RESEARCH TOPICS *in* EMPIRICAL PROCESSES

by

Christopher Ball

Supervisor : Estate Khmaladze

A thesis
submitted to Victoria University of Wellington
in fulfilment of the requirements for the degree of
Master of Science
in Stochastic Processes in Finance and Insurance.

Victoria University of Wellington
2013

Abstract

The first chapter consists of an overview of the theory of empirical processes, covering an introduction to empirical processes in $\mathbb{R}$, uniform empirical processes and function parametric empirical processes in SECTION 1.1. SECTION 1.2 contains an overview of the theory related to the law of the iterated logarithm for Brownian motion and the modulus of continuity for Brownian motion. SECTION 1.3 contains the theory of the limiting processes for the empirical process, most importantly Brownian motion, Brownian bridge and the connections and relationships between them, with distributions of selected statistics of Brownian motion and Brownian bridge derived from reflection principles. SECTION 1.4 contains an overview of the theory required to prove central limit results for the empirical processes, covering the theory of the space C and Donsker's theorem.

The second chapter covers research topics, starting with Fourier analysis of mixture distributions and associated theory in SECTION 2.1. SECTION 2.2 covers findings in a research problem about non-linear autoregressive processes. SECTION 2.3 introduces a martingale approach to testing a regression model. SECTION 2.4 links the theory of ranks and sequential ranks to the theory of empirical processes.

CONTENTS

1	GENERAL THEORY OF EMPIRICAL PROCESSES	1
1.1	The empirical distribution function	1
1.1.1	Glivenko-Cantelli theorem	1
1.1.2	Uniform empirical processes	8
1.1.3	Empirical processes indexed by functions	10
1.1.4	Multi-dimensional empirical process results	13
1.2	Law of the iterated logarithm and modulus of continuity	20
1.2.1	Law of the iterated logarithm	21
1.2.2	Levy modulus of continuity	30
1.3	Some limiting processes	33
1.3.1	Gaussian processes	34
1.3.2	Reflection principles	49
1.4	Central limit theorem for empirical processes	59
1.4.1	Theory of weak convergence	59
1.4.2	Convergence in distribution	62
1.4.3	Prohorov's theorem	65

1.4.4	The space C	68
1.4.5	Construction of the Wiener measure	73
1.4.6	Central limit theorem for empirical processes	77
2	COLLECTION OF RESEARCH PROBLEMS	85
2.1	Fourier Analysis of Mixtures	85
2.1.1	Characteristic Functions	85
2.1.2	Mixtures	91
2.2	Research Problem	99
2.3	New approach to testing regression	103
2.4	Ranks	110
2.4.1	Introduction to ranks	110
2.4.2	Applications	114
2.4.3	Sequential ranks	119
	BIBLIOGRAPHY	127

LIST OF FIGURES

FIGURE 1.1	LEMMA 1.2.4	23
FIGURE 1.2	Simulation of Brownian motion path	36
FIGURE 1.3	Simulation of Brownian Bridge path	36
FIGURE 1.4	THEOREM 1.3.29, FIGURE 1	52
FIGURE 1.5	THEOREM 1.3.29, FIGURE 2	53
FIGURE 1.6	THEOREM 1.3.30, FIGURE 1	54
FIGURE 1.7	THEOREM 1.3.30, FIGURE 2	54
FIGURE 1.8	THEOREM 1.3.32	57
FIGURE 2.1	Mixture of normal densities with $\mu_1 = -1.5, \mu_2 = 0, \mu_3 = 1.5$ and $\sigma_i^2 = 1$.	94
FIGURE 2.2	Mixture of normal distributions with $\mu_1 = -1.5, \mu_2 = 0, \mu_3 = 1.5$ and $\sigma_i^2 = 1$.	95
FIGURE 2.3	Mixture of normal densities with $\mu_1 = -2, \mu_2 = -1, \mu_3 = 1, \mu_4 = 2$ and $\sigma_i^2 = 1$.	96
FIGURE 2.4	Mixture of normal distributions with $\mu_1 = -2, \mu_2 = -1, \mu_3 = 1, \mu_4 = 2$ and $\sigma_i^2 = 1$.	97

FIGURE 2.5	Mixture of Cauchy densities with scale parameter = 1 for all components and shift parameters $\mu =$ $(-1, 0, 1)$	98
FIGURE 2.6	SECTION 2.2, FIGURE 1	100
FIGURE 2.7	SECTION 2.2, FIGURE 2	101
FIGURE 2.8	SECTION 2.2, FIGURE 3	102

GENERAL THEORY OF EMPIRICAL PROCESSES

1.1 THE EMPIRICAL DISTRIBUTION FUNCTION

The first section outlines properties of the empirical distribution function, which is of fundamental importance to the study of empirical processes. We outline some of the theory relating to the empirical distribution function, such as the Glivenko-Cantelli theorem, and some extensions to the uniform empirical process and empirical processes indexed by function. The first section concludes with some inequalities related to the uniform empirical process.

1.1.1 GLIVENKO-CANTELLI THEOREM

DEFINITION 1.1.1: The empirical distribution function $F_n(t)$ for a sample of size n is defined as

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq t\}} = \frac{1}{n} \# \{1 \leq i \leq n : X_i \leq t\}$$

It can be seen that $F_n(t)$ depends on both t and $\omega (\in \Omega)$. For fixed ω , $F_n(t)$ is a distribution function with a discrete measure assigning mass $\frac{1}{n}$ to each $X_i(\omega)$. For fixed $t \in \mathbb{R}$, $F_n(t)$ is a random variable taking values in $[0, 1]$.

LEMMA 1.1.2: The random variable $nF_n(t)$ has a binomial distribution with parameters $(n, F(t))$

PROOF: Consider an individual X_i .

$$\mathbb{P} \left\{ \mathbb{I}_{\{X_i \leq t\}} = 1 \right\} = \mathbb{P} \{X_i \leq t\} = F(t)$$

Therefore $nF_n(t)$ is the sum of n independent and identically distributed Bernoulli random variables $\mathbb{I}_{\{X_i \leq t\}}$, therefore it is binomial with parameters $(n, F(t))$. $\square$

COROLLARY 1.1.3:

$$E[F_n(t)] = F(t) \quad \text{and} \quad \text{Var}(F_n(t)) = \frac{1}{n} F(t) (1 - F(t))$$

PROOF: This follows directly from the empirical distribution function having a binomial distribution. $\square$

This implies that $F_n(t)$ is an unbiased estimator for $F(t)$. From the strong law of large numbers it follows

$$F_n(t) \rightarrow F(t) \quad \text{almost surely } (n \rightarrow \infty).$$

The next question is if this convergence can be made uniform in t , which is answered in the next theorem.

THEOREM 1.1.4: (Glivenko-Cantelli) For any distribution function $F(t)$

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \rightarrow 0 \quad \text{almost surely as } n \rightarrow \infty$$

We prove the above theorem in 2 different ways. The first proof is based on the strong law of large numbers; the second proof is a modern proof which can be extended in various ways. The original papers are [GLIVENKO \(1933\)](#) and [CANTELLI \(1933\)](#).

PROOF: We need to show that $\forall m \in \mathbb{N}$

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \leq \frac{1}{m+1} \quad \text{almost surely}$$

Choose $-\infty = t_{0,m} < t_{1,m} \leq \dots \leq t_{m,m} < t_{m+1,m} = \infty$ so that

$$F(t_{k,m-}) \leq \frac{k}{m+1} \leq F(t_{k,m}) \quad \forall k \in \{1, \dots, m\}$$

As

$$|F_n(t) - F(t)| = \max \{F_n(t) - F(t), F(t) - F_n(t)\}$$

for any $t \in [t_{k-1,m}, t_{k,m})$ and any $k \in \{1, \dots, m+1\}$ we have

$$\begin{aligned} F_n(t) - F(t) &\leq F_n(t_{k,m-}) - F(t_{k-1,m}) \\ &= F_n(t_{k,m-}) - F(t_{k,m-}) + F(t_{k,m-}) - F(t_{k-1,m}) \\ &\leq |F_n(t_{k,m-}) - F(t_{k,m-})| + \frac{1}{m+1} \end{aligned}$$

Similarly

$$F(t) - F_n(t) \leq |F_n(t_{k-1,m-}) - F(t_{k-1,m-})| + \frac{1}{m+1}$$

Because $F_n(-\infty) - F(-\infty) = F_n(\infty) - F(\infty) = 0$

$$\begin{aligned} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| &\leq \max_{1 \leq k \leq m} |F_n(t_{k,m-}) - F(t_{k,m-})| \\ &\quad + \max_{1 \leq k \leq m} |F_n(t_{k,m}) - F(t_{k,m})| + \frac{1}{m+1} \end{aligned}$$

Using the strong law of large numbers, as $n \rightarrow \infty$,

$$F_n(t_{k,m-}) \rightarrow F(t_{k,m-}) \text{ and } F_n(t_{k,m}) \rightarrow F(t_{k,m}) \quad \text{almost surely}$$

$$\Rightarrow \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \leq \frac{1}{m+1}$$

Since m was arbitrary, letting $m \rightarrow \infty$

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \rightarrow 0 \quad \text{almost surely}$$

as $n \rightarrow \infty$. □

For the second proof, we first need the Hoeffding inequality for Rademacher variables. Hoeffding's inequality holds for more general random variables, see [HOEFFDING \(1963\)](#). The next lemma is from [KOSOROK \(2008\)](#).

LEMMA 1.1.5: Let X_i be Rademacher variables for $i = 1, 2, \dots, n$, that is, $\mathbb{P}\{X = 1\} = \mathbb{P}\{X = -1\} = \frac{1}{2}$. Then for fixed $a = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$

$$\mathbb{P}\left\{\left|\sum_{i=1}^n a_i X_i\right| > \lambda\right\} \leq 2 \exp\left(-\frac{\lambda^2}{2\|a\|^2}\right)$$

where $\|a\|$ is the Euclidean norm.

PROOF: For all $\lambda \in \mathbb{R}$

$$\begin{aligned} E[\exp(\lambda X_i)] &= \frac{\exp(\lambda) + \exp(-\lambda)}{2} \\ &= \sum_{i=0}^{\infty} \left(\frac{\lambda^{2i}}{(2i)!}\right) \leq \sum_{i=0}^{\infty} \left(\frac{\lambda^{2i}}{2^i(i)!}\right) = \exp\left(\frac{\lambda^2}{2}\right) \end{aligned}$$

Therefore

$$\mathbb{P}\left\{\sum_{i=1}^n a_i X_i > \lambda\right\} = \mathbb{P}\left\{\exp\left(s \sum_{i=1}^n a_i X_i\right) > \exp(\lambda s)\right\}$$

By Markov's inequality

$$\begin{aligned} &\leq \exp(-\lambda s) E\left[\exp\left(s \sum_{i=1}^n a_i X_i\right)\right] \\ &\leq \exp(-\lambda s) \exp\left(\frac{s^2}{2}\|a\|^2\right) \end{aligned}$$

Choosing $s = \frac{\lambda}{\|a\|^2}$ gives

$$\mathbb{P}\left\{\sum_{i=1}^n a_i X_i > \lambda\right\} \leq \exp\left(-\frac{\lambda^2}{2\|a\|^2}\right)$$

$$\text{As } \mathbb{P} \left\{ \sum_{i=1}^n a_i X_i > \lambda \right\} = \mathbb{P} \left\{ - \sum_{i=1}^n a_i X_i > \lambda \right\}$$

$$\mathbb{P} \left\{ \left| \sum_{i=1}^n a_i X_i \right| > \lambda \right\} \leq 2 \exp \left(- \frac{\lambda^2}{2 \|a\|^2} \right)$$

□

THEOREM 1.1.6: (Glivenko-Cantelli). For any distribution function $F(t)$

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \rightarrow 0 \quad \text{almost surely as } n \rightarrow \infty$$

PROOF:(Second proof)

Let $X_1, \dots, X_n$ and $X'_1, \dots, X'_n$ be 2 independent samples from distribution F . Then for any $\epsilon > 0$ and $n > \frac{2}{\epsilon^2}$, using Chebyshev's inequality

$$\begin{aligned} \mathbb{P} \left\{ |F'_n(t) - F(t)| \leq \frac{\epsilon}{2} \right\} &\geq 1 - \frac{E[|F'_n(t) - F(t)|^2]}{\epsilon^2} \\ &= 1 - \frac{\text{Var}(|F'_n(t) - F(t)|)}{\epsilon^2} \\ &= 1 - \frac{\frac{1}{n} F(t) (1 - F(t))}{\epsilon^2} \\ &\geq 1 - \frac{1}{4n\epsilon^2} \\ \Rightarrow \mathbb{P} \left\{ |F'_n(t) - F(t)| \leq \frac{\epsilon}{2} \right\} &\geq 1 - \frac{1}{2} = \frac{1}{2} \end{aligned} \tag{1.1}$$

where F_n and F'_n are the respective empirical distribution functions. Now select a random variable τ , independent of F'_n , so that $|F_n(\tau) - F(\tau)| > \epsilon$ on the event $\left\{ \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| > \epsilon \right\}$. Note that EQUATION 1.1 remains true

when t is replaced by τ , even when we condition on F_n .

$$\begin{aligned}
& \frac{1}{2} \mathbb{P} \left\{ \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| > \epsilon \right\} \\
& \leq \mathbb{P} \left\{ |F'_n(\tau) - F(\tau)| \leq \frac{\epsilon}{2} \mid \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| > \epsilon \right\} \mathbb{P} \left\{ \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| > \epsilon \right\} \\
& = \mathbb{P} \left\{ |F'_n(\tau) - F(\tau)| \leq \frac{\epsilon}{2}, \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| > \epsilon \right\}
\end{aligned}$$

Using the definition of τ

$$\begin{aligned}
& = \mathbb{P} \left\{ |F'_n(\tau) - F(\tau)| \leq \frac{\epsilon}{2}, \quad |F_n(\tau) - F(\tau)| > \epsilon \right\} \\
& \leq \mathbb{P} \left\{ |F_n(\tau) - F'_n(\tau)| > \frac{\epsilon}{2} \right\} \\
& \leq \mathbb{P} \left\{ \sup_{t \in \mathbb{R}} |F_n(t) - F'_n(t)| > \frac{\epsilon}{2} \right\} \tag{1.2}
\end{aligned}$$

The next step is to define η_i , $i \in \{1, \dots, n\}$ as independent and identically distributed Rademacher variables,

$$\mathbb{P} \{ \eta_i = -1 \} = \mathbb{P} \{ \eta_i = 1 \} = \frac{1}{2}$$

where η_i is independent of X_i and X'_i . Observe that

$$\begin{aligned}
& \mathcal{L} \left(\mathbb{I}_{\{X_i \leq t\}} - \mathbb{I}_{\{X'_i \leq t\}}; i = 1, \dots, n; t \in \mathbb{R} \right) \\
& = \mathcal{L} \left(\eta_i \left(\mathbb{I}_{\{X_i \leq t\}} - \mathbb{I}_{\{X'_i \leq t\}} \right); i = 1, \dots, n; t \in \mathbb{R} \right)
\end{aligned}$$

where $\mathcal{L}()$ is the law. Hence

$$\begin{aligned}
& \mathbb{P} \left\{ \sup_{t \in \mathbb{R}} |F_n(t) - F'_n(t)| > \frac{\epsilon}{2} \right\} \\
& = \mathbb{P} \left\{ \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \left(\eta_i \left(\mathbb{I}_{\{X_i \leq t\}} - \mathbb{I}_{\{X'_i \leq t\}} \right) \right) \right| > \frac{\epsilon}{2} \right\} \\
& \leq \mathbb{P} \left\{ \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \eta_i \mathbb{I}_{\{X_i \leq t\}} \right| > \frac{\epsilon}{4} \right\} + \mathbb{P} \left\{ \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \eta_i \mathbb{I}_{\{X'_i \leq t\}} \right| > \frac{\epsilon}{4} \right\}
\end{aligned}$$

$$= 2\mathbb{P} \left\{ \sup_{t \in \mathbb{R}} |F_n^s(t)| > \frac{\epsilon}{4} \right\} \quad (1.3)$$

where $F_n^s(t) = \frac{1}{n} \sum_{i=1}^n \eta_i \mathbb{I}_{\{X'_i \leq t\}}$, $t \in \mathbb{R}$. Using EQUATION 1.2 and EQUATION 1.3 we get

$$\mathbb{P} \left\{ \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| > \epsilon \right\} \leq 4\mathbb{P} \left\{ \sup_{t \in \mathbb{R}} |F_n^s(t)| > \frac{\epsilon}{4} \right\}. \quad (1.4)$$

We now want to derive bounds for EQUATION 1.4. Conditioning on the values of X_i we see that

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{t \in \mathbb{R}} |F_n^s(t)| > \frac{\epsilon}{4} \mid X_i = x_i; i \in \{1, \dots, n\} \right\} \\ & \leq \sum_{i=1}^n \mathbb{P} \left\{ |F_n^s(t)| > \frac{\epsilon}{4} \mid X_i = x_i; i \in \{1, \dots, n\} \right\} \end{aligned} \quad (1.5)$$

Using LEMMA 1.1.5, with $\lambda = \frac{n\epsilon}{4}$ and $a = (\mathbb{I}_{\{X_1 \leq t\}}, \dots, \mathbb{I}_{\{X_n \leq t\}})$

$$\begin{aligned} \mathbb{P} \left\{ |F_n^s(t)| > \frac{\epsilon}{4} \mid X_i = x_i; i \in \{1, \dots, n\} \right\} & \leq 2 \exp \left(\frac{-\left(\frac{n\epsilon}{4}\right)^2}{2nF_n(t)} \right) \\ & \leq 2 \exp \left(\frac{-n\epsilon^2}{32} \right) \end{aligned}$$

Using EQUATION 1.5,

$$\mathbb{P} \left\{ \sup_{t \in \mathbb{R}} |F_n^s| > \frac{\epsilon}{4} \mid X_i = x_i; i \in \{1, \dots, n\} \right\} \leq 2n \exp \left(\frac{-n\epsilon^2}{32} \right)$$

Noting that this does not depend on X_i and using EQUATION 1.4

$$\mathbb{P} \left\{ \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| > \epsilon \right\} \leq 8n \exp \left(\frac{-n\epsilon^2}{32} \right)$$

$$\Rightarrow \sum_{n=1}^{\infty} \mathbb{P} \left\{ \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| > \epsilon \right\} < \infty \quad \forall \epsilon > 0$$

The second proof follows from an application of the Borel-Cantelli lemma. $\square$

1.1.2 UNIFORM EMPIRICAL PROCESSES

The motivation for this section is to transform X_i onto the unit interval. If X_i has a continuous distribution function F , then we can transform X_i onto the unit interval in such a way that it has the same distribution as any other continuous distribution transformed in the same way. We call the transformed process the uniform empirical process. This allows us to derive limit theorems for the uniform empirical process, and then apply the limit theorem to any empirical process with continuous underlying distribution F .

DEFINITION 1.1.7: Consider a transformation onto the closed unit interval $[0, 1]$ as $\xi = \Gamma(t)$ for $t \in \mathbb{R}$. The corresponding empirical distribution function at stage $n \in \mathbb{N}$ is

$$\Gamma_n(t) = \frac{1}{n} \# \{1 \leq i \leq n : \xi_i \leq t\}; \quad t \in [0, 1]$$

where $\xi_i = \Gamma(X_i)$.

DEFINITION 1.1.8: The reduced empirical process is defined as

$$U_{n,\Gamma}(t) = n^{\frac{1}{2}} (\Gamma_n(t) - \Gamma(t)); \quad t \in [0, 1]$$

This is the empirical process of the random variables $\{\xi_i\}_{i=1}^n = \{\Gamma(X_i)\}_{i=1}^n$

LEMMA 1.1.9: Let $\xi_i = F(X_i)$. If F is continuous then

$$\mathcal{L}(\xi_i) = \mathcal{U}(0, 1)$$

where $\mathcal{U}(0, 1)$ is the law of the uniform distribution on $(0, 1)$.

PROOF: For $0 \leq t \leq 1$:

$$\begin{aligned}\mathbb{P}\{\xi_i \leq t\} &= \mathbb{P}\{F(X_i) \leq t\} \\ &= \mathbb{P}\{X_i \leq F^{-1}(t)\} \\ &= F(F^{-1}(t)) = t = I(t)\end{aligned}$$

since F is continuous, F^{-1} is defined as $F^{-1}(t) = \inf\{x : F(x) \geq t\}$. $\square$

DEFINITION 1.1.10: If $\Gamma = I$ then we denote $U_{n,I} = U_n$. This is called the *uniform empirical process*.

THEOREM 1.1.11: For continuous F we have

$$\mathcal{L}\left(\sup_{x \in \mathbb{R}} n^{\frac{1}{2}} |F_n(x) - F(x)|\right) = \mathcal{L}\left(\sup_{t \in [0,1]} |U_n(t)|\right)$$

PROOF: By LEMMA 1.1.9 we know that $\xi_1, \dots, \xi_n$ is a random sample of size n from $\mathcal{U}(0, 1)$

$$\begin{aligned}\sup_{x \in \mathbb{R}} n^{\frac{1}{2}} |F_n(x) - F(x)| &= \sup_{x \in \mathbb{R}} n^{\frac{1}{2}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq x\}} - F(x) \right| \\ &= \sup_{x \in \mathbb{R}} n^{\frac{1}{2}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{F(X_i) \leq F(x)\}} - F(x) \right| \\ &= \sup_{t \in [0,1]} n^{\frac{1}{2}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\xi_i \leq t\}} - t \right| \\ &= \sup_{t \in [0,1]} |U_n(t)|\end{aligned}$$

$\square$

The proof of the next theorem assumes some results which are proved later in the text.

THEOREM 1.1.12:

$$\mathbb{P} \left\{ \sup_{t \in [0,1]} |U_n(t)| \leq x \right\} \rightarrow 1 + 2 \sum_{k=1}^{\infty} (-1)^k \exp(-2k^2 x^2) = H(x)$$

PROOF: We are going to assume that U_n weakly converges to Brownian bridge, which is the statement of THEOREM 1.4.34. Given weak convergence all that remains is to show that $\sup_{t \in [0,1]} |\mathbb{U}(t)|$ has the distribution specified above, for $\mathbb{U}$ Brownian bridge. This follows directly from the statement of THEOREM 1.3.33. $\square$

1.1.3 EMPIRICAL PROCESSES INDEXED BY FUNCTIONS

The aim of this section is to give an introduction to a generalisation of the empirical process given in the previous section and to also generalise the Glivenko-Cantelli theorem. This subsection follows the beginning of Part 2 of VAN DER VAART & WELLNER (1996).

The empirical process can be generalised by instead of considering not only indicator function on the real line, but considering collections of measurable functions mapping from the sample space into $\mathbb{R}$. So now X_i are not necessarily in $\mathbb{R}$, but are instead in an arbitrary measurable space. We consider functions $f: \mathcal{X} \mapsto \mathbb{R}$ where $\mathcal{X}$ is an arbitrary measurable space.

DEFINITION 1.1.13: The *empirical measure* $\mathbb{P}_n$ of a sample of random elements $X_1, X_2, \dots, X_n$ in a measurable space $(\mathcal{X}, \mathcal{A})$ is the discrete random measure given by

$$\mathbb{P}_n \{C\} = \frac{1}{n} \# \{1 \leq i \leq n: X_i \in C\}$$

It can be described as the random measure that assigns mass $\frac{1}{n}$ at each observation. It can also be given as the linear combination of the Dirac

measures at the observations

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

where

$$\delta_{X_i}(C) = \mathbb{I}_{\{X_i \in C\}}$$

the indicator function that X_i is an element of C .

Now consider a collection of measurable functions $\mathcal{F}$, where $f \in \mathcal{F}$ maps $\mathcal{X} \rightarrow \mathbb{R}$. The empirical measure induces a map from $\mathcal{F}$ to $\mathbb{R}$ given by

$$f \mapsto \int f d\mathbb{P}_n = \mathbb{P}_n f$$

Likewise the abbreviation $Qf = \int f dQ$ is used for a given measurable function f and signed measure Q .

DEFINITION 1.1.14: Let $\mathbb{P}$ be the common distribution of the X_i . The $\mathcal{F}$ -indexed empirical process $\mathbb{G}_n$ is given by

$$f \mapsto \mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - \mathbb{P})f = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(f(X_i) - \int f d\mathbb{P} \right)$$

We want to prove a generalisation of the Glivenko-Cantelli theorem for empirical processes indexed by functions. In the real case the (first) proof of the Glivenko-Cantelli theorem used the order on $\mathbb{R}$ to assert the existence of t_i so $\mathbb{P}\{t_{i-1} \leq X_i \leq t_i\} \geq \frac{1}{m}$ for each $i = 0, 1, \dots, m$ and $m \in \mathbb{N}$. No such ordering exists in general for a class of functions. We need an additional constraint on the size and the structure of the class, which is obtained using the following definitions.

DEFINITION 1.1.15: A *bracket* $[l, u]$, where $l, u \in \mathcal{F}$, is the set of all functions $f \in \mathcal{F} : l \leq f \leq u$.

DEFINITION 1.1.16: An ϵ -bracket is a bracket $[l, u]$ with $\|u - l\|_{\mathcal{F}} < \epsilon$

DEFINITION 1.1.17: The *bracketing number* $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$ is the minimum number of ϵ -brackets needed to cover $\mathcal{F}$.

DEFINITION 1.1.18: A Glivenko-Cantelli class has $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \rightarrow 0$ in outer probability or outer almost surely, where $\|Q\|_{\mathcal{F}} = \sup \{|Qf| : f \in \mathcal{F}\}$.

For the rest of this subsection we are assuming that everything is appropriately measurable. For an overview of some of the measurability issues that can arise, see [VAN DER VAART & WELLNER \(1996\)](#), among others.

THEOREM 1.1.19: Let $\mathcal{F}$ be a class of measurable functions, such that for all $\epsilon > 0$, $N_{[]}(\epsilon, \mathcal{F}, L_1(\mathbb{P})) < \infty$. Then $\mathcal{F}$ is Glivenko-Cantelli.

PROOF: Fix $\epsilon > 0$. As $N_{[]}(\epsilon, \mathcal{F}, L_1(\mathbb{P})) < \infty$ we can choose finitely many ϵ -brackets $[l_i, u_i]$ with $\mathcal{F} \subseteq \bigcup_{i=1}^n [l_i, u_i]$ and $\mathbb{P}u_i - \mathbb{P}l_i < \epsilon$ for all i . For every $f \in \mathcal{F}$ there is a bracket such that

$$\begin{aligned} (\mathbb{P}_n - \mathbb{P})f &\leq \mathbb{P}_n u_i - \mathbb{P}f && (\text{As } u_i \geq f) \\ &= (\mathbb{P}_n - \mathbb{P})u_i + \mathbb{P}(u_i - f) \\ &\leq (\mathbb{P}_n - \mathbb{P})u_i + \epsilon && (\text{As } f \in [l_i, u_i]) \end{aligned}$$

Therefore

$$\sup_{f \in \mathcal{F}} (\mathbb{P}_n - \mathbb{P})f \leq \max_{1 \leq i \leq n} (\mathbb{P}_n - \mathbb{P})u_i + \epsilon$$

Similarly

$$\begin{aligned} (\mathbb{P}_n - \mathbb{P})f &\geq \mathbb{P}_n l_i - \mathbb{P}f && (\text{As } l_i \leq f) \\ &= (\mathbb{P}_n - \mathbb{P})l_i + \mathbb{P}(l_i - f) \\ &\geq (\mathbb{P}_n - \mathbb{P})l_i - \epsilon && (\text{As } f \in [l_i, u_i]) \end{aligned}$$

Therefore

$$\inf_{f \in \mathcal{F}} (\mathbb{P}_n - \mathbb{P})f \geq \min_{1 \leq i \leq n} (\mathbb{P}_n - \mathbb{P})l_i - \epsilon$$

Using the strong law of large numbers gives

$$\limsup_{n \rightarrow \infty} \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \leq \epsilon$$

almost surely for all $\epsilon > 0$. Take a sequence of $\epsilon_m \downarrow 0$ to see that

$$\limsup_{n \rightarrow \infty} \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \rightarrow 0$$

almost surely. Therefore $\mathcal{F}$ is Glivenko-Cantelli. □

1.1.4 MULTI-DIMENSIONAL EMPIRICAL PROCESS RESULTS

This subsection considers inequalities related to the uniform empirical process, which can be used to prove results like the law of iterated logarithms for empirical processes and probability inequalities about the local fluctuation of the uniform empirical process. Most of the results in this subsection are given without proof, though the proofs can be found in [EINMAHL \(1993\)](#), among others.

Assume for this section that vectors take values in $(0, 1)^d$; $d \in \mathbb{N}$. The notation for coordinates of a vector t is $t = \langle t_1, \dots, t_d \rangle = \langle t_j \rangle$ and if $t_j = u$ for all $1 \leq j \leq d$ we simply write $\langle u \rangle$. For $s, t \in \mathbb{R}^d$ define $s \leq t$ as $s_j \leq t_j$ for all $1 \leq j \leq d$ and $s < t$ as $s_j < t_j$ for all $1 \leq j \leq d$.

DEFINITION 1.1.20: Let d be a fixed positive constant and fix $v \in \mathbb{N}^d$. For $t \in \mathbb{N}^d$ with $t \leq v$, let Y_t be a random variable with mean 0 and finite variance σ_t^2 and assume these random variables are independent. For arbitrary $s, u \in \mathbb{N}^d$

with $s \leq u \leq v$ define partial sums by

$$S_u = \sum_{t \leq u} Y_t \quad S_{s,u} = \sum_{s \leq t \leq v} Y_t$$

and write $\sigma^2 = \text{Var}(S_v) = \sum_{t \leq v} \sigma_t^2$

THEOREM 1.1.21: For all $\lambda \in \mathbb{R}$

$$\mathbb{P} \left\{ \max_{u \leq v} S_u \geq \lambda \right\} \leq 2^d \mathbb{P} \left\{ S_v \geq \lambda - d\sqrt{2\sigma^2} \right\}$$

PROOF: Let

$$A = \left\{ \max_{u \leq v} S_u \geq \lambda \right\}$$

and

$$B = \left\{ \max_{u \leq v; u_d = v_d} S_u \geq \lambda - \sqrt{2\sigma^2} \right\}$$

The first step is to prove

$$\mathbb{P} \{A\} \leq 2\mathbb{P} \{B\}$$

Let

$$C = \left\{ k \in \mathbb{N} : k \leq \prod_{j=1}^d v_j \right\}$$

be an index set for the possible vectors $u \leq v$ and let

$$D = \left\{ t \in \mathbb{N}^d : t_j \leq v_j ; \forall 1 \leq j \leq d \right\}$$

be the set of possible vectors $u \leq v$. Furthermore let $g : C \rightarrow D$ be a fixed bijective function with the property that $k < l$ implies $(g(k))_d \leq (g(l))_d$.

Write

$$\widetilde{g(k)} = \langle (g(k))_1, (g(k))_2, \dots, (g(k))_{d-1}, v_d \rangle$$

which can be seen as a projection of $g(k)$, replacing the last coordinate of $g(k)$ with the last coordinate of v . Also consider elementary events

$$E_k = \left\{ S_{g(k)} \geq \lambda \right\}$$

and the events

$$A_k = E_k \cup \left(\bigcup_{i=1}^{k-1} E_i^c \right),$$

which are the events that $S_{g(k)}$ is the only vector from indices 1 to k that is greater than λ . It is seen that A_k are pairwise disjoint and $A = \bigcup_{k \in C} A_k$. We see by Chebyshev's inequality that

$$\begin{aligned} \mathbb{P} \left\{ \widetilde{S_{g(k)}} - S_{g(k)} \geq -\sqrt{2\sigma^2} \right\} &\geq 1 - \frac{E \left[\left(\widetilde{S_{g(k)}} - S_{g(k)} \right)^2 \right]}{2\sigma^2} \\ &\geq 1 - \frac{\text{Var}(\widetilde{S_{g(k)}})}{2\sigma^2} \\ &\geq 1 - \frac{\sigma^2}{2\sigma^2} = \frac{1}{2} \end{aligned}$$

This implies that

$$\mathbb{P} \{A_k\} \leq 2\mathbb{P} \{A_k\} \mathbb{P} \left\{ \widetilde{S_{g(k)}} - S_{g(k)} \geq -\sqrt{2\sigma^2} \right\}$$

Using the independence of A_k and $\left\{ \widetilde{S_{g(k)}} - S_{g(k)} \geq -\sqrt{2\sigma^2} \right\}$

$$\begin{aligned} &= 2\mathbb{P} \left\{ A_k \cap \left\{ \widetilde{S_{g(k)}} - S_{g(k)} \geq -\sqrt{2\sigma^2} \right\} \right\} \\ &\leq 2\mathbb{P} \left\{ A_k \cap \left\{ \widetilde{S_{g(k)}} \geq \lambda - \sqrt{2\sigma^2} \right\} \right\} \\ &= 2\mathbb{P} \{A_k \cap B\} \end{aligned}$$

This implies

$$\begin{aligned} \mathbb{P} \{A\} &= \sum_{k \in C} \mathbb{P} \{A_k\} \\ &\leq 2 \sum_{k \in C} \mathbb{P} \{A_k B\} \\ &= 2\mathbb{P} \{AB\} \\ &\leq 2\mathbb{P} \{B\} \end{aligned}$$

This allows us to prove the result using induction. Letting $d = 1$, the base case is proved by a direct application of the above result. Under the induction hypothesis we can say it holds for

$$\mathbb{P} \left\{ \max_{u \leq v} S_u \geq \lambda \right\} \leq 2^k \mathbb{P} \left\{ S_v \geq \lambda - k\sqrt{2\sigma^2} \right\}$$

and need to prove that it holds for

$$\mathbb{P} \left\{ \max_{u \leq v} S_u \geq \lambda \right\} \leq 2^{k+1} \mathbb{P} \left\{ S_v \geq \lambda - (k+1)\sqrt{2\sigma^2} \right\}$$

where all dimensions increase in size by 1. This leads to

$$\begin{aligned} \mathbb{P} \left\{ \max_{u \leq v} S_u \geq \lambda \right\} &\leq 2 \mathbb{P} \left\{ \max_{u \leq v; u_d = v_d} S_u \geq \lambda - \sqrt{2\sigma^2} \right\} \\ &\leq 2 \left(2^k \mathbb{P} \left\{ \max_{u \leq v; u_d = v_d} S_u \geq \lambda - k\sqrt{2\sigma^2} - \sqrt{2\sigma^2} \right\} \right) \\ &\leq 2^{k+1} \mathbb{P} \left\{ S_v \geq \lambda - (k+1)\sqrt{2\sigma^2} \right\} \end{aligned}$$

□

DEFINITION 1.1.22: Let N_n be a Poisson process on $[0, 1]^d$, with $E[N_n(t)] = n\Gamma(t)$ for all $t \in [0, 1]^d$. Then

$$Z_n(t) = \frac{1}{\sqrt{n}} (N_n(t) - n\Gamma(t)) \quad t \in [0, 1]^d$$

The similarity between the above definition and [DEFINITION 1.1.8](#) should be apparent. This makes sense as conditional on $N_n(\langle 1 \rangle) = n$, the processes $U_{n,\Gamma}$ and Z_n are equal in law.

DEFINITION 1.1.23: The function $\psi: [-1, \infty) \mapsto [0, \infty)$ is defined as

$$\begin{aligned} \psi(\lambda) &= \frac{2}{\lambda^2} \int_0^\lambda \log(1 + \sigma) d\sigma \\ &= \begin{cases} 2\lambda^{-2} [(1 + \lambda) \log(1 + \lambda) - \lambda] & -1 < \lambda < 0 \text{ or } 0 < \lambda \\ 2 & \lambda = -1 \\ 1 & \lambda = 0 \end{cases} \end{aligned}$$

RESULT 1.1.24: For ψ

- ψ is decreasing and continuous.
- $\psi(\lambda) \rightarrow \frac{2\log(\lambda)}{\lambda}$ as $\lambda \rightarrow \infty$
- $\psi(\lambda) \geq \frac{1}{1 + \frac{\lambda}{3}}$ for $\lambda \geq -1$

DEFINITION 1.1.25: For the following two types of index sets are used.

- (a) Quadrants of the form $[0, t_1] \times \dots \times [0, t_d]; t = \langle t_j \rangle \in [0, 1]^d$
- (b) Half-open rectangles of the form $(s_1, t_1] \times \dots \times (s_d, t_d]; s = \langle s_j \rangle < t = \langle t_j \rangle; s, t \in [0, 1]^d$. Call the class of all such rectangles $\mathcal{R}$

RESULT 1.1.26: For any $R \in \mathcal{R}$ with $\mathbb{P}\{\mathcal{R}\} \leq \frac{1}{2}$, all $\lambda \in \mathbb{R}$ and either choice of sign

$$\begin{aligned} \mathbb{P}\left\{\sup_{S \subset R} \pm U_{n,\Gamma}(S) \geq \lambda\right\} &\leq 2\mathbb{P}\left\{\sup_{S \subset R} \pm Z_n(S) \geq \lambda\right\} \\ &\leq 2^{2d+1}\mathbb{P}\left\{\pm Z_n(R) \geq \lambda - 2d\sqrt{2\mathbb{P}\{R\}}\right\} \end{aligned}$$

LEMMA 1.1.27: Let $\tau > 0$. For a Poisson(τ) random variable Z and either choice of sign

$$\mathbb{P}\{\pm(Z - \tau) \geq \lambda\} \leq \exp\left(\frac{-\lambda^2}{2\tau}\psi\left(\frac{\pm\lambda}{\tau}\right)\right); \quad \lambda \geq 0$$

For the minus part the additional assumption of $\lambda \leq \tau$ is required by the definition of ψ .

PROOF: If $\lambda = 0$, then this inequality is trivially true for either choice of sign. The other boundary case is for the negative choice of sign when $\lambda = \tau$, which

is also trivially true. For all other cases, apply the Markov inequality

$$\begin{aligned}\mathbb{P}\{\pm(Z - \tau) \geq \lambda\} &= \mathbb{P}\{\exp(\pm s(Z - \tau)) \geq \exp(\lambda s)\} \\ &\leq \exp(-\lambda s) E[\exp(\pm s(Z - \tau))] \\ &\leq \exp(-s(\lambda \pm \tau)) \exp(\tau(\exp(\pm s) - 1))\end{aligned}$$

The value of s which minimizes this expression is $s = \pm \log\left(1 \pm \frac{\lambda}{\tau}\right)$, therefore

$$\begin{aligned}\mathbb{P}\{\pm(Z - \tau) \geq \lambda\} &\leq \exp\left(-(\tau \pm \lambda) \log\left(1 \pm \frac{\lambda}{\tau}\right) \mp \lambda\right) \\ &= \exp\left(\frac{-\lambda^2}{2\tau} \psi\left(\frac{\pm\lambda}{\tau}\right)\right)\end{aligned}$$

□

THEOREM 1.1.28: Let $n \in \mathbb{N}$ and $R \in \mathcal{R}$ with $0 < \mathbb{P}\{R\} \leq \frac{1}{2}$. Then for any $\epsilon \in (0, 1)$,

(a) For $\lambda \geq 0$;

$$\mathbb{P}\left\{\sup_{S \subset R} U_{n,\Gamma}(S) \geq \lambda\right\} \leq C^+ \exp\left(\frac{-(1-\epsilon)\lambda^2}{2\mathbb{P}\{R\}} \psi\left(\frac{\lambda}{\sqrt{n}\mathbb{P}\{R\}}\right)\right);$$

(b) For $0 \leq \lambda \leq \sqrt{n}\mathbb{P}\{R\}$;

$$\mathbb{P}\left\{\sup_{S \subset R} -U_{n,\Gamma}(S) \geq \lambda\right\} \leq C^- \exp\left(\frac{-(1-\epsilon)\lambda^2}{2\mathbb{P}\{R\}} \psi\left(\frac{-(1-\epsilon)\lambda}{\sqrt{n}\mathbb{P}\{R\}}\right)\right)$$

(c) For $\lambda \geq 0$;

$$\mathbb{P}\left\{\sup_{S \subset R} |U_{n,\Gamma}(S)| \geq \lambda\right\} \leq C \exp\left(\frac{-(1-\epsilon)\lambda^2}{2\mathbb{P}\{R\}} \psi\left(\frac{\lambda}{\sqrt{n}\mathbb{P}\{R\}}\right)\right)$$

where $S \in \mathcal{R}$ and C^+, C^- and $C \in (0, \infty)$ depend only on d and ϵ .

PROOF: Using RESULT 1.1.26 and LEMMA 1.1.27 and $\lambda \geq \frac{4d\sqrt{2\mathbb{P}\{R\}}}{\epsilon}$

$$\begin{aligned}
& \mathbb{P} \left\{ \sup_{S \subset R} \pm U_{n,\Gamma}(S) \geq \lambda \right\} \\
& \leq 2^{2d+1} \mathbb{P} \left\{ \pm Z_n(R) \geq \lambda - 2d\sqrt{2\mathbb{P}\{R\}} \right\} \\
& = 2^{2d+1} \mathbb{P} \left\{ \pm (N_n(t) - n\mathbb{P}\{R\}) \geq \sqrt{n} \left(\lambda - 2d\sqrt{2\mathbb{P}\{R\}} \right) \right\} \\
& \leq 2^{2d+1} \exp \left(\frac{-n \left(\lambda - 2d\sqrt{2\mathbb{P}\{R\}} \right)^2}{2n\mathbb{P}\{R\}} \psi \left(\frac{\pm \sqrt{n} \left(\lambda - 2d\sqrt{2\mathbb{P}\{R\}} \right)}{n\mathbb{P}\{R\}} \right) \right) \\
& \leq 2^{2d+1} \exp \left(\frac{-\lambda^2 \left(1 - \frac{\epsilon}{2} \right)^2}{2\mathbb{P}\{R\}} \psi \left(\frac{\pm \lambda \left(1 - \frac{\epsilon}{2} \right)}{\sqrt{n}\mathbb{P}\{R\}} \right) \right) \\
& \leq 2^{2d+1} \exp \left(\frac{-\lambda^2 (1 - \epsilon)^2}{2\mathbb{P}\{R\}} \psi \left(\frac{\pm \lambda \left(1 - \frac{\epsilon}{2} \right)}{\sqrt{n}\mathbb{P}\{R\}} \right) \right)
\end{aligned}$$

As ψ is decreasing

$$\begin{aligned}
& \psi \left(\frac{\lambda \left(1 - \frac{\epsilon}{2} \right)}{\sqrt{n}\mathbb{P}\{R\}} \right) \geq \psi \left(\frac{\lambda}{\sqrt{n}\mathbb{P}\{R\}} \right) \\
& \exp \left(\frac{-\lambda^2 (1 - \epsilon)^2}{2\mathbb{P}\{R\}} \psi \left(\frac{\lambda \left(1 - \frac{\epsilon}{2} \right)}{\sqrt{n}\mathbb{P}\{R\}} \right) \right) \leq \exp \left(\frac{-\lambda^2 (1 - \epsilon)^2}{2\mathbb{P}\{R\}} \psi \left(\frac{\lambda}{\sqrt{n}\mathbb{P}\{R\}} \right) \right)
\end{aligned}$$

which immediately gives (a). Furthermore

$$\begin{aligned} \psi\left(\frac{-\lambda\left(1-\frac{\epsilon}{2}\right)}{\sqrt{n}\mathbb{P}\{R\}}\right) &\geq \psi\left(\frac{-\lambda(1-\epsilon)}{\sqrt{n}\mathbb{P}\{R\}}\right) \\ \exp\left(\frac{-\lambda^2(1-\epsilon)^2}{2\mathbb{P}\{R\}}\psi\left(\frac{-\lambda\left(1-\frac{\epsilon}{2}\right)}{\sqrt{n}\mathbb{P}\{R\}}\right)\right) &\leq \exp\left(\frac{-\lambda^2(1-\epsilon)^2}{2\mathbb{P}\{R\}}\psi\left(\frac{-\lambda(1-\epsilon)}{\sqrt{n}\mathbb{P}\{R\}}\right)\right) \end{aligned}$$

which gives (b). (c) is a consequence of (a) and (b). $\square$

We finish this subsection by considering an inequality for the modulus of continuity (see DEFINITION 1.2.12) for the uniform empirical process in $\mathbb{R}$.

THEOREM 1.1.29: Let $0 < \epsilon \leq \frac{1}{2}$ and $0 < \alpha < \frac{1}{4}$. For $\lambda \geq 0$,

$$\mathbb{P}\left\{\omega_n(\alpha) \geq \lambda\alpha^{\frac{1}{2}}\right\} \leq \frac{C}{\alpha} \exp\left(\frac{-(1-\epsilon)\lambda^2}{2}\psi\left(\frac{\lambda}{(n\alpha)^{\frac{1}{2}}}\right)\right)$$

where $\omega_n(\alpha)$ is the modulus of continuity of the uniform empirical process.

PROOF: Can be found in EINMAHL (1993). $\square$

1.2 LAW OF THE ITERATED LOGARITHM AND MODULUS OF CONTINUITY

This section outlines the classical versions of the law of the iterated logarithm for Brownian motion, which describes the maximum fluctuation of Brownian motion, and a similar result for the modulus of continuity for Brownian motion, which describes the maximum local fluctuation of Brownian motion. As the limiting process of the uniform empirical process is Brownian bridge, we would expect that these results generalise to the uniform empirical process, and the generalisations are given at the end of the appropriate subsection.

1.2.1 LAW OF THE ITERATED LOGARITHM

This section describes the law of the iterated logarithm, which gives the order of the maximum fluctuation of an empirical process. This lemma is from **FELLER** (1971*a*), among others.

LEMMA 1.2.1: Let $\varphi(x)$ and $\Phi(x)$ be the density and distribution function of the standard normal distribution respectively. Then for $x > 0$

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \varphi(x) \leq 1 - \Phi(x) \leq \frac{1}{x} \varphi(x)$$

PROOF: Using integration by parts

$$\begin{aligned} 1 - \Phi(x) &= \int_x^\infty \frac{1}{t} (t\varphi(t)) dt \\ &= \frac{-1}{t} \varphi(t) \Big|_x^\infty - \int_x^\infty \frac{1}{t^2} \varphi(t) dt \\ &= \frac{1}{x} \varphi(x) - \int_x^\infty \frac{1}{t^2} \varphi(t) dt \end{aligned}$$

For the upper bound, as $x > 0$ it follows $\int_x^\infty \frac{1}{t^2} \varphi(t) dt > 0$, therefore

$$1 - \Phi(x) \leq \frac{1}{x} \varphi(x)$$

For the lower bound, as $x > 0$ and $\frac{1}{t^2} \downarrow$ as $t \uparrow$

$$\begin{aligned} 1 - \Phi(x) &\geq \frac{1}{x} \varphi(x) - \frac{1}{x^2} \int_x^\infty \varphi(t) dt \\ &= \frac{1}{x} \varphi(x) - \frac{1}{x^2} (1 - \Phi(x)) \\ &\geq \frac{1}{x} \varphi(x) - \frac{1}{x^2} \frac{1}{x} \varphi(x) = \left(\frac{1}{x} - \frac{1}{x^3}\right) \varphi(x) \end{aligned}$$

□

COROLLARY 1.2.2: As $x \rightarrow \infty$,

$$1 - \Phi(x) \sim \frac{1}{x} \varphi(x)$$

PROOF:

$$\frac{1}{x} - \frac{1}{x^3} \rightarrow \frac{1}{x}$$

as $x \rightarrow \infty$. Using LEMMA 1.2.1 the result follows. $\square$

The following theorem is from ITO & MCKEAN JR (1974), among others. Consider $h(t) \in C(0, 1] : h(t) \geq 0$, a non-negative continuous function on the interval $(0, 1]$. Using Blumenthal's 0-1 law implies that $\mathbb{P} \left\{ \limsup_{t \downarrow 0} \frac{\mathbb{S}(t)}{h(t)} < 1 \right\} \in \{0, 1\}$, where $\mathbb{S}(t)$ is Brownian motion (see DEFINITION 1.3.1).

DEFINITION 1.2.3: $h(t)$ belongs to the *upper class* if $\mathbb{P} \left\{ \limsup_{t \downarrow 0} \frac{\mathbb{S}(t)}{h(t)} < 1 \right\} = 1$.
 $h(t)$ belongs to the *lower class* if $\mathbb{P} \left\{ \limsup_{t \downarrow 0} \frac{\mathbb{S}(t)}{h(t)} < 1 \right\} = 0$.

LEMMA 1.2.4 (Kolmogorov's Test): If $h(t) \uparrow$ and $t^{-\frac{1}{2}}h(t) \downarrow$ for small $t > 0$, then $h(t)$ belongs to the upper class if

$$\int_0^b t^{-\frac{3}{2}} h(t) \exp \left(-\frac{h^2(t)}{2t} \right) dt = \int_{0^+} t^{-\frac{3}{2}} h(t) \exp \left(-\frac{h^2(t)}{2t} \right) dt$$

converges for some $b > 0$.

PROOF: Define $\mathbb{P}_0$ to be the distribution of the Brownian motion which has initial point 0. Let $T_a \leq t$ be the first passage time of level a for Brownian motion. It will be shown in THEOREM 1.3.28 that the first passage time distribution is

$$\mathbb{P}_0 \{T_a \leq t\} = \int_0^t \frac{a}{\sqrt{2\pi s^3}} \exp \left(-\frac{a^2}{2s} \right) ds$$

Let $0 < a = t_1 < t_2 < \dots < t_n = b \leq 1$ be increments of $(0, 1]$ such that $h(t) \uparrow$ for $t \leq b$. Then, from FIGURE 1.1,

FIGURE 1.1: If $\mathbb{S}(t)$ exceeds $h(t)$ on $(a, b]$, then it must have already exceeded the stepwise approximation shown.

$$\begin{aligned}
& \mathbb{P}_0 \{ \mathbb{S}(t) \geq h(t) \text{ for some } a \leq t \leq b \} \\
& \leq \mathbb{P}_0 \{ T_{h(a)} \leq a \} + \sum_{m \geq 2} \mathbb{P}_0 \{ t_{m-1} < T_{h(t_{m-1})} \leq t_m \} \\
& = \int_0^a \frac{h(a)}{\sqrt{2\pi t^3}} \exp \left(-\frac{h^2(a)}{2t} \right) dt + \sum_{m \geq 2} \int_{t_{m-1}}^{t_m} \frac{h(t_{m-1})}{\sqrt{2\pi t^3}} \exp \left(-\frac{h^2(t_{m-1})}{2t} \right) dt
\end{aligned} \tag{1.6}$$

As $\max \{ |t_m - t_{m-1}| \} \downarrow 0$ as $n \rightarrow \infty$ this becomes

$$\leq \int_0^{\frac{a}{h^2(a)}} \frac{\exp \left(-\frac{1}{2t} \right) dt}{\sqrt{2\pi t^3}} + \int_a^b \frac{h(t)}{\sqrt{2\pi t^3}} \exp \left(-\frac{h^2(t)}{2t} \right) dt$$

At this stage we need to show that $\lim_{t \downarrow 0} t^{-\frac{1}{2}} h(t) = \infty$. Assume that

$\lim_{t \downarrow 0} t^{-\frac{1}{2}} h(t) \rightarrow M$ for $M < \infty$. As $t^{-\frac{1}{2}} h(t) \downarrow$ and $h(t)$ is non-negative, it

follows that $\lim_{t \downarrow 0} t^{-\frac{1}{2}} h(t) \neq 0$ unless $h(t) = 0$, which is trivial. If

$\lim_{t \downarrow 0} t^{-\frac{1}{2}} h(t) \rightarrow M < \infty$ for $M > 0$, the integral

$$\int_{0+} t^{-\frac{3}{2}} h(t) \exp \left(-\frac{h^2(t)}{2t} \right) dt \rightarrow \int_{0+} \frac{M \exp \left(-\frac{M^2}{2} \right)}{t} dt = \infty,$$

which contradicts the assumption that $\int_{0+} t^{-\frac{3}{2}} h(t) \exp \left(-\frac{h^2(t)}{2t} \right) dt$ is finite.

Therefore $\lim_{t \downarrow 0} t^{-\frac{1}{2}} h(t) = \infty$ and $\frac{a}{h^2(a)} \rightarrow 0$ as $a \rightarrow 0$. Letting $a \downarrow 0$ gives

$$\mathbb{P}_0 \{ \mathbb{S}(t) \geq h(t) \text{ for some } 0 < t \leq b \} \leq \int_0^b \frac{h(t)}{\sqrt{2\pi t^3}} \exp \left(-\frac{h^2(t)}{2t} \right) dt \rightarrow 0$$

as $b \downarrow 0$. □

THEOREM 1.2.5 (Local Law of the Iterated Logarithm):

$$\mathbb{P}_0 \left\{ \limsup_{t \downarrow 0} \frac{\mathbb{S}(t)}{\sqrt{2t \log \log \left(\frac{1}{t} \right)}} = 1 \right\} = 1$$

PROOF: Consider the function $h(t) = (1 + \epsilon)\sqrt{2t \log \log \left(\frac{1}{t}\right)}$ for $\epsilon > 0$. As $h(t)$ is positive, it is increasing if

$$\begin{aligned} \frac{d}{dt} t \log \log \left(\frac{1}{t}\right) &= \log \log \left(\frac{1}{t}\right) - \frac{1}{\log \left(\frac{1}{t}\right)} > 0 \\ &\Leftrightarrow \log \left(\frac{1}{t}\right) \log \log \left(\frac{1}{t}\right) > 1 \\ &\quad t < e^{-1} \end{aligned}$$

Then $h(t)$ is increasing for $t < e^{-1}$ and $t^{-\frac{1}{2}}h(t)$ is decreasing for $t \leq 1$ as $\log \log \left(\frac{1}{t}\right)$ decreases as t increases. We also need to show the convergence of

$$\int_{0^+} t^{-\frac{3}{2}} h(t) \exp \left(-\frac{h^2(t)}{2t} \right) dt = \int_0^\delta (1 + \epsilon) \sqrt{2 \log \log \left(\frac{1}{t}\right)} \left(\log \left(\frac{1}{t}\right) \right)^{-(1+\epsilon)^2} \frac{dt}{t}.$$

Letting $v = \log \log \left(\frac{1}{t}\right)$ gives

$$e^v = \log \left(\frac{1}{t}\right) \quad \frac{1}{t} = \exp(e^v) \quad dt = -e^v \exp(-e^v) dv$$

$$= \int_{\log \log \left(\frac{1}{\delta}\right)}^\infty (1 + \epsilon) \sqrt{2v} (e^v)^{1-(1+\epsilon)^2} dv.$$

For $\epsilon > 0$ we get convergence and for $-1 < \epsilon < 0$ the integral diverges, so it follows from LEMMA 1.2.4 that

$$\mathbb{P}_0 \left\{ \limsup_{t \downarrow 0} \frac{S(t)}{\sqrt{2t \log \log \left(\frac{1}{t}\right)}} \leq 1 \right\} = 1$$

Given $0 < \epsilon < 1$ and ignoring the set of Brownian paths for which

$\limsup_{t \downarrow 0} \mathbb{S}(t) > \sqrt{2t \log \log \left(\frac{1}{t} \right)}$, which has probability 0, if

$$A_n = \left\{ B_{\epsilon^n} < \left(1 - 3\sqrt{\epsilon} \right) \sqrt{2\epsilon^n \log \log \epsilon^{-n}} \right\}$$

occurs as $n \uparrow \infty$, then

$\mathbb{S}(\epsilon^n) - \mathbb{S}(\epsilon^{n+1}) < \left(1 - 3\sqrt{\epsilon} \right) \sqrt{2\epsilon^n \log \log \epsilon^{-n}} + \frac{3}{2} \sqrt{2\epsilon^{n+1} \log \log \epsilon^{-n-1}}$ and $A_n \subseteq B_n$ as $n \rightarrow \infty$, where

$$B_n = \left\{ \mathbb{S}(\epsilon^n) - \mathbb{S}(\epsilon^{n+1}) < (1 - \sqrt{\epsilon}) \sqrt{2\epsilon^n \log \log \epsilon^{-n}} \right\}$$

If $\theta = \frac{1 - \sqrt{\epsilon}}{\sqrt{1 - \epsilon}} < 1$ and $c = \theta \sqrt{2 \log \log \epsilon^{-n}}$, then using COROLLARY 1.2.2

$$1 - \mathbb{P}_0 \{ B_n \} = \int_c^\infty \varphi(b) db \sim \frac{1}{\theta \log n + \log \log \epsilon^{-1}} \frac{1}{\sqrt{2\pi}} \left(n^{-\theta^2} + \left(\log(\epsilon^{-1}) \right)^{-\theta^2} \right)$$

so $1 - \mathbb{P}_0 \{ B_n \} = C \frac{n^{-\theta^2}}{\sqrt{\log n}}$ where C does not depend on n . Using Cauchy's condensation test

$$\sum_{n=2}^\infty \frac{n^{-\theta^2}}{\sqrt{\log n}} \rightarrow \infty \Leftrightarrow \sum_{n=2}^\infty 2^n \frac{(2^n)^{-\theta^2}}{\sqrt{n \log 2}} = \sum_{n=2}^\infty \frac{(2^{n(1-\theta^2)})}{\sqrt{n \log 2}} \rightarrow \infty$$

which is true as $\theta^2 < 1$ implies $\frac{(2^{n(1-\theta^2)})}{\sqrt{n \log 2}} \rightarrow \infty$ as $n \rightarrow \infty$. As $\sum_{n=2}^\infty \frac{n^{-\theta^2}}{\sqrt{\log n}}$ diverges and B_n are independent, using the Borel-Cantelli lemma

$$1 - \mathbb{P}_0 \{ B_n \} = 1 \Rightarrow \mathbb{P}_0 \{ B_n \} = 0$$

Therefore

$$\begin{aligned} \mathbb{P}_0 \left\{ \bigcup_{n \geq 1} \bigcap_{m \geq n} A_m \right\} &\leq \mathbb{P}_0 \left\{ \bigcup_{n \geq 1} \bigcap_{m \geq n} B_m \right\} \\ &= \lim_{n \uparrow \infty} \mathbb{P}_0 \left\{ \bigcap_{m \geq n} B_m \right\} = 0 \end{aligned}$$

This shows that

$$\begin{aligned} 1 &= \mathbb{P}_0 \left\{ \mathbb{S}(\epsilon^n) \geq (1 - 3\sqrt{\epsilon}) \sqrt{2\epsilon^n \log \log (\epsilon^{-n})}, \text{ i.o., } n \uparrow \infty \right\} \\ &\leq \mathbb{P}_0 \left\{ \limsup_{t \downarrow 0} \frac{\mathbb{S}(t)}{\sqrt{2 \log \log \left(\frac{1}{t} \right)}} \geq 1 - 3\sqrt{\epsilon} \right\} \end{aligned}$$

Letting $\epsilon \downarrow 0$ gives

$$\mathbb{P}_0 \left\{ \limsup_{t \downarrow 0} \frac{\mathbb{S}(t)}{\sqrt{2t \log \log \left(\frac{1}{t} \right)}} \geq 1 \right\} = 1$$

This implies that

$$\mathbb{P}_0 \left\{ \limsup_{t \downarrow 0} \frac{\mathbb{S}(t)}{\sqrt{2t \log \log \left(\frac{1}{t} \right)}} = 1 \right\} = 1$$

□

REMARK 1.2.6: Kolmogorov's test also establishes that if

$$\int_0^b t^{-\frac{3}{2}} h(t) \exp \left(-\frac{h^2(t)}{2t} \right) dt = \int_{0^+} t^{-\frac{3}{2}} h(t) \exp \left(-\frac{h^2(t)}{2t} \right) dt$$

diverges for all $b > 0$, then $h(t)$ must belong to the lower class. This is more difficult to prove than convergence implies upper class, but using Kolmogorov's test in this form means that proving the local law of the iterated logarithm is the same as showing

$$h(t) = (1 + \epsilon) \sqrt{2t \log \log \left(\frac{1}{t} \right)} \quad \text{is upper class for } \epsilon > 0$$

and

$$h(t) = (1 - \epsilon) \sqrt{2t \log \log \left(\frac{1}{t} \right)} \quad \text{is lower class for } \epsilon > 0$$

which was done in the proof given above.

COROLLARY 1.2.7: Let $\mathbb{U}$ be Brownian bridge. Then

$$\mathbb{P}_0 \left\{ \limsup_{t \downarrow 0} \frac{\mathbb{U}(t)}{\sqrt{2t \log \log \left(\frac{1}{t} \right)}} = 1 \right\} = 1$$

PROOF: Let $\mathbb{U}(t) = \mathbb{S}(t) - t\mathbb{S}(1)$. The result follows directly from THEOREM 1.2.5 □

We conclude this subsection by considering law of the iterated logarithm results for empirical processes. The theorem is Smirnov's law of the iterated logarithm (see SMIRNOV (1944)), which will be proven using the following theorem.

THEOREM 1.2.8: Let λ_n be increasing. If $\sum_{n=1}^{\infty} \frac{\lambda_n^2}{n} \exp(-2\lambda_n^2) < \infty$,

$$\mathbb{P} \{ \|\mathbb{U}_n\| \geq \lambda_n \text{ i.o.} \} = 0.$$

Alternatively, if $\sum_{n=1}^{\infty} \frac{\lambda_n^2}{n} \exp(-2\lambda_n^2) = \infty$,

$$\mathbb{P} \{ \|\mathbb{U}_n\| \geq \lambda_n \text{ i.o.} \} = 1.$$

PROOF: Refer to CHUNG (1949). □

THEOREM 1.2.9: Let $\mathbb{U}_n$ denote the uniform empirical process. Then

$$\limsup_{n \rightarrow \infty} \frac{\|\mathbb{U}_n\|}{\sqrt{2 \log \log n}} = \frac{1}{2} \quad \text{almost surely}$$

PROOF: This proof follows [SHORACK & WELLNER \(1986\)](#). The lower bound can be established by considering $\mathbb{U}_n\left(\frac{1}{2}\right)$, which is the average of n independent Bernoulli $\left(\frac{1}{2}\right)$ random variables, centered by subtracting the expected value $\frac{1}{2}$. The law of iterated logarithm for Bernoulli variables (see [FELLER \(1971b\)](#), among others) implies that

$$\limsup_{n \rightarrow \infty} \frac{\|\mathbb{U}_n\|}{\sqrt{2 \log \log n}} \geq \limsup_{n \rightarrow \infty} \frac{\mathbb{U}_n\left(\frac{1}{2}\right)}{\sqrt{2 \log \log n}} \geq \frac{1}{2} \quad \text{almost surely}$$

For the upper bound let $\lambda_n = (1 + \epsilon) \frac{\sqrt{2 \log \log(n)}}{2}$ for any $\epsilon > 0$. We consider

$$\begin{aligned} & \sum_{n=1}^{\infty} \frac{\lambda_n^2}{n} \exp(-2\lambda_n^2) \\ &= \sum_{n=1}^{\infty} \frac{(1 + \epsilon)^2}{2} \frac{\log \log(n)}{n} \exp\left(-(1 + \epsilon)^2 \log((\log(n))^{-(1+\epsilon)^2})\right) \\ &= \sum_{n=1}^{\infty} \frac{(1 + \epsilon)^2}{2} \frac{\log \log(n)}{n (\log n)^{(1+\epsilon)^2}} \end{aligned}$$

Using the Cauchy condensation test, this will converge if and only if

$$\sum_{n=1}^{\infty} \frac{(1 + \epsilon)^2}{2} \frac{\log(n \log 2)}{n^{(1+\epsilon)^2} (\log 2)^{(1+\epsilon)^2}}$$

converges, which converges if and only if

$$\sum_{n=1}^{\infty} C n 2^{1-(1+\epsilon)^2}$$

converges, for constant C which does not depend on n , which is true for $\epsilon > 0$.

Using [THEOREM 1.2.8](#), it follows that

$$\mathbb{P} \left\{ \frac{\|\mathbb{U}_n\|}{\sqrt{2 \log \log n}} > \frac{1}{2}(1 + \epsilon) \text{ i.o.} \right\} = 0$$

for all $\epsilon > 0$. Letting $\epsilon \downarrow 0$ completes the proof. $\square$

1.2.2 LEVY MODULUS OF CONTINUITY

LEMMA 1.2.10 (Chung, Erdos, Sirao): If $h(t) \in C(0, 1]$ is positive and if $h(t) \uparrow$ and $t^{-\frac{1}{2}}h(t) \downarrow$ as $t \downarrow 0$, then

$$\mathbb{P} \left\{ \lim_{\epsilon \downarrow 0} \sup_{t_2 - t_1 = \epsilon} \frac{|\mathbb{S}(t_2) - \mathbb{S}(t_1)|}{h(\epsilon)} < 1; 0 \leq t_1 < t_2 \leq 1 \right\} = 1$$

if for any $\delta > 0$

$$\int_{0+}^{\delta} t^{-\frac{7}{2}} h^3(t) \exp \left(-\frac{h^2(t)}{2t} \right) dt < \infty$$

and

$$\mathbb{P} \left\{ \lim_{\epsilon \downarrow 0} \sup_{t_2 - t_1 = \epsilon} \frac{|\mathbb{S}(t_2) - \mathbb{S}(t_1)|}{h(\epsilon)} < 1; 0 \leq t_1 < t_2 \leq 1 \right\} = 0$$

if for all $\delta > 0$

$$\int_{0+}^{\delta} t^{-\frac{7}{2}} h^3(t) \exp \left(-\frac{h^2(t)}{2t} \right) dt = \infty$$

PROOF: The reader is referred to **CHUNG ET AL. (1959)** for a proof of this result. $\square$

Using this lemma we intend to find a function $h(\epsilon)$ so that

$$\mathbb{P} \left\{ \lim_{\epsilon \downarrow 0} \sup_{t_2 - t_1 = \epsilon} \frac{|\mathbb{S}(t_2) - \mathbb{S}(t_1)|}{h(\epsilon)} = 1 \right\} = 1 \quad 0 \leq t_1 < t_2 \leq 1 \quad (1.7)$$

As the variance of the increment $\mathbb{S}(t_2) - \mathbb{S}(t_1)$ is ϵ (if $t_2 - t_1 = \epsilon$), a reasonable first choice for the function $h(\epsilon)$ could be $h(\epsilon) = M\sqrt{\epsilon}$, for some constant M . For M large enough we would initially expect that

$$\mathbb{P} \left\{ \lim_{\epsilon \downarrow 0} \sup_{t_2 - t_1 = \epsilon} \frac{|\mathbb{S}(t_2) - \mathbb{S}(t_1)|}{M\sqrt{\epsilon}} \leq 1 \right\} = 1 \quad 0 \leq t_1 < t_2 \leq 1$$

because for fixed ϵ , the probability of an increment of Brownian motion exceeding $M\sqrt{\epsilon} \rightarrow 0$ as $M \uparrow$. Using LEMMA 1.2.10

$$\int_{0+}^{\delta} t^{-\frac{7}{2}} M^3 t^{\frac{3}{2}} \exp\left(-\frac{M^2 t}{2}\right) dt = M^3 \exp\left(-\frac{M^2}{2}\right) \int_{0+}^{\delta} t^{-2} dt = \infty$$

so we conclude that

$$\mathbb{P} \left\{ \lim_{\epsilon \downarrow 0} \sup_{t_2 - t_1 = \epsilon} \frac{|\mathbb{S}(t_2) - \mathbb{S}(t_1)|}{M\sqrt{\epsilon}} \leq 1 \right\} = 0 \quad 0 \leq t_1 < t_2 \leq 1.$$

This means there will be infinitely many increments which cross $M\sqrt{\epsilon}$ as $\epsilon \rightarrow 0$. We need a different function for $h(\epsilon)$, which is given in the next theorem.

THEOREM 1.2.11 (P. Levy's Holder condition):

$$\mathbb{P}_0 \left\{ \lim_{\epsilon \downarrow 0} \sup_{t_2 - t_1 = \epsilon} \frac{|\mathbb{S}(t_2) - \mathbb{S}(t_1)|}{\sqrt{2\epsilon \log\left(\frac{1}{\epsilon}\right)}} = 1; 0 \leq t_1 < t_2 \leq 1 \right\} = 1$$

PROOF: Let $h(t) = (1 + \epsilon) \sqrt{2t \log\left(\frac{1}{t}\right)}$. We show equality in two parts. Firstly by showing that the integral diverges for $\epsilon \leq 0$, so that the increments of Brownian motion cross $\sqrt{2t \log\left(\frac{1}{t}\right)}$ infinitely often, which shows we have found a function that is not too large. Secondly by showing that for any $\epsilon > 0$ the integral diverges, so if the function is increased by any small multiple, the increments of Brownian motion will no longer cross infinitely often. It is seen that $h(t) \uparrow$ for $t < e^{-1}$, $t^{-\frac{1}{2}}h(t) \downarrow$ for $t \leq 1$. Using LEMMA 1.2.10

$$\begin{aligned} \int_0^{\delta} t^{-\frac{7}{2}} h^3(t) \exp\left(-\frac{h^2(t)}{2t}\right) dt &= 2^{\frac{3}{2}} (1 + \epsilon)^3 \int_0^{\delta} \left(\log\left(\frac{1}{t}\right)\right)^{\frac{3}{2}} t^{(1+\epsilon)^2-2} dt \\ &= \int_{\log(\frac{1}{\delta})}^{\infty} u^{\frac{3}{2}} (e^u)^{1-(1+\epsilon)^2} du \end{aligned}$$

which diverges for small $\epsilon \leq 0$ and converges for $\epsilon > 0$. □

We also want to show that other functions can also satisfy (1.7). Let

$$h(t) = \sqrt{2t \left[\log \left(\frac{1}{t} \right) + \frac{5}{2} \log_2 \left(\frac{1}{t} \right) + \sum_{j=3}^{n-1} \log_j \left(\frac{1}{t} \right) + (1 + \delta) \log_n \left(\frac{1}{t} \right) \right]}$$

Then

$$\begin{aligned} -\frac{h^2(t)}{2t} &= \log(t) + \log \left(\frac{1}{\left(\log \left(\frac{1}{t} \right) \right)^{\frac{5}{2}}} \right) + \sum_{i=2}^{n-2} \log \left(\frac{1}{\log_i \left(\frac{1}{t} \right)} \right) \\ &\quad + \log \left(\frac{1}{\left(\log_{n-1} \left(\frac{1}{t} \right) \right)^{1+\delta}} \right) \end{aligned}$$

and

$$\begin{aligned} &\lim_{\epsilon \downarrow 0} \int_0^\epsilon t^{-\frac{7}{2}} h^3(t) \exp \left(-\frac{h^2(t)}{2t} \right) dt \\ &= \lim_{\epsilon \downarrow 0} \int_0^\epsilon t^{-\frac{7}{2}} h^3(t) t \frac{1}{\left(\log \left(\frac{1}{t} \right) \right)^{\frac{5}{2}}} \frac{1}{\log_2 \left(\frac{1}{t} \right)} \cdots \frac{1}{\log_{n-2} \left(\frac{1}{t} \right)} \frac{1}{\left(\log_{n-1} \left(\frac{1}{t} \right) \right)^{1+\delta}} dt \end{aligned}$$

Let $u = \log_n \left(\frac{1}{t} \right)$, $du = \frac{-1}{t} \frac{1}{\left(\log \left(\frac{1}{t} \right) \right)} \prod_{j=2}^{n-2} \frac{1}{\log_j \left(\frac{1}{t} \right)} \frac{1}{\left(\log_{n-1} \left(\frac{1}{t} \right) \right)} dt$ and

$\frac{1}{t} = \exp_n(u)$. For ϵ small enough

$$\begin{aligned} &= \lim_{\epsilon \downarrow 0} \int_{\log_n \left(\frac{1}{\epsilon} \right)}^\infty \frac{\left(u(1 + \delta) + \exp(u) + \cdots + \frac{5}{2} \exp_{n-1}(u) + \exp_{n-1}(u) \right)^{\frac{3}{2}} du}{(\exp(u))^\delta (\exp_{n-1}(u))^{\frac{3}{2}}} \\ &= \lim_{\epsilon \downarrow 0} \int_{\log_n \left(\frac{1}{\epsilon} \right)}^\infty \left(\frac{u(1 + \delta)}{\exp_{n-1}(u)} + \cdots + 1 \right)^{\frac{3}{2}} \exp(-u\delta) du \end{aligned}$$

which converges for $\delta > 0$ and diverges for $\delta \leq 0$. Therefore, for $0 \leq t_1 < t_2 \leq 1$

$$\mathbb{P}_0 \left\{ \lim_{\epsilon \downarrow 0} \sup_{t_2 - t_1 = \epsilon} \frac{|\mathbb{S}(t_2) - \mathbb{S}(t_1)|}{\sqrt{2\epsilon \left[\log \left(\frac{1}{\epsilon} \right) + \frac{5}{2} \log_2 \left(\frac{1}{\epsilon} \right) + \cdots + \log_n \left(\frac{1}{\epsilon} \right) \right]}} = 1 \right\} = 1$$

We finish this section by considering an extension of this type of results to the uniform empirical process.

DEFINITION 1.2.12: The modulus of continuity ω_n of the uniform empirical process $\mathbb{U}_n$ is defined as

$$\omega_n(a) = \sup_{0 \leq h \leq a} \sup_{0 \leq t \leq 1-h} |\mathbb{U}_n(t+h) - \mathbb{U}_n(t)|$$

THEOREM 1.2.13: Let $a_n \downarrow 0$, $na_n \uparrow$, $\frac{\log(a_n^{-1})}{\log \log(n)} \rightarrow \infty$ and $\frac{\log(a_n^{-1})}{na_n} \rightarrow 0$. Then

$$\lim_{n \rightarrow \infty} \frac{\omega_n(a_n)}{\sqrt{2a_n \log \log \left(\frac{1}{a_n} \right)}} = 1 \quad \text{a.s.}$$

PROOF: Can be found in [SHORACK & WELLNER \(1986\)](#), among others. $\square$

Other limiting behaviours are possible depending on the exact rate which a_n converges to 0. An excellent reference for a discussion of this topic is [SHORACK & WELLNER \(1986\)](#).

1.3 SOME LIMITING PROCESSES

This section provides an introduction to the theory of Gaussian (or normal) processes, the natural limit of many empirical processes. The theory of Gaussian processes is of independent interest and has many applications

beyond empirical processes, but will mainly serve to motivate the next section. The aim of this section is to give an overview of the theory behind some of the processes which can arise as limits of empirical processes. The following section follows [SHORACK & WELLNER \(1986\)](#).

1.3.1 GAUSSIAN PROCESSES

DEFINITION 1.3.1: Let $\mathbb{S}$ be a Gaussian process. Then if

1. $E[\mathbb{S}(t)] = 0; \quad \forall t$
2. $\text{Cov}[\mathbb{S}(t), \mathbb{S}(t')] = \min\{t, t'\} = t \wedge t'; \quad \forall t, t'$

$\mathbb{S}$ is called Brownian motion.

DEFINITION 1.3.2: Let T be an interval in $(-\infty, \infty)$. Then

$$X(s, t] \equiv X(t) - X(s); \quad \forall s, t \in T$$

is an increment of X .

DEFINITION 1.3.3: If $X(t_0), X(t_0, t_1], \dots, X(t_{k-1}, t_k]$ are independent random variables for all $k \geq 1$ and all $t_0 < t_1 < \dots < t_k$ then X has independent increments.

DEFINITION 1.3.4: If

$$(X(t_1 + h), X(t_2 + h), \dots, X(t_k + h)) \cong (X(t_1), X(t_2), \dots, X(t_k))$$

for all $k \geq 1$, $h \geq 0$ and all time points in T , then X is a stationary process

LEMMA 1.3.5: Let $0 \equiv t_0 < t_1 < \dots < t_{k+1} \equiv 1$. Then

$$\mathbb{S}(t_1) - \mathbb{S}(t_0), \dots, \mathbb{S}(t_{k+1}) - \mathbb{S}(t_k)$$

are independent random variables distributed as $N(0, t_1 - t_0), \dots, N(0, t_{k+1} - t_k)$ for all $k \geq 1$. Also $\mathbb{S}(0) = 0$.

PROOF: Follows directly from the definition of Brownian motion. $\square$

DEFINITION 1.3.6: Let $\mathbb{U}$ be a Gaussian process. If $E[\mathbb{U}(t)] = 0$ and

$$\text{Cov}(\mathbb{U}(t), \mathbb{U}(t')) = t \wedge t' - tt'$$

for $0 \leq t, t' \leq 1$, then $\mathbb{U}$ is Brownian bridge.

DEFINITION 1.3.7: Let $\mathbb{X}$ be a Gaussian process. If $E[\mathbb{X}(t)] = 0$ and

$$\text{Cov}(\mathbb{X}(t), \mathbb{X}(t')) = \exp(-|t - t'|)$$

for all $t, t' \in \mathbb{R}$, then $\mathbb{X}$ is the *Ornstein-Uhlenbeck* process.

DEFINITION 1.3.8: Let $\mathbb{S}$ be a Gaussian process. If $E[\mathbb{S}(t_1, t_2)] = 0$ and

$$\text{Cov}(\mathbb{S}(t_1, t_2), \mathbb{S}(t'_1, t'_2)) = (t_1 \wedge t'_1)(t_2 \wedge t'_2)$$

for all $t_1, t_2, t'_1, t'_2 \in \mathbb{R}^+$, then $\mathbb{S}$ is the Brownian motion in 2-dimensional time.

DEFINITION 1.3.9: Let $\mathbb{K}$ be a Gaussian process. If $E[\mathbb{K}(t_1, t_2)] = 0$ and

$$\text{Cov}(\mathbb{K}(t_1, t_2), \mathbb{K}(t'_1, t'_2)) = (t_1 \wedge t'_1)(t_2 \wedge t'_2 - t_2 t'_2)$$

for all $t_1, t'_1 \in \mathbb{R}^+$ and $t_2, t'_2 \in [0, 1]$, then $\mathbb{K}$ is the *Keifer* process. Notice that this is Brownian motion in the first dimension and Brownian bridge in the second dimension.

LEMMA 1.3.10: Define a random element $\mathbb{V}$ to be

$$\mathbb{V}(t) = \mathbb{S}(t) - t\mathbb{S}(1); \quad 0 \leq t \leq 1$$

Then $\mathbb{V}$ is Brownian bridge if $\mathbb{S}$ is Brownian motion.

FIGURE 1.2: Simulation of Brownian motion path

FIGURE 1.3: Simulation of Brownian Bridge path

PROOF:

$$\begin{aligned}
E[\mathbb{V}(t)\mathbb{V}(t')] &= E[(\mathbb{S}(t) - t\mathbb{S}(1))(\mathbb{S}(t') - t'\mathbb{S}(1))] \\
&= E[\mathbb{S}(t)\mathbb{S}(t')] - E[t\mathbb{S}(1)\mathbb{S}(t')] - E[t'\mathbb{S}(t)\mathbb{S}(1)] + E[tt'\mathbb{S}(1)\mathbb{S}(1)] \\
&= t \wedge t' - tt' - tt' + tt' \\
&= t \wedge t' - tt'
\end{aligned}$$

Therefore $\mathbb{V}$ is Brownian bridge. $\square$

LEMMA 1.3.11: Let $\mathbb{U}$ be Brownian bridge on $[0, 1]$. Then

$$\mathbb{S}(t) = (1+t)\mathbb{U}\left(\frac{t}{1+t}\right); \quad t \geq 0$$

is Brownian motion on $\mathbb{R}^+$.

PROOF:

$$\begin{aligned}
E[\mathbb{S}(t)\mathbb{S}(t')] &= E\left[(1+t)\mathbb{U}\left(\frac{t}{1+t}\right)(1+t')\mathbb{U}\left(\frac{t'}{1+t'}\right)\right] \\
&= (1+t)(1+t')\left[\min\left\{\frac{t}{1+t}, \frac{t'}{1+t'}\right\} - \frac{tt'}{(1+t)(1+t')}\right] \\
&= \min\{t(1+t'), t'(1+t)\} - tt' \\
&= t \wedge t'
\end{aligned}$$

Therefore $\mathbb{S}(t)$ is Brownian motion. $\square$

LEMMA 1.3.12: If $\mathbb{S}$ is Brownian motion on $\mathbb{R}^+$ and $r > 0$, then

$$\sqrt{r}\mathbb{S}\left(\frac{t}{r}\right)$$

is Brownian motion.

PROOF:

$$\begin{aligned} \text{Cov} \left(\sqrt{r} \mathbb{S} \left(\frac{t}{r} \right), \sqrt{r} \mathbb{S} \left(\frac{t'}{r} \right) \right) &= E \left[r \mathbb{S} \left(\frac{t}{r} \right) \mathbb{S} \left(\frac{t'}{r} \right) \right] \\ &= r \min \left\{ \frac{t}{r}, \frac{t'}{r} \right\} = t \wedge t' \end{aligned}$$

Therefore $\sqrt{r} \mathbb{S} \left(\frac{t}{r} \right)$ is Brownian motion. □

LEMMA 1.3.13: If $\mathbb{S}$ is Brownian motion on $\mathbb{R}^+$ and $r > 0$, then

$$\mathbb{S}(t+r) - \mathbb{S}(r)$$

is Brownian motion.

PROOF:

$$\begin{aligned} &E[(\mathbb{S}(t+r) - \mathbb{S}(r))(\mathbb{S}(t'+r) - \mathbb{S}(r))] \\ &= E[\mathbb{S}(t+r)\mathbb{S}(t'+r)] - E[\mathbb{S}(r)\mathbb{S}(t'+r)] - E[\mathbb{S}(t+r)\mathbb{S}(r)] + E[\mathbb{S}(r)\mathbb{S}(r)] \\ &= \min\{t+r, t'+r\} - r - r + r \\ &= t \wedge t' \end{aligned}$$

Therefore $\mathbb{S}(t+r) - \mathbb{S}(r)$ is Brownian motion. □

LEMMA 1.3.14: Let $\mathbb{S}$ be Brownian motion on $\mathbb{R}^+$. Then

$$\mathbb{U}(t) = (1-t)\mathbb{S}\left(\frac{t}{1-t}\right)$$

is Brownian bridge.

PROOF:

$$\begin{aligned}
 E[\mathbb{U}(t)\mathbb{U}(t')] &= (1-t)(1-t')E\left[\mathbb{S}\left(\frac{t}{1-t}\right)\mathbb{S}\left(\frac{t'}{1-t'}\right)\right] \\
 &= (1-t)(1-t')\min\left\{\frac{t}{1-t}, \frac{t'}{1-t'}\right\} \\
 &= t \wedge t' - tt'
 \end{aligned}$$

Therefore $\mathbb{U}(t)$ is Brownian bridge. $\square$

LEMMA 1.3.15: Let $S : \mathbb{U}(t) \rightarrow \mathbb{S}(t) = (1+t)\mathbb{U}\left(\frac{t}{1+t}\right)$ and $T : \mathbb{S}(t) \mapsto \mathbb{U}(t) = (1-t)\mathbb{S}\left(\frac{t}{1-t}\right)$. Then S and T are inverse transformations of each other.

PROOF: Let $\mathbb{S}(t)$ be Brownian motion. Then $T(\mathbb{S}(t)) = (1-t)\mathbb{S}\left(\frac{t}{1-t}\right)$ is Brownian bridge and we will define $\mathbb{U}'(t) = T(\mathbb{S}(t))$. Then,

$$S(T(\mathbb{S}(t))) = (1+t)\mathbb{U}'\left(\frac{t}{1+t}\right) = (1+t)\left(1 - \frac{t}{1+t}\right)\mathbb{S}\left(\frac{\frac{t}{1+t}}{1 - \frac{t}{1+t}}\right) = \mathbb{S}(t)$$

Now let $\mathbb{U}(t)$ be Brownian bridge. Then $S(\mathbb{U}(t)) = (1+t)\mathbb{U}\left(\frac{t}{1+t}\right) = \mathbb{S}'(t)$ and

$$T(S(\mathbb{U}(t))) = (1-t)\mathbb{S}'\left(\frac{t}{1-t}\right) = (1-t)\left(1 + \frac{t}{1-t}\right)\mathbb{U}\left(\frac{\frac{t}{1-t}}{1 + \frac{t}{1-t}}\right) = \mathbb{U}(t)$$

$\square$

LEMMA 1.3.16: Define $\mathbb{Z}$ to be random element by

$$\mathbb{Z}(t) = \frac{\mathbb{U}(t) + \mathbb{U}(1-t)}{\sqrt{2}}; \quad 0 \leq t \leq 1$$

where $\mathbb{U}$ is Brownian bridge. Then $\mathbb{Z}$ is Brownian motion on $\left[0, \frac{1}{2}\right]$. Further

$$\mathbb{U}\left(\frac{t}{2}\right) - \mathbb{U}\left(1 - \frac{t}{2}\right); \quad 0 \leq t \leq 1$$

is Brownian bridge.

PROOF:

$$\begin{aligned} E[\mathbb{Z}(t)\mathbb{Z}(t')] &= \frac{1}{2} (E[\mathbb{U}(t)\mathbb{U}(t')] + E[\mathbb{U}(t)\mathbb{U}(1-t')] \\ &\quad + E[\mathbb{U}(1-t)\mathbb{U}(t')] + E[\mathbb{U}(1-t)\mathbb{U}(1-t')]) \\ &= \frac{1}{2} (\min\{t, t'\} + \min\{t, 1-t'\} + \min\{1-t, t'\} + \min\{1-t, 1-t'\} - 1) \end{aligned}$$

At this point it is convenient to assume $t < t'$. If $t < t'$, then if $\mathbb{Z}$ is Brownian motion, the covariance being calculated should be t .

$$\begin{aligned} &= \frac{1}{2} (t + t + t' + 1 - t' - 1) \\ &= \frac{1}{2} (2t) = t \end{aligned}$$

Therefore $\mathbb{Z}$ is Brownian motion.

$$\begin{aligned} &E\left[\left(\mathbb{U}\left(\frac{t}{2}\right) - \mathbb{U}\left(1 - \frac{t}{2}\right)\right)\left(\mathbb{U}\left(\frac{t'}{2}\right) - \mathbb{U}\left(1 - \frac{t'}{2}\right)\right)\right] \\ &= \frac{1}{2} \min\{t, t'\} - \frac{tt'}{4} - \min\left\{\frac{t}{2}, 1 - \frac{t'}{2}\right\} + \frac{t}{2} - \frac{tt'}{4} \\ &\quad - \min\left\{\frac{t'}{2}, 1 - \frac{t}{2}\right\} + \frac{t'}{2} - \frac{tt'}{4} + \min\left\{1 - \frac{t}{2}, 1 - \frac{t'}{2}\right\} - 1 + \frac{t}{2} + \frac{t'}{2} - \frac{tt'}{4} \end{aligned}$$

Assume $t < t'$. This would lead to $t - tt'$ if this process is indeed Brownian bridge

$$\begin{aligned} &= \frac{t}{2} - \frac{t}{2} + \frac{t}{2} - \frac{t'}{2} + \frac{t'}{2} + 1 - \frac{t'}{2} - 1 + \frac{t}{2} + \frac{t'}{2} - tt' \\ &= t - tt' \end{aligned}$$

Therefore $\mathbb{U}\left(\frac{t}{2}\right) - \mathbb{U}\left(1 - \frac{t}{2}\right)$ is Brownian bridge on $\left[0, \frac{1}{2}\right]$. $\square$

LEMMA 1.3.17: Let $\mathbb{U}$ and $\mathbb{V}$ be independent Brownian bridges. Then for $0 \leq a \leq 1$,

$$\sqrt{1-a}\mathbb{U} + \sqrt{a}\mathbb{V}$$

is also Brownian bridge.

PROOF: First, note that because $\mathbb{U}$ and $\mathbb{V}$ are independent, $E[\mathbb{U}(t)\mathbb{V}(t')] = 0$ for all t and t' . Therefore the cross terms can be ignored in the expansion below.

$$\begin{aligned} &E\left[\left(\sqrt{1-a}\mathbb{U}(t) + \sqrt{a}\mathbb{V}(t)\right)\left(\sqrt{1-a}\mathbb{U}(t') + \sqrt{a}\mathbb{V}(t')\right)\right] \\ &= (1-a)(\mathbb{U}(t)\mathbb{U}(t')) + a(\mathbb{V}(t)\mathbb{V}(t')) \\ &= (1-a)(t \wedge t' - tt') + a(t \wedge t' - tt') \end{aligned}$$

Therefore $\sqrt{1-a}\mathbb{U} - \sqrt{a}\mathbb{V}$ is Brownian bridge. $\square$

LEMMA 1.3.18: Let $\mathbb{U}$ be Brownian bridge and Z be a $N(0, 1)$ random variable independent of $\mathbb{U}$. Then $\mathbb{U} + tZ$ is Brownian motion on $[0, 1]$.

PROOF: First, note that because $\mathbb{U}$ and Z are independent, $E[\mathbb{U}Z] = 0$. Therefore the cross terms can be ignored in the expansion below.

$$\begin{aligned} E[(\mathbb{U}(t) + tZ)(\mathbb{U}(t') + t'Z)] &= E[\mathbb{U}(t)\mathbb{U}(t')] + tt'E[Z^2] \\ &= t \wedge t' - tt' + tt' = t \wedge t' \end{aligned}$$

Therefore $\mathbb{U} + tZ$ is Brownian motion. $\square$

LEMMA 1.3.19: $\mathbb{S}(1)$ and $\mathbb{U}(t)$ are independent.

PROOF: $E[\mathbb{S}(1)\mathbb{U}(t)] = E[\mathbb{S}(1)\mathbb{S}(t)] - tE[\mathbb{S}(1)\mathbb{S}(1)] = t - t = 0$ $\square$

LEMMA 1.3.20: Let $\mathbb{S}$ be Brownian motion on $[0, \infty)$. Then the *time reversal* process

$$t\mathbb{S}\left(\frac{1}{t}\right)$$

for $t \geq 0$ is also Brownian motion.

PROOF:

$$\begin{aligned} E\left[t\mathbb{S}\left(\frac{1}{t}\right)t'\mathbb{S}\left(\frac{1}{t'}\right)\right] &= tt' \min\left\{\frac{1}{t}, \frac{1}{t'}\right\} \\ &= t' \wedge t \end{aligned}$$

Therefore $t\mathbb{S}\left(\frac{1}{t}\right)$ is Brownian motion. $\square$

LEMMA 1.3.21: Let $\mathbb{S}$ be Brownian motion on $\mathbb{R}^+$. Then

$$\mathbb{X}(t) = e^{-t}\mathbb{S}(e^{2t})$$

is a Ornstein-Uhlenbeck process.

PROOF:

$$\begin{aligned} \text{Cov}\left(e^{-t}\mathbb{S}(e^{2t}), e^{-t'}\mathbb{S}(e^{2t'})\right) &= e^{-t-t'} E\left[\mathbb{S}(e^{2t})\mathbb{S}(e^{2t'})\right] \\ &= e^{-t-t'} \min\left\{e^{2t}, e^{2t'}\right\} \\ &= e^{-|t-t'|} \end{aligned}$$

Therefore $\mathbb{X}(t)$ is an Ornstein-Uhlenbeck process. $\square$

LEMMA 1.3.22: Let $\mathbb{S}$ be a Ornstein-Uhlenbeck process on $\mathbb{R}$. Then

$$\mathbb{U}(t) = \begin{cases} \sqrt{t(1-t)}\mathbb{X}\left(\frac{1}{2}\log\left(\frac{t}{1-t}\right)\right) & 0 \leq t \leq 1 \\ 0 & t = 0 \text{ or } t = 1 \end{cases}$$

is Brownian bridge on $[0, 1]$.

PROOF:

$$E[\mathbb{U}(t)\mathbb{U}(t')] = E\left[\sqrt{tt'(1-t)(1-t')}\mathbb{X}\left(\frac{1}{2}\log\left(\frac{t}{1-t}\right)\right)\mathbb{X}\left(\frac{1}{2}\log\left(\frac{t'}{1-t'}\right)\right)\right]$$

Assuming $t < t'$

$$\begin{aligned} &= \sqrt{tt'(1-t)(1-t')} \exp\left(-\log\left(\left(\frac{t(1-t')}{t'(1-t)}\right)^{\frac{1}{2}}\right)\right) \\ &= \sqrt{tt'(1-t)(1-t')} \sqrt{\frac{t(1-t')}{t'(1-t)}} = t - tt' \end{aligned}$$

Therefore it is Brownian bridge. $\square$

LEMMA 1.3.23: Let $\mathbb{U}$ be a Brownian bridge. Define

$$\mathbb{B}(t) = \frac{[\mathbb{U}(at) - t\mathbb{U}(a)]}{\sqrt{a}}$$

for fixed $0 < a \leq 1$. Then $\mathbb{B}$ is Brownian bridge on $[0, 1]$ and is independent of $\{\mathbb{U}(t) : a \leq t \leq 1\}$.

PROOF:

$$\begin{aligned}
E[\mathbb{B}(t)\mathbb{B}(t')] &= \frac{1}{a} (E[\mathbb{U}(at)\mathbb{U}(at') - \mathbb{U}(at)t'\mathbb{U}(a) - \mathbb{U}(at')t\mathbb{U}(a) + tt'\mathbb{U}(a)\mathbb{U}(a)]) \\
&= \frac{1}{a} (at \wedge at' - a^2tt' - t'(at - a^2t) - t(at' - at'^2) + tt'a(1 - a)) \\
&= t \wedge t' - tt'
\end{aligned}$$

To show independence consider the covariance between $\mathbb{B}$ and $\mathbb{U}$ ($0 \leq t \leq a \leq t' \leq 1$).

$$\begin{aligned}
E[\mathbb{B}(t)\mathbb{U}(t')] &= \frac{1}{\sqrt{a}} (E[\mathbb{U}(at)\mathbb{U}(t')] - tE[\mathbb{U}(a)\mathbb{U}(t')]) \\
&= \frac{1}{\sqrt{a}} (at - att' - t(a - att')) \\
&= 0
\end{aligned}$$

As they are uncorrelated and Gaussian, they are independent. $\square$

LEMMA 1.3.24: Let $\mathbb{S}$ be Brownian motion and $\mathbb{U}$ be Brownian bridge. Then

1. $\mathbb{S}(t) - \int_0^t \frac{\mathbb{S}(r)}{r} dr; \quad 0 \leq t \leq 1$
2. $\mathbb{U}(t) - \int_0^t \frac{\mathbb{U}(r)}{r} dr; \quad 0 \leq t \leq 1$
3. $\mathbb{U}(t) + \int_0^t \frac{\mathbb{U}(r)}{1-r} dr; \quad 0 \leq t \leq 1$

are all Brownian motion.

PROOF:

1. Assuming $t < t'$

$$\begin{aligned}
& E \left[\left(\mathbb{S}(t) - \int_0^t \frac{\mathbb{S}(r)}{r} dr \right) \left(\mathbb{S}(t') - \int_0^{t'} \frac{\mathbb{S}(r)}{r} dr \right) \right] \\
&= E [\mathbb{S}(t)\mathbb{S}(t')] - E \left[\mathbb{S}(t) \int_0^{t'} \frac{\mathbb{S}(r)}{r} dr \right] - E \left[\mathbb{S}(t') \int_0^t \frac{\mathbb{S}(r)}{r} dr \right] \\
&\quad + E \left[\int_0^t \frac{\mathbb{S}(r)}{r} dr \int_0^{t'} \frac{\mathbb{S}(r)}{r} dr \right] \\
&= t - \int_0^t \frac{E [\mathbb{S}(t)\mathbb{S}(r)]}{r} dr - \int_t^{t'} \frac{E [\mathbb{S}(t)\mathbb{S}(r)]}{r} dr - \int_0^t \frac{E [\mathbb{S}(t')\mathbb{S}(r)]}{r} dr \\
&\quad + \int_0^t \int_0^t \frac{E [\mathbb{S}(r)\mathbb{S}(r')]}{rr'} dr dr' + \int_t^{t'} \int_0^t \frac{E [\mathbb{S}(r)\mathbb{S}(r')]}{rr'} dr dr' \\
&= t - \int_0^t dr - t \int_t^{t'} \frac{dr}{r} - \int_0^t dr + 2 \int_0^t \int_0^{r'} \frac{1}{r'} dr dr' + \int_t^{t'} \int_0^t \frac{1}{r'} dr dr' \\
&= t - t - t \int_t^{t'} \frac{dr}{r} - t + 2t + t \int_t^{t'} \frac{1}{r'} dr' \\
&= t
\end{aligned}$$

Therefore $\mathbb{S}(t) - \int_0^t \frac{\mathbb{S}(r)}{r} dr$; $0 \leq t \leq 1$ is Brownian motion.

2.

$$\begin{aligned}
\mathbb{U}(t) - \int_0^t \frac{\mathbb{U}(r)}{r} dr &= \mathbb{S}(t) - t\mathbb{S}(1) - \int_0^t \frac{\mathbb{S}(r)}{r} dr - \mathbb{S}(1) \int_0^t dr \\
&= \mathbb{S}(t) - \int_0^t \frac{\mathbb{S}(r)}{r} dr
\end{aligned}$$

Therefore using 1, we see that $\mathbb{U}(t) - \int_0^t \frac{\mathbb{U}(r)}{r} dr$; $0 \leq t \leq 1$ is Brownian motion.

3. Assume $t < t'$

$$\begin{aligned}
& E \left[\left(\mathbb{U}(t) + \int_0^t \frac{\mathbb{U}(r)}{1-r} dr \right) \left(\mathbb{U}(t') + \int_0^{t'} \frac{\mathbb{U}(r)}{1-r} dr \right) \right] \\
&= E [\mathbb{U}(t)\mathbb{U}(t')] + \int_0^{t'} \frac{E [\mathbb{U}(t)\mathbb{U}(r)]}{1-r} dr + \int_0^t \frac{E [\mathbb{U}(t')\mathbb{U}(r)]}{1-r} dr \\
&\quad + \int_0^{t'} \int_0^t \frac{E [\mathbb{U}(r)\mathbb{U}(r')]}{(1-r)(1-r')} dr dr'
\end{aligned}$$

At this point it is convenient to consider the integrals separately.

$$\begin{aligned}
& \int_0^{t'} \frac{E[\mathbb{U}(t)\mathbb{U}(r)]}{1-r} dr \\
&= \int_0^t \frac{E[\mathbb{U}(t)\mathbb{U}(r)]}{1-r} dr + \int_t^{t'} \frac{E[\mathbb{U}(t)\mathbb{U}(r)]}{1-r} dr \\
&= (1-t) \int_0^t \frac{r}{1-r} dr + t(t'-t) \\
&= -t + \int_0^t \frac{1}{1-r} dr - t \int_0^t \frac{r}{1-r} dr + t(t'-t) \\
& \int_0^t \frac{E[\mathbb{U}(t')\mathbb{U}(r)]}{1-r} dr \\
&= (1-t') \int_0^t \frac{r}{1-r} dr \\
&= -t + \int_0^t \frac{1}{1-r} dr - t' \int_0^t \frac{r}{1-r} dr \\
& \int_0^{t'} \int_0^t \frac{E[\mathbb{U}(r)\mathbb{U}(r')]}{(1-r)(1-r')} dr dr' \\
&= 2 \int_0^t \int_{r'}^t \frac{r'}{(1-r')} dr dr' + \int_0^t \int_t^{t'} \frac{r'}{(1-r')} dr dr' \\
&= 2t \int_0^t \frac{r}{(1-r)} dr - 2 \int_0^t \frac{r^2}{(1-r)} dr + (t'-t) \int_0^t \frac{r}{(1-r)} dr \\
&= 2t \int_0^t \frac{r}{(1-r)} dr - 2 \int_0^t \frac{1}{(1-r)} dr + (t^2 + 2t) + (t'-t) \int_0^t \frac{r}{(1-r)} dr
\end{aligned}$$

Putting all this together gives the required result, that

$$\mathbb{U}(t) + \int_0^t \frac{\mathbb{U}(r)}{1-r} dr; \quad 0 \leq t \leq 1 \text{ is Brownian motion.}$$

□

LEMMA 1.3.25: The sequence of functions $\left\{ \frac{\sqrt{2} \sin(k\pi t)}{k\pi} \right\}_{k=0}^{\infty} = \{f_k(t)\}_{k=0}^{\infty}$ are orthogonal on $0 \leq t \leq 1$.

PROOF: Let $k \neq j$, then

$$\int_0^1 f_k(t) f_j(t) dt = \frac{2}{kj\pi^2} \int_0^1 \sin(k\pi t) \sin(j\pi t) dt$$

Using integration by parts

$$= \frac{2}{kj\pi^2} \left[-\frac{\sin(k\pi t) \cos(j\pi t)}{j\pi} \Big|_0^1 + \frac{k}{j} \int_0^1 \cos(k\pi t) \cos(j\pi t) dt \right]$$

Using integration by parts again and evaluating the first expression, letting $I = \int_0^1 f_k(t) f_j(t) dt$

$$\begin{aligned} I &= \frac{2}{kj\pi^2} \frac{k}{j} \left[\frac{\cos(k\pi t) \sin(j\pi t)}{j\pi} \Big|_0^1 + \frac{k}{j} \int_0^1 \sin(k\pi t) \sin(j\pi t) dt \right] \\ &= \frac{k^2}{j^2} I \end{aligned}$$

As $k \neq j$, $I = 0$. □

THEOREM 1.3.26: Let $Z_0, Z_1, \dots$ be i.i.d. $N(0, 1)$ random variables. Define

$$\mathbb{U}(t) \equiv \sum_{k=1}^{\infty} Z_k \frac{\sqrt{2} \sin(k\pi t)}{k\pi}; \quad 0 \leq t \leq 1$$

and

$$\mathbb{S}(t) \equiv Z_0 t + \sum_{k=1}^{\infty} Z_k \frac{\sqrt{2} \sin(k\pi t)}{k\pi}; \quad 0 \leq t \leq 1$$

then $\mathbb{U}$ is Brownian bridge and $\mathbb{S}$ is Brownian motion on $[0, 1]$. This establishes the existence of Brownian motion and Brownian bridge in 1-dimension.

PROOF:

$$\begin{aligned} &\text{Cov}(\mathbb{U}(t), \mathbb{U}(t')) \\ &= E \left[\left(\sum_{k=1}^{\infty} Z_k \frac{\sqrt{2} \sin(k\pi t)}{k\pi} \right) \left(\sum_{j=1}^{\infty} Z_j \frac{\sqrt{2} \sin(j\pi t')}{j\pi} \right) \right] \end{aligned}$$

As $Z_i Z_j$ are independent for $i \neq j$ the expected value of these is 0, so

$$\begin{aligned} &= \sum_{k=1}^{\infty} E \left[Z_k^2 \frac{2 \sin(k\pi t) \sin(k\pi t')}{k^2 \pi^2} \right] \\ &= \sum_{k=1}^{\infty} \frac{2 \sin(k\pi t) \sin(k\pi t')}{k^2 \pi^2} \end{aligned}$$

which is seen to be the bivariate Fourier series of $t \wedge t' - tt'$. To show $\mathbb{S}$ is Brownian motion use LEMMA 1.3.18 as

$$\mathbb{S}(t) \equiv Z_0 t + \sum_{k=1}^{\infty} Z_k \frac{\sqrt{2} \sin(k\pi t)}{k\pi} = Z_0 t + \mathbb{U}(t)$$

□

LEMMA 1.3.27: Let $\Pi: \mathbb{S}(t) \mapsto \mathbb{S}(t) - t\mathbb{S}(1)$ and $\Upsilon: \mathbb{U}(t) \mapsto (1+t)\mathbb{U}\left(\frac{t}{1+t}\right)$. Begin with $\mathbb{S}(t) = \mathbb{S}^0(t)$ Brownian motion and define inductively $\mathbb{U}^i(t) = \Pi(\mathbb{S}^i(t))$ for $i = 0, 1, \dots$ and $\mathbb{S}^i = \Upsilon(\mathbb{U}^{i-1}(t))$ for $i = 1, 2, \dots$. Then $\mathbb{U}^i(t)$ are Brownian bridges and $\mathbb{S}^i(t)$ are Brownian motions, with

$$\mathbb{S}^n(t) = (1+nt)\mathbb{S}\left(\frac{t}{1+nt}\right) - nt\mathbb{S}\left(\frac{1}{n}\right)$$

PROOF: Assuming that $\mathbb{S}^n(t) = (1+nt)\mathbb{S}\left(\frac{t}{1+nt}\right) - nt\mathbb{S}\left(\frac{1}{n}\right)$ then

$$\begin{aligned} \mathbb{U}^n(t) &= \Pi(\mathbb{S}^n(t)) \\ &= (1+nt)\mathbb{S}\left(\frac{t}{1+nt}\right) - nt\mathbb{S}\left(\frac{1}{n}\right) - (n+1)t\mathbb{S}\left(\frac{1}{n+1}\right) + nt\mathbb{S}\left(\frac{1}{n}\right) \\ &= (1+nt)\mathbb{S}\left(\frac{t}{1+nt}\right) - (n+1)t\mathbb{S}\left(\frac{1}{n+1}\right) \end{aligned}$$

$$\begin{aligned}
\mathbb{S}^{n+1}(t) &= \Upsilon(\mathbb{U}^n(t)) \\
&= (1+t) \left(1 + \frac{nt}{1+t}\right) \mathbb{S}\left(\frac{\frac{t}{1+t}}{1 + \frac{nt}{1+t}}\right) - (1+t)(n+1) \frac{t}{1+t} \mathbb{S}\left(\frac{1}{n+1}\right) \\
&= \mathbb{S}^{n+1}(t) = (1+(n+1)t) \mathbb{S}\left(\frac{t}{1+(n+1)t}\right) - (n+1)t \mathbb{S}\left(\frac{1}{n+1}\right)
\end{aligned}$$

Using LEMMA 1.3.11 and LEMMA 1.3.10 completes the proof by induction. $\square$

1.3.2 REFLECTION PRINCIPLES

The following subsection follows SHORACK & WELLNER (1986).

THEOREM 1.3.28: Let $\mathbb{S}$ be Brownian motion and T_x be the first moment the Brownian motion crosses level x . Then for $b > 0$

$$\mathbb{P}\{T_b < t\} = \int_0^t \frac{b}{\sqrt{2\pi s^3}} \exp\left(-\frac{b^2}{2s}\right) ds$$

PROOF: This proof follows the proof given in FELLER (1971a).

$$\mathbb{P}\{T_b < t\} = \mathbb{P}\{T_b < t, \mathbb{S}(t) > b\} + \mathbb{P}\{T_b < t, \mathbb{S}(t) \leq b\}$$

As $\mathbb{S}(t)$ is continuous, $\mathbb{P}\{T_b < t, \mathbb{S}(t) > b\} = \mathbb{P}\{\mathbb{S}(t) > b\}$. Consider the second term. At τ , using the strong Markov property we have a new Brownian motion $(\mathbb{S}(t) - \mathbb{S}(\tau))$ with respect to the starting point (τ, b) . Therefore the probability of $\mathbb{S}(t) \leq b$ is $\frac{1}{2}$, the same as the probability of Brownian motion starting at 0 being less than 0 at any given $t \geq 0$.

$$\begin{aligned}
&= \mathbb{P}\{\mathbb{S}(t) > b\} + \frac{1}{2}\mathbb{P}\{T_b < t\} \\
\Rightarrow \mathbb{P}\{T_b < t\} &= 2\mathbb{P}\{\mathbb{S}(t) > b\} \\
&= 2\left(1 - \Phi\left(\frac{b}{\sqrt{t}}\right)\right) \\
&= 2\int_0^t \frac{d}{ds}\left(1 - \Phi\left(\frac{b}{\sqrt{s}}\right)\right) ds \\
&= 2\int_0^t \frac{b}{2\sqrt{2\pi}s^3} \exp\left(-\frac{b^2}{2s}\right) ds = \int_0^t \frac{b}{\sqrt{2\pi}s^3} \exp\left(-\frac{b^2}{2s}\right) ds
\end{aligned}$$

□

THEOREM 1.3.29: Let $\mathbb{S}$ be Brownian motion. Then for $b > 0$

$$\begin{aligned}
\mathbb{P}\left\{\sup_{0 \leq s \leq t} |\mathbb{S}(s)| \geq b\right\} &= 4 \sum_{k=0}^{\infty} \mathbb{P}\{(4k-3)b < N(0, t) < (4k+1)b\} \\
&= 1 - \sum_{k=-\infty}^{\infty} (-1)^k \mathbb{P}\{(2k-1)b < N(0, t) < (2k+1)b\}
\end{aligned}$$

PROOF: To prove the main result it is first required to define

$$\begin{aligned}
\{A_+\} &\equiv \{\mathbb{S} \text{ crosses the level } b \text{ in } [0, t]\} \\
\{A_-\} &\equiv \{\mathbb{S} \text{ crosses the level } -b \text{ in } [0, t]\}
\end{aligned}$$

This allows us to inductively define

$$\begin{aligned}
\{A_{+-}\} &\equiv \{\mathbb{S} \text{ crosses the level } b \text{ then crosses level } -b \text{ in } [0, t]\} \\
\{A_{-+}\} &\equiv \{\mathbb{S} \text{ crosses the level } -b \text{ then crosses level } +b \text{ in } [0, t]\}
\end{aligned}$$

and similarly for $A_{+-+}, A_{-+-}, A_{+--+}, \dots$. The idea is to expand out the event $\left\{ \sup_{0 \leq s \leq t} |\mathbb{S}(s)| \geq b \right\}$ in terms of the events defined above. Clearly

$$\mathbb{P} \left\{ \sup_{0 \leq s \leq t} |\mathbb{S}(s)| \geq b \right\} < \mathbb{P} \{A_+\} + \mathbb{P} \{A_-\}$$

as the event where $\mathbb{S}$ crosses both b and $-b$ is counted twice. So then

$$\mathbb{P} \left\{ \sup_{0 \leq s \leq t} |\mathbb{S}(s)| \geq b \right\} > \mathbb{P} \{A_+\} - \mathbb{P} \{A_{+-}\} + \mathbb{P} \{A_-\} - \mathbb{P} \{A_{-+}\}$$

This leads us to conclude that

$$\begin{aligned} \mathbb{P} \left\{ \sup_{0 \leq s \leq t} |\mathbb{S}(s)| \geq b \right\} &= \mathbb{P} \{A_+\} - \mathbb{P} \{A_{+-}\} + \mathbb{P} \{A_{++}\} - \dots \\ &\quad + \mathbb{P} \{A_-\} - \mathbb{P} \{A_{-+}\} + \mathbb{P} \{A_{--}\} - \dots \end{aligned}$$

By symmetry

$$= 2 (\mathbb{P} \{A_+\} - \mathbb{P} \{A_{+-}\} + \mathbb{P} \{A_{++}\} - \dots)$$

Using the idea in FIGURE 1.4 and THEOREM 1.3.28

$$= 2 \sum_{k=1}^{\infty} (-1)^{k+1} 2\mathbb{P} \{N(0, t) > (2k-1)b\}$$

By considering the area remaining under the graph of the normal density after each k in FIGURE 1.5

$$\begin{aligned} &= 4 \sum_{k=1}^{\infty} \mathbb{P} \{(4k-3)b < N(0, t) < (4k-1)b\} \\ &= 1 - \sum_{k=-\infty}^{\infty} (-1)^k \mathbb{P} \{(2k-1)b < N(0, t) < (2k+1)b\} \end{aligned}$$

□

FIGURE 1.4: The red line is the reflection of the Brownian motion in the line $y = b$ after the Brownian motion hits b . The blue line is the reflection of the red line in the line $y = 3b$ after the Brownian motion hits $y = -b$. This justifies why $\mathbb{S}(t) \geq b$ and $\mathbb{S}(t) \leq -b$ is the same as $\mathbb{S}(t) \geq 3b$. An extension of this idea to when the Brownian motion repeatedly crosses the levels b and $-b$ is used in THEOREM [1.3.29](#)

FIGURE 1.5: $\sum_{k=1}^{\infty} \mathbb{P}\{(4k-3)b < N(0,1) < (4k-1)b\}$ is the blue area on the positive half axis. By symmetry this is equal to the blue area in the negative half axis and red area = 1 - blue area. Using symmetry and simplifying we get $2 \sum_{k=1}^{\infty} (-1)^{k+1} 2\mathbb{P}\{N(0,1) > (2k-1)b\} = 1 - \text{red area} + \text{blue area}$.

THEOREM 1.3.30: Let $\mathbb{S}(t)$ be Brownian motion. For $a \geq 0$ and $b > 0$

$$\mathbb{P}\left\{\sup_{t \geq 0} \frac{\mathbb{S}(t)}{at+b} \geq 1\right\} = \exp(-2ab)$$

PROOF: We first note some properties. Let $\phi(a, b) = \mathbb{P}\left\{\sup_{t \geq 0} \mathbb{S}(t) \geq at + b\right\}$. Then $\phi(a, b) \geq \mathbb{P}\{\mathbb{S}(t) \geq a + b\} > 0$ for all $a \geq 0$ and $b > 0$, $\phi(a, b)$ decreases in b and $\phi(a, b_1 + b_2) = \phi(a, b_1)\phi(a, b_2)$, which is shown in FIGURE 1.6. The only solution of the functional equation $\phi(a, b_1 + b_2) = \phi(a, b_1)\phi(a, b_2)$ in b is

$$\phi(a, b) = \exp(-\psi(a)b)$$

$$\begin{aligned} \exp(-\psi(a)b) &= \phi(a, b) \\ &= E_{\tau}[\phi(a, a\tau)] \end{aligned}$$

FIGURE 1.6: At the instant τ when $\mathbb{S}(\tau) = a\tau + b_1$, we have the same problem again, but with $\phi(a, b_2)$ relative to the point $(\tau, \mathbb{S}(\tau))$

FIGURE 1.7: Relative to (τ, b) , the sloping line has equation $y = at + a\tau$, where τ is the first instant $\mathbb{S}(t) = b$.

Using THEOREM 1.3.28

$$= \int_0^\infty \exp(-\psi(a)as) \frac{b}{\sqrt{2\pi s^3}} \exp\left(-\frac{b^2}{2s}\right) ds$$

Let $y^2 = \frac{b^2}{2s}$

$$= \frac{2}{\sqrt{\pi}} \int_0^\infty \exp\left(-\left(y^2 + \frac{c}{y^2}\right)\right) dy$$

where $c = \frac{b^2 a \psi(a)}{2}$.

$$= \exp\left(-b\sqrt{2a\psi(a)}\right)$$

Therefore $\psi(a) = 2a$ and

$$\mathbb{P}\left\{\sup_{t \geq 0} \frac{\mathbb{S}(t)}{at + b} \geq 1\right\} = \exp(-2ab)$$

□

THEOREM 1.3.31: Let $\mathbb{U}$ be Brownian bridge. For all $b > 0$

$$\mathbb{P}\left\{\sup_{0 \leq t \leq 1} \mathbb{U}(t) > b\right\} = \exp(-2b^2)$$

PROOF: Using LEMMA 1.3.14

$$\begin{aligned} \mathbb{P}\left\{\sup_{0 \leq t \leq 1} \mathbb{U}(t) > b\right\} &= \mathbb{P}\left\{\sup_{0 \leq t \leq 1} (1-t)\mathbb{S}\left(\frac{t}{1-t}\right) > b\right\} \\ &= \mathbb{P}\left\{\sup_{r > 0} \frac{\mathbb{S}(r)}{1+r} > b\right\} \end{aligned}$$

Using THEOREM 1.3.30

$$= \exp(-2b^2)$$

□

THEOREM 1.3.32: Let $\mathbb{S}(t)$ be Brownian motion. Then for $a \geq 0$ and $b > 0$

$$\mathbb{P} \left\{ \sup_{t \geq 0} \frac{|\mathbb{S}(s)|}{at + b} \geq 1 \right\} = 2 \sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2 ab)$$

PROOF: Following **DOOB (1949)**, consider the more general problem of Brownian motion intersecting a sequence of lines. Let $\{[u_n, v_n]\}_{n=1}^{\infty}$ be a sequence of lines in the form $u_n t + v_n$ for $u_n \geq 0, v_n \geq 0$. Let t_1 be the first instant that $\mathbb{S}(t)$ intersects $[u_1, v_1]$, if such a point exists; if t_1 is defined, then define t_2 to be the first instant $t > t_1$ that $\mathbb{S}(t)$ intersects $[-u_2, -v_2]$. Define $t_3, t_4, \dots$ similarly. Let π_n be the probability that t_n exists, that is, $\mathbb{S}(t)$ meets the lines $[u_1, v_1], [-u_2, -v_2], \dots, [(-1)^{n+1}u_n, (-1)^{n+1}v_n]$ in at least n successive points. We aim to show that π_n can be calculated from π_{n-1} for any n by changing the gradient and intercepts of the lines, so to show the dependence on the underlying parameters u_i and v_i the notation

$$\pi_n = \pi_n(u_1, v_1; \dots; u_n, v_n)$$

will be used.

Let Q be the point $(t_{n-1}, \mathbb{S}(t_{n-1}))$ and assume that n is even. The argument for n being odd follows by changing signs. From **FIGURE 1.3.2** we see that starting at Q we must hit a line $[-u_n, -u_{n-1}t_n - 1 - v_{n-1} - u_n t_{n-1} - v_n]$, which has probability, from **THEOREM 1.3.30**,

$$\exp(-2u_n(u_{n-1}t_{n-1} + v_{n-1} + u_n t_{n-1} + v_n)) \quad (1.8)$$

We can replace this line by any other line that has the same probability, still depending on t_{n-1} . Choose the line with slope $-(u_{n-1} + u_n)$ which is

FIGURE 1.8: Assuming Q is the first intersection of the Brownian motion and the line $u_{n-1}t + v_{n-1}$, the red line is the line with the same probability of intersection as $u_nt + v_n$ relative to the point Q , while the blue line is the red line reflected in the dashed line intersecting Q parallel to the t axis.

$\frac{u_n}{u_{n-1} + u_n} (u_{n-1}t_{n-1} + v_{n-1} + u_nt_{n-1} + v_n)$ units below Q at t_{n-1} . Using the strong Markov property we may reflect this line in the line parallel to the t axis going through the point Q . The reflected line has parameters

$$\left[u_{n-1} + u_n, \frac{u_{n-1}v_{n-1} + u_nv_n + 2u_nv_{n-1}}{u_{n-1} + u_n} \right],$$

so the reflected line is independent of t_{n-1} . Therefore

$$\begin{aligned} \pi_n(u_1, v_1; \dots; u_n, v_n) \\ = \pi_{n-1} \left(u_1, v_1; \dots; u_{n-2}, v_{n-2}; u_{n-1} + u_n, \frac{u_{n-1}v_{n-1} + u_nv_n + 2u_nv_{n-1}}{u_{n-1} + u_n} \right) \end{aligned}$$

As in THEOREM 1.3.29, define events

$$\{A_+\} \equiv \{\mathbb{S} \text{ crosses the level } at + b \text{ for } t \geq 0\}$$

$$\{A_-\} \equiv \{\mathbb{S} \text{ crosses the level } -at - b \text{ for } t \geq 0\}$$

This allows us to inductively define

$$\{A_{+-}\} \equiv \{\mathbb{S} \text{ crosses the level } at + b \text{ then crosses level } -at - b \text{ for } t \geq 0\}$$

$$\{A_{-+}\} \equiv \{\mathbb{S} \text{ crosses the level } -at - b \text{ then crosses level } at + b \text{ for } t \geq 0\}$$

and similarly for A_{++} , A_{--} , $\dots$

Therefore

$$\begin{aligned} \mathbb{P} \left\{ \sup_{t \geq 0} \frac{|\mathbb{S}(t)|}{at + b} \geq 1 \right\} &= 2 (\mathbb{P} \{A_+\} - \mathbb{P} \{A_{+-}\} + \mathbb{P} \{A_{++}\} - \dots) \\ &= 2 \sum_{k=1}^{\infty} (-1)^{k+1} \pi_k(\underbrace{a, b; a, b; \dots; a, b}_{k \text{ times}}) \\ &= 2 \sum_{k=1}^{\infty} (-1)^{k+1} \pi_1(ka, kb) \end{aligned}$$

Using THEOREM 1.3.30

$$= 2 \sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2 ab)$$

□

THEOREM 1.3.33: Let $\mathbb{U}$ be Brownian bridge. Then for all $b > 0$

$$\mathbb{P} \left\{ \sup_{0 \leq t \leq 1} |\mathbb{U}(t)| \geq b \right\} = 2 \sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2 b^2)$$

PROOF:

$$\mathbb{P} \left\{ \sup_{0 \leq t \leq 1} |\mathbb{U}(t)| \geq b \right\} = \mathbb{P} \left\{ \sup_{r > 0} \frac{|\mathbb{S}(r)|}{1+r} \geq b \right\}$$

Using THEOREM 1.3.32

$$= 2 \sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2 b^2)$$

□

1.4 CENTRAL LIMIT THEOREM FOR EMPIRICAL PROCESSES

This section outlines central limit theorem type results for the empirical process. We now are dealing with convergence of processes, so we need to consider the theory of convergence in metric spaces to prove the central limit theorem for empirical processes. The first four subsections introduce the theory of convergence in metric spaces, the fifth subsection consider a first application of convergence in metric spaces to prove Donsker's theorem, while showing the existence of the Weiner measure and the final subsection outlines results relevant to empirical processes.

1.4.1 THEORY OF WEAK CONVERGENCE

We will begin by outlining the theory of weak convergence. This is a generalisation of convergence in distribution to metric spaces.

DEFINITION 1.4.1: Let Ω be any set. A *metric space* (Ω, ρ) is a pair consisting of a set Ω and a metric $\rho: \Omega \times \Omega \mapsto \mathbb{R}$ such that:

- (a) $\rho(x, y) = 0$ iff $x = y$ for all $x, y \in \Omega$.
- (b) $\rho(x, y) = \rho(y, x)$ for all $x, y \in \Omega$.
- (c) $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ for all $x, y, z \in \Omega$.

DEFINITION 1.4.2: A metric space (Ω, ρ) is *complete* if every Cauchy (fundamental) sequence in Ω converges to a limit in Ω .

DEFINITION 1.4.3: A set $B \subseteq \Omega$ is *compact* if every open covering of B has a finite subcovering.

Let (S, ρ) be a metric space with $\mathcal{S}$ the Borel σ -field, generated by the open sets.

DEFINITION 1.4.4: A probability measure $\mathbb{P}$ on $(S, \mathcal{S})$ is *tight* if $\forall \epsilon > 0$ there exists compact set K such that $\mathbb{P}\{K\} > 1 - \epsilon$.

DEFINITION 1.4.5: If probability measures $\mathbb{P}_n$ and $\mathbb{P}$ satisfy $\mathbb{P}_n f \rightarrow \mathbb{P}f$ for every bounded, continuous real function f on S , then $\mathbb{P}_n$ *converges weakly* to $\mathbb{P}$, denoted $\mathbb{P}_n \Rightarrow \mathbb{P}$.

THEOREM 1.4.6: The following are equivalent:

- (i) $\mathbb{P}_n \Rightarrow \mathbb{P}$.
- (ii) $\mathbb{P}_n f \rightarrow \mathbb{P}f$ for all bounded, uniformly continuous f .
- (iii) $\limsup_{n \rightarrow \infty} \mathbb{P}_n F \leq \mathbb{P}F$ for all closed F .
- (iv) $\liminf_{n \rightarrow \infty} \mathbb{P}_n G \geq \mathbb{P}G$ for all open G .
- (v) $\mathbb{P}_n\{A\} \rightarrow \mathbb{P}\{A\}$ for all $\mathbb{P}$ -continuity sets A (sets where $\mathbb{P}\{\partial A\} = 0$).

PROOF:

(i) $\Rightarrow$ (ii) : If it holds for bounded and continuous f , then it clearly holds for bounded and uniformly continuous f .

(ii) $\Rightarrow$ (iii) : Choose $f(x) = \left(1 - \frac{\rho(x, F)}{\epsilon}\right) \vee 0$. It is bounded between 0 and 1 and is uniformly continuous as $|f(x) - f(y)| \leq \frac{\rho(x, y)}{\epsilon}$. Let

$F_\epsilon = \{x: \rho(x, F) < \epsilon\}$. Then it follows, as $x \in F$ implies $f(x) = 1$ and $x \notin F_\epsilon$ implies $f(x) = 0$,

$$\mathbb{I}_{\{F\}}(x) \leq f(x) \leq \mathbb{I}_{\{F_\epsilon\}}(x).$$

$$\limsup_{n \rightarrow \infty} \mathbb{P}_n F \leq \limsup_{n \rightarrow \infty} \mathbb{P}_n f = \mathbb{P} f \leq \mathbb{P} F_\epsilon$$

As F is closed, letting $\epsilon \downarrow 0$ gives the required result.

(iii) $\Rightarrow$ (iv) : Follows immediately by complementation of (iii).

(iv) & (iii) $\Rightarrow$ (v) : Let A° and A^- be the interior (largest open subset of A) and closure (smallest closed superset of A), it follows that

$$\begin{aligned} \mathbb{P}\{A^\circ\} &\leq \liminf_{n \rightarrow \infty} \mathbb{P}_n\{A^\circ\} \leq \liminf_{n \rightarrow \infty} \mathbb{P}_n\{A\} \\ &\leq \limsup_{n \rightarrow \infty} \mathbb{P}_n\{A\} \leq \limsup_{n \rightarrow \infty} \mathbb{P}_n\{A^-\} \leq \mathbb{P}\{A^-\} \end{aligned}$$

If A is a $\mathbb{P}$ -continuity set, then $\mathbb{P}\{A^\circ\} = \mathbb{P}\{A^-\}$ and (v) follows.

(v) $\Rightarrow$ (i) : By linearity we may assume that the bounded f satisfies $0 < f < 1$. So $\mathbb{P}f = \int_0^1 \mathbb{P}\{f > t\} dt$ and $\mathbb{P}_n f = \int_0^1 \mathbb{P}_n\{f > t\} dt$. As f is continuous, $\partial\{f > t\} \subseteq \{f = t\}$ and $\{f > t\}$ is a $\mathbb{P}$ -continuity set except for countably many t . From (v) it follows that $\mathbb{P}_n\{f > t\} \rightarrow \mathbb{P}\{f > t\}$ so using the bounded convergence theorem

$$\mathbb{P}_n f = \int_0^1 \mathbb{P}_n\{f > t\} dt \rightarrow \int_0^1 \mathbb{P}\{f > t\} dt = \mathbb{P}f$$

□

Let h map S to another metric space S' , with metric ρ' and Borel σ -field $\mathcal{S}'$. If h is measurable $\mathcal{S}/\mathcal{S}'$ ($A' \in \mathcal{S}'$ implies $A \in \mathcal{S}$), then each $\mathbb{P}$ on $(S, \mathcal{S})$ induces a probability measure $\mathbb{P}h^{-1}$ (defined by $\mathbb{P}h^{-1}\{A'\} = \mathbb{P}\{h^{-1}A'\}$).

THEOREM 1.4.7: If h is continuous, then $\mathbb{P}_n \Rightarrow \mathbb{P}$ implies $\mathbb{P}_n h^{-1} \Rightarrow \mathbb{P}h^{-1}$.

PROOF: Let f be bounded and continuous on S' , then as h is continuous, it follows that fh is continuous and bounded on S , so by change of variable it follows that

$$\begin{aligned} \int_{S'} f(y) d\mathbb{P}_n h^{-1}(y) &= \int_S f(h(x)) d\mathbb{P}_n(x) \\ &\xrightarrow{\mathbb{P}_n \Rightarrow \mathbb{P}} \int_S f(h(x)) d\mathbb{P}(x) = \int_{S'} f(y) d\mathbb{P} h^{-1}(y). \end{aligned}$$

□

1.4.2 CONVERGENCE IN DISTRIBUTION

We are considering mappings X from a probability space $(\Omega, \mathcal{F}, P)$ to a metric space S . X is a random element if it is measurable $\mathcal{F}/\mathcal{S}$ ($S \in \mathcal{S}$ implies $X^{-1}S \in \mathcal{F}$). The distribution of X is the probability measure $\mathbb{P} = PX^{-1}$ on $(S, \mathcal{S})$ defined for $A \in \mathcal{S}$ by

$$\mathbb{P}\{A\} = P\{X^{-1}(A)\} = P\{\omega: X(\omega) \in A\}.$$

This is also called the *law* of X , denoted $\mathcal{L}(X)$.

DEFINITION 1.4.8: A sequence $\{X_n\}_{n=1}^{\infty}$ of random elements *converges in distribution* to a random element X if $\mathbb{P}_n \Rightarrow \mathbb{P}$, where $\mathbb{P}_n$ and $\mathbb{P}$ are the distribution functions of X_n and X . This is denoted $X_n \xrightarrow{d} X$ and is also sometimes denoted $X_n \Rightarrow X$.

We require $X, X_1, X_2, \dots$ to map into the same metric space (S) with the same topology but the underlying probability spaces may all be different. This is because we are concerned with the distributions which the background probability spaces induce on S . As

$$E[f(X)] = \int_{\Omega} f(X(\omega)) dP(\omega) = \int_S f(x) d\mathbb{P}(x) = \mathbb{P}f$$

letting E_n and E denote expectation on $(\Omega_n, \mathcal{F}_n, P_n)$ and $(\Omega, \mathcal{F}, P)$, the background probability spaces for X_n and X , it follows that $\mathbb{P}_n f \Rightarrow \mathbb{P} f$ if and only if $E_n[f(X_n)] \rightarrow E[f(X)]$. Letting P and E refer to probabilities and expectations in the background probability space we are working in, it follows that $X_n \xrightarrow{d} X$ if and only if $E[f(X_n)] \rightarrow E[f(X)]$.

COROLLARY 1.4.9: The following are equivalent:

- (i) $X_n \xrightarrow{d} X$.
- (ii) $E[f(X_n)] \rightarrow E[f(X)]$ for all bounded, uniformly continuous f .
- (iii) $\limsup_{n \rightarrow \infty} P\{X_n \in F\} \leq P\{X \in F\}$ for all closed F .
- (iv) $\liminf_{n \rightarrow \infty} P\{X_n \in G\} \geq P\{X \in G\}$ for all open G .
- (v) $P\{X_n \in A\} \rightarrow P\{X \in A\}$ for all X -continuity sets A (sets where $P\{X \in \partial A\} = 0$).

PROOF: Follow directly from THEOREM 1.4.6 applied to the mapping X . $\square$

DEFINITION 1.4.10: Let a be an element of S . If, for all $\epsilon > 0$

$$P\{\rho(X_n, a) < \epsilon\} \rightarrow 1$$

then X_n converges in probability to a . This is denoted $X_n \xrightarrow{P} a$

RESULT 1.4.11: X_n converges in probability to a if and only if $X_n \xrightarrow{d} a$

PROOF: Given X_n converges in probability to a , let G be open and $a \in G$. For ϵ small enough,

$$\liminf_n P\{X_n \in G\} \geq \lim_n P\{\rho(X_n, a) < \epsilon\} = 1 = P\{a \in G\}.$$

Using COROLLARY 1.4.9 (iv), $X_n \xrightarrow{d} a$.

Now let $X_n \xrightarrow{d} a$. The set $G_2 = \{x: \rho(x, a) < \epsilon\}$ is open, so it follows that

$$\liminf_n P \{X_n \in G_2\} \geq P \{a \in G_2\} = 1.$$

Therefore it follows that $P \{\rho(X_n, a) < \epsilon\} \rightarrow 1$. □

Let (X_n, Y_n) be a random element of $S \times S$ and consider the random variable $\rho(X_n, Y_n)$ which takes value $\rho(X_n(\omega), Y_n(\omega))$ at point ω .

THEOREM 1.4.12: Let (X_n, Y_n) be random elements of $S \times S$. If $X_n \Rightarrow X$ and $\rho(X_n, Y_n) \xrightarrow{d} 0$, then $Y_n \Rightarrow X$.

PROOF: Let $F_\epsilon = \{x: \rho(x, F) \leq \epsilon\}$. Then either $Y_n \in F$ and $\rho(X_n, Y_n) < \epsilon$ which implies $X_n \in F$, or $\rho(X_n, Y_n) \geq \epsilon$

$$\mathbb{P} \{Y_n \in F\} \leq \mathbb{P} \{\rho(X_n, Y_n) \geq \epsilon\} + \mathbb{P} \{X_n \in F_\epsilon\}.$$

As F_ϵ is closed, using COROLLARY 1.4.9 (iii)

$$\limsup_{n \rightarrow \infty} \mathbb{P} \{Y_n \in F\} \leq \limsup_{n \rightarrow \infty} \mathbb{P} \{X_n \in F_\epsilon\} \leq \mathbb{P} \{X \in F_\epsilon\}.$$

If F is closed, then $F_\epsilon \downarrow F$ as $\epsilon \downarrow 0$. Using COROLLARY 1.4.9 (iii) again completes the proof. □

COROLLARY 1.4.13: Let (X, Y_n) be random elements of $S \times S$. If $\rho(X, Y_n) \xrightarrow{d} 0$, then $Y_n \xrightarrow{d} X$.

PROOF: Choose the X_n in THEOREM 1.4.12 to be the sequence $X_n = X$ for all n . □

We restrict our attention for the remainder of this subsection to random variables, where $S = \mathbb{R}$.

THEOREM 1.4.14: If $X_n \xrightarrow{d} X$, then $E|X| \leq \liminf_n E|X_n|$

PROOF: By THEOREM 1.4.7 $|X_n| \xrightarrow{d} |X|$. Using COROLLARY 1.4.9 (v) $P\{|X_n| > t\} \rightarrow P\{|X| > t\}$, which is open, so using COROLLARY 1.4.9 (iv) and Fatou's lemma

$$\begin{aligned} E|X| &= \int_0^\infty P\{|X| > t\} dt \\ &\leq \int_0^\infty \liminf_{n \rightarrow \infty} P\{|X_n| > t\} dt \\ &\leq \liminf_{n \rightarrow \infty} \int_0^\infty P\{|X_n| > t\} dt = \liminf_{n \rightarrow \infty} E|X_n| \end{aligned}$$

□

1.4.3 PROHOROV'S THEOREM

Let Π be a family of probability measures on $(S, \mathcal{S})$. We extend the definition of tightness for a probability measure to the family of probability measures in the following way.

DEFINITION 1.4.15: The family Π is *tight* if, for every ϵ there exists a compact set K such that $\mathbb{P}\{K\} > 1 - \epsilon$ for every $\mathbb{P}$ in Π .

DEFINITION 1.4.16: Π is *relatively compact* if every sequence of elements of Π contains a weakly convergent subsequence. This means that for every sequence $\{\mathbb{P}_n\}_{n=1}^\infty$ there exists a subsequence $\{\mathbb{P}_{n_i}\}_{i=1}^\infty$ and a probability measure Q , defined on $(S, \mathcal{S})$ but not necessarily an element of Π , such that $\mathbb{P}_{n_i} \xrightarrow{d} Q$ as $i \rightarrow \infty$. We also require that $Q\{S\} = 1$, so that probability cannot escape in the limit.

The reason to introduce relative compactness is that on the space $C[0, 1]$, the space of continuous functions on the unit interval, relative compactness and convergence of the finite-dimensional distributions implies weak convergence.

This idea also extends to other spaces. The idea is then to find a way of proving relative compactness. This is done by using Prohorov's theorem.

DEFINITION 1.4.17: A set A is *dense* in B , if $\forall x \in B$ such that $x \in J$ for open J , then $J \cap A \neq \emptyset$. This is equivalent to $B \subseteq A^-$

DEFINITION 1.4.18: The space S is *separable* if it contains a countable, dense subset. A subset M of S is *separable* if there is a countable set D that is dense in M ($M \subseteq D^-$)

LEMMA 1.4.19: Suppose that the subset M of S is separable. Then there is a countable class $\mathcal{A}$ of open sets with the property that, if $x \in G \cap M$ for G open, then $x \subseteq A \subseteq A^- \subseteq G$ for some $A \in \mathcal{A}$.

PROOF: Let D be a countable, dense subset of M and take $\mathcal{A}$ to be the balls $B(d, r)$ for $d \in D$ and r rational, so $\mathcal{A}$ is countable and open. If $x \in G \cap M$ and G is open, choose ϵ so that $B(x, \epsilon) \subseteq G$, which exists as G is open, then choose $d \in D$ so that $\rho(x, d) < \frac{\epsilon}{2}$, which exists as D is dense in D is dense. Finally choose a rational r so that $\rho(x, d) < r < \frac{\epsilon}{2}$. It immediately follows that $x \subseteq B(d, r) \subseteq B(d, r)^- \subseteq B(x, \epsilon) \subseteq G$ □

THEOREM 1.4.20 (Prohorov's Theorem): Let Π be a family of probability measures. Then

- (i) If Π is tight, then Π is relatively compact.
- (ii) If Π is relatively compact and S is separable and complete, then Π is tight.

PROOF:

- (i) Let $\{\mathbb{P}_n\}_{n=1}^\infty$ be a sequence in the tight family Π . Choose compact sets K_j so that $K_1 \subseteq K_2 \subseteq K_3 \cdots$ and $\mathbb{P}_j \{K_j\} > 1 - \frac{1}{j}$. The set

$\bigcup_{i=1}^n K_i$ is separable. Therefore we can construct $\mathcal{H}$ consisting of $\emptyset$ and finite unions of the sets $A^- \cap K_j$ for $j \geq 1$ and $A \in \mathcal{A}$, the class of open sets that exist by LEMMA 1.4.19. Using the diagonal method that can be found in BILLINGSLEY (1986), there exists a subsequence $\{\mathbb{P}_{n(i)}\}_{i=1}^\infty$ for which the limit $\alpha(H) = \lim_{i \rightarrow \infty} \mathbb{P}_{n(i)}\{H\}$ exists for each H in the countable class $\mathcal{H}$. Claim that the probability measure is $\mathbb{P}\{G\} = \sup_{H \subseteq G} \alpha(H)$ for all open sets G . If such a measure exists then for $H \subseteq G$ it follows that $\alpha(H) \leq \liminf_i \mathbb{P}_{n(i)}\{G\}$, which implies that $\mathbb{P}\{G\} \leq \liminf_i \mathbb{P}_{n(i)}\{G\}$, so by THEOREM 1.4.6 $\mathbb{P}_{n(i)} \xrightarrow{d} \mathbb{P}$. The remainder of the proof is to show that $\mathbb{P}$ is a probability measure. It follows as K_j has a finite covering of $\mathcal{A}$ sets by compactness, that K_j lies in $\mathcal{H}$. Therefore

$$1 \geq \mathbb{P}\{S\} \geq \sup_j \alpha(K_j) \geq \sup_u \left(1 - \frac{1}{u}\right) = 1.$$

Some measurability concerns still remain, which are dealt with in BILLINGSLEY (1999).

- (ii) Let G_n be open sets increasing to S . First it will be proved that $\forall \epsilon > 0$ there is an n such that $\mathbb{P}\{G_n\} > 1 - \epsilon$ for all $\mathbb{P}$ in Π . Assume not; then for each n there exists $\mathbb{P}_n$ in Π with $\mathbb{P}_n G_n \leq 1 - \epsilon$. As Π is relatively compact, there is a subsequence $\mathbb{P}_{n(i)} \Rightarrow \mathbb{Q}$ for some probability measure $\mathbb{Q}$. This is impossible as

$$\mathbb{Q}\{G_n\} \leq \liminf_i \mathbb{P}_{n(i)}\{G_n\} \leq \liminf_i \mathbb{P}_{n(i)}\{G_{n(i)}\} \leq 1 - \epsilon$$

but $G_n \uparrow S$. From separability there exists a sequence $A_{k_1}, A_{k_2}, \dots$ of open balls of radius $\frac{1}{k}$ which cover S . Then there exists an n_k such that for all $\mathbb{P}$ in Π , $\mathbb{P}\left\{\bigcup_{i \leq n_k} A_{ki}\right\} > 1 - \frac{\epsilon}{2^k}$ as the sequence $A_{k_1}, A_{k_2}, \dots$ covers S . The set $A = \bigcap_{k \geq 1} \bigcup_{i \leq n_k} A_{ki}$ is totally bounded

and as the space S is complete, the closure of A is compact. Let $K = A$. Then $\mathbb{P}\{K\} > 1 - \epsilon$ for all $\mathbb{P}$ in Π .

□

THEOREM 1.4.21: If the sequence $\{\mathbb{P}_n\}$ is relatively compact and the finite-dimensional distributions of $\mathbb{P}_n$ converge weakly to the finite-dimensional distributions of $\mathbb{P}$, then $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$

PROOF: Let $\mathbb{A}\pi_{t_1, \dots, t_k}^{-1}$ represent the finite-dimensional distributions of some probability measure $\mathbb{A}$. As $\{\mathbb{P}_n\}$ is relatively compact, each subsequence $\{\mathbb{P}_m\}$ contains a further subsequence $\{\mathbb{P}_{m(i)}\}$ converging weakly to some $\mathbb{Q}$. From THEOREM 1.4.7 it follows that $\mathbb{P}_{m(i)}\pi_{t_1, \dots, t_k}^{-1} \Rightarrow \mathbb{Q}\pi_{t_1, \dots, t_k}^{-1}$ and using the convergence of the finite-dimensional distributions it follows that $\mathbb{Q}\pi_{t_1, \dots, t_k}^{-1} = \mathbb{P}\pi_{t_1, \dots, t_k}^{-1}$ for all $t_1, \dots, t_k$. This implies that $\mathbb{P}_n \Rightarrow \mathbb{P}$ □

1.4.4 THE SPACE C

We now consider the space $C[0, 1]$ (abbreviated to C) with the uniform topology, with the metric

$$\rho(x, y) = \|x - y\| = \sup_{0 \leq t \leq 1} |x(t) - y(t)|.$$

Weak convergence does not follow from convergence of finite-dimensional distributions alone, but weak convergence does follow from relative compactness and convergence of finite-dimensional distributions. From Prohorov's theorem we have the following.

THEOREM 1.4.22: Let $\mathbb{P}_n, \mathbb{P}$ be probability measures on $(C, \mathcal{C})$. If the finite-dimensional distributions of $\mathbb{P}_n$ converge weakly to the finite-dimensional distributions of $\mathbb{P}$ and if $\{\mathbb{P}_n\}_{n=1}^\infty$ is tight, then $\mathbb{P}_n \Rightarrow \mathbb{P}$.

PROOF: Follows immediately from THEOREM 1.4.20 and THEOREM 1.4.21 $\square$

To use THEOREM 1.4.22 in C , we need to understand relative compactness in C . This is the statement of the Arzela-Ascoli theorem. Before the statement of the Arzela-Ascoli theorem, the modulus of continuity is defined.

DEFINITION 1.4.23: The *modulus of continuity* of an arbitrary function x (not necessarily in C) on $[0, 1]$ is

$$w_x(\delta) = \sup_{|s-t| \leq \delta} |x(s) - x(t)|, \quad 0 < \delta \leq 1.$$

THEOREM 1.4.24 (Arzela-Ascoli): The set A in C is relatively compact (each sequence in A contains a convergent subsequence) iff

$$\sup_{x \in A} |x(0)| < \infty \tag{1.9}$$

and

$$\lim_{\delta \rightarrow 0} \sup_{x \in A} w_x(\delta) = 0. \tag{1.10}$$

PROOF: Let A be relatively compact. As C is a metric space, this is equivalent to A^- is compact, which implies (1.9). Given fixed $\delta > 0$, $|w_x(\delta) - w_y(\delta)| \leq 2\rho(x, y)$, so $w_x(\delta)$ is continuous in x . As $w_x\left(\frac{1}{n}\right)$ is non-increasing in n , Dini's theorem applies, which states that for functions that are continuous and monotonically decreasing to 0, the convergence is uniform on compact sets and (1.10) follows as A^- is compact.

Now let A satisfy (1.9) and (1.10). Choose k large enough that $\sup_{x \in A} w_x\left(\frac{1}{k}\right)$ is finite. Using the triangle inequality

$$|x(t)| \leq |x(0)| + \sum_{i=1}^k \left| x\left(\frac{it}{k}\right) - x\left(\frac{(i-1)t}{k}\right) \right|$$

and it follows that

$$\sup_{0 \leq t \leq 1} \sup_{x \in A} |x(t)| < \infty. \quad (1.11)$$

We want to use (1.10) and (1.11) to prove that A is relatively compact. Let $\{t_k\}_{k=1}^\infty$ be an enumeration of the rationals in $[0, 1]$. Using (1.11) it follows that $\{x(t_k)\}_{x \in A}$ is bounded for given k , so there exists a subsequence (of a given sequence of elements in A) $\{x_{n_1}\}_{n_1=1}^\infty \in A$ such that $\{x_{n_1}(t_1)\}_{n_1=1}^\infty$ converges. Continuing for $t_2, t_3, \dots$ we find $\{x_{n_2}\}_{n_2=1}^\infty$ a subsequence of $\{x_{n_1}\}_{n_1=1}^\infty$ which converges on for t_2 and $\{x_{n_i}\}_{n_i=1}^\infty$ a subsequence of $\{x_{n_{i-1}}\}_{n_{i-1}=1}^\infty$ which converges on for t_i . Therefore the sequence of functions $\{x_n\}_{n=1}^\infty$, where x_j is the j -th function of the sequence $\{x_{n_j}\}_{n_j=1}^\infty$, converges for rational $t_k \in [0, 1]$. So there exists an N such that

$$|x_n(t_k) - x_m(t_k)| < \frac{\epsilon}{3} \quad \text{for } m, n \geq N$$

By (1.10), for given ϵ and any $t \in [0, 1]$ there exists a t_k such that $\sup_{x \in A} |x(t) - x(t_k)| < \frac{\epsilon}{3}$ when $|t - t_k| < \delta$. Further for given δ we only need a finite number of t_k , as $\bigcup_{k=1}^\infty B(t_k, \delta)$ form an open cover of $[0, 1]$, so there is a finite subcover, $B(t_{1'}, \delta), B(t_{2'}, \delta), \dots, B(t_{j'}, \delta)$. Therefore for any $t \in [0, 1]$ we have

$$\sup_{t \in [0, 1]} |x_n(t) - x_m(t)| \leq |x_n(t) - x_n(t_k)| + |x_n(t_k) - x_m(t_k)| + |x_m(t) - x_m(t_k)|$$

So for m, n greater than some finite N'

$$\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

So the sequence $\{x_n\}_{n=1}^\infty$ is Cauchy in C with respect to the uniform metric, therefore the limit is continuous as C is complete. Therefore A is relatively compact. $\square$

Now we want to consider probability measures on $(C, \mathcal{C})$.

THEOREM 1.4.25: The sequence of probability measures $\{\mathbb{P}_n\}$ is tight if and only if the following two conditions hold

(a) For all $\eta > 0$, there exists a and N such that

$$\mathbb{P}_n \{x: |x(0)| \geq a\} \leq \eta, \quad n \geq N \quad (1.12)$$

(b) For all $\eta > 0$ and $\epsilon > 0$, there exists a $0 < \delta < 1$ and an N such that

$$\mathbb{P}_n \{x: w_x(\delta) \geq \epsilon\} \leq \eta, \quad n \geq N \quad (1.13)$$

(b) is equivalent to

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}_n \{x: w_x(\delta) \geq \epsilon\} = 0$$

PROOF: Let $\{\mathbb{P}_n\}$ be tight. For given η , there is a compact K such that $\mathbb{P}_n \{K\} > 1 - \eta$ for all n . By THEOREM 1.4.24 for large enough a we have $K \subseteq \{x: |x(0)| \leq a\}$ and for small enough δ we have $K \subseteq \{x: w_x(\delta) \leq \epsilon\}$, so (a) and (b) hold for $N = 1$.

Let (a) and (b) hold. We want to show that any given probability measure on C is tight, and then extend this to the sequence of probability measures. Since C is separable, for each k there is a sequence of open $\frac{1}{k}$ balls $A_{k1}, A_{k2}, \dots$

covering C . Choose n_k large enough so that $\mathbb{P} \left\{ \bigcup_{i \leq n_k} A_{ki} \right\} > 1 - \frac{\epsilon}{2k}$. As C is

complete, the totally bounded set $\left\{ \bigcap_{k \geq 1} \bigcup_{i \leq n_k} A_{ki} \right\}$ has compact closure K ,

with $\mathbb{P} \{K\} > 1 - \epsilon$. Therefore using the necessity of the theorem, (a) and (b) hold for each $\mathbb{P}_n$ individually. By choosing a to be the maximum of a that work for the first j elements of $\{\mathbb{P}_n\}$, and choosing δ to be the minimum of

the δ that work for the first j elements of $\{\mathbb{P}_n\}$, we have that (a) and (b) hold for finitely many elements of $\{\mathbb{P}_n\}$, so we may assume that $N = 1$.

Given η , from (a) we may choose a so that, if $B = \{x: |x(0)| \leq a\}$, then $\mathbb{P}_n\{B\} \geq 1 - \eta$ for all n . Similarly, from (b) we may choose δ_k so that, if $B_k = \left\{x: w_x(\delta_k) < \frac{1}{k}\right\}$, then $\mathbb{P}_n\{B_k\} \geq 1 - \frac{\eta}{2^k}$ for all n . Let K be the closure of $A = B \cap \left(\bigcap_k B_k\right)$, then $\mathbb{P}_n\{K\} \geq 1 - 2\eta$ and from THEOREM 1.4.24 it follow that A is relatively compact and K is compact. Therefore $\{\mathbb{P}_n\}$ is tight. $\square$

THEOREM 1.4.26: Let $0 = t_0 < t_1 < \cdots < t_v = 1$ and

$$\min_{1 \leq i \leq v} (t_i - t_{i-1}) \geq \delta \quad (1.14)$$

Then for arbitrary x ,

$$w_x(\delta) \leq 3 \max_{1 \leq i \leq v} \left(\sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_i)| \right), \quad (1.15)$$

and for arbitrary $\mathbb{P}$,

$$\mathbb{P}\{x: w_x(\delta) \geq 3\epsilon\} \leq \sum_{i=1}^v \mathbb{P}\left\{x: \sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_i)| \geq \epsilon\right\} \quad (1.16)$$

PROOF: Let m be the maximum in (1.15). If s and t lie in the same interval $[t_{i-1}, t_i]$, then

$$|x(s) - x(t)| \leq |x(s) - x(t_{i-1})| + |x(t) - x(t_{i-1})| \leq 2m.$$

If s and t lie in adjacent intervals, then

$$|x(s) - x(t)| \leq |x(s) - x(t_{i-1})| + |x(t_i) - x(t_{i-1})| + |x(t) - x(t_{i-1})| \leq 3m.$$

Therefore (1.15) holds. Therefore

$$\begin{aligned} \mathbb{P}\{x: w_x(\delta) \geq 3\epsilon\} &\leq \mathbb{P}\left\{x: 3 \max_{1 \leq i \leq v} \left(\sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_i)| \right) \geq 3\epsilon\right\} \\ &\leq \sum_{i=1}^v \mathbb{P}\left\{x: \sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_i)| \geq \epsilon\right\} \end{aligned}$$

□

1.4.5 CONSTRUCTION OF THE WEINER MEASURE

We want to prove that the partial sum process weakly converges to Brownian motion. The first step is to prove the existence of the limiting process, while the second is to prove that this is indeed the correct limiting process. The ideas used in this section can then be generalised to empirical processes, defined on a suitable space to be discussed in the next subsection. This subsection follows BILLINGSLEY (1999).

DEFINITION 1.4.27: The *Wiener* measure, $\mathbb{W}$, is a probability measure on $(C, \mathcal{C})$ that satisfies the following:

- (i) Each x_t is normally distributed under $\mathbb{W}$ with mean 0 and variance t

$$\mathbb{W}\{x_t \leq a\} = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^a \exp\left(-\frac{u^2}{2t}\right) du$$

$$\text{with } \mathbb{W}\{x_t = 0\} = 1$$

- (ii) The stochastic process $\{x_t: 0 \leq t \leq 1\}$ has independent increments under $\mathbb{W}$

The Wiener measure is the measure associated with Brownian motion. What we want to do is prove the existence on $(C, \mathcal{C})$ of a probability measure having a specified finite-dimensional distribution.

DEFINITION 1.4.28: Given a sequence $\xi_1, \xi_2, \dots$ of independent and identically distributed random variables on some probability space having mean 0 and finite, positive variance σ^2 . Let $S_n = \sum_{i=1}^n \xi_i$, with $S_0 = 0$, and let $X^n(\omega)$ be an element of C defined by

$$X_t^n(\omega) = \frac{1}{\sigma\sqrt{n}}S_{[nt]}(\omega) + (nt - [nt])\frac{1}{\sigma\sqrt{n}}\xi_{[nt]+1}(\omega)$$

$X^n(\omega)$ is the linear interpolation between the values $X_{\frac{i}{n}}^n(\omega) = \frac{1}{\sigma\sqrt{n}}S_i(\omega)$ defined by the partial sums $S_i: i = 1, 2, \dots, n$.

THEOREM 1.4.29: There exists on $(C, \mathcal{C})$ a probability measure $\mathbb{W}$ having the finite-dimensional distribution specified by

$$\begin{aligned} & \mathbb{W} \left\{ x_{t_i} - x_{t_{i-1}} \leq \alpha_i : i = 1, \dots, k \right\} \\ &= \prod_{i=1}^k \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \int_{-\infty}^{\alpha_i} \exp \left(-\frac{u^2}{2(t_i - t_{i-1})} \right) du \end{aligned}$$

PROOF: The first step is to show that the finite-dimensional distributions of X^n converge to the finite-dimensional distributions of $\mathbb{W}$. Using Chebyshev's inequality $(nt - [nt])\frac{1}{\sigma\sqrt{n}}\xi_{[nt]+1}(\omega) \xrightarrow{d} 0$ as $n \rightarrow \infty$, and by the central limit theorem $\frac{1}{\sigma\sqrt{n}}S_{[nt]}(\omega) \Rightarrow \sqrt{t}N$, where N is the standard normal distribution.

Using THEOREM 1.4.12 $X_t^n \xrightarrow{d} \sqrt{t}N$. The same argument applied to the differences show that $(X_{t_1}^n, X_{t_2}^n - X_{t_1}^n, \dots, X_{t_k}^n - X_{t_{k-1}}^n)$ converge to $(N_1, N_2, \dots, N_k)$ with N_i having variance $t_i - t_{i-1}$. Letting $\mathbb{P}_n$ be the distribution on C of X^n , for each $t_1, \dots, t_k$ we have $\mathbb{P}_n \pi_{t_1, \dots, t_k} \Rightarrow \mathbb{W} \pi_{t_1, \dots, t_k}$, so the finite-dimensional distributions converge.

The second step is to show that X^n is tight, which implies that X^n is relatively compact, so $\mathbb{P}_n \xrightarrow{d} \mathbb{W}$. We will show that

$$\lim_{\lambda \rightarrow 0} \limsup_{n \rightarrow \infty} \lambda^2 P \left\{ \max_{k \leq n} |S_k| \geq \lambda \sigma \sqrt{n} \right\} = 0 \quad (1.17)$$

implies that X^n is tight. Using THEOREM 1.4.25, we see condition (a) is satisfied as $X_0^n = 0$ for all n . For condition (b) we show that, for $\epsilon > 0$

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P \{w(X^n, \delta) \geq \epsilon\} = 0.$$

Using THEOREM 1.4.26, if EQUATION (1.14) holds, then

$$P \{w(X^n, \delta) \geq 3\epsilon\} \leq \sum_{i=1}^v P \left\{ \sup_{t_{i-1} \leq s \leq t_i} |X_s^n - X_t^n| \geq \epsilon \right\} \quad (1.18)$$

Choosing $t_i = \frac{m_i}{n}$ for integer m_i satisfying $0 = m_0 < m_1 < \cdots < m_v = n$ allows us to evaluate the supremum in 1.18 as a maximum over the carefully chosen t_i :

$$\begin{aligned} P \{w(X^n, \delta) \geq 3\epsilon\} &\leq \sum_{i=1}^v P \left\{ \max_{m_{i-1} \leq k \leq m_i} \frac{|S_k - S_{m_{i-1}}|}{\sigma \sqrt{n}} \geq \epsilon \right\} \\ &= \sum_{i=1}^v P \left\{ \max_{k \leq m_i - m_{i-1}} |S_k| \geq \epsilon \sigma \sqrt{n} \right\} \quad (\text{By independence}) \end{aligned}$$

We want to further modify our choice of m_i by letting $m_i = im$ for some integer m , where m is a function of n and δ . Choose m as $m \lceil n\delta \rceil$ for $i < v$ and choose $v = \lceil \frac{n}{m} \rceil$, then

$$m_v - m_{v-1} \leq m, \quad v = \lceil \frac{n}{m} \rceil \xrightarrow{n} \frac{1}{\delta} < \frac{2}{\delta}, \quad \frac{n}{m} \xrightarrow{n} \frac{1}{\delta} > \frac{1}{2\delta}$$

so for large n

$$\begin{aligned} P \{w(X^n, \delta) \geq 3\epsilon\} &\leq v P \left\{ \max_{k \leq m} |S_k| \geq \epsilon \sigma \sqrt{n} \right\} \\ &\leq \frac{2}{\delta} P \left\{ \max_{k \leq m} |S_k| \geq \frac{\epsilon \sigma \sqrt{m}}{\sqrt{2\delta}} \right\} \end{aligned}$$

Let $\lambda = \frac{\epsilon}{\sqrt{2\delta}}$, so $\lambda \rightarrow \infty$ as $\delta \rightarrow 0$

$$\leq \frac{4\lambda^2}{\epsilon^2} P \left\{ \max_{k \leq m} |S_k| \geq \lambda \sigma m \right\}$$

Using (1.17), for given $\epsilon > 0$ and $\eta > 0$, there exists a λ such that

$$\frac{4\lambda^2}{\epsilon^2} \limsup_{m \rightarrow \infty} P \left\{ \max_{k \leq m} |S_k| \geq \lambda \sigma m \right\} < \eta$$

Letting $m \rightarrow \infty$ implies $n \rightarrow \infty$ and letting $\lambda \rightarrow \infty$ implies $\delta \rightarrow 0$ therefore

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P \{w(X^n, \delta) \geq \epsilon\} = \lim_{\lambda \rightarrow 0} \limsup_{n \rightarrow \infty} \lambda^2 P \left\{ \max_{k \leq n} |S_k| \geq \lambda \sigma \sqrt{n} \right\} = 0$$

The last step is to show that (1.17) holds. Using Etemadi's inequality implies that

$$P \left\{ \max_{u \leq m} |S_u| \geq \alpha \right\} \leq 3 \max_{u \leq m} P \left\{ |S_u| \geq \frac{\alpha}{3} \right\}$$

so (1.17) holds if

$$\lim_{\lambda \rightarrow 0} \limsup_{n \rightarrow \infty} \lambda^2 \max_{k \leq n} P \left\{ |S_k| \geq \lambda \sigma \sqrt{n} \right\} = 0.$$

To prove existence we may use whatever sequence of ξ_i we choose, so choose ξ_i to be independent standard normal distributions. This means $\frac{S_k}{\sqrt{k}}$ also has a normal distribution and as

$$P \{|N| \geq \lambda\} < \frac{E[N^4]}{\lambda^4} = \frac{3}{\lambda^4}$$

we have

$$P \left\{ |S_k| \geq \lambda \sigma \sqrt{n} \right\} = P \left\{ \sqrt{k} |N| \geq \lambda \sigma \sqrt{n} \right\} < \frac{3}{\lambda^4 \sigma^4}$$

which implies (1.17), which implies that X^n is tight, so using THEOREM 1.4.22, we have proved the existence of the Wiener measure. $\square$

We have shown that for ξ_i standard normal that the Wiener measure exists,

but it is well-known that X^n converges to a Wiener measure for any sequence of independent and identically distributed ξ_i with mean 0 and finite, positive variance σ^2 . This is the statement of Donsker's theorem, which is proved below.

THEOREM 1.4.30 (Donsker's Theorem): If $\xi_1, \xi_2, \dots$ are independent and identically distributed with mean 0 and finite, positive variance σ^2 , then if $\mathbb{P}_n$ is the probability measure associated with X^n , then $\mathbb{P}_n \xrightarrow{d} \mathbb{W}$

PROOF: Following from the proof of THEOREM 1.4.29 all that remains is to show that

$$\lim_{\lambda \rightarrow 0} \limsup_{n \rightarrow \infty} \lambda^2 \max_{k \leq n} P \{ |S_k| \geq \lambda \sigma \sqrt{n} \} = 0 \quad (1.19)$$

for $\xi_1, \xi_2, \dots$ not necessarily standard normal. Using the central limit theorem, if k_λ is large enough, then for $k_\lambda \leq k \leq n$ then

$$P \{ |S_k| \geq \lambda \sigma \sqrt{n} \} \leq P \{ |S_k| \geq \lambda \sigma \sqrt{k} \} < \frac{3}{\lambda^4}$$

For $k < k_\lambda$ we can use Chebyshev's inequality to get

$$P \{ |S_k| \geq \lambda \sigma \sqrt{n} \} \leq \frac{k_\lambda \sigma^2}{\lambda^2 \sigma^2 n} = \frac{k_\lambda}{\lambda^2 n}$$

Therefore the maximum in (1.19) is bounded by $\left(\frac{3}{\lambda^4} \right) \vee \left(\frac{k_\lambda}{\lambda^2 n} \right)$ □

1.4.6 CENTRAL LIMIT THEOREM FOR EMPIRICAL PROCESSES

We begin by outlining the central limit theorem for empirical processes in $\mathbb{R}$.

LEMMA 1.4.31: For $r = 1, 2, \dots$ and any choice of $0 \leq t_1 < t_2 < \dots < t_r \leq 1$,

$$\langle \mathbb{U}_n(t_1), \mathbb{U}_n(t_2), \dots, \mathbb{U}_n(t_r) \rangle \xrightarrow{d} \langle \mathbb{U}(t_1), \mathbb{U}(t_2), \dots, \mathbb{U}(t_r) \rangle$$

PROOF:

$$\begin{aligned}
E[\mathbb{U}(t_j)\mathbb{U}(t_k)] &= \frac{1}{n} E \left[\left(\sum_{i=1}^n (\mathbb{I}_{\{0,t_j\}}(X_i) - t_j) \right) \left(\sum_{i=1}^n (\mathbb{I}_{\{0,t_j\}}(X_i) - t_j) \right) \right] \\
&= \frac{1}{n} \sum_{i \neq l} E [\mathbb{I}_{\{0,t_j\}}(X_i) - t_j] E [\mathbb{I}_{\{0,t_k\}}(X_l) - t_k] \\
&\quad + \frac{1}{n} \sum_{i=1}^n E [(\mathbb{I}_{\{0,t_j\}}(X_i) - t_j) (\mathbb{I}_{\{0,t_j\}}(X_i) - t_j)] \\
&= E [(\mathbb{I}_{\{0,t_j\}}(X_1) - t_j) (\mathbb{I}_{\{0,t_j\}}(X_1) - t_j)] \\
&= t_j \wedge t_k - t_j t_k
\end{aligned}$$

Therefore the covariance matrices of $\mathbb{U}_n$ and $\mathbb{U}$ are the same, so using the multivariate central limit theorem it follows

$$\langle \mathbb{U}_n(t_1), \mathbb{U}_n(t_2), \dots, \mathbb{U}_n(t_r) \rangle \xrightarrow{d} \langle \mathbb{U}(t_1), \mathbb{U}(t_2), \dots, \mathbb{U}(t_r) \rangle$$

□

To show weak convergence of the uniform empirical process to Brownian bridge requires an extension to the theory developed so far in this section. The empirical process $\mathbb{U}_n(t)$ is not continuous, so we cannot apply the theory of the space C to establish weak convergence. Following [BILLINGSLEY \(1999\)](#) we define the space D and a metric for D .

DEFINITION 1.4.32: Let $D = D[0, 1]$ be the space of real functions x on $[0, 1]$ that are right-continuous and have left-hand limits:

- (i) For $0 \leq t < 1$, $x(t^+) = \lim_{s \downarrow t} x(s)$ exists and $x(t^+) = x(t)$.
- (ii) For $0 < t \leq 1$, $x(t^-) = \lim_{s \uparrow t} x(s)$.

Functions satisfying (i) and (ii) are called *cadlag* functions.

DEFINITION 1.4.33: Let Λ denote the class of strictly increasing, continuous mappings $[0, 1] \mapsto [0, 1]$. If $\lambda \in \Lambda$, then $\lambda(0) = 0$ and $\lambda(1) = 1$. Let $\|x\| = \sup_t |x(t)|$ (which is bounded $\forall x \in D$) and let I be the identity map on $[0, 1]$, then

$$d(x, y) = \inf_{\lambda \in \Lambda} \{ \|\lambda - I\| \vee \|x - y(\lambda)\| \}$$

This metric defines the Skohorod topology, and we state that under this metric the space D is separable and there exists an equivalent metric under which D is complete. The proof of the following theorem is modified from [EINMAHL \(1993\)](#).

THEOREM 1.4.34: In the space D , $\mathbb{U}_n \xrightarrow{d} \mathbb{U}$ in D , where $\mathbb{U}$ is Brownian bridge.

PROOF: For each $m \in \mathbb{N}$, define $A_m: D \mapsto D$ by

$$(A_m(x))(t) = \begin{cases} x\left(\frac{j-1}{m}\right), & \frac{j-1}{m} \leq t < \frac{j}{m} \\ x(1), & t = 1 \end{cases} \quad j = 1, \dots, m$$

Choose $f \in C_d^*(D)$, the real-valued functions on D that are bounded, uniformly continuous with respect to d and measurable. Then

$$0 \leq \|f\|_D \leq C < \infty \quad 0 \leq \delta_f(\epsilon) = \sup_{d(x,y) < \epsilon} |f(x) - f(y)| < \infty$$

$$\begin{aligned} \text{On } \{d(\mathbb{U}_n, A_m(\mathbb{U}_\infty)) < \epsilon\} \quad & |f(\mathbb{U}_n) - f(A_m(\mathbb{U}_n))| \leq \delta_d(\epsilon) \\ \text{On } \{d(\mathbb{U}_n, A_m(\mathbb{U}_\infty)) \geq \epsilon\} \quad & |f(\mathbb{U}_n) - f(A_m(\mathbb{U}_n))| \leq 2C \end{aligned}$$

with the same inequalities holding for $\mathbb{U}$. It follows that

$$\begin{aligned}
& |E[f(\mathbb{U}_n)] - E[f(\mathbb{U})]| \\
& \leq |E[f(\mathbb{U}_n)] - E[f(A_m(\mathbb{U}_n))]| + |E[f(A_m(\mathbb{U}_n))] - E[f(A_m(\mathbb{U}))]| \\
& \quad + |E[f(A_m(\mathbb{U}))] - E[f(\mathbb{U})]| \\
& \leq \delta_d(\epsilon) + 2C\mathbb{P}\{d(\mathbb{U}_n, A_m(\mathbb{U}_n)) \geq \epsilon\} + |E[f(A_m(\mathbb{U}_n))] - E[f(A_m(\mathbb{U}))]| \\
& \quad + \delta_d(\epsilon) + 2C\mathbb{P}\{d(\mathbb{U}, A_m(\mathbb{U})) \geq \epsilon\}
\end{aligned}$$

Letting $\omega_n(\iota)$ be the modulus of continuity for the uniform empirical process at stage n , $d(\mathbb{U}_n, A_m(\mathbb{U}_n)) \leq \omega_n\left(\frac{1}{m}\right)$. For $0 < \epsilon \leq \frac{1}{2}$, using THEOREM 1.1.29

$$\begin{aligned}
\mathbb{P}\{d(\mathbb{U}_n, A_m(\mathbb{U}_n)) \geq \epsilon\} & \leq \mathbb{P}\left\{\omega_n\left(\frac{1}{m}\right) \geq \epsilon\right\} \\
& \leq C'm \exp\left(\frac{-(1-\epsilon)\epsilon^2 m}{2} \psi\left(\frac{\epsilon m}{\sqrt{n}}\right)\right)
\end{aligned}$$

Therefore, using DEFINITION 1.1.23,

$$\limsup_{n \rightarrow \infty} \mathbb{P}\{d(\mathbb{U}_n, A_m(\mathbb{U}_n)) \geq \epsilon\} \leq C'm \exp\left(\frac{1}{2}(1-\epsilon)\epsilon^2 m\right) \rightarrow 0$$

as $m \rightarrow \infty$. As the sample paths of $\mathbb{U}$ are uniformly continuous on $[0, 1]$, we have $d(\mathbb{U}, A_m(\mathbb{U})) \rightarrow 0$ almost surely as $m \rightarrow \infty$. Therefore, for all $\epsilon > 0$,

$$\mathbb{P}\{d(\mathbb{U}, A_m(\mathbb{U})) \geq \epsilon\} \rightarrow 0$$

as $m \rightarrow \infty$.

Finally from LEMMA 1.4.31 it follows that $E[f(A_m(\mathbb{U}_n))] = E[f(A_m(\mathbb{U}))]$, therefore $E[f(\mathbb{U}_n)] \rightarrow E[f(\mathbb{U})]$ and by COROLLARY 1.4.9, $\mathbb{U}_n \xrightarrow{d} \mathbb{U}$. $\square$

The previous theorem can also be proved by similar ideas used for proving Donsker's theorem in C , by first deriving a theorem which generalises Arzela-Ascoli to D , then deriving a probabilistic equivalent and showing that the uniform empirical process satisfies these conditions. This approach is taken in

BILLINGSLEY (1999)

For the remainder of this subsection we are considering the $\mathcal{F}$ -indexed empirical process $\mathbb{G}_n$ introduced in SECTION 1.1.3. The remainder of this sections outlines extensions to the function parametric empirical process following **VAN DER VAART & WELLNER (1996)**. Let $f \in \mathcal{F}$, then we consider the map

$$f \mapsto \mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - \mathbb{P})f = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(f(X_i) - \int f d\mathbb{P} \right). \quad (1.20)$$

For a given $f \in \mathcal{F}$ it follows from the central limit theorem, if $\mathbb{P}f^2 < \infty$

$$\mathbb{G}_n f \xrightarrow{d} N\left(0, \mathbb{P}(f - \mathbb{P}f)^2\right)$$

The aim is to make this statement uniform over $\mathcal{F}$, that is,

$$\|\mathbb{G}_n - \mathbb{G}\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |(\mathbb{G}_n - \mathbb{G})f| \rightarrow 0$$

where $\mathbb{G}$ is an appropriate limiting process. It is assumed that $\sup_{f \in \mathcal{F}} |f(x) - \mathbb{P}f| < \infty$ for all x , which implies that $\|\mathbb{G}_n\|_{\mathcal{F}} < \infty$, so $\mathbb{G}_n$ is a bounded functional on the space $\mathcal{F}$.

DEFINITION 1.4.35: Let T be a metric space. Then $l^\infty(T)$ is defined to be the set of all uniformly bounded real functions on T . So for $f \in l^\infty(T)$, $\sup_{t \in T} |f(t)| < \infty$

This means that the empirical process $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ is an element of $l^\infty(\mathcal{F})$. We consider the nature of the limit distribution $\mathbb{G}$ by first considering the marginal distributions. Let $f_1, f_2, \dots, f_k \in \mathcal{F}$ be any finite set of functions. If $f_1, f_2, \dots, f_k$ are square integrable, then it follows from the multivariate

central limit theorem that

$$(\mathbb{G}_n f_1, \mathbb{G}_n f_2, \dots, \mathbb{G}_n f_k) \xrightarrow{d} N(\mathbf{0}, \Sigma)$$

where

$$\Sigma = \begin{bmatrix} \mathbb{P}(f_1 - \mathbb{P}f_1)^2 & \mathbb{P}f_1 f_2 - \mathbb{P}f_1 \mathbb{P}f_2 & \cdots & \mathbb{P}f_1 f_k - \mathbb{P}f_1 \mathbb{P}f_k \\ \mathbb{P}f_1 f_2 - \mathbb{P}f_1 \mathbb{P}f_2 & \mathbb{P}(f_2 - \mathbb{P}f_2)^2 & \cdots & \mathbb{P}f_2 f_k - \mathbb{P}f_2 \mathbb{P}f_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{P}f_1 f_k - \mathbb{P}f_1 \mathbb{P}f_k & \mathbb{P}f_2 f_k - \mathbb{P}f_2 \mathbb{P}f_k & \cdots & \mathbb{P}(f_k - \mathbb{P}f_k)^2 \end{bmatrix}$$

At this stage if a limit process exists, then $\{\mathbb{G}f: f \in \mathcal{F}\}$ must be a zero-mean Gaussian process with covariance

$$E[\mathbb{G}f_1 \mathbb{G}f_2] = \mathbb{P}f_1 f_2 - \mathbb{P}f_1 \mathbb{P}f_2,$$

which is the $\mathbb{P}$ -Brownian bridge. This requires tightness to establish $\mathbb{G}_n \xrightarrow{d} \mathbb{G}$ in $l^\infty(\mathcal{F})$, and such a class $\mathcal{F}$ for which this convergence is true is called a *Donsker* class.

Sufficient conditions for a class to be Donsker depend on the entropy numbers of the class, introduced in SECTION 1.1.3. A finite class of square integrable functions is always Donsker, while the class of all square integrable, uniformly bounded functions is almost never Donsker. It is shown in [VAN DER VAART & WELLNER \(1996\)](#) that

$$\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(\mathbb{P}))} d\epsilon < \infty$$

is a sufficient condition for a class $\mathcal{F}$ to be $\mathbb{P}$ -Donsker. To arrive at this requires an extension of the theory of weak convergence outlined so far of the space $l^\infty(\mathcal{F})$.

We would like to use the theory which has been developed so far, but an element of l^∞ is not necessarily continuous, so we cannot use the Arzela-Ascoli theorem used in C . However we know that Prohorov's theorem still applies, so

the idea is to characterise tightness in l^∞ in a similar way to the Arzela-Ascoli theorem. This is the idea of asymptotic equicontinuity. The condition which ensures asymptotic equicontinuity for $\mathbb{G}_n$ is, for every $\epsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{\rho_{\mathbb{P}}(f-g) < \delta} |\mathbb{G}_n(f-g)| > \epsilon \right\} = 0$$

where $\rho_{\mathbb{P}}(f) = \left(\mathbb{P}(f - \mathbb{P}f)^2 \right)^{\frac{1}{2}}$ is a seminorm.

Notice the similarity between the definition of asymptotic equicontinuity and condition (b) of THEOREM 1.4.25. This leads to the following theorem:

THEOREM 1.4.36: A class $\mathcal{F}$ is Donsker if and only if $\mathcal{F}$ is totally bounded in $\mathcal{L}_2(\mathbb{P})$ and is asymptotically equicontinuous.

PROOF: Can be found in, for example, VAN DER VAART & WELLNER (1996). □

COLLECTION OF RESEARCH PROBLEMS

2.1 FOURIER ANALYSIS OF MIXTURES

2.1.1 CHARACTERISTIC FUNCTIONS

DEFINITION 2.1.1: Let X be a random variable with distribution function F . The characteristic function of X (or F) is the function ψ defined for real t by

$$\psi(t) = E [\exp(itX)] = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

It can equivalently be defined as

$$\psi(t) = u(t) + iv(t)$$

where

$$u(t) = \int_{-\infty}^{\infty} \cos(tx) dF(x) \quad v(t) = \int_{-\infty}^{\infty} \sin(tx) dF(x)$$

For distributions F with a density f this is

$$\psi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx$$

LEMMA 2.1.2: Let $\psi(t)$ be the characteristic function of the random variable X with distribution F . Then

1. $\psi(t)$ is uniformly continuous
2. For constants a, b , $aX + b$ has the characteristic function

$$E [\exp(it(aX + b))] = e^{ibt} \psi(at)$$

PROOF:

1. It is useful to note that $|e^{itx}| = 1$, where $|(.)|$ denotes the complex modulus. Therefore

$$\begin{aligned} |e^{it(x+h)} - e^{itx}| &\leq |e^{itx}| |e^{ith} - 1| \\ &= |e^{ith} - 1| \\ &= \left| ith + \frac{(ith)^2}{2!} + \dots \right| = (ht)^2 + o(h^3) \end{aligned}$$

So for h sufficiently small this tends to 0. Therefore $\psi(t)$ is uniformly continuous.

2.

$$E [\exp(it(aX + b))] = e^{ibt} E [\exp(it(aX))] = e^{ibt} E [\exp(i(at)X)] = e^{ibt} \psi(at)$$

□

LEMMA 2.1.3: Let X_1 and X_2 be independent random variables. Then

$$E [\exp(it(X_1 + X_2))] = \psi_{X_1}(t) \psi_{X_2}(t)$$

where $\psi_{X_i}(t)$ represents the characteristic function of the random variable X_i .

PROOF:

$$\begin{aligned} E [\exp(it(X_1 + X_2))] &= E [\exp(itX_1) \exp(itX_2)] \\ &= E [\exp(itX_1)] E [\exp(itX_2)] \quad (\text{By independence}) \\ &= \psi_{X_1}(t) \psi_{X_2}(t) \end{aligned}$$

□

COROLLARY 2.1.4: 1. Let $X_1, X_2, \dots, X_n$ be a finite sequence of independent random variables, with distribution functions $F_1, F_2, \dots, F_n$

respectively. Then if

$$U = \sum_{i=1}^n X_i$$

the characteristic function of U is

$$E[\exp(itU)] = \prod_{i=1}^n \psi_{X_i}(t)$$

2. A characteristic function is real iff X is symmetric.

PROOF:

1. The result follows directly LEMMA 2.1.3.
2. Assuming that the characteristic function uniquely defines a distribution function

$$\psi(t) \in \mathbb{R} \Leftrightarrow \psi(t) = \overline{\psi(t)} \Leftrightarrow \psi_X(t) = \psi_{-X}(t)$$

Therefore a characteristic function is real iff X is symmetric.

□

EXAMPLE: Let X be the standard normal distribution, with density

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) \quad -\infty < x < \infty$$

The characteristic function can be obtained by

$$\begin{aligned}
 E[\exp(itX)] &= \int_{-\infty}^{\infty} \exp(itx) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\
 &= \exp\left(\frac{-t^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-it)^2}{2}\right) dx \\
 &= \exp\left(\frac{-t^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y)^2}{2}\right) dy \quad (\text{Let } y = x - it) \\
 &= \exp\left(\frac{-t^2}{2}\right)
 \end{aligned}$$

(As the above integral is the normal density over $\mathbb{R}$)

EXAMPLE: Let X be the uniform distribution on (a, b) for $a < b$, with density

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

The characteristic function is

$$\begin{aligned}
 E[\exp(itX)] &= \int_a^b \exp(itx) \frac{1}{b-a} dx \\
 &= \frac{1}{it(b-a)} \exp(itx) \Big|_a^b \\
 &= \frac{\exp(itb) - \exp(ita)}{it(b-a)}
 \end{aligned}$$

THEOREM 2.1.5: Let ψ be the characteristic function of the distribution F and suppose that $\psi \in L$. Then f has bounded continuous density f given by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \psi(t) dt$$

PROOF:

$$\begin{aligned}
 e^{-itx} \psi(t) &= \int_{-\infty}^{\infty} e^{it(y-x)} dF(y) \\
 \Rightarrow \int_{-\infty}^{\infty} e^{-itx} \psi(t) dG(t) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{it(y-x)} dF(y) dG(t)
 \end{aligned}$$

Reversing the order of integration and letting $\gamma(t)$ denote the characteristic function of G

$$= \int_{-\infty}^{\infty} \gamma(y-x) dF(y)$$

Define a function

$$N(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

and let G be the normal distribution with variance a^{-2} . This implies that $g(x) = aN(ax)$ and $\gamma(x) = \sqrt{2\pi}N\left(\frac{x}{a}\right)$

$$\begin{aligned} &\Rightarrow \int_{-\infty}^{\infty} e^{-itx} \psi(t) g(t) dt = \int_{-\infty}^{\infty} \sqrt{2\pi} N\left(\frac{y-x}{a}\right) \\ &\Rightarrow \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \psi(t) \exp\left(-\frac{a^2 t^2}{2}\right) dt = \int_{-\infty}^{\infty} \frac{1}{a} N\left(\frac{x-t}{a}\right) dF(y) \end{aligned}$$

The right hand side of the above expression is the convolution of F with a normal distribution with mean 0 and variance a^2 . Therefore

$$f(x) = \lim_{a \downarrow 0} \left\{ \int_{-\infty}^{\infty} \frac{1}{a} N\left(\frac{x-t}{a}\right) dF(x) \right\}$$

On the other hand, taking the limit of the left hand side gives

$$\lim_{a \downarrow 0} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \psi(t) e^{-\frac{a^2 t^2}{2}} dt \right\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \psi(t) dt$$

Therefore

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \psi(t) dt$$

□

THEOREM 2.1.6: Let X be a random variable with distribution function F

and characteristic function $\psi(t)$. For $a < b$

$$F(b) - F(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(t) \frac{e^{-iat} - e^{-ibt}}{it} dt$$

PROOF: Consider the convolution of F with a uniform random variable U distributed over the interval $(-b, -a)$, which is independent of X . Let g denote the density of this convolution. By conditioning on the value of X it is seen that

$$g(x) = \int_{x+a}^{x+b} \frac{1}{b-a} dF(x) = \frac{F(x+b) - F(x+a)}{b-a}$$

Let $\chi(t)$ be the characteristic function of G . As U is independent of X ,

$$\chi(t) = \psi(t) \left(\int_{-b}^{-a} \frac{e^{itx}}{b-a} dx \right) = \psi(t) \frac{e^{-iat} - e^{-ibt}}{it(b-a)}$$

Using Theorem 2.1.5 we know that

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi(t) e^{-itx} dt$$

Therefore

$$\begin{aligned} \frac{F(x+b) - F(x+a)}{b-a} &= g(x) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi(t) e^{-itx} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(t) \frac{e^{-iat} - e^{-ibt}}{it(b-a)} e^{-itx} dt \end{aligned}$$

Let $x = 0$ and cancel the $(b-a)^{-1}$ term on each side to get

$$F(b) - F(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(t) \frac{e^{-iat} - e^{-ibt}}{it} dt$$

□

2.1.2 MIXTURES

DEFINITION 2.1.7: Let $F_1, F_2, \dots$ be probability distributions with characteristic functions $\psi_1(t), \psi_2(t), \dots, \psi_n(t)$. If $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$, then the mixture is defined as

$$U = \sum_{i=1}^n p_i F_i$$

LEMMA 2.1.8: The mixture is a probability distribution with characteristic function

$$\psi_U(t) = \sum_{i=1}^n p_i \psi_{F_i}(t)$$

PROOF:

$$\begin{aligned} E[\exp(itx)] &= \int_{-\infty}^{\infty} \exp(itx) dU(x) \\ &= \int_{-\infty}^{\infty} \exp(itx) \left(\sum_{i=1}^n p_i dF_i(x) \right) \\ &= \sum_{i=1}^n p_i \left(\int_{-\infty}^{\infty} \exp(itx) dF_i(x) \right) \\ &= \sum_{i=1}^n p_i \psi_{F_i}(t) \end{aligned}$$

Let $t = 0$ and $\psi_U(0) = 1$, which implies that the mixture is a valid probability distribution. $\square$

This leads to an interesting problem. Given a mixture of normal distributions with parameters μ_i and σ_i^2 , if the characteristic function of the mixture was divided by a factor of $\exp\left(-\frac{\sigma_0^2}{2}\right)$, where $\sigma_0^2 < \min_{i=1, \dots, n} \{\sigma_i^2\}$, can we ‘remove the background noise’ to find the means of the underlying distributions. This corresponds to removing a normal component of mean 0 and variance σ_0^2 from each of the components of the mixture.

This was implemented using numerical methods to construct the characteristic function and the inverse of the characteristic function after division by $\exp\left(-\frac{\sigma_0^2}{2}\right)$. As the characteristic function is a complex integral, it was evaluated by approximating the real integral and imaginary integral separately. Let $\Delta = \frac{b_x - a_x}{n}$, where $[a_x, b_x]$ is the interval over the real axis for which the mixture distribution (X) is going to be numerically approximated. Let $a_x = x_0$, $x_k = x_{k-1} + \Delta$ for $k = 1, \dots, n$. Then

$$\int_{-\infty}^{\infty} \exp(itx) dF(x) = \int_{-\infty}^{\infty} \cos(tx) dF(x) + i \int_{-\infty}^{\infty} \sin(tx) dF(x)$$

So we approximate as follows

$$\begin{aligned} \int_{-\infty}^{\infty} \cos(tx) dF(x) &\approx \sum_{i=0}^n \cos(tx_i) f(x_i) \Delta \\ \int_{-\infty}^{\infty} \sin(tx) dF(x) &\approx \sum_{i=0}^n \sin(tx_i) f(x_i) \Delta \end{aligned}$$

where $f(x)$ is the density of the mixture. This gives an expression for the real and imaginary parts of the characteristic function as a function of t , so we must also evaluate these expressions over a set of values for t . Let $[a_t, b_t]$ be the interval where the characteristic function is going to be numerically evaluated on. Then $\Delta_t = \frac{b_t - a_t}{n}$ and $a_t = t_0$, $t_k = t_{k-1} + \Delta_t$ for $k = 1, \dots, n$. So this sum is evaluated over the set of values for t_i .

After this has been done, every value of the characteristic function is divided by $\exp\left(-\frac{\sigma_0^2 t^2}{2}\right)$. Because this is very large (for $|t| \geq 3$) and the error involved in numerically evaluating the integrals is significant when multiplying by such a large value, the characteristic function needs to be smoothed by saying any real or imaginary part with modulus $< \epsilon$ is now $= 0$, for ϵ sensible compared to the value of σ_0^2 . This is a sensible procedure as the error in evaluating the integral for the characteristic function is about 10^{-17} , so removing these terms is necessary when $t \geq 3$. As we know that the density is a function in $\mathbb{R}$, we may simplify the inversion formula in Theorem 2.1.5 and approximate the

inversion by

$$f'(x) \approx \frac{1}{2\pi} \sum_{i=1}^n (\cos(t_i x) \mathcal{R}(\psi(t_i)) \Delta_t + \sin(t_i x) \mathcal{I}(\psi(t_i)) \Delta_t)$$

where $\mathcal{R}()$ is the real part and $\mathcal{I}()$ is the imaginary part. Results are provided in the figures at the end of this section.

The last mixture distribution considered in this talk is the mixture of Cauchy random variables. Consider the artificial example of a number of light sources placed a given perpendicular distance from a screen. If a light is randomly chosen and then emits a light ray at a uniformly generated angle (on $(-\pi, \pi)$), then the resulting positions where the light rays hit the screen will have a distribution which is a mixture of Cauchy distributions. The modifications to the previous method that are needed to deal with the Cauchy distribution are to use the Cauchy distribution function instead of the normal and to multiply by $\exp(\gamma|t|)$ instead of $\exp\left(-\frac{\sigma^2 t^2}{2}\right)$, the characteristic function of the Cauchy distribution with scale parameter γ (the characteristic function of the Cauchy distribution with location parameter μ and scale parameter γ is $\exp(i\mu t - \gamma|t|)$, so the division of a central Cauchy distribution with parameter $\gamma' < \gamma$ will give a Cauchy distribution with a reduced scale parameter). Results are provided in the figures at the end of this section.

FIGURE 2.1: Mixture of normal densities with $\mu_1 = -1.5, \mu_2 = 0, \mu_3 = 1.5$ and $\sigma_i^2 = 1$.

FIGURE 2.2: Mixture of normal distributions with $\mu_1 = -1.5$, $\mu_2 = 0$, $\mu_3 = 1.5$ and $\sigma_i^2 = 1$.

FIGURE 2.3: Mixture of normal densities with $\mu_1 = -2$, $\mu_2 = -1$, $\mu_3 = 1$, $\mu_4 = 2$ and $\sigma_i^2 = 1$.

FIGURE 2.4: Mixture of normal distributions with $\mu_1 = -2$, $\mu_2 = -1$, $\mu_3 = 1$, $\mu_4 = 2$ and $\sigma_i^2 = 1$.

FIGURE 2.5: Mixture of Cauchy densities with scale parameter = 1 for all components and shift parameters $\mu = (-1, 0, 1)$

2.2 RESEARCH PROBLEM

The main reference for this section is [HOGNAS \(1986\)](#), with other background references being [PFLUG \(1991\)](#) and [O'BRIEN \(1972\)](#). Consider a sequence defined by

$$X_{n+1} = f(X_n) + \xi_{n+1}$$

where ξ_i are independent and identically distributed with the standard normal distribution. If $f(x) = ax$ for $|a| < 1$ constant, then X_∞ has a limit distribution of $N\left(0, \frac{1}{1-a^2}\right)$. The problem posed by Goran Hognas is:

Given an arbitrary function $f(x)$ that lies between two linear functions $f_1(x) = ax$ and $f_2(x) = bx$, for all x , with $0 < a < b < 1$, does the limit distribution lie between the limit distributions of the sequences generated by $f_1(x)$ and $f_2(x)$?

The first function considered is

$$f(x) = \frac{a+b}{2}x + \frac{b-a}{2}x \sin(x)$$

The distribution function was constructed by first constructing the empirical distribution function from a generated data set of 100,000 observations of this process. The results for $b = 0.75$ and $a = 0.25$ are given in [FIGURE 2.6](#). This answers the problem in the negative. Another choice of function is

$$f(x) = \frac{a+b}{2}x + \frac{b-a}{2}x \sin(4x)$$

FIGURE 2.6: The blue lines indicate the limit distributions of the linear functions ax and bx . The black line is the numerical estimate of the limit distribution using $f(x) = \frac{a+b}{2}x + \frac{b-a}{2}x \sin(x)$

with $a = 0.75$ and $b = 0.25$, which leads to the results in FIGURE 2.7. The final choice of function considered here is

$$f(x) = \frac{a+b}{2}x + \frac{b-a}{2}x \cos(x)$$

with $a = 0.75$ and $b = 0.25$, which leads to the results in FIGURE 2.8. These results show that general functions within the boundary of the linear

FIGURE 2.7: The blue lines indicate the limit distributions of the linear functions ax and bx . The black line is the numerical estimate of the limit distribution using $f(x) = \frac{a+b}{2}x + \frac{b-a}{2}x \sin(4x)$

functions ax and bx do not necessarily lead to a limit distribution between the limit distributions of the linear functions, however some non-linear functions could lead to limit functions between the limit distributions of the linear functions. Therefore, the actual behaviour of the function f changes the limit distribution. The current conjecture is that for functions f , if we consider $f'(x) = f(x) - \frac{a+b}{2}x$, if $f'(x)$ is odd, then the limit distribution will lie within the limit distribution of the linear functions.

FIGURE 2.8: The blue lines indicate the limit distributions of the linear functions ax and bx . The black line is the numerical estimate of the limit distribution using $f(x) = \frac{a+b}{2}x + \frac{b-a}{2}x \cos(x)$

2.3 NEW APPROACH TO TESTING REGRESSION

The reference for this section is [KHMALADZE & KOUL \(2004\)](#). The aim of this section is to introduce a way of testing regression problems based on transformations of the errors process.

We are assuming that

$$Y_i = m(X_i; \theta) + \epsilon_i$$

where $m(X_i; \theta)$ is the regression function and ϵ_i are independent and identically distributed, with mean 0 and variance σ^2 . If they do not have mean 0, then we can add the bias of the error terms into the regression function and consider the new error terms which have mean 0. If we have chosen $m(X_i; \theta)$ correctly, then we can consider the partial sum process

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{i \leq nt} \epsilon_i \quad 0 \leq t \leq 1$$

which, as $n \rightarrow \infty$, has independent increments, mean 0 and variance $t\sigma^2$. By the central limit theorem the distribution for a given t is Gaussian, so the limit is Brownian motion. Now consider estimating θ using Least-Squares. If

$m(X_i; \theta)$ is differentiable in θ then

$$\begin{aligned}
 \min_{\theta} \sum_{i=1}^n \epsilon_i^2 &= \min_{\theta} \sum_{i=1}^n (Y_i - m(X_i; \theta))^2 \\
 &= \frac{\partial}{\partial \theta} \sum_{i=1}^n (Y_i - m(X_i; \theta))^2 \\
 &= \sum_{i=1}^n \frac{\partial}{\partial \theta} (Y_i - m(X_i; \theta))^2 \\
 &= -2 \sum_{i=1}^n (Y_i - m(X_i; \theta)) m'(X_i; \theta) = 0
 \end{aligned} \tag{2.1}$$

Letting $\hat{\theta}$ be the solution of (2.1), we can consider the partial sum process of the errors $\hat{\epsilon}_i$ under $\hat{\theta}$,

$$\widehat{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{i \leq nt} (Y_i - m(X_i; \hat{\theta})) = \frac{1}{\sqrt{n}} \sum_{i \leq nt} \hat{\epsilon}_i \quad 0 \leq t \leq 1.$$

We cannot apply the same logic to deduce that $\widehat{W}_n(t)$ has a Brownian motion limit because the Least-Squares condition means that the errors are no longer independent. It follows from the Least-Squares condition that

$$\int_0^1 m'(X_{nt}; \hat{\theta}) d\widehat{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^n m'(X_k; \hat{\theta}) \hat{\epsilon}_k = 0$$

As an example

$$H_0: Y_i = a + \epsilon_i$$

for constant a , then the Least-Squares estimator of a is $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$, and it follows from (2.1) that $\widehat{W}_n(1) = 0$, as

$$\int_0^1 m'(X_{nt}; \hat{\theta}) d\widehat{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^n \hat{\epsilon}_k = 0.$$

This suggests that $\widehat{W}_n(t) \rightarrow \mathbb{V}$, where $\mathbb{V}$ is Brownian bridge, which is actually the case.

We want to modify $\hat{\epsilon}_i = Y_i - \bar{Y}$ in such a way that the partial sum process of the modified summands converges to Brownian motion. Consider the filtration $\mathcal{H} = \{\mathcal{H}_i\}$, where each $\mathcal{H}_i$ is the σ -algebra generated by $Y_1, Y_2, \dots, Y_i, \bar{Y}$.

$$E[Y_i | \mathcal{H}_{i-1}] = E\left[Y_i \left| \sum_{j=i}^n Y_j \right.\right]$$

This conditional expectation may be difficult to calculate, but we can consider a regression of Y_i on $\sum_{j=i}^n Y_j$, with $\xi_i = Y_i - \gamma \sum_{j=i}^n Y_j$ for regression parameter γ . The Least-Squares estimator of γ is

$$\begin{aligned} E\left[\left(Y_i - \gamma \sum_{j=i}^n Y_j\right) \sum_{j=i}^n Y_j\right] &= 0 \\ \Rightarrow E\left[Y_i^2 - \gamma \sum_{j=i}^n Y_j^2\right] &= 0 \\ \Rightarrow \sigma^2 &= (n - i + 1)\gamma\sigma^2 \\ \Rightarrow \gamma &= \frac{1}{n - i + 1} \end{aligned}$$

Letting $\xi_i = (Y_i - \bar{Y}) - \frac{1}{n - i + 1} \sum_{j=i}^n (Y_j - \bar{Y}) = Y_i - \frac{1}{n - i + 1} \sum_{j=i}^n Y_j$, we find that for $i < j$, and assuming, without loss of generality, that $E[Y_i] = 0$

$$\begin{aligned} E[\xi_i \xi_j] &= E\left[Y_i Y_j - \frac{Y_i}{n - j + 1} \sum_{l=j}^n Y_l - \frac{Y_j}{n - i + 1} \sum_{k=i}^n Y_k \right. \\ &\quad \left. + \left(\frac{1}{n - i + 1} \sum_{k=i}^n Y_k\right) \left(\frac{1}{n - j + 1} \sum_{l=j}^n Y_l\right)\right] \end{aligned}$$

As Y_i and Y_j are independent with mean 0 it follows that

$$E[\xi_i \xi_j] = 0 - 0 - \frac{\sigma^2}{n-i+1} + \frac{1}{n-i+1} \frac{1}{n-j+1} (n-j+1) \sigma^2 = 0$$

We can also calculate the variance of ξ_i as

$$\begin{aligned} E[\xi_i^2] &= E \left[Y_i^2 - 2Y_i \frac{1}{n-i+1} \sum_{j=i}^n Y_j + \left(\frac{1}{n-i+1} \right)^2 \left(\sum_{j=i}^n Y_j \right)^2 \right] \\ &= \sigma^2 - 2 \frac{1}{n-i+1} \sigma^2 + \left(\frac{1}{n-i+1} \right)^2 (n-i+1) \sigma^2 \\ &= \sigma^2 \left(1 - \frac{1}{n-i+1} \right) \end{aligned}$$

Therefore the partial sum process

$$\widetilde{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{i \leq nt} \xi_i$$

is a sum of uncorrelated summands. It is also a martingale with respect to $\mathcal{H}$.

Using martingale limit theory it is possible to show that $\widetilde{W}_n(t)$ converges to Brownian motion.

We can also reconstruct the original errors $\hat{\epsilon}_i$ from ξ_i . With $\xi_i = Y_i -$

$\frac{1}{n-i+1} \sum_{j=i}^n Y_j$, we find that $\xi_1 = \hat{\epsilon}_1$,

$$\hat{\epsilon}_k = \xi_k - \sum_{j=1}^{k-1} \left(\frac{1}{n-j} \xi_j \right) \quad k = 2, \dots, n-1$$

and finally $\hat{\epsilon}_n = - \sum_{j=1}^{n-1} \hat{\epsilon}_j$.

For the second example we will consider linear regression

$$Y_i = a + b(X_i - \bar{X}) + \epsilon_i$$

for ϵ_i independent and identically distributed. The Least-Squares estimators a and b must satisfy

$$\frac{\partial}{\partial a} \sum_{i=1}^n (Y_i - a - b(X_i - \bar{X}))^2 = - \sum_{i=1}^n (Y_i - a - b(X_i - \bar{X})) = 0 \quad (2.2)$$

$$\frac{\partial}{\partial b} \sum_{i=1}^n (Y_i - a - b(X_i - \bar{X}))^2 = - \sum_{i=1}^n (Y_i - a - b(X_i - \bar{X})) (X_i - \bar{X}) = 0 \quad (2.3)$$

From (2.2) it follows that

$$a = \frac{1}{n} \sum_{i=1}^n Y_i - \frac{b}{n} \sum_{i=1}^n (X_i - \bar{X}) = \bar{Y}$$

and from (2.3), substituting $a = \bar{Y}$

$$\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = b \sum_{i=1}^n (X_i - \bar{X})^2 \Rightarrow b = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Therefore

$$Y_i = \bar{Y} + \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} (X_i - \bar{X}) + \hat{\epsilon}_i$$

In this case, the limit of the partial sum process is subject to the conditions

$$\int_0^1 d\widehat{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^n \hat{\epsilon}_k = 0 \quad (2.4)$$

$$\int_0^1 (X_t - \bar{X}) d\widehat{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - \bar{X}) \hat{\epsilon}_k = 0 \quad (2.5)$$

so the limit is Brownian bridge subject to the extra condition (2.5). We want to modify the partial sum process so that the modified summands ξ_i are uncorrelated. As we have added a new variable to the regression function, we have a new filtration $\mathcal{H} = \{\mathcal{H}_i\}$, where each $\mathcal{H}_i$ is the σ -algebra generated by $Y_1, Y_2, \dots, Y_i, \bar{Y}, X_1, X_2, \dots, X_n, \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$. We want to again modify the partial sum process so that the modified partial sum process converges to Brownian motion.

$$E[\epsilon_i | \mathcal{H}_{i-1}] = E \left[\epsilon_i \left| \sum_{j=i}^n \epsilon_j, \sum_{j=i}^n (X_j - \bar{X}) \epsilon_j \right. \right]$$

If we know $\sum_{j=i}^n (X_j - \bar{X}) \epsilon_j$ and $\sum_{j=i}^n \epsilon_j$, then we also know $\sum_{j=i}^n X_j \epsilon_j$, from which we can calculate $\sum_{j=i}^n (X_j - \bar{X}_{i,n}) \epsilon_j$, where $\bar{X}_{i,n} = \frac{1}{n-i+1} \sum_{j=i}^n X_j$ are partial averages. It can be shown that $\sum_{j=i}^n \epsilon_j$ and $\sum_{j=i}^n (X_j - \bar{X}_{i,n}) \epsilon_j$ are uncorrelated.

$$= E \left[\epsilon_i \left| \sum_{j=i}^n \epsilon_j, \sum_{j=i}^n (X_j - \bar{X}_{i,n}) \epsilon_j \right. \right]$$

We can consider a regression of ϵ_i on $\sum_{j=i}^n \epsilon_j$ and $\sum_{j=i}^n (X_j - \bar{X}_{i,n})\epsilon_j$, where

$$\xi_i = \epsilon_i - \alpha_i \sum_{j=i}^n \epsilon_j - \beta_i \sum_{j=i}^n (X_j - \bar{X}_{i,n})\epsilon_j,$$

so using Least-Squares

$$\frac{\partial}{\partial \alpha_i} E \left[\sum_{k=1}^n \xi_k^2 \right] = 0 = E \left[\xi_i \left(\sum_{j=i}^n \epsilon_j \right) \right] = \sigma^2 - \alpha_i \sigma^2 (n - i + 1) \quad (2.6)$$

$$\begin{aligned} \frac{\partial}{\partial \beta_i} E \left[\sum_{k=1}^n \xi_k^2 \right] &= 0 = E \left[\xi_i \left(\sum_{j=i}^n (X_j - \bar{X}_{i,n}) \epsilon_j \right) \right] \\ &= \sigma^2 (X_i - \bar{X}_{i,n}) - \beta_i \sum_{j=i}^n (X_j - \bar{X}_{i,n})^2 \end{aligned} \quad (2.7)$$

From (2.6) it follows

$$\alpha_i = \frac{1}{n - i + 1} \quad \beta_i = \frac{X_i - \bar{X}_{i,n}}{\sum_{j=i}^n (X_j - \bar{X}_{i,n})^2}$$

Therefore

$$\xi_i = \epsilon_i - \frac{1}{n - i + 1} \sum_{j=i}^n \epsilon_j - \frac{X_i - \bar{X}_{i,n}}{\sum_{j=i}^n (X_j - \bar{X}_{i,n})^2} \sum_{j=i}^n (X_j - \bar{X}_{i,n}) \epsilon_j$$

We also note that $\epsilon_i - c$ can be substituted in place of ϵ_i in the above equation without changing ξ_i , so it follows that if

$$\xi'_i = \hat{\epsilon}_i - \frac{1}{n - i + 1} \sum_{j=i}^n \hat{\epsilon}_j - \frac{X_i - \bar{X}_{i,n}}{\sum_{j=i}^n (X_j - \bar{X}_{i,n})^2} \sum_{j=i}^n (X_j - \bar{X}_{i,n}) \hat{\epsilon}_j$$

and that

$$\widetilde{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{i \leq nt} \xi'_i$$

is Brownian motion.

2.4 RANKS

2.4.1 INTRODUCTION TO RANKS

For this section assume that $X_i \in \mathbb{R}$. Define

$$\min_{1 \leq i \leq n} X_i = X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n} = \max_{1 \leq i \leq n} X_i$$

as the order statistics for $i = 1, \dots, n$.

DEFINITION 2.4.1: Let $X_1, X_2, \dots, X_n$ be a sequence of random variables.

The *rank* of X_i , denoted R_i , is $R_i = \sum_{j=1}^n \mathbb{I}_{\{X_j \leq X_i\}}$

It is also convenient to assume that the underlying distribution function F is continuous on $\mathbb{R}$, as then the sample elements are almost surely different.

RESULT 2.4.2: For continuous F we have the following

$$F_n(X_i) = \frac{1}{n} R_{in}; \quad F_n(X_{i:n}) = \frac{i}{n}; \quad F_n^{-1}\left(\frac{i}{n}\right) = X_{i:n}$$

PROOF: The result follows immediately from the definition. □

For the next theorem we need to define

$$\mathbb{O}_n = \{\langle t_1, \dots, t_n \rangle \in \mathbb{R}^n : t_1 \leq \dots \leq t_n\}$$

and $\mathcal{R}_n$ as the set of permutations of $\langle 1, \dots, n \rangle$

THEOREM 2.4.3: For continuous F , $O \subset \mathbb{O}$ and $(r_1, \dots, r_n) \in \mathcal{R}_n$ we have

1. $\mathbb{P}\{\langle X_{1:n}, \dots, X_{n:n} \rangle \in O\} = n! \mathbb{P}\{\langle X_1, \dots, X_n \rangle \in O\}$
2. $\mathbb{P}\{\langle R_{1n}, \dots, R_{nn} \rangle = \langle r_1, \dots, r_n \rangle\} = \frac{1}{n!}$
3. $\langle X_{1:n}, \dots, X_{n:n} \rangle$ and $\langle R_{1n}, \dots, R_{nn} \rangle$ are independent.

PROOF: Let $\rho = \langle \rho_1, \dots, \rho_n \rangle$ be the inverse permutation of $\langle r_1, \dots, r_n \rangle$. Then it is clear that

$$\begin{aligned} & \mathbb{P}\{\langle R_{1n}, \dots, R_{nn} \rangle = \langle r_1, \dots, r_n \rangle, \langle X_{1:n}, \dots, X_{n:n} \rangle \in O\} \\ &= \mathbb{P}\{\langle X_{\rho_1}, \dots, X_{\rho_n} \rangle \in O\} \\ &= \mathbb{P}\{\langle X_1, \dots, X_n \rangle \in O\} \end{aligned}$$

Letting $O = \mathbb{O}$ we get

$$\begin{aligned}
 & \mathbb{P} \{ \langle R_{1n}, \dots, R_{nn} \rangle = \langle r_1, \dots, r_n \rangle \} \\
 &= \mathbb{P} \{ \langle R_{1n}, \dots, R_{nn} \rangle = \langle r_1, \dots, r_n \rangle, \langle X_{1:n}, \dots, X_{n:n} \rangle \in \mathbb{O} \} \\
 &= \mathbb{P} \{ \langle X_1, \dots, X_n \rangle \in \mathbb{O} \} \\
 &= \mathbb{P} \{ X_1 \leq \dots \leq X_n \} \\
 &= \frac{1}{n!}
 \end{aligned}$$

By summing over the possible permutations of the ranks it is clear that

$$\begin{aligned}
 & \mathbb{P} \{ \langle X_{1:n}, \dots, X_{n:n} \rangle \in O \} \\
 &= n! \mathbb{P} \{ \langle X_1, \dots, X_n \rangle \in O \}
 \end{aligned}$$

This leads to

$$\begin{aligned}
 & \mathbb{P} \{ \langle R_{1n}, \dots, R_{nn} \rangle = \langle r_1, \dots, r_n \rangle, \langle X_{1:n}, \dots, X_{n:n} \rangle \in O \} \\
 &= \mathbb{P} \{ \langle X_1, \dots, X_n \rangle \in O \} \\
 &= \frac{1}{n!} \mathbb{P} \{ \langle X_{1:n}, \dots, X_{n:n} \rangle \in O \} \\
 &= \mathbb{P} \{ \langle R_{1n}, \dots, R_{nn} \rangle = \langle r_1, \dots, r_n \rangle \} \mathbb{P} \{ \langle X_{1:n}, \dots, X_{n:n} \rangle \in O \}
 \end{aligned}$$

□

This leads us to consider the marginal distributions.

THEOREM 2.4.4:

$$\mathbb{P}\{R_{in} = j\} = \frac{1}{n}; \quad \forall j \in \{1, \dots, n\}$$

PROOF: This is clear by considering the number of possible values of R_{in} ($1, \dots, n$) and noticing that because the sample is independent and identically distributed that all values have equal probability. $\square$

THEOREM 2.4.5:

$$\begin{aligned} \mathbb{P}\{X_{i:n} \leq t\} &= \frac{n!}{(i-1)!(n-i)!} \int_0^{F(t)} s^{i-1} (1-s)^{n-i} ds \\ &= \sum_{k=i}^n \binom{n}{k} [F(t)]^k [1-F(t)]^{n-k} \end{aligned}$$

PROOF: The first expression assumes that there is a derivative f of F on $\mathbb{R}$.

Using the multinomial distribution, it is seen that

$$\begin{aligned} &\mathbb{P}\{t < X_{i:n} \leq t + \Delta\} \\ &= \frac{n!}{(i-1)!1!(n-i)!} [F(t)]^{i-1} [F(t+\Delta) - F(t)] [1-F(t)]^{n-i} + O(\Delta^2) \end{aligned}$$

Dividing by Δ and taking the limit for $\Delta \downarrow 0$ we find the density g for the distribution of $X_{i:n}$ as

$$g(t) = \frac{n!}{(i-1)!(n-i)!} [F(t)]^{i-1} [1-F(t)]^{n-i} f(t)$$

Taking the integral of $g(t)$ from 0 to s and making a substitution $s = F(t)$

gives

$$\begin{aligned}\mathbb{P}\{X_{i:n} \leq t\} &= \frac{n!}{(i-1)!(n-i)!} \int_0^{F(t)} s^{i-1} (1-s)^{n-i} ds \\ &= \sum_{k=i}^n \binom{n}{k} [F(t)]^k [1-F(t)]^{n-k}\end{aligned}$$

To derive the second expression it is useful to note that

$$\{X_{i:n} \leq t\} \Leftrightarrow \{F_n(t) \geq t\}$$

By conditioning on the number of $X_i \leq t$ and summing over the events the second expression immediately follows. $\square$

2.4.2 APPLICATIONS

EXAMPLE: (Two Sample Problem)

Suppose that $X_1, \dots, X_{n_1}$ is a random sample of size n_1 from a continuous distribution function F_1 on $\mathbb{R}$ and that $X_{n_1+1}, \dots, X_{n_1+n_2}$ is a random sample of size n_2 from a continuous distribution function F_2 on $\mathbb{R}$. The two sample problem tests the null hypothesis of $F_1 = F_2$ versus the alternative of inequality. Let $n = n_1 + n_2$ and $R_{1n}, \dots, R_{nn}$ be the ranks of the pooled sample elements. Then a good test statistic would be

$$\sum_{i=1}^{n_1} R_{in} \text{ or } \sum_{i=1}^{n_1} \frac{R_{in}}{n+1}$$

which would be rejected for small or large values. Assume from now that the null hypothesis is true and that F is the common distribution function.

THEOREM 2.4.6: Suppose for each n the real numbers $c_{1n}, \dots, c_{nn}$ satisfy

$$\sum_{i=1}^n c_{in} = 0, \quad \sum_{i=1}^n c_{in}^2 = 1, \quad \max_{1 \leq i \leq n} c_{in} \rightarrow 0$$

Furthermore let

$$W_n = (12)^{\frac{1}{2}} \sum_{i=1}^n c_{in} \frac{R_{in}}{n+1}, \quad S_n = (12)^{\frac{1}{2}} \sum_{i=1}^n c_{in} \xi_i$$

where $\xi_i = F(X_i)$ are independent with distribution $\mathcal{U}(0, 1)$ under the null hypothesis. Then

$$\mathbb{P}\{W_n \leq x\} \xrightarrow{n \rightarrow \infty} \Phi(x) \quad \forall x \in \mathbb{R}$$

PROOF: The first part is to show S_n converges to the standard normal distribution.

$$\begin{aligned} E[S_n] &= (12)^{\frac{1}{2}} \left(\sum_{i=1}^n c_{in} \right) E[\xi_i] = 0 \\ \text{Var}(S_n) &= 12 \left(\sum_{i=1}^n c_{in}^2 \right) \text{Var}(\xi_i) = 12(1) \left(\frac{1}{12} \right) = 1 \end{aligned}$$

Therefore S_n converges to the standard normal distribution using the central limit theorem. The next stage is to show that $W_n \rightarrow S_n$ in $\mathcal{L}^2$, which implies convergence in probability, which implies convergence in distribution. Let

$$(V_1, \dots, V_n) = \left(\frac{R_{1n}}{n+1} - \xi_1, \dots, \frac{R_{nn}}{n+1} - \xi_n \right)$$

which has exchangeable components. This implies that $\mathcal{L}(V_i) = \mathcal{L}(V_1)$ and $\mathcal{L}(V_i V_j) = \mathcal{L}(V_1 V_2); \forall i \neq j$.

$$\begin{aligned} \frac{1}{12} E [W_n - S_n]^2 &= E \left[\sum_{i=1}^n c_{in} V_i \right]^2 \\ &= \sum_{i \neq j} c_{in} c_{jn} E [V_i V_j] + \sum_{i=1}^n c_{in}^2 E [V_i^2] \\ &= \left(\sum_{i=1}^n c_{in} \right)^2 E [V_1 V_2] + \left(\sum_{i=1}^n c_{in}^2 \right) (E [V_1^2] - E [V_1 V_2]) \\ &\leq 2E [V_1^2] \end{aligned}$$

All that remains is to show that $E [V_1^2] \rightarrow 0$.

$$\begin{aligned} E [V_1^2] &= E \left[E \left[\left(\frac{R_{1n}}{n+1} - \xi_i \right)^2 \mid R_{1n} = i \right] \right] \\ &= \sum_{i=1}^n E \left[\left(\frac{R_{1n}}{n+1} - \xi_i \right)^2 \mid R_{1n} = i \right] \mathbb{P} \{R_{1n} = i\} \\ &= \frac{1}{n} \sum_{i=1}^n E \left[\left(\frac{i}{n+1} - \xi_{i:n} \right)^2 \right] \end{aligned}$$

So using THEOREM 2.4.4 and THEOREM 2.4.5 on ξ_i , we see the distribution function of ξ_i for $0 \leq t \leq 1$ is

$$\begin{aligned}\mathbb{P}\{\xi_{i:n} \leq t\} &= \frac{n!}{(i-1)!(n-i)!} \int_0^t s^{i-1}(1-s)^{n-i} dF(s) \\ &= \frac{n!}{(i-1)!(n-i)!} \int_0^t s^{i-1}(1-s)^{n-i} ds\end{aligned}$$

Which gives the first and second moments as

$$\begin{aligned}E[\xi_{i:n}] &= \frac{n!}{(i-1)!(n-i)!} \int_0^1 s^i(1-s)^{n-i} ds \\ &= \frac{n!}{(i-1)!(n-i)!} \frac{(i)!(n-i)!}{(n+1)!} = \frac{i}{n+1} \\ E[\xi_{i:n}^2] &= \frac{n!}{(i-1)!(n-i)!} \int_0^1 s^{i+1}(1-s)^{n-i} ds \\ &= \frac{n!}{(i-1)!(n-i)!} \frac{(i+1)!(n-i)!}{(n+2)!} = \frac{(i+1)i}{(n+2)(n+1)}\end{aligned}$$

Therefore

$$\begin{aligned}E\left[\left(\frac{i}{n+1} - \xi_{i:n}\right)^2\right] &= E[\xi_{i:n}^2] - (E[\xi_{i:n}])^2 \\ &= \frac{(i+1)i}{(n+2)(n+1)} - \left(\frac{i}{n+1}\right)^2 = \frac{i(n-i+1)}{(n+2)(n+1)^2}\end{aligned}$$

This gives an upper bound for $\frac{1}{12}E[W_n - S_n]^2$ as

$$\begin{aligned}\frac{1}{12}E[W_n - S_n]^2 &\leq 2E[V_1^2] \\ &= 2 \sum_{i=1}^n \frac{i(n-i+1)}{n(n+2)(n+1)^2} \downarrow 0 \quad \text{as } n \rightarrow \infty\end{aligned}$$

□

COROLLARY 2.4.7: (Wilcoxin's two-sample test) For the two sample problem

$$\mathbb{P} \left\{ \left(\frac{12n}{n_1 n_2} \right)^{\frac{1}{2}} \sum_{i=1}^{n_1} \left(\frac{R_{in}}{n+1} - \frac{1}{2} \right) \leq x \right\} \rightarrow \Phi(x) \quad \text{as } n_1 \wedge n_2 \rightarrow \infty; \forall x \in \mathbb{R}$$

PROOF: Let

$$c_{in} = \begin{cases} \left(\frac{n_2}{nn_1} \right)^{\frac{1}{2}} & \text{for } i = 1, \dots, n_1 \\ - \left(\frac{n_1}{nn_2} \right)^{\frac{1}{2}} & \text{for } i = n_1 + 1, \dots, n_1 + n_2 \end{cases}$$

Then we only need to show it satisfies the conditions for [THEOREM 2.4.6](#).

$$1. \sum_{i=1}^n c_{in} = 0$$

$$\begin{aligned} \sum_{i=1}^n c_{in} &= \sum_{i=1}^{n_1} \left(\frac{n_2}{nn_1} \right)^{\frac{1}{2}} + \sum_{i=1}^{n_2} \left(-\frac{n_1}{nn_2} \right)^{\frac{1}{2}} \\ &= n_1 \left(\frac{n_2}{nn_1} \right)^{\frac{1}{2}} + n_2 \left(-\frac{n_1}{nn_2} \right)^{\frac{1}{2}} \\ &= \left(\frac{n_1 n_2}{n} \right)^{\frac{1}{2}} + \left(-\frac{n_1 n_2}{n} \right)^{\frac{1}{2}} = 0 \end{aligned}$$

$$2. \sum_{i=1}^n c_{in}^2 = 0$$

$$\begin{aligned} \sum_{i=1}^n c_{in}^2 &= \sum_{i=1}^{n_1} \left(\frac{n_2}{nn_1} \right) + \sum_{i=1}^{n_2} \left(\frac{n_1}{nn_2} \right) \\ &= \left(\frac{n_2}{n} \right) + \left(\frac{n_1}{n} \right) = 1 \end{aligned}$$

$$3. \max_{1 \leq i \leq n} c_{in} \rightarrow 0$$

For $i = 1, \dots, n_1$

$$|c_{in}| \leq \frac{1}{(n_1)^{\frac{1}{2}}}$$

and for $i = n_1 + 1, \dots, n_1 + n_2$

$$|c_{in}| \leq \frac{1}{(n_2)^{\frac{1}{2}}}.$$

Both of these $\rightarrow 0$ as $n_1 \wedge n_2 \rightarrow \infty$, concluding the proof.

□

2.4.3 SEQUENTIAL RANKS

The reference for this subsection is [KHMALADZE \(2010\)](#). We are assumig that the underlying distribution of the X_i is continuous for this subsection.

DEFINITION 2.4.8: Let $X_1, X_2, \dots, X_n$ be a sequence of random variables.

The *sequential rank* of X_i , denoted S_i , is

$$S_i = \sum_{j=1}^i \mathbb{I}_{\{X_j \leq X_i\}}$$

The sequential rank is the rank of X_i among only the first i observations. At this stage it is useful to see a comparison of the properties of ranks and sequential ranks.

PROPOSITION 2.4.9: For sequential ranks $S_i \in \{1, \dots, i\}$ with $S_1 = 1$ and $S_n = R_n$, whereas for ranks $R_i \in \{1, \dots, n\}$

PROPOSITION 2.4.10: For $i \neq j$, $\mathbb{P}\{S_i = k | S_j = l\} = \mathbb{P}\{S_i = k\}$, so the sequential ranks are independent. For ranks $\mathbb{P}\{R_i = k | R_j = k\} = 0$ where $\mathbb{P}\{R_i = k\} \neq 0$ for all k , so ranks are not independent. Further, $\sum_{i=1}^n \phi(S_i)$ is the sum of independent random variables and is non-degenerate random variable, whereas $\sum_{i=1}^n \phi(R_i) = \sum_{i=1}^n \phi(i)$ is constant.

THEOREM 2.4.11: The relationship between the vector of ranks and the vector of sequential ranks is one-to-one.

PROOF: Fix a sample of size n . To get the sequential ranks from the ranks, we consider the sequential ranks of the sequence of random variables

$R_1, \dots, R_n$, so

$$S_i = \sum_{j=1}^i \mathbb{I}_{\{R_j \leq R_i\}} = \sum_{j=1}^i \mathbb{I}_{\{X_j \leq X_i\}}$$

where $X_j \leq X_i$ if and only if $R_j \leq R_i$.

To get the ranks from the sequential ranks we start with the known rank

$R_n = S_n$. Now consider S_{n-1} . If $S_n \leq S_{n-1}$, then $X_n < X_{n-1}$, so

$R_{n-1} = S_{n-1} + 1$. If $S_n > S_{n-1}$, then $X_n > X_{n-1}$, so X_n does not affect the rank of X_{n-1} , so $R_{n-1} = S_{n-1}$. It follows that $R_{n-1} = S_{n-1} + \mathbb{I}_{\{S_{n-1} \geq S_n\}}$.

Similar reasoning leads to

$$R_i = S_i + \sum_{j=i+1}^n \mathbb{I}_{\{S_i \geq S_j\}} \quad i = 1, 2, \dots, n-1$$

where $R_n = S_n$. □

When applying ranks in testing problems, such as the linear rank statistics and goodness of fit statistics, the statistics are usually based on the empirical field

$$z_R(t, u) = \sum_{i=1}^{nt} \left[\mathbb{I}_{\{R_{in} \leq u(n+1)\}} - \frac{[nu]}{n+1} \right], \quad (t, u) \in [0, 1]^2.$$

The linear rank statistics are linear functionals from the empirical field

$$\psi(\mathbf{R}_n) = \int \psi(t, u) z_R(dt, du) = \sum_{i=1}^n \left(\psi\left(\frac{i}{n}, \frac{R_{in}}{n+1}\right) - E \left[\psi\left(\frac{i}{n}, \frac{R_{in}}{n+1}\right) \right] \right)$$

where $\mathbf{R}_n$ is a vector of ranks of size n .

THEOREM 2.4.12: If, under the null hypothesis, the sample is independent and identically distributed, while under the alternative hypothesis the distribution A_i of each X_i is such that

$$\frac{dA_i(x)}{dF(x)} = 1 + \frac{1}{\sqrt{n}} a\left(\frac{i}{n}, F(x)\right) + \text{smaller terms}, \quad n \rightarrow \infty \quad (2.8)$$

where $\int_0^1 a(t, F(x)) dt = 0$, then the linear rank statistic, with $\psi = a$,

$$a(\mathbf{R}_n) = \sum_{i=1}^n a\left(\frac{i}{n}, \frac{R_{in}}{n+1}\right)$$

is asymptotically optimal against this alternative.

PROOF: See [HAJEK ET AL. \(1999\)](#), among others. □

Returning to sequential ranks, we would expect that we could use sequential ranks for testing instead of ranks, and still have asymptotic optimality. We can construct an empirical field based on sequential ranks

$$z_S(t, u) = \sum_{i=1}^{nt} \left[\mathbb{I}_{\{S_i \leq u(i+1)\}} - \frac{[iu]}{i+1} \right], \quad (t, u) \in [0, 1]^2.$$

and sequential linear rank statistics based on the empirical field

$$\phi(\mathbf{S}_n) = \int \phi(t, u) z_S(dt, du) = \sum_{i=1}^n \left(\psi\left(\frac{i}{n}, \frac{S_i}{i+1}\right) - E\left[\psi\left(\frac{i}{n}, \frac{S_i}{i+1}\right)\right] \right).$$

However, $a(\mathbf{S}_n) = \sum_{i=1}^n a\left(\frac{i}{n}, \frac{S_i}{i+1}\right)$ is not optimal against [\(2.8\)](#).

THEOREM 2.4.13: The empirical fields z_R and z_S are asymptotically linear transformations of each other. Further $\psi(\mathbf{R}_n)$ and $\phi(\mathbf{S}_n)$ have the same limiting distribution under the null hypothesis and under any alternative [\(2.8\)](#) if

$$\phi(t, u) = \psi(t, u) - \frac{1}{t} \int_0^t \psi(\tau, u) d\tau$$

or

$$\psi(t, u) = \phi(t, u) - \int_t^1 \frac{1}{\tau} \phi(\tau, u) d\tau$$

For the linear rank statistics,

$$\sum_{i=1}^n a\left(\frac{i}{n}, \frac{R_{in}}{n+1}\right) \quad \text{and} \quad \sum_{i=1}^n \left(a\left(\frac{i}{n}, \frac{S_{in}}{i+1}\right) - \frac{n}{i} \int_0^{\frac{i}{n}} a\left(\tau, \frac{S_{in}}{i+1}\right) d\tau \right) \quad (2.9)$$

are asymptotically optimal test statistics against alternative (2.8).

PROOF: Refer to [KHMALADZE & PARJANADZE \(1986\)](#) and [PARJANADZE & KHMALADZE \(1986\)](#). □

EXAMPLE: Returning to the Wilcoxon rank statistic, we consider the sequential rank version of this statistic. From (2.9) it follows that

$$- \sum_{i=n_1+1}^n \frac{n_1}{i} \frac{S_i}{i+1}$$

is asymptotically equivalent. We also note that if the size of the first sample is fixed and we keep adding new observations to the second sample, then for the sequential rank statistic we would only need to add the new summands to the existing statistic.

BIBLIOGRAPHY

Billingsley, P. (1986), *Probability and Measure*, Probability and statistics, 2nd edn, Wiley. Cited on page [67](#).

Billingsley, P. (1999), *Convergence of Probability Measures*, Probability and statistics, 2nd edn, Wiley. Cited on pages [67](#), [73](#), [78](#), and [81](#).

Cantelli, F. P. (1933), ‘Sulla determinazione empirica della legge di probabilita.’, *Giorn. Ist. Ital. Attuari* **4**, 421–424. Cited on page [2](#).

Chung, K., Erdos, P. & Sirao, T. (1959), ‘On the Lipschitz’s condition for Brownian motion’, *Journal of the Mathematical Society of Japan* **11**(4). Cited on page [30](#).

Chung, K.-L. (1949), ‘An estimate concerning the Kolmogoroff limit distribution’, *Transactions of the American Mathematical Society* **67**(1), 36–50. Cited on page [28](#).

Doob, J. (1949), ‘Heuristic approach to the Kolmogorov-Smirnov theorems’, *Annals of Mathematical Statistics* **20**(3), 394–403. Cited on page [56](#).

Einmahl, J. (1993), Empirical and related processes with applications, Technical report, Eindhoven University of Technology. Cited on pages [13](#), [20](#), and [79](#).

Feller, W. (1971a), *An Introduction to Probability Theory and its Applications*, Vol. 2, Wiley. Cited on pages [21](#) and [49](#).

Feller, W. (1971b), *An Introduction to Probability Theory and its Applications*, Vol. 1, Wiley. Cited on page [29](#).

Glivenko, V. (1933), ‘Sulla determinazione empirica della legge di probabilita.’, *Giorn. Ist. Ital. Attuari* **4**, 92–99. Cited on page [2](#).

- Hajek, J., Sidak, Z. & Sen, P. (1999), *Theory of Rank Tests*, Probability and mathematical statistics, 2nd edn, Academic press. Cited on page [122](#).
- Hoeffding, W. (1963), 'Probability inequalities for sums of bounded random variables', *Journal of the American Statistical Association* **58**(301), 13–30. Cited on page [4](#).
- Hognas, G. (1986), 'Comparison of some non-linear autoregressive processes', *Journal of Time Series Analysis* **7**(3), 205 – 211. Cited on page [99](#).
- Ito, K. & McKean Jr, H. P. (1974), *Diffusion Processes and their Sample Paths*, Springer-Verlag. Cited on page [22](#).
- Khmaladze, E. & Koul, H. (2004), 'Martingale transforms goodness-of-fit tests in regression models', *The annals of statistics* **32**(3), 995–1034. Cited on page [103](#).
- Khmaladze, E. & Parjanadze, A. (1986), 'Functional limit theorems for linear statistics of sequential ranks', *Notes Sci. Seminars POMI* **204**, 115–125. Cited on page [123](#).
- Khmaladze, E. V. (2010), Sequential ranks, in 'Encyclopedia of statistics'. To appear in 2010. Cited on page [119](#).
- Kosorok, M. R. (2008), *Introduction to Empirical Processes and Semiparametric Inference*, Statistics and Statistical Theory and Methods, Springer New York. Cited on page [4](#).
- O'Brien, G. (1972), 'A note on comparisons of Markov processes.', *Ann. Math. Statist.* **43**, 365–368. Cited on page [99](#).
- Parjanadze, A. & Khmaladze, E. (1986), 'On the asymptotic theory of statistics based on sequential ranks', *Theory Probability Applications* **31**, 669–682. Cited on page [123](#).
- Pflug, G. (1991), 'A note on the comparison of stationary laws of Markov processes.', *Statist. Probab. Lett.* (4), 331–334. Cited on page [99](#).

Shorack, G. R. & Wellner, J. A. (1986), *Empirical Processes with Applications to Statistics*, Wiley. Cited on pages [29](#), [33](#), [34](#), and [49](#).

Smirnov, N. N. (1944), ‘An approximation to distribution laws of random quantities determined by empirical data.’, *Uspekhi Mat. Nauk.* **10**, 179–206. Cited on page [28](#).

van der Vaart, A. & Wellner, J. (1996), *Weak Convergence and Empirical Processes*, Springer. Cited on pages [10](#), [12](#), [81](#), [82](#), and [83](#).