International Journal of Research and Reviews in Applied Sciences
ISSN: 2076-734X, EISSN: 2076-7366
Volume 1, Issue 2(November 2009)

NONPARAMETRIC DENSITY ESTIMATION WITH RESPECT TO THE LINEX LOSS FUNCTION

R. HASHEMI, S. REZAEI AND L. AMIRI
Department of Statistics, Faculty of Science,
Razi University, 67149, Kermanshah, Iran.

ABSTRACT

In this paper the leave-one-out and leave-p-out cross validation risk estimator for histogram and kernel density estimators, with respect to the Linex loss function are presented, and compared to the corresponding results under square loss function. The exact expectation and variance of each estimator is presented. Finally using the relation of density estimators and the regression method, a regression based density estimator under the Linex loss function is proposed.

KEYWORDS: *Density estimation; Linex loss; Cross-validation; Kernel estimator.*

1. INTRODUCTION:

The oldest and the most frequently used nonparametric density estimator is the histogram estimator. Histograms are in two general kinds: the regular and non regular histograms. In order to construct a histogram it is sufficient to partition the covering domain of data to distinct subintervals. If the length of all subintervals are equal, the histogram is called regular. The length of each subinterval is called smoothing parameter. The height of each rectangular is proportional to the number of data in each subinterval, so to construct a histogram, we need to have two factors: the length and the endpoints of each subinterval. Assume there exist D equal subintervals with the length h, so the histogram estimator can be written as follows:

$$\hat{f}_n(x) = \sum_{k=1}^{D} \frac{\hat{p}_k}{h} I(x \in B_k) \qquad (1)$$

where B_k is the k^{th} subintervals, n_k is the number of data which fall into the subinterval B_k and $\hat{P}_k = \frac{n_k}{n}$.

1.1 Characteristics of histogram estimator

Histogram estimators are not smooth, and they are essentially sensitive to the choice of smoothing parameters and highly dependent on the endpoints. Since a histogram estimator is an approximation of the density function so the density function is also non smooth and biased because of its dependence on the endpoints.

1.2 Kernel estimator

An estimator of density function which is more smooth and less dependent on endpoints than the histogram estimator, is the kernel estimator which has been presented by Rosenblat(1956) and then by Parlzen(1962). Its general form is as follows:

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K(\frac{x - X_i}{h}), \qquad \forall x \in \Re \qquad (2)$$

where k(.) is called kernel function or in other words a smooth function(that is a function whose high order derivatives are continuous), and h is the smoothing parameter. It can be seen that this estimator depends only on the smoothing parameter. The kernel function is centered on the observed data, hence the estimator is less sensitive to the choice of the endpoints of subintervals. Characteristics of kernel estimator can be briefly mentioned:

- $K(x) \geq 0 \qquad \forall x \in \Re$
- $\int K(x)dx = 1$
- $\int xK(x)dx = 0$
- $\sigma_k^2 = \int x^2 K(x)dx < \infty$

2 METHODS OF ESTIMATION:

As mentioned above, the histogram and Kernel estimators are sensitive to the choice of smoothing parameter h. Therefore, it is important to find the value of h which minimizes the risk of estimators. Previous research was based on squared error loss function in which the risk function was defined as follows:

$$R_x = R(f(x), \hat{f}_n(x)) = E(\hat{f}_n(x) - f(x))^2$$

Integrated risk of the estimators is as follows:

$$R = \int R_x dx = E_f[\int \hat{f}_n^2(x)dx - 2\int f(x)\hat{f}_n(x)dx + \int f^2(x)dx]$$

Minimizing the risk of estimators in respect to h is equivalent to minimizing

$$L(\hat{f}_n) = E_f[\int \hat{f}_n^2(x)dx - 2\int f(x)\hat{f}_n(x)dx]$$

Since the L(ˆ fn) is a function of the unknown f(x), cross validation methods are used to estimate it, and then the optimum h is obtained by minimizing the result. Cross-validation can be used in two different ways as follows:

- Leave-One-Out(L.O.O.)
- Leave-P-Out(L.P.O.)

Before presenting the main results, we need to review these two methods of cross-validation.

2.1 L.O.O. Risk Estimator

In the L.O.O. method, the data set is divided into two parts. Part one includes the first n − 1 observations, and part two includes only the last observation. The first part has been used to compute the estimation of the parameter, and the second part has been used to verify the accuracy of the estimation. This procedure has been conducted in n steps, because there is n choices for the first part. Finally, the average of these estimations will be used as L.O.O. cross-validation of the risk estimator.

$$\widehat{L_1}(\hat{f}_n) = \sum_{i=1}^{n} \int (\hat{f}^{(i)}(x))^2 dx - \frac{2}{n} \sum_{i=1}^{n} \hat{f}^{(i)}(x) \qquad (3)$$

2.2 L.P.O. Risk Estimator

As in the L.O.O. method, the data set is divided into two parts, but the first part includes the n−p observations, and the second part includes p remaining observations. The first part has been used to compute the estimation of the parameter, and the second part has been used to assess the accuracy of the estimation. In the L.P.O. cross-validation, this procedure has been conducted in the $\binom{n}{p}$ steps. Finally, the average of $\binom{n}{p}$ estimation will be used as L.P.O. cross-validation of the risk estimator.

$$\widehat{L_p}(\hat{f}_n) = \binom{n}{p}^{-1} \sum_{e \in \varepsilon_p} \{ \int (\hat{f}^{\bar{e}})^2 dx - \frac{2}{p} \sum_{i \in e} \hat{f}^{\bar{e}}(X_i) \} \qquad (4)$$

where $X_1, \cdots, X_n$ are independent and identically distributed random variables, and for each $P \in \{1, \cdots, n-1\}$, εp denotes the set of all p subsets of $\{1, \cdots, n\}$, and

$$\forall e \in \varepsilon_p, \bar{e}$$

denotes a subset of $\{1, \cdots, n\}$ which is not in e and

$$X^{\bar{e}} = \{X_i \; i \in \bar{e}\}, \text{ and also } \hat{f}^{(i)}$$

denotes the estimator whose computed from

$X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n$, as such as $\hat{f}^{\bar{e}}$ denotes the estimator computed from $X^{\bar{e}}$. It must be noted that in this procedure the larger the P the more bias and the less variation in L.P.O. risk estimator.

In this paper we replace the squared loss function with the linear exponential loss which is a nonsymmetric loss function and is defined as follows

$L_y = \exp(\alpha(y - y_0)) - \alpha(y - y_0) - 1$

where y_0 is the goal function.

Expanding this loss function on y0 we have

$$L_y = \exp(\alpha(y - y_0)) - \alpha(y - y_0) - 1$$

$$= \sum_{i=2}^{\infty} \frac{\alpha^i (y - y_0)^i}{i!} \simeq \frac{\alpha^2}{2}(y - y_0)^2$$

Hence it can be deduced that for small values of α the Linex loss is equivalent to the squared loss. Under the Linex loss risk of the estimator $\hat{f}$ at point x is defined as follows

$$R_x = E_f[\exp(\alpha(\hat{f}_n(x) - f(x)) - \alpha(\hat{f}_n(x) - f(x)) - 1]$$

and in the same way integrated risk estimator at point x is as follows

$$R = \int R_x dx = E_f[\int (\exp(\alpha(\hat{f}_n(x) - f(x)) - \alpha(\hat{f}_n(x) - f(x)) - 1) dx]$$

Therefor, to find the optimal smoothing parameter, this integrated risk must be minimized with respect to h, but this minimization is equivalent to minimizing the following equation

$$L(h) = \int \exp(\alpha(\hat{f}_n(x) - f(x))dx - \alpha \int \hat{f}_n(x)dx \qquad (5)$$

because f(x) does not depend on h. But produced h as such as l(h) by the above mentioned way will be a function of unknown f(x), so we have to minimize an estimation of l(h).

L.O.O risk estimator under linex loss function

$$\widehat{L_1}(h) = \int \exp\{\frac{\alpha}{n}\sum_{i=1}^{n} \hat{f}^{(i)}(X_i) - \alpha f(x)\}dx$$

$$-\frac{\alpha}{n}\sum_{i=1}^{n}\int \hat{f}^{(i)}(x)dx$$

We review here a familiar theorem of mathematical analysis:

Theorem

$$\int_0^1 e^{g(x)}dx \geq e^{\int_0^1 g(x)dx}$$

using the above theorem we can :

$$\widehat{L_1}(h) \geq \exp\{-\alpha + \frac{\alpha}{n}\sum_{i=1}^{n} \hat{f}^{(i)}(X_i)\} - \frac{\alpha}{n}\sum_{i=1}^{n}\int \hat{f}^{(i)}(x)dx$$

Now let

$$\widehat{L_1^*}(h) = \exp\{\alpha(\frac{1}{n}\sum_{i=1}^{n} \hat{f}^{(i)}(X_i) - 1)\} - \frac{\alpha}{n}\sum_{i=1}^{n}\int \hat{f}^{(i)}(x)dx \qquad (6)$$

So minimizing $\widehat{L_1}(h)$ with respect to the h is equivalent to minimization the $\widehat{L_1^*}(h),$ the produced estimator in this way is called L.O.O. estimator.

L.P.O. risk estimator under linex loss function

Set $X_1, ..., X_n$ as i.i.d. random variables. For $p \in \{1, ..., n\}$, let ε_p be the set of all possible p-subsets of {1, ..., n}. For any

$e \in \varepsilon_p, \quad \bar{e} = \{1, ..., n\} - e \quad \text{and } X^{\bar{e}} = \{X_i / i \in \bar{e}\}.$

With respect to theorem 1 and definition 3 we can write:

$$\widehat{L_p}(h) = \int \exp\{\frac{\alpha}{p\binom{n}{p}} \sum_{e\in\varepsilon_p}\sum_{i\in e} \hat{f}^{\bar{e}}(X_i)\} \exp\{-\alpha f(x)\}dx - \frac{\alpha}{\binom{n}{p}} \sum_{e\in\varepsilon_p} \int \hat{f}^{\bar{e}}(x)dx$$

$$\geq \exp\{\frac{\alpha}{p\binom{n}{p}} \sum_{e\in\varepsilon_p}\sum_{i\in e} \hat{f}^{\bar{e}}(X_i) - \alpha\} - \frac{\alpha}{\binom{n}{p}} \sum_{e\in\varepsilon_p} \int \hat{f}^{\bar{e}}(x)dx$$

In this equation the average of $\binom{n}{p}$ estimators of each steps is used to estimate the $f_n(x)$, so using one more time, the above theorem we can write

$$\widehat{L_p^*}(h) = \exp\{\alpha(\frac{1}{p\binom{n}{p}} \sum_{e\in\varepsilon_p}\sum_{i\in e} \hat{f}^{\bar{e}}(X_i) - 1)\} - \frac{\alpha}{\binom{n}{p}} \sum_{e\in\varepsilon_p} \int \hat{f}^{\bar{e}}(x)dx \quad (7)$$

Since

$\widehat{L_p}(h) \geq \widehat{L_p^*}(h),$

we use the minimized

$\widehat{L_p^*}(h)$

with respect to the h as L.P.O. risk estimator.

3. COMPARATIVE STUDY:

Since $e^{-\alpha} > \alpha$ for each α, hence minimization of

$\widehat{L_1^*}(h)$ is equivalent to minimization of the following relation

$$\widehat{L_1^{**}}(h) = -\alpha \exp\{\frac{\alpha}{p\binom{n}{p}} \sum_{e\in\varepsilon_p}\sum_{i\in e} \hat{f}^{\bar{e}}(X_i)\} - \frac{\alpha}{\binom{n}{p}} \sum_{e\in\varepsilon_p} \int \hat{f}^{\bar{e}}(x)dx \leq \widehat{L_1^*}(h) \quad (8)$$

So we compare

$\widehat{L_1}(h)$ with $\widehat{L_1^{**}}(h)$ let $(\hat{f}^{(i)}(x))^2 \leq \hat{f}^{(i)}(x)$ and $\alpha > 0$

then:

$$\frac{1}{n}\sum_{i=1}^{n}\int(\hat{f}^{(i)}(x))^2dx > \frac{-\alpha}{n}\sum_{i=1}^{n}\int\hat{f}^{(i)}(x)dx \quad (9)$$

Now, let $\alpha > 2$:

$$\frac{-2}{n}\sum_{i=1}^{n}\hat{f}^{(i)}(X_i) > -\alpha\exp\{\frac{\alpha}{n}\sum_{i=1}^{n}\hat{f}^{(i)}(X_i)\} \quad (10)$$

Hence, if $\alpha > 2$ and $(\hat{f}^{(i)}(x))^2 \leq \hat{f}^{(i)}(x)$ we can write

$$\frac{1}{n}\sum_{i=1}^{n}\int(\hat{f}^{(i)}(x))^2dx - \frac{2}{n}\sum_{i=1}^{n}\hat{f}^{(i)}(X_i) > \frac{-\alpha}{n}\sum_{i=1}^{n}\int\hat{f}^{(i)}(x)dx - \alpha\exp\{\frac{\alpha}{n}\sum_{i=1}^{n}\hat{f}^{(i)}(X_i)\}$$

in the other hand:

$$\widehat{L_1^{**}}(h) < \widehat{L_1}(h) \quad (11)$$

4. COMPUTATION OF DENSITY ESTIMATORS:

4.1 Risk Estimator for Histograms

$$\widehat{L_1}(h) = \frac{2}{(n-1)h} - \frac{n+1}{(n-1)h}\sum_{k=1}^{D}(\frac{n_k}{n})^2$$

$$\widehat{L_p}(h) = \frac{2n-p}{(n-1)(n-p)}\sum_{k=1}^{D}\frac{n_k}{h_k} - \frac{n(n-p+1)}{(n-1)(n-p)}\sum_{k=1}^{D}\frac{1}{h_k}(\frac{n_k}{n})^2$$

Theorem 2(L.O.O risk estimator for histograms):
in the case of regular D-piece histogram (h = 1/D),

$$\widehat{L_1^{*}}(h) = \exp\{-\alpha(1+\frac{1}{(n-1)h}) + \frac{\alpha}{n(n-1)h}\sum_{k=1}^{D}n_k^2\} - \alpha \quad (12)$$

proof: write

$$\hat{p_k}^{(i)} = \frac{|\{j\neq i/X_j\in B_k\}|}{n-1} \text{ and } \hat{f}^{(i)}(x) = \textstyle\sum_{k=1}^{D}\frac{\hat{p}_k^{(i)}}{h}I(x \in B_k)$$

Hence,

$$\begin{aligned}\sum_{i=1}^{n}\hat{f}^{(i)}(X_i) &= \sum_{i=1}^{n}\sum_{k=1}^{D}\frac{1}{h}\hat{p}_k^{(i)}I(X_i\in B_k)\\ &= \sum_{k=1}^{D}\frac{n_k^2}{(n-1)h} - \frac{n}{(n-1)h}\end{aligned}$$

$$
\begin{aligned}
\sum_{i=1}^{n} \int \hat{f}^{(i)}(x)dx &= \sum_{i=1}^{n} \int \sum_{k=1}^{D} \frac{\hat{p}_k^{(i)}}{h} I(x \in B_k)dx \\
&= \sum_{i=1}^{n} \sum_{k=1}^{D} \frac{\hat{p}_k^{(i)}}{h} \int I(x \in B_k)dx \\
&= n
\end{aligned}
$$

Then:

$$\widehat{L_1^*}(h) = \exp\{-\alpha(1 + \frac{1}{(n-1)h}) + \frac{\alpha}{n(n-1)h} \sum_{k=1}^{D} n_k^2\} - \alpha.$$

Lemma 1.

Set a D-partition

(B_k), $e \in \varepsilon_p$ and $X^e = \{X_i / i \in \bar{e}\}$.

For any

$k \in \{1, ..., D\}$, define $n_k^e = |\{i / X_i \in X^e \cap B_k\}|$ and $n_k^{\bar{e}} = n_k - n_k^e$.

Then for $p \in \{1, ..., n-1\}$,

$$\forall k \in \{1, ..., D\}, \qquad \sum_{e \in \varepsilon_p} n_k^e n_k^{\bar{e}} = n_k(n_k - 1) \begin{pmatrix} n-2 \\ p-1 \end{pmatrix},$$

$$\sum_{e \in \varepsilon_p} (n_k^{\bar{e}})^2 = n_k \begin{pmatrix} n-1 \\ p \end{pmatrix} + n_k(n_k - 1) \begin{pmatrix} n-2 \\ p \end{pmatrix}$$

Theorem 3 (L.P.O. risk estimator for histograms).

$$\widehat{L_p^*}(h) = \exp\{\alpha(\frac{1}{n(n-1)} \sum_{k=1}^{D} \frac{n_k}{h_k} + 1) + \frac{\alpha}{n(n-1)} \sum_{k=1}^{D} \frac{n_k^2}{h_k}\} - \alpha \qquad (13)$$

Proof:

$$\begin{aligned}\sum_{e\in\varepsilon_p}\sum_{i\in e}\hat{f}^{\bar{e}}(X_i) &= \sum_{e\in\varepsilon_p}\sum_{i\in e}\sum_{k=1}^{D}\frac{n_k^{\bar{e}}}{(n-p)h_k}I(X_i\in B_k)\\ &= \sum_{k=1}^{D}\sum_{e\in\varepsilon_p}\frac{n_k^{\bar{e}}n_k^{e}}{(n-p)h_k}\\ &= \frac{\begin{pmatrix} n-2\\ p-1\end{pmatrix}}{(n-p)}\{\sum_{k=1}^{D}\frac{n_k^2}{h_k}-\sum_{k=1}^{D}\frac{n_k}{h_k}\}\end{aligned}$$

$$\sum_{e\in\varepsilon_p}\int\hat{f}^{\bar{e}}(x)dx = \sum_{k=1}^{D}\sum_{e\in\varepsilon_p}\frac{n_k^{\bar{e}}}{(n-p)} = \frac{n\begin{pmatrix} n-1\\ p\end{pmatrix}}{(n-p)}$$

Then,

$$\widehat{L_p^*}(h) = \exp\{\alpha(\frac{1}{n(n-1)}\sum_{k=1}^{D}\frac{n_k}{h_k}+1)+\frac{\alpha}{n(n-1)}\sum_{k=1}^{D}\frac{n_k^2}{h_k}\}-\alpha$$

Lemma 2(Exact expectation and variance expression).

$$E[\widehat{L_1^*}(h)] = e^{-\alpha}\prod_{k=1}^{D}(p_k^2\quad e^{\frac{\alpha}{n(n-1)h}}+(1-p_k^2))^{n(n-1)}-\alpha \qquad (14)$$

$$V_p(\widehat{L_1^*}(h)) = e^{-2\alpha}\prod_{k=1}^{D}(p_k^2\quad e^{\frac{2\alpha}{n(n-1)h}}+(1-p_k^2))^{n(n-1)}$$

$$-e^{-2\alpha}\prod_{k=1}^{D}(p_k^2\quad e^{\frac{\alpha}{n(n-1)h}}+(1-p_k^2))^{2n(n-1)} \qquad (15)$$

Proof:
let $p(X_i \in B_k) = p_k$ and $X_1, ..., X_n$ be i.i.d.,Hence

$$E[\widehat{L_1^*}(h)] = E[\exp\{-\alpha(1+\frac{1}{(n-1)h}) + \frac{\alpha}{n(n-1)h}\sum_{k=1}^{D} n_k^2\} - \alpha]$$

$$= e^{-\alpha} E[\exp\{\frac{\alpha}{n(n-1)h})\sum_{k=1}^{D}(n_k^2 - n_k)\}] - \alpha$$

$$= e^{-\alpha}\prod_{k=1}^{D} E[\exp\{\frac{\alpha}{n(n-1)h})(n_k^2 - n_k)\}] - \alpha$$

$$= e^{-\alpha}\prod_{k=1}^{D}(p_k^2 \quad e^{\frac{\alpha}{n(n-1)h}} + (1-p_k^2))^{n(n-1)} - \alpha$$

Now consider the variance,

$$V_p(\widehat{L_1^*}(h)) = E[\widehat{L_1^*}(h)]^2 - E^2[\widehat{L_1^*}(h)]$$

$$= E[\exp\{-2\alpha + \frac{2\alpha}{n(n-1)h}\sum_{k=1}^{D}(n_k^2 - n_k)\}]$$
$$-E^2[\exp\{-\alpha + \frac{\alpha}{n(n-1)h}\sum_{k=1}^{D}(n_k^2 - n_k)\}]$$

$$= e^{-2\alpha}\prod_{k=1}^{D}(p_k^2 \quad e^{\frac{2\alpha}{n(n-1)h}} + (1-p_k^2))^{n(n-1)}$$
$$-e^{-2\alpha}\prod_{k=1}^{D}(p_k^2 \quad e^{\frac{\alpha}{n(n-1)h}} + (1-p_k^2))^{2n(n-1)}$$

4.2 Computation of Risk for Kernel Estimator

Theorem 4.
Assume that f is continuous at x and that $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.Then

$$\hat{f}_n(x) \longrightarrow^p f(x).$$

Theorem 5(L.O.O risk estimator for kernels).

For any h > 0,

$$\widehat{L_1^*}(h) = \exp\{-\alpha(1 + \frac{k(0)}{(n-1)}) + \frac{\alpha}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} K_h(X_i - X_j)\} - \alpha \quad (16)$$

proof. write $K_h(x) = \frac{1}{h}K(\frac{x}{h})$.Now,

$$\sum_{i=1}^{n} \hat{f}^{(i)}(X_i) = \frac{1}{(n-1)}\{\sum_{i=1}^{n}\sum_{j=1}^{n} K_h(X_i - X_j) - nK(0)\}$$

$$\sum_{i=1}^{n} \int \hat{f}^{(i)}(x)dx = \frac{1}{(n-1)} \sum_{i=1}^{n} \sum_{j \neq i=1}^{n} \int K_h(X_i - X_j)dx = n$$

Hence,

$$\widehat{L_1^*}(h) = \exp\{-\alpha - \frac{\alpha}{(n-1)}K(0) + \frac{\alpha}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} K_h(X_i - X_j)\} - \alpha$$

Theorem 6(L.P.O risk estimator for kernels).
Following the section (2).set
K a kernel and h > 0, its bandwidth. Hence, for any p ∈ {1, ..., n},

$$\widehat{L_p^*}(h) = \exp\{-\alpha(1 + \frac{K(0)}{(n-1)}) + \frac{\alpha}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} K_h(X_i - X_j)\} - \alpha \quad (17)$$

proof.

$$\sum_{e\in\varepsilon_p}\sum_{i\in e}\hat{f}^{\bar{e}}(X_i) = \sum_{e\in\varepsilon_p}\sum_{i=1}^{n}\sum_{j\neq i=1}^{n}\frac{1}{(n-p)}K_h(X_i-X_j)\mathrm{I}(i\in e)\mathrm{I}(j\in\bar{e})$$

$$= \frac{\binom{n-2}{p-1}}{(n-p)}\{\sum_{i=1}^{n}\sum_{j=1}^{n}K_h(X_i-X_j)-nK(0)\}$$

$$\sum_{e\in\varepsilon_p}\int\hat{f}^{\bar{e}}(x)dx = \frac{1}{(n-p)}\sum_{e\in\varepsilon_p}\sum_{j\in\bar{e}}\int K_h(x-X_j)dx = \binom{n}{p}$$

Then ,

$$\widehat{L_p^*}(h) = \exp\{-\alpha(1+\frac{K(0)}{(n-1)}) + \frac{\alpha}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h(X_i-X_j)\} - \alpha$$

5. LINKING THE DENSITY ESTIMATOR TO REGRESSION:

Using the relation of density estimator and regression methods we can use all relations in nonparametric estimators for density estimation. This method is used under squared loss function for the first time by Nussabun(1996). In this method assume that $X_1, \cdots, X-n$ be i.i.d. random variables which follows the distribution F and the density function f = F', and the data in [0, 1] interval, then:

$$Y_j = \sqrt{\frac{k}{n}}\quad\sqrt{N_j+\frac{1}{4}}$$

where k = n/10 and N_j is the number of observations in the j^{th} subinterval.
Then the regression equation is as follows:

$$Y_j \approx r(t_j) + \sigma\varepsilon_j$$

Where

$$\varepsilon_j \sim N(0,1), \quad \sigma = \sqrt{\tfrac{k}{4n}}, \quad r(x) = \sqrt{f(x)} \text{ and } t_j$$

is the midpoint of the jth subinterval.

$$\hat{f}_n(x) = \frac{(r^+(x))^2}{\int_0^1 (r^+(s))^2 ds}$$

Where

$$r^+(x) = max\{\hat{r}_n(x), 0\}.$$

$$g(x) = \sqrt{f(t_j)} + \frac{\alpha}{2}\sigma^2$$

$$\hat{f}_n(x) = \frac{(\hat{g}(x) - \frac{\alpha}{2}\sigma^2)^2}{\int_0^1 (\hat{g}(s) - \frac{\alpha}{2}\sigma^2)^2 ds}$$

where in this estimation ˆg(x) is the non parametric estimation of the equation

$$g(x) = \frac{1}{\alpha} logE[e^{\alpha y}|x]$$

which was presented by S. Anatolyev(2006) under
the Linex loss function.

REFERENCES:

[1] Anatolyev,S.(2006).Kernel estimation under linear-exponential loss .Economics letters 91,39-43.

[2] Clisse,A.,Robin,S.(2008).Nonparametric density estimation by exact leavep-out cross-validation. Comput.Statist.Data Anal 52,2350-2368.

[3] Chang chang,Y.,Liang hung,V.(2007).Linex loss function with applications to determining the optimum process parameters.Springer 41,291-301.

[4] Silverman ,B.W.(1986).Density estimation for statistical and data analysis. New york :Chapman and hall.

[5] Wasserman ,L .(2007).All of nonparametric statistics.Speringer,New york.