Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 124 (2014) 646–677

stochastic processes and their applications

www.elsevier.com/locate/spa

A martingale decomposition for quadratic forms of Markov chains (with applications)☆

Yves F. Atchadé[a,∗], Matias D. Cattaneo[b]

[a] University of Michigan, Department of Statistics, Ann Arbor, 48109, MI, United States
[b] University of Michigan, Department of Economics, Ann Arbor, 48109, MI, United States

Received 28 July 2011; received in revised form 31 August 2013; accepted 1 September 2013
Available online 16 September 2013

Abstract

We develop a martingale-based decomposition for a general class of quadratic forms of Markov chains, which resembles the well-known Hoeffding decomposition of U-statistics of i.i.d. data up to a reminder term. To illustrate the applicability of our results, we discuss how this decomposition may be used to studying the large-sample properties of certain statistics in two problems: (i) we examine the asymptotic behavior of lag-window estimators in time series, and (ii) we derive an asymptotic linear representation and limiting distribution of U-statistics with varying kernels in time series. We also discuss simplified examples of interest in statistics and econometrics.

MSC: 60J10; 62M10

Keywords: Central limit theorems; Markov chains; Markov chain Monte Carlo; Martingale approximations; Quadratic forms; U-statistics

☆ Note: A previous version of this paper circulated under the title "Limit theorems for quadratic forms of Markov chains". We are grateful to Michael Jansson, Lutz Kilian and Yizao Wang for helpful discussions and comments. This work was partly supported by NSF grant DMS-0906631 (Atchade) and NSF grants SES-0921505 & SES-1122994 (Cattaneo).
∗ Corresponding author.
E-mail addresses: yvesa@umich.edu (Y.F. Atchadé), cattaneo@umich.edu (M.D. Cattaneo).

0304-4149/$ - see front matter
http://dx.doi.org/10.1016/j.spa.2013.09.001

1. Introduction

This paper deals with quadratic forms of the type

$$U_n(h_n) = \sum_{\ell=1}^{n} \sum_{j=1}^{\ell} w_n(\ell, j) h_n(X_\ell, X_j), \quad n \geq 1, \tag{1}$$

for a stochastic process $\{X_n, n \geq 0\}$, weight matrices $w_n : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ and symmetric kernels $h_n : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Quadratic forms of possibly time-dependent random variables naturally arise in a variety of statistical and econometric problems, and their large-sample properties are of particular importance to develop asymptotically valid inference procedures.

For sequences of independent random variables $\{X_n, n \geq 0\}$, the well-known Hoeffding decomposition provides a useful approach to studying the asymptotic properties of $U_n(h_n)$ because it decomposes the statistic into two (uncorrelated) martingale sequences, which are then easily handled by standard martingale theory. See, e.g., [24] for a review. When the process $\{X_n, n \geq 0\}$ is time-dependent, however, the classical Hoeffding decomposition is not as useful because the resulting representation does not have the desirable martingale property in general. As a consequence, the large-sample properties of quadratic forms of time-dependent random variables are typically established in a less systematic way, and the most well understood special case of (1) is the standard U-statistic where h_n does not depend on n and $w_n(\ell, j) = 1$ if $\ell \neq j$ and 0 otherwise (e.g., [27,7,5]). Other special cases of (1) have been considered in the literature including [11] which studied $U_n(h)$ where neither h_n nor w_n depends on n for stationary processes, whereas [26] studied $U_n(h_n)$ when $h_n(x, y) = h(x, y) = xy$ for a martingale-difference sequence (see also, e.g., [2] for i.i.d. sequences).

In this paper, we develop a martingale approximation for $U_n(h_n)$ that allows for a general and systematic analysis of the quadratic form (1) when $\{X_n, n \geq 0\}$ is a Markov chain. Martingale approximation is a well established technique when dealing with linear partial sums of dependent processes [14,17], but has not been fully explored in dealing with quadratic forms (a notable exception is [26]). Our goal is to decompose $U_n(h)$ into two martingale sequences and a remainder term under general conditions, thereby offering an (approximate) analogue of the classical Hoeffding decomposition (whenever the remainder term is "small"). Specifically, we obtain under easy-to-interpret assumptions an approximating quadratic martingale to $U_n(h_n)$ from a solution of a bivariate analogue of the well known Poisson's equation. We also illustrate our approach with two main statistical applications.

In the first application we study the asymptotic behavior of lag-window estimators of long-run variance (asymptotic variance) for Markov chains (see, e.g., [21]). We obtain a decomposition of lag-window estimators that shed some new light on the asymptotic behavior of these estimators, particularly by contrasting the classical asymptotics and the so-called "fixed-b" asymptotics [19,12]. We derive two theorems that extend existing results. We obtain the consistency of lag-window estimators for non-geometrically ergodic Markov chains extending recent results of Flegal and Jones [8] and Atchade [1], and we extend the "fixed-b" asymptotics framework to handle non-stationary Markov chains. These results have important implications for Markov Chain Monte Carlo (MCMC) simulations, offering in particular new robust procedures for constructing Monte Carlo confidence intervals. We also offer a simple illustration of our results in the context of a simple GARCH(1, 1) model.

As another application of the martingale approximation method, we derive an analogue of an asymptotic linear representation for U-statistics with varying kernels without imposing stationarity and under easy-to-verify assumptions. In particular, we do not rely on mixing conditions. This result extends the "projection lemma" for U-statistics with varying kernels of Powell et al. [20, Lemma 3.1] to the context of time-dependent data, which may be easily used to derive a central limit theorem for these and related statistics under weak conditions. We illustrate the applicability of our results in the context of kernel-based semiparametric density-weighted average derivatives (for review see, e.g., [25,23,3]).

The remainder of the paper is organized as follows. The rest of the introduction outlines the general setup and main notation employed throughout, while Section 2 derives the main martingale approximation method. Section 3 studies the asymptotic properties of lag-window estimators and discusses a simple example in statistics and econometrics. We study U-statistics with varying kernels in Section 4, where we also illustrate our results by analyzing the asymptotic properties of semiparametric density-weighted average derivatives. All the proofs are presented in Section 5.

1.1. Setup and notation

We employ standard notation and results from the Markov chains literature; for a review see, e.g., [18] in general, and their Chapter 16 in particular. Throughout the paper, $\{X_n, n \geq 0\}$ denotes a Markov chain taking values in a general state space $(\mathcal{X}, \mathcal{B})$ equipped with a sigma-algebra $\mathcal{B}$. We denote by P the transition kernel of the Markov chain and μ its invariant distribution whose existence is assumed. Unless explicitly stated otherwise, $\{X_n, n \geq 0\}$ is a nonstationary Markov chain with initial distribution ρ. We write $\mathbb{E}$ (resp. $\mathbb{E}_x$) for the expectation operator induced by the Markov chain when $X_0 \sim \rho$ (resp. $X_0 = x$), where we use $Y \sim \mu$ to denote that a random variable Y has distribution μ.

We will rely on the following set of general notation. Suppose that $(\mathbb{T}, \mathcal{A})$ is an arbitrary measure space. If $W : \mathbb{T} \to [1, +\infty)$ is a function, the W-norm of a function $f : \mathbb{T} \to \mathbb{R}$ is defined as $|f|_W := \sup_{x\in\mathbb{T}} |f(x)|/W(x)$. The set of measurable functions $f : \mathbb{T} \to \mathbb{R}$ with finite W-norm is denoted by $\mathcal{L}_W(\mathbb{T})$ or simply $\mathcal{L}_W$ when there is no ambiguity on the space $\mathbb{T}$. For a finite real-valued signed measure ν on $\mathbb{T}$, we denote the W-norm of ν as

$$\|\nu\|_W := \int W(x)|\nu|(dx) = \sup_{|f|_W\leq 1} \left|\int f(x)\nu(dx)\right|,$$

where $|\nu|$ is the total variation measure of ν. We denote $\mathcal{M}_W(\mathbb{T})$ the space of all finite real-valued signed measures ν on $\mathbb{T}$ such that $\|\nu\|_W < \infty$. It is well-known that $(\mathcal{M}_W(\mathbb{T}), \|\cdot\|_W)$ is a Banach space. When the measure space $\mathbb{T}$ is understood, we simply write $\mathcal{M}_W$. We will use the notation $\nu(f)$ to denote the integral $\int f(x)\nu(dx)$. If μ, ν are two finite signed measures on $(\mathbb{T}, \mathcal{A})$, we denote their product by $\mu\nu$, and the product of a finite number k of finite signed measures $\nu_1, \ldots, \nu_k$ is denoted by $\prod_{j=1}^k \nu_j$.

If Q is a transition kernel on $(\mathbb{T}, \mathcal{A})$, its iterates are defined as: Q^0 is the identity kernel $(Q^0(x, A) = \mathbf{1}_A(x))$ and for $n \geq 1$, we define $Q^n(x, \cdot) = \int Q(x, dz)Q^{n-1}(z, \cdot)$. If $h : \mathbb{T}\times\mathbb{T} \to \mathbb{R}$ is a bivariate function then Qh is the bivariate function defined by the rule $Qh(x, y) = \int Q(x, dz)h(y, z)$ and Q^2h is defined as $Q^2h(x_1, x_2) = \int Q(x_1, dz_1) \int Q(x_2, dz_2)h(z_1, z_2)$. If $h : \mathbb{T} \to \mathbb{R}$ is univariate, Qh is defined similarly as $Qh(x) = \int Q(x, dz)h(z)$. Fix Q a Markov kernel, and $V : \mathbb{T} \times \mathbb{T} \to [1, \infty)$. For $p \geq 1$ and a function $h : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$, we define

$$|||h|||_{p,V} := \sup_{x,y\in\mathbb{T}} \frac{\left(\int Q(x, dz)|h(y, z)|^p\right)^{1/p}}{V(x, y)}.$$

For a univariate function $V : \mathbb{T} \to [1, \infty)$ and for $h : \mathbb{T} \to \mathbb{R}$, we also define

$$|||h|||_{p,V} := \sup_{x\in\mathbb{T}} V(x)^{-1} \left(\int Q(x, dz)|h(z)|^p \right)^{1/p}.$$

When we use the notation $|||h|||_{p,V}$ below, it will always be with respect to P, the Markov kernel of the reference process $\{X_n, n \geq 0\}$, unless stated otherwise. For any $a \geq 0$ and $b \geq 0$, let $a \lesssim b$ denote $a \leq Cb$ for some finite positive constant C that may depend solely on the kernel P (but not, e.g., on the family of functions $\{h_n, n \geq 1\}$ considered).

The following short-range dependence concept will play an important rule.

Definition. Fix $r \in \mathbb{N}$. For measurable functions $\overline{V}_r \leq \overline{W}_r : \mathbb{T}^r \to [1, \infty)$, we say that the transition kernel Q with invariant distribution μ satisfies the condition $\mathsf{C}(r, \overline{V}_r, \overline{W}_r)$ if

$$\sum_{\ell_1\geq 0} \cdots \sum_{\ell_r \geq 0} \left\| \prod_{j=1}^{r} \left(Q^{\ell_j}(x_j, \cdot) - \mu \right) \right\|_{\overline{V}_r} \lesssim \overline{W}_r(x_1, \ldots, x_r), \quad (x_1, \ldots, x_r) \in \mathcal{X}^r. \tag{2}$$

Finally, we let $\overset{\mathsf{Pr}}{\to}$ and $\overset{\mathsf{w}}{\to}$ denote convergence in probability and weak convergence, respectively. All limits are taken as $n \to \infty$ unless explicitly noted otherwise.

2. A martingale approximation for quadratic forms

For notational convenience, we shall write $\bar{\mu}$ to denote the product probability measure $\bar{\mu}(du, dv) = \mu(du)\mu(dv)$, where μ is the invariant distribution of the Markov kernel P. Consider the following assumption.

Assumption A. There exist symmetric measurable functions $\overline{V}_2 \leq \overline{W}_2 : \mathcal{X} \times \mathcal{X} \to [1, \infty)$ such that P satisfies $\mathsf{C}(2, \overline{V}_2, \overline{W}_2)$. Furthermore, $P^s \overline{W}_2(x, y) < \infty$ for all $x, y \in \mathcal{X}$ and for $s \in \{1, 2\}$.

This assumption restricts the class of time-series considered in this paper. Intuitively, it requires the Markov chain $\{X_n, n \geq 0\}$ (with transition kernel P) to satisfy an ergodicity-type condition, which in this case is expressed in terms of summability of the V-uniform norm of a bivariate, in multiplicative form, centered product Markov kernel. The univariate version of this assumption (i.e., P satisfies $\mathsf{C}(1, \overline{V}_1, \overline{W}_1)$ for short) takes the familiar form from the Markov chains literature:

$$\sum_{\ell\geq 0} \left\| Q^{\ell}(x, \cdot) - \mu \right\|_{\overline{V}_1} \lesssim \overline{W}_1(x), \quad x \in \mathcal{X}, \tag{3}$$

where $\overline{V}_1, \overline{W}_1 : \mathcal{X} \to [1, \infty)$ are univariate functions satisfying $\overline{V}_1 \leq \overline{W}_1$. We require a bivariate version of this ergodicity condition because we are interested in the bivariate quadratic form $U_n(h_n)$.

Assumption A may be verified in a variety of ways, as the following remarks discuss.

Remark 1. It is always possible to deduce Assumption A from the univariate version (3). Indeed, if P satisfies $\mathsf{C}(1, V_1, W_1)$ and $\mathsf{C}(1, V_2, W_2)$, and $PW_1 < \infty$, $PW_2 < \infty$, define $\overline{V}_2(x, y) = V_1(x)V_2(y)$ and $\overline{W}_2(x, y) = W_1(x)W_2(y)$, then

$$\sum_{n\geq 0} \sum_{m\geq 0} \left\| \left(P^n(x, \cdot) - \mu\right) \left(P^m(y, \cdot) - \mu\right) \right\|_{\overline{V}_2} \lesssim \overline{W}_2(x, y)$$

because $\|(P^n(x,\cdot)-\mu)(P^m(y,\cdot)-\mu)\|_{\overline{V}_2} = \|P^n(x,\cdot)-\mu\|_{V_1}\|P^m(y,\cdot)-\mu\|_{V_2}$, implying that Assumption A holds.

Remark 2. The univariate condition $\mathsf{C}(1, V, W)$ holds for geometrically ergodic Markov kernels (that is, kernels P for which $\|P^n(x,\cdot)-\mu\|_V$ converges to zero exponentially fast for some $V \geq 1$). It also holds for sub-geometrically ergodic Markov kernels ($\|P^n(x,\cdot)-\mu\|_V$ converges to zero sub-geometrically) for which the rate of convergence is summable. It is sometimes possible to check the condition $\mathsf{C}(1, V, W)$ using Lyapunov drift conditions and their extensions; see, e.g., [6,16,18] for several examples.

We show that whenever Assumption A holds, there exists a martingale approximation to $U_n(h_n)$ that offers a simple route to study the asymptotics of $U_n(h_n)$. The space $\mathcal{M}_{\overline{V}_2}(\mathcal{X}\times\mathcal{X})$ of all finite signed measure ν on $\mathcal{X}\times\mathcal{X}$ such that $\|\nu\|_{\overline{V}_2} < \infty$, equipped with the norm $\|\cdot\|_{\overline{V}_2}$, is a Banach space. Under Assumption A, and for any $x, y \in \mathcal{X}$,

$$\bar{R}_2(x, y; (du, dv)) := \sum_{n_1\geq 0}\sum_{n_2\geq 0}\left(P^{n_1}(x, du) - \mu(du)\right)\left(P^{n_2}(y, dv) - \mu(dv)\right)$$

is a finite signed measure that belongs to $\mathcal{M}_{\overline{V}_2}(\mathcal{X}\times\mathcal{X})$. Furthermore, $\|\bar{R}_2(x, y; \cdot)\|_{\overline{V}_2} \lesssim \overline{W}_2(x, y)$ for all $x, y \in \mathcal{X}$. Let $h : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ be a symmetric measurable function such that $\bar{\mu}(|h|) < \infty$. (Recall $\bar{\mu}(du, dv) = \mu(du)\mu(dv)$.) Let $\theta = \bar{\mu}(h) = \int\int h(x, y)\mu(dx)\mu(dy)$ and define

$$\bar{h}_1(x) := \int h(x, z)\mu(dz) - \theta, \qquad \bar{h}_2(x, y) = h(x, y) - \bar{h}_1(x) - \bar{h}_1(y) - \theta,$$

$$\bar{G}_2(x, y) := \int\int \bar{R}_2(x, y; dz_1, dz_2)\bar{h}_2(z_1, z_2), \quad x, y \in \mathcal{X}.$$

We say that h is degenerate when $\bar{h}_1$ is identically zero. Notice that $\bar{G}_2$ is defined from the reduced form $\bar{h}_2$ of h, not h itself. For $x \in \mathcal{X}$, δ_x denotes the Dirac measure at x.

Lemma 2.1. *Suppose Assumption A holds and $\bar{h}_2 \in \mathcal{L}_{\overline{V}_2}$. Then $\bar{G}_2$ is well-defined, $\bar{G}_2 \in \mathcal{L}_{\overline{W}_2}$, and $|\bar{G}_2|_{\overline{W}_2} \lesssim |\bar{h}_2|_{\overline{V}_2}$ and for all $x, y \in \mathcal{X}$,*

$$\bar{h}_2(x, y) = \int (\delta_x(dz_1) - P(x, dz_1)) \int \left(\delta_y(dz_2) - P(y, dz_2)\right) \bar{G}_2(z_1, z_2). \tag{4}$$

If, in addition, $P^s\bar{h}_2 \in \mathcal{L}_{\overline{V}_2}$, then $|P^s\bar{G}_2|_{\overline{W}_2} \lesssim |P^s\bar{h}_2|_{\overline{V}_2}$ for $s \in \{1, 2\}$.

Proof. See Section 5.1.1. □

Remark 3. To help put this result in perspective, we recall briefly the classical univariate Poisson equation. The Poisson equation is a well-known technique for deriving martingale approximation for partial sums of dependent processes, particularly for Markov chains. The idea is as follows. If $\{X_n, n \geq 0\}$ is a Markov chain with invariant distribution μ and transition kernel P, and h is a function such that $\mu(h) = 0$, set $g(x) = \sum_{j\geq 0} P^j h(x)$. If g is well defined, then it solves the equation

$$g(x) - Pg(x) = h(x), \quad x \in \mathcal{X}, \tag{5}$$

known as the Poisson equation (for h and P). The usefulness of g comes from the fact that we can use it to rewrite the partial sum $\sum_{k=1}^n h(X_k)$ as

$$\sum_{k=1}^n h(X_k) = \sum_{k=1}^n (g(X_k) - Pg(X_k)) = \sum_{k=1}^n (g(X_k) - Pg(X_{k-1})) + R_n,$$

where $R_n = Pg(X_0) - Pg(X_n)$ is negligible, and $\sum_{k=1}^n (g(X_k) - Pg(X_{k-1}))$ is a martingale that approximates $\sum_{k=1}^n h(X_k)$. Therefore, limit theorems for $\sum_{k=1}^n h(X_k)$, particularly central limit theorems, may be proved by deriving the corresponding results for the martingale $\sum_{k=1}^n (g(X_k) - Pg(X_{k-1}))$. More details and references on these ideas can be found in [14,17]. In the present case, Eq. (4) gives a bivariate version of the univariate Poisson equation (5), and we shall use it below to derive a martingale approximation for quadratic forms.

We introduce the function, for any $x_1, x_2, y_1, y_2 \in \mathcal{X}$,

$$\begin{aligned} \Lambda_2(x_1, x_2; y_1, y_2) &= \int \left(\delta_{y_1}(dz_1) - P(x_1, dz_1)\right) \int \left(\delta_{y_2}(dz_2) - P(x_2, dz_2)\right) \bar{G}_2(z_1, z_2) \\ &= \bar{G}_2(y_1, y_2) - P\bar{G}_2(x_2, y_1) - P\bar{G}(x_1, y_2) + P^2\bar{G}_2(x_1, x_2). \end{aligned} \tag{6}$$

Then (4) can be written as $\bar{h}_2(x, y) = \Lambda_2(x, y; x, y)$, with the key property of Λ_2:

$$\int P(x, dy)\Lambda_2(u, x; v, y) = \int P(u, dv)\Lambda_2(u, x; v, y) = 0, \quad x, y, u, v \in \mathcal{X}. \tag{7}$$

Now suppose that we have $\{h_n : \mathcal{X} \times \mathcal{X} \to \mathbb{R}\}$, a family of symmetric measurable functions such that $\bar{\mu}(|h_n|) < \infty$. We write θ_n, $\bar{h}_{n,1}$, $\bar{h}_{n,2}$, $\bar{G}_{n,2}$, and $\Lambda_{n,2}$ to denote respectively the quantities $\theta, \bar{h}_1, \bar{h}_2, \bar{G}_2$, and Λ_2 defined above with $h = h_n$. From the definition of these functions we have

$$h_n(x, y) = \theta_n + \bar{h}_{n,1}(x) + \bar{h}_{n,1}(y) + \bar{h}_{n,2}(x, y), \quad x, y \in \mathcal{X}.$$

This implies, after some trivial rearrangements that

$$U_n(h_n) = U_{n,0} + \sum_{\ell=1}^n w_{n,1}(\ell)\bar{h}_{n,1}(X_\ell) + \sum_{\ell=1}^n \sum_{j=1}^\ell w_n(\ell, j)\bar{h}_{n,2}(X_\ell, X_j), \tag{8}$$

where

$$U_{n,0} = \theta_n \sum_{\ell=1}^n \sum_{j=1}^\ell w_n(\ell, j), \qquad w_{n,1}(\ell) := \sum_{j=1}^\ell w_n(\ell, j) + \sum_{j=\ell}^n w_n(j, \ell).$$

For $1 \le j \le \ell \le n$, we introduce the random variables

$$Q_{n,\ell,j} := \Lambda_{n,2}\left(X_{j-1}, X_{\ell-1}, X_j, X_\ell\right).$$

For $j < \ell$, and by the Markov property and (7), we have

$$\mathbb{E}[Q_{n,\ell,j}|\mathcal{F}_{\ell-1}] = \int P(X_{\ell-1}, dz)\Lambda_{n,2}\left(X_{j-1}, X_{\ell-1}, X_j, z\right) = 0,$$

almost surely. This shows that $\{(\sum_{j=1}^{\ell-1} Q_{n,\ell,j}, \mathcal{F}_\ell), 2 \le \ell \le n\}$ is a martingale-difference array.

The representation (8), the Poisson equation (4), and (6), give the main result of the paper.

Lemma 2.2. *Suppose $\{X_n, n \geq 0\}$ is a Markov chain with transition kernel P satisfying Assumption A, and $\bar{h}_{n,2} \in \mathcal{L}_{\overline{V}_2}$ for each $n \geq 1$. Then, for the quadratic form $U_n(h_n)$ given in (1),*

$$U_n(h_n) = U_{n,0} + \hat{U}_n(h_n) + \sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1} w_n(\ell, j)Q_{n,\ell,j} + \zeta_n,$$

where

$$U_{n,0} = \theta_n \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell, j), \qquad \hat{U}_n(h_n) = \sum_{\ell=1}^{n} \left\{ w_{n,1}(\ell)\bar{h}_{n,1}(X_\ell) + w_n(\ell,\ell)Q_{n,\ell,\ell} \right\},$$

and

$$\begin{aligned}\zeta_n = & \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell, j)\left(P\bar{G}_{n,2}(X_{\ell-1}, X_j) - P\bar{G}_{n,2}(X_\ell, X_j)\right) \\ & + \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell, j)\left(P\bar{G}_{n,2}(X_{j-1}, X_\ell) - P\bar{G}_{n,2}(X_j, X_\ell)\right) \\ & + \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell, j)\left(P^2\bar{G}_{n,2}(X_j, X_\ell) - P^2\bar{G}_{n,2}(X_{j-1}, X_{\ell-1})\right),\end{aligned}$$

with $\bar{G}_{n,2}(x, y) := \int\int \bar{R}_2(x, y; dz_1, dz_2)\bar{h}_{n,2}(z_1, z_2)$, $x, y \in \mathcal{X}$.

Proof. See Section 5.1.2. □

This lemma shows that the quadratic form $U_n(h_n)$ admits a Hoeffding-type decomposition when $\{X_n, n \geq 0\}$ is a Markov chain satisfying Assumption A, a quite general short-range requirement. Specifically, this decomposition leads to four terms: a non-random overall mean ($U_{n,0}$), a random single-average ($\hat{U}_n(h_n)$), a random double-average ($\sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1} w_n(\ell, j)Q_{n,\ell,j}$), and a random reminder term (ζ_n). The single-average and double-average terms are (possibly correlated) martingale differences with respect to the natural filtration. This decomposition offers a simple and useful way to analyzing quadratic forms of Markov chains via conventional martingale results, whenever the additional "remainder" term ζ_n may be ignored.

Remark 4. The usefulness of the decomposition in Lemma 2.2 comes from the fact that the remainder ζ_n often can be shown to be negligible. In the upcoming sections we offer a few examples where this remainder is indeed shown to be asymptotically small. As a result, one can easily study the asymptotic behavior of $U_n(h_n)$ by focusing only on the linear term $\hat{U}_n(h_n)$ and the quadratic martingale $\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1} w_n(\ell, j)Q_{n,\ell,j}$.

Remark 5. The linear martingale $\hat{U}_n(h_n)$ gives an analogue of the well-known "Hájek projection" of a U-statistic, while the quadratic martingale $\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1} w_n(\ell, j)Q_{n,\ell,j}$ provides an analogue of the remainder of this projection. Unlike the case of independent random variables, this martingale decomposition is not exact and leads to linear and quadratic terms that are correlated in general.

3. Application: asymptotic variance estimation

In this section, we apply Lemma 2.2 to study the asymptotics of lag-windows estimators of asymptotic variance in time series. Let $h : \mathcal{X} \to \mathbb{R}$ be a measurable function such that $\mu(|h|^2) < \infty$. We assume without any loss of generality that $\mu(h) = 0$. We are interested in the estimation of the long-run variance (or the asymptotic variance) of h defined as:

$$\sigma^2(h) = \mathsf{Var}(h(\check{X}_0)) + 2\sum_{\ell\geq 1} \mathsf{Cov}(h(\check{X}_0), h(\check{X}_\ell)), \tag{9}$$

where $\{\check{X}_n, n \geq 0\}$ is the stationary Markov chain with transition kernel P and initial distribution μ (i.e., when $\rho = \mu$ in our notation). More generally, we employ the conventional notation $\mathsf{Var}_\mu(h(X_0)) := \int \mu(dx)h^2(x)$, and $\mathsf{Cov}_\mu(h(X_0), h(X_\ell)) := \int \mu(dx)\mathbb{E}_x[h(X_0)h(X_\ell)]$. The long-run variance plays an important role in time series analysis. The population parameter $\sigma^2(h)$ may be interpreted as the long-run variance of a covariance-stationary time series $\{h(X_n) : n \geq 0\}$. More precisely, under Assumption A, $\sigma^2(h) = \lim_{n\to\infty} n\mathsf{Var}_\mu(n^{-1}\sum_{\ell=1}^n h(X_\ell))$ when $\{X_n : n \geq 0\}$ is a Markov chain with invariant distribution μ and transition kernel P.

A classical estimator for $\sigma^2(h)$ is the lag-windows estimator defined as

$$\Gamma^2_{n,b}(h) := \gamma_{n,0} + 2\sum_{k=1}^{n-1} w_b(kc_n^{-1})\gamma_{n,k} = \sum_{k=-n+1}^{n-1} w_b\left(\frac{k}{c_n}\right)\gamma_{n,|k|},$$

where $\gamma_{n,k} := n^{-1}\sum_{j=1}^{n-k}\left(h(X_j) - \mu_n(h)\right)\left(h(X_{j+k}) - \mu_n(h)\right)$ is the k-th order sample autocovariance with $\mu_n(h) = n^{-1}\sum_{j=1}^n h(X_j)$, $w_b : \mathbb{R} \to [0, 1]$ is a weight function (with parameter $b > 0$) such that $w(-x) = w(x)$, and $\{c_n, n \geq 1\}$ is an increasing sequence of positive numbers. We refer the reader to [21] for detailed discussion on lag-windows estimators. By re-arranging the summations, $\Gamma^2_{n,b}(h)$ can also be written as

$$\begin{aligned}\Gamma^2_{n,b}(h) &= \frac{1}{n}\sum_{\ell=1}^n\sum_{j=1}^n w_b\left(\frac{\ell - j}{c_n}\right)(h(X_\ell) - \mu_n(h))\left(h(X_j) - \mu_n(h)\right)\\ &= \sum_{\ell=1}^n\sum_{j=1}^{\ell} w_{n,b}(\ell - j)h(X_\ell)h(X_j) + R_n,\end{aligned} \tag{10}$$

where

$$R_n = -\frac{2}{n}\left(\sum_{\ell=1}^n v_n(\ell)h(X_\ell)\right)\left(\sum_{j=1}^n h(X_j)\right) + \frac{u_n}{n}\left(\sum_{j=1}^n h(X_j)\right)^2, \tag{11}$$

and $w_{n,b}(k) = 2n^{-1}w_b\left(\frac{k}{c_n}\right)$, $w_{n,b}(0) = 1/n$, $v_n(\ell) = \frac{1}{n}\sum_{i=1}^n w_b\left(\frac{\ell-i}{c_n}\right)$, $u_n = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n w_b\left(\frac{j-i}{c_n}\right)$.

As we can see from the expression (10), $\Gamma^2_{n,b}(h)$ is a quadratic form of the type (1), up to the term R_n, to which Lemma 2.2 can be applied. We consider the following weight functions.

Assumption W. For some $b > 0$, $w_b : \mathbb{R} \to [0, 1]$ is a continuous function with support $[-b, b]$, of class $\mathcal{C}^2$ on the interval $(0, b)$, such that $w_b(b) = 0$ and $w_b(0) = 1$.

This assumption allows for the use of all commonly employed weighting functions, including the Bartlett and Parzen kernels. We impose the following ergodicity assumption. (Recall P denotes the transition kernel of the Markov chain; see Section 1.1 for details.)

Assumption B. There exist measurable functions $V_k : \mathcal{X} \to [1, \infty)$, $k = 1, 2, 3$, $V_1 \le V_2$, $V_2^2 \le V_3$, such that $|PV_2^2|_{V_2^2} < \infty$, $PV_3(x) < \infty$ for all $x \in \mathcal{X}$, and P satisfies the assumptions $\mathsf{C}(1, V_1, V_2)$ and $\mathsf{C}(1, V_2^2, V_3)$. Furthermore, there exists $q > 1$ such that $\sup_{n\ge 0} \mathbb{E}[V_3^q(X_n)] < \infty$.

Assumption B implies Assumption A with $\overline{V}_2(x, y) = V_1(x)V_1(y)$ and $\overline{W}_2(x, y) = V_2(x)V_2(y)$. Therefore, Lemma 2.2 applies with $h_n(x, y) = h(x)h(y)$. In this case, $h_{n,1}(x) = \int h(x)h(y)\mu(dy) = 0$, $\theta_n = 0$, and $h_{n,2}(x, y) = h(x)h(y)$. Define, for any $x \in \mathcal{X}$,

$$G(x) := \sum_{j\ge 0} P^j h(x), \quad \text{and} \quad PG(x) = \int P(x, dz)G(z).$$

Then $\bar{G}_2(x, y) = G(x)G(y)$, $P\bar{G}_2(x, y) = PG(x)G(y)$, and $P^2\bar{G}_2(x, y) = PG(x)PG(y)$. Therefore,

$$Q_{n,\ell,j} = Q_\ell Q_j \quad \text{with } Q_\ell = G(X_\ell) - PG(X_{\ell-1}). \tag{12}$$

As above, $\{(Q_\ell, \mathcal{F}_\ell), \ell \ge 1\}$ is a martingale: $\mathbb{E}[Q_\ell|\mathcal{F}_{\ell-1}] = 0$. The proof of the following result employs Lemma 2.2 as its main ingredient.

Theorem 3.1. *Suppose $\{X_n, n \ge 0\}$ is a Markov chain with transition kernel P satisfying Assumption B, Assumption W holds, and $h \in \mathcal{L}_{V_1}$. For all $n \ge 1$,*

$$\Gamma^2_{n,b}(h) = n^{-1} \sum_{\ell=1}^{n} Q_\ell^2 + \sum_{\ell=1}^{n} \sum_{j=1}^{\ell-1} w_{n,b}(\ell - j) Q_\ell Q_j + R_n + \zeta_n, \tag{13}$$

where $w_{n,b}(k) = 2n^{-1} w_b(k/c_n)$, $k \ne 0$. The exact form of ζ_n is given in the proof to save notation.

Furthermore, there exist $p > 1$ such that for all $n \ge 3$,

$$\mathbb{E}^{1/p}[|\zeta_n|^p] \lesssim c_n^{-1+\frac{1}{2}\vee\frac{1}{p}}, \qquad \mathbb{E}^{1/p}[|R_n|^p] \lesssim \frac{c_n}{n} + n^{-1/2},$$

$$\mathbb{E}^{1/p}\left[\left|\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1} w_{n,b}(\ell - j)Q_\ell Q_j\right|^p\right] \lesssim \left(\frac{c_n}{n}\right)^{\frac{1}{2}} n^{-\frac{1}{2}+\frac{1}{p}\vee\frac{1}{2}}.$$

Proof. See Section 5.2.1. □

A clearer picture of the behavior of the lag-window estimator emerges from this result. For $p \ge 2$,

$$\Gamma^2_{n,b}(h) = \underbrace{n^{-1} \sum_{\ell=1}^{n} Q_\ell^2}_{O_p(1)} + \underbrace{\sum_{\ell=1}^{n} \sum_{j=1}^{\ell-1} w_{n,b}(\ell - j) Q_\ell Q_j + R_n}_{O_p\left(\sqrt{\frac{c_n}{n}} + \frac{c_n}{n} + n^{-1/2}\right)} + \underbrace{\zeta_n}_{O_p(c_n^{-1/2})}. \tag{14}$$

By the law of large numbers for Markov chain the term $n^{-1}\sum_{\ell=1}^{n} Q_\ell^2 \stackrel{\mathsf{Pr}}{\to} \sigma^2(h)$. As the result, Theorem 3.1 implies that $\Gamma_{n,b}^2(h) \stackrel{\mathsf{Pr}}{\to} \sigma^2(h)$ provided $c_n \to \infty$, $c_n = o(n)$ and $p \geq 2$ (for $1 < p < 2$, specific rate assumption on c_n might be needed). The decomposition (14) also gives some insight into the well known fact that $\Gamma_{n,b}(h)$ often has poor finite-sample properties in estimating $\sigma^2(h)$, particularly for highly correlated time-series. Indeed, for $c_n = o(n)$, both terms $R_n + \sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1} w_{n,b}(\ell-j)Q_\ell Q_j$ and ζ_n converge to zero but at antagonistic rates. If $c_n \approx n$, then $\zeta_n \approx O_p(n^{-1/2})$ but $R_n + \sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1} w_{n,b}(\ell-j)Q_\ell Q_j \approx O_p(1)$. Whereas for $c_n \ll n$, the convergence of ζ_n is slow ($\zeta_n = O_p(c_n^{-1/2})$) but $R_n+\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1} w_{n,b}(\ell-j)Q_\ell Q_j$ vanishes quickly.

We now show how Theorem 3.1 can be used to derive confidence intervals for $\mu(h)$. When the goal is to construct a confidence interval for $\mu(h)$ (and one is not interested in estimating $\sigma^2(h)$ per se), it has been suggested to use the lag-window estimator $\Gamma_{n,b}^2(h)$ with $c_n = n$, the so-called "fixed-b asymptotics" [19,12]. With $c_n = n$, $\Gamma_{n,b}^2(h)$ no longer converges to $\sigma^2(h)$, but asymptotically valid confidence intervals can still be derived for $\mu(h)$.

Theorem 3.2. *Suppose the assumptions of Theorem 3.1 hold.*

(1) *If $p \geq 2$ and $c_n = o(n)$, then $\Gamma_{n,b}^2(h) \stackrel{\mathsf{Pr}}{\to} \sigma^2(h)$. Furthermore, assuming $\Gamma_{n,b}^2(h) > 0$ almost surely,*

$$\left\{n\Gamma_{n,b}^2(h)\right\}^{-1/2}\sum_{j=1}^{n}\left(h(X_j)-\mu(h)\right) \stackrel{\mathsf{w}}{\to} \mathcal{N}(0,1).$$

(2) *Let $\{B(t),\ 0 \leq t \leq 1\}$ be the standard Brownian motion. If $c_n = n$, then $\Gamma_{n,b}^2(h) \stackrel{\mathsf{w}}{\to} \sigma^2(h)\mathsf{K}_b$, where*

$$\begin{aligned}\mathsf{K}_b &= 1 + 2\int_0^1\int_0^t w_b(t-s)dB(s)dB(t)\\ &\quad - 2B(1)\int_0^1 g_b(t)dB(t) + B^2(1)\int_0^1 g_b(t)dt,\end{aligned}$$

where $g_b(t) = \int_0^1 w_b(t-u)du$. Furthermore, assuming $\Gamma_{n,b}^2(h) > 0$ almost surely,

$$\left\{n\Gamma_{n,b}^2(h)\right\}^{-1/2}\sum_{j=1}^{n}\left(h(X_j)-\mu(h)\right) \stackrel{\mathsf{w}}{\to} \frac{B(1)}{\sqrt{\mathsf{K}_b}}.$$

Proof. See Section 5.2.2. □

By Theorem 3.2(1) an asymptotically valid $(1-\alpha)$-confidence interval for $\mu(h)$ is

$$\mu_n(h) \pm z_{1-\alpha/2}\frac{\hat{\sigma}_n(h)}{\sqrt{n}}, \tag{15}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution and where $\hat{\sigma}_n(h) = \sqrt{\Gamma_{n,b}^2(h)}$, with $b = 1$, $c_n = o(n)$. Typical choice of c_n includes $c_n \propto n^\delta$, with $\delta \in (0,1)$ typically around 0.5. Theorem 3.2(2) justifies another asymptotically valid confidence interval

Table 1
0.975-quantile of the distribution of $B(1)/\sqrt{\mathsf{K}_b}$.

	$w_b(x) = 1 - x/b$	$w_b(x) = 1 - (x/b)^2$	$w_b(x) = \mathbf{1}_{(0,1)}(x/b)$
$b = 0.3$	2.828	4.134	5.496
$b = 0.5$	3.557	6.580	6.299
$b = 0.9$	4.735	12.575	13.045

for $\mu(h)$:

$$\mu_n(h) \pm t_{1-\alpha/2}\frac{\tilde{\sigma}_n(h)}{\sqrt{n}}, \tag{16}$$

where $t_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the distribution of $B(1)/\sqrt{\mathsf{K}_b}$ and where $\tilde{\sigma}_n(h) = \sqrt{\Gamma^2_{n,b}(h)}$, with $c_n = bn$, with $b \in (0, 1)$.

Although the limiting distribution $B(1)/\sqrt{\mathsf{K}_b}$ is non-standard, it can be simulated, for example by discretization of the stochastic integrals in K_b. We report in Table 1 the 95% quantiles of the distribution of $B(1)/\sqrt{\mathsf{K}_b}$ using $w_b(x) = \mathbf{1}_{(0,b)}(x)$, $w_b(x) = (1 - x/b)\mathbf{1}_{(0,b)}(x)$ and $w_b(x) = (1 - (x/b)^2)\mathbf{1}_{(0,b)}(x)$, and for different values of b, based on 10,000 replications of $B(1)/\sqrt{\mathsf{K}_b}$. The distribution departs further from the standard normal distribution as b increases.

Remark 6. These results can be employed in the context of Markov Chain Monte Carlo (MCMC), a popular computational tool to obtain random samples from intractable and high-dimensional distributions (for review see, e.g., [9,22]). To be concrete, suppose we are interested in sampling from the probability measure μ to compute the integral $\mu(h) = \int h(x)\mu(dx)$. Let $\{X_n, n \geq 0\}$ be Markov chains with transition kernel P, invariant distribution μ and initial distribution ρ. By simulating the Markov chain, we approximate $\mu(h)$ by the Monte Carlo average $\mu_n(h) = n^{-1}\sum_{k=1}^n h(X_k)$. Furthermore, under appropriate assumptions, $\lim_{n\to\infty} n^{1/2}\mathsf{Var}(\mu_n(h)) = \sigma^2(h)$, as given by (9), and a central limit theorem holds: $n^{-1/2}\sum_{k=1}^n (h(X_k) - \pi(h)) \overset{\mathsf{w}}{\to} \mathcal{N}(0, \sigma^2(h))$ (see, e.g., [18, Chapter 17]). Therefore (15) and (16) provide two valid confidence intervals for $\mu(h)$, under the weak conditions imposed in Theorem 3.1.

To illustrate the applicability of our results, we present next a simple simulation example that compares the finite-sample properties of the two alternative confidence intervals obtained from Theorem 3.2 in terms of coverage probability and interval length. All the simulations are performed using the Bartlett kernel $w(x) = 1 - x$.

3.1. Illustration: a GARCH(1, 1) model

Consider the following simple linear GARCH(1, 1) model. For $h_0 \in (0, \infty)$, $u_0 \sim \mathcal{N}(0, h_0)$ and $n \geq 1$,

$$u_n = h_n^{1/2}\epsilon_n$$
$$h_n = \omega + \beta h_{n-1} + \alpha u_{n-1}^2,$$

where $\{\epsilon_n, n \geq 0\}$ is i.i.d. $\mathcal{N}(0, 1)$ and $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$. We assume that the parameters α and β satisfy $\mathbb{E}\left[\left(\beta + \alpha Z^2\right)^{\nu}\right] < 1$, for $Z \sim \mathcal{N}(0, 1)$ and some $\nu > 0$.

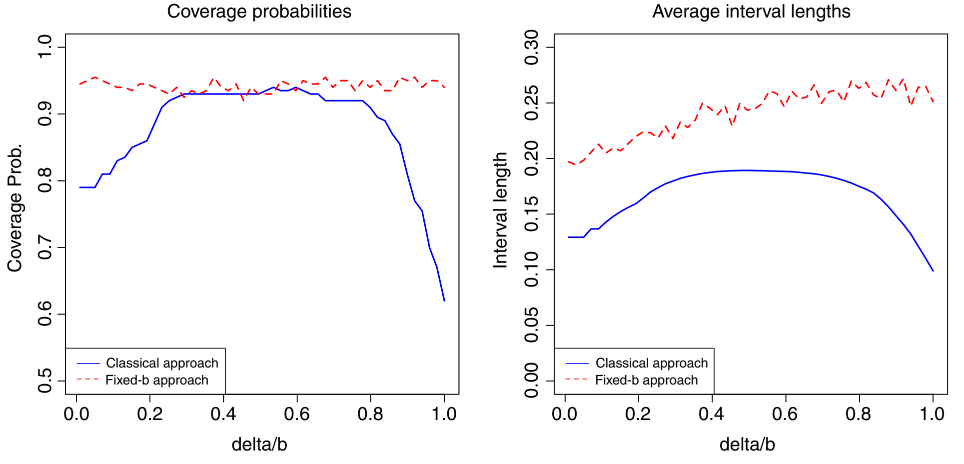

Fig. 1. Coverage probabilities plots for various values of δ (classical confidence interval) and b (fixed-b confidence interval).

Under these conditions, [16, Theorem 2] showed that the joint process $\{(u_n, h_n), n \geq 0\}$ is a phi-irreducible aperiodic Markov chain that admits an invariant distribution and is geometrically ergodic with a drift function $V(u,h) = 1 + h^\nu + |u|^{2\nu}$. Therefore, for $\nu \geq 2$, Assumption B holds with $V_1 = V_2 = V^{1/2}$ and $V_3 = V$. We are interested in a confidence interval for $\mu(h)$ with $h(u) = u^2$, which belongs to $\mathcal{L}_{V_1}$. The exact value is $\mu(h) = \omega(1-\alpha-\beta)^{-1}$. As a consequence, all the assumptions in Theorems 3.1 and 3.2 are satisfied in this example.

For the simulations we set $\omega = 1, \alpha = 0.1, \beta = 0.7$ which gives $\mu(h) = 5$. We compare the confidence intervals (15) and (16) by computing (by Monte Carlo) their coverage probabilities and average lengths. The comparison is performed using sample paths of length 60,000 from the GARCH(1, 1) Markov chain. The results are plotted in Fig. 1, which shows across the board better coverage probability of the fixed-b confidence interval but, as expected, at the expense of a slightly wider confidence intervals.

4. Application: U-statistics with varying kernels

As a second application of Lemma 2.2, we study the asymptotic behavior of U-statistics with varying kernels. These statistics are another special case of quadratic forms and correspond to setting $w_n(\ell,\ell) = 0$ and $w_n(\ell, j) = \binom{n}{2}^{-1}$ if $\ell \neq j$. We thus have

$$U_n(h_n) = \binom{n}{2}^{-1} \sum_{\ell=2}^{n} \sum_{j=1}^{\ell-1} h_n(X_\ell, X_j). \tag{17}$$

For U-statistics with a fix kernel ($h_n = h$), one typically distinguishes the cases h degenerate versus h non-degenerate. When h is non-degenerate the linear part of $U_n(h)$ is asymptotically leading ($U_n(h)$ is asymptotically linear), whereas when h is degenerate, the quadratic part is asymptotically leading ($U_n(h)$ is intrinsically quadratic). U-statistics with varying kernels have much richer behaviors, since in this case either terms (or both) may be asymptotically leading. U-statistics with varying kernels play an important role in nonparametric and semiparametric statistics and econometrics. As an example, we will illustrate further below how the results obtained here apply to the problem of kernel-based semiparametric density-weighted average derivatives.

In the present case, under Assumption A, Lemma 2.2 reduces to

$$U_n(h_n) = \theta_n + \frac{2}{n}\sum_{\ell=1}^{n} \bar{h}_{n,1}(X_\ell) + \binom{n}{2}^{-1}\sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1} Q_{n,\ell,j} + \zeta_n,$$

where $\theta_n = \int\int h_n(x,y)\mu(dx)\mu(dy)$, and by taking advantage of the telescoping terms in the expression of ζ_n in Lemma 2.2, we get

$$\binom{n}{2}\zeta_n = \sum_{\ell=1}^{n-1}\big(P\bar{G}_{n,2}(X_0, X_{\ell+1}) - P\bar{G}_{n,2}(X_\ell, X_{\ell+1}) + P\bar{G}_{n,2}(X_\ell, X_\ell) - P\bar{G}_{n,2}(X_n, X_\ell)\big) + \sum_{\ell=1}^{n-1}\Big(P^2\bar{G}_{n,2}(X_\ell, X_n) - P^2\bar{G}_{n,2}(X_0, X_\ell)\Big). \tag{18}$$

We impose the following two (moment) assumptions.

Assumption C1. With $\overline{V}_2$ and $\overline{W}_2$ as in Assumption A, $P\overline{V}_2 \lesssim \overline{V}_2$.
Furthermore, $\sup_{\ell,j\geq 0}\mathbb{E}[\overline{W}_2^p(X_\ell, X_j)] < \infty$ for $p = 2$.

Assumption C2. With $\overline{W}_2$ as in Assumption A, there exist measurable functions $\overline{\mathcal{U}}_0, \overline{\mathcal{U}}_1 \leq \overline{\mathcal{V}}_1 : \mathcal{X} \to [1,\infty)$, such that P satisfies $\mathsf{C}(1, \overline{\mathcal{U}}_1, \overline{\mathcal{V}}_1)$ and for all $m \geq 0$ and all $x_0, x_1 \in \mathcal{X}$,

$$\int P(x_0, dz)\mathbb{E}_z\left[\left(\overline{W}_2(X_m, z) + \overline{W}_2(X_m, x_0)\right)\overline{W}_2(X_m, x_1)\right] \lesssim \overline{\mathcal{U}}_1(x_0)\overline{\mathcal{U}}_0(x_1). \tag{19}$$

Furthermore, $\sup_{\ell\geq 1}\mathbb{E}[\overline{\mathcal{V}}_1(X_\ell)[\overline{\mathcal{U}}_0(X_\ell) + \overline{\mathcal{U}}_0(X_{\ell-1})]] < \infty$.

Remark 7. Assumptions C1 and C2 are similar to the assumptions imposed in [5, Theorem 1.8] to obtain a CLT for U-statistics of stationary dependent processes. If $\overline{W}_2$ has a multiplicative form, these assumptions boil down to a univariate moment assumption on the Markov chain. For example, if $\overline{W}_2(x,y) = W(x)W(y)$, then C1 holds if $\sup_{n\geq 0}\mathbb{E}[W^{2p}(X_n)] < \infty$ (with $p = 2$). Furthermore, if $\mathbb{E}_x[W^{2p}(X_m)] \lesssim M(x)$, for some function M for which $PM \lesssim M$, and $P\{MW\} \lesssim MW$, then the right hand side of (19) can be taken as $\{M(x_0)W(x_0)\}W(x_1)$. The latter conditions can be checked for example when the Markov chain P is geometrically with a known drift function [6,16,18].

The next theorem gives control over the higher order components of $U_n(h_n)$.

Theorem 4.1. *Suppose $\{X_n, n \geq 0\}$ is a Markov chain with transition kernel P satisfying Assumption A, and $\bar{h}_{n,2} \in \mathcal{L}_{\overline{V}_2}$ for each $n \geq 1$. (Recall that ζ_n is given in (18).)*

(1) *If Assumption C1 holds, then*

$$\mathbb{E}^{1/2}[|\zeta_n|^2] \lesssim n^{-1}|P\bar{h}_{n,2}|_{\overline{V}_2}, \quad \text{and} \quad \mathbb{E}^{1/2}[Q_{n,\ell,j}^2] \lesssim \left\|\bar{h}_{n,2}\right\|_{2,\overline{V}_2}.$$

(2) *If Assumptions C1 and C2 hold, then*

$$\mathbb{E}^{1/2}\left[\left(\sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)^2\right] \lesssim n\left\|\bar{h}_{n,2}\right\|_{2,\overline{V}_2}.$$

Proof. See Section 5.3.1. □

Theorem 4.1 leads to a Hoeffding-type decomposition for U-statistics with varying kernel:

$$U_n(h_n) - \theta_n = \underbrace{\frac{2}{n}\sum_{\ell=1}^{n} \bar{h}_{n,1}(X_\ell)}_{O_p\left(n^{-1/2}\|\|\bar{h}_{n,1}\|\|_{2,V_1}\right)} + \underbrace{\binom{n}{2}^{-1}\sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1} Q_{n,\ell,j}}_{O_p\left(n^{-1}\|\|\bar{h}_{n,2}\|\|_{2,\overline{V}_2}\right)} + \underbrace{\zeta_n}_{O_p\left(n^{-1}|P\bar{h}_{n,2}|_{\overline{V}_2}\right)} . \tag{20}$$

Clearly, we have $|P\bar{h}_{n,2}|_{\overline{V}_2} \le \left\|\left|\bar{h}_{n,2}\right|\right\|_{2,\overline{V}_2}$. But, in fact, $|P\bar{h}_{n,2}|_{\overline{V}_2} = o(\left\|\left|\bar{h}_{n,2}\right|\right\|_{2,\overline{V}_2})$ in many instances. For a simple example, suppose $X_n \overset{i.i.d.}{\sim} \mu$, then $P(x,\cdot)=\mu$ for all x, and $|P\bar{h}_{n,2}|_{\overline{V}_2}=0$. Another example is given in our illustration below; see Remark 9.

The next result gives conditions under which the asymptotically leading term in $U_n(h_n)$ is the linear term $\frac{2}{n}\sum_{\ell=1}^{n}\bar{h}_{n,1}(X_\ell)$. Define

$$\sigma_{n,1}^2 := \mathsf{Var}_\mu\left(h_{n,1}(X_0)\right) + 2\sum_{\ell\ge 1}\mathsf{Cov}_\mu\left(h_{n,1}(X_0), h_{n,1}(X_\ell)\right).$$

In addition, we introduce the following assumption to establish asymptotic normality of the linear term under high-level conditions. More easily verifiable conditions can be derived if the dependence of $h_{n,1}$ on n is made explicit.

Assumption C3. Suppose that $h_{n,1} \in \mathcal{L}_{\overline{\mathcal{U}}_1}$ for all $n \ge 0$, with $\overline{\mathcal{U}}_1$ as in Assumption C2. In addition, for $g_n(x) := \sum_{j\ge 0} P^j\bar{h}_{n,1}(x)$,

$$\frac{1}{n\sigma_{n,1}^2}\sum_{k=1}^{n}\left\{Pg_n^2(X_k) - (Pg_n(X_k))^2\right\} \overset{\mathsf{Pr}}{\to} 1,$$

and, for all $\epsilon > 0$,

$$\frac{1}{n\sigma_{n,1}^2}\sum_{k=1}^{n}\int P(X_{k-1},du)g_n^2(u)\mathbf{1}_{\{|g_n(u)|\ge\epsilon\sigma_{n,1}\sqrt{n}/2\}} \overset{\mathsf{Pr}}{\to} 0.$$

Theorem 4.2. *Suppose the assumptions of Theorem 4.1 hold. For the U-statistic in (17), if $\sigma_{n,1} > 0$ and $n^{-1/2}\sigma_{n,1}^{-1}\left\|\left|\bar{h}_{n,2}\right|\right\|_{2,\overline{V}_2} = o(1)$, then*

$$\frac{\sqrt{n}}{2\sigma_{n,1}}\left(U_n(h_n) - \theta_n\right) = \frac{1}{\sigma_{n,1}\sqrt{n}}\sum_{\ell=1}^{n}\bar{h}_{n,1}(X_\ell) + o_p(1).$$

If, in addition, Assumption C3 holds and $\sigma_{n,1}\sqrt{n} \to \infty$, then

$$\frac{\sqrt{n}}{2\sigma_{n,1}}\left(U_n(h_n) - \theta_n\right) \overset{\mathsf{w}}{\to} \mathcal{N}(0,1).$$

Proof. See Section 5.3.2. □

Remark 8. If h_n does not depend on n, the condition $n^{-1/2}\sigma_{n,1}^{-1}\left\|\left|\bar{h}_{n,2}\right|\right\|_{2,\overline{V}_2} = o(1)$ automatically holds and Theorem 4.2 implies a standard CLT for U-statistics [27,5]. But, unlike these previous works, Theorem 4.2 does not assume stationarity.

The assumption $\left\|\bar{h}_{n,2}\right\|_{2,V_2} = o(\sqrt{n}\sigma_{n,1})$, which makes the quadratic term asymptotically negligible, does not always hold. When this assumption fails, $U_n(h_n)$ can still be shown to satisfy a central limit theorem with a Gaussian limit, but in this case both the linear and quadratic parts of the decomposition are typically asymptotically non-negligible. We do not pursue this further here.

4.1. Illustration: density-weighted average derivatives

In this section $\mathcal{X} = \mathcal{Y} \times \mathcal{Z}$, where $\mathcal{Y} \subseteq \mathbb{R}$, $\mathcal{Z} \subseteq \mathbb{R}^d$. A generic element $x \in \mathcal{X}$ will be written $x = (y, z')'$, for $y \in \mathcal{X}$, and $z \in \mathcal{Z}$. Suppose that $\{X_n = (Y_n, Z_n')', n \geq 0\}$ is a Markov chain with $Y_n \in \mathcal{Y}$, $Z_n \in \mathcal{Z}$, and Z_n continuously distributed with (marginal, invariant) distribution $\mu_z(\cdot) = \int \mu(dy, \cdot)$ with (Lebesgue) density $f(\cdot)$. Let $s = (s_1, \ldots, s_d)' \in \mathbb{Z}_+^d$ be a multi-index with the usual notations (e.g., $|s| = s_1 + \cdots + s_d$, $z^s = z_1^{s_1} \cdots z_d^{s_d}$, $\partial^s f(z) = \partial^{|s|} f(z)/(\partial^{s_1} z_1 \cdots \partial^{s_d} z_d)$, etc.). For $s \in \mathbb{Z}_+^d$, the density-weighted average derivative is

$$\theta^{(s)} = \mathbb{E}_\mu \left[Y_n\ f^{(s)}(Z_n) \right], \qquad f^{(s)}(z) = \partial^s f(z).$$

This class of estimands are of interest in statistics and econometrics. See, e.g., [25,20] for early results concerning $|s| = 1$ with i.i.d. data, and [23,4] for early results concerning $\theta^{(s)}$ with $|s| \geq 1$ with time-dependent data. See [3] and references therein for some recent results and a literature review for the case $|s| = 1$ under random sampling.

In this section, we show how Theorem 4.2 may be used to obtain an analogue of an asymptotic linear representation for a kernel-based semiparametric estimator of $\theta^{(s)}$ ($|s| \geq 1$). To verify the conditions of our theorem we impose the following additional assumptions. Set $e(z) = g(z)f(z)$ with $g(Z_n) = \mathbb{E}_\mu[Y_n|Z_n]$ and $v(Z_n) = \mathbb{E}_\mu[Y_n^2|Z_n]$. In addition, define (recall $x = (y, z')'$)

$$4\sigma_1^2 = \mathbb{E}_\mu[\psi(X_0)^2] + 2\sum_{\ell \geq 1} \mathbb{E}_\mu[\psi(X_0)\psi(X_\ell)],$$
$$\psi(x) = y\partial^s f(z) + (-1)^{|s|}\partial^s e(z) - 2\theta.$$

The following assumption gives conditions on the distribution μ. Let $\mathcal{I}_s = \{\ell \in \mathbb{Z}_+^d : \ell \leq s\}$ and $\mathcal{I}(k) = \{\ell \in \mathbb{Z}_+^d : |\ell| \leq k\}$.

Assumption M. (a) $\mathbb{E}_\mu[|Y_n|^p] < \infty$ with $p \geq 2$, and $0 < \sigma_1^2$.
(b) $\partial^\ell f(z)$ and $\partial^\ell e(z)$ exist and are bounded for all $\ell \in \mathcal{I}_s$.
(c) For some $\varsigma \geq 2$: $\partial^{s+\ell} f(z)$ exists and is bounded for all $\ell \in \mathcal{I}(\varsigma)$.
(d) $\partial^{s+\ell} e(z)$ and $\partial^\ell f(z)$ exist and are bounded for all $\ell \in \mathcal{I}(1)$.
(e) $\lim_{\|z\| \to \infty} \left[|\partial^\ell f(z)| + |\partial^\ell e(z)| \right] = 0$ for all $\ell \in \mathcal{I}_s$.

We introduce the following assumption characterizing a class of kernel functions that will be used to nonparametrically estimate the unknown function $f(\cdot)$.

Assumption K. (a) $K : \mathbb{R}^d \to \mathbb{R}$ is even and bounded.
(b) $\partial^\ell K(u)$ exists and is bounded for all $\ell \in \mathcal{I}_s$.
(c) For some $\varrho \geq 2$: $\int (1 + \|u\|^\varrho)|K(u)|du < \infty$, $\int (1 + \|u\|^2)|\partial^s K(u)|du < \infty$, and $\int u^\ell K(u)du = \mathbf{1}_{\{|\ell|=0\}}$ for all $\ell \in \mathcal{I}(\varrho - 1)$.

A semiparametric plug-in leave-one-out kernel-based estimator of $\theta^{(s)}$ is

$$\hat{\theta}_n^{(s)} = \binom{n}{2}^{-1} \sum_{i=1}^{n} \hat{f}_i^{(s)}(Z_i) Y_i, \qquad \hat{f}_i^{(s)}(z) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} K_{b_n}^{(s)}(z - Z_j),$$

with $K_b^{(s)}(u) = \partial^s K_b(u)$, $K_b(u) = K(u/b)/b$, and where b_n is a positive bandwidth sequence. This class of kernel-based estimators can be recast as n-varying U-statistics as in (17), where

$$\hat{\theta}_n^{(s)} = U_n(h_n), \qquad h_n(x_1, x_2) = \frac{K_{b_n}^{(s)}(z_1 - z_2) y_1 + K_{b_n}^{(s)}(z_2 - z_1) y_2}{2}, \quad x_1, x_2 \in \mathcal{X}.$$

Finally, the following assumption imposes some restrictions on the Markov chain. Recall that $x = (y, z')'$, $x_1 = (y_1, z_1')'$ and $x_2 = (y_2, z_2')'$ denote generic (possibly distinct) points in $\mathcal{X} = \mathcal{Y} \times \mathcal{Z}$.

Assumption P. (a) $\{X_n = (Y_n, Z_n'), n \geq 0\}$ is a Markov chain with transition kernel $P(x_1, dx_2) = Q_z(z_1, dz_2) Q_{y|z}(z_2; dy_2)$ where $Q_z(z_1, dz_2) = q_z(z_1, z_2) dz_2$ is a transition kernel with invariant distribution μ_z.

(b) P satisfies Assumptions A, C1 and C2. Define $V(x_1) = \inf_{x_2 \in \mathcal{X}} \overline{V}(x_1, x_2)$.

(c) For some $p \in [1, 2]$ and $m_{y,p}(z) := \int |y|^p Q_{y|z}(z; dy)$, $q_z(z_1, \cdot)$ and $m_{y,p}(\cdot)$ are continuous, $\sup_{x \in \mathcal{X}} \frac{|y|}{V(x)} < \infty$ and $\sup_{x_1, x_2 \in \mathcal{X}^2} \frac{m_{y,p}(z_2) q_z(z_1, z_2)}{\overline{V}(x_1, x_2)^p} < \infty$.

(d) For some $p > 1$, $\psi \in \mathcal{L}_{\overline{\mathcal{U}}_1}$, and $\sup_n \mathbb{E}[\overline{\mathcal{V}}_1^{2p}(X_n)] < \infty$ where $\overline{\mathcal{U}}_1$ and $\overline{\mathcal{V}}_1$ are as in C2.

Assumption P rules out any potential "feedback" of the outcome variable Y_n. This restriction is imposed to simplify the exposition but may be relaxed. The short-range dependence imposed, however, is crucial to obtain asymptotic normality (c.f., [4]).

Under these conditions, we obtain the following result by an application of Theorem 4.2.

Theorem 4.3. *Suppose the assumptions of Theorem 4.2 and Assumptions M, K and P hold. If* $\sqrt{n} b_n^{\varrho \wedge \varsigma} \to 0$ *and* $n b_n^{d+2|s|} \to \infty$, *then*

$$\sqrt{n}\left(\hat{\theta}_n^{(s)} - \theta^{(s)}\right) = \frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} \psi(X_\ell) + o_p(1) \overset{\mathsf{w}}{\to} \mathcal{N}(0, 4\sigma_1^2), \qquad \sigma_{n,1}^2 = \sigma_1^2 + o(1).$$

Proof. See Section 5.3.3. □

This result shows that, under the bandwidths restrictions imposed, the kernel-based semiparametric density-weighted average derivative estimator $\hat{\theta}_n^{(s)}$ satisfies the analogue of an asymptotic linear representation with "influence function" $\psi(\cdot)$. Because the function $\psi(\cdot)$ is not n-varying in this case, the Gaussian distributional approximation follows easily. If the condition $n b_n^{d+2|s|} \to \infty$ is not satisfied, then the quadratic term in our martingale approximation will also be first-order, under appropriate regularity conditions, and the linear representation obtained in Theorem 4.3 will no longer be valid. See [3] for further discussion on this under i.i.d. sampling.

5. Proofs

5.1. Proofs of the results in Section 2

5.1.1. Proof of Lemma 2.1

Note first that $\bar{G}_2 \in \mathcal{L}_{\overline{W}_2}$ and $|\bar{G}_2|_{\overline{W}_2} \lesssim |\bar{h}_2|_{\overline{V}_2}$ because $\|\bar{R}_2(x, y; \cdot)\|_{\overline{V}_2} \lesssim \overline{W}_2(x, y)$. Since $P^s\overline{W}_2(x, y) < \infty$ for all $x, y \in \mathcal{X}$ and $s \in \{1, 2\}$ by Assumption A, we deduce that the right-hand-side of (4) is well-defined and can be written as $\bar{G}_2(x, y) - P\bar{G}_2(y, x) - P\bar{G}_2(x, y) + P^2\bar{G}_2(x, y)$. To save space, set $\pi_{n,m}(x, y; (du, dv)) = (P^n(x, du) - \mu(du))(P^m(y, dv) - \mu(dv))$. Notice that, for all $N, M \geq 1$ and for all $x, y \in \mathcal{X}$, we have

$$
\begin{aligned}
\sum_{n=0}^{N}\sum_{m=0}^{M} |\pi_{n,m}\bar{h}_2(x, y)| &\leq \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} |\pi_{n,m}\bar{h}_2(x, y)| \\
&\leq |\bar{h}_2|_{\bar{V}_1} \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \|(P^n(x, \cdot) - \mu)(P^m(y, \cdot) - \mu)\|_{\overline{V}_2} \\
&\leq |\bar{h}_2|_{\bar{V}_1} \overline{W}_2(x, y).
\end{aligned}
$$

And since $P^s\overline{W}_2(x, y) < \infty$ for all $x, y \in \mathcal{X}$ and $s \in \{1, 2\}$, by dominated convergence we have

$$
\begin{aligned}
&\lim_{N,M\to\infty} \sum_{n=0}^{N}\sum_{m=0}^{M} \Big\{ \int \delta_x(dz_1) \int \delta_y(dz_2)\pi_{n,m}\bar{h}_2(z_1, z_2) \\
&\quad - \int P(x, dz_1) \int \delta_y, (dz_2)\pi_{n,m}\bar{h}_2(z_1, z_2) - \int \delta_x(dz_1) \\
&\quad \times \int P(y, dz_2)\pi_{n,m}\bar{h}_2(z_1, z_2) + \int P(x, dz_1) \int P(y, dz_2)\pi_{n,m}\bar{h}_2(z_1, z_2) \Big\} \\
&= \int \delta_x(dz_1) \int \delta_y(dz_2) \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \pi_{n,m}\bar{h}_2(z_1, z_2) \\
&\quad - \int P(x, dz_1) \int \delta_y, (dz_2) \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \pi_{n,m}\bar{h}_2(z_1, z_2) \\
&\quad - \int \delta_x(dz_1) \int P(y, dz_2) \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \pi_{n,m}\bar{h}_2(z_1, z_2) \\
&\quad + \int P(x, dz_1) \int P(y, dz_2) \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \pi_{n,m}\bar{h}_2(z_1, z_2) \\
&= \bar{G}_2(x, y) - P\bar{G}_2(y, x) - P\bar{G}_2(x, y) + P^2\bar{G}_2(x, y).
\end{aligned}
$$

But, on the other hand, by telescoping the sum we have

$$
\begin{aligned}
&\sum_{n=0}^{N}\sum_{m=0}^{M} \Big\{ \int \delta_x(dz_1) \int \delta_y(dz_2)\pi_{n,m}\bar{h}_2(z_1, z_2) \\
&\quad - \int P(x, dz_1) \int \delta_y, (dz_2)\pi_{n,m}\bar{h}_2(z_1, z_2) - \int \delta_x(dz_1)
\end{aligned}
$$

$$\times \int P(y, dz_2)\pi_{n,m}\bar{h}_2(z_1, z_2) + \int P(x, dz_1) \int P(y, dz_2)\pi_{n,m}\bar{h}_2(z_1, z_2)\Big\}$$

$$= \sum_{n=0}^{N}\sum_{m=0}^{M} \left\{\pi_{n,m}\bar{h}_2(x, y) - \pi_{n+1,m}\bar{h}_2(x, y) - \pi_{n,m+1}\bar{h}_2(x, y) + \pi_{n+1,m+1}\bar{h}_2(x, y)\right\}$$

$$= \bar{h}_2(x, y) - \pi_{N+1,0}\bar{h}_2(x, y) - \pi_{0,M+1}\bar{h}_2(x, y) + \pi_{N+1,M+1}\bar{h}_2(x, y).$$

Let $N, M \to \infty$ and compare the two limits, to conclude that $\bar{h}_2(x, y) = \bar{G}_2(x, y) - P\bar{G}_2(y, x) - P\bar{G}_2(x, y) + P^2\bar{G}_2(x, y)$, which proves (4). The bound $|P^s\bar{G}_2|_{\overline{W}_2} \lesssim |P^s\bar{h}_2|_{\overline{V}_1}$ is obtained by showing in a similar way that

$$P^s\bar{G}_2(x, y) = \lim_{N,M\to\infty} \sum_{n=0}^{N}\sum_{m=0}^{M} \pi_{n,m}\{P^s\bar{h}_2\}(x, y)$$

$$= \int\int \bar{R}_2(x, y; dz_1, dz_2)\{P^s\bar{h}_2\}(z_1, z_2). \quad \square$$

5.1.2. Proof of Lemma 2.2

Recall that $h_n(x, y) = \theta_n + \bar{h}_{n,1}(x) + \bar{h}_{n,1}(y) + \bar{h}_{n,2}(x, y)$. After some simple algebra, we have

$$U_n(h_n) = U_{n,0} + \sum_{\ell=1}^{n} w_{n,1}(\ell)\bar{h}_{n,1}(X_\ell) + \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell, j)\bar{h}_{n,2}(X_\ell, X_j).$$

Using (4), we write

$$\begin{aligned}
\bar{h}_{n,2}(X_\ell, X_j) &= \Lambda_{n,2}(X_j, X_\ell, X_j, X_\ell) \\
&= Q_{n,\ell,j} + \Lambda_{n,2}(X_j, X_\ell, X_j, X_\ell) - \Lambda_{n,2}(X_{j-1}, X_{\ell-1}, X_j, X_\ell) \\
&= Q_{n,\ell,j} + \left(P\bar{G}_{n,2}(X_{\ell-1}, X_j) - P\bar{G}_{n,2}(X_\ell, X_j)\right) \\
&\quad + \left(P\bar{G}_{n,2}(X_{j-1}, X_\ell) - P\bar{G}_{n,2}(X_j, X_\ell)\right) \\
&\quad + \left(P^2\bar{G}_{n,2}(X_j, X_\ell) - P^2\bar{G}_{n,2}(X_{j-1}, X_{\ell-1})\right).
\end{aligned}$$

Rearranging the terms, it is easy to verify that

$$\sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell, j)\bar{h}_{n,2}(X_\ell, X_j) = \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell, j)Q_{n,\ell,j} + \zeta_n,$$

where ζ_n is as stated in the lemma. $\square$

5.2. Proof of the results in Section 3

We start with some general consequences of Assumption B that will be useful in the proof of Theorems 3.1 and 3.2. Since by Assumption B, $V_2^2 \le V_3$ and $\sup_n \mathbb{E}[V_3^q(X_n)] < \infty$, by setting $p = q$, we have

$$\sup_{n\ge0} \mathbb{E}[V_2^{2p}(X_n)] < \infty. \tag{21}$$

Take $h \in \mathcal{L}_{V_1}$ such that $\mu(h) = 0$. And consider the partial sum $\sum_{k=1}^{n} a_{n,k} h(X_k)$, for some arbitrary sequence of real numbers $\{a_{n,\ell}, 1 \le \ell \le n\}$. Proposition A.1 (Eq. (38)) applied with $f_n \equiv h$, implies that

$$\sum_{\ell=1}^{n} a_{n,\ell} h(X_\ell) = \sum_{\ell=1}^{n} a_{n,\ell} Q_\ell + \epsilon_{n,1} \tag{22}$$

for some remainder $\epsilon_{n,1}$, where Q_ℓ is as in (12). The sequence $\{(Q_\ell, \mathcal{F}_\ell), 1 \le \ell \le n\}$ is a martingale difference sequence with respect to $\mathcal{F}_\ell = \sigma(X_0, \dots, X_\ell)$. Furthermore, for any $1 < \alpha \le 2p$, the following inequalities hold true

$$\begin{aligned} &\sup_{n \ge 0} \mathbb{E}[|Q_n|^\alpha] < \infty, \quad \mathbb{E}\left[\left|\sum_{\ell=1}^{n} a_{n,\ell} Q_\ell\right|^\alpha\right] \lesssim \left(\sum_{\ell=1}^{n} |a_{n,\ell}|^{\alpha \wedge 2}\right)^{1 \vee \frac{\alpha}{2}} \quad \text{and} \\ &\mathbb{E}[|\epsilon_{n,1}|^\alpha] \lesssim \left(|a_{n,1}| + |a_{n,n}| + \sum_{\ell=2}^{n} |a_{n,\ell} - a_{n,\ell-1}|\right)^\alpha . \end{aligned} \tag{23}$$

The first inequality in (23) is given by (41) and (21), and the last two inequalities are given by (39), and using (21). The case $a_{n,k} = \frac{1}{\sigma(h)\sqrt{n}}$ is important, and we summarize it in the following lemma.

Lemma 5.1. *Suppose that the assumptions of Theorem 3.1 holds. Take $h \in \mathcal{L}_{V_1}$ such that $\mu(h) = 0$. For all $n \ge 1$,*

$$\frac{1}{\sigma(h)\sqrt{n}} \sum_{\ell=1}^{n} h(X_\ell) = \frac{1}{\sigma(h)\sqrt{n}} \sum_{\ell=1}^{n} Q_\ell + \epsilon_{n,1}, \tag{24}$$

where $\{(Q_\ell, \mathcal{F}_\ell), 1 \le \ell \le n\}$ is a martingale difference, and $\epsilon_{n,1} \stackrel{\text{Pr}}{\to} 0$.

Furthermore, define $B_n(t) = \frac{1}{\sigma(h)\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} Q_k$, with $0 \le t \le 1$, and where $\lfloor x \rfloor$ denote the largest integer smaller or equal to x. Then, $B_n \stackrel{\text{w}}{\to} B$, in $D([0,1], \mathbb{R})$ (the space of cadlag functions from $[0,1] \to \mathbb{R}$) equipped with the Skorohod topology, where B is the standard Brownian motion $B = \{B(t), 0 \le t \le 1\}$.

Proof. Note that $\sigma^2(h)$ can also be written as $\sigma^2(h) = \int \pi(dx) \int P(x,dy)(g(y) - Pg(x))^2$. The decomposition (24) is simply (22) with $a_{n,k} = \frac{1}{\sigma(h)\sqrt{n}}$. In that case, it follows from (23) that $\mathbb{E}^{1/p}[|\epsilon_{n,1}|^p] \lesssim n^{-1/2}$, hence $\epsilon_{n,1} \stackrel{\text{Pr}}{\to} 0$. For the invariance principle, we apply Corollary 3.8 in [15]. It suffices to check that for any $\epsilon > 0$, and $t \in [0,1]$,

$$\frac{1}{\sigma^2(h)n} \sum_{k=1}^{\lfloor nt \rfloor} \mathbb{E}[Q_k^2 | \mathcal{F}_{k-1}] \stackrel{\text{Pr}}{\to} t \quad \text{and} \quad \frac{1}{\sigma^2(h)n} \sum_{k=1}^{\lfloor nt \rfloor} \mathbb{E}[Q_k^2 \mathbf{1}_{\{|Q_k| > \epsilon\sigma(h)\sqrt{n}\}} | \mathcal{F}_{k-1}] \stackrel{\text{Pr}}{\to} 0. \tag{25}$$

Now, $\mathbb{E}[Q_k^2 | \mathcal{F}_{k-1}] = \int P(X_{k-1}, dz)\,(g(z) - Pg(X_{k-1}))^2 = U(X_{k-1})$ almost surely, where $U(x) = \int P(x,dz)\,(g(z) - Pg(x))^2$. Since $g \in \mathcal{L}_{V_2}$, and given Assumption B, we see that $U \in \mathcal{L}_{V_2^2}$. Furthermore, by Assumption B, P satisfies $\mathsf{C}(1, V_2^2, V_3)$, and $\sup_n \mathbb{E}[V_3^q(X_n)] < \infty$. Hence we can apply Proposition A.1, with V_1 (resp. V_2) set equal to V_2^2 (resp. V_3), and with $f_n \equiv U$, $p = q$, and $a_{n,k} = 1$. This yields that $\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[Q_k^2 | \mathcal{F}_{k-1}] \stackrel{\text{Pr}}{\to} \int U(z)\mu(dz) = \sigma^2(h)$.

Hence, as $n \to \infty$,

$$\frac{1}{\sigma^2(h)n}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}[Q_k^2|\mathcal{F}_{k-1}] = \frac{\lfloor nt\rfloor}{n}\frac{1}{\sigma^2(h)}\left(\frac{1}{\lfloor nt\rfloor}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}[Q_k^2|\mathcal{F}_{k-1}]\right) \to t.$$

To check the second part of (25), the conditional Lindeberg condition, it suffices to check the unconditional Lindeberg condition with $t = 1$ (i.e., $\frac{1}{\sigma^2(h)n}\sum_{k=1}^{n}\mathbb{E}[Q_k^2\mathbf{1}_{\{|Q_k|>\epsilon\sigma(h)\sqrt{n}\}}] \overset{\mathsf{Pr}}{\to} 0$ for all $\epsilon > 0$). By Holder's inequality and the Markov inequality, we have

$$\begin{aligned}\mathbb{E}\left[Q_k^2\mathbf{1}_{\{|Q_k|>\epsilon\sigma(h)\sqrt{n}\}}\right] &\le \mathbb{E}^{1/p}[|Q_k|^{2p}]\mathbb{P}^{1-1/p}[|Q_\ell| > \epsilon\sigma(h)\sqrt{n}]\\ &\le \mathbb{E}^{1/p}[|Q_k|^{2p}]\left\{\frac{\mathbb{E}[|Q_k|^{2p}]}{\epsilon^{2p}\sigma^{2p}(h)n^p}\right\}^{1-\frac{1}{p}} \lesssim \frac{n^{-(p-1)}}{(\epsilon^2\sigma^2(h))^{p-1}},\end{aligned}$$

since $p > 1$, and the result follows. □

5.2.1. Proof of Theorem 3.1

Without any loss of generality we assume that $\mu(h) = 0$. Recall that $w_{n,b}(k) = 2n^{-1}w_b(k/c_n)$ if $k \neq 0$ and $w_{n,b}(0) = n^{-1}$, and from (10) we have $\Gamma_{n,b}^2(h) = \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_{n,b}(\ell - j) h(X_\ell)h(X_j) + R_n$. By applying Lemma 2.2 to the term $\sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_{n,b}(\ell - j)h(X_j)h(X_\ell)$, we obtain directly (13) with R_n given in (11), so it only remains to give the expression of ζ_n:

$$\begin{aligned}\zeta_n &= \sum_{\ell=1}^{n} PG(X_{\ell-1})\sum_{j=1}^{\ell}\Delta_{n,b}^{(1)}(\ell - j)Q_j + \sum_{\ell=1}^{n} Q_\ell\sum_{j=1}^{\ell}\Delta_{n,b}^{(1)}(\ell - j + 1)PG(X_{j-1})\\ &+ \sum_{\ell=3}^{n} PG(X_{\ell-1})\sum_{j=1}^{\ell-2}\Delta_{n,b}^{(2)}(\ell - j)PG(X_{j-1})\\ &+ PG(X_0)\left\{\sum_{\ell=1}^{n} w_{n,b}(\ell)Q_\ell + \sum_{\ell=1}^{n}\Delta_{n,b}^{(1)}(\ell)PG(X_{\ell-1})\right\}\\ &- PG(X_n)\left\{\sum_{j=1}^{n} w_{n,b}(n - j)Q_j - \sum_{j=1}^{n}\Delta_{n,b}^{(1)}(\ell - j)PG(X_j)\right\}\\ &+ \Delta_{n,b}^{(2)}(0)\sum_{\ell=1}^{n}(PG(X_{\ell-1}))^2 - \left(w_{n,b}(0) - \Delta_{n,b}^{(2)}(1)\right)\sum_{\ell=1}^{n} PG(X_\ell)Q_\ell\\ &- \Delta_{n,b}^{(2)}(1)\,(PG(X_n))^2 - w_{n,b}(n - 1)PG(X_n)PG(X_0),\end{aligned}$$

where we set $\Delta_{n,b}^{(1)}(\ell) := w_{n,b}(\ell) - w_{n,b}(\ell - 1)$ and $\Delta_{n,b}^{(2)}(\ell) := 2w_{n,b}(\ell) - w_{n,b}(\ell + 1) - w_{n,b}(\ell - 1)$. Since w_b is bounded and has support $[-b, b]$, we have $|v_n(\ell)| \lesssim \frac{c_n}{n}$, uniformly in ℓ. Similarly, $|u_n| \le \frac{c_n}{n}$. With the same argument, and using the mean value theorem, we obtain that $|v_n(\ell) - v_n(\ell - 1)| \lesssim n^{-1}$, uniformly in ℓ. To summarize, we have:

$$|u_n| \le \frac{c_n}{n}, \qquad |v_n(\ell)| \lesssim \frac{c_n}{n}, \quad \text{and} \quad |v_n(\ell) - v_n(\ell - 1)| \lesssim n^{-1}, \quad \text{uniformly in } \ell. \tag{26}$$

We now deal with each of the terms in the decomposition of $\Gamma_{n,b}^2(h)$.

Term R_n. By the Cauchy–Schwartz inequality, we have

$$\mathbb{E}[|R_n|^p] \lesssim \frac{1}{n^p}\mathbb{E}^{1/2}\left[\left|\sum_{\ell=1}^{n} v_n(\ell)h(X_\ell)\right|^{2p}\right]\mathbb{E}^{1/2}\left[\left|\sum_{\ell=1}^{n} h(X_\ell)\right|^{2p}\right] + \left(\frac{u_n}{n}\right)^p \mathbb{E}\left[\left|\sum_{\ell=1}^{n} h(X_\ell)\right|^{2p}\right].$$

The bound in (23), applied with $a_{n,\ell} \equiv 1$, and with $a_{n,\ell} \equiv v_n(\ell)$, using (26) gives

$$\mathbb{E}\left[\left|\sum_{\ell=1}^{n} h(X_\ell)\right|^{2p}\right] \lesssim n^p \quad \text{and} \quad \mathbb{E}\left[\left|\sum_{\ell=1}^{n} v_n(\ell)h(X_\ell)\right|^{2p}\right] \lesssim 1 + \frac{c_n^{2p}}{n^p}. \tag{27}$$

We deduce that $\mathbb{E}[|R_n|^p] \lesssim (c_n/n)^p + n^{-p/2}$.

Term $\sum_{\ell=1}^{n} Q_\ell \sum_{j=1}^{\ell-1} w_{n,b}(\ell - j)Q_j$. Notice that $\sum_{j=1}^{\ell-1} w_{n,b}(\ell - j)Q_j \in \mathcal{F}_{\ell-1}$, hence $\{(Q_\ell \sum_{j=1}^{\ell-1} w_{n,b}(\ell - j)Q_j, \mathcal{F}_\ell), 1 \le \ell \le n\}$ is a martingale difference array. By the martingale inequality (42), we have

$$\mathbb{E}\left[\left|\sum_{\ell=1}^{n} Q_\ell \sum_{j=1}^{\ell-1} w_{n,b}(\ell - j)Q_j\right|^p\right] \lesssim \left\{\sum_{\ell=1}^{n} \mathbb{E}^{1\wedge\frac{2}{p}}\left[\left|Q_\ell \sum_{j=1}^{\ell-1} w_{n,b}(\ell - j)Q_j\right|^p\right]\right\}^{1\vee\frac{p}{2}}. \tag{28}$$

By the Cauchy–Schwartz inequality, and the martingale inequality (42) applied to $\{(w_{n,b}(\ell - j)Q_j, \mathcal{F}_j), 1 \le j \le \ell - 1\}$, we get

$$\begin{aligned}\mathbb{E}\left[\left|Q_\ell \sum_{j=1}^{\ell-1} w_{n,b}(\ell - j)Q_j\right|^p\right] &\lesssim \mathbb{E}^{1/2}[|Q_\ell|^{2p}]\mathbb{E}^{1/2}\left[\left|\sum_{j=1}^{\ell-1} w_{n,b}(\ell - j)Q_j\right|^{2p}\right] \\ &\lesssim \mathbb{E}^{1/2}\left[\left|\sum_{j=1}^{\ell-1} w_{n,b}(\ell - j)Q_j\right|^{2p}\right] \\ &\lesssim \left(\sum_{j=1}^{\ell-1} w_{n,b}^2(\ell - j)\mathbb{E}^{1/p}[|Q_j|^{2p}]\right)^{p/2} \\ &\lesssim \left(\sum_{j=1}^{\ell-1} w_{n,b}^2(\ell - j)\right)^{p/2} \lesssim c_n^{p/2}.\end{aligned}$$

Combining this bound with (28), it follows that $\mathbb{E}\left[\left|\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1} w_{n,b}(\ell - j)Q_\ell Q_j\right|^p\right] \lesssim c_n^{p/2} n^{1\vee\frac{p}{2}}$.

Remainder ζ_n. The remainder ζ_n is handled along the same lines as above. By the mean value theorem, $|\Delta_{n,b}^{(1)}(\ell)| \lesssim \frac{1}{nc_n}$, $|\Delta_{n,b}^{(2)}(\ell)| \lesssim \frac{1}{nc_n^2}$, uniformly in ℓ. Using this, and the same techniques as above, we derive that $\mathbb{E}^{1/p}[|\zeta_n|^p] \lesssim c_n^{-1+\frac{1}{2}\vee\frac{1}{p}}$, for all $n \geq 3$.

This concludes the proof. □

5.2.2. Proof of Theorem 3.2

Without any loss of generality we assume that $\mu(h) = 0$. If $c_n = o(n)$ and $p \geq 2$, then from Theorem 3.1,

$$\Gamma_{n,b}^2(h) = n^{-1}\sum_{\ell=1}^{n} Q_\ell^2 + o_p(1).$$

Given the ergodicity assumption $\mathsf{C}(1, V_2^2, V_3)$ and $\sup_{k\geq 0}\mathbb{E}[V_3^q(X_k)] < \infty$, it follows from Proposition A.1 that $n^{-1}\sum_{\ell=1}^{n} Q_\ell^2 \overset{\mathsf{Pr}}{\to} \sigma^2(h) = \int \mu(dx)\int P(x,dy)(G(y) - PG(x))^2$. The invariance principle in Lemma 5.1 implies that $\frac{1}{\sigma(h)\sqrt{n}}\sum_{\ell=1}^{n} h(X_\ell) \overset{\mathsf{w}}{\to} \mathcal{N}(0,1)$. This proves the first part of the theorem.

From now on, we assume that $c_n = n$. Define, for all t such that $0 \leq t \leq 1$,

$$W_{n,\ell} = \frac{Q_\ell}{\sqrt{n}\sigma(h)}, \qquad B_n(t) = \sum_{\ell=1}^{\lfloor nt\rfloor} W_{n,\ell},$$
$$Z_n(t) = \int_0^t w_b(t-u)dB_n(u) = \sum_{k=1}^{\lfloor nt\rfloor} w_b\left(t - \frac{k-1}{n}\right)W_{n,k}.$$

We have established in Lemma 5.1 that

$$\frac{1}{\sigma(h)\sqrt{n}}\sum_{\ell=1}^{n} h(X_\ell) = \sum_{\ell=1}^{n} W_{n,\ell} + r_{n,2} = B_n(1) + r_{n,1}, \quad \text{where } r_{n,2} \overset{\mathsf{Pr}}{\to} 0. \tag{29}$$

A similar result holds for $\sum_{\ell=1}^{n} v_n(\ell)h(X_\ell)$: from (22) with $a_{n,\ell} = v_n(\ell)$, and using (23) and (26),

$$\sum_{\ell=1}^{n} v_n(\ell)h(X_\ell) = \sum_{\ell=1}^{n} v_n(\ell)Q_\ell + \epsilon_{n,1}, \quad \text{where } \mathbb{E}^{1/p}[|\epsilon_{n,1}|^p] \lesssim 1. \tag{30}$$

Recall that $v_n(\ell) = n^{-1}\sum_{i=1}^{n} w_b\left(\frac{\ell-i}{n}\right) = n^{-1}\sum_{i=1}^{n} w_b\left(\frac{\ell-1}{n} - \frac{i-1}{n}\right)$, and hence $v_n(\ell)$ is a Riemann sum approximation of $g_b\left(\frac{\ell-1}{n}\right) = \int_0^1 w_b\left(\frac{\ell-1}{n} - s\right)ds$. Similarly, $u_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i=1}^{n} w((i-j)/n)$ is a Riemann sum approximation of $\int_0^1\int_0^1 w_b(s-t)dtds = \int_0^1 g_b(s)ds$. Therefore,

$$v_n(\ell) = g_b\left(\frac{\ell-1}{n}\right) + \delta_n(\ell), \qquad u_n = \int_0^1 g_b(t)dt + \delta_n', \quad |\delta_n(\ell)| + |\delta_n'| \lesssim n^{-1}, \tag{31}$$

where the latter bound is uniform in ℓ.

Using (30), (31), and

$$\mathbb{E}\left[\left(\frac{1}{\sigma(h)\sqrt{n}}\sum_{\ell=1}^{n}\left(v_n(\ell)-g_b\left(\frac{\ell-1}{n}\right)\right)Q_\ell\right)^2\right] = \frac{1}{\sigma^2(h)n}\sum_{\ell=1}^{n}|\delta_n(\ell)|^2\mathbb{E}[Q_\ell^2] \lesssim n^{-2},$$

which follows from the martingale property, we have

$$\begin{aligned}\frac{1}{\sigma(h)\sqrt{n}}\sum_{\ell=1}^{n}v_n(\ell)h(X_\ell) &= \frac{1}{\sigma(h)\sqrt{n}}\sum_{\ell=1}^{n}g_b\left(\frac{\ell-1}{n}\right)Q_\ell \\ &\quad + \frac{1}{\sigma(h)\sqrt{n}}\sum_{\ell=1}^{n}\left(v_n(\ell)-g_b\left(\frac{\ell-1}{n}\right)\right)Q_\ell \\ &\quad + \frac{\epsilon_{n,1}}{\sigma(h)\sqrt{n}} \\ &= \frac{1}{\sigma(h)\sqrt{n}}\sum_{\ell=1}^{n}g_b\left(\frac{\ell-1}{n}\right)Q_\ell + r_{n,2} \\ &= \int_0^1 g_b(t)dB_n(t) + r_{n,2} \quad \text{with } r_{n,2} \overset{\mathsf{Pr}}{\to} 0. \qquad (32)\end{aligned}$$

We combine (29) and (32) to rewrite R_n as

$$\begin{aligned}R_n &= -2\left(\frac{1}{\sqrt{n}}\sum_{\ell=1}^{n}v_n(\ell)h(X_\ell)\right)\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}h(X_j)\right) + u_n\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}h(X_j)\right)^2 \\ &= -2\sigma^2(h)\left(\int_0^1 g_b(t)dB_n(t) + r_{n,2}\right)\left(B_n(1)+r_{n,1}\right) \\ &\quad + \sigma^2(h)\left(\int_0^1 g_b(s)ds + \delta_n'\right)\left(B_n(1)+r_{n,1}\right)^2 \\ &= -2\sigma^2(h)B_n(1)\int_0^1 g_b(t)dB_n(t) + \sigma^2(h)\left(\int_0^1 g_b(t)dt\right)B_n^2(1) + r_{n,3}, \qquad (33)\end{aligned}$$

where, using (29), (32), (27) and (31), $r_{n,3} \overset{\mathsf{Pr}}{\to} 0$.

Putting all the previous results together, we can rewrite $\Gamma_{n,b}^2(h)$ as

$$\begin{aligned}\Gamma_{n,b}^2(h) &= n^{-1}\sum_{\ell=1}^{n}Q_\ell^2 + \frac{2}{n}\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1}w_b\left(\frac{\ell-j}{n}\right)Q_\ell Q_j + \zeta_n + R_n \\ &= n^{-1}\sum_{\ell=1}^{n}Q_\ell^2 + 2\sigma^2(h)\sum_{\ell=1}^{n}W_{n,\ell}\left\{\sum_{j=1}^{\ell-1}w_b\left(\frac{\ell-j}{n}\right)W_{n,j}\right\} + R_n + \zeta_n \\ &= \sigma^2(h)\sum_{\ell=1}^{n}W_{n,\ell}^2 + 2\sigma^2(h)\sum_{\ell=1}^{n}Z_n\left(\frac{\ell-1}{n}\right)W_{n,\ell} + R_n + \zeta_n\end{aligned}$$

$$= \sigma^2(h) \sum_{\ell=1}^{n} W_{n,\ell}^2 + 2\sigma^2(h) \int_0^1 Z_n(t) dB_n(t) + R_n + \zeta_n,$$

$$= \sigma^2(h) \sum_{\ell=1}^{n} W_{n,\ell}^2 + 2\sigma^2(h) \int_0^1 Z_n(t) dB_n(t) - 2\sigma^2(h) B_n(1) \int_0^1 g_b(t) dB_n(t)$$

$$+ \sigma^2(h) \left(\int_0^1 g_b(t) dt \right) B_n^2(1) + r_{n,4},$$

where $r_{n,4} = r_{n,3} + \zeta_n \overset{\mathsf{Pr}}{\to} 0$. It only remains to show that, for $Z(t) := \int_0^t w(t-s) dB(s)$,

$$\left(B_n(1), \int_0^1 Z_n(s) dB_n(s), \int_0^1 g_b(t) dB_n(t) \right)$$

$$\overset{\mathsf{w}}{\to} \left(B(1), \int_0^1 Z(s) dB(s), \int_0^1 g_b(t) dB(t) \right). \quad (34)$$

Once (34) is established, and since $r_{n,4} \overset{\mathsf{Pr}}{\to} 0$ and $\sum_{\ell=1}^{n} W_{n,\ell}^2 \overset{\mathsf{Pr}}{\to} 1$, we can employ the continuous mapping theorem and Slutsky's theorem to conclude that

$$\Gamma_{n,b}^2(h) \overset{\mathsf{w}}{\to} \sigma^2(h) \left(1 + 2 \int_0^1 Z(t) dB(t) - 2B(1) \int_0^1 g_b(t) dB(t) + B^2(1) \int g_b(t) dt \right),$$

and that $\sum_{j=1}^{n} h(X_j) / \sqrt{n \Gamma_{n,b}^2(h)} = \sigma^2(h) B_n(1) / \sqrt{\Gamma_{n,b}^2(h)} \overset{\mathsf{w}}{\to} \frac{B(1)}{\sqrt{\mathsf{K}_b}}$.

To show (34), we first show that $(B_n, Z_n) \overset{\mathsf{w}}{\to} (B, Z)$ in $D([0, 1], \mathbb{R}^2)$. We write d_q to denote the Skorohod metric in $D([0, 1], \mathbb{R}^q)$, the space of all cadlag functions from $[0, 1]$ to $\mathbb{R}^q$, where $\mathbb{R}^q$ is equipped with its Euclidean distance. That is,

$$d_q(x, y) = \inf_{\lambda} \max \left\{ \sup_{0 \le t \le 1} |\lambda(t) - t|, \sup_{0 \le t \le 1} \| x \circ \lambda(t) - y(t) \| \right\},$$

where the infimum is taken over Λ, the set of all increasing continuous functions λ from $[0, 1]$ onto $[0, 1]$. We have shown in Lemma 5.1 that $B_n \overset{\mathsf{w}}{\to} B$ in $D([0, 1], \mathbb{R})$. Hence by the Skorohod representation of weak convergence, there exists a version of $\{(B_n, B), n \ge 1\}$, and a sequence $\{\lambda_n, n \ge 1\}$ of Λ, such that $B \in C[0, 1]$ almost surely, and $\sup_{0 \le t \le 1} |\lambda_n(t) - t| \to 0$ and $\sup_{0 \le t \le 1} |B_n(t) - B \circ \lambda_n(t)| \to 0$ almost surely. We show next that for a given sample path for which the aforementioned convergence holds, and $B \in C[0, 1]$, we also have $d_2\left((B_n, Z_n), (B, Z)\right) \to 0$ and therefore $(B_n, Z_n) \overset{\mathsf{w}}{\to} (B, Z)$ in $D([0, 1], \mathbb{R}^2)$. Now, since

$$d_2((B_n, Z_n), (B, Z)) \le \sup_{0 \le t \le 1} \left\{ |\lambda_n(t) - t| + |B_n(t) - B \circ \lambda_n(t)| + |Z_n(t) - Z \circ \lambda_n(t)| \right\},$$

it suffices to show that $\sup_{0 \le t \le 1} |Z_n(t) - Z \circ \lambda_n(t)| \to 0$. Because w_b is of class $\mathcal{C}^2$, Z can be defined pathwise and, by integration by parts, we have $Z(t) = B(t) + \int_0^t B(u) w_b'(t-u) du$, $0 \le t \le 1$. Therefore,

$$Z \circ \lambda_n(t) = B(\lambda_n(t)) + \int_0^{\lambda_n(t)} B(u) w_b'(\lambda_n(t) - u) du$$

$$= B(\lambda_n(t)) + \int_0^{\lambda_n(t)} B(u)\left(w_b'(\lambda_n(t)-u) - w_b'(t-u)\right)du$$
$$+ \sum_{k=1}^{\lfloor nt\rfloor -1} \int_{\frac{k-1}{n}}^{\frac{k}{n}} B(u)w_b'(t-u)du + \int_{\frac{\lfloor nt\rfloor -1}{n}}^{\lambda_n(t)} B(u)w_b'(t-u)du,$$

and, similarly,

$$Z_n(t) = \int_0^t w_b(t-u)dB_n(t) = \sum_{k=1}^{\lfloor nt\rfloor} w_b\left(t - \frac{k-1}{n}\right) W_{n,k}$$
$$= w_b\left(t - \frac{\lfloor nt\rfloor - 1}{n}\right) B_n\left(\frac{\lfloor nt\rfloor}{n}\right)$$
$$- \sum_{k=1}^{\lfloor nt\rfloor -1} B_n\left(\frac{k}{n}\right)\left(w_b\left(t-\frac{k}{n}\right) - w_b\left(t - \frac{k-1}{n}\right)\right).$$

Comparing the corresponding terms in $Z \circ \lambda_n(t)$ and $Z_n(t)$, we deduce that $\sup_{0\le t\le 1} |Z_n(t) - Z \circ\lambda_n(t)|$ converges to zero. We give the details for the corresponding terms $\int_{\frac{k-1}{n}}^{\frac{k}{n}} B(u)w_b'(t-u)du$ and $B_n\left(\frac{k}{n}\right)\left(w_b\left(t-\frac{k}{n}\right) - w_b\left(t-\frac{k-1}{n}\right)\right)$. The other terms are handled similarly. By the mean value theorem, we can write $w_b\left(t-\frac{k}{n}\right) - w_b\left(t-\frac{k-1}{n}\right) = n^{-1}w_b'(t-\bar{k}_n)$ for some $\bar{k}_n \in [(k-1)/n, k/n]$. Hence

$$\int_{\frac{k-1}{n}}^{\frac{k}{n}} B(u)w_b'(t-u)du - B_n\left(\frac{k}{n}\right)\left(w_b\left(t-\frac{k}{n}\right) - w_b\left(t-\frac{k-1}{n}\right)\right)$$
$$= \int_{\frac{k-1}{n}}^{\frac{k}{n}} \left[B(u) - B\left(\lambda_n\left(\frac{k}{n}\right)\right)\right] w_b'(t-u)du$$
$$+ B\left(\lambda_n\left(\frac{k}{n}\right)\right)\left[\int_{\frac{k-1}{n}}^{\frac{k}{n}} w_b'(t-u)du - \frac{w_b'(t-\bar{k}_n)}{n}\right]$$
$$+ \left[B\left(\lambda_n\left(\frac{k}{n}\right)\right) - B_n\left(\frac{k}{n}\right)\right]\frac{w_b'(t-\bar{k}_n)}{n}.$$

Since B is continuous on $[0, 1]$ it is uniformly continuous. Using this and the fact that $\sup_{0\le t\le 1} |\lambda_n(t) - t| \to 0$, it follows that for any $\epsilon > 0$, we can find $n_0 \ge 1$ such that for all $n \ge n_0$:

$$\left|\int_{\frac{k-1}{n}}^{\frac{k}{n}} B(u)w_b'(t-u)du - B_n\left(\frac{k}{n}\right)\left(w_b\left(t-\frac{k}{n}\right) - w_b\left(t-\frac{k-1}{n}\right)\right)\right|$$
$$\lesssim \frac{1}{n}\left[\epsilon + \left(\sup_{0\le t\le 1}|B(t)|\right)\frac{1}{n} + \sup_{0\le t\le 1}|B\circ\lambda_n(t) - B_n(t)|\right].$$

Since g_b is deterministic, $(Z_n, B_n) \overset{w}{\to} (Z, B)$ implies that $(g_b, Z_n, B_n) \overset{w}{\to} (g_b, Z, B)$. To finish the argument, we apply Theorem 2.2 of [13] to conclude:

$$\left(g_b, Z_n, B_n, \int_0^1 g_b dB_n(t), \int_0^1 Z_n(t)dB_n(t)\right) \xrightarrow{\text{w}} \left(g_b, Z, B, \int_0^1 g_b(t)dB(t), \int_0^1 Z(t)dB(t)\right).$$

We apply the theorem with $Y_n = B_n$ and $X_n = (g_b, Z_n)$. In our case, the process B_n is a martingale wrt $\{\mathcal{F}_t^{(n)} = \sigma(W_k, 0 \le k \le \lfloor nt \rfloor)\}$. We set $\delta = +\infty$, hence $J_\delta = 0$ and $A_n^\delta = 0$. It only remains to check Kurtz and Protter's condition C2.2(i) [13]. Let $\tau_n^\alpha = 2\alpha$, $\alpha > 0$. Then, $\mathbb{P}[\tau_n^\alpha \le \alpha] = 0 \le 1/\alpha$ and

$$\sup_n \mathbb{E}\left[[M_n^\delta]_{t\wedge\tau_n^\alpha} + T_{t\wedge\tau_n^\alpha}(A_n^\delta)\right] = \sup_n \mathbb{E}\left[[B_n]_{t\wedge\tau_n^\alpha}\right] = \sup_n \mathbb{E}\left[\sum_{k=1}^{\lfloor n(t\wedge\tau_n^\alpha)\rfloor} W_{n,k}^2\right] \le \sup_n \mathbb{E}\left[\sum_{k=1}^{n} W_{n,k}^2\right] \le \frac{1}{n\sigma^2(h)} \sum_{k=1}^{n} \mathbb{E}[Q_k^2] \lesssim 1.$$

Therefore, condition C2.2(i) in [13] holds, and the proof is complete. □

5.3. Proofs of the results of Section 4

5.3.1. Proof of Theorem 4.1

`Proof of Part 1.` From Lemma 2.1, $|P^s \bar{G}_{n,2}(x,y)| \lesssim |P^s \bar{h}_{n,2}|_{\overline{V}_2} \overline{W}_2(x,y)$ with $s \in \{1,2\}$, and

$$|P^2\bar{h}_{n,2}|_{\overline{V}_2} = \sup_{x,y\in\mathcal{X}} \{\overline{V}_2(x,y)\}^{-1} \left|\int P(x,du)\int P(y,dv)\bar{h}_{n,2}(u,v)\right| \le |P\bar{h}_{n,2}|_{\overline{V}_2} \sup_{x,y\in\mathcal{X}} \{\overline{V}_2(x,y)\}^{-1} \int P(x,du)\overline{V}_2(u,y) \lesssim |P\bar{h}_{n,2}|_{\overline{V}_2},$$

using $P\overline{V}_2 \lesssim \overline{V}_2$ from Assumption C1. This implies that for $p = 2$, and $1 \le j, k \le n$,

$$\mathbb{E}^{1/p}[|P^s\bar{G}_{n,2}(X_k,X_j)|^p] \lesssim |P\bar{h}_{n,2}|_{\overline{V}_2} \mathbb{E}^{1/p}[|\bar{W}_2(X_k,X_j)|^p] \lesssim |P\bar{h}_{n,2}|_{\overline{V}_2}, \tag{35}$$

using the moment assumption in C1. By Minkowski's inequality, $\binom{n}{2}\mathbb{E}^{1/2}[\zeta_n^2] \lesssim n|P\bar{h}_{n,2}|_{\overline{V}_2}$. Recall that $Q_{n,\ell,j} = \bar{h}_{n,2}(X_\ell,X_j) - P\bar{G}_{n,2}(X_{\ell-1},X_j) - P\bar{G}_{n,2}(X_{j-1},X_\ell) + P^2\bar{G}_{n,2}(X_{\ell-1},X_{j-1})$, which gives

$$|Q_{n,\ell,j}| \lesssim |\bar{h}_{n,2}(X_\ell,X_j)| + |P\bar{h}_{n,2}|_{\overline{V}_2}(\overline{W}_2(X_{\ell-1},X_j) + \overline{W}_2(X_{j-1},X_\ell)) + |P^2\bar{h}_{n,2}|_{\overline{V}_2}\overline{W}_2(X_{\ell-1},X_{j-1}). \tag{36}$$

Thus, for $j < \ell$ and $\mathcal{F}_\ell = \sigma(X_0, X_1, \ldots, X_\ell)$,

$$\mathbb{E}[Q_{n,\ell,j}^2|\mathcal{F}_{\ell-1}] \lesssim \int P(X_{\ell-1},dz)\left|\bar{h}_{n,2}(z,X_j)\right|^2 + \|\bar{h}_{n,2}\|_{2,\overline{V}_2} \times \mathbb{E}\left[\overline{W}_2^2(X_{\ell-1},X_j) + \overline{W}_2^2(X_\ell,X_{j-1}) + \overline{W}_2^2(X_{\ell-1},X_{j-1})|\mathcal{F}_{\ell-1}\right].$$

Taking expectation on both sides, and using Assumption C1, $\mathbb{E}^{1/2}[Q^2_{n,\ell,j}] \lesssim \left\|\bar{h}_{n,2}\right\|_{2,\overline{V}_2}$, for all $n \geq 1$. This proves the first part of the theorem.

Proof of Part 2. Using the martingale property and Lemma A.2, we have

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)^2\right] &= \sum_{\ell=2}^{n}\mathbb{E}\left[\left(\sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)^2\right] \\ &= \sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1}\mathbb{E}[Q^2_{n,\ell,j}] + 2\sum_{\ell=2}^{n}\sum_{k=1}^{\ell-2}\sum_{j=k+1}^{\ell-1}\mathbb{E}[Q_{n,\ell,j}Q_{n,\ell,k}] \\ &\lesssim n^2\left\|\bar{h}_{n,2}\right\|^2_{2,\overline{V}_2} + n^2\left\|\bar{h}_{n,2}\right\|^2_{2,\overline{V}_2},\end{aligned}$$

by $\mathbb{E}[Q^2_{n,\ell,j}] \lesssim \left\|\bar{h}_{n,2}\right\|^2_{2,\overline{V}_2}$ and the following technical lemma.

Lemma 5.2. *Under the assumptions of Theorem 4.1(2),*

$$\left|\mathbb{E}\left[\sum_{k=1}^{\ell-2}\sum_{j=k+1}^{\ell-1} Q_{n,\ell,j}Q_{n,\ell,k}\right]\right| \lesssim n\left\|\bar{h}_{n,2}\right\|^2_{2,\overline{V}_2}, \quad 3 \leq \ell \leq n.$$

Proof. Fix $1 \leq k < \ell$ and define $T_k = T_{n,\ell,k} := \mathbb{E}[\sum_{j=k+1}^{\ell-1} Q_{n,\ell,j}Q_{n,\ell,k}|\mathcal{F}_{k-1}]$, so that

$$\left|\mathbb{E}\left[\sum_{k=1}^{\ell-2}\sum_{j=k+1}^{\ell-1} Q_{n,\ell,j}Q_{n,\ell,k}\right]\right| \leq \sum_{k=1}^{\ell-2}\mathbb{E}[|T_k|].$$

Hence, it suffices to show that $\mathbb{E}[|T_k|] \lesssim \left\|\bar{h}_{n,2}\right\|^2_{2,\bar{V}_2}$. For $m \geq 0$, define

$$\begin{aligned}\Upsilon_{2,m}(x_{j-1}, x_{k-1}, x_k) &:= \int P(x_{j-1}, dx_j)\int P^m(x_j, dx_{\ell-1}) \\ &\quad\times \int P(x_{\ell-1}, dx_\ell)\Lambda_{n,2}(x_{j-1}, x_{\ell-1}; x_j, x_\ell)\Lambda_{n,2}(x_{k-1}, x_{\ell-1}; x_k, x_\ell), \\ \Upsilon_{1,m}(x_{k-1}, x_k) &:= \int \left\{P^m(x_k, dx_{j-1}) - \mu(dx_{j-1})\right\}\Upsilon_{2,\ell-m-k-2}(x_{j-1}, x_k, x_{k-1}).\end{aligned}$$

We recall that $Q_{n,\ell,j} := \Lambda_{n,2}\left(X_{j-1}, X_{\ell-1}, X_j, X_\ell\right)$. Using this and the Markov property, we verify with some straightforward but cumbersome calculations that $T_k = \mathbb{E}[\sum_{j=1}^{\ell-k-1}\Upsilon_{1,j}(X_k, X_{k-1})|\mathcal{F}_{k-1}]$, almost surely. Bound (36) and the Cauchy–Schwartz inequality imply that

$$\begin{aligned}&\left|\int P(x_{\ell-1}, dx_\ell)\Lambda_{n,2}(x_{j-1}, x_{\ell-1}; x_j, x_\ell)\Lambda_{n,2}(x_{k-1}, x_{\ell-1}; x_k, x_\ell)\right| \\ &\quad\lesssim \left\|\bar{h}_{n,2}\right\|^2_{2,\overline{V}_2}\left(\overline{W}(x_{\ell-1}, x_j) + \overline{W}(x_{\ell-1}, x_{j-1})\right)\left(\overline{W}(x_{\ell-1}, x_k) + \overline{W}(x_{\ell-1}, x_{k-1})\right).\end{aligned}$$

We combine this with Assumption C2 to conclude that, for all $m \geq 0$,

$$\left|\Upsilon_{2,m}(x_{j-1}, x_{k-1}, x_k)\right| \lesssim \left\|\bar{h}_{n,2}\right\|^2_{2,\overline{V}_2}\overline{\mathcal{U}}_1(x_{j-1})\left(\overline{\mathcal{U}}_0(x_k) + \overline{\mathcal{U}}_0(x_{k-1})\right).$$

By the short-range dependence assumption $\mathsf{C}(1, \overline{\mathcal{U}}_1, \overline{\mathcal{V}}_1)$, it follows that

$$\left| \sum_{j=0}^{\ell-k-1} \Upsilon_{1,j}(X_k, X_{k-1}) \right|$$
$$\leq \sum_{j=0}^{\ell-k-1} \left| \int \left\{ P^j(X_k, dx_{j-1}) - \mu(dx_{j-1}) \right\} \Upsilon_{2,\ell-j-k-2}(x_{j-1}, X_{k-1}, X_k) \right|$$
$$\lesssim \left\| \bar{h}_{n,2} \right\|^2_{2,\overline{V}_2} \overline{\mathcal{V}}_1(X_k) \left(\overline{\mathcal{U}}_0(X_k) + \overline{\mathcal{U}}_0(X_{k-1}) \right).$$

Therefore, we obtain $\mathbb{E}[|T_k|] \lesssim \left\| \bar{h}_{n,2} \right\|^2_{2,\overline{V}_2} \mathbb{E}[\overline{\mathcal{V}}_1(X_k)(\overline{\mathcal{U}}_0(X_k) + \overline{\mathcal{U}}_0(X_{k-1}))] \lesssim \left\| \bar{h}_{n,2} \right\|^2_{2,\overline{V}_2}$, using Assumption C2. □

This completes the proof of Theorem 4.1. □

5.3.2. Proof of Theorem 4.2

We first notice that $|P\bar{h}_{n,2}|_{\bar{V}_2} \leq \left\| \bar{h}_{n,2} \right\|_{2,\bar{V}_2}$. Using this, Lemma 2.2, and Theorem 4.1 we obtain

$$\frac{\sqrt{n}}{2\sigma_{n,1}} (U_n(h_n) - \theta_n) = \frac{1}{\sigma_{n,1}\sqrt{n}} \sum_{\ell=1}^{n} \bar{h}_{n,1}(X_\ell) + O_p\left(n^{-1/2} \sigma_{n,1}^{-1} \left\| \bar{h}_{n,2} \right\|_{2,\overline{V}_2} \right),$$

and the first part of the theorem follows immediately under the assumptions of the theorem.

If Assumption C3 holds, and because P satisfies $\mathsf{C}(1, \overline{\mathcal{U}}_1, \overline{\mathcal{V}}_1)$, the function $g_n(x) = \sum_{j \geq 0} P^j \bar{h}_{n,1}(x)$ is well defined and belongs to $\mathcal{L}_{\overline{\mathcal{V}}_1}$. Therefore, the assumption $\mathbb{E}[\overline{\mathcal{V}}_1(X_n)] < \infty$ implies that the remainder term $\frac{1}{\sigma_{n,1}\sqrt{n}} (P g_n(X_0) - P g_n(X_n))$ is negligible in the martingale decomposition

$$\frac{1}{\sigma_{n,1}\sqrt{n}} \sum_{k=1}^{n} \bar{h}_{n,1}(X_k) = \frac{1}{\sigma_{n,1}\sqrt{n}} \sum_{k=1}^{n} (g_n(X_k) - P g_n(X_{k-1}))$$
$$+ \frac{1}{\sigma_{n,1}\sqrt{n}} (P g_n(X_0) - P g_n(X_n)).$$

Set $\Delta_{n,k} := \frac{1}{\sigma_{n,1}\sqrt{n}}(g_n(X_k) - P g_n(X_{k-1}))$, and note that $\mathbb{E}[\Delta_{n,k}|\mathcal{F}_{k-1}] = 0$. We apply the martingale CLT. Notice that $\mathbb{E}[\Delta^2_{n,k}|\mathcal{F}_{k-1}] = \frac{1}{\sigma^2_{n,1} n}(P g_n^2(X_{k-1}) - \left(P g_n(X_{k-1}))^2 \right)$, and using Dvoretzky's inequality, for any $\epsilon > 0$,

$$\mathbb{E}[\Delta^2_{n,k} \mathbf{1}_{\{|\Delta_{n,k}| > \epsilon\}}|\mathcal{F}_{k-1}] \lesssim \frac{1}{\sigma^2_{n,1} n} \mathbb{E}\left[g_n^2(X_k) \mathbf{1}_{\{|g_n(X_k)| > \epsilon \sigma_{n,1}\sqrt{n}/2\}} \middle| \mathcal{F}_{k-1} \right]$$
$$= \frac{1}{\sigma^2_{n,1} n} \int P(X_{k-1}, du) g_n^2(u) \mathbf{1}_{\{|g_n(X_k)| > \epsilon \sigma_{n,1}\sqrt{n}/2\}}.$$

Therefore, using Theorem 3.2 in [10] and Assumption C3, we conclude that the martingale term $\sum_{k=1}^{n} \Delta_{n,k} \stackrel{\mathsf{w}}{\to} \mathcal{N}(0, 1)$. The proof of the stated result follows. □

5.3.3. Proof of Theorem 4.3

The proof verifies the main condition in Theorem 4.2, $\sigma_{n,1}^{-1} \left\| \bar{h}_{n,2} \right\|_{2,\overline{V}_2} = o(n^{1/2})$, because all the other conditions are satisfied by assumption. The following lemma gives the main ingredients used in the proof. Let $p \in [1,2]$. Define $\theta_n^{(s)} = \int\int h_n(x_1,x_2)\mu(dx_1)\mu(dx_2) = \theta_n$ and $h_{n,1}(x_1) = \int h_n(x_1,x_2)\mu(dx_2)$.

Lemma 5.3. *If the assumptions of Theorem 4.3 hold, then* (i) $\theta_n^{(s)} - \theta^{(s)} = O(b_n^{\varrho\wedge\varsigma})$ *and* (ii) $\left\|\!\left|\bar{h}_{n,2}\right|\!\right\|_{p,\overline{V}_2}^p = O(b_n^{-(p-1)d-p|s|})$.

Proof. For (i), change of variables, integration by parts, and a Taylor's series expansion give

$$\theta_n^{(s)} = \int\int K_{b_n}^{(s)}(z_1-z_2)g(z_1)\mu(dz_1)\mu(dz_2)$$
$$= \int\left[\int K(u)\partial^s e(z_2-ub_n)du\right]f(z_2)dz_2 = \theta^{(s)} + O(b_n^{\varrho\wedge\varsigma}).$$

For (ii), first note that $|\theta_n^{(s)}| = O(1)$, $\sup_{x_1,x_2\in\mathcal{X}^2}(\overline{V}_2(x_1,x_2)^p)^{-1}\int P(x_1,dx_2)|h_{n,1}(x_2)|^p = O(1)$ and $\sup_{x_1,x_2\in\mathcal{X}^2}(\overline{V}_2(x_1,x_2)^p)^{-1}\int P(x_1,dx_2)|h_{n,1}(x_3)|^p = O(1)$. This implies that $\left\|\!\left|\bar{h}_{n,2}\right|\!\right\|_{p,\overline{V}_2}^p \lesssim \||h_n\||_{p,\overline{V}_2}^p + O(1)$. Then, elementary bounds and change of variables give

$$\int P(x_1,dx_2)|h_n(x_2,x_3)|^p \lesssim \frac{1}{b_n^{(p-1)d+p|s|}}\int|\partial^s K(u)|^p$$
$$\times[m_{y,p}(z_3-ub_n)+|y_3|^p]q_z(z_1,z_3-ub_n)du,$$

and therefore $\||h_n\||_{p,\overline{V}_2}^p = O(b_n^{-(p-1)d-p|s|})$, as desired. □

Remark 9. The previous calculations also show that $|P\bar{h}_{n,2}|_{\overline{V}_2}^p = O(b_n^{-p|s|})$, which implies that $|P\bar{h}_{n,2}|_{\overline{V}_2} = o(\left\|\!\left|\bar{h}_{n,2}\right|\!\right\|_{p,\overline{V}_2})$ in this example.

Therefore, under the conditions of the theorem, the proof of $\sigma_{n,1}^2 - \sigma_1^2 = o(1)$ follows from previous calculations and the Cauchy–Schwartz inequality, while the first conclusion in Theorem 4.3 follows because $\sqrt{n}(\theta_n^{(s)} - \theta^{(s)}) = O(\sqrt{n}b_n^{\varrho\wedge\varsigma}) = o(1)$, $\sigma_{n,1}^{-1}\left\|\!\left|\bar{h}_{n,2}\right|\!\right\|_{2,\overline{V}_2} = O(b_n^{-d/2-|s|}) = o(n^{1/2})$ for $p=2$, and $\frac{1}{\sqrt{n}}\sum_{\ell=1}^n \psi(X_\ell) \overset{\mathsf{w}}{\to} \mathcal{N}(0,4\sigma_1^2)$ because Assumption C3 is automatically satisfied given our assumptions. □

Appendix. A weak law of large numbers for Markov chains

Proposition A.1. *Let $\{X_n, n\geq 0\}$ be a Markov chain with invariant distribution μ and transition kernel P. Assume $V_1 \leq V_2 : \mathcal{X} \to [1,\infty)$ are measurable functions such that $\mu(V_1) < \infty$ and*

$$\sum_{k\geq 0}\|P^k(x,\cdot)-\mu\|_{V_1} \lesssim V_2(x), \quad x\in\mathcal{X}. \tag{37}$$

Suppose also that $\mathsf{v}_n := \mathbb{E}^{1/p}[V_2^p(X_n)] < \infty$ for each $n\geq 0$ and for some $p>1$. Let $\{f_n, n\geq 1\}$ be such that $f_n, Pf_n \in \mathcal{L}_{V_1}$ and let $\{a_{n,k}, 0\leq k\leq n\}$ be a sequence of real numbers. Then,

$$\sum_{k=1}^n a_{n,k}(f_n(X_k)-\mu(f_n)) = \sum_{k=1}^n a_{n,k}Q_{n,k} + \epsilon_n, \tag{38}$$

where $\{(Q_{n,k}, \mathcal{F}_k), 1 \leq k \leq n\}$ *is a martingale-difference array with respect to* $\mathcal{F}_k := \sigma(X_0, \ldots, X_k)$ *and, for* $g_n(x) := \sum_{j\geq 0}(P^j f_n(x) - \mu(f_n))$*,*

$$\epsilon_n = \sum_{k=1}^{n}(a_{n,k} - a_{n,k-1})Pg_n(X_{k-1}) + \left(a_{n,0}Pg_n(X_0) - a_{n,n}Pg_n(X_n)\right).$$

Furthermore, we have

$$\begin{aligned} &\mathbb{E}\left[\left|\sum_{k=1}^{n} a_{n,k}Q_{n,k}\right|^p\right] \leq \|\!| f_n |\!\|_{p,V_1}^p \left\{\sum_{k=1}^{n}\left(|a_{n,k}|(1+\mathsf{v}_p+\mathsf{v}_{p-1})\right)^{p\wedge 2}\right\}^{1\vee\frac{p}{2}}, \quad \textit{and} \\ &\mathbb{E}\left[|\epsilon_n|^p\right] \leq |Pf_n|_{V_1}^p \left(|a_{n,0}|\mathsf{v}_0 + |a_{n,n}|\mathsf{v}_n + \sum_{k=1}^{n}|a_{n,k} - a_{n,k-1}|\mathsf{v}_{k-1}\right)^p. \end{aligned} \tag{39}$$

If $1 < p \leq 2$*, and* $\{a_{n,k}, 0 \leq k \leq n\}$ *is such that*

$$\|\!| f_n |\!\|_{p,V_1}^p \left(\sum_{k=1}^{n}|a_{n,k}|\right)^{-p} \sum_{k=1}^{n}|a_{n,k}|^p(1+\mathsf{v}_{k-1}+\mathsf{v}_k)^p \to 0$$

and

$$|Pf_n|_{V_1}\left(\sum_{k=1}^{n}|a_{n,k}|\right)^{-1}\sum_{k=1}^{n}|a_{n,k} - a_{n,k-1}|\mathsf{v}_{k-1} \to 0,$$

then $\left(\sum_{k=1}^{n}|a_{n,k}|\right)^{-1}\sum_{k=1}^{n} a_{n,k}(f_n(X_k) - \mu(f_n)) \overset{\mathsf{Pr}}{\to} 0.$

Proof. Using (37), $|g_n(x)| \lesssim |f_n|_{V_1}V_2(x)$ and $|Pg_n(x)| \lesssim |Pf_n|_{V_1}V_2(x)$. Thus, for all $1 \leq k \leq n$,

$$\mathbb{E}^{1/p}[|Pg_n(X_k)|^p] \lesssim |Pf_n|_{V_1}\mathbb{E}^{1/p}[V_2^p(X_k)] = |Pf_n|_{V_1}\mathsf{v}_k. \tag{40}$$

By the Poisson equation, $f_n(x) - \mu(f_n) = g_n(x) - Pg_n(x)$, and hence

$$\begin{aligned} \sum_{k=1}^{n} a_{n,k}(f_n(X_k) - \mu(f_n)) &= \sum_{k=1}^{n} a_{n,k}\left(g_n(X_k) - Pg_n(X_{k-1})\right) \\ &\quad + \sum_{k=1}^{n}(a_{n,k} - a_{n,k-1})Pg_n(X_{k-1}) \\ &\quad + \left(a_{n,0}Pg_n(X_0) - a_{n,n}Pg_n(X_n)\right) = \sum_{k=1}^{n} a_{n,k}Q_{n,k} + \epsilon_n, \end{aligned}$$

where $Q_{n,k} := g_n(X_k) - Pg_n(X_{k-1})$. Using the Markov property, we have $\mathbb{E}[Q_{n,k}|\mathcal{F}_{k-1}] = 0$. Hence $\{(Q_{n,k}, \mathcal{F}_k), 1 \leq k \leq n\}$ is a martingale-difference array. g_n satisfies the Poisson equation $g_n - Pg_n = f_n - \mu(f_n)$. Hence $Q_{n,k} = f_n(X_k) - \mu(f_n) + Pg_n(X_k) - Pg_n(X_{k-1})$. Using the bounds established above for g_n and Pg_n, we get $|Q_{n,k}|^p \lesssim |f_n(X_k)|^p + |\mu(f_n)|^p + |Pf_n|_{V_1}^p (V_2(X_{k-1}) + V_2(X_k))^p$. Hence by conditioning first on $\mathcal{F}_{k-1}$, it follows that

$$\mathbb{E}[|Q_{n,k}|^p] \lesssim \|\!| f_n |\!\|_{p,V_1}^p (1+\mathsf{v}_k+\mathsf{v}_{k-1})^p. \tag{41}$$

The bound (41), combined with the martingale inequality (42) applied to $\{(Q_{n,k}, \mathcal{F}_k), 1 \le k \le n\}$ yields the first part of (39). The second part of the bound (39) follows easily from the expression of ϵ_n, and (40).

Under the stated assumptions, $\left(\sum_{k=1}^n |a_{n,k}|\right)^{-1} \sum_{k=1}^n a_{n,k}(f_n(X_k) - \mu(f_n))$ converges to zero in L^p, and hence in probability. □

Remark 10. An important special case is $a_{n,\ell} = 1$ and $\sup_{n \ge 0} \mathbb{E}[V_2^p(X_n)] < \infty$, which gives a law of large numbers if $n^{-1+1/p} \, |||f_n|||_{p,V_1} \to 0$. If, in addition, $\sup_{x \in \mathcal{X}} PV_1^p(x)/V_1^p(x) < \infty$, then $|||f_n|||_{p,V_1} \lesssim |f_n|_{V_1}$ and the law of large numbers holds if $n^{-1+1/p}|f_n|_{V_1} \to 0$.

We employ the following version of Burkholder's inequality for martingales.

Lemma A.2. *Let $\{(D_k, \mathcal{F}_k), 1 \le k \le n\}$ be a martingale difference sequence for some non-decreasing filtration $\{\mathcal{F}_k, 1 \le k \le n\}$. Then, for any $p > 1$,*

$$\mathbb{E}\left[\left|\sum_{k=1}^n D_k\right|^p\right] \le C \left(\sum_{k=1}^n \mathbb{E}^{1 \wedge \frac{2}{p}}[|D_k|^p]\right)^{1 \vee \frac{p}{2}}, \tag{42}$$

where C is a finite constant that only depends on p.

Proof. By Burkholder's inequality (Theorem 2.10 in [10]), we have

$$\mathbb{E}\left[\left|\sum_{k=1}^n D_k\right|^p\right] \le C\mathbb{E}\left[\left(\sum_{k=1}^n D_k^2\right)^{p/2}\right],$$

where C is a finite constant that only depends on p. If $p \ge 2$, by Minkowski's inequality we have

$$\mathbb{E}\left[\left(\sum_{k=1}^n D_k^2\right)^{p/2}\right] \le \left\{\sum_{k=2}^n \mathbb{E}^{2/p}[|D_k|^p]\right\}^{p/2}.$$

If $1 < p < 2$, the inequality $(\sum_{i=1}^n a_i)^{p/2} \le \sum_{i=1}^n a_i^{p/2}$, which holds true for all $a_i \ge 0$, gives

$$\mathbb{E}\left[\left(\sum_{k=1}^n D_k^2\right)^{p/2}\right] \le \sum_{k=1}^n \mathbb{E}[|D_k|^p].$$

The proof of the lemma follows from these bounds. □

References

[1] Y. Atchade, Kernel estimators of asymptotic variance for adaptive Markov chain Monte Carlo, Ann. Statist. 39 (2011) 990–1011.

[2] R.J. Bhansali, L. Giraitis, P.S. Kokoszka, Approximations and limit theory for quadratic forms of linear processes, Stochastic Process. Appl. 117 (2007) 71–95.

[3] M.D. Cattaneo, R.K. Crump, M. Jansson, Robust data-driven inference for density-weighted average derivatives, J. Amer. Statist. Assoc. 105 (2010) 1070–1083.

[4] B. Cheng, P.M. Robinson, Semiparametric estimation from time series with long-range dependence, J. Econometrics 64 (1994) 335–353.

[5] H. Dehling, M. Wendler, Central limit theorem and the bootstrap for U-statistics of strongly mixing data, J. Multivariate Anal. 101 (2010) 126–137.

[6] R. Douc, G. Fort, E. Moulines, P. Soulier, Practical drift conditions for subgeometric rates of convergence, Ann. Appl. Probab. 14 (2004) 1353–1377.

[7] G.K. Eagleson, Orthogonal expansions and U-statistics, Austral. J. Statist. 21 (1979) 221–237.

[8] J.M. Flegal, G.L. Jones, Batch means and spectral variance estimators in Markov chain Monte Carlo, Ann. Statist. 38 (2010) 1034–1070.

[9] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, second ed., in: Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, FL, 2004.

[10] P. Hall, C.C. Heyde, Martingale Limit Theory and its Application, Academic Press, New York, 1980.

[11] T. Hsing, W.B. Wu, On weighted U-statistics for stationary processes, Ann. Probab. 32 (2004) 1600–1631.

[12] N.M. Kiefer, T.J. Vogelsang, A new asymptotic theory for heteroskedasticity-autocorrelation robust tests, Econom. Theory 21 (2005) 1130–1164.

[13] T. Kurtz, P. Protter, Weak limit theorems for stochastic integrals and stochastic differential equations, Ann. Probab. 19 (1991) 1035–1070.

[14] M. Maxwell, M. Woodroofe, Central limit theorems for additive functional of Markov chains, Ann. Probab. 28 (2000) 713–724.

[15] D.L. McLeish, Dependent central limit theorems and invariance principles, Ann. Probab. 2 (1974) 620–628.

[16] M. Meitz, P. Saikkonen, Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models, Econom. Theory 24 (2008) 1291–1320.

[17] F. Merlevede, M. Peligrad, S. Utev, Recent advances in invariances principles for stationary sequences, Probab. Surv. 3 (2006) 1–36.

[18] S.P. Meyn, R.L. Tweedie, Markov Chains and Stochastic Stability, second ed., Cambridge University Press, London, 2009.

[19] H.R. Neave, An improved formula for the asymptotic variance of spectrum estimates, Ann. Math. Statist. 41 (1970) 70–77.

[20] J.L. Powell, J.H. Stock, T.M. Stoker, Semiparametric estimation of index coefficients, Econometrica 57 (1989) 1403–1430.

[21] M.B. Priestley, Spectral Analysis and Time Series, vol. 1, Academic Press Inc. (Harcourt Brace Jovanovich Publishers), London, 1981, Univariate Series, Probability and Mathematical Statistics.

[22] G.O. Roberts, J.S. Rosenthal, General state space Markov chains and MCMC algorithms, Probab. Surv. 1 (2004) 20–71. electronic.

[23] P.M. Robinson, Hypothesis testing in semiparametric and nonparametric models for econometric time series, Rev. Econom. Stud. 56 (1989) 511–534.

[24] R.J. Serfling, Approximation Theorems of Mathematical Statistics, Wiley, New York, 1980.

[25] T.M. Stoker, Consistent estimation of scaled coefficients, Econometrica 54 (1986) 1461–1481.

[26] W.B. Wu, X. Shao, A limit theorem for quadratic forms and its applications, Econom. Theory 23 (2007) 930–951.

[27] K.-i. Yoshihara, Limiting behavior of U-statistics for stationary, absolutely regular processes, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 35 (1976) 237–252.