ESTIMATION OF A MULTIVARIATE DENSITY

THEOPHILOS CACOULLOS

(Received Dec. 21, 1964)

1. Introduction and summary

Parzen [2] gave the asymptotic properties of a class of estimates $f_n(x)$ of a univariate density function $f(x)$ on the basis of a random sample $X_1, \cdots, X_n$ from $f(x)$. $f_n(x)$ is of the form

$$f_n(x)=\frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)$$

where $h=h(n)\to 0$ as $n\to\infty$ and $K(x)$ a bounded function such that $\int_{-\infty}^{+\infty} K(x)dx=1$ and $|y||K(y)|\to 0$ as $|y|\to\infty$. Our purpose is to indicate how the $f_n(x)$ can be adapted to provide estimates of a multivariate density. Actually here the extension is carried out in two directions corresponding to the two general forms of kernels K, as given in Theorems 2.1 and 4.1.

The results concerning the consistency, asymptotic unbiasedness, and bounds for bias and mean square error of f_n follow very easily by using Theorem 2.1 below and straightforward modifications of those in [2]. With respect to asymptotic normality we give here a stronger result, namely, the joint asymptotic normality of the estimates f_n at continuity points of f (Theorem 3.5). Finally, the interesting case of estimates based on product kernels is studied in Section 4.

2. Preliminary results

Motivated as in [2], we consider estimates $f_n(x)$ of the density function $f(x)$ of the following form:

$$f_n(x)=\int\frac{1}{h^p(n)}K\left(\frac{x-y}{h(n)}\right)dF_n(y)=\frac{1}{nh^p(n)}\sum_{j=1}^{n} K\left(\frac{x-X_j}{h(n)}\right), \tag{2.1}$$

where $F_n(x)$ denotes the empirical distribution function based on the sample of n independent observations $X_1, \cdots, X_n$ on the random

p-dimensional vector X with density $f(x)$; $K(y)$ is a kernel which is chosen to satisfy suitable conditions and $\{h(n)\}$ is a sequence of positive constants which in the sequel will always satisfy

$$\lim_{n\to\infty} h(n)=0\,; \tag{2.2}$$

here, and henceforth unless otherwise stated, the domain of integration is the entire range of the integrated variable.

Remark 2.1. A more general estimate than $f_n(x)$ is

$$f_n^*(x)=\frac{1}{n}\,\frac{1}{h_1\cdots h_p}\sum_{j=1}^{n} K\left(\frac{x_1-X_{j1}}{h_1},\ \cdots,\ \frac{x_p-X_{jp}}{h_p}\right) \tag{2.3}$$

where $X_j=(X_{j1},\cdots,X_{jp})$, $j=1,\cdots,n$, and the $h_i=h_i(n)>0$ satisfy

$$\lim_{n\to\infty} h_i(n)=0,\ i=1,\cdots,p\,. \tag{2.4}$$

For convenience in the exposition, we will restrict attention to the case of $h_1=\cdots=h_p$, and consider the $f_n^*(x)$ only in conjunction with the product kernels in Section 4.

The asymptotic properties of $f_n(x)$ depend rather heavily on the following multivariate analog of Theorem 1 A of [2].

THEOREM 2.1. *Suppose $K(y)$ is a Borel scalar function on E_p such that*

$$\sup_{y\in E_p} |K(y)|<\infty, \tag{2.5}$$

$$\int |K(y)|\,dy<\infty, \tag{2.6}$$

$$\lim_{|y|\to\infty} |y|^p K(y)=0, \tag{2.7}$$

where $|y|$ denotes the length of the vector y. Let $g(y)$ be another scalar function on E_p such that

$$\int |g(y)|\,dy<\infty,$$

and define

$$g_n(x)=\frac{1}{h^p(n)}\int K\left(\frac{y}{h(n)}\right) g(x-y)\,dy, \tag{2.8}$$

where $\{h(n)\}$ is a sequence of positive constants satisfying (2.2). Then at every point x of continuity of g

$$\lim_{n\to\infty} g_n(x)=g(x)\int K(y)dy\,. \tag{2.9}$$

PROOF. Choose $\delta>0$ and split the region of integration into two regions: $|y|\leqq\delta$ and $|y|>\delta$. Then we have

$$\left|g_n(x)-g(x)\int K(y)dy\right|=\left|\frac{1}{h^p}\int[g(x-y)-g(x)]K\left(\frac{y}{h}\right)dy\right|$$

$$\leqq \max_{|y|\leqq\delta}|g(x-y)-g(x)|\int_{|z|\leqq\delta/h(n)}|K(z)|dz+\int_{|y|>\delta}\frac{|g(x-y)|}{|y|^p}\frac{|y|^p}{h^p}\left|K\left(\frac{y}{h(n)}\right)\right|dy$$

$$+|g(x)|\int_{|y|>\delta}\frac{1}{h^p(n)}\left|K\left(\frac{y}{h(n)}\right)\right|dy$$

$$\leqq \max_{|y|\leqq\delta}|g(x-y)-g(x)|\int|K(z)|dz+\frac{1}{\delta^p}\sup_{|z|\geqq\delta/h(n)}|z|^p|K(z)|\int|g(y)|dy$$

$$+|g(x)|\int_{|z|\geqq\delta/h(n)}|K(z)|dz\,,$$

which tends to 0 if we let first $n\to\infty$ $(h(n)\to 0)$ and then $\delta\to 0$.

For our purposes we also require the following lemmas.

LEMMA 2.1 *Let*

$$\xi_n(x)=\frac{1}{h^p}K\left(\frac{x-X}{h}\right)\,. \tag{2.10}$$

Then, for every positive integer r, at every continuity point x of f

$$\lim_{n\to\infty} h^{p(r-1)}E[\xi_n^r(x)]=f(x)\int K^r(y)dy\,. \tag{2.11}$$

PROOF. By (2.5) and (2.6), $K^r(y)$ is bounded and absolutely integrable, and hence by Theorem 2.1

$$h^{p(r-1)}E[\xi_n^r(x)]=\int\frac{1}{h^p}K^r\left(\frac{x-y}{h}\right)f(y)dy \tag{2.12}$$

converges to $f(x)\int K^r(y)dy$ as $n\to\infty$.

LEMMA 2.2 *Let x and x^* be two continuity points of f. Then, under (2.2), the asymptotic covariance of $f_n(x)$ and $f_n(x^*)$ satisfies*

$$nh^p\,\mathrm{Cov}\,(f_n(x),\;f_n(x^*))\to 0 \tag{2.13}$$

as $n\to\infty$.

PROOF. From (2.1) we can write $f_n(x)$ as an average

$$f_n(x)=\frac{1}{n}\sum_{j=1}^{n}\xi_{nj}(x) \tag{2.14}$$

of independent random variables

$$\xi_{nj}(x)=\frac{1}{h^p}K\left(\frac{x-X_j}{h}\right)$$

identically distributed as $\xi_n(x)$ in (2.10). Hence the quantity in (2.13) is equal to

$$E\left[\frac{1}{h^p}K\left(\frac{x-X}{h}\right)K\left(\frac{x^*-X}{h}\right)\right]-h^pE[\xi_n(x)]E[\xi_n(x^*)]\,.$$

The second term $\to 0$ as $n\to\infty$ by (2.2) since by Theorem 2.1 as $n\to\infty$

$$E[\xi_n(x)]=\int\frac{1}{h^p}K\left(\frac{x-y}{h}\right)f(y)dy\to f(x)\int K(y)dy\,. \tag{2.15}$$

The first term after changing the variables can be written as

$$\int K(z)K\left(z+\frac{x^*-x}{h}\right)f(x-hz)dz\,. \tag{2.16}$$

To show that this also tends to zero as $h\to 0$ note that $K(y)$ is bounded by hypothesis and $K(y)\to 0$ as $|y|\to\infty$ by (2.7); hence, it is possible to split the region of integration into two regions; $|z|\leqq R$ and $|z|>R$ where R is sufficiently large so that for every $\varepsilon_1>0$ $|K(z)|<\varepsilon_1$ for $|z|>R$, and for every $\varepsilon_2>0$ there is an integer $N=N(\varepsilon_2)$ such that for all $n>N(\varepsilon_2)$ $h(n)$ makes $|K(z+(x^*-x)/h(n))|<\varepsilon_2$ for all $|z|\leqq R$. Since $|K(y)|<M<\infty$ it follows that the absolute value of the integral in (2.16) is smaller than

$$\varepsilon_2 M\int\frac{1}{h^p}\left|K\left(\frac{x^*-y}{h}\right)\right|f(y)dy+\varepsilon_1 M\int\frac{1}{h^p}\left|K\left(\frac{x-y}{h}\right)\right|f(y)dy\,,$$

which by Theorem 2.1, tends to 0 if we let $n\to\infty$.

3. Asymptotic properties of $f_n(x)$

Noting that $E[f_n(x)]=E[\xi_n(x)]$, we get from Lemma 2.1

THEOREM 3.1. (*Asymptotic unbiasedness of f_n*). *Suppose the kernel $K(y)$ satisfies, in addition to (2.5)-(2.7), the condition*

$$\int K(y)dy=1\,. \tag{3.1}$$

Then, under (2.2), at every continuity point x of f,

$$\lim_{n\to\infty} E[f_n(x)]=f(x).$$

Remark 3.1. Theorem 3.1 holds even if $K(y)$ is not positive for all y. However, it is desirable that the estimate $f_n(x)$ be nonnegative for every x and every n, and therefore, in the sequel, we make the more natural assumption that $K(y)\geqq 0$. This makes both $f_n(x)$ and $K(y)$ (cf. (3.1)) density functions. Moreover, the form of $f_n(x)$ as a function of the $(x-X_j)/h$ through K motivates the additional assumption from now on that

(3.2) $$K(y)=K(-y) \text{ for all } y.$$

The following two theorems can be proved in exactly the same manner as the corresponding ones in [2], by using the results of Section 2.

THEOREM 3.2. (*Consistency of f_n in quadratic mean*) *If the constants $h=h(n)$, in addition to (2.2), satisfy the condition*

(3.3) $$\lim_{n\to\infty} nh^p=0,$$

then

(3.4) $$\lim_{n\to\infty} E[f_n(x)-f(x)]^2=0$$

at every continuity point x of f.

THEOREM 3.3. (*Uniform consistency of f_n*). *If*

(*i*) *The constants $h=h(n)$, in addition to (2.2), satisfy*

(3.5) $$\lim_{n\to\infty} nh^{2p}=\infty,$$

(*ii*) *the Fourier transform*

$$k(u)=\int e^{iu'y}K(y)dy$$

of $K(y)$ is absolutely integrable, (hence $f(x)$ is uniformly continuous), then, for every $\varepsilon>0$,

(3.6) $$\lim_{n\to\infty} P[\sup_{E_p}|f_n(x)-f(x)|>\varepsilon]=0.$$

THEOREM 3.4. (*Evaluation of bias*). *If the probability density function $f(x)$ has continuous partial derivatives of third order in a neighborhood of x, then the bias*

(3.7) $$b[f_n(x)]=E[f_n(x)]-f(x)$$

satisfies

$$\lim_{n\to\infty} h^{-2}b[f_n(x)] = \frac{1}{2}\int d^2f(x\,;y)K(y)dy = \frac{1}{2}I \tag{3.8}$$

say, where

$$d^2f(x\,;y)=\sum_{i=1}^{p}\sum_{j=1}^{p}\frac{\partial^2 f(x)}{\partial x_i\partial x_j}y_iy_j\,,$$

provided the integral in (3.8) converges absolutely.

PROOF. By (2.1), (3.1) and (3.2), (3.7) may be written as

$$b[f_n(x)]=\int[f(x+hy)-f(x)]K(y)dy$$

and, expanding $f(x+hy)$ by Taylor's Theorem, (3.8) follows in view of the fact that, by (3.2),

$$\int y_iK(y)dy=0\ .$$

Under the assumptions of Theorem 3.3, an approximate expression may be given for the mean square error (m.s.e.):

$$E[f_n(x)-f(x)]^2\sim\frac{f(x)}{nh^p}\int K^2(y)dy+\frac{h^4}{4}I^2\,. \tag{3.9}$$

The value of h which minimizes the m.s.e. for a fixed value of n is easily found to be (c.f. Lemma 4a of [2])

$$h=\left[p(nI^2)^{-1}f(x)\int K^2(y)dy\right]^{1/(p+4)}, \tag{3.10}$$

from which the m.s.e.

$$E[f_n(x)-f(x)]^2\sim(p+4)\left(\frac{I^2}{4p}\right)^{p/(p+4)}\left(\frac{f(x)}{4n}\int K^2(y)dy\right)^{4/(p+4)}.$$

Therefore, $f_n(x)$ as an estimate of $f(x)$, is consistent of order $n^{4/(p+4)}$, i.e., its m.s.e.$=0(n^{-4/(p+4)})$ (cf. [2] and [3] for $p=1$).

THEOREM 3.5. *(The $\{f_n(t), t\in E_p\}$ is asymptotically a Gaussian process with independent components). Let $t_1,\cdots,t_k$ be any finite set of continuity points of the density f. If the constants satisfy (2.2) and (3.3), then the joint distribution of the random variables $f_n(t_1),\cdots,f_n(t_k)$ is asymptotically a k-variate normal in the following sense: For any real numbers $c_1,\cdots,c_k$,*

$$\lim_{n\to\infty} P[(nh^p)^{1/2}(f_n(t_i)-E[f_n(t_i)])\leqq c_i,\ 1\leqq i\leqq k]=\prod_{i=1}^{k}\Phi\left(\frac{c_i}{\sigma_i}\right)$$

where Φ denotes the standard normal distribution function, and

$$\sigma_i^2 = f(t_i)\int K^2(y)dy,\ i=1, \cdots, k. \tag{3.11}$$

PROOF. From (2.14) we have

$$(nh^p)^{1/2} f_n(t_i) = n^{-1/2} h^{p/2} \sum_{j=1}^{n} \xi_{nj}(t_i),$$

where, for each fixed i, $i=1, \cdots, k$, and each n, the $\xi_{nj}(t_i)$ are independent random variables identically distributed as

$$\xi_n(t_i) = h^{-p} K\left(\frac{t_i - X}{h}\right).$$

By Bernstein's multivariate central limit theorem (see, e.g., [1]) as applied to the sequence of independent and identically distributed random vectors

$$Z_{nj} = h^{p/2}(\xi_{nj}(t_1), \cdots, \xi_{nj}(t_k)),\ j=1, \cdots, n$$

it suffices to show that for r, $s=1, \cdots, k$

$$\lim_{n\to\infty} \operatorname{Cov}(h^{p/2}\xi_n(t_r), h^{p/2}\xi_n(t_s)) = \sigma_r^2 \delta_{rs} \tag{3.12}$$

where δ_{rs} is the Kronecker delta, and

$$\lim_{n\to\infty} n^{-1/2} \rho_n^3 = 0 \tag{3.13}$$

where

$$\rho_n^3 = \max_{1\le i\le k} E[h^{p/2}(\xi_n(t_i) - E[\xi_n(t_i)])]^3.$$

Now (3.12) follows immediately from (2.11) and (2.13), and for (3.13) it is enough to show that

$$n^{-1/2} E[h^{p/2}\xi_n(t_i)]^3 \to 0,\ i=1, \cdots, k \tag{3.14}$$

as $n\to\infty$. But by (2.11), for each i, the quantity in (3.14) is approximately equivalent to

$$(nh^p)^{-1/2} f(t_i)\int K^3(y)dy,$$

and hence (3.14) follows, in view of (3.3) and $\int K^3(y)dy < \infty$ since K is bounded and integrable.

Remark 3.2. In order to be able to replace $E[f_n(x)]$ by its limit

$f(x)$ in Theorem 3.5, so that we can state that $\sqrt{nh^p}\cdot f_n(x)$ is asymptotically normal with mean $f(x)$ and variance $f(x)\int K^2(y)dy$, it is necessary to impose some further restrictions on the rate of convergence of h to 0 as a function of n. Thus from Theorem 3.5, the bias of $f_n(x)$ must satisfy

$$(nh^p)^{1/2}b[f_n(x)]\to 0$$

as $n\to\infty$, which, under the assumptions of Theorem 3.4 and by (3.3), holds if

$$h=0(n^{-\alpha}), \quad (p+4)^{-1}<\alpha<p^{-1}.$$

It is interesting however to note that the above range of α does not include the optimum $\alpha^*=(p+4)^{-1}$ corresponding to the $h(n)$ of (3.10). Yet, α^* being the left end point of the above α interval, it suggests choosing h " just smaller " than the optimal h. This would make possible the above normal approximation of the distribution of $f_n(x)$ for " large " n, and, in such case, it is clear how this might be used, for example, in setting up a test for the hypothesis that $f(x)$ assumes a specified value. However, the discussion of this and related problems is outside the scope of our present investigation and we will not pursue it any more here.

4. Case of product kernels

In this section we indicate how the preceding results extend with respect to estimates of the form (2.3) in the special case that $K(y)$ is a product kernel in the sense that

$$K(y)=K_0(y_1)\cdots K_0(y_p) \tag{4.1}$$

where K_0 is a kernel on the real line E_1. The estimates (2.3) now take the form

$$f_n^0(x)=\frac{1}{n}\sum_{j=1}^{n}\left\{\prod_{i=1}^{p}\frac{1}{h_i}K_0\left(\frac{x_i-X_{ji}}{h_i}\right)\right\}. \tag{4.2}$$

The following theorem plays the same role as Theorem 2.1 in the preceding sections.

THEOREM 4.1. *Suppose $K(y)$ is a product kernel in the sense of (4.1) where, furthermore, K_0 is a positive bounded and even Borel function (cf. Remark 3.1) such that*

$$\int K_0(t)dt=1, \tag{4.3}$$

$$\lim_{t\to\infty} tK_0(t)=0\,. \tag{4.4}$$

Let $g(y)$ be as in Theorem 2.1 and define

$$g_n(x)=\int g(y)\left\{\prod_{i=1}^{p}\frac{1}{h_i}K_0\left(\frac{x_i-y_i}{h_i}\right)\right\}dy\,, \tag{4.5}$$

where the h_i are positive constants satisfying (2.4). Then, at every continuity point x of g, we have

$$\lim_{n\to\infty} g_n(x)=g(x)\,.$$

PROOF. For the sake of brevity and clarity, we give the proof for the bivariate case ($p=2$) since the general case requires only obvious but rather laborious modifications.

Let $\delta_1>0$, $\delta_2>0$ such that the rectangle R: $|x_1-y_1|<\delta_1$, $|x_2-y_2|<\delta_2$ is a neighborhood of $x=(x_1, x_2)$ where $g(y_1, y_2)$ is continuous. Let I_n denote the integral in (4.5) over R and I_n^* the same integral over the complement R^* of R. It suffices to show that

$$\lim_{n\to\infty} I_n=g(x_1, x_2)\,, \tag{4.6}$$

and

$$\lim_{n\to\infty} I_n^*=0\,. \tag{4.7}$$

For (4.6), note that, setting

$$x_i-y_i=h_i^{-1}z_i,\quad i=1, 2\,, \tag{4.8}$$

we can write

$$I_n=\int_{|z_1|<\delta_1/h_1}\int_{|z_2|<\delta_2/h_2} K(z_1)K(z_2)g\left(x_1-\frac{z_1}{h_1},\ x_2-\frac{z_1}{h_2}\right)dz_1\,dz_2\,,$$

which by the continuity of g and (4.3), tends to $g(x_1, x_2)$ if we let first $n\to\infty$ and then δ_1 and δ_2 go to zero.

To show (4.7), split the region of integration R^* into eight subregions as follows;

R_1: $-\infty<y_i<x_i-\delta_i$, $i=1, 2$; R_2: $-\infty<y_1<x_1-\delta_1$, $x_2+\delta_2<y_2<\infty$;
R_3: $x_1+\delta_1<y_1<\infty$, $-\infty<y_2<x_2-\delta_2$; R_4: $x_1+\delta_1\leqq y_1<\infty$, $x_2+\delta_2<y_2<\infty$;
R_5: $-\infty<y_1<x_1-\delta_1$, $|y_2-x_2|<\delta_2$; R_6: $|y_1-x_1|<\delta_1$, $-\infty<y_2<x_2-\delta_2$;
R_7: $x_1+\delta_1<y_1<\infty$, $|y_2-x_2|<\delta_2$; R_8: $|y_1-x_1|<\delta_1$, $x_2+\delta_2<y_2<\infty$.

Let the integral over R_i be denoted by I_i, $i=1, \cdots, 8$.

Now consider I_1; the I_2, I_3 and I_4 can be treated similarly. We have by (4.8) also

$$|I_1| \leqq \frac{1}{\delta_1\delta_2} \sup_{\substack{z_1>h_1^{-1}\delta_1 \\ z_2>h_2^{-1}\delta_2}} z_1 z_2 K_0(z_1)K_0(z_2) \iint |g(y_1, y_2)| dy_1\, dy_2 ,$$

which tends to zero as $n \to \infty$ by (4.4).

Finally, for I_5 (I_6, I_7 and I_8 can be treated likewise) we have

$$|I_5| \leqq \frac{1}{\delta_1} \sup_{z_1>\delta_1/h_1} z_1 K_0(z_1) \int_{|z_2|<\delta_2/h_2} K_0(z_2) g^*(x_2 - h_2 z_2) dz_2$$

where $g^*(y_2) = \int g(y_1, y_2) dy_1$. Now, in view of the uniform boundedness of $K(z_2)$ and the continuity of $g^*(y_2)$ in the interval $(x_2-\delta_2, x_2+\delta_2)$, $I_5 \to 0$ as $n \to \infty$ by (4.4).

The asymptotic properties of the estimates $f_n^0(x)$ can be obtained by using Theorem 4.1 in the same manner that Theorem 2.1 was used for $f_n(x)$ in Sections 2 and 3. Thus, for example, the condition (3.3) in Theorem 3.2 will be replaced by $n \prod_{i=1}^{p} h_i(n) \to \infty$ as $n \to \infty$, and (3.5) in Theorem 3.3 by $n \prod_{i=1}^{p} h_i^2(n) \to \infty$ as $n \to \infty$. Theorem 3.5 holds for f_n^0 if we replace h^p by $h_1 h_2 \cdots h_p$; note also that, since $K(y) = \prod_{i=1}^{p} K_0(y_i)$, the asymptotic variances in (3.11) become $\sigma_i^2 = f(t_i) \cdot \left[\int K_0^2(t)\, dt\right]^p$.

For an estimate of the bias of $f_n^0(x)$ we have the following analog of Theorem 3.4, which can be easily established.

THEOREM 4.2. *Suppose $f(x)$ satisfies the assumptions of Theorem 3.4, and $h = (h_1(n), \cdots, h_p(n))$, in addition to (2.4), satisfies*

$$\lim_{n\to\infty} \frac{h_i(n)}{h_j(n)} = r_{ij} > 0,\quad i \neq j,\ i, j = 1, \cdots, p;$$

then as $n \to \infty$

$$\frac{b[f_n^0(x)]}{|h|^2} \to \sum_i \frac{f_{ii}(x)}{r_i^2} \int t^2 K_0(t) dt \tag{4.9}$$

where

$$f_{ii}(x) = \frac{\partial^2 f(x)}{\partial x_i^2},\quad r_i^2 = \sum_{j=1}^{p} r_{ij}^2,\quad r_{ii} = 1,\quad i = 1, \cdots, p ,$$

provided the integral in (4.9) converges. Furthermore, it can be easily *verified, that for fixed n, the optimum choice of* $h_1(n), \cdots, h_p(n)$ *in order to minimize the approximate expression for the mean square error of* $f_n^0(x)$ *(cf. (3.9)) requires taking* $h_1(n)=h_2(n)=\cdots=h_p(n)=h_0(n)$*, say. It then follows that again* $h_0(n)$ *is of the same order of magnitude as* $h(n)$ *in (3.10).*

Finally, we should like to point out that the estimates $f_n^0(x)$ have a stronger invariance property than the one possessed by the $f_n(x)$, namely, whereas the $f_n(x)$ are invariant under the same scale transformation $X_i \to cX_i (c>0)$ of each of the components $X_1, \cdots, X_p$ of the observation vector X, the $f_n^0(x)$ are invariant under different scale transformations of the components of X, i.e., $X_i \to c_i X_i (c_i > 0)$. This property of $f_n^0(x)$ is more desirable from the practical point of view, since the components of X may represent incommensurable characteristics (e.g., height and weight).

UNIVERSITY OF MINNESOTA*)

REFERENCES

[1] W. Hoeffding and H. Robbins, "The central limit theorem for dependent random variables," *Duke Math. J.*, 15 (1948), 773-780.

[2] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, 33 (1962), 1065-1076.

[3] M. Rosenblatt, "Remarks on some non-parametric estimates of a density function," *Ann. Math. Statist.*, 27 (1956), 832-837.

*) Now at New York University