STRONG APPROXIMATION OF DENSITY ESTIMATORS FROM WEAKLY DEPENDENT OBSERVATIONS BY DENSITY ESTIMATORS FROM INDEPENDENT OBSERVATIONS

BY MICHAEL H. NEUMANN

Humboldt Universität

We derive an approximation of a density estimator based on weakly dependent random vectors by a density estimator built from independent random vectors. We construct, on a sufficiently rich probability space, such a pairing of the random variables of both experiments that the set of observations $\{X_1, \ldots, X_n\}$ from the time series model is nearly the same as the set of observations $\{Y_1, \ldots, Y_n\}$ from the i.i.d. model. With a high probability, all sets of the form $(\{X_1, \ldots, X_n\} \Delta \{Y_1, \ldots, Y_n\}) \cap ([a_1, b_1] \times \cdots \times [a_d, b_d])$ contain no more than $O(\{[n^{1/2} \prod(b_i - a_i)] + 1\} \log(n))$ elements, respectively. Although this does not imply very much for parametric problems, it has important implications in nonparametric statistics. It yields a strong approximation of a kernel estimator of the stationary density by a kernel density estimator in the i.i.d. model. Moreover, it is shown that such a strong approximation is also valid for the standard bootstrap and the smoothed bootstrap. Using these results we derive simultaneous confidence bands as well as supremum-type nonparametric tests based on reasoning for the i.i.d. model.

1. Introduction. Density estimation on the basis of i.i.d. observations is one of the most often studied problems in nonparametric statistics. Important asymptotic properties concerning the pointwise as well as the joint probabilistic behavior of commonly used estimators are now well known and allow for powerful methods of statistical inference such as tests for certain hypotheses or simultaneous confidence bands which guarantee asymptotically the desired error probability of the first kind and coverage probability, respectively.

In contrast, much less is known in the case of dependent observations. This case is very important from the practical point of view, since data from time series usually show some dependence. In order to develop analogous tools as in the independent case, it seems to be on first sight unavoidable to account for the dependence by specific corrections. This might, however, turn out to be quite a difficult and messy task. Hence, it is tempting to seek for conditions which ensure asymptotically the same behavior of certain statistics as known from the i.i.d. setting.

Whereas long-range dependence usually leads to phenomena essentially different from those under independence, there seems to be some hope for asymptotic similarities to the independent case under short-range dependence. Some commonly imposed conditions for weak dependence are strong (α-) mixing and

---

absolute regularity (β-mixing). Provided the corresponding mixing coefficients decay fast enough, commonly used nonparametric estimators converge with the same *rates* as in the independent case; compare [19]. The fact that desirable properties of the estimators remain valid in the dependent case provides a strong motivation for applying just the same estimation techniques as under the assumption of independence. However, some important tools for statistical inference require a more accurate knowledge of the asymptotic properties of the underlying estimators. Assuming mixing and some additional, not very restrictive, condition on the boundedness of the joint densities of consecutive random variables, it was shown in [39], [30] and [26] that certain nonparametric estimators have actually the same asymptotic variance as in the independent case. This phenomenon, which was described as "whitening by windowing" by Hart, is in sharp contrast to what happens in (finite-dimensional) parametric problems. For example, the asymptotic variance of the mean of time-series data does of course depend on the covariances as well. Results such as those of [39], [30] and [26] on the *pointwise* behavior of nonparametric estimators allow one, for example, to neglect the dependence structure when one establishes pointwise confidence intervals for the density function. Such an effect was also observed in [24], which showed that the mean integrated squared error (MISE) of a kernel density estimator from a MA(∞)-process may be expanded as the sum of the MISE of a kernel estimator based on an i.i.d. sample, plus a term $E(\overline{X} - X_1) \int (f')^2$ which is $O(n^{-1})$ under short-range dependence.

On the other hand, other problems of statistical inference require an even stronger notion of asymptotic equivalence. For example, the construction of simultaneous confidence bands or the determination of critical values for certain tests against a nonparametric alternative require knowledge about the *joint* distribution of the nonparametric estimator used to define the corresponding statistic. A first step in this direction has been done in [34]. The authors characterized the asymptotic equivalence of nonparametric autoregression and nonparametric regression through a strong approximation of a local polynomial estimator of the autoregression function by a local polynomial estimator in an appropriate regression setup. However, the nonparametric autoregressive model automatically imposes certain *structural* conditions on the data-generating process, which were essential for the approximation method used. Since this restricts the applicability of such a method in practice, it would be very desirable to develop similar results without any such structural assumptions.

In the present paper we show quite a surprising similarity between the observations that stem from a time-series model and a set of independent observations. Let $X_1, \ldots, X_n$ be d-dimensional, weakly dependent random vectors with a stationary density f. As a counterpart we consider i.i.d. random vectors $Y_1, \ldots, Y_n$ with the same density. Let $\Delta_n = \{X_1, \ldots, X_n\} \Delta \{Y_1, \ldots, Y_n\}$ be the symmetric difference of both sets of observations. We show that there exists, on a sufficiently rich probability space, a pairing of the random variables of both models, which preserves the respective joint distributions, such that the following fact is true. With a probability exceeding $1 - O(n^{-\lambda})$, the

relation

$$\#\big(\Delta_n \cap [a, b]\big) = O\bigg(\bigg\{\bigg[n^{1/2} \prod_{i=1}^{d}(b_i - a_i)\bigg] + 1\bigg\} \log(n)\bigg)$$

is simultaneously satisfied for all hyperrectangles $[a, b] = [a_1, b_1] \times \cdots \times [a_d, b_d]$, where $\lambda < \infty$ is an arbitrarily large constant. The link is achieved by embedding both the random variables from the time-series model and the i.i.d. model in a common Poisson process on $(0, \infty) \times \mathbb{R}^d$.

Let $\hat{f}_h(x) = (nh^d)^{-1} \sum_{i=1}^{n} K((x - X_i)/h))$ and $\tilde{f}_h(x) = (nh^d)^{-1} \sum_{i=1}^{n} K((x - Y_i)/h))$ be kernel estimators of $f(x)$, where K is a compactly supported kernel function. Then we see that, with a high probability,

$$\sup_{x}\big\{\#\big[\Delta_n \cap \operatorname{supp}(K((x - \cdot)/h))\big]\big\} = O\big(n^{1/2}h^d \log(n)\big),$$

and, therefore,

$$\sup_{x \in \mathbb{R}^d}\big\{\hat{f}_h(x) - \tilde{f}_h(x)\big\} = O\big(n^{-1/2} \log(n)\big).$$

In view of the fact that $\sup_x\{\operatorname{var}(\hat{f}_h(x))\} \asymp (nh^d)^{-1}$, we have a useful strong approximation of the kernel estimator $\{\hat{f}_h(x)\}_{x \in \mathbb{R}^d}$ by $\{\tilde{f}_h(x)\}_{x \in \mathbb{R}^d}$.

As some interesting applications we construct simultaneous confidence bands for f as well as tests based on the maximum absolute deviation between the above kernel estimator $\hat{f}_h$ and estimators corresponding to hypotheses of lower-dimensional parametric or semiparametric structures. To determine the required tuning parameters, that is, the width of the bands and the critical value for the test, respectively, we propose two bootstrap methods, both developed under the assumption of independence.

2. The approximation scheme. The main goal in this section is to establish a link between density estimation under weak dependence and density estimation based on independent observations. This will be achieved in a mainly constructive way, by embedding the random variables of both models in a common Poisson process indexed by time as well as spatial position in $\mathbb{R}^d$. The seemingly quite involved problem of finding a global (in x) connection between kernel estimators $\hat{f}_h(x)$ and $\tilde{f}_h(x)$ in these models will be reduced to a collection of one-dimensional problems, which can be analyzed separately from each other. Hence, in contrast to many other papers on strong approximations, the pleasant fact with our approximation method is that the technical part of the calculations becomes quite elementary.

2.1. *The model and basic assumptions.* Assume we have d-dimensional realizations $X_1, \ldots, X_n$ of a stationary process with a stationary density f. Let $\mathscr{F}_i^{\,j} = \sigma(X_i, X_{i+1}, \ldots, X_j)$ be the σ-field generated by $X_i, \ldots, X_j$. Throughout the paper we use the letter C to denote a generic constant which may attain different values at different places. Sometimes we use the letters $C_1, C_2, \ldots$ for constants whose exact value is important in subsequent

calculations. To obtain some kind of asymptotic equivalence to the case of i.i.d. random variables, we impose the following conditions.

ASSUMPTION 1. The coefficient of absolute regularity (β-mixing coefficient) is defined as

$$\beta(k) = \sup_i E \sup_{V \in \mathscr{F}_{i+k}^n} \{|P(V \mid \mathscr{F}_1^i) - P(V)|\}.$$

We suppose that the $\beta(k)$ decay with an exponential rate, that is,

$$\beta(k) \leq C \exp(-C_1 k).$$

ASSUMPTION 2. Let $f_{X_i \mid \mathscr{F}_j^{i-1}}$ be the density of the conditional distribution $\mathscr{L}(X_i \mid X_j, \ldots, X_{i-1})$. We assume that there exist constants $C_2, C_3 > 0$ such that

$$\sup_{i > \gamma+1} \left\{ P\left(\sup_x \{|f_{X_i \mid \mathscr{F}_1^{i-1}}(x) - f_{X_i \mid \mathscr{F}_{i-\gamma}^{i-1}}(x)|\} > C \exp(-C_2 \gamma) \right) \right\} \leq C \exp(-C_3 \gamma)$$

holds for all $\gamma \geq 1$ and

$$\sup_{i>1} \sup_{x \in \mathbb{R}^d} \{f_{X_i \mid \mathscr{F}_1^{i-1}}(x)\} \leq C.$$

REMARK 1. (i) Our assumption of exponentially decaying mixing coefficients is stronger than actually needed and can possibly be relaxed on the expense of a slightly larger error in our approximation. Nevertheless, many of the commonly used time series models describe processes which are geometrically absolutely regular under natural conditions. For example, sufficient conditions for geometric absolute regularity of multivariate MA(∞) processes and ARMA processes can be easily read off from results of [37]; see also [31] for geometric β-mixing of vector ARMA processes. This property was established in [36] for generalized random coefficient autoregressive models and bilinear models. Sufficient conditions are provided in [32], Theorem 2.1, for a Markov chain to be geometrically β-mixing. This result was used in [2] to derive sufficient conditions for a vector autoregressive process with conditional heteroscedasticity given as

$$X_{i+1} = m(X_i) + g(X_i)\varepsilon_{i+1},$$

ε_i i.i.d., to be geometrically ergodic, which implies geometrical β-mixing if the chain is stationary. In [17] this is extended to the case of not necessarily identical distributions of the innovations which may also have compact support. A survey on available results concerning mixing properties of popular time-series models is given in [9].

(ii) Some kind of mixing seems to be a minimal requirement which brings the time series model close to an i.i.d. situation. This is, however, not enough to get the desired asymptotic equivalence. We need some additional condition which ensures that closely neighbored (in time) observations do not behave too differently from an i.i.d. situation. Whereas [39], [30] and [26] imposed

2018 M. H. NEUMANN

a condition on the boundedness of the joint densities, we set this slightly stronger Assumption 2, which also reflects a rapidly decaying memory of the process $\{X_i\}$.

2.2. *Embedding the random variables in a common Poisson process.* Now we relate the random vectors $X_1, \ldots, X_n$ from the above setup to i.i.d. random vectors $Y_1, \ldots, Y_n$ having a density f. For that, we define on a sufficiently rich probability space copies $X'_1, \ldots, X'_n$ and $Y'_1, \ldots, Y'_n$ with the same joint distribution as $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively. As the connecting device, which determines both $X'_1, \ldots, X'_n$ and $Y'_1, \ldots, Y'_n$, we use a Poisson process N on $(0, \infty) \times \mathbb{R}^d$ with an intensity function equal to the Lebesgue measure. For details concerning the definition and construction of N, see [38], Section 2.1. In contrast to Reiss, we use the equivalent formulation of a set-valued process instead of a point measure-valued process. Furthermore, since it is unlikely that this causes any confusion, we do not distinguish between X_i and X'_i as well as Y_i and Y'_i and denote the versions of these random variables on the common probability space simply by X_i and Y_i, respectively.

First we describe in detail how the Poisson process N is used to generate the observations $X_1, \ldots, X_n$, retaining the joint distribution of these random vectors. The embedding of $Y_1, \ldots, Y_n$ is completely analogous, since independence is a special case of weak dependence.

A part of a realization of a Poisson process on $(0, \infty) \times \mathbb{R}$ is shown in Figure 1.

(i) *Embedding of X_1.* Let $\{(U_j, V_j), j = 1, 2, \ldots\}$ denote a realization of N, where $U_j \in (0, \infty)$ and $V_j \in \mathbb{R}^d$. The basic idea of how X_1 is represented by $N^{(1)} = N$ may be explained as follows: consider the graph $(t f_{X_1}(v), v)$ of the function $g_t(v) = t f_{X_1}(v)$, which spreads out, starting from $\{0\} \times \mathbb{R}^d$, with a velocity proportional to $f_{X_1}(v)$. We define

$$X_1 = V_{j_1},$$

where (U_{j_1}, V_{j_1}) is the first realization of $N^{(1)}$ hit by $(t f_{X_1}(v), v)$ as t grows from zero to infinity. In other words, we have

$$j_1 = \arg\inf\{U_j / f_{X_1}(V_j)\}.$$

Note that $\{(U_j / f_{X_1}(V_j), V_j), j = 1, 2, \ldots\}$ is a Poisson process on $(0, \infty) \times \mathbb{R}^d$ with intensity function $p(u, v) = f_{X_1}(v)$. Hence, it is clear that X_1 has just the desired density $f_{X_1} = f$.

To explain the following steps in a formally correct way, we introduce stopping times $\tau_v^{(i)}$ indexed by spatial position $v \in \mathbb{R}^d$, $i = 0, \ldots, n$. Define

$$\tau_v^{(0)} \equiv 0$$

and

$$\tau_v^{(1)} = \tau_v^{(0)} + \left[U_{j_1} / f_{X_1}(V_{j_1})\right] f_{X_1}(v).$$
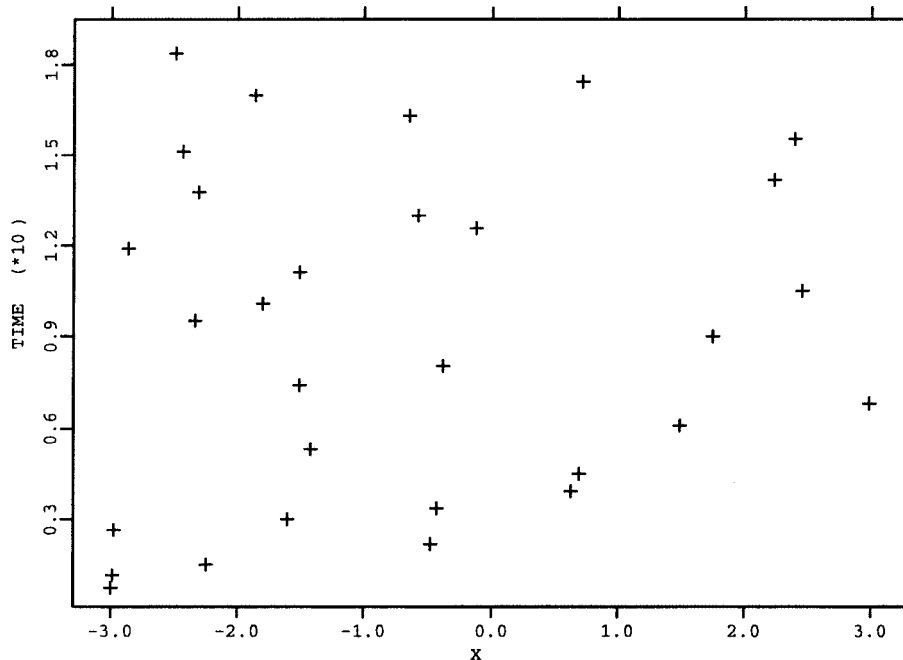
FIG. 1. *Realization of the Poisson process.*

The process of determining (U_{j_1}, V_{j_1}) is sketched in Figure 2. For the purpose of illustrating our embedding method, we generated a Gaussian time series $X_i = aX_{i-1} + \varepsilon_i$, where $\varepsilon_i \sim N(0, 1 - a^2)$ are i.i.d. and $a = 0.8$. Then (U_{j_1}, V_{j_1}) is marked by a star and the corresponding value of X_1 is marked by a circle. The graph of the stopping times $\tau_v^{(1)}$ is drawn as a solid line.

If we order $\{(U_j/f_{X_1}(V_j), V_j), j = 1, 2, \ldots\}$ with respect to the first component, we may alternatively construe this object as a marked Poisson point process where the second argument has the density f_{X_1}. If we denote the corresponding realizations of this process by (S_j, W_j), $S_1 < S_2 < \cdots$, then X_1 is just equal to W_1. By the strong Markov property of a marked Poisson point process, the remaining part of N,

$$N^{(2)} = \{(U_j - \tau_{V_j}^{(1)}, V_j)\} \cap ((0, \infty) \times \mathbb{R}^d),$$

is again a Poisson process on $(0, \infty) \times \mathbb{R}^d$.

(ii) *Embedding of X_i.* Assume that $X_1, \ldots, X_{i-1}$ have already been embedded in N, according to their conditional distributions $\mathscr{L}(X_k \mid X_{k-1}, \ldots, X_1)$. Moreover, assume that $\tau_v^{(1)}, \ldots, \tau_v^{(i-1)}$ are already defined. We embed X_i
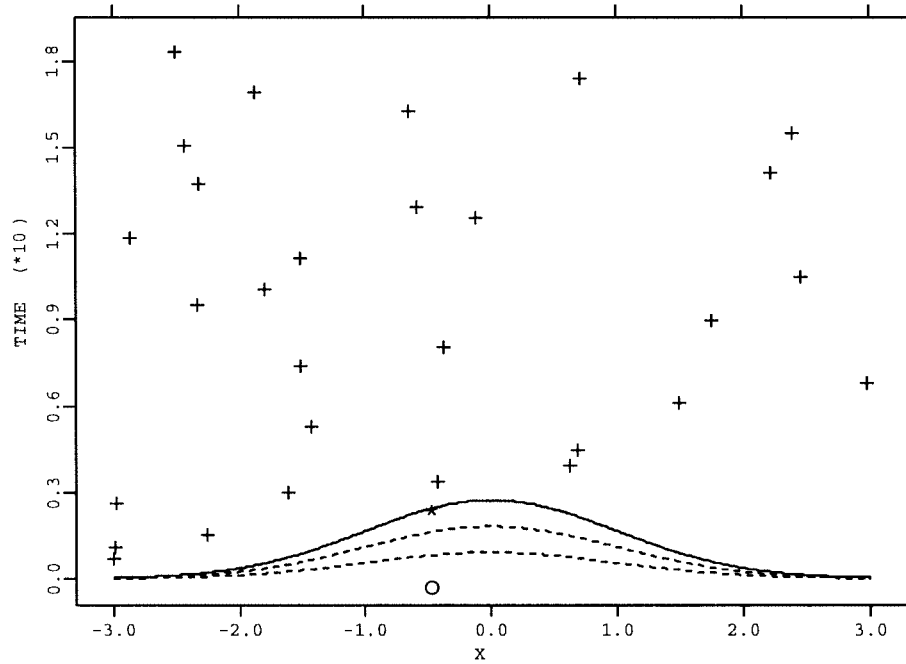
FIG. 2.   *Process of determining (U_{j_1}, V_{j_1}).*

in the remaining part of N, that is,

$$N^{(i)} = \left\{ \left(U_j - \tau_{V_j}^{(i-1)}, V_j \right) \right\} \cap \left((0, \infty) \times \mathbb{R}^d \right).$$

In other words, we use from the whole set of realizations $\{(U_j, V_j)\}$ of N only those from the subset $\{(U_j, V_j) \mid U_j > \tau_{V_j}^{(i-1)}\}$. By the strong Markov property of the corresponding marked Poisson point process, $N^{(i)}$ is again a Poisson process on $(0, \infty) \times \mathbb{R}^d$. Now we define

$$X_i = V_{j_i},$$

where

$$j_i = \arg\inf \left\{ (U_j - \tau_{V_j}^{(i-1)}) / f_{X_i \mid \mathscr{F}_1^{i-1}}(V_j), U_j > \tau_{V_j}^{(i-1)} \right\}.$$

Further, we set

$$\tau_v^{(i)} = \tau_v^{(i-1)} + \left[(U_{j_i} - \tau_{V_{j_i}}^{(i-1)}) / f_{X_i \mid \mathscr{F}_1^{i-1}}(V_{j_i}) \right] f_{X_i \mid \mathscr{F}_1^{i-1}}(v).$$

The process of determining (U_{j_2}, V_{j_2}) is sketched in Figure 3; (U_{j_2}, V_{j_2}) is marked by a star and the corresponding value of X_2 is marked by a circle. The graph of the stopping times $\tau_v^{(2)}$ is drawn as a solid line.
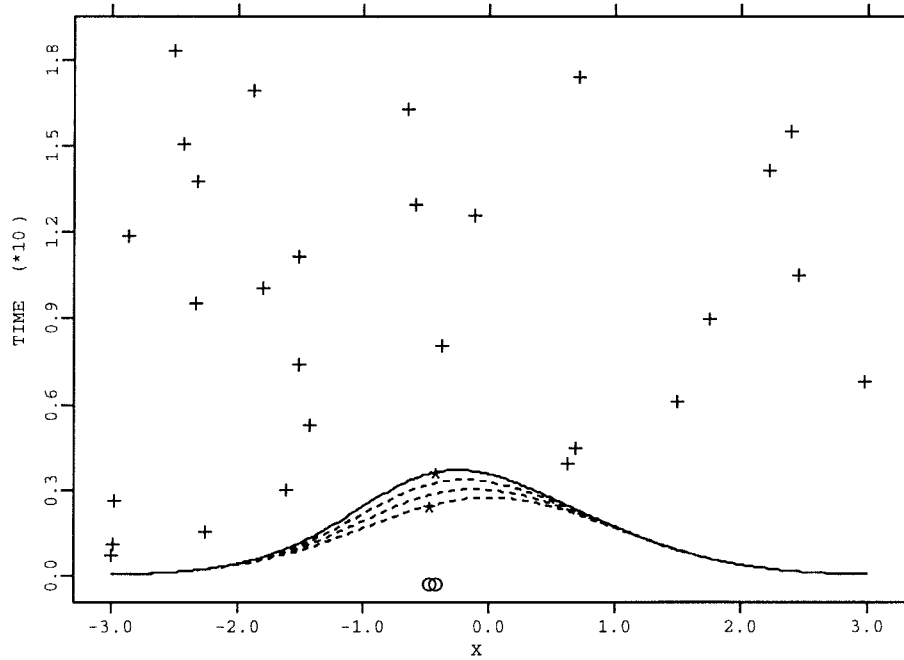
FIG. 3. *Process of determining* (U_{j_2}, V_{j_2}).

Finally, we obtain that

$$\{X_1, \ldots, X_n\} = \{V_j \mid U_j \leq \tau^{(n)}_{V_j}\}. \tag{2.1}$$

(iii) *Embedding of* $Y_1, \ldots, Y_n$. The embedding of $Y_1, \ldots, Y_n$ is completely analogous to that of $X_1, \ldots, X_n$. Since $\mathscr{L}(Y_i \mid Y_{i-1}, \ldots, Y_1) = \mathscr{L}(Y_1)$, we have to deform the time axis only once.

Let $\{(\tilde{T}_j, \tilde{W}_j), j = 1, 2, \ldots\}$ be the marked Poisson point process corresponding to $\{(U_j/f(V_j), V_j), j = 1, 2, \ldots\}$. That is, we have in particular $\tilde{T}_1 < \tilde{T}_2 < \cdots$. Then we define

$$Y_i = \tilde{W}_i, \qquad i = 1, \ldots, n.$$

We may introduce stopping times $\tilde{\tau}^{(i)}_v$ analogous to the $\tau^{(i)}_v$'s. We obtain $\tilde{\tau}^{(n)}_v = \tilde{T}_n f(v)$, which implies that

$$\{Y_1, \ldots, Y_n\} = \{V_j \mid U_j \leq \tilde{\tau}^{(n)}_{V_j}\}. \tag{2.2}$$
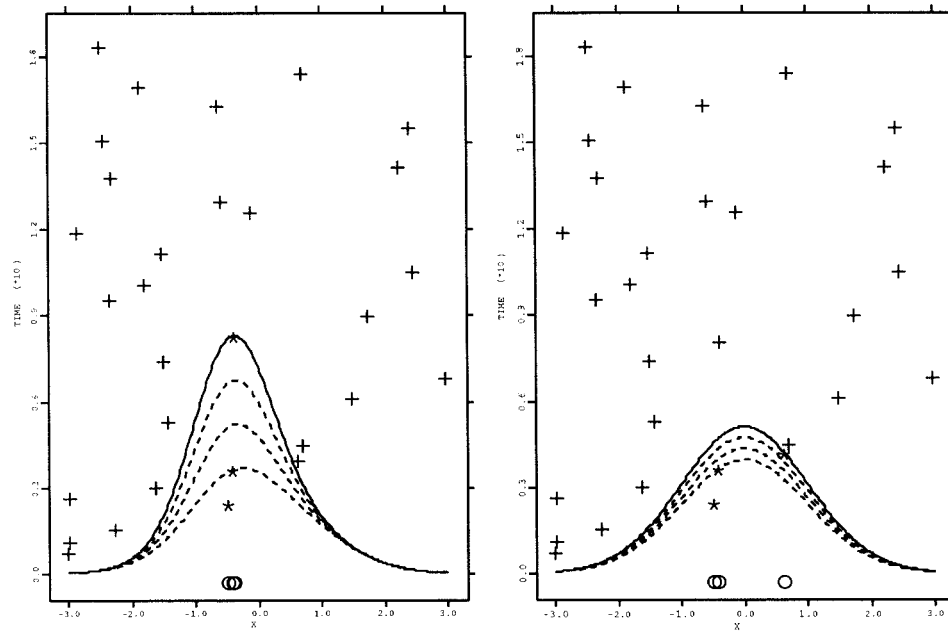
FIG. 4. X_1, X_2, X_3 *vs.* Y_1, Y_2, Y_3.

Figure 4 displays the first three realizations of the processes $\{X_i\}$ (left side) and $\{Y_i\}$ (right side).

REMARK 2.   (i) It may well happen that the X_i's emerge in a different chronological order than the Y_i's. Since the transition densities are usually different from the stationary density, the construction for the time-series model "borrows" some probability mass assigned to future time points in the i.i.d. model. This is just the reason why we introduce a "time axis" for our embedding method.

(ii) Poisson processes are occasionally used to generate other stochastic processes. A marked Poisson point process to generate a Poisson process with random intensity is used in [6]. However, apart from the common fact in both papers that a Poisson process is used to generate some other stochastic process, both the purpose as well as the method of embedding in their paper are completely different from ours. The author is not aware of any other work where time-series data are generated by a Poisson process in the way described here.

2.3. *Approximation results.*   To get estimates for the number of elements of Δ_n that fall in certain hyperrectangles, we derive first an estimate for the distance between $\tau_v^{(n)}$ and $\tilde{\tau}_v^{(n)}$, respectively, and their common expectation $nf(v)$.

Since many assertions in this article are of the type that a certain random variable is below some threshold with a high probability, we introduce the following notation.

DEFINITION 2.1. Let $\{Z_n\}$ be a sequence of random variables and let $\{\alpha_n\}$ and $\{\gamma_n\}$ be sequences of positive reals. We write

$$Z_n = \tilde{O}(\alpha_n, \gamma_n)$$

if

$$P(|Z_n| > C\alpha_n) \leq C\gamma_n$$

holds for $n \geq 1$ and some $C < \infty$.

This definition is obviously stronger than the usual O_P and it is well suited for our particular purposes of constructing confidence bands and nonparametric tests; see its application in Section 3.

Further, we make throughout the paper the convention that $\delta > 0$ will denote an arbitrarily small and $\lambda < \infty$ an arbitrarily large constant.

LEMMA 2.1. *Suppose Assumptions* 1 *and* 2 *are fulfilled. Then, for arbitrary fixed* $v \in \mathbb{R}^d$,

$$\left|\tau_v^{(n)} - nf(v)\right| + \left|\tilde{\tau}_v^{(n)} - nf(v)\right| = \tilde{O}\big(n^{1/2}\log(n), n^{-\lambda}\big).$$

Whereas the pointwise (in v) similar behavior of $\tau_v^{(n)}$ and $\tilde{\tau}_v^{(n)}$ does not imply anything essential, a uniform version of the result given in Lemma 2.1 will finally yield the desired result about the difference set Δ_n. To derive such a uniform version, we impose the following smoothness condition on the conditional densities.

ASSUMPTION 3. There exists some constant $C < \infty$ such that

$$\sup_i \sup_{F \in \mathscr{F}_1^{i-1}} \left\{\left|f_{X_i \mid F}(v) - f_{X_i \mid F}(v')\right|\right\} \leq C\|v - v'\|.$$

LEMMA 2.2. *Suppose that Assumptions* 1–3 *are fulfilled. Then we have, for any* fixed *hyperrectangle* $[a, b] = [a_1, b_1] \times \cdots \times [a_d, b_d]$, *that*

$$\sup_{v \in [a,b]} \left\{\left|\tau_v^{(n)} - nf(v)\right| + \left|\tilde{\tau}_v^{(n)} - nf(v)\right|\right\} = \tilde{O}\big(n^{1/2}\log(n), n^{-\lambda}\big).$$

Now we are in a position to relate both experiments to a common experiment given by the restriction of N to

$$S_n = \big\{(u, v) \mid \quad 0 < u \leq nf(v), v \in \mathbb{R}^d\big\}.$$

Let

(2.3) $$\{Z_1, \ldots, Z_\nu\} = \{V_j \mid U_j \leq nf(V_j)\}.$$

We obtain estimates for the cardinality of the sets $(\{X_1,\ldots,X_n\}\Delta\{Z_1,\ldots,Z_\nu\})\cap[a,b]$ as well as $(\{Y_1,\ldots,Y_n\}\Delta\{Z_1,\ldots,Z_\nu\})\cap[a,b]$ from Lemma 2.2 and an appropriate exponential inequality for Poisson processes.

PROPOSITION 2.1. *Suppose Assumptions 1–3 are fulfilled. Then, with a probability exceeding* $1-O(n^{-\lambda})$,

$$\left.\begin{array}{l}\#\{(\{X_1,\ldots,X_n\}\Delta\{Z_1,\ldots,Z_\nu\})\cap[a,b]\}\\[4pt]\#\{(\{Y_1,\ldots,Y_n\}\Delta\{Z_1,\ldots,Z_\nu\})\cap[a,b]\}\end{array}\right\}$$

$$=O\left(\left\{\left[n^{1/2}\prod_{i=1}^d(b_i-a_i)\right]+1\right\}\log(n)\right)$$

holds simultaneously for all hyperrectangles $[a,b]=[a_1,b_1]\times\cdots\times[a_d,b_d]$ *with* $\max_i\{b_i-a_i\}=O(n^C)$.

If additionally, $P(X_1\in[a,b])\le C\prod_{i=1}^d[(b_i-a_i)\wedge 1]$ *is satisfied, then the above assertion holds, with a probability exceeding* $1-O(n^{-\lambda})$, *for all* $[a,b]$.

Now we obtain, as an immediate consequence of Proposition 2.1, the desired strong approximation of a kernel estimator $\hat{f}_h$ in the time-series model by a kernel estimator $\tilde{f}_h$ in the i.i.d. model. Let

$$\hat{f}_h(x)=\frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$$

and

$$\tilde{f}_h(x)=\frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-Y_i}{h}\right).$$

For simplicity we impose the following condition.

ASSUMPTION 4. The kernel K is supported on $[-1,1]^d$ and

$$\sup_x\{|K(x)|\}\le K_0.$$

It is obvious that, with a probability exceeding $1-O(n^{-\lambda})$,

$$\begin{aligned}&\left|\hat{f}_h(x)-\tilde{f}_h(x)\right|\\&\quad\le\frac{K_0}{nh^d}\#\{(\{X_1,\ldots,X_n\}\Delta\{Y_1,\ldots,Y_n\})\\&\qquad\cap([x_1-h,x_1+h]\times\cdots\times[x_d-h,x_d+h])\}\\&\quad=\tilde{O}\left([n^{-1/2}+(nh^d)^{-1}]\log(n),n^{-\lambda}\right)\end{aligned}$$

(2.4)

holds simultaneously for all $x\in\mathbb{R}^d$. This is formalized by the following theorem.

THEOREM 2.1. *Suppose Assumptions* 1–4 *are fulfilled. Then*

$$\sup_{x\in\mathbb{R}^d}\big\{\big|\hat{f}_h(x) - \tilde{f}_h(x)\big|\big\} = \tilde{O}\big([n^{-1/2} + (nh^d)^{-1}]\log(n), n^{-\lambda}\big).$$

Now it becomes clear what we have achieved by our embedding of $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ in a common Poisson process: the seemingly quite difficult task of getting a *uniform* (in x) approximation of $\hat{f}_h(x)$ by $\tilde{f}_h(x)$ is reduced to the technically much simpler task of proving a *pointwise* result as in Lemma 2.1.

3. Application to simultaneous confidence bands and nonparametric tests. Theorem 2.1 in Section 2 provides an approximation of a kernel estimator in the time-series model by a kernel estimator in an i.i.d. model. Besides the more fundamental message that weak dependence is asymptotically negligible, the practical significance lies on the possibility of transfering methods of inference originally developed under the assumption of independence to the case of weakly dependent random variables. As two important applications, we propose in this section confidence bands and supremum-type tests based on a bootstrap approximation of the distribution of the L_∞-distance between $\hat{f}_h$ and $E\hat{f}_h$. We did not attempt to develop versions of these methods based on asymptotic theory. Although, at least in the one-dimensional case, the process $\{(\hat{f}_h(x) - E\hat{f}_h(x))/\sqrt{\mathrm{var}(\hat{f}_h(x))}\}_{x\in[a,b]}$ can be well approximated by a Gaussian process, the approximation of the supremum of the modulus of this Gaussian process by its limit, as proposed by [5], converges with the very slow rate $(\log(n))^{-1}$; compare [20]. In contrast, it will be shown that the bootstrap approximation converges with a certain algebraic rate.

3.1. *Two bootstrap proposals.* We consider two methods of bootstrapping the empirical process: the standard bootstrap and the smoothed bootstrap. Both versions were proposed in [12] in the context of i.i.d. observations.

Denote by P_n the empirical distribution based on $\{X_1, \ldots, X_n\}$. In the standard bootstrap, we draw with replacement n independent bootstrap resamples $X_1^*, \ldots, X_n^*$. That is, the unknown distribution P is replaced by its empirical analog P_n. In the smoothed bootstrap, we draw n independent bootstrap resamples $X_1^{*,g}, \ldots, X_n^{*,g}$ from a smoothed version $P_{n,g}$ of P_n. Here $P_{n,g}$ is the distribution function which corresponds to the kernel estimate

$$\hat{f}_g(x) = \frac{1}{ng^d}\sum_{i=1}^{n} L\left(\frac{x - X_i}{g}\right)$$

of $f(x)$. We use the letters L and g to indicate that one may use a kernel and a bandwidth different from K and h, respectively. It will turn out that there is very much freedom for the choice of g.

A discussion about the relative merits of the standard bootstrap and the smoothed bootstrap as well as some examples may be found in [12], [13],

[41], [23], [14] and [15]. A survey is given in [21], Appendix IV. Roughly speaking, smoothing does not improve the convergence rate of the bootstrap estimate, if that estimate can be expressed as (or is well approximated by) a smooth function of a vector sample mean. In other cases such as in estimating the distribution of a quantile estimate, the smoothed bootstrap can significantly outperform the unsmoothed one; compare [23] and [14]. Moreover, [15] showed that the smoothed bootstrap is consistent w.r.t. the variational distance, whereas the unsmoothed one is merely correct w.r.t. the Kolmogorov–Smirnov distance.

The derivation of asymptotic properties of the bootstrap methods goes again via strong approximations. We begin with the smoothed bootstrap and construct a pairing of $(Y_1, \ldots, Y_n)$ and $(X_1^{*, g}, \ldots, X_n^{*, g})$, which are both vectors of i.i.d. random variables, as follows. First we draw n independent Bernoulli random variables $B_i \sim \text{Bernoulli}(p)$, where $p = \int (f(x) \wedge \hat{f}_g(x)) \, dx$. If $B_i = 1$, then we generate Y_i according to the density $(f(x) \wedge \hat{f}_g(x))/p$, and set $X_i^{*, g} = Y_i$. If $B_i = 0$, then we draw independently Y_i according to the density $[f(x) - (f(x) \wedge \hat{f}_g(x))]/(1 - p)$ and $X_i^{*, g}$ with the density $[\hat{f}_g(x) - (f(x) \wedge \hat{f}_g(x))]/(1 - p)$. It is easy to see that $Y_1, \ldots, Y_n$ are i.i.d. with density f and $X_1^{*, g}, \ldots, X_n^{*, g}$ are i.i.d. with density $\hat{f}_g$. The next theorem shows that this construction actually leads to a useful approximation of $\{\hat{f}_h(x) - E\hat{f}_h(x)\}_{x \in \mathbb{R}^d}$ by $\{\hat{f}_h^{*, g}(x) - E\hat{f}_h^{*, g}(x)\}_{x \in \mathbb{R}^d}$, where

$$\hat{f}_h^{*, g}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i^{*, g}}{h}\right).$$

Since the proofs of the assertions of this section use approximations of the kernel estimators on fine grids, we impose the following additional conditions:

ASSUMPTION 5. The kernel K is Lipschitz continuous and of second order.

ASSUMPTION 6. The kernel L is Lipschitz continuous and of second order. Moreover, L is supported on $[-1, 1]^d$ and $\sup_x \{|L(x)|\} \leq L_0$.

THEOREM 3.1. *Suppose Assumptions* 1–6 *are fulfilled. Let*

$$\mu_n = g^2 + (ng^d)^{-1/2}\sqrt{\log(n)}.$$

Then there exists a pairing of the random variables $X_1, \ldots, X_n$ *and* $X_1^{*, g}, \ldots, X_n^{*, g}$ *such that*

$$\sup_{x \in \mathbb{R}^d} \{|(\hat{f}_h(x) - E\hat{f}_h(x)) - (\hat{f}_h^{*, g}(x) - E\hat{f}_h^{*, g}(x))|\}$$

$$= \tilde{O}\left(n^{-1/2}\log(n) + (nh^d)^{-1}\log(n) + (nh^d)^{-1/2}\mu_n^{1/2}\sqrt{\log(n)}, n^{-\lambda}\right).$$

Now we turn to the standard bootstrap. Here $\{\hat{f}_h(x) - E\hat{f}_h(x)\}_{x \in \mathbb{R}^d}$ is approximated by $\{\hat{f}_h^{*,\,0}(x) - E\hat{f}_h^{*,\,0}(x)\}_{x \in \mathbb{R}^d}$, where

$$\hat{f}_h^{*,\,0}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i^*}{h}\right).$$

In contrast to the case of the smoothed bootstrap, the distributions P and P_n are actually orthogonal. Hence, there is no hope of finding a pairing of both experiments such that enough random variables from them coincide. However, obviously one can define a pairing of $(X_1^{*,\,g}, \ldots, X_n^{*,\,g})$ and $(X_1^*, \ldots, X_n^*)$ such that $\|X_i^{*,\,g} - X_i^*\| \leq \sqrt{d}g$ for all i. Hence, for $g \ll h$, $\hat{f}_h^{*,\,0}(x)$ is well approximated by $\hat{f}_h^{*,\,g}(x)$, which finally provides the desired strong approximation of $\{\hat{f}_h(x) - E\hat{f}_h(x)\}_{x \in \mathbb{R}^d}$ by $\{\hat{f}_h^{*,\,0}(x) - E\hat{f}_h^{*,\,0}(x)\}_{x \in \mathbb{R}^d}$.

THEOREM 3.2. *Suppose Assumptions* 1–5 *are fulfilled. Then there exists a pairing of the random variables* $X_1, \ldots, X_n$ *and* $X_1^{*,\,0}, \ldots, X_n^{*,\,0}$ *such that*

$$\sup_{x \in \mathbb{R}^d} \left\{ \left| (\hat{f}_h(x) - E\hat{f}_h(x)) - (\hat{f}_h^{*,\,0}(x) - E\hat{f}_h^{*,\,0}(x)) \right| \right\}$$

$$= \tilde{O}\Big(n^{-1/2} \log(n) + (nh^d)^{-1} \log(n)$$

$$+ (nh^d)^{-1/2} \sqrt{\log(n)} \inf_g \left\{ (ng^d)^{-1/2} \sqrt{\log(n)} + g/h \right\}, n^{-\lambda} \Big).$$

In order to assess the significance of the above strong approximation results for the desired approximation of the distribution of the maximum absolute deviation of $\hat{f}_h$ from its expectation, we still need an upper bound for the probabilities that this supremum falls into small intervals.

PROPOSITION 3.1. *Suppose Assumptions* 1–5 *are fulfilled.*

$$P\Big(\sup_{x \in \mathbb{R}^d} \{ |\hat{f}_h(x) - E\hat{f}_h(x)| \} \in [c, d] \Big)$$

$$= O\Big((d - c)(nh^d)^{1/2}(\log(n))^{1/2} + h\log(n) + (nh^d)^{-1/4}(\log(n))^{5/4}$$

$$+ h^{d/2}(\log(n))^{3/2} + (nh^d)^{-1/2}(\log(n))^{3/2} \Big).$$

This estimate will finally imply, in conjunction with Theorems 2.1, 3.1 and 3.2, the validity of the bootstrap for the supremum functional. We apply this to the construction of simultaneous confidence bands and nonparametric tests in the following two subsections.

3.2. *Simultaneous confidence bands.* Confidence bands are an important universal tool which provide some impression about the exactness of a nonparametric estimator. Similarly to nonparametric tests, they can indicate

whether there is empirical evidence for certain conjectured features of the curve.

There already exists a considerable amount of literature on the construction of confidence bands in the context of independent observations. Work on simultaneous confidence bands in nonparametric density estimation dates back to the seminal paper [5], which used a first-order asymptotic approximation of the distribution of the supremum of a certain Gaussian process that approximates the deviation of the kernel estimator from its mean. The use of the bootstrap to determine an appropriate width for confidence bands for a univariate density was proposed in [16] on a heuristic level and investigated in more detail by [22]. One of the main messages in [20] and [22] is that the application of the bootstrap leads to much smaller errors in coverage probability than the approach of [5].

In contrast to the papers mentioned above, we consider confidence bands of uniform size rather than bands with a varying size, proportional to $(\mathrm{var}(\widehat{\hat{f}}_h(x)))^{1/2}$. The latter bands seem to be somewhat more natural and they work well as long as they are restricted to some compact set on which the density f is bounded away from zero. One has to exclude regions where the density is low, because the performance of the bootstrap approximation deteriorates there. Such a truncation is not necessary with uniform bands, because then the problematic regions are automatically faded out.

Let t_α^* be the $(1 - \alpha)$-quantile of the distribution of $\sup_x\{|\hat{f}_h^{*,\,g}(x) - E\hat{f}_h^{*,\,g}(x)|\}$, that is,

$$(3.1) \qquad P\Big(\sup_{x\in\mathbb{R}^d}\{|\hat{f}_h^{*,\,g}(x) - E\hat{f}_h^{*,\,g}(x)|\} > t_\alpha^*\Big|X_1,\dots,X_n\Big) = \alpha.$$

For simplicity, we restrict the following considerations to the smoothed bootstrap. Using Theorem 3.2 instead of Theorem 3.1, one may derive results similar to the following theorems for t_α^* based on the standard bootstrap.

Let K_h be the smoothing operator defined by

$$(3.2) \qquad K_h(f)(x) = \int \frac{1}{h^d} K\Big(\frac{x-z}{h}\Big) f(z)\,dz.$$

Although statisticians usually focus on confidence intervals or bands for the density itself, we consider first simultaneous confidence bands for $K_h(f)$. The reason is that this problem is much easier to deal with, and with bands for $K_h(f)$ we have also more freedom to choose h. Theorems 2.1 and 3.2 and Proposition 3.1 imply the following theorem.

THEOREM 3.3.   *Suppose Assumptions* 1–6 *are fulfilled. Then*

$$P\big(K_h(f)(x) \in \big[\hat{f}_h(x) - t_\alpha^*, \hat{f}_h(x) + t_\alpha^*\big] \text{ for all } x \in \mathbb{R}^d\big)$$

$$= 1 - \alpha + O\Big(h^{d/2}(\log(n))^{3/2} + (nh^d)^{-1/2}(\log(n))^{3/2} + \mu_n^{1/2}\log(n)$$

$$+ h\log(n) + (nh^d)^{-1/4}(\log(n))^{5/4}\Big).$$

If

(3.3)
$$h = o\big((\log(n))^{-((3/d)\vee 1)}\big),$$

(3.4)
$$(nh^d)^{-1} = o\big(\log(n)\big)^{-5}\big)$$

and

(3.5)
$$\mu_n = o\big((\log(n))^{-2}\big),$$

then the confidence band will have asymptotically the prescribed coverage probability for $K_h(f)$. Certain *qualitative* features of f such as unimodality or monotonicity in some region remain valid for the smoothed version $K_h(f)$ under mild regularity assumptions on the kernel K. Hence, the confidence band for $K_h(f)$ can also be used as a criterion to assess whether there is enough evidence for such a feature. This is, of course, closely related to the formal test proposed in Section 3.3.

Since density estimation is an ill-posed inverse problem, there are certain limitations for any kind of *pointwise* inference about $f(x)$. For example, one cannot consistently distinguish between two densities that differ only on an interval shrinking at a sufficiently fast rate. This is in some way reflected in the bias problem one necessarily encounters in the construction of confidence bands for f. Nevertheless, there seems to be considerable interest in such bands, because they provide an easily accessible quantitative characterization of the precision of a nonparametric estimator.

To determine the width of the confidence band, we will use again the $(1-\alpha)$-quantile t_α^* of the bootstrapped maximum absolute deviation of the density estimator from its mean. We will obtain an asymptotically correct coverage probability, if the bias of $\hat{f}_h$ is of smaller order of magnitude than its standard deviation. Hence, the nominal coverage probability is asymptotically attained for an undersmoothed estimator $\hat{f}_h$, which, however, excludes the usual mean-squared-error optimal choice of h.

THEOREM 3.4. *Suppose Assumptions* 1–6 *are fulfilled. Then*

$$P\big(f(x) \in [\hat{f}_h(x) - t_\alpha^*, \hat{f}_h(x) + t_\alpha^*] \ \text{ for all } \ x \in \mathbb{R}^d\big)$$

$$= 1 - \alpha + O\Big(h^{d/2}(\log(n))^{3/2} + (nh^d)^{-1/2}(\log(n))^{3/2} + \mu_n^{1/2}\log(n)$$

$$+ h\log(n) + (nh^d)^{-1/4}(\log(n))^{5/4} + h^2(nh)^{1/2}(\log(n))^{1/2}\Big).$$

We see from this theorem that the confidence band has asymptotically the desired coverage probability, if, besides (3.3), (3.4) and (3.5),

(3.6)
$$h^2 = o\big((nh^d)^{-1/2}(\log(n))^{-1/2}\big)$$

is satisfied. (3.6) means that we have to undersmooth in order to make the bias of $\hat{f}_h$, which was not mimicked by the bootstrap, negligible. A well-known alternative consists in an explicit bias correction, which allows then also band-widths $h = h_n$ decaying at the mean-squared-error optimal rate $n^{-1/(4+d)}$.

We do not dwell on the effect of a data-driven bandwidth choice which is important for a real application of this method. Usually data-driven bandwidths $\hat{h}$ are designed to approximate a certain nonrandom bandwidth h_n. If $(\hat{h} - h_n)/h_n$ converges at an appropriate rate, then the estimators $\hat{f}_{\hat{h}}$ and $\hat{f}_{h_n}$ are sufficiently close to each other such that the results obtained in this paper remain valid; see [33] for a detailed investigation of these effects for pointwise confidence intervals in nonparametric regression.

3.3. *A nonparametric test.* Tests against a nonparametric alternative are an important tool to assess the appropriateness of a parametric or a semiparametric model. In contrast to classical tests such as the Kolmogorov–Smirnov or the Cramér–von Mises test, our density-based test is more powerful for local deviations from the assumed model. Moreover, by considering the supremum statistic, we exploit the whitening-by-windowing principle, which allows one to neglect the dependence structure. We allow for a composite hypothesis, that is,

$$H_0\colon f \in \mathscr{F},$$

where the only requirement is that the functional class $\mathscr{F}$ allows a faster rate of convergence than the full nonparametric model. We will assume:

ASSUMPTION 7. There exists an estimator $\widehat{\widehat{f}}$ of f such that, for $f \in \mathscr{F}$,

$$\sup_{x \in \mathbb{R}^d} \left\{ \left| \int h^{-d} K\left(\frac{x-z}{h}\right) [\widehat{\widehat{f}}(z) - f(z)] \, dz \right| \right\}$$
$$= o_P\big((nh^d)^{-1/2}(\log(n))^{-1/2}\big).$$

Note that Assumption 7 is in particular fulfilled if

$$\sup_{x \in \mathbb{R}^d} \left\{ \left| \widehat{\widehat{f}}(x) - f(x) \right| \right\} = o_P\big((nh^d)^{-1/2}(\log(n))^{-1/2}\big).$$

In the case $d = 1$, this includes some parametric models,

$$\mathscr{F} = \{f_\theta, \theta \in \Theta\}.$$

In the higher-dimensional case, one may test for parametric but also for certain semiparametric models such as a multiplicative nonparametric model that corresponds to the assumption that the components of the X_i's are independent,

$$\mathscr{F} = \left\{ f(x) = \prod_{i=1}^{d} f_i(x_i) \,\Big|\, f_i \text{ "sufficiently smooth"} \right\},$$

or a semiparametric model proposed in [18],

$$\mathscr{F} = \left\{ f(x) = f_0(x) \prod_{i=1}^{M} f_i(\alpha_i' x) \,\Big|\, f_i \text{ "sufficiently smooth"} \right\}.$$

In accordance with our theory above, we consider the maximum absolute deviation between $\hat{f}_h$ and $K_h(\widehat{\hat{f}})$, that is,

$$T = \sup_{x \in \mathbb{R}^d} \left\{ \left\| \hat{f}_h(x) - \int h^{-d} K\left(\frac{x-z}{h} \right) \widehat{\hat{f}}(z) \, dz \right\| \right\}.$$

The next theorem shows that the prescribed error of the first kind is asymptotically guaranteed.

THEOREM 3.5. *Suppose Assumptions* 1–7 *as well as* (3.3), (3.4) *and* (3.5) *are fulfilled. Then*

$$P_{H_0}(T > t_\alpha^*) \to \alpha \quad as \ n \to \infty.$$

REMARK 1. It seems that L_2-tests such as those proposed by [5] for the density and by [25] in the regression setup, are the most popular ones among nonparametric statisticians. Such tests can be optimal for testing against smooth alternatives, whereas supremum-type tests have less power in such a situation. On the other hand, supremum-type tests can also outperform L_2-tests for testing against local alternatives having the form of sharp peaks; see [28] and [42] for more details.

The methodology in this paper is obviously restricted to supremum-type tests. The negligibility of weak dependence for L_2-tests is proved in [43] and [35] by arguments different from those in the present paper.

4. Discussion.

4.1. *Mixing plus extra conditions on joint densities.* By now, strong mixing and absolute regularity have been accepted as being benchmark conditions to characterize weak dependence. A lot of effort has been devoted to show that estimation problems under weak dependence allow the same *rates* of convergence as under independence.

However, as we see in this paper, as well as in [39], [30] and [26], suitable extra conditions on the joint densities lead to qualitatively much stronger results: then we obtain asymptotic equivalence on the level of *constants*. In many instances, such an extra condition is not very restrictive and leads to an immediate applicability of important statistical methods developed under the assumption of independence.

4.2. *Does a multiscale approach lead to a better approximation?* In many cases one obtains better rates for strong approximations by a multiscale approach based on a dyadic partition of the interval of interest. A classical example is the construction in [27]. A dyadic approximation scheme has also been employed in [34] for constructing a strong approximation of nonparametric autoregression by nonparametric regression. The simultaneous consideration of different resolution scales makes sense for the above examples, because the *relative* approximation rate deteriorates as one moves to smaller intervals.

However, in our context, the possibility of approximating density estimators under weak dependence by density estimators under independence is essentially based on the "whitening-by-windowing" principle. Therefore, the relative approximation rate becomes even better for finer scales. It seems to be unlikely that a multiscale approach leads to better approximation rates between kernel estimators from both models.

4.3. *Optimality of the approximation.* Our basic result (Proposition 2.1) is stronger than usual as well as stronger than necessary. For our particular purpose of constructing a strong approximation of kernel estimators it is not necessary at all that most of the observations coincide. Therefore, it is natural to ask whether our pairing on the level of exact coincidence of random variables is actually an appropriate method.

However, it seems that our pairing is indeed the closest possible between nonparametric estimators in both models, perhaps up to some logarithmic factor. Suppose, for example, that f has support $[0, 1]$ and we have such a pairing of $(X_1, \ldots, X_n)$ with $(Y_1, \ldots, Y_n)$ that the corresponding histogram estimators satisfy

$$\sup_{k=1, \ldots, [h^{-1}]+1} \left\{ (nh)^{-1} \big[\#\{i \mid X_i \in [(k-1)h, kh)\} - \#\{i \mid Y_i \in [(k-1)h, kh)\} \big] \right\}$$

$$= \tilde{O}\big(r_n, n^{-\lambda}\big).$$

Let $\tilde{X}_i = h[X_i/h]$ and $\tilde{Y}_i = h[Y_i/h]$. Then

$$n^{-1/2}\bigg[\sum_i \tilde{X}_i - \sum_i \tilde{Y}_i \bigg]$$

$$= n^{-1/2} \sum_{k=1}^{[h^{-1}]+1} (k-1)h\bigg[\#\{i \mid X_i \in [(k-1)h, kh)\}$$

$$- \#\{i \mid Y_i \in [(k-1)h, kh)\}\bigg]$$

$$= \tilde{O}\big(n^{1/2} r_n, n^{-\lambda}\big).$$

If r_n were of order $o(n^{-1/2})$, then the asymptotic distributions of $n^{-1/2} \sum \tilde{X}_i$ and $n^{-1/2} \sum \tilde{Y}_i$ would coincide, which is not necessarily the case under our conditions. Hence, although it is not impossible that one can find a closer pairing of the nonparametric estimators at *one single point*, it seems that there does not exist an essentially better approximation in the uniform norm.

4.4. *Are these nonstandard proofs really necessary?* Compared to existing literature on similar topics, the methods of proof in this paper are somehow nonstandard. In particular, all proofs are based on certain constructive pairing techniques instead of the commonly used first-order approximation by the supremum of the limiting Gaussian process. This is done for the following two reasons: First, a purely analytical derivation of the asymptotic distribution

of the maximal deviation between $\hat{f}_h$ and its expectation is presumably very technical and neither pleasant for the author nor for the reader. Second, it is well known that first-order asymptotic theory leads to poor rates of convergence in this context. Once we had used such an approximation at any point, we were not able to prove that the bootstrap actually leads to better rates of convergence.

There exists an extensive literature on strong approximations for empirical cumulative distribution functions by certain Gaussian processes. For example, [8] and [11] showed for absolutely regular processes that the cumulative distribution function can be approximated by a Gaussian process with an error of order $n^{-1/2-\lambda}$, for a certain $\lambda > 0$. Such a result can also be used to show that a kernel density estimator is approximated by a certain Gaussian process. However, in dependence on the value of λ, there are limitations for the significance of such results. Kernel estimators with small bandwidths h will require more localized approximations.

4.5. *Two stages of generating time-series data.* The successful simultaneous embedding of time-series data and i.i.d. data in a common Poisson process provides a new view on the generation of random variables from stochastic processes. Actually, our embedding shows that the generation of each new datum can be construed as a two-stage process: first, the influence of the past is reflected by the specific manner in which the graphs $(tf_{X_i\mid\mathscr{F}_1^{i-1}}(x), x)$ spread out as $t \to \infty$; and second, the remaining uncertainty can be driven by an independent process. The result of our embedding procedure is comparable to a result in an i.i.d. situation because the determining conditions are on average the same as those for the i.i.d. counterpart. This has of course similarities to well-known embeddings of martingales in Wiener processes, which then lead to strong approximations by partial sum processes of i.i.d. random variables.

4.6. *Alternative bootstrap methods.* Even if the effect of the dependence vanishes asymptotically, it is still present in higher order terms. Instead of neglecting it, one could also try to mimic the dependence structure by the bootstrap. One standard tool is the blockwise bootstrap introduced in [29]. In [7] it is shown that the blockwise bootstrap consistently estimates the distribution of a multivariate empirical process based on α-mixing observations, and this result is applied to a nonlinear estimator of a finite-dimensional parameter. On the other hand, the blockwise bootstrap requires the estimation of many more features of the data-generating process, which in turn leads to new fluctuations of the resulting estimates. It seems to be an important and challenging task to explore by how much such an approach can improve the rate of approximation.

4.7. *Existing results for nonparametric estimation in the supremum norm.* There already exists some literature on density estimation under weak dependence where the error is measured in the uniform norm. Under appropriate α- or β-mixing conditions, it has been shown that appropriate kernel estima-

tors can attain the same *rate* of uniform convergence that is optimal in the i.i.d. case; see [45], [44], [3] and [4]. The proofs of these results are based on blocking techniques which allow one to replace dependent blocks of observations by independent ones. For our purpose of constructing confidence bands and supremum-type tests we need more exact approximations of the distribution of the supremum deviation, which requires a different method of proof.

4.8. *Other nonparametric estimators.* The whitening-by-windowing principle, even in its global version described in this article, is closely connected with the occurence of rare events. It is quite obvious that it also applies to a variety of other nonparametric estimators such as histogram estimators, smoothed histogram estimators or linear wavelet estimators, provided the corresponding analogue to the bandwidth in kernel estimation tends to zero. Moreover, although first-order asymptotics of empirical versions of the Fourier coefficients does depend on the dependence structure, one can show that certain Fourier series estimators also obey the whitening-by-windowing principle. To be specific, suppose it is known that $\text{supp}(f) \subseteq [0, 1]$, which gives rise to the following Fourier series estimator:

$$\hat{f}_n(x) = 1 + \sum_{k=1}^{\infty} r_k \big[\hat{c}_k \cos(2\pi k x) + \hat{s}_k \sin(2\pi k x) \big],$$

where $\hat{c}_k = n^{-1} \sum 2 \cos(2\pi k X_i)$ and $\hat{s}_k = n^{-1} \sum 2 \sin(2\pi k X_i)$. Assume further that $1 \geq r_1 \geq r_2 \geq \cdots$ and $\sum_k |r_k| = O(c_n)$. It is easy to see that $\hat{f}_n$ can be rewritten as

$$\hat{f}_n(x) = 1 + n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{\infty} r_k 2 \cos(2\pi k(x - X_i)).$$

The kernel $K_n(x, z) = 1 + \sum r_k 2 \cos(2\pi k(x - z))$ does not have a shrinking support, however, by using the well-known fact $\sum_{k=1}^{N} \cos(2\pi k u) = \cos(\pi(N + 1)u) \sin(\pi N u)/\sin(\pi u)$, it can be shown that

$$\sum_{s: \; -1-1/c_n < (s-1)/c_n < 1} \; \sup_{(s-1)/c_n \leq x - z \leq s/c_n} \big\{ \big| K_n(x, z) \big| \big\} = O\big(c_n \log(c_n)\big).$$

Hence, we obtain analogously to the proof of Theorem 2.1 a strong approximation of $\hat{f}_n$ by its analogue $\tilde{f}_n$ in the i.i.d. model

$$\sup_{x \in [0, 1]} \big\{ \big| \hat{f}_n(x) - \tilde{f}_n(x) \big| \big\} = \tilde{O}\big([n^{-1/2} \log(n) + n^{-1} c_n \log(n)], n^{-\lambda} \big).$$

5. Proofs.

PROOF OF LEMMA 2.1. Define

$$T_i = \big(U_{j_i} - \tau_{V_{j_i}}^{(i-1)} \big) \big/ f_{X_i | \mathscr{F}_1^{i-1}}(V_{j_i}).$$

We split up

$$\tag{5.1} \tau_v^{(n)} = \sum_{i=1}^n T_i f_{X_i \mid \mathscr{F}_1^{i-1}}(v) = nf(v) + R_1 + R_2,$$

where

$$R_1 = \sum_{i=1}^n T_i \big[f_{X_i \mid \mathscr{F}_1^{i-1}}(v) - f_{X_i \mid \mathscr{F}_{i-\gamma_n}^{i-1}}(v) \big],$$

$$R_2 = \sum_{i=1}^n \big[T_i f_{X_i \mid \mathscr{F}_{i-\gamma_n}^{i-1}}(v) - f(v) \big]$$

and γ_n is chosen such that $\gamma_n \geq \max\{\log(n)/(4C_2), \lambda \log(n)/C_3\}$, $\gamma_n = O(\log(n))$, where C_2 and C_3 are given by Assumption 2. According to Assumption 2, we have that

$$\tag{5.2} \sup_v \big\{ \big| f_{X_i \mid \mathscr{F}_1^{i-1}}(v) - f_{X_i \mid \mathscr{F}_{i-\gamma_n}^{i-1}}(v) \big| \big\} = \tilde{O}(n^{-1/4}, n^{-\lambda}).$$

It is easy to see that the vector $(T_1, \ldots, T_n)$ is independent of $(X_1, \ldots, X_n)$ and that $T_i \sim \mathrm{Exp}(1)$ are i.i.d. [To see this, consider for a moment the situation where we start with independent vectors $(\tilde{T}_1, \ldots, \tilde{T}_n)$ and $(\tilde{X}_1, \ldots, \tilde{X}_n)$, where $\tilde{T}_i \sim \mathrm{Exp}(1)$ are i.i.d. and $\mathscr{L}(\tilde{X}_i \mid \tilde{X}_{i-1} = x_{i-1}, \ldots, \tilde{X}_1 = x_1) = \mathscr{L}(X_i \mid X_{i-1} = x_{i-1}, \ldots, X_1 = x_1)$. Now we easily see that the conditional distributions $\mathscr{L}((T_i, X_i) \mid (T_{i-1}, X_{i-1}), \ldots, (T_1, X_1))$ and $\mathscr{L}((\tilde{T}_i, \tilde{X}_i) \mid (\tilde{T}_{i-1}, \tilde{X}_{i-1}), \ldots, (\tilde{T}_1, \tilde{X}_1))$ coincide, which implies that $(T_1, \ldots, T_n)$ and $(X_1, \ldots, X_n)$ are actually independent.]

R_1 is a weighted sum of the T_i's, where $(T_1, \ldots, T_n)$ is independent of the weights $([f_{X_i \mid \mathscr{F}_1^{i-1}}(v) - f_{X_i \mid \mathscr{F}_{i-\gamma_n}^{i-1}}(v)], i = 1, \ldots, n)$. Hence, we obtain by Theorem 4 of [1] that

$$P\left(|R_1| \geq \kappa \sqrt{\sum [f_{X_i \mid \mathscr{F}_1^{i-1}}(v) - f_{X_i \mid \mathscr{F}_{i-\gamma_n}^{i-1}}(v)]^2} \sqrt{\log(n)} \Big| X_1, \ldots, X_n \right)$$

$$= \tilde{O}(n^{-\kappa^2/2}, n^{-\lambda})$$

holds for arbitrary $\kappa < \infty$ and uniformly in $X_1, \ldots, X_n$. This implies that

$$\tag{5.3} R_1 = \tilde{O}\left(n^{1/2} \sqrt{\log(n)}, n^{-\lambda} \right).$$

To estimate R_2, we consider blocks of observations $\{X_j, j \in J_i\}$, where $J_i = \{(i-1)\rho_n - \gamma_n + 1, \ldots, i\rho_n\}$ and $\rho_n \geq (\lambda+1)\log(n)/C_1 + \gamma_n - 1$, $\rho_n = O(\log(n))$. Without loss of generality, we consider the blocks with odd numbers. Note that we have

$$\beta\big(\sigma(\{X_j, j \in J_i\}), \sigma(\{X_j, j \in J_k\}, k = i+2, i+4, \ldots) \big)$$

$$\leq C \exp(-C_1(\rho_n - \gamma_n + 1)).$$

By Proposition 2 of [10], page 407, there exists a sequence of independent blocks $\{\tilde{X}_j, j \in J_i\}$, i odd, where the $\tilde{X}_j$'s are independent of the T_j's, $\mathscr{L}((\tilde{X}_j, j \in J_i)) = \mathscr{L}((X_j, j \in J_i))$, and

$$
\begin{aligned}
P((\tilde{X}_j, j \in J_i) \neq (X_j, j \in J_i)) &\leq C \exp(-C_1(\rho_n - \gamma_n + 1)) \\
&= O(n^{-\lambda-1}).
\end{aligned}
$$

(5.4)

Now we have

$$
\operatorname{var}\left(\sum_{j=(i-1)\rho_n+1}^{i\rho_n} T_j f_{\tilde{X}_j \mid \tilde{X}_{j-1}, \ldots, \tilde{X}_{j-\gamma_n}}(v) \right)
$$

$$
\leq \rho_n \sum_{j=(i-1)\rho_n+1}^{i\rho_n} \operatorname{var}\left(T_j f_{\tilde{X}_j \mid \tilde{X}_{j-1}, \ldots, \tilde{X}_{j-\gamma_n}}(v) \right),
$$

which implies, again by Theorem 4 of [1], that

$$
\sum_{i \text{ odd}} \sum_{j=(i-1)\rho_n+1}^{i\rho_n} \left(T_j f_{X_j \mid \mathscr{F}_{i-\gamma_n}^{i-1}}(v) - f(v) \right)
$$

$$
= \tilde{O}\left(\sqrt{\rho_n \sum_{i \text{ odd}} \sum_{j=(i-1)\rho_n+1}^{i\rho_n} \operatorname{var}\left(T_j f_{X_j \mid \mathscr{F}_{i-\gamma_n}^{i-1}}(v) \right) } \sqrt{\log(n)}, n^{-\lambda} \right)
$$

$$
= \tilde{O}\left(n^{1/2} \log(n), n^{-\lambda} \right).
$$

An analogous result can be shown for the blocks with even numbers, which implies, in conjunction with (5.4), that

(5.5) $$ R_2 = \tilde{O}\left(n^{1/2} \log(n), n^{-\lambda} \right). $$

The proof of the assertion about $\tilde{\tau}_v^{(n)}$ is analogous, which completes the proof. $\square$

PROOF OF LEMMA 2.2. We prove the assertion only for $\sup_{v \in [a, b]}\{|\tau_v^{(n)} - nf(v)|\}$. Let $\mathscr{N}_n = \{v_1, \ldots, v_{\gamma_n}\}$ be an $n^{-1/2}$-net for the hyperrectangle $[a, b]$ of cardinality $\#\mathscr{N}_n = \gamma_n = O(n^{d/2})$. It is clear from Lemma 2.1 that

(5.6) $$ \sup_{1 \leq j \leq \gamma_n} \{ |\tau_{v_j}^{(n)} - nf(v_j)| \} = \tilde{O}\left(n^{1/2} \log(n), n^{-\lambda} \right) $$

holds. Let $v \in [a, b]$ be arbitrary. Then there exists a $j(v) \in \{1, \ldots, \gamma_n\}$ such that $\|v - v_{j(v)}\| = O(n^{-1/2})$. Since the T_i are i.i.d., we have according to [1], Theorem 4, that

$$
\sum_{i=1}^{n} T_i = O(n) + \tilde{O}\left(n^{1/2} \sqrt{\log(n)}, n^{-\lambda} \right).
$$

Because of $\sup_i \sup_v \{|f_{X_i|\mathscr{F}_1^{i-1}}(v) - f_{X_i|\mathscr{F}_1^{i-1}}(v_{j(v)})|\} = O(n^{-1/2})$, we have that

(5.7)
$$\left|\tau_v^{(n)} - \tau_{v_{j(v)}}^{(n)}\right| \leq \sum_{i=1}^n T_i \left|f_{X_i|\mathscr{F}_1^{i-1}}(v) - f_{X_i|\mathscr{F}_1^{i-1}}(v_{j(v)})\right|$$
$$= \tilde{O}(n^{1/2}, n^{-\lambda}),$$

which yields, in conjunction with $|f(v) - f(v_{j(v)})| = O(n^{-1/2})$, the assertion. □

PROOF OF PROPOSITION 2.1. (i) *Proof of the assertion for fixed* $[a, b]$.
Let $\Delta_{n1} = \{X_1, \ldots, X_n\}\Delta\{Z_1, \ldots, Z_\nu\}$. According to Lemma 2.2, we have that

$$\Delta_{n1} \cap [a, b] \subseteq \left\{V_j \in [a, b]\big| nf(v) - C_\lambda n^{1/2}\log(n) \leq U_j \leq nf(v)\right.$$
$$\left. + C_\lambda n^{1/2}\log(n)\right\}$$

holds with a probability exceeding $1 - O(n^{-\lambda})$, where C_λ is an appropriate constant. To get an estimate for the cardinality of the latter set, we apply an exponential inequality to the restriction N_D of the Poisson process N to

$$D = \left\{(u, v)\big| nf(v) - C_\lambda n^{1/2}\log(n) \leq u \leq nf(v) + C_\lambda n^{1/2}\log(n), v \in [a, b]\right\}.$$

It is clear that N_D is a Poisson process with intensity $\mu(D) = O(n^{1/2}\log(n) \prod(b_i - a_i))$.

If $\mu(D) \geq (8/3)\lambda\log(n)$, then we obtain by Inequality 14.5.1 on page 569 and Proposition 11.1.1(10) on page 441 in [40] that

(5.8) $\quad P\big(N_D > 2\mu(D)\big) \leq \exp\left(-\frac{\mu(D)}{2}\psi(1)\right) \leq \exp\left(-\frac{\mu(D)}{2}\frac{3}{4}\right) = O(n^{-\lambda}).$

If $\mu(D) < (8/3)\lambda\log(n)$, then we obtain, again by inequality 14.5.1 and Proposition 11.1.1(10) of [40], that

$$P\left(N_D - \mu(D) > \frac{8}{3}\lambda\log(n)\right)$$

(5.9)
$$\leq \exp\left(-\frac{((8/3)\lambda\log(n))^2}{2\mu(D)}\psi\left(\frac{(8/3)\lambda\log(n)}{\mu(D)}\right)\right)$$
$$\leq \exp\left(-\frac{((8/3)\lambda\log(n))^2}{2\mu(D)}\frac{3}{4}\frac{\mu(D)}{(8/3)\lambda\log(n)}\right) = O(n^{-\lambda}).$$

Inequalities (5.8) and (5.9) imply

(5.10) $\quad \#(\Delta_{n1} \cap [a, b]) = \tilde{O}\left(\left[n^{1/2}\left(\prod_{i=1}^d (b_i - a_i)\right) + 1\right]\log(n), n^{-\lambda}\right).$

(ii) *Uniformity over the set of all* $[a, b]$ *with* $\max_i\{b_i - a_i\} \leq Cn^C$. Let F_k be the cumulative distribution function of the kth component of X_1. We consider the hyperrectangles

$$I_{s,t} = [u_{s_1}^{(1)}, u_{t_1}^{(1)}) \times \cdots \times [u_{s_d}^{(d)}, u_{t_d}^{(d)}),$$

where $0 \leq s_k < t_k \leq n$, $u_s^{(k)} = F_k^{-1}(s/n)$, $F_k^{-1}(0) = -\infty$ and $F_k^{-1}(1) = \infty$. (Without loss of generality, we prove the assertion for the case that the F_k are continuous. The result in the general case follows by simple modifications of the arguments.)

Since the number of these hyperrectangles is of an algebraic order, we obtain from (5.10) that

$$(5.11) \qquad \#\big(\Delta_{n1} \cap I_{s,\,t}\big) = \tilde{O}\Bigg(\bigg[n^{1/2} \prod_{k=1}^{d} \big(u_{t_k}^{(k)} - u_{s_k}^{(k)}\big) + 1\bigg]\log(n), n^{-\lambda}\Bigg)$$

holds for all s, t with $\max_k\{u_{t_k}^{(k)} - u_{s_k}^{(k)}\} \leq Cn^C$.

Now we consider the slices

$$I_j^{(k)} = (-\infty, \infty)^{k-1} \times \big[F_k^{-1}\big(u_{j-1}^{(k)}\big), F_k^{-1}\big(u_j^{(k)}\big)\big) \times (-\infty, \infty)^{n-k},$$

Since they are of unbounded size, we cannot use (5.10) to estimate the cardinality of the sets $\Delta_{n1} \cap I_j^{(k)}$. However, since $P(X_1 \in I_j^{(k)}) = O(n^{-1})$, we can find sufficiently small upper estimates via exponential inequalities.

For i.i.d. random variables $Z_1, \ldots, Z_n$ with $EZ_1 = 0$ and $|Z_1| \leq K_n$ we obtain by Bernstein's inequality (see, e.g., [40], page 855)

$$(5.12) \qquad \sum Z_i = \tilde{O}\big(\sqrt{n \operatorname{var}(Z_1)}\sqrt{\log(n)} + K_n \log(n), n^{-\lambda}\big).$$

Using this in conjunction with the same blocking technique as in the proof of Lemma 2.1, we obtain

$$(5.13) \qquad \begin{aligned} \#\big(\{X_1, \ldots, X_n\} \cap I_j^{(k)}\big) &= \sum_{i \text{ odd}}\bigg[\sum_{l=(i-1)\rho_n+1}^{i\rho_n} I\big(X_l \in I_j^{(k)}\big)\bigg] + \sum_{i \text{ even}}[\cdots] \\ &= \tilde{O}\big(\log(n), n^{-\lambda}\big). \end{aligned}$$

Further, we obtain from (5.8) and (5.9) that

$$(5.14) \qquad \#\big(\{Z_1, \ldots, Z_\nu\} \cap I_j^{(k)}\big) = \tilde{O}\big(\log(n), n^{-\lambda}\big).$$

According to (5.11), (5.13) and (5.14), there exists a set of events Ω_n with $P(\Omega_n) = 1 - O(n^{-\lambda})$,

$$\#\big(\Delta_{n1} \cap I_{s,\,t}\big) \leq C\bigg[n^{1/2} \prod_{k=1}^{d} (u_{t_k}^{(k)} - u_{s_k}^{(k)}) + 1\bigg]\log(n)$$

for all s, t with $\max_k\{u_{t_k}^{(k)} - u_{s_k}^{(k)}\} \leq Cn^C$, and

$$\#\big(\Delta_{n1} \cap I_j^{(k)}\big) \leq C \log(n).$$

Let now $[a, b]$ be arbitrary with $\max_k\{b_k - a_k\} \leq Cn^C$. Then there exist s, t such that

$$I_{s,\,t} \subseteq [a, b] \subseteq I_{s,\,t} \cup \bigg(\bigcup_{k=1}^{d} I_{s_k}^{(k)} \cup I_{t_{k+1}}^{(k)}\bigg),$$

which implies, for $\omega \in \Omega_n$,

$$\#\big(\Delta_{n1} \cap [a, b]\big) \le \#(\Delta_{n1} \cap I_{s,t}) + \sum_{k=1}^{d} \#\big(\Delta_{n1} \cap (I_{s_k}^{(k)} \cup I_{t_{k+1}}^{(k)})\big)$$

(5.15)

$$\le C\bigg[n^{1/2}\bigg(\prod_{k=1}^{d}(b_k - a_k)\bigg) + 1\bigg]\log(n).$$

(iii) *Uniformity over all hyperrectangles.* Under the additional condition $P(X_1 \in [a, b]) \le C \prod_{i=1}^{d}[(b_i - a_i) \wedge 1]$, we have that

$$u_{j+1}^{(k)} - u_j^{(k)} \ge Cn^{-1/2}.$$

If $[a, b]$ contains some $I_{s,t}$ with $\max\{u_{t_k}^{(k)} - u_{s_k}^{(k)}\} \ge n^{d/2}$, then

$$\prod_{k=1}^{d}(b_k - a_k) \ge \prod_{k=1}^{d}\big(u_{t_k}^{(k)} - u_{s_k}^{(k)}\big) \ge Cn^{1/2}.$$

In this case (5.15) is trivially fulfilled. Otherwise the assertion follows from (ii).

Finally, since independence is a special case of weak dependence, the second part of the assertion concerning $\{Y_1, \dots, Y_n\}\Delta\{Z_1, \dots, Z_\nu\}$ is also proved. $\square$

PROOF OF THEOREM 3.1. It is easy to show that

(5.16) $$\sup_{x \in \mathbb{R}^d}\big\{\big|\hat{f}_g(x) - f(x)\big|\big\} = \tilde{O}(\mu_n, n^{-\lambda}).$$

We show in this proof that there exists a pairing of the random variables $Y_1, \dots, Y_n$ and $X_1^{*,g}, \dots, X_n^{*,g}$ such that

(5.17)
$$\sup_{x \in \mathbb{R}^d}\big\{\big|[\tilde{f}_h(x) - E\tilde{f}_h(x)] - [\hat{f}_h^{*,g}(x) - E\hat{f}_h^{*,g}(x)]\big|\big\}$$
$$= \tilde{O}\Big((nh^d)^{-1/2}\mu_n^{1/2}\sqrt{\log(n)} + (nh^d)^{-1}\log(n), n^{-\lambda}\Big).$$

The assertion of the theorem follows then in conjunction with Theorem 2.1.

Since $Y_i = X_i^{*,g}$ if $B_i = 1$, we have

(5.18)
$$[\tilde{f}_h(x) - E\tilde{f}_h(x)] - [\hat{f}_h^{*,g}(x) - E\hat{f}_h^{*,g}(x)]$$
$$= \frac{1}{nh^d}\sum_i\bigg\{I(B_i = 0)\bigg[K\Big(\frac{x - Y_i}{h}\Big) - K\Big(\frac{x - X_i^{*,g}}{h}\Big)\bigg]$$
$$- \int K\Big(\frac{x - z}{h}\Big)[f(z) - \hat{f}_g(z)]\,dz\bigg\}.$$

To estimate the right-hand side of (5.18), we distinguish between two sets of hypercubes $I_k = [(k_1 - 1)h, k_1 h] \times \cdots \times [(k_d - 1)h, k_d h]$:

$$\mathcal{K}_1 = \left\{ k \,\bigg|\, \int_{I_k} |\hat{f}_g(z) - f(z)| \, dz \geq n^{-\tau} \right\},$$

$$\mathcal{K}_2 = \left\{ k \,\bigg|\, \int_{I_k} |\hat{f}_g(z) - f(z)| \, dz < n^{-\tau} \right\},$$

where $\tau > \lambda + 1$. The cardinality of the set $\mathcal{K}_1$ is $O(n^\tau)$.

First we investigate the case of $x \in \mathscr{X} = \bigcup_{k \in \mathcal{K}_1} \bigcup_{x \in I_k} \mathrm{supp}(K((\cdot - x)/h)))$. Let $\mathcal{N}_n = \{x_1, \ldots, x_{c_n}\}$ be an n^{-1}-net of $\mathscr{X}$, where $c_n = O(n^{d+\tau})$. Because of

$$\mathrm{var}([\tilde{f}_h(x) - E\tilde{f}_h(x)] - [\hat{f}_h^{*,g}(x) - E\hat{f}_h^{*,g}(x)])$$

$$= O\left(n(nh^d)^{-2} \int_{\mathrm{supp}(K((x-\cdot)/h))} |f(z) - \hat{f}_g(z)| \, dz \right) = O((nh^d)^{-1}\mu_n),$$

we obtain by (5.12) that

(5.19)
$$\sup_{x \in \mathcal{N}_n} \left\{ \left| [\tilde{f}_h(x) - E\tilde{f}_h(x)] - [\hat{f}_h^{*,g}(x) - E\hat{f}_h^{*,g}(x)] \right| \right\}$$
$$= \tilde{O}\left((nh^d)^{-1/2}\mu_n^{1/2}\sqrt{\log(n)} + (nh^d)^{-1}\log(n), n^{-\lambda} \right).$$

Let $x \in \mathscr{X}$ be arbitrary. Then there exists a $j(x) \in \{1, \ldots, c_n\}$ such that $\|x - x_{j(x)}\| = O(n^{-1})$. Since

$$\left| \tilde{f}_h(x) - \tilde{f}_h(x_{j(x)}) \right| + \left| \hat{f}_h^{*,g}(x) - \hat{f}_h^{*,g}(x_{j(x)}) \right| = O(h^{-d}n^{-1})$$

is satisfied with probability 1, we have that

(5.20)
$$\sup_{x \in \mathscr{X}} \left\{ \left| [\tilde{f}_h(x) - E\tilde{f}_h(x)] - [\hat{f}_h^{*,g}(x) - E\hat{f}_h^{*,g}(x)] \right| \right\}$$
$$= \tilde{O}\left((nh^d)^{-1/2}\mu_n^{1/2}\sqrt{\log(n)} + (nh^d)^{-1}\log(n), n^{-\lambda} \right).$$

Concerning the set $\mathcal{K}_2$, we show, analogously to the corresponding part of the proof of Proposition 2.1, that

$$\sup_{k \in \mathcal{K}_2} \left\{ \#(\{Y_1, \ldots, Y_n\} \Delta \{X_1^{*,g}, \ldots, X_n^{*,g}\}) \cap I_k \right\} = \tilde{O}(\log(n), n^{-\lambda}),$$

which implies

(5.21)
$$\sup_{x \notin \mathscr{X}} \left\{ \left| [\tilde{f}_h(x) - E\tilde{f}_h(x)] - [\hat{f}_h^{*,g}(x) - E\hat{f}_h^{*,g}(x)] \right| \right\}$$
$$= \tilde{O}((nh^d)^{-1}\log(n), n^{-\lambda}).$$

Equations (5.20) and (5.21) imply (5.17), which yields the assertion in conjunction with Theorem 2.1. □

PROOF OF THEOREM 3.2. As already mentioned, we cannot use the idea of the proof of Theorem 3.1, because the probability measures P and P_n are orthogonal. However, we may exploit the pairing of $X_1, \ldots, X_n$ and $X_1^{*, g}, \ldots,$ $X_n^{*, g}$ used for proving Theorems 2.1 and 3.1 as an intermediate step to show the closeness of $[\hat{f}_h(x) - E\hat{f}_h(x)]$ and $[\hat{f}_h^{*, 0}(x) - E\hat{f}_h^{*, 0}(x)]$. In addition to this paring we pair the $X_i^{*, 0}$'s with the $X_i^{*, g}$'s in such a way that

$$(5.22) \qquad \left\| X_i^{*, 0} - X_i^{*, g} \right\| \leq \sqrt{d}\, g$$

holds with probability 1. Since then

$$K\left(\frac{x - X_i^{*, 0}}{h} \right) - K\left(\frac{x - X_i^{*, g}}{h} \right) = O\left(\frac{g}{h} \right),$$

we obtain by an approximation on a sufficiently fine grid that

$$\sup_{x \in \mathbb{R}^d} \left\{ \left| [\hat{f}_h^{*, 0}(x) - E\hat{f}_h^{*, 0}(x)] - [\hat{f}_h^{*, g}(x) - E\hat{f}_h^{*, g}(x)] \right| \right\}$$

$$= O\left((nh^d)^{-1/2}(g/h)\sqrt{\log(n)}, n^{-\lambda} \right).$$

This yields, in conjunction with Theorem 3.1, that

$$\sup_{x \in \mathbb{R}^d} \left\{ \left| [\hat{f}_h(x) - E\hat{f}_h(x)] - [\hat{f}_h^{*, 0}(x) - E\hat{f}_h^{*, 0}(x)] \right| \right\}$$

$$= O\left(n^{-1/2}\log(n) + (nh^d)^{-1}\log(n) \right.$$

$$\left. + (nh^d)^{-1/2}\sqrt{\log(n)}\left[(ng^d)^{-1/2}\sqrt{\log(n)} + g/h \right], n^{-\lambda} \right). \qquad \square$$

PROOF OF PROPOSITION 3.1. *Upper estimates for Poisson probabilities.* Before we turn directly to the proof of the assertion, we first derive some technical results to be applied in the main part of this proof.

Let $P_s(\{k\}) = e^{-s}s^k/k!$ be a Poisson probability. Let $k = s \pm s^{1/2}\sqrt{r_s \log(s)}$ be an integer. Then we obtain by formula 11.9.19 in [40], page 486, that

$$P_s(\{k\}) = \frac{\exp(-a(k))}{\sqrt{2\pi s}} \frac{1}{\sqrt{1 \pm s^{-1/2}\sqrt{r_s \log(s)}}}$$

$$\times \exp\left(-\frac{r_s \log(s)}{2}\psi\left(\pm s^{-1/2}\sqrt{r_s \log(s)} \right) \right),$$

where $1/(12k + 1) < a(k) < 1/(12k)$. Using the estimate for $\psi(\cdot)$ given in Proposition 11.1.1(10) in [40], page 441, we get, for an appropriate c_λ,

$$(5.23) \qquad P_s(\{k\}) = O(s^{-\lambda}) \quad \text{if } |k - s| \geq s^{1/2}\sqrt{c_\lambda \log(s)}.$$

For $k < s + s^{1/2}\sqrt{c_\lambda \log(s)}$ we obtain that

$$\frac{P_s(\{k, k+1, \ldots\})}{P_s(\{k\})}$$

$$= 1 + \frac{s}{k+1} + \frac{s}{k+1}\frac{s}{k+2} + \cdots$$

$$(5.24) \quad \geq \left[\left(\frac{s}{c_\lambda \log(s)}\right)^{1/2}\right]\left(\frac{s}{s + \sqrt{sc_\lambda \log(s)} + [\sqrt{s/(c_\lambda \log(n))}]}\right)^{[\sqrt{s/(c_\lambda \log(s))}]}$$

$$\geq \left[\left(\frac{s}{c_\lambda \log(s)}\right)^{1/2}\right]\left(1 - \frac{\sqrt{sc_\lambda \log(s)} + \sqrt{s/(c_\lambda \log(s))}}{s}\right)^{[\sqrt{s/(c_\lambda \log(s))}]}$$

$$\geq C\sqrt{s/\log(s)}.$$

Analogously we get, for $k > s - s^{1/2}\sqrt{c_\lambda \log(s)}$, that

$$\frac{P_s(\{k, k-1, \ldots\})}{P_s(\{k\})}$$

$$= 1 + \frac{k}{s} + \frac{k}{s}\frac{k-1}{s} + \cdots$$

$$(5.25) \quad \geq \left[\left(\frac{s}{c_\lambda \log(s)}\right)^{1/2}\right]\left(\frac{s - \sqrt{sc_\lambda \log(s)} - \sqrt{s/(c_\lambda \log(s))}}{s}\right)^{[\sqrt{s/(c_\lambda \log(s))}]}$$

$$\geq C\sqrt{s/\log(s)}.$$

Equations (5.23) and (5.24) imply

$$(5.26) \qquad P_s(\{k\}) \leq C_\lambda\left[\sqrt{\log(s)/s}\, P_s(\{k, k+1, \ldots\}) + s^{-\lambda}\right],$$

and (5.23) and (5.25) yield

$$(5.27) \qquad P_s(\{k\}) \leq C_\lambda\left[\sqrt{\log(s)/s}P_s(\{k, k-1, \ldots\}) + s^{-\lambda}\right]$$

for all $k \in \mathbb{Z}$.

(ii) *Some preparatory considerations.* We consider instead of $\hat{f}_h$ the artificial quantity

$$(5.28) \qquad\qquad \bar{f}_h(x) = \frac{1}{nh^d}\sum_{i=1}^{\nu}K\left(\frac{x - Z_i}{h}\right),$$

where $\{Z_1, \ldots, Z_\nu\}$ were defined by (2.3).

The crucial point is that $\bar{f}_h$ is based on a Poisson process instead of an empirical process. Therefore, $\bar{f}_h(x_1)$ and $\bar{f}_h(x_2)$ are independent, if the supports of the corresponding kernels are disjoint.

We decompose the $\mathbb{R}^d$ into nonoverlapping hypercubes of sidelength $2h$; that is,

$$I_k = \big[2(k_1 - 1)h, 2k_1 h\big) \times \cdots \times \big[2(k_d - 1)h, 2k_d h\big).$$

Further, we divide the set $\mathbb{Z}^d$ into 2^d subsets,

$$\mathscr{K}_l = \big\{k = (k_1, \ldots, k_d) \,|\, k_i = 2j_i + l_i, \, j_i \in \mathbb{Z}\big\},$$

where $l = (l_1, \ldots, l_d) \in \{0, 1\}^d$. We fix l and consider

$$Z_l = \sup_{k \in \mathscr{K}_l} \sup_{x \in I_k}\big\{\big|\bar{f}_h(x) - E\hat{f}_h(x)\big|\big\}.$$

It can be seen from the following considerations that

(5.29) $$P\big(Z_l < C_\lambda (nh^d)^{-1/2}(\log(n))^{1/2}\big) = O(n^{-\lambda})$$

holds for sufficiently small C_λ.

Let

$$\mu_k = \sup_{x \in I_k}\bigg\{\int h^{-d} K\bigg(\frac{x-z}{h}\bigg)f(z)\,dz\bigg\}.$$

Similarly to the considerations in the proof of Proposition 2.1, we can show that

(5.30) $$P\bigg(\sup_{k:\,\mu_k<\bar{\mu}} \sup_{x \in I_k}\big\{\big|\bar{f}_h(x) - E\hat{f}_h(x)\big|\big\} \geq C_\lambda(nh^d)^{-1/2}(\log(n))^{1/2}\bigg)$$
$$= O(n^{-\lambda}),$$

for some $\bar{\mu}$ sufficiently small. Hence, with a probability exceeding $1 - O(n^{-\lambda})$, the supremum Z_l will be attained on one of the intervals I_k with $\mu_k \geq \bar{\mu}$. Let

$$\{k_1, \ldots, k_{\rho_l}\} = \big\{k \in \mathscr{K}_l \,|\, \mu_k \geq \bar{\mu}\big\}.$$

(iii) *Decomposition of* $\bar{f}_h(x) - E\hat{f}_h(x)$. Let

$$\mathscr{J}_k = I_k \oplus \mathrm{supp}\bigg(K\bigg(\frac{x-\cdot}{h}\bigg)\bigg)$$
$$= \big[(2k_1 - 3)h, (2k_1 + 1)h\big) \times \cdots \times \big[(2k_d - 3)h, (2k_d + 1)h\big).$$

Further, let Z_i^k be the ith variable of $Z_1, \ldots, Z_\nu$ that falls into $\mathscr{J}_k$, and let $\hat{\nu}_k = \#\{1 \leq i \leq \nu \mid Z_i \in \mathscr{J}_k\}$ be the number of them. Then

$$\bar{f}_h(x) = \frac{1}{nh^d} \sum_{i=1}^{\hat{\nu}_k} K\left(\frac{x - Z_i^k}{h}\right).$$

Let $\nu_k = E\hat{\nu}_k = nP(Z_1 \in \mathscr{J}_k)$.

Now we have, for $x \in I_k$, that

$$\bar{f}_h(x) - E\hat{f}_h(x)$$

$$= \frac{1}{nh^d} \sum_{i=1}^{\hat{\nu}_k} \left[K\left(\frac{x - Z_i^k}{h}\right) - \frac{1}{P(Z_1 \in \mathscr{J}_k)} \int K\left(\frac{x - z}{h}\right) f(z)\, dz \right]$$

$$+ \frac{1}{nh^d} \frac{\hat{\nu}_k - \nu_k}{P(Z_1 \in \mathscr{J}_k)} \int K\left(\frac{x - z}{h}\right) f(z)\, dz$$

$$= \frac{1}{nh^d} \frac{\hat{\nu}_k - \nu_k}{P(Z_1 \in \mathscr{J}_k)} \frac{1}{|I_k|} \int_{I_k} \int K\left(\frac{x - z}{h}\right) f(z)\, dz\, dx$$

(5.31)
$$+ \frac{1}{nh^d} \sum_{i=1}^{[\nu_k]} \left[K\left(\frac{x - Z_i^k}{h}\right) - \frac{1}{P(Z_1 \in \mathscr{J}_k)} \int K\left(\frac{x - z}{h}\right) f(z)\, dz \right]$$

$$+ \frac{1}{nh^d} \frac{\hat{\nu}_k - \nu_k}{P(Z_1 \in \mathscr{J}_k)} \left\{ \int K\left(\frac{x - z}{h}\right) f(z)\, dz \right.$$

$$\left. - \frac{1}{|I_k|} \int_{I_k} \int K\left(\frac{x - z}{h}\right) f(z)\, dz \right\}$$

$$+ \frac{1}{nh^d} \left\{ \sum_{i=1}^{\hat{\nu}_k} [\cdots] - \sum_{i=1}^{[\nu_k]} [\cdots] \right\}$$

$$= T_{k1} + T_{k2}(x) + R_{k1}(x) + R_{k2}(x).$$

The main purpose of this decomposition was to split $\bar{f}_h(x) - E\hat{f}_h(x)$ into a term T_{k1} proportional to the Poisson variable $\hat{\nu}_k \sim P_{\nu_k}$, a term $\{T_{k2}(x)\}_{x \in I_k}$ independent of T_{k1}, and two asymptotically negligible terms, $R_{k1}(x)$ and $R_{k2}(x)$.

(iv) *Proof of the assertion.* Next we show that

(5.32) $$P\left(\sup_{k \in \mathscr{K}_l}\{T_{k1} + T_{k2}\} \in [c, d] \right) = O\left((d - c)(nh^d)^{1/2} \sqrt{\log(n)} + n^{-\lambda}\right),$$

where

$$T_{k2} = \sup_{x \in I_k}\{T_{k2}(x)\}.$$

We keep for a moment $\{T_{k2}\}_{k \in \mathscr{K}_l}$ fixed. Since the T_{k1}'s are independent of the T_{k2}'s, we obtain, by (5.26), that

$$P\Big(\sup_{k \in \mathscr{K}_l} \{T_{k1} + T_{k2}\} \in [c, d] \Big| T_{k_1, 2}, \ldots, T_{k_{\rho_l}, 2} \Big)$$

$$\leq P\big(T_{k_1, 1} \in [c - T_{k_1, 2}, d - T_{k_1, 2}]\big)$$

$$+ P\big(T_{k_2, 1} \in [c - T_{k_2, 2}, d - T_{k_2, 2}]; T_{k_1, 1} < c - T_{k_1, 2}\big) + \cdots$$

$$+ P\big(T_{k_{\rho_l}, 1} \in [c - T_{k_{\rho_l}, 2}, d - T_{k_{\rho_l}, 2}]; \ T_{k_1, 1} < c - T_{k_1, 2}, \ldots,$$

$$T_{k_{\rho_l - 1}, 1} < c - T_{k_{\rho_l - 1}, 2}\big)$$

$$\leq (d - c)(nh^d)^{1/2}\sqrt{\log(n)}\big\{ P\big(T_{k_1, 1} > c - T_{k_1, 2}\big) + \ldots$$

$$+ P\big(T_{k_{\rho_l}, 1} > c - T_{k_{\rho_l}, 2}; T_{k_1, 1} \leq c - T_{k_1, 2}, \ldots, T_{k_{\rho_l - 1}, 1} \leq c - T_{k_{\rho_l - 1}, 2}\big)\big\}$$

$$+ O(n^{-\lambda})$$

$$= O\Big((d - c)(nh^d)^{1/2}\sqrt{\log(n)}$$

$$\times P\Big(\sup_{k \in \mathscr{K}_l} \{T_{k1} + T_{k2}\} > c \Big| T_{k_1, 2}, \ldots, T_{k_{\rho_l}, 2} \Big) \Big)$$

$$+ O(n^{-\lambda})$$

$$= O\Big((d - c)(nh^d)^{1/2}\sqrt{\log(n)} + n^{-\lambda} \Big).$$

Integrating over all realizations for $T_{k_1, 2}, \ldots, T_{k_{\rho_l}, 2}$, we get (5.32).

Since f is Lipschitz, we easily obtain that

(5.33) $$\sup_{x \in I_k}\{|R_{k1}(x)|\} = \tilde{O}\Big((nh^d)^{-1/2}h\sqrt{\log(n)}, n^{-\lambda} \Big).$$

Because of $\hat{\nu}_k - [\nu_k] = \tilde{O}((nh^d)^{1/2}(\log(n))^{1/2}, n^{-\lambda})$, we can readily show that

(5.34) $$\sup_{x \in I_k}\{|R_{k2}(x)|\} = \tilde{O}\big((nh^d)^{-3/4}(\log(n))^{3/4} + (nh^d)^{-1}\log(n), n^{-\lambda}\big).$$

By (5.31) to (5.34) we obtain, with $\kappa_n = C[(nh^d)^{-1/2}h(\log(n))^{1/2} + (nh^d)^{-3/4}(\log(n))^{3/4}]$, that

$$P\big(Z_l \in [c, d]\big)$$

(5.35) $$\leq P\Big(\sup_{k \in \mathscr{K}_l} \{T_{k1} + T_{k2}\} \in [c - \kappa_n, d + \kappa_n]\Big) + O(n^{-\lambda})$$

$$= O\big((d - c)(nh^d)^{1/2}(\log(n))^{1/2} + h\log(n) + (nh^d)^{-1/4}(\log(n))^{5/4}\big).$$

By analogous considerations, where we only have to use (5.27) instead of (5.26), we obtain

$$
(5.36) \quad P\Big(\inf_{k \in \mathscr{K}_l} \inf_{x \in I_k} \big\{ \bar{f}_h(x) - E\hat{f}_h(x) \big\} \in [-d, -c] \Big)
$$
$$
= O\big((d - c)(nh^d)^{1/2}(\log(n))^{1/2} + h \log(n) + (nh^d)^{-1/4}(\log(n))^{5/4} \big).
$$

This implies

$$
P\Big(\sup_{x \in \mathbb{R}^d} \big\{ |\bar{f}_h(x) - E\hat{f}_h(x)| \big\} \in [c, d] \Big)
$$
$$
(5.37) \quad \leq \sum_{l \in \{0, 1\}^d} P\Big(\sup_{k \in \mathscr{K}_l} \sup_{x \in I_k} \big\{ |\bar{f}_h(x) - E\hat{f}_h(x)| \big\} \in [c, d] \Big) + O(n^{-\lambda})
$$
$$
= O\big((d - c)(nh^d)^{1/2}(\log(n))^{1/2} + h \log(n) + (nh^d)^{-1/4}(\log(n))^{5/4} \big).
$$

Using

$$
\sup_{x \in \mathbb{R}^d} \big\{ |\bar{f}_h(x) - \hat{f}_h(x)| \big\} = \tilde{O}\big([n^{-1/2} + (nh^d)^{-1}] \log(n), n^{-\lambda} \big),
$$

we obtain the assertion. □

Theorems 3.3, 3.4 and 3.5 are straightforward implications of Theorem 3.1 and Proposition 3.1. We omit these proofs.

REFERENCES

[1] AMOSOVA, N. N. (1972). On limit theorems for probabilities of moderate deviations. *Vestnik Leningrad. Univ.* **13** 5–14 (in Russian).

[2] ANGO NZE, P. (1992). Critères d'ergodicité de quelques modèles à représentation markovienne. *C. R. Acad. Sci. Paris Sèr. I Math.* **315** 1301–1304.

[3] ANGO NZE, P. and DOUKHAN, P. (1993). Estimation fonctionnelle de séries temporelles mélangeantes. *C. R. Acad. Sci. Paris Sèr. I Math.* **317** 405–408.

[4] ANGO NZE, P. and RIOS, R. (1995). Estimation L^∞ de la fonction de densité d'un processus faiblement dépendant: les cas absolument régulier et fortement mélangeant. *C. R. Acad. Sci. Paris Sèr. I Math.* **320** 1259–1262.

[5] BICKEL, P. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

[6] BRÉMAUD, P. and MASSOULIÉ, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Probab.* **24** 1563–1588.

[7] BÜHLMANN, P. (1994). Blockwise bootstrapped empirical process for stationary sequences. *Ann. Statist.* **22** 995–1012.

[8] DHOMPONGSA, S. (1984). A note on the almost sure approximation of the empirical process of weakly dependent random vectors. *Yokohama Math. J.* **32** 113–121.

[9] DOUKHAN, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statist.* **85**. Springer, New York.

[10] DOUKHAN, P., MASSART, P. and RIO, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.* **31** 393–427.

[11] DOUKHAN, P. and PORTAL, F. (1987). Principe d'invariance faible pour la fonction de répartition empirique dans un cadre multidimensionnel et mélangeant. *Probab. Math. Statist.* **8** 117–132.

[12] EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** 1–26.

[13] EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia.

[14] FALK, M. and REISS, R.-D. (1989a). Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates. *Ann. Probab.* **17** 362–371.

[15] FALK, M. and REISS, R.-D. (1989b). Bootstrapping the distance between smooth bootstrap and sample quantile distribution. *Probab. Theory Related Fields* **82** 177–186.

[16] FARAWAY, J. J. and JHUN, M. (1990). Bootstrap choice of bandwidth for density estimation. *J. Amer. Statist. Assoc.* **85** 1119–1122.

[17] FRANKE, J., KREISS, J.-P., MAMMEN, E. and NEUMANN, M. H. (1998). Properties of the nonparametric autoregressive bootstrap. Discussion Paper 54/98, SFB 373, Humboldt Univ., Berlin.

[18] FRIEDMAN, J., STUETZLE, W. and SCHROEDER, A. (1984). Projection pursuit density estimation. *J. Amer. Statist. Assoc.* **79** 599–608.

[19] GYÖRFI, L., HÄRDLE, W., SARDA, P. and VIEU, P. (1989). *Nonparametric Curve Estimation from Time Series*. Springer, Berlin.

[20] HALL, P. (1991). On convergence rates of suprema. *Probab. Theory Related Fields* **89** 447–455.

[21] HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

[22] HALL, P. (1993). On Edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation. *J. R. Statist. Soc. Ser. B* **55** 291–304.

[23] HALL, P., DICICCIO, T. J. and ROMANO, J. P. (1989). On smoothing and the bootstrap. *Ann. Statist.* **17** 692–704.

[24] HALL, P. and HART, J. D. (1990). Convergence rates in density estimation for data from infinite-order moving average processes. *Probab. Theory Related Fields* **87** 253–274.

[25] HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947.

[26] HART, J. D. (1996). Some automated methods of smoothing time-dependent data. *J. Nonparametr. Statist.* **6** 115–142.

[27] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent rv's and the sample distribution function. *Z. Wahrsch. Verw. Gebiete* **32** 111–131.

[28] KONAKOV, V., LÄUTER, H. and LIERO, H. (1995). Nonparametric versus parametric goodness of fit. *Statistics*. To appear.

[29] KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241.

[30] MASRY, E. (1994). Probability density estimation from dependent observations using wavelet orthonormal bases. *Statist. Probab. Lett.* **21** 181–194.

[31] MOKKADEM, A. (1988). Mixing properties of ARMA processes. *Stochastic Process. Appl.* **29** 309–315.

[32] MOKKADEM, A. (1990). Propriétés de mélange des processus autorégessifs polynomiaux. *Ann. Inst. H. Poincaré Probab. Statist.* **26** 219–260.

[33] NEUMANN, M. H. (1995). Automatic bandwidth choice and confidence intervals in nonparametric regression. *Ann. Statist.* **23** 1937–1959.

[34] NEUMANN, M. H. and KREISS, J.-P. (1998). Regression-type inference in nonparametric autoregression. *Ann. Statist.* **26** 1570–1613.

[35] NEUMANN, M. H. and PAPARODITIS, E. (1998). A nonparametric test for the stationary density. Discussion Paper 58/98, SFB 373, Humboldt Univ., Berlin.

[36] PHAM, T. D. (1986). The mixing property of bilinear and generalised random coefficient autoregressive models. *Stochastic Process. Appl.* **23** 291–300.

2048 NEUMANN

M. H. NEUMANN

2048   M. H. NEUMANN

[37] PHAM, T. D. and TRAN, L. T. (1985). Some mixing properties of time series models. *Stochastic Process. Appl.* **19** 297–303.

[38] REISS, R.-D. (1993). *A Course on Point Processes.* Springer, New York.

[39] ROBINSON, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207.

[40] SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[41] SILVERMAN, B. W. and YOUNG, G. A. (1987). The bootstrap: to smooth or not to smooth? *Biometrika* **74** 469–479.

[42] SPOKOINY, V. G. (1996). Adaptive and spatially adaptive testing of a nonparametric hypothesis. *Math. Methods Statist.* To appear.

[43] TAKAHATA, H. and YOSHIHARA, K. (1987). Central limit theorems for integrated square error of nonparametric density estimators based on absolutely regular random sequences. *Yokohama Math. J.* **35** 95–111.

[44] TRAN, L. T. (1990). Density estimation for time series by histograms. *J. Statist. Plann. Inference* **40** 61–79.

[45] YU, B. (1993). Density estimation in the L^∞ norm for dependent data with applications to the Gibbs sampler. *Ann. Statist.* **21** 711–735.

HUMBOLDT UNIVERSITÄT
SONDERFORSCHUNGSBEREICH 373
SPANDAUER STRAßE 1
D-10178 BERLIN
GERMANY
E-MAIL: neumann@wiwi.hu-berlin.de