LIMIT THEOREMS FOR QUADRATIC FORMS OF MARKOV CHAINS

YVES F. ATCHADÉ AND MATIAS D. CATTANEO

(April 2011)

Abstract. We develop a martingale approximation approach to studying the limiting behavior of quadratic forms of Markov chains. We use the technique to examine the asymptotic behavior of lag-window estimators in time series and we apply the results to Markov Chain Monte Carlo simulation. As another illustration, we use the method to derive a central limit theorem for U-statistics with varying kernels.

1. Introduction

This paper deals with quadratic forms of the type

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

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 asymptotic properties are of particular importance to develop asymptotically valid inference procedures.

For independent sequences $\{X_n,\ n \geq 0\}$, 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., Serfling (1980) for a review. When the process $\{X_n,\ n \geq 0\}$ is time-dependent, however, the classical Hoeffding decomposition is not very 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. The most well understood case is the case of a standard U-statistics where h_n does not depend on n and $w_n(\ell, j) = 1$ if $\ell \neq j$ and 0 otherwise (Yoshihara (1976); Eagleson

2000 *Mathematics Subject Classification.* 60J10, 62M10.

Key words and phrases. Central limit theorems, Markov Chains, Markov Chain Monte Carlo, Martingale approximations, Quadratic forms, U-statistics.

Y. F. Atchadé: University of Michigan, Department of Statistics, Ann Arbor, 48109, MI, United States. *E-mail address:* yvesa@umich.edu.

Matias D. Cattaneo: University of Michigan, Department of Economics, Ann Arbor, 48109, MI, United States. *E-mail address:* cattaneo@umich.edu.

(1979); Dehling and Wendler (2010)). There has been some recent progress. Hsing and Wu (2004) considers $U_n(h)$ where neither h_n nor w_n depends on n, whereas Wu and Shao (2007) studies $U_n(h_n)$ when $h_n(x,y) = h(x,y) = xy$ for a martingale-difference sequence (see also Bhansali et al. (2007) for i.i.d. sequences).

We develops a martingale approximation for $U_n(h_n)$ which allows for a general and systematic analysis of $U_n(h_n)$ 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 (Maxwell and Woodroofe (2000); Merlevede et al. (2006)), but has not been fully explored in dealing with quadratic forms (a notable exception is Wu and Shao (2007)). In the present paper, we obtain an approximating quadratic martingale to $U_n(h_n)$ from a solution of a bivariate analog of the well known Poisson's equation.

As an application we study the asymptotic behavior of lag-window estimators of long-run variance (asymptotic variance) for Markov chains (see, e.g., Priestley (1981)). 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 (Neave (1970); Kiefer and Vogelsang (2005)). 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 (2010) and Atchade (2011); 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.

As another application of the martingale approximation method, we derive a central limit theorem for U-statistics with varying kernels without imposing stationarity and under assumptions that are more easily verifiable. In particular, we do not rely on mixing conditions.

The paper is organized as follows. The rest of the introduction outlines the general setup and introduces the main notation employed throughout, while Section 2 derives the main martingale approximation method. Section 3 derives the asymptotic properties of lag-window estimators and, in particular, applies these results to MCMC simulation. We study U-statistics with varying kernels in Section 4. All the proofs are presented in Section 5.

1.1. **Setup and Notation.** 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 countably generated 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 will rely on the following set of general notation. Suppose that $(\mathbb{T}, \mathcal{A})$ be 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

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

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 $\mu\nu$ or $\mu \bigotimes \nu$, and the product of a finite number k of finite signed measures $\nu_1, \dots, \nu_k$ is denoted by $\bigotimes_{j=1}^k \nu_j$.

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 $n \ge 1$, 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 $Qh(x, y) = \int Q(x, dz)h(z, y)$ 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 $p \ge 1$ and a function $h : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$, we define

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

For a univariate function $V : \mathbb{T} \to [1, \infty)$ and for $h : \mathbb{T} \to \mathbb{R}$, we define $|\!|\!| h |\!|\!|_{p,V}$ similarly as

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

When we use the notation $|\!|\!| h |\!|\!|_{p,V}$ below, it will always be with respect to P, the Markov kernel of the reference process $\{X_n,\ n \ge 0\}$, unless stated otherwise. The following short-range dependence concept will play an important rule.

Definition 1.1. *Fix $r \in \mathbb{N}$. For measurable functions $\overline{V}_r \le \overline{W}_r : \mathbb{T}^r \to [1, \infty)$, we say that the transition kernel Q with invariant distribution μ satisfies the condition $\mathsf{C}(r, \overline{V}_r, \overline{W}_r)$ if there exists a finite constant c such that*

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

Throughout the paper, we denote by c a finite constant which depends solely on the kernel P but whose actual value can change from one equation to the next. In particular c does not depend on the family of function $\{h_n,\ n \ge 1\}$ considered. Finally, 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 A1 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, $P^s \overline{W}_2(x) < \infty$ for all $x \in \mathcal{X}^2$ and for $s \in \{1, 2\}$.

Remark 1. It is always possible to deduce A1 from a univariate short-range dependence assumption. 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

$$\| \left(P^n(x, \cdot) - \mu\right) \bigotimes \left(P^m(y, \cdot) - \mu\right) \|_{\overline{V}_2} = \|P^n(x, \cdot) - \mu\|_{V_1} \|P^m(y, \cdot) - \mu\|_{V_2}.$$

Thus

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

and therefore A1 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 and this has been done for several time series Markov models (Douc et al. (2004); Meitz and Saikkonen (2008); Meyn and Tweedie (2009)).

We show that whenever A1 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 on $\mathcal{X} \times \mathcal{X}$ with finite $\|\cdot\|_{\overline{V}_2}$ norm, equipped with the norm $\|\cdot\|_{\overline{V}_2}$ is a Banach space. Under A1 and for any $x, y \in \mathcal{X}$,

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

is a finite signed measure that belongs to $\mathcal{M}_{\overline{V}_2}(\mathcal{X} \times \mathcal{X})$. Furthermore we have for all $x, y \in \mathcal{X}$,

$$\|\bar{R}_2(x, y; \cdot)\|_{\overline{V}_2} \leq c\overline{W}_2(x, y). \tag{3}$$

Let $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric measurable function such that $\bar{\mu}(|h|) < \infty$. Denote $\theta = \int \int h(x, y)\mu(dx)\mu(dy)$ and define

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

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

We say that h is degenerate when $\bar{h}_1$ is identically zero. For $x \in \mathcal{X}$, δ_x denotes the Dirac measure at x.

Lemma 2.1. *Assume A1. Suppose that $\bar{h}_2 \in \mathcal{L}_{\overline{V}_2}$. Then $\bar{G}_2$ is well-defined, $\bar{G}_2 \in \mathcal{L}_{\overline{W}_2}$, and $|\bar{G}_2|_{\overline{W}_2} \le c|\bar{h}_2|_{\overline{V}_2}$ and for all $x, y \in \mathcal{X}$,*

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

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

Proof. See Section 5.1. □

Remark 3. Equation (4) gives a bivariate Poisson's equation which extends the well known univariate Poisson's equation.

We introduce the function

$$\begin{aligned}\Lambda_2(x_1,x_2;y_1,y_2) &= \int (\delta_{y_1}(dz_1) - P(x_1,dz_1)) \int (\delta_{y_2}(dz_2) - P(x_2,dz_2))\, G_2(z_1,z_2) \\ &= \bar{G}_2(y_1,y_2) - P\bar{G}_2(x_2,y_1) - P\bar{G}(x_1,y_2) + P^2\bar{G}_2(x_1,x_2), \quad x_1,x_2,y_1,y_2 \in \mathcal{X}.\end{aligned}$$

Then (4) can be written as $\bar{h}_2(x,y) = \Lambda_2(x,y,x,y)$. A specially important property of Λ_2 that we rely on in the sequel is the following. For any $x,y,u,v \in \mathcal{X}$, it is easy to see that

$$\int P(x,dy)\Lambda_2(u,x,v,y) = \int P(u,dv)\Lambda_2(u,x,v,y) = 0. \quad (5)$$

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 θ, $\bar{h}_1$, $\bar{h}_2$, $\bar{G}_2$, and Λ_2 defined above with $h = h_n$.

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 (5), we have

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

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. We need the following sequences

$$w_{n,1}(\ell) := \left\{\sum_{j=1}^{\ell} w_n(\ell,j) + \sum_{j=\ell}^{n} w_n(j,\ell)\right\}, \quad \varpi_{n,1}(\ell) := w_n(\ell,j) - w_n(\ell-1,j)$$

$$\varpi_{n,2}(\ell,j) := w_n(\ell,j) - w_n(\ell,j-1),$$

$$\text{and} \quad \varpi_{n,3}(\ell,j) := w_n(\ell,j) + w_n(\ell-1,j-1) - w_n(\ell,j-1) - w_n(\ell-1,j).$$

Lemma 2.2. *Assume A1 and suppose that* $\bar{h}_{n,2} \in \mathcal{L}_{\overline{V}_2}$ *for each* $n \geq 1$. *Then*

$$U_n(h_n) = U_{n,0} + \sum_{\ell=1}^{n} \left\{ w_{n,1}(\ell)\bar{h}_{n,1}(X_\ell) + w_n(\ell,\ell)Q_{n,\ell,\ell} \right\} + \sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1} w_n(\ell,j)Q_{n,\ell,j} + \zeta_n, \quad (6)$$

where $U_{n,0} = \theta_n \sum_{\ell=1}^n \sum_{j=1}^{\ell} w_n(\ell,j)$, *and*

$$\zeta_n = \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} \varpi_n^{(1)}(\ell,j)\left(P\bar{G}_{n,2}(X_{\ell-1},X_j) - P^2\bar{G}_{n,2}(X_{\ell-1},X_{j-1})\right) + \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} \varpi_n^{(2)}(\ell,j)\left(P\bar{G}_{n,2}(X_{j-1},X_\ell) - P^2\bar{G}_{n,2}(X_{\ell-1},X_{j-1})\right) + \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} \varpi_n^{(3)}(\ell,j)P^2\bar{G}_{n,2}(X_{\ell-1},X_{j-1}) + \epsilon_n,$$

where

$$\epsilon_n = \sum_{\ell=1}^{n} \left\{ w_n(\ell,0)P\bar{G}_{n,2}(X_0,X_\ell) - w_n(\ell-1,0)P^2\bar{G}_{n,2}(X_0,X_{\ell-1}) \right) + \sum_{j=1}^{n} w_n(n,j)\left(P^2\bar{G}_{n,2}(X_n,X_j) - P\bar{G}_{n,2}(X_n,X_j)\right) + \sum_{\ell=1}^{n}\left(w_n(\ell-1,\ell)P\bar{G}_{n,2}(X_{\ell-1},X_\ell) - w_n(\ell,\ell)P\bar{G}_{n,2}(X_\ell,X_\ell)\right).$$

Proof. See Section 5.2. □

Remark 4. The usefulness of this decomposition comes from the fact that the remainder ζ_n involves either single summations or difference sequences of the weights w_n. As a result, these remainders are typically negligible compared to the other terms in the decomposition and one can easily study the asymptotic behavior of $U_n(h_n)$ by focusing on the linear term $\sum_{\ell=1}^n \left\{ w_{n,1}(\ell)\bar{h}_{n,1}(X_\ell) + w_n(\ell,\ell)Q_{n,\ell,\ell} \right\}$, and the quadratic martingale $\sum_{\ell=1}^n \sum_{j=1}^{\ell-1} w_n(\ell,j)Q_{n,\ell,j}$.

3. Application: Asymptotic variance estimation

In this section, we use the martingale approximation of Lemma 2.2 to study the asymptotics of lag-windows estimators of asymptotic variance in time series and we apply the results to Markov chain Monte Carlo. 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:

$$\sigma^2(h) = \mathsf{Var}_\mu(h(X_0)) + 2\sum_{\ell \geq 1} \mathsf{Cov}_\mu\left(h(X_0), h(X_\ell)\right), \quad (7)$$

which plays a role in time series analysis and in Markov Chain Monte Carlo. A classical estimator for $\sigma^2(h)$ is the lag-windows estimators defined as

$$\Gamma^2_{n,b}(h) := \gamma_{n,0} + 2\sum_{k=1}^{n-1} w_b(kc_n^{-1})\gamma_{n,k}, \tag{8}$$

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 is a weight function (with a parameter b) and $\{c_n,\ n\geq 1\}$ is an increasing sequence of positive numbers. We refer the reader to Priestley (1981) for detailed discussion on lag-windows estimators. We consider weight functions with the following properties.

Assumption W: For $b>0$, $w_b:\ [0,\infty)\to[0,1]$ is a continuous function with support $[0,b]$, 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. When $w_b(x)=w(x/b)$, an equivalent parametrization of $\Gamma^2_{n,b}(h)$ is $\Gamma^2_{n,b}(h) = \gamma_{n,0}+2\sum_{k=1}^{n-1} w(k/c_n)\gamma_{n,k}$, with $c_n \leftarrow bc_n$. We impose the following ergodicity assumption.

Assumption A2 There exist measurable functions $V_k:\mathcal{X}\to[1,\infty)$ $(k=1,2,3)$, $V_1\leq V_2$, $V_2^2\leq V_3$, such that $PV_3(x)<\infty$ 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

$$\sup_{n\geq 0}\mathbb{E}\left(V_3^q(X_n)\right)<\infty. \tag{9}$$

A2 implies A1 with $\overline{V}_2(x,y)=V_1(x)V_1(y)$ and $\overline{W}_2(x,y)=V_2(x)V_2(y)$. Define the partial sums $S_{n,k} := \sum_{j=k+1}^{k+n} h(X_j)$, and the weight $w_{n,b}(0)=n^{-1}$ and $w_{n,b}(k)=2n^{-1}w_b(kc_n^{-1})$ for $k>0$. We can rewrite $\gamma_{n,k}$ as

$$\gamma_{n,k} = n^{-1}\sum_{j=1}^{n-k} h(X_j)h(X_{j+k}) + n^{-3}(n-k)S_{n,0}^2 - n^{-2}S_{n,0}\left(S_{n-k,0}+S_{n-k,k}\right),$$

so that

$$\Gamma^2_{n,b}(h) = \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_{n,b}(\ell-j)h(X_j)h(X_\ell) + R_n, \tag{10}$$

where

$$R_n = 2n^{-2}S_{n,0}^2\sum_{k=1}^{n-1} w_b(kc_n^{-1})\left(1-\frac{k}{n}\right) \\ - 2n^{-2}S_{n,0}\left(\sum_{j=2}^{n} h(X_j)\sum_{k=1}^{j-1} w_b(kc_n^{-1}) + \sum_{j=1}^{n-1} h(X_j)\sum_{k=1}^{n-j} w_b(kc_n^{-1})\right) - n^{-2}S_{n,0}^2. \tag{11}$$

If we set aside the term R_n, the expression (10) is of the form (1) with $h_n(x,y)=h(x)h(y)$ and $w_n(\ell,j)=w_{n,b}(\ell-j)$. Here we have $h_{n,1}(x)=\int h(x)h(y)\mu(dy)=0$, $\theta_n=0$, and

$h_{n,2}(x,y) = h(x)h(y)$. Define

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

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

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

As above, $\{(Q_\ell, \mathcal{F}_\ell),\ \ell \geq 1\}$ is a martingale: $\mathbb{E}\left(Q_\ell | \mathcal{F}_{\ell-1}\right) = 0$. From Lemma 2.2 we obtain the following.

Theorem 3.1. *Assume (A2) and (W) and $h \in \mathcal{L}_{V_1}$. For all $n \geq 1$,*

$$\Gamma^2_{n,b}(h) = n^{-1}\sum_{\ell=1}^n Q_\ell^2 + \sum_{\ell=1}^n\sum_{j=1}^{\ell-1} w_{n,b}(\ell-j)Q_\ell Q_j + R_n + \zeta_n. \tag{12}$$

Furthermore, there exist $p > 1$ and a finite constant c such that for all $n \geq 3$,

$$\mathbb{E}^{1/p}\left(|\zeta_n|^p\right) \leq c c_n^{-1+\frac{1}{2}\vee\frac{1}{p}}, \quad \mathbb{E}^{1/p}\left(|R_n|^p\right) \leq c n^{-1} c_n,$$
$$\text{and} \quad \mathbb{E}^{1/p}\left(\left|\sum_{\ell=1}^n\sum_{j=1}^{\ell-1} w_{n,b}(\ell-j)Q_\ell Q_j\right|^p\right) \leq c\left(\frac{c_n}{n}\right)^{\frac{1}{2}} n^{-\frac{1}{2}+\frac{1}{p}\vee\frac{1}{2}}.$$

Proof. See Section 5.3. □

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

$$\Gamma^2_{n,b}(h) = \underbrace{n^{-1}\sum_{\ell=1}^n Q_\ell^2}_{O_p(1)} + \underbrace{\sum_{\ell=1}^n\sum_{j=1}^{\ell-1} w_{n,b}(\ell-j)Q_\ell Q_j + R_n}_{O_p\left(\sqrt{\frac{c_n}{n}}+\frac{c_n}{n}\right)} + \underbrace{\zeta_n}_{O_p(c_n^{-1/2})}, \tag{13}$$

By the law of large numbers for Markov chain the term $n^{-1}\sum_{\ell=1}^n Q_\ell^2$ converges to $\sigma^2(h)$. As the result, Theorem 3.1 implies that $\Gamma^2_{n,b}(h)$ converges in probability 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 (13) 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 then $R_n + \sum_{\ell=1}^n\sum_{j=1}^{\ell-1} w_{n,b}(\ell-j)Q_\ell Q_j \approx O(1)$. 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.

When the goal is to construct 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 $\Gamma^2_{n,b}(h)$ with $c_n = n$, the so-called "fixed-b asymptotics" (Neave (1970); Kiefer and Vogelsang (2005)). With $c_n = n$, $\Gamma^2_{n,b}(h)$ no longer converges to $\sigma^2(h)$, but as it turns out, asymptotically valid confidence intervals can still be derived for $\mu(h)$. We have the following.

Theorem 3.2. *Under the assumption of Theorem 3.1, the following holds true.*

(1) *If $p \geq 2$ and $c_n = o(n)$, then $\Gamma^2_{n,b}(h)$ converges in probability to $\sigma^2(h)$. Furthermore, assuming $\Gamma^2_{n,b}(h) > 0$ almost surely,*

$$\left\{n\Gamma^2_{n,b}(h)\right\}^{-1/2} \sum_{j=1}^{n} (h(X_j) - \mu(h)) \xrightarrow{w} \mathcal{N}(0,1).$$

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

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

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

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

Proof. See Section 5.4. □

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

$$\mu_n(h) \pm z_{1-\alpha/2}\frac{\hat{\sigma}_n(h)}{\sqrt{n}}, \tag{14}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution and where $\hat{\sigma}_n(h) = \sqrt{\Gamma^2_{n,b}(h)}$, with $b = 1$, $c_n = o(n)$. Typical choice of c_n includes $c_n = n^{-\delta}$, $\delta \in (0,1)$ typically around 0.5. Theorem 3.1 (2) provides another asymptotically valid confidence interval for $\mu(h)$:

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

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 Euler 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.

In the next simulation examples, we compare the finite sample properties of these two confidence intervals in terms of coverage probability and interval length. All the simulations are performed using the Bartlett kernel $w(x) = 1 - x$.

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

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

3.1. **Illustration: the Garch$(1,1)$ model.** Consider the linear GARCH$(1,1)$ model defined as follows. $h_0 \in (0,\infty)$, $u_0 \sim \mathcal{N}(0,h_0)$ and for $n \geq 1$

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

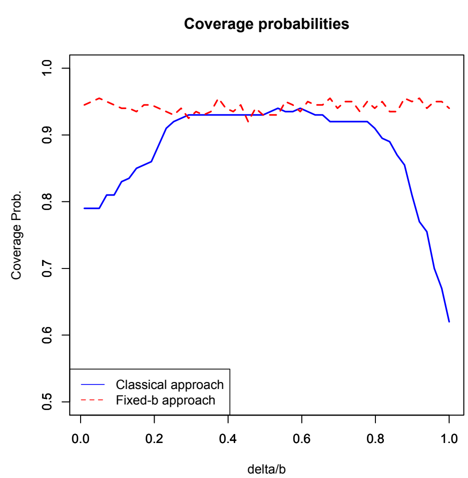
where $\{\epsilon_n,\ n \geq 0\}$ is i.i.d. $\mathcal{N}(0,1)$ and $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$. We assume that α, β satisfy

C1 There exists $\nu > 0$ such that

$$\mathbb{E}\left[\left(\beta + \alpha Z^2\right)^\nu\right] < 1, \quad Z \sim \mathcal{N}(0,1). \tag{16}$$

It is shown by Meitz and Saikkonen (2008) (Theorem 2) that under (16) 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 $\nu \geq 2$, A2 holds with $V_1 = V_2 = V^{1/2}$, and $V_3 = V$. We are interested in a confidence interval for $\mu(h)$ where $h(u) = u^2$ which belongs to $\mathcal{L}_{V_1}$. The exact value is $\mu(h) = \omega(1-\alpha-\beta)^{-1}$.

For the simulations we set $\omega = 1$, $\alpha = 0.1$, $\beta = 0.7$ which gives $\mu(h) = 5$. We compare the confidence intervals (14) and (15) 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 Figure 1 and 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.



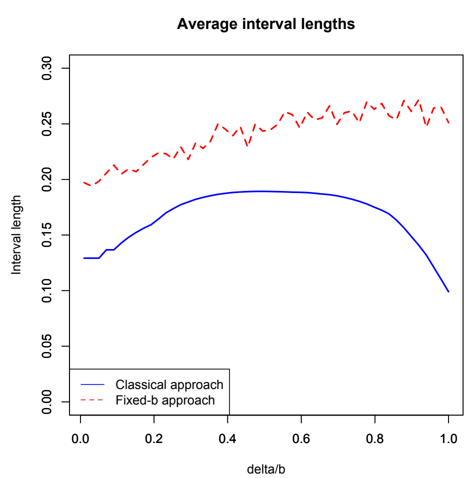


Figure 1: Coverage probabilities plots for various values of δ (classical confidence interval) and b (fixed-b confidence interval).

3.2. **Markov Chain Monte Carlo.** Markov Chain Monte Carlo (MCMC) is a popular computational tools to obtain random samples from intractable and high-dimensional distributions (see e.g. Roberts and Rosenthal (2004) for a survey and for additional references).

Suppose that we interested in sampling from the probability measure μ and compute the integral $\mu(h) = \int h(x)\mu(dx)$. Let $\{X_n,\ n \geq 0\}$ be a 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 A2, $\lim_{n\to\infty} n^{1/2}\mathsf{Var}\left(\mu_n(h)\right) = \sigma^2(h)$, as given by (7), and a central limit theorem holds: $n^{-1/2}\sum_{k=1}^n \left(h(X_k) - \pi(h)\right) \xrightarrow{w} N(0, \sigma^2(h))$. Therefore (14) and (15) provide two valid confidence intervals for $\mu(h)$. We compare the coverage probabilities and average interval lengths of these two confidence interval procedures with the following simulation example.

3.2.1. *Illustration: a Poisson regression model.* We undertake the comparison using a log-linear model taken from Gelman et al. (2004). For $e = 1, \ldots, N_e$ and $p = 1, \ldots, N_p$, the variables y_{ep} are conditionally independent given $(\{\beta_p\},\ \{\varepsilon_{ep}\}) \in \mathbb{R}^{N_p} \times \mathbb{R}^{N_e N_p}$, with conditional distribution

$$y_{ep} \sim \mathcal{P}\left(n_{ep} e^{\mu+\alpha_e+\beta_p+\varepsilon_{ep}}\right),\ \ e = 1, \ldots, N_e,\ \ p = 1, \ldots, N_p, \tag{17}$$

where $\mathcal{P}(\lambda)$ is the Poisson distribution with parameter λ. In the above display, $\{n_{ep}\}$ is a deterministic baseline covariate, and $\mu \in \mathbb{R}$, $\{\alpha_e\} \in \mathbb{R}^{N_e}$ are parameters. We assume that $\{\beta_p\}$ and $\{\varepsilon_{ep}\}$ are independent with distributions

$$\beta_p \overset{iid}{\sim} N(0, \sigma_\beta^2), \quad \varepsilon_{ep} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2),\ \ e = 1, \ldots, N_e,\ \ p = 1, \ldots, N_p, \tag{18}$$

for some parameters $\sigma_\beta^2 > 0$, $\sigma_\varepsilon^2 > 0$. We assume a diffuse prior for $(\mu, \alpha, \sigma_\beta^2, \sigma_\varepsilon^2)$ $(\sigma_\epsilon^2 > 0, \sigma_\beta^2 > 0)$ with the additional constraint that $\alpha_{N_e} = -\sum_{k=1}^{N_e-1}\alpha_k$. Let $\theta = (\mu, \alpha, \beta, \epsilon, \sigma_\epsilon^2, \sigma_\beta^2) \in \mathbb{R}^{3+N_e-1+(N_p+1)N_e}$. The posterior distribution of θ given $\mathcal{D} = (y_{ep}, n_{ep})$ takes the form

$$\pi(\theta|\mathcal{D}) \propto \exp\left\{\sum_{e,p} y_{e,p}(\mu + \alpha_e + \beta_p + \epsilon_{e,p}) - n_{ep}e^{\mu+\alpha_e+\beta_p+\epsilon_{ep}}\right.$$
$$\left. - \frac{N_e N_p}{2}\log\sigma_\epsilon^2 - \frac{N_p}{2}\log\sigma_\beta^2 - \frac{1}{2\sigma_\epsilon^2}\sum_{e,p}\epsilon_{e,p}^2 - \frac{1}{2\sigma_\beta^2}\sum_{p=1}^{N_p}\beta_p^2\right\}. \tag{19}$$

This posterior distribution is typical of probability distributions for which MCMC is useful. We set $N_e = 3$ and $N_p = 20$. Suppose that we are interested in a confidence interval for the posterior mean of the parameter α_1, i.e. $\int \alpha_1\pi(\theta|\mathcal{D})d\theta$. To compare the two confidence intervals methods described above, we generate an artificial dataset with

$(\alpha_1, \alpha_2, \mu, \sigma_\varepsilon^2, \sigma_\beta^2) = (0.35, 0.15, -1.0, 0.1, 0.3)$. We run a preliminary MCMC sampler for 6 millions (6×10^6) iterations and compute its sample mean. We obtain $\bar{\alpha}_1 = 0.3309$. We take this value to be $\int \alpha_1 \pi(\theta) d\theta$.

To compare the two confidence interval methods, we use a Random Walk Metropolis (RWM) algorithm with proposal kernel $\mathcal{N}(0, \kappa\Sigma)$ where κ and Σ are selected (from a preliminary simulation) to yield a reasonably good mixing of the chain. We run the MCMC sampler for $60,000$ iterations and discard the first $10,000$ iterations as burn-in. We repeat the simulations 200 times in order to estimate the coverage probabilities and interval lengths. The results are given in Figure 2. We find again that in terms of finite sample behavior, the fixed-b confidence interval is more robust to the choice of b.

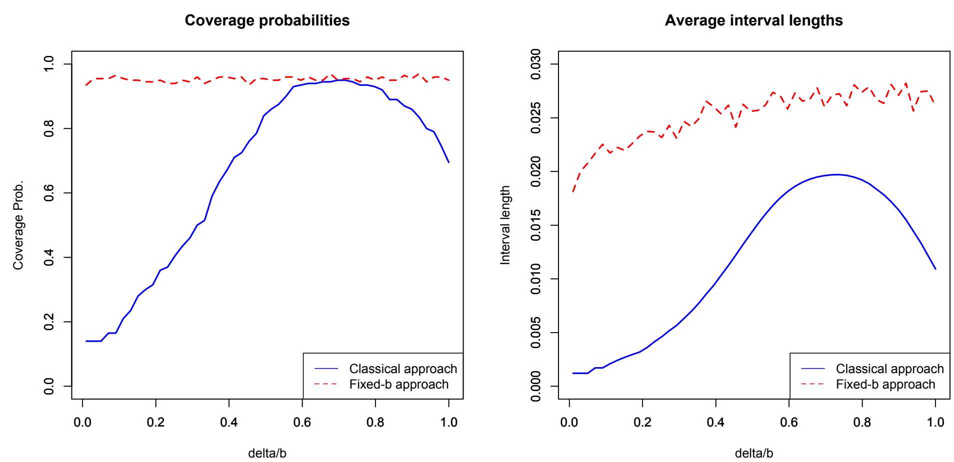


Figure 2: Coverage probability and confidence interval length for α_1 and for different values of b and δ.

4. Application: a CLT for U-statistics with varying kernels

U-statistics with varying kernels are a special case of quadratic forms and correspond to setting $w_n(\ell, \ell) = 0$ and $w_n(\ell, j) = 1$ if $\ell \neq j$. We thus have

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

We illustrate another application of Lemma 2.2, by deriving a CLT for $U_n(h_n)$. U-statistics and U-statistics with varying kernels play an important role in nonparametric and semi-parametric statistics. In the present case, under A1, Lemma 2.2 reduces to

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

We impose the following moment assumption

B1 With $\overline{V}_2$ and $\overline{W}_2$ as in A1, suppopse that $P\overline{V}_2 \leq c\overline{V}_2$, $P\overline{W}_2^2 \leq c\overline{W}_2^2$, and for $p = 2$,

$$\sup_{\ell,j\geq 0} \mathbb{E}\left(\overline{W}_2^p(X_\ell, X_j)\right) < \infty. \tag{20}$$

We recall from Lemma 2.1 that for $s \in \{1,2\}$, $|P^s \bar{G}_{n,2}(x,y)| \leq c|P^s \bar{h}_{n,2}|_{\overline{V}_2} \overline{W}_2(x,y)$, and

$$\begin{aligned}
|P^2\bar{h}_{n,2}|_{\overline{V}_2} &= \sup_{x,y\in\mathcal{X}} \{\overline{V}_2(x,y)\}^{-1} \left| \int P(x,du) \int P(y,dv)\bar{h}_{n,2}(u,v) \right| \\
&\leq |P\bar{h}_{n,2}|_{\overline{V}_2} \sup_{x,y\in\mathcal{X}} \{\overline{V}_2(x,y)\}^{-1} \int P(x,du)\overline{V}_2(u,y) \leq c|P\bar{h}_{n,2}|_{\overline{V}_2} \leq c \left|\!\left|\!\left| \bar{h}_{n,2} \right|\!\right|\!\right|_{p,\overline{V}_2},
\end{aligned}$$

for all $p \geq 1$, using the assumption $P\overline{V}_2 \leq c\overline{V}_2$ in B1. In combination with (20), and the expression of ζ_n in Lemma 2.2, it follows that

$$\mathbb{E}^{1/2}\left(|\zeta_n|^2\right) \leq c\, n \left(|P\bar{h}_{n,2}|_{\overline{V}_2} + |P^2\bar{h}_{n,2}|_{\overline{V}_2}\right) \leq c\, n \left|\!\left|\!\left| \bar{h}_{n,2} \right|\!\right|\!\right|_{2,\overline{V}_2}. \tag{21}$$

By definition, $Q_{n,\ell,j} = \bar{h}_{n,2}(X_\ell, X_j) - P\bar{G}_{n,2}(X_{\ell-1}, X_j) - P\bar{G}_{n,2}(X_{j-1}, X_\ell) + P^2\bar{G}_{n,2}(X_{\ell-1}, X_{j-1})$ and assuming that $P\bar{h}_{n,2}, \in \mathcal{L}_{\overline{V}_2}$, one obtains

$$\begin{aligned}
|Q_{n,\ell,j}| \leq \left|\bar{h}_{n,2}(X_\ell, X_j)\right| + c|P\bar{h}_{n,2}|_{\overline{V}_2} \left(\overline{W}_2(X_{\ell-1}, X_j) + \overline{W}_2(X_{j-1}, X_\ell)\right) \\
+ c|P^2\bar{h}_{n,2}|_{\overline{V}_2} \overline{W}_2(X_{\ell-1}, X_{j-1}).
\end{aligned} \tag{22}$$

Thus for $j < \ell$,

$$\begin{aligned}
\mathbb{E}\left(Q_{n,\ell,j}^2 | \mathcal{F}_{\ell-1}\right) &\leq 4 \int P(X_{\ell-1}, dz) \left|\bar{h}_{n,2}(z, X_j)\right|^2 \\
&+ 4 \left|\!\left|\!\left| \bar{h}_{n,2} \right|\!\right|\!\right|_{2,\overline{V}_2} \mathbb{E}\left(\overline{W}_2^2(X_{\ell-1}, X_j) + \overline{W}_2^2(X_\ell, X_{j-1}) + \overline{W}_2^2(X_{\ell-1}, X_{j-1}) | \mathcal{F}_{\ell-1}\right).
\end{aligned}$$

Taking the expectation on both side and using (20), it follows that for all $n \geq 1$,

$$\mathbb{E}^{1/2}\left(Q_{n,\ell,j}^2\right) \leq c \left|\!\left|\!\left| \bar{h}_{n,2} \right|\!\right|\!\right|_{2,\overline{V}_2}. \tag{23}$$

We impose an addition stability assumptions.

B2 With $\overline{W}_2$ as in A1, there exists measurable functions $\overline{\mathcal{U}}_1 \leq \overline{\mathcal{V}}_1 : \mathcal{X} \to [1,\infty)$, a symmetric measurable function $\overline{\mathcal{U}}_2 : \mathcal{X} \times \mathcal{X} \to [1,\infty)$ such that P satisfies $\mathsf{C}(1, \overline{\mathcal{U}}_1, \overline{\mathcal{V}}_1)$ and for all $m \geq 0$, all $x_1, x_2, x_3 \in \mathcal{X}$,

$$\begin{aligned}
\int P(x_1, dz_1) \int P^m(z_1, dz_2) \\
\times \left(\overline{W}_2(z_2, z_1) + \overline{W}_2(z_2, x_1)\right) \left(\overline{W}_2(z_2, x_2) + \overline{W}_2(z_2, x_3)\right) \\
\leq c\overline{\mathcal{U}}_1(x_1)\overline{\mathcal{U}}_2(x_2, x_3).
\end{aligned} \tag{24}$$

Futhermore,

$$\sup_{\ell\geq 1} \mathbb{E}\left(\overline{\mathcal{V}}_1(X_\ell)\overline{\mathcal{U}}(X_\ell, X_{\ell-1})\right) < \infty.$$

Remark 5. Assumptions B1-B2 are similar to the assumptions imposed in Dehling and Wendler (2010) (Theorem 1.8) to obtain a CLT for U-statistics of stationary dependent processes. Assumption B2 can be easy to check. For example if $\overline{W}_2$ is given by $\overline{W}_2(x,y) = W(x)W(y)$ and $P^m W \leq cW$ for all $m \geq 0$, then (24) holds with $\overline{\mathcal{U}}_1(x) = PW^2(x) + W(x)PW(x)$ and $\overline{\mathcal{U}}_2(x,y) = W(x) + W(y)$.

Define

$$\sigma_{n,1}^2 := \int \{\mu P\}(dx,dy) L_n^2(x,y)$$
$$= \mathsf{Var}_\mu\left(h_{n,1}(X_0)\right) + 2\sum_{\ell \geq 1} \mathsf{Cov}_\mu\left(h_{n,1}(X_0), h_{n,1}(X_\ell)\right), \quad \text{and} \quad \sigma_n^2 := n(n-1)^2\sigma_{n,1}^2.$$

Theorem 4.1. *Assume A1, B1-B2 and let $\{h_n,\, n \geq 1\}$ be such that $\bar{h}_{n,2} \in \mathcal{L}_{\overline{V}_2}$. Suppose also that*

$$|||\bar{h}_{n,2}|||_{2,\overline{V}_2} = o\left(n^{1/2}\sigma_{n,1}\right). \tag{25}$$

Then

$$\frac{1}{\sigma_n}\left(U_n(h_n) - \theta_n\binom{n}{2}\right) - \frac{1}{\sigma_{n,1}\sqrt{n}}\sum_{\ell=1}^n \bar{h}_{n,1}(X_\ell) \to 0, \quad \textit{in probab.}$$

Proof. See Section 5.5. □

Remark 6. From the above result, it is clear that if $\frac{1}{\sigma_{n,1}\sqrt{n}}\sum_{\ell=1}^n \bar{h}_{n,1}(X_\ell)$ converges weakly to $\mathcal{N}(0,1)$, then so does $\frac{1}{\sigma_n}\left(U_n(h_n) - \theta_n\binom{n}{2}\right)$. If h_n does not depend on n, (25) automatically holds and Theorem 4.1 implies a standard CLT for U-statistics (Yoshihara (1976); Dehling and Wendler (2010)). But unlike these previous works, Theorem 4.1 does not assume stationarity and for Markov chains, the weak dependence assumption A1 is sometimes easier to check than mixing assumptions.

The theorem describes the limiting behavior of $U_n(h_n)$ in the case where the kernels h_n are not degenerate and the quadratic term $\sum_{\ell=1}^n\sum_{j=1}^{\ell-1} Q_{n,\ell,j}$ is negligible. In general, the quadratic term needs not be negligible. In which case a correct account of the limiting behavior of $U_n(h_n)$ will then require a joint study the processes $\sum_{\ell=1}^n \bar{h}_{n,1}(X_\ell)$ and $\sum_{\ell=1}^n\sum_{j=1}^{\ell-1} Q_{n,\ell,j}$.

5. Proofs

5.1. **Proof Lemma 2.1.** That $\bar{G}_2 \in \mathcal{L}_{\overline{W}_2}$, and $|\bar{G}_2|_{\overline{W}_2} \leq c|\bar{h}_2|_{\overline{V}_2}$ follows from (3). Set

$$\pi_{n,m}(x,y;(du,dv)) = \left(P^n(x,du) - \mu(du)\right)\bigotimes\left(P^m(y,dv) - \mu(dv)\right).$$

Since $P^s\overline{W}_2(x,y) < \infty$ for all $x,y \in \mathcal{X}$ and $s \in \{1,2\}$ by A1, we deduce that the rhs of (4) is well-defined and can be written as $\bar{G}_2(x,y) - P\bar{G}_2(y,x) - P\bar{G}_2(x,y) + P^2\bar{G}_2(x,y)$.

By dominated convergence,

$$
\begin{aligned}
&\bar{G}_2(x,y) - P\bar{G}_2(y,x) - P\bar{G}_2(x,y) + P^2\bar{G}_2(x,y) \\
&\quad = \lim_{N,M\to\infty} \sum_{n=0}^{N}\sum_{m=0}^{M} \int (\delta_x(dz_1) - P(x,dz_1)) \int (\delta_y(dz_2) - P(y,dz_2))\, \pi_{n,m}\bar{h}_2(z_1,z_2), \\
&\quad = \lim_{N,M\to\infty} \left\{ \bar{h}_2(x,y) - \pi_{N+1,0}\bar{h}_2(x,y) - \pi_{0,M+1}\bar{h}_2(x,y) + \pi_{N+1,M+1}\bar{h}_2(x,y) \right\} \\
&\qquad = \bar{h}_2(x,y),
\end{aligned}
$$

proving (4). The bound $|P^s\bar{G}_2|_{\overline{W}_2} \le c|P^s\bar{h}_2|_{\overline{V}_1}$ is obtained by showing in a similar way that

$$
\begin{aligned}
P^s\bar{G}_2(x,y) &= \lim_{N,M\to\infty} \sum_{n=0}^{N}\sum_{m=0}^{M} \pi_{n,m}\{P^s\bar{h}_2\}(x,y) \\
&= \int\int \bar{R}_2(x,y;dz_1,dz_2)\{P^s\bar{h}_2\}(z_1,z_2).
\end{aligned}
$$

□

5.2. **Proof Lemma 2.2.** From the definition, we have $h_n(x,y) = \theta_n + \bar{h}_{n,1}(x) + \bar{h}_{n,1}(y) + \bar{h}_{n,2}(x,y)$, and we deduce after some rearrangements that

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

Using (4), we write

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

Rearranging the terms, it is easy to verify that

$$
\begin{aligned}
\sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell,j)\bar{h}_{n,2}(X_\ell,X_j) &= \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell,j)Q_{n,\ell,j} \\
&+ \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} \varpi_n^{(1)}(\ell,j)\left(P\bar{G}_{n,2}(X_{\ell-1},X_j) - P^2\bar{G}_{n,2}(X_{j-1},X_{\ell-1})\right) \\
&+ \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} \varpi_n^{(2)}(\ell,j)\left(P\bar{G}_{n,2}(X_{j-1},X_\ell) - P^2\bar{G}_{n,2}(X_{j-1},X_{\ell-1})\right) \\
&+ \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} \varpi_n^{(3)}(\ell,j)P^2\bar{G}_{n,2}(X_{j-1},X_{\ell-1}) + \epsilon_n,
\end{aligned}
$$

where ϵ_n is comprised of the remainder telescoping sums. We obtain

$$\epsilon_n = \sum_{\ell=1}^{n} \left\{ w_n(\ell,0) P\bar{G}_{n,2}(X_0, X_\ell) - w_n(\ell-1,0) P^2\bar{G}_{n,2}(X_0, X_{\ell-1}) \right) \\ + \sum_{j=1}^{n} w_n(n,j) \left(P^2\bar{G}_{n,2}(X_n, X_j) - P\bar{G}_{n,2}(X_n, X_j) \right) \\ + \sum_{\ell=1}^{n} \left(w_n(\ell-1,\ell) P\bar{G}_{n,2}(X_{\ell-1}, X_\ell) - w_n(\ell,\ell) P\bar{G}_{n,2}(X_\ell, X_\ell) \right).$$

□

5.3. **Proof theorem 3.1.** A2 implies that we can find $p > 1$ such that

$$\sup_{n\geq 0} \mathbb{E}\left(V_2^{2p}(X_n) \right) < \infty. \tag{26}$$

From (10), $\Gamma^2_{n,b}(h) = \sum_{\ell=1}^{n} \sum_{j=1}^{\ell} w_{n,b}(\ell-j) h(X_j) h(X_\ell) + R_n$, where $w_{n,b}(0) = n^{-1}$ and $w_{n,b}(\ell) = 2n^{-1} w_b(k c_n^{-1})$ (in particular $w_{n,b}(\ell) = 0$ for $\ell < 0$). Set $\Delta^{(1)}_{n,b}(\ell) = w_{n,b}(\ell) - w_{n,b}(\ell-1)$ and $\Delta^{(2)}_{n,b}(\ell) = 2w_{n,b}(\ell) - w_{n,b}(\ell+1) - w_{n,b}(\ell-1)$. Then Lemma 2.2 applied to $\sum_{\ell=1}^{n} \sum_{j=1}^{\ell} w_{n,b}(\ell-j) h(X_j) h(X_\ell)$ gives:

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

where

$$\zeta_n = \sum_{\ell=1}^{n} PG(X_{\ell-1}) \sum_{j=1}^{\ell} \Delta^{(1)}_{n,b}(\ell-j) Q_j + \sum_{\ell=1}^{n} Q_\ell \sum_{j=1}^{\ell} \Delta^{(1)}_{n,b}(\ell-j+1) PG(X_{j-1}) \\ + \sum_{\ell=3}^{n} PG(X_{\ell-1}) \sum_{j=1}^{\ell-2} \Delta^{(2)}_{n,b}(\ell-j) PG(X_{j-1}) \\ + PG(X_0) \left\{ \sum_{\ell=1}^{n} w_{n,b}(\ell) Q_\ell + \sum_{\ell=1}^{n} \Delta^{(1)}_{n,b}(\ell) PG(X_{\ell-1}) \right\} \\ - PG(X_n) \left\{ \sum_{j=1}^{n} w_{n,b}(n-j) Q_j - \sum_{j=1}^{n} \Delta^{(1)}_{n,b}(\ell-j) PG(X_j) \right\} \\ + \Delta^{(2)}_{n,b}(0) \sum_{\ell=1}^{n} (PG(X_{\ell-1}))^2 - \left(w_{n,b}(0) - \Delta^{(2)}_{n,b}(1) \right) \sum_{\ell=1}^{n} PG(X_\ell) Q_\ell \\ - \Delta^{(2)}_{n,b}(1) \left(PG(X_n) \right)^2 - w_{n,b}(n-1) PG(X_n) PG(X_0).$$

Using A2, (26), the martingale-difference property of $\{Q_\ell,\ \ell \geq 1\}$, the smoothness of w_b, we derive that for $p > 1$ as in (26),

$$\mathbb{E}^{1/p}\left(|\zeta_n|^p \right) \leq c\, c_n^{-1+\frac{1}{2}\vee\frac{1}{p}}, \quad n \geq 3, \tag{27}$$

for some finite constant c.

By martingale approximation for linear partial sums (see e.g. the proof of Proposition 6.1 below), for any sequence of real numbers $\{a_{n,\ell},\ 1 \le \ell \le n\}$,

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

$$\text{where} \quad \mathbb{E}\left(\left|\sum_{\ell=1}^{n} a_{n,\ell} Q_\ell\right|^{\alpha}\right) \le c \left(\sum_{\ell=1}^{n} |a_{n,\ell}|^{\alpha\wedge 2}\right)^{1\vee\frac{\alpha}{2}},$$

$$\text{and} \quad \mathbb{E}\left(|\epsilon_{n,1}|^{\alpha}\right) \le c \left(|a_{n,1}| + |a_{n,n}| + \sum_{\ell=2}^{n} |a_{n,\ell} - a_{n,\ell-1}|\right)^{\alpha}, \quad (28)$$

provided $\sup_{n\ge 0} \mathbb{E}\left(V_2^{\alpha}(X_n)\right) < \infty$. We use (28) to bound the term R_n as given in (11) and obtain for all $n \ge 1$:

$$\mathbb{E}\left(|R_n|^p\right) \le c n^{-p} c_n^p,$$

for some finite constant c.

By standard martingale inequalities, we obtain the bound

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

□

5.4. **Proof theorem 3.2.** If $c_n = o(n)$ and $p \ge 2$, then from Theorem 3.1, $\Gamma^2_{n,b}(h) = n^{-1}\sum_{\ell=1}^{n} Q_\ell^2 + o_P(1)$. Given the ergodicity assumption $\mathsf{C}(1, V_2^2, V_3)$ and $\sup_{k\ge 0} \mathbb{E}(V_3^q(X_k)) < \infty$, it follows from Proposition 6.1 that the term $n^{-1}\sum_{\ell=1}^{n} Q_\ell^2$ converges in probability to the limit

$$\int \mu(dx) \int P(x, dy) \left(G(y) - PG(x)\right)^2,$$

which is easily seen to be equal to $\sigma^2(h)$. This proves the first part of the theorem.

From now on, we assume that $c_n = n$. Define $W_{n,\ell} = \frac{Q_\ell}{\sqrt{n}\sigma(h)}$. Then by (28) with $a_{n,\ell} \equiv \frac{1}{\sqrt{n}\sigma(h)}$,

$$\sum_{\ell=1}^{n} \frac{h(X_\ell)}{\sqrt{n}\sigma(h)} = \sum_{\ell=1}^{n} W_{n,\ell} + \epsilon_{n,2}, \quad \text{where} \quad \epsilon_{n,2} \text{ converges in probability to zero.}$$

Define $\lfloor x \rfloor$ as the largest integer smaller or equal to x and for $0 \le t \le 1$, we introduce

$$B_n(t) = \sum_{\ell=1}^{\lfloor nt \rfloor} W_{n,\ell}, \quad \text{and} \quad Z_n(t) = \int_0^t w_b(t-u) dB_n(u).$$

Since B_n has jumps only at times $\ell/n = \ell/c_n$, we see that $Z_n(\ell c_n^{-1}) = \sum_{j=0}^{\ell-1} w_b((\ell - j)c_n^{-1}) W_{n,j+1}$. It is also easy to see that the term R_n in (11) can be written as

$$R_n = 2B_n^2(1)\int_0^1 (1-u) w_b(u) du$$
$$- 2B_n(1)\int_0^1 \left(\int_0^t w_b(u)du + \int_0^{1-t} w_b(u)du\right) dB_n(t) + \epsilon_{n,3},$$

where $\epsilon_{n,3}$ converges in probability to zero. Thus

$$\Gamma^2_{n,b}(h) = n^{-1}\sum_{\ell=1}^n Q_\ell^2 + 2\sigma^2(h)\sum_{\ell=1}^n \frac{Q_\ell}{\sigma(h)\sqrt{n}}\sum_{j=1}^{\ell-1} w_b\left(\frac{\ell-j}{c_n}\right)\frac{Q_j}{\sigma(h)\sqrt{n}} + R_n + \zeta_n$$
$$= \sigma^2(h)\sum_{\ell=1}^n W_{n,\ell}^2 + 2\sigma^2(h)\sum_{\ell=1}^n W_{n,\ell} Z_n\left((\ell-1)c_n^{-1}\right) + R_n + \zeta_n$$
$$= \sigma^2(h)\sum_{\ell=1}^n W_{n,\ell}^2 + 2\sigma^2(h)\int_0^1 Z_n(t)dB_n(t) + 2B_n^2(1)\sigma^2(h)\int_0^1 w_b(u)(1-u)du$$
$$- 2B_n(1)\sigma^2(h)\int_0^1 g_b(u)dB_n(t) + \epsilon_{n,4},$$

where $g_b(u) = \int_0^t w_b(u)du + \int_0^{1-t} w_b(u)du$, and $\epsilon_{n,4}$ converges in probability to zero.

From the assumptions, $\sup_{\ell\geq 0}\mathbb{E}\left(|Q_\ell|^{2+\epsilon}\right) < \infty$, for some $\epsilon > 0$. Therefore, by the functional central limit theorem for martingales, $B_n \xrightarrow{w} B$, where $B = \{B(t),\ 0\leq t\leq 1\}$ is the standard Brownian motion. By the continuous mapping theorem, $(B_n, Z_n) \xrightarrow{w} (B, Z)$, where $Z(t) = \int_0^t w(t-u)dB(u)$. And by the weak convergence of stochastic integrals (see, e.g., Theorem 2.2 in Kurtz and Protter (1991)),

$$\left\{\left(B_n(t), \int_0^t Z_n(t)dB_n(t), \int_0^1 g(u)dB_n(u), \sum_{\ell=1}^n W_{n,\ell}^2\right), 0\leq t\leq 1\right\}$$

converges weakly to the stochastic process

$$\left\{\left(B(t), \int_0^t Z(u)dB(u), \int_0^1 g(u)dB(u), 1\right),\ 0\leq t\leq 1\right\}.$$

As the remainders $(\epsilon_{n,2}, \epsilon_{n,4})$ converges in probability to 0, this entails that $\left(\sum_{\ell=1}^n \frac{h(X_\ell)}{\sqrt{n}\sigma(h)}, \Gamma^2_{n,b}(h)\right)$ converges weakly to the limit

$$(B(1),$$
$$\sigma^2(h)\left(1 + 2\int_0^1 Z(u)dB(u) + 2B^2(1)\int_0^1 (1-u)w(u)du - 2B(1)\int_0^1 g(u)dB(t)\right)\Bigg).$$

The conclusion of the theorem follows by the continuous mapping theorem.

□

5.5. **Proof Theorem 4.1.** By (6), we have:

$$\sigma_n^{-1}\left[U_n(h_n) - \theta_n\binom{n}{2}\right] = \sigma_n^{-1}(n-1)\sum_{\ell=1}^n h_n(X_\ell) + \sigma_n^{-1}\sum_{\ell=2}^n\sum_{j=1}^{\ell-1} Q_{n,\ell,j} + \sigma_n^{-1}\zeta_n.$$

(21) gives

$$\sigma_n^{-2}\mathbb{E}\left(|\zeta_n|^2\right) \leq cn^{-1}\sigma_{n,1}^{-2}\left|\!\left|\!\left|\bar{h}_{n,2}\right|\!\right|\!\right|^2_{2,\overline{V}_2} = o(1)$$

by (25). This shows that $\sigma_n^{-1}\zeta_n$ converges in probability to zero.

Now, by the martingale property, we have

$$\mathbb{E}\left[\left(\sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1}Q_{n,\ell,j}\right)^2\right]=\sum_{\ell=2}^{n}\mathbb{E}\left[\left(\sum_{j=1}^{\ell-1}Q_{n,\ell,j}\right)^2\right]=\sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1}\mathbb{E}\left(Q_{n,\ell,j}^2\right)$$
$$+2\sum_{\ell=2}^{n}\sum_{k=1}^{\ell-2}\sum_{j=k+1}^{\ell-1}\mathbb{E}\left(Q_{n,\ell,j}Q_{n,\ell,k}\right)=o(\sigma_n^2),$$

by (23), Lemma 5.1 and (25).

□

Lemma 5.1. *Under the assumptions of Theorem 4.1,*

$$\left|\mathbb{E}\left(\sum_{k=1}^{\ell-2}\sum_{j=k+1}^{\ell-1}Q_{n,\ell,j}Q_{n,\ell,k}\right)\right|\le cn\left\|\bar{h}_{n,2}\right\|_{2,\overline{V}_2}^2,\quad 3\le\ell\le n.$$

Proof. Fix $1\le k<\ell$ and define

$$T_k=T_{n,\ell,k}:=\mathbb{E}\left(\sum_{j=k+1}^{\ell-1}Q_{n,\ell,j}Q_{n,\ell,k}|\mathcal{F}_{k-1}\right),$$

so that

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

For $m\ge 0$, define

$$\Upsilon_{2,m}(x_{j-1},x_{k-1},x_k)=\int P(x_{j-1},dx_j)\int P^m(x_j,dx_{\ell-1})$$
$$\times\int P(x_{\ell-1},dx_\ell)\Lambda_{n,2}(x_{j-1},x_{\ell-1};x_j,x_\ell)\Lambda_{n,2}(x_{k-1},x_{\ell-1};x_k,x_\ell),$$

$$\Upsilon_{1,m}(x_{k-1},x_k):=\int\left\{P^m(x_k,dx_{j-1})-\mu(dx_{j-1})\right\}\Upsilon_{2,\ell-m-k-2}(x_{j-1},x_k,x_{k-1}).$$

Then almost surely we have:

$$T_k=\mathbb{E}\left(\sum_{j=0}^{\ell-k-1}\Upsilon_{1,j}(X_k,X_{k-1})|\mathcal{F}_{k-1}\right).$$

The bound (22) and the Cauchy-Schwartz inequality imply that

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

We combine this with B2 to conclude that for all $m\ge 0$,

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

By the short-range dependence assumption $\mathsf{C}(1,\overline{\mathcal{U}}_1,\overline{\mathcal{V}}_1)$, it follows that for any $n \geq 0$,

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

for some finite constant c. We conclude that

$$|T_k| \leq c \left\|\bar{h}_{n,2}\right\|_{2,\overline{V}_2}^2 \mathbb{E}\left(\overline{\mathcal{V}}_1(X_k)\overline{\mathcal{U}}_2(X_k, X_{k-1})|\mathcal{F}_{k-1}\right).$$

The lemma follows. □

6. Appendix A: A weak law of large numbers for Markov chains

Proposition 6.1. *Let $\{X_n,\ n \geq 0\}$ be a Markov chain with invariant distribution μ and transition kernel P. Suppose that there exist measurable functions $V_1 \leq V_2 :\ \mathcal{X} \to [1, \infty)$ such that*

$$\sum_{k\geq 0} \|P^k(x,\cdot) - \mu\|_{V_1} \leq cV_2(x),\ x \in \mathcal{X}, \tag{29}$$

for some finite constant c. Suppose also that $\mathsf{v}_n := \mathbb{E}(V_2^p(X_n)) < \infty$ for each $n \geq 0$ and for some $p \in (1, 2]$. 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 such that

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

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

Then, as $n \to \infty$, $\left(\sum_{k=1}^n |a_{n,k}|\right)^{-1} \sum_{k=1}^n a_{n,k}\left(f_n(X_k) - \mu(f_n)\right)$ converges in probability to zero.

Proof. Define $S_n = \sum_{k=1}^n a_{n,k}\left(f_n(X_k) - \mu(f_n)\right)$ and $g_n(x) = \sum_{j\geq 0}(P^j f_n(x) - \mu(f_n))$. Under (29), $|g_n(x)| \leq c|f_n|_{V_1} V_2(x)$ and $|Pg_n(x)| \leq c|Pf_n|_{V_1} V_2(x)$. By the Poisson equation, $f_n(x) - \mu(f_n) = g_n(x) - Pg_n(x)$ which implies that

$$\begin{aligned}S_n = \sum_{k=1}^n a_{n,k}\left(g_n(X_k) - Pg_n(X_{k-1})\right) + \sum_{k=1}^n (a_{n,k} - a_{n,k-1})Pg_n(X_{k-1}) \\ + \left(a_{n,0}Pg_n(X_0) - a_{n,n}Pg_n(X_n)\right).\end{aligned}$$

where the martingale array $\sum_{k=1}^n a_{n,k}\left(g_n(X_k) - Pg_n(X_{k-1})\right)$ satisfies

$$\mathbb{E}\left(|\sum_{k=1}^n a_{n,k}\left(g_n(X_k) - Pg_n(X_{k-1})\right)|^p\right) \leq C\, \|f_n\|_{p,V_1}^p \sum_{k=1}^n |a_{n,k}|^p \mathsf{v}_k^p.$$

The last inequality follows by noting that $g_n(x)-Pg_n(y) = f_n(x)-\mu(f_n)-Pg_n(x)-Pg_n(y)$ and by conditioning on $\mathcal{F}_{k-1}$. Thus, under the stated assumptions, $\left(\sum_{k=1}^n |a_{n,k}|\right)^{-1} S_n$ converges in probability to zero. $\square$

Remark 7. An important special case is the case where $a_{n,\ell} = 1$ and $\sup_{n\geq 0} \mathbb{E}\left(V_2^p(X_n)\right) < \infty$. In this case it is enough to have $n^{-1+1/p} \left\|\!\left| f_n \right|\!\right\|_{p,V_1} \to 0$. If in addition it is true that $\sup_{x\in\mathcal{X}} PV_1^p(x)/V_1^p(x) < \infty$, then clearly $\left\|\!\left| f_n \right|\!\right\|_{p,V_1} \leq c|f_n|_{V_1}$ and the law of large number holds if $n^{-1+1/p}|f_n|_{V_1} \to 0$.

References

Atchade, Y. (2011). Kernel estimators of asymptotic variance for adaptive Markov Chain Monte Carlo. *Annals of Statistics* **39** 990–1011.

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

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

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

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

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

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian data analysis*. 2nd ed. Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, FL.

Gordin, M. I. (1969). The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR* **188** 739–741.

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

Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* **21** 1130–1164.

Kurtz, T. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* **19** 1035–1070.

Maxwell, M. and Woodroofe, M. (2000). Central limit theorems for additive functional of markov chains. *Annals of Probability* **28** 713–724.

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

Merlevede, F., Peligrad, M. and Utev, S. (2006). Recent advances in invariances principles for stationary sequences. *Probability surveys* **3** 1–36.

MEYN, S. P. and TWEEDIE, R. L. (2009). *Markov chains and stochastic stability.* Cambridge University Press; 2 edition, London.

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

PRIESTLEY, M. B. (1981). *Spectral analysis and time series. Vol. 1.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London. Univariate series, Probability and Mathematical Statistics.

ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71 (electronic).

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

WU, W. B. and SHAO, X. (2007). A limit theorem for quadratic forms and its applications. *Econometric Theory* **23** 930–951.

YOSHIHARA, K.-I. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35** 237–252.