Weak Convergence of Stationary Empirical Processes

Dragan Radulović

Department of Mathematics, Florida Atlantic University

Marten Wegkamp

Department of Mathematics & Department of Statistical Science, Cornell University

September 12, 2017

Abstract

We offer an umbrella type result which extends weak convergence of classical empirical processes on the line to that of more general processes indexed by functions of bounded variation. This extension is not contingent on the type of dependence of the underlying sequence of random variables. As a consequence we establish weak convergence for stationary empirical processes indexed by general classes of functions under α-mixing conditions.

Running title: Weak convergence of stationary empirical processes
MSC2000 Subject classification: Primary 60F17; secondary 60G99.
Keywords and phrases: Integration by parts, bounded variation, empirical processes, stationary distributions, weak convergence.

1 Introduction

We consider the empirical process

$$\bar{\mathbb{Z}}_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{g(X_i) - \mathbb{E}g(X_i)\}, \ g \in \mathcal{G}, \tag{1}$$

indexed by some class $\mathcal{G}$ of functions $g : \mathbb{R} \to \mathbb{R}$. It is an obvious generalization of the classical process

$$\mathbb{G}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ 1_{(-\infty,t]}(X_i) - F(t) \right\}, \tag{2}$$

in which case the indexing class is $\mathcal{G} = \left\{ g(x) = 1_{(-\infty,t]}(x) : \ t \in \mathbb{R} \right\}$. If the underlying sequence $\{X_i\}_{i\geq 1}$ is i.i.d., then the limiting behavior (as $n \to \infty$) of the empirical processes $\{\bar{\mathbb{Z}}_n(g), \ g \in \mathcal{G}\}$ is well understood. The same is true for the bootstrap counterpart $\{\bar{\mathbb{Z}}_n^*(g), \ g \in \mathcal{G}\}$ based on an i.i.d. bootstrap sample $\{X_{1,n}^*, \ldots, X_{m_n,n}^*\}$, see, for instance, Van der Vaart and Wellner (1996) and Dudley (1999). The theory of weak convergence of empirical processes based on independent sequences has yielded a wealth of statistical applications and, in particular, it was instrumental for establishing the weak convergence of numerous novel statistics. Often the limiting distributions of these statistics do not allow for closed form solutions, in which case the bootstrap version of the process is utilized. As is standard practice nowadays, see the thorough exposition in the first chapter of Van der Vaart and Wellner (1996), weak convergence in this paper should be understood in the sense of Hoffmann-Jørgensen and expectations should be interpreted as outer expectations.

For empirical processes based on stationary sequences $\{X_i\}_{i\geq 1}$ the situation is rather different. The classical process $\{\mathbb{G}_n(t), \, t \in \mathbb{R}\}$ has been treated by numerous authors, who established weak convergence under sharp mixing assumptions, see, for instance, Rio (2000). However, nothing similar exists for more general processes. The only work that we could find in the literature, Andrews and Pollard (1994), treats more general indexing classes, but imposes rather restrictive assumptions on the decay of α-mixing coefficients. This discrepancy between the conditions needed for $\bar{\mathbb{Z}}_n(g)$ and those for $\mathbb{G}_n(t)$, is due to fact that the typical approach for proving the uniform limiting theorems heavily relies on the estimation of entropy numbers, which in turn require good exponential maximal inequalities. Only β-mixing, via decoupling, allows for such an estimate. This is the reason why one can find in the literature the treatment of $\bar{\mathbb{Z}}_n(g)$ only for β-mixing sequences, see Arcones and Yu (1994) and Doukhan, Massart and Rio (1995).

The situation for bootstrap of stationary empirical processes is even worse. Although introduced more than twenty years ago, we only found three papers that study the bootstrap for stationary empirical processes indexed by general classes $\mathcal{G}$, see Künsch (1989), Liu and Singh (1992) and Sengupta, Volgushev and Shao (2016). All these results operate under β-mixing assumptions. Moreover, only in the case of VC-classes do we have the sharp conditions, see Radulović (1996). Bracketing classes were considered by Bühlmann (1995), but this was established under restrictive assumptions on both β-mixing coefficients and bracketing numbers. It is worth mentioning that the covariance function of the limiting Gaussian process is unknown, and consequently in most actual applications of these results, we heavily rely on the bootstrap version of the process for which adequate results are sorely lacking. In short, the most general α-mixing sequences have never been considered for stationary bootstrap processes $\{\bar{\mathbb{Z}}_n^*(g),\ g \in \mathcal{G}\}$, while for the non-bootstrap version $\{\bar{\mathbb{Z}}_n(g),\ g \in \mathcal{G}\}$ we have only one example in the literature.

In what follows we prove two general results that allow us to extend weak convergence of $\{\mathbb{G}_n(t),\ t \in \mathbb{R}\}$ to that of $\{\bar{\mathbb{Z}}_n(g), g \in \mathcal{G}\}$, where $\mathcal{G}$ is a class of functions of uniformly bounded total variation (called BV_T in this paper). We would like to point out that although there are examples of Donsker classes of infinite variation (for instance, $f : [0,1] \to [0,1]$ with $|f(x) - f(y)| \leq |x - y|^\alpha$, $1/2 < \alpha < 1$), such cases are rather the exception than the norm. The majority of examples of bounded Donsker classes that are given in the literature are subsets of the class BV_T.

This enlargement from $\{\mathbb{G}_n(t), t \in \mathbb{R}\}$ to $\{\bar{\mathbb{Z}}_n(g),\ g \in \mathcal{G}\}$ is not contingent on the dependence structure of underlying sequence $\{X_i\}_{i\geq 1}$ (or $\{X_i^*\}_{i\geq 1}$) and the only requirement is that $\mathbb{G}_n(t)$ converges weakly to a Gaussian process. This allows us to derive weak convergence of $\{\bar{\mathbb{Z}}_n(g),\ g \in \mathcal{G}\}$ and $\{\bar{\mathbb{Z}}_n^*(g),\ g \in \mathcal{G}\}$ for α-mixing sequences. The same extension applies for the short memory causal processes considered in Doukhan and Surgailis (1998), as well as processes treated in Dehling, Durieu and Volny (2009). The technique we employ is a simple application of the integration-by-parts formula, followed by the continuous mapping theorem. Arguably this approach may have been known before, although we could not find any communication of it in the literature, and certainly not among the research related to stationary empirical processes. A follow-up paper (Radulović, Wegkamp and Zhao (2017)) extends the results obtained in this work to classes indexed by *multivariate* functions of bounded variation, with an emphasis on empirical copula processes. An important technical difference with the follow-up paper is that here we allow for general stationary distribution functions of the underlying $\{X_i\}_{i\geq 1}$. The proof for general, non-continuous processes, is not trivial caused by technical complications related to the interplay between the atoms of the limiting process $\mathbb{Z}(t)$ and discontinuities of $g \in \mathcal{G}$.

The paper is organized as follows. Section 2 contains the statement of the main result (Theorem 1) and related discussions, while Section 3 presents the bootstrap version (Theorem 6) of this result. The proofs of these two main results are collected in Section 4. For completeness, the well-known integration by parts formula can be found in the appendix.

2 Main results

For the total variation norm of a function $g : \mathbb{R} \to \mathbb{R}$, we use the notation

$$||g||_{TV} = \sup_{\Pi} \sum_{x_i, x_{i+1} \in \Pi} |g(x_{i+1}) - g(x_i)|.$$

The supremum is taken over all countable partitions $\Pi = \{x_1 < x_2 < \ldots\}$ of $\mathbb{R}$. We set, for $T > 0$,

$$BV_T := \{g : \mathbb{R} \to \mathbb{R} :\ \|g\|_{TV} \leq T,\ \|g\|_\infty \leq T\}.$$

We let $BV_T' \subset BV_T$ be the class of all functions g in BV_T that are right-continuous. Finally, we let $\{\mathbb{Z}_n(t),\ t \in \mathbb{R}\}$ be an arbitrary stochastic process such that

A1: $\lim_{|t|\to\infty} \mathbb{Z}_n(t) = 0$;

A2: *The sample paths of $\mathbb{Z}_n$ are right-continuous and of bounded variation.*

Clearly, both requirements A1 and A2 are met for the canonical empirical process $\{\mathbb{G}_n(t),\ t \in \mathbb{R}\}$. In this paper, we study the limit distribution (as $n \to \infty$) of the sequence of processes

$$\left\{ \overline{\mathbb{Z}}_n(g) := \int g(x)\, \mathrm{d}\mathbb{Z}_n(x), \quad g \in \mathcal{G} \right\}$$

for some class $\mathcal{G} \subseteq BV_T'$, for some finite T.

Although the motivation as well as the most notable applications of our results are related to the canonical case $\mathbb{Z}_n(x) = \mathbb{G}_n(x)$, the actual proof carries over for any process $\mathbb{Z}_n$ as long as assumptions A1 and A2 are satisfied.
Our main result is the following theorem.

Theorem 1 *Assume that $\{\mathbb{Z}_n(t),\ t \in \mathbb{R}\}$ converges weakly, as $n \to \infty$, to a Gaussian process $\{\mathbb{Z}(t),\ t \in \mathbb{R}\}$, that has uniformly continuous sample paths with respect to the distance $d(s,t) = |F_0(s) - F_0(t)|$ for some distribution function F_0.*
Then, for any $T < \infty$ and $\mathcal{G} \subseteq BV_T'$, $\{\bar{\mathbb{Z}}_n(g),\ g \in \mathcal{G}\}$ converges weakly to a Gaussian process in $\ell^\infty(\mathcal{G})$, that has uniformly $L_1(F_0)$-continuous sample paths.

Remark. Usually, an envelope condition $\sup |g(x)| \leq G(x)$ and $G \in L_2(F_0)$ is imposed. Since such a condition, coupled with $\|g\|_{TV} \leq T$, implies that the functions g are uniformly bounded, we assume $\|g\|_\infty \leq T$ in our definition of BV_T.

Theorem 1 allows us to derive weak convergence of $\bar{\mathbb{Z}}_n(g)$ via $\mathbb{Z}_n(t)$, regardless of the structure of the latter process. For instance, taking $\mathbb{Z}_n$ as the standard empirical processes $\mathbb{G}_n$ based on stationary sequences $\{X_i\}_{i\geq 1}$, we obtain the following corollary as an immediate consequence of Theorem 1.

Corollary 2 *Let $\{X_k\}_{k\geq 1}$ be a stationary sequence of random variables with distribution F and α-mixing coefficients α_n satisfying $\alpha_n = O(n^{-r})$, for some $r > 1$ and all $n \geq 1$. Then, for any $\mathcal{G} \subseteq BV_T'$,*

- *$\{\int g \, \mathrm{d}\mathbb{G}_n, \, g \in \mathcal{G}\}$ converges weakly to a Gaussian process with uniformly $L_1(F)$-continuous sample paths in $\ell^\infty(\mathcal{G})$.*

- *For continuous F, $\mathcal{G}$ can be enlarged to BV_T.*

Proof. It is well known, see Theorem 7.2, page 96 in Rio (2000), that $\alpha_n = O(n^{-r})$ with $r > 1$, implies that the standard empirical process $\mathbb{G}_n(t)$ converges weakly to a Brownian bridge process with uniformly continuous paths with respect to the distance $d(s,t) = |F(s) - F(t)|$, for the stationary distribution F of X_k. The result for $\mathcal{G} \subseteq BV_T'$ now follows trivially from Theorem 1. If F is continuous, then the limiting Brownian bridge has uniformly continuous sample paths with respect to the Lebesgue measure on $\mathbb{R}$. This, combined with Lemma 3 below, implies Corollary 2. ■

We used the following lemma.

Lemma 3 *We have, for $\bar{\mathbb{Z}}_n(g)$ based on the canonical process $\mathbb{Z}_n(x) = \mathbb{G}_n(x)$,*

$$\sup_{g \in BV_T} \inf_{h \in BV_T'} \left| \bar{\mathbb{Z}}_n(g) - \bar{\mathbb{Z}}_n(h) \right| \leq T \sup_x |\mathbb{G}_n(x) - \mathbb{G}_n(x^-)|.$$

Proof. Let g be an arbitrary function in BV_T. We denote its countable many discontinuities by a_i. Let $\overline{g}$ be the right-continuous version of g, that is, $\overline{g}(x) = g(x)$ for all $x \neq a_i$ and $\overline{g}(a_i) = g(a_i^+)$ for all i. Then

$$\begin{aligned} \left| \int g \, \mathrm{d}\mathbb{G}_n - \int \overline{g} \, \mathrm{d}\mathbb{G}_n \right| &\leq \sum_i |g(a_i) - \overline{g}(a_i)| |\mathbb{G}_n(a_i) - \mathbb{G}_n(a_i^-)| \\ &\leq \|g\|_{TV} \sup_x |\mathbb{G}_n(x) - \mathbb{G}_n(x^-)| \end{aligned}$$

and the conclusion follows. ■

We would like to point out that α-mixing is the least restrictive form of available mixing assumptions in the literature. To the best of our knowledge, there are actually very few results that treat processes $\{\int g \, \mathrm{d}\mathbb{G}_n, \ g \in \mathcal{G}\}$ indexed by functions, and they all require very stringent conditions on the entropy numbers of $\mathcal{G}$ and on the rate of decay for α_k. See, for instance, Andrews and Pollard (1994). This is due to fact that α-mixing does not allow for sharp exponential inequalities for partial sums. Consequently, the only known cases for which we have sharp conditions are under more restrictive, β-mixing dependence. Indeed, β-mixing allows for decoupling and it does yield exponential inequalities not unlike the i.i.d. case. The current state-of-the art results, Arcones and Yu (1994), Doukhan, Massart and Rio (1995), applied to bounded sequences, require $\sum_n \beta_n < \infty$.

While it is correct that α-mixing is the weakest of the known mixing concepts, a referee pointed out that functionals of mixing processes constitutes a much weaker concept of short range dependence. For instance, Billingsley (1968, Theorem 22.2) establishes the empirical process CLT for certain functionals of ϕ-mixing processes. The same referee made us aware that Theorem 1 also applies to the empirical process of long-range dependent data, established by Dehling and Taqqu (1989) for subordinate Gaussian processes, and by Ho and Hsing (1996) for linear processes. We provide an example to demonstrate that Theorem 1 goes beyond dependence defined via mixing conditions. It applies to short memory causal linear sequences $\{X_i\}_{i\geq 1}$ that are defined by $X_i = \sum_{j=0}^{\infty} a_j \xi_{i-j}$ based on i.i.d. random variables ξ_i and constants a_i. While the X_i form a stationary sequence, they do not necessarily satisfy any mixing condition. Weak convergence of the empirical processes $\{\mathbb{G}_n(t), \ t \in \mathbb{R}\}$ was established under sharp conditions in Doukhan and Surgailis (1998) on the weights a_j ($\sum_{j\geq 0} |a_j|^\gamma < \infty$ for some $0 < \gamma \leq 1$), characteristic function of ξ_0 ($|\mathbb{E}[\exp(iu\xi_0)]| \leq C(1+|u|)^{-p}$ for all $u \in \mathbb{R}$, some $C < \infty$ and $2/3 < p \leq 1$), and a moment condition on ξ_0 ($\mathbb{E}[|\xi_0|^{4\gamma}] < \infty$). To the best of our knowledge, there are no extensions to the more general processes $\{\bar{\mathbb{Z}}_n(g), \ g \in \mathcal{G}\}$. Theorem 1 and the Doukhan and Surgailis (1998) result combined imply the following:

Corollary 4 *Let $\{X_i = \sum_{j\geq 0} a_j \xi_{i-j}\}_{i\geq 1}$ be such that conditions of Doukhan and Surgailis (1998, pp 87–88) are satisfied and let F be the stationary distribution of X_i.*

Then, for any $\mathcal{G} \subseteq BV_T$, $\{\int g \, \mathrm{d}\mathbb{G}_n, \; g \in \mathcal{G}\}$ converges weakly to a Gaussian process with uniformly $L_1(F)$-continuous sample paths in $\ell^\infty(\mathcal{G})$.

Proof. Doukhan and Surgailis (1998) proved that $\{\mathbb{G}_n(t), \; t \in \mathbb{R}\}$ converges weakly to a Gaussian process in the Skorohod space. Since F is continuous under their assumptions, see Doukhan and Surgailis (1998, pp 88), the sample paths of limiting process of $\mathbb{G}_n$ are uniformly continuous with respect to the distance $d(s,t) = |F(s) - F(t)|$ based on the stationary distribution F. The proof now follows trivially from Theorem 1 and Lemma 3. ∎

The recent papers by Dehling, Durieu and Volny (2009) and Dehling, Durieu and Tusche (2014) offer yet another, clever way to prove the weak limit of the standard empirical processes $\mathbb{G}_n$ based on stationary sequences that are not necessarily mixing. Their technique uses finite dimensional convergence coupled with a bound on the higher moments of partial sums, which in turn controls the dependence structure. Dehling, Durieu and Volny (2009) establishes the weak convergence of $\mathbb{G}_n$, while Dehling, Durieu and Tusche (2014) extends this idea to $\bar{\mathbb{Z}}_n$ indexed by more general classes of functions. However, these authors impose cumbersome entropy conditions and only manage to marginally extend the classes. For example, they prove weak convergence of the process $\int f_t \, \mathrm{d}\mathbb{G}_n$, indexed by functions $f_t(x)$ which constitute an one-dimensional monotone class (under the restrictive requirement that $s \leq t \Rightarrow f_s \leq f_t$). Theorem 1 applied in their setting, yields a more general result.

Corollary 5 *Let $\mathcal{G} \subseteq BV_T$ and let F be the stationary distribution of the sequence $\{X_i\}_{i \geq 1}$. Under assumptions (i) and (ii) in Section 1 of Dehling, Durieu and Volny (2009), $\{\int g \, \mathrm{d}\mathbb{G}_n, \; g \in \mathcal{G}\}$ converges weakly to a Gaussian process with uniformly $L_1(F)$-continuous sample paths in $\ell^\infty(\mathcal{G})$.*

Proof. The underlying distribution function F of X_k in Dehling, Durieu and Volny (2009) is continuous, see their display (3) at page 3702. Moreover, these authors establish weak convergence of $\mathbb{G}_n$ to a Gaussian process that has uniformly continuous sample paths with respect to the distance $d(s,t) = |F(s) - F(t)|$. The proof follows trivially from Theorem 1 and Lemma 3. ∎

3 Bootstrap

The weak limit of $\bar{\mathbb{Z}}_n$ based on $\mathbb{G}_n$ in Theorem 1 is a Gaussian process $\bar{\mathbb{Z}}$ with complicated covariance structure

$$\begin{aligned}\mathbb{E}[\bar{\mathbb{Z}}(f)\bar{\mathbb{Z}}(g)] &= \operatorname{Cov}(f(X_0), g(X_0)) + \sum_{k=1}^{\infty} \operatorname{Cov}(f(X_0), g(X_k)) \\ &\quad + \sum_{k=1}^{\infty} \operatorname{Cov}(f(X_k), g(X_0))\end{aligned}$$

for $f, g \in \mathcal{G} \subset BV_T$. A closed form solution is seldom available, so that actual applications of weak limit results of $\bar{\mathbb{Z}}_n$ are hard to implement. This situation calls for the bootstrap principle. Given the sample of first n observations $\{X_1, \ldots, X_n\}$, we let $\mathbb{G}_n^*$ be the bootstrap empirical process

$$\mathbb{G}_n^*(t) = \sqrt{m_n}\left(\frac{1}{m_n}\sum_{i=1}^{m_n} 1_{(-\infty,t]}(X_i^*) - \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,t]}(X_i)\right), \ t \in \mathbb{R},$$

based on a bootstrap sample $\{X_{1,n}^*, \ldots, X_{m_n,n}^*\}$. We stress that no additional assumption on the structure of the variables $X_{i,n}^*$ is required for our purposes. Analogous to $\bar{\mathbb{Z}}_n(g) = \int g \, \mathrm{d}\mathbb{G}_n$, we define $\bar{\mathbb{Z}}_n^*(g) := \int g \, \mathrm{d}\mathbb{G}_n^*$ for any $g \in BV_T'$ with $T < \infty$. We recall that the bounded Lipschitz distance

$$d_{BL}(\mathbb{Z}_n^*, \mathbb{Z}) = \sup_{h \in BL_1} |\mathbb{E}^*[h(\mathbb{Z}_n^*)] - \mathbb{E}[h(\mathbb{Z})]|$$

between two processes $\mathbb{Z}_n^*$ and $\mathbb{Z}$ metrizes weak convergence. The symbol $\mathbb{E}^*$ stands for the expectation over the randomness of the bootstrap sample $X_{1,n}^*, \ldots, X_{m_n,n}^*$, conditionally given the original sample $X_1, \ldots, X_n$, while BL_1 denotes the space of Lipschitz functionals $h : \ell^\infty(\mathcal{G}) \to \mathbb{R}$ with $|h(x)| \leq 1$ and $|h(x) - h(y)| \leq \|x - y\|_\infty$ for all $x, y \in \mathcal{G} \subset BV_T$. As is customary in the literature, we speak of weak convergence in probability if the random variable $d_{BL}(\mathbb{Z}_n^*, \mathbb{Z})$ converges to zero in probability; if it converges to zero almost surely, we speak of weak convergence almost surely.

Theorem 6 *Let $\mathcal{G} \subseteq BV_T'$. Assume that, conditionally on $X_1, \ldots, X_n$, $\{\mathbb{G}_n^*(t), \ t \in \mathbb{R}\}$ converges weakly to a Gaussian process, in probability, that has unifomly continuous* *sample paths with respect to the distance $d(s,t) = |F(s) - F(t)|$ based on the stationary distribution F of X_i. The following three statements hold true:*

1. *$\left\{\int g \, \mathrm{d}\mathbb{G}_n^*, \; g \in \mathcal{G}\right\}$ converges weakly to a Gaussian process with uniformly $L_1(F)$-continuous sample paths in $\ell^\infty(\mathcal{G})$.*

2. *If the weak convergence of $\{\mathbb{G}_n^*(t), \; t \in \mathbb{R}\}$ holds almost surely, then the conclusion that $\left\{\int g \, \mathrm{d}\mathbb{G}_n^*, \; g \in \mathcal{G}\right\}$ converges weakly in $\ell^\infty(\mathcal{G})$ holds almost surely as well.*

3. *If $\{\mathbb{G}_n(t), \; t \in \mathbb{R}\}$ and $\{\mathbb{G}_n^*(t), \; t \in \mathbb{R}\}$ converge to the same limit, then so do $\left\{\int g \, \mathrm{d}\mathbb{G}_n, \; g \in \mathcal{G}\right\}$ and $\left\{\int g \, \mathrm{d}\mathbb{G}_n^*, \; g \in \mathcal{G}\right\}$.*

The literature offers numerous bootstrapping techniques for stationary data, such as moving block bootstrap, stationary bootstrap, sieved bootstrap, Markov chain bootstrap, to name a few, but their validity is proved for specific cases/statistics only. Due to complications with entropy calculations for dependent triangular arrays, almost all results treat the standard empirical processes $\mathbb{G}_n^*$ with few notable exceptions. The moving block bootstrap was justified for VC-type classes, but only under rather restrictive β-mixing conditions on X_i (Radulović, 1996). Bracketing classes were considered by Bühlmann (1995), but his conditions are even more restrictive. In contrast, the process $\{\mathbb{G}_n(t), \, t \in \mathbb{R}\}$ is rather easy to bootstrap. This coupled with Theorem 6 offers the following result.

Corollary 7 *Let $\{X_j\}_{j\geq 1}$ be a stationary sequence of random variables with continuous stationary distribution function F and α-mixing coefficients satisfying $\sum_{k\geq n} \alpha_k \leq Cn^{-\gamma}$, for some $0 < \gamma < 1/3$ and $C < \infty$. Let $\mathbb{G}_n^*$ be the bootstrapped standard empirical process based on the moving block bootstrap, with block sizes b_n, specified in Peligrad (1998, page 882). Then, for $\mathcal{G} \subset BV_T$, $\{\int g \, \mathrm{d}\mathbb{G}_n^*, \; g \in \mathcal{G}\}$ converges weakly to Gaussian process, almost surely, with uniformly $L_1(F)$-continuous sample paths.*

Proof. Theorem 2.3 of Peligrad (1998) establishes the convergence of $\mathbb{G}_n^*$ to a Gaussian process with uniformly continuous sample paths with respect to the distance $d(s,t) = |F(s) - F(t)|$ based on the stationary distribution F of X_i. Invoke Theorem 6 and Lemma 3 to conclude the proof. ∎

Just as for weak convergence of the empirical process based on stationary sequences, there are numerous results that consider bootstrap for stationary, non-mixing sequences. For example, El Ktaibi, Ivanoff and Weber (2014) study short memory causal linear sequences, and prove weak convergence of $\mathbb{G}_n^*$ under conditions (on the growth of the weights a_j, the characteristic function of ξ_0 and moments of ξ_0) akin to the ones required for its non-bootstrap counterpart $\mathbb{G}_n$ (Doukhan and Surgailis, 1998). Again, Theorem 6 easily extends their result.

Corollary 8 *Let $\{X_i = \sum_{j\geq 0} a_j \xi_{i-j}\}_{i\geq 1}$ be a sequence of random variables with stationary distribution F such that conditions of El Ktaibi, Ivanoff and Weber (2014) are satisfied. Then, for any $\mathcal{G} \subset BV_T$, $\{\int g \,\mathrm{d}\mathbb{G}_n^*,\ g \in \mathcal{G}\}$ converges weakly to a Gaussian limit with uniformly $L_1(F)$-continuous sample paths, almost surely.*

Proof. El Ktaibi, Ivanoff and Weber (2014) establish weak convergence of $\mathbb{G}_n^*$ in the Skorohod space, for continuous F, which in turn implies that the limiting process has uniformly continuous sample paths with respect to distance $d(s,t) = |F(s) - F(t)|$. Again, invoke Theorem 6 and Lemma 3 to conclude the proof. ■

4 Proofs for Theorem 1 and Theorem 6

We first give a short proof of Theorem 1 if the limit $\mathbb{Z}(t)$ of $\mathbb{Z}_n(t)$ has continuous sample paths with respect to the Lebesgue measure. Let d_{BL} be the bounded Lipschitz metric that metrizes weak convergence, see, e.g., Van der Vaart and Wellner (1996, page 73) for the definition. Set $\bar{\mathbb{Z}}_n(g) = \int g \,\mathrm{d}\mathbb{Z}_n$ for any $g \in BV_T'$. Assumptions A1 and A2 imply that the Lebesgue-Stieltjes integrals

$$\widetilde{\mathbb{Z}}_n(g) = -\int \mathbb{Z}_n \,\mathrm{d}g$$

are well defined. Next, by the integration by parts formula in Lemma A, we have

$$\bar{\mathbb{Z}}_n(g) = \tilde{\mathbb{Z}}_n(g) + R_n(g)$$

with

$$R_n(g) \leq T \sup_x |\mathbb{Z}_n(x) - \mathbb{Z}_n(x^-)| \to 0,$$

in probability, as $n \to \infty$, since $\mathbb{Z}$ has uniformly continuous sample paths. For fixed g, $\tilde{\mathbb{Z}}_n(g)$ converges weakly to $\tilde{\mathbb{Z}}(g) := -\int \mathbb{Z}\,\mathrm{d}g$ by the continuous mapping theorem and weak convergence of $\mathbb{Z}_n$. The continuous mapping theorem also guarantees that the limit $\tilde{\mathbb{Z}}$ is tight in $\ell^\infty(\mathcal{G})$ as the map $X \mapsto \Gamma_g(X) = -\int X\,\mathrm{d}g$, $g \in \mathcal{G}$, is continuous. By the triangle inequality,

$$\begin{aligned} d_{BL}(\bar{\mathbb{Z}}_n, \tilde{\mathbb{Z}}) &\leq d_{BL}(\bar{\mathbb{Z}}_n, \tilde{\mathbb{Z}}_n) + d_{BL}(\tilde{\mathbb{Z}}_n, \tilde{\mathbb{Z}}) \\ &= \sup_{H \in BL_1} |\mathbb{E}H(\bar{\mathbb{Z}}_n) - \mathbb{E}H(\tilde{\mathbb{Z}}_n)| + \sup_{H \in BL_1} |\mathbb{E}H(\tilde{\mathbb{Z}}_n) - \mathbb{E}H(\tilde{\mathbb{Z}})| \\ &\leq T\mathbb{E}[\sup_x |\mathbb{Z}_n(x) - \mathbb{Z}_n(x^-)|] + T \sup_{H' \in BL_1} |\mathbb{E}H'(\mathbb{Z}_n) - \mathbb{E}H'(\mathbb{Z})| \\ &= T\mathbb{E}[\sup_x |\mathbb{Z}_n(x) - \mathbb{Z}_n(x^-)|] + T d_{BL}(\mathbb{Z}_n, \mathbb{Z}) \end{aligned}$$

The second inequality follows since the map $\Gamma_f(X) := \int X\ \mathrm{d}f$ is linear and Lipschitz with Lipschitz constant $\int |\mathrm{d}f| \leq T$ and the suprema are taken over all Lipschitz functionals $H : \ell^\infty(\mathcal{G}) \to \mathbb{R}$ with $\|H\|_\infty \leq 1$ and $|H(X) - H(Y)| \leq \|X - Y\|_\infty$ and $H' : \ell^\infty(\mathbb{R}) \to \mathbb{R}$ with $\|H'\|_\infty \leq 1$ and $|H'(X) - H'(Y)| \leq \|X - Y\|_\infty$, respectively. Together with the tightness of the limit $\tilde{\mathbb{Z}}$, this implies that the empirical process $\bar{\mathbb{Z}}_n(g)$ indexed by $g \in \mathcal{G} \subset BV'_T$ converges weakly.

The above proof for continuous limit processes $\mathbb{Z}$ (that is, processes $\mathbb{Z}$ with uniformly continuous sample paths) is rather simple. Nevertheless, we could not find an actual publication of this trick. We would like to stress that the extension of this proof to general, non-continuous processes, is not trivial. Technical complications related to the interplay between the atoms of the limiting process $\mathbb{Z}(t)$ and discontinuities of $g \in \mathcal{G}$, require some care.

Lemma 9 *Assume that $\mathbb{Z}_n(t)$ converges weakly to a Gaussian process $\mathbb{Z}(t)$, as $n \to \infty$. Then, for any right continuous function $g : \mathbb{R} \to \mathbb{R}$ of bounded variation $\overline{\mathbb{Z}}_n(g) = \int g\,\mathrm{d}\mathbb{Z}_n$ is well defined, and converges to a normal distribution on $\mathbb{R}$.*

Proof. Let g be an arbitrary right-continuous function of bounded variation. First, we notice that $\int g\,\mathrm{d}\mathbb{Z}_n$ and $\int \mathbb{Z}_n\,\mathrm{d}g$ are indeed well defined as Lebesgue-Stieltjes integrals.

Recall that g can have only countably many discontinuities which we will denote a_i. By the integration by parts formula, Lemma A in the appendix, we have

$$\int g(x)\,\mathrm{d}\mathbb{Z}_n(x) = T_1(\mathbb{Z}_n) + T_2(\mathbb{Z}_n)$$

with operators $T_1, T_2 : \ell^\infty(\mathbb{R}) \to \mathbb{R}$

$$\begin{aligned} T_1(\mathbb{Z}_n) &:= -\int \mathbb{Z}_n(x)\,\mathrm{d}g(x) \\ T_2(\mathbb{Z}_n) &:= \int\int 1_{x=y}\,\mathrm{d}g(x)\,\mathrm{d}\mathbb{Z}_n(y) = \sum_i (g(a_i) - g(a_i^-))(\mathbb{Z}_n(a_i) - \mathbb{Z}_n(a_i^-)). \end{aligned}$$

Since g has finite variation, it is bounded with $\sum_i |\alpha_i| < \infty$ for the jumps $\alpha_i := g(a_i) - g(a_i^-)$. Hence, the operator T_2, given by

$$T_2(\mathbb{Z}_n) = \sum_i \alpha_i(\mathbb{Z}_n(a_i) - \mathbb{Z}_n(a_i^-))$$

is linear and continuous. We conclude the proof by observing that the continuity of the operators T_1 and T_2 and the weak convergence of $\mathbb{Z}_n(t)$ to a Gaussian process $\mathbb{Z}$ ensures that these sequences $Y_n := \int g\,\mathrm{d}\mathbb{Z}_n = T_1(\mathbb{Z}_n) + T_2(\mathbb{Z}_n)$ converge weakly to a normal distribution via the continuous mapping theorem. ■

For any distribution function F_0 and any $\beta > 0$, we define

$$\begin{aligned} \Psi_{\beta,F_0}(\mathbb{Z}_n) &:= \sup_{|F_0(s)-F_0(t)|\le\beta} |\mathbb{Z}_n(s) - \mathbb{Z}_n(t)|\,, \\ \eth_{\beta,F_0}(\mathbb{Z}_n) &:= \sup_{F_0(s)-F_0(s^-)>\beta} \left|\mathbb{Z}_n(s) - \mathbb{Z}_n(s^-)\right|, \end{aligned}$$

with the convention that the supremum taken over the empty set is equal to zero. Clearly, $\eth_{\beta,F_0}(\mathbb{Z}_n) = 0$ for all $\beta > 0$ if F_0 is continuous. In general, for arbitrary F_0, the quantity $\eth_{\beta,F_0}(\mathbb{Z}_n)$ is bounded in probability, for all $\beta > 0$, by the continuous mapping theorem, as long as $\mathbb{Z}_n$ converges weakly. The following lemma plays an instrumental role and it could be perhaps of independent interest.

Lemma 10 (Decoupling Lemma) *For any distribution function F_0 on $\mathbb{R}$, any right-continuous function $g : \mathbb{R} \to \mathbb{R}$ and any $\beta > 0$, we have*

$$|\bar{\mathbb{Z}}_n(g)| \le \left\{2\beta^{-1}||g||_{L_1(F_0)} + 6||g||_{TV}\right\} \Psi_{2\beta,F_0}(\mathbb{Z}_n) + \beta^{-1}\|g\|_{L_1(F_0)}\eth_{\beta,F_0}(\mathbb{Z}_n).$$

Here $\|g\|_{L_1(F_0)} = \int |g|\,\mathrm{d}F_0$ and $\mathbb{Z}_n$ satisfies assumptions A1 and A2.

Proof. Without loss of generality we can assume that $||g||_{L_1(F_0)} + ||g||_{TV} < \infty$. Since F_0 is a distribution function we can construct, for any $0 < \beta < 1$, a finite grid $-\infty = s_0 < s_1 < ... < s_M < \infty$ such that

$$F_0(s_j) - F_0(s_{j-1}) \geq \beta,\ F_0(s_M) < 1 - 2\beta$$

and

$$F_0(s_j^-) - F_0(s_{j-1}) \leq 2\beta,$$

leaving the possible jumps $F_0(s_j) - F_0(s_j^-)$ unspecified. Based on this grid we approximate $\mathbb{Z}_n(t)$ by

$$\widetilde{\mathbb{Z}}_n(t) := \sum_{i=1}^{M} \mathbb{Z}_n(s_{j-1}) 1_{[s_{j-i}, s_j)}(t)$$

and we set $\widetilde{\mathbb{Z}}_n(\pm\infty) = 0$. We observe that by construction

$$\begin{aligned} \sup_x |\mathbb{Z}_n(x) - \widetilde{\mathbb{Z}}_n(x)| &\leq \max_{1 \leq j \leq M} \sup_{x \in [s_{j-1}, s_j)} |\mathbb{Z}_n(x) - \mathbb{Z}_n(s_{j-1})| + \sup_{x \geq s_M} |\mathbb{Z}_n(x)| \\ &\leq \sup_{|F_0(x) - F_0(y)| \leq 2\beta} |\mathbb{Z}_n(x) - \mathbb{Z}_n(y)| = \Psi_{\beta, F_0}(\mathbb{Z}_n) \end{aligned} \tag{3}$$

Since the process $\widetilde{\mathbb{Z}}_n$ inherits the bounded variation properties of $\mathbb{Z}_n$,

$$\int g(s)\, \mathrm{d}\mathbb{Z}_n(s) = \int g(s)\, \mathrm{d}\widetilde{\mathbb{Z}}_n(s) + \int g(s)\, \mathrm{d}(\mathbb{Z}_n - \widetilde{\mathbb{Z}}_n)(s).$$

is well defined for any right-continuous function g of bounded variation. Using the integration by parts formula (Lemma A in the appendix) we obtain for the last term on the right

$$\int g(s)\, \mathrm{d}(\mathbb{Z}_n - \widetilde{\mathbb{Z}}_n)(s) = -\int (\mathbb{Z}_n - \widetilde{\mathbb{Z}}_n)(s)\, \mathrm{d}g(s) + \int\int 1_{x=y}\, \mathrm{d}g(x)\, \mathrm{d}(\mathbb{Z}_n - \widetilde{\mathbb{Z}}_n)(y).$$

Since g is of bounded variation, it has countably many discontinuities a_i. Using (3), we obtain

$$\begin{aligned} \left| \int\int 1_{x=y}\, \mathrm{d}g(x)\, \mathrm{d}(\mathbb{Z}_n - \widetilde{\mathbb{Z}}_n)(y) \right| &\leq \sum_i |g(a_i) - g(a_i^-)| \left| (\mathbb{Z}_n - \widetilde{\mathbb{Z}}_n)(a_i) - (\mathbb{Z}_n - \widetilde{\mathbb{Z}}_n)(a_i^-) \right| \\ &\leq 2\, \|g\|_{TV}\, \Psi_{2\beta, F_0}(\mathbb{Z}_n). \end{aligned}$$

Consequently,

$$\left|\int g(s)\,\mathrm{d}(\mathbb{Z}_n - \widetilde{\mathbb{Z}}_n)(s)\right| \le 3\,\|g\|_{TV}\,\Psi_{2\beta,F_0}(\mathbb{Z}_n)$$

Next we deal with the finite dimensional approximation

$$\int g(s) d\widetilde{\mathbb{Z}}_n(s) = \sum_{j=1}^{M} g(s_j)(\mathbb{Z}_n(s_j) - \mathbb{Z}_n(s_{j-1})).$$

Clearly,

$$\left|\int g(s)d\widetilde{\mathbb{Z}}_n(s)\right| \le \left|\sum_{j=1}^{M} g(s_j)(\mathbb{Z}_n(s_j^-) - \mathbb{Z}_n(s_{j-1}))\right| + \left|\sum_{j=1}^{M} g(s_j)(\mathbb{Z}_n(s_j) - \mathbb{Z}_n(s_j^-))\right| \quad (4)$$

For the first term in (4) we introduce the step function

$$g^*(t) = \sum_{j=1}^{M} 1_{(s_{j-1},s_j]}(t) \inf_{s_{j-1}<s\le s_j} |g(s)|.$$

designed to approximate $|g(t)|$. Clearly $0 \le g^*(t) \le |g(t)|$ and

$$\begin{aligned}\sum_{j=1}^{M} g^*(s_j) &\le \beta^{-1}\sum_{j=1}^{M} g^*(s_j)(F_0(s_j) - F_0(s_{j-1}))\\ &= \beta^{-1}\int g^*\,\mathrm{d}F_0\\ &\le \beta^{-1}\int |g|\,\mathrm{d}F_0 = \beta^{-1}||g||_{L_1(F_0)}\end{aligned}$$

Hence, by (3) we have

$$\begin{aligned}\left|\sum_{j=1}^{M} g(s_j)(\mathbb{Z}_n(s_j^-) - \mathbb{Z}_n(s_{j-1}))\right| &\le \Psi_{2\beta,F_0}(\mathbb{Z}_n)\left[\sum_{j=1}^{M}(|g(s_j)| - g^*(s_j)) + \sum_{j=1}^{M} g^*(s_j))\right]\\ &\le \Psi_{2\beta,F_0}(\mathbb{Z}_n)(\|g\|_{TV} + \beta^{-1}||g||_{L_1(F_0)}) \qquad (5)\end{aligned}$$

For the second term in (4), we have

$$\begin{aligned}&\left|\sum_{j=1}^{M} g(s_j)\left(\mathbb{Z}_n(s_j)) - \mathbb{Z}_n(s_j^-)\right)\right|\\ &\le \left|\sum_{j:\ F_0(s_j)-F_0(s_j^-)\le\beta} g(s_j)\left(\mathbb{Z}_n(s_j)) - \mathbb{Z}_n(s_j^-)\right)\right| + \left|\sum_{j:\ F_0(s_j)-F_0(s_j^-)>\beta} g(s_j)\left(\mathbb{Z}_n(s_j)) - \mathbb{Z}_n(s_j^-)\right)\right|\\ &\le \Psi_{2\beta,F_0}(\mathbb{Z}_n)\left(2||g||_{TV} + \beta^{-1}||g||_{L_1(F_0)}\right) + \eth_{\beta,F_0}(\mathbb{Z}_n)\beta^{-1}\|g\|_{L_1(F_0)}, \qquad (6)\end{aligned}$$

using for the last inequality

$$\sum_{j=1}^{M} |g(s_j)| \leq 2\|g\|_{TV} + \beta^{-1}\|g\|_{L_1(F_0)}$$

and

$$\begin{aligned}\sum_{j=1}^{M} |g(s_j)| 1\{F_0(s_j) - F_0(s_j^-) > \beta\} &\leq \beta^{-1}\sum_{j=1}^{M} |g(s_j)||(F_0(s_j) - F_0(s_j^-)| \\ &\leq \beta^{-1}\|g\|_{L_1(F_0)}.\end{aligned}$$

Lemma 10 now follows by combining the estimates (3) and (4) and (5). ■

An immediate corollary is the following result.

Corollary 11 *For any distribution function F_0 and for all $T < \infty$, $\delta > 0$ and $p \geq 1$, we have*

$$\sup_g \left| \int g \, \mathrm{d}\mathbb{Z}_n \right| \leq (2\sqrt{\delta} + 6T)\Psi_{2\sqrt{\delta},F_0}(\mathbb{Z}_n) + \eth_{\sqrt{\delta},F_0}(\mathbb{Z}_n)\sqrt{\delta},$$

where the sup *is taken over all right-continuous functions g with $\|g\|_{TV} \leq T$ and $\|g\|_{L_p(F_0)} = \left(\int |g|^p \mathrm{d}F_0\right)^{1/p} \leq \delta$.*

Proof. The proof follows trivially from Lemma 10 by taking $\beta = \sqrt{\delta}$ and observing that $\|g\|_{L_1(F_0)} \leq \|g\|_{L_p(F_0)}$ for $p \geq 1$. ■

Proof of Theorem 1.

First we recall, see, for instance, Chapter 1.5 in Van der Vaart and Wellner (1996), that $\{\bar{\mathbb{Z}}_n(g), \, g \in \mathcal{G}\}$ converges weakly to a tight limit in $\ell^\infty(\mathcal{G})$, provided

(a) the marginals $(\bar{\mathbb{Z}}_n(g_1), \ldots, \bar{\mathbb{Z}}_n(g_k))$ converge weakly for every finite subset $g_1, \ldots, g_k \in \mathcal{G}$, and

(b) there exists a semi-metric ρ on $\mathcal{G}$ such that $(\mathcal{G}, \rho)$ is totally bounded and $\bar{\mathbb{Z}}_n(g)$ is ρ-stochastically equicontinuous, that is,

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathbb{P}\left\{ \sup_{\rho(f,g) \leq \delta} |\bar{\mathbb{Z}}_n(f) - \bar{\mathbb{Z}}_n(g)| > \varepsilon \right\} = 0$$

for all $\varepsilon > 0$.

The finite dimensional convergence (a) follows trivially from Lemma 9, linearity of the process $\overline{\mathbb{Z}}_n(g)$ and the Cramèr-Wold device. As for the stochastic equicontinuity (b) of $\overline{Z}_n(g)$, it is sufficient to show that, for every $\varepsilon > 0$,

$$\lim_{\delta\to 0}\lim\sup_{n\to\infty}\mathbb{P}\left\{\sup_{\|h\|_{L_1(F_0)}\le\delta}\left|\int h(t)\,\mathrm{d}\mathbb{Z}_n(t)\right| > \varepsilon\right\} = 0,$$

where the supremum is taken over all differences $h = g - g'$ with $g, g' \in \mathcal{G}$ and $\|h\|_{L_1(F_0)} = \int |h| dF_0 \le \delta$. Since h is also right-continuous and $\|h\|_{TV} \le 2T$, Corollary 11 implies that

$$\sup_{\|h\|_{L_1(F_0)}\le\delta}\left|\int h(t)\,\mathrm{d}\mathbb{Z}_n(t)\right| \le (2\sqrt{\delta}+12T)\Psi_{2\sqrt{\delta},F_0}(\mathbb{Z}_n) + \eth_{\sqrt{\delta},F_0}(\mathbb{Z}_n)\sqrt{\delta}.$$

Let $f_t(x) = 1\{x \le t\}$, so that $\mathbb{Z}_n(t) = \overline{\mathbb{Z}}_n(f_t)$ and

$$d(s,t) = |F_0(s) - F_0(t)|$$

and observe that

$$\sup_{d(s,t)\le\delta}|\mathbb{Z}_n(s) - \mathbb{Z}_n(t)| = \Psi_{\delta,F_0}(\mathbb{Z}_n).$$

The weak convergence of $\mathbb{Z}_n(t)$, or equivalently, $\overline{\mathbb{Z}}_n(f_t)$, to a continuous (with respect to $d(.,.)$) process $\mathbb{Z}(t)$ implies that

$$\Psi_{2\sqrt{\delta},F_0}(\mathbb{Z}_n) \xrightarrow{P} 0 \text{ as } \delta\to 0 \text{ and } n\to\infty.$$

Moreover, the weak convergence of $\mathbb{Z}_n(t)$ implies that $\eth_{\sqrt{\delta},F_0}(\mathbb{Z}_n)$ is bounded in probability, so $\eth_{\sqrt{\delta},F_0}(\mathbb{Z}_n)\sqrt{\delta} \to 0$ as $\delta \to 0$ and $n \to \infty$.

Summarizing, $\bar{\mathbb{Z}}_n(g)$, converges for each $g \in \mathcal{G}$ to a Gaussian limit and $\bar{\mathbb{Z}}_n$ is uniformly $L_1(F_0)$-equicontinuous, in probability. Moreover, $(\mathcal{G}, L_1(F_0))$ is totally bounded (for any distribution F_0). This well-known fact can be found in Example 2.6.21 in Van der Vaart and Wellner (1996, page 149). It implies the weak convergence of $\bar{\mathbb{Z}}_n(g)$ to a process with uniformly $L_1(F_0)-$continuous sample paths (see Theorems 1.5.4 and 1.5.7 in Van der Vaart and Wellner 1996).

Proof of Theorem 6.

Here we only prove the "in probability" case. The almost sure statement follows after straightforward changes in the proof of Theorem 1. A simple modification of Lemma 9 yields

$$\int g(t)\,\mathrm{d}\mathbb{G}_n^*(t) = T_1(\mathbb{G}_n^*) + T_2(\mathbb{G}_n^*)$$

for the same operators T_1 and T_2 as defined in Lemma 9. Since $\mathbb{G}_n^*$ converges to a Gaussian limit the finite dimensional convergence follows by repeating the computation presented in the proof of Lemma 9 after replacing $\mathbb{G}_n(t)$ with $\mathbb{G}_n^*(t)$. As for stochastic equicontinuity of $\mathbb{Z}_n^*(g)$ we find, analogous to Corollary 11, that for $\delta > 0$

$$\sup_{\|g\|_{L_p(F_0)}\le\delta,\ \|g\|_{TV}\le T} \left| \int g\,\mathrm{d}\mathbb{G}_n^* \right| \le (2\sqrt{\delta} + 12T)\Psi_{2\sqrt{\delta},F_0}(\mathbb{G}_n^*) + \eth_{\sqrt{\delta},F_0}(\mathbb{G}_n^*)\sqrt{\delta}.$$

Now the weak convergence of $\mathbb{Z}_n^*$ follows from the convergence of $\mathbb{G}_n^*$, as in the proof of Theorem 1, conditionally given the sample $X_1, \ldots, X_n$. Moreover, if $\mathbb{G}_n(t)$ and $\mathbb{G}_n^*(t)$ converge to the same Gaussian process, then Lemma 9 coupled with the Cramèr-Wold device implies that the finite dimensional distribution of the limiting process of $\mathbb{Z}_n(g)$ and $\mathbb{Z}_n^*(g)$ are the same. This concludes the proof.

A Appendix

For completeness, we state the following classical result and give a simple elementary proof which was communicated to us by David Pollard. Theorem 18.4 in Billingsley (1986) states the result for functions on a bounded interval, yet its proof can be easily extended to the entire real line.

Lemma A. *Let f and g be right-continuous functions of bounded variation and define measures μ and ν as $\mu(-\infty, x] = f(x) - f(-\infty)$ and $\nu(-\infty, y] = g(y) - g(-\infty)$. Then*

$$\int f(x)\,\mathrm{d}g(x) + \int g(x)\,\mathrm{d}f(x) = (fg)(\infty) - (fg)(-\infty) + \int\int 1_{x=y}\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y).$$

Moreover, if either $f(\pm\infty) = 0$ *or* $g(\pm\infty) = 0$*, then*

$$\int f(x)\,\mathrm{d}g(x) + \int g(x)\,\mathrm{d}f(x) = \int\int 1_{x=y}\,\mathrm{d}\mu\,\mathrm{d}\nu.$$

Proof. Set $H(x,y) = 1\{x \le y\}$ and observe that by the very definition of Lebesgue integral

$$f(y) = \int H(x,y)\,\mathrm{d}\mu(x) + f(-\infty)$$

and

$$g(x) = \int H(y,x)\,\mathrm{d}\nu(y) + g(-\infty).$$

Hence

$$\begin{aligned}
&\int f(y)\,\mathrm{d}\nu(y) + \int g(x)\,\mathrm{d}\mu(x) \\
&= \int \left(\int H(x,y)\,\mathrm{d}\mu(x)\right)\,\mathrm{d}\nu(y) + \int \left(\int H(y,x)\,\mathrm{d}\nu(y)\right)\,\mathrm{d}\mu(x) \\
&\quad + f(-\infty)(g(\infty) - g(-\infty)) + g(-\infty)(f(\infty) - f(-\infty))
\end{aligned}$$

Next we apply Fubini

$$\begin{aligned}
&\int \left(\int H(x,y)\,\mathrm{d}\mu(x)\right)\,\mathrm{d}\nu(y) + \int \left(\int H(y,x)\,\mathrm{d}\nu(y)\right)\,\mathrm{d}\mu(x) \\
&= \int\int (H(x,y) + H(y,x))\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) = \int\int (1_{x\le y} + 1_{y\le x})\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) \\
&= \int\int \mathrm{d}\mu(x)\,\mathrm{d}\nu(y) + \int\int 1_{x=y}\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)
\end{aligned}$$

to prove the lemma. ■

Acknowledgements.

We would like to thank both referees for their careful reading and many constructive remarks.

The research of Wegkamp was supported in part by NSF grants DMS-1310119 and DMS-1712709.

References

[1] D.W.K. Andrews and D.P. Pollard (1994). An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes. *International Statistical Review*, **62**, 119–132.

[2] M.A. Arcones, and B. Yu (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *Journal of Theoretical Probability*, **7**, 47–71.

[3] P. Billingsley (1968). *Convergence of probability measures.* J. Wiley & Sons, New York.

[4] P. Billingsley (1986). *Probability and Measure, 2nd edition.* J. Wiley & Sons, New York.

[5] P. Bühlmann (1995). The blockwise bootstrap for general empirical processes of stationary sequences. *Stochastic Processes and their Applications*, **58**, 247–265.

[6] R. M. Dudley, (1999), *Uniform Central Limit Theorems.* Cambridge University Press.

[7] H. Dehling, O. Durieu and D. Volny (2009). New techniques for empirical processes of dependent data. *Stochastic Processes and their Applications*, **119**, 3699–3718.

[8] H. Dehling, O. Durieu and M. Tusche (2014). Approximating class approach for empirical processes of dependent sequences indexed by functions. *Bernoulli*, **20**, 1372–1403.

[9] H. Dehling and M.S. Taqqu (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Annals of Statistics*, **17**, 1767–1783.

[10] P. Doukhan, and D. Surgailis (1998). Functional central limit theorem for the empirical process of short memory linear processes. *Comptes Rendus de l'Académie des Sciences, Mathématique*, **326**, 87–92.

[11] P. Doukhan, P. Massart, and E. Rio (1995). Invariance principles for absolutely regular empirical processes. *Annales de l'Institute Henri Poincaré Probababilités et Statististiques*, **31**, 393–427.

[12] F. El Ktaibi, B.G. Ivanoff, N.C. Weber (2014). Bootstrapping the empirical distribution of a linear process. *Statistics & Probability Letters*, **93**, 134–142.

[13] T.H. Hildebrandt (1963). *Introduction to the Theory of Integration*. Academic Press, New York, London.

[14] H.C. Ho and T. Hsing (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Annals of Statistics*, **24**, 992–1024.

[15] R.Y. Liu and K. Singh (1992), Moving blocks jackknife and bootstrap capture weak dependence, in: R. LePage and L. Billard, eds., *Exploring the Limits of Bootstrap*, Wiley, New York, pp. 225-248.

[16] H.R. Künsch (1989), The jackknife and the bootstrap for general stationary observations, *Annals of Statistics*, **17**, 1217-1241.

[17] M. Peligrad (1998). On the blockwise bootstrap for empirical processes for stationary sequences. *Annals of Probability*, **26**, 887–901.

[18] D. Radulović (1996). The bootstrap for empirical processes based on stationary observations. *Stochastic Processes and Their Applications*, **65**, 259–279.

[19] D. Radulović, M.H. Wegkamp and Y. Zhao (2017). Weak convergence of empirical copula processes indexed by functions. *Bernoulli*, **23**, 3346-3384.

[20] E. Rio (1998). Processus empiriques absolument réguliers et entropie universelle. *Probability Theory and Related Fields*, **111**, 585–608.

[21] E. Rio (2000). *Théorie asymptotique des processus aléatoires faiblement dependants*, Collection Mathématiques and Applications, **31**. Springer.

[22] S. Sengupta, S. Volgushev and X. Shao (2016). A Subsampled Double Bootstrap for Massive Data, *Journal of the American Statistical Association*, **111**, 1222–1232.

[23] G.R. Shorack and J.A. Wellner (2009). *Empirical Processes with Applications to Statistics*, 2nd Edition, SIAM.

[24] A.W. van der Vaart and J. Wellner (2000). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer.

[25] D. Viennet (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probability Theory and Related Fields*, **107**, 467–492.