\epsilon^2)$ is obviously convex, it follows that $A(v,\epsilon)$ is convex.

Now note that $A(v,\epsilon)\to\infty$ as $v\to 0$, $\epsilon\to 0$, as well as $v\to\infty$, $\epsilon\to\infty$. This fact, coupled with the convexity of $A(v,\epsilon)$, shows that $A(v,\epsilon)$ has unique global minimizers (v_n^*,ϵ_n^*).

Optimal order of AMISE. By the preceding arguments, the minimizers (v_n^*,ϵ_n^*) may be found by solving $(\partial/\partial v)A=0$, $(\partial/\partial\epsilon)A=0$. This leads to

$$\begin{aligned} C_1(nv^3)+C_2(nv\epsilon) &= \frac{C_0}{4v^2}\int_0^\infty \frac{f(x)}{x+\epsilon}dx, \\ C_2(nv^3)+C_3(nv\epsilon) &= \frac{C_0}{2}\int_0^\infty \frac{f(x)}{(x+\epsilon)^2}dx, \end{aligned} \tag{4.7}$$

and

$$(\frac{C_1}{\epsilon}+\frac{C_3}{v^2})\int_0^\infty \frac{f(x)}{x+\epsilon}dx = 2(C_1v^2+C_2)\int_0^\infty \frac{f(x)}{(x+\epsilon)^2}dx.$$

It follows that in the optimal solution (v_n^*,ϵ_n^*), $\epsilon_n^* = O((v_n^*)^2)$. To determine the order of $AMISE(v_n^*,\epsilon_n^*)$ consider the following two cases:

(i) $\int_0^\infty \frac{f(x)}{x}\,dx$ **exists.** For instance, consider $f(x) = 2xe^{-x^2}$. In this case $\int_0^\infty \frac{f(x)}{x+\epsilon}dx = O(1)$ as $\epsilon \longrightarrow 0$, hence from the top part of Eq.(4.7) and using $\epsilon_n^* = O((v_n^*)^2)$, we have

$$n(v_n^*)^5 = O(1),\ \ A(v_n^*, \epsilon_n^*) = O((nv_n^*)^{-1}),$$

which leads to the usual optimal rate in density estimation.

(ii) $\int_0^\infty \frac{f(x)}{x}\,dx$ **does not exist, i.e.,** $\int_0^\infty \frac{f(x)}{x}\,dx = \infty$**.** Consider, for instance, the standard Exponential density $f(x) = e^{-x}$. In this case, we need to make assumptions about the order of $\int_0^\infty \frac{f(x)}{x+\epsilon}\,dx$ as $\epsilon \longrightarrow 0$. In the case of standard Exponential it is $(-\ln \epsilon)$, i.e.,

$$\lim_{\epsilon\to 0} \frac{1}{-\ln\epsilon} \int_0^\infty \frac{\mathrm{e}^{-x}}{x+\epsilon}\,dx = 1$$

Thus in the case of standard Exponential we have

$$n(v_n^*)^5 = O(-2\log v_n^*),\ \ A(v_n^*, \epsilon_n^*) = O(-2(nv_n^*)^{-1}\log v_n^*).$$

This slightly suboptimal order is needed to take care of the jump at the boundary $x = 0$.

For a *data-driven* optimal choice of (v_n, ϵ_n) there is no problem in using the 'plug-in' (i.e., empirical) version of Eq.(4.5), as we shall see below.

Below we describe optimal choice of (v_n, ϵ_n) by minimizing *unbiased* and *biased cross-validation functions*, and present numerical results as well as graphs of estimated densities based on simulated data. The cross-validation methods are adapted from Scott (1992) and Wand and Jones (1995).

Unbiased cross-validation. Consider the *integrated squared error*

$$\begin{aligned} ISE(v_n, \epsilon_n) &= \int_0^\infty [f_n^+(x) - f(x)]^2\,dx \\ &= \int_0^\infty f_n^{+2}(x)dx - 2\int_0^\infty f_n^+(x)f(x)dx + \int_0^\infty f^2(x)dx \end{aligned}$$

Subtracting the constant term, and replacing the second term by its estimate we obtain the *unbiased cross-validation function*

$$UCV(v_n, \epsilon_n) = \int_0^\infty f_n^{+2}(x)dx - \frac{2}{n}\sum_{i=1}^n f_{n,i}^+(X_i), \tag{4.8}$$

where the second term is the *leave-one-out estimate* given by

$$f_{n,i}^+(X_i) = \frac{1}{(n-1)} \cdot \frac{1}{(X_i+\epsilon_n)^2} \sum_{j\neq i} X_j\, q_v\left(\frac{X_j}{X_i+\epsilon_n}\right) \tag{4.9}$$

Our numerical results show that this method performs poorly compared to biased cross-validation and introduces a lot of noise in the resulting density estimate.

Biased cross-validation. This technique is based on the idea of direct estimation of each term involving $f(\cdot)$ in the $AMISE$. Recalling the $AMISE$ from Eq.(4.5) and replacing $f(x)$ and $f'(x)$ by $f_n^+(x)$ and $f_n^{+'}(x)$, respectively, we have the *biased cross-validation function* given by

$$BCV(v_n, \epsilon_n) = \frac{I_2(q)}{n v_n} \int_0^\infty \frac{f_n^+(x)}{x + \epsilon_n}\, dx + \int_0^\infty [(x\, v_n^2 + \epsilon_n) f_n^{+'}(x)]^2\, dx \tag{4.10}$$

which is then minimized with respect to (v_n, ϵ_n). For our study, we take $q_v(\cdot)$ to be the Gamma $(\alpha = 1/v^2, \beta = v^2)$ density, so that $I_2(q) = 1/\sqrt{4\pi}$ (see Remark 3).

First, we take the underlying density to be *Weibull* with pdf:

$$f(x) = 2\, x\, e^{-x^2}, \quad x \geq 0. \tag{4.11}$$

Our computations showed that the optimal ϵ is very close to 0 (note that $\int_0^\infty \frac{f(x)}{x}\, dx$ exists). Hence we let $\epsilon = 0$ and find only the optimal v^2.

For each sample-size , optimal v^2 was obtained by minimizing, respectively, the UCV, BCV and the exact ISE — the latter for comparison. The minimization was done over a grid of v values using the software Mathematica. The following tables give the results:

Table 1: Unbiased Cross-Validation for Weibull distribution

Size	Minimum UCV	v^2	ISE (v^2)
100	-0.672500	0.09	0.010382
200	-0.636767	0.04	0.004924
500	-0.658883	0.04	0.005404

Table 2: Biased Cross-Validation for Weibull distribution

Size	Minimum BCV	v^2	ISE (v^2)
100	0.0628264	0.15	0.016479
200	0.0333457	0.10	0.006290
500	0.0175793	0.06	0.005138

Table 3: Integrated Squared Error for Weibull distribution

Size	Minimum ISE	v^2
100	0.009732	0.07
200	0.003964	0.06
500	0.005091	0.05

Second, we assume that the underlying density is Exponential with pdf

$$f(x) = e^{-x}, \quad x \geq 0. \tag{4.12}$$

Since $f(0) = 1$ and $\int_0^\infty \frac{f(x)}{x}\,dx$ is ∞, we need to calculate the optimal v_n and ϵ_n.

The following tables present the results:

Table 4: Unbiased Cross-Validation for Exponential distribution

Size	Minimum UCV	v^2	ϵ	ISE (v^2, ϵ)
100	-0.499494	0.275	0.035	0.007969
200	-0.486211	0.05	0.13	0.008351
500	-0.502629	0.11	0.03	0.002884

Table 5: Biased Cross-Validation for Exponential distribution

Size	Minimum BCV	v^2	ϵ	ISE (v^2, ϵ)
100	0.0660912	0.20	0.13	0.006063
200	0.0365576	0.13	0.12	0.008308
500	0.0241541	0.13	0.09	0.002782

Table 6: Integrated Squared Error for Exponential distribution

Size	Minimum ISE	v^2	ϵ
100	0.003136	0.12	0.10
200	0.005413	0.11	0.06
500	0.002530	0.14	0.02

Thus we observe that for the Weibull density unbiased or biased cross-validation do not make much of a difference, whereas for the Exponential, the former creates a noisy estimate near the boundary but the latter is very accurate. Also, as the sample size increases, the cross-validation-based choice of $(v_n,\ \epsilon_n)$ get closer to the ones that minimize the integrated squared error.

Acknowledgement: The authors thank Hongdan Zhu, graduate student at Concordia University, for his extremely thorough job of producing the cross-validation results and the density estimates in Section 4. Y. Chaubey and A. Sen would like to acknowledge the research support from NSERC of Canada.

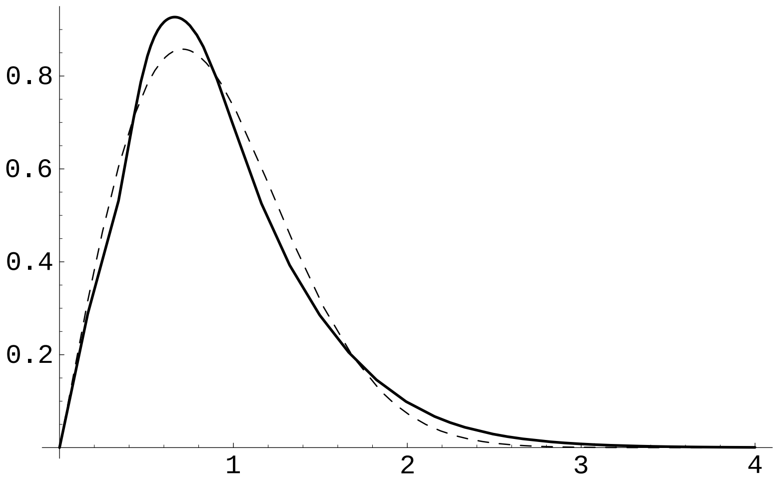

Figure 1: Weibull density (dashed line) and its estimate (solid line), $n = 500$, with unbiased cross-validation

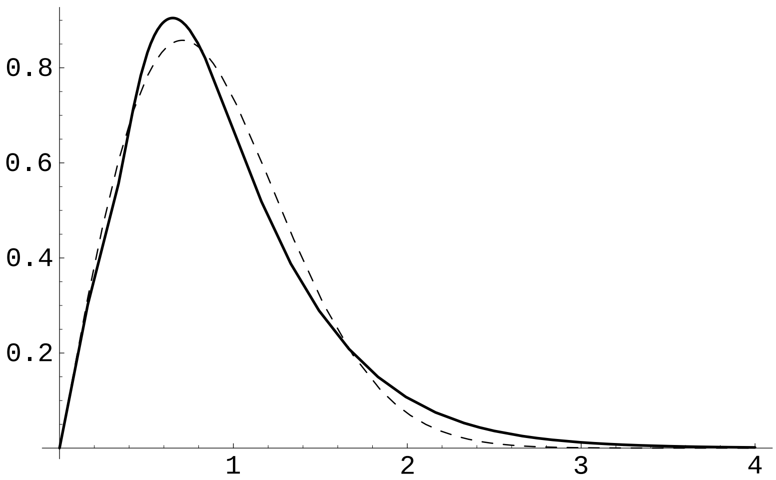

Figure 2: Weibull density (dashed line) and its estimate (solid line), $n = 500$, with biased cross-validation

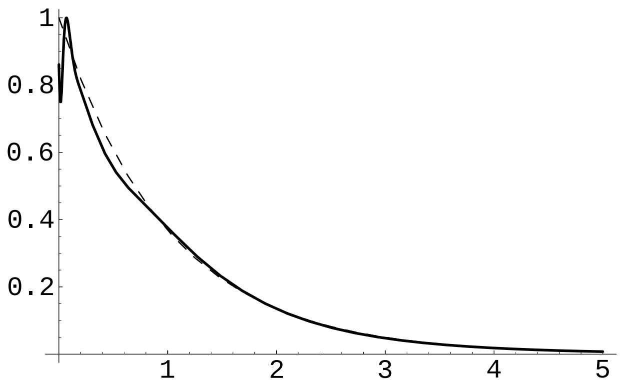

Figure 3: Exponential density (dashed line) and its estimate (solid line), $n = 500$, with unbiased cross-validation

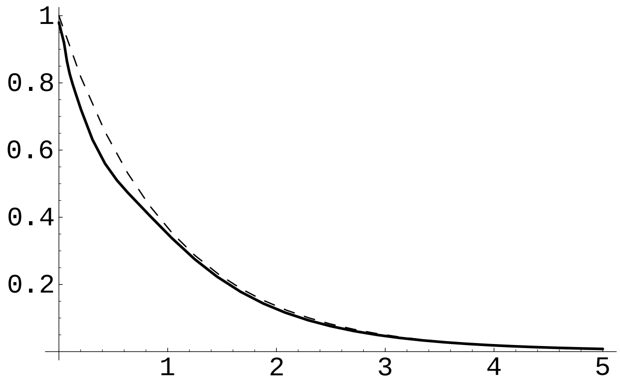

Figure 4: Exponential density (dashed line) and its estimate (solid line), $n = 500$, with biased cross-validation

APPENDIX: PROOFS

The proofs of Theorems 1 and 2 follow along the same lines as those of Theorems 3.1 and 3.2 respectively of Chaubey and Sen (1996) and therefore are omitted.

Proof of Theorem 3. Using the representation in Eq.(3.3), we have

$$\begin{aligned} f_n^+(x) - f(x) &= \int_0^\infty (F_n(t) - F(t)) \left[\frac{d}{dx} g_{x+\epsilon_n,n}(t)\right] dt \\ &\quad + \int_0^\infty F(t) \left[\frac{d}{dx} g_{x+\epsilon_n,n}(t)\right] dt - f(x). \end{aligned}$$

Now look at the second term in the above expression. We can write

$$\int_0^\infty F(t) \left(\frac{d}{dx} g_{x+\epsilon_n,n}(t)\right) dt = E(f_n^+(x)).$$

Hence

$$\begin{aligned} |f_n^+(x) - f(x)| &\le \sup_t |F_n(t) - F(t)| \int_0^\infty | \left[\frac{d}{dx} g_{x+\epsilon_n,n}(t)\right] | dt \\ &\quad + \int_0^\infty |f(t(x+\epsilon_n)) - f(x)| t q_v(t) dt. \end{aligned}$$

The first term in the above inequality converges to zero *a.s.* under the condition given in the theorem. The second term also converges to zero as can be seen as follows. For any $M > 0$,

$$\sup_{x\ge 0} \int_0^\infty |f(t(x+\epsilon_n)) - f(x)| t q_v(t) dt = \max\{\sup_{0\le x\le M} \int_0^\infty |\cdots|, \sup_{x>M} \int_0^\infty |\cdots|\} = \max\{a_M, b_M\}, \text{ say.}$$

Now take any $\varepsilon > 0$. Then we can get $M > 0$ such that

$$\begin{aligned} b_M &\le \sup_{x>M} \int_0^\infty f(t) \frac{(t/(x+\epsilon_n)) q_v(t/(x+\epsilon_n))}{x+\epsilon_n} dt + \sup_{x>M} f(x) \int_0^\infty t q_v(t) dt \\ &\le \varepsilon/2 + \varepsilon/2, \end{aligned}$$

using a dominated convergence argument in the *first* term (by Assumption C) and the fact that $f(t) \to 0$ as $t \to \infty$ in the *second*. Further, for this M we have by Assumption D and Cauchy-Schwartz inequality,

$$\begin{aligned} a_M &\le M \int_0^\infty |t-1| t q_v(t) dt + \epsilon_n \int_0^\infty t^2 q_v(t) dt \\ &\le M \sqrt{\int_0^\infty (t-1)^2 q_v(t) dt} \sqrt{\int_0^\infty t^2 q_v(t) dt} + O(\epsilon_n) \\ &= M.O(v) + O(\epsilon_n) \to 0 \text{ as } v \to 0,\ \epsilon_n \to 0. \end{aligned}$$

Thus

$$\sup_{x\geq 0}\int_0^\infty |f(t(x+\epsilon_n))-f(x)|tq_v(t)dt \to 0 \text{ as } v\to 0.$$

□

Proof of Theorem 4.

(a) Fix $x>0$. Write $\sqrt{nv}(f_n^+(x)-f(x)) = \sqrt{nv}(f_n^+(x)-E(f_n^+(x)))+\sqrt{nv}(E(f_n^+(x))-f(x))$. Note that $\sqrt{nv}(f_n^+(x)-E(f_n^+(x))) = \sum_{i=1}^n Z_{in}$, where

$$\begin{aligned} Z_{in} &= \sqrt{\frac{v}{n}}\left[\frac{X_i}{(x+\epsilon_n)^2}q_v(\frac{X_i}{x+\epsilon_n}) - E\left(\frac{X_1}{(x+\epsilon_n)^2}q_v(\frac{X_1}{x+\epsilon_n})\right)\right] \\ &= \sqrt{\frac{v}{n}}(Y_{in}-E(Y_{in})),\ 1\leq i\leq n,\ n\geq 1, \end{aligned}$$

Y_{in} being defined as

$$Y_{in} = \frac{X_i}{(x+\epsilon_n)^2}q_v(\frac{X_i}{x+\epsilon_n}).$$

We may verify the *Lyapounov condition* (see, for instance, Chung (1974), Theorem 7.1.2):

$$\frac{\sum_{i=1}^n E|Z_{in}|^3}{(\sum_{i=1}^n EZ_{in}^2)^{3/2}} \to 0,\ \text{as } n\to\infty,$$

which will imply that

$$\frac{\sum_{i=1}^n Z_{in}}{(\sum_{i=1}^n EZ_{in}^2)^{1/2}} = \frac{\sqrt{nv}(f_n^+(x)-E(f_n^+(x)))}{(v.\mathrm{var}(Y_{1n}))^{1/2}} \to \text{Normal } (0,1)$$

in distribution, as $n\to\infty$.

Since $Y_{in},\ 1\leq i\leq n$, are non-negative, we have,

$$\begin{aligned} \frac{\sum_{i=1}^n E|Z_{in}|^3}{(\sum_{i=1}^n EZ_{in}^2)^{3/2}} &\leq \frac{E[Y_{1n}+E(Y_{1n})]^3}{\sqrt{n}[\mathrm{var}(Y_{1n})]^{3/2}} \\ &= \frac{E(Y_{1n}^3)+3E(Y_{1n}^2).E(Y_{1n})+4(E(Y_{1n}))^3}{\sqrt{n}[E(Y_{1n}^2)-(E(Y_{1n}))^2]^{3/2}} \end{aligned} \qquad (A.1)$$

Next, one can show that

$$E(Y_{1n}^3) = O(v^{-2}),\ \text{as } v\to 0,\ \epsilon_n\to 0, \qquad (A.2)$$

$$vE(Y_{1n}^2)\to I_2(q)\frac{f(x)}{x}, \qquad (A.3)$$

so that

$$E(Y_{1n}^2) = O(v^{-1}) \qquad (A.4)$$

and

$$E(Y_{1n}) = O(1) \qquad (A.5)$$

which together with the last expression in Eq.(A.1), imply that

$$\frac{\sum_{i=1}^{n} E|Z_{in}|^3}{(\sum_{i=1}^{n} EZ_{in}^2)^{3/2}} = O(\frac{1}{\sqrt{nv}}),$$

so that the Lyapounov condition holds by Condition (H).

Now from Eq. (A.3) and using (F),

$$\begin{aligned}
\sqrt{nv}|E(f_n^+(x)) - f(x)| &= \sqrt{nv}|E(Y_{1n}) - f(x))| \\
&\leq \sqrt{nv}\int_0^\infty [(\text{const.})(x+\epsilon_n)t|t-1| + |t-1|f(x+\epsilon_n)]q_v(t)dt \\
&\quad +\sqrt{nv}|f(x+\epsilon_n) - f(x)| \\
&\leq \sqrt{nv}(\text{const.})(x+\epsilon_n)\sqrt{\left(\int_0^\infty t^2 q_v(t)dt\right)}\sqrt{\left(\int_0^\infty (t-1)^2 q_v(t)dt\right)} \\
&\quad +\sqrt{nv}f(x+\epsilon_n)\sqrt{\int_0^\infty (t-1)^2 q_v(t)dt} + \text{ const.}\sqrt{nv}\epsilon_n, \\
&\quad \text{by Cauchy-Schwartz inequality,} \\
&= \sqrt{nv}(O(v) + O(\epsilon_n)) = O(\sqrt{nv^3}) + O(\sqrt{nv\epsilon_n^2})
\end{aligned}$$

This together with Eq. (A.3) and (A.4) (using Condition (H)) completes the proof of Theorem 4, part (a).

(b) For $x = 0$, we have

$$\begin{aligned}
\sqrt{nv\epsilon_n}(f_n^+(0) - E(f_n^+(0))) &= \sum_{i=1}^{n}\sqrt{\frac{v\epsilon_n}{n}}\left[\frac{X_i}{\epsilon_n^2}q_v(\frac{X_i}{\epsilon_n}) - E\left(\frac{X_1}{\epsilon_n^2}q_v(\frac{X_1}{\epsilon_n})\right)\right] \\
&= \sum_{i=1}^{n}\sqrt{\frac{v\epsilon_n}{n}}(Y'_{in} - E(Y'_{1n})) \text{ (say) } = \sum_{i=1}^{n} Z'_{in} \text{ (say)},
\end{aligned}$$

and

$$\begin{aligned}
\sqrt{nv\epsilon_n}(E(f_n^+(0)) - f(0)) &= \sqrt{nv\epsilon_n}\int_0^\infty [t(f(\epsilon_n t) - f(\epsilon_n)) + (t-1)f(\epsilon_n)]q_v(t)dt \\
&\quad +\sqrt{nv\epsilon_n}(f(\epsilon_n) - f(0)) \\
&= O(\sqrt{nv^3\epsilon_n^3} + \sqrt{nv^3\epsilon_n} + \sqrt{nv\epsilon_n^3}) \to 0, \text{ as } n \to \infty,
\end{aligned}$$

by Condition H. Now exactly as in Part (a) (details omitted), we establish that as $n \to \infty$,

(i) $\sum_{i=1}^{n} E|Z'_{in}|^3 / (\sum_{i=1}^{n} E(Z'_{in})^2)^{3/2} = O((nv\epsilon_n)^{-1/2}) \to 0$, so that

$$\frac{\sum_{i=1}^{n} Z'_{in}}{(\sum_{i=1}^{n} E(Z'_{in})^2)^{1/2}} = \frac{\sqrt{nv\epsilon_n}(f_n^+(x) - E(f_n^+(x)))}{(v\epsilon_n \text{var}(Y'_{1n}))^{1/2}} \to \text{Normal } (0,1)$$

in distribution,

(ii) $v\epsilon_n \text{var}(Y'_{1n}) \to I_2(q) f(0)$.

The proof of Theorem 4, Part (b) follows. □

REFERENCES

Bagai, I. and Prakasa Rao, B.L.S. (1996). Kernel Type Density Estimates for Positive Valued Random Variables. *Sankhya*, **A57** 56-67.

Bouezmarni, Taoufik; Scaillet, Olivier (2005). Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory*, **21**, 390-412.

Chaubey, Y. P., and Sen, P. K. (1996). On smooth estimation of survival and density functions. *Statist. Decisions*, **14**, 1-22.

Chen, S. X. (2000). Probability density function estimation using Gamma kernels. *An.n. Inst. Statist. Meth.* **52,** 471-480.

Chung, K. L. (1974). *A course in probability theory (2nd ed.).* Academic Press, New York and London.

Devroye, L. (1989) *A Course in Density Estimation.* Birkhäuser, Boston.

Eubank, R.L. (1988) *Spline Smoothing and Nonparametric Regression,* Marcel Dekker, New York.

Feller, W. (1965) *An Introduction to Probability Theory and its Applications, Vol. II.* John Wiley and Sons, New York.

Hille, E. (1948)*Functional Analysis and Semigroups*, Am. Math. Colloq. Pub 31, New York.

Marron, J. S., Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *J. Roy. Statist. Soc.* Ser. B **56,** 653-671.

Parzen, E. (1962) On estimation of probability density and mode. *Ann. Math Statist.*, **33** 1065-1070.

Rosenblatt, M. (1956) Remarks on some nonparametric estimates of density functions. *Ann. Math Statist.*, **27** 832-837.

Ruppert, D. and Wand, M. P. (1992). Correcting for kurtosis in density estimation. *Australian Journal of Statistics,* **34,** 1929.

Scaillet, O. (2004). Density estimation using inverse Gaussian and reciprocal inverse Gaussian kernels. *Journal of Nonparametric Statistics,* **16,** 217-226.

Scott, D. W. (1992). *Multivariate Density Estimation. Theory, Practice and Visualizations.* New York: John Wiley and Sons.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

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

Wand, M.P., Marron, J.S. and Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association,* **86,** 343-361.

List of Recent Technical Reports

46. Yogendra P. Chaubey and Debaraj Sen, *An investigation into properties of an estimator of mean of an inverse Gaussian population*, September, 2004.

47. Steven N. Evans and Xiaowen Zhou, *Balls-in-boxes duality for coalescing random walks and coalescing Brownian motions*, September, 2004.

48. Qihe Tang, *Asymptotic ruin probabilities of the renewal model with constant interest force and regular variation*, November, 2004.

49. Xiaowen Zhou, *On a classical risk model with a constant dividend barrier*, November, 2004.

50. K. Balasubramanian and M.I. Beg, *Three isomorphic vector spaces–II*, December, 2004.

51. Michael A. Kouritzin and Wei Sun, *Rates for branching particle approximations of continuous-discrete filters*, December, 2004.

52. Rob Kaas and Qihe Tang, *Introducing a dependence structure to the occurrences in studying precise large deviations for the total claim amount* , December, 2004.

53. Qihe Tang and Gurami Tsitsiashvili, *Finite and infinite time ruin probabilities in the presence of stochastic returns on investments*, December, 2004.

54. Alexander Melnikov and Victoria Skornyakova, *Efficient hedging methodology applied to equity-linked life insurance*, February, 2005.

55. Qihe Tang, *The finite time ruin probability of the compound Poisson model with constant interest force*, June, 2005.

56. Marc J. Goovaerts, Rob Kaas, Roger J.A. Laeven, Qihe Tang and Raluca Vernic, *The tail probability of discounted sums of Pareto-like losses in insurance*, August, 2005.

57. Yogendra P. Chaubey and Haipeng Xu, *Smooth estimation of survival functions under mean residual life ordering*, August, 2005.

58. Xiaowen Zhou, *Stepping-stone model with circular Brownian migration*, August, 2005.

59. José Garrido and Manuel Morales, *On the expected discounted penalty function for Lévy risk processes*, November, 2005.

60. Z. Hu, Z. Ma and W. Sun, *Extensions of Lévy-Khintchine formula and Beurling-Deny formula in semi-dirichlet forms setting*, February, 2006.

61. Z. Hu, Z. Ma and W. Sun, *Formulae of Beurling-Deny and LeJan for non-symmetric dirichlet forms*, February, 2006.

62. Z. Hu and W. Sun, *A note on exponential stability of the nonlinear filter for denumerable Markov chains*, February, 2006.

63. H. Brito-Santana, R. Rodríguez-Ramos, R. Guinovart-Díaz, J. Bravo-Castillero and F. J. Sabina, *Variational bounds for multiphase transversely isotropic composites*, August, 2006.

64. José Garrido and Jun Zhou, *Credibility theory for generalized linear and mixed models*, December, 2006.

65. Daniel Dufresne, José Garrido and Manuel Morales, *Fourier inversion formulas in option pricing and insurance*, December, 2006.

66. Xiaowen Zhou, *A superprocess involving both branching and coalescing* , December, 2006.

67. Yogendra P. Chaubey, Arusharka Sen and Pranab K. Sen, *A new smooth density estimator for non-negative random variables*, January 2007.

Copies of technical reports can be requested from:

Prof. Xiaowen Zhou
Department of Mathematics and Statistics
Concordia University
1455 de Maisonneuve Blvd West
Montréal (QC) H3G 1M8 CANADA