NONPARAMETRIC ESTIMATION IN NULL RECURRENT TIME SERIES

By Hans Arnfinn Karlsen and Dag Tjøstheim

University of Bergen

Department of Mathematics, Johannes Bruns gt.12, 5007 Bergen, Norway

17th May 1998

Abstract

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 in 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 series 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 behaviour of the recurrence time.

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

AMS 1991 subject classification. Primary 62M10, 62G07; secondary 60J05.

Contents

1 Introduction **3**

2 Markov theory **5**
2.1 Notation 5
2.2 Basic conditions and the split chain 6
2.3 The invariant measure 8
2.3.1 A more general split chain 9
2.4 Notation for functions in several variables 10
2.5 Regularity concepts 11
2.6 β-null recurrence and tail behaviour of recurrence times 11
2.7 Weak limits for the number of regenerations 14
2.8 Regeneration and some strong laws 19
2.9 A central limit result 22

3 Asymptotics with a smoothing parameter **24**
3.1 Basic notation 24
3.2 Basic conditions 24
3.3 Properties of the number of regenerations 26
3.4 Weak limits 27
3.5 Consistency 34

4 Asymptotics for some nonparametric statistics **37**

A Appendix **49**

B Appendix **55**

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 et al, 1998a), and, more importantly, nonlinear processes are not excluded.

With the single exception of the work by Yakowitz (1993) on consistency of nearest neighbour estimates, as far as we know, the estimation theory of null recurrent processes has been confined to the parametric case. Asymptotics of parametric (usually non-time series) models have been treated by Höpfner (1990, 1994), Höpfner et al (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 & Kac (1957) and Kallianpur & 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 super-efficiency 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 neighbourhood 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, e.g. 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 discussed, and we refer to Myklebust et al (1998a) for some examples and details on practical implementations. 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 overwhelming 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 non-normal and the parameter estimates are super-efficient (Dickey & 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 & Escribano, 1997), but once more 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 Myklebust et al (1998a).

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 et al (1998), but there are also many hard problems connected with the basic estimation theory itself. A few of the latter ones are looked at in Myklebust et al (1998b).

Since our paper draws quite heavily on Markov theory for recurrent chains, we start in Section 2 by stating 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 3. Applications to nonparametric estimation of an invariant density and conditional mean/variance functions are given in Section 4, where we derive consistency and asymptotic normality of these estimates in a null recurrent situation. Some details of the technical derivations are relegated to two appendices.

2 Markov theory

2.1 Notation

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 $(E, \mathcal{E})$ with transition probability P. The sigma algebra of measurable sets, $\mathcal{E}$, 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 non-negative measurable functions with ϕ-positive support by $\mathcal{E}^+$. For a set $A \in \mathcal{E}$ we write $A \in \mathcal{E}^+$ if the indicator function $1_A \in \mathcal{E}^+$. The chain is Harris recurrent if for all $A \in \mathcal{E}^+$

$$\mathrm{P}(S_A < \infty \mid X_0 = x) \equiv 1 \quad \text{where,} \quad S_A = \min\{n \geq 1 \colon X_n \in A\} \ . \tag{2.1}$$

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 non-negative measurable function and λ is a measure, then the kernel $\eta \otimes \lambda$ is defined by

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

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 $A \in \mathcal{E}$ 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 $\mathcal{G}_r^d = \{f\colon\ (E^r, \mathcal{E}^r) \mapsto (R^d, \mathcal{B}(R^d)\}$ where $\mathcal{B}(R^d)$ is the class of Borel sets on R^d. If $r = 1$ or $d = 1$, we drop the subscript or superscript.

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

$$P^m \geq b\eta \otimes \lambda \ . \tag{2.2}$$

If λ satisfies (2.2) for some η, b and m, then λ is a small measure.

2.2 Basic conditions and the split chain

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

$$P^{m_0} \geq s \otimes \nu \ . \tag{2.3}$$

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

$$P \geq s \otimes \nu \tag{2.4}$$

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 (2.4) 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 2.1 *Assume that*

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

where $\{Z_t, \ t \geq 0\}$ 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

$$p^{(n)}(y \mid x) = \int p^{(n-1)}(y \mid u)\zeta(u - f(x))du, \quad n \geq 2 \ . \tag{2.5}$$

Let C be a compact set with positive Lebesgue measure. Define $\rho_0(y) = \inf_{x \in C} \zeta(y - f(x))$, $a = \int \rho_0(y)dy$, $\rho = a^{-1}\rho_0$, $s = a1_C$. 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 (2.4) is satisfied .

In the nonparametric estimation theory an important role will be played by the split chain, which can be constructed once the minorization condition is fulfilled. It permits splitting the chain into separate and identical parts which are building blocks in the subsequent analysis.

We introduce an auxiliary chain $\{Y_t\}$, where Y_t can only take the values 0 and 1. The split chain, $\{X_t, Y_t, t \geq 0\}$ is defined on an extension of the basic probability space

so that $\alpha = E \times \{1\}$ is a proper atom. The simplest description of this construction is given by an algorithm. Let $\widetilde{E} = E \times \{0,1\}$ and $\widetilde{\mathcal{E}}$ the corresponding extension of $\mathcal{E}$. Define

$$\eta_y(x) = s(x)y + (1 - s(x))(1 - y) = \begin{cases} s(x), & y = 1; \\ 1 - s(x), & y = 0 \end{cases} \tag{2.6}$$

with s as in (2.4). For each fixed $y \in \{0,1\}$, η_y is a function defined on E. For an arbitrary measure λ and an arbitrary function f, defined on $(E, \mathcal{E})$ let $\widetilde{\lambda}$ and $\widetilde{f}$ denote the extension to $(\widetilde{E}, \widetilde{\mathcal{E}})$ given by

$$\begin{aligned} \widetilde{f}(x,y) &= f(x)\eta_y(x), \\ \widetilde{\lambda}(dx, \{y\}) &= \lambda(dx)\eta_y(x) \ . \end{aligned}$$

Let

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

and

$$\Theta_y(x,A) = \nu(A)y + Q(x,A)(1-y) = \begin{cases} \nu(A), & y = 1; \\ Q(x,A), & y = 0. \end{cases} \tag{2.7}$$

By (2.4) the kernel Θ_y is a probability kernel on $(E, \mathcal{E})$ for each fixed y.

Let λ denote an arbitrary initial distribution on E , let $\mathcal{F}_{-1}^Y$ be the trivial sigma algebra and define $\{(X_t, Y_t), t \geq 0\}$ by

$$\begin{aligned} \mathrm{P}(X_0 \in A) &= \lambda(A), \\ \mathrm{P}(Y_t = y \mid \mathcal{F}_t^X \vee \mathcal{F}_{t-1}^Y) &= \eta_y(X_t), \quad t \geq 0; \\ \mathrm{P}(X_t \in A \mid \mathcal{F}_{t-1}^X \vee \mathcal{F}_{t-1}^Y) &= \Theta_{Y_{t-1}}(X_{t-1}, A), \quad t \geq 1, \end{aligned} \tag{2.8}$$

where $\mathcal{F}_t^X$ and $\mathcal{F}_t^Y$ are the σ-algebras generated by $\{X_j,\ j \leq t\}$ and $\{Y_j,\ j \leq t\}$. We observe that the distribution of $\{(X_n, Y_n),\ n \geq 0\}$ 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 the actual expressions involved are independent of λ, then we may drop the subscript.

Lemma 2.1 *The split chain defined by* (2.8) *is a Markov chain with state space* $\widetilde{E}$, *initial distribution* $\widetilde{\lambda}$ *and transition probability function* $\widetilde{P}$ *given by*

$$\widetilde{P}\Big((x_0, y_0), (dx, \{y\})\Big) = \Theta_{y_0}(x_0, dx)\eta_y(x) \ . \tag{2.9}$$

The set $\alpha = E \times \{1\}$ *is a proper atom for this chain; i.e.,* $\widetilde{P}((x,y), \cdot)$ *is independent of* (x,y) *when* $(x,y) \in \alpha$. *The* X*- marginal process of the compound chain has the same properties as the original chain, moreover*

$$\mathbf{P}(X_t \in A \mid \mathcal{F}_{t-1}^X \vee \mathcal{F}_{t-2}^Y) = P(X_{t-1}, A) \ . \tag{2.10}$$

<u>Proof:</u> See Nummelin (1984, p. 61 and Th.4.2, p. 62).

The compound chain is $\widetilde{\phi}$-irreducible, aperiodic and Harris recurrent (cf. Nummelin, 1984, Ch. 4.4).

The distribution of $\{(X_t, Y_t),\ t \geq 0\}$ given by (2.8) can be written

$$\begin{aligned}\mathbf{P}_{\widetilde{\lambda}}&(X_0 \in dx_0, Y_0 = y_0, X_1 \in dx_1, Y_1 = y_1, \ldots, X_k \in dx_k, Y_k = y_k, \ldots)\\ &= \widetilde{\lambda}(dx_0, \{y_0\})\Theta_{y_0}(x_0, dx_1)\eta_{y_1}(x_1) \cdots \Theta_{y_{k-1}}(x_{k-1}, dx_k)\eta_{y_k}(x_k)\cdots. \quad (2.11)\end{aligned}$$

We simplify the notation and write $\mathbf{P}_\lambda$ to denote this distribution. 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 $\delta_\alpha(x, y)$, i.e. $Y_0 = 1,\ X_0 = x$ arbitrary, then

$$\begin{aligned}\mathbf{P}_\alpha&(X_1 \in dx_1, Y_1 = y_1, \ldots, X_k \in dx_k, Y_k = y_k, \ldots)\\ &= \nu(dx_1)\eta_{y_1}(x_1)\,\Theta_{y_1}(x_1, dx_2)\eta_{y_2}(x_2) \cdots \Theta_{y_{k-1}}(x_{k-1}, dx_k)\eta_{y_k}(x_k) \cdots\\ &= \mathbf{P}_\nu(X_0 \in dx_1, Y_0 = y_1, X_1 \in dx_2, Y_1 = y_2, \ldots, X_k \in dx_{k+1}, Y_k = y_{k+1}, \ldots). (2.12)\end{aligned}$$

Let ζ be a non-negative measurable function defined on $(E \times \{0,1\})^\infty$. Then by (2.12)

$$\mathrm{E}_\alpha\Big[\zeta(X_1, Y_1, \ \ldots)\Big] = \mathrm{E}_\nu\Big[\zeta(X_0, Y_0, \ \ldots)\Big] = \mathrm{E}_\nu(\zeta)\ . \quad (2.13)$$

2.3 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.

Let $\tau = \tau_\alpha = \min\{n \geq 0\colon Y_n = 1\}$ and $S_\alpha = \min\{n \geq 1\colon Y_n = 1\}$. Since $\{S_\alpha = n\} = \{\cap_{j=1}^{n-1}(Y_j = 0),\ Y_n = 1\}$ and $\{\tau = n\} = \{S_\alpha = n\} \cap \{Y_0 = 0\}$, it follows from (2.11) (cf. Nummelin 1984, p. 63) that

$$\begin{aligned}\mathbf{P}_x(\tau = n) &= (P - s \otimes \nu)^n s(x), \quad n \geq 0,\\ \mathbf{P}_\alpha(S_\alpha = n) &= \nu(P - s \otimes \nu)^{n-1} s, \quad n \geq 1 \quad . \end{aligned} \quad (2.14)$$

Define π_s by

$$\pi_s(A) = \pi_s 1_A = \mathrm{E}_\alpha\Big[\sum_{n=1}^{S_\alpha} 1_A(X_n)\Big], \quad A \in \mathcal{E}\ . \quad (2.15)$$

Then by (2.14), (2.8) and (2.11)

$$\begin{aligned}\pi_s(A) &= \sum_{n=1}^{\infty} \mathrm{E}_\alpha\Big[1_A(X_n)1(S_\alpha \geq n)\Big]\\ &= \sum_{n=1}^{\infty} \nu(P - s \otimes \nu)^{n-1} 1_A\\ &= \nu G_{s,\nu} 1_A \end{aligned} \quad (2.16)$$

where

$$G_{s,\nu} \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} (P - s \otimes \nu)^n \ . \tag{2.17}$$

This means that $\pi_s = \nu G_{s,\nu}$ and by (2.14) $\pi_s(s) = \mathrm{P}_\alpha(S_\alpha < \infty)$. Since the split chain is Harris recurrent,

$$G_{s,\nu} s(x) = \mathrm{P}_x(\tau < \infty) \equiv 1 \ . \tag{2.18}$$

Thus $\pi_s(s) = 1$. From (2.17) we get

$$G_{s,\nu} = I + G_{s,\nu} P - G_{s,\nu} s \otimes \nu \tag{2.19}$$

which implies that $\pi_s = \pi_s P$. Thus π_s is an invariant measure. The results stated below can be found in Nummelin (1984).

Remark 2.1 *If π is another invariant measure then $\pi = \pi(s)\pi_s$ (p. 73). The invariant measure π_s is equivalent to ϕ, $\pi_s(C) < \infty$ for all small sets C and it is σ-finite (Prp. 5.6., p. 72).*

The chain is positive recurrent if and only if $\pi_s 1_E < \infty$ (p. 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.

Remark 2.2 *It is seen from (2.15) that $\{X_t\}$ is positive recurrent if and only if $\mathrm{E}_\alpha S_\alpha < \infty$.*

It is seen from Remark 2.1 that although the representation of π_s given by (2.16) does depend of the atom (s, ν), the measure π_s itself is independent of ν and only depends on s through a constant.

We may extend π to the compound chain by $\widetilde{\pi}_s(dx, \{y\}) \stackrel{\text{def}}{=} \pi_s(dx)\eta_y(x)$, and we have $\widetilde{\pi} = \widetilde{\pi}\widetilde{P}$.

Suppose that the original chain has a proper atom α. Let $s = 1_\alpha$ and $\nu = P_\alpha = P(\alpha, \cdot)$. Then $P(x, dy) \geq 1_\alpha(x) P(x, dy) = s(x)\nu(dy)$. Hence (2.4) is satisfied and all the formulae in this sub-section are still true if we define the auxiliary process $\{Y_t\}$ by $Y_t = 1_\alpha(X_t)$. It is common to denote $G_{s,\nu}$ by G_α in this case.

2.3.1 A more general split chain

If $m_0 > 1$ in (2.3), then the m_0-step chain $\{X_{tm_0}\}$ satisfies (2.4) with transition probability P^{m_0} and (s, ν) is an atom for this chain. The corresponding definitions of $G_{s,\nu}$ and π_s are given by

$$G_{m_0,s,\nu} = \sum_{t=0}^{\infty} (P^{m_0} - s \otimes \nu)^t, \quad \pi_s = \nu G_{m_0,s,\nu} \ . \tag{2.20}$$

But π_s is still the unique invariant measure for the original chain which satisfies $\pi_s(s) = 1$.

2.4 Notation for functions in several variables

It is necessary to extend the notation of the preceding sub-sections to functions of several variables. All integrals involved will be assumed to be well-defined.

Recall that for $g \in \mathcal{G}_1$, $\pi_s(g) = \int \pi_s(dx)g(x)$ and (cf. (2.15) and (2.17))

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

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

Definition 2.1 *Let $r \geq 1$, and let $g \in \mathcal{G}_r$. 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*

$$\widetilde{I}_g(x,dy)(r) = \int P(x,dx_2) \cdots P(x_{r-1},dy)g(x,x_2,\ldots,x_{r-1},y) \tag{2.21}$$

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

$$\widetilde{g} = \widetilde{I}_g 1 \ . \tag{2.22}$$

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

$$\mathrm{E}_x\Big[g(X_0,X_1,\ldots,X_{r-1})\Big] = \widetilde{I}_g 1(x) = \widetilde{g}(x) \tag{2.23}$$

and

$$\mathrm{E}_x\Big[\sum_{j=0}^{\tau} g(X_j,X_{j+1},\ldots,X_{j+r-1})\Big] = G_{s,\nu}\widetilde{g}(x) \tag{2.24}$$

which is easily verified (cf. (2.15) and (2.17)). The right hand sides of (2.23) and (2.24) 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 $g \in \mathcal{G} = \mathcal{G}_1$ then $\widetilde{I}_g = I_g P$ and $\widetilde{g} = I_g P 1 = g$. In order to reduce the notation further we extend π_s to $\cup_{r=1}^{\infty}\mathcal{G}_r$ by

$$\pi_s g \stackrel{\text{def}}{=} \pi_s \widetilde{g} = \int \pi_s(dx_1)P(x_1,dx_2) \cdots P(x_{r-1},dx_r)g(x_1,\ldots,x_r), \quad g \in \mathcal{G}_r \ . \tag{2.25}$$

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

$$L_r^p(\pi_s) \stackrel{\text{def}}{=} \Big\{g \in \mathcal{G}_r \colon \|g\|_{p,\pi_s}^p \stackrel{\text{def}}{=} \pi_s \widetilde{I}_{|g|^p} 1 < \infty\Big\}, \quad p \in (0,\infty),\ r \geq 1 \ . \tag{2.26}$$

If we define $\widetilde{\pi}_{s,r}(dx_1,\ldots,dx_r) = \pi_s(dx_1)P(x_1,dx_2) \cdots P(x_{r-1},dx_r)$, then $L_r^p(\pi_s) = L^p(\widetilde{\pi}_{s,r})$.

All of the notation in this sub-section is trivially extended to $\mathcal{G}_r^d$.

2.5 Regularity concepts

We wish to formulate the regularity conditions to be stated in Section 3 and 4 in terms of standard Markov chain concepts, and we therefore include the definition of a special function and the concept of a f-regular measure. Theorem 2.1 will be used repeatedly in Section 3.

Definition 2.2 *Let f be a non-negative measurable function, defined on E, which is π_s-integrable. The kernels $\{V_A, A \in \mathcal{E}^+\}$ are defined by*

$$V_A f(x) = \mathrm{E}_x\Big[\sum_{n=1}^{S_A} f(X_n)\Big],$$

where $S_A = \min\{n \geq 1,\ X_n \in A\}$. Let λ be a finite measure on $\mathcal{E}$ and $g \in L_r^1(\pi_s)$. The measure λ is g-regular if $\widetilde{I}_{|g|}1$ is λ-integrable and $\lambda V_A \widetilde{I}_{|g|}1$ is finite for all A. The function g is special if $g \in L_r(\pi_s) \cap L_r^2(\pi_s)$, $\sup \widetilde{I}_{|g|}1$ is finite and $\sup V_A \widetilde{I}_{|g|}1$ is finite for all A. The set $D \in \mathcal{E}^+$ is a special set if 1_D is a special function.

To indicate that the restriction $m_0 = 1$ in (2.3) can be relaxed we state the following theorem for a general m_0. In this paper it will only be used with $m_0 = 1$, however.

Theorem 2.1 [Nummelin 1984, Prp. 5.13, p. 80)] *Assume that $\{X_t\}$ is aperiodic. Let $g \in L_r^1(\pi_s)$ and $\Delta_{m_0} = I + P + \cdots + P^{m_0-1}$. A finite measure λ is g-regular if and only if $\lambda G_{m_0,s,\nu} \Delta_{m_0} \widetilde{I}_{|g|}1$ is finite. Assume that $\sup \widetilde{I}_{|g|}1$ is finite and $g \in L_r(\pi_s) \cap L_r^2(\pi_s)$. Then g is a special function if and only if $\sup G_{m_0,s,\nu} \Delta_{m_0} \widetilde{I}_{|g|}1$ is finite. In particular, for $r = 1$ all small functions are special.*

Remark 2.3 *Nummelin (1984) only treats $r = 1$, but since $\widetilde{I}_{|g|}1$ is a non negative function in one variable, the extension of the theorem to $r > 1$ is trivial. However, the condition that $\sup \widetilde{I}_{|g|}1$ is finite is in general a weaker condition than $\sup |g| < \infty$ when $r > 1$. In particular, even if g satisfies that $\sup \widetilde{I}_{|g|}1$ and $\pi_s|g|$ are finite, higher order moments of $|g|$ with respect to $\pi_s(x_0)P(x_0, dx_1) \cdots P(x_{r-2}, dx_{r-1})$ may not be finite.*

2.6 β-null recurrence and tail behaviour of recurrence times

To carry asymptotic theory through we need a regularity condition for the tail behaviour 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 (2.4) holds.

A positive function L defined on $[a, \infty)$, where $a \geq 0$, is slowly varying at infinity (Bingham et. al. 1989, p. 6) if

$$\lim_{x \uparrow \infty} \frac{L(\kappa x)}{L(x)} = 1 \quad \text{for all } \kappa > 0 \text{ and for all } x \in [a, \infty) \ . \tag{2.27}$$

Definition 2.3 *The Markov chain $\{X_t\}$ is β-null recurrent if there exists a small non-negative function h, an initial measure λ, a constant $\beta \in (0,1)$ and a slowly varying function L_h so that*

$$\mathrm{E}_\lambda \Big[\sum_{t=0}^{n} h(X_t) \Big] \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_h(n) \tag{2.28}$$

as $n \to \infty$.

Remark 2.4 *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 application of slowly varying functions they are only unique up to equivalence. Hence, when (2.28) is true, without any loss of generality we assume that L_h is normalized (Bingham et. al. , 1989, pp 15, 24), i.e., the function $x^\beta L_h(x)$ is strictly increasing and continuous in the interval $[x_0, \infty)$ for some x_0.*

Let

$$G^{(n)} = \sum_{t=0}^{n} P^t \ . \tag{2.29}$$

The left hand side of (2.28) can be written as $\lambda G^{(n)} h$. We first prove that for a fixed parameter β (2.28) is actually a global property shared by all non-negative special functions.

Lemma 2.2 *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 special functions f the asymptotic relation (2.28) holds with $L_f = \pi_s(f) L_s$ where π_s is defined by (2.20).*

<u>Proof:</u> Let λ and h be given by Definition 2.3 and the atom (s, ν) be fixed. Let

$$L_s \stackrel{\text{def}}{=} \frac{L_h}{\pi_s h} \ . \tag{2.30}$$

Using a null recurrent ratio limit theorem (Nummelin, 1984, Cor. 7.2(i), p. 131) and (2.28),

$$\frac{\lambda G^{(n)} h}{\nu G^{(n)} s} = \pi_s(h)(1 + o(1)) \ . \tag{2.31}$$

Using (2.28) again and the above expression, it follows that

$$\nu G^{(n)} s \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_s(n) \ . \tag{2.32}$$

Let f be a given special function. Then by (2.31) with f instead of h and by (2.32) it follows that

$$\lambda G^{(n)} f \sim \frac{1}{\Gamma(1+\beta)} n^{\beta} \pi_s(f) L_s(n) \ . \tag{2.33}$$

□

Remark 2.5 *If the atom (s, ν) in (2.4) is changed to (s', ν') then (cf. Remark 2.1)*

$$\pi_{s'} = \frac{\pi_s}{\pi_s(s')}, \quad L_{s'} \sim \pi_s(s') L_s \ . \tag{2.34}$$

The asymptotic expression (2.32) is closely connected to the Tauberian theorem (Feller, 1971, p. 447):

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

$$\sum_{k=0}^{n} d_k \sim \frac{1}{\Gamma(1+\rho)} n^{\rho} L_1(n) \Longleftrightarrow d(r) \sim (1-r)^{-\rho} L_1\Big(\frac{1}{1-r}\Big) \tag{2.35}$$

when $n \to \infty$ and $r \uparrow 1^-$, respectively. If $\{d_n\}$ is monotone and $\rho > 0$, then each of the conditions given by (2.35) is equivalent with

$$d_n \sim \frac{n^{\rho-1}}{\Gamma(\rho)} L_1(n) \ . \tag{2.36}$$

If (2.4) 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 behaviour of the distribution of the recurrence time S_α.

Theorem 2.2 *Assume (2.4) is true. Then $\{X_t\}$ is β-null recurrent if and only if*

$$\mathbf{P}_\alpha(S_\alpha > n) = \frac{1}{\Gamma(1-\beta) n^{\beta} L_s(n)} \Big(1 + o(1)\Big) \ . \tag{2.37}$$

Remark 2.6 *If (2.37) is true, then it is not difficult to show that*

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

Thus, even though $\mathrm{E}_\alpha S_\alpha = \infty$ for a null recurrent process, if (2.4) and (2.37) hold, then $\mathrm{E}_\alpha S_\alpha^p$ is finite for p small enough. For an ordinary random walk $\beta = 1/2$ (Kallianpur and Robbins, 1954) and hence $\mathrm{E}_\alpha S_\alpha^p < \infty$ for $0 \leq p < 1/2$. Some other examples of β-null recurrent processes are given in Myklebust et al (1998a).

Proof of Theorem 2.2: 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 (2.32) and (2.35) (with $\rho = \beta$ and $L_1 = L_s$), β-null recurrence is equivalent with

$$\nu G(r) s \sim (1-r)^{-\beta} L_s\Big(\frac{1}{1-r}\Big) . \tag{2.38}$$

We have $B_n \stackrel{\text{def}}{=} \mathrm{P}_\alpha(S_\alpha > n) = \nu(P - s \otimes \nu)^n 1$. If (2.37) holds, then by (2.35) and (2.36) (with $\rho = 1 - \beta$ and $L_1 = L_0 = 1/L_s$)

$$B(r) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} r^k B_k = \nu G_{s,\nu}(r) 1 \sim (1-r)^{\beta-1} L_0\Big(\frac{1}{1-r}\Big) . \tag{2.39}$$

Let $b_n = \mathbf{P}_\alpha(S_\alpha = n)$, $w_n = \mathrm{P}_\alpha(Y_n = 1)$ 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

$$\begin{aligned} w_n &= \mathrm{P}_\alpha(Y_n = 1, S_\alpha \geq n) + \sum_{k=1}^{n-1} \mathbf{P}_\alpha(Y_{n-k} = 1)\mathrm{P}_\alpha(S_\alpha = k) \\ &= \sum_{k=0}^{n} w_{n-k} b_k, \quad n \geq 1 \end{aligned} \tag{2.40}$$

which shows that $\{w_n\}$ is an undelayed renewal sequence corresponding to the increment sequence $\{b_n\}$. By (2.40) 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

$$w(r) = \frac{1}{B(r)(1-r)} . \tag{2.41}$$

By (2.8) and (2.13) we find that $w_n = \mathrm{P}_\alpha(Y_n = 1) = \mathbf{E}_\nu\Big[s(X_{n-1})\Big] = \nu P^{n-1} s$ when $n \geq 1$. This gives

$$w(r) = 1 + r\Big[\nu G(r) s\Big] . \tag{2.42}$$

Combining (2.41) and (2.42) we finally obtain

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

This identity in conjunction with (2.38) and (2.39) show the equivalence. □

2.7 Weak limits for the number of regenerations

We assume aperiodicity and that (2.4) and (2.37) hold in this section. We also assume that L_s is normalized in the β-null case which implies that the function

$$u(z) \stackrel{\text{def}}{=} z^\beta L_s(z), \quad z \in R_+ \tag{2.43}$$

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

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

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

A weak limit result can be derived from (2.37). Let $T(n)$ denote the complete number of regenerations in the time interval $[0, n]$, i.e.,

$$T(n) = \begin{cases} \max\{k\colon \tau_k \le n\}, & \text{if } \tau_0 \le n; \\ 0, & \text{otherwise,} \end{cases} \tag{2.44}$$

where

$$\tau_k = \begin{cases} \inf\{n \ge 0\colon Y_n = 1\}, & k = 0; \\ \inf\{n > \tau_{k-1}\colon Y_n = 1\}, & \text{when } k \ge 1. \end{cases} \tag{2.45}$$

Then $\mathrm{E}_\alpha\Big[T(n)\Big] = \mathrm{E}_\alpha(\sum_{j=1}^n Y_j) = \sum_{k=0}^{n-1} \nu P^k s$, and it follows by (2.28) (with $h = s$),

$$\mathrm{E}_\alpha\left[\frac{T(n)}{u(n)}\right] = \frac{1}{\Gamma(1+\beta)} + \mathrm{o}(1)\ . \tag{2.46}$$

In some respects $T(n)$ corresponds to the total number of observations for a positive recurrent process, and it plays a crucial role in the asymptotics of Section 3 and 4. Our next task is to derive a functional limit theorem for $T(n)$.

Let $\mathcal{D}[0,\infty)$ denote the space of right continuous real valued functions with finite left hand limits, i.e. this is the space of cadlag functions defined on $[0,\infty)$ (cf. Appendix B). We write $\xrightarrow{\mathcal{L}_{\mathcal{D}[0,\infty)}}$ for weak convergence in $\mathcal{D}[0,\infty)$ and $\xrightarrow{\text{fd}}$ for convergence of finite dimensional laws. A Levy process is a stochastic process with stationary independent increments and sample paths in $\mathcal{D}[0,\infty)$. Consider the process

$$S_{\beta,z}(t) \stackrel{\text{def}}{=} \frac{1}{v(z)} \sum_{k=1}^{[zt]} (\tau_k - \tau_{k-1}), \quad t \in [0,\infty), \quad z \in R_+\ . \tag{2.47}$$

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

By (2.37) (cf. Bingham et. al., 1989, p. 349) it follows that

$$S_{\beta,z} \xrightarrow[z]{\text{fd}} S_\beta\ . \tag{2.48}$$

where S_β is the one-sided stable Levy 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)

$$S_{\beta,z} \xrightarrow[z]{\mathcal{L}_{\mathcal{D}[0,\infty)}} S_\beta\ . \tag{2.49}$$

The Mittag Leffler process (cf. Kasahara, 1984) with parameter β, $M_\beta = \{M_\beta(t),\ t \ge 0\}$ 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

$$\mathrm{E}\Big[\exp^{\{i\zeta M_\beta(t)\}}\Big] = \sum_{k=0}^{\infty} \frac{(i\zeta t^\beta)^k}{\Gamma(1+k\beta)}, \quad \zeta \in R, \quad t \ge 0\ . \tag{2.50}$$

An alternative description is given by

$$E(M_\beta^m(1)) = \frac{m!}{\Gamma(1+m\beta)}, \quad m \geq 0, \quad M_\beta(t) \stackrel{\mathrm{d}}{=} t^\beta M_\beta(1) \ . \tag{2.51}$$

We need the continuous extension of $T(n)$. Let

$$T_n = \Big\{\frac{T([nt])}{u(n)}, \quad t \geq 0\Big\} \ . \tag{2.52}$$

The next theorem establishes a weak limit result, which will be of use in the asymptotic theory to be established in the sequel.

Theorem 2.3 *Let λ be any initial measure. Assume that* (2.37) *holds. Then*

$$\mathrm{E}_\lambda\Big[T(n)\Big]^m \sim \frac{n^{m\beta} L_s^m(n)}{\Gamma(1+m\beta)} \tag{2.53}$$

and

$$T_n \stackrel{\mathcal{L}_{\mathcal{D}[0,\infty)}}{\underset{n}{\longrightarrow}} M_\beta \ . \tag{2.54}$$

<u>Proof:</u> Let

$$\widetilde{T}(n) = \sum_{k=0}^n Y_k, \tag{2.55}$$

then

$$T(n) = (\widetilde{T}(n) - 1)1(\widetilde{T}(n) > 0) \ . \tag{2.56}$$

The normalized continuous version is also denoted by $\widetilde{T}_n$. It has the same properties as T_n. By Lemma A.3 in Appendix A we can write

$$\mathrm{E}_\alpha\Big[\widetilde{T}(n)\Big]^m = \sum_{k=1}^m \sum_{\ell \in \Delta_{m,k}} \binom{m}{\ell} J_{n,k,\ell} \tag{2.57}$$

where $\Delta_{m,k} = \{\ell = (\ell_1, \ldots, \ell_k) \in \mathcal{N}_+^k \colon \sum \ell_i = m\}$ for $k \geq 1$, and where

$$\begin{aligned} J_{n,k,\ell} &= \sum_{h_1=0}^n \sum_{h_2=1}^{n-h_1} \cdots \sum_{h_k=1}^{n-h_1-\cdots-h_{k-1}} E_\alpha\Big[Y_{h_1}^{\ell_1} Y_{h_1+h_2}^{\ell_2} \cdots Y_{h_1+\cdots+h_k}^{\ell_k}\Big] \\ &= \sum_{h_1=0}^n \sum_{h_2=1}^{n-h_1} \cdots \sum_{h_k=1}^{n-h_1-\cdots-h_{k-1}} w_{h_1} \cdots w_{h_k}, \quad w_h = \mathbf{P}_\alpha(Y_h = 1) \ . \end{aligned} \tag{2.58}$$

We can write $J_{n,k,\ell} = J_{n,k}$ since (2.58) shows that this quantity is independent of ℓ. Let

$$J_k(r) \stackrel{\mathrm{def}}{=} \sum_{n=0}^\infty J_{n,k} r^n, \quad r \in [0,1) \ . \tag{2.59}$$

Then it can be shown from (2.58), (2.39) and (2.41) 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

$$J_k(r) \sim w^k(r)(1-r)^{-1} \sim (1-r)^{-k\beta-1} L_s^k\Big(\frac{1}{1-r}\Big) \tag{2.60}$$

as $r \uparrow 1^-$. From (2.35), (2.36) with $\rho = k\beta + 1$, $L_1 = L_s^k$, since $\{J_{n,k},\ n \geq 1\}$ is a monotone sequence in n, (2.60) implies

$$J_{n,k} \sim \frac{n^{k\beta} L_s^k(n)}{\Gamma(1+k\beta)} = \frac{u^k(n)}{\Gamma(1+k\beta)} \tag{2.61}$$

as $n \to \infty$. Inserting (2.61) into (2.57) gives

$$\mathrm{E}_\alpha\big[\widetilde{T}(n)\big]^m \sim \sum_{k=1}^{m} \Big\{ \sum_{\ell \in \Delta_{m,k}} \binom{m}{\ell} \Big\} \frac{n^{k\beta} L_s^k(n)}{\Gamma(1+k\beta)} \,. \tag{2.62}$$

and since $\Delta_{m,m} = \{\underline{1}\} = \{(1,\ldots,1)\}$ and $\binom{m}{\underline{1}} = m!$ (cf Lemma A.3, Appendix A), we finally get by (2.62) that

$$\mathrm{E}_\alpha\left[\frac{\widetilde{T}(n)}{u(n)}\right]^m \sim \frac{m!}{\Gamma(1+m\beta)} \tag{2.63}$$

and (2.53) is proved.

We turn to the proof of (2.54). By the method of moments and (2.50) we find that for each t

$$\widetilde{T}_n(t) \xrightarrow[n]{d} M_\beta(t) \,.$$

However, it is difficult to establish a functional weak convergence from the marginal convergences since M_β is not a Levy process. In order to prove (2.54) it is an advantage to use a continuous index; i.e., $T_z(t) \stackrel{\text{def}}{=} T([zt])/u(z)$. By (2.48) and the proof of Theorem B.1 of Appendix B with $S_\beta = A$ in that proof,

$$S_{\beta,z}^{(-1)} \xrightarrow[z]{\mathcal{L}_{\mathcal{D}[0,\infty)}} M_\beta, \quad \text{where} \quad S_{\beta,z}^{(-1)}(t) = \inf\{x\colon\ S_{\beta,z}(x) > t\} \,. \tag{2.64}$$

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 (2.54) it is sufficient to prove that

$$\sup_{0<t\leq K} |T_z(t) - S_z^{(-1)}(t)| = o_P(1) \tag{2.65}$$

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

$$\{\sum_{j=1}^{n} Y_j > m\} = \{\tau_m < n\}, \quad \{\sum_{j=1}^{n} Y_j < m\} = \{\tau_m > n\} . \tag{2.66}$$

Let $\eta > 0$. From (2.47) and (2.66) we get

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

and in the same way we get

$$\begin{aligned} \{S_{u(z)}^{(-1)}(t) > \eta\} &\subseteq \{S_{u(z)}(\eta) \le t\} \\ &= \{T_z(t) \ge u^{-1}(z)[u(z)\eta]\} . \end{aligned}$$

Let $\epsilon_1 \in (0,1)$ be arbitrary. Then for $\eta_1 < \eta_2$,

$$\begin{aligned} \{\eta_1 \le S_{u(z)}^{(-1)}(t) < \eta_2\} &\subseteq \{\eta_1(1-\epsilon_1) < S_{u(z)}^{(-1)}(t) < \eta_2\} \\ &\subseteq \{u^{-1}(z)[u(z)\eta_1(1-\epsilon_1)] \le T_z(t) < u^{-1}(z)[u(z)\eta_2]\} \end{aligned}$$

which gives

$$|T_z(t) - S_{u(z)}^{(-1)}(t)| < (\eta_2 - \eta_1) + \epsilon_1\eta_1 + \frac{1}{u(z)} \quad \text{when} \quad S_{u(z)}^{(-1)}(t) \in [\eta_1, \eta_2) . \tag{2.67}$$

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

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

By (2.64)

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

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 $\epsilon > 0$, we can choose $\eta_0, \ldots, \eta_L$, z_1, ϵ_1 with $\eta_0 = 0$ and $\eta_L = s_0$ so that $\max_k(\eta_{k+1} - \eta_k) < \epsilon/3$, $\epsilon_1 < s_0^{-1}\epsilon/3$ and $z_1 > v(3\epsilon^{-1})$. Then by (2.67)

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

and therefore

$$P\Big(\sup_{t\le K} |T_z(t) - S_{u(z)}^{(-1)}(t)| > \epsilon\Big) < \delta, \quad z > z_1 . \tag{2.68}$$

The function u is unbounded continuous and strictly increasing in an interval $[x_0, \infty)$ and therefore by (2.64),

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

Hence (2.68) implies (2.65).

□

2.8 Regeneration and some strong laws

The decomposition result of Lemma 2.4 and the strong law of Lemma 2.6 are both key elements in the asympotic theory of Section 3 and 4.

Lemma 2.3 *Assume that* (2.4) *holds and let* τ_k *be defined by* (2.45). *Then*

$$\Big\{\Big(\tau_k - \tau_{k-1},\ X_{\tau_{k-1}+1}, , \ldots, X_{\tau_k}\Big),\ k \ge 1\Big\} \tag{2.69}$$

are iid random elements which are independent of (X_0, Y_0).

Proof: This is true since $\alpha = E \times \{1\}$ is a proper atom for the split chain (cf. Nummelin 1984, p. 76).

□

The regeneration method is essential in decomposing the sum

$$S_n(g) = \sum_{j=0}^{n} g(X_j, \ldots, X_{j+r-1}) . \tag{2.70}$$

with $g \in \mathcal{G}_r^d$. Sums of this type will be needed in the nonparametric analysis. Let $\{U_k, k \ge 0\}$, $U_{(n)}$ be defined by

$$U_k = \begin{cases} \sum_{j=0}^{\tau_0} g(X_j, \ldots, X_{j+r-1}), & \text{when } k = 0; \\ \sum_{j=\tau_{k-1}+1}^{\tau_k} g(X_j, \ldots, X_{j+r-1}), & \text{when } k \ge 1; \\ \sum_{j=\tau_{T(n)}+1}^{n} g(X_j, \ldots, X_{j+r-1}), & \text{when } k = (n) \end{cases} \tag{2.71}$$

Lemma 2.4 *Assume that* (2.4) *holds and that* $g \in \mathcal{G}_r^d$ *for* $r \ge 1$. *Then*

$$S_n(g) = U_0 + V_{T(n)} + U_{(n)}, \quad V_n = \sum_{k=1}^{n} U_k, \quad V_0 = 0, \tag{2.72}$$

where $\{U_k,\ k \ge 1\}$ *is a* $(r-1)$*-dependent stationary sequence, which is independent of the initial distribution of the Markov chain. We have for* $k \ge 1$;

$$\begin{aligned} (U_k, U_{k+1}) &\stackrel{d}{=} \mathcal{L}_\alpha\Big\{\sum_{j=1}^{S_\alpha(0)} g(X_j, \ldots, X_{j+r-1}), \sum_{j=S_\alpha(0)+1}^{S_\alpha(1)} g(X_j, \ldots, X_{j+r-1})\Big\} \\ &\stackrel{d}{=} \mathcal{L}_\nu\Big\{\sum_{j=0}^{\tau_0} g(X_j, \ldots, X_{j+r-1}), \sum_{j=\tau_0+1}^{\tau_1} g(X_j, \ldots, X_{j+r-1})\Big\} \end{aligned} \tag{2.73}$$

where $\mathcal{L}_\nu$ denote the distribution of $\{(X_n, Y_n),\ n \geq 0\}$ with initial distribution ν (cf. (2.8), (2.12)) *and $S_\alpha(k) = \min\{n > S_\alpha(k-1)\colon\ \ Y_n = 1\}$ and $S_\alpha(-1) \stackrel{\text{def}}{=} 0$.*

Proof: This is a consequence of Lemma 2.3 and of (2.71) (cf. Nummelin 1984, p. 135).

□

We note that U_0 has the same distribution as one of the U_k's, $k \geq 1$, when ν is the initial measure. We use the notation

$$U_g = U_0 \tag{2.74}$$

when this is the case. From (2.15), (2.24) and (2.73) we have

$$\mu_g \stackrel{\text{def}}{=} \pi_s g = \mathbf{E} U_k = \mathbf{E}_\nu U_g, \quad k \geq 1\ . \tag{2.75}$$

Assume that U_g has a finite covariance matrix. The asymptotic covariance matrix for $n^{-1/2}\sum_{k=1}^n U_k$, given by

$$\Sigma_g = \lim_{n\to\infty} \mathrm{Var}\Big(n^{-1/2}\sum_{k=1}^n U_k\Big) = \sum_{h=-(r-1)}^{r-1} \gamma_U(h), \tag{2.76}$$

is important in a central limit theorem for $S_n(g)$. Here γ_U is the autocovariance function for the stationary process $\{U_k,\ k \geq 1\}$. When g real-valued, we write $\sigma_g^2 = \Sigma_g$. We note that if A is a $d_1 \times d$ dimensional matrix, then

$$U_{Ag} = AU_g, \quad \Sigma_{Ag} = A\Sigma_g A', \tag{2.77}$$

where A' is the transposed of A.

More explicit expressions for μ_g, σ_g^2 and Σ_g when $r = 2$ are given in Appendix A, Lemma A.1 and Lemma A.2.

We need a result of general character:

Lemma 2.5 *Let $\{X_n,\ n \geq 0\}$ and $\{\mathcal{T}_k,\ k \geq 0\}$ be stochastic sequences defined on a common probability space where $\mathcal{T}_k \in N_+$ Assume that $\mathcal{T}_k \uparrow \infty$* a.s. *and let X denote a stochastic variable which may be infinite with positive probability. If $X_n \to X$ a.s. then $X_{\mathcal{T}_k} \to X$ a.s.*

Proof of Lemma 2.5: Let ω be a fixed outcome so that $x_n = X_n(\omega)$ converges towards $x = X(\omega)$ and $t_k = \mathcal{T}_k(\omega)$ goes to infinity. Then the subsequence $\{x_{t_k}\}$ converges to x. Since this is true for all ω outside a null set, the assertion is true. □

Lemma 2.6 *Assume that* (2.4) *holds. Let* $r \geq 1$, $g \in \mathcal{G}_r^d$ *and* $\|g\| \in L_r^1(\pi_s)$, *and let the process have an arbitrary initial distribution* λ. *Then*

$$\frac{S_n(g)}{T(n)} \xrightarrow[n]{\text{a.s.}} \pi_s(g) \ . \tag{2.78}$$

Proof: Assume that $d = 1$ and $g \geq 0$. By (2.72) we can write $S_n(g) = U_0 + V_{T(n)} + U_{(n)}$. Since $g \geq 0$ we get

$$V_{T(n)} \leq S_n(g) \leq U_0 + V_{T(n)+1} \ .$$

By the definition of U_0, $\{\tau < \infty\} \subseteq \{U_0 < \infty\}$. From (2.18), $\mathbf{P}_x(\tau < \infty) \equiv 1$. Hence $\mathrm{P}_\lambda(|U_0| < \infty) = 1$. The convergence result in (2.78) is then a consequence of Lemma 2.5 and 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 non-negative g-functions. □

Corollary 2.1 *Assume that* (2.4) *holds. Let the process have an arbitrary initial distribution* λ. *Let* $g \in \mathcal{G}_r^d$, $g \in L_r^1(\pi_s)$ *and* $\pi_s(g)$ *be non-zero. If* (2.37) *is true, then*

$$\frac{S_n(g)}{n^\beta L_s(n)} \xrightarrow[n]{d} \pi_s(g) M_\beta(1) \ . \tag{2.79}$$

Moreover, let $g_1 \in L_r^1(\pi_s)$ *and* $\pi_s(g_1) \neq 0$. *Then*

$$\frac{S_n(g)}{S_n(g_1)} \xrightarrow[n]{\text{a.s.}} \frac{\pi_s(g)}{\pi_s(g_1)} \ . \tag{2.80}$$

Proof: By (2.54) and (2.78) it follows that

$$\frac{S_n(g)}{n^\beta L_s(n)} = \left\{\frac{S_n(g)}{T(n)}\right\}\left\{\frac{T(n)}{n^\beta L_s(n)}\right\} \xrightarrow[n]{d} \pi_s(g) M_\beta(1) \tag{2.81}$$

when $\pi_s(g)$ is finite. The last part is a consequence of

$$\frac{S_n(g)}{S_n(g_1)} = \left\{\frac{S_n(g)}{T(n)}\right\}\left\{\frac{S_n(g_1)}{T(n)}\right\}^{-1} .$$

□

Remark 2.7 *If* $C \in \mathcal{E}^+$, *the number of times the process is visiting* C *up to the time* n *is denoted by*

$$T_C(n) = \sum_{k=0}^{n} 1_C(X_k) = S_n(1_C) \ . \tag{2.82}$$

From (2.78) *we have that* $T_C(n)/T(n) \xrightarrow{\text{a.s.}} \pi_s 1_C$. *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 Section* 3 *and* 4.

Remark 2.8 *Corollary* 2.1 *shows that there is essentially only one limit variable* $M_\beta(1)$*; i.e., if* $g = (g_1, g_2)$ *is a two dimensional vector in* (2.79)*, then the weak limit is degenerate.*

Remark 2.9 *If* $\{X_t\}$ *is positive recurrent with* π *as its unique stationary probability measure then* $\pi = \pi(s)\pi_s$ *and* $g = 1$ *is* π*-integrable. It follows by Lemma* 2.6 *that* $T(n)/n = T(n)/S_n(1_E) \overset{\text{a.s.}}{\to} 1/\pi_s(1) = \pi(s)$*, and consequently by the same lemma* $T_C(n)/n \to \pi(C)$ *a.s..*

2.9 A central limit result

We assume aperiodicity, that (2.4) is fulfilled and that $g \in \mathcal{G}_2^d$; i.e., g is a function of two variables. The latter restriction is not essential but simplifies notation.

In Theorem 2.4 we state without proof a central limit result for null recurrent Markov chains. In Section 3 this result is proved and at the same time extended to include a smoothing parameter. The reason for stating Theorem 2.4 separately is that it is intimately connected with the concepts developed in the present section, and that it is more transparent than the general result. Also we give conditions on g (Lemma 2.7) guaranteeing the finiteness of the variance parameter in the central limit expression. These conditions will be referred to in Section 3. We use the standard Euclidean norm $\|g\| = [\sum_{i=1}^d g_i^2(x,y)]^{1/2}$ for $g \in \mathcal{G}_2^d$. We start with the following equivalences for $d = 1$.

Lemma 2.7 *Let* $g \in \mathcal{G}_2$ *and* U_g *be given by* (2.74)*. Then the following conditions are equivalent*

$$\mathrm{E}_\nu U_{|g|}^2 < \infty \ . \tag{2.83}$$

$$g \in L_2(\pi_s) \cap L_2^2(\pi_s) \quad \textit{and} \quad \pi_s \widetilde{I}_{|g|} G_{s,\nu} \widetilde{I}_{|g|} 1 < \infty \ . \tag{2.84}$$

$$\textit{The measure } \pi_s \widetilde{I}_{|g|} \textit{ is } \widetilde{I}_{|g|}1\textit{-regular and } g \in L_2(\pi_s). \tag{2.85}$$

Proof: This is straightforward using Theorem 2.1 for (2.84) and (2.85), and Lemma A.1 of Appendix A and its proof for (2.83) and (2.84).

□

Remark 2.10 *Note that*

$$\pi_s \widetilde{I}_{|g|} G_{s,\nu} \widetilde{I}_{|g|} 1 = \mathrm{E}_\nu \Big[\sum_{n=0}^{\tau} |g|(X_0, X_1) |g|(X_n, X_{n+1}) \Big] \ .$$

Remark 2.11 *If g is a special function then by Definition 2.2, Lemma 2.7 and Theorem 2.1, (2.83) is fulfilled.*

Theorem 2.4 *Let $g \in \mathcal{G}_2^d$ and assume that $\|g\|$ satisfies (2.84) and Σ_g is positive definite. If $\{X_n\}$ is null recurrent and satisfies (2.37) then*

$$\left(\frac{1}{\sqrt{T(n)}}\Big\{S_n(g) - T(n)\mu_g\Big\}, \quad \frac{T(n)}{n^\beta L_s(n)}\right) \xrightarrow[n]{d} \left(\Sigma_g^{1/2} Z,\ M_\beta(1)\right) \tag{2.86}$$

where $Z \sim \mathcal{N}(0, I)$ is of dimension d, and where Z is independent of $M_\beta(1)$. A formula for Σ_g is given by (A.14) in Appendix A. The result is true for any initial probability distribution λ. If $\{X_n\}$ is positive recurrent, then (2.86) is true with $\beta = 1$, $L_s(n) \equiv \pi(s)$ and $M_\beta(1) \equiv 1$.

Proof: By Corollary 3.1, Corollary 3.4 and Remark 3.9 we can easily deduce the result.

□

Remark 2.12 *The statement of the normality part of (2.86) does not depend explicitly on the parameter β. Note that the proof of the positive recurrent part does not use the condition (2.37). Indeed, $\beta = 1$ does not belong to the allowable range of values in (2.37).*

Theorem 2.5 *Assume that the conditions in Theorem 2.4 are true and $\pi_s g = 0$. If there exists another atom (s', ν') (cf. (2.4)) then*

$$\Sigma_g' = \pi_{s'}(s)\Sigma_g \tag{2.87}$$

where Σ_g' is defined by (A.14) in terms of (s', ν') . In particular if $\pi_s = \pi_{s'}$ then $\Sigma_g' = \Sigma_g$.

Proof: Let $C \in \mathcal{E}^+$ so that $\pi_s 1_C$ is finite. Let T_C be defined as in (2.82). Then by Lemma 2.6 and Theorem 2.4

$$\frac{S_n(g)}{T_C^{1/2}(n)} = \left\{\frac{T(n)}{T_C(n)}\right\}^{1/2} \frac{S_n(g)}{T^{1/2}(n)} = \pi_s^{-1/2} 1_C \cdot \Sigma_g^{1/2} Z_n + o_p(1)$$

where $\{Z_n\}$ converges in distribution towards $Z \sim \mathcal{N}(0, I)$. Since the left hand side of this equation is independent of the actual atom we get by a corresponding "primed "right hand side that

$$\pi_s^{-1/2} 1_C \cdot \Sigma_g^{1/2} = \pi_{s'}^{-1/2} 1_C \cdot \Sigma_g'^{1/2} .$$

Since $\pi_s = \pi_{s'}^{-1}(s)\pi_{s'}$, the theorem follows.

□

A similar simple transformation formula when $\pi_s g \neq 0$ does not exist.

3 Asymptotics with a smoothing parameter

In this section some asymptotic results are derived for partial sums of $g_h(X_t, \ldots, X_{t+r-1})$ where $h = h_n$ is a design parameter which is assumed to be a function of n, the number of observations. In our applications in the next section h_n is the bandwidth rate in a nonparametric estimation problem. We assume aperiodicity and (2.4).

3.1 Basic notation

We assume that a regeneration sequence $\{\tau_0, \tau_1 \ldots\}$ is defined on the same probability space as the Harris recurrent Markov process $\{X_t\}$ so that $\delta_k = \tau_k - \tau_{k-1}$ for $k \geq 1$ are iid. Let $g_h \in \mathcal{G}_r^d$ and define $\{U_{h,k}\}$ by (2.71) and $U_h \stackrel{\text{def}}{=} U_{g_h}$ by (2.74). The random variables $\{U_{h,k},\ k \geq 1\}$ are $(r-1)$ - dependent. We assume that the pair (s, ν) satisfies (2.4) with a corresponding invariant measure π_s so that $\pi_s g_h = \mathbf{E}_\nu U_{g_h}$. To ease notation we first specialize to the scalar case where $d = 1$. A multivariate extension is given in Corollary 3.4.

The mean $\mu_h \stackrel{\text{def}}{=} \mu_{g_h}$ and the variance parameter, $\sigma_h^2 \stackrel{\text{def}}{=} \sigma_{g_h}^2$ are defined by (2.75) and (2.76) with $d = 1$. We write $\sigma_{U_h}^2 = \mathrm{Var}(U_h)$, $\mu_{C,h} = \mu_h/\pi_s 1_C$ and $\sigma_{C,h}^2 = \sigma_h^2/\pi_s 1_C$. We define $W_h = \sigma_h^{-1}(U_h - \mu_h)$ and in the same way $\{W_{h,k},\ k \geq 1\}$ and $\sigma_{W_h} = \sigma_h^{-1}\sigma_{U_h}$. As in Section 2, let

$$S_n(g_h) = \sum_{j=0}^{n} g_h(X_j, \ldots, X_{j+r-1}) \ .$$

Recall that $u(n) = n^\beta L_s(n)$, $T(n)$ is defined by (2.44) and $T_C(n)$ by (2.82).

We use $c_1, c_2, \ldots$ as a sequence of generic constants in our proofs. All primed versions of U, W, etc. refers to $g_h' = |g_h|$ in the scalar case. If $g_h \in \mathcal{G}_r^d$ with components $g_{i,h}$, then g_h' has components $|g_{i,h}|$, and as before $\|g_h\|$ is the standard Euclidean vector norm of g_h in R^d.

Finally if $\{a_n\}$ and $\{b_n\}$ are two real-valued strictly positive sequences, then we write $a_n \ll b_n$ if $a_n = o(b_n)$.

3.2 Basic conditions

Subsets of the following conditions will be used. They are of general nature, but it will be shown in the next section how they can be used to obtain specific and explicit conditions in the asymptotics of nonparametric estimation. The conditions are stated for $d = 1$; a modification to $d > 1$ is given in Corollary 3.4.

R_0: The expectation $\mu_{|g_h|}$ and the variance $\sigma_{U_{|g_h|}}^2$ are finite for each h.

R_1: $\varliminf_{h \downarrow 0} h\sigma_{g_h}^2 > 0$.

R_1': $\overline{\lim_{h\downarrow 0}}\, h\sigma^2_{U_{|g_h|}} = 0$.

R_2: There exists finite numbers μ and μ' so that
$\mu_{g_h} = \mu + o(1), \quad \mu_{|g_h|} = \mu' + o(1)$ as $h \downarrow 0$.

R_2': $h^{1/2}\mu_{|g_h|} = o(1)$ as $h \downarrow 0$.

R_2'': $h\mu_{|g_h|} = O(1)$ as $h \downarrow 0$.

R_2''': R_2 holds and $\mu_{g_h} = \mu + O(h^2)$.

R_3: $\mathbf{E}_\nu |U_{g_h} - \mu_{g_h}|^{2m} \le d_m h^{-2m+v}$ for some $m \ge 1$, $d_m > 0$ and for a $v \in \{0,1\}$.

R_4: $\sigma^{-1}_{g_h}\sigma_{|g_h|} = O(1)$.

R_5: $\sigma^{-1}_{g_h}\sigma_{U_{g_h}} = O(1)$.

R_6: $|g_h|$ satisfies R_1, R_3 and R_5.

R_7: The Markov process satisfies the tail condition (2.37), or it is positive recurrent.

R_7': $\beta \stackrel{\mathrm{def}}{=} \sup\{p \ge 0\colon \mathbf{E}_\alpha S_\alpha^p < \infty\} > 0$.

R_8: For some $\epsilon > 0$, $m > 1$, $v \in \{0,1\}$, $h_n^{-1} \ll n^{\beta\delta_m - \epsilon}$, $\delta_m = (m-v)^{-1}(m-1)$, and where β is determined by an appropriate tail condition on the distribution of S_α.

R_8': The sequence $\{h_n\}$ satisfies R_8 with $\delta_m = (2m-v)^{-1}(m-2)$ and $m > 2$.

R_8'': $h_n^{-1} = o(u(n))$ where u is defined in (2.43).

R_9: There exists a function g_0 so that $|g_h| \le h^{-1}g_0, \quad \mathbf{P}_x(U_{g_0} < \infty) \equiv 1$.

R_9': If $\{h_n\}$ satisfies R_8'' then $h_n U'_{h_n,0} = O_P(1)$ where $U'_{h,0} = \sum_{j=0}^{\tau} |g_h(X_j, \ldots, X_{j+r-1})|$.

R_{10}: The set C in T_C is special.

Remark 3.1 *Since $\epsilon > 0$ in R_8, this condition implies R_8''.*

Remark 3.2 *From the definitions we have*

$$(2r-1)^{-1}\sigma^2_{g_h} \le \sigma^2_{U_{g_h}} \le \sigma^2_{U_{|g_h|}} + \mu^2_{|g_h|}, \quad \sigma^2_{|g_h|} \ge \sigma^2_{U_{|g_h|}} - (2r-1)\mu^2_{|g_h|}$$

and

$$\sigma^2_{|g_h|} \ge \sigma^2_{g_h} - (2r-1)\mu^2_{|g_h|}\,.$$

Remark 3.3 *We have*

$$\sigma^{-2}_{g_h}\sigma^2_{U_{g_h}} \le \sigma^{-2}_{g_h}\{\sigma^2_{|g_h|} + (2r-1)\mu^2_{|g_h|}\}$$

which shows that R_1, R_2' and R_4 imply R_5.

Remark 3.4 *If both $\lim_{h\downarrow 0} h\sigma^2_{g_h}$ and $\lim_{h\downarrow 0} h\sigma^2_{|g_h|}$ exist as finite positive numbers, then R_1 and R_4 are fulfilled. If in addition R_2' holds, then also R_5 is true. An important example results if*

$$h\sigma^2_{|g_h|} = h\pi_s g_h^2 + o(1) = h\sigma^2_{g_h} + o(1) \tag{3.1}$$

and $\lim_{h\downarrow 0} h\sigma^2_{|g_h|}$ exists as a finite positive number.

Remark 3.5 *If* R_2' *holds then,*

$$h\sigma_{g_h}^2 = \sum_{k=1}^{2r-1} \mathrm{E}_\nu\Big[U_{h,r}U_{h,k}\Big] + o(1), \tag{3.2}$$

and if R_5 *and* R_{10} *hold then* $h\sigma_{g_h-\mu_{C,h}1_C}^2 = h\sigma_{g_h}^2 + o(1)$.

Remark 3.6 *If* R_1 *and* R_3 *hold*

$$\mathrm{E}_\nu(W_h^{2m}) = (h\sigma_h^2)^{-m}h^m\mathrm{E}_\nu(U_h-\mu_h)^{2m} \le d_m' h^{-m+v}\ .$$

If R_2' *holds, then* R_3 *is equivalent with* $\mathrm{E}_\nu(U_h^{2m}) \le d_m'' h^{-2m+v}$.

Remark 3.7 *Application of* R_{10} *requires that one is able to specify a special set for a given process* $\{X_t\}$. *However, note that (cf. Nummelin 1984, pp. 15, 80) compact sets are small, and hence special, under quite wide assumptions. For Example* 2.1 *it suffices that* ζ *is equivalent with the Lebesgue measure.*

3.3 Properties of the number of regenerations

In Section 2.7 we alluded to the link between the number of regenerations $T(n)$ and the total number of observations n. The next lemma shows that the tail condition (2.37) gives the connection between the two.

Lemma 3.1 *If* R_7 *holds, then* $n^{\beta-\epsilon} \ll T(n) \ll n^{\beta+\epsilon}$ a.s. *for all* $\epsilon > 0$. *This is also true for* $T_C(n)$. *The lower bound requires only* R_7'.

<u>Proof of Lemma 3.1:</u> By Remark 2.6 it follows that R_7 implies R_7'. Let $\epsilon \in (0,\beta)$ and define $p = \beta - \epsilon/2$. By R_7', $\mathrm{E}_\alpha S_\alpha^p < \infty$ for all $\epsilon > 0$ which again entails that $k^{-1/p}S_\beta(k) \to 0$ a.s. (cf. Chow & Teicher, 1988, p. 125) where $S_\beta(k) \stackrel{\text{def}}{=} \sum_{j=1}^k(\tau_j - \tau_{j-1})$. 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 relation, when $\tau_0 = 0$,

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

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 R_7. Then, for all $\eta > 0$, we have by the Markov inequality and the convergence of all moments of $T(n)/u(n)$ given by (2.53) that

$$\mathbf{P}_\alpha\Big(T(n) \ge \eta n^{\beta+\epsilon}\Big) \le \eta^{-m}\{u(n)/n^{\beta+\epsilon}\}^m\mathrm{E}\Big\{T(n)/u(n)\Big\}^m \le C_m n^{-m\epsilon/2}, \quad m \ge 1\ .$$

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

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

□

Remark 3.8 *Lemma* 3.1 *shows that* $\widehat{\beta} \stackrel{\text{def}}{=} \ln\{T_C(n)\}/\ln(n)$ *is a strongly consistent estimator for* β *if* R_7 *is fulfilled. Due to the slow convergence rate it is of limited practical use.*

We need to evaluate the difference in growth of $T_C(n)/\pi_s(C)$ and $T(n)$.

Lemma 3.2 *Assume* R_{10}. *Then for all* $p \in (1/2, 1)$ *we have*

$$T^{1-p}(n)\Big\{T^{-1}(n)T_C(n) - \pi_s(C)\Big\} = o(1) \quad \text{a.s.} \tag{3.3}$$

and for all $\epsilon > 0$,

$$P\Big(\pi_s^{-1}(C)T_C(n) \notin [T(n) \pm \epsilon T^p(n)] \quad \text{i.o.}\Big) = 0 \ . \tag{3.4}$$

<u>Proof of Lemma 3.2:</u> Without loss of generality we may assume that $\pi_s(C) = 1$. Define U_k with $g = 1_C$ in Lemma 2.4. Then

$$T_C(n) = S_n(1_C) = U_0 + V_{T(n)} + U_{(n)} \ .$$

By a result of Marcinkiewicz-Zygmund (Chow & Teicher, p. 125), $n^{-p}(V_n - n) \to 0$ a.s. By Lemma 2.5 we have $T^{-p}(n)(V_{T(n)} - T(n)) = o(1)$ a.s. The correction terms U_0 and $U_{(n)}$ can be ignored since $n^{-p}(V_{n+1} - V_n) = o(1)$ a.s.. Hence $T^{-p}(n)(T_C(n) - T(n)) = o(1)$ a.s., which implies both (3.3) and (3.4). □

3.4 Weak limits

Following (2.52) 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 $\mathcal{D}^d[0,\infty)$ is written $\mathcal{L}_{\mathcal{D}^d[0,\infty)}$ (cf. Appendix B).

We can now state our main theorem and three corollaries, whose proofs are given at the end of this sub-section.

Theorem 3.1 *Assume that* R_0, R_1, R_2'', R_3 *with* $m \geq 2$ *and* $v \in \{0,1\}$, R_4-R_8 *and* R_9' *hold. Then, with*

$$\Delta_{n,h_n}(t) = u^{-1/2}(n)\sigma_{h_n}^{-1}(S_{[nt]}(g_{h_n}) - \mu_{h_n}T([nt])), \tag{3.5}$$

$$\Big(\Delta_{n,h_n},\ T_n\Big) \xrightarrow[n]{\mathcal{L}_{\mathcal{D}^2[0,\infty)}} \Big(B \circ M_\beta, M_\beta\Big), \quad B \text{ and } M_\beta \text{ are independent.} \tag{3.6}$$

Corollary 3.1 *Assume that $g_h \equiv g$. If* R_0 *and* R_7 *hold and $\sigma_g^2 > 0$, $\sigma_{|g|}^2 > 0$, then* (3.6) *holds.*

Remark 3.9 *Corollary* 3.1 *implies the scalar version of Theorem* 2.4 *since* (2.84) *implies* R_0.

Definition 3.1 *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.*

Corollary 3.2 *Assume all conditions in Theorem* 3.1 *are true and that* R_2'' *is strengthened to* R_2' *and in addition* R_{10} *holds. Then, with $\Delta_{C,n,h_n}(t) = u^{-1/2}(n)\sigma_{h_n}^{-1}$ $(S_{[nt]}(g_{h_n}) - \mu_{C,h_n}T_C([nt]))$, the sequences $\{\left(\Delta_{n,h_n}, T_n\right)\}$ and $\{\left(\Delta_{C,n,h_n}, T_{C,n}/\pi_s 1_C\right)\}$ are equivalent.*

The next corollary is used in the proof of Theorem 4.4.

Corollary 3.3 *If* R_0, R_1', R_2', R_7, R_8'' *and* R_9' *hold, then*

$$u^{-1/2}(n){h_n}^{1/2}(S_{[nt]}(g_{h_n}) - \mu_{h_n}T([nt]))$$

is negligible. If also R_{10} *holds, we can replace $\mu_{h_n}T([nt])$ by $\mu_{C,h_n}T_C([nt])$.*

Remark 3.10 *In general we have that $\Delta_{C,n,h_n} = u^{-1/2}(n)\sigma_{h_n}^{-1}S_{[nt]}(g_h - \mu_{C,h}1_C)$ and $\pi_s(g_h - \mu_{C,h}1_C) = 0$.*

The proof of Theorem 3.1 consists of several lemmas. Let the sequence $\{h_n\}$ be chosen according to R_8. Define $q_n = h_{u^{(-1)}(n)}$ so that $h_n = q_{u(n)}$. Then by R_8, $q_n^{-1} \ll n^{\delta_m}$.

By Lemma 2.4 we can write

$$S_n(g_h) = U_{h,0} + V_{T(n),h} + U_{h,(n)} \tag{3.7}$$

where $|U_{h,(n)}| \le U'_{h,(n)} \le U'_{h,T(n)+1}$ and $V_{n,h} = \sum_{k=1}^n U_{h,k}$.

The array $\{U_{h,k}, 1 \le k \le [nt]\}$ fulfills a functional CLT when we neglect the dependence on the sequence of stochastic stopping times τ_k. This is not a surprise since the sequence is $(r-1)$ - dependent.

Lemma 3.3 *Assume that* R_0, R_1, R_3 *with* $m \geq 2$, R_5 *hold and that* $q_n^{-1} \ll n^{\delta_m}$, $\delta_m = (m-v)^{-1}(m-1)$. *Then* $Q_{n,q_n} \xrightarrow[n]{\mathcal{L}_{\mathcal{D}[0,\infty)}} B$ *where*

$$Q_{n,h}(t) = n^{-1/2} \sum_{k=1}^{[nt]} W_{h,k}, \quad t \geq 0 . \tag{3.8}$$

<u>Proof of Lemma 3.3:</u> The proof is based upon an ordinary mixing array CLT and a tightness argument. Let t be fixed. By the definition of W and R_1, R_3 with $m \geq 2$, Remark 3.6 and the condition on the rate of $\{q_n\}$,we have

$$n^{-m} \sum_{k=1}^{[nt]} \mathrm{E}(W_{q_n,k}^{2m}) \leq d'_m t\, n^{-(m-1)} q_n^{-(m-v)} = o(1) . \tag{3.9}$$

Thus the array satisfies a Liapounov condition. Since $\mathrm{E}(Q_{n,q_n}^2(t)) = t(1+o(1))$ we have by an appropriate CLT (Bergstrøm, 1981, Th. 1, p. 161) that $Q_{n,q_n}(t) \xrightarrow[n]{d} \mathcal{N}(0,t)$ 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, pp. 68-69), we find that

$$Q_{n,q_n} \xrightarrow[n]{\mathrm{fd}} B . \tag{3.10}$$

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

$$\chi_{h,2k-j}^{i} = \delta_{ij} \eta_h W_{h,2k-i}, \quad 0 \leq i,j \leq 1, \quad k \geq 1, \tag{3.11}$$

where $\eta_h = 2^{1/2} \sigma_{W_h}^{-1}$ and δ_{ij} is the Kronecker symbol. Note that $\eta_h = O(1)$ by R_5.

Next, define

$$Q_n^i(t) \stackrel{\mathrm{def}}{=} n^{-1/2} \sum_{k=1}^{[nt]} \chi_{q_n,k}^{i} = \eta_{q_n} n^{-1/2} \sum_{k=1}^{[(nt+i)/2]} W_{q_n,2k-i}, \quad i = 0,1 .$$

Now, the marginal arrays $\{\chi_{q_n,k}^{i}\}$ consist of independent variables. By (3.9) and an ordinary multivariate CLT we have that (3.10) is fulfilled for Q_n^i. If $\psi_n^i(s,t) \stackrel{\mathrm{def}}{=} \mathrm{E}\{Q_n^i(t) - Q_n^i(s)\}^2$ satisfies, for all fixed $t_0 > 0$,

$$\lim_{\delta \downarrow 0} \overline{\lim_{n}} \sup_{\substack{s \leq t_0 \\ |t-s| \leq \delta}} \psi_n^i(s,t) = 0 \tag{3.12}$$

then $Q_n^i \xrightarrow[n]{\mathcal{L}_{\mathcal{D}[0,\infty)}} B$, $i = 0,1$ by standard theory (Pollard, 1984, Th. 19, p. 104). Now, using R_5 and $|t-s| \leq \delta$,

$$\psi_n^i(s,t) \leq c_1 n^{-1} \eta_{q_n}^2 (n|t-s|+2) \mathrm{E}_\nu\{W_{q_n}^2\} \leq c_2 |t-s+2/n| \leq c_3(\delta + 2n^{-1}),$$

which shows that (3.12) holds. Hence $\{Q_n^i,\ i = 1,2\}$ are tight in $\mathcal{D}[0,\infty)$. Since B is a continuous process and $Q_{n,q_n} = \eta_{q_n}^{-1} \sum_{i=0}^{1} Q_n^i$ and η_h^{-1} is bounded, by R_5 we can conclude that $\{Q_{n,q_n}\}$ is tight (cf. Jacod & Shiryaev, 1987, p. 317). □

The next lemma takes care of the edge terms.

Lemma 3.4 *Assume* R_0, R_1, R_2'', R_4, R_6 *with* $m \geq 2$ *in* R_3, R_7, R_8 *and* R_9' *are true. Then* $\{\Delta_{n,h_n}\}$, *defined in* (3.5), *is equivalent to* $\{Z_{n,h_n}\}$ *where*

$$Z_{n,h}(t) = u^{-1/2}(n) \sum_{k=1}^{T([nt])} W_{h,k} \ . \tag{3.13}$$

Proof of Lemma 3.4: By (3.7) we have with $U_{h,0}(t) = \sum_{j=0}^{\tau_0 \wedge [nt]} g_h(X_j, \ldots, X_{j+r-1})$, $U_{h,(n)}(t) = U_{h,([nt])}$ that it is enough to show that

$$\begin{aligned} \delta_{h,n}(t) &\stackrel{\text{def}}{=} u^{-1/2}(n)\sigma_h^{-1}\Big\{U_{h,0}(t) + U_{h,(n)}(t)\Big\} \\ &= u^{-1/2}(n)\sigma_h^{-1}U_{h,0}(t) + u^{-1/2}(n)\sigma_h^{-1}U_{h,(n)}(t) \end{aligned} \tag{3.14}$$

is negligible. By R_1, R_8 and R_9'

$$|u^{-1/2}(n)\sigma_{h_n}^{-1}U_{h_n,0}(t)| \leq \{h_n u(n) h_n \sigma_{h_n}^2\}^{-1/2} h_n U_{g_h}' = o_P(1) \tag{3.15}$$

independent of t. Hence we can neglect this term. Assume now that g_h is replaced by $|g_h|$ in the definition of $U_{h,(n)}$ in the primed versions. Then since $T_n([nt]) = T_n(t)u(n)$,

$$\begin{aligned} |U_{h,(n)}(t)| \leq U_{h,T([nt])+1}' &= \sigma_h' W_{h,T([nt])+1}' + \mu_h' \\ &= \sigma_h' u^{1/2}(n)\Big\{Q_{u(n),h}'(T_n(t) + 1/u(n)) - Q_{u(n),h}'(T_n(t))\Big\} + \mu_h' \ . \end{aligned} \tag{3.16}$$

The tail condition R_7 guarantees that T_n has a specified asymptotic distribution. By the continuous mapping theorem, R_6, R_8 and Lemma 3.3, $\xi_n(t) \stackrel{\text{def}}{=} Q_{n,q_n}'(t + 1/n) - Q_{n,q_n}'(t)$ is negligible. Again by the continuous mapping teorem 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

$$u^{-1/2}(n)\sigma_{h_n}^{-1}|U_{h_n,(n)}(\cdot)| \leq \{\sigma_{h_n}^{-1}\sigma_{h_n}'\}(\xi_{u(n)} \circ T_n) + \{u(n)h_n\sigma_{h_n}^2 h_n\}^{-1/2} h_n \mu_{h_n}' \xrightarrow[n]{\mathcal{L}_{\mathcal{D}[0,\infty)}} 0 \tag{3.17}$$

due to R_1, R_2'', R_4 and R_8. □

We wish to be able to replace T_n by $T_{C,n}$, which is a function of the original chain $\{X_t\}$.

Lemma 3.5 *Assume that* R_7 *is true. Then* $T_{C,n}/\pi_s 1_C$ *is equivalent with* T_n.

Proof of Lemma 3.5: Assume without loss of generality that $\pi_s(C) = 1$. It is enough to prove that $\sup_{t \leq t_0} \xi_n(t) = o_p(1)$ where $\xi_n(t) \stackrel{\text{def}}{=} |T_{n,C}(t) - T_n(t)|$ for all t_0. Let $0 < \delta < 1$ and $n_\delta = n^{-(1-\delta)}$. We have $\sup_{t \leq n_\delta} \xi_n(t) \leq u^{-1}(n)\Big\{T_C(n^\delta) + T(n^\delta)\Big\} = o_P(1)$ and when $t \geq n_\delta$,

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

since $T_C(n)/T(n)$ converges to $\pi_s 1_C$ with probability one by Remark 2.7. □

Proof of Theorem 3.1: By Lemma 3.4 we can neglect the edge terms and it is enough to consider $(Z_{n,h_n}, T_n) = (Q_{u(n),q_{u(n)}} \circ T_n, T_n)$. The proof is completed by Lemma 3.3, Theorem 2.3 and Theorem B.1 in Appendix B with $(B_n, A_n) = (Q_{u(n),q_{u(n)}}, S_{\beta,n})$, $A^{(-1)} = M_\beta$ in that theorem. The process $S_{\beta,n}$ is defined in Section 2.7 by (2.47). By the proof of Theorem 2.3 we have that $S_{\beta,n}^{(-1)}$ and T_n are equivalent stochastic processes. □

Proof of Corollary 3.1: We proceed to the simplified situation in which there is no bandwidth involved and the U_k's are independent of n and strictly stationary. The existence of the fourth order moment in the proof of Lemma 3.3 is superfluous, Lemma 3.4 is still true, and hence the conclusion is true.

□

Proof of Corollary 3.2: To prove the equivalence, note that

$$\begin{aligned}\Delta_{C,n,h_n}(t) - \Delta_{n,h_n}(t) &= \sigma_{h_n}^{-1}\mu_{h_n} u^{-1/2}(n)\Big\{T_C([nt])\pi_s^{-1}(C) - T([nt])\Big\} \\ &= \{h_n\sigma_{h_n}^2\}^{-1/2} h_n^{1/2}\mu_{h_n}\sigma\Delta_n(t),\end{aligned}$$

where $\Delta_n(t) \stackrel{\text{def}}{=} u^{-1/2}(n)\sigma^{-1}\Big\{S_{[nt]}(1_C/\pi_s 1_C) - T([nt])\Big\}$ and σ^2 is defined by (2.76) with $d = 1$ and with $g = 1_C/\pi_s 1_C$. By R_{10}, R_0 is fulfilled for $g = 1_C/\pi_s 1_C$. By the previous corollary $\Delta_n(t) \xrightarrow[n]{\mathcal{L}_{\mathcal{D}[0,\infty)}} B \circ M_\beta$. Since $h_n^{1/2}\mu_{h_n} = o(1)$ by R_2', it follows that $\Delta_{C,n,h_n} - \Delta_{n,h_n} \xrightarrow[n]{\mathcal{L}_{\mathcal{D}[0,\infty)}} 0$. By Lemma 3.5, we have that $T_{C,n}/\pi_s(C)$ and T_n are equivalent. □

Proof of Corollary 3.3: Let

$$\widetilde{\Delta}_{n,h_n} = u^{-1/2}(n) {h_n}^{1/2}(S_{[nt]}(g_{h_n}) - \mu_{h_n} T([nt]))\,.$$

First assume that $r = 1$, i.e., independence. Choose $\{h_n\}$ according to R_8'', let $h_n = q_{u(n)}$ and analogous to (3.8),

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

Then

$$\mathrm{E}\{\widetilde{Q}_n(t)\}^2 = n^{-1}q_n[nt]\sigma_{U_{g_{q_n}}}^2 \le t\Big[q_n\sigma_{U_{|g_{q_n}|}}^2 + q_n\mu_{|g_{q_n}|}^2\Big],$$

and using R_1' and R_2',

$$\sup_{t\le K}\mathrm{E}\{\widetilde{Q}_n(t)\}^2 = o(1)\,. \tag{3.18}$$

Since $\widetilde{Q}_n$ is a normalized sum of independent variables we can conclude that $\{\widetilde{Q}\}$ is negligible (cf. Pollard, 1984, p 104). This is also true when we replace g_h by $|g_h|$.

The edge effects are tackled in a similar way as in the proof of Lemma 3.4. We have

$$\delta_{h_n,n}(t) = u^{-1/2}(n)h_n^{1/2}U_{h_n,0}(t) + u^{-1/2}(n)h_n^{1/2}U_{h_n,(n)}(t)\,.$$

The first term is of order $o_P(1)$ uniformly with respect to t using R_8'' and R_9'. For the second term

$$\begin{aligned} u^{-1/2}(n)h_n^{1/2}|U_{h,(n)}(t)| &\leq u^{-1/2}(n)h_n^{1/2}\Big\{(U'_{h,T([nt])+1} - \mu'_h) + \mu'_h\Big\} \\ &= \Big\{\widetilde{Q}'_{u(n),h}(T_n(t) + 1/u(n)) - \widetilde{Q}'_{u(n),h}(T_n(t))\Big\} + u^{-1/2}(n)h_n^{1/2}\mu'_h \\ &\leq \xi_{u(n)}(T_n(t)) + u^{-1/2}(n)h_n^{1/2}\mu'_h, \end{aligned}$$

where $\xi_n(t) \stackrel{\text{def}}{=} |\widetilde{Q}'_{n,q_n}(t + 1/n)| + |\widetilde{Q}'_{n,q_n}(t)|$. By the preceding calculations we can conclude that

$$\sup_{t \leq t_0} \xi_n(t) = o_P(1) \ .$$

Thus the process $\xi_n \circ T_n$ is negligible (cf. the proof of Theorem 3.1), and we have that the edge effects can be ignored.

Hence it is sufficient to look at

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

since $\widetilde{\Delta}_{n,h_n}$ is equivalent to $\widetilde{Z}_{n,h_n}$. But

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

and $\widetilde{Q}_{u(n),q_n} \circ T_n$ is easily shown to be negligible from the first part ot the proof.

If $r > 1$ then we can write $\widetilde{Q}$ as a finite sum of sums of independent variables and use Remark 3.2.

□

Theorem 3.2 *Assume that* R_0, R_1, R_2', R_3 *with* $m \geq 2$ *and* $v \in \{0, 1\}$, R_4-R_8, R_9' *and* R_{10} *are true. Then*

$$T_C^{1/2}(n)\sigma_{C,h_n}^{-1}\Big\{T_C^{-1}(n)S_n(g_h) - \mu_{C,h_n}\Big\} \xrightarrow[n]{d} \mathcal{N}(0,1) \ . \tag{3.19}$$

If R_2''' *holds and* $h_n^{-1} \gg n^{\beta/5+\epsilon}$, *then* μ_{C,h_n} *can be replaced by* $\mu_C = \mu/\pi_s 1_C$.

<u>Proof of Theorem 3.2:</u> Denote the left hand side of (3.19) by Λ_{n,h_n}. We have that $\Lambda_{n,h_n} = (\pi_s 1_C)^{1/2}\Big\{T_C(n)/u(n)\Big\}^{-1/2}\Delta_{C,n,h_n}(1)$. By Lemma 3.2, Theorem 3.1 and Corollary 3.2 the main portion of the statements follows. Assume that R_2''' holds then

$$T_C^{1/2}(n)h_n^{1/2}|\mu_{C,h_n} - \mu_C| \leq c_1\{T_C(n)h_n^5\}^{1/2} = o_P(1) \ .$$

□

A multivariate extension of Theorem 3.1 and Theorem 3.2 is also useful. We use the notation $\Sigma_{c,h} = \Sigma_{g_h}/\pi_s 1_C$.

Corollary 3.4 *Let $g_h \in \mathcal{G}_r^d$. Assume that $\|g_h\|$ and the process $\{X_t\}$ satisfy* R_0, R_1, R_2', R_3 *with $m \geq 2$ and $v \in \{0,1\}$,* R_4-R_8, R_9', R_{10}. *In addition we assume that*

$$\underline{\lim}\, h\Sigma_{g_h} \geq \Sigma_0 \tag{3.20}$$

for some positive definite matrix Σ_0 and

$$\|\Sigma_{g_h}\|^{-1}\sigma^2_{\|g_h\|} = O(1)\ . \tag{3.21}$$

Then (3.19) *holds with σ^{-1}_{C,h_h} replaced by $\Sigma^{-1/2}_{C,h_n}$, and the limit is the d-dimensional standard multivariate normal distribution. Moreover, a multivariate extension of* (3.6) *holds where B is replaced by a d-dimensional Brownian motion.*

If R_2''' *holds and $h_n^{-1} \gg n^{\beta/5+\epsilon}$, then μ_{C,h_n} can be replaced by μ_C.*

Proof of Corollary 3.4: Analogous to (3.5) and (3.14) we define

$$\Delta_{n,g_{h_n}}(t) = u^{-1/2}(n)\Sigma^{-1/2}_{g_{h_n}}(S_{[nt]}(g_{h_n}) - \mu_{h_n}T([nt])), \tag{3.22}$$

and

$$\delta_{g_h,n}(t) \stackrel{\text{def}}{=} u^{-1/2}(n)\Sigma^{-1/2}_{g_h}\Big\{U_{g_h,0}(t) + U_{g_h,(n)}(t)\Big\}\ . \tag{3.23}$$

Now since $|g_{i,h}| \leq \|g_h\|$ we have $|U_{g_{i,h}}| \leq U_{\|g_h\|}$ for $i = 1,\ldots,d$. Using this inequality we find that $\|U_{g_h}\| \leq d^{1/2}U_{\|g_h\|}$ and

$$\|\delta_{g_h,n}(t)\| \leq d^{1/2}\|\Sigma_{g_h}\|^{-1/2}\sigma_{\|g_h\|}\delta_{\|g_h\|,n}(t)\ . \tag{3.24}$$

From the conditions assumed in the corollary it follows that the conditions in Theorem 3.1 for $\|g_h\|$ are satisfied. In particular this implies that $\delta_{\|g_{h_n}\|,n}$ is negligible. Together with (3.21), (3.23) and (3.24) this implies that $\|\delta_{g_{h_n},n}\|$ is negligible.

We have thus shown that $\Delta_{n,g_{h_n}}$ is equivalent with $Z_{n,g_{h_n}}$ defined by

$$\begin{aligned} Z_{n,g_h}(t) &= u^{-1/2}(n)\sum_{k=1}^{T([nt])} W_{g_h,k}, \quad W_{g_h,k} \stackrel{\text{def}}{=} \Sigma^{-1/2}_{g_h}\Big(U_{g_h,k} - \mu_{g_h}\Big) \\ &= Q_{u(n),q_{u(n)}}(T_n(t)) \end{aligned}$$

where Q_{n,q_n} is defined by (3.8). It is enough to prove that

$$a'Q_{n,q_n} \xrightarrow[n]{\mathcal{L}_{\mathcal{D}[0,\infty)}} a'B, \quad \forall a \in R^d, \tag{3.25}$$

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

$$\mu_{f_h} = b_h'\mu_{g_h}, \quad \sigma^2_{f_h} = h^{-1}\|a\|^2\ .$$

Moreover,

$$\begin{aligned} a'Q_{n,q_n} &= n^{-1/2}\sum_{k=1}^{[nt]} h^{1/2}h^{-1/2}a'\Sigma^{-1/2}_{g_h}\Big(U_{g_h,k} - \mu_{g_h}\Big) \\ &= \|a\|n^{-1/2}\sum_{k=1}^{[nt]}\sigma^{-1}_{f_h}\Big(U_{f_h,k} - \mu_{f_h}\Big) \\ &= \|a\|Q^f_{n,q_n}, \quad \text{say,} \end{aligned}$$

where Q^f_{n,q_n} is defined by (3.8) with $W_{h,k} = W_{f_h,k}$. We have by R_1, R_2' and R_3 for $\|g_h\|$ that

$$\begin{aligned}
\mathrm{E}(W^{2m}_{f_h,k}) &= \sigma_{f_h}^{-2m}\mathrm{E}\Big[b_h'(U_{g_h} - \mu_{g_h})\Big]^{2m} \\
&\le \sigma_{f_h}^{-2m}\mathrm{E}\Big[\|b_h\|^{2m}\|U_{g_h} - \mu_{g_h}\|^{2m}\Big] \\
&\le \sigma_{f_h}^{-2m}\|b_h\|^{2m}\mathrm{E}\Big[\|U_{g_h} - \mu_{g_h}\|^{2m}\Big] \\
&\le c_1\sigma_{f_h}^{-2m}\|b_h\|^{2m}\Big[\mathrm{E}\|U_{g_h}\|^{2m} + \|\mu_{g_h}\|^{2m}\Big] \\
&\le c_2\sigma_{f_h}^{-2m}\|b_h\|^{2m}\Big[\mathrm{E}(U^{2m}_{\|g_h\|}) + \mu^{2m}_{\|g_h\|}\Big] \\
&\le c_3\sigma_{f_h}^{-2m}h^{-(2m-v)} \\
&\le c_4 h^m h^{-(2m-v)} \\
&\le c_4 h^{-(m-v)} \ .
\end{aligned}$$

We have also used that $\|b_h\|$ is bounded with respect to h; in fact it follows from (3.20) that $\overline{\lim}\,\|b_h\|^2 \le \|a\|^2\rho$ where ρ denotes the spectral radius of Σ_0^{-1}. Hence the analogue of (3.9) is satisfied and (3.25) is proved using the method of proof of Lemma 3.3. By (3.20), (3.21), R_2' and Remark 3.3 we have that, R_0, R_1, R_2', R_3 with $m \ge 2$, R_4 -R_6 and R_9' are fulfilled for f_h. Hence Theorem 3.2 can be applied to f_h and by a Cramér-Wold argument the first statement of the corollary is proved.

Moreover, we easily get the multivariate extension of Theorem 3.1 claimed in the second statement of the corollary: let B_d denote a d-dimensional Brownian motion. It is enough to prove that

$$\Big(a'\Delta_{n,h_n},\ T_n\Big) \xrightarrow[n]{\mathcal{L}_{\mathcal{D}^2[0,\infty)}} \Big(\|a\|B \circ M_\beta, M_\beta\Big), \quad B \text{ and } M_\beta \text{ are independent}, \tag{3.26}$$

since $a'B_d$ has the same distribution as $\|a\|B$. But (3.26) follows from (3.25) and the proof of Theorem 3.1.

The last statement of the corollary is proved as in the proof of Theorem 3.2.

□

Remark 3.11 *Proceeding as in Corollary* 3.1 *and Remark* 3.9*, Theorem* 2.4 *can now be derived as a special case of Corollary* 3.4*.*

3.5 Consistency

Weak consistency may be deduced from the asymptotic distributional limit.

Corollary 3.5 *Suppose that all the conditions in Theorem* 3.1 *are fulfilled, that* R_2' *is strengthened to* R_2 *and in addition that* R_{10} *holds. Then*

$$\begin{aligned}
T_C^{-1}(n)S_n(g_{h_n}) &= \mu_{C,h_n} + O_P\Big(T_C^{-1/2}(n)\sigma_{h_n}\Big) \\
&= \mu_C + o_P(1) \ .
\end{aligned} \tag{3.27}$$

Proof: From definitions, $S_n(g_{h_n}) = T_C(n)\mu_{C,h_n} + \Delta_{C,n,h_n}(1)u^{1/2}(n)\sigma_{h_n}$ and the first line of (3.27) follows from Corollary 3.2. The last statement is achieved by R_1, R_8 and Lemma 3.1.

□

It is more difficult to obtain strong consistency.

Theorem 3.3 *Assume that* R_2, R_3 *with* $m > 2$, R_7', R_8', *and* R_9 *hold. Then* $T_C^{-1}(n)S_n(g_{h_n})$ *converges with probability one towards* μ_C.

The proof starts with a lemma.

In order to obtain strong consistency with a deterministic bandwidth we require that the rate is related to the lower bound of the sequence $\{T(n)\}$.

Lemma 3.6 *Assume that* $q_n^{-1} \ll n^{\delta_m - \epsilon}$ *with* δ_m *as in* R_8, *and that* R_3 *holds with an* $m > 2$. *Let*

$$S_{l,p.r} \stackrel{\text{def}}{=} \frac{1}{lp+r-1} \sum_{k=1}^{lp+r-1} U_{q_l,k}, \quad (l,p,r) \in N_+^3, \quad r \in [1,l]\,. \tag{3.28}$$

Then

$$\lim_{|lp|\uparrow\infty} \sup_{1\le r\le l} |S_{l,p.r} - \mu_{q_l}| = o(1) \quad \text{a.s.} \quad . \tag{3.29}$$

Proof of Lemma 3.6: By the Borel-Cantelli lemma it is enough to show that $\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\sum_{r=1}^{l} \mathrm{E}(S_{l,p.r} - \mu_{q_l})^{2m} < \infty$. By $(r-1)$ - dependence and R_3,

$$\begin{aligned}\mathrm{E}(S_{l,p.r} - \mu_{q_l})^{2m} \le c_1(lp+r-1)^{-m}\mathrm{E}(U_{q_l} - \mu_{q_l})^{2m} &\le c_2(lp)^{-m}q_l^{-(2m-v)} \\ &\le c_3 p^{-m} l^{-\{m-(\delta_m-\epsilon)(2m-v)\}}\end{aligned}$$

which gives

$$\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\sum_{r=1}^{l} \mathrm{E}(S_{l,p.r} - \mu_{q_l})^{2m} \le c_4 \sum_{l=1}^{\infty} l^{-\{m-1-(\delta_m-\epsilon)(2m-v)\}} < \infty$$

since $m - 1 - \delta_m(2m - v) \ge 1$. □

Proof of Theorem 3.3: Let $\{h_n\}$ be fixed. Recall that $\mu_{h_n} = \mu_C \pi_s 1_C$. Without loss of generality we may assume that $h_n = q_{u_0(n)}$ where $u_0(n) = [n^{\beta-\epsilon}]$ for an $\epsilon > 0$ and $\{q_n\}$ satisfies the rate inequality given by Lemma 3.6. We use the representation of $S_n(g_{h_n})$ given by (3.7) for each n. By Lemma 3.1, $T(n) \gg u_0(n)$ 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

$$\left|T(n)^{-1}V_{T(n),h_n} - \mu_{h_n}\right| = \left|\frac{1}{lp+r-1}\sum_{k=1}^{lp+r-1} U_{q_l,k} - \mu_{q_l}\right| = |S_{l,p.r} - \mu_{q_l}|\,.$$

By Lemma 3.6 and R_2, $T(n)^{-1}V_{T(n),h_n}$ converges to μ with probability one (cf. the proof of Lemma 2.5). The arguments also show that $T(n)^{-1}V_{T(n)+1,h_n}$ converges to μ with probability one. Hence $U_{h,(n)}$ is negligible. By R_9 we get rid of $U_{h_n,0}$ since

$$|T^{-1}(n)U_{h_n,0}| \leq \{h_n T(n)\}^{-1} U_{g_0} = o(1) \quad \text{a.s.}$$

using that U_{g_0} is finite a.s. and $h_n T(n) \geq n^{-(\beta-\epsilon)}T(n) \uparrow \infty$ a.s. by Lemma 3.1. Since $T_C(n)/T(n) = \pi_s 1_C + o(1)$ a.s., the proof is finished.

□

It turns out that consistency results are easier to obtain with a stochastic bandwidth sequence, although weak limits are considerable harder in this case.

Theorem 3.4 *Assume that* R_2, R_3 *with* $m \geq 2$, R_7' *and* R_9 *are satisfied. If* $q_n^{-1} < n^{\delta_m - \epsilon}$ *where* $\delta_m = (m-1)/(2m-v)$ *then* $T_C^{-1}(n)S_n(g_{H_n})$ *converges almost surely towards* μ_C *where* $H_n = q_{T_C(n)}$

<u>Proof:</u> Neglecting the edge terms, it is enough to show that

$$T_C^{-1}(n)V_{T_C(n),q_{T_C(n)}} \xrightarrow[n]{\text{a.s.}} \mu_C \ .$$

By Lemma 2.5 this convergence is implied by

$$n^{-1}V_{n,q_n} \xrightarrow[n]{\text{a.s.}} \mu_C \ .$$

But the latter is proved by R_3, the Markov inequality and the Borel-Cantelli Lemma since

$$\sum_n n^{-m} q_n^{-2m+v} < \infty \ .$$

□

4 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.

We assume aperiodicity and (2.4). 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 $K_{x,h}(y) = h^{-1}K((y-x)/h)$, $\mathcal{N}_x(h) = \{y\colon K_{x,h}(y) \neq 0\}$ and $\mathcal{N}_x = \mathcal{N}_x(1)$. In our context a locally bounded function will be taken to mean bounded in a neighbourhood of x and a locally continuous function is continuous at the point x. Without loss of generality we may assume that this neighbourhood equals $\mathcal{N}_x$, and that local continuity implies local boundedness. This is so since $\mathcal{N}_x(h) = x \oplus h\mathcal{N}_0$. Again, we use $c_1, c_2, \ldots$ as a sequence of generic constants in our proofs.

The following condition is always assumed.

K_0: The kernel K is non-negative, $\int K(u)du < \infty$ and $\int K^2(u)du < \infty$.

Subsets of the conditions listed below are used according to need, where the point x is fixed.

K_1: The support of the kernel, $\mathcal{N}_0$, is contained in a compact set.

K_2: The kernel is bounded and $\mathcal{N}_x$ is a special set.

K_3: The kernel is normalized so that $\int K(u)du = 1$.

K_4: The kernel satisfies $\int uK(u)du = 0$.

P_1: The invariant measure π_s has a locally bounded density p_s.

P_2: The density p_s is locally continuous.

P_3: The density p_s possesses locally continuous partial derivatives of a given specified order.

P_4: The density is locally strictly positive, i.e., $\underline{\lim}_{y\to x}\, p_s(y) > 0$.

P_5: For all $\{A_h\} \in \mathcal{E}$ so that $A_h \downarrow \emptyset$ when $h \downarrow 0$, $\lim_{h\downarrow 0} \overline{\lim}_{y\to x}\, P(y, A_h) = 0$.

P_6: The tail condition (2.37) holds, or the chain is positive recurrent.

P'_6: $\beta \stackrel{\text{def}}{=} \sup\{p \geq 0\colon \mathrm{E}_\alpha S_\alpha^p < \infty\} \wedge 1 > 0$.

T_1: The set C in T_C is a special set.

Remark 4.1 *The conditions* K_1-K_4 *are regularity conditions for the kernel. The first three P-conditions represent different levels of smoothness for a local density of* π_s. *The last one of these is used to get rid of a bias term. In order to have a strictly positive asymptotic variance,* P_4 *is necessary in some cases. It turns out that a sort of uniform local continuity property for the transition probability is needed in some weak limit theorems. This is expressed by* P_5. *The tail condition* (2.37) *is crucial for weak limits in the null recurrent case.*

To ease notation, as in Section 2.9 and Appendix A, we will restrict ourselves to functions contained in $\mathcal{G}_r^d$ with $r \le 2$, and in the following ξ_0, ξ are arbitrary functions in $\mathcal{G}_2^d$ and $\mathcal{G}^d$, respectively. Our first lemma shows that the condition K_2 in conjunction with local boundedness for a function in one variable ensure finiteness of moments. For functions of two variables the appropriate local boundedness also depends upon global properties; i.e., local boundedness of $\widetilde{\xi}_0$ depends upon global properties of P and ξ_0.

Let

$$g_h(u,w) = K_{x,h}(u)\xi_0(u,w), \quad \xi_0 \in \mathcal{G}_2^d, \tag{4.1}$$

and $S_n(g_h) = \sum_{t=0}^n g_h(X_t, X_{t+1})$. Our discussion is restricted to statistics based on $S_n(g_h)$. Recall that

$$\widetilde{I}_{\xi_0}(y,dz) = P(y,dz)\xi_0(y,z), \quad \widetilde{\xi}_0(y) = \widetilde{I}_{\xi_0}1(y) = \int P(y,dz)\xi_0(y,z)\,. \tag{4.2}$$

As before, when we write $\widetilde{\xi}_0(y)$, the integral in (4.2) is implicitly assumed to exist. Also note that $\|\widetilde{\xi}_0\|^p = \left\{\widetilde{I}_{\|\xi_0\|^2}1\right\}^{p/2}$ when $p \ge 0$. Lemma 2.4 leads to the decomposition of $S_n(g_h)$ in a sum of $U_{h,k}$'s which are strictly stationary, 1-dependent and marginally distributed as

$$U_{g_h} = \sum_{j=0}^{\tau} K_{x,h}(X_j)\xi_0(X_j, X_{j+1}), \tag{4.3}$$

with ν as its initial measure.

If $\xi_0(u,w) = \xi(u)$, i.e., $\xi \in \mathcal{G}^d$, then $\widetilde{\xi}_0 = \xi$.

Lemma 4.1 *Let g_h be given by (4.1) with $g_h \in \mathcal{G}_2$. Assume K_1, K_2, P_1, let $m \ge 1$ and define $\psi_m = \widetilde{I}_{|\xi_0|^{2m}}1$. Assume that ψ_m is locally bounded. If $m \ge 1$, then corresponding to R_3 in Section 3,*

$$\mathrm{E}_\nu U_{|g_h|}^{2m} \le d'_m h^{-2m+v} \tag{4.4}$$

with $v = 1$. Without P_1 the above inequality still holds but with $v = 0$. Moreover, $\mu_{|g_h|} = O(1)$ and $h\sigma^2_{|g_h|} = O(1)$.

<u>Proof of Lemma 4.1:</u> Without any loss of generality we assume that $\xi_0 \in \mathcal{G}_2$ and $\xi_0 \ge 0$. We use the simple inequality $\xi_0^r \le 1 + \xi_0^{2m}$ when $r \le 2m$. To show (4.4), by Theorem A.1 in Appendix A, it is enough to consider

$$J \stackrel{\text{def}}{=} \nu G_{s,\nu}\widetilde{I}_{g_h^{\ell_1}} \cdots G_{s,\nu}\widetilde{I}_{g_h^{\ell_r}}1$$

where $r \le 2m$, $\ell_i \ge 1$ and $\sum \ell_i = 2m$, and where $g \ge 0$ implies that $\check{I}_g \le \widetilde{I}_g$ with $\check{I}_g$ defined by (A.18). Now, by K_1, K_2 and the local boundedness of ψ_m,

$$\widetilde{I}_{g_h^{\ell_r}}1 \le K_{x,h}^{\ell_r}(1+\psi_m) \le c_1 K_{x,h}^{\ell_r} \le c_2 h^{-\ell_r} 1_{\mathcal{N}_x}\,.$$

By K_2, Definition 2.2 and Theorem 2.1 with $m_0 = 1$, the function $G_{s,\nu}1_{\mathcal{N}_x}$ is bounded so that

$$G_{s,\nu}\widetilde{I}_{g_h^{\ell_r}} \leq c_3 h^{-\ell_r} 1 \ .$$

Hence, using K_1 and the fact that $\nu G_{s,\nu} = \pi_s$,

$$J \leq c_4 h^{-\sum_{j=2}^{r} \ell_j} \pi_s K_{x,h}^{\ell_1} \leq c_5 h^{-2m+v}$$

where $v = 1$ by P_1. The proof of the statements concerning $\mu_{|g_h|}$ and $h\sigma^2_{|g_h|}$ is similar, albeit simpler.

□

Next we introduce our nonparametric estimators. If the invariant measure π_s has a density p_s, we define $p_C = p_s/\pi_s 1_C$ which is the normalized density relative to the set C, i.e., $\int_A p_C(x)dx = \pi_C 1_A = \pi_s 1_A/\pi_s 1_C$. The density p_C does not depend upon s as is easily verified from the non-uniqueness property stated in Remark 2.1 of Section 2.3.

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

$$\widehat{p}_C(x) = \widehat{p}_{C,n,h}(x) = T_C^{-1}(n) \sum_{t=0}^{n} K_{x,h}(X_t) \ . \tag{4.5}$$

Let $\xi \in \mathcal{G}^d$. Note that $P\xi(x) = \mathrm{E}\Big[\xi(X_t) \mid X_{t-1} = x\Big]$ is the conditional mean of $\xi(X_t)$ at the point x. A kernel estimator for the conditional mean function $P\xi$ is given by

$$\widehat{P}\xi(x) = \widehat{P}\xi_{n,h}(x) = \Big\{\sum_{t=0}^{n} K_{x,h}(X_t)\Big\}^{-1} \Big\{\sum_{t=0}^{n} \xi(X_{t+1}) K_{x,h}(X_t)\Big\} \ . \tag{4.6}$$

The corresponding conditional variance function is given by

$$V\xi = P(\xi \otimes \xi) - P\xi \otimes P\xi \ . \tag{4.7}$$

If we replace the two terms of the right hand side of (4.7) by estimates defined by (4.6), we obtain a conditional variance estimator $\widehat{V}\xi(x)$. Moreover, (4.6) coincides, and (4.5) essentially coincides, with respectively the Nadaraya - Watson estimator and the kernel estimator in the positive recurrent case, since then $T_C(n)/n \to \pi 1_C$ a.s..

The next lemma is central in establishing the properties of the above estimates. It gives conditions for convergence of μ_{g_h} and Σ_{g_h}. It turns out that the asymptotic variance is independent of μ_{g_h}. (The tensor symbols involved are defined just prior to Lemma A.2 in the Appendix).

Lemma 4.2 *Let $g_h(u,w) = K_{x,h}(u)\xi_0(u,w)$ where $\xi_0 \in \mathcal{G}_2^d$. Assume that* K_1, K_2, P_1, P_2 *are true, $\widetilde{I}_{||\xi_0||^2}1$ is locally bounded and $\widetilde{I}_{\xi_0}1$, $\widetilde{I}_{\xi_0\otimes\xi_0}1$ are locally continuous. Then*

$$\mu_{g_h} = p_s(x)\widetilde{\xi}_0(x)\Big\{\int K(u)du\Big\} + o(1), \tag{4.8}$$

$$\Sigma_{g_h} = p_s(x)\widetilde{I}_{\xi_0\otimes\xi_0}1(x)\Big\{\int K^2(u)du\Big\} + hA_h + hA_h' + o(1), \tag{4.9}$$

where $A_h = \pi_s \widetilde{I}_{g_h} \otimes G_{s,\nu} \widetilde{I}_{g_h} 1$. *If* $hA_h = o(1)$, *then* $h\Sigma_{g_h} = h\Sigma_{U_{g_h}} + o(1)$. *If* $\widetilde{\xi}_0(x) = 0$ *or* P_4 *and* P_5 *are true, then* $hA_h = o(1)$. *If* $\xi_0(u,w) = \xi(w) - P\xi(u)$ *then*

$$h\Sigma_{g_h} = p_s(x) V\xi(x) \int K^2(u)du + o(1) \ . \tag{4.10}$$

If $\xi_0(u,w) = \xi(u)$ *and* $\xi(x) = 0$, *then* $h\Sigma_{g_h} = o(1)$.

<u>Proof:</u> By P_1, P_2, K_1, the continuity of $\widetilde{I}_{\xi_0} 1$ and Bochners Lemma (cf. Wheeden & Zygmund, 1977, Ch. 9), (4.8) easily follows.

By (A.14) we have

$$\Sigma_{g_h} = \pi_s(g_h \otimes g_h) + A_h + A_h' - B_h - B_h'$$

where

$$B_h = \frac{1}{2}(\pi_s g_h \otimes \pi_s g_h) + \pi_s g_h \otimes \pi_s s \cdot g_{h,\nu}$$

with g_ν defined in (A.1) and A_h is expressed above. Again by P_1, P_2, K_1, the continuity of $\widetilde{I}_{\xi_0} 1$, and Bochner's Lemma,

$$\pi_s g_h = \pi_s I_{K_{x,h}} \widetilde{\xi}_0 = p_s(x) \widetilde{\xi}_0(x) \int K(u)du + o(1) \ .$$

The term $\pi_s s \cdot g_{h,\nu}$ can be treated likewise and hence $hB_h = o(1)$. 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 (4.9) follows.

By the multivariate extension of (A.6) and (A.9) we have that

$$\Sigma_{U_{g_h}} = \pi_s(g_h \otimes g_h) + A_h + A_h' - (1 \otimes \mu_{g_h}) - (1 \otimes \mu_{g_h})' - \mu_{g_h} \otimes \mu_{g_h} \ .$$

If $hA_h = o(1)$, then by (4.8) and (4.9) it follows that

$$h\Sigma_{U_{g_h}} = p_s(x) V\xi(x) \int K^2(u)du + o(1)$$

so that $h\Sigma_{g_h} = h\Sigma_{U_{g_h}} + o(1)$.

We can write

$$hA_h = \pi_s I_{K_{x,h}} \widetilde{I}_{\xi_0} \otimes \psi_h, \quad \psi_h \stackrel{\text{def}}{=} G_{s,\nu} h I_{K_{x,h}} \widetilde{\xi}_0 \ .$$

By K_2, Theorem 2.1 with $m_0 = 1$ and the local boundedness of $\widetilde{I}_{||\xi_0||}$, we have $\|\psi_h\| \le c_1$ and it follows that $\|hA_h\| = O(1)$. If $\widetilde{\xi}_0(x) = 0$, then $\sup \|\psi_h\| = o(1)$ since $\sup_{y \in \mathcal{N}_x(h)} \|\widetilde{\xi}_0\|(y) = o(1)$ by the local continuity of $\widetilde{\xi}_0$, and hence $\|hA_h\| = o(1)$.

Assume that P_4 and P_5 are true. Then, since $\widetilde{I}_{||\xi_0||} 1$ is locally bounded and since K_2 holds,

$$\|G_{s,\nu} h I_{K_{x,h}} \widetilde{\xi}_0(z)\| \le G_{s,\nu} c_2 1_{\mathcal{N}_x(h)}(z) = c_2 G_{s,\nu}(z, \mathcal{N}_x(h)) \ .$$

This gives

$$\|hA_h\| \le c_3 \int \pi_s(dy) K_{x,h}(y) \int P(y,dz) \|\xi_0(y,z)\| G_{s,\nu}(z, \mathcal{N}_x(h))$$

$$= c_3 \int p_s(y) K_{x,h}(y) \Big\{ \int P(y,dz) \|\xi_0(y,z)\| G_{s,\nu}(z, \mathcal{N}_x(h)) \Big\} dy,$$

and by the Cauchy Schwartz inequality,

$$\le c_4 \int p_s(y) K_{x,h}(y) \Big\{ \int P(y,dz) \|\xi_0(y,z)\|^2 \Big\}^{1/2}$$

$$\Big\{ \int P(y,dz) \big[G_{s,\nu}(z, \mathcal{N}_x(h)) \big]^2 \Big\}^{1/2} dy$$

$$= c_4 \int p_s(y) K_{x,h}(y) \big\{ \widetilde{I}_{\|\xi_0\|^2} 1(y) \big\}^{1/2} \Big\{ \int P(y,dz) \big[G_{s,\nu}(z, \mathcal{N}_x(h)) \big]^2 \Big\}^{1/2} dy$$

$$= c_4 \int p_s(x+hu) K(u) \big\{ \widetilde{I}_{\|\xi_0\|^2} 1(x+hu) \big\}^{1/2}$$

$$\Big\{ \int P(x+hu,dz) \big[G_{s,\nu}(z, \mathcal{N}_x(h)) \big]^2 \Big\}^{1/2} du$$

and again by the Cauchy Schwartz inequality, P_1, and the local boundedness of $\widetilde{I}_{\|\xi_0\|^2}$,

$$\le c_5 \Big\{ \int p_s(x+hu) K(u) \int P(x+hu,dz) \big[G_{s,\nu}(z, \mathcal{N}_x(h)) \big]^2 du \Big\}^{1/2}$$

and by K_2, Theorem 2.1 with $m_0 = 1$, and P_1,

$$\le c_6 \Big\{ \int K(u) \int P(x+hu,dz) G_{s,\nu}(z, \mathcal{N}_x(h)) du \Big\}^{1/2} \tag{4.11}$$

At this stage we need the following fact:

$$\int_{\mathcal{N}_0} \sum_{n=1}^{\infty} P^n(x+hu, \{x\}) du = 0, \tag{4.12}$$

which follows from P_4 since

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

Let

$$\delta(z) = \lim_{h \downarrow 0} G_{s,\nu}(z, \mathcal{N}_x(h)), \quad \eta_h(z) = G_{s,\nu}(z, \mathcal{N}_x(h)) - \delta(z) \ .$$

Then by K_2 and Theorem 2.1 with $m_0 = 1$, the functions δ and η_h are bounded and

$$\forall z: \quad \lim_{h\downarrow 0} \eta_h(z) \downarrow 0, \quad \delta(z) = \sum_{n=0}^{\infty} (P - s \otimes \nu)^n (z, \{x\}) \leq \sum_{n=0}^{\infty} P^n(z, \{x\}) \ . \tag{4.13}$$

By (4.11) we get

$$\|hA_h\|^2 \leq c_7 \int K(u) \int P(x + hu, dz)(\eta_h(z) + \delta(z))du \ .$$

By (4.12), (4.13), K_1 and K_2,

$$\int K(u) \int P(x + hu, dz)\delta(z)du \leq c_8 \sum_{n=1}^{\infty} \int_{\mathcal{N}_0} P^n(x + hu, \{x\})du = 0 \ .$$

Let $\epsilon > 0$ and $\mathcal{A}_h = \{\eta_h > \epsilon\}$ then by P_5, $\overline{\lim}_{h\downarrow 0} \sup_{y\in\mathcal{N}_x(h)} P(y, \mathcal{A}_h) = 0$, which gives

$$\|hA_h\|^2 \leq c_7\epsilon + o(1)$$

where c_7 is independent of ϵ. Hence $\|hA_h\| = o(1)$.

If $\xi_0(u, w) = \xi(w) - P\xi(u)$ then $\widetilde{\xi}_0 \equiv 0$, $\widetilde{I}_{\xi_0\otimes\xi_0}1(x) = V\xi(x)$ and $\widetilde{\xi}_0(x) = 0$. Thus (4.10) is true. The last statement of the theorem is trivial to prove. □

The next result is designed to take care of the edge effect due to the initial measure as stated in R_9.

Lemma 4.3 *Let $g_h(u, w) = \xi_0(u, w)K_{x,h}(u)$ where $\xi_0 \in \mathcal{G}_2^d$. Assume K_1, K_2 and that $\widetilde{I}_{\|\xi_0\|}1$ is locally bounded. Then $|g_h| \leq h^{-1}g_0$, where*

$$\mathbf{P}_y(U_{g_0} < \infty) \equiv 1, \quad g_0(u, w) \stackrel{\text{def}}{=} \|\xi_0\|(u, w)\{\sup K\}1_{\mathcal{N}_x}(u) \ . \tag{4.14}$$

<u>Proof:</u> We have $U_{g_0} = \sum_{j=0}^{\tau} g_0(X_j, X_{j+1})$ and

$$\mathbf{E}_y(U_{g_0}) = G_{s,\nu}\widetilde{g}_0(y), \quad \widetilde{g}_0(y) = 1_{\mathcal{N}_x}(y) \cdot \widetilde{I}_{\|\xi_0\|}1(y) \ .$$

Hence, using K_2, $\widetilde{g}_0$ is a special function, and by Theorem 2.1, $G_{s,\nu}\widetilde{g}_0$ is bounded and (4.14) holds. □

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 4.1 *Let $\epsilon > 0$. Assume K_1-K_3 and P_1, P_2, P_6' and $h_n^{-1} \ll n^{\beta/2-\epsilon}$ or $h_n = q_{T_C(n)}$ where $q_n^{-1} \ll n^{1/2-\epsilon}$. Then $\widehat{p}_C$ is a strongly consistent estimator of p_C at the point x.*

Proof: We have $\widehat{p}_C(x) = T_C^{-1}(n)S_n(g_h)$ with $g_h = K_{x,h}$. By (4.8) and Lemma 4.1 with $\xi_0 \equiv 1$, R_3 holds for all $m \geq 1$, R'_8 is fulfilled since δ_m in this condition approaches $1/2$ when $m \uparrow \infty$. By Lemma 4.2, K_0 and K_3, $\mu_{g_h} = p_s(x) + o(1)$. By Lemma 4.2 and P_2, R_2 is true. The condition R'_7 is implied by P'_6. We use $\xi_0 \equiv 1$ in Lemma 4.3 to verify R_9. We conclude that the conditions in Theorem 3.3 are fulfilled when h_n is deterministic, and equally the conditions in Theorem 3.4 when the bandwidth is stochastic.

□

Theorem 4.2 *Let $\epsilon > 0$ and $m > 2$. Assume* K_1-K_3, P_1, P_2, P_4, P'_6, *$P\|\xi\|^{2m}$ is locally bounded and $P\xi$, $P\|\xi\|$ are locally continuous, $h_n^{-1} \ll n^{\beta\delta_m - \epsilon}$ with $\delta_m = (m-2)/(2m-1)$ or $h_n = q_{T_C(n)}$ with $q_n^{-1} \ll n^{\delta_m - \epsilon}$ and $\delta_m = (m-1)/(2m-1)$. Then $\widehat{P}\xi$ is a strongly consistent estimator of $P\xi$ at the point x.*

Proof: In this case we have $\widehat{P}\xi(x) = T_C^{-1}(n)S_n(g_h)/\widehat{p}_C^{-1}(x)$ with $g_h(u,w) = \xi(w)K_{x,h}(u)$. The previous theorem together with P_4 guarantees that the denominator converges almost surely to $p_C(x) > 0$. Hence it is enough to consider the numerator. By Lemma 4.2 with $\widetilde{\xi}_0 = \widetilde{I}_\xi 1 = P\xi$ and K_3, $\mu_{g_h} = p_s(x)P\xi(x) + o(1)$. By Lemma 4.1 moments of $|g_h|$ up to order $2m$ can be used and R_3 holds. The condition R_9 is satisfied by Lemma 4.3 since $\widetilde{I}_{\|\xi\|}1 = P\|\xi\|$ is assumed to be locally bounded. The rest of the conditions in Theorem 3.3 and Theorem 3.4 are verified in a similar way as in the previous theorem, and the numerator converges with probability one to $p_C(x)P\xi(x)$.

□

Theorem 4.3 *Let $\epsilon > 0$. Assume* K_1, K_2, P_1, P_2, P_4- P_6 *and* T_1. *If $h_n^{-1} \ll n^{\beta-\epsilon}$, then*

$$T_C^{1/2}(n)h_n^{1/2}\Big\{\widehat{p}_C(x) - p_C(x) - (\pi_C(K_{x,h_n}) - p_C(x))\Big\} \xrightarrow[n]{d} \mathcal{N}(0, p_C(x)\int K^2(u)du) \ . \quad (4.15)$$

Without the conditions P_2 *and* P_5 *we still have that*

$$T_C^{1/2}(n)\sigma_{C,h_n}^{-1}\Big\{\widehat{p}_C(x) - p_C(x) - (\pi_C(K_{x,h_n}) - p_C(x))\Big\} \xrightarrow[n]{d} \mathcal{N}(0,1) \ . \quad (4.16)$$

where $\sigma_{C,h}$ is defined in Section 3.1 with $g_h = K_{x,h}$, and where

$$\begin{aligned} h\sigma_{C,h}^2 &= \int p_C(x+hu)K^2(u)du \\ &\quad +2\int p_C(x+hu)K(u)\Big\{\int PG_{s,\nu}(x+hu,dy)K(h^{-1}(y-x))\Big\}du + o(1) \ . \quad (4.17) \end{aligned}$$

If K_4 *and* P_3 *holds with order 2, then the bias term $\pi_C(K_{x,h_n}) - p_C(x)$ of* (4.15) *and* (4.16) *is negligible if $h_n^{-1} \gg n^{\beta/5+\epsilon}$.*

Proof: We need to verify the conditions of Theorem 3.2. We have $\xi_0 \equiv 1$ which is trivially bounded, and the support of $g_h = K_{x,h}$ is a special set by K_2. The conditions

in Lemma 4.1 are fulfilled for $m \geq 1$, and in particular Lemma 4.1 then implies that R_0 and R_3 hold for $m = 2$ and $v = 1$. Moreover, the condition on the bandwidth means that R_8 is true with $v = 1$. The conditions R_7 and R_{10} are implied by P_6 and T_1. By P_1 and K_1, K_2,

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

which implies R_2'. Since $g_h \geq 0$, it follows by Remark 3.3 that R_4 and R_5 are true. If K_1, K_2 and P_1, P_4 hold, then since $A_h \geq 0$, reasoning as above for $\mu_{|g_h|}$ by the proof of Lemma 4.2 with $\xi_0 \equiv 1$, $\underline{\lim}\, h\sigma_{g_h}^2 > 0$. Hence R_1 and R_6 are true. By Lemma 4.3 with $\xi_0 \equiv 1$, R_9 is true. Moreover, R_8, R_9 imply R_9'. The conditions in Theorem 3.2 are then fulfilled and (4.16) is true.

The asymptotic expression for $\sigma_{C,h}^2$ given by (4.17) follows from Lemma A.1 in Appendix A and the first part of the proof of Lemma 4.2.

If also P_2 and P_5 hold, by Lemma 4.2,

$$\begin{aligned} \pi_s g_h &= p_s(x) + o(1) \\ h\sigma_g^2 &= p_s(x) \int K^2(u)du + o(1) \ . \end{aligned} \tag{4.18}$$

Thus (4.15) holds.

It remains to consider the bias term. By K_4 and P_3

$$\begin{aligned} \pi_C(K_{x,h_n}) - p_C(x) &= \int (p_C(x+hu) - p_C(x))K(u)du \\ &= \Big\{\frac{d}{dx} p_C(x) h\Big\} \int uK(u)du + O(h^2) \\ &= O(h^2) \ . \end{aligned}$$

Hence R_2''' holds, and by Theorem 3.2 the bias term is negligible when $h_n^{-1} \geq n^{\beta/5+\epsilon}$.

□

Theorem 4.4 *Let $\epsilon > 0$ and $\psi_x(y) = P\xi(y) - P\xi(x)$. Assume* K_1, K_2, K_4, P_1, P_2, P_4,P_6. *Moreover, assume that $P\xi$, $V\xi$ are locally continuous, $V\xi(x)$ is positive definite and $P\|\xi\|^{2m}$ is locally bounded for some $m \geq 2$. If $h_n^{-1} \ll n^{\beta-\epsilon}$, then*

$$\Big\{h_n \sum_{t=0}^{n} K_{x,h_n}(X_t)\Big\}^{1/2} \Big\{\widehat{P}\xi(x) - P\xi(x) - \frac{\pi_s \psi_x \cdot K_{x,h_n}}{\pi_s K_{x,h_n}}\Big\} \xrightarrow[n]{d} \mathcal{N}(0, V\xi(x) \int K^2(u)du) \ . \tag{4.19}$$

If K_4 *and* P_3 *hold with order 2 and $P\xi$ possesses continuous derivatives of second order, then the bias term $\pi_s\psi_x \cdot K_{x,h_n}/\pi_s K_{x,h_n}$ is negligible if $h_n^{-1} \gg n^{\beta/5+\epsilon}$.*

Proof: Denote the left hans side of (4.19) by Λ_{n,h_n}. Since $\widehat{P}\xi(x)$ is a ratio, we get some additional bias terms.

Let $\mu_{p,h} = \pi_s K_{x,h}$, $\mu_{\psi,h} = \pi_s I_{K_{x,h}} \psi_x$. Now,

$$\begin{aligned} \xi(X_{t+1}) &= \{\xi(X_{t+1}) - P\xi(X_t)\} + (P\xi(X_t) - P\xi(x)) + P\xi(x) \\ \xi(X_{t+1})K_{x,h}(X_t) &= g_h(X_t, X_{t+1}) + \psi_x(X_t)K_{x,h}(X_t) + P\xi(x)K_{x,h}(X_t) \end{aligned} \tag{4.20}$$

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

$$\widehat{P}\xi(x) - P\xi(x) = S_n^{-1}(K_{x,h})\Big\{S_n(g_h) + S_n(\psi_x \cdot K_{x,h})\Big\} . \tag{4.21}$$

The last term on the right hand side represents the bias. It contains a stochastic quantity, and we want to replace it by a deterministic bias term. Let $a_h \stackrel{\text{def}}{=} \mu_{p,h}^{-1}\mu_{\psi,h}$. Then

$$\begin{aligned}\widehat{P}\xi(x) - P\xi(x) - a_h = & \; S_n^{-1}(K_{x,h})\Big\{S_n(g_h) + S_n(\psi_x \cdot K_{x,h}) - a_h S_n(K_{x,h})\Big\} \\ = & \; S_n^{-1}(K_{x,h})\Big\{S_n(g_h) + S_n(f_h)\Big\},\end{aligned} \tag{4.22}$$

where

$$f_h = (\psi_x - a_h) \cdot K_{x,h} . \tag{4.23}$$

We note that $\mu_{g_h} = 0$ and $\mu_{f_h} = 0$. Now, we can write

$$\Lambda_{n,h} = \Delta_{n,h}^1 + \Delta_{n,h}^2 \tag{4.24}$$

where

$$\begin{aligned}\Delta_{n,h}^1 = & \; \Big\{\widehat{p}_C(x)\Big\}^{-1/2} T_C^{-1/2}(n) h^{1/2} S_n(g_h) \\ \Delta_{n,h}^2 = & \; \Big\{\widehat{p}_C(x)\Big\}^{-1/2} T_C^{-1/2}(n) h^{1/2} S_n(f_h) ,\end{aligned} \tag{4.25}$$

and where C is a purely auxiliary set chosen such that T_1 holds. Then conditions are fulfilled for the second part of Theorem 4.3 , and thus (4.16) and (4.17) imply

$$\widehat{p}_C(x) = \pi_C K_{x,h_n} + o_P(1) . \tag{4.26}$$

Hence $\widehat{p}_C(x) \to p_C(x)$ in probability. The function f_h can be written as

$$f_h(y) = \Big(P\xi(y) - P\xi(x) - a_h\Big) \cdot K_{x,h}(y) .$$

The next step consists in applying Corollary 3.3 to f_h in order to show that $\Delta_{n,h}^2$ in (4.24) can be neglected.

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

$$\sup_{y \in \mathcal{N}_x(h)} \|P\xi(y) - P\xi(x) - a_h\| = o(1) . \tag{4.27}$$

Moreover, f_h and $\widetilde{I}_{f_h \otimes f_h}$ are locally continuous and special functions by K_2 and $\pi_s f_h = 0$. By Lemma A.1 (cf. A.12) we have that

$$\sigma_{\|f_h\|}^2 \le \pi_s \|f_h\|^2 + 2\pi_s I_{\|f_h\|} P G_{s,\nu} \|f_h\| + 3\mu_{\|f_h\|}^2$$

and by local continuity of $P\xi$, K_2, P_2 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 (4.27) and K_2,

$$\sup_z G_{s,\nu} h\|f_h\|(z) \le c_1 \sup_z G_{s,\nu}(z, \mathcal{N}_x(h)) \le c_2,$$

and it follows that

$$\begin{aligned} h\pi_s I_{\|f_h\|} G_{s,\nu} \|f_h\| &\le c_2 \mu_{\|f_h\|} = o(1) \\ h\pi_s \|f_h\|^2 &\le c_3 \Big\{ \sup_z h \|f_h(z)\| \Big\} \pi_s \|f_h\| \\ &\le c_4 \mu_{\|f_h\|} = o(1) \ . \end{aligned}$$

By Remark 3.3 we can infer that $\|f_h\|$ satisfies R_0, R_1', R_2'. By Lemma 4.3 with $\xi_0(u,w) = P\xi(u) - P\xi(x) - a_h$ it follows that R_9 is true for f_h. The conditions R_7 and R_8'' of Corollary 3.3 are trivially verified. Hence the conditions in this corollary are true for each component of f_h, and it follows from Lemma 3.1 and 3.2 that

$$T_C^{-1/2}(n) h_n^{1/2} S_n(f_{h_n}) = o_p(1) \ . \tag{4.28}$$

Hence by (4.24), (4.25) and (4.28),

$$\Lambda_{n,h_n} = \Delta^1_{n,h_n} + o_P(1) \tag{4.29}$$

It remains to verify that the conditions in Corollary 3.4 are fulfilled for Δ^1_{n,h_n}.

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 $\|P\xi\|^{2m} \le P\|\xi\|^{2m}$ which gives that

$$\widetilde{I}_{\|\xi_0\|^{2m}} 1 \le 2^m P\|\xi\|^{2m}$$

because

$$\begin{aligned} \widetilde{I}_{\|\xi_0\|^{2m}} 1(x) &= \int P(x,dy) \|\xi_0(x,y)\|^{2m} \\ &= \mathrm{E}_x \Big[\|\xi_0(X_0,X_1)\|^{2m} \mid X_0 = x \Big] \\ &\le \mathrm{E}_x \Big[\Big\{ \|\xi(X_1)\| + \|P\xi(X_0)\| \Big\}^{2m} \mid X_0 = x \Big] \\ &\le 2^{m-1} \Big\{ \mathrm{E}_x \Big[\|\xi(X_1)\|^{2m} \mid X_0 = x \Big] + \mathrm{E}_x \Big[\|P\xi(X_0)\|^{2m} \mid X_0 = x \Big] \Big\} \\ &\le 2^{m-1} \Big\{ P\|\xi\|^{2m}(x) + \|P\xi\|^{2m}(x) \Big\} \\ &\le 2^{m-1} \Big\{ P\|\xi\|^{2m}(x) + P\|\xi\|^{2m}(x) \Big\} \\ &\le 2^m P\|\xi\|^{2m}(x) \ . \end{aligned}$$

Hence the conditions in Lemma 4.1 are fulfilled since the right hand side of the above inequality is by assumption locally bounded. Thus $\|g_h\|$ satisfies R_0, R_2' and R_3 (with $m = 2$ and $v = 1$).

Likewise, since $P\xi$, $P(\xi \otimes \xi)$ are continuous at the point x and $\widetilde{\xi}_0 = P\xi - P\xi \equiv 0$, the conditions in Lemma 4.2 are fulfilled. In particular (4.10) holds and

$$h\Sigma_{g_h} = p_C(x) V\xi(x) \int K^2(u)du + o(1) \ . \tag{4.30}$$

Since $A_h \ge 0$ for $\|g_h\|$, by Lemma 4.2 applied to $\|g_h\|$ we have that R_1 holds for $\|g_h\|$, and by (4.30) it follows that (3.20) and (3.21) are fulfilled with $\Sigma_0 = p_C(x) V\xi(x) \int K^2(u)du$. Further, the conditions R_4 - R_8 are easily verified for $\|g_h\|$. We use Lemma 4.3 to verify

R_9 for $\|g_h\|$. By our choice of the set C, we have that T_1, R_{10} hold. Thus the conditions in Corollary 3.4 hold and (4.19) follows.

The bias term is negligible if

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

Introducing the set C as in (4.25), by P_4 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 $h_n^{-1} \geq n^{\beta/5+\epsilon}$. Then from Lemma 3.1 and Lemma 3.2

$$h_n^5 T_C(n) = o(1) \quad \text{a.s.} \ .$$

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

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

and

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

This gives

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

Hence the theorem is proved.

□

Remark 4.2 *Note that the set C only plays an auxiliary role in the proof of Theorem 4.4. Actually, we might have used $T(n)$ instead of $T_C(n)$. This is in contradistinction to Theorem 4.3, where C is an essential part of the theorem and its proof. The convergence rate in both theorems is $(T(n)h)^{-1/2}$ or equivalently $(T_C(n)h)^{-1/2}$, and it reduces to the familiar rate $(nh)^{-1/2}$ in the positive recurrent case. Some examples of Markov models (both linear and nonlinear) satisfying the assumptions of Theorem 4.3 and Theorem 4.4 are given in Myklebust et al (1998a), where we also report on simulation experiments checking the validity of the two theorems.*

A Appendix

In this appendix we derive an expression for σ_g^2 and Σ_g given by (2.76). We also give formulae for higher order moments. We assume (2.4) and aperiodicity.

Lemma A.1 *Let $g \in \mathcal{G}_2$,*

$$g_\nu(u) = \int \nu(dz) g(u,z) \tag{A.1}$$

and

$$\gamma_0(g) = \pi_s(g^2) - 2\mu_g \pi_s(s g_\nu) + \mu_g^2 \ . \tag{A.2}$$

Then

$$\sigma_g^2 = \gamma_0 + 2\Big\{\pi_s \widetilde{I}_g G_{s,v} \widetilde{g} - \mu_g^2\Big\} \ . \tag{A.3}$$

Proof: The notation is in accordance with Lemma 2.3 with initial measure ν so that U_0 has the same distribution as U_1, and the process $\{U_k,\ k \geq 0\}$ is stationary. We have τ_0 and τ_1 defined by (2.45). The variables U_0 and U_1 are given by (2.71). We define

$$Z_j = g(X_j, X_{j+1}), \quad \widetilde{Z}_j = \widetilde{g}(X_j) \ .$$

Recall that $\mu_g = \mathrm{E}_\nu U_0 = \pi_s g$. Let $A_j = 1(Y_j = 0)$, $B_{j,j+k} = \prod_{s=j}^{j+k-1} A_s$ when $k \geq 1$ and $B_{j,j} = 1$, $B_j = B_{0,j}$. Then $1(\tau_0 \geq j) = B_j$ is $\mathcal{F}_{j-1}^Y$-measurable and

$$U_0 = \sum_{j=0}^{\infty} B_j Z_j \ . \tag{A.4}$$

Let $g_l(x) = \mathrm{E}_x\Big[1(Y_0 = l) Z_0\Big]$, $\psi_l(x) = \mathrm{E}_x\Big[1(Y_0 = l) Z_0 \cdot \sum_{k=1}^{S_\alpha} Z_k\Big]$ for $l = 0, 1$ and $\psi = \psi_0 + \psi_1$. By (A.4) and since $B_j B_{j+1+k} = B_j A_j B_{j+1,j+1+k}$, we can write

$$\begin{aligned} U_0^2 &= \sum_{j=0}^{\infty} B_j Z_j^2 + 2\sum_{j=0}^{\infty} B_j A_j W_j \\ W_j &\stackrel{\text{def}}{=} Z_j \sum_{k=0}^{\infty} B_{j+1,j+1+k} g(X_{j+1+k}, X_{j+2+k}) \ . \end{aligned} \tag{A.5}$$

By (2.10), $\mathrm{E}(A_j W_j \mid \mathcal{F}_j^X \vee \mathcal{F}_{j-1}^Y) = \psi_0(X_j)$ and using (A.5) we obtain

$$\mathrm{E}_\nu(U_0^2) = \pi_s g^2 + 2\pi_s \psi_0 \ . \tag{A.6}$$

Let $U_{01} = \sum_{j=0}^{\tau_0 - 1} Z_j$ and $U_{02} = g(X_{\tau_0}, X_{\tau_0+1})$. Then $U_0 = U_{01} + U_{02}$ and U_{01} is independent of U_1. Representations for the variables are given below, where we have used that $1(\tau_0 = j) = B_j(1 - A_j)$ and $1(\tau_1 \geq j + 1 + k,\ \tau_0 = j) = B_j(1 - A_j) B_{j+1,j+1+k}$,

$$\begin{aligned} U_{01} &= \sum_{j=0}^{\infty} B_j A_j Z_j \\ U_{02} &= \sum_{j=0}^{\infty} B_j (1 - A_j) Z_j \\ U_1 &= \sum_{j=0}^{\infty} B_j (1 - A_j) \sum_{k=0}^{\infty} B_{j+1,j+1+k} Z_{j+1+k} \ . \end{aligned} \tag{A.7}$$

This gives $U_{02}U_1 = \sum_{j=0}^{\infty} B_j(1-A_j)W_j$ and by (2.10) $\mathbf{E}((1-A_j)W_j \mid \mathcal{F}_j^X \vee \mathcal{F}_{j-1}^Y) = \psi_1(X_j)$. Thus $\mathrm{E}_\nu(U_{01}) = \pi_s g_0$ and $\mathrm{E}_\nu(U_{02}U_1) = \pi_s\psi_1$. Hence

$$\begin{aligned}\mathbf{E}_\nu(U_0U_1) &= \mathbf{E}_\nu(U_{01})\mathbf{E}_\nu(U_1) + \mathbf{E}_\nu(U_{02}U_1)\\ &= \pi_s g_0 \cdot \pi_s g + \pi_s\psi_1 \ . \end{aligned} \tag{A.8}$$

Combining (A.6) and (A.8), and using (2.76) and the one-dependence of the process $\{U_k,\ k \geq 1\}$ we get

$$\begin{aligned}\sigma_g^2 &= E_\nu(U_0^2) + 2E_\nu(U_0U_1) - 3\mathrm{E}_\nu(U_0)\mathrm{E}_\nu(U_1)\\ &= \pi_s g^2 + 2\pi_s\psi_0 + 2\pi_s\psi_1 + 2\pi_s g_0 \cdot \pi_s g - 3\pi_s g \cdot \pi_s g\\ &= \{\pi_s g^2 - \pi_s^2 g\} + 2\pi_s\psi - 2\pi_s g \cdot \{\pi_s g - \pi_s g_0\}\\ &= \pi_s g^2 - \pi_s^2 g - 2\pi_s g \cdot \pi_s s g_\nu + 2\pi_s\psi \ . \end{aligned} \tag{A.9}$$

where we have used that $g_1 = sg_\nu$ with g_ν defined by (A.1) so that $\pi_s g - \pi_s g_0 = \pi_s(sg_\nu)$. From

$$\sum_{j=1}^{S_\alpha} Z_j = \sum_{j=0}^{\infty} B_{1,1+j}Z_{j+1} \tag{A.10}$$

we obtain $\mathrm{E}_x\Big\{\sum_{k=1}^{S_\alpha} Z_j \mid \mathcal{F}_1^X \vee \mathcal{F}_0^Y\Big\} = G_{s,v}\tilde{g}(x)$ which gives with the aid of (2.10) that

$$\psi(x) = \mathbf{E}_x\Big\{g(X_0,X_1)G_{s,v}\tilde{g}(X_1)\Big\}, \quad \pi_s\psi = \pi_s\tilde{I}_g G_{s,v}\tilde{g} \ . \tag{A.11}$$

□

If $g(x,y) = g(x)$ then $g_\nu = g$ since $\nu(E) = 1$ and $\pi_s(g_1) = \pi_s(sg)$. The expression for σ_g^2 simplifies to

$$\sigma_g^2 = \pi_s(g^2) - \pi_s^2(g) + 2\pi_s I_g P G_{s,\nu} g - 2\pi_s(g)\pi_s(sg) \ . \tag{A.12}$$

We note that if $\pi_s(g) = 0$ then $\sigma_g^2 = \pi_s(g^2) + 2\pi_s(\psi)$ which is in accordance with Nummelin(1984, p 139). However, his conditions are different since he assumes that $g \in \mathcal{G}$ and at the same time avoids $m_0 = 1$.

In the multivariate case we also need an expression for the asymptotic covariance matrix. In order to have a compact notation we will use a d dimensional kernel. Recall that

$$\tilde{I}_g(x,dy) = P(x,dy)g(x,y), \quad g = (g_1,\ldots,g_d)' \in \mathcal{G}_2^d \tag{A.13}$$

and note that if $f = (f_1,\ldots,f_d)' \in \mathcal{G}^d$ then $\tilde{I}_g \otimes f(u)$ is the matrix $\{\tilde{I}_{g_i}f_j(u),\ 1 \leq i,j \leq d\}$. Moreover, $\tilde{I}_{g\otimes g}$ is defined by replacing g in (A.13) with $g \otimes g$. We also use the standard Euclidian norm $\|g\| = [\sum_{i=1}^d g_i^2(x,y)]^{1/2}$.

Lemma A.2 *Suppose $g \in \mathcal{G}_2^d$ then*

$$\Sigma_g = \pi_s(g \otimes g) + A_g + A_g' - B_g - B_g' \tag{A.14}$$

where

$$A_g = \pi_s\tilde{I}_g \otimes G_{s,v}\tilde{g}, \quad B_g = \frac{1}{2}(\pi_s g \otimes \pi_s g) + \pi_s g \otimes \pi_s s \cdot g_\nu \ . \tag{A.15}$$

Proof: Let $g_a = a'g$ where $a \in R^d$. By (2.77) it is enough to show that $\sigma^2_{g_a} = a\Sigma_g a'$. We have $\pi_s(g_a^2) = a'\pi_s(g \otimes g)a$, $\widetilde{I}_{g_a} = a'\widetilde{I}_g$, $G_{s,\nu}\widetilde{I}_{g_a} = a'G_{s,\nu}\widetilde{I}_g$ and

$$\widetilde{I}_{g_a} \otimes G_{s,v}\widetilde{g}_a = a'(\widetilde{I}_g \otimes G_{s,v}\widetilde{g})a$$

with corresponding results for all parts of (A.14). Thus $\sigma^2_{g_a} = a'\Sigma_g a$, and the lemma follows by comparing to (A.9) and (A.11).

It is of interest to find expression of higher order moments of U. We start with the following lemma which is well-known, but is stated for completeness.

Lemma A.3 *Let $\{a_k\}$ be a real sequence and let $\Delta_{m,r} = \{\ell = (\ell_1, \ldots, \ell_r) \in \mathcal{N}_+^r\colon \ \sum \ell_k = m\}$ for $r \geq 1$. Then*

$$\Big\{\sum_{k=0}^{n} a_k\Big\}^m = \sum_{r=1}^{m} \sum_{\ell \in \Delta_{m,r}} \binom{m}{\ell} J_{n,r,\ell} \tag{A.16}$$

where $\binom{m}{\ell} = (\prod_{j=1}^r \ell_j!)^{-1} m!$ and

$$\begin{aligned} J_{n,r,\ell} &= \sum_{0 \leq j_1 < j_2 < \cdots < j_r \leq n} a_{j_1}^{\ell_1} \cdots a_{j_r}^{\ell_r} \\ &= \sum_{h_1=0}^{n} \sum_{h_2=1}^{n-h_1} \cdots \sum_{h_r=1}^{n-h_1-\cdots-h_{r-1}} a_{h_1}^{\ell_1} \cdots a_{h_1+\cdots+h_r}^{\ell_r} \ . \end{aligned} \tag{A.17}$$

Proof: Let $\mathcal{N}_m^{n+1} = \{\alpha \in \mathcal{N}^{n+1}\colon \sum_{j=0}^n \alpha_j = m\}$, $\Delta'_{m,r} = \{\alpha \in N_m^{n+1}\colon \#\{j\colon \alpha_j > 0\} = r\}$ and $C_\alpha = \{j\colon \alpha_j > 0\}$. Then, since $\mathcal{N}_m^{n+1} = \cup_{r=1}^m \Delta_{m,r}$ and $\binom{m}{\alpha}$ is a symmetric function of $(\alpha_0, \ldots, \alpha_n)$

$$\begin{aligned} \Big\{\sum_{k=0}^{n} a_k\Big\}^m &= \sum_{\alpha \in \mathcal{N}_m^{n+1}} \binom{m}{\alpha} a_0^{\alpha_0} \cdots a_n^{\alpha_n} \\ &= \sum_{r=1}^{m} \sum_{\alpha \in \Delta'_{m,r}} \binom{m}{\alpha} a_0^{\alpha_0} \cdots a_n^{\alpha_n} \\ &= \sum_{r=1}^{m} \sum_{\alpha \in \Delta'_{m,r}} \binom{m}{\alpha} \prod_{j \in C_\alpha} a_j^{\alpha_j} \\ &= \sum_{r=1}^{m} \sum_{\ell \in \Delta_{m,r}} \binom{m}{\ell} \sum_{0 \leq i_1 < i_2 < \cdots < i_r \leq n} \prod_{j=1}^{r} a_{i_j}^{\ell_j} \ . \end{aligned}$$

□

Let

$$\check{I}_g(x, dy) = (P - s \otimes \nu)(x, dy) g(x, y) \tag{A.18}$$

then refering to the notation of the proof of Lemma A.1 we can write

$$\psi_0 = \check{I}_g G_{s,\nu}\widetilde{g}, \quad \psi_1 = (\widetilde{I}_g - \check{I}_g) G_{s,\nu}\widetilde{g}, \quad \psi = \widetilde{I}_g G_{s,\nu}\widetilde{g} \tag{A.19}$$

and

$$g_0 = \check{I}_g 1, \quad g_1 = (\widetilde{I}_g - \check{I}_g)1, \quad g_0 + g_1 = \widetilde{I}_g 1 \ .$$

Our next result gives an exact description of an arbitrary central moment of U. This is subsequently applied to give bounds of a given moment.

Theorem A.1 *Let $g \in \mathcal{G}_2$ and let $U_g = \sum_{j=0}^{\tau} g(X_j, X_{j+1})$. Let $m \geq 1$. Then*

$$\mathrm{E}_\nu U_g^m = \sum_{r=1}^{m} \sum_{\ell \in \Delta_{m,r}} \binom{m}{\ell} \nu G_{s,\nu} \check{I}_{g^{\ell_1}} \cdots G_{s,\nu} \check{I}_{g^{\ell_{r-1}}} G_{s,\nu} \widetilde{I}_{g^{\ell_r}} 1 \tag{A.20}$$

in the notation of Lemma A.3. *The right hand side of* (A.20) *is well-defined if it is finite when U_g is replaced by $U_{|g|}$. If $g(x,y) = f(x)h(y)$, then $\check{I}_g = I_f(P - s \otimes \nu)I_h$ and* (A.20) *simplifies.*

Remark A.1 *When $m = 2$ we get*

$$\mathrm{E}_\nu(U_g^2) = \pi_s \widetilde{I}_{g^2} 1 + 2\pi_s \check{I}_g G_{s,\nu} \widetilde{I}_g 1, \tag{A.21}$$

which coincides with the right hand side of (A.6).

<u>Proof:</u> Since $U_g = \sum_{j=0}^{\infty} B_j g(X_j, X_{j+1})$, we can apply Lemma A.3 with $a_j = B_j g(X_j, X_{j+1})$ and $n = \infty$. We can apply this lemma with $n = \infty$ since we assume that $\mathrm{E}(U_{|g|})$ is finite and therefore

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

Let $r \leq m$, $\ell \in \Delta_{m,r}$ be fixed and let $g^s(x,y) = g^{\ell_s}(x,y)$, $Z_j^s = g^s(X_j, X_{j+1})$, for $s = 1, \ldots, r$. Define $J_k = J_{r,\ell,k}$ for $k = 1, \ldots, r$,

$$\begin{aligned} J_k &\stackrel{\text{def}}{=} \sum_{h_1=0}^{\infty} \sum_{h_2=1}^{\infty} \cdots \sum_{h_k=1}^{\infty} B_{h_1 + \cdots h_k} Z_{h_1}^{r-k+1} \cdots Z_{h_1+h_2+\cdots+h_k}^{r} \\ &= \sum_{h_1=0}^{\infty} B_{h_1} A_{h_1} Z_{h_1}^{r-k+1} W_{h_1+1}^{(k-1)}, \quad \text{when } k \geq 2, \end{aligned} \tag{A.22}$$

where according to the above $Z^{r-k+1} = g^{\ell_{r-k+1}}(X_j, X_{j+1})$ and where

$$W_{h_1}^{(k-1)} \stackrel{\text{def}}{=} \sum_{h_2=0}^{\infty} \sum_{h_3=1}^{\infty} \cdots \sum_{h_k=1}^{\infty} B_{h_1, h_1+h_2+\cdots+h_k} Z_{h_1+h_2}^{r-k+2} \cdots Z_{h_1+h_2+\cdots+h_k}^{r} \ . \tag{A.23}$$

Let

$$\begin{aligned} \phi_k(x) &\stackrel{\text{def}}{=} e_x J_k, \quad k = 1, \ldots, r \\ \psi_k(x) &\stackrel{\text{def}}{=} \begin{cases} \mathrm{E}_x g^r(X_0, X_1), & \text{when } k = 1 \\ \mathrm{E}_x A_0\, g^{r-k+1}(X_0, X_1)\phi_{k-1}(X_1), & \text{when } k = 2, \ldots, r. \end{cases} \end{aligned} \tag{A.24}$$

By calculation we find that

$$\psi_k = \begin{cases} \widetilde{I}_{g^r} 1, & \text{if } k = 1; \\ \check{I}_{g^{r-k+1}} \phi_{k-1} 1, & \text{if } k = 2, \dots, r. \end{cases} \tag{A.25}$$

From (A.23) and (A.24) we get

$$E\Big\{ W_{h_1+1}^{(k-1)} \mid \mathcal{F}_{h_1+1}^X \vee \mathcal{F}_{h_1}^Y \Big\} = \phi_{k-1}(X_{h_1+1}), \quad k = 2, \dots, r, \tag{A.26}$$

and by (A.22), (A.26) we obtain

$$\phi_k = G_{s,\nu} \psi_k, \quad k = 1, \dots, r, \tag{A.27}$$

which gives by (A.25)

$$\psi_k = \begin{cases} \widetilde{I}_{g^r} 1, & \text{if } k = 1; \\ \check{I}_{g^{r-k+1}} G_{s,\nu} \psi_{k-1}, & \text{if } k = 2, \dots, r, \end{cases} \tag{A.28}$$

and

$$\mathrm{E}_\nu(J_r) = \nu \phi_r = \nu G_{s,\nu} \psi_r = \pi_s \psi_r \ . \tag{A.29}$$

The equation (A.25) is of the form

$$\psi_k = K_{r-k+1} \psi_{k-1}, \quad k = 2, \dots, r, \quad K_j \stackrel{\text{def}}{=} \check{I}_{g^j} G_{s,\nu}, \quad j = 1, \dots, r-1$$

hence

$$\begin{aligned} \psi_r &= K_1 K_2 \ \cdots \ K_{r-1} \psi_1 \\ &= \check{I}_{g^1} G_{s,\nu} \check{I}_{g^2} G_{s,\nu} \ \cdots \ \check{I}_{g^{r-1}} G_{s,\nu} \widetilde{I}_{g^r} 1 \ . \end{aligned} \tag{A.30}$$

The proof is finished by (A.29) and (A.30) and by Lemma A.3 since

$$\mathrm{E}_\nu U^m = \sum_{r=1}^m \sum_{\ell \in \Delta_{m,r}} \binom{m}{\ell} \cdot \mathrm{E}_\nu J_{r,\ell,r} \ .$$

□

Corollary A.1 *If*

$$\sup G_{s,\nu} \widetilde{I}_{|g|} 1 + \sup G_{s,\nu} \widetilde{I}_{|g|^m} 1 < \infty \tag{A.31}$$

then $\mathrm{E}_\nu U_{|g|}^m$ *is finite. The assumption* (A.31) *holds if* $\widetilde{I}_{|g|} 1$ *and* $\widetilde{I}_{|g|^m} 1$ *are special functions or if* g *is a bounded special function.*

<u>Proof:</u> Note that $0 \le \check{I}_{|g|} \le \widetilde{I}_{|g|}$ and using the simple inequality $|g|^j \le |g| + |g|^m$, $1 \le j \le m$, it follows from (A.31) that both $\sup G_{s,\nu} \widetilde{I}_{|g|^k}$ and $\sup G_{s,\nu} \check{I}_{|g|^k} 1 < \infty$ are finite for $k = 1, \dots, m$.

Let the notation be as in the proof of Theorem A.1 and let c denote an upper bound for the left hand side of (A.31). Then for

$$\psi_r = \check{I}_{|g|^{\ell_1}} G_{s,\nu} \cdots G_{s,\nu} \check{I}_{|g|^{\ell_{r-1}}} G_{s,\nu} \widetilde{I}_{|g|^{\ell_r}} 1$$

we have

$$\pi_s \psi_r \leq c \cdot \pi_s \psi_{r-1} \leq c^r \ .$$

By (A.29) and Theorem A.1 the first part is established. If g is bounded and special then by Theorem 2.1

$$\sup G_{s,v} \widetilde{I}_{|g|^j} 1 \leq \{\sup |g|\}^{j-1} \sup G_{s,v} \widetilde{I}_{|g|} 1 < \infty, \quad j = 1, \ldots, m$$

and (A.31) is fulfilled. The same is true if $\widetilde{I}_{|g|} 1$ and $\widetilde{I}_{|g|^m} 1$ are special functions.

□

B Appendix

A very brief summary of the properties of $D^d[0,\infty)$, the space of all R^d-valued functions defined on R_+ which have left limits and are right continuous (cadlag), is given below. We refer to Jacod & Shiryaev(1987, p 288-322) for a complete description of this space, and the concept of weak convergence of stochastic processes with sample paths herein.

The topology of $D^d[0,\infty)$ is defined by the Skorokhod J_1-metric which may be described in the following way. A time change λ is a strictly increasing continuous bijection of R_+. A sequence $\{x_n\} \in D^d[0,\infty)$ converges to x if and only if there exists a sequence of time changes $\{\lambda_n\}$ so that both $\{x_n \circ \lambda_n\}$ and $\{\lambda_n\}$ converges locally uniformly (uniformly on compact sets) towards x and $\lambda(t) \equiv t$. If x happens to be continuous, then convergence of $\{x_n\}$ to x is locally uniform convergence without any time changes. The space $D^d[0,\infty)$ with the J_1-metric is a Polish space (separable and complete). The σ-field induced by the projection maps coincides with the Borel σ-field induced by the J_1-metric. This means that if $\xi = \{\xi(t),\ t \in R_+\}$ is a collection of random vectors in R^d defined on some common probability space, and each sample path is cadlag, then ξ is stochastic process with values in $D^d[0,\infty)$. If $\{\xi_n\}$ is a sequence of stochastic processes with values in $D^d[0,\infty)$ then $\xi_n \overset{\mathcal{L}_{\mathcal{D}^d[0,\infty)}}{\underset{n}{\longrightarrow}} \xi$ 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, i.e., $P_n(C^d[0,\infty)) = 1$ 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.

Theorem B.1 *For each n let (B_n, A_n) be a pair of stochastic processes which are cadlag, where A_n is non-negative and nondecreasing. Let B denote a Brownian motion defined for $t \in R_+$ and let A denote a strictly increasing non-negative process with independent increments, $A(0) \equiv 0$ and with no fixed jumps. Assume that $B_n \overset{\mathcal{L}_{\mathcal{D}[0,\infty)}}{\underset{n}{\longrightarrow}} B$ and $A_n \overset{\mathcal{L}_{\mathcal{D}[0,\infty)}}{\underset{n}{\longrightarrow}} A$. Then*

$$(B_n, A_n, A_n^{(-1)}) \overset{\mathcal{L}_{\mathcal{D}^3[0,\infty)}}{\underset{n}{\longrightarrow}} (B, A, A^{(-1)}) \tag{B.1}$$

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

$$(A_n^{(-1)}, B_n \circ A_n^{(-1)}) \overset{\mathcal{L}_{\mathcal{D}^2[0,\infty)}}{\underset{n}{\longrightarrow}} (A^{(-1)}, B \circ A^{(-1)})\ . \tag{B.2}$$

For all $\epsilon > 0$

$$\left(A_n^{(-1)}, \frac{B_n \circ A_n^{(-1)}}{\sqrt{A_n^{(-1)}}}\psi_\epsilon(A_n^{(-1)})\right) \overset{\mathcal{L}_{\mathcal{D}^2[0,\infty)}}{\underset{n}{\longrightarrow}} \left(A^{-1}, \frac{B \circ A^{(-1)}}{\sqrt{A^{(-1)}}}\psi_\epsilon(A^{(-1)})\right) \tag{B.3}$$

where $\psi_\epsilon(x) = \epsilon^{-1/2}x^{1/2}1(x \le \epsilon) + 1(x > \epsilon)$. *If we let* $\psi_\epsilon \equiv 1$ *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 & Shiryaev, 1987, Cor. 3.33, p. 317). If (B', A') is a limit point for this sequence, then necessarily $B' \stackrel{\mathrm{d}}{=} B$ and $A' \stackrel{\mathrm{d}}{=} A$. 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 $a \in \mathcal{C} \stackrel{\mathrm{def}}{=} \{x\colon x \text{ is strictly increasing}\}$. By the continuous mapping theorem we find that $A_n^{(-1)}$ converges weakly to $A^{(-1)}$ since $A \in \mathcal{C}$. 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 (B.1). The map $(b, x) \mapsto b \circ x$ is continuous at all points where b is continuous and x is non-negative. Again, by the continuous mapping theorem we can conclude that (B.2) is true. The reasoning is similar for (B.3) where the function ψ_ϵ guards against a discontinuity at zero. By Jacod & Shiryaev(1987, Prp. 3.14, p. 313) we have that (B.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 $t > 0$.

□

References

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

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

Bingham, N. H., Goldie, C. M., Teugels, J. L. (1989) *Regular Variation.* Cambridge University Press, Cambridge.

Billingsley, P. (1968) *Convergence of Probability.* John Wiley & Sons, New York.

Chow, Y. S. & Teicher, H. (1988) *Probability Theory. Second Edition.* Springer Verlag, New York.

Darling, D. A. & Kac, M. (1957) On occupation times for Markoff processes. *Transactions Amer. Math. Soc.* **84** 444-458.

Dickey, D. A. & Fuller, W. A. (1979) Distribution of the estimators for autoregressive time series with unit root. *Econometrica* **49** 1057-1072.

Feller, W. R. (1971) *An Introduction to Probability Theory and Its Application II, second edition.* John Wiley & Sons, New York.

Granger, C. W. J. & Hallman, J. (1991) Long memory series with attractors. Oxford Bulletin of Economics and Statistics **53**, 11-26.

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

Hjellvik, V., Yao, Q., Tjøstheim, D. (1998) Linearity testing using local polynomial approximation. *J. Stat. Plan. Inf.*, to appear.

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. *Statistics & Decisions* **12** 149-160.

Höpfner, R., Jacod, J., Ladelli, L. (1990) Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Th. Rel. Fields* **86** 105-129.

Jacod, J. & Shiryaev, A. N. (1987) *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin.

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

Kallianpur, G. & Robbins, H. (1954) The sequence of sums of independent random variables. *Duke Math. J.* **21** 285-307.

Kasahara, Y. (1982) A limit theorem for slowly increasing occupation times. *Ann. Prob.* **10** 728-736.

KASAHARA, Y. (1984) Limit theorems for Levy 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 & TJØSTHEIM, D. (1995) Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* **11** 258-289.

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

MYKLEBUST, T., KARLSEN, H., TJØSTHEIM, D. (1998a) Recurrence properties of unit-root models and a nonlinear generalization. In preparation.

MYKLEBUST, T., KARLSEN, H., TJØSTHEIM, D. (1998b) Extension of some central limit theorems in null recurrent Markov chains related to nonparametric estimation. In preparation.

NUMMELIN, E. (1984) *General Irreducible Markov Chains and Non-negative Operators.* Cambridge University Press, Cambridge.

POLLARD, D. (1984) *Convergence of Stochastic Processes.* Springer-Verlag, 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. Stat.* **25** 2054- 2083.

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

TOUATI, P. A. (1990) Loi fonctionnelle du logarithme itere pour les processus de Markov recurrents. *Ann. Prob.* **18** 140-159.

TJØSTHEIM, D. & 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* **4** 2843 - 2918. Engle, R. F. & McFadden, D.L. editors. Elsevier, Amsterdam.

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

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