The Annals of Statistics
2001, Vol. 29, No. 2, 372–416

NONPARAMETRIC ESTIMATION IN NULL RECURRENT TIME SERIES

BY HANS ARNFINN KARLSEN AND DAG TJØSTHEIM

University of Bergen

We develop a nonparametric estimation theory in a nonstationary environment, more precisely in the framework of null recurrent Markov chains. An essential tool is the split chain, which makes it possible to decompose the times series under consideration into independent and identical parts. A tail condition on the distribution of the recurrence time is introduced. This condition makes it possible to prove weak convergence results for sums of functions of the process depending on a smoothing parameter. These limit results are subsequently used to obtain consistency and asymptotic normality for local density estimators and for estimators of the conditional mean and the conditional variance. In contradistinction to the parametric case, the convergence rate is slower than in the stationary case, and it is directly linked to the tail behavior of the recurrence time. Applications to econometric, and in particular to cointegration models, are indicated.

1. Introduction. Work on nonparametric estimation has so far with very few exceptions been carried out in a stationary strongly mixing framework [see, e.g., Robinson (1983), Masry and Tjøstheim (1995), and references therein]. Recently asymptotics for processes with long-range dependence have been covered [Robinson (1997)], but still no systematic theory exists for a nonstationary situation.

The main purpose of this paper is to try to fill this gap by establishing a nonparametric estimation theory that can be used in a nonstationary environment. Clearly the collection of all nonstationary processes is much too wide, but in our opinion an appropriate framework for working with such problems is the class of null recurrent Markov chains, or possibly regime models including null recurrent states. It is true that this requires the model to be stated as a Markov chain, but this is a mild restriction. The random walk model and many of the related unit-root processes belong to this class [Myklebust, Karlsen and Tjøstheim (2001)], and, more important, nonlinear processes are not excluded.

With the single exception of the work by Yakowitz (1993) on consistency of nearest neighbor estimates, as far as we know, the estimation theory of null recurrent processes has been confined to the parametric case.

Independent of our work, however, nonparametric estimation has been considered in a random walk situation by Phillips and Park (1998), who use local

Received May 1998; revised January 2001.

AMS 2000 *subject classifications*. Primary 62M10, 62G07; secondary 60J05.

Key words and phrases. Nonstationary time series models, null recurrent Markov chain, nonparametric kernel estimators, split chain.

time argument to derive asymptotic distributions and by Xia (1998), who in his doctoral thesis gives a proof of consistency in a transfer function case.

Asymptotics of parametric null recurrent (usually nontime series) models have been treated by Höpfner (1990, 1994), Höpfner, Jacod and Ladelli (1990), Kasahara (1982, 1984, 1985), Touati (1990), and we will exploit some of their techniques. For two early contributions in this field we refer to Darling and Kac (1957) and Kallianpur and Robbins (1954). However, there are important differences between the parametric and nonparametric situations. A parametric estimate is strongly influenced by the large values of the process, and for unit-root processes superefficiency is obtained with a faster rate of convergence than in the stationary case. In contradistinction, a nonparametric estimator depends heavily on observations which are confined to a neighborhood of a given point, and the rate of convergence turns out, not unexpectedly, to be slower than in the stationary case. This means that series with large or very large sample sizes are required.

Long series are becoming increasingly available, for example, in finance and econometrics. There is therefore also a practical motivation behind our work. The particulars of this motivation are much the same as for the stationary case: it is desirable to have greater flexibility in the initial stage of modelling than that offered by a fixed parametric or semiparametric model, for example using nonparametric estimates as a guide in choosing a parametric (linear or nonlinear) model. Since the present paper is directed towards establishing a theory, specific practical aspects are not much discussed, and we refer to Myklebust, Karlsen and Tjøstheim (2001) for some more examples. We would like to mention very briefly potential implications for econometric time series modelling, though, since such series are often thought to be nonstationary. The kind of nonstationarity that has been built into the *parametric* econometric modelling has overwhelmingly been of linear unit-root type, leading to ARIMA models and, in the multivariate case, to linear cointegration models. For such models a very considerable body of literature exists [cf. the review papers by Stock (1994), Watson (1994) and the book by Johansen (1995)]. Asymptotic distributions are typically nonnormal and the parameter estimates are superefficient [Dickey and Fuller (1979), Johansen (1995)]. The need for models combining features of nonlinearity and nonstationarity has been emphasized [see, e.g., Granger and Hallman (1991), Granger (1995), Aparicio and Escribano (1997)], but no systematic estimation theory exists. Again, we believe that the class of null recurrent processes constitutes an adequate framework for posing such problems. The technique used in this paper is general, and although we focus on nonparametric estimation, it is in principle possible to develop an analogous theory covering nonlinear and nonstationary parametric time series models. Finally, it should be mentioned that there are challenging and interesting connections to attempts having been made to construct a nonlinear cointegration theory. We look at some of these in Karlsen, Myklebust and Tjøstheim (2000).

There are a number of open problems and possibilities for further research. These are related to exploratory problems such as those examined by Tjøstheim and Auestad (1994), Masry and Tjøstheim (1997) and Hjellvik, Yao and Tjøstheim (1998), but there are also many hard problems connected with the basic estimation theory itself.

Since our paper draws quite heavily on Markov theory for recurrent chains, in the beginning of Section 3 we briefly state some main facts stemming from that theory. Much of the material is based on the book by Nummelin (1984), but since to our knowledge, it has not been utilized before in the context of nonparametric estimation, it has been included to make the paper more self-contained. In fact, we consider the merger of the recurrence theory of Markov chains, in particular use of the split chain, and the asymptotic theory of sums depending on a smoothing parameter to be a main contribution of the paper. This synthesis is achieved in Section 4. Applications to nonparametric estimation of an invariant density and conditional mean–variance functions are given in Section 5, where we derive consistency and asymptotic normality of these estimates in a null recurrent situation. A simple example is given in Section 6.

2. Motivation and intuition. Some of the proofs of this paper are quite technical and draw on results from disciplines that are not usually brought together: recurrence theory of Markov processes, functional limit theorems and nonparametric estimation in a time series context. Before we start on the technical derivations we will therefore try to provide a brief nontechnical and intuitive overview of the paper.

For a time series $\{X_t, t \geq 0\}$, the traditional kernel estimator of the conditional mean $M(x) = \mathbf{E}(X_t \mid X_{t-1} = x)$ is given by

$$\widehat{M}(x) = \frac{\sum_t X_{t+1} K\big(\frac{X_t - x}{h}\big)}{\sum_t K\big(\frac{X_t - x}{h}\big)}, \tag{2.1}$$

where K is a kernel function whose argument depends on the bandwidth h. In the stationary case, mixing results are often used to provide asymptotics for $\widehat{M}(x)$. Such arguments are not available in the situation we consider. A nonparametric estimate is a local estimate. A necessary condition for an asymptotic theory is therefore that $\{X_t\}$ returns to any neighborhood around x infinitely often, that is, the process $\{X_t\}$ should be recurrent in this sense. A suitable framework then seems to be the class of recurrent Markov chains. This class contains both stationary and nonstationary processes.

Fortunately, there is an extensive theory for recurrent Markov chains, and what we will use in particular is the device of the split chain. It is well known that for an irreducible aperiodic finite state Markov chain, the process can be decomposed into independent identically distributed parts by taking as a regeneration point any of its states. The processes we consider have continuous state space, and we cannot use $x = x_0$, say, as a regeneration point since for a continuous variable, $P(X_t = x_0) = 0$. However, it is possible to get around this point by replacing x_0 by an appropriate set containing x_0 and a randomization mechanism. This is the splitting technique of a continuous state Markov chain, and it will be briefly reviewed in Section 3. The crux of the methodology is that a sum such as $\sum_t K(\frac{X_t - x}{h})$ appearing in the denominator of (2.1) can essentially be decomposed into a sum of independent components

$$\sum_{k=1}^{T(n)} \sum_{t=\tau_{k-1}+1}^{\tau_k} K\left(\frac{X_t - x}{h}\right), \tag{2.2}$$

where $T(n)$ is the number of regenerations at time n, and where the τ_k are the regeneration time points. The sums $\sum_{t=\tau_{k-1}+1}^{\tau_k} K(\frac{X_t - x}{h})$, $k = 1, \dots$ are independent random variables, each consisting of a random number of addends. In the numerator of (2.1) the corresponding terms are one dependent due to the presence of both X_t and X_{t+1}.

A decomposition result makes it possible to use a functional central limit theorem to derive the asymptotic distribution of sums of type (2.2) when appropriately scaled. The derivation is relatively difficult because of the dependence between $T(n)$ and the sums $\sum_{t=\tau_{k-1}+1}^{\tau_k} K(\frac{X_t - x}{h})$ and the presence of the smoothing parameter. A functional CLT is studied in Section 4 under high level regularity conditions. A prerequisite for the developments of Section 4 is the existence of an asymptotic theory for $T(n)$. This is provided in Section 3 in conjunction with some other ingredients needed from the Markov chain recurrence theory. For example, an important role is played by a condition on the tail of the interrecurrence times $\tau_k - \tau_{k-1}$, $k = 1, \dots$.

What remains is to adapt the general nature CLT of Section 4 to the nonparametric estimation situation. Among other things the high level conditions of Section 4 have to be translated into low level conditions on the kernel function and related quantities. It should be noted that even though recurrence does not imply stationarity it does imply the existence of a (nonunique) invariant measure. Suitably normalized on a so-called small set (cf. beginning of Section 5) this results in a probability density that can be estimated using a kernel estimator.

As stated above we think that recurrence is a natural tool for attacking the asymptotic theory of nonparametric estimates in this situation, and we believe that the theory can be extended to, say, local polynomial estimation. The splitting technique can also be generalized to a transfer function situation

$$Z_t = f(X_t) + W_t,$$

where the aim is to estimate the transfer function f for a stationary noise process $\{W_t\}$ and a recurrent input process $\{X_t\}$. We refer to Karlsen, Myklebust and Tjøstheim (2000). This gives a nonlinear cointegration type relationship between Z_t and X_t.

3. Markov theory.

3.1. *Notation and the split chain.* We adopt the notation used by Nummelin (1984). We denote by $\{X_t, t \geq 0\}$ a ϕ-irreducible Markov chain on a general state space $(\mathbf{E}, \mathscr{E})$ with transition probability P. The sigma algebra of measurable sets $\mathscr{E}$ is countably generated and we assume that ϕ is maximal in the sense that if ϕ' is another irreducible measure then ϕ' is absolutely continuous with respect to ϕ. We denote the class of nonnegative measurable functions with ϕ-positive support by $\mathscr{E}^+$. For a set $A \in \mathscr{E}$ we write $A \in \mathscr{E}^+$ if the indicator function $1_A \in \mathscr{E}^+$. The chain is Harris recurrent if for all $A \in \mathscr{E}^+$,

$$\mathbf{P}(S_A < \infty \mid X_0 = x) \equiv 1 \quad \text{where } S_A = \min\{n \geq 1\colon X_n \in A\}. \tag{3.1}$$

In the following $\{X_t,\ t \geq 0\}$ will always be assumed to be ϕ-irreducible Harris recurrent. The chain is positive recurrent if there exists an initial probability measure such that $\{X_t, t \geq 0\}$ is strictly stationary, and the process is null recurrent otherwise.

If η is a nonnegative measurable function and λ is a measure, then the kernel $\eta \otimes \lambda$ is defined by

$$\eta \otimes \lambda(x, A) = \eta(x)\lambda(A), \qquad (x, A) \in (E, \mathscr{E}).$$

If K is a general kernel, the function $K\eta$, the measure λK and the number $\lambda\eta$ are defined by

$$K\eta(x) = \int K(x, dy)\eta(y), \quad \lambda K(A) = \int \lambda(dx)K(x, A), \quad \lambda\eta = \int \lambda(dx)\eta(x).$$

Sometimes we write $\lambda(\eta)$ instead of $\lambda\eta$. The convolution of two kernels K_1 and K_2 gives another kernel defined by

$$K_1K_2(x, A) = \int K_1(x, dy)K_2(y, A).$$

Due to associative laws, the number $\lambda K_1 K_2 \eta$ is uniquely defined. If $A \in \mathscr{E}$ and 1_A is the corresponding indicator variable, then $K1_A(x) = K(x, A)$. The kernel I_η is defined by $I_\eta(x, A) = \eta(x)1_A(x)$ [and $I_\eta(x, dy) = \eta(x)\delta_x(dy)$ where δ_x is the Dirac delta measure at the point x]. We abbreviate the identity function 1_E by 1. We let $\mathscr{G}_r^d = \{f\colon (E^r, \mathscr{E}^r) \mapsto (R^d, \mathscr{B}(R^d)\}$ where $\mathscr{B}(R^d)$ is the class of Borel sets on R^d. If $r = 1$ or $d = 1$, we drop the subscript or superscript.

We define $\eta \in \mathscr{E}^+$ to be small if there exists a measure λ, a positive constant b and an integer $m \geq 1$ so that

$$P^m \geq b\eta \otimes \lambda. \tag{3.2}$$

A set A is said to be small if 1_A is small. Under quite wide conditions [cf. Feigin and Tweedie (1985)] a compact set will be small. In this case it follows from (3.2) that a ϕ-positive subset of a compact set would be small. If λ satisfies (3.2) for some η, b and m, then λ is a small measure.

A fundamental fact for ϕ-irreducible Markov chains is the existence of a minorization inequality [Nummelin (1984), Theorem 2.1 and Proposition 2.6, pages 16–19]: there exists a small function s, a probability measure ν and an integer $m_0 \geq 1$ so that

$$P^{m_0} \geq s \otimes \nu. \tag{3.3}$$

It creates some technical difficulties to have $m_0 > 1$ because it necessitates the m_0-step chain, and it is not a severe restriction to assume $m_0 = 1$. Therefore, unless otherwise stated, in the sequel we assume that the minorization inequality

$$P \geq s \otimes \nu \tag{3.4}$$

holds, where s and ν are small and $\nu(E) = 1$. In particular, this implies that $0 \leq s(x) \leq 1$, $x \in E$. If (3.4) holds, then the pair (s, ν) is called an atom (for P).

We illustrate what the minorization inequality means in the case of a nonlinear autoregressive process.

EXAMPLE 3.1. Assume that

$$X_t = \begin{cases} X_0, & \text{when } t = 0, \\ f(X_{t-1}) + e_t, & \text{when } t \geq 1, \end{cases}$$

where $\{e_t, t \geq 0\}$ are iid random variables with zero mean and with density ζ with respect to the Lebesgue measure on $E = R$. Assume that the function f is bounded on compact sets and $\inf_{x \in C} \zeta(x)$ is strictly positive for all compact sets C. The transition probability is given by

$$P(x, dy) = p^{(1)}(y \mid x) dy \stackrel{\text{def}}{=} \zeta(y - f(x))\, dy$$

and the n step transition function is $P^n(x, dy) = p^{(n)}(y \mid x)\, dy$ where

$$p^{(n)}(y \mid x) = \int p^{(n-1)}(y \mid u) \zeta(u - f(x))\, du, \qquad n \geq 2.$$

Let C be a compact set with positive Lebesgue measure. Define $\rho_0(y) = \inf_{x \in C} \zeta(y - f(x))$, $a = \int \rho_0(y)\, dy$, $\rho = a^{-1}\rho_0$, $s = a 1_C$. Then

$$\begin{aligned} P(x, dy) &\geq 1_C(x) \rho_0(y)\, dy \\ &= s(x) \nu(dy), \end{aligned}$$

where $\nu(dy) = \rho(y)\, dy$ and $\nu(E) = 1$. Thus (3.4) is satisfied.

Note that a pth order model with $X_t = f(X_{t-1}, \ldots, X_{t-p}) + e_t$ would require the use of (3.3) with $m_0 = p$. With $p = 1$ and $f(x) \equiv x$ the random walk is obtained as a special case (cf. Section 6).

In our approach to the nonparametric estimation theory a vital role will be played by the split chain, which can be constructed once the minorization condition is fulfilled. It permits a decomposition of the chain into separate and identical parts which are building blocks in the subsequent analysis.

To understand the significance of (3.4), consider the identity

$$P(x, A) = (1 - s(x))\{(1 - s(x))^{-1}(P(x, A) - s(x)\nu(A))1(s(x) < 1) + 1_A(x)1(s(x) = 1)\} + s(x)\nu(A)$$
$$\stackrel{\text{def}}{=} (1 - s(x))Q(x, A) + s(x)\nu(A).$$

If (3.4) holds, Q is a transition probability, and because $0 \leq s(x) \leq 1$ and $\nu(E) = 1$, the transition probability P can be thought of as a mixture of the transition probability Q and the small measure ν. Since ν is independent of x, this means that the chain regenerates each time ν is chosen. This occurs with probability $s(x)$.

This reasoning can be formalized by introducing the split chain $\{X_t, Y_t, t \geq 0\}$, where the auxiliary chain $\{Y_t\}$ can only take the values 0 and 1. Given that $X_t = x, Y_{t-1} = y_{t-1}$, Y_t takes the value 1 with probability $s(x)$, so that $\alpha = E \times \{1\}$ is a proper atom of the split chain. More precisely, let λ denote an arbitrary initial distribution on E , let $\mathscr{F}_t^X$ and $\mathscr{F}_t^Y$ denote the σ-algebras generated by $\{X_j, j \leq t\}$ and $\{Y_j, j \leq t\}$, $\mathscr{F}_{-1}^Y$ being the trivial σ-algebra; then the split chain $\{X_t, Y_t, t \geq 0\}$ is defined by

$$\begin{aligned} &\mathbf{P}(X_0 \in A) = \lambda(A), \\ &\mathbf{P}(Y_t = y \mid \mathscr{F}_t^X \vee \mathscr{F}_{t-1}^Y) = s(X_t)y + (1 - s(X_t))(1 - y), \qquad t \geq 0, \\ &\mathbf{P}(X_t \in A \mid \mathscr{F}_{t-1}^X \vee \mathscr{F}_{t-1}^Y) = \nu(A)y + Q(x, A)(1 - y), \qquad t \geq 1. \end{aligned} \tag{3.5}$$

For the properties of the split chain we refer to Nummelin [(1984), Chapter 4].

We observe that the distribution of $\{(X_t, Y_t), t \geq 0\}$ is determined by λ, P and (s, ν). We use $\mathbf{P}_\lambda$ as generic symbol for the distribution of the Markov chain with initial distribution λ, and the corresponding expectation is denoted by $\mathbf{E}_\lambda$. If $\lambda = \delta_x$ we write $\mathbf{P}_x$, which is the conditional distribution of $(Y_0, \{(X_t, Y_t), t \geq 1\})$ given that $X_0 = x$. If the initial distribution is equal to $\delta_\alpha(x, y)$, that is, $Y_0 = 1,\ X_0 = x$ arbitrary, then we write $\mathbf{P}_\alpha$ and $\mathbf{E}_\alpha$.

3.2. *The invariant measure.* In a general null recurrent chain $\{X_t\}$ no marginal distribution function exists that can be estimated nonparametrically. There is a generalization of the distribution function in the invariant measure, however, and in Section 4 we will show that if there is an associated density function, then it can be estimated.

Let $\tau = \tau_\alpha = \min\{n \geq 0\colon Y_n = 1\}$ and $S_\alpha = \min\{n \geq 1\colon Y_n = 1\}$. Since $\{S_\alpha = n\} = \{Y_j = 0,\ 1 \leq j < n,\ Y_n = 1\}$ and $\{\tau = n\} = \{S_\alpha = n,\ Y_0 = 0\}$, it follows [cf. Nummelin (1984), page 63] that

$$\begin{aligned} \mathbf{P}_x(\tau = n) &= (P - s \otimes \nu)^n s(x), \qquad n \geq 0, \\ \mathbf{P}_\alpha(S_\alpha = n) &= \nu(P - s \otimes \nu)^{n-1} s, \qquad n \geq 1. \end{aligned} \tag{3.6}$$

Define π_s by

$$\pi_s(A) = \pi_s 1_A = \mathbf{E}_\alpha\left[\sum_{n=1}^{S_\alpha} 1_A(X_n)\right], \qquad A \in \mathscr{E}. \tag{3.7}$$

Then by (3.6) and (3.5),

$$\pi_s(A) = \sum_{n=1}^{\infty} \mathbf{E}_\alpha[1_A(X_n)1(S_\alpha \geq n)] = \sum_{n=1}^{\infty} \nu(P - s\otimes\nu)^{n-1}1_A = \nu G_{s,\nu}1_A,$$

where

$$G_{s,\nu} \stackrel{\text{def}}{=} \sum_{n=0}^{\infty}(P - s\otimes\nu)^n. \tag{3.8}$$

This means that $\pi_s = \nu G_{s,\nu}$ and by (3.6) $\pi_s(s) = \mathbf{P}_\alpha(S_\alpha < \infty)$. Since the split chain is Harris recurrent, it follows that $\pi_s(s) = 1$. From (3.8) it is not difficult to prove that $\pi_s = \pi_s P$. Thus π_s is an invariant measure. The results stated below can be found in Nummelin (1984).

REMARK 3.1. If π is another invariant measure, then $\pi = \pi(s)\pi_s$ (page 73). The invariant measure π_s is equivalent to ϕ, $\pi_s(C) < \infty$ for all small sets C and it is σ-finite (Proposition 5.6, page 72).

The chain is positive recurrent if and only if $\pi_s 1_E < \infty$ (page 68). In the positive recurrent case $\pi \stackrel{\text{def}}{=} \pi_s/\pi_s 1_E$ is the unique stationary probability measure for $\{X_t\}$. In the latter situation, when the initial distribution of X_0 is given by π, $\{X_t\}$ will evolve as a strictly stationary process having π as its marginal distribution. It is seen from (3.7) that $\{X_t\}$ is positive recurrent if and only if $\mathbf{E}_\alpha S_\alpha < \infty$.

3.3. *Notation for functions in several variables.* Since the kernel estimator typically involves more than one variable, it is necessary to extend the notation of Sections 3.1 and 3.2 to functions of several variables. All integrals will be assumed to be well defined.

Recall that for $g \in \mathscr{G}_1$, $\pi_s(g) = \int \pi_s(dx)g(x)$ and [cf. (3.7) and (3.8)]

$$G_{s,\nu}g(x) = \int G_{s,\nu}(x, dy)g(y) = \mathbf{E}_x\left[\sum_{n=0}^{\tau} g(X_n)\right].$$

We introduce a useful transformation from $\mathscr{G}_r$ to $\mathscr{G}_1$.

DEFINITION 3.1. Let $r \geq 1$ and let $g \in \mathscr{G}_r$. For $r = 1$ and $r = 2$ we define $\widetilde{I}_g(x, dy)(1) = P(x, dy)g(x)$ and $\widetilde{I}_g(x, dy)(2) = P(x, dy)g(x, y)$, respectively. For $r > 2$ let

$$\widetilde{I}_g(x, dy)(r) = \int P(x, dx_2)\cdots P(x_{r-1}, dy)g(x, x_2, \ldots, x_{r-1}, y),$$

where the integration is with respect to $x_2, \dots, x_{r-1}$ and whenever the right-hand side is well defined. Furthermore, define

$$\tilde{g} = \widetilde{I}_g 1.$$

Since $\mathscr{G}_{r-1} \subset \mathscr{G}_r$ for $g \in \mathscr{G}_{r-1}$, when $r \geq 2$ we can write $\widetilde{I}_g(x, dy)(r) = \widetilde{I}_g(x, dy)(r-1)P$. An interpretation of $\tilde{g}$ is given by

$$\mathbf{E}_x[g(X_0, X_1, \dots, X_{r-1})] = \tilde{g}(x),$$
$$\mathbf{E}_x\left[\sum_{j=0}^{\tau} g(X_j, X_{j+1}, \dots, X_{j+r-1})\right] = G_{s,\nu}\tilde{g}(x), \tag{3.9}$$

which is easily verified [cf. (3.7) and (3.8)]. The right-hand sides of (3.9) can be seen as convenient and compact ways of writing the conditional expectations on the left-hand side. In the following we omit r in $\widetilde{I}_g(x, dy)(r)$.

If $g \in \mathscr{G} = \mathscr{G}_1$, then $\widetilde{I}_g = I_g P$ and $\tilde{g} = I_g P1 = g$. In order to reduce the notation further we extend π_s to $\cup_{r=1}^{\infty}\mathscr{G}_r$ by

$$\pi_s g \stackrel{\text{def}}{=} \pi_s \tilde{g} = \int \pi_s(dx_1)P(x_1, dx_2)\cdots P(x_{r-1}, dx_r)g(x_1, \dots, x_r), \qquad g \in \mathscr{G}_r.$$

We also extend the L^p spaces generated by π_s:

$$L_r^p(\pi_s) \stackrel{\text{def}}{=} \{g \in \mathscr{G}_r \colon \|g\|_{p,\pi_s}^p \stackrel{\text{def}}{=} \pi_s \widetilde{I}_{|g|^p} 1 < \infty\}, \qquad p \in (0, \infty),\ r \geq 1.$$

All of the notation in this subsection is trivially extended to $\mathscr{G}_r^d$. The notation is somewhat unfamiliar in a time series context, but it has the advantage of leading to compact derivations and expressions.

3.4. *β-null recurrence and tail behavior of recurrence times.* We have not been able to carry through the asymptotic theory of nonparametric estimation for a general null recurrent chain. We need a regularity condition for the tail behavior of the distribution of the recurrence time S_α. Since this condition is crucial for most of what we are doing, we introduce it in a rather general way and then specialize to the case when (3.4) holds.

A positive function L defined on $[a, \infty)$, where $a \geq 0$, is slowly varying at infinity [Bingham, Goldie and Teugels (1989), page 6] if

$$\lim_{x\uparrow\infty} \frac{L(\kappa x)}{L(x)} = 1 \quad \text{for all } \kappa > 0 \quad \text{and for all } x \in [a, \infty). \tag{3.10}$$

DEFINITION 3.2. The Markov chain $\{X_t\}$ is β-null recurrent if there exists a small nonnegative function h, an initial measure λ, a constant $\beta \in (0, 1)$ and a slowly varying function L_h so that

$$\mathbf{E}_\lambda\left[\sum_{t=0}^{n} h(X_t)\right] \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_h(n) \quad \text{as } n \to \infty. \tag{3.11}$$

REMARK 3.2. If L and L' are two slowly varying functions at infinity, then they are said to be equivalent if $\lim_{x\uparrow\infty} L(x)/L'(x) = 1$. In all of our applications of slowly varying functions they are only unique up to equivalence. Hence, when (3.11) is true, without any loss of generality we assume that L_h is normalized [Bingham, Goldie and Teugels (1989), pages 15, 24]; that is, the function $x^\beta L_h(x)$ is strictly increasing and continuous in the interval $[x_0, \infty)$ for some x_0.

Let $G^{(n)} = \sum_{t=0}^{n} P^t$. The left-hand side of (3.11) can be written as $\lambda G^{(n)}h$. We first prove that for a fixed parameter β (3.11) is actually a global property shared by all nonnegative small functions.

LEMMA 3.1. *Assume that $\{X_t\}$ is β-null recurrent and aperiodic. Let (s,ν) be a fixed atom. Then we can find an L_s so that for all small functions f the asymptotic relation* (3.11) *holds with $L_f = \pi_s(f)L_s$ where π_s is defined by* (3.7).

PROOF. Let λ and h be given by Definition 3.2 and the atom (s,ν) be fixed. Let $L_s \stackrel{\text{def}}{=} L_h/\pi_s h$. Using a null recurrent ratio limit theorem [Nummelin (1984), Corollary 7.2(i), page 131] (where it should be noted that a small function is a special function) and (3.11),

$$\frac{\lambda G^{(n)}h}{\nu G^{(n)}s} = \pi_s(h)(1+o(1)). \tag{3.12}$$

Using (3.11) again and the above expression, it follows that

$$\nu G^{(n)}s \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_s(n). \tag{3.13}$$

Let f be a given small function. Then by (3.12) with f instead of h and by (3.13) it follows that

$$\lambda G^{(n)}f \sim \frac{1}{\Gamma(1+\beta)} n^\beta \pi_s(f) L_s(n). \qquad \square$$

REMARK 3.3. If the atom (s,ν) in (3.4) is changed to (s',ν') then (cf. Remark 3.1)

$$\pi_{s'} = \frac{\pi_s}{\pi_s(s')}, \qquad L_{s'} \sim \pi_s(s')L_s.$$

The asymptotic expression (3.13) is connected to the Tauberian theorem [Feller (1971), page 447].

TAUBERIAN THEOREM. *Let $\{d_n, n \geq 0\}$ be any nonnegative sequence and let $d(r) = \sum_{n=0}^{\infty} r^n d_n$ be finite when $|r|$ is less than* 1. *Moreover, let L_1 be slowly varying and $\rho \in [0,\infty)$. Then*

$$\sum_{k=0}^{n} d_k \sim \frac{1}{\Gamma(1+\rho)} n^\rho L_1(n) \quad \Longleftrightarrow \quad d(r) \sim (1-r)^{-\rho} L_1\left(\frac{1}{1-r}\right), \tag{3.14}$$

when $n \to \infty$ *and* $r \uparrow 1^-$, *respectively. If* $\{d_n\}$ *is monotone and* $\rho > 0$, *then each of the conditions given by* (3.14) *is equivalent with*

$$d_n \sim \frac{n^{\rho-1}}{\Gamma(\rho)} L_1(n). \tag{3.15}$$

If (3.4) is true, the Tauberian theorem can be used to show that then the concept of β-null recurrence implies a regularity condition for the tail behavior of the distribution of the recurrence time S_α.

THEOREM 3.1. *Assume* (3.4) *is true. Then* $\{X_t\}$ *is* β-*null recurrent if and only if*

$$\mathbf{P}_\alpha(S_\alpha > n) = \frac{1}{\Gamma(1-\beta) n^\beta L_s(n)} (1 + o(1)). \tag{3.16}$$

REMARK 3.4. *If* (3.16) *is true, then it is not difficult to show that*

$$\sup\{p \geq 0 \colon \mathbf{E}_\alpha S_\alpha^p < \infty\} = \beta.$$

Thus, even though $\mathbf{E}_\alpha S_\alpha = \infty$ *for a null recurrent process, if* (3.4) *and* (3.16) *hold, then* $\mathbf{E}_\alpha S_\alpha^p$ *is finite for* p *small enough. For an ordinary random walk* $\beta = 1/2$ [*Kallianpur and Robbins* (1954)] *and hence* $\mathbf{E}_\alpha S_\alpha^p < \infty$ *for* $0 \leq p < 1/2$. *Some other examples of* β-*null recurrent processes are given in Myklebust, Karlsen and Tjøstheim* (2001).

PROOF. Let $G(r) = \sum_{k=0}^\infty r^k P^k$ and $G_{s,\nu}(r) = \sum_{k=0}^\infty r^k (P - s \otimes \nu)^k$. Then by (3.13) and (3.14) (with $\rho = \beta$ and $L_1 = L_s$), β-null recurrence is equivalent with

$$\nu G(r) s \sim (1-r)^{-\beta} L_s\left(\frac{1}{1-r}\right). \tag{3.17}$$

We have $B_n \stackrel{\text{def}}{=} \mathbf{P}_\alpha(S_\alpha > n) = \nu(P - s \otimes \nu)^n 1$. If (3.16) holds, then by (3.14) and (3.15) (with $\rho = 1 - \beta$ and $L_1 = L_0 = 1/L_s$),

$$B(r) \stackrel{\text{def}}{=} \sum_{k=0}^\infty r^k B_k = \nu G_{s,\nu}(r) 1 \sim (1-r)^{\beta-1} L_0\left(\frac{1}{1-r}\right). \tag{3.18}$$

Let $b_n = \mathbf{P}_\alpha(S_\alpha = n)$, $w_n = \mathbf{P}_\alpha(Y_n = 1)$ for $n \geq 1$ and $b_0 = 0$, $w_0 \stackrel{\text{def}}{=} 1$ and define the corresponding generating functions $w(r)$ and $b(r)$. By a first entrance decomposition

$$\begin{aligned} w_n &= \mathbf{P}_\alpha(Y_n = 1,\ S_\alpha \geq n) + \sum_{k=1}^{n-1} \mathbf{P}_\alpha(Y_{n-k} = 1) \mathbf{P}_\alpha(S_\alpha = k) \\ &= \sum_{k=0}^{n} w_{n-k} b_k, \qquad n \geq 1, \end{aligned} \tag{3.19}$$

which shows that $\{w_n\}$ is an undelayed renewal sequence corresponding to the increment sequence $\{b_n\}$. By (3.19) we get $w(r) = 1 + w(r)b(r)$. Since $b_n = B_{n-1} - B_n$ when $n \geq 1$, we find that $b(r) = 1 - B(r)(1-r)$. Hence

$$w(r) = \frac{1}{B(r)(1-r)}. \tag{3.20}$$

By (3.5) we find that $w_n = \mathbf{P}_\alpha(Y_n = 1) = \mathbf{E}_\nu[s(X_{n-1})] = \nu P^{n-1}s$ when $n \geq 1$. This gives

$$w(r) = 1 + r[\nu G(r)s]. \tag{3.21}$$

Combining (3.20) and (3.21) we finally obtain

$$1 + r[\nu G(r)s] = \frac{1}{[\nu G_{s,\nu}(r)1](1-r)}.$$

This identity in conjunction with (3.17) and (3.18) shows the equivalence. □

3.5. *Decomposing the chain and the number of regenerations.* The kernel estimator, ignoring the bandwidth parameter h for the moment, involves sums of type

$$S_n(g) = \sum_{j=0}^{n} g(X_j, \dots, X_{j+r-1}) \tag{3.22}$$

with $g \in \mathscr{G}_r^d$. Using the splitting technique this sum can be decomposed in a form which makes it amenable to a CLT argument.

Let $T(n)$ denote the complete number of regenerations in the time interval $[0, n]$; that is,

$$T(n) = \begin{cases} \max\{k\colon \tau_k \leq n\}, & \text{if } \tau_0 \leq n, \\ 0, & \text{otherwise,} \end{cases}$$

with

$$\tau_k = \begin{cases} \inf\{n \geq 0\colon Y_n = 1\}, & k = 0, \\ \inf\{n > \tau_{k-1}\colon Y_n = 1\}, & \text{when } k \geq 1, \end{cases}$$

and where $\tau_0 = \tau$ in (3.6). Then $S_n(g)$ can be written as

$$S_n(g) = U_0 + \sum_{k=1}^{T(n)} U_k + U_{(n)}, \tag{3.23}$$

where $\{U_k, k \geq 0\}$, $U_{(n)}$ are defined by

$$U_k = \begin{cases} \displaystyle\sum_{j=0}^{\tau_0} g(X_j, \dots, X_{j+r-1}), & \text{when } k = 0, \\ \displaystyle\sum_{j=\tau_{k-1}+1}^{\tau_k} g(X_j, \dots, X_{j+r-1}), & \text{when } k \geq 1, \\ \displaystyle\sum_{j=\tau_{T(n)}+1}^{n} g(X_j, \dots, X_{j+r-1}), & \text{when } k = (n). \end{cases} \tag{3.24}$$

Since [cf. Nummelin (1984), page 76]

$$\{(\tau_k - \tau_{k-1}, X_{\tau_{k-1}+1}, \dots, X_{\tau_k}),\ k \geq 1\}$$

are iid random elements which are independent of (X_0, Y_0), it follows that [cf. Nummelin (1984), page 135] $\{U_k, k \geq 1\}$ is a $(r-1)$-dependent stationary sequence, which is independent of the initial distribution of the Markov chain. We have for $k \geq 1$ [using $S_\alpha(0) = S_\alpha$ and $S_\alpha(1) = \inf\{n > S_\alpha\colon Y_n = 1\}$],

$$\begin{aligned}(U_k,\ U_{k+1}) & \\ &\overset{d}{=} \mathscr{L}_\alpha\left(\sum_{j=1}^{S_\alpha(0)} g(X_j, \dots, X_{j+r-1}),\ \sum_{j=S_\alpha(0)+1}^{S_\alpha(1)} g(X_j, \dots, X_{j+r-1})\right) \\ &\overset{d}{=} \mathscr{L}_\nu\left(\sum_{j=0}^{\tau_0} g(X_j, \dots, X_{j+r-1}),\ \sum_{j=\tau_0+1}^{\tau_1} g(X_j, \dots, X_{j+r-1})\right).\end{aligned} \tag{3.25}$$

We prove a law of large numbers for the sum $S_n(g)$, where $T(n)$ plays the role of the number of observations, and where the expected value is replaced by an integral with respect to the invariant measure as defined in Section 3.3.

LEMMA 3.2. *Assume that* (3.4) *holds. Let* $r \geq 1$, $g \in \mathscr{G}_r^d$ *and* $\|g\| \in L_r^1(\pi_s)$, *and let the process have an arbitrary initial distribution* λ. *Then*

$$\frac{S_n(g)}{T(n)} \underset{n}{\overset{\text{a.s.}}{\to}} \pi_s(g). \tag{3.26}$$

PROOF. Assume that $d = 1$ and $g \geq 0$. By (3.23), since $g \geq 0$, we can write

$$\sum_{k=1}^{T(n)} U_k \leq S_n(g) \leq U_0 + \sum_{k=1}^{T(n)+1} U_k.$$

By the definition of U_0, $\{\tau < \infty\} \subseteq \{U_0 < \infty\}$. Since the chain is Harris recurrent $\mathbf{P}_x(\tau < \infty) \equiv 1$. Hence $\mathbf{P}_\lambda(|U_0| < \infty) = 1$. From (3.7), (3.9) and (3.25) we have that $\mathbf{E}(U_k) = \pi_s g$. Because $T(n) \uparrow \infty$ a.s. as $n \to \infty$, the convergence result in (3.26) is then a consequence of the strong law of large numbers for $(r-1)$-dependent stationary variables. The rest of the proof is obvious since a general component of g can be written as a difference between two nonnegative g-functions. □

REMARK 3.5. The result of Lemma 3.2 is of somewhat academic character since $T(n)$ is not observable. However, it can be used to link $T(n)$ with a directly observable hitting time. Indeed, if $C \in \mathscr{E}^+$, the number of times the process is visiting C up to the time n is denoted by

$$T_C(n) = \sum_{k=0}^{n} 1_C(X_k) = S_n(1_C).$$

From (3.26) we have that $T_C(n)/T(n) \overset{\text{a.s.}}{\to} \pi_s 1_C$. In contrast to $T(n)$, the variable $T_C(n)$ is observable, and it is essential in stating applicable versions of the limit theorems of Sections 4 and 5.

To obtain a central limit result for $S_n(g)$ is more difficult, and the derivation is postponed to Section 4, where the bandwidth parameter is also included. However, such a result requires quite precise knowledge about the asymptotic behavior of the number of regenerations $T(n)$, and the rest of this section will be concerned with this matter.

We assume aperiodicity and that (3.4) and (3.16) hold with L_s of (3.16) normalized, which implies that the function

$$u(z) \overset{\text{def}}{=} z^\beta L_s(z), \qquad z \in R_+ \tag{3.27}$$

is strictly increasing in the interval $[z_0, \infty)$ for some z_0.

Then $\mathbf{E}_\alpha T(n) = \mathbf{E}_\alpha(\sum_{j=1}^n Y_j) = \sum_{k=0}^{n-1} \nu P^k s$, and it follows by (3.11) (with $h = s$),

$$\mathbf{E}_\alpha\left[\frac{T(n)}{u(n)}\right] = \frac{1}{\Gamma(1+\beta)} + o(1).$$

We need to extend this result to higher order moments.

LEMMA 3.3. *Let λ be any initial measure. Assume that* (3.16) *holds. Then*

$$\mathbf{E}_\lambda[T(n)]^m \sim \frac{m!\, n^{m\beta} L_s^m(n)}{\Gamma(1+m\beta)}. \tag{3.28}$$

PROOF. Let $\widetilde{T}(n) = \sum_{k=0}^n Y_k$, so that $T(n) = (\widetilde{T}(n) - 1)1(\widetilde{T}(n) > 0)$. Moreover, let $\mathcal{N}_+$ be the set of all positive integers and let $\mathcal{N}_+^k$ be the corresponding k-fold Cartesian product. By a straightforward calculation we can write

$$\mathbf{E}_\alpha\big[\widetilde{T}(n)\big]^m = \sum_{k=1}^m \sum_{l \in \Delta_{m,k}} \binom{m}{l} J_{n,k,l} \tag{3.29}$$

where $\Delta_{m,k} = \{l = (l_1, \ldots, l_k) \in \mathcal{N}_+^k\colon \sum l_i = m\}$ for $k \geq 1$, and where

$$\begin{aligned} J_{n,k,l} &= \sum_{h_1=0}^{n} \sum_{h_2=1}^{n-h_1} \cdots \sum_{h_k=1}^{n-h_1-\cdots-h_{k-1}} \mathbf{E}_\alpha\Big[Y_{h_1}^{l_1} Y_{h_1+h_2}^{l_2} \cdots Y_{h_1+\cdots+h_k}^{l_k}\Big] \\ &= \sum_{h_1=0}^{n} \sum_{h_2=1}^{n-h_1} \cdots \sum_{h_k=1}^{n-h_1-\cdots-h_{k-1}} w_{h_1} \cdots w_{h_k}, \end{aligned} \tag{3.30}$$

with $w_h \overset{\text{def}}{=} \mathbf{P}_\alpha(Y_h = 1)$. We can write $J_{n,k,l} = J_{n,k}$, since (3.30) shows that this quantity is independent of l. Let

$$J_k(r) \overset{\text{def}}{=} \sum_{n=0}^\infty J_{n,k} r^n, \qquad r \in [0, 1).$$

Then it can be shown from (3.30), (3.18) and (3.20) that

$$\begin{aligned}\sum_{n=0}^{\infty} J_{n,k} r^n &= \sum_{n=0}^{\infty}\sum_{h_1=0}^{n}\sum_{h_2=1}^{n-h_1}\cdots\sum_{h_k=1}^{n-h_1-\cdots-h_{k-1}} w_{h_1}\cdots w_{h_k} r^n \\ &= \sum_{h_1=0}^{\infty}\sum_{n=h_1}^{\infty}\sum_{h_2=1}^{n-h_1}\cdots\sum_{h_k=1}^{n-h_1-\cdots-h_{k-1}} w_{h_1}\cdots w_{h_k} r^n \\ &= \sum_{h_1=0}^{\infty}\sum_{h_2=1}^{\infty}\cdots\sum_{h_k=1}^{\infty} w_{h_1}\cdots w_{h_k}\sum_{n=1}^{\infty} r^{n+h_1+\cdots h_k} \\ &= w(r)(w(r)-1)^{k-1} r(1-r)^{-1} \\ &= w^k(r)(1-r)^{-1}(1+o(1))\end{aligned}$$

and hence

$$J_k(r) \sim w^k(r)(1-r)^{-1} \sim (1-r)^{-k\beta-1} L_s^k\left(\frac{1}{1-r}\right) \tag{3.31}$$

as $r \uparrow 1^-$. From (3.14), (3.15) with $\rho = k\beta + 1$, $L_1 = L_s^k$, since $\{J_{n,k}, n \ge 1\}$ is a monotone sequence in n, (3.31) implies

$$J_{n,k} \sim \frac{n^{k\beta} L_s^k(n)}{\Gamma(1+k\beta)} = \frac{u^k(n)}{\Gamma(1+k\beta)} \tag{3.32}$$

as $n \to \infty$. Inserting (3.32) into (3.29) gives

$$\mathbf{E}_\alpha[\widetilde{T}(n)]^m \sim \sum_{k=1}^{m}\left\{\sum_{l\in\Delta_{m,k}}\binom{m}{l}\right\}\frac{n^{k\beta}L_s^k(n)}{\Gamma(1+k\beta)} \tag{3.33}$$

and since $\Delta_{m,m} = \{\underline{1}\} = \{(1,\dots,1)\}$ and $\binom{m}{\underline{1}} = m!$, we finally obtain (3.28) by (3.33) and the relationship between $\widetilde{T}(n)$ and $T(n)$. □

REMARK 3.6. *In the rest of the paper when we write $a_n \ll b_n$ for two real-valued strictly positive sequences $\{a_n\}$ and $\{b_n\}$, this means that $a_n = o(b_n)$.*

Lemma 3.3 suggests that $T(n)$ behaves approximately as n^β. This is made precise in the following lemma.

LEMMA 3.4. *If the tail condition* (3.16) *holds, then $n^{\beta-\varepsilon} \ll T(n) \ll n^{\beta+\varepsilon}$ a.s. for all $\varepsilon > 0$. This is also true for $T_C(n)$. The lower bound requires only that*

$$\sup\{p > 0\colon \mathbf{E}_\alpha S_\alpha^p < \infty\} = \beta > 0, \tag{3.34}$$

which (*cf. Remark* 3.4) *is a weakening of* (3.16).

PROOF. Let $\varepsilon \in (0, \beta)$ and define $p = \beta - \varepsilon/2$. Assume that (3.34) holds. Then $\mathbf{E}_\alpha S_\alpha^p < \infty$ which again entails that $k^{-1/p} S_\beta(k) \to 0$ a.s. [cf. Chow and Teicher (1988), page 125] where $S_\beta(k) =^{\text{def}} \sum_{j=1}^k (\tau_j - \tau_{j-1})$. Let ω be an outcome so that this holds. Then there exists a finite constant $c = c(\omega)$ so that $S_\beta(k) \le ck^{1/p}$ for all k and by this inequality, when $\tau_0 = 0$,

$$T(n) = \max\{k\colon S_\beta(k) \le n\} \ge \max\{k\colon ck^{1/p} \le n\} = \left[\frac{n^p}{c^p}\right],$$

where $[\cdot]$ is the integer function. Hence for this outcome the lower bound is satisfied for $T(n)$. If $\tau_0 > 0$ some minor modifications of this argument are needed.

To prove the upper bound we assume (3.16). Then, for all $\eta > 0$, we have by the Markov inequality and the convergence of all moments of $T(n)/u(n)$ given by (3.28) that

$$\begin{aligned}\mathbf{P}_\alpha(T(n) \ge \eta n^{\beta+\varepsilon}) &\le \eta^{-m}\{u(n)/n^{\beta+\varepsilon}\}^m \mathbf{E}_\alpha\{T(n)/u(n)\}^m \\ &\le C_m n^{-m\varepsilon/2}, \qquad m \ge 1.\end{aligned}$$

Choosing $m > 2\varepsilon^{-1}$, the upper bound is implied by the Borel–Cantelli lemma.

Since $T_C(n)/T(n)$ converges with probability 1 to $\pi_s(C)$, the two variables must have the same bounds of this type. □

REMARK 3.7. Lemma 3.4 shows that $\hat{\beta} \stackrel{\text{def}}{=} \ln\{T_C(n)\}/\ln(n)$ is a strongly consistent estimator for β if (3.16) is fulfilled. Due to the slow convergence rate it is of limited practical use.

For our derivations in Section 4 we need to evaluate the difference in growth between $\pi_s^{-1}(C)T_C(n)$ and $T(n)$.

LEMMA 3.5. *Assume that C is small. Then for all $p \in (1/2, 1)$ we have*

$$T^{1-p}(n)\{T^{-1}(n)T_C(n) - \pi_s 1_C\} = o(1) \quad a.s. \tag{3.35}$$

and for all $\varepsilon > 0$,

$$P\big(\pi_s^{-1} 1_C T_C(n) \notin [T(n) \pm \varepsilon T^p(n)] \quad i.o.\big) = 0. \tag{3.36}$$

PROOF. From Remark 3.1 we have that $\pi_s 1_C < \infty$ and therefore without loss of generality we may assume that $\pi_s(C) = 1$. Define U_k with $g = 1_C$ in (3.23), (3.24). Then

$$T_C(n) = S_n(1_C) = U_0 + \sum_{k=1}^{T(n)} U_k + U_{(n)}.$$

By a result of Marcinkiewicz–Zygmund [Chow and Teicher (1988), page 125], $n^{-p} \times (\sum_{k=1}^n U_k - n) \to 0$ a.s. Because $T(n) \uparrow \infty$ a.s. as $n \uparrow \infty$ we have that $T^{-p}(n)(\sum_{k=1}^{T(n)} U_k - T(n)) = o(1)$ a.s. The correction terms U_0 and $U_{(n)}$ can be ignored since $n^{-p}U_n = o(1)$ a.s. Hence $T^{-p}(n)(T_C(n) - T(n)) = o(1)$ a.s., which implies both (3.35) and (3.26). □

Our next task is to derive a functional limit theorem for $T(n)$, which will be used extensively in the asymptotics of Sections 4 and 5. Recall the definition of the function u in (3.27) which is strictly increasing on $[z_0, \infty)$. Let

$$v(z) = u^{(-1)}(z) = \inf\{s: u(s) > z\}.$$

Then $v(u(z)) = u(v(z)) = z$ for all $z \in [z_0, \infty)$.

Consider the space $\mathscr{D}[0, \infty)$ of right continuous real-valued functions with finite left-hand limits, that is, the space of cadlag functions defined on $[0, \infty)$ [cf. Jacod and Shiryaev (1987), pages 288–322]. We write $\xrightarrow{\mathscr{L}_{\mathscr{D}[0,\infty)}}$ for weak convergence in $\mathscr{D}[0, \infty)$ and $\xrightarrow{\text{fd}}$ for convergence of finite-dimensional laws. A Lévy process is a stochastic process with stationary independent increments and sample paths in $\mathscr{D}[0, \infty)$. Consider the process

$$S_{\beta, z}(t) \stackrel{\text{def}}{=} \frac{1}{v(z)} \sum_{k=1}^{[zt]} (\tau_k - \tau_{k-1}), \qquad t \in [0, \infty), \; z \in R_+, \tag{3.37}$$

where $[zt]$ is the integer value of zt, that is, the largest integer not exceeding zt.

By (3.16) [cf. Bingham, Goldie and Teugels (1989), page 349] it follows that

$$S_{\beta, z} \underset{z}{\xrightarrow{\text{fd}}} S_\beta, \tag{3.38}$$

where S_β is the one-sided stable Lévy process defined by the marginal characteristic function $E[\exp^{\{i\zeta S_\beta(t)\}}] = \exp^{\{i\zeta^\beta t\}}$ for $\zeta \in R$ and $t \in [0, \infty)$. Moreover [cf. Kasahara (1984)] $S_{\beta, z} \underset{z}{\xrightarrow{\mathscr{L}_{\mathscr{D}[0,\infty)}}} S_\beta$. The Mittag–Leffler process [cf. Kasahara (1984)] with parameter β, $M_\beta = \{M_\beta(t), t \geq 0\}$ is defined as the inverse of S_β. It is a strictly increasing continuous stochastic process, and the characteristic functions describing the marginal distributions are given by

$$\mathbf{E}\Big[\exp^{\{i\zeta M_\beta(t)\}}\Big] = \sum_{k=0}^{\infty} \frac{(i\zeta t^\beta)^k}{\Gamma(1 + k\beta)}, \qquad \zeta \in R, \; t \geq 0. \tag{3.39}$$

An alternative description is given by

$$E(M_\beta^m(1)) = \frac{m!}{\Gamma(1 + m\beta)}, \qquad m \geq 0, \quad M_\beta(t) \stackrel{\text{d}}{=} t^\beta M_\beta(1).$$

We need the continuous scaled extension

$$T_n = \left\{ \frac{T([nt])}{u(n)}, \; t \geq 0 \right\} \tag{3.40}$$

of $T(n)$. The next theorem establishes a weak limit result.

THEOREM 3.2. *Let λ be any initial measure. Assume that the tail condition* (3.16) *holds. Then*

$$T_n \overset{\mathscr{L}_{\mathscr{D}[0,\infty)}}{\underset{n}{\longrightarrow}} M_\beta. \tag{3.41}$$

PROOF. By the method of moments and (3.39) we find that for each t,

$$T_n(t) \overset{d}{\underset{n}{\to}} M_\beta(t).$$

However, it is difficult to establish a functional weak convergence from the marginal convergences since M_β is not a Lévy process. In order to prove (3.41) it is an advantage to use a continuous index; that is, $T_z(t) \overset{\text{def}}{=} T([zt])/u(z)$. By (3.38) and the proof of Theorem A.1 in the Appendix with $S_\beta = A$ in that proof,

$$S_{\beta,z}^{(-1)} \overset{\mathscr{L}_{\mathscr{D}[0,\infty)}}{\underset{z}{\longrightarrow}} M_\beta \quad \text{where} \quad S_{\beta,z}^{(-1)}(t) = \inf\{x\colon\ S_{\beta,z}(x) > t\}. \tag{3.42}$$

In the rest of the proof we omit the index β and write $S_z = S_{\beta,z}$ and $S_z^{(-1)} = S_{\beta,z}^{(-1)}$.

To prove (3.41) it is sufficient to prove that

$$\sup_{0<t\le K} |T_z(t) - S_z^{(-1)}(t)| = o_P(1) \tag{3.43}$$

for all finite K. Assume that $\tau_0 = 0$ without loss of generality. Then

$$\left\{\sum_{j=1}^{n} Y_j > m\right\} = \{\tau_m < n\}, \qquad \left\{\sum_{j=1}^{n} Y_j < m\right\} = \{\tau_m > n\}. \tag{3.44}$$

Let $\eta > 0$. From (3.37) and (3.44) we have

$$\begin{aligned}\left\{S_{u(z)}^{(-1)}(t) < \eta\right\} &\subseteq \{S_{u(z)}(\eta) > t\} \\ &= \left\{\tau_{[u(z)\eta]} > zt\right\} = \left\{\sum_{j=1}^{[zt]} Y_j < [u(z)\eta]\right\} \\ &= \{T_z(t) < u^{-1}(z)[u(z)\eta]\}.\end{aligned}$$

In the same way we get

$$\left\{S_{u(z)}^{(-1)}(t) > \eta\right\} \subseteq \{S_{u(z)}(\eta) \le t\} = \{T_z(t) \ge u^{-1}(z)[u(z)\eta]\}.$$

Let $\varepsilon_1 \in (0,1)$ be arbitrary. Then for $\eta_1 < \eta_2$,

$$\begin{aligned}\left\{\eta_1 \le S_{u(z)}^{(-1)}(t) < \eta_2\right\} &\subseteq \left\{\eta_1(1-\varepsilon_1) < S_{u(z)}^{(-1)}(t) < \eta_2\right\} \\ &\subseteq \{u^{-1}(z)[u(z)\eta_1(1-\varepsilon_1)] \le T_z(t) < u^{-1}(z)[u(z)\eta_2]\},\end{aligned}$$

which gives

$$|T_z(t) - S_{u(z)}^{(-1)}(t)| < (\eta_2 - \eta_1) + \varepsilon_1 \eta_1 + \frac{1}{u(z)} \quad \text{when} \quad S_{u(z)}^{(-1)}(t) \in [\eta_1, \eta_2). \tag{3.45}$$

Let $\varepsilon > 0$ be given. For all s we have

$$\begin{aligned} &P\Big(\sup_{t\le K} |T_z(t) - S_{u(z)}^{(-1)}(t)| > \varepsilon\Big) \\ &\quad \le P\Big(\sup_{t\le K} |T_z(t) - S_{u(z)}^{(-1)}(t)| > \varepsilon, \sup_{t\le K} S_{u(z)}^{(-1)}(t) < s\Big) \\ &\qquad + P\Big(\sup_{t\le K} S_{u(z)}^{(-1)}(t) \ge s\Big). \end{aligned}$$

By (3.42),

$$\lim_{s\uparrow\infty} \lim_{z\to\infty} P\Big(\sup_{t\le K} S_{u(z)}^{(-1)}(t) \ge s\Big) = 0.$$

Hence for all $\delta > 0$ we can choose s_0 so large that

$$P\Big(\sup_{t\le K} S_{u(z)}^{(-1)}(t) \ge s_0\Big) < \delta$$

for all z large enough. For fixed $\varepsilon > 0$, we can choose $\eta_0, \dots, \eta_L$, z_1, ε_1 with $\eta_0 = 0$ and $\eta_L = s_0$ so that $\max_k(\eta_{k+1} - \eta_k) < \varepsilon/3$, $\varepsilon_1 < s_0^{-1}\varepsilon/3$ and $z_1 > v(3\varepsilon^{-1})$. Then by (3.45),

$$P\Big(\sup_{t\le K} |T_z(t) - S_{u(z)}^{(-1)}(t)| > \varepsilon, \sup_{t\le K} S_{u(z)}^{(-1)}(t) < s_0\Big) = 0, \qquad z > z_1$$

and therefore

$$P\Big(\sup_{t\le K} |T_z(t) - S_{u(z)}^{(-1)}(t)| > \varepsilon\Big) < \delta, \qquad z > z_1. \tag{3.46}$$

The function u is unbounded continuous and strictly increasing in an interval $[x_0, \infty)$ and therefore by (3.42),

$$\sup_{t\le K} |S_z^{(-1)}(t) - S_{u(z)}^{(-1)}(t)| = o_P(1).$$

Hence (3.46) implies (3.43). □

Our final result of this section again concerns the replacement of $T(n)$ by $T_C(n)$. In analogy with (3.40) we define $T_{n,C} = \{T_C([nt])/u(n),\ t \ge 0\}$.

DEFINITION 3.3. If $\{X_n\}$ and $\{\widetilde{X}_n\}$ are random elements of $D^d[0, \infty)$ they are said to be equivalent if the difference converges weakly to the zero-process.

If $\{X_n\}$ converges weakly in $D^d[0,\infty)$ to the zero-process, then we say that $\{X_n\}$ is negligible.

LEMMA 3.6. *Assume that the tail condition* (3.16) *holds. Then* $T_{C,n}/\pi_s 1_C$ *is equivalent with* T_n.

PROOF. Assume without loss of generality that $\pi_s(C) = 1$. It is enough to prove that $\sup_{t\le t_0} \xi_n(t) = o_p(1)$ where $\xi_n(t) \overset{\text{def}}{=} |T_{n,C}(t) - T_n(t)|$ for all t_0. Let $0 < \delta < 1$ and $n_\delta = n^{-(1-\delta)}$. We have $\sup_{t\le n_\delta} \xi_n(t) \le u^{-1}(n)\{T_C(n^\delta)+T(n^\delta)\} = o_P(1)$ and when $t \ge n_\delta$,

$$\begin{aligned}\sup_{n_\delta\le t\le t_0} \xi_n(t) &\le \{T_n(t_0)\} \sup_{n_\delta\le t\le t_0} \{T^{-1}([nt])(T_C([nt]) - T([nt]))\}\\ &\le \{T_n(t_0)\} \sup_{m\ge n^\delta} \{T^{-1}(m)(T_C(m) - T(m))\}\\ &= o_P(1)\end{aligned}$$

since $T_C(n)/T(n)$ converges to $\pi_s 1_C$ with probability 1 by Remark 3.5. □

4. Asymptotics with a smoothing parameter. So far we have neglected the smoothing bandwidth h_n of the kernel estimator when we have looked at the split chain decomposition of the sum $S_n(g)$ of (3.22)–(3.24). Now, to bring it closer to the actual form of the kernel estimator we let $g = g_h$ depend on h and write

$$\begin{aligned}S_n(g_h) &= \sum_{j=0}^{n} g_h(X_j,\dots,X_{j+r-1})\\ &= U_0(g_h) + \sum_{k=1}^{T(n)} U_k(g_h) + U_{(n)}(g_h),\end{aligned} \tag{4.1}$$

where $\{U_k(g_h), k \ge 1\}$ for each fixed h is a sequence of $(r-1)$ dependent identically distributed random variables. In the following we will sometimes use the symbol $U = U(g_h)$ to denote a random variable having this common distribution. Note that also $U_0(g_h)$ has this distribution if the chain has initial measure ν.

The purpose of this section is to prove a functional limit theorem for $S_n(g_h)$ and a corresponding CLT for a properly normalized quantity. The conditions that we will assume for these theorems are high level conditions mainly formulated in terms of the moments of $U(g_h)$. In Section 5 they will be translated into workable conditions in a kernel estimation situation.

To state the conditions we need some notation. We treat the scalar case when $g_h \in \mathscr{G}_r$ first and extend to the case where $g_h \in \mathscr{G}_r^d$ toward the end of the section. The common mean and variance (assuming that they exist) of the random variables $\{U_k(g_h), k \ge 1\}$ are denoted by

$$\mu = \mu(g_h) = \mathbf{E}\,U(g_h) = \pi_s(g_h) \tag{4.2}$$

and

$$\sigma^2 = \sigma^2(g_h) = \operatorname{Var}(U(g_h)). \tag{4.3}$$

The quantity

$$\bar{\sigma}^2 = \bar{\sigma}^2(g_h) = \sum_{k=-(r-1)}^{r-1} \operatorname{Cov}\left(U_{1+|k|}(g_h), U_1(g_h)\right) \tag{4.4}$$

often appears in our derivations and $\bar{\sigma}^2 \leq (2r-1)\sigma^2$. Note that $\bar{\sigma}^2 = \operatorname{Var}(n^{-1/2}\sum_{k=1}^n U_k) + o(1)$ and $\bar{\sigma}^2 = \sigma^2$ when the U_k's are iid. Finally, recall that [cf. (3.27)] $u(n) = n^\beta L_s(n)$. This quantity will be much used in the following.

The first condition simply states that μ and σ exist for each fixed h.

A_0. (i) $\mu(|g_h|) < \infty$, (ii) $\sigma^2(|g_h|) < \infty$.
The next condition concerns the relationship between $\sigma(g_h)$ and $\sigma(|g_h|)$ and the corresponding quantities for $\bar{\sigma}$ as $h \downarrow 0$.

A_1. (i) $\bar{\sigma}^{-1}(g_h)\bar{\sigma}(|g_h|) = O(1)$, (ii) $\bar{\sigma}^{-1}(g_h)\sigma(|g_h|) = O(1)$, (iii) $\bar{\sigma}^{-1}(|g_h|)\sigma(|g_h|) = O(1)$.
The third condition concerns the relationship between μ, σ and h.

A_2. (i) $h\mu(|g_h|) = O(1)$, (ii) $\liminf_{h\downarrow 0} h\bar{\sigma}^2(g_h) > 0$, (iii) $\liminf_{h\downarrow 0} h\bar{\sigma}^2(|g_h|) > 0$.
A bound for higher order moments is imposed in

A_3. (i) $\mathbf{E}|U(g_h) - \mu(g_h)|^{2m} \leq d_m h^{-2m+v}$, (ii) $\mathbf{E}|U(|g_h|) - \mu(|g_h|)|^{2m} \leq d_m h^{-2m+v}$ for some $m \geq 1$, $d_m > 0$ and for a $v \in \{0, 1\}$.
The two final conditions impose further restrictions on $\{h_n\}$.

A_4. (i) For some $\varepsilon > 0$, $m > 1$, $v \in \{0, 1\}$, $h_n^{-1} \ll n^{\beta\delta_m - \varepsilon}$, $\delta_m = (m-v)^{-1}(m-1)$, where β is determined by the tail condition (3.16).

A_5. (i) If $h_n^{-1} = o(u(n)) = o(n^\beta L_s(n))$, then $h_n U_0(|g_{h_n}|) = O_P(1)$.

We are ready to state and prove a basic functional limit result.

Following (3.40) we denote the process $\{T([nt])/u(n), t \geq 0\}$ by T_n and likewise for $T_{C,n}$. The standard Brownian motion defined for $t \geq 0$ is denoted by B, and $B \circ M_\beta$ denotes $\{B[M_\beta(t)], t \geq 0\}$. Weak convergence in $D^d[0, \infty)$ is written D^d.

Our theorems also hold for the positive recurrent case which corresponds to $\beta = 1$, $M_1(t) \equiv t$ and $u(n) = n$.

THEOREM 4.1. *Assume that the tail condition* (3.16) *is fulfilled or that* $\{X_t\}$ *is positive recurrent. If the conditions* A_0–A_5 *hold with* $m \geq 2$ *and* $v \in \{0, 1\}$, *then with*

$$\Delta_{n,h}(t) = u^{-1/2}(n)\bar{\sigma}^{-1}(g_h)\left\{S_{[nt]}(g_h) - \mu(g_h)T([nt])\right\}, \tag{4.5}$$

$$(\Delta_{n,h_n}, T_n) \xrightarrow[n]{\mathscr{D}^2} (B \circ M_\beta, M_\beta), \quad B \text{ and } M_\beta \text{ are independent.} \tag{4.6}$$

PROOF. The proof would be quite easy if $T(n)$ were independent of $\{U_k(g_h), k \geq 1\}$. It is the dependence of $\{U_k(g_h)\}$ on the sequence of stopping times τ_k which makes it intricate. The idea of the proof is first to establish a functional

CLT neglecting this dependence, and subsequently incorporating the stopping times via the asymptotic theory for $T(n)$ developed in the previous section. Accordingly, the proof is subdivided into four parts. In the first part we prove a functional CLT for $\{U_k(g_h),\ k = 1, \dots, [nt]\}$ just by exploiting that this is an $(r-1)$-dependent sequence. The remainder terms $U_0(g_h)$ and $U_{(n)}(g_h)$ are shown to be of smaller order in the second part, and in the third part $T(n)$ and its asymptotic theory is brought into the picture. The last part is devoted to the positive recurrent case.

Part 1. We introduce the scaled variables

$$W_k(g_h) = \bar{\sigma}^{-1}(g_h)(U_k(g_h) - \mu) \quad \text{and} \quad W = \bar{\sigma}^{-1}(U - \mu(g_h)),$$

where clearly $\sigma_W = \bar{\sigma}^{-1}\sigma$. Moreover, using A_2 and A_3,

$$\mathbf{E}\, W^{2m} = (h\bar{\sigma}^2)^{-m} h^m \mathbf{E}(U-\mu)^{2m} \le d'_m h^{-m+\nu} \tag{4.7}$$

for some constant d'_m. It is also convenient to rescale the bandwidth sequence $\{h_n\}$. Let $\{h_n\}$ be chosen according to A_4. Define $q_n = h_{u^{-1}(n)}$ so that $h_n = q_{u(n)}$. Then by A_4, $q_n^{-1} \ll n^{\delta_m}$. We consider $\{W_k(g_h), k \ge 1\}$, and we will prove that

$$Q_{n,q_n} \xrightarrow[n]{\mathscr{D}} B, \qquad Q_{n,h}(t) \stackrel{\text{def}}{=} n^{-1/2} \sum_{k=1}^{[nt]} W_k(g_h). \tag{4.8}$$

This can be proved by using an ordinary mixing array CLT and a tightness argument. Let t be fixed. By (4.7) and the condition on the rate of $\{q_n\}$, we have for some $m \ge 1$,

$$n^{-m} \sum_{k=1}^{[nt]} \mathbf{E}\, W_k^{2m}(g_{q_n}) \le d'_m t n^{-(m-1)} q_n^{-(m-v)} = o(1). \tag{4.9}$$

Thus the array satisfies a Liapounov condition. Since $\mathbf{E}\, Q^2_{n,\,q_n}(t) = t(1+o(1))$ we have by an appropriate CLT [Bergstrøm (1981), Theorem 1, page 161] that $Q_{n,\,q_n}(t) \xrightarrow[n]{d} \mathscr{N}(0,t)$ for all t. Using a standard argument, the same CLT and the fact that $Q_{n,\,q_n}$ has asymptotically independent increments [cf. Billingsley (1968), pages 68 and 69], we find that

$$Q_{n,\,q_n} \xrightarrow[n]{\text{fd}} B. \tag{4.10}$$

It remains to prove tightness. Without loss of generality we can assume that $r = 2$. Define

$$\chi^i_{2k-j,h} = \delta_{ij} \eta_h W_{2k-i}(g_h), \qquad 0 \le i,\ j \le 1,\ k \ge 1, \tag{4.11}$$

where $\eta_h = 2^{1/2}\sigma_W^{-1}(g_h)$ and δ_{ij} is the Kronecker symbol. Note that $\eta_h = O(1)$ by A_1.

Next, define

$$Q^i_n(t) \stackrel{\text{def}}{=} n^{-1/2} \sum_{k=1}^{[nt]} \chi^i_{k,q_n} = \eta_{q_n} n^{-1/2} \sum_{k=1}^{[(nt+i)/2]} W_{2k-1}(g_{q_n}), \qquad i = 0, 1.$$

The marginal arrays $\{\chi^i_{k,\,q_n}\}$ consist of independent variables. By (4.9) and an ordinary multivariate CLT we have that (4.10) is fulfilled for Q^i_n. If $\psi^i_n(s,t) \overset{\text{def}}{=} \mathbf{E}\{Q^i_n(t) - Q^i_n(s)\}^2$ satisfies, for all fixed $t_0 > 0$,

$$\lim_{\delta\downarrow 0} \sup_n \sup_{\substack{s\le t_0 \\ |t-s|\le\delta}} \psi^i_n(s,t) = 0, \tag{4.12}$$

then $Q^i_n \xrightarrow{\mathscr{D}} B$, $i = 0, 1$ by standard theory [Pollard (1984), Theorem 19, page 104]. Now, using A_1 and $|t-s| \le \delta$,

$$\psi^i_n(s,t) \le c_1 n^{-1}\eta^2_{q_n}(n|t-s|+2)\mathbf{E}\, W^2(g_{q_n}) \le c_2|t-s+2/n| \le c_3(\delta + 2n^{-1}),$$

which shows that (4.12) holds. Hence Q^i_n, $i = 1, 2$ are tight in $D[0,\infty)$. Since B is a continuous process and $Q_{n,\,q_n} = \eta^{-1}_{q_n}\sum_{i=0}^{1} Q^i_n$ and η^{-1}_h is bounded, by A_1 we can conclude that $\{Q_{n,\,q_n}\}$ is tight [cf. Jacod and Shiryaev (1987), page 317].

The next part takes care of the edge terms.

Part 2. By (4.1) and the definition of $\Delta_{n,\,h_n}$ in the statement of the theorem we have that with $U_{0,\,h}(t) = \sum_{j=0}^{\tau_0\wedge[nt]} g_h(X_j,\dots,X_{j+r-1})$, $U_{h,(n)}(t) = U_{([nt])}(g_h)$, it is enough to show that

$$\delta_{g_h,n}(t) \overset{\text{def}}{=} u^{-1/2}(n)\bar\sigma^{-1}(g_h)\{U_{h,\,0}(t) + U_{h,\,(n)}(t)\} \tag{4.13}$$

is negligible in the sense of Definition 3.3. By A_2, A_4 and A_5,

$$\left|u^{-1/2}(n)\bar\sigma^{-1}(g_{h_n})U_{h_n,\,0}(t)\right| \le \left\{h_n u(n) h_n\bar\sigma^2(g_{h_n})\right\}^{-1/2} h_n U(|g_{h_n}|) = o_P(1)$$

independent of t. Hence we can neglect this term. Moreover, since $T_n([nt]) = T_n(t)u(n)$,

$$\begin{aligned}|U_{h,\,(n)}(t)| &\le U_{T([nt])+1}(|g_h|)\\ &= \bar\sigma(|g_h|)W_{T([nt])+1}(|g_h|) + \mu(|g_h|)\\ &= \bar\sigma(|g_h|)u^{1/2}(n)\big\{Q'_{u(n),h}(T_n(t)+1/u(n)) - Q'_{u(n),\,h}(T_n(t))\big\}\\ &\quad + \mu(|g_h|),\end{aligned}$$

where

$$Q'_{n,\,h}(t) = n^{-1/2}\sum_{k=1}^{[nt]} W_k(|g_h|).$$

The tail condition (3.16) guarantees by Theorem 3.2 that T_n has a specified asymptotic distribution. Since the conditions A_0–A_4 hold for $|g_h|$ as well as for g_h Part 1 of the proof, with g_h replaced by $|g_h|$, and the continuous mapping theorem, implies that $\xi_n(t) \overset{\text{def}}{=} Q'_{n,\,q_n}(t+1/n) - Q'_{n,\,q_n}(t)$ is negligible. Again by the continuous mapping theorem with the map $D^2[0,\infty) \mapsto D[0,\infty)$ given by $(a,b) \mapsto a \circ b$, the process $\xi_n \circ T_n$ converges to zero. This gives

$$u^{-1/2}(n)\bar{\sigma}^{-1}(g_{h_n})|U_{h_n,(n)}(\cdot)| \\ \le \frac{\bar{\sigma}(|g_h|)}{\bar{\sigma}(g_h)}(\xi_{u(n)} \circ T_n) + \{u(n)h_n\bar{\sigma}^2(g_h)h_n\}^{-1/2}h_n\mu(|g_h|) \xrightarrow[n]{\mathscr{D}} 0$$

because of A_1, A_2 and A_4.

Part 3. By Part 2 of the proof we can neglect the edge terms and (Δ_{n,h_n}, T_n) is equivalent with

$$(Z_{n,h_n}, T_n) = (Q_{u(n),q_{u(n)}} \circ T_n, T_n),$$

where Q is defined in (4.8). Let $(B_n, A_n) \overset{\text{def}}{=} (Z_{n,h_n}, S_{\beta,n})$ where $S_{\beta,n}$ is defined in (3.37). By the proof of Theorem 3.2 we have that $S^{(-1)}_{\beta,n}$ and T_n are equivalent processes. The proof of (4.6) and the asymptotic independence of B and M_β now follow by Theorem 3.2, (4.8) and Theorem A.1 in the Appendix.

Part 4. In the positive recurrent case $\{X_t\}$ is ergodic which implies that $T_n \xrightarrow{\mathscr{D}} I$ where $I(t) \equiv t$. Thus by (4.8) the proof is complete. □

For later applications it is necessary to restate Theorem 4.1 for the observable stopping process $T_{C,n} = \{T_C([nt])/u(n), t \ge 0\}$.

COROLLARY 4.1. *Assume that the conditions of Theorem* 4.1 *hold with* A_2(i) *strengthened to* $h^{1/2}\mu(|g_h|) = o(1)$ *as* $h \downarrow 0$. *Let* C *be a small set and define* $\Delta_{C,n,h_n}(t) = u^{-1/2}(n)\bar{\sigma}^{-1}(g_{h_n})\{S_{[nt]}(g_{h_n}) - \pi_s^{-1}(C)\mu(g_{h_n})T_C([nt])\}$. *Then the sequences* $\{(\Delta_{n,h_n}, T_n)\}$ *and* $\{(\Delta_{C,n,h_n}, \pi_s^{-1}(C)T_{C,n})\}$ *are equivalent.*

PROOF. By definitions we have

$$\Delta_{C,n,h}(t) - \Delta_{n,h}(t) = \bar{\sigma}^{-1}(g_h)\mu(g_h)u^{-1/2}(n)\{\pi_s^{-1}(C)T_C([nt]) - T([nt])\} \\ = \{h\sigma^2(g_{h_n})\}^{-1/2}h^{1/2}\mu(g_h)\bar{\sigma}(\pi_s^{-1}(C)1_C)\Delta_n(t),$$

where $\Delta_n(t) \overset{\text{def}}{=} u^{-1/2}(n)\bar{\sigma}^{-1}(\pi_s^{-1}(C)1_C)\{S_{[nt]}(\pi_s^{-1}(C)1_C) - T([nt])\}$. We now apply Theorem 4.1 to $\Delta_n(t)$ in the simplified situation with $g = \pi_s^{-1}(C)1_C$. It follows straightforwardly that $\Delta_n \xrightarrow{\mathscr{D}} B \circ M_\beta$. Since $h^{1/2}\mu(g_h) \le h^{1/2}\ \mu(|g_h|) = o(1)$, it follows by A_2(ii) that $\Delta_{C,n,h_n} - \Delta_{n,h_n} \xrightarrow{\mathscr{D}} 0$. The equivalence of $\pi_s^{-1}(C)T_{C,n}$ and T_n is the content of Lemma 3.6. □

The limit distributions obtained in Theorem 4.1 and Corollary 4.1 are non-Gaussian. However, a Gaussian distribution can be obtained by a stochastic normalization.

THEOREM 4.2. *Assume that the conditions of Theorem* 4.1 *hold with* A_2(i) *strengthened to* $h^{1/2}\mu(|g_h|) = o(1)$ *as* $h \downarrow 0$. *Let* C *be a small set. Then*

$$T_C^{1/2}(n)\pi_s^{1/2}(C)\bar{\sigma}^{-1}(g_{h_n})\Big\{T_C^{-1}(n)S_n(g_{h_n}) - \pi_s^{-1}(C)\mu(g_{h_n})\Big\} \xrightarrow[n]{d} \mathcal{N}(0,1). \tag{4.14}$$

If it is assumed in addition that there exists a constant μ_0 *such that* $\mu(g_h) = \mu_0 + O(h^2)$ *and* $h_n^{-1} \gg n^{\beta/5+\varepsilon}$, *then* $\mu(g_{h_n})$ *can be replaced by* μ_0 *in* (4.14).

PROOF. Denote the left-hand side of (4.14) by Λ_{n,h_n}. We have that $\Lambda_{n,h} = \pi_s^{1/2}(C)\{T_C(n)/u(n)\}^{-1/2}\Delta_{C,n,h}(1)$. By Lemma 3.5, Theorems 3.2 and 4.1 and Corollary 4.1 the implication of (4.14) follows. The last statement of the theorem follows by A_2 because

$$T_C^{1/2}(n)\Big\{h_n\bar{\sigma}^2(g_{h_n})\Big\}^{-1/2} h_n^{1/2}|\mu(g_{h_n}) - \mu_0| \le c_1\Big\{T_C(n)h_n^5\Big\}^{1/2} = o_P(1). \qquad \square$$

A multivariate extension to $g_h \in \mathcal{G}_r^d$ with $d > 1$ is useful. The asymptotic covariance matrix of $n^{-1/2}\sum_{k=1}^n U_k(g_h)$ is given in complete analogy with $\bar{\sigma}^2$ by

$$\begin{aligned}\overline{\Sigma} = \overline{\Sigma}(g_h) &= \tfrac{1}{2}\sum_{k=-(r-1)}^{r-1} \operatorname{Cov}(U_{1+|k|}(g_h), U_1(g_h)) \\ &\quad + \operatorname{Cov}\big(U_1(g_h), U_{1+|k|}(g_h)\big).\end{aligned} \tag{4.15}$$

The norm of g_h is defined by $\|g_h(x_1,\dots,x_r)\| = \{\sum_{i=1}^d g_{i,h}^2(x_1,\dots,x_r)\}^{1/2}$ where $g_{i,h}$ is the ith component of g_h.

COROLLARY 4.2. *Let* $g_h \in \mathcal{G}_r^d$. *Assume that the conditions of Theorem* 4.1 *hold for* $\|g_h\|$ *and* $\{X_t\}$ *with* A_2(i) *strengthened to* $h^{1/2}\mu(\|g_h\|) = o(1)$ *as* $h \downarrow 0$. *In addition we assume that*

$$\limsup h\overline{\Sigma}(g_h) \ge \Sigma_0 \tag{4.16}$$

for some positive definite matrix Σ_0 *and*

$$\|\overline{\Sigma}(g_h)\|^{-1}\,\overline{\sigma}^2(\|g_h\|) = O(1). \tag{4.17}$$

If C *is a small set, then* (4.14) *holds with* $\bar{\sigma}^{-1}(g_h)$ *replaced by* $\overline{\Sigma}^{-\frac{1}{2}}(g_h)$, *and the limit is the* d*-dimensional standard multivariate normal distribution. Moreover, a multivariate extension of* (4.6) *holds where* B *is replaced by a* d*-dimensional Brownian motion.*

If there exists a constant μ_0 *such that* $\mu(g_h) = \mu_0 + O(h^2)$ *and* $h_n^{-1} \gg n^{\beta/5+\varepsilon}$, *then* $\mu(g_{h_n})$ *can be replaced by* μ_0.

PROOF. Analogous to (4.5) and (4.13), define

$$\Delta_{n,h}(t) = u^{-1/2}(n)\overline{\Sigma}^{-1/2}(g_h)\Big\{S_{[nt]}(g_h) - \mu(g_h)T([nt])\Big\}$$

and

$$\delta_{g_h,n}(t) \stackrel{\text{def}}{=} u^{-1/2}(n)\overline{\Sigma}^{-1/2}(g_h)\Big\{U_{0,h}(t) + U_{(n),h}(t)\Big\} \tag{4.18}$$

Now since $|g_{i,h}| \le \|g_h\|$ we have $|U(g_{i,h})| \le U(\|g_h\|)$ for $i = 1, \dots, d$. Using this inequality we find that $\|U(g_h)\| \le d^{1/2}U(\|g_h\|)$ and

$$\|\delta_{g_h,n}(t)\| \le d^{1/2}\|\overline{\Sigma}(g_h)\|^{-1/2}\bar{\sigma}(\|g_h\|)\delta_{\|g_h\|,n}(t). \tag{4.19}$$

From the conditions assumed in the corollary it follows that the conditions in Theorem 4.1 for $\|g_h\|$ are satisfied. In particular this implies that $\delta_{\|g_{h_n}\|,n}$ is negligible. Together with (4.17), (4.18) and (4.19) this means that $\|\delta_{h_n}, n\|$ is negligible.

We have thus shown that Δ_{n,h_n} is equivalent with Z_{n,h_n} defined by

$$Z_{n,h_n}(t) = u^{-1/2}(n) \sum_{k=1}^{T([nt])} W_k(g_{h_n}) = Q_{u(n),q_{u(n)}}(T_n(t))$$

with $Q_{n,h}$ defined by (4.8) and $W_k(g_h) \stackrel{\text{def}}{=} \bar{\Sigma}^{-\frac{1}{2}}(g_h)(U_k(g_h) - \mu(g_h))$. It is enough to prove that

$$a'Q_{n,q_n} \xrightarrow[n]{\mathscr{D}} a'B \qquad \forall\, a \in R^d, \tag{4.20}$$

where a' means the transpose of a. Let $f_h = b_h' g_h$ where $b_h = (h\overline{\Sigma}(g_h))^{-1/2}a$. Then

$$\mu(f_h) = b_h'\mu(g_h), \qquad \bar{\sigma}^2(f_h) = h^{-1}\|a\|^2.$$

Moreover,

$$\begin{aligned} a'Q_{n,h} &= n^{-1/2} \sum_{k=1}^{[nt]} h^{1/2}h^{-1/2}a'\overline{\Sigma}^{-\frac{1}{2}}(g_h)\big(U_k(g_h) - \mu(g_h)\big) \\ &= \|a\|n^{-1/2} \sum_{k=1}^{[nt]} \overline{\sigma}^{-1}(f_h)\big(U_k(f_h) - \mu(f_h)\big) \\ &= \|a\|\,Q^f_{n,h} \quad \text{say,} \end{aligned}$$

where $Q^f_{n,n}$ is defined by (4.8) with $W_k(g_h)$ replaced by $W_k(f_h)$. Using $h^{1/2}$ $\mu(\|g_h\|) = o(1)$ as $h \downarrow 0$ and applying A_1 and A_3 to $\|g_h\|$ we have

$$\begin{aligned} \mathbf{E}W_k^{2m}(f_h) &= \bar{\sigma}^{-2m}(f_h)\mathbf{E}\big[b_h'(U(g_h) - \mu(g_h))\big]^{2m} \\ &\le c_2\bar{\sigma}^{-2m}(f_h)\|b_h\|^{2m}\big[\mathbf{E}\ U^{2m}(\|g_h\|) + \mu^{2m}(\|g_h\|)\big] \\ &\le c_3h^m h^{-(2m-v)} \\ &= c_3h^{-(m-v)}. \end{aligned}$$

We have also used that $\|b_h\|$ is bounded with respect to h; in fact it follows from (4.16) that $\limsup \|b_h\|^2 \le \|a\|^2\rho$ where ρ denotes the spectral radius of Σ_0^{-1}. Hence the analogue of (4.9) is satisfied and (4.20) is proved using the method of Part 1 of the proof of Theorem 4.1. We now easily get the multivariate extension of Theorem 4.1 claimed in the second statement of the corollary by a Cramér–Wold argument: let B_d denote a d-dimensional Brownian motion. It is enough to prove that

$$(4.21)\qquad \left(a'\Delta_{n,h_n}, T_n\right) \xrightarrow[n]{\mathscr{D}} \left(\|a\| B \circ M_\beta, M_\beta\right) \quad B \text{ and } M_\beta \text{ are independent,}$$

since $a'B_d$ has the same distribution as $\|a\|B$. But (4.21) follows from (4.20) and the proof of Theorem 4.1.

The first and the last statement of the corollary are proved as in the proof of Theorem 4.2. □

5. Asymptotics for some nonparametric statistics. The objective of this section is to extend nonparametric kernel estimation from the traditional stationary case [see, e.g., Robinson (1983)] to the null recurrent case. This will be done using the result established in the preceding section.

We assume aperiodicity and (3.4). In addition we assume that the state space $E \subseteq R$ so that X_t is one-dimensional. All of these conditions can be relaxed. The bandwidth $h = h_n$ is assumed to satisfy $h_n \downarrow 0$, and with no loss of generality we also assume that $h_n \leq 1$. Let $K\colon R \to R$ be a kernel function and for a fixed x let $K_{x,h}(y) = h^{-1}K\{(y-x)/h\}$. We will consider estimation of both the conditional mean and an analogue of the density function. Concerning the latter we look at the normalized density relative to a small set C; that is, if the invariant measure π_s has a density p_s, we define $p_C = \pi_s^{-1}(C)p_s$. The density p_C does not depend upon s as is easily verified from the nonuniqueness property stated in Remark 3.1.

By analogy with the ordinary kernel estimator for a density in the positive recurrent case, an estimator for p_C is defined by

$$(5.1)\qquad \hat{p}_C(x) = \hat{p}_{C,n,h}(x) = T_C^{-1}(n)\sum_{t=0}^{n} K_{x,h}(X_t).$$

Next, coming to the conditional expectation, note that for a function $\xi \in \mathscr{G}^d$, in the Markov chain terminology adopted in this paper the conditional mean of $\xi(X_t)$ given $X_{t-1} = x$ is written $P\xi(x) = \mathbf{E}[\xi(X_t)|X_{t-1} = x]$. A kernel estimator for the conditional mean function $P\xi$ is given by

$$(5.2)\qquad \widehat{P}\xi(x) = \widehat{P}\xi_{n,h}(x) = \left\{\sum_{t=0}^{n} K_{x,h}(X_t)\right\}^{-1}\left\{\sum_{t=0}^{n} \xi(X_{t+1})K_{x,h}(X_t)\right\}$$

and is seen to coincide with the traditional Nadaraya–Watson estimator in the positive recurrent case. The corresponding conditional variance function is given by

$$(5.3)\qquad V\xi = P(\xi \otimes \xi) - P\xi \otimes P\xi.$$

If we replace the two terms of the right-hand side of (5.3) by estimators defined by (5.2), we obtain a conditional variance estimator $\widehat{V}\xi(x)$.

The expressions (5.1)–(5.3) can be put into the framework of Section 4 by considering

$$S_n(g_h) = \sum_{t=0}^{n} g_h(X_t, X_{t+1}) = U_0(g_h) + \sum_{k=1}^{T(n)} U_k(g_h) + U_{(n)}(g_h)$$

with

$$g_h(u, w) = K_{x,h}(u)\xi_0(u, w) \quad \text{or} \quad g_h(u, w) = K_{x,h}(u)\xi(w) \tag{5.4}$$

with ξ_0 and ξ being arbitrary functions in $\mathscr{G}_2^d$ and $\mathscr{G}^d$. There are no difficulties in extending the theory to $\mathscr{G}_r^d$ with $r > 2$, but it is notationally more complex.

We will prove consistency and supply results for the asymptotic distribution of kernel estimates at a fixed point $x \in E$. Both types of result require an adaption of the theory of Section 4 in the form of two lemmas. Basically, these lemmas secure the existence of the first two moments as required in assumption A_0, and supply bounds of higher order moments of the kind contained in A_3. To be able to prove the lemmas and the subsequent theorems we need to impose some (relatively mild) conditions on the kernel function K and the invariant density p_s. To this end let $\mathscr{N}_x(h) = \{y: K_{x,h}(y) \neq 0\}$ and $\mathscr{N}_x = \mathscr{N}_x(1)$. In our context a locally bounded function will be taken to mean bounded in a neighborhood of x and a locally continuous function is continuous at the point x. Without loss of generality we may assume that this neighborhood equals $\mathscr{N}_x$, and local continuity implies local boundedness. This is so since $\mathscr{N}_x(h) = x \oplus h\mathscr{N}_0$. Again, we use $c_1, c_2, \dots$ as a sequence of generic constants in our proofs.

The following condition is always assumed:

B_0. The kernel K is nonnegative, $\int K(u)\,du < \infty$ and $\int K^2(u)\,du < \infty$.

The next condition is standard:

B_1. (i) $\int K(u)\,du = 1$, (ii) $\int uK(u)\,du = 0$.

Condition B_2, however, is more specific. Note that the last part of (ii) follows from (i) under wide conditions [cf. comment after (3.2)].

B_2. (i) The support $\mathscr{N}_0$ of the kernel is contained in a compact set.
(ii) The kernel is bounded, and $\mathscr{N}_x$ is a small set.

The next condition is a smoothness condition on p_s.

B_3. The invariant measure π_s has a locally continuous density p_s which is locally strictly positive; that is, $p_s(x) > 0$.

The last condition is a sort of uniform local continuity condition on the transition probability.

B_4. For all $\{A_h\} \in \mathscr{E}$ so that $A_h \downarrow \varnothing$ when $h \downarrow 0$, $\lim_{h\downarrow 0} \limsup_{y\to x} P(y, A_h) = 0$.

The following result is needed so often in our derivations that we state it as a separate remark.

REMARK 5.1. If B_2(ii) holds and $\xi \in \mathscr{G}$ is a locally bounded function, then, since a small function is a special function, by Propostion 5.13 in Nummelin with $m_0 = 1$ we have $\sup G_{s,\nu} 1_{\mathscr{N}_x} \xi < \infty$.

The first lemma is a generalization to the null recurrent case of the well-known results in the one-dimensional positive recurrent case that $\mathbf{E}\{K_{x,h}(X_t)\xi_0(X_t, X_{t+1})\} = p(x) \int P(x, dz)\xi_0(x, z) \int K(u)\,du + o(1)$ and $h \operatorname{Var}\{K_{x,h}(X_t)\xi_0(X_t, X_{t+1})\} = p(x) \int P(x, dz)\xi_0^2(x, z) \int K^2(u)\,du + o(1)$, which hold under some mild regularity conditions. Recall that in the notation of Section 3.3,

$$\widetilde{I}_{\xi_0}(y, dz) = P(y, dz)\xi_0(y, z), \quad \tilde{\xi}_0(y) = \widetilde{I}_{\xi_0} 1(y) = \int P(y, dz)\xi_0(y, z). \tag{5.5}$$

As before, when we write $\tilde{\xi}_0(y)$, the integral in (5.5) is implicitly assumed to exist. Also note that $\|\tilde{\xi}_0\|^p \le \{\widetilde{I}_{\|\xi_0\|^2} 1\}^{p/2}$ when $p \ge 0$. If $\xi_0(u, w) = \xi(u)$; that is, $\xi \in \mathscr{G}^d$, then $\tilde{\xi}_0 = \xi$.

LEMMA 5.1. *Let $g_h(u, w) = K_{x,h}(u)\xi_0(u, w)$ where $\xi_0 \in \mathscr{G}_2^d$, and let $\mu(g_h)$ and $\overline{\Sigma}(g_h)$ be defined by* (4.2) *and* (4.15).

(a) *Assume that B_2 and B_3 hold, that $\widetilde{I}_{\|\xi_0\|^2} 1$ is locally bounded and $\widetilde{I}_{\xi_0} 1$, $\widetilde{I}_{\xi_0 \otimes \xi_0} 1$ are locally continuous. Then*

$$\mu(g_h) = p_s(x)\tilde{\xi}_0(x) \int K(u)\,du + o(1) \tag{5.6}$$

$$h\overline{\Sigma}(g_h) = p_s(x)\widetilde{I}_{\xi_o \otimes \xi_0} 1(x) \int K^2(u)\,du + hA(g_h) + hA'(g_h) + o(1), \tag{5.7}$$

where $A(g_h) = \pi_s \widetilde{I}_{g_h} \otimes G_{s,\nu} \widetilde{I}_{g_h} 1$. Let $\Sigma(g_h)$ be the multivariate analogue of $\sigma^2(g_h)$ of (4.3). *If $hA(g_h) = o(1)$, then $h\overline{\Sigma}(g_h) = h\Sigma(g_h) + o(1) = h\pi_s(g_h \otimes g_h) + o(1)$.*

(b) *If in addition $\tilde{\xi}_0(x) = 0$ or B_3 and B_4 hold, then $hA(g_h) = o(1)$. If $\xi_0(u, w) = \xi(w) - P\xi(u)$, then*

$$h\overline{\Sigma}(g_h) = p_s(x) V\xi(x) \int K^2(u) du + o(1). \tag{5.8}$$

PROOF. The proof of (5.6) follows immediately from B_2(i), B_3, (3.7), the continuity of $\widetilde{I}_{\xi_0} 1$ and Bochner's lemma [cf. Wheeden and Zygmund (1977), Chapter 9]. The rest of the proof is more intricate and is subdivided into two steps according to the subdivision of the lemma in parts (a) and (b).

PROOF OF THE REST OF PART (a). We start by the scalar case ($d = 1$), because, as will be indicated, the general case can be deduced from this case.

We assume with no loss of generality that the chain is started with initial measure ν, so that $\{U_k, k \geq 0\}$ is a stationary sequence of one-dependent variables. Then

$$\bar{\sigma}^2(g_h) = \mathbf{E}_\nu U_0^2 - \mathbf{E}_\nu^2 U_0 + 2\mathbf{E}_\nu(U_0 U_1) - 2\mathbf{E}_\nu^2 U_0.$$

Some manipulation of this expression [cf. Appendix A, Karlsen and Tjøstheim (1998), pages 49 and 50] yields

$$\bar{\sigma}^2(g_h) = \pi_s g_h^2 - \pi_s^2 g_h - 2\pi_s g_h \pi_s(s g_{h,\nu}) + 2\pi_s \psi, \tag{5.9}$$

where $g_{h,\nu}(u) = \int \nu(dz) g_h(u,z)$ and $\psi(x) = \mathbf{E}_x\{g_h(X_0, X_1)\sum_{j=1}^{S_\alpha} g_h(X_j, X_{j+1})\}$.

Using Nummelin [(1984), (4.16a), page 61] and (3.9) we find that

$$\mathbf{E}\left\{\sum_{j=1}^{S_\alpha} g_h(X_j, X_{j+1}) | \mathscr{F}_1^X \vee \mathscr{F}_0^Y\right\} = G_{s,\nu}\tilde{g}_h(X_1),$$

and we obtain

$$\psi(x) = \widetilde{I}_{g_h} G_{s,\nu}\tilde{g}_h(x) \quad \text{and} \quad \pi_s\psi = \pi_s \widetilde{I}_{g_h} G_{s,\nu}\tilde{g}_h. \tag{5.10}$$

The appropriate generalization of (5.9) and (5.10) to the multivariate case is given by

$$\overline{\Sigma}(g_h) = \pi_s(g_h \otimes g_h) + A(g_h) + A'(g_h) - B(g_h) - B'(g_h), \tag{5.11}$$

$$\begin{aligned} A(g_h) &= \pi_s \widetilde{I}_{g_h} \otimes G_{s,\nu}\tilde{g}_h, \\ B(g_h) &= 2^{-1}(\pi_s g_h \otimes \pi_s g_h) + \pi_s g_h \otimes \pi_s(s g_{h,\nu}). \end{aligned} \tag{5.12}$$

This can be verified by looking at $g_{h,a} = a' g_h$ and noting that $\bar{\sigma}^2(g_{h,a}) = a'\overline{\Sigma}(g_h)a$ with $a \in R^d$. We can then identify terms in (5.11), (5.12) with corresponding terms in (5.9), (5.10).

It is now straightforward to obtain (5.7). Indeed, by (5.6) $\mu(g_h) = O(1)$. The term $\pi_s(s \cdot g_{h,\nu})$ can be treated likewise, and hence $hB(g_h) = o(1)$. Similarly, we have

$$h\pi_s(g_h \otimes g_h) = p_s(x)\widetilde{I}_{\xi_0 \otimes \xi_0} 1(x) \int K^2(u)\, du + o(1)$$

and (5.7) follows.

Using the same reasoning as when deriving (5.11) and (5.12), we have

$$\begin{aligned} \Sigma(g_h) = \pi_s(g_h \otimes g_h) &+ A(g_h) + A'(g_h) \\ &- 1 \otimes \pi_s g_h - (1 \otimes \pi_s g_h)' - \pi_s g_h \otimes \pi_s g_h. \end{aligned}$$

If $hA(g_h) = o(1)$, it therefore follows that $h\overline{\Sigma}(g_h) = h\Sigma(g_h) + o(1) = h\pi_s(g_h \otimes g_h) + o(1)$, and part (a) of the lemma is proved.

PROOF OF PART (b). We can write

$$hA(g_h) = \pi_s I_{K_{x,h}} \widetilde{I}_{\xi_0} \otimes \psi_h, \qquad \psi_h \stackrel{\text{def}}{=} G_{s,\nu} h I_{K_{x,h}} \tilde{\xi}_0.$$

Because a small set is a special set it follows by B_2 (ii) (cf. Remark 5.1) and the local boundedness of $\widetilde{I}_{\|\xi_0\|}$, that we have $\|\psi_h\| \leq c_1$, and hence that $\|hA(g_h)\| = O(1)$. If $\tilde{\xi}_0(x) = 0$, then $\sup \|\psi_h\| = o(1)$ since $\sup_{y\in\mathscr{N}_x(h)} \|\tilde{\xi}_0\|(y) = o(1)$ by local continuity of $\tilde{\xi}_0$, and hence $\|hA(g_h)\| = o(1)$.

Assume that B_3 and B_4 are true. Then, since $\widetilde{I}_{\|\xi_0\|}1$ is locally bounded and since B_2 (ii) holds, $\|\psi_h\| \leq c_1 f_h$ where $f_h = G_{s,\nu} 1_{\mathscr{N}_x(h)}$. This gives

$$\|hA(g_h)\| \leq c_3 \pi_s I_{K_{x,h}} \widetilde{I}_{\|\xi_0\|} f_h. \tag{5.13}$$

By the Cauchy–Schwarz inequality,

$$\begin{aligned} \int P(y,dz)\|\xi_0(y,z)\| f_h(z) &\leq c_4 \left\{ \int P(y,dz)\|\xi_0(y,z)\|^2 \right\}^{1/2} \left\{ P f_h^2(y) \right\}^{1/2} \\ &= c_4 \left\{ \widetilde{I}_{\|\xi_0\|^2} 1(y) \right\}^{1/2} \left\{ P f_h^2(y) \right\}^{1/2}. \end{aligned} \tag{5.14}$$

By the local boundedness of $\widetilde{I}_{\|\xi_0\|^2}$,

$$\sup_{y\in\mathscr{N}_x} \widetilde{I}_{\|\xi_0\|^2}(y) < \infty. \tag{5.15}$$

Inserting (5.14) into (5.13) and by B_3, (5.15) and Remark 5.1 applied once to f_h we obtain

$$\begin{aligned} \|hA(g_h)\| &\leq c_5 \int p_s(x+hu) K(u) \{ P f_h^2(x+hu) \}^{1/2} \, du \\ &\leq c_6 \left\{ \int K(u) P f_h(x+hu) \, du \right\}^{1/2}. \end{aligned} \tag{5.16}$$

Let $\delta = \lim_{h\downarrow 0} f_h$ and $\eta_h = f_h - \delta$. Then (cf. Remark 5.1) the functions δ and η_h are bounded and

$$\lim_{h\downarrow 0} \eta_h \downarrow 0, \qquad \delta = G_{s,\nu} 1_{\{x\}} \leq G 1_{\{x\}}, \tag{5.17}$$

where $G = \sum_{n=0}^{\infty} P^n$. By (5.16) and B_2 we get

$$\|hA(g_h)\|^2 \leq c_7 \int_{\mathscr{N}_0} \{ P\eta_h(x+hu) + PG(x+hu, \{x\}) \} \, du,$$

and we need to show that the integral in the above expression is $o(1)$. We first prove

$$\int_{\mathscr{N}_0} PG(x+hu, \{x\}) \, du = 0, \tag{5.18}$$

which follows from B_3 since $\pi_s(\{x\}) = 0$ and for fixed n,

$$\begin{aligned}\pi_s(\{x\}) &\geq \int_{\mathcal{N}_x(h)} \pi_s(dz)P^n(z, \{x\}) \\ &= \int_{\mathcal{N}_x(h)} p_s(z)P^n(z, \{x\})\,dz \\ &= \int_{\mathcal{N}_0} p_s(x+hu)P^n(x+hu, \{x\})h\,du \\ &\geq h \inf_{y\in\mathcal{N}_x(h)} p_s(y) \int_{\mathcal{N}_0} P^n(x+hu, \{x\})\,du.\end{aligned}$$

Let $\varepsilon > 0$ and $\mathscr{A}_h = \{\eta_h > \varepsilon\}$. Then by B_4, $\limsup_{h\downarrow 0}\sup_{y\in\mathcal{N}_x(h)} P(y, \mathscr{A}_h) = 0$, which gives $\|hA(g_h)\|^2 \leq c_8\varepsilon + o(1)$ where c_8 is independent of ε. Hence $\|hA(g_h)\| = o(1)$.

If $\xi_0(u, w) = \xi(w) - P\xi(u)$, then $\tilde{\xi}_0 \equiv 0$, $\widetilde{I}_{\xi_0\otimes\xi_0}1(x) = V\xi(x)$ and $\tilde{\xi}_0(x) = 0$. Thus (5.8) is true. □

In the second lemma we obtain bounds of higher order moments of the type required in A_3.

LEMMA 5.2. *Let g_h be given by* (5.4) *with $g_h \in \mathscr{G}_2$. Assume that* B_2 *holds, and that $I_{\xi_0}1$ is locally continuous and that the invariant measure π_s has a locally bounded density p_s. Let $m \geq 1$ and assume that $\widetilde{I}_{|\xi_0|^{2m}}1$ is locally bounded. Then corresponding to* A_3 *in Section* 4,

$$\mathbf{E}U^{2m}(|g_h|) \leq d_m h^{-2m+1}$$

for some $d_m > 0$. Moreover, $h\bar{\sigma}^2(g_h) = O(1)$.

PROOF. From definitions we have

$$\begin{aligned}\mathbf{E}U^{2m}(|g_h|) &= \mathbf{E}_\nu\left\{\sum_{j=0}^{\tau} |g_h|(X_j, X_{j+1})\right\}^{2m} \\ &= \mathbf{E}_\nu\left\{\sum_{j=0}^{\infty}\left\{\prod_{i=0}^{j-1} 1(Y_i = 0)\right\}|g_h|(X_j, X_{j+1})\right\}^{2m} \\ &= \mathbf{E}_\nu\left\{\sum_{j=0}^{\infty} B_j|g_h|(X_j, X_{j+1})\right\}^{2m}\end{aligned}$$

with $B_j = \prod_{i=0}^{j-1} 1(Y_i = 0)$. As in the proof of Lemma 3.3, let $\Delta_{2m,k} = \{l = (l_1, \dots, l_k) \in \mathcal{N}_+^k\colon \sum l_i = 2m\}$ for $k \geq 1$. By B_2 and since $\tilde{\xi}_0$ is locally bounded we have that $\mathbf{E}U(|g_h|) = \pi_s K_{x,h}|\tilde{\xi}_0|$ is finite, and consequently,

$$\sum_{j=0}^{\infty} B_j|g_h|(X_j, X_{j+1}) < \infty \qquad \text{a.s.}$$

Hence we can use the technique of Lemma 3.3 with $n = \infty$ to obtain

$$\mathbf{E}U^{2m}(|g_h|) = \sum_{k=1}^{2m} \sum_{l \in \Delta_{2m,k}} \frac{(2m)!}{l_1! \cdots l_k!} E_\nu J_{k,l}, \tag{5.19}$$

where, using the compressed notation $j_s \overset{\text{def}}{=} i_1 + \cdots + i_s$, $s = 1, \ldots, k$,

$$J_{k,l} = \sum_{i_1=0}^{\infty} \sum_{i_2=1}^{\infty} \cdots \sum_{i_k=1}^{\infty} B_{j_1} \cdots B_{j_k} |g_h|^{l_1}(X_{j_1}, X_{j_1+1}) \cdots |g_h|^{l_k}(X_{j_k}, X_{j_k+1}).$$

A tedious calculation, the details of which can be found in Karlsen and Tjøstheim [(1998), pages 52 and 53], shows that

$$\mathbf{E}_\nu J_{k,l} = \nu G_{s,\nu} \check{I}_{|g_h|^{l_1}} G_{s,\nu} \check{I}_{|g_h|^{l_2}} \cdots G_{s,\nu} \check{I}_{|g_h|^{l_{k-1}}} G_{s,\nu} \widetilde{I}_{|g_h|^{l_k}} 1,$$

where the kernel $\check{I}$ is defined by

$$\check{I}_f(x, dy) = (P - s \otimes \nu)(x, dy) f(x, y).$$

Because of (5.19) and the fact that $\check{I}_f \leq \widetilde{I}_f$ when $f \geq 0$, it suffices to consider the expression

$$\nu G_{s,\nu} \widetilde{I}_{|g_h|^{l_1}} G_{s,\nu} \widetilde{I}_{|g_h|^{l_2}} \cdots G_{s,\nu} \widetilde{I}_{|g_h|^{l_{k-1}}} G_{s,\nu} \widetilde{I}_{|g_h|^{l_k}} 1,$$

where $k \leq 2m$, $l_i \geq 1$ and $\sum l_i = 2m$. Now we use the simple inequality $|\xi_0|^{l_i} \leq 1 + |\xi_0|^{2m}$, when $l_i \leq 2m$, B_2 and the local boundedness of $\widetilde{I}_{|\xi_0|^{2m}} 1$ to obtain

$$\widetilde{I}_{|g_h|^{l_i}} 1 \leq K_{x,h}^{l_i}(1 + \widetilde{I}_{|\xi_0|^{2m}} 1) \leq c_1 K_{x,h}^{l_i} \leq c_2 h^{-l_i} 1_{\mathcal{N}_x}.$$

By Remark 5.1 $G_{s,\nu} 1_{\mathcal{N}_x}$ is bounded, so that $G_{s,\nu} \widetilde{I}_{|g_h|^{l_i}} \leq c_3 h^{-l_i} 1$. Hence, using the above inequalities successively for $i = k, k-1, \ldots, 2$, it follows from B_2 and the fact that $\nu G_{s,\nu} = \pi_s$,

$$\mathbf{E}_\nu J_{k,l} \leq c_4 h^{-\sum_{i=2}^{k} l_i} \pi_s K_{x,h}^{l_1} \leq c_5 h^{-2m+v},$$

where $v = 1$, since π_s has a locally bounded density p_s and $l_1 \geq 1$. The proof of the statement concerning $h\bar{\sigma}^2(|g_h|)$ is similar, albeit simpler. □

We are now in a position to state and prove the main results in this section. The first two theorems give conditions for consistency, and the last two theorems deal with weak convergence.

THEOREM 5.1. *Assume that the tail condition* (3.16) *and* B_1–B_3 *hold. Let* $\varepsilon > 0$ *and assume that* $h_n^{-1} \ll n^{\beta/2-\varepsilon}$ *or* $h_n = q_{T_C(n)}$, *where* $q_n^{-1} \ll n^{1/2-\varepsilon}$. *Then* $\hat{p}_C$ *defined in* (5.1) *is a strongly consistent estimator of* p_C *at the point* x.

PROOF. We have $\hat{p}_C(x) = T_C^{-1}(n)S_n(g_h)$ with $g_h = K_{x,h}$. From Lemma 5.1, $\mu(g_h) = p_s(x) + o(1)$. The case where $h_n^{-1} \ll n^{\beta/2-\varepsilon}$ will be treated first. The stochastic bandwidth case then follows easily as will be seen at the end of the proof.

Let $\{h_n\}$ be fixed. Without loss of generality we may assume $h_n = q_{u_0(n)}$, where $u_0(n) = [n^{\beta-\varepsilon}]$, and where $q_n^{-1} \ll n^{1/2-\varepsilon}$. We use the representation

$$S_n(g_h) = U_0(g_h) + \sum_{k=1}^{T(n)} U_k(g_h) + U_{(n)}(g_h)$$

for each n. By Lemma 3.4, $T(n) \gg u_0(n)$ a.s. Let $t = T(n)$, $l = u_0(n)$, $p = [t/l]$, $r = t - lp + 1$ so that $t = lp + r - 1$, $1 \le r \le l$. Then

$$\left|T^{-1}(n)\sum_{k=1}^{T(n)} U_k(g_{h_n}) - \mu(g_{h_n})\right| = \left|\frac{1}{lp+r-1}\sum_{k=1}^{lp+r-1} U_k(g_{q_l}) - \mu(g_{q_l})\right|$$
$$\stackrel{\text{def}}{=} |S_{l,p,r} - \mu(g_{q_l})|.$$

By the Borel–Cantelli lemma we have $|T^{-1}(n)\sum_{k=1}^{T(n)} U_k(g_{h_n}) - \mu(g_{h_n})| \to 0$ a.s. if we can show that for some m,

$$\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\sum_{r=1}^{l} \mathbf{E}\{S_{l,p,r} - \mu(g_{q_l})\}^{2m} < \infty. \tag{5.20}$$

By Lemma 5.2 and the independence properties of $\{U_k(g_{q_l}), k \ge 1\}$,

$$\begin{aligned}\mathbf{E}\{S_{l,p,r} - \mu(g_{q_l})\}^{2m} &\le c_1(lp+r-1)^{-m}\mathbf{E}\{U_k(g_{q_l}) - \mu(g_{q_l})\}^{2m}\\ &\le c_2(lp)^{-m} q_l^{-(2m-1)} \le p^{-m} l^{-\{m-1-(\frac{1}{2}-\varepsilon)(2m-1)\}} l^{-1}\end{aligned}$$

and the assertion in (5.20) follows because $m - 1 - (\frac{1}{2} - \varepsilon)(2m-1) = \varepsilon(2m-1) + \frac{1}{2} - 1 > 1$ for m large enough. Since $T(n) \to \infty$ a.s. as $n \to \infty$, and $T_C(n)/T(n) \to \pi_s 1_C$ a.s., it follows from the definition of $p_C(x)$ that

$$T_C^{-1}(n)\sum_{k=1}^{T(n)} U_k(g_{h_n}) \to p_C(x) \quad \text{a.s.}$$

It remains to consider the edge terms $U_0(g_h)$ and $U_{(n)}(g_h)$ of $S_n(g_h)$. First, note that the above arguments also show that $T^{-1}(n)\sum_{k=1}^{T(n)+1} U_k(g_{h_n}) \to p_s(x)$ a.s. Hence we can neglect $U_{(n)}(g_{h_n})$. Since $g_h(u,w) = K_{x,h}(u)$, it follows from B_2 that $g_h(u,w) \le h^{-1}g_0(u)$ with $g_0(u) \stackrel{\text{def}}{=} \{\sup K\}1_{\mathcal{N}_x}(u)$. Then

$$T^{-1}(n)U_0(g_{h_n}) \le h_n^{-1}T^{-1}(n)U(g_0).$$

But $U(g_0) = \sum_{j=0}^{\tau} g_0(X_j)$ and $\mathbf{E}_y\{U(g_0)\} = G_{s,\nu}g_0(y)$. It follows (cf. Remark 5.1) that $G_{s,\nu}g_0$ is bounded and hence $\mathbf{P}_y(U(g_0) < \infty) \equiv 1$, and therefore $\{h_n\, T(n)\}^{-1}U(g_0) = o(1)$ a.s., and the first part of the proof is complete. If $h_n = q_{T_C(n)}$ with $q_n^{-1} \ll n^{1/2-\varepsilon}$, and neglecting the edge terms as above, it is enough to show

$$T_C^{-1}(n) \sum_{k=1}^{T_C(n)} U_k(g_{q_{T_C(n)}}) \to p_C(x) \quad \text{a.s.}$$

But since $T_C(n) \to \infty$ a.s., this convergence is implied by $n^{-1} \sum_{k=1}^{n} U_k(g_{q_n}) \to \infty$ a.s. $p_C(x)$, and the latter is proved by the Borel–Cantelli lemma in complete analogy with the first part of the proof. □

REMARK 5.2. It will be noted that only the lower bound of Lemma 3.4 is required, so that in fact only the weakened form of the tail condition stated in that lemma is needed. This is the case for the next theorem as well, where we prove strong consistency of the conditional mean estimator $\widehat{P}\xi(x)$ defined in (5.2).

THEOREM 5.2. *Assume that the weakened version* (3.34) *of the tail condition* (3.16) *and* B_1–B_3 *hold. In addition assume that* $P\xi$ *and* $P\|\xi\|$ *are locally continuous,* $P\|\xi\|^{2m}$ *is locally bounded for some integer* m, $h_n^{-1} \ll n^{\beta\delta_m-\varepsilon}$ *with* $\delta_m = (m-2)/(2m-1)$ *for some* $\varepsilon > 0$, *or* $h_n^{-1} = q_{T_c(n)}$ *with* $q_n^{-1} \ll n^{\delta_m-\varepsilon}$. *Then* $\widehat{P}\xi$ *is a strongly consistent estimator of* $P\xi$ *at the point* x.

PROOF. First note that $\delta_m = (m-2)/(2m-1) \le \frac{1}{2}$. Hence, $h_n^{-1} \ll n^{\beta\delta_m-\varepsilon}$ implies $h_n^{-1} \ll n^{\beta/2-\varepsilon}$, and it follows from Theorem 5.1 that the denominator of

$$\widehat{P}\xi(x) = \left(T_C^{-1}(n) \sum_{t=0}^{n} K_{x,h}(X_t)\right)^{-1} \left(T_C^{-1}(n) \sum_{t=0}^{n} \xi(X_{t+1}) K_{x,h}(X_t)\right) \tag{5.21}$$

converges almost surely to $p_C(x)$. For the numerator we can again use the proof of Theorem 5.1 (for one dependence in this case), the only change being that the convergence in Lemma 5.1 is to $p_C(x)P\xi(x)$ and that in the Borel–Cantelli argument $\sum_{l=1}^{\infty} l^{-\{m-1-(\frac{1}{2}-\varepsilon)(2m-1)\}}$ is now replaced by

$$\sum_{l=1}^{\infty} l^{-\{m-1-(\delta_m-\varepsilon)(2m-1)\}} < \infty,$$

the finiteness of the sum being true since $m-1-\delta_m(2m-1) \ge 1$ and $\varepsilon(2m-1) > 0$. □

In the expression (5.24) of the following theorem we obtain a β-null recurrent generalization of the weak convergence theorem for the kernel density estimator.

THEOREM 5.3. *Assume that the tail condition* (3.16) *holds or that the chain is positive recurrent. Let* C *be a small set. If* B_2 *and* B_3 *hold and* $h_n^{-1} \ll n^{\beta-\varepsilon}$ *for some* $\varepsilon > 0$, *then*

$$\begin{aligned} &T_C^{1/2}(n)\pi_s^{1/2}(C)\bar{\sigma}^{-1}(K_{x,h_n}) \\ &\quad \times \left\{\hat{p}_C(x) - p_C(x) - \left(\pi_s^{-1}(C)\mu(K_{x,h_n}) - p_C(x)\right)\right\} \xrightarrow[n]{d} \mathscr{N}(0,1), \end{aligned} \tag{5.22}$$

where

$$\pi_s^{-1}(C)h\bar{\sigma}^2(K_{x,h}) = \int p_C(x+hu)K^2(u)du + 2\int K(u)PG_{s,\nu}K_{x,h}(x+hu) + o(1). \tag{5.23}$$

If in addition B_4 *holds, then*

$$T_C^{1/2}(n)h_n^{1/2}\Big\{\hat{p}_C(x) - p_C(x) - \Big(\pi_s^{-1}(C)\mu(K_{x,h_n}) - p_C(x)\Big)\Big\} \xrightarrow[n]{d} \mathcal{N}\Big(0,\, p_C(x)\int K^2(u)du\Big). \tag{5.24}$$

Moreover, if p_C *has a locally continuous second-order derivative and if* B_1 *holds, then the bias term* $\pi_s^{-1}(C)\mu(K_{x,h_n}) - p_C(x)$ *is negligible if* $h_n^{-1} \gg n^{\beta/5+\varepsilon}$.

PROOF. We first prove (5.22) by using Theorem 4.2, and then establish the more explicit expression (5.24).

We have to verify the conditions of Theorem 4.2 with $g_h = K_{x,h}$. The main tools for doing this are Lemmas 5.2 and 5.1. In Lemma 5.2 the local boundedness assumptions for p_s and ξ_0 follow from B_3 and the fact that $\xi_0 \equiv 1$. Moreover, since $g_h \geq 0$, conditions A_0 and A_3 (with $m = 2$ and $v = 1$) of Theorem 4.2 hold. Condition A_1 follows from the fact that $|g_h| = g_h$ and that due to independence of $\{U_k(g_h), k \geq 1\}$ in the present case $\bar{\sigma}(g_h) = \sigma(g_h)$.

Since by definition $A(g_h) \geq 0$ in the formulation of Lemma 5.1, it follows from that lemma that lim inf $h\bar{\sigma}^2(g_h) > 0$, and hence A_2(ii) and A_2(iii) hold. The strengthening of A_2(i) required in Theorem 4.2 follows from

$$\mu(|g_h|) = \mu(g_h) = \pi_s(K_{x,h}) \leq \sup_{u \in \mathcal{N}_x} |p_s(u)| \int K(u)\,du < \infty.$$

Furthermore, condition A_4 ($v = 1$) follows directly from our assumption $h_n^{-1} \ll n^{\beta-\varepsilon}$ on the bandwidth. It remains to prove A_5 concerning the edge term $U_0(|g_h|)$. We use the same reasoning as in the proof of Theorem 5.1; that is, it follows from B_2 that $g_h \leq h^{-1}g_0$ with $g_0 \stackrel{\text{def}}{=} \{\sup K\}1_{\mathcal{N}_x}(u)$ and $\mathbf{P}_y(U(g) < \infty) \equiv 1$, and A_5 now easily follows from this and A_4 with $v = 1$.

The conditions of Theorem 4.2 are then fulfilled and (5.22) is true.

The expression for $\bar{\sigma}^2(K_{x,h})$ essentially is a consequence of a simplification of (5.11) in the proof of Lemma 5.1.

If also B_4 holds, by Lemma 5.1,

$$\mu(g_h) = p_s(x) + o(1) \quad \text{and} \quad h\bar{\sigma}^2(g_h) = p_s(x)\int K^2(u)\,du + o(1).$$

This implies (5.24). It remains to consider the bias term. By B_1 and the local continuity of the second-order derivative of p_C,

$$\begin{aligned}\pi_s^{-1}(C)\mu(K_{x,h}) - p_C(x) &= \int\{p_C(x+hu) - p_C(x)\}K(u)\,du \\ &= \left\{\frac{d}{dx}p_C(x)h\right\}\int uK(u)\,du + O(h^2) \\ &= O(h^2),\end{aligned}$$

and the desired conclusion is a direct consequence of Theorem 4.2. □

A weak limit theorem for the conditional mean estimator of (5.2) can be obtained along the same lines, but requires some more work.

THEOREM 5.4. *Assume that the tail condition* (3.16) *holds or that the chain is positive recurrent. Moreover, assume that* B_1–B_3 *hold, that the conditional mean* $P\xi$ *is locally continuous, the conditional variance* $V\xi$ *defined by* (5.3) *is locally continuous and positive and* $P\|\xi\|^{2m}$ *is locally bounded for some* $m \geq 2$. *If* $h_n^{-1} \ll n^{\beta-\varepsilon}$ *for some* $\varepsilon > 0$, *then*

$$\left\{h_n \sum_{t=0}^{n} K_{x,h_n}(X_t)\right\}^{1/2}\left\{\widehat{P}\xi(x) - P\xi(x) - a_{h_n}\right\} \xrightarrow[n]{d} \mathscr{N}\left(0, V\xi(x)\int K^2(u)du\right), \tag{5.25}$$

where $a_h \stackrel{\text{def}}{=} \pi_s(\psi_x \cdot K_{x,h})/\pi_s K_{x,h}$, *with* $\psi_x(y) \stackrel{\text{def}}{=} P\xi(y) - P\xi(x)$, *is the bias term. If* p_s *and* $P\xi$ *have locally continuous second-order derivatives, then the bias term is negligible if* $h_n^{-1} \gg n^{\beta/5+\varepsilon}$.

PROOF. We start by deriving an expression for $\widehat{P}\xi(x) - P\xi(x)$. Note that

$$\begin{aligned}\xi(X_{t+1}) &= \{\xi(X_{t+1}) - P\xi(X_t)\} + \{P\xi(X_t) - P\xi(x)\} + P\xi(x), \\ \xi(X_{t+1})K_{x,h}(X_t) &= g_h(X_t, X_{t+1}) + \psi_x(X_t)K_{x,h}(X_t) + P\xi(x)K_{x,h}(X_t),\end{aligned}$$

where $g_h(u, w) = (\xi(w) - P\xi(u))K_{x,h}(u)$. This gives, by (5.2),

$$\begin{aligned}\widehat{P}\xi(x) - P\xi(x) - a_h &= S_n^{-1}(K_{x,h})\{S_n(g_h) + S_n(\psi_x \cdot K_{x,h})\} - a_h \\ &= S_n^{-1}(K_{x,h})\{S_n(g_h) + S_n(f_h)\},\end{aligned}$$

where $f_h = (\psi_x - a_h)\cdot K_{x,h}$. We note that $\mu(g_h) = \mu(f_h) = 0$. Denote the left-hand side of (5.25) by Δ_{n,h_n}. Then

$$\Delta_{n,h} = \Delta_{n,h}^1 + \Delta_{n,h}^2, \tag{5.26}$$

where

$$\begin{aligned}\Delta_{n,h}^1 &= \{\hat{p}_C(x)\}^{-1/2}T_C^{-1/2}(n)h^{1/2}S_n(g_h), \\ \Delta_{n,h}^2 &= \{\hat{p}_C(x)\}^{-1/2}T_C^{-1/2}(n)h^{1/2}S_n(f_h),\end{aligned} \tag{5.27}$$

and where C is a purely auxiliary small set. The conditions are fulfilled for the second part of Theorem 5.3, and thus (5.22) and (5.23) imply that $\hat{p}_C(x) \to p_C(x)$ in probability. The next step consists in showing that Δ^2_{n,h_n} in (5.26) can be neglected in the sense of Definition 3.3. We first establish that $\limsup_{h\downarrow 0} h\mathbf{E}U^2(\|f_h\|) = 0$, and then we use this to prove negligibility.

Since $P\xi$ is locally continuous, it follows that $a_h = o(1)$ and

$$\sup_{y\in\mathscr{N}_x(h)} \|P\xi(y) - P\xi(x) - a_h\| = o(1). \tag{5.28}$$

Moreover, f_h and $\widetilde{I}_{f_h\otimes f_h}$ are locally continuous and small functions by B_2(ii), and $\pi_s f_h = 0$. By a simplified version of (5.9) [$g(x, y) = g(x)$, which implies $g_\nu = g$] in the proof of Lemma 5.1,

$$\mathbf{E}U^2(\|f_h\|) \le \pi_s\|f_h\|^2 + 2\pi_s I_{\|f_h\|} PG_{s,\nu}\|f_h\| + 2\mu^2(\|f_h\|)$$

and by local continuity of $P\xi$, B_2(ii), B_3 and the Lebesgue dominated convergence theorem,

$$\mu(\|f_h\|) = \int p_s(x+hu)\|P\xi(x+hu) - P\xi(x)\|K(u)\,du + o(1) = o(1).$$

By (5.28) and B_2(ii),

$$\sup_z G_{s,\nu} h\|f_h\|(z) \le \sup_z G_{s,\nu}(z, \mathscr{N}_x(h)) \le c_2,$$

and it follows that

$$\begin{aligned} &h\pi_s I_{\|f_h\|} PG_{s,\nu}\|f_h\| \le c_3\mu(\|f_h\|) = o(1), \\ &\qquad h\pi_s\|f_h\|^2 \le c_4\{\sup_z h\|f_h(z)\|\}\pi_s\|f_h\| \le c_5\mu(\|f_h\|) = o(1). \end{aligned} \tag{5.29}$$

Hence we have that $\limsup_{h\downarrow 0} \mathbf{E}\,U^2(\|f_h\|) = 0$.

To show that Δ^2_{n,h_n} is negligible it is sufficient to show that $\Delta^3_{n,h_n} = T_C^{1/2}(n)$ $h_n^{1/2} S_n(f_{h_n})$ is negligible. Consider therefore $\widetilde{\Delta}^3_{n,h} = u^{-1/2}(n)h^{1/2}S_{[nt]}(f_h)$, where u is defined in (3.27). Since $h_n^{-1} \ll n^{\beta-\varepsilon}$, we have $h_n^{-1} = o(u(n))$.

Define $q_n = h_{u^{-1}(n)}$ so that $h_n = q_{u(n)}$. Neglecting for the moment the edge terms of $\widetilde{\Delta}^3_{n,h_n}$, and reasoning as in Part 1 of the proof of Theorem 4.1, analogously to (4.8) we consider

$$\widetilde{Q}^3_n(t) = \{nq_n^{-1}\}^{-1/2} \sum_{k=1}^{[nt]} U_k(f_{q_n}).$$

It is enough to prove that each of the components of $\widetilde{Q}^3_n$ is negligible. Hence, without loss of generality, we can assume that f_h is real valued. Then

$$\mathbf{E}\Big\{\widetilde{Q}^3_n(t)\Big\}^2 \le c_6 n^{-1} q_n [nt] \mathbf{E}U^2(\|f_{q_n}\|) \le c_7 t q_n \mathbf{E}U^2(\|f_{q_n}\|)$$

and hence, since $\limsup_{h\downarrow 0} h\mathbf{E}U^2(\|f_h\|) = 0$, we have for any finite M, $\sup_{t\le M}$ $\mathbf{E}\{\widetilde{Q}^3_n(t)\}^2 = o(1)$.

Since $\{U_k, k \geq 1\}$ is a sequence of independent random variables, we can conclude [cf. Pollard (1984), page 104] that $\{\widetilde{Q}_n^3\}$ is negligible.

The edge effects are tackled in the same way as in Part 2 of the proof of Theorem 4.1. Hence it is sufficient to look at

$$\widetilde{Z}_{n,h}^3(t) = u^{-1/2}(n) \sum_{k=1}^{T([nt])} h^{1/2} U_k(f_h)$$

since $\widetilde{\Delta}_{n,h_n}^3(t)$ is equivalent to $\widetilde{Z}_{n,h_n}^3(t)$. But

$$\widetilde{Z}_{n,h_n}^3(t) = \widetilde{Q}_{u(n)}^3(T_n(t))$$

and $\widetilde{Q}_{u(n)}^3 \circ T_n$ is easily shown to be negligible from the negligibility of $\{\widetilde{Q}_n^3\}$.

It follows from Lemmas 3.4 and 3.5 that

$$T_C^{-1/2}(n) h_n^{1/2} S_n(f_{h_n}) = o_P(1). \tag{5.30}$$

Hence by (5.26), (5.27) and (5.30),

$$\Delta_{n,h_n} = \Delta_{n,h_n}^1 + o_P(1).$$

The next part of the proof consists in verifying that the conditions of Corollary 4.2 are fulfilled for Δ_{n,h_n}^1. It is patterned after the proof of Theorem 5.3, and Lemmas 5.1 and 5.2 are the main tools.

In the present case we have that $g_h(u, w) = \xi_0(u, w) K_{x,h}(u)$ with $\xi_0(u, w) = \xi(w) - P\xi(u)$. By Jensen's inequality $\|P\xi\|^{2m} \leq P\|\xi\|^{2m}$, which gives that

$$\begin{aligned}
\widetilde{I}_{\|\xi_0\|^{2m}} 1(x) &= \int P(x, dy) \|\xi_0(x, y)\|^{2m} \\
&\leq \int P(x, dy) \Big\{ \|\xi(y)\| + \|P\xi(x)\| \Big\}^{2m} \\
&\leq 2^{2m-1} \Big\{ P\|\xi\|^{2m}(x) + \|P\xi\|^{2m}(x) \Big\} \\
&\leq 2^{2m} P\|\xi\|^{2m}(x).
\end{aligned}$$

Hence the conditions of Lemma 5.2 are satisfied since the right-hand side of the above inequality is by assumption locally bounded. Thus $\|g_h\|$ satisfies A_0 and A_3 ($m = 2, v = 1$). Moreover, $h^{1/2}\mu(\|g_h\|) = o(1)$, so that $\|g_h\|$ satisfies the strengthening of A_2(i) required in Corollary 4.2.

Likewise, since $P\xi$, $P(\xi \otimes \xi)$ are continuous at the point x and $\tilde{\xi}_0 = P\xi - P\xi \equiv 0$, the conditions in Lemma 5.1 are fulfilled. In particular (5.8) holds and

$$\overline{\Sigma}(g_h) = p_C(x) V\xi(x) \int K^2(u)\, du + o(1). \tag{5.31}$$

Since $A(\|g_h\|) \geq 0$ in the formulation of Lemma 5.1, we have that A_2(ii) and A_2(iii) hold for $\|g_h\|$, and by (5.31) it follows that (4.16) and (4.17) are fulfilled with $\Sigma_0 = p_C(x) V\xi(x) \int K^2(u)\, du$. The remaining conditions A_1, A_4 and A_5 of Corollary 4.2 are verified in exactly the same manner as in the proof of Theorem 5.3, and (5.25) follows.

The proof is concluded by considering the bias term. This term is negligible if

$$\left\{\sum_{t=0}^{n} K_{x,h_n}(X_t)\right\}^{1/2} h_n^{1/2}\left|\frac{\pi_s\psi_x\cdot K_{x,h_n}}{\pi_s K_{x,h_n}}\right| = o_P(1).$$

Introducing the set C as in (5.27); by B_3 this is equivalent with

$$T_C^{1/2}(n)h_n^{1/2}\left|\frac{\pi_s\psi_x\cdot K_{x,h_n}}{\pi_s K_{x,h_n}}\right| = o_P(1).$$

Assume that $h_n^{-1} \gg n^{\beta/5+\varepsilon}$. Then from Lemma 3.4 and Lemma 3.5, $h_n^5 T_C(n) = o_P(1)$ a.s. Hence it is enough to verify that

$$h^{-2}\left|\frac{\pi_s\psi_x\cdot K_{x,h}}{\pi_s K_{x,h}}\right| = O(1).$$

Assume without loss of generality that $d = 1$. Then by a Taylor expansion we can write

$$p_s(x+hu) = p_s(x) + hR_1(x, hu, h), \qquad \sup_{h\le 1}\sup_{y\in\mathcal{N}_0}|R_1(x,y,h)| < \infty$$

and

$$P\xi(x+hu) = P\xi(x) + h\frac{d}{dx}P\xi(x)u + h^2R_2(x,hu,h),$$
$$\sup_{h\le 1}\sup_{y\in\mathcal{N}_0}|R_2(x,y,h)| < \infty.$$

This gives

$$\begin{aligned}
\pi_s\psi_x K_{x,h} &= \int \pi_s(dy)(P\xi(y) - P\xi(x))K_{x,h}(y)\\
&= \int p_s(x+hu)(P\xi(x+hu) - P\xi(x))K(u)\,du\\
&= p_s(x)\int (P\xi(x+hu) - P\xi(x))K(u)\,du\\
&\quad + \int huR_1(x,hu,h)(P\xi(x+hu) - P\xi(x))K(u)\,du\\
&= p_s(x)h\frac{d}{dx}P\xi(x)\int uK(u)\,du + O(h^2)\\
&\quad + h^2\frac{d}{dx}P\xi(x)\int u^2R_1(x,hu,h)K(u)\,du + O(h^3)\\
&= O(h^2).
\end{aligned}$$

Hence the theorem is proved. □

6. An example. An important task is to find good examples of β-null recurrent processes. Some such examples are given in Myklebust, Karlsen and Tjøstheim (2001). But we hasten to add that this is really an open problem, and it is difficult to apply the drift criterion supplied in the book of Meyn and

Tweedie (1994) since we need to verify the tail condition (3.16) in addition. The present paper is directed towards deriving asymptotic theory, and we will therefore satisfy ourselves by providing a very simple example, that of a random walk $\{X_t, t \geq 0\}$ where $X_0 = x_0$ and for $t \geq 1$,

$$X_t = X_{t-1} + e_t, \tag{6.1}$$

where $\{e_t, t \geq 0\}$ is a series of zero-mean independent identically distributed random variables. If $\{e_t\}$ is Gaussian, say, it is well known that $\{X_t\}$ is what we have termed β-null recurrent with $\beta = 1/2$ and with Lebesgue measure as an invariant measure. It should be noted that this instantly gives rise to a whole class of β-null recurrent processes, since if $\{X_t\}$ is β-null recurrent, and if f is a one-to-one function, then $Y_t = f(X_t)$ is β-null recurrent with the same β.

We will consider the problem of estimating the conditional quantity $P\xi(x)$ with $\xi(X_t) = X_t$, which means $P\xi(x) = M(x) = \mathrm{E}(X_{t+1}|X_t = x) = x$ in the case of (6.1). We will examine the finite sample properties of the estimator (5.2). A difficult and largely unresolved problem is that of choosing a proper bandwidth. Theorem 5.4 only gives the allowable rate as n tends to infinity, and these rates are based on the asymptotic situation, where, in the random walk case, the observations are evenly scattered over the real line according to Lebesgue measure. In practice, the data are unevenly distributed, and we have found it useful to employ cross-validation, and to let the bandwidth depend on x. In fact we typically let h_n be proportional to $T_C(n)\hat{p}_C(x)^{1/5}$, where $\hat{p}_C(x)$ could be thought of as a locally estimated density according to Theorem 5.3.

We look briefly at two aspects of the estimation problem. In Figure 1 is shown an estimate of $M(x)$ with the corresponding scatter diagram for $n = 500$ observations and with $\{e_t\}$ being standard normal. It is seen that $M(x)$ is well estimated. Only a single realization is shown, but it is representative for the quality of the estimate. In Figure 2 we turn to the finite sample approximation of the asymptotic standard normal distribution for the estimator

$$\left\{\int K^2(u)\,du\right\}^{-1/2}\left\{h_n \sum_{t=0}^{n} K_{x,h_n}(X_t)\right\}^{1/2}\{\widehat{M}(x) - x\} \tag{6.2}$$

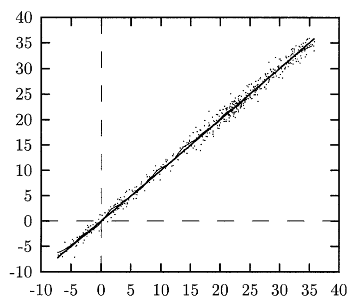

FIG. 1. *Estimated conditional mean $\hat{M}(x)$.*

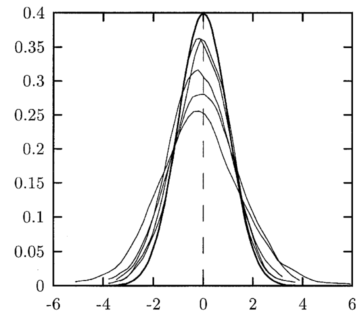

FIG. 2. *Finite sample approximation.*

related to (5.25), using that $V\xi(x) = 1$ and assuming that the bias can be neglected.

The quality of the sample approximation has to be judged using a multitude of realizations. A problem not encountered in the stationary case is that the simulated realizations may cover very different x-regions. (This is one reason for only using one realization in Figure 1.) Hence, for a fixed $x = x'$, close to the starting value $X_0 = 0$, say, of each realization, some realizations may have many observations in the neighborhood of x', whereas other realizations may have a few, or even no observations in the vicinity of x' for the sample size we are considering. This kind of behavior does not occur in the stationary case, where the expected time until the process reaches x' is always finite and in practice small, when $|x'|$ is small. This means that in the verification of Theorem 5.4 we can either keep x fixed and wait until we have sufficiently many observations close to x, the other realizations being discarded, or we can choose a central realization-dependent value, for example the modal value of the sample, for studying the normalized ratio (6.2).

In Figure 2 is shown the approximation to normality as a function of sample size for the first procedure at the point $x = 7.5$. We have used 1000 realizations, and a particular realization is admitted into the evaluation as, respectively, 100, 200, 300, 500 and 800 observations are accumulated in the interval (5,10). Considering the modest sample sizes in the "relevant" region surrounding x and the approximations involved, the convergence towards normality is quite satisfactory.

A much more extensive set of simulation experiments in the transfer function case is carried out in Karlsen, Myklebust and Tjøstheim (2001). These experiments also involve larger sample sizes and the second procedure mentioned above.

APPENDIX

We denote by $D^d[0, \infty)$, the space of all R^d-valued functions defined on R_+ which have left limits and are right continuous (cadlag). We refer to Jacod and

Shiryaev [(1987), pages 288–322] for a complete description of this space, and the concept of weak convergence of stochastic processes with sample paths therein.

If $\{\xi_n\}$ is a sequence of stochastic processes with values in $D^d[0,\infty)$ then $\xi_n \xrightarrow{\mathscr{D}} \xi$ if the corresponding sequence of induced measures converges weakly to P_ξ, where $P_\xi(\cdot) = P(\xi \in \cdot)$. We have finite-dimensional convergence if for each finite set $F \subset [0,\infty)$, the vector sequence $\{\xi_n(t), t \in F\}$ converges in distribution to the vector $\{\xi(t), t \in F\}$. The sequence $\{\xi_n\}$ is C-tight if $\{\xi_n\}$ is tight and all limits points of $\{P_n\}$ charges only $C^d[0,\infty)$, the space of all continuous R^d-valued functions, that is, $P_n(C^d[0,\infty)) = 1$ for all n. In particular, if ξ_n converges weakly to a ξ, which has continuous sample paths, then $\{\xi_n\}$ is C-tight.

The inverse of a function f is denoted by $f^{(-1)}$. For $x \in D[0,\infty)$ and x increasing, we define $x^{(-1)}(t) = \inf\{s\colon x(s) > t\}$. If x is strictly increasing, then $x^{(-1)}$ is continuous and nondecreasing.

The following theorem is essentially due to Kashara (1984). A key factor in this result is the fact that a Brownian motion B and an increasing process A are independent.

THEOREM A.1. *For each n let (B_n, A_n) be a pair of stochastic processes which are cadlag, where A_n is nonnegative and nondecreasing. Let B denote a Brownian motion defined for $t \in R_+$ and let A denote a strictly increasing nonnegative process with independent increments,* $A(0) \equiv 0$ *and with no fixed jumps. Assume that $B_n \xrightarrow{\mathscr{D}} B$ and $A_n \xrightarrow{\mathscr{D}} A$. Then*

$$(B_n, A_n, A_n^{(-1)}) \xrightarrow[n]{\mathscr{D}^3} (B, A, A^{(-1)}), \tag{A.1}$$

where B is independent of $(A, A^{(-1)})$ and

$$\left(A_n^{(-1)}, B_n \circ A_n^{(-1)}\right) \xrightarrow[n]{\mathscr{D}^2} \left(A^{(-1)}, B \circ A^{(-1)}\right). \tag{A.2}$$

For all $\varepsilon > 0$,

$$\left(A_n^{(-1)}, \frac{B_n \circ A_n^{(-1)}}{\sqrt{A_n^{(-1)}}} \psi_\varepsilon\big(A_n^{(-1)}\big)\right) \xrightarrow[n]{\mathscr{D}^2} \left(A^{-1}, \frac{B \circ A^{(-1)}}{\sqrt{A^{(-1)}}} \psi_\varepsilon\big(A^{(-1)}\big)\right), \tag{A.3}$$

where $\psi_\varepsilon(x) = \varepsilon^{-1/2} x^{1/2} 1(x \le \varepsilon) + 1(x > \varepsilon)$. If we let $\psi_\varepsilon \equiv 1$ and put $0/0$ *equal to* 0, *then still finite-dimensional convergence holds. In this case we have for each fixed t that the limit vector is distributed as $(A^{(-1)}(t), Z)$ where Z is a standard normal variable independent of $A^{(-1)}(t)$.*

PROOF. By assumption $\{B_n\}$ is C-tight and $\{A_n\}$ is tight. Hence $\{(B_n, A_n)\}$ is tight [cf. Jacod and Shiryaev (1987) Corollary 3.33, page 317]. If (B', A') is a limit point for this sequence, then necessarily $B' \stackrel{d}{=} B$ and $A' \stackrel{d}{=} A$. But since A is strictly increasing, B' and A' are independent [cf. Kasahara (1984)]. Hence $(B', A') = (B, A)$.

The map given by $a \mapsto a^{(-1)}$ is continuous when $a \in \mathscr{C}_0 \stackrel{\text{def}}{=} \{x : x \text{ is strictly increasing}\}$. By the continuous mapping theorem we find that $A_n^{(-1)}$ converges weakly to $A^{(-1)}$ since $A \in \mathscr{C}_0$. Now, $\{(A_n^{(-1)}, A_n)\}$ is tight since $\{A_n^{(-1)}\}$ is C-tight and $\{A_n\}$ is tight. Again by the same argument it follows that $\{(B_n, A_n^{(-1)}, A_n)\}$ is tight which implies (A.1). Generally, the map $(b, x) \mapsto b \circ x$ is continuous at all points where b is continuous and x is nonnegative. Again, by the continuous mapping theorem we can conclude that (A.2) is true. The reasoning is similar for (A.3) where the function ψ_ε guards against a discontinuity at zero. By Jacod and Shiryaev [(1987), Proposition 3.14, page 313] we have that (A.3) implies finite-dimensional convergence when ψ_ε is not present. Let $\xi(t) = B(A^{(-1)}(t))/\sqrt{A^{(-1)}(t)}$. Since $B(s)/\sqrt{s} \sim Z$ for all $s > 0$, and since B and $A^{(-1)}$ are independent, we have that $\xi(t) \sim Z$ for all $t > 0$. $\square$

Acknowledgments. We are grateful to the editor Hans Künsch, an Associate Editor and two referees for a number of very constructive comments, which have greatly improved the presentation of our results. We are indebted to Terje Myklebust for carrying out the simulations for the example in Section 6.

REFERENCES

APARICIO, F. M. and ESCRIBANO, A. (1997). Searching for linear and nonlinear cointegration: a new approach. Working paper 97-65, Univ. Carlos III de Madrid, Statistics and Economics Series.

BERGSTRØM, H. (1981). *Weak Convergence.* Academic Press, New York.

BILLINGSLEY, P. (1968) . *Convergence of Probability.* Wiley, New York.

BINGHAM, N. H., GOLDIE, C. M. AND TEUGELS, J. L. (1989). *Regular Variation.* Cambridge Univ. Press.

CHOW, Y. S. and TEICHER, H. (1988). *Probability Theory, 2nd. ed.* Springer, New York.

DARLING, D. A. and KAC, M. (1957). *On occupation times for Markoff processes. Trans. Amer. Math. Soc.* **84** 444–458.

DICKEY, D. A. and FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with unit root. *Econometrica* **49** 1057–1072.

FEIGIN, P. D. and TWEEDIE, R. L. (1985). Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.* **6** 1–14.

FELLER, W. R. (1971). *An Introduction to Probability Theory and Its Application* 2, 2nd. ed. Wiley, New York.

GRANGER, C. W. J. (1995). Modelling nonlinear relationships between extended-memory variables. *Econometrica* **63** 265–279.

GRANGER, C. W. J. and HALLMAN, J. (1991). Long memory series with attractors. *Oxford Bull. Econom. Statist.* **53** 11–26.

HJELLVIK, V., YAO, Q. and TJØSTHEIM, D. (1998). Linearity testing using local polynomial approximation. *J. Statist. Plann. Inference* **68** 295–321.

HÖPFNER, R. (1990). Null recurrent birth and death processes: limits for certain martingales and local asymptotic mixed normality. *Scand. J. Statist.* **17** 201–215.

HÖPFNER, R. (1994). Estimating a parameter in a birth-and-death process model. *Statist. Decisions* **12** 149–160.

HÖPFNER, R., JACOD, J. and LADELLI, L. (1990). Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Theory Related Fields* **86** 105–129.

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes.* Springer, Berlin.

JOHANSEN, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models.* Oxford Univ. Press.

KALLIANPUR, G. and ROBBINS, H. (1954). The sequence of sums of independent random variables. *Duke Math. J.* **21** 285–307.

KARLSEN, H. A., MYKLEBUST, T. and TJØSTHEIM, D. (2000). Nonparametric estimation in a nonlinear cointegration type model. Sonderforschungsbereich 373 **33**. Humboldt Univ., Berlin.

KARLSEN, H. A. and TJØSTHEIM, D. (1998). Nonparametric estimation in null recurrent time series. Sonderforschungsbereich 373 **50**. Humboldt Univ., Berlin.

KASAHARA, Y. (1982). A limit theorem for slowly increasing occupation times. *Ann. Probab.* **10** 728–736.

KASAHARA, Y. (1984). Limit theorems for Lévy processes and Poisson point processes and their applications to Brownian excursions. *J. Math. Kyoto Univ.* **24** 521–538.

KASAHARA, Y. (1985). A limit theorem for sums of random number of i.i.d. random variables and its application to occupation times of Markov chains. *J. Math. Soc. Japan* **37** 197–205.

MASRY, E. and TJØSTHEIM, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* **11** 258–289.

MASRY, E. and TJØSTHEIM, D. (1997). Additive nonlinear ARX time series and projection estimators. *Econometric Theory* **13** 214–252.

MEYN, S. P. and TWEEDIE, R. L. (1994). *Markov Chains and Stochastic Stability*. Springer, London.

MYKLEBUST, T., KARLSEN, H. A. and TJØSTHEIM, D. (2001). A Markov chain characterization of unit-root processes and the problem of nonlinear cointegration. Unpublished manuscript.

NUMMELIN, E. (1984). *General Irreducible Markov Chains and Nonnegative Operators.* Cambridge Univ. Press.

PHILLIPS, P. C. B. and PARK, J. (1998). Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper 1181, Yale Univ.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

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

ROBINSON, P. M. (1997). Large sample inference for nonparametric regression with dependent errors. *Ann. Statist.* **25** 2054–2083.

STOCK, J. H. (1994). Unit roots, structural breaks and trends. In *Handbook of Econometrics* (R. F. Engle and D. L. McFadden, eds.) **4** 2740–2843. North-Holland, Amsterdam.

TOUATI, P. A. (1990) . Loi fonctionnelle du logarithme iteré pour les processus de Markov recurrents. *Ann. Probab.* **18** 140–159.

TJØSTHEIM, D. and AUESTAD, B. (1994). Nonparametric identification of nonlinear time series; projections and selecting significant lags. *J. Amer. Statist. Assoc.* **89** 1398–1419.

WATSON, M. W. (1994). Vector autoregression and cointegration. In *Handbook of Econometrics* (R. F. Engle and D. L. McFadden, eds.) **4** 2843–2918. North-Holland, Amsterdam.

WHEEDEN, R. L. and ZYGMUND, A. (1977). *Measure and Integral.* Dekker, New York.

XIA, Y. (1998). Doctoral thesis, Univ. Hong Kong.

YAKOWITZ, S. (1993). Nearest neigbour regression estimation for null-recurrent Markov times series. *Stochastic Process. Appl.* **48** 311–318.

DEPARTMENT OF MATHEMATICS
JOHANNES BRUNS GT.12
5008 BERGEN
NORWAY