SEQUENTIAL METHODS IN STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 16
PWN-POLISH SCIENTIFIC PUBLISHERS
WARSAW 1985

RECURSIVE AND SEQUENTIAL DENSITY ESTIMATION

W. WERTZ

Institute of Statistics, Technical University,
Vienna, Austria

1. Introduction and notation

1.1. This paper intends to give a concise survey over the field of sequential density estimation without claiming completeness. Nevertheless, compared with the literature about density estimation in general, there has been done only little work about the sequential part of the theory, and many open problems are left open until now.

In Chapter 2 we give a general (negative) result about *uniform consistency* of sequential density estimators due to Farrell. In Chapter 3 the most usual *recursive estimators* are discussed; there we have very weak regularity assumptions in view; additionally, some material about the nonrecursive kernel estimators is included in order to compare the efficiency, the speed of convergence of the estimators etc. Under somewhat stronger conditions this is done at some length in Chapter 4; we use some different sets assumptions systematically. For simplicity of notation we restrict ourselves to the case of one-dimensional observations from Chapter 2 on, but most of the results can be extended to higher dimensions. In order to make the results more transparent, we give the asymptotic expressions for the variance and bias of the estimators separately (see, e.g., 4.4, 4.5 and 4.6). In these two chapters stress is laid on the local estimation problem.

For the investigation of the asymptotic behaviour of a global error of the estimator (mean integrated square error) the method of *stochastic approximation* in Hilbert spaces can be applied; some results of this type are included in Chapter 5.

Chapter 6 treats *sequential methods*, for the local estimation problem as well as for the global one. In particular, asymptotic fixed-length confidence intervals and confidence regions are given.

1.2. We use the following *notation*:

$\boldsymbol{R}(\boldsymbol{N})$	set of real (natural) numbers;
$\boldsymbol{1}$	indicator function;
$\|\cdot\|_\infty$	essential supremum norm of a function;
f	probability density;
$\boldsymbol{E}_f(\boldsymbol{V}_f)$	expectation (variance) computed with density f;
$\xrightarrow{\mathscr{L}}$	convergence in distribution;
$\mathscr{N}(\xi, \sigma^2)$	normal distribution;
$\mu^n, p^n, \ldots$	product measures.

1.3. Let there be given a statistical model $(X, \mathfrak{X}, \mathfrak{P})$ and suppose that each $P \in \mathfrak{P}$ has a Radon–Nikodym density f with respect to a given measure μ. ξ denotes a random variable defined on some probability space $(\Omega, \mathfrak{A}, \boldsymbol{P})$ mapping into X and having distribution $P \in \mathfrak{P}$, i.e., $\boldsymbol{P}[\xi \in A] = \int_A f\, d\mu$ for every $A \in \mathfrak{X}$. $\xi_1, \ldots, \xi_n, \ldots$ denote independent random variables distributed like ξ, $\xi^n := (\xi_1, \ldots, \xi_n)$; x_n denotes values of ξ_n, $x^n := (x_1, \ldots, x_n)$ realizations of ξ^n. (By the way, some of the results can be generalized to the case of dependent random variables under suitable conditions.) Further we denote by $\mathfrak{F}$ the class of densities corresponding to $\mathfrak{P}$.

1.4. A stopping variable ν is a measurable function ν: $\Omega \to \boldsymbol{N} \cup \{\infty\}$ such that $\boldsymbol{P}[\nu = \infty] = 0$ and $\{\omega \in \Omega: \nu(\omega) = n\}$ belongs to the σ-algebra generated by $\xi_1, \ldots, \xi_n$, for every n. For details and references we refer, e.g., to [8]. Obviously, it is possible to redefine ν as a function on X^∞.

1.5. For given finite sample size n, every measurable function $\hat{f}$: $X^n \times X \to \boldsymbol{R}$ is called a *(density) estimator* of f. Usually further restrictions are imposed on $\hat{f}$, e.g., $\hat{f}$ belongs to some suitable function space. The estimation problem can be formalized as follows (for details see [26], [27]): let B a Banach space (or more generally, a locally convex topological linear space), Φ a nonnegative convex function on B. Let d: $\mathfrak{F} \to B$ be an arbitrary function, D a subspace of B serving as decision space;

$$L(g, f) := \Phi(g - d(f)) \qquad (g \in D, f \in \mathfrak{F})$$

is used as loss and an appropriate space of strongly measurable functions h: $X^n \to B$ as space of estimators. Then

$$R(h, f) := \boldsymbol{E}_f L(h(\xi_1, \ldots, \xi_n; \cdot), f)$$

is the risk of h at f.

Taking B as a proper space of measurable functions defined on $(X, \mathfrak{X}, \mu)$ and $\Phi(g) := \int \varphi \circ g\, d\lambda$, where φ is a nonnegative convex function and λ an appropriate measure on $(X, \mathfrak{X})$, gives the density estimation problem as a special case. Choosing λ as a point measure yields a local problem, choosing

λ as a measure, equivalent to μ yields a global problem. Mostly we use $\varphi(t):= t^2$. From the general theory complete class theorems, existence theorems for optimal solutions etc. for density estimation can be obtained.

For example, the risk of a density estimator $\hat{f}$ of f with $\lambda =$ point measure concentrated in x yields the mean square error

$$R(\hat{f}, f) = E_f[\hat{f}(\xi^n; x) - f(x)]^2.$$

1.6. A sequential density estimator $\hat{f}_\nu$ is a pair $((\hat{f}_n), \nu)$ consisting of a sequence of fixed sample size density estimators $\hat{f}_n$ and a stopping variable ν. The corresponding sequential risk is

$$R(\hat{f}_\nu, f) = \sum_{n=1}^{\infty} E_f\left[\Phi(\hat{f}_n(\xi^n; \cdot) - f) \cdot 1_{\{\nu = n\}}\right]$$
$$= \sum_{n=1}^{\infty} \int_{\{x^n: \nu = n\}} \Phi(\hat{f}_n(x^n; \cdot) - f) \cdot \prod_{i=1}^{n} f(x_i)\, d\mu^n(x^n). \tag{1.1}$$

2. Nonexistence of uniformly consistent sequential estimators

2.1. In the following chapters several types of consistent density estimators will be discussed. However, it is not possible to construct uniformly consistent sequential density estimators, even pointwise, under rather weak assumptions. The following results is due to Farrell ([20]):

2.2. THEOREM. *Let $\alpha \geq 3$ and $\mathfrak{C}_\alpha^*$ be the set of all one-dimensional probability densities f which are piecewise continuously differentiable and fulfil $\|f\|_\infty \leq \alpha$. Let $\mathfrak{C}_\alpha$ be the subclass of all $f \in \mathfrak{C}_\alpha^*$ which are continuously differentiable. Let $\hat{f}_\nu$ be a sequential estimator of f at the point 0 and the risk be based on quadratic loss,* i.e., *the sequential risk* (1.1) *is given by*

$$R(\hat{f}_\nu, f) = \sum_{n=1}^{\infty} \int_{\{\nu = n\}} [\hat{f}_n(x^n; 0) - f(0)]^2 \prod_{i=1}^{n} f(x_i)\, dx^n.$$

Then

(i) *there is a two-parameter subclass $\mathfrak{F}$ of $\mathfrak{C}_\alpha^*$ with* $\sup_{f \in \mathfrak{F}} R(\hat{f}_\nu, f) \geq 1/16$ *provided* $\sup_{f \in \mathfrak{F}} E_f \nu < \infty$;

(ii) $\sup_{f \in \mathfrak{C}_\alpha} R(\hat{f}_\nu, f) \geq 1/16$ *provided* $\sup_{f \in \mathfrak{C}_\alpha} E_f \nu < \infty$.

The idea of the proof is the following: let for $\gamma > 0$ and $\theta \in \boldsymbol{R}$

$$f(t; \theta, \gamma) := \begin{cases} C(\gamma) \cdot e^{-|t+\theta|} & \text{if } |t+\theta| \geq \gamma, \\ C(\gamma) \cdot \left[e^{-\gamma} + 1 - \dfrac{|t+\theta|}{\gamma}\right] & \text{if } |t+\theta| \leq \gamma \end{cases}$$

with a norming constant $C(\gamma)$. Clearly, $\mathfrak{F} := \{f(\cdot\,; \theta, \gamma)\colon \theta \in \boldsymbol{R}, \gamma > 0\}$ is a subclass of $\mathfrak{C}_\alpha^*$. Application of Wolfowitz' sequential version of the Fréchet–Cramér–Rao inequality (see [65]) yields a lower bound for $R(\hat{f}_\nu; f(\cdot\,; \theta, \gamma))$ and some computations yield the first result. Since every element of $\mathfrak{C}_\alpha^*$ can be approximated almost everywhere by functions of $\mathfrak{C}_\alpha$, a simple application of Fatou's lemma gives (ii).

2.3. Theorem 2.2 yields of course a lower bound also for the maximum risk of fixed-sample-size estimators.

3. Recursive estimators – fundamental properties

3.1. One of the most popular estimation methods for curves is the *kernel method*. It has been introduced in 1956 in a paper of Rosenblatt ([35]) and further investigated by Parzen ([31]); since that time a great deal of work has been done about this kind of estimators (the literature up to 1977 is listed in [42], a survey over the most usual methods is given in [60]).

The original idea was to estimate the density by a two-sided difference quotient of the empirical distribution function; this can be expressed in the form

$$\hat{f}_n(x_1, \ldots, x_n; x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{b_n} K\left(\frac{x - x_i}{b_n}\right) \tag{3.1}$$

where b_n is a positive real number and $K = \frac{1}{2} \cdot \boldsymbol{1}_{[-1,1]}$.

More generally, estimators of the form (3.1), where (b_n) is a sequence with

$$b_n > 0, \qquad \lim_{n\to\infty} b_n = 0, \qquad \lim_{n\to\infty} n b_n = \infty \tag{3.2}$$

and $K\colon \boldsymbol{R} \to \boldsymbol{R}$ is a measurable function with

$$K(t) = K(-t) \qquad \text{for every} \quad t \in \boldsymbol{R}, \quad \int K(t)\,dt = 1 \tag{3.3}$$

are called *kernel estimators* (of the density). Additionally we shall always assume K to be nonnegative. Under these conditions, $x \mapsto \hat{f}_n(x_1, \ldots, x_n; x)$ is a probability density itself for every set of observations. b_n is called *bandwidth* and K *kernel*. We shall refer to this type of estimator as *Rosenblatt–Parzen estimator* (RPE).

3.2. Let us suppose that the loss is measured by some mean integrated square error formed with measure λ on $\boldsymbol{R}$; then, under suitable conditions,

$$\begin{aligned} R(\hat{f}_n, f) &= \boldsymbol{E}_f \int [\hat{f}_n(\xi^n; x) - f(x)]^2\, d\lambda(x) \\ &= \int \boldsymbol{E}_f [\hat{f}_n(\xi^n; x) - \boldsymbol{E}_f \hat{f}_n(\xi^n; x)]^2\, d\lambda(x) + \int [\boldsymbol{E}_f \hat{f}_n(\xi^n; x) - f(x)]^2\, d\lambda(x). \end{aligned}$$

The integrand of the first integral is $V_f \hat{f}_n(\xi^n; x)$, the integrand of the second one is $[B(\hat{f}_n, f)]^2$, the square of the *bias of* $\hat{f}_n$. The problem turns out to make the variance and the bias of the estimator small simultaneously. (We remark that unbiased density estimators do not exist if the statistical model is large enough, see [35] and, for a refinement, [3]; concerning problems of unbiasedness, Lehmann's concept (see [28]) is more appropriate in nonparametric statistics than the classical one.)

In case $\hat{f}_n$ is a kernel estimator with fixed kernel K, the choice of a large bandwidth makes the variance small (the estimators are smooth then). It will be discussed below that

$$t \mapsto K_n(t) := \frac{1}{b_n} K\left(\frac{t}{b_n}\right)$$

is an approximate identity under a few regularity conditions and

$$E_f \hat{f}_n(\xi_1, \ldots, \xi_n; x) = (K_n * f)(x);$$

small b_n's give a good approximation, hence make the bias small. The problem is to find a medium size for b_n, meeting both requirements.

3.3. The estimators (3.1) can be easily generalized to the r-dimensional case. One possibility is

$$\hat{f}_n(x_1, \ldots, x_n; x) = \sum_{i=1}^{n} \frac{1}{b_n^r} \cdot K\left(\frac{x - x_i}{b_n}\right) \tag{3.4}$$

with $K: \boldsymbol{R}^r \to \boldsymbol{R}$; condition (3.2) is substituted by

$$b_n > 0, \qquad \lim_{n \to \infty} b_n = 0, \qquad \lim_{n \to \infty} n b_n^r = 0 \tag{3.5}$$

(see [5]). A more general definition is due to Văduva ([51]): let $x_i = (x_i^1, \ldots, x_i^r)$, $x = (x^1, \ldots, x^r)$, $b_n = (b_n^1, \ldots, b_n^r)$ and

$$b_n^j > 0, \qquad \lim_{n \to \infty} \|b_n\| = 0, \qquad \lim_{n \to \infty} n b_n^1 \cdot \ldots \cdot b_n^r = 0; \tag{3.6}$$

then the estimator is defined by

$$\hat{f}_n(x_1, \ldots, x_n; x) = \frac{1}{n} \sum_{i=1}^{n} (b_n^1 \cdot \ldots \cdot b_n^r)^{-1} \cdot K\left(\frac{x^1 - x_i^1}{b_n^1}, \ldots, \frac{x^r - x_i^r}{b_n^r}\right). \tag{3.7}$$

Kernel estimators for more general sample space (topological groups, homogeneous spaces, etc.) are dealt with, e.g., in [9], [59] and [61].

3.4. Further generalizations are treated by several authors. Let (K_n) be a sequence of measurable functions; under different conditions ensuring (K_n) to be an approximate identity, Watson and Leadbetter in [56], [57], Văduva in [52] and others investigate estimators of the form

$$\hat{f}_n(x_1, \ldots, x_n; x) = \frac{1}{n} \sum_{i=1}^{n} K_n(x - x_i). \tag{3.8}$$

Watson and Leadbetter ([55]), Hašimov ([68]), Mirzahmedov and Hašimov ([30]) study estimators (3.8) by means of Fourier analytic methods. Whittle ([64]), Földes and Révész ([22]), Földes ([21]), Walter and Blum ([54]), and Susarla and Walter ([50]) investigate the even more general class

$$\hat{f}_n(x_1, \ldots, x_n; x) = \frac{1}{n} \sum_{i=1}^{n} K_n(x, x_i) \tag{3.9}$$

with bimeasurable functions K_n; their crucial property should be

$$\lim_{n \to \infty} \int K_n(x, y) \cdot \varphi(y)\, dy = \varphi(x) \tag{3.10}$$

for every sufficiently smooth function φ. (Some authors call estimators of the type (3.8) and (3.9) delta-sequence estimators.) Many kinds of estimators, e.g., (3.1), (3.4), (3.7), orthogonal series estimators, histograms, certain interpolation estimators, turn out to be special cases of (3.9).

For purposes of sequential analysis and recursive estimation theory, a general definition, due to Deheuvels ([15]) is advantageous:

$$\hat{f}_n(x_1, \ldots, x_n; x) = \sum_{k=1}^{n} K_{n,k}(x, x_k) \tag{3.11}$$

with $K_{n,k}$ fulfilling conditions similar to (3.10). The following special case of (3.11) is convenient for our purposes:

$$\hat{f}_n(x_1, \ldots, x_n; x) = \Big[\sum_{i=1}^{n} b_i \cdot H(b_i)\Big]^{-1} \sum_{k=1}^{n} H(b_k) \bar{K}\left(x, \frac{x - x_k}{b_k}\right),$$

where (b_n) fulfils (3.2), H is some fixed positive function and $y \mapsto \bar{K}(x, y)$ is a probability density for every x (see [13]). We shall consider in particular

$$\Big[\sum_{i=1}^{n} b_i H(b_i)\Big]^{-1} \sum_{k=1}^{n} H(b_k) K\left(\frac{x - x_k}{b_k}\right), \tag{3.12}$$

and special cases of this estimator given in (3.15) and (3.16).

3.5. Deheuvels ([16]) has given the most general form of recursive estimators of the form (3.11): a sequence $(\hat{f}_n)$ of estimators is called *recursive* (or, par abus du langage, $\hat{f}_n$ is a recursive estimator) if

$$\hat{f}_{n+1}(x_1, \ldots, x_{n+1}; x) = R_n(\hat{f}_n(x_1, \ldots, x_n; x), x_{n+1}, x) \tag{3.13}$$

for every $n \in N$, x and every sequence (x^∞) of observations with appropriately measurable functions R_n.

THEOREM ([16]). *Let $\hat{f}_n$ be a recursive estimator of the form* (3.11), $u \mapsto R_n(u, x_{n+1}, x)$ *and* $y \mapsto K_{n,k}(x, y)$ *be differentiable, the latter function not constant. Then $\hat{f}_n$ can be expressed in the form*

$$\hat{f}_n(x_1, \ldots, x_n; x) = c_n(x) \cdot \sum_{k=1}^{n} K_k^*(x, x_k) \tag{3.14}$$

with certain constants $c_n(x)$ and kernels K_k^*.

3.6. In the sequel we shall mainly consider the following types of kernel estimators:

(i) the RPE (3.1);

(ii) Deheuvels' estimator (3.12), abbreviated DE;

(iii) the estimator defined by Wolverton and Wagner ([66]), and independently by Yamato ([67]), abbreviated by WWYE:

$$\hat{f}_n(x_1, \ldots, x_n; x) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{b_k} K\left(\frac{x - x_k}{b_k}\right); \tag{3.15}$$

(iv) a more special estimator defined by Deheuvels ([13]), called DE*:

$$\hat{f}_n(x_1, \ldots, x_n; x) = \left[\sum_{i=1}^{n} b_i\right]^{-1} \cdot \sum_{k=1}^{n} K\left(\frac{x - x_k}{b_k}\right). \tag{3.16}$$

(iii) and (iv) are special cases of (ii), corresponding to the choice $H(b) = 1/b$ and $H \equiv 1$, respectively. As will be seen below, these both cases have certain asymptotic optimality properties.

3.7. The following lemma, due to Parzen (part (3)), and Devroye, is fundamental for establishing asymptotic properties of kernel estimators:

LEMMA ([31], [17]). *Let K and f: $\mathbf{R}^r \to \mathbf{R}$ be integrable functions,* $\|K\|_\infty < \infty$, $b_n > 0$ *and* $\lim_{n \to \infty} b_n = 0$.

We say that condition [D] *is satisfied if one of the following conditions holds:*

(1) *f is almost everywhere bounded and x is Lebesgue point of f;*

(2) *K has compact support and x is Lebesgue point of f;*

(3) $\lim_{\|t\| \to \infty} \|t\|^r K(t) = 0$ *and f is continuous at x;*

(4) $\int_0^\infty u^{r-1} L(u)\, du < \infty$ *and x is Lebesgue point of f;*

(5) $\int_{\mathbf{R}^r} L^*(y)\, dy < \infty$, *$x$ is Lebesgue point of f; here*

$$L(u) := \sup_{\|x\| \geq u} |K(x)| \qquad \text{and} \qquad L^*(y) := \sup_{\|x\| \geq \|y\|} |K(x)|.$$

Let $K_n(t) := \frac{1}{b_n} K\left(\frac{t}{b_n}\right)$. *Then, under condition* [D],

$$\lim_{n\to\infty} (K_n * f)(x) = f(x) \cdot \int K(y)\, dy$$

holds.

From a statistical point of view, conditions (2), (4) and (5) are most convenient, because there are no assumptions about the underlying function f; it is well known that almost every point x is Lebesgue point of the integrable function f.

We further remark that $\lim_{\|t\|\to\infty} \|t\|^r \cdot K(t) = 0$ (in condition (3)) is equivalent to $\lim_{u\to\infty} u^r L(u) = 0$ which is a little weaker than $\int_0^\infty u^{r-1} L(u)\, du < \infty$ (in (4)).

The proofs are in [17], [18], [46]. As mentioned earlier, for sake of lucidity, we shall only consider the case $r = 1$.

3.8. COROLLARY. *Let ξ be a random variable distributed with probability density f, K an essentially bounded probability density and condition* [D] *hold with $r = 1$. Then*

$$\lim_{n\to\infty} E_f \frac{1}{b_n} K\left(\frac{x-\xi}{b_n}\right) = f(x).$$

3.9. COROLLARY. *Let K be an essentially bounded probability density and* [D] *with $r = 1$ be satisfied. Then*

$$\lim_{n\to\infty} b_n \cdot V_f \left[\frac{1}{b_n} K\left(\frac{x-\xi}{b_n}\right)\right] = f(x) \cdot \int K^2(y)\, dy.$$

3.10. Next we apply these results to the RPE, WWYE and DE. By Toeplitz' lemma we get:

THEOREM. *Let K be an essentially bounded probability density, condition* [D] *be satisfied and (b_n) a sequence of bandwidths fulfilling* (3.2). *Then we have*

(i) *For the Rosenblatt–Parzen estimator* (3.1)

$$\lim_{n\to\infty} E_f \hat{f}_n(\xi_1, \ldots, \xi_n; x) = f(x) \tag{3.17}$$

and

$$\lim_{n\to\infty} n b_n \cdot V_f \hat{f}_n(\xi_1, \ldots, \xi_n; x) = f(x) \cdot \int K^2(y)\, dy \tag{3.18}$$

hold.

(ii) *Let*

$$\lim_{n\to\infty} \frac{nb_n}{[\sum_{i=1}^{n} b_i H(b_i)]^2} \cdot \sum_{k=1}^{n} b_k H^2(b_k) = \alpha \qquad \textit{exist with} \quad 0 < \alpha < \infty. \tag{3.19}$$

Then for the Deheuvels estimator (3.12), (3.17) *and*

$$\lim_{n\to\infty} nb_n \cdot V_f \hat{f}_n(\xi_1, \ldots, \xi_n; x) = \alpha \cdot f(x) \cdot \int K^2(y)\, dy \tag{3.20}$$

is valid.

(iii) *If*

$$\lim_{n\to\infty} \left(\frac{b_n}{n}\right) \cdot \sum_{k=1}^{n} \frac{1}{b_k} = \alpha \qquad \textit{exists with} \quad 0 < \alpha < \infty, \tag{3.21}$$

then (3.17) *and* (3.20) *hold for the Wolverton–Wagner–Yamato estimator* (3.15).

(iv) *Let*

$$\lim_{n\to\infty} nb_n \left(\sum_{i=1}^{n} b_i\right)^{-1} = \alpha \qquad \textit{exist with} \quad 0 < \alpha < \infty; \tag{3.22}$$

then (3.17) *and* (3.20) *are valid for the special Deheuvels estimator* (3.16).

Of course, (3.21) and (3.22) are special cases of (3.19).

Remark. Under much more restrictive conditions, (i) has been proved by Parzen ([31]) and (iii) by Yamato ([67]), (ii) and (iv) by Deheuvels ([15]). A popular choice of the bandwidth is $b_n = \gamma \cdot n^{-\beta}$ ($\gamma > 0$, $0 < \beta < 1$); with this b_n, $\alpha = 1/(1+\beta)$ in case (iii) and $\alpha = 1-\beta$ in case (iv). Further by Schwarz' inequality

$$\left(\sum_{i=1}^{n} b_i\right)^{-1} \leqslant \frac{\sum_{i=1}^{n} b_i H^2(b_i)}{[\sum_{i=1}^{n} b_i H(b_i)]^2};$$

hence $H \equiv 1$ yields the minimal asymptotic variance in the class of all DE's (3.12). Hence the variances of the estimators satisfy asymptotically

$$V(\text{iv}) \leqslant V(\text{iii}) \leqslant V(\text{i}) \qquad \text{and} \qquad V(\text{iv}) \leqslant V(\text{ii}).$$

Banon ([1]) constructs a (rather complicated) recursive estimator with asymptotically smaller variance than (3.16). Considerations involving also the bias term are treated below.

3.11. We remark that there is a great variety of possible choices of b_n, e.g., $b_n = \gamma \cdot [n^\beta \cdot \log n \cdot \log\log n \cdot \ldots \cdot \log_l n]^{-1}$ with $0 \leqslant \beta \leqslant 1$, $l \in \boldsymbol{Z}$, $l \geqslant 0$, $\beta + l > 0$. Again condition (3.21) yields $\alpha = 1/(1+\beta)$ and (3.22) gives $\alpha = 1-\beta$.

Similar choices of bandwidths, having interesting effects on the mean square error, are discussed by Deheuvels ([16]).

3.12. By the preceding results, one can prove the following weak and strong consistency results due to Devroye; since in the next chapter even the exact rate of convergence will be given (Theorem 4.8), for the sake of simplicity we only deal with the WWYE.

THEOREM ([17]): *Let K be an essentially bounded probability density, (b_n) satisfy* (3.2), *the assumptions of Lemma* 3.7 *be satisfied and $\hat{f}_n$ the* WWYE. *Then $\hat{f}_n(\xi_1, \ldots, \xi_n; x) \to f(x)$ in probability. If additionally*

$$\lim_{n\to\infty} \frac{nb_n}{\log\log n} = \infty, \tag{3.23}$$

then $\boldsymbol{P}\left[\lim_{n\to\infty} \hat{f}_n(\xi_1, \ldots, \xi_n; x) = f(x)\right] = 1.$

The proof makes use of Lemma 3.7 and Bennett's inequality ([2]). For the strong formulation an appropriate version of the strong law of large numbers (see [29], p. 253) is applied.

3.13. THEOREM. *Let the conditions of Theorem* 3.12 *be fulfilled, but instead of $nb_n \to \infty$ we assume the (weaker) condition*

$$\frac{1}{n^2} \cdot \sum_{i=1}^{n} \frac{1}{b_i} \to 0. \tag{3.24}$$

Then $\boldsymbol{E}_f\,[\hat{f}_n(\xi^n; x) - f(x)]^2 \to 0$, i.e., *mean square consistency. If the (stronger) condition*

$$\sum_{n=1}^{\infty} \frac{1}{n^2 b_n} < \infty \tag{3.25}$$

is satisfied, $\boldsymbol{P}\left[\lim_{n\to\infty} \hat{f}_n(\xi^n, x) = f(x)\right] = 1$ *follows.*

Remark. Conditions (3.25) and (3.23) do not imply each other.

3.14. For the RPE (3.1) and the DE* (3.16) analogous results are valid (see [18] and [17], respectively) under corresponding conditions (in particular assumptions of Lemma 3.7). For the RPE weak pointwise consistency holds if (b_n) satisfies (3.2), strong pointwise consistency if additionally $\sum_{n=1}^{\infty} e^{-\alpha n b_n} < \infty$ for all $\alpha > 0$. If (b_n) satisfies $\sum_{n=1}^{\infty} b_n = \infty$, strong pointwise consistency holds for the DE* (3.16).

3.15. Application of a lemma of Glick ([23]), which is a generalization of Scheffé's convergence theorem ([38]), yields:

COROLLARY ([17]). *Let the conditions of Lemma 3.7 be fulfilled and $\hat{f}_n$ the* WWYE; *for (b_n) we assume $b_n \to 0$. If (3.24) holds, then*

$$\int |\hat{f}_n(\xi^n; x) - f(x)|\, dx \to 0 \qquad \text{in probability.}$$

If (3.23) or (3.25) is valid, then

$$\boldsymbol{P}\big[\lim_{n\to\infty} \int |\hat{f}_n(\xi^n; x) - f(x)|\, dx = 0\big] = 1 \tag{3.26}$$

follows.

Remark. The use of the L_1-norm as a loss is quite natural when estimating densities globally: if P and Q are probability measures and f and g their Radon–Nikodym densities with respect to a dominating measure μ, then

$$\sup_A |P(A) - Q(A)| = \tfrac{1}{2} \int |f-g|\, d\mu.$$

Hence (3.26) can be interpreted in the following way: if P_f denotes the "true" probability distribution and $P_{\hat{f}(\xi^n;\cdot)}$ the estimator defined by $P_{\hat{f}(\xi^n;\cdot)}(A) := \int_A \hat{f}(\xi^n; x)\, d\mu(x)$, then

$$\boldsymbol{P}\big[\lim_{n\to\infty} \sup_A |P_{\hat{f}(\xi^n;\cdot)}(A) - P_f(A)| = 0\big] = 1.$$

3.16. A recursive series estimator. The series method for density estimation has been established by Čencov ([69]). Rutkowski gives the following recursive version: assume $f \in L_2$ has a Fourier series with respect to an orthonormal base $\{\varphi_j\}$:

$$f(x) \sim \sum_{j=0}^{\infty} a_j \cdot \varphi_j(x) \qquad \text{with} \quad a_j = \boldsymbol{E}_f\, \varphi_j(\xi).$$

Then

$$\hat{f}_n(x_1, \ldots, x_n; x) := \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{k(i)} \varphi_j(x_i) \cdot \varphi_j(x),$$

where $(k(i))$ is a truncation sequence with $\lim_{i\to\infty} k(i) = \infty$, defines a recursive estimator.

Assuming $|\varphi_j(x)| \leqslant c_j$ for all x and every index j, convergence almost everywhere of these estimators is proved in [37] under the following condition:

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \big[\sum_{j=0}^{k(n)} c_j^2\big]^2 < \infty.$$

4. Pointwise asymptotic properties of kernel estimators

4.1. For a more detailed analysis of the asymptotic behaviour of the estimators discussed in the preceding chapter, more stringent conditions have to be imposed on f. As before, we treat also the RPE in order to compare it with the recursive estimators.

In the following we shall consider mainly two sets of assumptions:

$$\begin{aligned} &f \text{ is differentiable, } f' \text{ is essentially bounded,} \\ &K,\ z \mapsto zK(z) \text{ integrable, } \textstyle\int zK(z)\,dz = 0; \end{aligned} \tag{4.1}$$

and

$$\begin{aligned} &f \text{ is twice differentiable, } f'' \text{ essentially bounded,} \\ &K,\ z \mapsto z^2 K(z) \text{ integrable, } K \text{ even.} \end{aligned} \tag{4.2}$$

4.2. LEMMA. *Under condition* (4.1),

$$E_f\left[\frac{1}{b_k} K\left(\frac{x-\xi}{b_k}\right)\right] - f(x)\cdot \int K(z)\,dz = o(b_k) \qquad as \quad k \to \infty \tag{4.3}$$

and

$$b_k V_f\left[\frac{1}{b_k} K\left(\frac{x-\xi}{b_k}\right)\right] - f(x)\cdot \int K^2(z)\,dz + b_k f^2(x) = o(b_k) \tag{4.4}$$

If the condition $\int zK(z)\,dz = 0$ *is dropped, we get* $b_k \|f'\|_\infty \cdot \int |zK(z)|\,dz = O(b_k)$ *as a bound in* (4.3) *and*

$$b_k \|f'\|_\infty \cdot \int |zK^2(z)|\,dz + O(b_k^2) \qquad in\ (4.4).$$

Under condition (4.2)

$$E_f\left[\frac{1}{b_k} K\left(\frac{x-\xi}{b_k}\right)\right] - f(x)\cdot \int K(z)\,dz = \tfrac{1}{2} b_k^2 f''(x)\cdot \int z^2 K(z)\,dz + o(b_k^2) \tag{4.5}$$

and

$$\begin{aligned} b_k V_f\left[\frac{1}{b_k} K\left(\frac{x-\xi}{b_k}\right)\right] &- f(x) \int K^2(z)\,dz + b_k f^2(x) \\ &= \tfrac{1}{2} b_k^2 f''(x)\cdot \int z^2 K^2(z)\,dz + o(b_k^2). \end{aligned} \tag{4.6}$$

Anyhow,

$$V_f\left[\frac{1}{b_k} K\left(\frac{x-\xi}{b_k}\right)\right] \sim \frac{f(x)}{b_k} \int K^2(z)\,dz.$$

Remark. In view of the application of this lemma, it should be remarked that improvement of the order of convergence by imposing stronger differentiability conditions on f is only possible if the nonnegativeness of the kernel K is dropped. This is obvious by (4.5).

4.3. We further need the following

LEMMA. *Let $K \geqslant 0$ and $\|K\|_\infty < \infty$, x and y points in $\mathbf{R}$, $x \neq y$, $b_j \to 0$, and one of the following conditions be fulfilled:*

(1) *f is essentially bounded;*

(2) $\lim\limits_{z \to \pm\infty} zK(z) = 0$;

(3) *$f \in L_2$, x and y are Lebesgue points of f,* $\int\limits_0^\infty \sup\limits_{|u| \geqslant z} K(u)\,dz < \infty$.

Then

$$\lim_{k \to \infty} \mathbf{E}_f\left[\frac{1}{b_k} K\left(\frac{x-\xi}{b_k}\right) \cdot K\left(\frac{y-\xi}{b_k}\right)\right] = 0.$$

Proof. We choose $0 < \varepsilon < |x-y|/2$. Then obviously,

$$\mathbf{E}\left[\frac{1}{b_k} K\left(\frac{x-\xi}{b_k}\right) \cdot K\left(\frac{y-\xi}{b_k}\right)\right]$$

$$= (*) = \int\limits_{|z-x| > \varepsilon} + \int\limits_{\substack{|z-x| \leqslant \varepsilon \\ |z-y| > \varepsilon}} \frac{1}{b_k} K\left(\frac{x-z}{b_k}\right) \cdot K\left(\frac{y-z}{b_k}\right) \cdot f(z)\,dz.$$

Now

$$I_k(\varepsilon) := \int\limits_{|z-x| > \varepsilon} \frac{1}{b_k} K\left(\frac{x-z}{b_k}\right) \cdot K\left(\frac{y-z}{b_k}\right) \cdot f(z)\,dz$$

$$\leqslant \|K\|_\infty \cdot \|f\|_\infty \int\limits_{|z-x| > \varepsilon} \frac{1}{b_k} K\left(\frac{x-z}{b_k}\right) dz \to 0$$

in case (1), because $z \mapsto \dfrac{1}{b_k} K\left(\dfrac{z}{b_k}\right)$ is an approximate identity.

In case (2) we write

$$I_k(\varepsilon) = \int\limits_{|z-x| > \varepsilon} \frac{1}{|x-z|} \cdot \frac{|x-z|}{b_k} \cdot K\left(\frac{x-z}{b_k}\right) \cdot K\left(\frac{y-z}{b_k}\right) \cdot f(z)\,dz$$

$$\leqslant \frac{1}{\varepsilon} \cdot \|K\|_\infty \cdot \sup_{|z-x| > \varepsilon} \left[\frac{|x-z|}{b_k} \cdot K\left(\frac{x-z}{b_k}\right)\right] \cdot \int\limits_{|z-x| > \varepsilon} f(z)\,dz$$

$$\leqslant \frac{1}{\varepsilon} \cdot \|K\|_\infty \cdot \sup_{|t| > \varepsilon/b_k} |t|\,K(t) \to 0 \qquad \text{by assumption.}$$

In case (3),

$$[I_k(\varepsilon)]^2 \leqslant \int_{|z-x|>\varepsilon} \frac{1}{b_k}\cdot K^2\left(\frac{x-z}{b_k}\right)dz\cdot \int_{|z-x|>\varepsilon} \frac{1}{b_k}\cdot K^2\left(\frac{y-z}{b_k}\right) f^2(z)\,dz.$$

The first integral tends to 0 because K is integrable and bounded almost everywhere, and the second one converges to $f^2(y)\int K^2(z)\,dz$ by Lemma 3.7. The second integral in $(*)$ is treated similarly.

4.4. In what follows we shall always assume (3.2) for (b_n) and (3.3) for K, further $\|K\|_\infty < \infty$.

THEOREM. Asymptotic unbiasedness.

Under condition (4.1) *we have*

$$E_f \hat{f}_n(\xi^n; x) - f(x) = o(b_n) \quad \textit{for the } \mathrm{RPE};$$

$$= o\left(\frac{\sum_{i=1}^n b_i^2 H(b_i)}{\sum_{i=1}^n b_i H(b_i)}\right), \quad \textit{provided } \sum_{k=1}^\infty b_k^2 H(b_k) = \infty \textit{ for the } \mathrm{DE};$$

$$= o\Big(\sum_{i=1}^n b_i/n\Big), \quad \textit{provided } \sum_{k=1}^\infty b_k = \infty \textit{ for the } \mathrm{WWYE}.$$

(The remarks of Lemma 4.2 hold correspondingly. For the big-O versions the additional conditions about b_k for the DE and WWYE resp. are superfluous.)

Under conditions (4.2) *the following is valid*:

$$E_f \hat{f}_n(\xi^n; x) - f(x) = \tfrac{1}{2} f''(x)\cdot \int z^2 K(z)\,dz \cdot b_n^2 + o(b_n^2) \qquad \textit{for the } \mathrm{RPE};$$

$$= \tfrac{1}{2} f''(x)\cdot \int z^2 K(z)\,dz \cdot \frac{\sum_{i=1}^n b_i^3 H(b_i)}{\sum_{i=1}^n b_i H(b_i)}\cdot[1+o(1)],$$

$$\textit{provided } \sum_{k=1}^\infty b_k^3 H(b_k) = \infty \textit{ for the } \mathrm{DE};$$

$$= \tfrac{1}{2} f''(x)\cdot \int z^2 K(z)\,dz \cdot \frac{\sum_{i=1}^n b_i^2}{n}\cdot[1+o(1)],$$

$$\textit{provided } \sum_{k=1}^\infty b_k^2 = \infty \textit{ for the } \mathrm{WWYE}.$$

(For sake of brevity, we shall write in the following $c \cdot a_n + o$ instead of $c \cdot a_n + o(a_n)$).

4.5. THEOREM. Asymptotic variance.
Let (4.1) *be valid. Then*

$$V_f \hat{f}_n(\xi^n; x)$$

$$= \frac{1}{nb_n} \cdot f(x) \cdot \int K^2(z)\,dz - \frac{1}{n} f^2(x) + o\left(\frac{1}{n}\right) \quad \textit{for the } \text{RPE};$$

$$= \frac{\sum_{i=1}^{n} b_i H^2(b_i)}{[\sum_{i=1}^{n} b_i H(b_i)]^2} \cdot f(x) \cdot \int K^2(z)\,dz - \frac{\sum_{i=1}^{n} b_i^2 H^2(b_i)}{[\sum_{i=1}^{n} b_i H(b_i)]^2} \cdot f^2(x) + o,$$

$$\textit{provided } \sum_{k=1}^{\infty} b_k^2 H^2(b_k) = \infty \textit{ for the } \text{DE};$$

$$= \frac{1}{n^2} \cdot \sum_{i=1}^{n} \frac{1}{b_i} \cdot f(x) \cdot \int K^2(z)\,dz - \frac{1}{n} \cdot f^2(x) + o \textit{ for the } \text{WWYE}.$$

Let (4.2) *be satisfied. Then*

$$V_f \hat{f}_n(\xi^n; x) =$$

$$= \frac{1}{nb_n} \cdot f(x) \cdot \int K^2(z)\,dz - \frac{1}{n} \cdot f^2(x) + \frac{b_n}{2n} \cdot f''(x) \cdot \int z^2 K^2(z)\,dz + o\left(\frac{b_n}{n}\right)$$

$$\textit{for the } \text{RPE};$$

$$= [\sum_{i=1}^{n} b_i H(b_i)]^{-2} \cdot [\sum_{k=1}^{n} b_k H^2(b_k) \cdot f(x) \cdot \int K^2(z)\,dz - \sum_{k=1}^{n} b_k^2 H^2(b_k) \cdot f^2(x) +$$

$$+ \sum_{k=1}^{n} b_k^3 H^2(b_k) \cdot \tfrac{1}{2} f''(x) \cdot \int z^2 K^2(z)\,dz] + o,$$

$$\textit{provided } \sum_{k=1}^{\infty} b_k^3 H^2(b_k) = \infty, \textit{ for the } \text{DE};$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \frac{1}{b_i} \cdot f(x) \cdot \int K^2(z)\,dz - \frac{1}{n} \cdot f^2(x) + \frac{1}{n^2} \sum_{k=1}^{n} b_k \cdot \tfrac{1}{2} f''(x) \cdot \int z^2 K^2(z)\,dz +$$

$$+ o\left(\frac{1}{n^2} \sum_{k=1}^{n} b_k\right), \quad \textit{provided } \sum_{k=1}^{\infty} b_k = \infty, \textit{ for the } \text{WWYE}.$$

4.6. Combination of Theorems 4.4 and 4.5 yields expressions for the mean square error (MSE):

THEOREM. *Let (4.2) be satisfied. We introduce the notation*

$$c_1 := \left[\tfrac{1}{2} f''(x)\cdot \int z^2 K(z)\,dz\right]^2 \qquad \textit{and} \qquad c_2 := f(x)\cdot \int K^2(z)\,dz.$$

Then

$$\mathrm{MSE} = E_f\,[\hat{f}_n(\xi^n;\, x) - f(x)]^2 =$$

$$= b_n^4\cdot c_1 + (nb_n)^{-1} c_2 - \frac{1}{n}\cdot f^2(x) + o \qquad \textit{for the } \mathrm{RPE};$$

$$\mathrm{MSE} = \left[\frac{\sum_{k=1}^{n} b_k^3 H(b_k)}{\sum_{k=1}^{n} b_k H(b_k)}\right]^2 c_1 +$$

$$+ \left[\sum_{k=1}^{n} b_i H(b_i)\right]^{-2}\cdot\left[\sum_{k=1}^{n} b_k H^2(b_k)\cdot c_2 - f^2(x)\cdot \sum_{k=1}^{n} b_k^2 H^2(b_k)\right] + o$$

$$\textit{provided } \sum_{k=1}^{\infty} b_k^3 H^2(b_k) = \infty,\ \sum_{k=1}^{\infty} b_k^3 H(b_k) = \infty \ \textit{for the } \mathrm{DE};$$

$$\mathrm{MSE} = \left(\sum_{k=1}^{n} \frac{b_k^2}{n^2}\right)\cdot c_1 +$$

$$+ \frac{1}{n^2}\sum_{k=1}^{n} \frac{1}{b_k}\cdot c_2 - \frac{1}{n}\cdot f^2(x) + o\left(\frac{1}{n}\right) + o\left(\left(\frac{\sum_{k=1}^{n} b_k^2}{n}\right)^2\right)$$

$$\textit{provided } \sum_{k=1}^{\infty} b_k^2 = \infty, \ \textit{for the } \mathrm{WWYE}.$$

Remark. Under different conditions, expansions of this type have been given by Deheuvels, see [15].

4.7. Now we are going to discuss the (asymptotically) optimal choice of the bandwidth b_n, the function H and the kernel K for the types of estimators considered before. We restrict b_k to have the form $b_k = \gamma\cdot k^{-\beta}$ ($\gamma > 0$, $\beta > 0$) and H to have the form

$$H(b_k) = b_k^{\varrho} = \gamma^{\varrho} k^{-\beta\varrho} \qquad (\varrho \in \mathbf{R}).$$

(i) *The RPE*

$$\mathrm{MSE} = c_1\cdot\gamma^4\cdot n^{-4\beta} + c_2\cdot\gamma^{-1}\cdot n^{-(1-\beta)} + o(n^{-4\beta}) + o(n^{-(1-\beta)}).$$

The best rate of convergence is achieved by taking $\beta = 1/5$, the best choice of γ is $\gamma = (c_2/(4c_1))^{1/5}$, hence the best rate of the MSE is $c_1^{1/5}\cdot c_2^{4/5}\cdot 5\cdot 2^{-8/5}\cdot n^{-4/5} = 1{,}65938\cdot c_1^{1/5}\cdot c_2^{4/5}\cdot n^{-4/5}$ (see [36]; the factor "4" in Rosenblatt's formula (21) has to be deleted, in expression (22) "$2^{3/5}$" is superfluous). Now the problem remains to minimize $c_1^{1/5}\cdot c_2^{4/5}$ by appropriate choice of K. In order to make the problem well-defined we only take kernels K, fulfilling the norming condition $\int z^2 K(z)\,dz = 1$ into consideration. Then the minimum of $K \mapsto \int K^2(z)\,dz$ is attained at

$$K_0(z) = \begin{cases} \dfrac{3}{4\sqrt{5}}\left(1-\dfrac{z^2}{5}\right) & \text{for } |z| \leqslant \sqrt{5}, \\ 0 & \text{otherwise} \end{cases}$$

(see [19], [36]), and this yields the solution for minimizing the MSE.

(ii) *The* DE

$$\mathrm{MSE} = c_1\cdot\gamma^4\cdot\left[\frac{1-\beta(1+\varrho)}{1-\beta(3+\varrho)}\right]^2\cdot n^{-4\beta}+c_2\cdot\gamma^{-1}\cdot\frac{[1-\beta(1+\varrho)]^2}{1-\beta(1+2\varrho)}\cdot n^{-1+\beta}+o$$

Again $\beta = 1/5$ yields the best rate of convergence, and $\gamma = \left(\dfrac{c_2}{c_1}\cdot\dfrac{2-\varrho}{40}\right)^{1/5}$ is the optimal choice of γ, giving a

$$\mathrm{MSE} = 5\cdot 40^{-4/5}\cdot\frac{(4-\varrho)^2}{(2-\varrho)^{6/5}}\cdot c_1^{1/5}\cdot c_2^{4/5}\cdot n^{-4/5}.$$

The optimal ϱ is $\varrho = -1$, hence the WWYE is the best of the DE's with respect to local MSE (cf. [16]). The optimal kernel is again K_0.

For the optimal value of the MSE of the RPE and DE's we can write $\Delta\cdot c_1^{1/5}\cdot c_2^{4/5}\cdot n^{-4/5}$; here

$$\Delta = \begin{cases} 1{,}66 & \text{for the RPE,} \\ 1{,}75 & \text{for the WWYE,} \\ 1{,}82 & \text{for the DE* (3.16) } (H \equiv 1). \end{cases}$$

It is not surprising that the RPE yields the best results. On the other hand, the superiority of the WWYE to the DE* does not contradict the remark in 3.10: whereas the variance of DE* (3.16) is asymptotically smaller than the variance of WWYE (3.15), the MSE behaviour of the WWYE is better.

4.8. A law of the iterated logarithm.

THEOREM. *Let $f(x) > 0$. Let (3.19) be fulfilled,*

$$\lim_{n\to\infty}\frac{\log\log\{(nb_n)^{-1}[\sum_{k=1}^{n} b_k H(b_k)]^2\}}{\log\log n} = \delta \tag{4.7}$$

exist with $0 < \delta < \infty$ and

$$\lim_{n\to\infty} nb_n\cdot[\sum_{k=1}^{n} b_k H(b_k)]^{-2}\cdot H^2(b_n)\cdot\log\log\frac{[\sum_{k=1}^{n} b_k H(b_k)]^2}{nb_n} = 0. \tag{4.8}$$

Then for the DE,

$$P\left[\limsup_{n\to\infty}\left(\frac{nb_n}{\log\log n}\right)^{1/2}\cdot\frac{\hat{f}_n(\xi^n;x)-E_f\hat{f}_n(\xi^n;x)}{(2\alpha\delta f(x)\cdot\int K^2(z)\,dz)^{1/2}}=1\right]=1.$$

If additionally (4.1) *and*

$$\left(\frac{nb_n}{\log\log n}\right)^{1/2}\cdot\frac{\sum_{k=1}^{n} b_k^2 H(b_k)}{\sum_{k=1}^{n} b_k H(b_n)}\quad\textit{is bounded,}\qquad \sum_{k=1}^{\infty} b_k^2 H(b_k)=\infty; \tag{4.9}$$

or (4.2) *and*

$$\lim_{n\to\infty}\left(\frac{nb_n}{\log\log n}\right)^{1/2}\cdot\frac{\sum_{k=1}^{n} b_k^3 H(b_k)}{\sum_{k=1}^{n} b_k H(b_k)}=0, \tag{4.10}$$

are fulfilled, then for the DE *even*

$$P\left[\limsup_{n\to\infty}\left(\frac{nb_n}{\log\log n}\right)^{1/2}\cdot\frac{\hat{f}_n(\xi^n;x)-f(x)}{(2\alpha\delta f(x)\cdot\int K^2(z)\,dz)^{1/2}}=1\right]=1 \tag{4.11}$$

holds.

For the WWYE, *the conditions are*: (3.21), *further*:

$$\lim_{n\to\infty}\frac{\log\log(n/b_n)}{\log\log n}=\delta\qquad(0<\delta<\infty); \tag{4.7}$$

$$\lim_{n\to\infty}(nb_n)^{-1}\log\log(n/b_n)=0; \tag{4.8}$$

$$\left(\frac{b_n}{n\log\log n}\right)^{1/2}\cdot\sum_{k=1}^{n} b_k\quad\textit{is bounded and}\quad\sum_{k=1}^{\infty} b_k=\infty; \tag{4.9}$$

$$\lim_{n\to\infty}\left(\frac{b_n}{n\log\log n}\right)^{1/2}\cdot\sum_{k=1}^{n} b_k^2=0. \tag{4.10}$$

Choosing b_k and H as in 4.7, the restrictions are: $\beta(1+2\varrho)<1$, $\beta(1+\varrho)<1$, $\beta(1-\varrho)<1$ and $\beta\geqslant 1/3$ if (4.9) whereas $\beta\geqslant 1/5$ if (4.10) is assumed. The optimal choice of β is $\beta=1/3$ for case (4.1)+(4.9) and $\beta=1/5$ for case (4.2)+(4.10). In all these cases, $\delta=1$ and $\alpha=\dfrac{[1-\beta(1+\varrho)]^2}{1-\beta(1+2\varrho)}$. Hence (4.11) has the following form:

$$P\left[\limsup_{n\to\infty}\left(\frac{\gamma\cdot n^{1-\beta}}{\log\log n}\right)^{1/2}\frac{\hat{f}_n(\xi^n;x)-f(x)}{\left\{2\cdot\dfrac{[1-\beta(1+\varrho)]^2}{1-\beta(1+2\varrho)}\cdot f(x)\cdot\int K^2(z)\,dz\right\}^{1/2}}=1\right]=1. \tag{4.12}$$

(Under slightly different conditions a similar theorem is proved in [15]).

Remarks. 1. This theorem gives the *exact rate of convergence* almost surely; The optimal choice of b_n is $\gamma \cdot n^{-1/3}$ and $\gamma \cdot n^{-1/5}$ in case (4.1) and (4.2) respectively. (4.12) is independent of the choice of γ, hence great γ give quicker convergence. This is due to the fact that the factor "$\log\log n$" makes the bias term asymptotically negligible and, according to 4.5, great values of b_n give a good convergence of $\hat{f}_n - E\hat{f}_n$. The optimal value of β is the same as for the MSE (cf. 4.7); hence a resonable choice of γ is, e.g., to make the MSE asymptotically minimal.

2. The condition $f(x) > 0$ is essential to get an exact rate of convergence; it is easy to see that arbitrary good rates of convergence are available if f vanishes in a neighbourhood of x. Using Theorem 4.5, however, in case $f''(x) \neq 0$ an exact rate of convergence can be established.

3. In [58] a similar result is claimed for the WWYE under several technical assumptions concerning (b_n); but in fact, the method of proof of [58], using a strong approximation theorem applied to an auxiliary estimator, yields a weaker result than (4.12) (namely only an inequality). Further, for an estimate of the bias term $E\hat{f}_n - f$, the authors use a condition about the characteristic function of f, which essentially implies f'' to exist and to be bounded. (Obviously, a relation as (4.11) with "lim" instead of "lim sup" (as claimed in [58]) cannot hold in general).

4. There are results about the exact rate of convergence of $\sup_{x \in I} |\hat{f}_n(\xi^n; x) - f(x)|/\sqrt{f(x)}$ (I a compact interval) for nonrecursive estimators (e.g., [10], Theorem 6.2.5, [48], [49], [13]). The author has not yet been successful in proving a pointwise result like (4.11) for the RPE, but conjectures that in this case (4.11) holds with $\alpha = \delta = 1$.

4.9. Asymptotic normality

THEOREM. *Let* $f(x) > 0$. *Then for the* RPE

$$\lim_{n\to\infty} \sup_{y \in \mathbf{R}} \left| \mathbf{P}\left\{ \frac{\sqrt{nb_n}\,[\hat{f}_n(\xi^n; x) - E_f \hat{f}_n(\xi^n; x)]}{[f(x) \cdot \int K^2(z)\,dz]^{1/2}} \leqslant y \right\} - \Phi(y) \right| = 0 \qquad (4.13)$$

holds. (Φ *denotes the distribution function of the normal distribution* $\mathcal{N}(0, 1)$). *If*

$$\lim_{n\to\infty} (nb_n)^{1+\delta/2} \frac{\sum_{k=1}^{n} b_k H^{2+\delta}(b_k)}{[\sum_{k=1}^{n} b_k H(b_k)]^{2+\delta}} = 0 \quad \textit{for some} \quad \delta > 0 \qquad (4.14)$$

and (3.19) *hold, then* (4.13) *is true for the* DE.

For the WWYE, (4.14) *is equivalent to*

$$\lim_{n\to\infty} (b_n/n)^{1+\delta} \cdot \sum_{k=1}^{n} b_k^{-(1+\delta)} = 0 \qquad \textit{for some} \quad \delta > 0.$$

If moreover, the conditions of Lemma 4.3 *are satisfied and* $x_1, \ldots, x_k$ *are distinct points with* $f(x_j) > 0$, *then* $(\eta_1, \ldots, \eta_k) \xrightarrow{\mathscr{L}} \mathcal{N}(\mathrm{o}, I)$. *Here* $\mathcal{N}(\mathrm{o}, I)$ *denotes the k-dimensional normal distribution with mean vector* o *and unit covariance matrix I and*

$$\eta_j := \frac{\sqrt{nb_n}\,[\hat{f}_n(\xi^n; x_j) - E_f \hat{f}_n(\xi^n; x_j)]}{[f(x_j) \cdot \int K^2(z)\,dz]^{1/2}},$$

$\hat{f}_n(\xi^n; x)$ *the* RPE. *If additionally* $\sum_{k=1}^{\infty} b_k H^2(b_k) = \infty$ *is valid, then*

$$\left(\frac{\eta_1}{\alpha}, \ldots, \frac{\eta_k}{\alpha}\right) \xrightarrow{\mathscr{L}} \mathcal{N}(\mathrm{o}, I)$$

holds for the DE.

If additionally (4.1) *is satisfied and* nb_n^3 *bounded or* (4.2) *fulfilled and* $\lim_{n\to\infty} nb_n^5 = 0$, *then for the* RPE

$$\lim_{n\to\infty} \sup_{y\in\mathbf{R}} \left| \boldsymbol{P}\left\{ \sqrt{nb_n} \frac{\hat{f}_n(\xi^n; x) - f(x)}{[f(x) \cdot \int K^2(z)\,dz]^{1/2}} \leqslant y \right\} - \Phi(y) \right| = 0$$

holds.

In case of the DE *the following conditions are assumed:*

Let (4.1) *be satisfied,*

$$\sqrt{nb_n} \frac{\sum_{k=1}^{n} b_k^2 H(b_k)}{\sum_{k=1}^{n} b_k H(b_k)} \quad \textit{be bounded} \qquad \textit{and} \qquad \sum_{k=1}^{\infty} b_k^2 H(b_k) = \infty \tag{4.15}$$

or (4.2) *be fulfilled and*

$$\lim_{n\to\infty} \sqrt{nb_n} \frac{\sum_{k=1}^{n} b_k^3 H(b_k)}{\sum_{k=1}^{n} b_k H(b_k)} = 0. \tag{4.16}$$

Then

$$\lim_{n\to\infty} \sup_{y\in\mathbf{R}} \left| \boldsymbol{P}\left\{ \sqrt{nb_n} \frac{\hat{f}_n(\xi^n; x) - f(x)}{[\alpha \cdot f(x) \cdot \int K^2(z)\,dz]^{1/2}} \leqslant y \right\} - \Phi(y) \right| = 0$$

holds.

In case of the WWYE, (4.15) *is equivalent to*

$$(b_n/n)^{1/2}\cdot\sum_{k=1}^{n} b_k \quad \text{bounded} \qquad \text{and} \qquad \sum_{k=1}^{n} b_k = \infty$$

and (4.16) *is equivalent to*

$$(b_n/n)^{1/2}\cdot\sum_{k=1}^{n} b_k \to 0.$$

Under appropriate conditions (cf. in particular Lemma 4.3!) multivariate versions hold also for the DE's.

5. Some further results concerning recursive estimators

5.1. The recursive character of the WWYE suggests to apply the method of stochastic approximation in Hilbert function spaces (see [39], [40] and [53]). Using properties of the Robbins–Monro process, Schmetterer proves the following

THEOREM ([41]). *Let f and K be square integrable densities and $b_n \downarrow 0$. $\|\cdot\|$ denotes L_2-norm. Then*

$$\boldsymbol{P}_f\,[\lim_{n\to\infty}\|\hat{f}_n(\xi^n;\,\cdot)-f(\cdot)\| = 0] = 1 \tag{5.1}$$

and for every $\varepsilon > 0$ there is a constant $C(\varepsilon)$ with

$$\boldsymbol{P}_f\,[\|\hat{f}_n(\xi^n;\,\cdot)-f(\cdot)\| \geqslant \varepsilon] \leqslant e^{-nC(\varepsilon)}. \tag{5.2}$$

(A slightly weaker inequality – with $e^{-nb_nC(\varepsilon)}$ as right hand side of (5.2) – has been proved by Révész ([34]); earlier this inequality has been established by Rejtő and Révész by a direct method under considerably more stringent conditions, see [33]).

5.2. Applying similars ideas, Györfi proves (5.1) for dependent observations, by this generalizing a result in [25]:

THEOREM ([24]). *Let $\ldots,\xi_{-n},\ldots,\xi_{-1},\ \xi_0,\ \xi_1,\ldots,\xi_n,\ldots$ be a sequence of identically distributed random variables with density f. Suppose (ξ_n) to be strictly stationary and ergodic. We denote $Q_n(A) := \boldsymbol{P}[\xi_n\in A|\xi_{n-1},\,\xi_{n-2},\ldots,\xi_1,\,\xi_0,\,\xi_{-1},\ldots]$, assume Q_n to have a density f_n for every n and $\boldsymbol{E}|f_1| < \infty$ (as before, $\|\cdot\|$ is L_2-norm). If K is a square-integrable probability density, $b_n \to 0$, $\sum_{n=1}^{\infty}(nb_n)^{-1} < \infty$ and*

$$\sup_n \frac{1}{n}\cdot\sum_{k=1}^{n-1} k\cdot\int\left|\frac{b_{k+1}}{b_k}K\left(\frac{b_{k+1}}{b_k}x\right)-K(x)\right|dx < \infty, \tag{5.3}$$

then (5.1) *holds for the* WWYE.

(Condition (5.3) is discussed in [24]).

5.3. For simplicity of exposition we only mention the work about *uniform consistency* of the WWYE and DE (see [11], [13], [15]). The corresponding results for the RPE are proved by Schuster ([42]).

6. Sequential density estimators

6.1. Apparently the first paper dealing explicitly with sequential density estimators is [45]; however, the stopping rules considered there are independent of the observations and, as Carroll ([6]) remarks, several details are not correct.

Later Davies and Wegman ([12]) consider stopping rules depending on the observations, in the sense of the definitions of Chapter 1; they define

$$\nu_\varepsilon := \inf\{nM\colon |\hat{f}_{nM}(\xi^{nM}; x) - \hat{f}_{(n-1)M}(\xi^{(n-1)M}; x)| < \varepsilon\}, \tag{6.1}$$

where $\varepsilon > 0$ and $M \in N$ are fixed and $\hat{f}_n$ denotes the RPE. In [12] a strong consistency theorem is proved under rather stringent conditions:

$$\boldsymbol{P}_f\left[\lim_{\varepsilon \downarrow 0} \hat{f}_{\nu_\varepsilon}(\varepsilon^{\nu_\varepsilon}; x) = f(x)\right] = 1. \tag{6.2}$$

6.2. We start from the stopping rule (6.1) (for simplicity with $M = 1$) and apply it to a recursive estimator, say the WWYE. Then

$$\hat{f}_n(x^n; x) - \hat{f}_{n-1}(x^{n-1}; x) = \frac{1}{nb_n} K\left(\frac{x - x_n}{b_n}\right) - \frac{1}{n(n-1)} \cdot \sum_{k=1}^{n-1} \frac{1}{b_k} K\left(\frac{x - x_k}{b_k}\right);$$

the first term is of order $(nb_n)^{-1}$, the second one $\sim \frac{1}{n} f(x)$ and hence asymptotically small; this leads to the new stopping rule

$$\nu_\varepsilon := \inf\left\{n\colon \frac{1}{nb_n} K\left(\frac{x - \xi_n}{b_k}\right) < \varepsilon\right\} \tag{6.3}$$

which is unsatisfactory intuitively, in so far as it depends only on one observation. Nevertheless the following can be proved for the WWYE:

THEOREM ([58]). *Let* $\lim_{n\to\infty} b_n = 0$, K *satisfy* (3.3), $\|K\|_\infty < \infty$ *and* $\lim_{|t|\to\infty} tK(t) = 0$. *Then* (6.3) *defines a stopping variable,* $\boldsymbol{E}\nu_\varepsilon^k < \infty$ *for all* $k \in N$ *and* $\boldsymbol{E}e^{t\nu_\varepsilon} < \infty$ *for all* $t \in \boldsymbol{R}$. *Further, if* $K > 0$, $\lim_{\varepsilon\downarrow 0} \nu_\varepsilon = \infty$ *almost surely and, under the assumptions of Theorem* 3.12, (6.2) *holds.*

6.3. The following lemma is a slight generalization of a theorem due to Chow and Robbins ([7]) and enables one to construct a variety of stopping rules, in particular the stopping variable (6.13):

LEMMA. *Let* $G\colon \boldsymbol{N}\times(0,\infty)\to\boldsymbol{R}$ *be a function such that* $k\mapsto G(k,\varepsilon)$ *is monotonically increasing for every* $\varepsilon>0$; $\varepsilon\mapsto G(k,\varepsilon)$ *is monotonically decreasing for every* $k\in\boldsymbol{N}$; $\lim_{k\to\infty} G(k,\varepsilon)=\infty$ *and* $\lim_{k\to\infty}\dfrac{G(k,\varepsilon)}{G(k-1,\varepsilon)}=1$ *for every* $\varepsilon>0$; $\lim_{\varepsilon\to 0} G(k,\varepsilon)\leqslant 0$ *for every* $k\in\boldsymbol{N}$.

Let (η_k) *be a sequence of random variables such that* $\eta_k>0$ *almost surely and* $\boldsymbol{P}([\lim_{k\to\infty}\eta_k=\gamma])=1$ *with an appropriate* $\gamma>0$ *and* $v_\varepsilon:=\inf\{k\colon \eta_k\leqslant G(k,\varepsilon)\}$.

Then v_ε *is a stopping variable (finite almost surely),* $\varepsilon\to v_\varepsilon$ *is monotonically increasing a.s. if* $\varepsilon\downarrow 0$;

$$\lim_{\varepsilon\downarrow 0} v_\varepsilon=\infty \quad a.s., \qquad \lim_{\varepsilon\downarrow 0}\boldsymbol{E}v_\varepsilon=\infty \qquad and \qquad \lim_{\varepsilon\downarrow 0} G(v_\varepsilon,\varepsilon)=\gamma \quad a.s.$$

6.4. Carroll studies in [6] the problem of estimating the density f at a point x_0 with $f(x_0)>0$, which may be given or may be unknown, e.g., some quantile of the unknown distribution. Using the Chow–Robbins approach, fixed-width sequential confidence intervals for $f(x_0)$ are established.

THEOREM ([6]). *Let* K *be a bounded probability density fulfilling a Lipschitz condition of order* 1. *Let the following conditions be satisfied:* $b_n\to 0$, $nb_n^2\to\infty$,

for every $c>0$ *there is a constant* $M(c)>0$ *such that*

$$\boldsymbol{E}_f\left|K\left(\frac{x_0-\xi+t_1a_n}{b_{ns_1}}\right)-K\left(\frac{x_0-\xi+t_1a_n}{b_{ns_2}}\right)-K\left(\frac{x_0-\xi+t_2a_n}{b_{ns_1}}\right)+\right.$$
$$\left.+K\left(\frac{x_0-\xi+t_2a_n}{b_{ns_2}}\right)\right|^r$$
$$\leqslant M(c)\cdot b_n\cdot|s_2-s_1|^{r/2}\cdot|t_2-t_1|^{r/2} \qquad \textit{for each } s_1, s_2\geqslant c$$
$$\textit{with } ns_1,\ ns_2\in\boldsymbol{N} \textit{ and for } r=1,2,3,4; \tag{6.4}$$

$$\lim_{\substack{n\to\infty\\ u\to s}}\frac{1}{b_n}\boldsymbol{E}_f\left[K\left(\frac{x_0-\xi+t\alpha_n}{b_{[ns]}}\right)-K\left(\frac{x_0-\xi}{b_{[nu]}}\right)\right]=0. \tag{6.5}$$

Let (h_n) *be a sequence of estimators of* x_0 *such that*

$$h_n=x_0+o(a_n) \textit{ as } n\to\infty \textit{ almost surely}, \qquad \lim_{n\to\infty} a_n/b_n=0 \tag{6.6}$$

and v_ε *such that*

$$\lim_{\varepsilon\downarrow 0}\frac{v_\varepsilon}{n_\varepsilon}=1 \textit{ for a sequence } (n_\varepsilon) \textit{ of constants with } \lim_{\varepsilon\downarrow 0} n_\varepsilon=\infty. \tag{6.7}$$

Then for the RPE $\hat{f}_n$ *we have*

$$\sqrt{v_\varepsilon \cdot b_{v_\varepsilon}} \cdot \left[\hat{f}_{v_\varepsilon}(\xi^{v_\varepsilon}; h_{v_\varepsilon}(\xi^{v_\varepsilon})) - \frac{1}{b_{v_\varepsilon}} \cdot \int K\left(\frac{h_{v_\varepsilon}(\xi^{v_\varepsilon}) - y}{b_{v_\varepsilon}}\right) \cdot f(y)\, dy \right] \xrightarrow{\mathscr{L}} \mathscr{N}\left(0, f(x_0) \int K^2(z)\, dz\right).$$

6.5. A similar result holds for the WWYE:

THEOREM ([6]). *Let the conditions of Theorem* 4.9, *regarding the asymptotic normality of the* WWYE *and conditions* (6.6) *and* (6.7) *be satisfied. We further assume*: $nb_n^2 \to \infty$, *and*

$$K \text{ is Lipschitz of order } 1 \quad \text{and} \quad \lim_{n\to\infty} a_n/b_n = 0$$

or

K is continuously differentiable with bounded derivative K',

$$\lim_{n\to\infty} a_n^4/b_n^5 = 0 \quad \text{and} \quad \limsup_{n\to\infty} \int [K'(y(1+c_n))]^{2r} \cdot f(x - yc_n)\, dy < \infty$$

for every sequence $c_n \to 0$ *and for* $r = 1, 2$;

further $a_n/(nb_n^2) = O(n^{-\delta})$ *for some* $\delta > 0$ *and* $\lim_{n\to\infty} na_n^2 b_n = 0$.

Then

$$\sqrt{v_\varepsilon \cdot b_{v_\varepsilon}} \cdot \left[\hat{f}_{v_\varepsilon}(\xi^{v_\varepsilon}; h_{v_\varepsilon}(\xi^{v_\varepsilon})) - \frac{1}{v_\varepsilon} \sum_{k=1}^{v_\varepsilon} \frac{1}{b_k} \cdot \int K\left(\frac{h_{v_\varepsilon}(\xi^{v_\varepsilon}) - y}{b_k}\right) f(y)\, dy \right] \xrightarrow{\mathscr{L}} \mathscr{N}\left(0, \alpha \cdot f(x_0) \cdot \int K^2(z)\, dz\right).$$

6.6. COROLLARY ([6]). *Let* (4.2) *be satisfied*, f'' *bounded*, $\lim_{n\to\infty} nb_n a_n^2 = 0$ *and in case of the* RPE $\lim_{n\to\infty} nb_n^5 = 0$, *whereas in case of the* WWYE, $\lim_{n\to\infty} (b_n/n)^{1/2} \sum_{k=1}^{n} b_k = 0$ *is assumed* (cf. (4.16)!).

Let further the conditions of Theorem 6.3 *and* 6.4 *respectively be satisfied. Then*

$$\frac{\sqrt{v_\varepsilon b_{v_\varepsilon}} \cdot [\hat{f}_{v_\varepsilon}(\xi^{v_\varepsilon}; h_{v_\varepsilon}(\xi^{v_\varepsilon})) - f(x_0)]}{[\alpha \cdot f(x_0) \cdot \int K^2(z)\, dz]^{1/2}} \xrightarrow{\mathscr{L}} \mathscr{N}(0, 1) \tag{6.8}$$

holds for the RPE *and the* WWYE (*for the first one*, $\alpha := 1$).

Remark. In case x_0 is known, the assumptions, in particular (6.4) and (6.5), are simplified considerably, namely, one takes $a_n \equiv 0$. For simplicity of notation we shall write "$\hat{f}_n(h_n)$" instead of "$\hat{f}_n(\xi^n; h_n(\xi^n))$" etc. from now on.

6.7. As examples for stopping variables ν_ε fulfilling (6.7) we treat a local and a global one (see [6]):

Local stopping rule:

$$\nu_\varepsilon := \inf\{n \geqslant n_0 \colon nb_n \geqslant (c/\varepsilon)^2 \hat{f}_n(\xi^n; h_n(\xi^n))\}$$
$$\text{with} \quad c := \left[\alpha \cdot \textstyle\int K^2(z)\,dz\right]^{1/2} \cdot \Phi^{-1}(1-\tfrac{1}{2}\gamma) \qquad (0<\gamma<1). \tag{6.9}$$

THEOREM ([6]). *Let ν_ε be defined by* (6.9) *and $\hat{f}_n$ be the* RPE *or the* WWYE.

1. *Suppose $P_f[\lim_{n\to\infty} \hat{f}_n(h_n) = f(x_0)] = 1$. Then*

$$P_f\left[\lim_{\varepsilon\downarrow 0} \frac{\nu_\varepsilon}{b_{\nu_\varepsilon}} \cdot \frac{\varepsilon^2}{c^2 f(x_0)} = 1\right] = 1.$$

2. *Let $\lim_{n\to\infty} b_n/b_{n-1} = 1$ and let the conditions of Theorem* 6.4 *or* 6.5 *be satisfied, with exception of* (6.7). *Then* (6.8) *holds for the* RPE *and* WWYE, *respectively. In particular*

$$\lim_{\varepsilon\downarrow 0} P_f[|\hat{f}_{\nu_\varepsilon}(h_{\nu_\varepsilon}) - f(x_0)| \leqslant \varepsilon] = 1-\gamma \tag{6.10}$$

yields a fixed-width sequential confidence interval.

3. *If, additionally to the assumptions of part* 1, $\lim_{n\to\infty} E\hat{f}_n(h_n) = f(x_0)$, *if there is a constant $C>0$ with $\sum_{n=1}^{\infty} P_f[|\hat{f}_n(h_n) - E\hat{f}_n(h_n)| > C] < \infty$ and if $b_n = n^{-\beta}$ $(0<\beta<1)$, then $\lim_{\varepsilon\downarrow 0} E_{\nu_\varepsilon}/\varepsilon = 1$ holds with $n_\varepsilon = (c^2 f(x_0)/\varepsilon^2)^{1/(1-\beta)}$.*

4. *Under the conditions of part* 1 *and for $b_n = n^{-\beta}$ $(0<\beta<1)$,*

$$\frac{f^2(x_0)\left[\alpha \cdot \int K^2(z)\,dz\right]^{1/2}}{n_\varepsilon^{(1-\beta)/2}} \cdot [\nu_\varepsilon^{1-\beta} - n_\varepsilon^{1-\beta}] \xrightarrow{\mathscr{L}} \mathscr{N}(0,1)$$

holds.

Remark. The conclusions of this theorem do not essentially depend on the special form of $\hat{f}_n$; a more general formulation can be found in [6].

Global stopping rule: This stopping rule is of interest when estimating f at its unknown mode.

THEOREM ([6]). *Let f be unimodal, x_0 the unique mode: $f(x_0) = \max_x f(x)$, and let us assume that*

$$P_f[\lim_{n\to\infty} \sup_x \hat{f}_n(\xi^n; x) = f(x_0)] = 1$$

(*sufficient conditions can be found in* [31], [43], [52], [67]). *Let*

$$\nu_\varepsilon := \inf\{n \geqslant n_0 \colon \ nb_n \geqslant (c/\varepsilon)^2 \cdot \sup_x \hat{f}_n(\xi^n; x)\}. \tag{6.11}$$

(*c as before*).

If $b_n = n^{-\beta}$ $(0 < \beta < 1)$ *and* $n_\varepsilon = (c^2 f(x_0)/\varepsilon^2)^{1/(1-\beta)}$, *then*

$$\lim_{\varepsilon \downarrow 0} \frac{\nu_\varepsilon}{n_\varepsilon} = 1 \qquad a.s.$$

If additionally a $C > 0$ *exists such that*

$$\sum_{n=1}^{\infty} P_f[\sup_x |\hat{f}_n(\xi^n; x) - f(x)| > C] < \infty$$

(*sufficient conditions for this are given*, e.g., *in* [43]), *then*

$$\lim_{\varepsilon \downarrow 0} \frac{E\nu_\varepsilon}{n_\varepsilon} = 1.$$

Whereas it is possible to achieve a result like (6.10) also for this stopping variable, nothing is known about the limit distribution of ν_ε until now.

6.8. Under much more pleasant conditions than those used above, Stute derives a fixed-width sequential confidence interval for $f(x_0)$ at a given point x_0:

THEOREM ([47]). *Let* f *be twice differentiable and* $f'(x_0) > 0$, K *be a bounded probability density with compact support and differentiable in the interior of its support, admitting one-sided derivatives at the boundary,* $\int zK(z)\,dz = 0$. *Let* $b_n = n^{-1/5}$, $\varepsilon > 0$, $0 < \gamma < 1$, *further*

$$\delta_n := \left(\frac{\int K^2(z)\,dz}{nb_n}\right)^{1/2} \cdot \frac{[\hat{f}_n(\xi^n; x_0)]^{1/2}}{\hat{f}'_n(\xi^n; x_0)} \cdot \Phi^{-1}(1 - \tfrac{1}{2}\gamma),$$

$$I_n := [\hat{f}_n(x_0 - \delta_n), \hat{f}_n(x_0 + \delta_n)] \qquad and \qquad \nu_\varepsilon := \inf\{n \geqslant 1 \colon \ |I_n| \leqslant 2\varepsilon\}$$

($|I|_n$ *denotes the length of the interval* I_n). *Then*

$$\lim_{\varepsilon \downarrow 0} \frac{\nu_\varepsilon}{\varepsilon^{-5/2}} = \lim_{\varepsilon \downarrow 0} \frac{E\nu_\varepsilon}{\varepsilon^{-5/2}} = f(x_0)^{5/4} \cdot [\textstyle\int K^2(z)\,dz]^{5/4} [\Phi^{-1}(1 - \tfrac{1}{2}\gamma)]^{5/2} \tag{6.12}$$

(*both a.s. and in the mean*) *and*

$$\lim_{\varepsilon \downarrow 0} P[f(x_0) \in I_{\nu_\varepsilon}] = 1 - \gamma.$$

Remark. (6.12) allows to define the notion of efficiency of one kernel K_1 with respect to another one, K_2, under the norming restriction $\int z^2 K_i(z)\,dz = 1$:

$$\mathrm{eff}(K_1 : K_2) := \left[\frac{\int K_2^2(z)\,dz}{\int K_1^2(z)\,dz}\right]^{5/4}.$$

The optimal kernel is K_0 discussed in 4.7.

6.9. The following theorem treats a global stopping variable; the author is sure that it is possible to relax the conditions imposed on the density f considerably.

THEOREM ([62]). *Let*

$$\lim_{n\to\infty} b_n/b_{n-1} = 1, \qquad b_n = o(n^{-2/9}),$$

$$n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4} = o(b_n)$$

(e.g., $b_n = \gamma \cdot n^{-\beta}$ with $\gamma > 0$ and $2/9 < \beta < 1/4$). Let f vanish outside a known compact interval $[a, b]$ and let $f(x) > 0$ for $x \in (a, b)$; let f be twice differentiable, f'' and $f'/\sqrt{f}$ be bounded. Let K be a symmetric probability density, $z \mapsto z^2 K(z)$ be integrable and either: K have compact support $[-A, A]$ and be differentiable in $(-A, A)$ or : K be differentiable, $\int |K'(z)|\,dz < \infty$ and $\int |K'(z)|^2\,dz < \infty$. Let $\hat{f}_n$ be the RPE *and*

$$\nu_\varepsilon := \inf\left\{n \geqslant 1 : \left[\frac{b_n}{n} \cdot \sum_{k=1}^{n} \hat{f}_n(\xi^n; \xi_k)\right]^{1/2} \leqslant \frac{nb_n\varepsilon - \int K^2(z)\,dz}{\sqrt{2\int[\int K(z+t)K(t)\,dt]^2\,dz} \cdot \Phi^{-1}(1-\frac{1}{2}\gamma)}\right\} \tag{6.13}$$

Then

$$\lim_{\varepsilon \downarrow 0} P_f\left[\int[\hat{f}_{\nu_\varepsilon}(\xi^{\nu_\varepsilon}; x) - f(x)]^2\,dx \leqslant \varepsilon\right] = 1 - \gamma;$$

by this relation a global sequential fixed-width sequential confidence band for f with covering probability $1-\gamma$ is established.

The proof is given in [62]; the idea is to stop, if a certain degree of smoothness of the estimator is achieved, in the sense that $\int \hat{f}_n^2(\xi^n; x)\,dx$ is compared with $\int f^2(x)\,dx$; this is possible, because the rate of convergence of $\int \hat{f}_n^2(\xi^n; x)\,dx \approx n^{-1} \sum_{k=1}^{n} \hat{f}_n(\xi^n; \xi_k)$ to $\int f^2(x)\,dx$ is known (see [44]).

Now the limiting distribution of $\int [\hat{f}_n(\xi^n; x) - f(x)]^2\,dx$ (see [4] and [36]) and Lemma 6.3 is used to finish the proof.

Acknowledgment. The author is indebted to Jacek Koronacki for several discussion improving the first version of this paper.

References

[1] G. Banon, *Sur un estimateur non paramétrique de la densité de probabilité*, Rev. Statist. Appl. **24** (1976), 61–73.

[2] G. Bennett, *Probability inequalities for the sum of independent random variables*, J. Amer. Statist. Assoc. **57** (1962), 33–45.

[3] P. Bickel and E. Lehmann, *Unbiased estimation in convex families*, Ann. Math. Statist. **40** (1969), 1523–1535.

[4] P. J. Bickel and M. Rosenblatt, *On some global measures of the deviations of density function estimates*, Ann. Statist. **1** (1973), 1071–1095, Corrections: Ann. Statist. **3** (1975), 1370.

[5] T. Cacoullos, *Estimation of a multivariate density*, Ann. Inst. Statist. Math. **18** (1966), 179–189.

[6] R. J. Carroll, *On sequential density estimation*, Z. Wahrsch. Verw. Gebiete **36** (1976), 137–151.

[7] Y. S. Chow and H. Robbins, *On the asymptotic theory of fixed-width sequential confidence intervals for the mean*, Ann. Math. Statist. **36** (1965), 457–462.

[8] Y. S. Chow and H. Teicher, *Probability Theory*, Springer-Verlag, New York–Heidelberg–Berlin, 1978.

[9] K. Craswell, *Density estimation in a topological group*, Ann. Math. Statist. **36** (1965), 1047–1048.

[10] M. Csörgő and P. Révész, *Strong Approximations in Probability and Statistics*, Akademiai Kiadó, Budapest, 1981.

[11] H. I. Davies, *Strong consistency of a sequential estimator of a probability density function*, Bull. Math. Statist. **15** (1973), 49–54.

[12] H. I. Davies and E. J. Wegman, *Sequential density estimation*, IEEE Trans. Inform. Theory **21** (1975), 619–628.

[13] P. Deheuvels, *Sur l'estimation séquentielle de la densité*, C.R. Acad. Sci. Paris Sér. A **276** (1973), 1119–1121.

[14] –, *Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité*, C.R. Acad. Sci. Paris Sér. A **278** (1974), 1217–1220.

[15] –, *Estimation séquentielle de la densité*, Thèse présentée à l'Université Paris VI, 1974.

[16] –, *Estimation séquentielle de la densité*, in *Contribuciones en Probabilidad y Estadistica Matemática, Esenanza de la Matemática y Análisis*, Grindley Ediciones, Granada 1979, 156–168.

[17] L. Devroye, *On the pointwise and integral convergence of recursive kernel estimates of probability densities*, Utilitas Math. **15** (1979), 113–128.

[18] L. Devroye and T. J. Wagner, *The L_1 convergence of kernel density estimates*, Ann. Statist. **7** (1979), 1136–1139.

[19] V. A. Epanechnikov, *Nonparametric estimation of a multidimensional probability density*, Theory Probab. Appl. **14** (1969), 153–158.

[20] R. H. Farrell, *On the lack of a uniformly consistent sequence of estimators of a density function in certain cases*, Ann. Math. Statist. **38** (1967), 471–474.

[21] A. Földes, *Density estimation for dependent sample*, Studia Sci. Math. Hungar. **9** (1974), 443–452.

[22] A. Földes and P. Révész, *A general method for density estimation*, Studia Sci. Math. Hungar. **9** (1974), 81–92.

[23] N. Glick, *Consistency conditions for probability estimators and integrals of density estimators*, Utilitas Math. **6** (1974), 61–74.

[24] L. Györfi, *Strong consistent density estimate from ergodic sample*, J. Multivariate Anal. **11** (1981), 81–84.

[25] L. Györfi, Z. Györfi and I. Vajda, *A strong law of large numbers and some applications*, Studia Sci. Math. Hungar. **12** (1977), 233–244.

[26] A. Kozek and W. Wertz, *Estimators with values in Banach spaces under convex loss and applications to curve estimation problems*, Colloq. Math. Soc. János Bolyai, vol. 32 North-Holland, Amsterdam 1980 567–586.

[27] –, –, *On the existence of minimal complete classes of estimators*, in: P. Révész, L. Schmetterer, V. M. Zolotarev (eds.), *The First Pannonian Symposium on Mathematical Statistics*, Lecture Notes in Statist. **8**, Springer-Verlag, New York–Heidelberg–Berlin 1981. 140–155.

[28] E. L. Lehmann, *A general concept of unbiasedness*, Ann. Math. Statist. **22** (1951), 587–592.

[29] M. Loève, *Probability Theory*, 3rd ed., D. van Nostrand, Princeton, New Jersey, 1963.

[30] M. A. Mirzachmedov and Š. A. Hašimov, *On some properties of estimators of a probability density*, Kybernetika (Prague) **9** (1973), 242–250.

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

[32] V. V. Petrov, *Sums of Independent Random Variables*, Springer-Verlag, Berlin–Heidelberg–New York 1975.

[33] L. Rejtő and P. Révész, *Density estimation and pattern classification*, Problems Control Inform. Theory **2** (1973), 67–80.

[34] P. Révész, *Robbins–Monro procedure in a Hilbert space and its application in the theory of learning processes, I*, Studia Sci. Math. Hungar. **8** (1973), 391–398.

[35] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist. **27** (1956), 832–837.

[36] –, *Curve estimates*, Ann. Math. Statist. **42** (1971), 1815–1842.

[37] L. Rutkowski, *Sequential estimates of probability densities by orthogonal series and their application in pattern classification*, IEEE Trans. Systems Man Cybernet. **10** (1980), 918–920.

[38] H. Scheffé, *A useful convergence theorem for probability distributions*, Ann. Math. Statist. **18** (1947), 434–438.

[39] L. Schmetterer, *Sur l'itération stochastique. Le calcul des probabilités et ses applications*, Colloq. Internat. CNRS **87**, (1959), 55–67.

[40] –, *Sur quelques résultats asymptotiques pour le processus de Robbins–Monro*, Ann. Sci. Univ. Clermont Math. **12** (1976), 166–176.

[41] –, *From stochastic approximation to the theory of optimization*, in 11. *Steiermärkisches Math. Symposium. Ber. Math.-Statist. Sektion im Forschungszentrum Graz*, Graz 1979, Ber. Nr. 127.

[42] B. Schneider and W. Wertz, *Statistical density estimation: a bibliography*, Internat. Statist. Rev. **47** (1979), 155–175.

[43] E. F. Schuster, *Estimation of a probability density function and its derivatives*, Ann. Math. Statist. **40** (1969), 1187–1195.

[44] –, *On the rate of convergence of an estimate of a functional of a probability density*, Scand. Actuar. J. **1974** (1974), 103–107.

[45] R. C. Srivastava, *Estimation of probability density function based on random number of observations with applications*, Internat. Statist. Rev. **41** (1973), 77–86.

[46] E. L. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, New Jersey, 1970.

[47] W. Stute, *Sequential fixed-width confidence intervals for a nonparametric density function*, Math. Forschungsberichte, Univ.-Gesamthochschule Siegen, Nr. 72/1981.

[48] –, *The oscillation behavior of empirical processes*, Ann. Probab. **10** (1982), 86–107.

[49] –, *A law of the logarithm for kernel estimators*, Ann. Probab. **10** (1982), 414–422.

[50] V. Susarla and G. Walter, *Estimation of a multivariate density function using delta sequences*, Ann. Statist. **9** (1981), 347–355.

[51] I. Văduva, *Estimaţia unei densităţi de repartitie k-dimensională*, Stud. Cerc. Mat. **14** (1963), 653–660.

[52] —, *Contribuţii la teoria estimaţiilor statistice ale densităţilor de repartiţie şi aplicaţii*, Stud. Cerc. Mat. **20** (1968), 1207–1276.

[53] J. H. Venter, *On Dvoretzky stochastic approximation theorems*, Ann. Math. Statist. **37** (1966), 1534–1544.

[54] G. Walter and J. Blum, *Probability density estimation using delta sequences*, Ann. Statist. **7** (1979), 328–340.

[55] G. Watson and M. Leadbetter, *On the estimation of the probability density, I*, Ann. Math. Statist. **34** (1963), 480–491.

[56] G. Watson and M. Leadbetter, *Hazard analysis, I*, Biometrika **51** (1964), 175–184.

[57] —, —, *Hazard analysis, II*, Sankhyā Ser. A. **26** (1964), 101–116.

[58] E. J. Wegman and H. I. Davies, *Remarks on some recursive estimators of a probability density function*, Ann. Statist. **7** (1979), 316–327.

[59] W. Wertz, *Schätzfolgen für Dichten auf topologischen Gruppen*, Oper. Res. Verf. **14** (1972), 247–257.

[60] —, *Statistical Density Estimation. A Survey*, Vandenhoeck und Ruprecht, Göttingen 1978.

[61] —, *Nonparametric density estimation in abstract and homogeneous spaces*, in: P. Révész, L. Schmetterer, V. M. Zolotarev (eds.), *The First Pannonian Symposium on Mathematical Statistics*, Lecture Notes in Statist. **8**, Springer-Verlag, New York–Heidelberg–Berlin 1981, 290–301.

[62] —, *A sequential density estimator*, unpublished manuscript.

[63] —, *On some asymptotic properties of recursive density estimators*, to appear 1984.

[64] P. Whittle, *On the smoothing of probability density functions*, J. Roy Statist. Soc. Ser. B. **20** (1958), 334–343.

[65] J. Wolfowitz, *The efficiency of sequential estimates and Wald's equation for sequential processes*, Ann. Math. Statist. **18** (1974), 215–230.

[66] C. T. Wolverton and T. J. Wagner, *Recursive estimates of probability densities*, IEEE Trans. Systems Sci. Cybernet. **5** (1969), 246–247.

[67] H. Yamato, *Sequential estimation of a continuous probability density function and mode*, Bull. Math. Statist. **14** (1971), 1–12.

[68] Ш.А. Хашимов, *Равномерная среднеквадратическая сходимость оценки плотности вероятности*, в *Случайные Процессы и Статистические Выводы*, Издат. „Фан" Узбекской ССР, Ташкент 1974, 185-193.

[69] Н.Н. Ченцов, *Оценка неизвестной плотности распределения по наблюдениям*, ДАН СССР, **147** (1962), 45-48.

Presented to the semester
Sequential Methods in Statistics
September 7–December 11, 1981