CENTRAL LIMIT THEOREMS FOR EMPIRICAL PROCESSES BASED ON STOCHASTIC PROCESSES

A Dissertation

by

YUPING YANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Joel Zinn
Committee Members,	William B. Johnson
	Dmitry Panchenko
	Daren B.H. Cline
Department Head,	Emil Straube

August 2013

Major Subject: Mathematics

ABSTRACT

In this thesis, we study time-dependent empirical processes, which extend the classical empirical processes to have a time parameter; for example the empirical process for a sequence of independent stochastic processes $\{Y_i : i \in \mathbb{N}\}$:

$$\nu_n(t,y) = n^{-1/2}\sum_{i=1}^{n}[1_{Y_i(t)\le y} - \mathrm{P}(Y_i(t) \le y)], \quad t \in E, \quad y \in \mathbb{R}. \tag{1}$$

In the case of independent identically distributed samples (that is $\{Y_i(t) : i \in \mathbb{N}\}$ are iid), Kuelbs et al. (2013) proved a Central Limit Theorem for $\nu_n(t, y)$ for a large class of stochastic processes.

In Chapter 3, we give a sufficient condition for the weak convergence of the weighted empirical process for iid samples from a uniform process:

$$\alpha_n(t,y) := n^{-1/2}\sum_{i=1}^{n} w(y)(1_{X_i(t)\le y} - y), \quad t \in E, \quad y \in [0,1] \tag{2}$$

where $\{X(t), X_1(t), X_2(t), \cdots\}$ are independent and identically distributed uniform processes (for each $t \in E$, $X(t)$ is uniform on $(0, 1)$) and $w(x)$ is a "weight" function satisfying some regularity properties. Then we give an example when $X(t) := F_t(B_t) : t \in E = [1, 2]$, where B_t is a Brownian motion and F_t is the distribution function of B_t.

In Chapter 4, we investigate the weak convergence of the empirical processes for non-iid samples. We consider the weak convergence of the empirical process:

$$\beta_n(t,y) := n^{-1/2}\sum_{i=1}^{n}(1_{Y_i(t)\le y} - F_i(t,y)), \quad t \in E \subset \mathbb{R}, \quad y \in \mathbb{R} \tag{3}$$

where $\{Y_i(t) : i \in \mathbb{N}\}$ are independent processes and $F_i(t, y)$ is the distribution function of $Y_i(t)$. We also prove that the covariance function of the empirical process for non-iid samples indexed by a uniformly bounded class of functions necessarily uniformly converges to the covariance function of the limiting Gaussian process for a CLT.

To my parents

ACKNOWLEDGMENTS

I would first like to thank my thesis adviser Joel Zinn for his guidance, help and patience during my graduate studies. I also like to thank professors Johnson, Panchenko, and Cline for serving on the committee and their classes I have been in, which broadened my mathematical knowledge. Dr. Panchenko's critical reading of the manuscript leads to great improvement of it. Finally, I appreciate the talks, seminars and workshops in the department.

NOMENCLATURE

$\mathbb{N}$	the set of positive integers $1, 2, 3, \cdots$
$\mathbb{Q}$	the set of the rational numbers
$\mathbb{R}$	the set of the real numbers
iid	independent and identically distributed
a.s.	almost surely
rv	random variable, which is measurable
re	random element, which need not be measurable
$(\Omega, \mathcal{A}, \mathrm{P})$	the underlying probability space; i.e. the canonical product space of the laws of all independent rv's or re's
E	expectation on $(\Omega, \mathcal{A}, \mathrm{P})$
E^*, P^*	outer expectation on $(\Omega, \mathcal{A}, \mathrm{P})$
E_*, P_*	inner expectation on $(\Omega, \mathcal{A}, \mathrm{P})$
X, X_{nj}, $X(t)$, X_t, $X_{nj}(t)$	random variables or processes on $(\Omega, \mathcal{A}, \mathrm{P})$
P, P_{nj}	laws of X, X_{nj} or laws of $X(t)$, $X_{nj}(t)$ if defined
Pf, $P_{nj}f$	$\int f\, dP$, $\int f\, dP_{nj}$
$F(\cdot)$	distribution function of X
$F(t, \cdot)$, $F_t(\cdot)$	distribution function of $X(t)$
$\mathrm{E}(X\|Y)$, $\mathrm{E}_Y(X)$	conditional expectation of X given Y
$\mathcal{F}$	a collection of functions on the sample (state) space
$\ell^\infty(T)$	all bounded functions: $T \to \mathbb{R}$
1_A	indicator function of A
$\nu_n(f)$	$n^{-1/2} \sum_1^n (f(X_j) - \mathrm{E}f(X_j))$
$G_P(f)$, $f \in \mathcal{F}$	a Gaussian process indexed by $\mathcal{F}$ with covariance $\mathrm{P}(fg) - \mathrm{P}f\mathrm{P}g$

$\rho_P(f,g)$	the variance (pseudo) distance $(\mathrm{P}((f-g)-\mathrm{P}(f-g))^2)^{1/2}$
$e_P(f,g)$	the L_2 (pseudo) distance $(\mathrm{P}(f-g)^2)^{1/2}$
$:=$	equal by definition
$\xrightarrow{\mathcal{L}}$; $\Rightarrow$	convergence in law; weak convergence
CLT	Central Limit Theorem
FCLT	Functional Central Limit Theorem
df	distribution function

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Introduction

Empirical process theory is not only a rich part of probability theory as for example it has properties “close” to that of a Brownian motion but also a foundation of many statistical procedures as many statistics, e.g. the Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling statistics, can be seen as functionals of the empirical processes. The weak convergence or the Central Limit Theorem (CLT) says that the empirical process is close to a Gaussian process “in distribution” when n is big enough. Thus in many situations, the empirical process and the limiting Gaussian process can be exchangeable whichever is more convenient in practice.

Let $\{X, X_1, X_2, \cdots\}$ be independent real valued random variables (rv's) with distribution function (df) F. Let

$$F_n(x) := F_n(x;\omega) = \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,x]}(X_i(\omega)), \quad -\infty < x < \infty$$

be the (random) empirical distribution function, which is the distribution function of the (random) empirical measure

$$P_n(\cdot) := P_n(\cdot;\omega) := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}(\cdot)$$

where δ_x assigns mass one at x, i.e. $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ otherwise for any $A \subset \mathbb{R}$. In the literature as well as in this thesis, the ω in $F_n(\cdot;\omega)$ and $P_n(\cdot;\omega)$ will be omitted. The empirical process based on the samples $\{X_1, X_2, \cdots\}$

with common df $F(x)$ is

$$\nu_n(x) := n^{1/2}(F_n(x) - F(x)), \quad -\infty < x < \infty.$$

Given a uniform random variable ξ on $(0,1)$ and a df F, if let

$$F^{-}(t) := \inf\{x : F(x) \geq t\}, \quad t \in (0,1),$$

and $X = F^{-1}(\xi)$, then

$$\{\omega : X \leq x\} = \{\omega : \xi \leq F(x)\}$$

(cf. Shorack and Wellner (1986) p. 4).

Let $\{\xi, \xi_1, \xi_2, \cdots\}$ denote independent uniform $(0,1)$ rv's, and let $U_n(t)$ be the empirical process of these rv's:

$$U_n(t) := n^{-1/2} \sum_{i=1}^{n} [1_{\xi_i \leq t} - t], \quad t \in [0,1],$$

which will be called the uniform empirical process. Then (cf. Shorack and Wellner (1986) p. 4)

THEOREM 1.1.1 (cf. Shorack and Wellner (1986), Theorem 2, p. 4). *The sequence of random functions*

$$\{n^{1/2}(F_n(\cdot) - F(\cdot)) : n = 1, 2, \cdots\}$$

and the sequence of

$$\{U_n(F(\cdot)) : n = 1, 2, \cdots\}$$

on $(-\infty, \infty)$ *have identical probabilistic behavior.*

See more detail in Chapter 4, Section 1.

Hence by replacing the argument in $U_n(\cdot)$ with the deterministic function $F(\cdot)$, we can reduce the empirical process for general iid random variables to the uniform empirical process. In this sense, the uniform empirical process should be of central importance.

Doob's heuristic Doob (1949) says, when n is big enough, one can replace functionals of the process $\{U_n(t) : t \in [0,1]\}$ by the corresponding functionals of the Brownian bridge process $[0,1]$, which is a Gaussian process with covariance $\min(s,t) - st$ for $s,t \in [0,1]$. This is rigorously proved by Donsker (cf. Donsker (1952)).

Dudley's paper Dudley (1978) extended Donsker's theorem to empirical processes indexed by a class, $\mathcal{C}$, of Borel measurable sets:

$$\nu_n(C) := n^{-1/2} \sum_{i \le n} (1_C(X_i) - \mathrm{E}1_C(X_i)),\ C \in \mathcal{C}.$$

In turn it is extended to the so called Functional Central Limit Theorem (FCLT). Let $\{X, X_1, X_2, \cdots\}$ be independent random elements taking values in an arbitrary measure space $(S, \mathcal{S})$ with $P_i = \mathcal{L}(X_i)$. The empirical measure of $\{X, X_1, X_2, \cdots\}$ on S is

$$\mathrm{P}_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i}.$$

The corresponding empirical process is

$$\nu_n = n^{-1/2} \sum_{i=1}^{n} (\delta_{X_i} - P_i).$$

For a function f on S, write $\mu(f) = \int f \, d\mu$ for any measure or signed measure μ.

Further if given a class of functions $\mathcal{F}$ on S, ν_n can be viewed as a process indexed by $\mathcal{F}$:

$$\nu_n(f) = n^{-1/2} \sum_{i=1}^{n} (f(X_i) - \mathrm{E}f(X_i)), \quad f \in \mathcal{F}.$$

For example, for the set indexed empirical process, the class of functions is $\mathcal{F} := \{1_C : C \in \mathcal{C}\}$; and for the classical empirical process

$$n^{-1/2} \sum_{i=1}^{n} (1_{(-\infty,x]}(X_i) - F(x)),$$

for iid samples $\{X_1, X_2, \cdots\}$ with common df $F(x)$, the class of functions is $\mathcal{F} = \{1_{(-\infty,x]} : x \in \mathbb{R}\}$. We will review some FCLT results in Chapter 2, Section 4.

Another important aspect of the classical empirical process theory is the weighted empirical process:

$$w_n(t) := n^{-1/2} \sum_{i=1}^{n} w(t)(1_{[0,t]}(U_i) - t), \quad t \in [0,1]$$

where $w(t)$ is a positive function on $(0,1)$, symmetric about $1/2$, non-increasing near 0, and $\lim_{x\downarrow 0} w(x) = \lim_{x\uparrow 1} w(x) = \infty$ and where the rv's $\{U_1, U_2, \cdots\}$ are iid uniform $(0,1)$. The weighted empirical process is, for example, related to the Anderson-Darling statistic:

$$\int_{-\infty}^{\infty} n(F_n(x) - F(x))^2 w(x)^2 \, dx$$

where $w(x) = (F(x)(1 - F(x)))^{-1/2}$. When $w(x) \equiv 1$, we get the Cramér-von Mises statistic.

In the probability aspect, the weighted empirical process is related to the upper class functions for Brownian motion. Let $q(t)$ (throughout this thesis $w(t) = 1/q(t)$)

be a function on $(0, 1/2]$ such that $\inf_{\delta \le s \le 1/2} q(s) > 0$ for all $0 < \delta < 1/2$ (then we call it positive) and non-decreasing in a neighborhood of 0. Let

$$\beta := \limsup_{t \downarrow 0} |B(t)|/q(t). \tag{1.1}$$

By Blumenthal's 0-1 law (cf. Itô and McKean (1965), p. 25), β is a constant a.s. If $\beta < \infty$, then $q(t)$ is called a Erdös-Feller-Kolmogorov-Petrovski (EFKP) upper-class function of the Brownian motion (equivalently of the Brownian bridge) at 0 (cf. Csörgő et al. (1986)); if $\beta = 0$, then $q(t)$ is called a Chibisov-O'Reilly function (weight function). If let $\{B(t) : t \in [0, 1]\}$ be the Brownian bridge, then $B(t)$ has variance $t(1-t)$, hence near 0, we can roughly say it grows about $t^{1/2}$. One question is to characterize the class of functions $q(t)$ such that the weighted Brownian bridge $B(t)/q(t)$ is continuous (so called pre-Gaussian property for the CLT of the weighted empirical process). Since the uniform empirical process, say $\{U_n(t) : t \in [0, 1]\}$, converges weakly to $B(t)$, another question is to ask whether such a $q(t)$ also suffice the weak convergence of $U_n(t)/q(t)$ to $B(t)/q(t)$? These questions are first investigated Čibisov (1964b)), then O'Reilly (1974), and many others, e.g. Einmahl et al. (1988), Alexander (1987b), Csörgő et al. (1986). In the literature, it is called the Chibisov-O'Reilly theorem. There are various results by many authors; Theorem 4.2.1 (3.1.9) in Csörgő et al. (1986) gives a unified answer. we will review them in Chapter 3, Section 1.

1.2 Summary of the thesis

In many situations, the samples (data) are functions as in the stock market. In [Swa07], Swanson proved the following. If we write $M_N(t)$ for the median of N independent standard Brownian motion, then $N^{1/2} M_N(t)$ converges weakly to a continuous Gaussian process $\{G_t : 0 \le t < \infty\}$ with covariance

$$\mathrm{E}G_sG_t = \sin^{-1}(\frac{s \wedge t}{\sqrt{st}}).$$

Inspired by this result, Kuelbs et al. (2013) considered a empirical process involving time dependent data. Given independent and identically distributed stochastic processes $\{Y(t), Y_1(t), Y_2(t), \cdots\}$ for $t \in E$, form the empirical process

$$\nu_n(t, y) = n^{-1/2} \sum_1^n (1_{Y_i(t) \le y} - \mathrm{P}(Y_i(t) \le y)), \quad t \in E, \quad y \in \mathbb{R}.$$

If the process $Y(t)$ satisfies the so called L-condition (see Definition 3.2.5), then $\nu_n(t, y)$ converges weakly to a mean zero Gaussian process with covariance

$$\mathrm{E}G(s, x)G(t, y) = \mathrm{P}(Y_s \le x, Y_t \le y) - \mathrm{P}(Y_s \le x)\mathrm{P}(Y_t \le y).$$

As in the classical empirical process theory, there are two important questions that one can study: one is the weighted empirical process and the other is the empirical process for non-iid samples. Thus in this thesis, we pursue in these two directions but for time-dependent data. We call a process $\{X_t : t \in E\}$ a uniform process if $X(t)$ is uniformly distributed on $(0, 1)$ for each $t \in E$. We consider the time dependent weighted empirical process:

$$\alpha_n(t, y) := n^{-1/2} \sum_{i=1}^n w(y)(1_{X_i(t) \le y} - y), \quad t \in E, \quad y \in [0, 1] \tag{1.2}$$

where $\{X(t), X_1(t), X_2(t), \cdots\}$ are independent and identically distributed uniform processes. We proved that (Theorem 3.3.3) given iid uniform processes $\{X(t), X_1(t), X_2(t), \cdots\}$ on E and a weight function $w(y)$, under the WL-condition (see Theorem 3.3.3) on the process and the weight function, the empirical process $\alpha_n(t, y)$ converges weakly to a mean zero Gaussian process G_w with covariance

$$\mathrm{E}G_w(s, x)G_w(t, y) = w(x)w(y)[\mathrm{P}(X_s \leq x, X_t \leq y) - xy].$$

When we specialize that $w(x) \equiv 1$, our result gives a different proof of Theorem 3 in Kuelbs et al. (2013) (Corollary 3.3.7) provided the df $F_t(\cdot)$ of X_t for each $t \in E$ is strictly increasing; without this assumption the CLT only holds in $\ell^\infty(T_0)$ for any countable set $T_0 \subset E \times \mathbb{R}$.

In the second part of this thesis, we consider non-iid samples. Since there is no canonical law as in the iid case, which is the common law P, the average of the first n individual laws $\{P_j : 1 \leq j \leq n\}$ will take this role. We usually need a uniform bound in n on these averages to suffice a CLT. Given independent processes $\{Y_1(t), Y_2(t), \cdot\}$ on E, let

$$\nu_{n,t}(\cdot) = n^{-1} \sum_{j=1}^{n} \mathrm{P}(Y_j(t) \leq \cdot).$$

If each Y_j satisfies the L-condition in addition to some assumptions about $\nu_{n,t}(\cdot)$ for all $t \in E$ ($\nu_{n,t}$-condition), a CLT for the time-dependent empirical process for non-iid samples (Theorem 4.2.1) follows.

In this section, we also prove it is necessary for the weak convergence of empirical processes for non-iid samples indexed by a uniformly bounded class of functions that the covariance function of the empirical process converges uniformly to that of the limiting process.

Next we briefly discuss the proofs. The definition of weak convergence we will use is as in Hoffmann-Jørgensen's theory about general convergence of stochastic processes (Hoffmann-Jørgensen (1991)), which uses outer expectation with respect to the underlying probability space (cf. Andersen et al. (1988), p. 274). The definition will be introduced in detail in the second section. In our setting, the sample space is $\ell^\infty(E \times \mathbb{R})$ of all bounded functions on $E \times \mathbb{R}$ and the limiting process takes values with probability one in the separable space $C_b^u(E \times \mathbb{R})$ of all bounded and uniformly continuous functions with respect to the L_2 (pseudo) metric of the limiting process on $E \times \mathbb{R}$.

We will use two theorems in Andersen et al. (1988) to prove our results; one is for iid samples (Andersen et al. (1988), Theorem 4.4) and the other is for non-iid samples (Andersen et al. (1988), Theorem 4.1). They are given in Chapter 2, Section 5. According to them, two main parts need to be established.

In order to have a CLT (weak convergence to a Gaussian limit), "a limit" has to exist. One main part of the proof of a CLT is to establish some regularity property of the limiting Gaussian process, which is called pre-Gaussian property (or continuity of a Gaussian process); see precise definitions in Chapter 2. There are different methods for continuity of a Gaussian process, for example entropy methods (cf. Dudley (1967)), majorizing measures (cf. Fernique (1975b)), and generic chaining (cf. Talagrand (2005)). Generic Chaining method is a metric characterization of continuity of a Gaussian process (see Theorem 2.3.6). This characterization gives a comparison theorem (Theorem 2.3.7), which roughly says if a Gaussian process is dominated by a continuous Gaussian process, then it is continuous. In the proofs of our main results, we find continuous dominating Gaussian processes under assumptions (like L-condition, WL-condition or "$\nu_{n,t}$-condition"). Hence the comparison theorem gives the continuity of the limiting Gaussian processes of the time dependent empirical processes; see Theorem 3.3.9, which gives a simpler proof of the pre-Gaussian property than in Kuelbs et al. (2013), and Corollary 3.3.13, and Lemma 4.2.9.

The other is some form of small oscillation of the index set (or finite approximation of it), which is in Andersen et al. (1988) called the local modulus condition. It roughly says the weak L_2 norm of the local modulus (not uniform modulus) on each "ε-Gaussian ball" (ball in the index set of radius ε with respect to a continuous Gaussian metric) is $O(\varepsilon)$. Under the assumptions (the same L-condition, WL-condition, or "$\nu_{n,t}$-condition"!) of our theorems, there exist continuous Gaussian distances and with respect to which the local modulus condition holds.

2 OVERVIEW OF FUNCTIONAL CENTRAL LIMIT THEOREM

2.1 Weak convergence in $\ell^\infty(T)$

The definition of weak convergence we will use is a special case of weak convergence in $\ell^\infty(T)$ of all bounded functions on an arbitrary set T, endowed with the uniform norm: for $z \in \ell^\infty(T)$

$$\|z\| := \|z\|_T := \sup_{t\in T} |z(t)|.$$

Let $f : (\Omega, \mathcal{A}, \mathrm{P}) \to \mathbb{R}$ be any function. Define the outer integral of f by

$$\mathrm{P}^* f = \inf\{\mathrm{P}g : g \geq f \text{ and } g \text{ is } (\mathcal{A}, \mathcal{B}(\mathbb{R})) \text{ measurable}\},$$

and the inner integral

$$\mathrm{P}_* f = \sup\{\mathrm{P}g : g \leq f \text{ and } g \text{ is } (\mathcal{A}, \mathcal{B}(\mathbb{R})) \text{ measurable}\}.$$

For any set $A \subset \Omega$, define the outer probability $\mathrm{P}^*(A) = \mathrm{P}^*(1_A)$ and inner probability $\mathrm{P}_*(A) = \mathrm{P}_*(1_A)$.

Since $(\ell^\infty(T), \|\cdot\|)$ is generally non-separable, we first review some definitions and results about weak convergence in a (generally non-separable) metric space, say (S,d). Let $\mathcal{S}_0$ be the σ-algebra generated by open ball in (S,d). Let $\mathcal{B}(S)$ be the Borel σ-algebra (generated by open sets in (S,d)). Dudley (1966) and Wichura (1968) defined the weak convergence of a net of probability measures defined on different σ-algebras, say P_α defined on $\mathcal{S}_\alpha$ with $\mathcal{S}_0 \subset \mathcal{S}_\alpha \subset \mathcal{B}(S)$, to a tight Borel measure P_0

on (S,d) if

$$\lim_{\alpha} P_{\alpha}^{*}(f) = P_0(f) = \lim_{\alpha}(P_{\alpha})_{*}(f)$$

for each bounded continuous $f : S \to \mathbb{R}$ where $P_{\alpha}^{*}(f)$ and $(P_{\alpha})_{*}(f)$ are the outer and inner integral of f respect to P_{α}.

Hoffmann-Jørgensen (1991) suggested a weak convergence for random elements in (S,d).

DEFINITION 2.1.1 (cf. Gaenssler (1992)). Let (S,d) be an arbitrary (non-separable in general) metric space and $(\Omega_n, \mathcal{A}_n, \mathrm{P}_n)$ be a sequence of probability spaces. Further for $n = 1, 2, \cdots$ let $\eta_n : \Omega_n \to S$ be arbitrary maps and $\eta_0 : \Omega_0 \to S$ be $(\mathcal{A}_0, \mathcal{B}(S))$-measurable (where $\mathcal{B}(S)$ denotes the Borel-σ-algebra in (S,d)).
Let $C_b(S) := \{f : S \to \mathbb{R} : f$ bounded and d-continuous$\}$. Then

$$\eta_n \xrightarrow{\mathcal{L}} \eta_0 \text{ if and only if } \lim_{n\to\infty} \mathrm{E}^{*} f(\eta_n) = \mathrm{E} f(\eta_0) \text{ for all } f \in C_b(S).$$

If, in addition, for some separable subspace S_0 of S with $S_0 \in \mathcal{B}(S), \mathrm{P}(\eta_0 \in S_0) = 1$, we write $\eta_n \underset{sep}{\xrightarrow{\mathcal{L}}} \eta_0$ or $\eta_n \Rightarrow \eta_0$ and say η_n converges weakly to η_0. For weak convergence of the empirical measure, we always require the limiting map η_0 with probability one taking values in a separable subspace $S_0 \subset S$.

In this thesis, the metric space (S,d) is $(\ell^{\infty}(T), \|\cdot\|)$.

In this section, $X_n : (\Omega_n, \mathcal{A}_n, \mathrm{P}_n) \to (S,d)$ for $n = 1, 2, \cdots$ will denote a sequence of arbitrary maps, where $(\Omega_n, \mathcal{A}_n, \mathrm{P}_n)$ is a probability space and (S,d) is a metric space. We use E^{*} or P^{*} denote the outer expectation on $(\Omega_n, \mathcal{A}_n, \mathrm{P}_n)$ instead of P_n^{*} if no confusion occurs. The same applies to E_{*} or P_{*}.

To characterize the weak convergence in $(\ell^{\infty}(T), \|\cdot\|)$, we need a concept of tightness of a family of random elements.

Definition 2.1.2. A sequence of random elements X_n in (S,d) is asymptotically tight if for every $\varepsilon > 0$ there exists a compact set K such that

$$\liminf_{n\to\infty} \mathrm{P}_*(X_n \in K^\delta) \geq 1-\varepsilon, \qquad \text{for every } \delta > 0.$$

Here $K^\delta = \{y \in S : d(y,K) < \delta\}$.

In the classical theory when (S,d) is $C([0,1])$ with uniform metric or $D([0,1])$ with Skorokhod metric, the uniform tightness is used instead.

Definition 2.1.3. A sequence of Borel measurable maps X_n in (S,d) is uniformly tight if for every $\varepsilon > 0$ there exists a compact set K such that

$$\mathrm{P}(X_n \in K) \geq 1-\varepsilon, \qquad \text{for all } n.$$

Remark 2.1.4 (see van der Vaart and Wellner (1996), p. 27 Exercise 9). For Borel measurable sequences in a Polish space, uniform tightness and asymptotic tightness are the same.

Theorem 2.1.5 (cf. van der Vaart and Wellner (1996), Theorem 1.5.4). *Let $X_n : \Omega_n \to \ell^\infty(T)$ be arbitrary. Then X_n converges weakly to a tight limit if and only if X_n is asymptotically tight and the marginals $X_n(t_1), \cdots, X_n(t_k)$ converge weakly to a limit for ever finite subset $t_1, \cdots, t_k$ of T. If X_n is asymptotically tight and its marginals converges weakly to the marginals $(X(t_1), \cdots, X(t_k))$ of a stochastic process X, then there is a version of X with uniformly bounded sample paths and X_n converges weakly to X.*

To prove the weak convergence, the main task is to establish tightness. The following are two characterizations.

THEOREM 2.1.6 (cf. van der Vaart and Wellner (1996), Theorem 1.5.6). *A sequence $X_n : \Omega_n \mapsto \ell^\infty(T)$ is asymptotically tight if and only if $X_n(t)$ is asymptotically tight in $\mathbb{R}$ for every t and, for every $\varepsilon > 0$, $\eta > 0$, there exists a finite partition $T = \cup_{i=1}^k T_i$ such that*

$$\limsup_n \mathrm{P}^*[\sup_i \sup_{s,t\in T_i} |X_n(s) - X_n(t)| > \varepsilon] < \eta.$$

THEOREM 2.1.7 (cf. van der Vaart and Wellner (1996), Theorem 1.5.7). *A sequence $X_n : \Omega_n \mapsto \ell^\infty(T)$ is asymptotically tight if and only if $X_n(t)$ is asymptotically tight in $\mathbb{R}$ for every t and there exists a totally bounded pseudo-metric ρ on T such that the asymptotically uniformly ρ-equicontinuous in probability condition: for every $\varepsilon > 0$, $\eta > 0$, there is a $\delta > 0$ so that*

$$\limsup_n \mathrm{P}^*(\sup_{\rho(s,t)<\delta} |X_n(s) - X_n(t)| > \varepsilon) < \eta$$

holds.

Since we only consider the limiting process, say X, is Gaussian (i.e. the marginals $(X(t_1), \cdots, X(t_k))$ for any finite subset $t_1, \cdots, t_k$ of T are jointly normal), we have the following specific characterization.

In the following, $\rho_p(s,t) := (\mathrm{E}|X(s) - X(t)|^p)^{1/(p\vee 1)}$ for $0 < p < \infty$, where X is understood as the limiting process.

THEOREM 2.1.8 (van der Vaart and Wellner (1996), p. 40 Example 1.5.10). *Let X be a Gaussian process and let X_n be a sequence of random elements with values in $\ell^\infty(T)$. Then there exists a version of X which is a tight Borel measurable map into $\ell^\infty(T)$, and X_n converges weakly to X if and only if for some p (and then for all p):*

(i) the marginals of X_n converges weakly to the corresponding marginals of X;

(ii) X_n is asymptotically equicontinuous in probability with respect to ρ_p: for every $\varepsilon, \eta > 0$, there exists a $\delta > 0$ such that

$$\limsup_n \mathrm{P}^*\left(\sup_{\rho_p(s,t)<\delta} |X_n(s) - X_n(t)| > \varepsilon \right) < \eta.$$

(iii) T is totally bounded for ρ_p

THEOREM 2.1.9 (van der Vaart and Wellner (1996), p. 41). *A Gaussian process X in $\ell^\infty(T)$ is tight if and only if (T, ρ_p) is totally bounded and almost all paths $t \to X(t,\omega)$ are uniformly ρ_p-continuous for some p and (and then for all p).*

2.2 Weak convergence of $\nu_n(\cdot)$ in $\ell^\infty(\mathcal{F})$

As in the introduction, the general empirical process, $\nu_n(\cdot)$, indexed by a classes of functions, can be seen as an random element in $\ell^\infty(\mathcal{F})$, thus the general theory of weak convergence in $\ell^\infty(T)$ in the previous section applies to the sequence of random elements $(\nu_n(\cdot))_{n\geq 1}$ in $\ell^\infty(\mathcal{F})$.

Since we always work in this setting, we state the definition of weak convergence separately following the definitions as in Andersen et al. (1988).

Let $(S, \mathcal{S})$ be a measure space. Let $\{P_{nj} : j = 1, \cdots, n, n \in \mathbb{N}\}$ be probability measures on $(S, \mathcal{S})$ and let $\mathcal{F} \subset \bigcap_{n,j} \mathcal{L}_1(S, \mathcal{S}, P_{nj})$ such that

$$\sup_{f\in\mathcal{F}} |f(s)| < \infty \qquad \text{for all } s \in S.$$

Let also $(\Omega_n, \Sigma_n, \mathrm{Pr}_n) = (S^n, \mathcal{S}^n, P_{n1} \otimes \cdots \otimes P_{nn}) \times ([0,1], \mathcal{B}, \lambda)$, let $X_{nj} : \Omega_n \to S$ be the coordinate projections and let $\{a_n\}$ be a sequence of real positive numbers. For the following, let E^* or P^* denotes Pr_n^*, the outer expectation on $(\Omega_n, \Sigma_n, \mathrm{Pr}_n)$.

Recall

$$\mathrm{Pr}_n^*(f) := \inf\{\mathrm{Pr}_n(g) : g \geq f,\ g \text{ is } (\Sigma_n, \mathcal{B}(\mathbb{R})) \text{ measurable }\}.$$

Then:

Definition 2.2.1. $\mathcal{F}$ satisfies the CLT with centering at expectation with respect to $\{P_{nj}; a_n\}$ – $\mathcal{F} \in \mathrm{CLT}\{P_{nj}; a_n\}$ for short – if there exists a (centered) Radon (tight) measure γ on $\ell^\infty(\mathcal{F})$ such that for all $H : \ell^\infty(\mathcal{F}) \to \mathbb{R}$ bounded and continuous,

$$\mathrm{E}^*(H(a_n^{-1}\sum_{j=1}^{n}(\delta_{X_{nj}} - P_{nj})) \underset{n\to\infty}{\to} \int H\, d\gamma.$$

In this case, we say $\mathcal{F}$ is $\{P_{nj}; a_n\}$-Donsker; say $\mathcal{F}$ is P-Donsker for iid case with $a_n = n^{1/2}$ and write $\mathcal{F} \in \mathrm{CLT}(P)$.

For the following $\rho_P(f,g) := (P((f - g - P(f-g))^2)^{1/2}$. Following is a consequence in the context of FCLT of Theorem 2.1.8 with $p = 2$. Also see Giné and Zinn (1984), Theorem 2.12.

Theorem 2.2.2. *The following are equivalent*

(*i*) $\mathcal{F}$ *is* P*-Donsker*

(*ii*) $(\mathcal{F}, \rho_P)$ *is totally bounded and* ν_n *is asymptotically uniformly* ρ_P*-equicontinuous in probability with respect to* ρ_P*: for every* $\varepsilon > 0$

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \mathrm{P}^*(\sup_{\rho_P(f,g)<\delta} |\nu_n(f) - \nu_n(g)| > \varepsilon) = 0$$

(iii) *$(\mathcal{F}, \rho_P)$ is totally bounded and ν_n is asymptotically uniformly ρ_P-equicontinuous in mean with respect to ρ_P:*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathrm{E}^*[\sup_{\rho_P(f,g)<\delta} |\nu_n(f) - \nu_n(g)|] = 0$$

2.3 Pre-Gaussian

In order to have a FCLT (a weak convergence to a limit of a mean zero Gaussian process), a limit, say $G := \{G(f) : f \in \mathcal{F}\}$, with some regularity property, has to exist at first place. The covariance and L_2 distance of the limiting Gaussian process are the corresponding limits of the covariance and L_2 distance of the empirical process as $n \to \infty$. For iid samples with law P, the covariance and square of the L_2 distance of G are

$$\mathrm{E}G(f)G(g) = P(fg) - PfPg,$$
$$\mathrm{E}(G(f) - G(g))^2 = P[f - g - P(f - g)]^2;$$

and for the non-iid samples with laws $\{P_1, P_2, \cdots\}$, they are

$$\mathrm{E}G(f)G(g) = n^{-1} \sum_{i=1}^{n} [P_i(fg) - P_i f P_i g],$$
$$\mathrm{E}(G(f) - G(g))^2 = n^{-1} \sum_{i=1}^{n} [P_i(f - g - P(f - g))^2].$$

If a covariance function is given, by Kolmogorov's consistency theorem, we can construct a Gaussian process with the given covariance. If two Gaussian processes have the same covariance function, then we say they are versions to each other.

Given a law P on a measure space $(S, \mathcal{S})$, recall $\rho_P(f, g) := (P(f - g - P(f -$

$g))^2)^{1/2}$ for $f, g \in \mathcal{L}^2(S, P)$.

DEFINITION 2.3.1. We say $\mathcal{F} \subset L_2(P)$ is P-pre-Gaussian if

(i) the mean zero Gaussian process $\{G(f) : f \in \mathcal{F}\}$ with covariance $\mathrm{P}fg - \mathrm{P}f\mathrm{P}g$ has a version with all the sample functions bounded and uniformly ρ_P-continuous and

(ii) the pseudo-metric space $(\mathcal{F}, \rho_P)$ is totally bounded.

We say such a Gaussian process is continuous (or tight); and we say its L_2 distance $d_G(f, g) := (\mathrm{E}(G(f) - G(g))^2)^{1/2}$ is a continuous (or tight) Gaussian distance.

DEFINITION 2.3.2. A Radon (tight) measure μ on $\ell^\infty(\mathcal{F})$ is a finite Borel measure μ which is regular from below by compacts, i.e. $\mu(B) = \sup\{\mu(K) : K \subset \ell^\infty(\mathcal{F}), K \text{ is compact}\}$.

This is a rephrase in the context of FCLT of Theorem 2.1.9; also cf. Proposition 2.3.5 and Theorem 2.3.4.

THEOREM 2.3.3. *Let $\mathcal{F} \subset L_2(P)$. The following are equivalent.*

(i) $\mathcal{F}$ is P-pre-Gaussian.

(ii) the mean zero Gaussian process $\{G(f) : f \in \mathcal{F}\}$ with covariance $\mathrm{P}fg - \mathrm{P}f\mathrm{P}g$ is tight in $\ell^\infty(\mathcal{F})$; that is for every $\varepsilon > 0$, there is a compact set $K \subset \ell^\infty(\mathcal{F})$ such that $\mathrm{P}(G \in K) \geq 1 - \varepsilon$.

The following is a criterion for a subset in $\ell^\infty(T, \rho)$ with ρ totally bounded to be compact.

THEOREM 2.3.4 (Arzela-Ascoli; cf. van der Vaart and Wellner (1996), p. 41 Exercise 1). *Let (T, ρ) be a totally bounded (pseudo) metric space and $K \subset \ell^\infty(T)$.*

Suppose that for some $\varepsilon > 0$, $\delta > 0$ and every $z \in K$: $|z(s) - z(t)| < \varepsilon$ whenever $\rho(s,t) < \delta$. Moreover, suppose $\{z(t) : z \in K\}$ is bounded for every t. Then

(*i*) *K is uniformly bounded;*

(*ii*) *K can be covered with finitely many balls of radius 2ε with centers in $C^u(T,\rho)$, where $C^u(T,\rho)$ is the class of all uniformly continuous functions on (T,ρ).*

Consequently, a uniformly bounded, uniformly ρ-equicontinuous subset of $\ell^\infty(T)$ is totally bounded.

It follows that

PROPOSITION 2.3.5 (cf. Gaenssler and Schneemeier (1990), Corollary 2). *Let (S,d) be a (pseudo) metric space and $C_b^u(S,d)$ be the set of all bounded and uniformly continuous functions on (S,d). Define $\|z\|_S = \sup_s |z(s)|$ for $z \in C_b^u(S,d)$. Then $(C_b^u(S,d), \|\cdot\|_S)$ is separable if and only if (S,d) is totally bounded.*

To determine a mean zero Gaussian process has a version with bounded and uniform continuous sample paths a.s. with respect to its L_2 metric, the following is a metric characterization. We call a sequence of partitions $\{\mathcal{A}_n : n \in \mathbb{N}\}$ of a set T is admissible if the cardinality, $|\mathcal{A}_n|$, of $\mathcal{A}_n$ is less or equal 2^{2^n} and the partitions are increasing. Let $\Delta_d(A_n(t))$ be the diameter of the unique element in $\mathcal{A}_n$ which contains t.

THEOREM 2.3.6 (Talagrand (2005)). *Let $\{X_t : t \in T\}$ be a centered Gaussian process with L_2 metric d_X where T is countable, and (T,d_X) is totally bounded. Then the following are equivalent:*

(*i*) *X_t is uniformly continuous on (T,d_X) with probability one.*

(*ii*) $\lim_{\varepsilon\to 0} \mathrm{E}(\sup_{d(s,t)\le\varepsilon} (X_s - X_t)) = 0$.

(iii) There exists an admissible sequence of partitions of T such that

$$\lim_{k\to\infty} \sup_{t\in T} \sum_{n\geq k} 2^{n/2}\Delta_{d_X}(A_n(t)) = 0.$$

For a complete proof of this theorem, see Kuelbs et al. (2013), Theorem 4.

Using this characterization, the following comparison theorem is immediate, which gives a criterion of pre-Gaussian property.

THEOREM 2.3.7 (cf. Kuelbs et al. (2013), Proposition 1). *Let H_1 and H_2 be mean zero Gaussian processes with L_2 distances d_{H_1}, d_{H_2}, respectively, on T. Furthermore, assume T is countable, and $d_{H_1}(s,t) \leq d_{H_2}(s,t)$ for all $s,t \in T$. Then, H_2 sample bounded and uniformly continuous on (T, d_{H_2}) with probability one, implies H_1 is sample bounded and uniformly continuous on (T, d_{H_1}) with probability one.*

The assumption that T is countable is not much a restriction provided that T is given a totally bounded metric. We call following the extension lemma.

LEMMA 2.3.8. *Let $\{G(t) : t \in T\}$ be a mean zero Gaussian process with covariance function ρ. Further assume $\sup_{t\in T} \mathrm{E}G(t)^2 < \infty$. Let $d_G(s,t) := d_\rho(s,t) := (\mathrm{E}(G(s) - G(t))^2)^{1/2}$. Then, if T_0 is a dense set in (T, d_G) and the restricted process $\{G(t) : t \in T_0\}$ is sample bounded and uniformly d_G-continuous, then $\{G(t) : t \in T\}$ has a version with bounded and uniformly d_G-continuous sample paths.*

Before the proof of this lemma, we need the following.

LEMMA 2.3.9. *Let $\{L_t : t \in T\}$ be a centered Gaussian process with covariance function $\gamma(s,t)$. Then*

$$|\gamma(f,g) - \gamma(h,l)| \leq M(d_L(f,h) + d_L(g,l)),$$

where $M = \sup_{t\in T}(\mathrm{E}L_t^2)^{1/2}$.

PROOF. First note that $d_G^2(f,g) = \gamma(f,f) - 2\gamma(f,g) + \gamma(g,g)$.

$$\begin{aligned}
|\gamma(f,g) - \gamma(h,l)| &= |\mathrm{E}L(f)L(g) - \mathrm{E}L(h)L(l)| \\
&= |\mathrm{E}L(f)L(g) - \mathrm{E}L(h)L(g) + \mathrm{E}L(h)L(g) - \mathrm{E}L(h)L(l)| \\
&\le |\mathrm{E}L(f)L(g) - \mathrm{E}L(h)L(g)| + \mathrm{E}|L(h)L(g) - \mathrm{E}L(h)L(l)| \\
&\le (\mathrm{E}L(g)^2)^{1/2}(\mathrm{E}(L(f) - L(h))^2)^{1/2} \\
&\quad + (\mathrm{E}L(h)^2)^{1/2}(\mathrm{E}(L(g) - L(l))^2)^{1/2} \\
&\le M(d_L(f,h) + d_L(g,l)). \qquad \square
\end{aligned}$$

PROOF OF LEMMA 2.3.8. We denote the restricted process $\{G(t) : t \in T_0\}$ by G_0. Then almost surely its sample paths are uniformly continuous on T_0. Each sample path can be extended to a uniformly continuous sample path on T. Indeed, if let $G_0(\omega)$ is a sample path and $t \in T$, then there is a sequence, say $(t_m) \subset T_0$, such that $d_G(t_m, t) \to 0$ as $m \to \infty$ and define $\tilde{G}(t)(\omega) := \lim_{m\to\infty} G(t_m)(\omega)$. It's easy to see it's well defined and is uniformly d_G continuous on T. Let $\tilde{\rho}$ is the covariance of $\tilde{G}$. It remains to show $\rho = \tilde{\rho}$. But that ρ and $\tilde{\rho}$ coincide on $T_0 \times T_0$ together with Lemma 2.3.9 implies they coincide on $T \times T$. $\square$

For the boundedness and continuity of a Gaussian process, see, e.g., Dudley (1967), Dudley (1978), Fernique (1975a), Talagrand (2005) etc.

2.4 Finite Approximation and some Theorems about FCLT

For a FCLT to hold, the size of the class of functions can't be too big. We review some different ways to control the size and the related results.

2.4.1 Vapnik-Červonenkis classes

Let $(S, \mathcal{S})$ be a measurable space. Let $\mathcal{C} \subset \mathcal{S}$. We recall $\mathcal{C} \subset \mathcal{S}$ is a Vapnik-Červonenkis (VC) class if the VC dimension of the class:

$$V(\mathcal{C}) := \inf\{n \in \mathbb{N} : \text{ for any } n\text{-subset } F \subset S, |F \cap \mathcal{C}| < 2^n\}$$

is finite.

THEOREM 2.4.1. *If $V(\mathcal{C}) < \infty$, then for any probability measure on $(S, \mathcal{S})$, $\mathcal{C}$ is a P-Donsker class.*

2.4.2 Entropy conditions

Let P be a probability on the measure space $(S, \mathcal{S})$. Let $\|f\|_r := (\int |f|^r \, dP)^{1/r}$. Define the covering number of $\mathcal{F}$ with respect to $\|\cdot\|_{L_r(P)} := \|\cdot\|_r$

$$N_r(\varepsilon, \mathcal{F}, P) := \min\{k : \text{ for some } f_1, \cdots, f_k \in L_r(P), \min_{1 \le i \le k} \|f - f_i\|_r < \varepsilon \text{ for all } f \in \mathcal{F}\}.$$

Then function

$$H_r(\varepsilon) := H_r(\varepsilon, \mathcal{F}, P) := \log N_r(\varepsilon, \mathcal{F}, P)$$

is called the metric entropy of $\mathcal{F}$ in $L_r(P)$.

Another type of entropy of $\mathcal{F}$ is called metric entropy with bracketing. For $f, g \in L_0(S, \mathcal{S})$, the set of all real-valued $\mathcal{S}$-measurable functions on $\mathcal{S}$, let

$$[f, g] := \{h \in L_0(S, \mathcal{S}) : f \le h \le g\}.$$

Given $\varepsilon > 0$, $r > 0$, and a probability measure P on $(S, \mathcal{S})$, define

$$N_r^{[]}(\varepsilon, \mathcal{F}, P) = \min\{k : \text{ for some } f_1, \cdots, f_k \in L_r(P), \mathcal{F} \subset \cup_{i,j}\{[f_i, f_j] : \|f_i - f_j\| \le \varepsilon\}.$$

The function

$$H_r^{[]}(\varepsilon) := H_r^{[]}(\varepsilon, \mathcal{F}, P) := \log N_r^{[]}(\varepsilon, \mathcal{F}, P)$$

is called metric entropy with bracketing.

THEOREM 2.4.2 (Ossiander's CLT). *Let $\mathcal{F} \subset L_2(P)$, if*

$$\int_0^1 [H_2^{[]}(u, \mathcal{F}, P)]^{1/2}\, du < \infty,$$

then $\mathcal{F} \in \mathrm{CLT}(P)$.

The third (combinatorial) entropy (an extension of VC-dimension for classes of subsets to "VC-dimension" for classes of functions) is called the Kolčinskii-Pollard entropy. As before let $(S, \mathcal{S})$ be a measurable space and $\mathcal{F} \subset L_0(S, \mathcal{S})$, the space of all real-valued measurable functions on S. Write $F_{\mathcal{F}}(x) := \sup_{f \in \mathcal{F}} |f(x)|$. A measurable function F with $F \ge F_{\mathcal{F}}$ will be called an envelope function for $\mathcal{F}$. If a law P is given on S, then $F^* := F_{\mathcal{F}}^*$ for P will be called the envelope function of $\mathcal{F}$ for P, defined up to equality P-a.s. Let Γ be the set of all laws on $(S, \mathcal{S})$ of the form $n^{-1} \sum_{i \le n} \delta_{s(j)}$ for some $s(j) \in S : j = 1, \cdots, n$, and $n \in \mathbb{N}$ where the $s(j)$'s need not be distinct. For $\delta > 0$, $0 < p < \infty$, and $\gamma \in \Gamma$ let

$$D_F^{(p)}(\delta, \gamma, \mathcal{F}) := \sup\{m : \text{ for some } f_1, \cdots, f_m \in \mathcal{F} \text{ and all } i \ne j, \int |f_i - f_j|^p\, d\gamma > \delta^p \int F^p\, d\gamma\}.$$

Let

$$D_F^{(p)}(\delta, \mathcal{F}) := \sup_{\gamma \in \Gamma} D_F^{(p)}(\delta, \gamma, \mathcal{F}).$$

THEOREM 2.4.3 (Pollard's CLT). *Let $(S, \mathcal{S}, P)$ be a probability space and $\mathcal{F} \subset \mathcal{L}^2(S, \mathcal{S}, P)$. Let $F_{\mathcal{F}}^* \leq F \in \mathcal{L}^2(S, \mathcal{S}, P)$. Suppose that*

$$\int_0^1 (\log D_F^{(2)}(x, \mathcal{F})^{1/2}\, dx < \infty.$$

Then $\mathcal{F}$ is a P-Donsker class.

2.5 AGOZ (Two theorems for FCLT under local conditions)

In the literature, all the theorems about FCLT assume sufficient conditions. The following two theorems (one for iid-P samples; the other for non-iid samples) are sharp in some sense. If (cf. Ossiander's CLT)

$$\int_0^1 [H_2^{[]}(u, \mathcal{F}, P)]^{1/2}\, du < \infty,$$

then (1) $P^* F^2 < \infty$ where $F(\cdot) := \sup_{f \in \mathcal{F}} |f(\cdot)|$ and (2) if G_P is the limiting Gaussian process, then its associated distance $d_G := [E(G_P(f) - G_P(g))^2]^{1/2}$ satisfies the metric entropy condition, $\int_0 \log[N(\varepsilon, \mathcal{F}, d_G)]^{1/2} < \infty$, where $N(\varepsilon, \mathcal{F}, d_G)$ is the covering number of $\mathcal{F}$ by d_G (the minimal number of balls in $\mathcal{F}$ with d_G-radius less or equal to ε). Neither $P^* F^2 < \infty$ (although in case $\sup_{f \in \mathcal{F}} |Pf| < \infty$, the weak L_2 norm, $t^2 P^*(F > t) \to 0$, as $t \to$ is (cf. Giné and Zinn (1986), Proposition 2.7) nor is the entropy condition necessary for G_P to have a version with bounded d_G-uniformly continuous sample paths. But the first two conditions in Theorem 2.5.1 are necessary. About condition (iii) in this theorem, improved from Ossiander's CLT, the L_2-brackets are replaced by $\Lambda_{2,\infty}$-brackets (here $\Lambda_{2,\infty}(X) := [\sup_{u>0} u^2 P(X > u)]^{1/2}$

for a rv X) and the entropy condition (a sufficient condition for continuity of the limiting Gaussian process) is replaced by the weaker majorizing measure condition (which is a sufficient and necessary condition for continuity of a Gaussian process). Also, Theorem 2.5.1 gives Jain-Marcus central limit theorem for $C(S)$-valued random variables, the Chibisov-O'Reilly theorem without continuity assumption on the weight, etc.

We will use the following theorem to prove the main result in Chapter 3 for iid samples. Note that the Gaussian process in (*iii*) of the following theorem or in (*ii*) of the one after it is not necessarily the limiting Gaussian process of the empirical process; any continuous (tight) Gaussian process would do the job. In the proofs of our main theorems (Theorem 3.3.3, Theorem 4.2.1), we provide such ones under assumptions.

THEOREM 2.5.1 (Theorem 4.4, Andersen et al. (1988)). *Let* $\mathcal{F} \subset \mathcal{L}_2(S, \mathcal{S}, P)$ *and let* $F(x) := \sup\{|f(x)| : f \in \mathcal{F}\}$. *Assume* $\sup_{f\in\mathcal{F}} |Pf| < \infty$ *and*

$$(i) \qquad \lim_{t\to\infty} t^2 P^*(F > t) = 0.$$

Assume further that

$$(ii) \qquad \mathcal{F} \text{ is } P\text{-pre-Gaussian}$$

and that

(iii) *there exists a bounded and uniformly d_G-continuous centered Gaussian process G such that for all $\varepsilon > 0$ and all $f \in \mathcal{F}$*

$$\lim_{\alpha \to \infty} \alpha^2 P^*([\sup_{g \in B_{d_G}(f,\varepsilon)} |f - g|] > \alpha) \leq C\varepsilon^2.$$

Then $\mathcal{F} \in \mathrm{CLT}(P)$.

For non-iid samples in Chapter 4, we will use

THEOREM 2.5.2 (AGOZ, 1988). *Let $\mathcal{F} \subset \mathcal{L}_1(S, \mathcal{S}, P_{nj})$ and such that $\sup_{f \in \mathcal{F}} |f(s)| < \infty$, and let F be the envelope function of $\mathcal{F}$. Assume*

(i) For every $k \in \mathbb{N}$, and $f_1, \ldots, f_k \in \mathcal{F}$, the finite dimensional distributions

$$\Big\{\mathcal{L}\big[a_n^{-1} \sum_{j=1}^{n} (f_i(X_{nj}) - P_{nj} f_i)\big]_{i=1}^{k}\Big\}_{n=1}^{\infty}$$

converges weakly;

(ii) $\sum_{j=1}^n P_{nj}^*\{F > ta_n\} \underset{n\to\infty}{\longrightarrow} 0$ for all $t > 0$ and that there exits a pseudo-distance ρ on $\mathcal{F}$ dominated by the distance d_G of a centered Gaussian process G on $\mathcal{F}$ with bounded uniformly d_G-continuous paths, such that

(iii) $a_n^{-2} \sum_{j=1}^n P_{nj}(f-g)^2 \leq \rho^2(f,g)$ for all $f, g \in \mathcal{F}$ and

(iv) for all $f \in \mathcal{F}$ and $\varepsilon > 0$,

$$\sup_{t>0} t^2 \sum_{j=1}^{n} P_{nj}^* \{ \sup_{g \in B_\rho(f,\varepsilon)} |f - g| > ta_n\} \leq \varepsilon^2.$$

Then $\mathcal{F} \in \mathrm{CLT}\{P_{nj}; a_n\}$ *and the limiting measure is Gaussian.*

Note that by the comparison theorem, conditions (iii) in the above theorem ensures the pre-Gaussian condition for a CLT.

3 WEAK CONVERGENCE OF WEIGHTED EMPIRICAL PROCESSES BASED ON UNIFORM PROCESSES

3.1 Introduction

Given a sequence of iid uniform $(0,1)$ random variables $\{X_i : i \in \mathbb{N}\}$, we can form the uniform empirical process $U_n(x) = n^{-1/2} \sum_{i=1}^n (1_{X_i \le x} - x)$. Donsker's theorem says $U_n(x)$ converges weakly to the Brownian bridge process $B(x)$, on $[0,1]$. Weighted empirical processes consider suitable weight functions $w(x)$ such that $w(x)U_n(x)$ converges weakly to the weighted Brownian bridge process $w(x)B(x)$; in the literature, such a result is called the Chibisov-O'Reilly theorem. There is a long history of it.

Denote by $\mathcal{L}(X)$ the distribution of a real random variable X. That $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X_n)$ is understood as weak convergence of distributions on the real line; that is, if F_{X_n}, F_X denote the df's of X_n, X respectively, then $F_{X_n}(u)$ converges to $F_X(u)$ for each $u \in \mathbb{R}$ at which $F_X(u)$ is continuous.

For following Let the function $w(t) \ge 0$ is continuous on the interval $(0,1)$ and is regularly varying in the neighborhoods of 0 and 1; define a metric for functions $x(t)$, $y(t)$ on $[0,1]$ by

$$\rho_w(x(t), y(t)) = \sup_{0<t<1} w(t)|x(t) - y(t)|.$$

Let R_w be the metric space of functions on $[0,1]$ without 2nd-type discontinuity (i.e. only jump or removable discontinuities) with the metric ρ_w. Let $h_1(t) = 1/(t^{1/2}w(t))$ and $h_2(t) = 1/(t^{1/2}w(1-t))$.

THEOREM 3.1.1 (Čibisov (1964a), Theorem 3). *A necessary and sufficient con-*

dition that for any continuous functional on R_ϕ

$$\mathcal{L}[f(\beta_n(u))] \to \mathcal{L}[f(\beta(u))]$$

is the convergence of the integral

$$\int_{0+} \exp(-l^2 h_i^2(u)/2)\,\frac{du}{u}, \qquad i = 1, 2$$

for each $l > 0$*. Here and following* $\int_{0+}$ *is understood as integration from* 0 *to any small constant.*

THEOREM 3.1.2 (O'Reilly (1974)). *Let* $q(t)$ *be a continuous, nonnegative function on* $[0,1]$*, bounded away from zero on* $[\gamma, 1-\gamma]$ *for some* $\gamma > 0$*, non-decreasing (non-increasing) on* $[0,\gamma]$ $([1-\gamma, 1])$*. Then*

$$\int_0^1 t^{-1} \exp(-\varepsilon h(t)^2)\,dt < \infty \text{ for all } \varepsilon > 0, i = 1, 2 \tag{3.1}$$

is both necessary and sufficient for the weak convergence of $U_n(t)$ *to* $B(t)$ *in* $(D[0,1], \rho_q)$ *where* $\rho_q(x,y) = \sup\{|x(t) - y(t)|/q(t) : 0 \le t \le 1\}$ *and* $h_1(t) = t^{-1/2}q(t)$ *and* $h_2(t) = t^{-1/2}q(1-t)$*.*

A function $q(s)$ on $(0,1)$ is going (only in this section) to be called positive if $\inf_{\delta \le s \le 1-\delta} q(s) > 0$ for all $0 < \delta < 1/2$.

A function $q(s)$ on $(0,1/2]$ is going (only in this section) to be called positive if $\inf_{\delta \le s \le 1/2} q(s) > 0$ for all $0 < \delta < 1/2$.

DEFINITION 3.1.3. Let $q(t)$ be a positive function on $(0,1/2]$, non-decreasing in a neighborhood of 0. Such a q will be called an Erdös-Feller-Kolmogorov-Petrovski

(EFKP) upper-class function of a Brownian bridge $\{B(s); 0 \leq s \leq 1\}$ if

$$\limsup_{t \downarrow 0} |B(t)|/q(t) < \infty, \quad a.s. \tag{3.2}$$

REMARKS 3.1.4. (1) By the usual representation of a Brownian bridge in terms of a standard Wiener process, q is an EFKP upper-class function of a Brownian bridge if and only if it is an EFKP upper-class function of a standard Brownian motion (Wiener process).

(2) By Blumenthal's 0-1 law (cf. Itô and McKean (1965), p. 25), (3.2) holds if and only if there is a constant $0 \leq \beta < \infty$ such that

$$\limsup_{t \downarrow 0} |B(t)|/q(t) = \beta, \quad a.s. \tag{3.3}$$

DEFINITION 3.1.5. A function $q(t)$ $(t > 0)$ belongs to the upper class for a Brownian motion $\{B(t) : t \in [0, 1]\}$ if

$$\lim_{\tau \to 0} \mathrm{P}\{\sup_{0<t\leq\tau} \frac{B(t)}{q(t)} > 1\} = 0.$$

DEFINITION 3.1.6. A function $q(t)$ $(t > 0)$ belongs to the lower class for a Brownian motion $\{B(t) : t \in [0, 1]\}$ if for all τ

$$\mathrm{P}\{\sup_{0<u\leq\tau} \frac{B(t)}{q(t)} > 1\} = 1.$$

THEOREM 3.1.7 (Kolmogorov's test). *Let $h \in C(0, 1]$ and $h > 0$ on $(0, 1]$. Assume h is nondecreasing and $h(t)/t^{1/2}$ is nonincreasing for small $t > 0$. Then, h is in the upper class or lower class according as $\int_{0+} t^{-3/2} h e^{-h^2/(2t)}\, dt$ converges or diverges.*

The following is a connection between Kolmogorov's test and Chibisov-O'Reilly Theorem.

THEOREM 3.1.8 (Csörgő et al. (1986), Theorem 3.4). *Let $q(t)$ be a positive function on* $(0, 1/2]$, *nondecreasing in a neighborhood of zero. The following are equivalent.*

(*i*) $\limsup_{t\downarrow 0} |B(t)|/q(t) = 0$ a.s.

(*ii*) $\int_0^{1/2} s^{-3/2} q(s) \exp(-cs^{-1}q^2(s))\, ds < \infty$ *for all* $c > 0$
and $q(s)/s^{1/2} \to \infty$ *as* $s \downarrow 0$.

(*iii*) $\int_0^{1/2} s^{-1} \exp(-cs^{-1}q^2(s))\, ds < \infty$ *for all* $c > 0$.

THEOREM 3.1.9 (Csörgő et al. (1986), Theorem 4.2.1). *Let $q(t)$ be a positive function on* $(0, 1)$ *such that it is nondecreasing in a neighborhood of zero and nonincreasing in a neighborhood of one. On the* (common) *probability space of* Theorem 1.1 in Csörgő et al. (1986), *we have, as* $n \to \infty$,

$$\sup_{0<t<1} |U_n(t) - B_n(t)|/q(t) = o_P(1)$$

if and only if

$$\beta = 0 \quad \textit{in } (3.3),$$

that is, if and only if

$$\int_0^1 (t(1-t))^{-1} \exp(-c(t(1-t))^{-1}q(t)^2)\, dt < \infty \textit{ for all } c > 0, \tag{3.4}$$

where $\{B_i(t) : i = 1, 2, \cdots\}$ *is a sequence of Brownian bridges.*

It's easy to check the functions

(*i*) $q(t) = t^\alpha, \quad \alpha < 1/2,$

(*ii*) $q(t) = t^\alpha L(t)$, $\alpha < 1/2$ and $L(t)$ is slowly varying

(*iii*) and $q(t) = t^{1/2}(\log(1/t))^\alpha)$ for any $\alpha \geq 1/2$

satisfy (*ii*) and (*iii*) in Theorem 3.1.8; but $q(t) = t^{1/2}(\log\log(1/t))^{1/2}$ doesn't.

3.2 KKZ

Given a sequence of independent real rv's $\{Y, Y_1, Y_2, \cdots\}$, we can form the classical empirical process

$$\nu_n(y) := n^{-1/2} \sum_{i=1}^{n} [1_{Y_i \leq y} - \mathrm{P}(Y_i \leq y)],\ y \in \mathbb{R}.$$

Kuelbs et al. (2013) considered instead iid processes $\{Y(t), Y_1(t), Y_2(t), \cdots\}$ for $t \in E$ where E is a general parameter set. Under the L-condition (see Definition 3.2.5), they proved a CLT for the empirical process

$$\nu_n(t, y) := n^{-1/2} \sum_{i=1}^{n} [1_{Y_i(t) \leq y} - \mathrm{P}(Y_i(t) \leq y)], \quad t \in E,\ y \in \mathbb{R}.$$

The research in this thesis starts from this result.

Theorem 3.2.1 (Kuelbs et al. (2013)). *Let ρ be given by $\rho(s,t)^2 = \mathrm{E}(H(s) - H(t))^2$, for some centered Gaussian process H that is sample bounded and uniformly continuous on (E, ρ) with probability one. Further, assume that for some $L < \infty$, and all $\varepsilon > 0$, the L-condition holds, and $D(E)$ is a collection of real valued functions on E such that* $\Pr(X(\cdot) \in D(E)) = 1$. *If*

$$\mathcal{C} = \{C_{s,x} : s \in E, x \in \mathbb{R}\},$$

where

$$C_{s,x} = \{z \in D(E) : z(s) \leq x\}$$

for $s \in E$, $x \in \mathbb{R}$, *then* $\mathcal{C} \in \mathrm{CLT}(P)$.

In this case, we also say the empirical process based on $\{1_{Y(t)\leq y} - \mathrm{P}(Y(t) \leq y) : t \in E,\ y \in \mathbb{R}\}$ satisfies the CLT or write $\{1_{Y(t)\leq y} - \mathrm{P}(Y(t) \leq y) :\ t \in E,\ y \in \mathbb{R}\} \in$ CLT.

3.2.1 L-condition

Definition 3.2.2. Let $F(x)$ be a distribution function on $\mathbb{R}$. The (randomized) distributional transform of $F(x)$ is defined as

$$\tilde{F}(x) := \tilde{F}(x, V) := F(x-) + (F(x) - F(x-))V,$$

where V is a uniform random variable on $[0, 1]$.

Next we record some simple properties of the distributional transform.

Lemma 3.2.3. *(i)* $\tilde{F}(x) \leq F(x)$ *for all* $x \in \mathbb{R}$.

(ii) If $x < y$, *then* $F(x) \leq \tilde{F}(y)$.

(iii) If $x \leq y$, *then* $\tilde{F}(x) \leq \tilde{F}(y)$.

(iv) If $x < y$ *and* $F(\cdot)$ *is strictly increasing, then* $F(x) < \tilde{F}(y)$.

Proof. By definition, (i) is obvious. For (ii), take $x < z < y$, hence $F(x) \leq F(z)$. Since $F(z) \leq F(y-)$ and $F(y-) \leq \tilde{F}(y)$, hence $F(x) \leq \tilde{F}(y)$. For (iii), if $x = y$, there is nothing to prove; assume $x < y$. By (i) and (ii), we get (iii). For $(iiii)$, take $x < z < y$. Since $F(\cdot)$ is strictly increasing, $F(x) < F(z)$. But by (ii), $F(z) \leq \tilde{F}(y)$. Hence $F(x) < \tilde{F}(y)$. □

For a continuous df F of a random variable X, the random variable $F(X)$ is uniform on $[0,1]$; but for a general df F, this might not be the case. However using the (randomized) distributional transform overcomes this.

LEMMA 3.2.4. *If $F(x)$ is the distribution function of a random variable X, then $\tilde{F}(X) := \tilde{F}(X,V)$ is uniform on $[0,1]$. Here V is a uniform random variable on $[0,1]$ independent of X.*

PROOF. For a proof, see Rüschendorf (2009). □

DEFINITION 3.2.5 (L-condition for a stochastic process). Let $X = \{X_t : t \in E\}$ be a stochastic process. We say the process X satisfies the L-condition if there exists a continuous Gaussian distance ρ on E such that for every $\varepsilon > 0$

$$\sup_{t\in E} \mathrm{P}^*\Big(\sup_{s:\rho(s,t)\le\varepsilon} |\tilde{F}_t(X(t)) - \tilde{F}_t(X(s))| > \varepsilon^2\Big) \le L\varepsilon^2, \tag{3.5}$$

where $\tilde{F}_t(\cdot)$ is the distributional transform of the distribution function $F_t(\cdot)$ of X_t.

3.3 Weak convergence of the time-dependent weighted empirical process

In view of KKZ result (a time-dependent empirical process taking processes as samples) and the classical weighted empirical process (taking independent uniform $(0,1)$ samples), a natural direction is to put these two together in some way. Hence if we are given a uniform process (see Definition 3.3.1) and "weight" functions, we consider the following process, which we call the time-dependent weighted uniform empirical process,

$$\alpha_n(t,y) := n^{-1/2} \sum_{i\le n} w(y)(1_{X_i(t)\le y} - y), t \in E,\ y \in \mathbb{R}$$

where $\{X(t), X_1(t), X_2(t), \cdots\}$ are iid uniform processes. We are interested in conditions to ensure a CLT for the empirical process α_n.

DEFINITION 3.3.1. We call a process $X = \{X(t) : t \in E\}$ a uniform process if for each $t \in E$, $X(t)$ is uniformly distributed on $(0, 1)$.

We call the main condition in our theorem the WL-condition.

DEFINITION 3.3.2. [WL-condition for $(X; w)$] Given a uniform process $X := \{X_t : t \in E\}$ where E is a general parameter set and a function $w := w(x) > 0$ on $(0, 1)$, we say $(X; w)$ satisfies the WL-condition if for some constant L (depending on $w(x)$, but not on x), some continuous Gaussian distance ρ on E and all $\varepsilon > 0$, $0 < x < 1$, we have

$$\text{(WL-condition)} \qquad \begin{aligned} \sup_t \mathrm{P}^*\big(\sup_{s:\rho(s,t)\le\varepsilon} 1_{X_s \le x < X_t} > 0\big) &\le \frac{L\varepsilon^2}{w(x)^2}, \\ \sup_t \mathrm{P}^*\big(\sup_{s:\rho(s,t)\le\varepsilon} 1_{X_t \le x < X_s} > 0\big) &\le \frac{L\varepsilon^2}{w(x)^2}. \end{aligned}$$

The following is the main result of this chapter.

THEOREM 3.3.3. *Let $X := \{X_t : t \in E\}$ be a uniform process on a parameter set E. Let $w := w(x) > 0$ be a measurable function symmetric about $1/2$ for which there exists $\gamma \in (0, 1/2]$ such that w is non-increasing and $xw(x)^2$ is non-decreasing on $(0, \gamma)$ and such that w is uniformly bounded on $[\gamma, 1/2]$. Further, assume that $w(x)$ is regularly varying in a neighborhood of zero and satisfies the integral condition*

$$\int_0^\gamma t^{-1} \exp[-c/(tw(t)^2)]\, dt < \infty \text{ for all } c > 0.$$

If

$$\lim_{\alpha\to\infty} \alpha^2 \mathrm{P}^*(\sup_{t\in E} w(X_t) > \alpha) = 0$$

and the WL-condition for $(X;w)$ (3.3.2) *are satisfied, then the empirical process based on* $\{w(x)(1_{X_t \le x} - x) : t \in E,\, x \in [0,1]\}$ *converges weakly in* $\ell^\infty(E \times [0,1])$.

Remarks 3.3.4. (1) We require the function $w(x)$ is symmetric about $1/2$ is no loss of generality. As the Brownian bridge has the same behavior at 0 and 1. Moreover we only give the proof of the theorem for $0 < x < 1/2$. Indeed, if let $\tilde{X}_t := 1 - X_t$, then $(\tilde{X};w)$ satisfies the WL-condition. The result for $\tilde{X}$ for $0 < x \le 1/2$ gives a result of X for $1/2 < x \le 1$. The fact that if $\mathcal{F}_1$ and $\mathcal{F}_2$ are Donsker classes, then $\mathcal{F} := \mathcal{F}_1 \cup \mathcal{F}_2$ are Donsker class gives the result for $\mathcal{F} = E \times [0,1]$.

(2) For a general process $Y := \{Y_t : t \in E\}$, if define $X := X_t := \tilde{F}_t(Y_t)$, where $\tilde{F}_t(\cdot)$ is the (randomized) distributional transform of the df F_t of X_t, then X is a uniform process (see Lemma 3.2.4). Such a process X is called a copula process. If we have a CLT for the X process, then we have a CLT for the Y process; see Proposition 3.3.5 for precise statement. In case of $w \equiv 1$, this theorem gives a proof of Kuelbs et al. (2013), Theorem 3 provided that $F_t(\cdot)$ for each $t \in E$ is strictly increasing; see Corollary 3.3.7.

(3) If the uniform process X comes in such a way in (2), then the weight function $w(x)$ for which the time dependent weighted empirical process satisfies a CLT, is a measure of the local modulus of the process Y; we give an example in Section 3.4 when Y is a Brownian motion.

The proof of the theorem is given at the end of this section.

The following is a possible way that a CLT for the time dependent empirical process for Y can be obtained from proving a CLT for the process X.

Proposition 3.3.5. *Let* $w(x)$ *be any function on* $(0,1)$. *Let* $\{Y_t : t \in E\}$ *be a process and* $F_t(\cdot)$ *is the df of* Y_t. *Let* $X_t := \tilde{F}_t(Y_t)$. *Then the following hold*:

(i) *If* $F_t(\cdot)$ *is strictly increasing for each* $t \in E$*, then*

$$\{w(x)(1_{X_t \le x} - x) : t \in E,\, x \in [0,1]\} \in \mathrm{CLT}$$

implies

$$\{w(F_t(y))(1_{Y_t \le y} - F_t(y)) : t \in E,\, y \in \mathbb{R}\} \in \mathrm{CLT}.$$

(ii)

$$\{w(x)(1_{X_t \le x} - x) : t \in E,\, x \in [0,1]\} \in \mathrm{CLT}$$

implies

$$\{w(F_t(y))(1_{Y_t \le y} - F_t(y)) : (t,y) \in T_0\} \in \mathrm{CLT}$$

where T_0 *is any countable subset of* $E \times \mathbb{R}$*.*

PROOF. *Proof of* (i). Recall that $\tilde{F}(x) \le \tilde{F}(y)$ for $x \le y$ and $\tilde{F}(x) \le F(x)$ for all $x \in \mathbb{R}$ and for any df F (see Lemma 3.2.3). Hence $Y_t \le y$ implies that $\tilde{F}_t(Y_t) \le F_t(y)$; i.e.

$$1_{Y_t \le y} \le 1_{\tilde{F}_t(Y_t) \le F_t(y)}, \text{ uniformly in } t \in E,\, y \in \mathbb{R}. \tag{3.6}$$

Since $F_t(\cdot)$ is strictly increasing, by the same lemma if $x < y$, then $F(x) < \tilde{F}(y)$. Now if $\tilde{F}_t(Y_t) \le F_t(y)$ and $Y_t > y$ for some $t \in E$ and $y \in \mathbb{R}$, then $F_t(y) < \tilde{F}_t(Y_t)$. We have a contradiction: $F_t(y) < F_t(y)$. Thus $\tilde{F}_t(Y_t) \le F_t(y)$ implies $Y_t \le y$; i.e.

$$1_{Y_t \le y} \ge 1_{\tilde{F}_t(Y_t) \le F_t(y)}, \text{ uniformly in } t \in E,\, y \in \mathbb{R}.$$

Combining the two displays, we have

$$1_{Y_t \le y} = 1_{\tilde{F}_t(Y_t) \le F_t(y)}, \text{ uniformly in } t \in E,\, y \in \mathbb{R}. \tag{3.7}$$

Since $\{F_t(y) : t \in E,\, y \in \mathbb{R}\}$ is a subset of $[0,1]$, thus if the empirical process based on $\{w(x)(1_{\tilde{F}_t(Y_t) \le x} - x) : t \in E,\, x \in [0,1]\}$ satisfies CLT in $\ell^\infty(E \times [0,1])$, then, by substituting x with $F_t(y)$ and using (3.7), the empirical process based on $\{w(F_t(y))(1_{Y_t \le F_t(y)} - F_t(y)) : t \in E,\, y \in \mathbb{R}\}$ satisfies the CLT in $\ell^\infty(E \times \mathbb{R})$.

Proof of (ii). Fix $t \in E$ and $y \in \mathbb{R}$. If $\tilde{F}_t(Y_t) \le F_t(y)$, since $\tilde{F}_t(Y_t) = F_t(y)$ has probability zero, then, after throwing out this null set, $\tilde{F}_t(Y_t) < F_t(y)$, which will imply $Y_t \le y$. If not, then $Y_t > y$, by Lemma 3.2.3, hence $F_t(y) \le \tilde{F}_t(Y_t)$. Again we have a contradiction $F_t(y) < F_t(y)$. Thus almost surely $1_{\tilde{F}_t(Y_t) \le F_t(y)} \le 1_{Y_t \le y}$. Combining this with 3.6 gives, almost surely,

$$1_{Y_t \le y} = 1_{\tilde{F}_t(Y_t) \le F_t(y)}, \text{ uniformly in } (t, y) \in T_0, \tag{3.8}$$

where T_0 is any countable set in $E \times \mathbb{R}$. Restricting to the countable set, we have the stated implication as in (i). □

Remark 3.3.6. We need the assumption that $F_t(\cdot)$ is strictly increasing for each t in order to have equation (3.7). Without this assumption, we only have 3.8. If we can show T_0 is dense in $E \times \mathbb{R}$ with respect to the L_2 distance of the limiting Gaussian, then a version of the empirical process based on $\{w(F_t(y))(1_{Y_t \le F_t(y)} - F_t(y)) : (t, y) \in E \times \mathbb{R}\}$ converges weakly to the limiting Gaussian process.

Corollary 3.3.7 (cf. Kuelbs et al. (2013), Theorem 3). *Let $Y := \{Y_t : t \in E\}$ be a process. Let F_t be the df of Y_t. In addition, assume that $F_t(\cdot)$ is strictly increasing* *for each* $t \in E$ *and that* Y *satisfies the L-condition*:

$$\sup_{t\in E} \mathrm{P}^*(\sup_{s:\rho(s,t)\le\varepsilon} |\tilde{F}_t(Y_t) - \tilde{F}_t(Y_s)| > \varepsilon^2) \le L\varepsilon^2, \tag{3.9}$$

for a constant L *and a continuous Gaussian metric* $\rho(s,t)$ *on* E*. Then*

$$\{1_{Y_t\le y} - \mathrm{P}(Y_t \le y) : t \in E,\, y \in \mathbb{R}\} \in \mathrm{CLT}.$$

REMARK 3.3.8. Under the L-condition, we will see from the proof of Theorem 3.3.9 that there is a countable dense set in $E \times \mathbb{R}$ with respect to the L_2 distance, say ρ_P, of the limiting Gaussian process. Hence without the restriction that $F_t(\cdot)$ is strictly increasing, we still have a CLT but on a countable dense set; since this set is ρ_P-dense in $E \times \mathbb{R}$, we have a version of the empirical process, which converges weakly in $\ell^\infty(E \times \mathbb{R})$.

The following is a simpler proof than in Kuelbs et al. (2013) of the pre-Gaussian property of the empirical process considered there under the L-condition for a process.

THEOREM 3.3.9. *Let* $\{Y(t), Y_1(t), Y_2(t), \cdots\}$ *on* E *are iid and* $\{Y(t) : t \in E\}$ *satisfies the L-condition, then the centered Gaussian process on* $E\times\mathbb{R}$ *with covariance either*

$$\mathrm{P}(Y_s \le x, Y_t \le y) - \mathrm{P}(Y_s \le x)\mathrm{P}(Y_t \le y)$$

or

$$\mathrm{P}(Y_s \le x, Y_t \le y)$$

has a version, which is sample bounded and uniformly continuous with respect to its L_2 *distance.*

PROOF. Let $\{G_1(t,y) : t \in E,\, y \in \mathbb{R}\}$ and $\{G_2(t,y) : t \in E,\, y \in \mathbb{R}\}$ be the

Gaussian processes on $E \times \mathbb{R}$ with covariance $\mathrm{P}(Y_s \le x, Y_t \le y) - \mathrm{P}(Y_s \le x)\mathrm{P}(Y_t \le y)$ and $\mathrm{P}(Y_s \le x, Y_t \le y)$, respectively. Let d_{G_1} and d_{G_2} be their L_2 distances, respectively; i.e,

$$
\begin{aligned}
(3.10) \qquad & d_{G_1}((s,x),(t,y))^2 = \mathrm{E}(1_{Y_s \le x} - 1_{Y_t \le y})^2 - (\mathrm{E}(1_{Y_s \le x} - 1_{Y_t \le y}))^2, \\
& d_{G_2}((s,x),(t,y))^2 = \mathrm{E}(1_{Y_s \le x} - 1_{Y_t \le y})^2.
\end{aligned}
$$

And,

$$
\begin{aligned}
(3.11) \quad d_{G_2}((s,x),(t,y))^2 &= \mathrm{E}(1_{Y_s \le x} - 1_{Y_t \le y})^2 \\
&= \mathrm{E}(1_{Y_s \le x} - 1_{Y_t \le x} + 1_{Y_t \le x} - 1_{Y_t \le y})^2 \\
&\le 2\mathrm{E}(1_{Y_s \le x} - 1_{Y_t \le x})^2 + \mathrm{E}(1_{Y_t \le x} - 1_{Y_t \le y})^2 \\
&\le 2(\mathrm{P}(Y_s \le x < Y_t) + \mathrm{P}(Y_t \le x < Y_s)) + |F_t(y) - F_t(x)| \\
&\le 6(L+1)\rho(s,t)^2 + |F_t(y) - F_s(x)|,
\end{aligned}
$$

where in the last line of the above display, we used Lemma 4.2.5.

Let $W(\cdot)$ be a Brownian motion on $[0, \infty)$. Define the centered Gaussian process

$$
H_2(t,y) := W(F_t(y)) : t \in E,\ y \in \mathbb{R},
$$

where $F_t(\cdot)$ be the df of Y_t. Then its L_2 distance $d_{H_2}((s,x),(t,y)) = |F_t(y) - F_s(x)|^{1/2}$. It follows that H_2 is sample bounded and uniformly continuous with respect to d_{H_2}. By the L-condition, let $\{H_1(t) : t \in E\}$, independent from H_2, be a Gaussian process with bounded and uniformly continuous sample paths with it's L_2 distance ρ. Define $H(t,y) = H_2(t,y) + (6L+6)^{1/2}H_1(t)$. Then $\{H(t,y) : t \in E,\ y \in \mathbb{R}\}$ is sample bounded and uniformly continuous with respect to it's L_2 distance d_H. Total boundedness of d_{H_1} and d_{H_2} implies that of d_H as can be seen from the equation

$$d_H((t_1, y_1), (t_2, y_2))^2 = d_{H_2}((t_1, y_1), (t_2, y_2))^2 + (6L + 6)d_{H_1}(t_1, t_2)^2.$$

Thus let T_0 be a countable dense subset in $(E \times \mathbb{R}, d_H)$. Since $d_{G_1} \le d_H$ in view of (3.10) and (3.11), by the comparison theorem 2.3.7, $\{G_1(s, x) : (s, x) \in T_0\}$ is sample bounded and uniformly continuous with respect to d_{G_1}. Since T_0 is also dense in $(E \times \mathbb{R}, d_{G_1})$, by the extension lemma, Lemma 2.3.8, $\{G_1(s, x) : (s, x) \in E \times \mathbb{R}\}$ is samples bounded and uniformly d_{G_1}-continuous. □

PROOF OF COROLLARY 3.3.7. By Proposition 3.3.5, we only need to check the conditions in Theorem 3.3.3 with $w(x) \equiv 1$.

Under the L-condition, we have (cf. Kuelbs et al. (2013), Lemma 1)

$$\sup_x |F_t(x) - F_s(x)| \le 2(L + 1)\rho(s, t)^2.$$

Consequently by passing to the limit,

$$\sup_x |F_t(x-) - F_s(x-)| \le 2(L + 1)\rho(s, t)^2.$$

Recalling that $\tilde{F}_s(x) = F_s(x-) + V(F_s(x) - F_s(x-))$, we obtain

$$\begin{aligned}\sup_x |\tilde{F}_t(x) - \tilde{F}_s(x)| &\le \sup_x |F_t(x-) - F_s(x-)| + \sup_x |V(F_t(x) - F_s(x))| \\ &\quad + \sup_x |V(F_t(x-) - F_s(x-))| \\ &\le 6(L + 1)\rho(s, t)^2.\end{aligned}$$

For $t \in E$ fixed, let $A := \{\sup_{s:\rho(s,t)\le\varepsilon} |\tilde{F}_t(Y_t) - \tilde{F}_t(Y_s)| > \varepsilon^2\}$.

On A^c, we have for all s with $\rho(s,t) \le \varepsilon$,

$$\begin{aligned}|\tilde{F}_s(Y_s) - \tilde{F}_t(Y_t)| &\le |\tilde{F}_s(Y_s) - \tilde{F}_t(Y_s)| + |\tilde{F}_t(Y_s) - \tilde{F}_t(Y_t)| \\ &\le 6(L+1)\rho(s,t)^2 + \varepsilon^2 \\ &\le (6L+7)\varepsilon^2.\end{aligned}$$

Hence

$$\begin{aligned}\mathrm{P}(\sup_{s:\rho(s,t)\le\varepsilon} 1_{\tilde{F}_s(Y_s)\le x<\tilde{F}_t(Y_t)} > 0) &= \mathrm{P}(A^c, \sup_{s:\rho(s,t)\le\varepsilon} 1_{\tilde{F}_s(Y_s)\le x<\tilde{F}_t(Y_t)} > 0) \\ &\quad + \mathrm{P}(A, \sup_{s:\rho(s,t)\le\varepsilon} 1_{\tilde{F}_s(Y_s)\le x<\tilde{F}_t(Y_t)} > 0) \\ &\le \mathrm{P}(E^c, 1_{\tilde{F}_t(Y_t)-(6(L+1)\varepsilon^2+\varepsilon^2)\le x<\tilde{F}_t(Y_t)} > 0) + L\varepsilon^2 \\ &\le (7L+7)\varepsilon^2.\end{aligned}$$

Similarly,

$$\mathrm{P}(\sup_{s:\rho(s,t)\le\varepsilon} 1_{\tilde{F}_t(Y_t)\le x<\tilde{F}_s(Y_s)} > 0) \le (7L+7)\varepsilon^2.$$

In addition, obviously for $w(x) \equiv 1$

$$\lim_{\alpha\to\infty} \alpha^2 \mathrm{P}(\sup_{t\in E} w(\tilde{F}_t(Y_t)) > \alpha) = 0.$$

Thus we have verified the conditions in Theorem 3.3.3. □

We will prove Theorem 3.3.3 only for $0 < x < 1/2$ as explained in Remark 3.3.4. We will check the pre-Gaussian condition (ii) and the local modulus condition (iii) in Theorem 2.5.1

3.3.1 Pre-Gaussian

Let $\{G_0((s,x)) : s \in E,\ x \in [0,1]\}$ be the mean zero Gaussian process with covariance

$$\mathrm{E}G_0(s,x)G_0(t,y)) = w(x)w(y)\mathrm{P}(X_s \le x, X_t \le y). \tag{3.12}$$

Under the assumptions of Theorem 3.3.3, we will prove $G_0(s,x)$ has a version with bounded and uniformly continuous sample paths with its L_2 distance d_{G_0} a.s. by comparing it with some other continuous Gaussian distance.

LEMMA 3.3.10 (see Andersen et al. (1988), Example 4.8). *Let $W(y)$ be a Brownian motion and $w(y)$ as in Theorem 3.3.3. Then the Gaussian process $\{w(y)W(y) : y \in [0,1]\}$ is sample bounded and uniformly continuous w.r.t. its L_2 distance, which is given by*

(3.13)

$$d(x,y)^2 := \mathrm{E}(w(y)W(y) - w(x)W(x))^2 = w(x \vee y)^2|y-x| + (x \wedge y)(w(x) - w(y))^2.$$

LEMMA 3.3.11. *If $xw(x)^2$ is non-decreasing and $w(x)$ is non-increasing for $0 < x < \delta$, then*

$$d(x,y) \le d(x,z)$$

for $0 < x \le y \le z \le \delta$.

PROOF. Let $0 < x \le y \le z \le \delta$. Using definition (3.13) and the monotonicity of $xw(x)^2$ and $w(x)$, we obtain

$$\begin{aligned} d(x,y)^2 &= w(y)^2(y-x) + x(w(y)-w(x))^2 \\ &= xw(x)^2 + yw(y)^2 - 2xw(x)w(y) \\ &\le xw(x)^2 + zw(z)^2 - 2xw(x)w(z) \\ &= d(x,z)^2. \end{aligned}$$

□

Next we give an upper bound for d_{G_0} under WL-condition in Theorem 3.3.3.

LEMMA 3.3.12. *Let $d(x,y)$ be as in* (3.13) *and $d_{G_0}((s,x),(t,y))$ the L_2 distance of the Gaussian process G_0 in 3.12. Then under the WL-condition, we have*

$$d_{G_0}^2 \le 2d^2 + 4L\rho(s,t)^2.$$

PROOF. First observe that for $t \in E$

$$d(x,y)^2 = \mathrm{E}(w(y)W(y) - w(x)W(x))^2 = \mathrm{E}|w(x)1_{X_t \le x} - w(y)1_{X_t \le y}|^2. \tag{3.14}$$

Using by WL-assumption for fixed s and t

$$\mathrm{P}(X_s \le x < X_t) \le \frac{L\rho(s,t)^2}{w(x)^2} \text{ and } \mathrm{P}(X_t \le x < X_s) \le \frac{L\rho(s,t)^2}{w(x)^2}, \tag{3.15}$$

we obtain

$$
\begin{aligned}
(3.16) \qquad & d_{G_0}((s,x),(t,y))^2 \\
&= \mathrm{E}|w(x)1_{X_s \le x} - w(y)1_{X_t \le y}|^2 \\
&= \mathrm{E}|w(x)1_{X_s \le x} - w(x)1_{X_t \le x} + w(x)1_{X_t \le x} - w(y)1_{X_t \le y}|^2 \\
&\le 2\mathrm{E}|w(x)1_{X_s \le x} - w(x)1_{X_t \le x}|^2 + 2\mathrm{E}|w(x)1_{X_t \le x} - w(y)1_{X_t \le y}|^2 \\
&= 2w(x)^2\mathrm{E}|1_{X_s \le x} - 1_{X_t \le x}|^2 + 2d(x,y)^2 \qquad \text{by (3.14)} \\
&\le 2w(x)^2(\mathrm{P}(X_s \le x < X_t) + \mathrm{P}(X_t \le x < X_s)) + 2d(x,y)^2 \\
&\le 4L\rho(s,t)^2 + 2d(x,y)^2 \qquad \text{by (3.15).}
\end{aligned}
$$

□

COROLLARY 3.3.13. *Under the WL-assumption, the process $G_0(t,y)$ is sample bounded and uniformly continuous with its L_2 distance.*

PROOF. By assumption, ρ is the L_2 distance of a mean zero Gaussian process on E, say $\{H_0(t) : t \in E\}$, with bounded and uniformly ρ-continuous sample paths. Let the metric d on $[0,1]$ as given in 3.13 with the corresponding Gaussian process $w(x)W(x)$, which is sample bounded and uniformly d-continuous. Let $H_2((t,y)) := 2^{1/2}w(y)W(y) + 2L^{1/2}H_0(t) : t \in E, y \in [0,1]$. Then the L_2 distance, $d_{H_2}((s,x),(t,y))$, of H_2 is $2^{1/2}d(x,y) + 2L^{1/2}\rho(s,t)$. Total boundedness of d and ρ implies that of d_{H_2}. Thus let T_0 be a dense subset in $(E \times [0,1], d_{H_2})$; since $d_{G_0} \le d_{H_2}$ by (3.16), T_0 is also a dense subset in $(E \times [0,1], d_{G_0})$; Using the comparison Theorem 2.3.7 with $H_1 := G_0$ and that $d_{G_0} \le d_{H_2}$, the Gaussian process $\{G_0 : (s,x) \in T_0\}$ is sample bounded and uniformly d_{G_0} continuous. By the extension lemma, Lemma 2.3.8, $\{G_0 : (s,x) \in E \times R\}$ is sample bounded and uniformly d_{G_0}-continuous. □

3.3.2 Local modulus

Recall that a positive function $L(x)$ defined on $(0, \infty)$ is slowly varying at infinity (in a neighborhood of zero) if $L(\lambda x)/L(x) \to 1, x \to \infty$ $(x \to 0)$ for every $\lambda > 0$ (see (Feller, 1971, p. 276)). One says a function $U(x)$ is regularly varying at infinity (in a neighborhood of zero) if $U(x) = x^\rho L(x)$ for some $-\infty < \rho < \infty$, and some slowly varying at infinity (in a neighborhood of zero) function $L(x)$; ρ is called the exponent (see (Feller, 1971, p. 275)).

LEMMA 3.3.14. *Let $w(x) > 0$ for $0 < x \le 1/2$ and is regularly varying in a neighborhood of 0 with nonzero exponent. Let $\theta_0 > 0$ be small enough such that $w(x)$ is non-increasing for $0 < x < \theta_0$. Then for $0 < \theta < \theta_0$*

$$\sum_{k=0}^{\infty} \frac{1}{w(2^{-k}\theta)^2} \le \frac{C}{w(\theta)^2},$$

where C depends only on the weight function $w(x)$, but not on the argument x.

PROOF. Since $w(x)$ is non-increasing for $0 < x < \theta_0$,

$$(\ln 2) \sum_{k=1}^{\infty} \frac{1}{w(2^{-k}\theta)^2} \le \int_0^\theta \frac{1}{w(y)^2} \frac{dy}{y} \le (\ln 2) \sum_{k=0}^{\infty} \frac{1}{w(2^{-k}\theta)^2}.$$

By Theorem 1 in (Feller, 1971, p. 281), we have

$$\frac{\frac{1}{w(\theta)^2}}{\int_0^\theta \frac{1}{w(y)^2} \frac{dy}{y}} \to \alpha, \qquad \text{as } \theta \to 0,$$

where $\alpha \ne 0$ is the exponent of the regularly varying function $1/w(x)^2$ (note that if $w(x)$ is regularly varying, so is $1/w(x)^2$). Therefore, there is a constant $C(w)$ such that

$$\left|\frac{\int_0^\theta \frac{1}{w(y)^2}\frac{dy}{y}}{\frac{1}{w(\theta)^2}}\right| \leq C(w), \quad 0 < \theta < \theta_0. \qquad \square$$

LEMMA 3.3.15. *Given $\varepsilon > 0$, under the assumptions of Theorem 3.3.3, we have for $0 < a < b < 1$ and t fixed*

$$\mathrm{P}(\exists s, \rho(s,t) \leq \varepsilon, \exists x \in (a,b] : X_s \leq x < X_t) \leq \frac{C\varepsilon^2}{w(b)^2} + (b-a),$$

and

$$\mathrm{P}(\exists s, \rho(s,t) \leq \varepsilon, \exists x \in (a,b] : X_t \leq x < X_s) \leq \frac{C\varepsilon^2}{w(b)^2} + (b-a),$$

where C is a constant depending only on the function $w(x)$.

PROOF. Let $N \geq 0$ be the biggest integer such that $b/2^N \geq a$. Then,

$$
\begin{aligned}
&\mathrm{P}(\exists s, \rho(s,t) \leq \varepsilon, \exists x \in (a,b] : X_s \leq x < X_t) \\
&\leq \sum_{k=0}^{N-1} \mathrm{P}(\exists s, \rho(s,t) \leq \varepsilon, \exists x \in (2^{-k-1}b, 2^{-k}b] : X_s \leq x < X_t) \\
&\quad + \mathrm{P}(\exists s, \rho(s,t) \leq \varepsilon, \exists x \in (a, 2^{-N}b] : X_s \leq x < X_t) \\
&\leq \sum_{k=0}^{N-1} \mathrm{P}(\exists s, \rho(s,t) \leq \varepsilon, \exists x \in (2^{-k-1}b, 2^{-k}b] : X_s \leq 2^{-k}b < X_t) + \sum_{k=0}^{N-1} \mathrm{P}(2^{-k-1}b < X_t \leq 2^{-k}b) \\
&\quad + \mathrm{P}(\exists s, \rho(s,t) \leq \varepsilon, \exists x \in (a, 2^{-N}b] : X_s \leq 2^{-N}b < X_t) + \mathrm{P}(a < X_t \leq 2^{-N}b) \\
&\leq \sum_{k=0}^{N-1} \mathrm{P}(\exists s, \rho(s,t) \leq \varepsilon, \exists x \in (2^{-k-1}b, 2^{-k}b] : X_s \leq 2^{-k}b < X_t) + \sum_{k=0}^{N-1} (2^{-k}b - 2^{-k-1}b) \\
&\quad + \mathrm{P}(\exists s, \rho(s,t) \leq \varepsilon, \exists x \in (a, 2^{-N}b] : X_s \leq 2^{-N}b < X_t) + 2^{-N}b - a \\
&\leq \sum_{k=0}^{N} \mathrm{P}(\exists s, \rho(s,t) \leq \varepsilon : X_s \leq 2^{-k}b < X_t) + \sum_{k=0}^{N-1} (2^{-k}b - 2^{-k-1}b) + 2^{-N}b - a \\
&\leq \sum_{k=0}^{\infty} \frac{L\varepsilon^2}{w(2^{-k}b)^2} + (b-a) \quad \text{using WL-condition to bound the probabilities} \\
&\leq \frac{C\varepsilon^2}{w(b)^2} + (b-a) \quad \text{by Lemma 3.3.14.}
\end{aligned}
$$

The proof for the second part is similar; just change from $X_t \leq x < X_s$ for $2^{-k-1}b < x \leq 2^{-k}b$ to $X_t \leq 2^{-k-1}b < X_s$, with the same exceptional probability $(2^{-k}b - 2^{-k-1}b)$. □

For the following, we use C to denote a constant which may change from line to line and depends only on the weight function $w(x)$.

Let $e((s,x),(t,y)) := \max\{d(x,y), \rho(s,t)\}$, which is bounded by the Gaussian distance $(d(x,y)^2 + \rho(s,t)^2)^{1/2}$ on $E \times (0,1)$ and will be used as the 'ρ' in (iii) of Theorem 2.5.1.

LEMMA 3.3.16. *For* $t \in E$, $y \in (0,1)$, *let* $x_0 := \inf\{x : \text{ for some } s, e((s,x),(t,y)) < \varepsilon\}$, *then*

$$d(x_0, y) \leq \varepsilon. \tag{3.17}$$

PROOF. Indeed there exist a sequence $(s_n, x_n)_{n\in\mathbb{N}}$ in the set over which the infimum is taken such that $|x_n - x_0| \to 0$ as $n \to \infty$ and that $d(x_n, y) \leq \varepsilon$. By the continuity of $w(x)$, we have $d(x_n, y) \to d(x_0, y)$ as $n \to \infty$. Hence we have obtained $d(x_0, y) \leq \varepsilon$. □

REMARK. The finiteness of $d(x_0, y)$ implies that x_0 can't be zero in view of (3.13) since $w(x) \to \infty$ and $xw(x)^2 \to 0$ as $x \to 0$.

LEMMA 3.3.17. *For* $t \in E$, $y \in (0,1)$, *let* $x_1 := \sup\{x : \text{ for some } s, e((s,x),(t,y)) < \varepsilon\}$, *then*

$$d(y, x_1) \leq \varepsilon. \tag{3.18}$$

PROOF. By a similar argument as in the proof of the previous lemma. □

LEMMA 3.3.18. *Under the assumptions of Theorem 3.3.3, we have for all* $\varepsilon > 0$ *and* $(t,y) \in E \times [0,1]$,

$$w(y)^2 \mathrm{P}(\sup_{(s,x): e<\varepsilon, x\leq y} |1_{X_s \leq x} - 1_{X_t \leq x}| > 0) \leq C\varepsilon^2.$$

Proof.

$$
\begin{aligned}
& w(y)^2 \mathrm{P}\Big(\sup_{(s,x):e<\varepsilon, x\le y} |1_{X_s\le x} - 1_{X_t \le x}| > 0\Big) \\
& = w(y)^2 \Big(\mathrm{P}(\exists (s,x), e((s,x),(t,y)) < \varepsilon, x \le y : X_s \le x < X_t) \\
& \qquad + \mathrm{P}(\exists (s,x), e((s,x),(t,y)) < \varepsilon, x \le y : X_t \le x < X_s) \Big) \\
& = w(y)^2 \Big(\mathrm{P}(\exists (s,x), e((s,x),(t,y)) < \varepsilon, x \in (x_0, y] : X_s \le x < X_t) \\
& \qquad + \mathrm{P}(\exists (s,x), e((s,x),(t,y)) < \varepsilon, x \in (x_0, y] : X_t \le x < X_s) \Big) \\
& \le w(y)^2 (C\varepsilon^2 / w(y)^2 + (y - x_0)) \text{ by Lemma 3.3.15} \\
& \le C\varepsilon^2.
\end{aligned}
$$

For the last inequality, we used

$$
w(y)^2 (y - x_0) \le d(x_0, y)^2 \le \varepsilon^2 \quad \text{by (3.17).} \qquad \square
$$

Lemma 3.3.19. *Under the assumptions of Theorem 3.3.3, we have for all* $\varepsilon > 0$ *and* $(t,y) \in E \times [0,1]$,

$$
w(x_1)^2 \mathrm{P}\Big(\sup_{(s,x):e<\varepsilon, x>y} |1_{X_s\le x} - 1_{X_t \le x}| > 0\Big) \le C\varepsilon^2.
$$

PROOF.

$$
\begin{aligned}
& w(x_1)^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon, x>y} |1_{X_s \le x} - 1_{X_t \le x}| > 0) \\
& = w(x_1)^2 \Big(\mathrm{P}(\exists (s,x), e((s,x),(t,y)) < \varepsilon, x > y : X_s \le x < X_t) \\
& \qquad + \mathrm{P}(\exists (s,x), e((s,x),(t,y)) < \varepsilon, x > y : X_t \le x < X_s) \Big) \\
& = w(x_1)^2 \Big(\mathrm{P}(\exists (s,x), e((s,x),(t,y)) < \varepsilon, x \in (y, x_1] : X_s \le x < X_t) \\
& \qquad + \mathrm{P}(\exists (s,x), e((s,x),(t,y)) < \varepsilon, x \in (y, x_1] : X_t \le x < X_s) \Big) \\
& \le w(x_1)^2 (C\varepsilon^2 / w(x_1)^2 + (x_1 - y)) \text{ by Lemma 3.3.15} \\
& \le C\varepsilon^2.
\end{aligned}
$$

For the last inequality, we used

$$
w(x_1)^2 (x_1 - y) \le d(y, x_1)^2 \le \varepsilon^2 \quad \text{by (3.18).} \qquad \square
$$

In the following lemma, for fixed $(t,y) \in E \times [0,1]$, we write $\sup_{(s,x):e<\varepsilon}$ for $\sup_{\{(s,x):e((s,x),(t,y))<\varepsilon\}}$ and the same token applies to other similar quantities.

LEMMA 3.3.20. *Under the assumptions of Theorem 3.3.3, we have for all* $\varepsilon > 0$ *and* $(t,y) \in E \times [0,1]$,

$$
\sup_{\alpha > 0} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon} |w(x) 1_{X_s \le x} - w(y) 1_{X_t \le y}| > \alpha) \le C\varepsilon^2.
$$

PROOF. We split the quantity:

$$
w(x) 1_{X_s \le x} - w(y) 1_{X_t \le y} = [w(x) 1_{X_t \le x} - w(y) 1_{X_t \le y}] + [w(x)(1_{X_s \le x} - 1_{X_t \le x})].
$$

Consider the weak L_2 norms of the components:

$$A := \sup_{\alpha>0} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon} |w(x)1_{X_t \le x} - w(y)1_{X_t \le y}| > \alpha) \tag{3.19}$$

$$B := \sup_{\alpha>0} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon} w(x)|1_{X_s \le x} - 1_{X_t \le x}| > \alpha). \tag{3.20}$$

First we estimate A. Since

$$\sup_{\alpha>0} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon} |w(x)1_{X_t \le x} - w(y)1_{X_t \le y}| > \alpha)$$
$$\le \sup_{\alpha>0} \alpha^2 \mathrm{P}(\sup_{x:d(x,y)<\varepsilon} |w(x)1_{X_t \le x} - w(y)1_{X_t \le y}| > \alpha)$$

and t is fixed, this is the case in Example 4.9 in Andersen et al. (1988). Hence we have

$$A := \sup_{\alpha>0} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon} |w(x)1_{X_t \le x} - w(y)1_{X_t \le y}| > \alpha) \le C\varepsilon^2. \tag{3.21}$$

Now we consider B. Since

$$\sup_{\alpha>0} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon} w(x)|1_{X_s \le x} - 1_{X_t \le x}| > \alpha) \le \sup_{\alpha>0} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon, x\le y} w(x)|1_{X_s \le x} - 1_{X_t \le x}| > \alpha)$$
$$+ \sup_{\alpha>0} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon, x>y} w(x)|1_{X_s \le x} - 1_{X_t \le x}| > \alpha),$$

it suffices to consider the to bound the later two quantities. Without loss of generality, we assume $w(x)$ is monotone on $(0, 1/2]$. Let $x_\alpha = \sup\{x \in [0, 1/2] : w(x) > \alpha\}$ for $\alpha > 0$.

Case $x \le y$.

Recall $x_0 = \inf\{x : e((s, x), (t, y)) < \varepsilon\}$. First we consider the extreme cases for x_α.

(1). If $x_\alpha > y$, then $\alpha < w(y)$, consequently

$$\sup_{\alpha < w(y)} \alpha^2 \mathrm{P}\big(\sup_{(s,x):e<\varepsilon, x\le y} w(x)|1_{X_s\le x} - 1_{X_t\le x}| > \alpha\big)$$
$$\le w(y)^2 \mathrm{P}\big(\sup_{(s,x):e<\varepsilon, x\le y} |1_{X_s\le x} - 1_{X_t\le x}| > 0\big) \le C\varepsilon^2.$$

by Lemma 3.3.18. consider $\alpha \ge w(y)$, i.e. $x_\alpha \le y$.

(2). If $x_\alpha \le x_0$, then $w(x_0) \le \alpha$, hence $w(x) \le \alpha$ for $x_0 \le x$. For such α, the event under the probability of (4.14) is empty.

(3). Now $x_0 < x_\alpha \le y$. In this case, $w(y) \le \alpha < w(x_0)$. Take $\varepsilon > 0$. We have

$$\begin{aligned}
B :=& \sup_{w(y)\le\alpha<w(x_0)} \alpha^2 \mathrm{P}\big(\sup_{(s,x):e<\varepsilon} w(x)|1_{X_s\le x} - 1_{X_t\le x}| > \alpha\big) \\
&\le 4 \sup_{w(y)\le\alpha<w(x_0)} \alpha^2 \mathrm{P}\big(\sup_{(s,x):e<\varepsilon} w(x) 1_{X_s\le x<X_t} > \alpha\big) \\
&\quad + 4 \sup_{w(y)\le\alpha<w(x_0)} \alpha^2 \mathrm{P}\big(\sup_{(s,x):e<\varepsilon} w(x) 1_{X_t\le x<X_s} > \alpha\big) \\
&= 4I + 4II.
\end{aligned}$$

For I,

$$\begin{aligned}
I =& \sup_{w(y)\le\alpha<w(x_0)} \alpha^2 \mathrm{P}\big(\sup_{(s,x):e<\varepsilon} w(x) 1_{X_s\le x<X_t} > \alpha\big) \\
&\le \sup_{x_0<x_\alpha\le y} w(x_\alpha)^2 \mathrm{P}\big(\sup_{(s,x):e<\varepsilon} w(x) 1_{X_s\le x<X_t} > \alpha\big) \\
&\le \sup_{x_0<x_\alpha\le y} w(x_\alpha)^2 \mathrm{P}\big(\sup_{(s,x):e<\varepsilon} 1_{X_s\le x<X_t, x\le x_\alpha} > 0\big) \\
&\le \sup_{x_0<x_\alpha\le y} w(x_\alpha)^2 \big(C\varepsilon^2/w(x_\alpha)^2 + (x_\alpha - x_0)\big) \text{ using Lemma 3.3.15} \\
&\le C\varepsilon^2.
\end{aligned}$$

For the last inequality, we used

$$w(x_\alpha)^2(x_\alpha - x_0) \le d(x_0, x_\alpha)^2 \le d(x_0, y)^2 \le \varepsilon^2.$$

by Lemma 3.3.11 and Lemma 3.3.16.

II can be handled in the same way.

Case $x > y$.

Recall $x_1 = \sup\{x : e((s,x),(t,y)) < \varepsilon\}$. First we consider the extreme cases for x_α.

(1). If $x_\alpha > x_1$, then $\alpha < w(x_1)$, consequently

$$\sup_{\alpha < w(x_1)} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon, x>y} w(x)|1_{X_s \le x} - 1_{X_t \le x}| > \alpha)$$
$$\le w(x_1)^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon, x>y} |1_{X_s \le x} - 1_{X_t \le x}| > 0) \le C\varepsilon^2.$$

by Lemma 3.3.19. consider $\alpha \ge w(y)$, i.e. $x_\alpha \le y$.

(2). If $x_\alpha \le y$, then $w(y) \le \alpha$, hence $w(x) \le \alpha$ for $y \le x$. For such α, the event under the probability of (4.14) is empty.

(3). Now $y < x_\alpha \le x_1$. In this case, $w(x_1) \le \alpha < w(y)$. Take $\varepsilon > 0$. We have

$$\begin{aligned} B &:= \sup_{w(x_1) \le \alpha < w(y)} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon} w(x)|1_{X_s \le x} - 1_{X_t \le x}| > \alpha) \\ &\le 4 \sup_{w(x_1) \le \alpha < w(y)} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon} w(x) 1_{X_s \le x < X_t} > \alpha) \\ &\quad + 4 \sup_{w(x_1) \le \alpha < w(y)} \alpha^2 \mathrm{P}(\sup_{(s,x):e<\varepsilon} w(x) 1_{X_t \le x < X_s} > \alpha) \\ &= 4I + 4II. \end{aligned}$$

For I,

$$\begin{aligned}
I &= \sup_{w(x_1)\leq\alpha<w(y)} \alpha^2 \mathrm{P}\Big(\sup_{(s,x):e<\varepsilon} w(x)1_{X_s\leq x<X_t} > \alpha\Big) \\
&\leq \sup_{y<x_\alpha\leq x_1} w(x_\alpha)^2 \mathrm{P}\Big(\sup_{(s,x):e<\varepsilon} w(x)1_{X_s\leq x<X_t} > \alpha\Big) \\
&\leq \sup_{y<x_\alpha\leq x_1} w(x_\alpha)^2 \mathrm{P}\Big(\sup_{(s,x):e<\varepsilon} 1_{X_s\leq x<X_t, x\leq x_\alpha} > 0\Big) \\
&\leq \sup_{y<x_\alpha\leq x_1} w(x_\alpha)^2 \big(C\varepsilon^2/w(x_\alpha)^2 + (x_\alpha - y)\big) \text{ using Lemma 3.3.15} \\
&\leq C\varepsilon^2.
\end{aligned}$$

For the last inequality, we used

$$w(x_\alpha)^2(x_\alpha - y) \leq d(y, x_\alpha)^2 \leq d(y, x_1)^2 \leq \varepsilon^2.$$

by Lemma 3.3.11 and Lemma 3.3.17.

II can be handled in the same way. Hence we have $B \leq C\varepsilon^2$. This together with (3.21) completes the proof. □

Proof of Theorem 3.3.3. We use Theorem 2.5.1. In this case, the random element is the uniform process $\{X_s : s \in E\}$, the class of functions is $\mathcal{F} = \{f_{s,x}(\cdot) = w(x)1_{\delta_s(\cdot)\leq x} : s \in E,\ x \in [0, 1/2]\}$, where $\delta_s(\cdot)$ is the evaluation at s, hence $f_{s,x}(X) = w(x)1_{X_s\leq x}$. Since $X(s)$ takes values on $(0, 1)$ and $xw(x) \to 0$ as $x \to 0$, almost surely

$$\sup_{s\in E,\, x\in[0,1/2]} f_{s,x}(X) = \sup_{s\in E,\, x\in[0,1/2]} w(x)1_{X_s\leq x} < \infty.$$

Also we observe for each $s \in E, x \in [0, 1/2]$

$$\mathrm{P}f_{s,x}(X)^2 = \mathrm{P}(w(x)1_{X_s\leq x})^2 = w(x)^2 x < \infty,$$

and

$$\sup_{s\in E, x\in(0,1/2]} |\mathrm{P} f_{s,x}| = \sup_{s\in E, x\in(0,1/2]} |\mathrm{P} w(x) 1_{X_s\le x}| = w(x)x < \infty.$$

Thus we have $\mathcal{F} \subset L^2(\mathrm{P})$ and $\sup_{f\in\mathcal{F}} |\mathrm{P}f| < \infty$.

Since $w(x)$ is decreasing near 0,

$$\begin{aligned}\lim_{\alpha\to\infty} \alpha^2 \mathrm{P}(\sup_{s\in E, x\in(0,1/2]} w(x)1_{X_s\le x} > \alpha) &\le \lim_{\alpha\to\infty} \alpha^2 \mathrm{P}(\sup_{s\in E} w(X_s) > \alpha) \\ &= 0 \text{ by assumption of Theorem 3.3.3,}\end{aligned}$$

which verifies (i) in Theorem 2.5.1. This together with Proposition 3.3.13 for the Pre-Gaussian condition (ii) and Lemma 3.3.20 for the local modulus condition (iii) completes the proof. □

3.4 An example

A special class of uniform processes (copula processes) can be obtained from distributional transform. Specifically, given a process $Y := \{Y_t : t \in E\}$. Define $X := X_t := \tilde{F}_t(Y_t)$, where $\tilde{F}_t(\cdot)$ is the distributional transform of the df of Y_t. Now, we give an example as an application of Theorem 3.3.3 when $\{Y_t : t \in E\} = \{B_t : t \in [1, 2]\}$.

Theorem 3.4.1. *Let $\{B_t : t \ge 0\}$ be a Brownian motion and $F_t(x)$ be the distribution function of B_t. Let $w(x) = x^{-\alpha}L(x)$, for $0 < x < 1$, $0 < \alpha < 1/2$, and $L(x)$ slowly varying at 0 and assume $w(x)$ is symmetric about $1/2$. Further assume that $xw(x)^2$ is non-decreasing near 0. Then*

$$\{w(F_t(y))(1_{B_t\le y} - F_t(y)) : t \in [1, 2],\ y \in \mathbb{R}\} \in \mathrm{CLT}.$$

Remark 3.4.2. The interval $[1, 2]$ is not special; the theorem remains true for any interval $[a, b]$ provided $a > 0$.

We will verify the conditions in Proposition 3.3.5 to prove this theorem at the end of this section. To this end, we start with some lemmas. For the following, let $\phi(x) = (2\pi)^{-1/2} e^{-x^2/2}$ and $\Phi(y) := (2\pi)^{-1/2} \int_{-\infty}^{y} e^{-s^2/2}\, ds$.

LEMMA 3.4.3 (Feller (1968), p. 175). *For $y > 0$,*

$$y^{-1}(1 - y^{-2})(2\pi)^{-1/2} e^{-y^2/2} \leq \Phi(-y) \leq y^{-1}(2\pi)^{-1/2} e^{-y^2/2}.$$

In particular, we have for $y > \sqrt{2}$,

$$2^{-1} y^{-1}(2\pi)^{-1/2} e^{-y^2/2} \leq \Phi(-y) \leq y^{-1}(2\pi)^{-1/2} e^{-y^2/2}.$$

LEMMA 3.4.4 (Seneta (1976), p. 18). *Let $L(x)$ be a slowly varying function at 0, then for any $\gamma > 0$,*

$$x^{\gamma} L(x) \to 0, x^{-\gamma} L(x) \to \infty \text{ as } x \to 0.$$

Consequently, for $0 < \gamma_1 < 2\alpha < \gamma_2 < 1$ and a function $L(x)$ slowly varying (at 0), *there are constants c_1, c_2,*

$$c_1 x^{\gamma_2} \leq x^{2\alpha}/L(x) \leq c_2 x^{\gamma_1}, \quad 0 < x < 1/2.$$

For $c > 0$, let $L_c(x) = \exp(c\sqrt{\ln(1/x)})$.

LEMMA 3.4.5. *The function $L_c(x)$ is slowly varying at 0; that is for all $\lambda > 0$*

$$\lim_{x \to 0} \frac{L_c(\lambda x)}{L_c(x)} = 1.$$

PROOF. By definition. □

LEMMA 3.4.6. *For $0 < x < 1/4$, let $y = -\Phi^{-1}(x)$. Then*

$$y \le \sqrt{2\ln(1/x)}$$

and

$$\phi(-\Phi^{-1}(x) + c) \le CxL_C(x) \qquad \text{for } c < 0,$$

$$\phi(-\Phi^{-1}(x) + c) \le 2^{3/2}x\sqrt{\ln(1/x)} \qquad \text{for } c \ge 0,$$

where C depends only on c.

PROOF. By Lemma 3.4.3, for $y > (2\pi)^{-1/2}$, $x \le e^{-y^2/2}$; hence $y \le \sqrt{2\ln(1/x)}$.

$$\begin{aligned}\phi(-\Phi^{-1}(x) + c) &= (2\pi)^{-1/2}\exp(-\tfrac{(y+c)^2}{2})\\ &= (2\pi)^{-1/2}\exp(-\tfrac{y^2}{2})\exp(-yc)\exp(-c^2/2)\\ &\le 2y\Phi(-y)\exp(-yc) \qquad \text{by Lemma 3.4.3}\\ &\le 2xy\exp(-yc).\end{aligned}$$

The statement for $c > 0$ follows from that $\exp(-yc) \le 1$ and $y \le \sqrt{2\ln(1/x)}$. For $c \le 0$ the statement follows from that $y \le C\exp(yC)$ for some constant C. □

THEOREM 3.4.7 (Borell, see also Ledoux (2001), Theorem 7.1). *Let $G = (G_t)_{t\in T}$ be a centered Gaussian process indexed by countable set T such that $\sup_{t\in T} G_t < \infty$ almost surely. Then, $\mathrm{E}(\sup_{t\in T} G_t) < \infty$ and for every $r > 0$*

$$\mathrm{P}(\{\sup_{t\in T} G_t \ge \mathrm{E}(\sup_{t\in T} G_t) + r\}) \le e^{-r^2/2\sigma^2},$$

where $\sigma = \sup_{t\in T}(\mathrm{E}G_t^2)^{1/2}$.

For the following, let B_t be a Brownian motion and $F_t(x)$ the distribution function of B_t, which is $\Phi(\frac{x}{\sqrt{t}})$. Also for $1 \le t \le 2$, $0 < \varepsilon < 1/2$

$$\begin{aligned} D &:= \sup_{t<s\le t+\varepsilon} \tfrac{B_s - B_t}{\sqrt{s}} \\ m &:= m(t,\varepsilon) := \mathrm{E} \sup_{t<s\le t+\varepsilon} \tfrac{B_s - B_t}{\sqrt{s}} \\ m_0 &:= \sup\{m(t,\varepsilon) : 1 \le t \le 2, 0 < \varepsilon < 1/2\}. \end{aligned}$$

We use C to denote a constant, which may vary in each occurrence.

LEMMA 3.4.8. *For $1 \le t \le 2$, $0 < \varepsilon < 1/2$*

$$m \le 2(2/\pi)^{1/2}\varepsilon^{1/2}.$$

PROOF.

$$\begin{aligned} m &= \mathrm{E} \sup_{t<s\le t+\varepsilon} \tfrac{B_s - B_t}{\sqrt{s}} \\ &\le \mathrm{E} \sup_{t<s\le t+\varepsilon} \tfrac{|B_s - B_t|}{\sqrt{t}} \\ &\le \mathrm{E}\varepsilon^{1/2} 2|N(0,1)| \\ &\le 2(2/\pi)^{1/2}\varepsilon^{1/2}. \end{aligned}$$

□

LEMMA 3.4.9. *Let $d := \mathrm{E}(\sup_{1\le t\le 2} \frac{B_t}{\sqrt{t}}) > 0$, then*

$$\mathrm{P}(\inf_{1\le t\le 2} F_t(B_t) \le x) \le (2\pi)^{1/2}\phi(-\Phi^{-1}(x) - d).$$

Proof.

$$\begin{aligned}
\mathrm{P}(\inf_{1\le t\le 2} F_t(B_t) \le x) &= \mathrm{P}(\inf_{1\le t\le 2} \tfrac{B_t}{\sqrt{t}} \le \Phi^{-1}(x)) \\
&= \mathrm{P}(\sup_{1\le t\le 2} \tfrac{-B_t}{\sqrt{t}} \ge -\Phi^{-1}(x)) \\
&= \mathrm{P}(\sup_{1\le t\le 2} \tfrac{-B_t}{\sqrt{t}} \ge d - \Phi^{-1}(x) - d)
\end{aligned}$$

by Theorem 3.4.7 and for x so that $-\Phi^{-1}(x) - d > 0$

$$\begin{aligned}
&\le \exp(-\tfrac{(-\Phi^{-1}(x)-d)^2}{2}) \\
&= (2\pi)^{1/2}\phi(-\Phi^{-1}(x) - d).
\end{aligned}$$

Note that here $\sigma^2 = \sup_{1\le t\le 2} \mathrm{E}(\frac{-B_t}{\sqrt{t}})^2 = 1$. □

Lemma 3.4.10. *Let $w(x) = x^{-\alpha}L(x)$, $0 < \alpha < 1/2$ and $L(x)$ a slowly varying function (growing to infinity as $x \downarrow 0$). Assume $w(x)$ is decreasing near 0. Then*

$$\lim_{\lambda\to\infty} \lambda^2 \mathrm{P}(\sup_{1\le t\le 2} w(F_t(B_t)) > \lambda) = 0.$$

Proof. Let $\lambda = w(x)$.

$$\begin{aligned}
\lim_{\lambda\to\infty} \lambda^2 \mathrm{P}(\sup_{1\le t\le 2} w(F_t(B_t)) > \lambda) &= \lim_{\lambda\to\infty} \lambda^2 \mathrm{P}(w(\inf_{1\le t\le 2} F_t(B_t)) > \lambda) \\
&= \lim_{x\to 0} w(x)^2 \mathrm{P}(\inf_{1\le t\le 2} F_t(B_t) \le x) \\
&\le \lim_{x\to 0} w(x)^2 (2\pi)^{1/2} \phi(-\Phi^{-1}(x) - d) \\
&\le \lim_{x\to 0} x^{-2\alpha} L(x)^2 (2\pi)^{1/2} C x L_C(x) \\
&= 0 \qquad \text{by Lemma 3.4.4, Lemma 3.4.9.}
\end{aligned}$$

□

Lemma 3.4.11. *For $1 \le t \le 2$, $0 < \varepsilon < 1/2$, and $l > m$,*

$$P(\tfrac{B_t}{\sqrt{t}} < l \le \sup_{t<s\le t+\varepsilon} \tfrac{B_s}{\sqrt{s}}) \le C_t \varepsilon^{1/2} \phi(l-m)^{\frac{t+\varepsilon}{t+2\varepsilon}},$$

where C_t is a constant depending only on t. In particular, if let $C := \sup_{1\le t\le 2} C_t$, and recall $m_0 := \sup\{m(t,\varepsilon) : 1 \le t \le 2, 0 < \varepsilon < 1/2\}$, then for $l > m_0$

$$P(\tfrac{B_t}{\sqrt{t}} < l \le \sup_{t<s\le t+\varepsilon} \tfrac{B_s}{\sqrt{s}}) \le C \varepsilon^{1/2} \phi(l-m_0)^{\frac{t+\varepsilon}{t+2\varepsilon}}. \tag{3.22}$$

Proof. Since $\sigma^2 := \sup_{t<s\le t+\varepsilon} E(\frac{B_s - B_t}{\sqrt{s}})^2 = \frac{\varepsilon}{t+\varepsilon}$, by Borell's concentration inequality [3.4.7], it follows that for $r > 0$

$$P(D > m + r) \le e^{-r^2(t+\varepsilon)/(2\varepsilon)}. \tag{3.23}$$

Hence,

$$\begin{aligned}
P(\tfrac{B_t}{\sqrt{t}} < l \le \sup_{t<s\le t+\varepsilon} \tfrac{B_s}{\sqrt{s}}) &\le P(\tfrac{B_t}{\sqrt{t}} < l \le \sup_{t<s\le t+\varepsilon} (\tfrac{B_s}{\sqrt{s}} - \tfrac{B_t}{\sqrt{s}}) + \sup_{t<s\le t+\varepsilon} \tfrac{B_t}{\sqrt{s}}) \\
&= E_{\frac{B_t}{\sqrt{t}}} P(\tfrac{B_t}{\sqrt{t}} < l \le \sup_{t<s\le t+\varepsilon} (\tfrac{B_s}{\sqrt{s}} - \tfrac{B_t}{\sqrt{s}}) + \sup_{t<s\le t+\varepsilon} \tfrac{B_t}{\sqrt{s}} | \tfrac{B_t}{\sqrt{t}}) \quad \text{conditioning on } \tfrac{B_t}{\sqrt{t}}
\end{aligned}$$

by the independence of $\{B_s - B_t : s > t\}$ and B_t

$$\begin{aligned}
&= \int_{-\infty}^{\infty} P\big((y < l \le D + \sup_{t<s\le t+\varepsilon} \{(t/s)^{1/2} y\}\big) \tfrac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \\
&= \int_{0}^{\infty} P\big(y < l \le D + \sup_{t<s\le t+\varepsilon} \{(t/s)^{1/2} y\}\big) \tfrac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \\
&\quad + \int_{-\infty}^{0} P\big(y < l \le D + \sup_{t<s\le t+\varepsilon} \{(t/s)^{1/2} y\}\big) \tfrac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \\
&= I + II. \qquad (3.24)
\end{aligned}$$

Note

$$\sup_{t<s\le t+\varepsilon}\{(t/s)^{1/2}y\} = y \qquad \text{for } y>0,$$
$$\sup_{t<s\le t+\varepsilon}\{(t/s)^{1/2}y\} = ((t/(t+\varepsilon))^{1/2}y =: ay \qquad \text{for } y\le 0.$$

Therefore,

$$\begin{aligned} I &= \int_0^\infty \mathrm{P}\big(y<l\le D+y\big)\tfrac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy \\ &= \int_0^{l-m} \mathrm{P}\big(y<l\le D+y\big)\tfrac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy + \int_{l-m}^{l} \mathrm{P}\big(y<l\le D+y\big)\tfrac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy \\ &= \int_0^{l-m} \mathrm{P}\big(D\ge l-y\big)\tfrac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy + \int_{l-m}^{l} \mathrm{P}\big(y<l\le D+y\big)\tfrac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy \end{aligned}$$

by inequality (3.23) for the first summand and noting $r := l-y-m>0$

$$\le \int_0^{l-m} e^{-\frac{(l-y-m)^2(t+\varepsilon)}{2\varepsilon}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}}\,dy + \int_{l-m}^{l} \mathrm{P}\big(y<l\le D+y\big)\tfrac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy$$

by completing the square in y for the first summand

$$\le (\tfrac{\varepsilon}{t+2\varepsilon})^{1/2}e^{-\frac{(l-m)^2}{2}\frac{t+\varepsilon}{t+2\varepsilon}} + m\phi(l-m)$$

bounding m using Lemma 3.4.8

(3.25)
$$\le (\tfrac{\varepsilon}{t+2\varepsilon})^{1/2}(2\pi)^{1/2}\phi(l-m)^{\frac{t+\varepsilon}{t+2\varepsilon}} + 2(2/\pi)^{1/2}\varepsilon^{1/2}\phi(l-m).$$

For II,

$$II = \int_{-\infty}^{0} \mathrm{P}\big(y < l \le D + ay\big) \tfrac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$$
$$\le \int_{-\infty}^{0} \mathrm{P}\big(D \ge l - ay\big) \tfrac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$$

by equation (3.23)

$$\le \int_{-\infty}^{0} e^{-\frac{(l-ay-m)^2(t+\varepsilon)}{2\varepsilon}} \tfrac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$$

by completing the square in y

$$\le (\tfrac{\varepsilon}{t+\varepsilon})^{1/2} e^{-\frac{(l-m)^2}{2}}$$
$$= (\tfrac{\varepsilon}{t+\varepsilon})^{1/2} (2\pi)^{1/2} \phi(l-m). \tag{3.26}$$

Combining (3.24), (3.25), and(3.26) completes the proof. □

LEMMA 3.4.12. *For* $1 \le t \le 2$, $0 < \varepsilon \le 1/2$, *there is a universal constant* C *such that for* $0 < x < 1/4$

$$\mathrm{P}(\tfrac{B_t}{\sqrt{t}} \le \Phi^{-1}(x) < \sup_{t<s\le t+\varepsilon} \tfrac{B_s}{\sqrt{s}}) \le C\varepsilon^{1/2}(x \ln \tfrac{1}{x}).$$

PROOF.

$$\mathrm{P}(\tfrac{B_t}{\sqrt{t}} \le \Phi^{-1}(x) < \sup_{t<s\le t+\varepsilon} \tfrac{B_s}{\sqrt{s}})$$
$$= \mathrm{P}(\tfrac{B_t}{\sqrt{t}} \le \Phi^{-1}(x) < [\sup_{t<s\le t+\varepsilon} \tfrac{B_s}{\sqrt{s}} - \tfrac{B_t}{\sqrt{s}}] + \sup_{t<s\le t+\varepsilon} \tfrac{B_t}{\sqrt{s}})$$

letting $D = \sup_{t<s\leq t+\varepsilon} \frac{B_s}{\sqrt{s}} - \frac{B_t}{\sqrt{s}}$ and noting $B_t < 0$ inside the probability above

$$\leq \mathrm{E}_D \mathrm{P}(\tfrac{B_t}{\sqrt{t}} \leq \Phi^{-1}(x) \leq D + \tfrac{B_t}{\sqrt{t+\varepsilon}} | D)$$

by independence of $\{B_s - B_t : s > t\}$ and B_t

$$= \mathrm{E}_D \mathrm{P}(\tfrac{B_t}{\sqrt{t}} \leq \Phi^{-1}(x) \leq \tfrac{B_t}{\sqrt{t+\varepsilon}} + D)$$
$$= \mathrm{E}_D \mathrm{P}((\tfrac{t+\varepsilon}{t})^{1/2}(\Phi^{-1}(x) - D) \leq \tfrac{B_t}{\sqrt{t}} \leq \Phi^{-1}(x))$$

bounding the density of $\frac{B_t}{\sqrt{t}}$ from above by $\phi(\Phi^{-1}(x))$

$$\leq \mathrm{E}_D \phi(\Phi^{-1}(x))[(1 - (\tfrac{t+\varepsilon}{t})^{1/2})\Phi^{-1}(x) + (\tfrac{t+\varepsilon}{t})^{1/2} D]$$
$$\leq \phi(\Phi^{-1}(x))(-\Phi^{-1}(x))(\varepsilon/t) + \phi(-\Phi^{-1}(x))(\tfrac{t+\varepsilon}{t})^{1/2} \mathrm{E}_D D$$
$$\leq C(x \ln \tfrac{1}{x})(\varepsilon/t) + Cx(\ln \tfrac{1}{x})^{1/2} 8\varepsilon^{1/2} \text{ by Lemma 3.4.6 and Lemma 3.4.8}$$
$$\leq C\varepsilon^{1/2}(x \ln \tfrac{1}{x}). \qquad \square$$

PROPOSITION 3.4.13. *For $1 \leq t \leq 2$, $0 < \varepsilon \leq 1/2$, there is a universal constant C such that for $0 < x < 1/4$*

$$\mathrm{P}(F_t(B_t) \leq x < \sup_{s:|s-t|\leq\varepsilon} F_s(B_s)) \leq C\varepsilon(x \ln \tfrac{1}{x}) + C\varepsilon^{1/2}\phi(-\Phi^{-1}(x) - m_0)^{\frac{t}{t+\varepsilon}}.$$

Proof.

$$
\begin{aligned}
&\mathrm{P}(F_t(B_t) \le x < \sup_{\{s:|s-t|\le\varepsilon\}} F_s(B_s)) \\
&= \mathrm{P}(\Phi(\tfrac{B_t}{\sqrt{t}}) \le x < \sup_{\{s:|s-t|\le\varepsilon\}} \Phi(\tfrac{B_s}{\sqrt{s}})) \\
&= \mathrm{P}(\tfrac{B_t}{\sqrt{t}} \le \Phi^{-1}(x) < \sup_{\{s:|s-t|\le\varepsilon\}} \tfrac{B_s}{\sqrt{s}}) \\
&\le \mathrm{P}(\tfrac{B_t}{\sqrt{t}} \le \Phi^{-1}(x) < \sup_{t<s\le t+\varepsilon} \tfrac{B_s}{\sqrt{s}}) + \mathrm{P}(\tfrac{B_t}{\sqrt{t}} \le \Phi^{-1}(x) < \sup_{t-\varepsilon\le s<t} \tfrac{B_s}{\sqrt{s}}) \\
&= I + II.
\end{aligned}
$$

By Lemma 3.4.12,

$$
I \le C\varepsilon^{1/2}(x \ln \tfrac{1}{x}). \tag{3.27}
$$

Now we consider II.

$$
\begin{aligned}
II &= \mathrm{P}(\tfrac{B_t}{\sqrt{t}} \le \Phi^{-1}(x) < \sup_{t-\varepsilon\le s<t} \tfrac{B_s}{\sqrt{s}}) \\
&= \mathrm{P}(\tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}} \le \Phi^{-1}(x),\ \tfrac{B_t}{\sqrt{t}} \le \Phi^{-1}(x) < \sup_{t-\varepsilon\le s<t} \tfrac{B_s}{\sqrt{s}}) \\
&\qquad + \mathrm{P}(\tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}} > \Phi^{-1}(x),\ \tfrac{B_t}{\sqrt{t}} \le \Phi^{-1}(x) < \sup_{t-\varepsilon\le s<t} \tfrac{B_s}{\sqrt{s}}) \\
&\le \mathrm{P}(\tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}} \le \Phi^{-1}(x) < \sup_{t-\varepsilon<s\le t} \tfrac{B_s}{\sqrt{s}}) + \mathrm{P}(\tfrac{B_t}{\sqrt{t}} \le \Phi^{-1}(x) < \tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}}) \\
&\le \mathrm{P}(\tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}} \le \Phi^{-1}(x) < \sup_{t-\varepsilon<s\le t} \tfrac{B_s}{\sqrt{s}}) + \mathrm{P}(\tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}} \le -\Phi^{-1}(x) < \tfrac{B_t}{\sqrt{t}})
\end{aligned}
$$

(3.28)

$$
\le C\varepsilon^{1/2}(x \ln \tfrac{1}{x}) + C\varepsilon^{1/2}\phi(-\Phi^{-1}(x) - m_0)^{\frac{t}{t+\varepsilon}} \quad \text{by Lemmas 3.4.12 and 3.4.11.}
$$

□

Proposition 3.4.14. *For $1 \le t \le 2$, $0 < \varepsilon \le 1/2$, there is a universal constant C such that for $0 < x < 1/4$*

$$P(\inf_{\{s:|s-t|\le\varepsilon\}} F_s(B_s) \le x < (F_t(B_t)) \le C\varepsilon^{1/2}\phi(-\Phi^{-1}(x) - m_0)^{\frac{t}{t+\varepsilon}} + C\varepsilon^{1/2}(x \ln \tfrac{1}{x}).$$

Proof. First we consider the case $\{s > t : |s-t| \le \varepsilon\}$. Let $D = \sup_{t<s\le t+\varepsilon} \frac{B_s}{\sqrt{s}} - \frac{B_t}{\sqrt{s}}$.

$$\begin{aligned}
&P(\inf_{t<s\le t+\varepsilon} F_s(B_s) \le x < F_t(B_t)) \\
&= P(\inf_{t<s\le t+\varepsilon} \tfrac{B_s}{\sqrt{s}} \le \Phi^{-1}(x) < \tfrac{B_t}{\sqrt{t}}) \\
&= P(\tfrac{B_t}{\sqrt{t}} < -\Phi^{-1}(x) \le \sup_{t<s\le t+\varepsilon} \tfrac{B_s}{\sqrt{s}}) \\
&\le C\varepsilon^{1/2}\phi(-\Phi(x) - m_0)^{\frac{t+\varepsilon}{t+2\varepsilon}} \qquad \text{by Lemma 3.4.11.}
\end{aligned}$$

For the the case $\{s < t : |s-t| \le \varepsilon\}$,

$$\begin{aligned}
&P(\inf_{t-\varepsilon\le s<t} F_s(B_s) \le x < F_t(B_t)) \\
&= P(\tfrac{B_t}{\sqrt{t}} < -\Phi^{-1}(x) \le \sup_{t-\varepsilon\le s<t} \tfrac{B_s}{\sqrt{s}}) \\
&= P(\tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}} < -\Phi^{-1}(x), \tfrac{B_t}{\sqrt{t}} < -\Phi^{-1}(x) \le \sup_{t-\varepsilon\le s<t} \tfrac{B_s}{\sqrt{s}}) \\
&\quad + P(\tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}} \ge -\Phi^{-1}(x), \tfrac{B_t}{\sqrt{t}} < -\Phi^{-1}(x) \le \sup_{t-\varepsilon\le s<t} \tfrac{B_s}{\sqrt{s}}) \\
&= P(\tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}} < -\Phi^{-1}(x) \le \sup_{t-\varepsilon\le s<t} \tfrac{B_s}{\sqrt{s}}) + P(\tfrac{B_t}{\sqrt{t}} < -\Phi^{-1}(x) \le \tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}}) \\
&= P(\tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}} < -\Phi^{-1}(x) \le \sup_{t-\varepsilon\le s<t} \tfrac{B_s}{\sqrt{s}}) + P(\tfrac{B_{t-\varepsilon}}{\sqrt{t-\varepsilon}} \le \Phi^{-1}(x) < \tfrac{B_t}{\sqrt{t}}) \\
&\le C\varepsilon^{1/2}\phi(-\Phi(x) - m_0)^{\frac{t}{t+\varepsilon}} + C\varepsilon(x \ln \tfrac{1}{x}) \text{ by Lemmas 3.4.11 and 3.4.12 .} \qquad \square
\end{aligned}$$

Proof of Theorem 3.4.1. Let $0 < \varepsilon < 1/2$ and $1 \leq t \leq 2$. Choose $\theta > 4$ big enough such that $\frac{t}{t+\varepsilon^\theta} > 2\alpha$ uniformly in t and ε. Let $\rho(s,t) = |s-t|^{1/\theta}$. Then $\rho(s,t)$ is a continuous Gaussian metric on $[0,1]$ (indeed it is the L_2 distance of the fractional Brownian motion with Hurst index $1/\theta$). By Lemmas 3.4.4, 3.4.5, and 3.4.6, it follows that

$$\phi(-\Phi^{-1}(x) - m_0)^{\frac{t}{t+\varepsilon^\theta}} \leq [CxL_C(x)]^{\frac{t}{t+\varepsilon^\theta}} \leq Cx^{2\alpha}/L(x) = \frac{C}{w(x)^2}.$$

Hence Propositions [3.4.13] and [3.4.14] verify the WL-condition in Theorem 3.3.3 and Lemma 3.4.10 verifies the envelope function condition therein. Hence by Proposition 3.3.5 and noting the distribution functions F_t of B_t are strictly increasing , we conclude the proof. □

4 WEAK CONVERGENCE OF EMPIRICAL PROCESSES BASED ON NON-IID STOCHASTIC PROCESSES

4.1 Introduction

Since there is no common law P for non iid samples, the n-th average of the laws $\{P_1, P_2, \cdots, P_n\}$ would take this role. Thus to have a CLT for non iid samples, it would be advisable to have assumptions on this average. The following also illustrates, in real rv case, how empirical processes on general df's (for X_{ni}'s here) can be reduced (in some sense and in terms of probabilistic behavior) to empirical processes for continuous df's or to the uniform empirical process by changing the indexing argument.

Let $X_{n1}, X_{n2}, \cdots X_{nn}$ be independent rv's with arbitrary df's $F_{n1}, F_{n2}, \cdots F_{nn}$ and let

$$\mathbb{F}_n(x) := n^{-1} \sum_{i=1}^{n} 1_{(-\infty,x]}(X_{ni}), \tag{4.1}$$

$$\bar{F}_n(x) := n^{-1} \sum_{i=1}^{n} F_{ni}(x). \tag{4.2}$$

In case that all $F_{n1}, F_{n2}, \cdots F_{nn}$ are continuous, if

$$\alpha_{ni} := \bar{F}_n(X_{ni}), \quad G_{ni} := F_{ni} \circ \bar{F}_{ni}^{-1}, \text{ and } \xi_{ni} := G_{ni}(\alpha_{ni}),$$

then the following hold (see Shorack and Wellner (1986), p. 99):

- α_{ni} has absolutely continuous df G_{ni} on $[0, 1]$
- ξ_{ni} are uniform $(0, 1)$

- $n^{-1} \sum_{i=1}^{n} G_{ni} = t, \quad$ for $0 \le t \le 1$,

and it is almost surely true that

- $[\alpha_{ni} \le t] = [\xi_{ni} \le G_{ni}] \quad$ for all $0 \le\le t \le 1$
- $[X_{ni} \le x] = [\alpha_{ni} \le \bar{F}_n(x)] = [\xi_{ni} \le G_{ni}(\bar{F}_n(x))] \quad$ for all $-\infty < x < \infty$

The empirical df based on α_{ni}'s,

$$\mathbb{G}_n(t) := \frac{1}{n} \sum_{i=1}^{n} 1_{\alpha_{ni} \le t},$$

is called the reduced empirical df; and

$$\mathbb{X}_n(t) := n^{1/2}[\mathbb{G}_n(t) - t] = n^{-1/2} \sum_{i=1}^{n} (1_{\alpha_{ni} \le t} - t) = n^{-1/2} \sum_{i=1}^{n} (1_{\xi_{ni} \le G_{ni}} - G_{ni}) \quad \text{for } 0 \le t \le 1$$

the reduced empirical process.

We have the fundamental equation (see Shorack and Wellner (1986), Section 3.2)

$$\sqrt{n}[\mathbb{F}_n - \bar{F}_n] = \mathbb{X}_n(\bar{F}_n) \text{ on } (-\infty, \infty) \text{ for all } n, \text{ holds } a.s. \tag{4.3}$$

The point of this equation is that $\mathbb{X}_n$ is the empirical process based on *continuous* df's (α_{ni}'s are absolutely continuous for continuous df's); when X_{ni}'s are not continuous rv's, the fundamental equation remains true for some changed but still continuous α_{ni}'s. See more detail in Shorack and Wellner (1986), p. 102-103. There says " the empirical process $n^{1/2}[\mathbb{F}_n - \bar{F}_n]$ only 'looks in on' the reduced empirical process $\mathbb{X}_n$ of the associated array of continuous rv's at points in the range of $\bar{F}_n$ ".

For a FCLT to hold for non-iid samples, the covariance, say ρ_n, of the nth empirical process $\{\nu_n(f) : f \in \mathcal{F}\}$ (which is $n^{-1}\sum_{i\le n}(P_i(fg) - P_i f P_i g)$) has to converge point-wise on $\mathcal{F}\times\mathcal{F}$. Consequently the L_2 metric of the empirical process converges point-wise on $\mathcal{F}\times\mathcal{F}$ as we can write the L_2 metric in terms of the covariance function. Actually we will prove that the convergence is uniform if the $L_2(P)$ norm of the functions are uniformly bounded. Before the proof, we need the following lemma.

LEMMA 4.1.1 (Kolmogorov-Prokhorov exponential inequality; cf Stout (1974), Theorem 5.2.2). *Let $\{X_1, X_2, \cdots\}$ be a sequence of independent mean zero random variables. Let $S_n = \sum_{i\le n} X_i$ and $s_n^2 = \sum_{i\le n} \mathrm{E}X_i^2$. Assume for a constant $c > 0$,*

$$|X_i| \le cs_n \ a.s$$

for each $1 \le i \le n$ and $n \ge 1$. Suppose $\varepsilon > 0$ and $\gamma > 0$. Then there exist constants $\varepsilon(\gamma) > 0$ and $\pi(\gamma) > 0$ such that if $\varepsilon(\gamma) \le \varepsilon \le \pi(\gamma)/c$, then

$$\mathrm{P}(S_n/s_n > \varepsilon) \ge \exp(-(\varepsilon^2/2)(1+\gamma)).$$

The following theorem is an extension of Theorem 3.1 in Alexander (1987b) for empirical processes for non-iid samples indexed by sets to a index set consisting of uniformly bounded functions (actually we only need the functions are uniformly L_2 bounded).

THEOREM 4.1.2. *Let $\{X_1, X_2, \cdots\}$ be independent random elements in $(S, \mathcal{S})$. Let $\mathcal{F}$ be a class of functions each of which uniformly bounded by 1. Let $\{\nu_n(f) : f \in \mathcal{F}\}$ be the empirical process indexed by $\mathcal{F}$. Assume it converges weakly to a mean zero Gaussian process $G(\cdot)$. Let $\rho_n(f,g) = \mathrm{E}\nu_n(f)\nu_n(g)$ and $\rho(f,g) = \mathrm{E}G(f)G(g)$.*

Then

$$\rho_n \text{ converges to } \rho \text{ uniformly on } \mathcal{F} \times \mathcal{F}.$$

PROOF. Let d_{ρ_n} and d_ρ be the L_2 distances of the process ν_n and G; i.e. $d_{\rho_n}(f,g)^2 = \rho_n(f,f) - 2\rho_n(f,g) + \rho_n(g,g)$ and $d_\rho(f,g)^2 = \rho(f,f) - 2\rho(f,g) + \rho(g,g)$.

First we prove for all $f, g, h, l \in \mathcal{F}$

$$|\rho(f,g) - \rho(h,l)| \leq d_\rho(f,h) + d_\rho(g,l). \tag{4.4}$$

$$\begin{aligned} |\rho(f,g) - \rho(h,l)| &= |\mathrm{E}G(f)G(g) - \mathrm{E}G(h)G(l)| \\ &= |\mathrm{E}G(f)G(g) - \mathrm{E}G(h)G(g) + \mathrm{E}G(h)G(g) - \mathrm{E}G(h)G(l)| \\ &\leq |\mathrm{E}G(f)G(g) - \mathrm{E}G(h)G(g)| + \mathrm{E}|G(h)G(g) - \mathrm{E}G(h)G(l)| \\ &= (\mathrm{E}G(g)^2)^{1/2}(\mathrm{E}(G(f) - G(h))^2)^{1/2} \\ &\quad + (\mathrm{E}G(h)^2)^{1/2}(\mathrm{E}(G(g) - G(l))^2)^{1/2} \\ &\leq (\mathrm{E}(G(f) - G(h))^2)^{1/2} + (\mathrm{E}(G(g) - G(l))^2)^{1/2} \\ &= d_\rho(f,h) + d_\rho(g,l), \end{aligned} \tag{4.5}$$

where we used that $(\mathrm{E}G(f)^2)^{1/2} \leq 1$ for any $f \in \mathcal{F}$ as each f is uniformly bounded by 1. Thus we proved (4.4); it remains true if ρ is replaced by ρ_n. By definition, $(\mathcal{F}, \rho)$ is totally bounded; hence for any $\alpha > 0$, let $\mathcal{F}^\alpha$ be a α dense set in $\mathcal{F}$ with $|\mathcal{F}^\alpha| < \infty$. Write $f^\alpha \in \mathcal{F}^\alpha$ be within α for $f \in \mathcal{F}$. Since $\{\nu_n(f) : f \in \mathcal{F}\}$ converges weakly to $\{G(f) : f \in \mathcal{F}\}$, by continuous mapping theorem, $\{\nu_n(f) : f \in \mathcal{F}^\alpha\}$ converges weakly to $\{G(f) : f \in \mathcal{F}^\alpha\}$. Hence ρ_n converges uniformly to ρ on $\mathcal{F}^\alpha \times \mathcal{F}^\alpha$.

Suppose on the contrary that ρ_n does not converges uniformly to ρ on $\mathcal{F} \times \mathcal{F}$. Then there exist $\eta > 0$, subsequences $(n_k)_{k\geq 1}$ and $(f_{n_k})_{k\geq 1} \subset \mathcal{F}$ and $(g_{n_k})_{k\geq 1} \subset \mathcal{F}$

such that

$$|\rho_{n_k}(f_{n_k}, g_{n_k}) - \rho(f_{n_k}, g_{n_k})| \geq \eta.$$

By passing to the subsequence $\{n_k : k = 1, 2, \cdots\}$, we write the full sequence $\{n : n = 1, 2, \cdots\}$. Fix $\alpha < \eta/4$. Using (4.4), and the uniform convergence of ρ_n on $\mathcal{F}^\alpha \times \mathcal{F}^\alpha$, we have for all n big enough

$$\begin{aligned}
\eta &\leq |\rho_n(f_n, g_n) - \rho(f_n, g_n)| \\
&\leq |\rho_n(f_n, g_n) - \rho_n(f_n^\alpha, g_n^\alpha)| + |\rho_n(f_n^\alpha, g_n^\alpha) - \rho(f_n, g_n)| \\
&\quad + |\rho(f_n^\alpha, g_n^\alpha) - \rho(f_n^\alpha, g_n^\alpha)| \\
&\leq d_{\rho_n}(f_n, f_n^\alpha) + d_{\rho_n}(g_n, g_n^\alpha) + \eta/4 \\
&\quad + d_\rho(f_n, f_n^\alpha) + d_\rho(g_n, g_n^\alpha) \\
&\leq d_{\rho_n}(f_n, f_n^\alpha) + d_{\rho_n}(g_n, g_n^\alpha) + 3\eta/4.
\end{aligned}$$

We thereby obtain sequence (f_n), (g_n) for which $d_\rho(f_n, g_n) < \alpha$ but $d_{\rho_n}(f_n, g_n) > \eta/8$. By applying Lemma 4.1.1 with $S_n = \nu_n(f - g)$, there exist $\delta = \delta(\eta) > 0$ and $\varepsilon = \varepsilon(\eta) > 0$ such that $d_{\rho_n}(f, g) > \eta/8$ implies $\mathrm{P}(|\nu_n(f) - \nu_n(g)| > \varepsilon) \geq \delta$. Therefore

$$\limsup_n \mathrm{P}^*(\sup\{|\nu_n(f) - \nu_n(g)| : f, g \in \mathcal{F}, d_\rho(f, g) < \alpha\} > \varepsilon) \geq \delta.$$

This contradicts the asymptotic equicontinuity condition in Theorem 2.2.2. □

4.2 Weak convergence

Following is the main result of this section.

THEOREM 4.2.1. *Let $\{Y_j(t) : t \in E\}$ for $j \in \mathbb{N}$ be a sequence of indepen-*

dent processes and let $F_j(t,y)$ *be the distribution functions of* $Y_j(t)$*, i.e.* $F_j(t,y) = \mathrm{P}(Y_j(t) \le y)$*. Further let* $\nu_{n,t}(y) = n^{-1}\sum_{j=1}^{n} F_j(t,y)$*. Assume the following conditions hold.*

(i) *there are continuous distribution functions* $\nu_t(\cdot)$ *for* $t \in E$ *such that* $|\nu_{n,s}(x) - \nu_{n,t}(y)| \le |\nu_s(x) - \nu_t(y)|$ *for all* $x, y \in \mathbb{R}$ *and all* $n \in \mathbb{N}$*,*

(ii) *there is a continuous Gaussian distance* $\rho(s,t)$ *on* E*, constants* L_j *for* $j \in \mathbb{N}$ *and* L *such that for all* $\epsilon > 0$

$$\sup_{t\in E} \mathrm{P}^*\Big(\sup_{s:\rho(s,t)\le\epsilon} |\tilde{F}_j(t, Y_j(t)) - \tilde{F}_j(t, Y_j(s))| > \epsilon^2\Big) \le L_j\epsilon^2, \quad j \in \mathbb{N} \tag{4.6}$$

and

$$\sup_n n^{-1}\sum_{j=1}^{n} L_j \le L \tag{4.7}$$

(iii) *the covariances of the empirical process*

$$\beta_n(t,y) := n^{-1/2}\sum_{j=1}^{n}(1_{Y_j(t)\le y} - F_j(t,y)),\ t \in E,\ y \in \mathbb{R}$$

converge.

Then the empirical process β_n *converges weakly to a Gaussian process.*

When the input processes $\{Y(t), Y_1(t), Y_2(t), \cdots\}$ on E are iid with continuous marginal distributions, we have

COROLLARY 4.2.2. *Given iid processes* $\{Y_1(t), Y_2(t), \cdots,\}$ *on* E*, assume* $F(t,\cdot)$*, the distribution function of* $Y(t)$ *are continuous for each* $t \in E$*. Further assume* *the process $Y(t)$ satisfies the L-condition. Then the empirical process based on $\{1_{Y(t)\le y} - F(t,y) : t \in E,\, y \in \mathbb{R}\}$ converges weakly.*

Proof. In Theorem 4.2.1, $\nu_n(t,y)$ becomes $F(t,y)$, thus we can take $\nu(t,y) := F(t,y)$ and the other conditions in the theorem are satisfied trivially. $\square$

When specializing the time set E to have only one point, we obtain

Corollary 4.2.3. *Let $X_1, X_2, \cdots$ be a sequence of independent real valued random variables. Let $\nu_n(x) = n^{-1}\sum_{i=1}^n \mathrm{P}(X_i \le x)$. Let $Z_n(x) = n^{-1/2}\sum_{i=1}^n (1_{X_i \le x} - \mathrm{P}(X_i \le x))$ and $K_n(x,y) = \mathrm{E}Z_n(x)Z_n(y)$. If*

(i) there is a continuous distribution function $\nu(x)$ for which

$$|\nu_n(x) - \nu_n(y)| \le |\nu(x) - \nu(y)|$$

and

(ii) $K_N(x,y) \to$ some $K(x,y)$ as $n \to \infty$ for all $x, y \in \mathbb{R}$,

then $Z_n(x)$ converges weakly to a Gaussian process with the limiting covariance function $K(x,y)$.

Remarks 4.2.4. (1) By the quantile transformation, this corollary is Shorack (1979), Theorem 1.1.

(2) It has an extension to empirical processes indexed by VC-class (Alexander (1987a), Corollary 5.2).

Before the proof of Theorem 4.2.1, we collect some lemmas.

LEMMA 4.2.5 (Kuelbs et al. (2013), Lemma 1). *Let L_j be as in the L-condition for processes $\{Y_j(t) : t \in E\}$ for $j \in \mathbb{N}$. Then*

$$\mathrm{P}(Y_j(s) \leq x < Y_j(t)) \leq (L_j + 1)\rho(s,t)^2, \tag{4.8}$$

and

$$\sup_x |F_j(t,x) - F_j(s,x)| \leq 2(L_j + 1)\rho(s,t)^2. \tag{4.9}$$

LEMMA 4.2.6. *Let $W(y)$ be a Wiener process. Let $Z(t,y) = W(\nu_t(y))$. Then $Z(t,y)$ is a continuous Gaussian process w.r.t. its L_2 distance*

$$d_Z((s,x),(t,y)) = |\nu_s(x) - \nu_t(y)|^{1/2}.$$

PROOF. Only note that $W(\cdot)$ has continuous sample paths and $(\mathrm{E}(W(u) - W(v))^2)^{1/2} = |u-v|^{1/2}$. □

LEMMA 4.2.7.

$$\sup_{x \in \mathbb{R}} |\nu_{n,s}(x) - \nu_{n,t}(x)| \leq 2(L+1)\rho(s,t)^2.$$

Proof.

$$\begin{aligned}|\nu_{n,s}(x)-\nu_{n,t}(x)| &= |n^{-1}\sum_{j=1}^{n} F_j(s,x)-n^{-1}\sum_{j=1}^{n} F_j(t,x)| \\ &\le n^{-1}\sum_{j=1}^{n}|F_j(s,x)-F_j(t,x)| \\ &\le n^{-1}\sum_{j=1}^{n} 2(L_j+1)\rho(s,t)^2 \text{ by (4.9)} \\ &\le 2(L+1)\rho(s,t)^2. \qquad \square\end{aligned}$$

In the setting of Theorem 4.2.1, let $\tau_j((s,x),(t,y)) = (\mathrm{E}(1_{Y_j(s)\le x}-1_{Y_j(t)\le y})^2)^{1/2}$ and $\lambda_n((s,x),(t,y))^2 := \frac{1}{n}\sum_{j=1}^{n}\tau_j((s,x),(t,y))^2$.

Lemma 4.2.8. *Under assumptions of Theorem 4.2.1, we have*

$$\lambda_n((s,x),(t,y))^2 \le 2|\nu_s(x)-\nu_t(y)| + 4(L+1)\rho(s,t)^2.$$

Consequently, if let τ be the L_2 distance of the limiting Gaussian process of the empirical process β_n in Theorem 4.2.1, then

$$\tau((s,x),(t,y)) \le 2^{1/2} d_Z((s,x),(t,y))| + 2(L+1)^{1/2}\rho(s,t), \tag{4.10}$$

where d_Z is the L_2 distance of the Gaussian process in Lemma 4.2.6.

PROOF.

$$
\begin{aligned}
\lambda_n((s,x),(t,y))^2 &= n^{-1}\sum_{j=1}^{n} \mathrm{E}(1_{Y_j(s)\le x} - 1_{Y_j(t)\le y})^2 \\
&\le n^{-1}\sum_{j=1}^{n} \mathrm{E}\{2(1_{Y_j(s)\le x} - 1_{Y_j(t)\le x})^2 + 2(1_{Y_j(t)\le x} - 1_{Y_j(t)\le y})^2\} \\
&\le 2n^{-1}\sum_{j=1}^{n} \big[(\mathrm{P}(Y_j(s)\le x < Y_j(t)) + \mathrm{P}(Y_j(t)\le x < Y_j(s))) \\
&\quad + |F_j(t,x) - F_j(t,y)|\big] \\
&\le 4n^{-1}\sum_{j=1}^{n}(L_j+1)\rho(s,t)^2 + 2|\nu_{n,s}(x) - \nu_{n,t}(y)| \\
&\le 4(L+1)\rho(s,t)^2 + 2|\nu_s(x) - \nu_t(y)|.
\end{aligned}
$$

Since $\tau((s,x),(t,y))^2 \le \limsup_{n\to\infty} \lambda_n((s,x),(t,y))$, and by Lemma 4.2.6, the last statement in the Lemma follows. $\square$

LEMMA 4.2.9. *Under the assumptions of Theorem 4.2.1, the limiting Gaussian process of the empirical process β_n in Theorem 4.2.1 has a version whose sample paths are bounded and uniformly continuous with respect to its L_2 metric.*

PROOF. Let $\{H_1 : t \in E\}$ be a mean zero Gaussian process with its L_2 distance $\rho(s,t)$ on E and have bounded and uniformly ρ-continuous sample paths. Since (E,ρ) is totally bounded, let E_0 be a countable dense subset in (E,ρ). Let $\{Z(t,y) : (t,y) \in E \times \mathbb{R}\}$, independent from H_1, be the Gaussian process in Lemma 4.2.6. Observe that it is sample bounded and uniformly continuous with respect to its L_2 distance. Define $H(t,y) := 2^{1/2}Z(t,y) + 2(L+1)^{1/2}H_1(s,t)$ for $t \in E$, $y \in \mathbb{R}$. Then,

$$
d_H((s,x),(t,y))^2 = 2|\nu_s(x) - \nu_t(y)| + 4(L+1)\rho(s,t)^2.
$$

Let $\mathbb{Q}$ denote the set of rational numbers. By the continuity of $\nu_t(\cdot)$ for all $t \in E$ from the assumption, $E_0 \times \mathbb{Q}$ is a (countable) dense subset in $(E \times \mathbb{R}, d_H)$. Since $\tau \leq d_H$ by Lemma 4.10, (recall τ is the L_2 distance of the limiting Gaussian process, say G, in theorem 4.2.1), then $E_0 \times \mathbb{Q}$ is countable dense in $(E \times \mathbb{R}, \tau)$. Hence by the comparison theorem 2.3.7, $\{G((s,x)) : (s,x) \in E_0 \times \mathbb{Q}\}$ has bounded and uniformly τ-continuous sample paths; by the extension lemma, Lemma 2.3.8, $\{G((s,x)) : (s,x) \in E \times \mathbb{R}\}$ has a version with bounded and uniformly τ-continuous sample paths.

□

LEMMA 4.2.10. *If $\lambda_n((s,x),(t,y)) \leq \epsilon$ and $\rho(s,t) \leq \epsilon$, then*

$$|\nu_{n,s}(x) - \nu_{n,t}(y)| \leq (6L+8)\epsilon^2.$$

PROOF.

$$\begin{aligned}
|\nu_{n,s}(x) - \nu_{n,t}(y)| &\leq |\nu_{n,s}(x) - \nu_{n,t}(x)| + |\nu_{n,t}(x) - \nu_{n,t}(y)| \\
&\leq 2(L+1)\rho(s,t)^2 + |n^{-1}\sum_{j=1}^{n} F_j(t,x) - n^{-1}\sum_{j=1}^{n} F_j(t,y)| \\
&\leq 2(L+1)\rho(s,t)^2 + n^{-1}\sum_{j=1}^{n} |(F_j(t,x) - F_j(t,y))| \\
&= 2(L+1)\rho(s,t)^2 + n^{-1}\sum_{j=1}^{n} \mathrm{E}(1_{Y_j(t)\leq x} - 1_{Y_j(t)\leq y})^2 \\
&\leq 2(L+1)\rho(s,t)^2 + n^{-1}\sum_{j=1}^{n} \mathrm{E}\{2(1_{Y_j(t)\leq x} - 1_{Y_j(s)\leq x})^2 \\
&\qquad + 2(1_{Y_j(s)\leq x} - 1_{Y_j(t)\leq y})^2\} \\
&\leq (6L+6)\rho(s,t)^2 + 2\lambda_n((s,x),(t,y))^2 \\
&\leq (6L+8)\epsilon^2.
\end{aligned}$$

□

Next we prove a lemma about the local modulus.

Let $e_n((s,x),(t,y)) = \max\{\lambda_n((s,x),(t,y)), \rho(s,t)\}$.

LEMMA 4.2.11. *Under the assumptions in Theorem 4.2.1, we have for all* (t,y) *and* $\epsilon > 0$,

$$\frac{1}{n}\sum_{j=1}^{n} \mathrm{P}^*(\sup_{(s,x):e_n((s,x),(t,y))\le\epsilon} |1_{Y_j(t)\le y} - 1_{Y_j(s)\le x}| > 0) \le (18L+22)\epsilon^2.$$

PROOF. First we observe that

$$\frac{1}{n}\sum_{j=1}^{n} \mathrm{P}^*(\sup_{(s,x):e_n((s,x),(t,y))\le\epsilon} |1_{Y_j(t)\le y} - 1_{Y_j(s)\le x}| > 0)$$
$$= \frac{1}{n}\sum_{j=1}^{n} \mathrm{P}^*(\sup_{(s,x):e_n((s,x),(t,y))\le\epsilon} 1_{Y_j(t)\le y, Y_j(s)>x} + 1_{Y_j(t)>y, Y_j(s)\le x} > 0).$$

By the fact that $x < y$ implies $F_j(t,x) \le \tilde{F}_j(t,y)$, we have

$$\frac{1}{n}\sum_{j=1}^{n} \mathrm{P}^*(\sup_{(s,x):e_n((s,x),(t,y))\le\epsilon} |1_{Y_j(t)\le y} - 1_{Y_j(s)\le x}| > 0)$$
$$\le \frac{1}{n}\sum_{j=1}^{n} \mathrm{P}^*(\sup_{(s,x):e_n((s,x),(t,y))\le\epsilon} 1_{\tilde{F}_j(t,Y_j(t))\le\tilde{F}_j(t,y), \tilde{F}_j(t,Y_j(s))\ge F_j(t,x)} > 0)$$
$$+ \frac{1}{n}\sum_{j=1}^{n} \mathrm{P}^*(\sup_{(s,x):e_n((s,x),(t,y))\le\epsilon} 1_{\tilde{F}_j(t,Y_j(t))\ge F_j(t,y), \tilde{F}_j(t,Y_j(s))\le\tilde{F}_j(t,x)} > 0) =: I + II.$$

Using the L_j condition, i.e. $|\tilde{F}_j(t,Y_j(s)) - \tilde{F}_j(t,Y_j(t))| \le \rho(s,t)^2$ with probability at least $1 - L_j\epsilon^2$, and that $\tilde{F}_j(t,y) \le F_j(t,y)$ we obtain

$$
\begin{aligned}
I &\le \frac{1}{n}\sum_{j=1}^{n}\Big[\mathrm{P}^*(\sup_{(s,x):e_n((s,x),(t,y))\le\epsilon} 1_{F_j(t,x)-\epsilon^2\le\tilde{F}_j(t,Y_j(t))\le F_j(t,y)} > 0) + L_j\epsilon^2\Big] \\
&\le \frac{1}{n}\sum_{j=1}^{n}\Big[\mathrm{P}^*(\inf_{(s,x):e_n((s,x),(t,y))\le\epsilon}(F_j(t,x)-\epsilon^2) \le \tilde{F}_j(t,Y_j(t)) \le F_j(t,y)) + L_j\epsilon^2\Big] \\
&\le \frac{1}{n}\sum_{j=1}^{n}\Big[F_j(t,y) - \inf_{(s,x):e_n((s,x),(t,y))\le\epsilon}(F_j(t,x)-\epsilon^2) + L_j\epsilon^2\Big] \\
&\le \frac{1}{n}\sum_{j=1}^{n}\Big[\sup_{(s,x):e_n((s,x),(t,y))\le\epsilon}(F_j(t,y)-F_j(t,x))\Big] + (L+1)\epsilon^2
\end{aligned}
$$

(4.11)

$$
\le \frac{1}{n}\sum_{j=1}^{n}\Big[\sup_{\{x:\nu_{n,t}(x,y]\le\epsilon^2\}}(F_j(t,y)-F_j(t,x))\Big] + (L+1)\epsilon^2
$$

The second inequality follows from that $\tilde{F}_j(t,Y_j(t))$ is uniform on $[0,1]$. For $j = 1,2,\cdots,n$, let x_{nj} be the limit point in the set over which the supremum is taken in the last line of the last display. The supremum is actually maximum because of right continuity of $F_j(t,\cdot)$. Indeed, for fixed n and $1 \le j \le n$ there are sequences $(s_{nj}^{(m)}, x_{nj}^{(m)})_{m\in\mathbb{N}}$ such that $e_n((s_{nj}^{(m)}, x_{nj}^{(m)}),(t,y)) \le \epsilon$ for $m \in \mathbb{N}$ and that $x_{nj}^{(m)} \downarrow x_{nj}$ as $F_j(t,\cdot)$ is monotone increasing.

Therefore

$$\frac{1}{n}\sum_{j=1}^{n}\sup_{(s,x):e_n((s,x),(t,y))\le\epsilon}(F_j(t,y)-F_j(t,x))$$

$$\le \frac{1}{n}\sum_{j=1}^{n}F_j(t,y)-F_j(t,x_{nj})$$

$$\le \max_{1\le j\le n}(\nu_{n,t}(y)-\nu_{n,t}(x_{nj}))$$

(4.12)

$$\le \max_{1\le j\le n}\{\,[\nu_{n,t}(y)-\nu_{n,s_{nj}^{(m)}}(x_{nj}^{(m)})]+[\nu_{n,s_{nj}^{(m)}}(x_{nj}^{(m)})-\nu_{n,s_{nj}^{(m)}}(x_{nj})]+[\nu_{n,s_{nj}^{(m)}}(x_{nj})-\nu_{n,t}(x_{nj})]\,\}$$

Let

$$A := \nu_{n,t}(y)-\nu_{n,s_{nj}^{(m)}}(x_{nj}^{(m)}),$$

$$B := \nu_{n,s_{nj}^{(m)}}(x_{nj}^{(m)})-\nu_{n,s_{nj}^{(m)}}(x_{nj}),$$

$$C := \nu_{n,s_{nj}^{(m)}}(x_{nj})-\nu_{n,t}(x_{nj}).$$

By Lemma 4.2.10 and note that $e_n((s_{nj}^{(m)},x_{nj}^{(m)}),(t,y))\le\epsilon$, we have

$$A\le(6L+8)\epsilon^2. \tag{4.13}$$

By continuity of $\nu_t(\cdot)$ for each t,

$$\lim_{m\to\infty}B=0. \tag{4.14}$$

By Lemma 4.2.7,

$$C\le(2L+2)\rho(s_{nj}^{(m)},t)^2\le 2(L+1)\epsilon^2. \tag{4.15}$$

Combining (4.11), (4.12),(4.13), (4.14) (4.15) gives

$$I \leq (9L + 19)\epsilon^2. \tag{4.16}$$

Similarly $II \leq (9L + 11)\epsilon^2$. □

Proof of Theorem 4.2.1. Since Lemma 4.2.9 and Lemma 4.2.11 verify the pre-Gaussian and local modulus conditions respectively in Theorem 2.5.2, it suffices to check the finite dimensional convergences. But this follows from the classical CLT by noting that the summands (indicators) are bounded and the covariances converge by assumption. □

5 CONCLUSION

Using the L-condition and WL-condition, we extended some classical empirical process theorems to time dependent ones starting from the paper Kuelbs et al. (2013). Specifically, we have obtained weak convergence theorems for time dependent weighted empirical processes and time dependent empirical processes for independent and not necessarily identically distributed processes.

In the recent work of Mason and Kevei, they proved some strong approximation results for time dependent empirical processes based on fractional Brownian motions. Kuelbs and Zinn Kuelbs and Zinn (2013) proved empirical quantile CLTs involving time dependent data.

In the first part of this thesis, we considered empirical processes for uniform processes on E. A uniform process $\{X(t) : t \in E\}$ can be obtained, e.g., from any stochastic process by the (randomized) distributional transform: $\{X(t) := \tilde{F}(t, Y(t)) : t \in E\}$ where $Y(t)$ is a stochastic process on E and $\tilde{F}(t, \cdot)$ is the randomized distributional transform of the distribution of $Y(t)$. Such a uniform process is called copula process. If further given a weight function, under the WL-condition, we obtained the weak convergence of the time dependent weighted empirical process. The weight function $w(y)$ in Theorem 3.3.3 doesn't dependent on t. In the future work, we could consider $w(t, y)$. In particular, to investigate the continuity of $\{w(t, y)G(t, y) : t \in E, y \in \mathbb{R}\}$ where the Gaussian process $G(t, y)$ is the limiting Gaussian process from the time dependent empirical process of Kuelbs et al. (2013). Also, we can consider the LIL among others for the time dependent empirical process for some class of stochastic processes.

In the second part of this thesis, we considered independent and not necessarily identically distributed stochastic processes (non-iid samples). We assume each process satisfies the L-condition for the same Gaussian metric on E and the averages of these L constants uniformly bounded in n. Further assume some uniform bound on the averages of the distribution functions of these processes. Then we obtained a weak convergence theorem for the time dependent empirical process for non-iid stochastic processes. In the future work, we can consider the case that in each L_j condition, the Gaussian metric is different, say ρ_j, for the process $\{Y_j(t) : t \in E\}$. Then consider what conditions lead to a CLT. For example, if we have some uniform convergence for the nth averages of ρ_j's, the weak convergence might holds.

For some other possible extensions, we can consider the case that the input processes (samples) take values in $\mathbb{R}^d$.

REFERENCES

K. S. Alexander. Central limit theorems for stochastic processes under random entropy conditions. *Probab. Theory Related Fields*, 75(3):351–378, 1987a. ISSN 0178-8051.

K. S. Alexander. The central limit theorem for weighted empirical processes indexed by sets. *J. Multivariate Anal.*, 22(2):313–339, 1987b. ISSN 0047-259X.

N. T. Andersen, E. Giné, M. Ossiander, and J. Zinn. The central limit theorem and the law of iterated logarithm for empirical processes under local conditions. *Probab. Theory Related Fields*, 77(2):271–305, 1988. ISSN 0178-8051.

D. M. Čibisov. Some theorems on the limiting behaviour of an empirical distribution function. *Trudy Mat. Inst. Steklov.*, 71:104–112, 1964a. ISSN 0371-9685.

D. M. Čibisov. Some theorems on the limiting behaviour of an empirical distribution function. *Trudy Mat. Inst. Steklov.*, 71:104–112, 1964b. ISSN 0371-9685.

M. Csörgő, S. Csörgő, L. Horváth, and D. M. Mason. Weighted empirical and quantile processes. *Ann. Probab.*, 14(1):31–85, 1986. ISSN 0091-1798.

M. D. Donsker. Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems. *Ann. Math. Statistics*, 23:277–281, 1952. ISSN 0003-4851.

J. L. Doob. Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics*, 20:393–403, 1949. ISSN 0003-4851.

R. M. Dudley. Weak convergences of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.*, 10:109–126, 1966. ISSN 0019-2082.

R. M. Dudley. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis*, 1:290–330, 1967.

R. M. Dudley. Central limit theorems for empirical measures. *Ann. Probab.*, 6(6): 899–929 (1979), 1978. ISSN 0091-1798.

J. Einmahl, F. Ruymgaart, and J. Wellner. A characterization of weak convergence of weighted multivariate empirical processes. Open access publications from tilburg university, Tilburg University, 1988.

W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

X. Fernique. Des résultats nouveaux sur les processus gaussiens. In *Séminaire de Probabilités, IX (Seconde Partie, Univ. Strasbourg, Strasbourg, années universitaires 1973/1974 et 1974/1975)*, pages 318–335. Lecture Notes in Math., Vol. 465. Springer, Berlin, 1975a.

X. Fernique. Regularité des trajectoires des fonctions aléatoires gaussiennes. In *École d'Été de Probabilités de Saint-Flour, IV-1974*, pages 1–96. Lecture Notes in Math., Vol. 480. Springer, Berlin, 1975b.

P. Gaenssler. On weak convergence of certain processes indexed by pseudo-metric parameter spaces with applications to empirical processes. In *Trans. 11th Prague Conference*, volume 49, page 78, 1992.

P. Gaenssler and W. Schneemeier. On functional limit theorems for a class of stochastic processes indexed by pseudo-metric parameter spaces (with applications to empirical processes). Probability in Banach spaces 6, Proc. 6th Int. Conf., Sandbjerg/Denmark 1986, Prog. Probab. Stat. 20, 82-100 (1990)., 1990.

E. Giné and J. Zinn. Some limit theorems for empirical processes. *Ann. Probab.*, 12 (4):929–998, 1984. ISSN 0091-1798. With discussion.

E. Giné and J. Zinn. Lectures on the central limit theorem for empirical processes. In *Probability and Banach spaces (Zaragoza, 1985)*, volume 1221 of *Lecture Notes in Math.*, pages 50–113. Springer, Berlin, 1986.

J. Hoffmann-Jørgensen. *Stochastic processes on Polish spaces*, volume 39 of *Various Publications Series (Aarhus)*. Aarhus Universitet Matematisk Institut, Aarhus, 1991.

K. Itô and H. P. McKean, Jr. *Diffusion processes and their sample paths.* Die Grundlehren der Mathematischen Wissenschaften, Band 125. Academic Press Inc., Publishers, New York, 1965.

J. Kuelbs and J. Zinn. Empirical quantile clts for time dependent data. *Progress in Probab.*, 66:169–196, 2013.

J. Kuelbs, T. Kurtz, and J. Zinn. A CLT for empirical processes involving time dependent data. *Ann. Probab.*, 41(2):785–816, 2013.

M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001. ISBN 0-8218-2864-9.

N. E. O'Reilly. On the weak convergence of empirical processes in sup-norm metrics. *Ann. Probability*, 2:642–651, 1974.

L. Rüschendorf. On the distributional transform, Sklar's theorem, and the empirical copula process. *J. Statist. Plann. Inference*, 139(11):3921–3927, 2009. ISSN 0378-3758.

E. Seneta. *Regularly varying functions*. Lecture Notes in Mathematics, Vol. 508. Springer-Verlag, Berlin, 1976.

G. R. Shorack. Weak convergence of empirical and quantile processes in sup-norm metrics via KMT-constructions. *Stochastic Process. Appl.*, 9(1):95–98, 1979. ISSN 0304-4149.

G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. ISBN 0-471-86725-X.

W. F. Stout. *Almost sure convergence*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1974. Probability and Mathematical Statistics, Vol. 24.

M. Talagrand. *The generic chaining*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005. ISBN 3-540-24518-9. Upper and lower bounds of stochastic processes.

A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York, 1996. ISBN 0-387-94640-3. With applications to statistics.

M. J. Wichura. *On the weak convergence of non-borel probabilities on a metric space.* ProQuest LLC, Ann Arbor, MI, 1968. Thesis (Ph.D.)–Columbia University.