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:

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

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
\begin{equation}
\lim_{n \to \infty} h(n) = 0;
\end{equation}
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
\begin{equation}
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)
\end{equation}
where $X_j = (X_{j1}, \cdots, X_{jp})$, $j=1, \cdots, n$, and the $h_i = h_i(n) > 0$ satisfy
\begin{equation}
\lim_{n \to \infty} h_i(n) = 0, \quad i=1, \cdots, p.
\end{equation}
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 1A of [2].

**Theorem 2.1.** Suppose $K(y)$ is a Borel scalar function on $E_p$ such that
\begin{align}
\sup_{y \in E_p} |K(y)| &< \infty, \\
\int |K(y)| \, dy &< \infty, \\
\lim_{|y| \to \infty} |y|^p K(y) &= 0,
\end{align}
where $|y|$ denotes the length of the vector $y$. Let $g(y)$ be another scalar function on $E_p$ such that
\begin{equation}
\int |g(y)| \, dy < \infty,
\end{equation}
and define
\begin{equation}
g_n(x) = \frac{1}{h^p(n)} \int K \left( \frac{y}{h(n)} \right) g(x-y) \, dy,
\end{equation}
where $\{h(n)\}$ is a sequence of positive constants satisfying (2.2). Then at every point $x$ of continuity of $g$
\[(2.9) \lim_{n \to \infty} g_n(x) = g(x) \int K(y) \, dy.\]

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

\[
|g_n(x) - g(x)| \int K(y) \, dy \leq \frac{1}{h^n} \int |g(x - y) - g(x)| K\left(\frac{y}{h}\right) \, dy \\
\leq \max_{|y| \leq \delta} |g(x - y) - g(x)| \int |K(z)| \, dz + \int_{|y| > \delta} \frac{|g(x - y)|}{|y|^p} \left|\frac{y}{h^n}\right| K\left(\frac{y}{h^n}\right) \, dy \\
\leq \max_{|y| \leq \delta} |g(x - y) - g(x)| \int |K(z)| \, dz + \sup_{|y| > \delta} |g(y)| \int |K(y)| \, dy \\
\Rightarrow \lim_{n \to \infty} \frac{1}{h^n} \sup_{|z| > \delta} |K(z)| \int |g(y)| \, dy
\]

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

\[(2.10) \quad \xi_n(x) = \frac{1}{h^n} K\left(\frac{x-X}{h}\right).\]

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

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

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

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

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

\[(2.13) \quad nh^n \text{Cov} (f_n(x), f_n(x^*)) \to 0
\]

as $n \to \infty$.

**Proof.** From (2.1) we can write $f_n(x)$ as an average
(2.14) \[ f_n(x) = \frac{1}{n} \sum_{j=1}^{n} \xi_{n_j}(x) \]

of independent random variables \[ \xi_{n_j}(x) = \frac{1}{h^n} K\left(\frac{x-X_j}{h}\right) \]

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

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

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

(2.15) \[ E[\xi_n(x)] = \frac{1}{h^n} K\left(\frac{x-y}{h}\right) f(y) dy + f(x) \int K(y) dy. \]

The first term after changing the variables can be written as

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

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; \( |z| \leq R \) and \( |z| > R \) where \( R \) is sufficiently large so that for every \( \varepsilon > 0 \) \( |K(z)| < \varepsilon \) for \( |z| > R \), and for every \( \varepsilon > 0 \) there is an integer \( N = N(\varepsilon) \) such that for all \( n > N(\varepsilon) \) \( h(n) \) makes \( |K(z + (x^* - x)/h(n))| < \varepsilon \) for all \( |z| \leq R \). Since \( |K(y)| < M < \infty \) it follows that the absolute value of the integral in (2.16) is smaller than

\[ \varepsilon M \left[ \frac{1}{h^n} K\left(\frac{x^* - y}{h}\right) f(y) dy + \varepsilon M \left[ \frac{1}{h^n} K\left(\frac{x - y}{h}\right) f(y) dy , \right. \right. \]

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

(3.1) \[ \int K(y) dy = 1. \]

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 \( K(y) \geq 0 \). 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 \((x - X_i)/h\) through \( K \) motivates the additional assumption from now on that

\[
K(y) = K(-y) \quad \text{for all } y.
\]

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

\[
\lim_{n \to \infty} nh^p = 0,
\]

then

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

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

\[
\lim_{n \to \infty} nh^{2p} = \infty,
\]

(ii) the Fourier transform

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

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

\[
\lim_{n \to \infty} P[\sup_{x \in \mathbb{R}} |f_n(x) - f(x)| > \epsilon] = 0.
\]

**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

\[
b[f_n(x)] = E[f_n(x)] - f(x)
\]

satisfies
\[
\lim_{n \to \infty} h^{-1} b[f_n(x)] = \frac{1}{2} \int d^2 f(x \cdot y)K(y)dy = \frac{1}{2} I
\]
say, where
\[
d^2 f(x \cdot y) = \sum_{i=1}^{p} \sum_{j=1}^{p} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} y_i y_j,
\]
provided the integral in (3.8) converges absolutely.

**Proof.** By (2.1), (3.1) and (3.2), (3.7) may be written as
\[
b[f_n(x)] = \int \left[ f(x + hy) - f(x) \right] K(y)dy
\]
and, expanding \( f(x + hy) \) by Taylor’s Theorem, (3.8) follows in view of the fact that, by (3.2),
\[
\int y_i K(y)dy = 0.
\]

Under the assumptions of Theorem 3.3, an approximate expression may be given for the mean square error (m.s.e.):
\[
E[f_n(x) - f(x)]^2 \sim \frac{f(x)}{nh^2} \int K^2(y)dy + \frac{h^4}{4} I.
\]

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 \([2]\))
\[
h = \left[ p(nI)^{-1} f(x) \int K^2(y)dy \right]^{1/(p+4)},
\]
from which the m.s.e.
\[
E[f_n(x) - f(x)]^2 \sim (p+4) \left( \frac{I}{4p} \right)^{p/(p+4)} \left( \frac{f(x)}{4n} \int K^2(y)dy \right)^{4/(p+4)}.
\]
Therefore, \( f_n(x) \) as an estimate of \( f(x) \), is consistent of order \( n^{4/(p+4)} \), i.e., its m.s.e. \( = 0(n^{-4/(p+4)}) \) (c.f. \([2]\) 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 \( t_1, \ldots, t_k \) 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 \( f_n(t_1), \ldots, f_n(t_k) \) is asymptotically a \( k \)-variate normal in the following sense: For any real numbers \( c_1, \ldots, c_k \),
\[
\lim_{n \to \infty} P[(nh^{\alpha})^{1/2}(f_n(t_i) - E[f_n(t_i)]) \leq c_i, 1 \leq i \leq k] = \prod_{i=1}^{k} \Phi \left( \frac{c_i}{\sigma_i} \right)
\]
where \( \Phi \) denotes the standard normal distribution function, and

\[
\sigma_i^2 = f(t_i) \int K^i(y) dy, \quad i = 1, \ldots, k.
\]

**Proof.** From (2.14) we have

\[
(nh^n)^{1/2} f_n(t_i) = n^{-1/2} h^{p/2} \sum_{j=1}^n \xi_n(t_i),
\]

where, for each fixed \( i, \ i = 1, \ldots, k \), and each \( n \), the \( \xi_n(t_i) \) 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

\[
Z_{nj} = h^{p/2} (\xi_{nj}(t_1), \ldots, \xi_{nj}(t_k)), \quad j = 1, \ldots, n
\]

it suffices to show that for \( r, s = 1, \ldots, k \)

\[
\lim_{n \to \infty} \text{Cov} \left( h^{p/2} \xi_n(t_r), h^{p/2} \xi_n(t_s) \right) = \sigma_r^2 \delta_{rs},
\]

where \( \delta_{rs} \) is the Kronecker delta, and

\[
\lim_{n \to \infty} n^{-1/2} \rho_n^2 = 0
\]

where

\[
\rho_n^2 = \max_{1 \leq i \leq k} E \left[ h^{p/2} (\xi_n(t_i) - E[\xi_n(t_i)])^2 \right].
\]

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

\[
n^{-1/2} E[h^{p/2} \xi_n(t_i)] \to 0, \quad i = 1, \ldots, k
\]

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

\[
(nh^n)^{-1/2} f(t_i) \int K^i(y) dy,
\]

and hence (3.14) follows, in view of (3.3) and \( \int K^i(y) dy < \infty \) 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 $f(x)\int K'(y)dy$, 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

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

It is interesting however to note that the above range of $\alpha$ does not include the optimum $\alpha^* = (p+4)^{-1}$ corresponding to the $h(n)$ of (3.10). Yet, $\alpha^*$ being the left end point of the above $\alpha$ 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

$$(4.1) \quad K(y) = K_1(y_1) \cdots K_p(y_p)$$

where $K_i$ is a kernel on the real line $E_i$. The estimates (2.3) now take the form

$$(4.2) \quad f_n^*(x) = \frac{1}{n} \sum_{j=1}^{n} \left\{ \prod_{i=1}^{p} \frac{1}{h_i} K_i \left( \frac{x_i - X_{ii}}{h_i} \right) \right\}.$$ 

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, $K_i$ is a positive bounded and even Borel function (cf. Remark 3.1) such that

$$(4.3) \quad \int K_i(t)dt = 1,$$
(4.4) \[ \lim_{t \to \infty} tK_t(t) = 0. \]

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

(4.5) \[ g_n(x) = \left( \prod_{i=1}^p \frac{1}{h_i} K_0 \left( \frac{x_i - y_i}{h_i} \right) \right) dy, \]

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

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

and

(4.7) \[ \lim_{n \to \infty} I_n^* = 0. \]

For (4.6), note that, setting

(4.8) \[ x_i - y_i = h_i^{-1} x_i, \quad i = 1, 2, \]

we can write

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

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 \( \delta_1 \) and \( \delta_2 \) go to zero.

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

\[ R_1: \quad -\infty < y_1 < x_1 - \delta_1, \quad i = 1, 2; \quad R_4: \quad -\infty < y_1 < x_1 - \delta_1, \quad x_2 + \delta_2 < y_2 < \infty; \]
\[ R_2: \quad x_1 + \delta_1 < y_2 < \infty, \quad -\infty < y_1 < x_2 - \delta_2; \quad R_3: \quad x_1 + \delta_1 \leq y_1 < \infty, \quad x_2 + \delta_2 < y_2 < \infty; \]
\[ R_5: \quad -\infty < y_1 < x_1 - \delta_1, \quad |y_1 - x_2| < \delta_2; \quad R_4: \quad |y_1 - x_1| < \delta_1, \quad -\infty < y_2 < x_2 - \delta_2; \]
\[ R_6: \quad x_1 + \delta_1 < y_1 < \infty, \quad |y_2 - x_2| < \delta_2; \quad R_5: \quad |y_1 - x_1| < \delta_1, \quad x_2 + \delta_2 < y_2 < \infty. \]

Let the integral over \( R_i \) be denoted by \( I_i, \quad i = 1, \ldots, 8. \)
Now consider $I_i$; the $I_i$, $I$, and $I_i$ can be treated similarly. We have by (4.8) also

$$
|I_i| \leq \frac{1}{\bar{d}_i} \sup_{x_i, \gamma_i > \delta_i} z_i, K_i(z_i) K_i(z) \int \left| g(y_1, y_i) \right| dy_1, dy_i,
$$

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

Finally, for $I_i$ ($I_i$, $I$, and $I_i$ can be treated likewise) we have

$$
|I_i| \leq \frac{1}{\bar{d}_i} \sup_{x_i, \gamma_i > \delta_i} z_i, K_i(z_i) \int_{x_1 < y_i} K_i(z_1) g^*(x_1 - h_1 z_1) dx_1
$$

where $g^*(y_i) = \int g(y_1, y_i) dy_1$. Now, in view of the uniform boundedness of $K(z_i)$ and the continuity of $g^*(y_i)$ in the interval $(x_1 - \delta_i, x_1 + \delta_i)$, $I_i \to 0$ as $n \to \infty$ by (4.4).

The asymptotic properties of the estimates $f_n^*(x)$ 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 $n \prod_{i=1}^{p} h_i^*(n) \to \infty$ as $n \to \infty$. Theorem 3.5 holds for $f_n^*$ if we replace $h_i^*$ by $h_1 h_2 \cdots h_p$; note also that, since $K(y) = \prod_{i=1}^{p} K_i(y_i)$, the asymptotic variances in (3.11) become $\sigma_i^2 = f(t_i) \left[ \int K_i(t) dt \right]^p$.

For an estimate of the bias of $f_n^*(x)$ 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 $h = (h_1(n), \ldots, h_p(n))$, in addition to (2.4), satisfies

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

then as $n \to \infty$

$$
(4.9) \quad \frac{b[f_n^*(x)]}{|h|^i} \to \sum_{i} \frac{f_{ii}(x)}{r_i^2} \int t^i K_i(t) dt
$$

where

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

provided the integral in (4.9) converges. Furthermore, it can be easily
verified, that for fixed $n$, the optimum choice of $h_1(n), \ldots, h_p(n)$ in order to minimize the approximate expression for the mean square error of $f_n^2(x)$ (cf. (3.9)) requires taking $h_1(n) = h_2(n) = \cdots = h_p(n) = h(n)$, say. It then follows that again $h(n)$ is of the same order of magnitude as $h(n)$ in (3.10).

Finally, we should like to point out that the estimates $f_n^2(x)$ 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 $X_1 \rightarrow cX_1 (c > 0)$ of each of the components $X_1, \ldots, X_p$ of the observation vector $X$, the $f_n^2(x)$ are invariant under different scale transformations of the components of $X$, i.e., $X_1 \rightarrow c_i X_1 (c_i > 0)$. This property of $f_n^2(x)$ 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


*) Now at New York University