Central limit theorems under non-stationarity via relative weak convergence

Nicolai Palm and Thomas Nagler

May 3, 2025

Statistical inference for non-stationary data is hindered by the lack of classical central limit theorems (CLTs), not least because there is no fixed Gaussian limit to converge to. To address this, we introduce *relative weak convergence*, a mode of convergence that compares a statistic or process to a sequence of evolving processes. Relative weak convergence retains the main consequences of classical weak convergence while accommodating time-varying distributional characteristics. We develop concrete relative CLTs for random vectors and empirical processes, along with sequential, weighted, and bootstrap variants, paralleling the state-of-the-art in stationary settings. Our framework and results offer simple, plug-in replacements for classical CLTs whenever stationarity is untenable, as illustrated by applications in nonparametric trend estimation and hypothesis testing.

Contents

1. Introduction

In many application areas, data are naturally recorded over time. The processes generating these data often have evolving characteristics. Examples are environmental data, which are subject to evolving climatic conditions; economic and financial data, which are influenced by changing market fundamentals and political environment; and network data in epidemiology and social sciences, which are subject to changes in societal behavior. Given the significant role these data play in our society, there is a strong need for statistical methods that can handle such non-stationary data with formal guarantees.

Most inferential procedures in statistics rely on central limit theorems (CLTs) as a fundamental principle. This includes simple low-dimensional problems such as testing for equality of means, as well as modern high-dimensional and infinite-dimensional regression problems, where empirical process theory (including related uniform CLTs via weak convergence) has become the workhorse (Van der Vaart and Wellner, 2023, Dehling and Philipp, 2002, Kosorok, 2008). However, weak convergence and, by extension, CLTs are fundamentally tied to stationarity. In general non-stationary settings, there may be no fixed Gaussian limit to converge to because the data-generating process evolves over time.

Existing approaches A common ad-hoc fix to this conundrum is to de-trend, difference, or otherwise transform the data to remove the most glaring effects of non-stationarity (e.g., Shumway et al., 2000). The pre-processed data is assumed to be stationary, and classical CLTs are applied. This approach is sometimes successful in practice, but it has its pitfalls. The pre-processing is unlikely to remove all stationarity issues, and potential uncertainty and data dependence in the pre-processing are not accounted for in the subsequent analysis. A likely reason for the popularity of these heuristics is the apparent scarcity of limit theorems for non-stationary data that could support the development of inferential methods.

A few approaches have been proposed to establish CLTs for non-stationary processes, but they have significant limitations. Merlevede and Peligrad (2020) establish a univariate (uniform-in-time) CLT by standardizing the sample average by its standard deviation, which ensures a fixed standard Gaussian limit. Extending this approach to multivariate settings is straightforward by additionally de-correlating the coordinates. However, this is doomed to fail in infinite-dimensional settings, where a decorrelated Gaussian limit process has unbounded sample paths. Another recent linear of work (Karmakar and Wu, 2020, Mies and Steland, 2023, Bonnerjee et al., 2024) couples the sample average to an explicit sequence of Gaussian variables on an enriched probability space. Such results are stronger than necessary for most statistical applications and require substantially more effort to establish than classical CLTs. Furthermore, they apply only to multivariate sample averages with extensions to more complex quantities such as empirical processes being an open problem. The most developed approach thus far relies on the *local stationarity* assumption, as summarized in Dahlhaus (2012) and recently extended to empirical processes by Phandoidaen and Richter (2022). Here, asymptotic results become possible by considering a hypothetical sequence of data-generating processes providing increasingly many, increasingly stationary observations in a given time window. This approach is tailored to procedures that localize estimates in a small window, and the asymptotics do not concern the actual process generating the data. It does not apply, for example, to a simple (non-localized) sample average over non-stationary random variables.

Contribution 1: Relative weak convergence This paper takes a different approach. Fundamentally, a CLT is a comparison between the distribution of a sample quantity to a fixed Gaussian distribution. Since, in the case of non-stationary data, there is no fixed distribution to converge to, we may simply compare to an appropriate sequence of Gaussian distributions. The approximating Gaussians can vary with the sample size, reflecting the evolving structure of the underlying process. This approach is strong enough to substitute classical (uniform) CLTs in statistical methods and simple enough such that the required conditions are not substantially stronger than those required for stationary CLTs. To formalize the idea in general terms, we introduce the following concept.

Definition. *We say X_n, Y_n are* relatively weakly convergent *if*

$$|\mathbb{E}^* [f(X_n)] - \mathbb{E}^*[f(Y_n)]| \to 0$$

for all f bounded and continuous.

Note that if $Y_n = Y$ is constant, this is simply the definition of weak convergence. If T is finite-dimensional and the distributions of Y_n suitably continuous, relative weak convergence is equivalent to convergence of the difference of distribution functions. Our general notion of asymptotic normality looks as follows.

Definition. *We say that a sequence $\{Y_n(t) \colon T\}$ satisfies a* relative central limit theorem *(relative CLT) if there exists a relatively compact sequence of centered, tight, and Borel measurable Gaussian processes N_{Y_n} with*

$$\mathbb{C}\mathrm{ov}[Y_n(s), Y_n(t)] = \mathbb{C}\mathrm{ov}[N_n(s), N_n(t)]$$

such that Y_n and N_{Y_n} are relatively weakly convergent.

Let us explain the definition. The covariance structure of the approximating GP N_{Y_n} is the natural Gaussian process to compare Y_n to. Relatively compact sequences are the natural analog of tight limits in classical weak convergence and defined formally in Section 2. Under relative compactness, relative weak convergence is equivalent to weak convergence on subsequences, which makes it straightforward to transfer many useful tools, such as the continuous mapping theorem or the functional delta method, to relative weak convergence.

Contribution 2: Relative CLTs for sample averages and empirical processes The cases which interest us in this paper are both finite- and infinite-dimensional relative CLTs for the empirical process over a non-stationary time series. It turns out that they hold under high-level assumptions similar to those in classical stationary CLTs. One of our main results looks as follows (Theorem 3.6).

Theorem. *Under some moment, mixing, and bracketing entropy conditions the weighted empirical process* $\mathbb{G}_n \in \ell^\infty(S \times \mathcal{F})$ *defined by*

$$\mathbb{G}_n(s, f) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} w_{n,i}(s)\big(f(X_{n,i}) - \mathbb{E}[f(X_{n,i})]\big)$$

satisfies a relative CLT for every family of functions $w_{n,i} : S \to \mathbb{R}$ *with* $n, i \in \mathbb{N}$ *and* $\sup_{n,i,x} |w_{n,i}(x)| < \infty$.

The weights $w_{n,i}$ can be specified to cover a variety of different CLT flavors. For example, taking $w_{n,i}(s) = \mathbb{1}_{i \leq \lfloor sk_n \rfloor}$, the theorem yields a relative sequential CLT (Corollary 3.8); taking $w_{n,i}(s) = K((i/k_n - s)/b_n)$ for some kernel K and bandwidth b_n, we obtain localized CLTs similar to those established under local stationary by Phandoidaen and Richter (2022).

To prove the theorem, we rely on a characterization of relative CLTs in terms of relative compactness and marginal relative CLTs of $\mathbb{G}_n$, similar to classical empirical process theory. Regarding the marginals, we provide a relative version of Lyapunov's CLT (Theorem 3.2). For relative compactness, we introduce a version of bracketing entropy tailored to non-stationary β-mixing sequences. A coupling and chaining argument then yields asymptotic tightness of $\mathbb{G}_n$, and Prohorov's theorem implies relative compactness. Generally, such entropy conditions also ensure the existence of the sequence of approximating Gaussians.

Contribution 3: Bootstrap inference under non-stationarity Many inferential methods rely on a tractable Gaussian approximation in a weak sense, but whether that Gaussian changes with the sample size plays a subordinate role. A key difficulty remains, however: the covariance structure of the approximating Gaussian process is not known in practice and difficult to estimate. A common solution to this problem is to use the bootstrap. Considerable effort has been devoted to deriving consistent bootstrap schemes for special cases of non-stationary data. Because of the lack of CLTs, no bootstrap scheme currently exists for general non-stationary settings. Extending the work of Bücher and Kojadinovic (2019), bootstrap consistency can be established in terms of relative weak convergence and CLTs. Combined with our relative CLTs, this yields a consistent multiplier bootstrap for non-stationary time series solely under moment and mixing assumptions.

Summary and outline In summary, our main contributions are as follows:

- We introduce a new mode of convergence, which is general enough to explain the asymptotics of non-stationary processes and convenient to establish and use.
- We derive asymptotic normality of empirical processes for non-stationary time series under the same high-level assumptions as classical stationary CLTs.
- We demonstrate the practical use of our theory by proving a characterization of bootstrap consistency in terms of relative weak convergence and CLTs and derive a consistent multiplier bootstrap for non-stationary time series.

This provides a new perspective on the asymptotic theory of non-stationary processes and a comprehensive framework of concepts and tools for statistical inference with general non-stationary data. Our results are applicable to a wide range of statistical methods and effectively usable as drop-in replacements for classical CLTs when data are not stationary.

The remainder of this paper is structured as follows. Relative weak convergence and CLTs are developed in Section 2. We provide tools for proving relative CLTs and foster their intuition by connecting relative with classical CLTs. Section 3 develops relative CLTs for non-stationary β-mixing sequences. Section 4 develops multiplier bootstrap consistency for non-stationary time series, with applications to non-parametric trend estimation and hypothesis testing discussed in Section 5.

2. Relative Weak Convergence and CLTs

This section introduces definitions, characterizations, and basic properties of relative weak convergence and relative CLTs.

2.1. Background and notation

Throughout this paper, we make use of Hoffmann-Jørgensen's theory of weak convergence. For this reason, we recall some definitions and central statements about weak converge in general metric spaces and in the space of bounded functions. The material and references of this section are found in Chapter 1 of Van der Vaart and Wellner (2023).

Weak convergence in general metric spaces In what follows we denote by $X_n : \Omega_n \to \mathbb{D}$ sequences of (not necessarily measurable) maps with Ω_n probability spaces and $\mathbb{D}$ some metric space. We will assume $\Omega_n = \Omega$ without loss of generality (see discussion above Theorem 1.3.4). To avoid clutter, we omit the domain Ω and write $X_n \in \mathbb{D}$ whenever it is clear from context.

Definition 2.1 (Definition 1.3.3). *The sequence X_n converges weakly to some Borel measurable map $X \in \mathbb{D}$ if*

$$\mathbb{E}^*[f(X_\alpha)] \to \mathbb{E}[f(Y)]$$

for all bounded and continuous functions $f : \mathbb{D} \to \mathbb{R}$ *where* $\mathbb{E}^*[f(X_n)]$ *denotes the outer expectation of* $f(X_n)$ *(chapter 1.2). We write*

$$X_n \to_d X.$$

Weak convergence of (measurable) random vectors agrees with the usual notation in terms of distribution functions.

Definition 2.2. *The net* Y_α *is* asymptotically measurable *if*

$$\mathbb{E}^*[f(Y_\alpha)] - \mathbb{E}_*[f(Y_\alpha)] \to 0$$

for all $f : \mathbb{D} \to \mathbb{R}$ *bounded and continuous where* $\mathbb{E}^*$ *and* $\mathbb{E}_*$ *denote outer and inner expectation, respectively.The net is* asymptotically tight *if for all* $\varepsilon > 0$ *there exists a compact set* $K \subseteq \mathbb{D}$ *such that*

$$\liminf_{\alpha} P_*(Y_\alpha \in K^\delta) \geq 1 - \varepsilon$$

for all $\delta > 0$ *with* $K^\delta = \{y \in \mathbb{D} \colon \exists x \in K \text{ s.t. } d(x, y) < \delta\}$ *where* P_* *denotes the inner probability.*

Weak convergence to some tight Borel measure implies asymptotic measurability and asymptotic tightness. Conversely, Prohorov's theorem asserts weak convergence along subsequences whenever the sequence is asymptotically tight and measurable.

Weak convergence of stochastic processes A stochastic process indexed by a set T is a collection $\{Y(t) \colon t \in T\}$ of random variables $Y(t) : \Omega \to \mathbb{R}$ defined on the same probability space. A *Gaussian process (GP)* is a stochastic process $\{N(t) \colon t \in T\}$ such that $(N(t_1), \ldots, N(t_k))$ is multivariate Gaussian for all $t_1, \ldots, t_k \in T$.

Weak convergence of empirical processes is typically considered in (a subspace of) the space of bounded functions

$$\ell^\infty(T) = \left\{ f : T \to \mathbb{R} \colon \|f\|_T = \sup_{t \in T} |f(t)| < \infty \right\}$$

equipped with the uniform metric $d(f, g) = \sup_{t \in T} |f(t) - g(t)| = \|f - g\|_T$. If the sample paths $t \mapsto Y(t)(\omega)$ of a stochastic process Y are bounded for all $\omega \in \Omega$, it induces a map

$$Y : \Omega \to \ell^\infty(T),\ \omega \mapsto (t \mapsto X_t(\omega)).$$

To abbreviate notation, we say $Y \in \ell^\infty(T)$ is a stochastic process (with bounded sample paths) indicating that $Y(t) : \Omega \to \mathbb{R}$ are random variables with common domain.

A sequence $Y_n \in \ell^\infty(T)$ converges weakly to some tight and Borel measurable $Y \in \ell^\infty(T)$ if and only if Y_n is asymptotically tight, asymptotically measurable and all marginals converge weakly, i.e.,

$$(Y_n(t_1), \ldots, Y_n(t_k)) \to_d (Y(t_1), \ldots, Y(t_k))$$

in $\mathbb{R}^k$ for all $t_1, \ldots, t_k \in T$.

2.2. Non-stationary CLTs via weak convergence along subsequences

Let us first explain some difficulties with non-stationary CLTs. Consider a sequence of stochastic processes $X_i \in \ell^\infty(T)$. CLTs establish weak convergence of the empirical process $\mathbb{G}_n \in \ell^\infty(T)$ defined by

$$\mathbb{G}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i(t) - \mathbb{E}[X_i(t)]$$

to some tight and measurable Gaussian process $\mathbb{G} \in \ell^\infty(T)$. The convergence is characterized in terms of asymptotic tightness and marginal CLTs. The former is independent of stationarity (e.g., Theorems 2.11.1 and 2.11.9 of Van der Vaart and Wellner (2023)). The latter involves weak convergence of the marginals; in particular,

$$\mathbb{G}_n(t) \to_d \mathbb{G}(t) \quad \text{for all } t \in T.$$

This also implies convergence of the variances under mild conditions. Here a certain degree of stationarity of the samples is required. For non-stationary observations, the convergence

$$\mathbb{V}\mathrm{ar}[\mathbb{G}_n(t)] \to \mathbb{V}\mathrm{ar}[\mathbb{G}(t)]$$

fails in general.

Example 2.3. *Assume* $\mathbb{V}\mathrm{ar}[X_i(t)] \in \{\sigma_1^2, \sigma_2^2\}$ *and* $X_i(t)$ *are independent. Write* $N_{k,n} = \#\{i \leq n \colon \mathbb{V}\mathrm{ar}[X_i(t)] = \sigma_k^2\}$. *Then,*

$$\mathbb{V}\mathrm{ar}[\mathbb{G}_n(t)] = n^{-1} \sum_{i=1}^n \mathbb{V}\mathrm{ar}[X_i(t)] = \sigma_1^2 \frac{N_{1,n}}{n} + \sigma_1^2 \frac{N_{2,n}}{n} = \sigma_2^2 + (\sigma_1^2 - \sigma_2^2) \frac{N_{1,n}}{n}$$

converges if and only if $\frac{N_{1,n}}{n}$ *converges.*

For this reason, some non-stationary CLTs rely on standardization such that the covariance matrix equals the identity. Such standardization can only work in finite dimensions, however. The intuitive reason is a decorrelated Gaussian limit process $\mathbb{G}$ cannot have bounded sample paths (nor be tight) unless T is finite.

Lemma 2.4. *If a Gaussian process* $\mathbb{G} \in \ell^\infty(T)$ *satisfies* $\mathbb{C}\mathrm{ov}[\mathbb{G}(s), \mathbb{G}(t)] = 0$ *for* $s \neq t$ *and* $\mathbb{V}\mathrm{ar}[\mathbb{G}(s)] = 1$, *then,* T *is finite.*

Proof. Appendix F. □

In summary, the problem with uniform CLTs lies in the fact that marginal CLTs generally require a degree of stationarity - which cannot be bypassed by (naive) standardization.

The basic result which we want to put into a larger context is a version of Lyapunov's CLT. Write

$$S_n = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} X_{n,i} - \mathbb{E}[X_{n,i}].$$

for $X_{n,1} \dots X_{n,k_n} \in \mathbb{R}$ a triangular array of independent random variables. Assume the general moment condition $\sup_{n,i} \mathbb{E}[|X_{n,i}|^{2+\delta}] < \infty$ for some $\delta > 0$. Lyapunov's CLT asserts asymptotic normality whenever variances converge.

Proposition 2.5. *Suppose* $\sup_{n,i} \mathbb{E}[|X_{n,i}|^{2+\delta}] < \infty$ *for some* $\delta > 0$*. Then* $\mathbb{V}\mathrm{ar}[S_n] \to \sigma_\infty^2$ *if and only if* $S_n \to_d \mathcal{N}(0, \sigma_\infty^2)$*.*

Proof. The sufficiency is a special case of Proposition 2.27 of van der Vaart (2000) and the necessity is Example 1.11.4 of Van der Vaart and Wellner (2023). □

When $\mathbb{V}\mathrm{ar}[S_n]$ doesn't converge, the moment condition still implies that $\mathbb{V}\mathrm{ar}[S_n]$ is uniformly bounded. Thus, every subsequence of $\mathbb{V}\mathrm{ar}[S_n]$ contains a converging subsequence $\mathbb{V}\mathrm{ar}[S_{n_k}]$ along which Lyapunov's CLT implies asymptotic normality of S_{n_k}. The weak limit of S_{n_k} however depends on the limit of $\mathbb{V}\mathrm{ar}[S_{n_k}]$. In other words, along subsequences S_n converges weakly to some Gaussian, yet not globally. To obtain a global Gaussian to compare S_n to, the complementary observation is the following fact

$$\mathbb{V}\mathrm{ar}[S_{n_k}] \to \sigma_\infty^2 \qquad \text{if and only if} \qquad \mathcal{N}(0, \mathbb{V}\mathrm{ar}[S_{n_k}]) \to_d \mathcal{N}(0, \sigma_\infty^2).$$

Thus, along subsequences, S_n and $\mathcal{N}(0, \mathbb{V}\mathrm{ar}[S_n])$ have the same weak limits. Because any subsequence of S_n contains a weakly convergent subsequence, i.e., S_n is relatively compact, some thought reveals their difference of distribution functions converges to zero.

Proposition 2.6 (Relative Lyapunov's CLT)**.** *Denote by* F_n *resp.* Φ_n *the distribution function of* S_n *resp.* $\mathcal{N}(0, \mathbb{V}\mathrm{ar}[S_n])$*. Then,* $|F_n(t) - \Phi_n(t)| \to 0$ *for all* $t \in [0, 1]$*.*

Proof. This is a special case of Theorem A.7. □

In other words, even in non-stationary settings, the scaled sample average remains approximately Gaussian, but with the notion of a limiting distribution replaced by a sequence of Gaussians. This mode of convergence and the resulting type of asymptotic normality extends naturally to stochastic processes.

2.3. Relative weak convergence

All proofs of the remaining section are found in Appendix A. Let X_n, Y_n be sequences of arbitrary maps from probability spaces Ω_n, Ω_n' into a metric space $\mathbb{D}$. Throughout the rest of this paper it is tacitly understood that all such maps have a common probability space as their domain.

Definition 2.7 (Relative weak convergence)**.** *We say X_n and Y_n are* relatively weakly convergent *if*

$$|\mathbb{E}^* [f(X_n)] - \mathbb{E}^*[f(Y_n)]| \to 0$$

for all $f : \mathbb{D} \to \mathbb{R}$ bounded and continuous. We write

$$X_n \leftrightarrow_d Y_n.$$

Remark 2.8. *For X a Borel law, $X_n \leftrightarrow_d X$ if and only if $X_n \to_d X$.*

Contrary to weak convergence, relative weak convergence implies neither measurability nor tightness. Any sequence is relatively weakly convergent to itself. For the purpose of this paper, we restrict to relative weak convergence to relatively compact sequences. Those turn out to be the natural analog of tight, measurable limits in classical weak convergence.

Definition 2.9 (Relative asymptotic tightness and compactness)**.** *We call X_n* relatively asymptotically tight *if every subsequence contains a further subsequence which is asymptotically tight. We call X_n* relatively compact *if every subsequence contains a further subsequence which converges weakly to a tight Borel law.*

From the definition we see that if $X_n \leftrightarrow_d Y_n$ and $Y_n \to_d Y$ then $X_n \to_d Y$. Thus, if Y_n is relatively compact, $X_n \leftrightarrow_d Y_n$ essentially states that X_n and Y_n have the same weak limits along subsequences.

Proposition 2.10. *Assume that X_n is relatively compact. The following are equivalent:*

(i) $X_n \leftrightarrow_d Y_n$.

(ii) For all subsequences n_k such that $X_{n_k} \to_d X$ with X a tight Borel law it follows $Y_{n_k} \to_d X$.

(iii) For all subsequences n_k there exists a further subsequence n_{k_i} such that both $X_{n_{k_i}}$ and $Y_{n_{k_i}}$ converge weakly to the same tight Borel law.

In such case, Y_n is relatively compact as well.

The characterization of relative weak convergence via weak convergence on subsequences is very convenient. In particular, many useful properties of weak convergence can be transferred to relative weak convergence. The following results are particularly relevant for statistical applications and indicate that relative weak convergence allows for similar conclusions than weak convergence.

Proposition 2.11 (Relative continuous mapping)**.** *If $X_n \leftrightarrow_d Y_n$ and $g : \mathbb{D} \to \mathbb{E}$ is continuous then $g(X_n) \leftrightarrow_d g(Y_n)$.*

Proposition 2.12 (Extended relative continuous mapping). *Let $g_n : \mathbb{D} \to \mathbb{E}$ be a sequence of functions with $\mathbb{E}$ another metric space. Assume that for all subsequences of n there exists another subsequence n_k and some $g : \mathbb{D} \to \mathbb{E}$ such that $g_{n_k}(x_k) \to g(x)$ for all $x_k \to x$ in $\mathbb{D}$. Let Y_n be relatively compact. Then,*

(i) $g_n(Y_n)$ is relatively compact.

(ii) if $X_n \leftrightarrow_d Y_n$ then $g_n(X_n) \leftrightarrow_d g_n(Y_n)$.

Proposition 2.13 (Relative delta-method). *Let $\mathbb{D}, \mathbb{E}$ be metrizable topological vector spaces and $\theta_n \in \mathbb{D}$ be relatively compact. Let $\phi : \mathbb{D} \to \mathbb{E}$ be continuously Hadamard-differentiable (see Definition A.2) in an open subset $\mathbb{D}_0 \subset \mathbb{D}$ with $\theta_n \in \mathbb{D}_0$ for all n. Assume*

$$r_n\left(X_n - \theta_n\right) \leftrightarrow_d Y_n$$

for some sequence of constants $r_n \to \infty$ with Y_n relatively compact. Then,

$$r_n\big(\phi(X_n) - \phi(\theta_n)\big) \leftrightarrow_d \phi'_{\theta_n}(Y_n).$$

Let $S^\delta = \{x \in \mathbb{D} \colon d(x, S) < \delta\}$ denote the δ-enlargement and ∂S the boundary (closure minus interior) of a set S.

Proposition 2.14. *Let $X_n \leftrightarrow_d Y_n$ with Y_n relatively compact, and S_n be Borel sets such that every subsequence of S_n contains a further subsequence S_{n_k} such that $\lim_{k\to\infty} S_{n_k} = S$ for some S with*

$$\liminf_{\delta\to 0} \limsup_{k\to\infty} \mathbb{P}^*(Y_{n_k} \in (\partial S)^\delta) = 0. \tag{1}$$

Then

$$\lim_{n\to\infty} |\mathbb{P}^*(X_n \in S_n) - \mathbb{P}^*(Y_n \in S_n)| = 0.$$

Condition (1) ensures that S_n are continuity sets of Y_n asymptotically in a strong form. This prevents the laws from accumulating too much mass near the boundary of S_n. In statistical applications, Y_n is typically (the supremum of) a non-degenerate Gaussian process, for which appropriate anti-concentration properties can be guaranteed for any set S (e.g., Giessing, 2023).

A fundamental result in empirical process theory is the characterization of weak convergence in terms of asymptotic tightness and marginal weak convergence (Van der Vaart and Wellner, 2023, Theorem 1.5.7). This characterization underpins uniform CLTs. Because relative weak convergence to some relatively compact sequence is equivalent to weak convergence along subsequences, a similar result holds for relative weak convergence.

Theorem 2.15. *Let $X_n, Y_n \in \ell^\infty(T)$ with Y_n relatively compact. The following are equivalent:*

(i) $X_n \leftrightarrow_d Y_n$.

(ii) X_n *is relatively asymptotically tight and all marginals satisfy*

$$\big(X_n(t_1), \dots, X_n(t_d)\big) \leftrightarrow_d \big(Y_n(t_1), \dots, Y_n(t_d)\big)$$

for all finite subsets $\{t_1, \dots, t_d\} \subset T$.

2.4. Relative central limit theorems

Fix a sequence of stochastic processes $Y_n \in \ell^\infty(T)$ with $\sup_{t \in T} \mathbb{E}[Y_n(t)^2] < \infty$. CLTs assert weak convergence $Y_n \to_d N$ with N some tight and measurable GP. Naturally, we define (relative) asymptotic normality as $Y_n \leftrightarrow_d N_n$ with N_n some sequences of GPs. Contrary to CLTs with a fixed limit, there is no unique sequence of 'limiting' GPs, however. It shall be convenient to specify the covariance structure of N_n to mirror that of Y_n. This allows for a tractable Gaussian approximation.

Definition 2.16 (Corresponding GP)**.** *Let Y be a map with values in $\ell^\infty(T)$. A* Gaussian process (GP) corresponding to Y *is a map N_Y with values in $\ell^\infty(T)$ such that*

$$\{N_Y(t) \colon t \in T\}$$

is a centered GP with covariance function given by $(s,t) \mapsto \mathbb{C}\mathrm{ov}[Y(s), Y(t)]$.

Similar to classical CLTs and in view of Proposition 2.10, we further restrict to relatively compact sequences of tight and Borel measurable GPs and define a relative CLT as follows.

Definition 2.17 (Relative CLT)**.** *We say that the sequence Y_n satisfies a* relative central limit theorem *if exists a relatively compact sequence of tight and Borel measurable GPs N_{Y_n} corresponding to Y_n with $Y_n \leftrightarrow_d N_{Y_n}$.*

Theorem 2.15 characterizes relative in terms of marginal relative CLTs.

Corollary 2.18. *The sequence Y_n satisfies a relative CLT if and only if*

(i) there exist tight and Borel measurable GPs N_{Y_n} corresponding to Y_n,

(ii) Y_n and N_{Y_n} are relatively asymptotically tight and

(iii) all marginals $(Y_n(t_1), \dots, Y_n(t_k)) \in \mathbb{R}^k$ satisfy a relative CLT.

The restriction to relatively compact sequences of tight and Borel measurable GPs enables inference just as in classical weak convergence theory (see previous section and Section 5). Such sequences exist under mild assumptions. In finite dimensions, corresponding tight Gaussians always exist. Any such sequence converges iff its covariances converge. Hence, a sequence of Gaussians is relatively compact iff covariances converge along subsequences. Equivalently, the sequences of variances are uniformly bounded (Corollary A.6). As a result, relative CLTs can be characterized as follows.

Proposition 2.19. *Let Y_n be a sequence of $\mathbb{R}^d$-valued random variables. Denote by Σ_n the covariance matrix of Y_n. Then, the following are equivalent:*

(i) Y_n satisfies a relative CLT.

(ii) for all subsequences n_k such that $\Sigma_{n_k} \to \Sigma$ converges it holds

$$Y_{n_k} \to_d \mathcal{N}(0, \Sigma)$$

and $\sup_{n \in \mathbb{N}, i \leq d} \mathbb{V}\mathrm{ar}[Y_n^{(i)}] < \infty$ where $Y_n^{(i)}$ denotes the i-th component of Y_n.

(iii) all subsequences n_k contain a subsequence n_{k_i} such that $\Sigma_{n_{k_i}} \to \Sigma$ and

$$Y_{n_{k_i}} \to_d \mathcal{N}(0, \Sigma).$$

In other words, multivariate relative CLTs are essentially equivalent to CLTs along subsequences where covariances converge. From this it is straightforward to generalize classical multivariate CLTs to relative multivariate CLTs (e.g., Lindeberg's CLT, Theorem A.7).

For infinite dimensional index sets T, Kolmogorov's extension theorem implies existence of GPs $\{N_n(t) \colon t \in T\}$ with

$$\mathbb{C}\mathrm{ov}[N_n(s), N_n(t)] = \mathbb{C}\mathrm{ov}[Y_n(s), Y_n(t)],$$

but possibly unbounded sample paths. Tightness can be guaranteed under entropy conditions, as shown in the following. Define the covering number $N(\epsilon, T, d)$ with respect to some semi-metric space (T, d) as the minimal number of ϵ-balls needed to cover T. Denote by

$$\rho_n(s, t) = \mathbb{V}\mathrm{ar}[Y_n(s) - Y_n(t)]^{1/2}, \quad s, t \in T,$$

the standard deviation semi-metric on T induced by Y_n.

Proposition 2.20. *If for all n it holds*

$$\int_0^\infty \sqrt{\ln N(\epsilon, T, \rho_n)} d\epsilon < \infty,$$

there exists a sequence of tight and Borel measurable GPs N_{Y_n} corresponding to Y_n.

Proposition 2.21. *Let N_{Y_n} be a sequence of Borel measurable GPs corresponding to Y_n. Assume that there exists a semi-metric d on T such that*

(i) (T, d) is totally bounded.

(ii) $\lim_{n\to\infty} \int_0^{\delta_n} \sqrt{\ln N(\epsilon, T, \rho_n)} d\epsilon = 0$ *for all* $\delta_n \downarrow 0$.

(iii) $\lim_{n\to\infty} \sup_{d(s,t)<\delta_n} \rho_n(s,t) = 0$ *for every* $\delta_n \downarrow 0$.

If further $\sup_n \mathbb{V}\mathrm{ar}[Y_n(t)] < \infty$ *for all* $t \in T$*, the sequence* N_{Y_n} *is asymptotically tight.*

Condition (iii) requires the existence of a global semi-metric with respect to which the sequence of standard deviation semi-metrics are asymptotically uniformly continuous. In many practical settings, one can simply take $d(t,s) = \sup_n \rho_n(t,s)$.

3. Relative CLTs for non-stationary time-series

Now that we know what relative CLTs are, this section provides specific instances of such results for non-stationary time-series.

The assumptions of these CLTs are relatively weak, making them applicable in a wide range of statistical problems. Of course, no CLT can be expected to hold under arbitrary dependence structures. Several measures exist to constrain the dependence between observations, such as α-mixing or ϕ-mixing coefficents (Bradley, 2005), or the functional dependence measure of Wu (2005). In the following, we will focus on β-mixing, because it is widely applicable and allows for sharp coupling inequalities.

Definition 3.1. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space and* $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ *sub-* σ*-algebras. Define the* β-mixing coefficient

$$\beta(\mathcal{F}_1, \mathcal{F}_2) = \frac{1}{2} \sup \sum_{(i,j)\in I\times J} |P(A_i \cap B_j) - P(A_i)P(B_j)|$$

where the supremum is taken over all finite partitions $\cup_{i\in I} A_i = \cup_{j\in J} B_j = \Omega$ *with* $A_i \in \mathcal{A}_1, B_j \in \mathcal{A}_2$*. For* $p, n \in \mathbb{N}$ *and* $p < k_n$ *define*

$$\beta_n(p) = \sup_{k\le k_n - p} \beta\left(\sigma(X_{n,1}, \dots, X_{n,k}), \sigma(X_{n,k+p}, \dots, X_{n,k_n})\right).$$

The β-mixing coefficients quantify how independent events become as one moves further into the sequence. If the β-mixing coefficients become zero as n and p approach infinity, the events become close to independent. Note that the mixing coefficients themselves are indexed by n, because they may change over time in the nonstationary setting.

3.1. Multivariate relative CLT

We start with a multivariate relative CLT for triangular arrays of random variables. Proposition 2.19 extends classical to relative multivariate CLTs. The following result builds on Lyapunov's CLT in combination with a coupling argument. Let $X_{n,1}, \dots, X_{n,k_n}$ be a triangular array of $\mathbb{R}^d$-valued random variables.

Theorem 3.2 (Multivariate relative CLT)**.** *Let $\alpha \in [0, 1/2)$ and $1 + (1 - 2\alpha)^{-1} < \gamma$. Assume*

(i) $k_n^{-1} \sum_{i,j=1}^{k_n} |\mathbb{C}\mathrm{ov}[X_{n,i}^{(l_1)}, X_{n,j}^{(l_2)}]| \leq K$ *for all n and $l_1, l_2 = 1, \ldots d$.*

(ii) $\sup_{n,i} \mathbb{E}\left[|X_{n,i}^{(l)}|^\gamma\right] < \infty$ *for all $l = 1, \ldots, d$.*

(iii) $k_n \beta_n (k_n^\alpha)^{\frac{\gamma-2}{\gamma}} \to 0$.

Then, the scaled sample average $k_n^{-1/2} \sum_{i=1}^{k_n} (X_{n,i} - \mathbb{E}[X_{n,i}])$ satisfies a relative CLT.

Proof. Appendix C.1 □

The summability condition (i) on the covariances can be seen as a minimal requirement for any general CLT under mixing conditions (Bradley, 1999). Condition (i) and (iii) restrict the dependence where (i) essentially bounds the variances of the scaled sample average. Condition (ii) excludes heavy tails of $X_{n,i}$. Note that (ii) and (iii) exhibit a trade-off. Universal bounds on higher moments weaken the conditions on the mixing coefficients' decay rate.

3.2. Asymptotic tightness under bracketing entropy conditions

In order to extend the multivariate CLT to an empirical process CLT, we need a way to ensure relative compactness. Consider the following general setup:

- $X_{n,1}, \ldots, X_{n,k_n}$ is a triangular array of random variables in some Polish space $\mathcal{X}$.
- $\mathcal{F}_n = \{f_{n,t} \colon t \in T\}$ is a set of measurable functions from $\mathcal{X}$ to $\mathbb{R}$ for all n.
- $\cup_{n \in \mathbb{N}} \mathcal{F}_n$ admits a finite envelope function $F : \mathcal{X} \to \mathbb{R}$, i.e., $\sup_{n \in \mathbb{N}, f \in \mathcal{F}_n} |f(x)| \leq F(x)$ for all $x \in \mathcal{X}$.

Define the empirical process $\mathbb{G}_n$ on T by

$$\mathbb{G}_n(t) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} f_{n,t}(X_{i,n}) - \mathbb{E}[f_{n,t}(X_{i,n})].$$

Because we have an envelope, $\mathbb{G}_n$ has bounded sample paths and we obtain a map $\mathbb{G}_n$ with values in $\ell^\infty(T)$. This empirical process can be seen as a triangular version of the (classical) empirical process indexed by a single set of functions.

To ensure asymptotic tightness, we rely on bracketing entropy conditions, with norms tailored to the nonstationary time-series setting. Given a semi-norm $\|\cdot\|$ on (an extension of) $\mathcal{F}$, recall that the bracketing numbers $N_{[]}(\epsilon, \mathcal{F}, d) \in \mathbb{N}$ are the minimal number of brackets

$$[l_i, u_i] = \{f \in \mathcal{F} \colon l_i \leq f \leq u_i\}$$

such that $\mathcal{F} = \cup_{i=1}^{N_\epsilon}[l_i, u_i]$ with $l_i, u_i : \mathcal{X} \to \mathbb{R}$ measurable and $\|l_i - u_i\| < \epsilon$. For stationary observations $X_{n,i} \sim P$, bracketing entropy is usually measured with respect to an $L_p(P)$-norm, $p \geq 2$. In case of non-stationary observations the brackets need to be measured with respect to all underlying laws of the samples. It turns out that a scaled average of $L_p(P_{X_{n,i}})$-norms is sufficient.

Definition 3.3. *Let $\gamma \geq 1$. Define the semi-norms $\|\cdot\|_{\gamma,n}$ resp. $\|\cdot\|_{\gamma,\infty}$ on $\mathcal{F}$ by*

$$\|h\|_{\gamma,n} = \left(\frac{1}{k_n}\sum_{i=1}^{k_n}\mathbb{E}\left[|h(X_{n,i})|^\gamma\right]\right)^{1/\gamma} \qquad \|h\|_{\gamma,\infty} = \sup_{n\in\mathbb{N}}\|h\|_{\gamma,n}.$$

Note that $\|\cdot\|_{\gamma,n}$ is a composition of semi-norms, hence, a semi-norm itself (Lemma B.3). By a classical chaining argument, $\|\cdot\|_{\gamma,n}$-bracketing entropy conditions imply asymptotic tightness of $\mathbb{G}_n$ under mixing assumptions (Appendix B). Because $\|\cdot\|_{\gamma,n}$-bracketing entropy bounds covering entropy (Remark B.8) the existence of an approximating sequence of GPs is also guaranteed.

Theorem 3.4. *Assume that for some $\gamma > 2$*

(i) $\|F\|_{\gamma,\infty} < \infty$.

(ii) $\sup_{n\in\mathbb{N}}\max_{m\leq k_n} m^\rho\beta_n(m) < \infty$ *for some* $\rho > \gamma/(\gamma-2)$

(iii) $\int_0^{\delta_n}\sqrt{\ln N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{\gamma,n})}d\epsilon \to 0$ *for all* $\delta_n \downarrow 0$ *and are finite for all* n.

Denote by

$$d_n(s,t) = \|f_{n,s} - f_{n,t}\|_{\gamma,n}$$

for $s, t \in T$. Assume that there exists a semi-metric d on T such that

$$\lim_{n\to\infty}\sup_{d(s,t)<\delta_n} d_n(s,t) = 0$$

for all $\delta_n \downarrow 0$ and (T, d) is totally bounded. Then,

- *$\mathbb{G}_n$ is asymptotically tight.*
- *there exists an asymptotically tight sequence of tight Borel measurable GPs N_n corresponding to $\mathbb{G}_n$.*

The proof, given in Appendix B, relies on a chaining argument with adaptive coupling and truncation. An important intermediate step is a new maximal inequality (see Theorem B.6) that may be of independent interest:

Theorem 3.5. *Let $\mathcal{F}$ be a class of functions $f\colon \mathcal{X} \to \mathbb{R}$ with envelope F, and*

$$\|f\|_{\gamma,n} \leq \delta \qquad \frac{1}{n}\sum_{i,j=1}^{n}|\mathbb{C}\mathrm{ov}[h(X_i), h(X_j)]| \leq K_1\|h\|_{\gamma,n}^2$$

for some $\gamma > 2$, all $f \in \mathcal{F}$ and $h : \mathcal{X} \to \mathbb{R}$ bounded and measurable. Suppose that $\max_m \beta_n(m) \leq K_2 m^{-\rho}$ *for some $\rho \geq \gamma/(\gamma-2)$. Then, for any $n \geq 5$ and $\delta \in (0,1)$,*

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \int_0^\delta \sqrt{\ln_+ N_{[]}(\epsilon)}d\epsilon + \frac{\|F\|_{\gamma,n}[\ln N_{[]}(\delta)]^{[1-1/(\rho+1)](1-1/\gamma)}}{n^{-1/2+[1-1/(\rho+1)](1-1/\gamma)}} + \sqrt{n}N_{[]}^{-1}(e^n).$$

In many applications, $\|\cdot\|_{\gamma,n}$-bracketing numbers can be replaced by $L_\gamma(Q)$-bracketing numbers whenever all $P_{X_{n,i}}$ are simultaneously dominated by some measure Q (Appendix F.1), or simply the L_∞-bracketing numbers (which coincide with L_∞-covering numbers) whenever the function class is uniformly bounded. Many bounds on $L_\gamma(Q)$-bracketing numbers are well known (Section 2.7 of Van der Vaart and Wellner (2023)).

3.3. Weighted uniform relative CLT

We now specialize the general uniform relative CLT to the case of weighted empirical processes with a single function class $\mathcal{F}$. Let $\gamma > 2$ and let $\mathcal{F}$ be a set of measurable functions with finite envelope F. Recall that an envelope of $\mathcal{F}$ is a function $F : \mathcal{X} \to \mathbb{R}$ with $\sup_{f\in\mathcal{F}} |f(x)| \leq F(x)$ for all $x \in \mathcal{X}$. Assume the family of weights $w_{n,i} : S \to \mathbb{R}$, $n \in \mathbb{N}, i \leq k_n$ satisfies $\sup_{n,i,x} |w_{n,i}(x)| < \infty$ and recall the definition of the weighted empirical process $\mathbb{G}_n \in \ell^\infty(S \times \mathcal{F})$

$$\mathbb{G}_n(s,f) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} w_{n,i}(s)\big(f(X_{n,i}) - \mathbb{E}[f(X_{n,i})]\big).$$

Note that the envelope implies bounded sample paths of $\mathbb{G}_n$.

A relative CLT for $\mathbb{G}_n$ can be established in terms of bracketing entropy for the function class

$$\mathcal{W}_n = \{g_{n,s} \colon \{1,\ldots,k_n\} \to \mathbb{R}, i \mapsto w_{n,i}(s) \colon s \in S\}.$$

Denote by d_n semi-metrics on S defined by

$$d_n^w(s,t) = \|g_{n,s} - g_{n,t}\|_{\gamma,n} = \left(\frac{1}{k_n}\sum_{i=1}^{k_n} |w_{n,i}(s) - w_{n,i}(t)|^\gamma\right)^{1/\gamma}$$

and assume the following entropy conditions on the weights

W1 $\int_0^{\delta_n} \sqrt{\ln N_{[]}\left(\epsilon, \mathcal{W}_n, \|\cdot\|_{\gamma,n}\right)}d\epsilon \to 0$ for all $\delta_n \downarrow 0$ and finite for all n.

W2 there exists a semi-metric d^w on S such that for all $\delta_n \downarrow 0$

$$\lim_{n\to\infty} \sup_{d^w(s,t)<\delta_n} d_n^w(s,t) = 0.$$

W3 (S,d) is totally bounded.

These assumptions, cover the constant case $w_{n,i}(s) = 1$ as well as more sophisticated weights; see the next sections.

Theorem 3.6 (Weighted relative CLT)**.** *Assume* ***W1***–***W3*** *and the following hold:*

(i) $\|F\|_{\gamma,\infty} < \infty$.

(ii) $\sup_{n\in\mathbb{N}} \max_{m\leq k_n} m^\rho \beta_n(m) < \infty$ *for some* $\rho > 2\gamma(\gamma-1)/(\gamma-2)^2$

(iii) $\int_0^\infty \sqrt{\ln N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{\gamma,\infty})} d\epsilon < \infty$.

Then, $\mathbb{G}_n$ *satisfies a relative CLT in* $\ell^\infty(S \times \mathcal{F})$.

Proof. Appendix C.2 □

The moment condition (i) ensures that all γ-moments $\mathbb{E}[|f(X_{n,i})|^\gamma]$ are uniformly bounded. Again, (i) and (ii) display a trade-off. Higher moments allow for a slower polynomial decay of the mixing coefficients.

Example 3.7. *Assuming the moment condition with* $\gamma = 4$ *the* β*-mixing coefficients must decay polynomial of degree greater than* 6*, i.e.,* $\sup_n \beta_n(m) \leq Cm^{-\rho}$ *for some* $\rho > 6$ *and all* m.

3.4. Sequential relative CLT

A famous result by Donsker asserts asymptotic normality of the partial sum process over *iid* data. A generalization to sequential empirical processes $\mathbb{Z}_n \in \ell^\infty([0,1] \times \mathcal{F})$ indexed by functions, defined as

$$\mathbb{Z}_n(s, f) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{\lfloor sk_n \rfloor} f(X_{n,i}) - \mathbb{E}[f(X_{n,i})],$$

can be found in, e.g., Theorem 2.12.1 of Van der Vaart and Wellner (2023). Beyond the *iid* case, asymptotic normality of $\mathbb{Z}_n$ is hard to prove and requires additional technical assumptions even for finite function classes (Dahlhaus et al., 2019, Merlevède et al., 2019). Specifying $w_{n,i}(s) = \mathbb{1}\{i \leq \lfloor sn \rfloor\}$, relative sequential CLTs are simple corollaries of weighted relative CLTs.

Corollary 3.8 (Sequential relative CLT)**.** *Under conditions (i)–(iii) of Theorem 3.6, the sequential empirical process* $\mathbb{Z}_n \in \ell^\infty([0,1] \times \mathcal{F})$ *satisfies a relative CLT.*

Proof. Appendix C.2 □

4. Bootstrap inference

To make practical use of relative CLTs, we need a way to approximate the distribution of limiting GPs. Their covariance operators are a moving target, however, and generally difficult to estimate. The bootstrap is a convient way to approximate the distribution of limiting GPs, and easy to implement in practice. This section provides some general results on the consistency of multiplier bootstrap schemes for non-stationary time-series.

4.1. Bootstrap consistency and relative weak convergence

To define what bootstrap consistency means in the context of empirical processes, we follow the setup and notation of Bücher and Kojadinovic (2019). In fact, we shall see that the usual definition of bootstrap consistency can be equivalently expressed in terms of relative weak convergence. Let $\mathbb{X}_n$ be some sequence of random variables with values in $\mathcal{X}_n$ and $\mathbb{W}_n$ an additional sequence of random variables, independent of $\mathbb{X}_n$, with values in $\mathcal{W}_n$ with $\mathbb{W}_n^{(j)}$ denoting independent copies of $\mathbb{W}_n$. Denote by $\mathbb{G}_n = \mathbb{G}_n(\mathbb{X}_n)$ resp. $\mathbb{G}_n^{(j)} = \mathbb{G}_n(\mathbb{X}_n, \mathbb{W}_n^{(j)})$ a sequence of maps constructed from $\mathbb{X}_n$ resp. $\mathbb{X}_n, \mathbb{W}_n^{(j)}$ with values in $\ell^\infty(T)$ such that each $\mathbb{G}_n(t), \mathbb{G}_n(t)^{(j)}$ is measurable.

Proposition 4.1. *Assume that $\mathbb{G}_n$ is relatively compact. Then, the following are equivalent:*

(i) for $n \to \infty$

$$\sup_{h \in \mathrm{BL}_1(\ell^\infty(\mathcal{F}))} \left| \mathbb{E}\left[h(\mathbb{G}_n^{(1)})|\mathbb{X}_n\right] - \mathbb{E}\left[h(\mathbb{G}_n)\right]\right| \overset{P^*}{\to} 0$$

and $\mathbb{G}_n^{(1)}$ is asymptotically measurable

(ii) it holds

$$\left(\mathbb{G}_n, \mathbb{G}_n^{(1)}, \mathbb{G}_n^{(2)}\right) \leftrightarrow_d \mathbb{G}_n^{\otimes 3}$$

where $\mathrm{BL}_1(\ell^\infty(\mathcal{F}))$ denotes the set of 1-Lipschitz continuous functions from $\ell^\infty(\mathcal{F})$ to $\mathbb{R}$. Call $\mathbb{G}_n^{(j)}$ a consistent bootstrap scheme *in any such case.*

Classically, $\mathbb{G}_n$ is some (transformation of an) empirical process and consistency of the bootstrap is derived from CLTs for $\mathbb{G}_n$ and $\mathbb{G}_n^{(j)}$. In view of Proposition 4.1, this approach generalizes to relative CLTs (Corollary D.1).

4.2. Multiplier bootstrap

Now fix some triangular array $\mathbb{X}_n = (X_{n,1}, \ldots, X_{n,k_n}) \in \mathcal{X}^n$ of random variables with values in a Polish space $\mathcal{X}$ and some family of uniformly bounded functions $w_{n,i} : S \to \mathbb{R}$. Let $\mathcal{F}$ be a set of measurable functions from $\mathcal{X}$ to $\mathbb{R}$ with finite envelope F. Denote by $\mathbb{V}_n = (V_{n,1}, \dots V_{n,k_n}) \in \mathbb{R}^{k_n}$ a triangular array of random variables and by $\mathbb{V}_n^{(i)} = (V_{n,1}^{(i)}, \dots, V_{n,k_n}^{(i)})$ independent copies of $\mathbb{V}_n$. Define $\mathbb{G}_n, \mathbb{G}_n^{(j)} \in \ell^\infty(S \times \mathcal{F})$ by

$$\mathbb{G}_n(s, f) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} w_{n,i}(s)\big(f(X_{n,i}) - \mathbb{E}[f(X_{n,i})]\big),$$
$$\mathbb{G}_n^{(j)}(s, f) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} V_{n,i}^{(j)} w_{n,i}(s)\big(f(X_{n,i}) - \mathbb{E}[f(X_{n,i})]\big).$$

Proposition 4.2. *Let $X_{n,i}$ satisfy the conditions of Theorem 3.6 for some $\gamma > 2$ and ρ. For every $\epsilon > 0$, let $\nu_n(\epsilon)$ be such that*

$$\max_{|i-j| \le \nu_n(\epsilon)} |\mathbb{C}\mathrm{ov}[V_{n,i}, V_{n,j}] - 1| \le \epsilon.$$

Assume that

(i) $V_{n,1}, \dots V_{n,k_n}$ are identically distributed and independent of $(X_{n,i})_{i \in N}$.

(ii) $\mathbb{E}[V_{n,i}] = 0$, $\mathbb{V}\mathrm{ar}[V_{n,i}] = 1$, and $\sup_n \mathbb{E}[|V_{n,i}|^\gamma] < \infty$

(iii) $k_n \beta_n^X\big(\nu_n(\epsilon)\big)^{\frac{\gamma-2}{\gamma}}, k_n \beta_n^V\big(k_n^\alpha\big)^{\frac{\gamma-2}{\gamma}} \to 0$ for every $\epsilon > 0$ and some $1 + (1-2\alpha)^{-1} < \gamma$.

Then, $\mathbb{G}_n^{(j)}$ is a consistent bootstrap scheme.

Proof. Appendix D.2 □

Example 4.3 (Block bootstrap with exponential weights)**.** *Let $\xi_i \sim \mathrm{Exp}(1)$ be* iid *for $i \in \mathbb{Z}$ and define*

$$V_{n,i} = \frac{1}{\sqrt{2m_n}} \sum_{j=i-m_n}^{i+m_n} (\xi_j - 1).$$

Then $V_{n,i}$ are m_n-dependent and it holds $|\mathbb{C}\mathrm{ov}[V_{n,i}, V_{n,j}] - 1| \le |i-j|/m_n$. Choosing $\nu_n(\varepsilon) = \lfloor \varepsilon m_n \rfloor$, we see that if

(i) $m_n < k_n^\alpha$ for some $1 + (1-2\alpha)^{-1} < \gamma$ and

(ii) $k_n \beta_n^X(\epsilon m_n)^{(\gamma-2)/\gamma} \to 0$ for every $\epsilon > 0$,

conditions (i)–(iii) of Proposition 4.2 are satisfied. Continuing Example 3.7 with $\gamma = 4$ and $\sup_n \beta_n(m) \le Cm^{-\rho}$ for some $\rho > 6$, we can pick $m_n = k_n^{1/3}$.

4.3. Practical inference

The bootstrap process $\mathbb{G}_n^{(j)}$ in the previous section depends on the unknown quantity $\mu_n(i, f) = \mathbb{E}[f(X_{n,i})]$. In many testing applications, we have $\mathbb{E}[f(X_{n,i})] = 0$ at least under the null hypothesis; see Section 5. If this is not the case, estimating $\mu_n(i, f)$ consistently may still be possible in simple problems (e.g., fix-degree polynomial trend), or under triangular array asymptotics where $\mu_n(i, f)$ approaches a simple function (e.g., local stationarity). For a general, observed non-stationary process $(X_i)_{i\in\mathbb{N}}$, it is impossible to distinguish a random series $(X_i)_{i\in\mathbb{N}}$ with $\mathbb{E}[f(X_i)] = 0$ from a deterministic one with $\mathbb{E}[X_i] = \mu(i)$. As a consequence, it is generally impossible to quantify the uncertainty in $\mathbb{G}_n$ consistently. This is a fundamental problem in non-stationary time series analysis, which the relative CLT framework makes transparent.

A modified bootstrap can still provide valid, but possibly conservative, inference in practice. Let $\widehat{\mu}_n(i, f)$ be a potentially non-consistent estimator of $\mu_n(i, f)$, $\overline{\mu}_n(i, f) = \mathbb{E}[\widehat{\mu}_n(i, f)]$ its expectation, and define the processes

$$\widehat{\mathbb{G}}_n^*(s, f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n V_{n,i} w_{n,i}(s)(f(X_i) - \widehat{\mu}_n(i, f)),$$

$$\overline{\mathbb{G}}_n^*(s, f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n V_{n,i} w_{n,i}(s)(f(X_i) - \overline{\mu}_n(i, f)).$$

We assume that $\widehat{\mu}_n(i, f)$ converges to $\overline{\mu}_n(i, f)$ in the following sense.

Proposition 4.4. *Suppose the conditions of Proposition 4.2 are satisfied, $\widehat{\mathbb{G}}_n^* - \overline{\mathbb{G}}_n^*$ is relatively compact, and for every $\epsilon > 0$,*

$$\sup_{f\in\mathcal{F}} \max_{1\leq i\leq n} \mathbb{V}\mathrm{ar}[\widehat{\mu}_n(i, f)] = o(\nu_n(\epsilon)^{-1}).$$

Then $\|\widehat{\mathbb{G}}_n^* - \overline{\mathbb{G}}_n^*\|_{\mathcal{S}\times\mathcal{F}} \to_p 0$.

The variance condition is fairly mild: it must vanish, but can do so at a rate much slower than $1/n$. If the bias also vanishes at an appropriate rate, the approximated bootstrap process $\widehat{\mathbb{G}}_n^*$ is, in fact, consistent.

Proposition 4.5. *Suppose the conditions of Proposition 4.2 and Proposition 4.4 are satisfied, $\overline{\mathbb{G}}_n^*$ is relatively compact, and*

$$\sup_{f\in\mathcal{F}} \max_{1\leq i\leq n} |\overline{\mu}_n(i, f) - \mu_n(i, f)|^2 = o(\nu_n(\epsilon)^{-1}),$$

Then $\widehat{\mathbb{G}}_n^*$ is consistent for $\mathbb{G}_n$.

In particular, this allows to derive bootstrap consistency under local stationarity asymptotics under standard conditions. As explained above, this type of consistency should not be expected for the asymptotics of the observed process. Even without consistency, the bootstrap still provides valid, but conservative inference, as shown in the following propositions.

Proposition 4.6. *Define $\widehat{q}^*_{n,\alpha}$ as the $(1-\alpha)$-quantile of $\|\widehat{\mathbb{G}}^*_n\|_{\mathcal{S}\times\mathcal{F}}$. Suppose that the conditions of Proposition 4.4 hold, and that $\overline{\mathbb{G}}^*_n$ is relatively compact and satisfies a relative CLT with tight corresponding GPs and $\mathbb{V}\mathrm{ar}[\overline{\mathbb{G}}^*_n(s,f)], \mathbb{V}\mathrm{ar}[\mathbb{G}_n(s,f)] \geq \underline{\sigma} > 0$ for all $(s,f) \in \mathcal{S}\times\mathcal{F}$ and n large. Then,*

$$\liminf_{n\to\infty} \mathbb{P}(\|\mathbb{G}_n\|_{\mathcal{S}\times\mathcal{F}} \leq \widehat{q}^*_{n,\alpha}) \geq 1-\alpha.$$

If, on the other hand, $\overline{\mathbb{G}}^*_n$ is *not* relatively compact, it usually holds

$$\mathbb{P}(\|\overline{\mathbb{G}}^*_n\|_{\mathcal{S}\times\mathcal{F}} > t_n) \to 1, \tag{2}$$

for some $t_n \to \infty$. In this case, the bootstrap is over-conservative.

Proposition 4.7. *If (2) and the conditions of Theorem 3.6 and Proposition 4.4 hold, then*

$$\lim_{n\to\infty} \mathbb{P}(\|\mathbb{G}_n\|_{\mathcal{S}\times\mathcal{F}} \leq \widehat{q}^*_{n,\alpha}) = 1.$$

Although the bootstrap quantiles may be conservative, they are still informative: as n tends to infinity, it usually holds $\widehat{q}^*_{n,\alpha}/\sqrt{n} \to 0$ (see the proof of Lemma D.3). In this sense, the bootstrap quantiles yield potentially conservative, yet asymptotically vanishing, uniform confidence intervals for the weighted mean $n^{-1}\sum_{i=1}^n w_{n,i}\mu(i,\cdot)$.

5. Applications

To end, we explore some exemplary applications that cannot be handled by previous results. The methods are illustrated by monthly mean temperature anomalies in the Northern Hemisphere from 1880 to 2024 provided by NASA (GISTEMP Team, 2025, Lenssen et al., 2024), and shown in Fig. 1.

5.1. Uniform confidence bands for a nonparametric trend

Nonparametric estimation of the trend function $\mu(i) = \mathbb{E}[X_i]$ is a key problem in nonstationary time series analysis. As explained in Section 4, $\mu(i) = \mathbb{E}[X_i]$ is not estimable consistently in general, at least not under the asymptotics of the observed process $(X_i)_{i\in\mathbb{N}}$. The local stationarity assumption (Dahlhaus, 2012) resolves this issue by considering the asymptotics of a hypothetical sequence of time series $(X_{n,i})_{i\in\mathbb{N}}$ in which $\mu_n(i) = \mathbb{E}[X_{n,i}]$ becomes flat as $n\to\infty$.

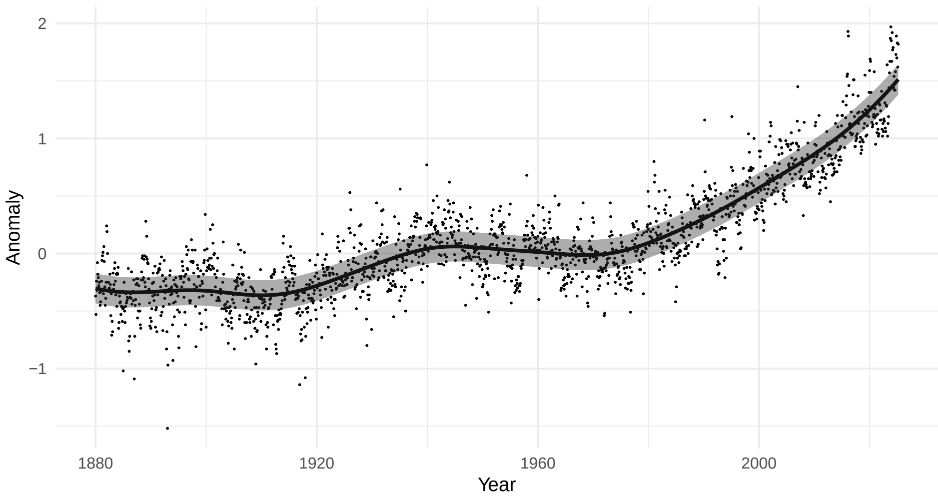

Figure 1: Monthly mean temperature anomalies in the Northern Hemisphere from 1880 to 2024 (dots), kernel estimate of the trend (solid line), and uniform 90%-confidence interval (shaded area).

The relative CLT framework allows us to consider the asymptotics of the observed process $(X_i)_{i\in\mathbb{N}}$, but we have to accept the fact that $\mu(i) = \mathbb{E}[X_i]$ cannot be estimated consistently. Instead, we aim for estimating the smoothed mean

$$\mu_b(i) = \frac{1}{nb}\sum_{j=1}^{n} K\left(\frac{j-i}{nb}\right)\mu(j) \quad \text{by} \quad \widehat{\mu}_b(i) = \frac{1}{nb}\sum_{j=1}^{n} K\left(\frac{j-i}{nb}\right)X_j,$$

where K is a kernel function supported on $[-1, 1]$ and $b > 0$ is the bandwidth parameter. In our setting, bandwidth b plays a different role than in the local stationarity framework. Since $\mathbb{E}[\widehat{\mu}_n(i)] = \mu_b(i)$ for every value of b, we do not require $b \to 0$ as $n \to \infty$. Instead, we may choose b to be a fixed value, quantifying the scale (as a fraction of the overall sample) at which we look at the series. Fig. 1 shows the kernel estimate of the trend function $\mu_b(i)$ for $b = 0.05$ as a solid line. Here, $b = 0.05$ means that the smoothing window spans $0.1 \times n$ months points, or roughly 15 years. Holding b fixed conveniently allows for uniform-in-time asymptotics as shown in the following.

Corollary 5.1. *Suppose the kernel K is a four times continuously differentiable probability density function supported on $[-1, 1]$, $\mathbb{E}[|X_i|^5] < \infty$, and $\sup_n \beta_n^X(k) \lesssim k^{-7}$. Then for any $b > 0$, the process*

$$s \mapsto \sqrt{n}\,(\widehat{\mu}_b(sn) - \mu_b(sn))$$

satisfies a relative CLT in $\ell^\infty([0, 1])$.

To quantify uncertainty of the estimator, we can use the bootstrap. Specifically, let

$$\widehat{\mu}_b^*(i) = \frac{1}{b}\sum_{j=1}^{n} K\left(\frac{j-i}{b}\right)V_{n,i}(X_j - \widehat{\mu}_b(j)),$$

with block multipliers $V_{n,i}$ as in Example 4.3. Let $\widehat{q}_{n,\alpha}$ be the $(1-\alpha)$-quantile of the distribution of $\sup_{s\in[0,1]} \sqrt{n}|\widehat{\mu}_b^*(sn)|$. Then, for $\alpha \in (0,1)$, we can construct a uniform confidence interval for $\mu_b(sn)$ by

$$\widehat{\mathcal{C}}_n(\alpha) = \left[\widehat{\mu}_b - \widehat{q}_{n,\alpha}/\sqrt{n}, \widehat{\mu}_b + \widehat{q}_{n,\alpha}/\sqrt{n}\right].$$

Corollary 5.2. *Suppose the condition of Corollary 5.1 hold,* $\sup_i |\mu(i)| < \infty$*, and there is* $s \in [0,1]$ *such that*

$$\overline{\sigma}_n^2(s) = \mathbb{V}\mathrm{ar}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n V_{n,i}K\left(\frac{i-sn}{nb}\right)(\mu_b(i)-\mu(i))\right] \to \infty.$$

Then

$$\liminf \mathbb{P}(\mu_b \in \widehat{\mathcal{C}}_n(\alpha)) \geq 1-\alpha.$$

The condition on $\overline{\sigma}_n^2(s)$ is usually satisfied in the time series setting, where a diverging number of the covariances $\mathbb{C}\mathrm{ov}[V_{n,i}, V_{n,j}]$ are close to 1. If this is not the case, a similar result could be established using Proposition 4.6. The confidence interval $\widehat{\mathcal{C}}_n(\alpha)$ is shown as a shaded area in Fig. 1. The confidence interval is uniformly valid for all $s \in [0,1]$ and shows a significant, strongly increasing trend in the last 50 years.

5.2. Testing for time series characteristics

Suppose $Z_1, Z_2, \ldots$ is a non-stationary time series and we want to test

$$H_0\colon \sup_i \sup_{f\in\mathcal{F}} |\mathbb{E}[f(Z_i)]| = 0 \qquad \text{against} \qquad H_1\colon \sup_i \sup_{f\in\mathcal{F}} |\mathbb{E}[f(Z_i)]| \neq 0.$$

The functions $f \in \mathcal{F}$ determine which characteristics of the time series we want to control. This framework includes many important applications, two of which are discussed below.

Example 5.3 (Equal characteristics of two series)**.** *Suppose* $Z_i = (X_i, Y_i)$*,* $i \in \mathbb{N}$*, and we want to test whether the two time series* $X_1, X_2, \ldots$ *and* $Y_1, Y_2, \ldots$ *have the same characteristics. To do so, let* $\mathcal{F} = \{f\colon g(x) - g(y), g \in \mathcal{G}\}$*, so that*

$$H_0\colon \sup_i \sup_{g\in\mathcal{G}} |\mathbb{E}[g(X_i)] - \mathbb{E}[g(Y_i)]| = 0.$$

Here $\mathcal{G}$ *describes the characteristics of the two time series* X_i, Y_i *that we want to match. Common choices are monomials or indicator functions for testing equality of moments or distribution, respectively.*

Example 5.4 (Deterministic trends). *Suppose we want to test for a deterministic trend in a time series $(X_i)_{i\in\mathbb{N}}$. Let $\Delta_h X_i = X_{i+h} - X_i$ be the h-step forward difference operator, and define $\Delta_h^r = \Delta_h^{r-1} X_{i+h} - \Delta_h^{r-1} X_i$, for $r \geq 2$. The null hypothesis is $H_0 : \mathbb{E}[\Delta_h^r X_i] = 0$ for all $i \in \mathbb{N}$ and $1 \leq r \leq R$, and fixed $h, R \in \mathbb{N}$. The parameter R determines the order of the polynomial trend we want to test for. The step-size h allows focusing on long-term trends in the presence of deterministic seasonality. This fits into the above framework by letting $Z_i = (\Delta_h X_i, \dots, \Delta_h^R X_i)$, with the convention $\Delta_h^r X_i = 0$ for $hr \geq i$, and*

$$\mathcal{F} = \{f \colon f(z_1, \dots, z_R) = z_j, 1 \leq r \leq R\}.$$

The multiplier bootstrap allows to construct a test for the general null hypothesis above. Define the test statistic and its bootstrap version

$$T_n = \sup_{s\in[0,1], f\in\mathcal{F}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n w_{n,i}(s) f(Z_i) \right|, \qquad T_n^* = \sup_{s\in[0,1], f\in\mathcal{F}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n V_{n,i} w_{n,i}(s) f(Z_i) \right|,$$

where $w_{n,i}(s)$ are some weights. For example, $w_{n,i}(s) = K((i - sn)/nb)$ allows focusing on time-local deviations from the null hypothesis.

Let $\alpha \in (0, 1)$, and $c_n^*(\alpha)$ be the $(1 - \alpha)$-quantile of the distribution of T_n^*. We reject H_0 iff $T_n > c_n^*(\alpha)$. Level and consistency of the test can be straightforwardly derived from our general results.

Corollary 5.5. *Let the sequence of weights $w_{n,i}(s)$ and $\mathcal{F}$ satisfy the conditions of Theorem 3.6. It holds $\mathbb{P}(T_n > c_n^*(\alpha)) \to \alpha$ under H_0, and $\mathbb{P}(T_n > c_n^*(\alpha)) \to 1$ whenever*

$$\liminf_{n\to\infty} \sup_{s\in\mathcal{S}, f\in\mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n w_{n,i}(s) \mathbb{E}[f(Z_i)] \right| = \delta > 0.$$

Because $\sup_i |\mathbb{E}[f(Z_i)]| \neq 0$ under H_1, the distribution of the bootstrap statistic T_n^* does not resemble the distribution of T_n under the alternative. Consistency still follows from the fact that $T_n/\sqrt{n}$ is bounded away from zero with probability tending to 1, and $T_n^*/\sqrt{n} \to_p 0$. The power of the test can be improved if we center by some (non-consistent) estimator $\widehat{\mu}_n(i, f)$ as discussed in Section 4.

As an illustration, we apply the above procedure to test for nonstationarity of the monthly mean anomalies. For example, let $Z_i = (X_i, X_{i-120})$ be a pair of anomolies 10 years apart, $\mathcal{F} = \{f \colon f(x, y) = \mathbb{1}(x < t) - \mathbb{1}(y < t) \colon t \in [-5, 5]\}$, and $w_{n,i}(s) = K_b((i - sn)/nb)$. This gives a Kolmogorov-Smirnov-type statistic

$$T_n = \sup_{t\in[-5,5], s\in[0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n K\left(\frac{i - sn}{bn}\right) \left(\mathbb{1}_{X_i<t} - \mathbb{1}_{X_{i-120}<t}\right) \right|.$$

It is straightforward to show that the conditions of Corollary 5.5 hold, and we can use the multiplier bootstrap to construct a test for the null hypothesis of no nonstationarity. Using $b = 0.05$ and a kernel estimator for $\widehat{\mu}_n(s, f)$ as in the previous section, we get $T_n = 0.69$ and $c_n^*(0.05) = 0.30$, and a bootstrapped p-value smaller than 0.0001, providing strong evidence against the null hypothesis of stationarity.

References

Bonnerjee, S., Karmakar, S., and Wu, W. B. (2024). Gaussian approximation for non-stationary time series with optimal rate and explicit construction. *The Annals of Statistics*, 52(5):2293 – 2317.

Bradley, R. C. (1999). On the growth of variances in a central limit theorem for strongly mixing sequences. *Bernoulli*, 5(1):67–80.

Bradley, R. C. (2005). Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys*, 2(none):107 – 144.

Bücher, A. and Kojadinovic, I. (2019). A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *Journal of Theoretical Probability*, 32(3):1145–1165.

Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of statistics*, volume 30, pages 351–413. Elsevier.

Dahlhaus, R., Richter, S., and Wu, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli*, 25(2):1013 – 1044.

Dehling, H. and Philipp, W. (2002). Empirical process techniques for dependent data. In *Empirical process techniques for dependent data*, pages 3–113. Springer.

Doukhan, P. (2012). *Mixing: Properties and Examples.* Lecture Notes in Statistics. Springer New York.

Giessing, A. (2023). Anti-concentration of suprema of gaussian processes and gaussian order statistics.

GISTEMP Team (2025). GISS Surface Temperature Analysis (GISTEMP), version 4. https://data.giss.nasa.gov/gistemp/. Dataset accessed 2025-05-01.

Karmakar, S. and Wu, W. B. (2020). Optimal gaussian approximation for multiple time series. *Statistica Sinica*, 30(3):1399–1417.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*, volume 61. Springer.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*, volume 23. Springer Science & Business Media.

Lenssen, N., Schmidt, G. A., Hendrickson, M., Jacobs, P., Menne, M., and Ruedy, R. (2024). A GISTEMPv4 observational uncertainty ensemble. *J. Geophys. Res. Atmos.*, 129(17):e2023JD040179.

Merlevede, F. and Peligrad, M. (2020). Functional clt for nonstationary strongly mixing processes. *Statistics & Probability Letters*, 156:108581.

Merlevède, F., Peligrad, M., and Utev, S. (2019). Functional CLT for martingale-like nonstationary dependent structures. *Bernoulli*, 25(4B):3203 – 3233.

Mies, F. and Steland, A. (2023). Sequential gaussian approximation for nonstationary time series in high dimensions. *Bernoulli*, 29(4):3114–3140.

Phandoidaen, N. and Richter, S. (2022). Empirical process theory for locally stationary processes. *Bernoulli*, 28(1):453 – 480.

Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. Probability Theory and Stochastic Modelling. Springer Berlin Heidelberg.

Shumway, R. H., Stoffer, D. S., and Stoffer, D. S. (2000). *Time series analysis and its applications*, volume 3. Springer.

van der Vaart, A. (2000). *Asymptotic Statistics*. Asymptotic Statistics. Cambridge University Press.

Van der Vaart, A. and Wellner, J. (2023). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer International Publishing.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154.

A. Proofs for relative weak convergence and CLTs

Lemma A.1. *The sequence X_n is relatively compact if and only if it is asymptotically measurable and relatively asymptotically tight.*

Proof. If X_n is relatively compact, it converges to some tight Borel law along subsequences. Along such subsequences n_k, X_{n_k} is asymptotically tight and measurable by Lemma 1.3.8 of Van der Vaart and Wellner (2023). We obtain the sufficiency. Note that asymptotic measurability along subsequences implies (global) asymptotic measurability.

For the necessity, for any subsequence there exists a further subsequence n_k such that X_{n_k} is asymptotically tight and measurable. By Prohorov's theorem (Van der Vaart and Wellner, 2023, Theorem 1.3.9), there exists a further subsequence n_{k_i} such that $X_{n_{k_i}}$ converges weakly to some tight Borel law. This implies relative compactness of X_n. □

Proof of Proposition 2.10. Recall that X_n converges weakly to a Borel law X iff

$$\mathbb{E}^*[f(X_n)] \to \mathbb{E}[f(X)]$$

for all $f : \mathbb{D} \to \mathbb{R}$ bounded and continuous.

1. $\Rightarrow$ 2. :

Assume that $X_n \leftrightarrow_d Y_n$. Then, for all $f : \mathbb{D} \to \mathbb{R}$ bounded and continuous

$$\begin{aligned}|\mathbb{E}^*[f(Y_{n_k})] - \mathbb{E}[f(X)]| &\leq |\mathbb{E}^*[f(X_{n_k})] - \mathbb{E}^*[f(Y_{n_k})]| + |\mathbb{E}^*[f(X_{n_k})] - \mathbb{E}[f(X)]| \\ &\to 0\end{aligned}$$

whenever $X_{n_k} \to_d X$ and X Borel measurable.

2. $\Rightarrow$ 3. :

Since X_n is relatively compact, every subsequence X_{n_k} contains a weakly convergent subsequence $X_{n_{k_i}} \to_d X$ with X tight and Borel measurable. By assumption also $Y_{n_k} \to_d X$.

3. $\Rightarrow$ 1. :

Given f, it suffices to prove that for all subsequences n_k there exists a further subsequence n_{k_i} such that

$$\left|\mathbb{E}^*[f(X_{n_{k_i}})] - \mathbb{E}^*[f(Y_{n_{k_i}})]\right| \to 0.$$

Pick n_{k_i} such that both

$$X_{n_{k_i}} \to_d X \text{ and } Y_{n_{k_i}} \to_d X$$

with X tight and Borel measurable. Then,

$$\begin{aligned}\left|\mathbb{E}^*[f(Y_{n_{k_i}})] - \mathbb{E}^*[f(Y_{n_{k_i}})]\right| &\leq \left|\mathbb{E}[f(X)] - \mathbb{E}^*[f(Y_{n_{k_i}})]\right| + \left|\mathbb{E}^*[f(X_{n_{k_i}})] - \mathbb{E}[f(X)]\right| \\ &\to 0\end{aligned}$$

At last, characterization (iii) implies relative compactness of Y_n. □

Proposition 2.11. For all $f : \mathbb{E} \to \mathbb{R}$ bounded and continuous, the composition $f \circ g : \mathbb{D} \to \mathbb{R}$ is bounded and continuous. Thus, $|\mathbb{E}^* [f \circ g(X_n)] - \mathbb{E}^*[f \circ g(Y_n)]| \to 0$ for all such f by definition of $X_n \leftrightarrow_d Y_n$. This yields the claim. $\square$

Proof of Proposition 2.12. Any subsequence of n contains a further subsequence such that $Y_{n_k} \to_d Y$ and there exists $g : \mathbb{D} \to \mathbb{E}$ such that $g_{n_k}(x_k) \to g(x)$ for all $x_k \to x$ in $\mathbb{D}$. Theorem 1.11.1 of Van der Vaart and Wellner (2023) implies $g_n(Y_{n_k}) \to_d g(Y)$. In particular, $g_n(Y_n)$ is relatively compact and Proposition 2.10 yields the second claim. $\square$

Proof of Proposition 2.14. We prove this statement by contradiction. Suppose that

$$\limsup_{n\to\infty} \mathbb{P}^*(X_n \in S_n) - \mathbb{P}^*(Y_n \in s_n) > 0.$$

Then there is a subsequence n_k of n such that

$$\lim_{i\to\infty} \mathbb{P}^*(X_{n_{k_i}} \in S_{n_{k_i}}) - \mathbb{P}^*(Y_{n_{k_i}} \in S_{n_{k_i}}) > 0, \tag{3}$$

for every subsequence n_{k_i} of n_k. By Proposition 2.10, n_k has a subsequence n_{k_i} on which $X_{n_{k_i}} \to_d Y$ and $Y_{n_{k_i}} \to_d Y$ for some tight Borel law Y. We may further assume that $S_{n_{k_i}}$ converges to S on the same subsequence. It holds

$$\begin{aligned}
&\limsup_{i\to\infty} \mathbb{P}^*(X_{n_{k_i}} \in S_{n_{k_i}}) - \mathbb{P}^*(Y_{n_{k_i}} \in S_{n_{k_i}}) \\
&\leq \limsup_{i\to\infty} \mathbb{P}^*(X_{n_{k_i}} \in S_{n_{k_i}}) - \liminf_{i\to\infty} \mathbb{P}^*(Y_{n_{k_i}} \in S_{n_{k_i}}) \\
&\leq \limsup_{i\to\infty} \mathbb{P}^*\left(X_{n_{k_i}} \in \limsup_{i\to\infty} S_{n_{k_i}}\right) - \liminf_{i\to\infty} \mathbb{P}^*\left(Y_{n_{k_i}} \in \liminf_{i\to\infty} S_{n_{k_i}}\right) \\
&= \limsup_{i\to\infty} \mathbb{P}^*(X_{n_{k_i}} \in S) - \liminf_{i\to\infty} \mathbb{P}^*(Y_{n_{k_i}} \in S).
\end{aligned}$$

Further, the Portmanteau theorem (Van der Vaart and Wellner, 2023, Theorem 1.3.4) gives

$$\mathbb{P}^*(Y \in \partial S) \leq \mathbb{P}^*(Y \in (\partial S)^\delta) \leq \limsup_{i\to\infty} \mathbb{P}^*(Y_{n_{k_i}} \in (\partial S)^\delta).$$

Taking $\delta \to 0$, we obtain $\mathbb{P}(Y \in \partial S) = 0$, so S is a continuity set of Y. Now the Portmanteau theorem implies

$$\limsup_{i\to\infty} \mathbb{P}^*(X_{n_{k_i}} \in S) - \liminf_{i\to\infty} \mathbb{P}^*(Y_{n_{k_i}} \in S) = \mathbb{P}^*(Y \in S) - \mathbb{P}^*(Y \in S) = 0,$$

which contradicts (3). The case where (3) holds with reverse sign is treated analogously. $\square$

Definition A.2. *Let $\mathbb{D}, \mathbb{E}$ be metrizable topological vector spaces, this is metric spaces such that addition and scalar multiplication are continuous. A map $\phi : \mathbb{D} \to \mathbb{E}$ is called*

Hadamard-differentiable at $\theta \in \mathbb{D}$ if there exists $\phi'_\theta : \mathbb{D} \to \mathbb{E}$ continuous and linear such that

$$\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \to \phi_\theta(h)$$

for all $t_n \to t$ in $\mathbb{R}$, $h_n \to h$ such that $\theta + t_n h_n \in \mathbb{D}$ for all n. ϕ is continuously Hadamard-differentiable in an open subset $U \subset \mathbb{D}$ if ϕ is Hadamard-differentiable for all $\theta \in U$ and ϕ'_θ is continuous in $\theta \in U$.

Proof of Proposition 2.13. Note that $g_n : \mathbb{D} \to \mathbb{E}, x \mapsto \phi'_{\theta_n}(x)$ satisfies the condition of Proposition 2.12 since ϕ has continuous Hadamard-differentials. Thus, $\phi'_{\theta_n}(Y_n)$ is relatively compact since Y_n is. By (iii) of Proposition 2.10 and descending to subsequences, we may assume $Y_n \to_d Y$, $\theta_n \to \theta$ and $\phi'_{\theta_n}(Y_n) \to_d \phi'_\theta(Y)$. By Theorem 3.10.4 in Van der Vaart and Wellner (2023), we obtain

$$r_n\left(\phi(X_n) - \phi(\theta_n)\right) \to_d \phi'_\theta(Y).$$

Then, (iii) of Proposition 2.10 yields the claim. □

A.1. Relative weak convergence in $\ell^\infty(T)$

Proof of Theorem 2.15. If $X_n \leftrightarrow_d Y_n$ then X_n is relatively compact by Proposition 2.10. This is equivalent to relative asymptotic tightness and asymptotic measurability by Lemma A.1. The continuous mapping theorem then implies marginal relative weak convergence.

For the reverse direction, let n_k be as subsequence. Let n_{k_i} be a subsequence of n_k such that $X_{n_{k_i}} \to_d X$ with X a tight Borel law and $Y_{n_{k_i}}$ is asymptotically tight. Since all marginals of X_n and Y_n are relatively weakly convergent, this implies the convergence of all marginals of Y_{n_k} to the marginals of X by characterization (ii) of Proposition 2.10. Note that all marginals of X_n are relatively compact, e.g., by the continuous mapping theorem. Together with asymptotic tightness of $Y_{n_{k_i}}$, this implies the convergence $Y_{n_{k_i}} \to_d X$ by Theorem 1.5.4 of Van der Vaart and Wellner (2023). By characterization (iii) of Proposition 2.10 we obtain $X_n \leftrightarrow_d Y_n$. □

Lemma A.3. *The sequence X_n is relatively compact if and only if it is relatively asymptotically tight and $X_n(t)$ is asymptotically measurable for all $t \in T$.*

Proof. By Lemma A.1, X_n is relatively compact if and only if it is relatively asymptotically tight and asymptotically measurable. By definition, any sequence X_n is asymptotically measurable if and only if any subsequence n_k contains a further subsequences n_{k_i} such that $X_{n_{k_i}}$ is asymptotically measurable. By Lemma 1.5.2 of Van der Vaart and Wellner (2023) being asymptotically measurable is equivalent to $X_n(t)$ being asymptotically measurable for all $t \in T$ whenever X_n is relatively asymptotically tight. All together, this implies the equivalence. □

Corollary A.4 (Relative Cramer-Wold device)**.** *Let X_n and Y_n be two sequences of $\mathbb{R}^d$-valued random variables. If X_n is uniformly tight, then,*

$$X_n \leftrightarrow_d Y_n \text{ if and only if } t^T X_n \leftrightarrow_d t^T Y_n$$

for all $t \in \mathbb{R}^d$.

Proof. Restricting to functions of the form $f(t^T \cdot)$ with $f : \mathbb{R} \to \mathbb{R}$ bounded and continuous, the only if part follows by definition.

For the other direction, assume $t^T X_n \leftrightarrow_d t^T Y_n$ for all $t \in \mathbb{R}^d$. Note that $t^T X_n$ is uniformly tight for all $t \in \mathbb{R}^d$. We use characterization (iii) of Proposition 2.10. Let n_k be a subsequence. Since X_n is uniformly tight, there exists a subsequence $X_{n_{k_j}} \to_d X$. Then, also $t^T X_{n_{k_j}} \to_d t^T X$. By characterization (ii) of Proposition 2.10 and $t^T X_n \leftrightarrow_d t^T Y_n$, it follows $t^T Y_{n_{k_j}} \to_d t^T X$ for all $t \in \mathbb{R}^d$. By the Cramer-Wold device, we derive $Y_{n_{k_j}} \to_d X$. This proves the claim by characterization (iii) of Proposition 2.10. □

A.2. Relative central limit theorems

Proof of Corollary 2.18. Note that $(Y_n(t_1), \ldots, Y_n(t_k))$ satisfies a relative CLT if and only if

$$(Y_n(t_1), \ldots, Y_n(t_k)) \leftrightarrow_d (N_{Y_n}(t_1), \ldots, N_{Y_n}(t_k))$$

since corresponding Gaussians are unique in distribution. The necessity then follows by Lemma A.3 (for (ii)) and Theorem 2.15 (for (iii)).

For the sufficiency observe that $N_{Y_n}(t)$ and $Y_n(t)$ are measurable by assumption. Thus, (ii) is equivalent to Y_n and N_{Y_n} being relatively compact by Lemma A.3. Then, Theorem 2.15 provides the claim. □

Proof of Proposition 2.19. By Corollary A.6, there exists an asymptotically tight sequence of tight Borel measurable GPs corresponding to Y_n if and only if

$$\sup_{n \in \mathbb{N}, i \leq d} \mathbb{V}\mathrm{ar}[Y_n^{(i)}] < \infty.$$

Equivalently, if all subsequences n_k contain a further subsequence n_{k_i} such that $\Sigma_{n_{k_i}}$ converges. Combined with the fact that a sequence of centered Gaussians converges weakly if and only if its corresponding sequence of covariances converges, the equivalences follows from Proposition 2.10. □

A.3. Existence and tightness of corresponding GPs

Proposition A.5. *If there exists a relatively asymptotically tight sequence of tight and Borel measurable GPs corresponding to Y_n, then, every subsequence of n contains a further subsequence n_{k_i} such that $\mathbb{C}\mathrm{ov}[Y_{n_{k_i}}(t), Y_{n_{k_i}}(s)]$ converges for all $s, t \in T$.*

Proof. Denote by N_{Y_n} a relatively asymptotically tight sequence of GPs corresponding to Y_n. Any subsequence contains a further subsequence n_{k_i} such that $N_{Y_{n_{k_i}}}$ converges weakly to some tight GP. In particular, all marginals of $N_{Y_{n_{k_i}}}$ converge weakly. Recall that a sequence of centered multivariate Gaussians converges weakly if and only if their corresponding covariances converges. Thus, we obtain convergence of all covariances

$$\mathbb{C}\mathrm{ov}[N_{Y_{n_{k_i}}}(t), N_{Y_{n_{k_i}}}(s)] = \mathbb{C}\mathrm{ov}[Y_{n_{k_i}}(t), Y_{n_{k_i}}(s)].$$

□

Corollary A.6. *If T is finite, the following are equivalent:*

(i) there exists a relatively asymptotically tight sequence of tight and Borel measurable GPs corresponding to Y_n.

(ii) $\sup_n \mathbb{V}\mathrm{ar}[Y_n(t)] < \infty$ *for all $t \in T$.*

Proof. For the sufficiency, Proposition A.5 implies that all sequences of covariances $\mathbb{C}\mathrm{ov}[Y_n(s), Y_n(t)]$ are relatively compact or, equivalently, bounded.

For the necessaty, identify Y_n with $(Y_n(t_1), \ldots, Y_n(t_d))$ for $T = \{t_1, \ldots, t_d\}$. Construct $N_{Y_n} \sim \mathcal{N}(0, \Sigma_n)$ with Σ_n the covariance matrix of Y_n. Then, each N_{Y_n} is measurable and tight. Then, $\sup_n \mathbb{V}\mathrm{ar}[Y_n(t)] < \infty$ implies that all covariance $\mathbb{C}\mathrm{ov}[Y_n(s), Y_n(t)]$ are relatively compact. Thus, every subsequence n_k contains a further subsequence n_{k_i} such that all covariances $\mathbb{C}\mathrm{ov}[Y_{n_{k_i}}(s), Y_{n_{k_i}}(t)]$ converge. This is equivalent to the weak convergence of N_{Y_n}. We obtain relative compactness, hence, asymptotic tightness of N_{Y_n}. □

Proof of Proposition 2.20. By Kolmogorov's extension theorem, there exist centered GPs $\{N_{Y_n}(t) : t \in T\}$ with covariance function given by $(s,t) \mapsto \mathbb{C}\mathrm{ov}[Y_n(s), Y_n(t)]$. Since (T, ρ_n) is totally bounded (by finitenss of the covering numbers), (T, ρ_n) is separable and, thus there exists a separable version of $\{N_{Y_n}(t) : t \in T\}$ with the same marginal distributions (Section 2.3.3 of Van der Vaart and Wellner (2023)). Without loss of generality, assume that $\{N_{Y_n}(t) : t \in T\}$ is separable.

Then,

$$\mathbb{E}[\|N_{Y_n}\|_T] \le C \int_0^\infty \sqrt{\ln N(\epsilon, T, \rho_n)} d\epsilon < \infty$$

$$\mathbb{E}\left[\sup_{\rho_n(s,t) \le \delta} |N_{Y_n}(t) - N_{Y}{}_n(s)|\right] \le C \int_0^\delta \sqrt{\ln N(\epsilon, T, \rho_n)} d\epsilon$$

for some constant C by Corollary 2.2.9 of Van der Vaart and Wellner (2023). The first inequality implies that each $\{N_{Y_n}(t) : t \in T\}$ has bounded sample paths almost surely, hence, without loss of generality $\{N_{Y_n}(t) : t \in T\}$ induces a map N_{Y_n} with values in $\ell^\infty(T)$. The second inequality implies that all sample paths of N_{Y_n} are uniformly ρ_n-equicontinuous in probability. Hence, each N_{Y_n} is tight and Borel measurable (Example 1.5.10 of Van der Vaart and Wellner (2023)). □

Proof of Proposition 2.21. We derive

$$\mathbb{E}\left[\sup_{\rho_n(s,t)\leq\delta}|N_{Y_n}(t)-N_n(s)|\right]\leq C\int_0^\delta\sqrt{\ln N(\epsilon,T,\rho_n)}d\epsilon$$

for some constant C independent of n by Corollary 2.2.9 of Van der Vaart and Wellner (2023). By (iii), for every sequence $\delta\to 0$ there exists $\epsilon(\delta)\to 0$ such that $d(s,t)<\delta$ implies $\rho_n(s,t)<\epsilon(\delta)$. Accordingly,

$$\begin{aligned}\limsup_n \mathbb{E}\left[\sup_{d(s,t)\leq\delta}|N_{Y_n}(t)-N_{Y_n}(s)|\right] &\leq \limsup_n \mathbb{E}\left[\sup_{\rho_n(s,t)\leq\epsilon(\delta)}|N_{Y_n}(t)-N_n(s)|\right]\\ &\leq \limsup_n C\int_0^{\epsilon(\delta)}\sqrt{\ln N(\epsilon,T,\rho_n)}d\epsilon\end{aligned}$$

Taking the limit $\delta\to 0$ the right hand side converges to zero by (ii). Together with Markov's inequality, we obtain that N_{Y_n} is asymptotically uniformly d-equicontinuous in probability.

By Corollary A.6 for fixed $t\in T$ the sequence $N_{Y_n}(t)$ is relatively asymptotically tight for all $t\in T$ if $\sup_n \mathbb{V}\mathrm{ar}[Y_n(t)]<\infty$. Since this sequence is $\mathbb{R}$-valued, relative asymptotic tightness, relative compactness and asymptotic tightness agree. Then, Theorem 1.5.7 of Van der Vaart and Wellner (2023) proves that N_{Y_n} is asymptotically tight. □

A.4. Relative Lindeberg CLT

Theorem A.7 (Relative Lindeberg CLT). *Let $X_{n,1},\dots,X_{n,k_n}$ be a triangular array of independent random vectors with finite variance. Assume*

$$\frac{1}{k_n}\sum_{i=1}^{k_n}\mathbb{E}\left[\|X_{n,i}\|^2\mathbb{1}_{\{\|X_{n,i}\|^2>k_n\epsilon\}}\right]\to 0$$

for all $\varepsilon>0$ and for all $n\in\mathbb{N}, l=1,\dots,d$

$$\frac{1}{k_n}\sum_{i=1}^{k_n}\mathbb{V}\mathrm{ar}\left[X_{n,i}^{(l)}\right]\leq K\in\mathbb{R}.$$

Then, the scaled sample average $\sqrt{k_n}\left(\overline{X}_n-\mathbb{E}\left[\overline{X}_n\right]\right)$ satisfies a relative CLT.

Proof. Let k_{l_n} be a subsequence of k_n such that

$$\frac{1}{k_{l_n}}\sum_{i=1}^{k_{l_n}}\mathbb{C}\mathrm{ov}[X_{l_n,i}]\to\Sigma$$

converges. Observe that the Lindeberg condition implies

$$\frac{1}{k_{l_n}} \sum_{i=1}^{k_{l_n}} \mathbb{E}\left[\|X_{l_n,i}\|^2 1_{\{\|X_{l_n,i}\|^2 > k_{l_n}\epsilon\}}\right] \to 0$$

for all $\epsilon > 0$. We apply Proposition 2.27 of van der Vaart (2000) to the triangular array $Y_{n,1}, \ldots Y_{n,k_{l_n}}$ with $Y_{n,i} = k_{l_n}^{-1/2} X_{l_n,i}$ to derive

$$\sqrt{k_{l_n}}\left(\overline{X}_{l_n} - \mathbb{E}\left[\overline{X}_{l_n}\right]\right) \to_d \mathcal{N}(0, \Sigma).$$

By (ii) of Proposition 2.19 we derive the claim. □

B. Tightness under bracketing entropy conditions

The proof of Theorem 3.4 is based on a long sequence of well-known arguments: We group the observations $X_{n,i}$ in alternating blocks of equal size and apply maximal coupling. This yields random variables $X^*_{n,i}$ which corresponding blocks are independent and we obtain the empirical process $\mathbb{G}^*_n$ where $X_{n,i}$ is replaced by $X^*_{n,i}$. Because $\mathbb{G}^*_n$ consists of independent blocks, we can derive a Bernstein type inequality bounding the first moment of $\mathbb{G}^*_n$ in terms of $\mathcal{F}_n$ provided that $\mathcal{F}_n$ is finite (Lemma B.2). For any fixed n, we use a chaining argument in order to reduce to finite $\mathcal{F}_n$ which, in combination with the Bernstein inequality, yields a bound of the first moment of $\mathbb{G}^*_n$ in terms of the bracketing entropy (Theorem B.6). Under the conditions of Theorem 3.4, this yields asymptotic equicontinuity of $\mathbb{G}_n$ which implies relative compactness of $\mathbb{G}_n$.

B.1. Coupling

Let m_n be a sequence in $\mathbb{N}$. Suppose for simplicity that k_n is a multiple of $2m_n$ and group the observations $X_{n,1}, \ldots, X_{n,k_n}$ in alternating blocks of size m_n. By maximal coupling (Rio, 2017, Theorem 5.1), there are random vectors

$$U^*_{n,j} = \left(X^*_{n,(j-1)m_n+1}, \ldots, X^*_{n,jm_n}\right) \in \mathcal{X}^n$$

such that

- $U_{n,j} = \left(X_{n,(j-1)m_n+1}, \ldots, X_{n,jm_n}\right) \overset{d}{=} U^*_{n,j}$ for every $j = 1, \ldots, m_n$,
- each of the sequences $\left(U^*_{n,2j}\right)_{j=1,\ldots,n/(2m_n)}$ and $\left(U^*_{n,2j-1}\right)_{j=1,\ldots,n/(2m_n)}$ are independent,
- $P\left(X_{n,j} \neq X^*_{n,j}\right) \leq \beta_n\left(m_n\right)$ for all j,
- $P\left(\exists j \colon U_{n,j} \neq U^*_{n,j}\right) \leq \left(k_n/m_n\right)\beta_n\left(m_n\right)$.

Define the coupled empirical process $\mathbb{G}_n^* \in \ell^\infty(T)$ as $\mathbb{G}_n$, but with all $X_{n,j}$ replaced by $X_{n,j}^*$.

In what follows, we will replace $\mathbb{E}^*$ (indicating the outer expectation) by $\mathbb{E}$ for better readability. We will provide an upper bound on $\mathbb{E}\|\mathbb{G}_n^*\|_T$ for any fixed n. In such case, we identify $\mathbb{G}_n^*$ with the empirical process

$$\mathbb{G}_n^*(f) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} f\left(X_{n,i}^*\right) - \mathbb{E}\left[f\left(X_{n,i}^*\right)\right]$$

indexed by the function class $\mathcal{F}_n$. For fixed n, we drop the index n, i.e., write $\mathcal{F}_n = \mathcal{F}$, $X_{n,i} = X_i$, $m_n = m$ etc. and assume without loss of generality that $k_n = n$.

Lemma B.1. *For any class $\mathcal{F}$ of functions $f\colon \mathcal{X} \to \mathbb{R}$ with $\sup_{f\in\mathcal{F}} \|f\|_\infty \le B$ and any integer $1 \le m \le n/2$, it holds*

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \le \mathbb{E}\|\mathbb{G}_n^*\|_{\mathcal{F}} + B\sqrt{n}\beta_n(m).$$

Proof. We have

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \le \mathbb{E}\sup_{f\in\mathcal{F}} |\mathbb{G}_n^* f| + \frac{B}{\sqrt{n}}\mathbb{E}\left[\sum_{i=1}^n \mathbb{1}(X_j \ne X_j^*)\right] \le \mathbb{E}\|\mathbb{G}_n^*\|_{\mathcal{F}} + B\sqrt{n}\beta_n(m).$$

□

B.2. Bernstein inequality

Lemma B.2. *Let $\mathcal{F}$ be a finite set of functions $f\colon \mathcal{X} \to \mathbb{R}$ with*

$$\|f\|_\infty \le B, \quad \frac{1}{n}\sum_{i,j=1}^n |\mathbb{C}\mathrm{ov}[f(X_i), f(X_j)]| \le K\delta^2$$

for all $f \in \mathcal{F}$. Then, for any $1 \le m \le n/2$ it holds

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \delta\sqrt{\ln_+(|\mathcal{F}|)} + \frac{mB\ln_+(|\mathcal{F}|)}{\sqrt{n}} + B\sqrt{n}\beta_n(m)$$

where the constant only depends on K and $\ln_+(x) = \ln(1+x)$.

Proof. We have

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \le \mathbb{E}\|\mathbb{G}_n^*\|_{\mathcal{F}} + B\sqrt{n}\beta_n(m),$$

by Lemma B.1.

Defining

$$A_{j,f} = \sum_{i=1}^{m} f\left(X^*_{(j-1)m+i}\right) - \mathbb{E}\left[f\left(X^*_{(j-1)m+i}\right)\right],$$

we can write

$$\sqrt{n}|\mathbb{G}^*_n(f)| = \left|\sum_{j=1}^{n/m} A_{j,f}\right| \leq \left|\sum_{j=1}^{n/(2m)} A_{2j,f}\right| + \left|\sum_{j=1}^{n/(2m)} A_{2j-1,f}\right|.$$

The random variables in the sequence $(A_{2j,f})_{j=1}^{n/(2m)}$ are independent and so are those in $(A_{2j-1,f})_{j=1}^{n/(2m)}$. We apply Bernstein's inequality (Van der Vaart and Wellner, 2023, Lemma 2.2.10). Note that $|A_{j,f}| \leq 2mB$, hence $\mathbb{E}\left[|A_{j,f}|^k\right] \leq (2mB)^{k-2}\,\mathbb{V}\mathrm{ar}\left[A_{j,f}\right]$ for $k \geq 2$. We obtain

$$\begin{aligned}
\frac{2m}{n}\sum_{i=1}^{n/(2m)} \mathbb{E}\left[|A_{j,f}|^k\right] &\leq (2mB)^{k-2}\,\frac{2m}{n}\sum_{i=1}^{n/(2m)} \mathbb{V}\mathrm{ar}\left[A_{j,f}\right] \\
&\leq (2mB)^{k-2}\,\frac{2m}{n}\sum_{i=1}^{n} |\mathbb{C}\mathrm{ov}[f(X_i), f(X_j)]| \\
&\leq (2mB)^{k-2}\, 2Km\delta^2.
\end{aligned}$$

Using Bernstein's inequality for independent random variables gives

$$P\left(\left|\sum_{j=1}^{n/(2m)} A_{2j,f}\right| > t\right) \leq 2\exp\left(-\frac{1}{2}\frac{t^2}{Kn\delta^2 + 2tmB}\right)$$

and the same bounds holds for the odd sums. Altogether we get

$$\begin{aligned}
P\left(|\mathbb{G}^*_n(f)| > t\right) &\leq P\left(\left|\sum_{j=1}^{n/(2m)} A_{2j,f}\right| > t\sqrt{n}/2\right) + P\left(\left|\sum_{j=1}^{n/(2m)} A_{2j-1,f}\right| > t\sqrt{n}/2\right) \\
&\leq 4\exp\left(-\frac{1}{8}\frac{t^2}{K\delta^2 + tmB/\sqrt{n}}\right)
\end{aligned}$$

The result follows upon converting this to a bound on the expectation (e.g., Van der Vaart and Wellner, 2023, Lemma 2.2.13). □

B.3. Chaining

We will abbreviate

$$\|f\|_{a,n} = \left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[|f\left(X_i\right)|^a\right]\right)^{1/a}$$

$$N_{[]}(\epsilon) = N_{[]}\left(\epsilon, \mathcal{F}, \|\cdot\|_{\gamma,n}\right).$$

Let us first collect some properties of $\|\cdot\|_{\gamma,n}$ and $\|\cdot\|_{\gamma,\infty}$.

Lemma B.3. *The following holds:*

(i) $\|\cdot\|_{a,n}$ *defines a semi-norm.*

(ii) $\|\cdot\|_{a,n} \leq \|\cdot\|_{b,n}$ *for* $a \leq b$.

(iii) $\|f\mathbb{1}_{|f|>K}\|_{a,n} \leq K^{a-b}\|f\|_{b,n}^{b/a}$ *for all* $K > 0$ *and* $a \leq b$.

Proof. Positivity and homogeinity of $\|\cdot\|_{\gamma,n}$ follow clearly and the triangle inequality follows by

$$\begin{aligned}
\|h+g\|_{\gamma,n} &= \left(\frac{1}{n}\sum_{i=1}^{n}\|(h+g)(X_i)\|_\gamma^\gamma\right)^{1/\gamma} \\
&\leq \left(\frac{1}{n}\sum_{i=1}^{n}\left[\|h(X_i)\|_\gamma + \|g(X_i)\|_\gamma\right]^\gamma\right)^{1/\gamma} \\
&\leq \left(\frac{1}{n}\sum_{i=1}^{n}\|h(X_i)\|^\gamma\right)^{1/\gamma} + \left(\frac{1}{n}\sum_{i=1}^{n}\|g(X_i)\|_\gamma^\gamma\right)^{1/\gamma} \qquad \text{Minkowski's inequality} \\
&= \|h\|_{\gamma,n} + \|g\|_{\gamma,n}
\end{aligned}$$

for all h, g. Next,

$$\begin{aligned}
\|f\|_{a,n} &= \left(\frac{1}{n}\sum_{i=1}^{n}\|f\left(X_i\right)\|_a^a\right)^{1/a} \\
&\leq \left(\frac{1}{n}\sum_{i=1}^{n}\|f\left(X_i\right)\|_b^a\right)^{1/a} \\
&\leq \left(\frac{1}{n}\sum_{i=1}^{n}\|f\left(X_i\right)\|_b^b\right)^{1/b} \qquad \text{by Jensen's inequality.}
\end{aligned}$$

Lastly, note

$$K^{b-a}|f(X_i)\mathbb{1}_{|f(X_i)|>K}|^a \leq |f(X_i)|^b.$$

Thus,

$$\begin{aligned}
\|f\mathbb{1}_{|f|\geq K}\|_{a,n} &= \left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[|f(X_i)\mathbb{1}_{|f(X_i)|>K}|^a\right]\right)^{1/a} \\
&\leq K^{a-b}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[|f(X_i)|^b\right]\right)^{1/a} \\
&= K^{a-b}\|f\|_{b,n}^{b/a}
\end{aligned}$$

□

Theorem B.4. *Let $\mathcal{F}$ be a class of functions $f\colon \mathcal{X} \to \mathbb{R}$ with envelope F and for some $\gamma \geq 2$,*

$$\|f\|_{\gamma,n} \leq \delta \qquad \frac{1}{n}\sum_{i,j=1}^{n} |\mathbb{C}\mathrm{ov}[h(X_i), h(X_j)]| \leq K_1 \|h\|_{\gamma,n}^2$$

for all $f \in \mathcal{F}$ and $h : \mathcal{X} \to \mathbb{R}$ bounded and measurable. Suppose that $\max_m \beta_n(m) \leq K_2 m^{-\rho}$ for some $\rho \geq \gamma/(\gamma-2)$. Then, for any $n \geq 5$, $m \geq 1, B \in (0,\infty)$,

$$\begin{aligned}\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim &\int_0^\delta \sqrt{\ln_+ N_{[]}(\epsilon)}d\epsilon \\ &+ \frac{mB \ln_+ N_{[]}(\delta)}{\sqrt{n}} + \sqrt{n}B\beta_n(m) + \sqrt{n}\|F\mathbb{1}\{F > B\}\|_{1,n} + \sqrt{n}N_{[]}^{-1}(e^n)\end{aligned}$$

with constants only depending on K_1, K_2. If the integral is finite, then $\sqrt{n}N_{[]}^{-1}(e^n) \to 0$ for $n \to \infty$.

Let us first derive some useful corollaries.

Corollary B.5. *Let $\mathcal{F}$ be a class of functions $f\colon \mathcal{X} \to \mathbb{R}$ with envelope F, $X_{n,i}$ are independent and $\|f\|_{2,n} \leq \delta$ for all $f \in \mathcal{F}$. Then, for any $n \geq 5$, $B \in (0,\infty)$,*

$$\begin{aligned}\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim &\int_0^\delta \sqrt{\ln_+ N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{2,n})}d\epsilon \\ &+ \frac{B \ln_+ N_{[]}(\delta, \mathcal{F}, \|\cdot\|_{2,n})}{\sqrt{n}} + \sqrt{n}\|F\mathbb{1}\{F > B\}\|_{1,n} + \sqrt{n}N_{[]}^{-1}(e^n)\end{aligned}$$

with constants only depending on K_1, K_2.

Proof. It holds

$$\frac{1}{n}\sum_{i,j=1}^{n} |\mathbb{C}\mathrm{ov}[h(X_i), h(X_j)]| = \frac{1}{n}\sum_{i=1}^{n} \mathbb{V}\mathrm{ar}[h(X_i)] \leq 2\|h\|_{2,n}$$

and the β-coefficients are 0 for all $m \geq 1$. Applying Theorem B.4 with $m = 1$ yields the claim. $\square$

Theorem B.6. *Let $\mathcal{F}$ be a class of functions $f\colon \mathcal{X} \to \mathbb{R}$ with envelope F and for some $\gamma > 2$,*

$$\|f\|_{\gamma,n} \leq \delta \qquad \frac{1}{n}\sum_{i,j=1}^{n} |\mathbb{C}\mathrm{ov}[h(X_i), h(X_j)]| \leq K_1 \|h\|_{\gamma,n}^2$$

for all $f \in \mathcal{F}$ and $h : \mathcal{X} \to \mathbb{R}$ bounded and measurable. Suppose that $\max_m \beta_n(m) \leq K_2 m^{-\rho}$ *for some $\rho \geq \gamma/(\gamma-2)$. Then, for any $n \geq 5$,*

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \int_0^\delta \sqrt{\ln_+ N_{[]}(\epsilon)} d\epsilon + \frac{\|F\|_{\gamma,n}[\ln N_{[]}(\delta)]^{[1-1/(\rho+1)](1-1/\gamma)}}{n^{-1/2+[1-1/(\rho+1)](1-1/\gamma)}} + \sqrt{n} N_{[]}^{-1}(e^n).$$

with constants only depending on K_1, K_2.

In particular, if the integral is finite, $\|F\|_{\gamma,\infty} < \infty$, $\rho > \gamma/(\gamma-2)$ and K_1, K_2 can be chosen independent of n, then

$$\limsup_{n\to\infty} \mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \int_0^\delta \sqrt{\ln N_{[]}(\epsilon)} d\epsilon.$$

Proof. It holds

$$\|F\mathbb{1}\{F > B\}\|_{1,n} \leq \frac{\sqrt{n}\|F\|_{\gamma,n}^\gamma}{B^{\gamma-1}}.$$

By Theorem B.4

$$\mathbb{E}[\|\mathbb{G}_n\|_{\mathcal{F}}] \lesssim \int_0^\delta \sqrt{\ln_+ N_{[]}(\epsilon)} d\epsilon + \frac{mB\ln_+ N_{[]}(\delta)}{\sqrt{n}} + \sqrt{n}B\beta_n(m) + \frac{\sqrt{n}\|F\|_{\gamma,n}^\gamma}{B^{\gamma-1}} + \sqrt{n}N_{[]}^{-1}(e^n).$$

Choose $m = (n/\ln_+ N_{[]}(\delta))^{1/(\rho+1)}$, which gives

$$\frac{m\ln_+ N_{[]}(\delta)}{\sqrt{n}} + \sqrt{n}\beta_n(m) \lesssim n^{-1/2+1/(\rho+1)}[\ln_+ N_{[]}(\delta)]^{1-1/(\rho+1)},$$

and, thus,

$$\mathbb{E}[\|\mathbb{G}_n\|_{\mathcal{F}}] \lesssim \int_0^\delta \sqrt{\ln_+ N_{[]}(\epsilon)} d\epsilon + \frac{B[\ln_+ N_{[]}(\delta)]^{1-1/(\rho+1)}}{n^{1/2-1/(\rho+1)}} + \frac{\sqrt{n}\|F\|_{\gamma,n}^\gamma}{B^{\gamma-1}} + \sqrt{n}N_{[]}^{-1}(e^n).$$

Next, choose

$$B = \left(\frac{n^{[1-1/(\rho+1)]}\|F\|_{\gamma,n}^\gamma}{[\ln_+ N_{[]}(\delta)]^{1-1/(\rho+1)}}\right)^{1/\gamma}$$

This gives

$$\frac{B[\ln_+ N_{[]}(\delta)]^{1-1/(\rho+1)}}{n^{1/2-1/(\rho+1)}} + \frac{\sqrt{n}\|F\|_{\gamma,n}^\gamma}{B^{\gamma-1}} = \frac{\|F\|_{\gamma,n}[\ln_+ N_{[]}(\delta)]^{[1-1/(\rho+1)](1-1/\gamma)}}{n^{-1/2+[1-1/(\rho+1)](1-1/\gamma)}}.$$

Lastly, $\rho > \gamma/(\gamma-2)$, then

$$-1/2 + [1 - 1/(\rho+1)](1-1/\gamma) > 0,$$

so the second term in the first statement vanishes as $n \to \infty$ and the last term vanishes since the bracketing integral is finite. □

Proof of Theorem B.4. Let us first deduce the last statement. If the bracketing integral exists, then it must hold $\sqrt{\ln_+ N_{[]}(\delta)} \lesssim \delta^{-1}/(1+|\ln(\delta)|)$ for $\delta \to 0$, because the upper bound is not integrable. For $\delta^{-1} = \sqrt{n\ln n}$, we have $\ln_+ N_{[]}(\delta) \lesssim n\ln n/(1+\ln n)^2 = o(n)$. So for large n, it must hold $N_{[]}^{-1}(e^n) \lesssim 1/\sqrt{n\ln n} \to 0$.

We now turn to the proof of the first statement.

Truncation

We first truncate the function class $\mathcal{F}$ in order to apply Bernstein's inequality in combination with a chaining argument. It holds

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \leq \mathbb{E}\big[\|\mathbb{G}_n(f\mathbb{1}\{|F| \leq B\})\|_{\mathcal{F}}\big] + \mathbb{E}\big[\|\mathbb{G}_n\left(f\mathbb{1}\{|F| > B\}\right)\|_{\mathcal{F}}\big],$$

and

$$\mathbb{E}\big[\|\mathbb{G}_n(f\mathbb{1}\{|F| > B\})\|_{\mathcal{F}}\big] \leq 2\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}\big[F(X_i)\mathbb{1}\{|F(X_i)| > B\}\big] = 2\sqrt{n}\|F\mathbb{1}\{F > B\}\|_{1,n}.$$

because In summary,

$$\mathbb{E}\|\mathbb{G}_n(f)\|_{\mathcal{F}} \leq \mathbb{E}\big[\|\mathbb{G}_n(f\mathbb{1}\{|F| \leq B\})\|_{\mathcal{F}}\big] + 2\sqrt{n}\|F\mathbb{1}\{F > B\}\|_{1,n}.$$

Note that $|f\mathbb{1}\{|F| \leq B\}| \leq F\mathbb{1}\{|F| \leq B\} \leq B$. By replacing $\mathcal{F}$ with

$$\mathcal{F}_{trun} = \left\{f\mathbb{1}\left\{|F| \leq B\right\} : f \in \mathcal{F}\right\},$$

we may without loss of generality assume that $\mathcal{F}$ has an envelope with $\|F\|_\infty \leq B$. Observe that the conditions of the theorem remain true for $\mathcal{F}_{trun}$ and that the bracketing numbers with respect to $\mathcal{F}_{trun}$ are bounded above by the bracketing numbers with respect to $\mathcal{F}$.

Lastly, we may assume $\ln_+ N_{r_0} \leq n$: If $n < \ln_+ N_{r_0} \leq \ln_+ N_{[]}(\delta)$ then

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \sqrt{n}B \leq \frac{mB\ln_+ N_{[]}(\delta)}{\sqrt{n}}$$

which still implies the claim.

Chaining setup

Fix integers $r_0 \leq r_1$ such that $2^{-r_0-1} < \delta \leq 2^{-r_0}$. For $r \geq r_0$ we construct a nested sequence of partitions $\mathcal{F} = \bigcup_{k=1}^{N_r}\mathcal{F}_{r,k}$ of $\mathcal{F}$ into N_r disjoint subsets such that for each $r \geq r_0$

$$\left\|\sup_{f,f'\in\mathcal{F}_{r,k}}|f - f'|\right\|_{\gamma,n} < 2^{-r}.$$

Clearly, we can choose the partition such that

$$N_{r_0} \leq N_{[]}\left(2^{-r_0}\right) \leq N_{[]}\left(\delta\right).$$

As explained in the proof of Theorem 2.5.8 of Van der Vaart and Wellner (2023), we may assume without loss of generality that

$$\sqrt{\ln_+ N_r} \leq \sum_{k=r_0}^{r}\sqrt{\ln_+ N_{[]}\left(2^{-k}\right)}.$$

Then by reindexing the double sum,

$$\begin{aligned}
\sum_{r=r_0}^{r_1} 2^{-r}\sqrt{\ln_+ N_r} &\leq \sum_{r=r_0}^{r_1} 2^{-r} \sum_{k=r_0}^{r} \sqrt{\ln_+ N_{[]}\left(2^{-k}\right)} \\
&= \sum_{k=r_0}^{r_1} \sqrt{\ln_+ N_{[]}\left(2^{-k}\right)} \sum_{r=k}^{r_1} 2^{-r} \\
&= \sum_{k=r_0}^{r_1} 2^{-k}\sqrt{\ln_+ N_{[]}\left(2^{-k}\right)} \sum_{r=k}^{r_1} 2^{-(r-k)} \\
&\lesssim \sum_{k=r_0}^{r_1} 2^{-k}\sqrt{\ln_+ N_{[]}\left(2^{-k}\right)} \\
&\lesssim \int_0^{\delta} \sqrt{\ln_+ N_{[]}(\epsilon)} d\epsilon.
\end{aligned}$$

Decomposition

For a given f, suppose that $\mathcal{F}_{r,k}$ is the element of the partition that contains f. Note that such $\mathcal{F}_{r,k}$ is unique since all $\mathcal{F}_{r,1}, \ldots, \mathcal{F}_{r,N_r}$ are disjoint. Define $\pi_r(f)$ as some fixed element of this set and define

$$\Delta_r(f) = \sup_{f_1, f_2 \in \mathcal{F}_{r,k}} |f_1 - f_2|.$$

Set

$$\tau_r = \frac{2^{-r}}{m_{r+1}}\sqrt{\frac{n}{\ln_+ N_{r+1}}}, \quad m_r = \min\left\{\sqrt{\frac{\ln_+ N_r}{n}}, 1\right\}^{-(\gamma-2)/(\gamma-1)}, \tag{4}$$

and

$$r_1 = -\log_2 N_{[]}^{-1}(e^n).$$

Then $\ln_+ N_r \leq n$ for all $r \leq r_1$. We will frequently apply Bernstein's inequality with $m = m_r$. Here, note

$$m_r \leq \sqrt{n/\ln_+ N_r} \leq \sqrt{n/\ln(2)} \leq n/2$$

for all $n \geq 5$.

The following (in-)equalities are the reason for the choices of τ_r and m_r: for r such that $\sqrt{\ln_+ N_{r+1}} \le n$, i.e., $r < r_1$ it holds

$$\begin{aligned}
\frac{m_r \tau_{r-1}}{\sqrt{n}} &= 2^{-r+1}\sqrt{\ln_+ N_r} \\
\sqrt{n}2^{-r\gamma}\tau_r^{-(\gamma-1)} &= 2^{-r}\sqrt{n}\left(\frac{1}{m_r}\sqrt{\frac{n}{\ln_+ N_{r+1}}}\right)^{-(\gamma-1)} \\
&= 2^{-r}\left(\sqrt{n}\right)^{2-\gamma}\left(\sqrt{\ln_+ N_{r+1}}\right)^{\gamma-1} m_{r+1}^{\gamma-1} \\
&= 2^{-r}\sqrt{\ln_+ N_{r+1}}\left(\sqrt{\frac{\ln_+ N_{r+1}}{n}}\right)^{\gamma-2} m_{r+1}^{\gamma-1} \\
&= 2^{-r}\sqrt{\ln_+ N_{r+1}} \\
\sqrt{n}\tau_{r-1}\beta_n(m_r) &= \sqrt{n}2^{-r+1}\frac{1}{m_r}\sqrt{\frac{n}{\ln_+ N_r}}\beta_n(m_r) \\
&\lesssim 2^{-r+1}\frac{1}{m_r}\frac{n}{\sqrt{\ln_+ N_r}} m_r^{-\rho} \\
&= 2^{-r+1}\sqrt{\ln_+ N_r}\frac{n}{\ln_+ N_r} m_r^{-\rho-1} \\
&= 2^{-r+1}\sqrt{\ln_+ N_r}\, m_r^{\frac{2(\gamma-1)}{\gamma-2}} m_r^{-\rho-1} \\
&\le 2^{-r+1}\sqrt{\ln_+ N_r}
\end{aligned}$$

where the last inequality holds because $1 \le m_r$ and $\gamma/(\gamma-2) \ge \rho$, hence, $m_r^{\frac{2(\gamma-1)}{\gamma-2}-\rho-1} \le 1$.

Decompose

$$\begin{aligned}
f &= \pi_{r_0}(f) + [f - \pi_{r_0}(f)]\,\mathbb{1}\{\Delta_{r_0}(f)/\tau_{r_0} > 1\} \\
&\quad + \sum_{r=r_0+1}^{r_1} [f - \pi_r(f)]\,\mathbb{1}\left\{\max_{r_0 \le k < r} \Delta_k(f)/\tau_k \le 1, \Delta_r(f)/\tau_r > 1\right\} \\
&\quad + \sum_{r=r_0+1}^{r_1} [\pi_r(f) - \pi_{r-1}(f)]\,\mathbb{1}\left\{\max_{r_0 \le k < r} \Delta_k(f)/\tau_k \le 1\right\} \\
&\quad + [f - \pi_{r_1}(f)]\,\mathbb{1}\left\{\max_{r_0 \le k \le r_1} \Delta_k(f)/\tau_k \le 1\right\} \\
&= T_1(f) + T_2(f) + T_3(f) + T_4(f).
\end{aligned}$$

To see this, note that if $\Delta_{r_0}(f)/\tau_{r_0} > 1$ all terms but $T_1(f)$ vanish and $T_1(f) = f$. Otherwise, define $\hat{r}$ as the maximal number $r_0 \le r \le r_1$ such that $\max_{r_0 \le k \le r} \Delta_k(f)/\tau_k \le 1$. Then,

$$T_1(f) = \pi_{r_0}(f)$$

and if $\hat{r} < r_1$, then,

$$T_2(f) = f - \pi_{\hat{r}+1}(f) \qquad T_3(f) = \pi_{\hat{r}+1}(f) - \pi_{r_0}(f) \qquad T_4(f) = 0.$$

If $\hat{r} = r_1$, then,

$$T_2(f) = 0 \qquad T_3(f) = \pi_{r_1}(f) - \pi_{r_0}(f) \qquad T_4(f) = f - \pi_{r_1}(f).$$

We prove the theorem by separately bounding the four terms $\mathbb{E}\|\mathbb{G}_n T_j\|_{\mathcal{F}}$. Note that $\mathbb{G}_n$ is additive by construction, i.e., $\mathbb{G}_n(f+g) = \mathbb{G}_n(f) + \mathbb{G}_n(g)$.

Bounding T_1

Note that for every $|g| \leq h$ it follows

$$|\mathbb{G}_n(g)| \leq |\mathbb{G}_n(h)| + 2\sqrt{n}\|h\|_{1,n}.$$

In combination with the triangle inequality we obtain

$$\|\mathbb{G}_n T_1\|_{\mathcal{F}} \leq \|\mathbb{G}_n \pi_{r_0}\|_{\mathcal{F}} + \|\mathbb{G}_n \Delta_{r_0}\|_{\mathcal{F}} + 2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_{r_0}(f)\mathbb{1}\{\Delta_{r_0}(f)/\tau_{r_0} > 1\}\|_{1,n}.$$

The sets $\{\Delta_{r_0}(f) : f \in \mathcal{F}\}$ and $\{\pi_{r_0}(f) : f \in \mathcal{F}\}$ contain at most N_{r_0} different functions each. The construction implies

$$\|\pi_{r_0}(f)\|_{\gamma,n} \leq \delta, \quad \|\pi_{r_0}(f)\|_\infty \lesssim B, \quad \|\Delta_{r_0}(f)\|_{\gamma,n} \leq 2\delta, \quad \|\Delta_{r_0}(f)\|_\infty \lesssim B.$$

Now the Bernstein bound from Lemma B.2 gives

$$\begin{aligned}\mathbb{E}\|\mathbb{G}_n \pi_{r_0}\|_{\mathcal{F}} + \mathbb{E}\|\mathbb{G}_n \Delta_{r_0}\|_{\mathcal{F}} &\lesssim \delta\sqrt{\ln_+ N_{r_0}} + \frac{mB}{n}\ln_+ N_{r_0} + \sqrt{n}B\beta_n(m).\\ &\leq \delta\sqrt{\ln_+ N_{[]}(\delta)} + \frac{mB}{n}\ln_+ N_{[]}(\delta) + \sqrt{n}B\beta_n(m).\end{aligned}$$

Since the bracketing numbers are decreasing,

$$\sqrt{\ln_+ N_{[]}(\delta)} \leq \delta^{-1}\int_0^\delta \sqrt{\ln_+ N_{[]}(\epsilon)}d\epsilon$$

so

$$\mathbb{E}\|\mathbb{G}_n \pi_{r_0}\|_{\mathcal{F}} + \mathbb{E}\|\mathbb{G}_n \Delta_{r_0}\|_{\mathcal{F}} \lesssim \int_0^\delta \sqrt{\ln_+ N_{[]}(\epsilon)}d\epsilon + \frac{mB\ln_+ N_{[]}(\delta)}{\sqrt{n}} + \sqrt{n}B\beta_n(m).$$

Recall $\ln_+ N_{r+1} \leq n$ for any $r < r_1$. For any such r, (iii) of Lemma B.3 gives

$$\begin{aligned}\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_r(f)\mathbb{1}\{\Delta_r(f)/\tau_r > 1\}\|_{1,n} &\leq \sqrt{n}\tau_r^{-(\gamma-1)}\sup_{f\in\mathcal{F}}\|\Delta_r(f)\|_{\gamma,n}^{\gamma}\\ &\leq \sqrt{n}\tau_r^{-(\gamma-1)}2^{-r\gamma}\end{aligned}$$

so that the final upper bound becomes

$$\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_r(f)\mathbb{1}\{\Delta_r(f)/\tau_r > 1\}\|_{1,n} \lesssim 2^{-r}\sqrt{\ln_+ N_{r+1}}, \tag{5}$$

for any $r < r_1$. In particular, using $\delta \leq 2^{-r_0}$, we get

$$\sqrt{n} \sup_{f \in \mathcal{F}} \|\Delta_{r_0}(f) \mathbb{1}\{\Delta_{r_0}(f)/\tau_{r_0} > 1\}\|_{1,n} \lesssim \delta \sqrt{\ln_+ N_{[]}(\delta)} \leq \int_0^\delta \sqrt{\ln_+ N_{[]}(\epsilon)} d\epsilon.$$

Combined,

$$\mathbb{E}\|\mathbb{G}_n T_1\|_{\mathcal{F}} \lesssim \int_0^\delta \sqrt{\ln_+ N_{[]}(\epsilon)} d\epsilon + \frac{mB \ln_+ N_{[]}(\delta)}{\sqrt{n}} + \sqrt{n} B \beta_n(m).$$

Bounding T_2

Next,

$$\begin{aligned}
\mathbb{E}\|\mathbb{G}_n T_2\|_{\mathcal{F}} &\leq \sum_{r=r_0+1}^{r_1} \mathbb{E} \left\| \mathbb{G}_n \Delta_r \mathbb{1}\left\{ \max_{r_0 \leq k < r} \Delta_k/\tau_k \leq 1, \Delta_r/\tau_r > 1 \right\} \right\|_{\mathcal{F}} \\
&\quad + 2\sqrt{n} \sum_{r=r_0+1}^{r_1} \sup_{f \in \mathcal{F}} \left\| \Delta_r(f) \mathbb{1}\left\{ \max_{r_0 \leq k < r} \Delta_k(f)/\tau_k \leq 1, \Delta_r/\tau_r > 1 \right\} \right\|_{1,n} \\
&= T_{2,1} + T_{2,2}.
\end{aligned}$$

We start by bounding the first term. It holds

$$\left\| \Delta_r(f) \mathbb{1}\left\{ \max_{r_0 \leq k < r} \Delta_k(f)/\tau_k \leq 1, \Delta_r(f)/\tau_r > 1 \right\} \right\|_{\gamma,n} \leq 2^{-r}$$

by construction of $\Delta_r(f)$. Since the partitions are nested $\Delta_r \leq \Delta_{r-1}$. Thus,

$$\left\| \Delta_r \mathbb{1}\left\{ \max_{r_0 \leq k < r} \Delta_k/\tau_k \leq 1, \Delta_r/\tau_r > 1 \right\} \right\|_{\mathcal{F}} \leq \tau_{r-1}.$$

Since there are at most N_r functions in $\{\Delta_r(f) : f \in \mathcal{F}\}$, the Bernstein bound from Lemma B.2 yields

$$\begin{aligned}
T_{2,1} &\lesssim \sum_{r=r_0+1}^{r_1} \left[2^{-r} \sqrt{\ln_+ N_r} + \frac{m_r \tau_{r-1}}{\sqrt{n}} \ln_+ N_r + \sqrt{n} \tau_{r-1} \beta_n(m_r) \right] \\
&\lesssim \sum_{r=r_0+1}^{r_1} 2^{-r} \sqrt{\ln_+ N_r} \\
&\lesssim \int_0^\delta \sqrt{\ln_+ N_{[]}(\epsilon)} d\epsilon.
\end{aligned}$$

Further, (5) gives

$$\sqrt{n} \sup_{f \in \mathcal{F}} \left\| \Delta_r(f) \mathbb{1}\left\{ \max_{r_0 \leq k < r} \Delta_k(f)/\tau_k \leq 1, \Delta_r/\tau_r > 1 \right\} \right\|_{1,n} \lesssim 2^{-r} \sqrt{\ln_+ N_r}.$$

for $r < r_1$ and

$$\sqrt{n} \sup_{f \in \mathcal{F}} \left\| \Delta_{r_1}(f) \mathbb{1} \left\{ \max_{r_0 \leq k < r_1} \Delta_k(f)/\tau_k \leq 1, \Delta_{r_1}/\tau_{r_1} > 1 \right\} \right\|_{1,n} \leq \sqrt{n} \sup_{f \in \mathcal{F}} \|\Delta_{r_1}(f)\|_{\gamma,n}^{\gamma}$$
$$\leq \sqrt{n} 2^{-r_1}$$
$$= \sqrt{n} N_{[]}^{-1}(e^n)$$

by the definition of r_1. Thus,

$$T_{2,2} \lesssim \sum_{r=r_0+1}^{r_1-1} 2^{-r} \sqrt{\ln_+ N_r} + \sqrt{n} N_{[]}^{-1}(e^n) \lesssim \int_0^{\delta} \sqrt{\ln_+ N_{[]}(\epsilon)} d\epsilon + \sqrt{n} N_{[]}^{-1}(e^n),$$

and, in summary,

$$\mathbb{E}\|\mathbb{G}_n T_2\|_{\mathcal{F}} \lesssim \int_0^{\delta} \sqrt{\ln_+ N_{[]}(\epsilon)} d\epsilon + \sqrt{n} N_{[]}^{-1}(e^n).$$

Bounding T_3

Next,

$$\mathbb{E}\|\mathbb{G}_n T_3\|_{\mathcal{F}} \leq \sum_{r=r_0+1}^{r_1} \mathbb{E} \left\| \mathbb{G}_n \left[\pi_r - \pi_{r-1}\right] \mathbb{1} \left\{ \max_{r_0 \leq k < r} \Delta_k/\tau_k \leq 1 \right\} \right\|_{\mathcal{F}}.$$

There are at most N_r functions $\pi_r(f)$ and at most N_{r-1} functions $\pi_{r-1}(f)$ as f ranges over $\mathcal{F}$. Since the partitions are nested, $|\pi_r(f) - \pi_{r-1}(f)| \leq \Delta_{r-1}(f)$ and

$$|\pi_r(f) - \pi_{r-1}(f)| \mathbb{1} \left\{ \max_{r_0 \leq k < r} \Delta_k(f)/\tau_k \leq 1 \right\} \leq |\Delta_{r-1}(f)| \mathbb{1} \{\Delta_{r-1}(f)/\tau_{r-1} \leq 1\} \leq \tau_{r-1}.$$

Further,

$$\|\pi_r(f) - \pi_{r-1}(f)\|_{\gamma,n} \leq \|\Delta_{r-1}(f)\|_{\gamma,n} \leq 2^{-r+1}.$$

Just as for $T_{2,1}$, the Bernstein bound (Lemma B.2) gives

$$\mathbb{E}\|\mathbb{G}_n T_3\|_{\mathcal{F}} \lesssim \int_0^{\delta} \sqrt{\ln_+ N_{[]}(\epsilon) d\epsilon}.$$

Bounding T_4

Finally,

$$\begin{aligned}
\mathbb{E}\|\mathbb{G}_n T_4\|_{\mathcal{F}} &= \mathbb{E} \left\| \mathbb{G}_n \left[f - \pi_{r_1}(f)\right] \mathbb{1} \left\{ \max_{r_0 \leq k \leq r_1} \Delta_k(f)/\tau_k \leq 1 \right\} \right\|_{f \in \mathcal{F}} \\
&\lesssim \mathbb{E} \left\|\mathbb{G}_n \Delta_{r_1} \mathbb{1} \{\Delta_{r_1} \leq \tau_{r_1}\}\right\|_{\mathcal{F}} + \sqrt{n} \sup_{f \in \mathcal{F}} \left\|\Delta_{r_1}(f) \mathbb{1} \{\Delta_{r_1}(f) \leq \tau_{r_1}\}\right\|_{1,n} \\
&\lesssim \mathbb{E} \left\|\mathbb{G}_n \Delta_{r_0}\right\|_{\mathcal{F}} + \sqrt{n} \tau_{r_1}.
\end{aligned}$$

and the first term is bounded by T_1. Finally, observe that, by the definition of r_1,

$$\sqrt{n}\tau_{r_1} \leq \sqrt{n}2^{-r_1} = \sqrt{n}N_{[]}^{-1}(e^n).$$

□

Lemma B.7. *For $\gamma > 2$ and*

$$\sum_{i=1}^{n} \beta_n\,(i)^{\frac{\gamma-2}{\gamma}} \leq K$$

it holds

$$\frac{1}{n}\sum_{i,j=1}^{n} |\mathbb{C}\mathrm{ov}[h(X_i), h(X_j)]| \leq 4K\|h\|_{\gamma,n}^2$$

for all $h : \mathcal{X} \to \mathbb{R}$ measurable.

Furthermore, if $\sup_{n\in\mathbb{N}} \max_{m\leq n} m^\rho \beta_n(m) < \infty$ for some $\rho > \gamma/(\gamma-2)$, then,

$$\sup_n \sum_{i=1}^{n} \beta_n\,(i)^{\frac{\gamma-2}{\gamma}} < \infty.$$

Proof. Let us first prove the latter claim. Note that if $\sup_{n\in\mathbb{N}} \max_{m\leq n} m^\rho \beta_n(m) < \infty$ then $\sum_{i=1}^{n} \beta_n\,(i)^{\frac{\gamma-2}{\gamma}} \lesssim \sum_{i=1}^{n} m^{-\rho\frac{\gamma-2}{\gamma}}$ and the latter display converges for $\rho > \gamma/(\gamma-2)$.

For the first claim, it holds

$$\begin{aligned}
&\frac{1}{n}\sum_{i,j=1}^{n} |\mathbb{C}\mathrm{ov}[h(X_i), h(X_j)]| \\
&\leq \frac{1}{n}\sum_{i,j=1}^{n} \beta_n(|i-j|)^{\frac{\gamma-2}{\gamma}} \|h(X_i)\|_\gamma \|h(X_j)\|_\gamma \\
&\leq \frac{1}{n}\sum_{i,j=1}^{n} \beta_n(|i-j|)^{\frac{\gamma-2}{\gamma}} (\|h(X_i)\|_\gamma^2 + \|h(X_j)\|_\gamma^2) \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \|h(X_i)\|_\gamma^2 \sum_{j=1}^{n} \beta_n(|i-j|)^{\frac{\gamma-2}{\gamma}} + \frac{1}{n}\sum_{j=1}^{n} \|h(X_j)\|_\gamma^2 \sum_{i=1}^{n} \beta_n(|i-j|)^{\frac{\gamma-2}{\gamma}} \\
&\leq \frac{4K}{n}\sum_{i=1}^{n} \|h(X_i)\|_\gamma^2 \\
&\leq 4K\|h\|_{\gamma,n}^2.
\end{aligned}$$

by Theorem 3 of Doukhan (2012) and where the last inequality follows from

$$\begin{aligned}
\left(\frac{1}{n}\sum_{i=1}^{n} \|h(X_i)\|_\gamma^2\right)^{1/2} &= \left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[|h(X_i)|^\gamma]^{2/\gamma}\right)^{1/2} \\
&\leq \left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[|h(X_i)|^\gamma]\right)^{1/\gamma} \qquad \text{by } 2/\gamma \leq 1 \text{ and Jensen's inequality.}
\end{aligned}$$

$\square$

Remark B.8. *Given a semi-metric d on $\mathcal{F}$ induced by a semi-norm $\|\cdot\|$ satisfying*

$$|f| \leq |g| \Rightarrow \|f\| \leq \|g\|,$$

any 2ε-bracket $[f, g]$ is contained in the ϵ-ball around $(f-g)/2$. Then,

$$N(\varepsilon, \mathcal{F}, d) \leq N_{[]}(2\varepsilon, \mathcal{F}, \|\cdot\|).$$

Both, $\|\cdot\|_{\gamma,n}$ and $\|\cdot\|_{\gamma,\infty}$, satisfy this property.

B.4. Proof of Theorem 3.4

We first prove the existence of an asymptotically tight sequence of GPs. We will conclude by Propositions 2.20 and 2.21: There exists $K \in \mathbb{R}$ such that

$$\sum_{i=1}^{k_n} \beta_n(i)^{\frac{\gamma-2}{\gamma}} \leq K$$

for all n by Lemma B.7. It holds

$$\begin{aligned}
\rho_n(s,t)^2 &= \mathbb{V}\mathrm{ar}[\mathbb{G}_n(s) - \mathbb{G}_n(t)] \\
&= \frac{1}{k_n}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{k_n}(f_{n,s} - f_{n,t})(X_{n,i})\right] \\
&\leq \frac{1}{k_n}\sum_{i,j=1}^{k_n} |\mathbb{C}\mathrm{ov}\left[(f_{n,s} - f_{n,t})(X_{n,i}), (f_{n,s} - f_{n,t})(X_{n,j})\right]| \\
&\leq 4K\|f_{n,s} - f_{n,t}\|_{\gamma,n}^2 \\
&= 4K d_n(s,t)^2
\end{aligned}$$

by Lemma B.7. Thus,

$$N(2\sqrt{K}\epsilon, T, \rho_n) \leq N(\epsilon, T, d_n) \leq N_{[]}(2\epsilon, \mathcal{F}_n, \|\cdot\|_{\gamma,n})$$

by Remark B.8. Next, observe that

$$\ln N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{\gamma,n}) = 0$$

for all $\varepsilon \geq 2\|F\|_{\gamma,n}$. Thus,

$$\int_0^\infty \sqrt{\ln N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{\gamma,n})} d\epsilon < \infty$$

for all n by the entropy condition (iii). In summary,

- (T, d) is totally bounded.
- $\lim_{n\to\infty} \int_0^{\delta_n} \sqrt{\ln N(\epsilon, T, \rho_n)} d\epsilon \lesssim \lim_{n\to\infty} \int_0^{\delta_n} \sqrt{\ln N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{\gamma,n})} d\epsilon = 0$ for all $\delta_n \downarrow 0$.
- $\lim_{n\to\infty} \sup_{d(s,t)<\delta_n} \rho_n(s,t) \le \lim_{n\to\infty} \sup_{d(s,t)<\delta_n} 2\sqrt{K} d_n(s,t) = 0$ for every $\delta_n \downarrow 0$.
- $\int_0^{\infty} \sqrt{\ln N(\epsilon, T, \rho_n)} d\epsilon \lesssim \int_0^{\infty} \sqrt{\ln N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{\gamma,n})} d\epsilon < \infty$ for all n.

by the assumptions. Lastly,

$$\sup_n \mathbb{V}\mathrm{ar}[\mathbb{G}_n(t)] \lesssim \sup_n \|F\|_{\gamma,n}^2 = \|F\|_{\gamma,\infty}^2 < \infty$$

for all $t \in T$ by the same argument as above. Combined, we derive the claim by Propositions 2.20 and 2.21.

Next, we prove asymptotic tightness of $\mathbb{G}_n$. We derive that $\sup_n \mathbb{E}\left[|\mathbb{G}_n(s)|^2\right] < \infty$ for all $s \in T$ again by the moment condition (i) and the summability condition Lemma B.7. Thus, each $\mathbb{G}_n(s)$ is asymptotically tight. By Markov's inequality and Theorem 1.5.7 of Van der Vaart and Wellner (2023), it suffices to prove uniform d-equicontinuity, i.e. that

$$\limsup_{n\to\infty} \mathbb{E}^* \sup_{d(f,g)<\delta_n} |\mathbb{G}_n(s) - \mathbb{G}_n(t)| = 0$$

for all $\delta_n \downarrow 0$. By

$$\lim_{n\to\infty} \sup_{d(s,t)<\delta_n} d_n(s,t) = 0$$

for all $\delta_n \downarrow 0$, for every sequence $\delta \to 0$ there exists a sequence $\epsilon(\delta) \to 0$ such that $d(s,t) < \delta$ implies $d_n(s,t) < \epsilon(\delta)$. Thus,

$$\limsup_{n\to\infty} \mathbb{E}^* \sup_{d(s,t)<\delta_n} |\mathbb{G}_n(s) - \mathbb{G}_n(t)| \le \limsup_{n\to\infty} \mathbb{E}^* \sup_{d_n(s,t)<\epsilon(\delta_n)} |\mathbb{G}_n(s) - \mathbb{G}_n(t)|.$$

Accordingly, it suffices to prove that

$$\limsup_{n\to\infty} \mathbb{E}^* \sup_{d_n(s,t)<\delta_n} |\mathbb{G}_n(s) - \mathbb{G}_n(t)| = 0.$$

For fixed n we again identify $\mathbb{G}_n$ with the empirical process $\mathbb{G}_n$ indexed by $\mathcal{F}_n$ and similarly for d_n. Note that $\mathbb{G}_n(f) - \mathbb{G}_n(g) = \mathbb{G}_n(f-g)$ and the bracketing number with respect to the function class

$$\mathcal{F}_{n,\delta} = \{f - g \colon f, g \in \mathcal{F}_n, \|f-g\|_{\gamma,n} < \delta\}$$

satisfies

$$N_{[]}(\epsilon, \mathcal{F}_{n,\delta}, \|\cdot\|_{\gamma,n}) \le N_{[]}(\epsilon/2, \mathcal{F}_n, \|\cdot\|_{\gamma,n})^2.$$

Indeed, given $\epsilon/2$ brackets $[l_f, u_f]$ and $[l_g, u_g]$ for f and g, $[l_f - u_g, u_f - l_g]$ is an ε-bracket for $f - g$. By Theorem B.6,

$$\begin{aligned}\limsup_n \mathbb{E}^* \sup_{d_n(s,t)<\delta_n} |\mathbb{G}_n(s) - \mathbb{G}_n(t)| &\lesssim \lim_{n\to\infty} \int_0^{2\delta_n} \sqrt{\ln N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{\gamma,n})} d\epsilon \\ &= 0\end{aligned}$$

by (iii) which proves the claim. □

C. Proofs for relative CLTs under mixing conditions

C.1. Proof of Theorem 3.2

We first restrict to univariate random variables which, in combination with the relative Cramer-Wold device (Corollary A.4), yields Theorem 3.2. The idea is to split the scaled sample average

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(X_{n,i}-\mathbb{E}\left[X_{n,i}\right]\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{r_n}\left(Z_{n,i}-\mathbb{E}\left[Z_{n,i}\right]+\tilde{Z}_{n,i}-\mathbb{E}\left[\tilde{Z}_{n,i}\right]\right)$$

into alternating long and short block sums. By considering a small enough length of the short blocks, the short block sum are asymptotically negligible. It then suffices to prove a relative CLT for the sequence of long block sums (Lemma C.1) . By maximal coupling and Lemma C.1, the sequence of long block sums can be considered independent and Lindeberg's CLT (Theorem A.7) applies.

Lemma C.1. *Let Y_n and Y_n^* be sequences of random variables in $\mathbb{R}$ such that*

(i) $|Y_n - Y_n^*| \xrightarrow{P} 0$,

(ii) $\sup_n \mathbb{V}\mathrm{ar}\left[Y_n\right] < \infty$ *and*

(iii) $|\mathbb{V}\mathrm{ar}\left[Y_n\right] - \mathbb{V}\mathrm{ar}\left[Y_n^*\right]| \to 0$.

Then, Y_n satisfies a relative CLT if and only if Y_n^* does.

Proof. It suffices to prove the if direction since the statement is symmetric. Let k_n be a subsequence of n and l_n be a further subsequence such that $Y_{l_n}^* \to_d N$ converges weakly to some Gaussian with

$$\mathbb{V}\mathrm{ar}\left[Y_{l_n}^*\right] \to \mathbb{V}\mathrm{ar}\left[N\right].$$

Such l_n exists by (iii) of Proposition 2.19. Since $|Y_n - Y_n^*| \xrightarrow{P} 0$, we obtain $Y_{l_n} \to_d N$. Note

$$\mathbb{V}\mathrm{ar}\left[N\right] = \lim_{n\to\infty} \mathbb{V}\mathrm{ar}\left[Y_{l_n}^*\right] = \lim_{n\to\infty} \mathbb{V}\mathrm{ar}\left[Y_{l_n}\right]$$

by assumption. Thus, Y_n satisfies a relative CLT by (iii) of Proposition 2.19. □

Theorem C.2 (Univariate relative CLT)**.** *Let $X_{n,1},\dots,X_{n,k_n}$ be a triangular array of univariate random variables. Let $\alpha \in [0, 1/2)$ and $1 + (1-2\alpha)^{-1} < \gamma$. Assume that*

(i) $k_n^{-1}\sum_{i,j=1}^{k_n}|\mathbb{C}\mathrm{ov}\left[X_{n,i}, X_{n,j}\right]| \le K$ *for all n.*

(ii) $\sup_{n,i} \mathbb{E}\left[|X_{n,i}|^\gamma\right] < \infty$.

(iii) $k_n\beta_n\left(k_n^{\alpha}\right)^{\frac{\gamma-2}{\gamma}} \to 0$.

Then, the scaled sample average $\sqrt{k_n}\left(\overline{X}_n - \mathbb{E}\left[\overline{X}_n\right]\right)$ satisfies a relative CLT.

Proof. There exists some δ with $0 < \delta < 1/2 - \alpha$ and $1 + (1-2\alpha)^{-1} < 1 + (2\delta)^{-1} < \gamma$. Define $q_n = k_n^{\alpha}$, $p_n = k_n^{1/2-\delta} - q_n$ and $r_n = k_n^{1/2+\delta}$. Group the observations in alternating blocks of size p_n resp. q_n, i.e.

$$
\begin{aligned}
U_{n,i} &= \left(X_{n,1+(i-1)(p_n+q_n)}, \ldots, X_{n,p_n+(i-1)(p_n+q_n)}\right) \in \mathbb{R}^{p_n} && \text{(long blocks)} \\
\tilde{U}_{n,i} &= \left(X_{n,1+ip_n+(i-1)q_n}, \ldots, X_{n,q_n+ip_n+(i-1)q_n}\right) \in \mathbb{R}^{q_n} && \text{(short blocks)}
\end{aligned}
$$

Define

$$
\begin{aligned}
Z_{n,i} &= \sum_{j=1}^{p_n} U_{n,i}^{(j)} && \text{(long block sums)} \\
\tilde{Z}_{n,i} &= \sum_{j=1}^{q_n} \tilde{U}_{n,i}^{(j)} && \text{(short block sums)}
\end{aligned}
$$

where the upper index j denotes the j-th component. Then,

$$
\sum_{i=1}^{k_n} \left(X_{n,i} - \mathbb{E}\left[X_{n,i}\right]\right) = \sum_{i=1}^{r_n} \left(Z_{n,i} - \mathbb{E}\left[Z_{n,i}\right] + \tilde{Z}_{n,i} - \mathbb{E}\left[\tilde{Z}_{n,i}\right]\right)
$$

and it holds

$$
\begin{aligned}
k_n^{-1}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{k_n} X_{n,i}\right] &\le k_n^{-1} \sum_{i,j=1}^{k_n} \left|\mathbb{C}\mathrm{ov}\left[X_{n,i}, X_{n,j}\right]\right| \le K \\
k_n^{-1}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{r_n} Z_{n,i}\right] &\le K \\
k_n^{-1}\mathbb{C}\mathrm{ov}\left[\sum_{i=1}^{r_n} \tilde{Z}_{n,i}, \sum_{i=1}^{r_n} Z_{n,i}\right] &= \mathcal{O}\left(r_n q_n / k_n\right) = \mathcal{O}\left(k_n^{1/2+\delta+\alpha-1}\right) = o(1) \\
k_n^{-1}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{r_n} \tilde{Z}_{n,i}\right] &= \mathcal{O}\left(r_n q_n / k_n\right) = o(1)
\end{aligned}
$$

by assumption. Thus,

$$
k_n^{-1/2} \sum_{i=1}^{r_n} \left(\tilde{Z}_{n,i} - \mathbb{E}\left[\tilde{Z}_{n,i}\right]\right) \xrightarrow{P} 0
$$

by Markov's inequality. Hence,

$$
\left| k_n^{-1/2} \sum_{i=1}^{k_n} \left(X_{n,i} - \mathbb{E}\left[X_{n,i}\right]\right) - k_n^{-1/2} \sum_{i=1}^{r_n} \left(Z_{n,i} - \mathbb{E}\left[Z_{n,i}\right]\right) \right| \xrightarrow{P} 0.
$$

Furthermore,

$$\left|\mathbb{V}\mathrm{ar}\left[k_n^{-1/2}\sum_{i=1}^{k_n} X_{n,i}\right] - \mathbb{V}\mathrm{ar}\left[k_n^{-1/2}\sum_{i=1}^{r_n} Z_{n,i}\right]\right| = k_n^{-1}\left|\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{r_n}\tilde{Z}_{n,i}\right] - 2\mathbb{C}\mathrm{ov}\left[\sum_{i=1}^{r_n}\tilde{Z}_{n,i}, \sum_{i=1}^{r_n} Z_{n,i}\right]\right| \to 0.$$

By the previous lemma, $k_n^{-1/2}\sum_{i=1}^{k_n}\left(X_{n,i} - \mathbb{E}\left[X_{n,i}\right]\right)$ satisfies a relative CLT if and only if $k_n^{-1/2}\sum_{i=1}^{r_n} Z_{n,i} - \mathbb{E}\left[Z_{n,i}\right]$ does.

By maximal coupling (Theorem 5.1 of Rio (2017)), for all $i = 1, \ldots, r_n$ there exist random vectors $U_{n,i}^* \in \mathbb{R}^{p_n}$ such that

- $U_{n,i} \overset{d}{=} U_{n,i}^*$.
- the sequence $U_{n,i}^*$ is independent.
- $P\left(\exists U_{n,i} \neq U_{n,i}^*\right) \leq r_n \beta_n\left(q_n\right)$.

Define the coupled long block sums

$$Z_{n,i}^* = \sum_{j=1}^{p_n} U_{n,i}^{*(j)}.$$

For all $\varepsilon > 0$ we obtain

$$\begin{aligned}
P\left(k_n^{-1/2}\left|\sum_{i=1}^{r_n} Z_{n,i}^* - \sum_{i=1}^{r_n} Z_{n,i}\right| > \epsilon\right) &\leq P\left(\exists U_{n,i} \neq U_{n,i}^*\right) \\
&\leq r_n \beta_n\left(q_n\right) \\
&= k_n^{1/2+\delta}\beta_n\left(k_n^\alpha\right) \\
&\leq k_n \beta_n\left(k_n^\alpha\right) \to 0.
\end{aligned}$$

Next,

$$\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{r_n} Z_{n,i}\right] = \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{r_n} Z_{n,i}^*\right] + \sum_{i \neq j}^{r_n}\mathbb{C}\mathrm{ov}\left[Z_{n,i}, Z_{n,j}\right]$$

by independence of $Z_{n,i}^*$ and $P_{Z_{n,i}} = P_{Z_{n,i}^*}$. Since $\sup_{n,k}\|X_{n,k}\|_\gamma < \infty$ and

$$\left|\mathbb{C}\mathrm{ov}\left[X_{n,i}, X_{n,j}\right]\right| \leq \beta_n\left(|i-j|\right)^{\frac{\gamma-2}{\gamma}}\sup_{n,k}\|X_{n,k}\|_\gamma$$

by Theorem 3. of Doukhan (2012), for $i \neq j$ we obtain

$$\begin{aligned}
\left|\mathbb{C}\mathrm{ov}\left[Z_{n,i}, Z_{n,j}\right]\right| &\leq \mathcal{O}\left(p_n^2 \beta_n\left(q_n\right)^{\frac{\gamma-2}{\gamma}}\right) \\
\left|\frac{1}{k_n}\sum_{i \neq j}^{r_n}\mathbb{C}\mathrm{ov}\left[Z_{n,i}, Z_{n,j}\right]\right| &\leq \mathcal{O}\left(k_n \beta_n\left(q_n\right)^{\frac{\gamma-2}{\gamma}}\right) = o\left(1\right).
\end{aligned}$$

Thus,

$$\frac{1}{k_n}\left|\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{r_n} Z_{n,i}\right] - \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{r_n} Z_{n,i}^*\right]\right| \to 0.$$

Combined with $P_{Z_{n,i}} = P_{Z_{n,i}^*}$, hence $k_n^{-1}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{r_n} Z_{n,i}^*\right] \leq K$ and $\mathbb{E}\left[Z_{n,i}\right] = \mathbb{E}\left[Z_{n,i}^*\right]$, the previous lemma yields that $k_n^{-1/2}\sum_{i=1}^{r_n} Z_{n,i} - \mathbb{E}\left[Z_{n,i}\right]$ satisfies a relative CLT if and only if $k_n^{-1/2}\sum_{i=1}^{r_n} Z_{n,i}^* - \mathbb{E}\left[Z_{n,i}^*\right]$ does.

Next, the moment assumption together with $P_{Z_{n,i}} = P_{Z_{n,i}^*}$ imply that the sequence $(r_n/k_n)^{1/2} Z_{n,i}^*$ satisfies the Lindeberg condition given in Theorem A.7. More specifically,

$$\begin{aligned}
\frac{1}{r_n}\sum_{i=1}^{r_n}\mathbb{E}\left[\left|\left(r_n/k_n\right)^{1/2} Z_{n,i}^*\right|^2 \mathbb{1}_{\{|(r_n/k_n)^{1/2} Z_{n,i}|^2 > r_n\varepsilon^2\}}\right] &= \frac{1}{k_n}\sum_{i=1}^{r_n}\mathbb{E}\left[|Z_{n,i}|^2\mathbb{1}_{\{|Z_{n,i}|^2 > k_n\varepsilon^2\}}\right] \\
&\leq \epsilon^{1-\gamma/2}\frac{1}{k_n^{\gamma/2}}\sum_{i=1}^{r_n}\mathbb{E}\left[|Z_{n,i}|^\gamma\right] \\
&\leq \epsilon^{1-\gamma/2} C\frac{r_n p_n^\gamma}{k_n^{\gamma/2}}
\end{aligned}$$

for $C = \sup_i \mathbb{E}\left[|X_{n,i}|^\gamma\right] < \infty$ where we used

$$\begin{aligned}
\mathbb{E}\left[|Z_{n,1}|^\gamma\right] &= \|Z_{n,1}\|_\gamma^\gamma \\
&= \left\|\sum_{i=1}^{p_n} X_{n,i}\right\|_\gamma^\gamma \\
&\leq \left(\sum_{i=1}^{p_n}\|X_{n,i}\|_\gamma\right)^\gamma \\
&\leq C p_n^\gamma
\end{aligned}$$

and similarly $\mathbb{E}\left[|Z_{n,i}|^\gamma\right] \leq C p_n^\gamma$. It holds

$$\frac{r_n p_n^\gamma}{k_n^{\gamma/2}} \leq \frac{r_n p_n^\gamma}{(r_n p_n)^{\gamma/2}} = \frac{p_n^{\gamma/2}}{r_n^{\gamma/2-1}} \leq k_n^{1/2+\delta-\delta\gamma} \to 0.$$

Thus,

$$k_n^{-1/2}\sum_{i=1}^{r_n}\left(Z_{n,i}^* - \mathbb{E}\left[Z_{n,i}^*\right]\right) = r_n^{-1/2}\sum_{i=1}^{r_n}\left(r_n/k_n\right)^{1/2}\left(Z_{n,i}^* - \mathbb{E}\left[Z_{n,i}^*\right]\right)$$

satisfies a relative CLT which finishes the proof. □

Proof of Theorem 3.2. Write

$$S_n = \frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}\left(X_{n,i} - \mathbb{E}\left[X_{n,i}\right]\right)$$

for the scaled sample average and Σ_n for its covariance matrix. Let

$$N_n \sim \mathcal{N}(0, \Sigma_n).$$

By assumption, Σ_n is componentwise a bounded sequence, hence, N_n is relatively compact. Thus, it suffices to prove $S_n \leftrightarrow_d N_n$. By Corollary A.4, this is equivalent to $t^T S_n \leftrightarrow_d t^T N_n$ for all $t \in \mathbb{R}^d$. Note that

$$t^T N_n \sim \mathcal{N}\left(0, t^T \Sigma_n\right)$$

and $t^T S_n$ is the scaled sample average associated to $t^T X_{n,i}$. Accordingly, it suffices to check the conditions of Theorem C.2 for $t^T X_{n,i}$: The moment and mixing conditions ((ii) and (iii)) of Theorem C.2 follow by assumption. Lastly,

$$k_n^{-1} \sum_{i,j=1}^{k_n} \left|\mathbb{C}\mathrm{ov}\left[t^T X_{n,i}, t^T X_{n,j}\right]\right| \leq t^T t \max_{l_1,l_2} k_n^{-1} \sum_{i,j=1}^{k_n} \left|\mathbb{C}\mathrm{ov}\left[X_{n,i}^{(l_1)}, X_{n,j}^{(l_2)}\right]\right| \leq t^T t K$$

for all n. This proves (i) of Theorem C.2 and combined we derive the claim. □

C.2. Proof of Theorem 3.6

We derive Theorem 3.6 from a more general result. Fix some triangular array $X_{n,1}, \ldots, X_{n,k_n}$ of random variables with values in a Polish space $\mathcal{X}$. For each $n \in \mathbb{N}$ let

$$\mathcal{F}_n = \{f_{n,t} \colon t \in T\}$$

be a set of measurable functions from $\mathcal{X}$ to $\mathbb{R}$. Assume that $\cup_{n\in\mathbb{N}} \mathcal{F}_n$ admits a finite envelope $F : \mathcal{X} \to \mathbb{R}$.

Theorem C.3. *Assume that for some $\gamma > 2$*

(i) $\|F\|_{\gamma,\infty} < \infty$.

(ii) $\sup_{n\in\mathbb{N}} \max_{m\leq k_n} m^\rho \beta_n(m) < \infty$ *for some* $\rho > 2\gamma(\gamma-1)/(\gamma-2)^2$

(iii) $\int_0^{\delta_n} \sqrt{\ln N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{\gamma,n})} d\epsilon \to 0$ *for all* $\delta_n \downarrow 0$ *and are finite for all* n.

Denote by

$$d_n(s,t) = \|f_{n,s} - f_{n,t}\|_{\gamma,n}$$

for $s,t \in T$. Assume that there exists a semi-metric d on T such that

$$\lim_{n\to\infty} \sup_{d(s,t)<\delta_n} d_n(s,t) = 0$$

for all $\delta_n \downarrow 0$ and (T,d) is totally bounded. Then, the empirical process $\mathbb{G}_n$ defined by

$$\mathbb{G}_n(t) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} f_{n,t}(X_{n,i}) - \mathbb{E}\left[f_{n,t}(X_{n,i})\right]$$

satisfies a relative CLT in $\ell^\infty(T)$.

Proof. We apply Theorem 3.4 to derive relative compactness of $\mathbb{G}_n$ and the existence of an asymptotically tight sequence of tight Borel measurable GPs $N_{\mathbb{G}_n}$ corresponding to $\mathbb{G}_n$. Note that each $N_{\mathbb{G}_n}(s)$ is measurable for all $s \in T$. Accordingly, $N_{\mathbb{G}_n}$ is asymptotically measurable, hence, relatively compact by Lemma A.3.

According to Corollary 2.18, it remains to prove relative CLTs of the marginals $(\mathbb{G}_n(t_1), \dots, \mathbb{G}_n(t_d))$ for all $d \in \mathbb{N}$, $t_1, \dots, t_d \in T$. We apply Theorem 3.2 to the triangular array $Y_{n,1}, \dots, Y_{n,k_n}$ with $Y_{n,k} = (f_{n,t_1}(X_k), \dots, f_{n,t_d}(X_k))$.

(ii) of Theorem 3.2 follows by $\|F\|_{\gamma,\infty} < \infty$. Next, pick

$$\rho^{-1}\frac{\gamma}{\gamma-2} < \alpha < \frac{\gamma-2}{2(\gamma-1)}.$$

Such α exists since

$$\rho^{-1}\frac{\gamma}{\gamma-2} < \frac{\gamma-2}{2(\gamma-1)}.$$

Then, $1 + (1-2\alpha)^{-1} < \gamma$ and $k_n\beta_n(k_n^\alpha)^{\frac{\gamma-2}{\gamma}} \lesssim k_n^{1-\rho\alpha\frac{\gamma-2}{\gamma}} \to 0$ since $\frac{\gamma}{\gamma-2} < \rho\alpha$. Lastly, the summability condition on the covariances follows by the summability condition on the β-mixing coefficients (Lemma B.7). Combined, we obtain the claim. $\square$

Proof of Theorem 3.6. Define the random variables $Y_{n,i} = (X_{n,i}, i) \in \mathcal{X} \times \mathbb{N}$. Note that $\mathcal{X} \times \mathbb{N}$ is Polish since $\mathcal{X}$ and $\mathbb{N}$ are. Define $T = S \times \mathcal{F}$, $\mathcal{F}_n = \{h_{n,t} \colon t \in T\}$ with

$$h_{n,(s,f)} : \mathcal{X} \times \mathbb{N} \to \mathbb{R}, (x,k) \mapsto w_{n,k}(s)f(x)$$

for every $(s,f) \in T$ with $w_{n,k} = 0$ for $k > k_n$. Note that each $h_{n,(s,f)}$ is measurable. Then, the empirical process associated to $Y_{n,i}$ and T is given by $\mathbb{G}_n$.

We apply Theorem C.3: set $K = \max\{\sup_{n,i,x} |w_{n,i}(x)|, \|F\|_{\gamma,\infty}\}$. Note that

$$F' : \mathcal{X} \times \mathbb{N} \to \mathbb{R}, (x,k) \mapsto KF(x)$$

is an envelope of $\cup_{n\in\mathbb{N}}\mathcal{F}_n$ satisfying condition (i) of Theorem C.3. Since we have $\sigma(X_{n,i}) = \sigma((X_{n,i}, i))$, the β-mixing coefficients w.r.t. $Y_{n,i}$ are equal to the β-mixing coefficients w.r.t. $X_{n,i}$. Thus, (ii) of Theorem C.3 are satisfied by assumption.

Define the semi-metric d on T by

$$d\big((s_1,f_1),(s_2,f_2)\big) = d^w(s_1,s_2) + \|f_1 - f_2\|_{\gamma,\infty}.$$

By the entropy condition (iii) and **W3** we derive that (T,d) is totally bounded. By

Minkowski's inequality, we get

$$\begin{aligned}
d_n\big((s_1,f_1),(s_2,f_2)\big) &= \|h_{n,(s_1,f_1)} - h_{n,(s_2,f_2)}\|_{\gamma,n} \\
&= \left(\frac{1}{k_n}\sum_{i=1}^{k_n} \|w_{n,i}(s_1)f_1(X_{n,i}) - w_{n,i}(s_2)f_2(X_{n,i})\|_\gamma^\gamma\right)^{1/\gamma} \\
&\le \left(\frac{1}{k_n}\sum_{i=1}^{k_n} \big[\|F(X_{n,i})\|_\gamma |w_{n,i}(s_1) - w_{n,i}(s_2)| + \|w_{n,i}\|_\infty \|(f_1-f_2)(X_{n,i})\|_\gamma\big]^\gamma\right)^{1/\gamma} \\
&\le K\left(\frac{1}{k_n}\sum_{i=1}^{k_n} |w_{n,i}(s_1) - w_{n,i}(s_2)|^\gamma\right)^{1/\gamma} + K\left(\frac{1}{k_n}\sum_{i=1}^{k_n} \|(f_1-f_2)(X_{n,i})\|_\gamma^\gamma\right)^{1/\gamma} \\
&\le K d_n^w(s_1,s_2) + K\|f_1 - f_2\|_{\gamma,\infty}
\end{aligned}$$

which implies

$$\lim_{n\to\infty} \sup_{d(s,t)<\delta_n} d_n(s,t) = 0$$

for all $\delta_n \downarrow 0$.

Define $g_{n,s}(i) = w_{n,i}(s)$ and $\mathcal{G}_n = \{g_{n,s} : \mathbb{N} \to \mathbb{R} : s \in S\}$. Given $f \in \mathcal{F}$, $s \in S$ and ε-brackets $\underline{g} \le g_{n,s} \le \overline{g}$ and $\underline{f} \le f \le \overline{f}$, set the centers $f_c = (\overline{f} + \underline{f})/2$ and $g_c = (\overline{g} + \underline{g})/2$. Then,

$$\begin{aligned}
|fg_{n,s} - f_cg_c| &\le |fg_{n,s} - fg_c| + |fg_c - f_cg_c| \\
&\le F\frac{\overline{g} - \underline{g}}{2} + K\frac{\overline{f} - \underline{f}}{2}.
\end{aligned}$$

Thus, we obtain a bracket

$$f_cg_c - \left(F\frac{\overline{g} - \underline{g}}{2} + K\frac{\overline{f} - \underline{f}}{2}\right) \quad \le \quad fg_{n,s} \quad \le \quad f_cg_c + \left(F\frac{\overline{g} - \underline{g}}{2} + K\frac{\overline{f} - \underline{f}}{2}\right)$$

with

$$\begin{aligned}
\|F(\overline{g} - \underline{g}) + K(\overline{f} - \underline{f})\|_{\gamma,n} &\le \|F\|_{\gamma,\infty}\|\overline{g} - \underline{g}\|_{\gamma,n} + K\|\overline{f} - \underline{f}\|_{\gamma,n} \\
&\le 2K\varepsilon
\end{aligned}$$

hence, an $2K\varepsilon$-bracket for $fg_{n,s}$. This implies

$$N_{[]}\left(2K\epsilon, \mathcal{F}_n, \|\cdot\|_{\gamma,n}\right) \le N_{[]}\left(\epsilon, \mathcal{F}, \|\cdot\|_{\gamma,\infty}\right) N_{[]}\left(\epsilon, S, \|\cdot\|_{\gamma,n}\right).$$

Since

$$\int_0^{\delta_n} \sqrt{\ln N_{[]}\left(\epsilon, \mathcal{F}, \|\cdot\|_{\gamma,\infty}\right)}d\epsilon, \qquad \int_0^{\delta_n} \sqrt{\ln N\left(\epsilon, S, d\right)}d\epsilon \to 0$$

for all $\delta_n \downarrow 0$, this implies the entropy condition (iii) of Theorem C.3 and, combined, we derive the claim. □

Proof of Corollary 3.8. In Theorem 3.6 set $w_{n,i} : [0,1] \to \{0,1\}, s \mapsto \mathbb{1}\{i \le \lfloor sn \rfloor\}$. Now note that $|w_{n,i}(s) - w_{n,i}(t)| \le 1$ can only be non-zero for $|\lfloor sk_n \rfloor - \lfloor tk_n \rfloor| \le 2k_n|s-t|$ many i's. Thus,

$$\begin{aligned} d_n^w(s,t) &= \left(\frac{1}{k_n} \sum_{i=1}^{k_n} |w_{n,i}(s) - w_{n,i}(t)|^\gamma \right)^{1/\gamma} \\ &\le (2|s-t|)^{1/\gamma} \end{aligned}$$

and **W2** and **W3** are satisfied for $d^w(s,t) = |s-t|^{1/\gamma}$. In combination with $w_{n,i}(s) \le w_{n,i}(t)$ for all $s \le t$,

$$N_{[]}\left(\epsilon, \mathcal{W}_n, \|\cdot\|_{\gamma,n}\right) \le \lfloor 1/\varepsilon \rfloor^{1/\gamma}$$

which implies **W1**. Theorem 3.6 gives the claim. $\square$

C.3. Asymptotic tightness of the multiplier empirical process

Let $V_{n,1}, \dots, V_{n,k_n}$ be a triangular array of identically distributed random variables. Define $\mathbb{G}_n^V \in \ell^\infty(S \times \mathcal{F})$ by

$$\mathbb{G}_n^V(s,f) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} V_{n,i} w_{n,i}(s) \big(f(X_{n,i}) - \mathbb{E}[f(X_{n,i})] \big).$$

We will derive asymptotic tightness of the multiplier empirical process $\mathbb{G}_n^V$ in terms of bracketing entropy conditions with respect to $\mathcal{F}$. Here, we apply a coupling argument and apply Corollary B.5 to the coupled empirical process.

We will argue similar to the proof of Theorem C.3. To avoid clutter we assume $w_{n,i} = 1$, but the argument is similar for the general case. Set $f_V : \mathbb{R} \times \mathcal{X} \to \mathbb{R}, (v,x) \mapsto vf(x)$. For $f \in \mathcal{F}$ define

$$f_{V,+,n} : (\mathbb{R} \times \mathcal{X})^{m_n} \to \mathbb{R}, (v,x) \mapsto m_n^{-1/2} \sum_{i=1}^{m_n} f_V(v_i, x_i).$$

Define $\mathcal{F}_{V,+,n} = \{f_{V,+,n} \colon f \in \mathcal{F}\}$. We use $U_{n,i}$ resp. $U_{n,i}^*$ from the maximal coupling paragraph but with $X_{n,i}$ replaced by $(V_{n,i}, X_{n,i})$. Define $r_n = k_n/(2m_n)$ and $\mathbb{G}_{n,1}, \mathbb{G}_{n,2} \in \ell^\infty(\mathcal{F})$ by

$$\begin{aligned} \mathbb{G}_{n,1}(f) &= \frac{1}{\sqrt{r_n}} \sum_{i=1}^{r_n} f_{V,+,n}(U_{n,2i-1}) - \mathbb{E}[f_{+,n}(U_{n,2i-1})], \\ \mathbb{G}_{n,2}(f) &= \frac{1}{\sqrt{r_n}} \sum_{i=1}^{r_n} f_{V,+,n}(U_{n,2i}) - \mathbb{E}[f_{+,n}(U_{n,2i})]. \end{aligned}$$

and $\mathbb{G}_{n,j}^*$ as $\mathbb{G}_{n,j}$ but with $U_{n,i}$ replaced by $U_{n,i}^*$. Note $\mathbb{G}_n^V = \mathbb{G}_{n,1} + \mathbb{G}_{n,2}$.

Lemma C.4. *Denote by $\beta_n^{(V,X)}$ the β-coefficients associated with the triangular array $(V_{n,i}, X_{n,i})$ If*

$$\frac{k_n}{m_n}\beta_n^{(V,X)}(m_n) \to 0,$$

and $\mathbb{G}_{n,1}^*, \mathbb{G}_{n,2}^*$ are asymptotically tight then, $\mathbb{G}_n^V$ is asymptotically tight.

Proof. It holds

$$\begin{aligned} P^*\left(\|\mathbb{G}_{n,1} - \mathbb{G}_{n,1}^*\|_{\mathcal{F}} \neq 0\right) &\leq r_n P\left(\exists i\colon U_i^* \neq U_i\right) \\ &\leq k_n/m_n\beta_n^{(V,X)}(m_n) \to 0 \end{aligned}$$

and similarly for $G_{n,2}$. Thus, $\mathbb{G}_{n,1} - \mathbb{G}_{n,1}^* \xrightarrow{P^*} 0$ and if $\mathbb{G}_{n,1}^*, \mathbb{G}_{n,2}^*$ are asymptotically tight, so are $\mathbb{G}_{n,1}, \mathbb{G}_{n,2}$. Since finite sums of asymptotically tight sequences remain asymptotically tight we obtain $\mathbb{G}_n^V = \mathbb{G}_{n,1} + \mathbb{G}_{n,2}$ is asymptotically tight. □

Theorem C.5. *For some $\gamma > 2$ and $\alpha < (\gamma-2)/2(\gamma-1)$, assume*

(i) $\|F\|_{\gamma,\infty} < \infty$.

(ii) $\sup_n \|V_{n,1}\|_\gamma < \infty$

(iii) $k_n^{1-\alpha}\beta_n^V(k_n^\alpha) + k_n^{1-\alpha}\beta_n^X(k_n^\alpha) \to 0$ *where β_n^V denotes the β-coefficients of the $V_{n,i}$.*

(iv) $\sup_n \sum_{i=1}^{k_n} \beta_n^X(i)^{\frac{\gamma-2}{\gamma}} < \infty$.

(v) $\int_0^{\delta_n} \sqrt{\ln N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{\gamma,\infty})}d\epsilon \to 0$ *for all $\delta_n \to 0$.*

Then, $\mathbb{G}_n^V$ is asymptotically tight.

Proof. First, we may without loss of generality assume $\mathbb{E}[f(X_{n,i})] = 0$. To see this define the function class $\mathcal{G}_n = \{g_f\colon f \in \mathcal{F}\}$ with

$$g_f : \mathcal{X} \times \{1, \ldots, k_n\} \to \mathbb{R}, (x, i) \mapsto f(x) - \mathbb{E}[f(X_{n,i})].$$

For $Y_{n,i} = (X_{n,i}, i)$ it holds $\mathbb{E}[g_f(Y_{n,i})] = 0$. Note that $\mathcal{X} \times \{1, \ldots, k_n\}$ remains Polish and the β-mixing coefficients with respect to $Y_{n,i}$ and $X_{n,i}$ are equal. Further, given an ε-bracket $[\underline{f}, \overline{f}]$ with respect to $\|\cdot\|_{\gamma,n}$, define

$$u(x,i) = \overline{f}(x) - \mathbb{E}[\underline{f}(X_{n,i})], l(x,i) = \underline{f} - \mathbb{E}[\overline{f}(X_{n,i})].$$

Clearly, $f \in [\underline{f}, \overline{f}]$ implies $g_f \in [l, u]$. Further,

$$\begin{aligned} |u - l| &\leq |\underline{f} - \overline{f}| + \mathbb{E}[|\underline{f} - \overline{f}|] \\ &\leq |\underline{f} - \overline{f}| + \|\underline{f} - \overline{f}\|_{\gamma,n} \\ &\leq |\underline{f} - \overline{f}| + \varepsilon. \end{aligned}$$

Thus, $\|u-l\|_{\gamma,n} \leq 2\varepsilon$ and $N_{[]}(2\varepsilon, \mathcal{G}_n, \|\cdot\|_{\gamma,n}) \leq N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{\gamma,n})$. Replacing $\mathcal{F}$ by $\mathcal{G}_n$ and $X_{n,i}$ by $Y_{n,i}$ we may assume $\mathbb{E}[f(X_{n,i})] = 0$.

Next, recall the function class $\mathcal{F}_V = \{f_V : f \in \mathcal{F}\}$ with $f_V : \mathbb{R}\times\mathcal{X} \to \mathbb{R}, (v,x) \mapsto vf(x)$ and set $Y_{n,i} = (V_{n,i}, X_{n,i})$. Then, the empirical process $\mathbb{G}_n^V \in \ell^\infty(\mathcal{F})$ is the empirical process associated to $\mathcal{F}_V$ and $Y_{n,i}$. Again, $\mathbb{R}\times\mathcal{X}$ is Polish and the β-coefficients β_n^Y associated to $Y_{n,i}$ satisfy

$$\beta_n^Y(m) \leq \beta_n^X(m) + \beta_n^V(m)$$

by Theorem 5.1 (c) of Bradley (2005). Set $m_n = k_n^\alpha$. Then,

$$\frac{k_n}{m_n}\beta_n^Y(m_n) = k_n^{1-\alpha}\beta_n^Y(k_n^\alpha) \to 0.$$

We apply Lemma C.4 to $Y_{n,i}$ and $\mathcal{F}_V$. Thus, it suffices to show that $\mathbb{G}_{n,1}^*$ and $\mathbb{G}_{n,1}^*$ are asymptotically tight. We will prove the statement for $\mathbb{G}_{n,1}^*$. The arguments for $\mathbb{G}_{n,1}^*$ are the same.

Similar to the proof of Theorem 3.4, it suffices to show

$$\limsup_{\delta\to 0}\limsup_{n\to\infty} \mathbb{E}^*\|\mathbb{G}_{n,1}^*\|_{\mathcal{F}_{V,\delta}} = 0,$$

where

$$\mathcal{F}_{V,\delta} = \{f_V - g_V : f,g \in \mathcal{F}, \|f-g\|_{\gamma,\infty} < \delta\}.$$

Recall

$$\mathbb{G}_{n,1}^*(f_V) = \frac{1}{r_n}\sum_{i=1}^{k_n} f_{V,+,n}(U_{n,2i-1}^*)$$

since $\mathbb{E}[f(X_{n,i})] = 0$. Note that for fixed n, $\mathbb{G}_{n,1}^*$ can be identified with an empirical process $\mathbb{G}_{n,1}^* \in \ell^\infty(\mathcal{F}_{V,+,n})$ indexed by the function class $\mathcal{F}_{V,+,n} = \{f_{V,+,n} : f \in \mathcal{F}\}$. Denote by $\|\cdot\|_{2,n,+}$ the $\|\cdot\|_{2,n}$-seminorm on $\mathcal{F}_{V,+,n}$ induced by $U_{n,i}^*$.

Next, let $f \in \mathcal{F}$ and an ε-bracket $f \in [\underline{f}, \overline{f}]$ with respect to $\|\cdot\|_{\gamma,n}$ be given. Clearly $f_{V,+,n} \in [\underline{f}_{V,+,n}, \overline{f}_{V,+,n}]$. Since $\mathbb{E}[f(X_{n,i})] = 0$ it holds

$$\begin{aligned}
\|f_{V,+,n}\|_{2,n,+}^2 &= \frac{1}{r_n}\sum_{i=1}^{r_n}\mathbb{E}[f_{V,+,n}(U_{n,i}^*)^2] \\
&= \frac{1}{r_n}\sum_{i=1}^{r_n}\mathbb{V}\mathrm{ar}[f_{V,+,n}(U_{n,i}^*)] \\
&= \frac{1}{m_n r_n}\sum_{i=1}^{r_n}\sum_{j,k=1}^{p_n}\mathbb{C}\mathrm{ov}[V_{n,(i-1)m_n+j}f(X_{n,(i-1)m_n+j}), V_{n,(i-1)m_n+k}f(X_{n,(i-1)m_n+k})] \\
&\lesssim \frac{2}{k_n}\sum_{i,j=1}^{k_n}|\mathbb{C}\mathrm{ov}[f(X_{n,i}), f(X_{n,j})] \\
&\lesssim \|f\|_{\gamma,n}^2,
\end{aligned}$$

by Lemma B.7. This yields

$$N_{[]}(C\varepsilon, \mathcal{F}_{V,+,n}, \|\cdot\|_{2,n,+}) \leq N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{\gamma,n}),$$

for some constant C independent of n and

$$\|f_{V,+,n} - g_{V,+,n}\|_{2,n,+} \leq C\|f - g\|_{\gamma,n} \leq C\delta,$$

for all $f_V - g_V \in \mathcal{F}_{V,\delta}$. Without loss of generality, $C = 1$. Lastly, note that $F_{V,+,n}$ are envelopes for $\mathcal{F}_{V,+,n}$.

Putting everything together, and in combination with Corollary B.5, we obtain

$$\mathbb{E}\|\mathbb{G}_{n,1}\|_{\mathcal{F}_{V,\delta}} \lesssim \int_0^{2\delta} \sqrt{\ln_+ N_{[]}(\epsilon)}d\epsilon + \frac{B \ln_+ N_{[]}(\delta)}{\sqrt{r_n}} + \sqrt{r_n}\|F_{V,+,n}\mathbb{1}\{F_{V,+,n} > B\}\|_{1,n} + \sqrt{r_n}N_{[]}^{-1}(e^{r_n})$$

with $N_{[]}(\epsilon) = N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{\gamma,\infty})$. Let $B = a_n\sqrt{r_n}$, with $a_n \to 0$ arbitrarily slowly. We obtain

$$\frac{B \ln_+ N_{[]}(\delta)}{\sqrt{r_n}} = a_n \ln_+ N_{[]}(\delta) \to 0,$$

for every fixed $\delta > 0$ due to condition (v). Further,

$$\begin{aligned}
\sqrt{r_n}\|F_{V,+,n}\mathbb{1}\{F_{V,+,n} > a_n\sqrt{r_n}\}\|_{1,n} &\leq \sqrt{r_n m_n}\|F_V\mathbb{1}\{F_V > a_n\sqrt{r_n/m_n}\}\|_{1,n} \\
&= \sqrt{r_n m_n (r_n/m_n)^{1-\gamma}}\|F_V\|_{\gamma,n}^{\gamma} a_n^{1-\gamma} \\
&= \sqrt{r_n m_n (r_n/m_n)^{1-\gamma}}\|V_{n,1}\|_{\gamma}\|F\|_{\gamma,n}^{\gamma} a_n^{1-\gamma} \\
&\lesssim \sqrt{\frac{m_n^{\gamma}}{r_n^{\gamma-2}}}\|F\|_{\gamma,n}^{\gamma} a_n^{1-\gamma} \\
&\lesssim \sqrt{\frac{m_n^{\gamma+\gamma-2}}{k_n^{\gamma-2}}}\|F\|_{\gamma,n}^{\gamma} a_n^{1-\gamma} \\
&= \sqrt{k_n^{2\alpha(\gamma-1)-(\gamma-2)}}\|F\|_{\gamma,n}^{\gamma} a_n^{1-\gamma} \\
&\to 0,
\end{aligned}$$

for $a_n \to 0$ sufficiently slowly, where we used that $V_{n,i}$ are identically distributed with $\sup_n \|V_{n,i}\|_{\gamma} < \infty$ in the third and fourth step, and our condition on α and (i) in the last. Combined, we obtain

$$\limsup_n \mathbb{E}\|\mathbb{G}_{n,1}\|_{\mathcal{F}_{V,\delta}} \lesssim \int_0^{\delta} \sqrt{\ln_+ N_{[]}(\epsilon)}d\epsilon$$

for all $\delta > 0$. With $\delta_n \to 0$, we obtain

$$\limsup_n \mathbb{E}\|\mathbb{G}_{n,1}\|_{\mathcal{F}_{V,\delta_n}} = \limsup_n \int_0^{\delta_n} \sqrt{\ln_+ N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{\gamma,\infty})}d\epsilon,$$

and the right hand side converges to 0 as $\delta \to 0$ by condition (v), completing the proof. □

D. Proofs for the bootstrap

We will use the notation introduced in Section 4 without further mentioning.

D.1. Proof of Proposition thm:bstrap-rwc and a corollary

Proof of Proposition 4.1. Since $\mathbb{G}_n$ is relatively compact, so is $\mathbb{G}_n^{\otimes 3}$ (Van der Vaart and Wellner, 2023, Example 1.4.6). Because $\mathbb{G}_n$ and $\mathbb{G}_n^{\otimes 3}$ are relatively compact, both statements can be checked at the level of subsequences and we may assume that $\mathbb{G}_n \to_d \mathbb{G}$ converges weakly to some tight Borel law. By (asymptotic) independence, we obtain $\mathbb{G}_n^{\otimes 3} \to_d \mathbb{G}^{\otimes 3}$. Then (ii) is equivalent to

$$\left(\mathbb{G}_n, \mathbb{G}_n^{(1)}, \mathbb{G}_n^{(2)}\right) \to_d \mathbb{G}^{\otimes 3}$$

in $\ell^\infty(\mathcal{F})^3$. Thus, (ii) is equivalent to (a) of Lemma 3.1 in Bücher and Kojadinovic (2019) and we obtain the claim. $\square$

Corollary D.1. *Assume that $\mathbb{G}_n$ satisfies a relative CLTs and $\mathbb{G}_n^{(i)}$ are relatively compact. Then, $\mathbb{G}_n^{(1)}$ is a consistent bootstrap scheme if*

(i) all marginals of $\left(\mathbb{G}_n, \mathbb{G}_n^{(1)}, \mathbb{G}_n^{(2)}\right)$ satisfy a relative CLT.

(ii) for $n \to \infty$

$$\begin{aligned}\operatorname{Cov}\left[\mathbb{G}_n^{(i)}(s), \mathbb{G}_n^{(i)}(t)\right] - \operatorname{Cov}\left[\mathbb{G}_n(s), \mathbb{G}_n(t)\right] &\to 0\\ \operatorname{Cov}\left[\mathbb{G}_n^{(i)}(s), \mathbb{G}_n^{(j)}(t)\right] &\to 0.\end{aligned}$$

for all $i, j = 0, 1, 2$, $i \neq j$ where we set $\mathbb{G}_n^{(0)} = \mathbb{G}_n$.

Proof. We conclude by Proposition 4.1, i.e., we prove

$$\left(\mathbb{G}_n, \mathbb{G}_n^{(1)}, \mathbb{G}_n^{(2)}\right) \to_d \mathbb{G}_n^{\otimes 3}.$$

Since $\mathbb{G}_n$ satisfies a relative CLTs resp. $\mathbb{G}_n^{(i)}$ is relatively compact and satisfies marginal relative CLTs, any subsequence of n contains a further subsequence such that both, $\mathbb{G}_n$ and $\mathbb{G}_n^{(i)}$, converge weakly to some tight and measurable GP. By Proposition 2.10 and (ii) of Proposition 4.1, we may assume that $\mathbb{G}_n$ and $\mathbb{G}_n^{(i)}$ converge weakly to some tight and measurable GP. By the latter condition, such limiting GPs are equal in distribution, i.e., $\mathbb{G}_n, \mathbb{G}_n^{(i)} \to_d N$ with N some tight and measurable GP. Denote by $N^{(i)}$ iid copies of N. It suffices to prove

$$\left(\mathbb{G}_n, \mathbb{G}_n^{(1)}, \mathbb{G}_n^{(2)}\right) \to_d N^{\otimes 3}.$$

By the middle condition,

$$\left(\mathbb{G}_n(t_1), \dots, \mathbb{G}_n(t_k), \mathbb{G}_n^{(1)}(t_{k+1}), \dots, \mathbb{G}_n^{(1)}(t_{k+m}), \mathbb{G}_n^{(2)}(t_{k+m+1}), \dots, \mathbb{G}_n^{(2)}(t_{m+k+l})\right)$$

converges weakly to

$$\big(N(t_1),\ldots,N(t_k),N^{(1)}(t_{k+1}),\ldots,N^{(1)}(t_{k+m}),N^{(2)}(t_{k+m+1}),\ldots,N^{(2)}(t_{m+k+l})\big).$$

Thus,

$$\big(\mathbb{G}_n,\mathbb{G}_n^{(1)},\mathbb{G}_n^{(2)}\big)\to_d N^{\otimes 3}$$

(Van der Vaart and Wellner, 2023, Section 1.5 Problem 3.) which finishes the proof. □

D.2. Proof of Proposition 4.2

Lemma D.2. *For every $\epsilon>0$ define $\nu_n(\varepsilon)$ as the maximal natural number such that*

$$\max_{|i-j|\le\nu(\varepsilon)}|\mathbb{C}\mathrm{ov}[V_{n,i},V_{n,j}]-1|\le\epsilon.$$

Assume that for some $\gamma>2$

(i) $\sup_{n,i}\mathbb{E}[|F(X_{n,i})|^\gamma]<\infty$,

(ii) $k_n\beta_n^X\big(\nu_n(\varepsilon)\big)^{\frac{\gamma-2}{\gamma}}\to 0$ *and*

(iii) $k_n^{-1}\sum_{i,j=1}^{k_n}|\mathbb{C}\mathrm{ov}[f(X_{n,i}),g(X_{n,j})]|\le K$ *for all* $f,g\in\mathcal{F}$.

Then,

$$\mathbb{C}\mathrm{ov}\big[\mathbb{G}_n^{(j)}(s,f),\mathbb{G}_n^{(j)}(t,g)\big]-\mathbb{C}\mathrm{ov}\big[\mathbb{G}_n(s,f),\mathbb{G}_n(t,g)\big]\to 0$$

for all $(s,f),(t,g)\in S\times\mathcal{F}$.

Proof. Recall

$$|\mathbb{C}\mathrm{ov}[f(X_{n,i}),g(X_{n,i})]|\le\sup_{n,i}\mathbb{E}[|F(X_{n,i})|^\gamma]^2\beta_n^X(|i-j|)^{\frac{\gamma-2}{\gamma}}$$

by Theorem 3 of Doukhan (2012).

Under the assumptions

$$\begin{aligned}\left|\sum_{\substack{i,j\le k_n\\|i-j|\le\nu_n(\varepsilon)}}[\mathbb{C}\mathrm{ov}[V_{n,i},V_{n,j}]-1]\,\mathbb{C}\mathrm{ov}[f(X_{n,i}),g(X_{n,j})]\right|&\lesssim\epsilon\sum_{\substack{i,j\le k_n\\|i-j|\le\nu_n(\varepsilon)}}|\mathbb{C}\mathrm{ov}[f(X_{n,i}),g(X_{n,j})]|\\&\lesssim k_n\epsilon\end{aligned}$$

$$\begin{aligned}k_n^{-1}\left|\sum_{\substack{i,j\le k_n\\|i-j|>\nu_n(\varepsilon)}}[\mathbb{C}\mathrm{ov}[V_{n,i},V_{n,j}]-1]\,\mathbb{C}\mathrm{ov}[f(X_{n,i}),g(X_{n,j})]\right|&\le k_n^{-1}\sum_{\substack{i,j\le k_n\\|i-j|>\nu_n(\varepsilon)}}|\mathbb{C}\mathrm{ov}[f(X_{n,i}),g(X_{n,j})]|\\&\lesssim k_n\beta_n^X\big(\nu_n(\varepsilon)\big)^{\frac{\gamma-2}{\gamma}}\to 0\end{aligned}$$

Thus,

$$\limsup_n \left| k_n^{-1} \sum_{i,j=1}^{k_n} [\mathbb{C}\mathrm{ov}[V_{n,i}, V_{n,j}] - 1] \,\mathbb{C}\mathrm{ov}[f(X_{n,i}), g(X_{n,j})] \right| \lesssim \epsilon,$$

with constant independent of ϵ. Hence,

$$\begin{aligned}
&\limsup_n \left| \mathbb{C}\mathrm{ov}\big[\mathbb{G}_n^{(j)}(s,f), \mathbb{G}_n^{(j)}(t,g)\big] - \mathbb{C}\mathrm{ov}\big[\mathbb{G}_n(s,f), \mathbb{G}_n(t,g)\big] \right| \\
&\leq \sup_{n,i,s} |w_{n,i}(s)|^2 \left[\limsup_n \left| \frac{1}{k_n} \sum_{i,j=1}^{k_n} [\mathbb{C}\mathrm{ov}[V_{n,i}, V_{n,j}] - 1] \,\mathbb{C}\mathrm{ov}[f(X_{n,i}), g(X_{n,j})] \right| \right] \\
&\lesssim \epsilon.
\end{aligned}$$

Since this is true for all $\epsilon > 0$, taking $\epsilon \to 0$ yields the claim. □

Proof of Proposition 4.2. We check the conditions of Corollary D.1. By Theorem 3.6 $\mathbb{G}_n$ satisfies a relative CLT.

As in the proof of Theorem C.3, there exists some $\alpha' < \frac{\gamma-2}{2(\gamma-1)}$ such that $k_n\beta_n(k_n^{\alpha})^{\frac{\gamma-2}{\gamma}} \to 0$. Taking the maximum of α and α' we may assume $\alpha' = \alpha$. By Theorem C.5 $\mathbb{G}_n^{(j)}$ is asymptotically tight hence relatively compact.

To prove marginal relative CLTs, assume without loss of generality $\mathbb{E}[f(X_{n,i})] = 0$ and note that the β-coefficients associated to the triangular arrays

$$(w_{n,i}(s_1)f_1(X_{n,i}), \dots, V_{n,i}^{(2)} w_{n,i}(s_k)f_k(X_{n,i}))$$

are bounded above by $\beta_n^X + \beta_n^V$ by Theorem 5.1 (c) of Bradley (2005). In particular, the assumptions on the β-coefficients' decay rate given in Theorem 3.2 is satisfied for the marginals. Next,

$$\mathbb{E}[|V_{n,i}^{(k)} w_{n,i}(s) f(X_{n,i})|^\gamma] = \mathbb{E}[|V_{n,i}^{(k)}|^\gamma]\mathbb{E}[|w_{n,i}(s) f(X_{n,i})|^\gamma]$$

by independence of $\mathbb{X}_n$ and $\mathbb{V}_n^{(k)}$. In particular,

$$\sup_{n,i} \mathbb{E}[|V_{n,i}^{(k)} w_{n,i}(s) f(X_{n,i})|^\gamma] < \infty$$

for all $(s,f) \in S \times \mathcal{F}$ by assumption. Note that by the law of total covariances and $\mathbb{E}[V_{n,i}^{(k)}] = 0$

$$\begin{aligned}
\left| \mathbb{C}\mathrm{ov}\big[V_{n,i}^{(k)} f(X_{n,i}), V_{n,j}^{(k)} g(X_{n,j})\big] \right| &= \left| \mathbb{E}\left[V_{n,i}^{(k)} V_{n,j}^{(k)}\right] \mathbb{C}\mathrm{ov}\big[f(X_{n,i}), g(X_{n,j})\big] \right| \\
&\lesssim \left| \mathbb{C}\mathrm{ov}\big[f(X_{n,i}), g(X_{n,j})\big] \right|.
\end{aligned}$$

Thus, (i) of Theorem 3.2 follows by the summability of the β-coefficients of $X_{n,i}$ (Lemma B.7). and Theorem 3.2 implies marginal relative CLTs of $(\mathbb{G}_n, \mathbb{G}_n^{(1)}, \mathbb{G}_n^{(2)})$.

By the law of total covariances, independence of $\mathbb{V}_n^{(k)}, \mathbb{X}_n$ and $\mathbb{E}[V_{n,i}^{(k)}] = 0$ we derive

$$\mathbb{Cov}\left[w_{n,i}(s)f(X_{n,i}), V_{n,j}^{(k)}w_{n,j}(t)g(X_{n,j})\right] = 0$$
$$\mathbb{Cov}\left[V_{n,i}^{(k)}w_{n,i}(s)f(X_{n,i}), V_{n,j}^{(l)}w_{n,j}(t)g(X_{n,j})\right] = 0$$

for all $k \neq l$ and $(s, f), (t, g) \in S \times \mathcal{F}$. By the above computation we obtain

$$\mathbb{Cov}[\mathbb{G}_n^{(k)}(s, f), \mathbb{G}_n^{(l)}(t, g)] = \mathbb{Cov}[\mathbb{G}_n(s, f), \mathbb{G}_n^{(k)}(t, g)] = 0$$

for $k \neq l$. Lastly,

$$\mathbb{Cov}[\mathbb{G}_n^{(1)}(s, f), \mathbb{G}_n^{(1)}(t, g)] - \mathbb{Cov}[\mathbb{G}_n(s, f), \mathbb{G}_n(t, g)] \to 0$$

by Lemma D.2. Then, Corollary D.1 provides the claim. □

D.3. Proof of Proposition 4.4

Observe that for every $\epsilon > 0$,

$$\begin{aligned}
&\mathbb{Var}[\widehat{\mathbb{G}}_n^*(s, f) - \overline{\mathbb{G}}_n^*(s, f)] \\
&= \mathbb{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n V_{n,i}w_{n,i}(s)(\overline{\mu}_n(i, f) - \widehat{\mu}_n(i, f))\right] \\
&= \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \mathbb{Cov}[V_{n,i}, V_{n,j}]w_{n,i}(s)w_{n,j}(s)\mathbb{Cov}[(\overline{\mu}_n(i, f) - \widehat{\mu}_n(i, f)), (\overline{\mu}_n(j, f) - \widehat{\mu}_n(j, f))] \\
&\lesssim \nu_n(\epsilon)\sup_i \mathbb{Var}[\widehat{\mu}_n(i, f)] + \epsilon,
\end{aligned}$$

using the same arguments as in the proof of Proposition 4.2. Taking $\epsilon \to 0$ and our assumption on the variance give $\mathbb{Var}[\widehat{\mathbb{G}}_n^*(s, f) - \overline{\mathbb{G}}_n^*(s, f)] \to_p 0$. Since also $\mathbb{E}[\widehat{\mathbb{G}}_n^*(s, f) - \overline{\mathbb{G}}_n^*(s, f)] = 0$, it follows $\widehat{\mathbb{G}}_n^*(s, f) - \overline{\mathbb{G}}_n^*(s, f) \to_p 0$ for every s, f. Since $\widehat{\mathbb{G}}_n^* - \overline{\mathbb{G}}_n^*$ is relatively compact, this implies $\|\widehat{\mathbb{G}}_n^* - \overline{\mathbb{G}}_n^*\|_{\mathcal{S}\times\mathcal{F}} \to_p 0$. □

D.4. Proof of Proposition 4.5

Similar to the proof of Proposition 4.4, we have

$$\begin{aligned}
\mathbb{Var}[\overline{\mathbb{G}}_n^*(s, f) - \mathbb{G}_n^*(s, f)] &= \mathbb{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n V_{n,i}w_{n,i}(s)(\overline{\mu}_n(i, f) - \mu_n(i, f))\right] \\
&\lesssim \nu_n(\epsilon)\sup_i |\overline{\mu}_n(i, f) - \mu_n(i, f)|^2 + \epsilon,
\end{aligned}$$

Now conclude by the same arguments. □

D.5. Proof of Proposition 4.6

Define $\overline{\mathbb{Z}}_n^*$ as the Gaussian process corresponding to $\overline{\mathbb{G}}_n^*$ and $\overline{q}_{n,\alpha}^*$ as the α-quantile of $\|\overline{\mathbb{Z}}_n^*\|_{\mathcal{S}\times\mathcal{F}}$. For every $\epsilon > 0$, Proposition 4.4 yields

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}(\|\widehat{\mathbb{G}}_n^*\|_{\mathcal{S}\times\mathcal{F}} \leq \overline{q}_{n,\alpha}^* - \epsilon) \\
&\leq \limsup_{n\to\infty} \mathbb{P}(\|\overline{\mathbb{G}}_n^*\|_{\mathcal{S}\times\mathcal{F}} \leq \overline{q}_{n,\alpha}^*) + \limsup_{n\to\infty} \mathbb{P}(\|\widehat{\mathbb{G}}_n^* - \overline{\mathbb{G}}_n^*\|_{\mathcal{S}\times\mathcal{F}} \geq \epsilon) \\
&\leq \limsup_{n\to\infty} \mathbb{P}(\|\overline{\mathbb{G}}_n^*\|_{\mathcal{S}\times\mathcal{F}} \leq \overline{q}_{n,\alpha}^*),
\end{aligned}$$

Taking $\epsilon \to 0$ implies

$$\limsup_{n\to\infty} \mathbb{P}(\|\widehat{\mathbb{G}}_n^*\|_{\mathcal{S}\times\mathcal{F}} < \overline{q}_{n,\alpha}^*) \leq \limsup_{n\to\infty} \mathbb{P}(\|\overline{\mathbb{G}}_n^*\|_{\mathcal{S}\times\mathcal{F}} \leq \overline{q}_{n,\alpha}^*).$$

Next, observe that Giessing (2023, Proposition 1) and our assumption give

$$\liminf_{n\to\infty} \mathbb{V}\mathrm{ar}[\|\overline{\mathbb{Z}}_n^*\|_{\mathcal{S}\times\mathcal{F}}] \gtrsim \liminf_{n\to\infty} \inf_{(s,f)\in\mathcal{S}\times\mathcal{F}} \mathbb{V}\mathrm{ar}[\overline{\mathbb{G}}_n^*(s,f)] > 0.$$

Let n_k be any subsequence of such that $\overline{\mathbb{Z}}_{n_k}^*$ converges weakly to a Gaussian process $\overline{\mathbb{Z}}^*$ with α-quantile q_α. Then the above implies that $\overline{q}_{n_k,\alpha}^* \to q_\alpha > 0$, and $\mathbb{P}(\|\overline{\mathbb{Z}}^*\|_{\mathcal{S}\times\mathcal{F}} = q_\alpha) = 0$. Thus, Proposition 2.14 gives

$$\limsup_{n\to\infty} \mathbb{P}(\|\overline{\mathbb{G}}_n^*\|_{\mathcal{S}\times\mathcal{F}} \leq \overline{q}_{n,\alpha}^*) \leq \limsup_{n\to\infty} \mathbb{P}(\|\overline{\mathbb{Z}}_n^*\|_{\mathcal{S}\times\mathcal{F}} \leq \overline{q}_{n,\alpha}^*) = 1-\alpha.$$

We have shown that

$$\limsup_{n\to\infty} \mathbb{P}(\|\widehat{\mathbb{G}}_n^*\|_{\mathcal{S}\times\mathcal{F}} < \overline{q}_{n,\alpha}^*) \leq 1-\alpha,$$

which implies $\widehat{q}_{n,\alpha}^* \geq \overline{q}_{n,\alpha}^*$ with probability tending to 1. This further implies

$$\liminf_{n\to\infty} \mathbb{P}(\|\mathbb{G}_n\|_{\mathcal{S}\times\mathcal{F}} \leq \widehat{q}_{n,\alpha}^*) \geq \liminf_{n\to\infty} \mathbb{P}(\|\mathbb{G}_n\|_{\mathcal{S}\times\mathcal{F}} \leq \overline{q}_{n,\alpha}^*) \geq \liminf_{n\to\infty} \mathbb{P}(\|\mathbb{Z}_n\|_{\mathcal{S}\times\mathcal{F}} \leq \overline{q}_{n,\alpha}^*),$$

using $\mathbb{G}_n \leftrightarrow_d \mathbb{Z}_n$ and a similar continuity argument as above. Decompose

$$\overline{\mathbb{G}}_n^*(s,f) - \overline{\mathbb{G}}_n^*(t,g) = \mathbb{G}_n(s,f) - \mathbb{G}_n(t,g) + [\overline{\mathbb{G}}_n^*(s,f) - \mathbb{G}_n(s,f) - \overline{\mathbb{G}}_n^*(t,g) + \mathbb{G}_n(t,g)],$$

and observe that $\mathbb{G}_n(s,f) - \mathbb{G}_n(t,g)$ and $[\overline{\mathbb{G}}_n^*(s,f) - \mathbb{G}_n(s,f) - \overline{\mathbb{G}}_n^*(t,g) + \mathbb{G}_n(t,g)]$ are uncorrelated for every $(s,f),(t,g) \in \mathcal{S}\times\mathcal{F}$. Thus,

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}[\mathbb{Z}_n(s,f) - \mathbb{Z}_n(t,g)] \\
&= \mathbb{V}\mathrm{ar}[\mathbb{G}_n(s,f) - \mathbb{G}_n(t,g)] \\
&\leq \mathbb{V}\mathrm{ar}[\mathbb{G}_n(s,f) - \mathbb{G}_n(t,g)] + \mathbb{V}\mathrm{ar}[\mathbb{G}_n^*(s,f) - \mathbb{G}_n^*(t,g) - \mathbb{G}_n(s,f) + \mathbb{G}_n(t,g)] \\
&= \mathbb{V}\mathrm{ar}[\overline{\mathbb{G}}_n^*(s,f) - \overline{\mathbb{G}}_n^*(t,g)] \\
&= \mathbb{V}\mathrm{ar}[\overline{\mathbb{Z}}_n^*(s,f) - \overline{\mathbb{Z}}_n^*(t,g)].
\end{aligned}$$

Since both $\mathbb{Z}_n(s,f)$ and $\overline{\mathbb{Z}}_n^*(s,f)$ are asymptotically tight, they are separable for large enough n. The version of Fernique's inequality given by Ledoux and Talagrand (1991, eq. 3.11 and following paragraph) implies

$$\liminf_{n\to\infty} \mathbb{P}(\|\mathbb{Z}_n\|_{\mathcal{S}\times\mathcal{F}} \leq \overline{q}_{n,\alpha}^*) \geq \liminf_{n\to\infty} \mathbb{P}(\|\overline{\mathbb{Z}}_n^*\|_{\mathcal{S}\times\mathcal{F}} \leq \overline{q}_{n,\alpha}^*) = 1-\alpha.$$

Altogether, we have shown that

$$\liminf_{n\to\infty} \mathbb{P}(\|\mathbb{G}_n\|_{\mathcal{S}\times\mathcal{F}} \leq \widehat{q}_{n,\alpha}^*) \geq 1-\alpha,$$

as claimed. □

D.6. Proof of Proposition 4.7

Proposition 4.4 and (2) imply that $\widehat{q}_{n,\alpha}^* \geq t_n$ with probability tending to 1. Then every subsequence of the sets $S_n = \{z : |z| \leq t_n\}$ converges to $\mathbb{R}$, whose boundary has probability zero under every tight Gaussian law. Proposition 2.14 and Theorem 3.6 give

$$\liminf_{n\to\infty} \mathbb{P}(\|\mathbb{G}_n\|_{\mathcal{S}\times\mathcal{F}} \leq \widehat{q}_{n,\alpha}^*) \geq \liminf_{n\to\infty} \mathbb{P}(\|\mathbb{G}_n\|_{\mathcal{S}\times\mathcal{F}} \leq t_n) \geq \liminf_{n\to\infty} \mathbb{P}(\|\mathbb{Z}_n\|_{\mathcal{S}\times\mathcal{F}} \leq t_n) = 1,$$

as claimed. □

D.7. A useful lemma

Lemma D.3. *Let $V_{n,1},\dots,V_{n,n}$ be a sequence of m_n-dependent random variables with $m_n = o(n^{1/2})$, $\mathbb{E}[V_{n,i}] = 0$, $\mathbb{V}\mathrm{ar}[V_{n,i}] = 1$, and $\sup_{i,n}\mathbb{E}[|V_{n,i}|^a] < \infty$ for any $a \in \mathbb{N}$. Let $\mathcal{F}_n$ be a sequence of functions satisfying conditions (i) and (iv) of Theorem 3.6 with $\gamma = 1$. Then*

$$\sup_{t\in T}\left|\frac{1}{n}\sum_{i=1}^n V_{n,i}\mathbb{E}[f_{n,t}(X_i)]\right| \to_p 0,$$

Proof. Let $N(\varepsilon)$ be the number of ϵ-brackets of $\mathcal{F}_n$ with respect to the $\|\cdot\|_{n,1}$-norm. As in the proof of Theorem 3.2, we can construct a $C\varepsilon$-bracketing with respect to the $\|\cdot\|_{n,1}$-norm for the class

$$\mathcal{G}_n = \{g_{n,t}(v,i) = v\mathbb{E}[f_{n,t}(X_i)] : t \in T\},$$

for some $C < \infty$ and size $N(\varepsilon)$. Let $\mathcal{G}_{n,k} = [\underline{g}_n^{(k)}, \overline{g}_n^{(k)}]$ be the k-th $C\epsilon$-bracket. Recall that

$$\mathbb{P}_n g_{n,t} = \frac{1}{n}\sum_{i=1}^n g_{n,t}(V_{n,i}, i) = \frac{1}{n}\sum_{i=1}^n V_{n,i}\mathbb{E}[f_{n,t}(X_i)],$$

and

$$Pg_{n,t} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[g_{n,t}(V_{n,i}, i)] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[V_{n,i}]\mathbb{E}[f_{n,t}(X_i)] = 0.$$

We thus get

$$\begin{aligned}
&\sup_{t\in T}|\mathbb{P}_n g_{n,t}| \\
&= \sup_{t\in T}|(\mathbb{P}_n - P)g_{n,t}| \\
&\le \max_{1\le k\le N(\varepsilon)}|(\mathbb{P}_n - P)\underline{g}_n^{(k)}| + \sup_{g\in\mathcal{G}_{n,k}}|(\mathbb{P}_n - P)g - (\mathbb{P}_n - P)\underline{g}_n^{(k)}| \\
&\le \max_{1\le k\le N(\varepsilon)}|(\mathbb{P}_n - P)\underline{g}_n^{(k)}| + \sup_{g\in\mathcal{G}_{n,k}}|\mathbb{P}_n(g - \underline{g}_n^{(k)})| + \sup_{g\in\mathcal{G}_{n,k}}|P(g - \underline{g}_n^{(k)})| \\
&\le \max_{1\le k\le N(\varepsilon)}|(\mathbb{P}_n - P)\underline{g}_n^{(k)}| + |\mathbb{P}_n(\overline{g}_n^{(k)} - \underline{g}_n^{(k)})| + \sup_{g\in\mathcal{G}_{n,k}}|P(g - \underline{g}_n^{(k)})| \\
&\le \max_{1\le k\le N(\varepsilon)}|(\mathbb{P}_n - P)\underline{g}_n^{(k)}| + |(\mathbb{P}_n - P)(\overline{g}_n^{(k)} - \underline{g}_n^{(k)})| + 2\sup_{g\in\mathcal{G}_{n,k}}|P(g - \underline{g}_n^{(k)})| \\
&\le 3\max_{1\le k\le N(\varepsilon)}|(\mathbb{P}_n - P)\underline{g}_n^{(k)}| + 2C\varepsilon,
\end{aligned}$$

where in the last step, we assumed without loss of generality that any upper bound of the brackets also appears as a lower bound. By the moment condition on $V_{n,i}$, we have $\max_{1\le i\le n}|V_{n,i}| \le a_n$ with probability tending to 1 for any $a_n \to \infty$ arbitrarily slowly. On this event, Bernstein's inequality Lemma B.2 with $m = m_n + 1$ gives

$$\mathbb{E}\left[\max_{1\le k\le N(\varepsilon)}|(\mathbb{P}_n - P)\underline{g}_n^{(k)}|\right] \lesssim \sqrt{\frac{m_n \ln N(\varepsilon)}{n}} + \frac{a_n m_n \ln N(\varepsilon)}{n},$$

where we used that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{C}\mathrm{ov}[V_{n,i}, V_{n,j}]\mathbb{E}[f_{n,t}(X_i)]\mathbb{E}[f_{n,t}(X_j)] \lesssim \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|\mathbb{C}\mathrm{ov}[V_{n,i}, V_{n,j}]| \lesssim m_n,$$

where the constant depends on the envelope of $\mathcal{F}_n$. Since $\ln N(\varepsilon) \le C\varepsilon^{-2}$, we can choose $\varepsilon = n^{-1/5}$ to get

$$\mathbb{E}\left[\max_{1\le k\le N(\varepsilon)}|(\mathbb{P}_n - P)\underline{g}_n^{(k)}|\right] + \varepsilon \to 0. \qquad \square$$

E. Proofs for the applications

E.1. Proof of Corollary 5.1

Theorem 3.2 implies marginal relative CLTs for every $s \in [0, 1]$. For the uniform result, we apply Theorem 3.6 holds. Condition (i)–(iii) follow immediately from the assumptions. For (iii), consider the class of functions

$$\mathcal{F}_n = \left\{(j, x) \mapsto \frac{1}{b} K\left(\frac{j - sn}{nb}\right) x \colon s \in [0, 1]\right\}.$$

Since the kernel is L-Lipschitz, we have

$$\sup_j \left|\frac{1}{b} K\left(\frac{j - sn}{nb}\right) - \frac{1}{b} K\left(\frac{j - s'n}{nb}\right)\right| \leq \frac{L|s - s'|}{b^2},$$

which also implies

$$\left(\frac{1}{n}\sum_{j=1}^n \mathbb{E}\left[\left|\frac{1}{b} K\left(\frac{j - sn}{nb}\right) X_j - \frac{1}{b} K\left(\frac{j - s'n}{nb}\right) X_j\right|^\gamma\right]\right)^{1/\gamma} \leq \frac{C|s - s'|}{b^2},$$

for $C = L \sup_j \mathbb{E}[|X_i|^\gamma]^{1/\gamma} < \infty$, $\gamma = 5$. Let $s_k = k\epsilon b^2/C$ for $k = 1, \dots, N(\varepsilon)$ with $N(\varepsilon) = \lceil(\epsilon b^2/C)^{-1}\rceil$. Then the functions

$$\underline{K}_k(j) = \frac{1}{b} K\left(\frac{j - s_k n}{nb}\right) - \epsilon/2, \qquad \overline{K}_k(j) = \frac{1}{b} K\left(\frac{j - s_k n}{nb}\right) + \epsilon/2,$$

form an ϵ-bracketing of $\mathcal{F}_n$ with respect to the $\|\cdot\|_{n,\gamma}$-norm. Thus,

$$\int_0^{\delta_n} \sqrt{\ln N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{n,\gamma})} d\epsilon \leq \int_0^{\delta_n} \sqrt{-\ln(\epsilon b^2/C)} d\epsilon \lesssim \delta_n \sqrt{\log \delta_n^{-1}},$$

which implies that condition (iii) of Theorem 3.6 holds. □

E.2. Proof of Corollary 5.2

We first apply Proposition 4.4. By assumption, the function

$$s \mapsto \mu_b(sn) = \frac{1}{nb}\sum_{j=1}^n K\left(\frac{j - sn}{nb'}\right)\mu(j),$$

has two continuous derivatives, uniformly bounded by some $C \in (0, \infty)$. Further, the first two derivatives of K are Lipschitz. Then the same arguments as in the proof of Corollary 5.1 show that

$$\sup_{s \in [0,1]} |\widehat{\mu}_b'(sn)| + \sup_{s \in [0,1]} |\widehat{\mu}_b''(sn)| + \sup_{s \in [0,1]} |\widehat{\mu}_b'''(sn)| \leq 4C.$$

with probability tending to 1. Thus, the functions $\widehat{\mu}_b(sn)$ are included in the smoothness class $C^3_{4C}[0, 1]$ with probability tending to 1. This class has sup-norm log-covering numbers (and, hence, $\|\cdot\|_{n,\gamma}$-log-bracketing numbers for any n, γ) of order $\varepsilon^{-3/2}$ (Van der

Vaart and Wellner, 2023, Theorem 2.7.1). Combining this with the arguments of the previous propostion, we get that

$$\mathcal{F}_n = \left\{ (j, x) \mapsto \frac{1}{b} K\left(\frac{j - sn}{nb}\right)(x - f(j/n)) \colon s \in [0,1], f \in C^3_{4C}[0,1] \right\}$$

has

$$\int_0^{\delta_n} \sqrt{\ln N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{n,\gamma})} d\epsilon \lesssim \int_0^{\delta_n} \varepsilon^{-3/4} d\epsilon \lesssim \delta_n^{1/4} \to 0,$$

for every $\delta_n \to 0$. For

$$\overline{\mu}_b^*(sn) = \frac{1}{nb} \sum_{j=1}^n V_{n,i} K\left(\frac{j - sn}{nb}\right)(X_i - \mu_b(j)),$$

Proposition 4.2 shows that $\sqrt{n}(\widehat{\mu}_b^*(sn) - \overline{\mu}_b^*(sn))$ satisfies a relative CLT in $\ell^\infty([0,1])$. Since further,

$$\sup_j \mathbb{V}\mathrm{ar}\left[\widehat{\mu}_b(j)\right] = O(n^{-1}),$$

the conditions of Proposition 4.4 hold and

$$\sqrt{n} \sup_{s \in [0,1]} |\widehat{\mu}_b^*(sn) - \overline{\mu}_b^*(sn)| \to_p 0.$$

Now consider the process $\sqrt{n}\overline{\mu}_b^*(sn)$. As in the proof of Proposition 4.6, we have

$$\mathbb{V}\mathrm{ar}\left[\sqrt{n}\overline{\mu}_b^*(sn)\right] - \mathbb{V}\mathrm{ar}\left[\sqrt{n}\widehat{\mu}_b(sn)\right] = \overline{\sigma}_n^2(s) \to \infty,$$

for some $s \in [0,1]$, which implies that

$$\mathbb{V}\mathrm{ar}\left[\sqrt{n}\overline{\mu}_b^*(sn)\right] / \overline{\sigma}_n^2(s) \to 1.$$

The univariate relative CLT (Theorem 3.2) with $\alpha > 1/3$ gives

$$\sqrt{n}\overline{\mu}_b^*(sn)/\overline{\sigma}_n(s) \to_d \mathcal{N}(0,1).$$

Since $\sup_{s\in[0,1]} |\overline{\mu}_b^*(sn)| \geq |\overline{\mu}_b^*(sn)|$ for any $s \in [0,1]$, this implies that $\mathbb{P}(\sup_{s\in[0,1]} \sqrt{n}|\overline{\mu}_b^*(sn)| > t_n) \to 1$ for some sequence $t_n \to \infty$. Now the result follows from Proposition 4.7. □

E.3. Proof of Corollary 5.5

Observe that

$$T_n = \sup_{f \in \mathcal{F}, s \in \mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n w_{n,i}(s)(f(Z_i) - \mathbb{E}[f(Z_i)]) \right|,$$

The relative CLT (Theorem 3.6) gives

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n w_{n,i}(s)(f(Z_i) - \mathbb{E}[f(Z_i)]) \leftrightarrow_d \mathbb{Z}_n(s,f) \quad \text{in } \ell^\infty(\mathcal{S}\times\mathcal{F}),$$

where $\{\mathbb{Z}_n(s,f)\colon f \in \mathcal{F}\}$ is a relatively compact, mean-zero Gaussian process. Under the null hypothesis, $\mathbb{E}[f(Z_i)] = 0$ for all $f \in \mathcal{F}$, and the relative bootstrap CLT (Theorem 3.6 and Example 4.3) gives

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n V_{n,i}w_{n,i}(s)(f(Z_i) - \mathbb{E}[f(Z_i)]) \leftrightarrow_d \mathbb{Z}_n(s,f) \quad \text{in } \ell^\infty(\mathcal{S}\times\mathcal{F}),$$

The relative continuous mapping theorem now implies

$$T_n \leftrightarrow_d \sup_{f\in\mathcal{F},s\in\mathcal{S}} |\mathbb{Z}_n(s,f)| \qquad \text{and} \qquad T_n^* \leftrightarrow_d \sup_{f\in\mathcal{F},s\in\mathcal{S}} |\mathbb{Z}_n(s,f)|,$$

which proves that $\mathbb{P}(T_n > c_n^*(\alpha)) \to \alpha$ under H_0.

Under the alternative, we have

$$\begin{aligned} T_n &= \sup_{f\in\mathcal{F},s\in\mathcal{S}} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n w_{n,i}(s)(f(Z_i) - \mathbb{E}[f(Z_i)]) + \frac{1}{\sqrt{n}}\sum_{i=1}^n w_{n,i}(s)\mathbb{E}[f(Z_i)]\right| \\ &\geq \sup_{f\in\mathcal{F},s\in\mathcal{S}} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n w_{n,i}(s)\mathbb{E}[f(Z_i)]\right| - \sup_{f\in\mathcal{F},s\in\mathcal{S}} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n w_{n,i}(s)(f(Z_i) - \mathbb{E}[f(Z_i)])\right|, \end{aligned}$$

which implies

$$T_n/\sqrt{n} \geq \sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n w_{n,i}(s)\mathbb{E}[f(Z_i)]\right| \geq \delta > 0,$$

with probability tending to 1. For the bootstrap statistic,

$$\frac{T_n^*}{\sqrt{n}} \leq \sup_{f\in\mathcal{F},s\in\mathcal{S}} \left|\frac{1}{n}\sum_{i=1}^n V_{n,i}w_{n,i}(s)(f(Z_i) - \mathbb{E}[f(Z_i)])\right| + \sup_{f\in\mathcal{F},s\in\mathcal{S}} \left|\frac{1}{n}\sum_{i=1}^n V_{n,i}w_{n,i}(s)\mathbb{E}[f(Z_i)]\right|.$$

The first term on the right converges to 0 in probability by Example 4.3, the second by Lemma D.3. Thus, $T_n^*/\sqrt{n} \to_p 0$, which implies $c_n^*(\alpha)/\sqrt{n} \to 0$ and the claim follows. □

F. Auxiliary results

Proof of Lemma 2.4. For any $M > 0$ and $t_1, \dots, t_n \in T$ it holds

$$P(\max_{1\leq i\leq n} |\mathbb{G}(t_i)| < M) = \prod_{i=1}^n P(|\mathbb{G}(t_i)| < M) = [\phi(M) - \phi(-M)]^n$$

where ϕ denotes the standard normal CDF. For $n \to \infty$, i.e., if T is infinite, the right side converges to zero. Thus, $P^*(\|\mathbb{G}\|_T < M) = 0$ for all M In other words, $\mathbb{G}$ does not have bounded samples paths. □

F.1. Bracketing numbers under non-stationarity

Fix some triangular array $X_{n,1}, \ldots, X_{n,k_n}$ of random variables with values in a Polish space $\mathcal{X}$. Denote by $\mathcal{F}$ a set of measurable functions $f : \mathcal{X} \to \mathbb{R}$.

Lemma F.1. *Assume that there exists some probability measure Q on $\mathcal{X}$ and a constant $K \in \mathbb{R}$ such that $P_{X_{n,i}}(A) \leq KQ(A)$ for all i and measurable sets A. Then,*

$$\|f\|_{\gamma,n} \leq \sup_{n\in\mathbb{N}, i\leq k_n} \|f(X_{n,i})\|_\gamma \leq K^{1/\gamma}\|f\|_{L_\gamma(Q)}$$

for all $\gamma > 0$. Hence,

$$N_{[]}\left(\epsilon, \mathcal{F}, \|\cdot\|_{\gamma,n}\right) \leq N_{[]}\left(K^{1/p}\epsilon, \mathcal{F}, \|\cdot\|_{L_\gamma(Q)}\right).$$

Proof. The condition

$$P_{X_{n,i}}(A) \leq KQ(A)$$

for all measurable sets A is equivalent to $P_{X_{n,i}}$ being absolutely continuous with respect to Q and all Radon-Nikodyn derivatives are bounded by K, i.e.,

$$\frac{\partial P_{X_{n,i}}}{\partial Q} \leq K$$

(upto zero sets of Q) for all i. For all i it holds

$$\begin{aligned}
\|f(X_{n,i})\|_\gamma^\gamma &= \int |f(X_{n,i})|^\gamma dP \\
&= \int |f|^\gamma \frac{\partial P_{X_{n,i}}}{\partial Q} dQ \\
&\leq \int |f|^\gamma K dQ \\
&= K\|f\|_{L_\gamma(Q)}^\gamma.
\end{aligned}$$

This proves the claim. $\square$