Detecting Jumps in Stochastic Processes

Jason Leung

Master of Science

October, 2012
Dedicated to my grandmother,

who taught me how to
use the microscope.
Acknowledgement

First and foremost, I would like to thank my Supervisor, Professor Konstantin Borovkov for his valuable advice and the effort he made to help me not just finishing this thesis, but also improving it.

Besides, I would also like to give my thanks to my friends, sharing an office with me or not, for helping me to solve many problems I encountered in \LaTeX and simulations.

Finally I would like to express my gratitude to my parents, especially my mother, who supported me with everything they have.
Introduction

Detecting the presence of jumps in a discretely monitored continuous time process is becoming a more and more important problem. Especially in finance, as we see the advance of modern technology leading to fully automated high frequency trading. Analysis of high frequency data can improve trading strategies and relevant methods such as risk modelling and hedging.

We are interested in a specific class of process called the Merton-Kou model, that is, a jump diffusion model commonly used in the financial industry. Our goal is to construct a test statistic for hypothesis testing, which will decide if a continuous time process has continuous trajectories.

This thesis is organised as follows:

Chapter 1 reviews definitions and results from measure and probability theory, weak convergence and semimartingales, which we shall need in later chapters.

Chapter 2 is devoted to the construction of a test statistic, proof of Central Limit Theorems, formulation of the asymptotic distribution of our test statistic. Then we propose our null and alternative hypotheses. Two tests proposed in Aït-Sahalia and Jacod (2009) and Bharath et al. (2012) were also reviewed in the first section of Chapter 2.

Finally, in Chapter 3, we simulate the Merton-Kou Model and conduct hypothesis tests under both hypotheses to provide numerical evidence for our results in Chapter 2.
Contents

1 Preliminaries 7
  1.1 Mathematical foundation ......................................... 7
    1.1.1 Measures, Integrals and Derivatives ......................... 7
    1.1.2 Product Measures ........................................... 12
  1.2 Probability theory ................................................. 15
    1.2.1 Probabilities, Expectations, and Independence ................. 15
    1.2.2 Conditional Expectations .................................... 17
    1.2.3 Characteristic Functions .................................... 18
    1.2.4 Limit Theorems ............................................. 21
  1.3 Weak Convergence of Probability Measures ....................... 24
    1.3.1 Weak convergence in metric spaces, tightness and relative compactness ........................................ 24
    1.3.2 The space $C$ ............................................. 27
    1.3.3 The Functional Central Limit Theorem ....................... 29
    1.3.4 The space $D$ ............................................. 32
    1.3.5 Skorokhod Representation Theorem ........................... 34
  1.4 Semimartingales .................................................... 35
    1.4.1 Basic definitions ............................................ 35
    1.4.2 Quadratic variation of semimartingales ....................... 36

2 Detecting jumps in a discretely observed process 38
  2.1 Literature review .................................................. 38
    2.1.1 The powered variation method ................................ 39
    2.1.2 The split function method .................................. 43
  2.2 Another test ........................................................ 48
    2.2.1 Setting and assumptions .................................. 48
    2.2.2 The asymptotic behaviour of $S_n^{(e)}$ ....................... 49
    2.2.3 The test statistic ........................................... 59
    2.2.4 Approximating the distribution of $T_n$ ....................... 63
3 Simulation results 66
  3.1 The Central Limit Theorem 66
  3.2 Testing for the presence of jumps 68
    3.2.1 Under the null hypothesis $H_0$ 69
    3.2.2 Under the alternative hypothesis $H_1$ 69
    3.2.3 Conclusion 72

A Appendix 73
  A.1 Miscellaneous results from real analysis 73
  A.2 Mill’s Ratio 76
Chapter 1

Preliminaries

This chapter is devoted to presenting of definitions and results that will be used in later chapters and can be used as a quick reference for the reader. We will briefly review some basic measure theory in Section 1, then probability theory in the Section 2. Finally, in Section 3, we will go through some results on weak convergence.

1.1 Mathematical foundation

1.1.1 Measures, Integrals and Derivatives

Measures

In this section, we readily present the essential definitions and results needed in later chapters. The proofs are omitted as there are many existing texts, such as Koliha (2008) and Athreya and Lahiri (2006), in which detailed presentation of the proofs can be found.

Let us first define the notion of “measure of a set” on some suitable space of subsets.

Definition 1.1.1. Let \( S \) be a set and \( \mathcal{A} \) be a collection of subsets of \( S \). Then \( \mathcal{A} \) is an algebra if

(i) \( S \in \mathcal{A} \),

(ii) \( A \in \mathcal{A} \) implies \( A^c \in \mathcal{A} \),

(iii) \( A, B \in \mathcal{A} \) implies \( A \cup B \in \mathcal{A} \).
Definition 1.1.2. Let \( S \) be a set and \( \Sigma \) be a collection of subsets of \( S \). Then \( \Sigma \) is a \( \sigma \)-algebra if it is an algebra such that

\[
A_n \in \Sigma \text{ for all } n \geq 1 \text{ implies } \bigcup_{n \geq 1} A_n \in \Sigma.
\]

The pair \((S, \Sigma)\) is called a measurable space.

Suppose \( \mathcal{A} \) is a collection of subsets, then the smallest \( \sigma \)-algebra containing \( \mathcal{A} \) is called the \( \sigma \)-algebra generated by \( \mathcal{A} \), and is denoted by \( \sigma(\mathcal{A}) \).

Example 1.1.1. Let \((X, \mathcal{T})\) be a topological space, then the \( \sigma \)-algebra generated by \( \mathcal{T} \) is called the Borel \( \sigma \)-algebra of \( X \) and is denoted by \( \mathcal{B}(X) \). Elements of \( \mathcal{B}(X) \) are called Borel sets.

Definition 1.1.3. Let \( S \) be a set and \( \mathcal{A} \) be an algebra of subsets of \( S \). A measure on \( \mathcal{A} \) is a function \( \mu : \mathcal{A} \to \mathbb{R} \) satisfying

(i) \( \mu(A) \geq 0 \) for all \( A \in \mathcal{A} \),

(ii) \( \mu(\emptyset) = 0 \),

(iii) suppose \( A_1, A_2, A_3, \ldots \in \mathcal{A} \) is a disjoint collection of subsets such that \( \bigcup_{n \geq 1} A_n \in \mathcal{A} \), then

\[
\mu \left( \bigcup_{n \geq 1} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n).
\]

If \((S, \Sigma)\) is a measurable space, the triplet \((S, \Sigma, \mu)\) is called a measure space. If \( \mu(S) < \infty \), \( \mu \) is called a finite measure.

Definition 1.1.4. Let \( S \) be a set and \( \mathcal{A} \) be an algebra of subsets of \( S \). A measure \( \mu \) is \( \sigma \)-finite if there exist \( A_1, A_2, A_3, \ldots \in \Sigma \) such that

(i) \( S = \bigcup_{n \geq 1} A_n \),

(ii) \( \mu(A_n) \leq \infty \) for all \( n \geq 1 \).

Theorem 1.1.1. (Carathéodory’s Theorem) Let \( S \) be a set, \( \mathcal{A} \) be an algebra of subsets of \( S \) and \( \mu_0 \) be a \( \sigma \)-finite measure on \( \mathcal{A} \). Then there exists a unique measure \( \mu \) on the measurable space \((S, \sigma(\mathcal{A}))\) which is an extension of \( \mu_0 \), that is,

\[
\mu(A) = \mu_0(A), \quad \forall A \in \mathcal{A}.
\]
Measurable functions

**Definition 1.1.5.** Let $(S_1, \Sigma_1)$ and $(S_2, \Sigma_2)$ be measurable spaces. A function $f : S_1 \to S_2$ is $(\Sigma_1, \Sigma_2)$-measurable if

$$f^{-1}(A) \in \Sigma_1, \quad \forall A \in \Sigma_2.$$ 

Sometimes we write $f : (S_1, \Sigma_1) \to (S_2, \Sigma_2)$ to identify the respective $\sigma$-algebras. We may simply say $f$ is measurable if it is clear what underlying $\sigma$-algebras are used.

**Example 1.1.2.** Let $f : X \to Y$ be a $(\Sigma_1, B(Y))$-measurable function where $B(Y)$ is the Borel $\sigma$-algebra on $Y$. Then $f$ is called a Borel measurable function.

**Lemma 1.1.1.** Let $(S_1, \Sigma_1)$ and $(S_2, \Sigma_2)$ be measurable spaces and $\Sigma_2 = \sigma(A)$, where $A$ is some collection of subsets of $S_2$. Then a function $f : S_1 \to S_2$ is measurable if and only if

$$f^{-1}(A) \in \Sigma_1, \quad \forall A \in A.$$

The following result is an immediate consequence of the above lemma.

**Corollary 1.1.1.** Let $(X, B(X)), (Y, B(Y))$ be measurable spaces. If $f : X \to Y$ is a continuous function, then $f$ is measurable.

Integration

We will state some basic results for Lebesgue Integrals. The relevant definitions thereof are not given here as it may not be necessary. For the definition of the Lebesgue Integral, see any texts on measure theory such as Koliha (2008).

**Definition 1.1.6.** Let $f_n$ be functions defined on a measure space $(S, \Sigma, \mu)$. One say that $f_n$ converges almost everywhere to a function $f$ if there exists a set $A \in \Sigma$ such that $\mu(A^c) = 0$ and

$$f_n(x) \to f(x), \quad \forall x \in A.$$ 

The almost everywhere convergence of $f_n$ is denoted by $f_n \to f \ \mu$-a.e or $f_n \xrightarrow{\text{a.e.}} f$.

**Theorem 1.1.2.** (Monotone Convergence Theorem) Let $\{f_n\}$ be a monotone increasing sequence of non-negative measurable functions defined on a measure space $(S, \Sigma, \mu)$ that is, $f_{n+1} \geq f_n$ for $n \geq 1$, such that $f_n \to f \ \mu$-a.e. Then

$$\int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu.$$
**Theorem 1.1.3.** (Fatou’s lemma) Let \( \{f_n\} \) be a sequence of non-negative measurable functions on the measure space \((S, \Sigma, \mu)\). Then
\[
\liminf_{n \to \infty} \int f_n \, d\mu \geq \int \liminf_{n \to \infty} f_n \, d\mu.
\]

**Theorem 1.1.4.** (Lebesgue’s dominated convergence theorem) Let \( \{f_n\} \) be a sequence of measurable functions on the measure space \((S, \Sigma, \mu)\) such that \( f_n \to f \) \( \mu \)-a.e for some function \( f \). If there exists a measurable function \( g \) such that \( |f_n| \leq g \) \( \mu \)-a.e for all \( n \geq 1 \) and \( \int g \, d\mu < \infty \), then \( f \in L^1 \),
\[
\lim_{n \to \infty} \int f_n \, d\mu = \int f \, d\mu \quad \text{and} \quad \lim_{n \to \infty} \int |f_n - f| \, d\mu = 0.
\]

**More on convergence**

In this section, we will present different modes of convergence and the relevant results that are commonly used in probability theory.

**Definition 1.1.7.** Let \( \{f_n\} \) be a sequence of measurable functions on the measure space \((S, \Sigma, \mu)\). The sequence \( f_n \) converge to \( f \) in measure as \( n \to \infty \), denoted by \( f_n \overset{m}{\rightharpoonup} f \), if for every \( \varepsilon > 0 \),
\[
\lim_{n \to \infty} \mu(\{|f_n - f| > \varepsilon\}) = 0.
\]

**Theorem 1.1.5.** Let \( \{f_n\} \) be a sequence of measurable functions on the measure space \((S, \Sigma, \mu)\) such that \( f_n \overset{a.e.}{\to} f \) as \( n \to \infty \) and \( \mu(S) < \infty \). Then \( f_n \overset{m}{\rightharpoonup} f \).

**Theorem 1.1.6.** Let \( \{f_n\} \) be a sequence of measurable functions on the measure space \((S, \Sigma, \mu)\). If \( f_n \overset{m}{\rightharpoonup} f \), then there exists a subsequence, \( \{f_{n'}\} \), of \( \{f_n\} \) such that \( f_{n'} \overset{a.e.}{\to} f \).

**Definition 1.1.8.** Let \( \{f_n\} \) be a sequence of measurable functions on the measure space \((S, \Sigma, \mu)\) and \( p > 0 \). Then \( f_n \) converges to \( f \) in \( L^p \) as \( n \to \infty \), denoted by \( f_n \overset{L^p}{\to} f \), if
\[
\lim_{n \to \infty} \int |f_n - f|^p \, d\mu = 0.
\]

We denote \( L^p(S, \Sigma, \mu) = \{ f : S \to \bar{\mathbb{R}} | \int |f|^p \, d\mu < \infty \} \).

**Lemma 1.1.2.** (Markov inequality) Let \( f \) be a measurable function on the measure space \((S, \Sigma, \mu)\). Then, for any \( \varepsilon > 0 \),
\[
\mu(|f| > \varepsilon) \leq \frac{\int |f| \, d\mu}{\varepsilon}.
\]
From the Markov inequality, we immediately obtain the following,

**Theorem 1.1.7.** Let \( \{f_n\} \) be a sequence of measurable functions on the measure space \((S, \Sigma, \mu)\). If, as \( n \to \infty \), \( f_n \overset{L^p}{\to} f \) for some \( p > 0 \), then \( f_n \overset{m}{\to} f \).

**Differentiation**

In the following section, we will quickly review *Radon-Nikodym derivative* of a measure.

**Definition 1.1.9.** Let \((S, \Sigma)\) be a measurable space and let \( \mu \) and \( \nu \) be measures on \((S, \Sigma)\). The measure \( \mu \) is dominated by \( \nu \), or absolutely continuous with respect to \( \nu \), denoted by \( \mu \ll \nu \) if, for every \( A \in \Sigma \)

\[ \nu(A) = 0 \quad \Rightarrow \quad \mu(A) = 0. \]

**Example 1.1.3.** Let \( f \) be a nonnegative measurable function on a measure space \((S, \Sigma, \nu)\). Define

\[ \mu(A) = \int_A f \, d\nu, \quad A \in \Sigma. \]

Then \( \mu \) is a measure on \((s, \Sigma)\) and \( \nu(A) = 0 \) implies \( \mu(A) = 0 \) for all \( A \in \Sigma \). Therefore \( \mu \ll \nu \).

**Theorem 1.1.8.** *(Radon-Nikodym Theorem)* Let \((S, \Sigma)\) be a measurable space, \( \mu \) and \( \nu \) be \( \sigma \)-finite measures on \((S, \Sigma)\) such that \( \mu \ll \nu \). Then there exists a nonnegative measurable function \( f \) such that

\[ \mu(A) = \int_A f \, d\nu, \quad \forall A \in \Sigma. \]

**Definition 1.1.10.** Let \( \mu, \nu \) be measures on a measurable space \((S, \Sigma)\) and \( f \) be a function such that

\[ \mu(A) = \int_A f \, d\nu, \quad \forall A \in \Sigma. \]

Then \( f \) is the *Radon-Nikodym derivative* of \( \mu \) with respect to \( \nu \) and is denoted by

\[ \frac{d\mu}{d\nu}. \]
1.1.2 Product Measures

We will review product measurable spaces, product measures and the Fubini-Tonelli theorems in this section.

Definition 1.1.11. Let \((S, \Sigma), (T, \Omega)\) be measurable spaces.

(i) \(S \times T = \{(s, t) : s \in S, t \in T\}\), the set of all ordered pairs, is the 
Cartesian product of \(S\) and \(T\).

(ii) A measurable rectangle is a subset of \(S \times T\) of the following form, 
\(A_1 \times A_2\), such that \(A_1 \in \Sigma, A_2 \in \Omega\).

(iii) The product \(\sigma\)-algebra of \(\Sigma\) and \(\Omega\) on \(S \times T\), denoted by \(\Sigma \times \Omega\) is the 
\(\sigma\)-algebra generated by all the measurable rectangles, that is, 
\[\Sigma \times \Omega = \sigma(\{A_1 \times A_2 : A_1 \in \Sigma, A_2 \in \Omega\})\].

The pair \((S \times T, \Sigma \times \Omega)\) is called the product measurable space of \((S, \Sigma)\) 
and \((T, \Omega)\).

Definition 1.1.12. Let \((S \times T, \Sigma \times \Omega)\) be the product measurable space of 
\((S, \Sigma)\) and \((T, \Omega)\) and \(A \in \Sigma \times \Omega\). The s-section of \(A\), denoted by \(A_{s,T}\), is 
defined as 
\[A_{s,T} = \{t \in T : (s, t) \in A\}\].

Likewise, the t-section of \(A\), denoted by \(A_{S,t}\), is defined as 
\[A_{S,t} = \{s \in S : (s, t) \in A\}\].

The following shows that the section of a measurable set is measurable.

Proposition 1.1.1. Let \((S \times T, \Sigma \times \Omega)\) be the product measurable space of 
\((S, \Sigma)\) and \((T, \Omega)\) and \(A \in \Sigma \times \Omega\). Then \(A_{s,T} \in \Omega\) and \(A_{S,t} \in \Sigma\).

Theorem 1.1.9. Let \((S, \Sigma, \mu), (T, \Omega, \nu)\) be \(\sigma\)-finite measure spaces.

(i) For every \(A \in \Sigma \times \Omega\), the functions \(s \mapsto \nu(A_{s,T})\) and \(t \mapsto \mu(A_{S,t})\) are 
\((\Sigma, \mathcal{B}(\mathbb{R}))\)-measurable and \((\Omega, \mathcal{B}(\mathbb{R}))\)-measurable respectively.
(ii) Define, for every \( A \in \Sigma \times \Omega \),

\[
\pi_1(A) = \int_S \nu(A_{s,T})d\mu, \quad \pi_2(A) = \int_T \mu(A_{s,t})d\nu.
\]

Then \( \pi_1, \pi_2 \) are \( \sigma \)-finite measures on \( (S \times T, \Sigma \times \Omega) \) and \( \pi_1(A) = \pi_2(A) \).

Moreover, \( \pi = \pi_1 = \pi_2 \) is the only measure on \( (S \times T, \Sigma \times \Omega) \) satisfying

\[
\pi(A_1 \times A_2) = \mu(A_1)\nu(A_2) \quad \text{for all } A_1 \in \Sigma, A_2 \in \Omega.
\]

**Definition 1.1.13.** Let \((S, \Sigma, \mu), (T, \Omega, \nu)\) be measure spaces. The measure \( \pi \) on the measurable space \((S \times T, \Sigma \times \Omega)\) satisfying

\[
\pi(A_1 \times A_2) = \mu(A_1)\nu(A_2) \quad \text{for all } A_1 \in \Sigma, A_2 \in \Omega
\]

is called the **product measure** and is denoted by \( \mu \otimes \nu \).

We know from the preceding theorem that such a measure is unique.

The Fubini-Tonelli’s theorems presented below will allow us to change the order of integrations.

**Theorem 1.1.10.** *(Tonelli’s theorem)* Let \((S, \Sigma, \mu), (T, \Omega, \nu)\) be \( \sigma \)-finite measure spaces and \( f : S \times T \to \mathbb{R}_{\geq 0} \) be a \((\Sigma \times \Omega, \mathcal{B}(\mathbb{R}_{\geq 0}))\)-measurable function. Define \( g_1 : S \to \mathbb{R} \) and \( g_2 : T \to \mathbb{R} \) by

\[
g_1(s) = \int_T f(s, t)\nu(dt) \quad \text{and} \quad g_2(t) = \int_S f(s, t)\mu(ds).
\]

Then \( g_1 \) and \( g_2 \) are \((\Sigma, \mathcal{B}(\mathbb{R}))\)-measurable and \((\Omega, \mathcal{B}(\mathbb{R}))\)-measurable, respectively. Moreover,

\[
\int_{S \times T} f d(\mu \otimes \nu) = \int_S g_1 d\mu = \int_T g_2 d\nu.
\]

**Theorem 1.1.11.** *(Fubini’s theorem)* Let \((S, \Sigma, \mu), (T, \Omega, \nu)\) be \( \sigma \)-finite measure spaces and \( f \in L^1(S \times T, \Sigma \times \Omega, \mu \otimes \nu) \). Then there exists sets \( A_1 \in \Sigma \) and \( A_2 \in \Omega \) such that

(i) \( \mu(S \setminus A_1) = \nu(T \setminus A_2) = 0 \).

(ii) For any fixed \( s \in S \) and \( t \in T \), \( f(s, \cdot) \in L^1(T, \Omega, \nu) \) and \( f(\cdot, t) \in L^1(S, \Sigma, \mu) \).
(iii) Define $g_1 : S \to \mathbb{R}$, $g_2 : T \to \mathbb{R}$ by

$$g_1(s) = \begin{cases} \int_T f(s,t)\nu(dt) & \text{for } s \in A_1, \\ 0 & \text{for } s \in A_1^c, \end{cases}$$

and

$$g_2(t) = \begin{cases} \int_S f(s,t)\mu(dt) & \text{for } t \in A_2, \\ 0 & \text{for } t \in A_2^c. \end{cases}$$

Then $g_1$ and $g_2$ are $(\Sigma, \mathcal{B}(\mathbb{R}))$-measurable and $(\Omega, \mathcal{B}(\mathbb{R}))$-measurable, respectively, and

$$\int_{S \times T} f d(\mu \otimes \nu) = \int_S g_1 d\mu = \int_T g_2 d\nu.$$
1.2 Probability theory

This section is devoted to a summery of the basics of probability theory. The proofs of the results presented in this section can be found in many existing texts such as Shiryaev (1995), Chung (2000) and Athreya and Lahiri (2006).

1.2.1 Probabilities, Expectations, and Independence

First we will introduce probability spaces and probability measures then random elements and expectations.

Definition 1.2.1. Let $(\Omega, \mathcal{F}, P)$ be a measure space such that $P(\Omega) = 1$. Then $(\Omega, \mathcal{F}, P)$ is called a probability space, $P$ is a probability measure, and elements of $\mathcal{F}$ are called events.

Definition 1.2.2. Let $(\Omega, \mathcal{F}), (E, \mathcal{E})$ be measurable spaces. An $(\mathcal{F}, \mathcal{E})$-measurable function $X : \Omega \rightarrow E$ is called a random element

- If $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}(\mathbb{R})$, then $X$ is called a random variable and it is $\mathcal{F}$-measurable.
- If $E = \mathbb{R}^n$ and $\mathcal{E} = \mathcal{B}(\mathbb{R}^n)$, then $X$ is called a random vector and it is $\mathcal{F}$-measurable.
- If $E = \bar{\mathbb{R}}$ and $\mathcal{E} = \mathcal{B}(\bar{\mathbb{R}})$, then $X$ is called an extended random variable and it is $\mathcal{F}$-measurable.

Proposition 1.2.1. Let $X$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. Define

$$ P_X(B) = P(X \in B), \quad \forall B \in \mathcal{B}(\mathbb{R}). \quad (1.2.1) $$

Then $P_X$ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

The measure defined by (1.2.1) is called the measure induced by the random variable $X$ or just simply the induced measure. The function $F$ defined by $F(x) = P_X((\infty, x])$ is called the distribution function of $X$.

Suppose $X_1, X_2, X_3, \ldots$ is a sequence of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. Then $P$-almost everywhere convergence of $X_n$ to $X$ as $n \rightarrow \infty$ is also called almost sure convergence, denoted by $X_n \xrightarrow{a.s.} X$, or convergence with probability 1, denoted by $X_n \rightarrow X \text{ w.p.} 1$. Likewise, convergence in measure of $X_n$ to $X$ is also called convergence in probability, denoted by $X_n \xrightarrow{P} X$. 


**Definition 1.2.3.** Let $X$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. Then the **expectation of $X$**, denoted by $EX$, is defined by

$$EX = \int XdP.$$  

The **variance of $X$**, denoted by $\text{Var}(X)$, is defined by

$$\text{Var}(X) = E(X - EX)^2.$$  

Since the expectation of a random variable $X$ is defined as a Lebesgue integral, therefore it has all the properties that the Lebesgue integral possesses such as monotonicity and linearity.

We finish this section by reviewing the concept of independence.

**Definition 1.2.4.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and $B_1, B_2, \ldots, B_n \in \mathcal{F}$. Then $B_1, B_2, \ldots, B_n$ are **independent events w.r.t. $P$** if,

$$P\left(\bigcap_{j=1}^{k} B_{i_j}\right) = \prod_{j=1}^{k} P(B_{i_j}), \quad \forall 1 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq n.$$  

**Definition 1.2.5.** Let $(\Omega, \mathcal{F}, P)$ be a probability space.

(i) Let $\{B_\alpha : \alpha \in A\} \subseteq \mathcal{F}$ be a collection of events. Then events $\{B_\alpha : \alpha \in A\}$ are **independent w.r.t. $P$** if, for every finite subset $J \subseteq A$,

$$P\left(\bigcap_{\alpha \in J} B_\alpha\right) = \prod_{\alpha \in J} P(B_\alpha).$$  

(ii) Let $\{X_\alpha : \alpha \in A\}$ be a collection of random variables defined on $(\Omega, \mathcal{F})$. Then $\{X_\alpha : \alpha \in A\}$ is **independent w.r.t. $P$** if for every finite subset $\{\alpha_1, \alpha_2, \ldots, \alpha_k\} \subseteq A$,

$$P(X_{\alpha_1} \in B_{\alpha_1}, X_{\alpha_2} \in B_{\alpha_2}, \ldots, X_{\alpha_k} \in B_{\alpha_k}) = P(X_{\alpha_1} \in B_{\alpha_1})P(X_{\alpha_2} \in B_{\alpha_2}) \cdots P(X_{\alpha_k} \in B_{\alpha_k})$$

for all $B_{\alpha_1}, B_{\alpha_2}, \ldots, B_{\alpha_k} \in \mathcal{F}$.

$\{X_\alpha : \alpha \in A\}$ are called **independent variables**.

**Definition 1.2.6.** Let $(\Omega, \mathcal{F})$ be a measurable space and $\{A_n\}_{n \geq 1} \subseteq \mathcal{F}$. Then

$$\limsup_{n \to \infty} = \bigcap_{k \geq 1} \bigcup_{n \geq k} A_n \quad \text{and} \quad \liminf_{n \to \infty} = \bigcup_{k \geq 1} \bigcap_{n \geq k} A_n.$$  

16
1.2.2 Conditional Expectations

We briefly review the concept of conditional expectation with respect to a \( \sigma \)-algebra as the Radon-Nikodym derivative.

**Definition 1.2.7.** Let \((\Omega, \mathcal{F}, P)\) be a probability space. The *conditional expectation of a nonnegative random variable* \(X\) *with respect to a \(\sigma\)-algebra* \(\mathcal{G}\) is a nonnegative extended random variable, denoted by \(E(X|\mathcal{G})\), such that

(i) \(E(X|\mathcal{G})\) is \(\mathcal{G}\)-measurable,

(ii) for every \(A \in \mathcal{G}\),

\[
\int_A X dP = \int_A E(X|\mathcal{G}) dP.
\]

It is clear that the existence of the conditional expectation of a nonnegative random variable is guaranteed by the Radon-Nikodym theorem. To see this, we write

\[
Q(A) = \int_A X dP, \quad \forall A \in \mathcal{G}.
\]

Then \(Q\) is a measure and it is dominated by \(P\). Therefore the Radon-Nikodym theorem applies.

**Definition 1.2.8.** Let \((\Omega, \mathcal{F}, P)\) be a probability space. The *conditional expectation of a random variable* \(X\) *with respect to a \(\sigma\)-algebra* \(\mathcal{G}\), denoted by \(E(X|\mathcal{G})\), is defined provided that

\[
\min\{E(X^+|\mathcal{G}), E(X^-|\mathcal{G})\} < \infty \quad \text{almost surely},
\]

and is given by

\[
E(X|\mathcal{G}) = \begin{cases} 
E(X^+|\mathcal{G})(\omega) - E(X^-|\mathcal{G})(\omega) & \text{if } \min\{E(X^+|\mathcal{G}), E(X^-|\mathcal{G})\}(\omega) < \infty \\
\text{any arbitrary values} & \text{otherwise}.
\end{cases}
\]

**Theorem 1.2.1.** (Properties of conditional expectations) Let \((\Omega, \mathcal{F}, P)\) be a probability space \(\mathcal{G} \subseteq \mathcal{F}\) a \(\sigma\)-algebra and \(X, Y\) random variables defined on \((\Omega, \mathcal{F})\).

(i) If \(X = C\) a.s. for some \(C \in \mathbb{R}\), then \(E(X|\mathcal{G}) = C\) a.s.

(ii) If \(X \leq Y\) a.s., then \(E(X|\mathcal{G}) \leq E(Y|\mathcal{G})\) a.s.
(iii) $|E(X|\mathcal{G})| \leq E(|X||\mathcal{G})$ a.s.

(iv) If $a, b \in \mathbb{R}$ and $aEX + bEY$ is defined, then

$$E(aX + bY|\mathcal{G}) = aE(X|\mathcal{G}) + bE(Y|\mathcal{G}) \quad \text{a.s.}$$

(v) If $\mathcal{F} = \{\emptyset, \Omega\}$ is the trivial $\sigma$-algebra, then

$$E(X|\mathcal{F}) = EX \quad \text{a.s.}$$

(vi) $E(X|\mathcal{F}) = X$ a.s.

(vii) $EE(X|\mathcal{G}) = EX$.

(viii) If $\mathcal{G}_1, \mathcal{G}_2$ are $\sigma$-algebras such that $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then

$$E(E(X|\mathcal{G}_2)|\mathcal{G}_1) = E(X|\mathcal{G}_1) \quad \text{a.s.}$$

(ix) If $\mathcal{G}_1, \mathcal{G}_2$ are $\sigma$-algebras such that $\mathcal{G}_2 \subseteq \mathcal{G}_1$, then

$$E(E(X|\mathcal{G}_2)|\mathcal{G}_1) = E(X|\mathcal{G}_2) \quad \text{a.s.}$$

(x) Suppose $X$ is independent of the $\sigma$-algebra $\mathcal{G}$, that is, $\sigma(\{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\})$ and $\mathcal{G}$ are independent, then

$$E(X|\mathcal{G}) = EX \quad \text{a.s.}$$

(xi) Suppose $Y$ is $\mathcal{G}$-measurable, $E|X| < \infty$ and $E|XY| < \infty$ then

$$E(XY|\mathcal{G}) = YE(X|\mathcal{G}) \quad \text{a.s.}$$

1.2.3 Characteristic Functions

In the present section, some basic results mainly concerning one dimensional characteristic functions are given.

Definition 1.2.9. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X$ be a random variable defined on $(\Omega, \mathcal{F})$. Then its characteristic function $\varphi_X : \mathbb{R} \to \mathbb{C}$ is

$$\varphi_X(t) = Ee^{itX}.$$
Since

\[ E e^{itX} = \int e^{itX} dP = \int \mathbb{R} e^{itx} P_X(dx) \]

where \( P_X \) is the induced measure, \( \varphi \) can also be written as

\[ \varphi_X(t) = \int \mathbb{R} e^{itx} dF_X(x), \]

where \( F_X \) is the distribution function of \( X \).

If the density of \( X \), \( f_X \), exists, then

\[ \varphi_X(t) = \int \mathbb{R} e^{itx} f_X(x) dx. \]

In other words, \( \varphi_X \) is the Fourier Transform of density \( f_X \).

**Definition 1.2.10.** Let \((\Omega, \mathcal{F}, P)\) be a probability space and \( X = (X_1, X_2, \ldots, X_n) \) be a random vector defined on \((\Omega, \mathcal{F})\). Then its characteristic function \( \varphi_X : \mathbb{R}^n \to \mathbb{C} \), is defined by

\[ \varphi_X(t) = E e^{it(X)} = \int_{\mathbb{R}^n} e^{itx} dF(x), \]

where \( F = F(x_1, x_2, \ldots, x_n) \) is the distribution function of \((X_1, X_2, \ldots, X_n)\) and \( \langle \cdot, \cdot \rangle : \mathbb{R}^2 \to \mathbb{R} \) is the usual Euclidean scalar product.

**Theorem 1.2.2.** (Properties of the characteristic function) Let \( X \) be a random variable with distribution function \( F \) and \( \varphi \) be its characteristic function. Then

(i) \( |\varphi(t)| \leq \varphi(0) = 1 \).

(ii) \( \varphi(t) \) is uniformly continuous for \( t \in \mathbb{R} \).

(iii) \( \varphi(t) = \overline{\varphi(-t)} \).

(iv) \( \varphi(t) \) is real-valued if and only if \( F \) is symmetric.

(v) If the \( 2n \)-th derivative \( \varphi^{(2n)}(t) \) exists and is finite, then \( EX^{2n} < \infty \) for \( n \geq 0 \).

The uniqueness theorem presented below states that a distribution function is uniquely determined by its characteristic function.
Theorem 1.2.3. (Uniqueness) Let \( F \) and \( G \) be distribution functions such that
\[
\int_{-\infty}^{\infty} e^{itx} dF(x) = \int_{-\infty}^{\infty} e^{itx} dG(x), \quad \forall t \in \mathbb{R},
\]
that is, they have the same characteristic function. Then \( F = G \).

The next theorem allows us to calculate the distribution function from its characteristic function.

Theorem 1.2.4. (Inversion formula) Let \( F \) be a distribution function and \( \varphi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x) \) be its characteristic function.

(i) For points \( a < b \) such that at which \( F \) is continuous,
\[
F(b) - F(a) = \lim_{c \to \infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt.
\]

(ii) If \( \int_{-\infty}^{\infty} |\varphi(t)| dt < \infty \), the distribution function \( F \) has a density \( f \),
\[
F(x) = \int_{-\infty}^{x} f(y) dy
\]
and
\[
f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt.
\]

From the uniqueness theorem, we can readily deduce the following necessary and sufficient condition for independence of random variables.

Theorem 1.2.5. A necessary and sufficient condition for the components of a random vector \( X = (X_1, X_2, \ldots, X_n) \) to be independent is:
\[
E e^{i(t_1 X_1 + t_2 X_2 + \cdots + t_n X_n)} = \prod_{k=1}^{n} E e^{i_k X_k}, \quad \forall (t_1, t_2, \ldots, t_n) \in \mathbb{R}^n.
\]
1.2.4 Limit Theorems

We will present the Laws of Large Numbers and the Central Limit Theorem in this section. Let us start with the formal definition of “convergence in distribution”.

**Definition 1.2.11.** Let \( \{X_n\}_{n \geq 0} \) be a sequence of random variables with distribution functions \( F_n, n \geq 1 \). Then \( X_n \) converges in distribution to \( X \) with distribution \( F \), denoted by \( X_n \xrightarrow{d} X \) as \( n \to \infty \), if

\[
F_n(x) \to F(x), \quad \forall x \in C(F)
\]

where \( C(F) = \{x \in \mathbb{R} : F \text{ is continuous at } x\} \).

**Theorem 1.2.6** (Strong Law of Large Numbers). Let \( \{X_n\}_{n \geq 1} \) be a sequence of i.i.d. random variables such that \( E|X_1| < \infty \). Then

\[
\overline{X}_n \xrightarrow{a.s.} EX_1
\]

where \( \overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k \).

The following result is on the uniform convergence of the empirical cdf to the true cdf.

**Theorem 1.2.7** (Glivenko-Cantelli’s theorem). Let \( \{X_n\}_{n \geq 1} \) be a sequence of i.i.d. random variables with distribution function \( F \). Let \( F_n \) be the empirical distribution function of \( X_1, X_2, \ldots, X_n \), defined by

\[
F_n(x) := \frac{1}{n} \sum_{j=1}^{n} 1(X_j \leq x), \quad x \in \mathbb{R}.
\]

Then,

\[
\sup_x |F_n(x) - F(x)| \xrightarrow{a.s.} 0, \quad n \to \infty.
\]

**Definition 1.2.12.** For every \( n \geq 1 \), let \( \{X_{n1}, X_{n2}, \ldots, X_{nr_n}\} \) be a collection of random variables defined on a probability space \( (\Omega_n, \mathcal{F}_n, P_n) \) such that \( \{X_{n1}, X_{n2}, \ldots, X_{nr_n}\} \) are independent and \( r_n \to \infty \) as \( n \to \infty \). Then \( \{X_{nj} : 1 \leq j \leq r_n\}_{n \geq 1} \) is called a triangular array of row-wise independent random variables.
Example 1.2.1.

\[
X_{1,1}, X_{2,1}, X_{2,2}, X_{3,1}, X_{3,2}, X_{3,3}, \ldots, X_{n,1}, X_{n,2}, X_{n,3}, \ldots, X_{n,n}
\]

where for each \(i \geq 1\), the random variables \(X_{i,1}, X_{i,2}, X_{i,3}, \ldots\) are i.i.d.

**Definition 1.2.13.** Let \(\{X_{n1}, X_{n2}, \ldots, X_{nr_n}\}\) be a triangular array of row-wise independent random variables such that

\[
EX_{nj} = 0, \quad EX^2_{nj} = \sigma^2_{nj} < \infty \quad \forall \ 1 \leq j \leq r_n, \ n \geq 1
\]

Then \(\{X_{n1}, X_{n2}, \ldots, X_{nr_n}\}\) is said to satisfy the Lindeberg condition if for every \(\varepsilon > 0\),

\[
\lim_{n \to \infty} s_n^{-2} \sum_{j=1}^{r_n} EX^2_{nj} 1(|X_{nj}| > \varepsilon s_n) = 0, \tag{1.2.2}
\]

where \(s^2_n = \sum_{nj}^r \sigma^2_{nj}, \ n \geq 1\).

**Theorem 1.2.8.** (Lindeberg Central Limit Theorem) Let \(\{X_{n1}, X_{n2}, \ldots, X_{nr_n}\}\) be a triangular array of row-wise independent random variables satisfying the Lindeberg condition (1.2.2). Then

\[
\frac{S_n}{s_n} \xrightarrow{d} Z, \quad n \to \infty,
\]

where \(Z \sim N(0, 1)\), \(S_n = \sum_{j=1}^{r_n} X_{nj}\) and \(s^2_n = Var(S_n) = \sum_{nj}^r X_{nj} \sigma^2_{nj}\).

**Definition 1.2.14.** Let \(\{X_{n1}, X_{n2}, \ldots, X_{nr_n}\}\) be a triangular array of row-wise independent random variables such that

\[
EX_{nj} = 0, \quad EX^2_{nj} = \sigma^2_{nj} < \infty \quad \forall \ 1 \leq j \leq r_n, \ n \geq 1
\]

Then \(\{X_{n1}, X_{n2}, \ldots, X_{nr_n}\}\) is said to satisfy Lyapounov’s condition if there exists a \(\delta > 0\) such that

\[
\lim_{n \to \infty} s_n^{-(2+\delta)} \sum_{j=1}^{r_n} E|X_{nj}|^{2+\delta} = 0. \tag{1.2.3}
\]
**Theorem 1.2.9.** (Lyapounov’s Central Limit Theorem) Let \( \{X_{n1}, X_{n2}, \ldots, X_{nr_n}\} \) be a triangular array of row-wise independent random variables satisfying Lyapounov’s condition (1.2.3). Then, as \( n \to \infty \),

\[
\frac{S_n}{s_n} \xrightarrow{d} Z
\]

where \( Z \sim N(0, 1) \), \( S_n = \sum_{j=1}^{r_n} X_{nj} \) and \( s_n^2 = \text{Var}(S_n) = \sum_{nj} \sigma_{nj}^2 \).
1.3 Weak Convergence of Probability Measures

We will study the weak convergence of probability measures on metric spaces roughly following Billingsley (1968). The aim is to present results of the space $C$ and the Functional Central Limit Theorem. The space $D$ and results analogous to those of the space $C$ will also be briefly reviewed. Proofs are not provided as this is intended to be a summary and quick reference for the reader. See Billingsley (1968) for the proofs and further details.

1.3.1 Weak convergence in metric spaces, tightness and relative compactness

We devote this section to the concepts of weak convergence, tightness and relative compactness.

Let $(S, \rho)$ be a metric space and $\mathcal{T}$ be the topology generated by the metric $\rho$. We call the the $\sigma$-algebra generated by $\mathcal{T}$, denoted by $\sigma(\mathcal{T})$, the Borel $\sigma$-algebra of $S$ and its elements Borel sets. In this section, $\mathcal{S}$ will be defined as the Borel $\sigma$-algebra of $S$ unless otherwise specified.

Denoted $C(S)$ to be the set of bounded continuous functions from $S$ to $\mathbb{R}$, where $\mathbb{R}$ is equipped with the usual Euclidean metric $d$. Corollary 1.1.1 tells us that elements of $C(S)$ are measurable so it makes sense to introduce the following definition.

**Definition 1.3.1.** The sequence of probability measures $\{P_n\}_{n \geq 1}$ on $(S, \mathcal{S})$ converges weakly to $P$ as $n \to \infty$, (denoted as $P_n \xrightarrow{w} P$), if

$$\int_S fdP_n \to \int_S fdP, \quad \forall f \in C(S).$$

**Definition 1.3.2.** A probability measure on $(S, \mathcal{S})$ is regular if for any $A \in \mathcal{S}$ and $\varepsilon > 0$, there exist a closed set $F$ and an open set $G$ such that $F \subseteq A \subseteq G$ and

$$P(G \setminus F) < \varepsilon.$$

**Theorem 1.3.1.** Every probability measure on $(S, \mathcal{S})$ is regular.

The preceding theorem shows that the value of $P(A)$ for any $A \in \mathcal{S}$ can be taken as the limit of $P(F_n)$ for $F_n \in \mathcal{S}$ closed, $n \geq 1$. Therefore it is sufficient to know all the values of $P(F)$ for $F$ closed.

The following theorem provides a sufficient condition of two measures to coincide.
Theorem 1.3.2. Two probability measures $P$, $Q$ on $(S, \mathcal{S})$ coincide if

$$\int f \text{d}P = \int f \text{d}Q, \quad \forall f \in C(S)$$

Introduce the notion of “tightness” for a collection of probability measures.

Definition 1.3.3. A set of probability measures $\Pi$ on $(S, \mathcal{S})$ is tight if for any $\varepsilon > 0$, there exists a compact set $K \in S$ such that

$$\inf_{P \in \Pi} P(K) > 1 - \varepsilon.$$ 

It will be natural to ask if a single probability measure is tight. Indeed, the following result shows that provided the sample space is “nice” enough, any probability measure on it is tight.

Theorem 1.3.3. If $S$ is a Polish space, then a probability measure $P$ on $(S, \mathcal{S})$ is tight.

Note that weak convergence depends on the topology of $S$. As shown in Theorem 1.3.2, $P$ is determined by the integrals $\int f \text{d}P$ for $f \in C(S)$ and continuity is a topological property of a function. Therefore two metrics generating the same topology give the same class of bounded continuous functions and the same Borel $\sigma$-algebra.

The theorem presented below provides sufficient and necessary conditions for weak convergence.

Theorem 1.3.4. (Portmanteau theorem) Let $\{P_n\}_{n \geq 1}$ be a sequence of probability measures on $(S, \mathcal{S})$. Then the followings are equivalent.

(i) $P_n \xrightarrow{w} P$, $n \to \infty$.

(ii) $\int f \text{d}P_n \to \int f \text{d}P$ for all $f \in C(S)$ as $n \to \infty$.

(iii) $\limsup_{n \to \infty} P_n(F) \leq P(F)$ for all closed sets $F$.

(iv) $\liminf_{n \to \infty} P_n(G) \geq P(G)$ for all open sets $G$.

(v) $\lim_{n \to \infty} P_n(A) = P(A)$ for all $A \in \mathcal{S}$ such that $P(\partial A) = 0$.

We call a set $A \in \mathcal{S}$ such that $P(\partial A) = 0$ a $P$-continuity set where $\partial A$ denotes the boundary of $A$.

Note that as $\partial A = \overline{A} \setminus A^o$ and since the closure of $A$, $\overline{A}$, is closed and the interior if $A$, $A^o$, is open we have $\partial A \in \mathcal{S}$. So $\partial A$ is measurable.

We now give the definition of relative compactness.
**Definition 1.3.4.** Let $\Pi$ be a family of probability measures on $(S, S)$. Then $\Pi$ is relatively compact if every sequence in $\Pi$ has a weakly convergent subsequence.

That is, if $\{P_n\}_{n \geq 1} \subseteq \Pi$, then there exists $\{P_{n'}\}_{n' \geq 1} \subseteq \{P_n\}_{n \geq 1}$ and a probability measure $P$ on $(S, S)$ such that $P_n \xrightarrow{w} P$. Note that the weak limit $P$ need not to be in $\Pi$.

**Example 1.3.1.** Consider $P_n = \delta_n$, the Dirac delta measure on $(\mathbb{R}, B(\mathbb{R}))$. That is, for $A \in B(\mathbb{R})$,

$$
\delta_n(A) = \begin{cases} 
1 & \text{if } n \in A \\
0 & \text{if } n \notin A
\end{cases}
$$

Then $\{P_n\}_{n \geq 1}$ is neither tight nor relatively compact.

To see this, first suppose it is tight. Then there is a compact set $[-K, K] \subset \mathbb{R}$ such that $P_n([-K, K]) = 1$ for all $n \geq 1$. This contradicts the fact that for $n > K$, $P_n([-K, K]) = 0$.

To see that $\{P_n\}_{n \geq 1}$ is not relatively compact, suppose it is relatively compact. There there is a subsequence $\{P_{n'}\}_{n' \geq 1} \subseteq \{P_n\}_{n \geq 1}$ and $P$ such that $P_{n'} \xrightarrow{w} P$ as $n' \to \infty$. Since

$$
\liminf_{n' \to \infty} P_{n'}((-k, k)) = 0 \quad \forall k,
$$

and by Portmanteau theorem (Theorem 1.3.4) we know that

$$
P((-k, k)) \leq \liminf_{n' \to \infty} P_{n'}((-k, k)) \leq 0 \quad \forall k.
$$

This contradicts the fact that $P$ is a probability measure, so the relative compactness cannot hold.

In the above example, the probability is concentrated on a “point” and such point “escapes to infinity” so that we can not capture it.

The following theorem shows that the condition we need for having no “escaping mass” is tightness.

**Theorem 1.3.5.** (Prokhorov’s theorem) Let $\Pi$ be a collection of probability measures on $(S, S)$. If $\Pi$ is tight, then $\Pi$ is relatively compact.

We also want to know under what conditions relative compactness will be equivalent to tightness. The following result provides such conditions and acts as a converse to Prokhorov’s theorem.

**Theorem 1.3.6.** Suppose $S$ is a Polish space and the family to probability measures $\Pi$ is relatively compact. Then $\Pi$ is tight.
1.3.2 The space $C$

In this section, we consider the space

$$C = \{ f \mid f : [0, 1] \to \mathbb{R}, f \text{ continuous} \}$$

Since $[0, 1]$ is compact in $\mathbb{R}$, one has $\sup_{x \in [0,1]} |f(x) - g(x)| < \infty$, in fact, this defines a metric on $C$.

**Lemma 1.3.1.** Let $\rho : C \times C \to \mathbb{R}$ be such that

$$\rho(f, g) = \sup_{x \in [0,1]} |f(x) - g(x)|.$$

Then $\rho$ is a metric.

Furthermore, we can show that $(C, \rho)$ is a complete and separable metric space.

**Theorem 1.3.7.** $(C, \rho)$ is a Polish space.

Since the metric space $(C, \rho)$ is Polish, every probability measure on the Borel $\sigma$-algebra of $C$ is tight.

Next, introduce the notion of projection and finite dimensional sets of $(C, \rho)$.

**Definition 1.3.5.**

(i) A projection of $(C, \rho)$ is a function $\pi_{t_1, t_2, \ldots, t_k} : C \to \mathbb{R}^k$, $0 \leq t_1 < t_2 < \ldots < t_k \leq 1$, $k \geq 1$, such that

$$\pi_{t_1, t_2, \ldots, t_k}(x) = (x(t_1), x(t_2), \ldots, x(t_k)).$$

(ii) A subset $A$ of $C$ is called a finite dimensional set if there is a projection $\pi_{t_1, t_2, \ldots, t_k}$ and a subset $H$ of $\mathbb{R}^k$ such that

$$A = \pi_{t_1, t_2, \ldots, t_k}^{-1}(H).$$

Suppose $\mathbb{R}^k$ is equipped with the ordinary Euclidean metric $d$, which we assume throughout this section. We have the following.

**Lemma 1.3.2.** $\pi_{t_1, t_2, \ldots, t_k} : C \to \mathbb{R}^k$ is continuous.

Denote the Borel $\sigma$-algebra of $C$ and that of $\mathbb{R}^k$ on $C$ and $\mathcal{B}(\mathbb{R}^k)$, respectively.

As a consequence of the continuity of $\pi_{t_1, t_2, \ldots, t_k}^{-1}$, we obtain the following

**Corollary 1.3.1.** $\pi_{t_1, t_2, \ldots, t_k} : C \to \mathbb{R}^k$ is $(C, \mathcal{B}(\mathbb{R}^k))$-measurable.
Now, introduce finite dimensional distributions of a probability measure.

**Definition 1.3.6.** Let $P$ be a probability measure on $(C, C)$. A finite dimensional distribution of $P$ is a probability measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ defined by

$$P_{\pi^{-1}_{t_1, t_2, \ldots, t_k}}(B) = P(\pi^{-1}_{t_1, t_2, \ldots, t_k}(B)), \quad \forall B \in \mathcal{B}(\mathbb{R}^k),$$

for some projection $\pi^{-1}_{t_1, t_2, \ldots, t_k}$.

Consider a sequence of probability measures $\{P_n\}_{n \geq 1}$ on $(C, C)$. Suppose $P_n \xrightarrow{w} P$ as $n \to \infty$ for some probability measure $P$ and $\pi^{-1}_{t_1, t_2, \ldots, t_k} : C \to \mathbb{R}^k$ is a projection map on $C$. Since $\pi_{t_1, t_2, \ldots, t_k}$ is measurable, $\pi^{-1}_{t_1, t_2, \ldots, t_k}(B) \in C$ for $B \in \mathcal{B}(\mathbb{R}^k)$. So the finite dimensional distribution of a probability measure is itself a probability measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, and for all bounded continuous functions $f : \mathbb{R}^k \to \mathbb{R}$,

$$\int fdP_n \pi^{-1}_{t_1, t_2, \ldots, t_k} \to \int f \circ \pi^{-1}_{t_1, t_2, \ldots, t_k} dP = \int fdP \pi^{-1}_{t_1, t_2, \ldots, t_k},$$

as $f \circ \pi^{-1}_{t_1, t_2, \ldots, t_k} : C \to \mathbb{R}$ is bounded continuous and $P_n \xrightarrow{w} P$. Therefore $P_n \pi^{-1}_{t_1, t_2, \ldots, t_k} \xrightarrow{w} P \pi^{-1}_{t_1, t_2, \ldots, t_k}$ as $n \to \infty$.

Summarising the above discussion, we obtain the following result.

**Theorem 1.3.8.** Let $\{P_n\}_{n \geq 1}$, $P$ be probability measures on $(C, C)$. If $P_n \xrightarrow{w} P$ as $n \to \infty$, then their respective finite dimensional distributions converge weakly.

Note that the converse of the above theorem is not always true.

Suppose we have two probability measures $P$ and $Q$ on $(C, C)$, such that all their finite dimensional distributions coincide. It is natural to raise the question of whether $P$ and $Q$ coincide. The following result answers such question.

**Theorem 1.3.9.** If the finite dimensional distribution of two probability measures $P$ and $Q$ on $(C, C)$ coincide, then $P$ and $Q$ coincide.

The following results give a criterion for weak convergence via finite dimensional distributions.

**Theorem 1.3.10.** Let $\{P_n\}_{n \geq 1}$, $P$ be probability measures on $(C, C)$. If the finite dimensional distributions of $P_n$ converge weakly to that of $P$ as $n \to \infty$ and $\{P_n\}_{n \geq 1}$ is tight, then $P_n$ converge weakly to $P$. 

28
1.3.3 The Functional Central Limit Theorem

In this section, we will give the definition of the Wiener Measure and review the functional central limit theorem.

Random Functions

Let \((\Omega, \mathcal{F}, P)\) be a probability space, \(X : C \rightarrow \mathbb{R}^k\) be a \((\mathcal{F}, \mathcal{C})\)-measurable function and \(\pi : C \rightarrow \mathbb{R}^k\) be a projection map \(\pi(x) = \pi_{t_1, t_2, \ldots, t_k} = (x(t_1), x(t_2), \ldots, x(t_k))\) where \(0 \leq t_1 < t_2 < \ldots, t_k \leq 1\).

Since \(\pi\) is continuous and hence \((C, \mathcal{B}(\mathbb{R}^k))\)-measurable, \(\pi^{-1}(B) \in \mathcal{C}\) for \(B \in \mathcal{B}(\mathbb{R}^k)\). Thus \(X^{-1}(\pi^{-1}(B)) \in \mathcal{C}\) for \(B \in \mathcal{B}(\mathbb{R}^k)\).

\[\Omega \xrightarrow{X} C \xrightarrow{\pi} \mathbb{R}^k\]

Figure 1.1: Fix \(\omega \in \Omega, X(\omega) \in C\), hence \(\pi_t \circ X(\omega) : [0, 1] \rightarrow \mathbb{R}\).

Let us agree to call \(X\) a Random Function and denote \(\pi_t \circ X\) by \(X_t\) or \(X(t)\). Obviously, for all \(0 \leq t \leq 1\), \(X_t\) is a random variable if \(X\) is a random function. The following shows that the converse is true if \(X_t\) is a random variable for all \(t \in [0, 1]\).

**Theorem 1.3.11.** Let \((\Omega, \mathcal{F}, P)\) be a probability space and \(X : \Omega \rightarrow C\). The \(X\) is a random function if and only if \(X_t : \Omega \rightarrow \mathbb{R}\) is a random variable for all \(0 \leq t \leq 1\).

Wiener Measure

Since the projection map \(\pi_t : C \rightarrow \mathbb{R}\) is \((C, \mathcal{B}(\mathbb{R}))\)-measurable, \(\pi_t\) is a random variable defined on \(C\) for \(0 \leq t \leq 1\). Therefore, it makes sense to define the following

**Definition 1.3.7.** The Wiener Measure, \(W\), is a probability measure on the measurable space \((C, \mathcal{C})\) such that

1. \(W(\pi_0 = 0) = 1\)
(ii) \( \mathcal{W}(\pi_t \leq \alpha) = \int_{-\alpha}^{\alpha} \frac{1}{\sqrt{2\pi t}} e^{-\frac{u^2}{2t}} du \)

(iii) \( \pi_{t_1} - \pi_{t_0}, \pi_{t_1} - \pi_{t_1}, \ldots, \pi_{t_k} - \pi_{t_{k-1}} \) are independent random variables for \( 1 \leq t_0 \leq t_0 \leq \ldots \leq t_k \leq 1 \).

where \( \pi_t : C \to \mathbb{R} \) is a projection map.

**Theorem 1.3.12.** Wiener measure exists.

We now define a Wiener process and a Brownian Bridge.

**Definition 1.3.8.** A **Wiener Process** is a stochastic process \( \{W_t\}_{t \geq 0} \) that satisfies the following,

(i) \( W(0) = 0, \)

(ii) \( W(t) \sim N(0, t), \) where \( N(\mu, \sigma^2) \) is the normal distribution with mean \( \mu \) and variance \( \sigma^2, \)

(iii) \( W(t_1) - W(t_0), W(t_2) - W(t_1), \ldots, W(t_n), W(t_{n-1}) \) are independent random variables for \( 0 \leq t_0 < t_1 < \ldots < t_n. \)

A process is **Gaussian** if its finite dimensional distributions are multivariate normally distributed.

The Wiener process is Gaussian, has continuous trajectories and

\[
E(W_t) = 0, \quad Cov(W_t, W_s) = \min\{s, t\}. \tag{1.3.1}
\]

**Definition 1.3.9.** A **Brownian Bridge process** is a stochastic process \( \{W^\circ_t\}_{t \in [0,1]} \) defined by

\[
W^\circ_t = W_t - tW_1 \tag{1.3.2}
\]

where \( \{W_t\}_{t \in [0,1]} \) is a Wiener process.

Likewise, a Brownian Bridge process is Gaussian, has continuous trajectories and

\[
E(W^\circ_t) = 0, \quad Cov(W^\circ_s, W^\circ_t) = \min\{s, t\}(1 - \max\{s, t\}). \tag{1.3.3}
\]
Donsker’s Invariance Principle

Let \((\Omega, \mathcal{F}, P)\) be a probability space and \(\xi_1, \xi_2, \ldots, \xi_n\) be i.i.d. random variables on \(\Omega\) such that \(E\xi_j = 0\) and \(Var\xi_j = \sigma^2\) for all \(j \geq 1\).

Define for each \(n \geq 1\), \(S_n := \xi_1 + \xi_2 + \ldots + \xi_n\) with \(S_0 = 0\) and \(X_n(t) : \Omega \to \mathbb{R}\) be such that

\[
X_n(t) = \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) \frac{1}{\sqrt{n}} \xi_{\lfloor nt \rfloor + 1}, \quad t \in [0, 1].
\]

That is, \(X_n\) is the scaled linear interpolation of the partial sums.

Since \(\xi_1, \xi_2, \ldots, \xi_n\) are random variables, therefore \(X_n(t)\) is a random variable for all \(t \in [0, 1]\). Hence \(X_n(\cdot) : \Omega \to C\) is a random function as \(X_n(\cdot) \in C\) for all \(\omega \in \Omega\).

Suppose \(\eta_1, \eta_2, \ldots, \eta_n\) are i.i.d. random variables with \(E\eta_1 = \mu\) and \(Var\eta_1 = \sigma^2\). By the central limit theorem, as \(n \to \infty\),

\[
\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z,
\]

where \(Z\) is a standard normal random variable.

It is natural to expect that the analogous version of central limit theorem holds for processes with trajectories defined by (1.3.4). This is the Donsker’s Invariance Principle or the so called Functional Central Limit Theorem, stated as below.
Theorem 1.3.13. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\xi_1, \xi_2, \ldots, \xi_n$ be independent and identically distributed random variables with $E\xi_1 = 0$ and $\text{Var}\xi_1 = \sigma^2$. For $n \geq 1$, define the random function $X_n : \Omega \to \mathcal{C}$ as in (1.3.4). Then the induced measure $PX_n^{-1}$ converge weakly to the Wiener Measure $W$ as $n \to \infty$.

Denote the weak convergence of the induced measures by $X_n \overset{D}{\to} W$.

where $W$ denotes a Wiener measure.

The theorem above can be stated informally as the scaled linear interpolations of partial sums on the interval $[0, 1]$ will eventually become a Wiener process, or a Brownian Motion.

1.3.4 The space $D$

We will give a summary of some results on the space $D$ in this section.

Unlike the Wiener process, many processes contain jumps, such as the Poisson process. So it is necessary to expand the space of continuous functions to the set of càdlàg functions, that is, functions continuous from the right and having the left limits at all points.

Denote by $D$ the set of all càdlàg functions on $[0, 1]$. Obviously, $C \subset D$.

The Skorokhod Topology

In the present section, we introduce the Skorokhod metric and the Skorokhod topology.

Let $d : D \times D \to \mathbb{R}$ to be a function defined by

$$d(x, y) = \inf \left\{ \varepsilon \geq 0 : \exists \lambda \in \Lambda, \sup_{t \in [0, 1]} |\lambda(t) - t| \leq \varepsilon, \rho(x(t) - y(\lambda(t))) \leq \varepsilon \right\}$$

(1.3.5)

where $\Lambda = \{ \lambda : [0, 1] \to [0, 1] : \lambda \text{ is a homeomorphism, } \lambda(0) = 0, \lambda(1) = 1 \}$.

Lemma 1.3.3. $(D, d)$ is a metric space.

$d$ allows us to “deform the time scale” by a small amount, so we can not measure time with perfect accuracy any more than we can position.

Since $d$ is a metric on the space $D$, it generates a topology. This topology is called the Skorokhod topology and $d$ is called the Skorokhod metric. In the space $C$, convergence under the Skorokhod metric implies convergence under the uniform metric $\rho$ as defined in the previous section. In fact, they generate the same topology on the space $C$. 

32
Theorem 1.3.14. The Skorokhod topology relativised to $C$ coincides with the topology generated by the uniform metric.

Lemma 1.3.4. The metric space $(D, d)$ is separable.

However, unlike the metric space $(C, \rho)$, the space $(D, d)$ is not complete. We shall introduce another metric with which $D$ is complete and which generates the same topology as $d$ does.

Define $d_0 : D \to D$ to be a function such that

$$d_0(x, y) = \inf \left\{ \varepsilon \geq 0 : \sup_{s, t \in [0, 1], s \neq t} \left| \frac{\lambda(t) - \lambda(s)}{t - s} \right| \leq \varepsilon, \sup_{0 \leq t \leq 1} |x(t) - y(\lambda(t))| \leq \varepsilon \right\}$$

(1.3.6)

Lemma 1.3.5. $(D, d_0)$ is a metric space.

Furthermore, convergence under $d_0$ implies convergence under $d$. Therefore, convergence of a sequence in $C$ under $d_0$ implies convergence under the uniform metric.

Lemma 1.3.6. Metrics $d$ and $d_0$ generate the same topology on $D$.

It can be shown that $(D, d_0)$ is a complete metric space.

Theorem 1.3.15. $(D, d_0)$ is a Polish metric space.

Donsker’s Invariance Principle

The Functional Central Limit Theorem can be formulated for the space $D$.

Theorem 1.3.16. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\xi_1, \xi_2, \ldots, \xi_n$ be i.i.d. random variables with $E\xi_1 = 0$, $\text{Var}\xi_1 = \sigma^2$ and $X_n$ be a random function defined by

$$X_n(t) = \frac{1}{\sigma \sqrt{n}} S_{\lfloor nt \rfloor}.$$  

Then

$$PX_n^{-1} \xrightarrow{w} W, \quad n \to \infty,$$

where $W$ is a Wiener process.
Empirical Distribution Functions

**Theorem 1.3.17.** Let $\xi_1, \xi_2, \ldots, \xi_n$ be i.i.d. random variables with distribution function $F$ and empirical distribution function $F_n$. Then

$$\sqrt{n}(F_n(t) - F(t)) \xrightarrow{D} W^\circ,$$

where $W^\circ = \{W^\circ_{F(t)}\}$ is a Brownian Bridge process.

### 1.3.5 Skorokhod Representation Theorem

The following result was proved in Skorokhod (1956).

**Theorem 1.3.18** (Skorokhod Representation Theorem). Let $(X, \Sigma)$ be a measurable space where $X$ is Polish under some metric $\rho$ and $\{\xi_n\}_{n \geq 1}$ be random elements taking values on $X$. Suppose $\xi_n \xrightarrow{D} \xi$ as $n \to \infty$ for some random element $\xi_0$. Then there exist a probability space $(\Omega, \mathcal{F}, P)$ and $\{x_n\}_{n \geq 0}$ from $\Omega$ to $X$ such that

(i) $\xi_n \overset{d}{=} x_n$ in distribution for all $n \geq 0$.

(ii) $x_n$ converge to $x_0$ $P$-a.e. as $n \to \infty$.

This result provides us a way to represent a sequence of random elements, which converge in distribution, by a sequence of almost sure convergent sequence of random elements.
1.4 Semimartingales

Definitions and results regarding to semimartingales and its quadratic variation are given in this section. For more detailed exposure of this topic, see Jacod and Shiryaev (1987), Protter (2005).

1.4.1 Basic definitions

**Definition 1.4.1.** Let $\mathcal{F}$ be a $\sigma$-algebra of subsets of $\Omega$. A filtration $(\mathcal{F})_{t \geq 0}$ is a family of $\sigma$-algebras on $\Omega$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $s < t$.

$(\Omega, \mathcal{F}, (\mathcal{F})_{t \geq 0}, P)$, where $P$ is a probability measure, is called a filtered space.

A stochastic process $\{X_t\}_{t \geq 0}$ defined on this filtered space is adapted if for every $t \geq 0$, $X_t$ is $\mathcal{F}_t$-measurable.

**Definition 1.4.2.** A stochastic process $\{X_t\}_{t \geq 0}$ adapted to the filtration $(\mathcal{F})_{t \geq 0}$ is a martingale if $E|X_t| < \infty$ for every $t \geq 0$ and for any $s < t$,

$$E(X_t|\mathcal{F}_s) = X_s.$$ 

**Definition 1.4.3.** A non-negative random variable $\tau$ defined on a $(\Omega, \mathcal{F}, (\mathcal{F})_{t \geq 0}, P)$ is a stopping time if $P(\tau < \infty) = 1$ and for any $t \geq 0$

$$\{\tau \leq t\} \in \mathcal{F}_t.$$ 

**Definition 1.4.4.** Let $(\Omega, \mathcal{F}, (\mathcal{F})_{t \geq 0}, P)$ be a filtered space and $\{X_t\}_{t \geq 0}$ is a process adapted to $(\mathcal{F})_{t \geq 0}$. $\{X_t\}_{t \geq 0}$ is a local martingale if there exists a sequence of stopping time $\{\tau_n\}_n$ such that

(i) $P(\tau_k \leq \tau_k + 1) = 1$ for all $k \geq 1$.

(ii) $\tau_n$ diverge almost surely as $n \to \infty$.

(iii) The stopped process $X_t^{\tau_k} := X_{\min(t, \tau_k)}$ is a martingale for all $k \geq 1$.

**Definition 1.4.5.** Let $f: \mathbb{R} \to \mathbb{R}$ be a function. The variation of $f$ on $[a, b]$ is defined by

$$V_f([a, b]) := \sup \sum_{i=1}^n |f(t^n_i) - f(t^n_{i-1})|,$$

where the supremum is taken over all the partitions $a = t^n_0 \leq t^n_1 \leq \cdots \leq t^n_n = b$. 

35
Definition 1.4.6. A process $X$ is of *finite variation* if its sample paths are right-continuous and have finite variation over every compact time interval, almost surely.

Example 1.4.1. Define a process $X$ by

$$X_t = \sum_{k=0}^{N_t} J_k, \quad (1.4.1)$$

where $J_1, J_2, J_3, \ldots$ are i.i.d. random variables, $N_t$ a Poisson process with rate $\lambda$, all random variables are independent of each other. Then $X$ is of finite variation.

Define a semimartingale as follow,

Definition 1.4.7. A *semimartingale* $X$ is a càdlàg adapted process such that $X$ can be represented as

$$X_t = X_0 + M_t + A_t, \quad (1.4.2)$$

where $M_t$ is a local martingale and $A_t$ is a process of finite variation with $M_0 = 0$ and $A_0 = 0$.

Semimartingales form the largest class of processes with respect to which the Itô integral can be defined and they are regarded as “good integrators”.

Example 1.4.2. A Wiener process $W$ is a semimartingale.

1.4.2 Quadratic variation of semimartingales

Definition 1.4.8. Let $X, Y$ be semimartingales defined on a filtered space, then the *quadratic covariance process* is defined by

$$[X, Y]_t = \lim_{n \to \infty} \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}), \quad (1.4.3)$$

where the limit is in probability and is taken over partition $\{t_i^n\}_{i=0}^n$ of the interval $[0, t]$ such that $\max_{0 \leq i \leq n} (t_{i+1} - t_i) \to 0$ as $n \to \infty$.

If $Y = X$, then $[X, X]_t$ is called the *quadratic variation process* of $X$.

Example 1.4.3. Let $W$ be a Wiener process, then $[W, W]_t = t$.

It follows from definition that the quadratic variations are symmetric and linear.
Proposition 1.4.1. Let \(X,Y,Z\) be semimartingales defined on a filtered space, then

(i) (Symmetry) \([X,Y] = [Y,X]\).

(ii) (Linearity) \([\alpha X + \beta Y, Z] = \alpha [X, Z] + \beta [Y, Z]\), \(\alpha, \beta \in \mathbb{R}\).

Lemma 1.4.1. Let \(X, Y\) be semimartingales defined on a filtered space and \(X\) is of finite variation, then

\[[X, Y]_t = \sum_{s \leq t} \Delta X_s \Delta Y_s,
\]

where \(\Delta X_s = X_s - X_{s-}\) and \(\Delta Y_s = Y_s - Y_{s-}\).

Example 1.4.4. Let \(X\) be a processes defined by (1.4.1), then

\[[X, X]_t = \sum_{k=0}^{N_t} J_k^2.
\]

Example 1.4.5 (The Merton/Kou Model). Let \(X\) be a process,

\[X_t = \mu t + \sigma W_t + \sum_{i=1}^{N_t} J_i, \quad 0 \leq t \leq T,\]

where \(\mu \in \mathbb{R}, \sigma \in \mathbb{R}_{>0}, N_t\) a Poisson process with rate \(\lambda\), \(J_1, J_2, J_3, \ldots\) are i.i.d. random variables, and \(\{W_t\}_{t \geq 0}, \{N_t\}_{t \geq 0}, \{J_i : i \geq 1\}\) are all independent of each other. Then

\[[X, X]_t = \sigma^2 t + \sum_{k=0}^{N_t} J_k^2.
\]
Chapter 2

Detecting jumps in a discretely observed process

Detecting the presence of jumps in a discretely monitored continuous time process is becoming a more and more important problem. Especially in finance, as we see the advance of modern technology leading to fully automated high frequency trading. Analysis of high frequency data can improve trading strategies and relevant methods such as risk modelling. One fundamental problem in this context is to determine whether the set of data we received coming from a process with continuous sample path, or a process with jumps.

Suppose we have a continuous time process $X_t$, and we observe the values of this process up to time $T$ at times $\Delta_n, 2\Delta_n, 3\Delta_n, \ldots$ where $\Delta_n = T/n$, $n \in \mathbb{N}$. We want to construct a statistical test for the presence of jumps. This problem was considered by a number of authors using different approaches and assumptions made about the underlying processes.

2.1 Literature review

Aït-Sahalia and Jacod (2009) proposed a test which is valid for all Itô semi-martingales which depends neither on the law of the process nor on any of the parameters, such as drift and diffusion coefficients, specifying the process. Therefore no preliminary estimations of parameters is needed. On the other hand, Bharath et al. (2012) suggested another test based on a clustering framework and showed the validity of such a test on the popular Merton and Kou model for asset pricing.

In this section, we will briefly review the methods suggested in Aït-Sahalia and Jacod (2009) and Bharath et al. (2012) to be called “The powered variation method” and “The split function method”.

38
2.1.1 The powered variation method

1. Setting and assumptions

Aït-Sahalia and Jacod (2009) assumed that the underlying model $X$ is an Itô semimartingale on some filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. That is, $X_t$ can be written as

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_E \kappa \circ \delta(s, x)(\mu - \nu)(ds, dx) + \int_0^t \int_E \kappa' \circ \delta(s, x)\mu(ds, dx),$$

where $W$ is a Wiener process, $\mu$ a Poisson random measure on $\mathbb{R}_+ \times E$ (with $(E, \mathcal{E})$ an auxiliary measurable space) and no assumptions about the dependence of $\mu$ and $W$ were made. Further, $\nu(ds, dx) = ds \otimes \lambda(dx)$, where $\lambda$ is a given finite or $\sigma$-finite measure on $(E, \mathcal{E})$, being the compensator of $\mu$. As usual, “$\circ$” denotes composition of functions.

Moreover, we need $\kappa$ to be a continuous function with compact support and $\kappa(x) = 0$ on a neighbourhood of 0, $\kappa'(x) = x - \kappa(x)$. The functions $b_t(\omega), \delta(\omega, t, x)$ and $\sigma_t(\omega)$ are such that the above integrals are well defined. Furthermore, $\sigma_t$ is also assumed to be an Itô semimartingale of the form

$$\sigma_t = \sigma_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \tilde{\sigma}'_s dW'_s + \int_0^t \int_E \kappa \circ \tilde{\delta}(s, x)(\mu - \nu)(ds, dx) + \int_0^t \int_E \kappa' \circ \tilde{\delta}(s, x)(\mu - \nu)(ds, dx),$$

where $W'$ is another Wiener process independent of $W$. Additional technical assumptions made on $b_t, \tilde{\sigma}_t, \tilde{\sigma}'_t$ and $\delta$ can be found in Aït-Sahalia and Jacod (2009).

2. Measuring the variability of $X_t$

Since our goal is to detect the presence of jumps in a process, it makes sense to first look at the variability of the process we are testing. The authors introduced a number of processes which all provide some kind of measure for variability. For $p > 0$, let

$$A(p)_t = \int_0^t |\sigma_s|^p ds, \quad B(p)_t = \sum_{s \leq t} |\Delta X_s|^p,$$
where $\Delta X_s = X_s - X_{s-}$. Since there are at most countable number of discontinuities, the sum above is well defined.

So if $X$ has a jump in $(0, t]$ then we have $B(p)_t > 0$. Therefore detecting the presence of jumps is equivalent to determining whether $B(p)_t > 0$. The authors of the paper suggested a natural estimator for $B(p)_t$, of the form

$$
\hat{B}(p, \Delta n)_t := \sum_{i=1}^{\lfloor t/\Delta n \rfloor} |\Delta n,i X|_p,
$$

(2.1.3)

where $\Delta n,i X = X_{i\Delta n} - X_{(i-1)\Delta n}$, $\Delta n \to 0$ as $n \to \infty$ for all $1 \leq i \leq \lfloor t/\Delta n \rfloor$ and $\lfloor x \rfloor$ denotes the integer part of $x$.

It is proved in the paper that the asymptotic behaviour for $\hat{B}(p, \Delta n)_t$ as $n \to \infty$ will depend on the value of $p$:

$$
\begin{align*}
\hat{B}(p, \Delta n)_t & \xrightarrow{P} B(p)_t & \text{if } p \geq 2, \\
\Delta n^{-1/2} \frac{p}{m_p} B(p, \Delta n)_t & \xrightarrow{P} A(p)_t & \text{if } p < 2, \\
\Delta n^{-1/2} \frac{p}{m_p} \hat{B}(p, \Delta n)_t & \xrightarrow{P} A(p)_t & \text{for any } p > 0 \text{ and if } X_t \text{ is continuous},
\end{align*}
$$

(2.1.4)

where $m_r = \mathbb{E}(|U|^r) = \pi^{-1/2} 2^{r/2} \Gamma(r+1)$ for $r \in (0, \infty), U \sim N(0, 1)$

Intuitively, suppose our process $X_t$ has jumps. The first convergence in (2.1.4) can be interpreted as follows: when we have high powers ($p \geq 2$), large jumps are “magnified” while the small jumps are being “diminished”. In the second convergence relation, small jumps are “magnified” at the expense of large jumps. Therefore we see that the second convergence gives the same limit (in probability) as in the third convergence, which takes place when our process $X_t$ is continuous.

3. The test statistics and its asymptotic behaviour

Observe that the second and third convergence require a normalisation which depends on $\Delta n$ while the first convergence does not. From this observation, the authors proposed the following test statistics,

$$
\hat{S}(p, k, \Delta n)_t = \frac{\hat{B}(p, k\Delta n)_t}{\hat{B}(p, \Delta n)_t}
$$

(2.1.5)

Introduce notation

$$
\begin{align*}
\Omega^j_t & = \{ \omega : s \mapsto X_s(\omega) \text{ is discontinuous on } [0, t] \}, \\
\Omega^c_t & = \{ \omega : s \mapsto X_s(\omega) \text{ is continuous on } [0, t] \}.
\end{align*}
$$

(2.1.6)
That is, if we are on $\Omega^j_t$, the process trajectory has jumps on $[0,t]$. From (2.1.4) we immediately get

**Theorem 2.1.1.** Let $t > 0, p > 2$ and $k \geq 2$. Then $\hat{S}(p,k,\Delta_n)_t$ converges in probability to $S(p,k)_t$ defined by

$$ S(p,k)_t = \begin{cases} 1 & \text{on the set } \Omega^j_t, \\ k^{p/2-1} & \text{on the set } \Omega^c_t. \end{cases} $$

Then the authors established the Central Limit Theorem for the test statistics, stated as follows

**Theorem 2.1.2.** (a) Let $p > 3$ and $t > 0$. Then, as $n \to \infty$, $\Delta_n^{-1/2}(\hat{S}(p,k,\Delta_n)_t - 1)$ converges in law, in restriction to the set $\Omega^j_t$, to a variable $S(p,k)_t$ defined on an extension $(\Omega, \tilde{F}, (\tilde{F}_t)_{t \geq 0}, \tilde{P})$ of the original filtered space $(\Omega, F, (F_t)_{t \geq 0}, P)$, which, conditionally on $F$, is centred with variance

$$ \tilde{E}((S(p,k)_t)^2|F) = \frac{(k - 1)p^2 D(2p - 2) - 2}{B(p^2)_t}, $$

where

$$ D(p)_t = \sum_{s \leq t} |\Delta X_s|^p (\sigma^2_{s-} + \sigma^2_s). $$

Moreover, if the processes $\sigma_t$ and $X_t$ have no common jumps, the variable $S(p,k)_t$ is $F$-conditionally Gaussian.

(b) Assume in addition that $X_t$ is continuous and let $p \geq 2$ and $t > 0$. Then $\Delta_n^{-1/2}(\hat{S}(p,k,\Delta_n)_t - k^{p/2-1})$ converges stably in law to a variable $S(p,k)_t$ which, conditionally on $F$, is centred normal with variance

$$ \tilde{E}((S(p,k)_t)^2|F) = M(p,k) \frac{A(2p)_t}{A(p^2)_t}, $$

where

$$ M(p,k) = \frac{1}{m^2_p} (k^{p-2}(1+k)m_{2p}^2 + k^{p-2}(k-1)m_p^2 - 2k^{p/2-1}m_{k,p}), $$

$$ m_{k,p} = \mathbb{E}(|U|^p|U + \sqrt{k-1}V|^p), $$

$U, V \sim N(0,1)$ being independent random variables.

To evaluate the level of tests based on the test statistics proposed above, we need consistent estimators for $A(p)_t$ and $D(p)_t$. The detail of estimating $A(p)_t$ and $D(p)_t$ and their respective asymptotic behaviour is not discussed.
here and can be found in Aït-Sahalia and Jacod (2009). However, we provide an extract of the numerical results from 5000 simulations, with the null hypothesis that $X$ is continuous on $[0, T]$.

\[ \hat{S}(4, k, \Delta_n) \]

<table>
<thead>
<tr>
<th>$\Delta_n$</th>
<th>$n$</th>
<th>$k$</th>
<th>Rejection rate (0.1)</th>
<th>Rejection rate (0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 sec</td>
<td>23400</td>
<td>2</td>
<td>0.098</td>
<td>0.046</td>
</tr>
<tr>
<td>1 sec</td>
<td>23400</td>
<td>3</td>
<td>0.099</td>
<td>0.048</td>
</tr>
<tr>
<td>5 sec</td>
<td>4680</td>
<td>2</td>
<td>0.099</td>
<td>0.044</td>
</tr>
<tr>
<td>5 sec</td>
<td>4680</td>
<td>3</td>
<td>0.099</td>
<td>0.043</td>
</tr>
<tr>
<td>10 sec</td>
<td>2340</td>
<td>2</td>
<td>0.090</td>
<td>0.041</td>
</tr>
<tr>
<td>15 sec</td>
<td>1560</td>
<td>2</td>
<td>0.090</td>
<td>0.043</td>
</tr>
<tr>
<td>30 sec</td>
<td>780</td>
<td>2</td>
<td>0.085</td>
<td>0.035</td>
</tr>
</tbody>
</table>

As $n$ grows, the rejection rate converge to the prescribe level of the test.
2.1.2 The split function method

Instead of the more “natural” $p$-powered variation approach introduced in Aït-Sahalia and Jacod (2009), Bharath et al. (2012) developed a clustering framework for testing the presence of jumps in a continuous time model. This framework is based on the so-called “Split function” introduced in Hartigan (1978).

Let $\xi_1, \xi_2, \ldots, \xi_n$ be continuous i.i.d. random variables with cumulative distribution function $F$ such that

1. $F$ is invertible for $0 < p < 1$ and absolutely continuous with density $f$.
2. $E(\xi_1) = 0$ and $E(\xi_1^2) = 1$.
3. $F^{-1}$ is twice differentiable at any $p \in (0, 1)$.

(Invertible means $F^{-1}$ exists and $F \circ F^{-1} = F^{-1} \circ F = I$, $I$ is the identity map.)

1. Split function and cross-over function

As in Hartigan (1978), the split function of $F^{-1}$ at $p$ is defined as

$$B(F, p) = p(F^{-1}_i(p))^2 + (1 - p)(F^{-1}_u(p))^2 - \left( \int_0^1 F^{-1}(q)dp \right)^2,$$

where

$$F^{-1}_i(p) = \frac{1}{p} \int_{q \leq p} F^{-1}(q)dq \quad \text{and} \quad F^{-1}_u(p) = \frac{1}{1 - p} \int_{q > p} F^{-1}(q)dq.$$

and the split point, $p_0$, is arg max $B(F, p)$

The split function corresponds to a measure of distance between two clusters, and such a distance can be interpreted as the between cluster sum of squares. Note that $F(\xi_i) \sim U(0, 1)$ for $i = 1, 2, \ldots, n$, where $U(0, 1)$ is the uniform distribution. So $\frac{1}{p} \int_{q \leq p} F^{-1}(q)dq$ can be thought as the “centre of mass” distributed, with distribution $F$, on the line $[0, p]$ and likewise $\frac{1}{1 - p} \int_{q > p} F^{-1}(q)dq$ measures the “centre of mass” on the line $[p, 1]$. If the mass is “equally distributed”, the centre of mass would be in the mid-point.

Note that $F_i(p) = E(\xi_1|\xi_1 \leq q_p(\xi_1))$, where $q_p(\xi_1)$ is the $p$th quantile of $W$.

The purpose of introducing the split function was to construct a tool value for cluster separation. If the value of $B(F, p)$ is large at point $p$, it means the separation between clusters is large. Differentiating $B(F, p)$ with respect
to $p$, we see that the value $p_0$ maximising or minimising $B(F, p)$ satisfies the following equation:

\[(F_u^{-1}(p_0) - F_l^{-1}(p_0)) [F_l^{-1}(p_0) + F_u^{-1}(p_0) - 2F^{-1}(p_0)] = 0. \quad (2.1.14)\]

Since $F_u^{-1}(p_0) - F_l^{-1}(p_0) > 0$, we only have to consider the cross-over function

\[G(p) = F_l^{-1}(p_0) + F_u^{-1}(p_0) - 2F^{-1}(p_0) \quad (2.1.15)\]

and solve the equation $G(p) = 0$ for $p$.

In order to estimate the cross-over function, the authors proposed an estimator, called the Empirical Cross-over Function, defined as follows:

\[G_n(p) = \frac{1}{k} \sum_{j=1}^{k} W(j) - W(k) + \frac{1}{n-k} \sum_{j=k+1}^{n} W(j) - W(k+1), \quad (2.1.16)\]

where $W(j)$ is the $j$th order statistic of $W_1, W_2, \ldots, W_n$.

Moreover, we want to estimate the split point $p_0$, which can be done by the Empirical Split Point introduced by the authors as

\[p_n = \frac{1}{n} \max \left\{ 1 \leq k \leq n : G_n \left( \frac{k}{n} \right) \geq 0 \right\}. \quad (2.1.17)\]

2. The Asymptotic behaviour of $G_n$ and $p_n$

In Bharath et al. (2011), the authors showed the consistency and asymptotic normality of $G_n$.

**Theorem 2.1.3.** (i) Under the assumptions A1 and A2, as $n \to \infty$,

\[G_n(p) \xrightarrow{P} G(p).\]

(ii) Under the assumptions A1 – A3, as $n \to \infty$,

\[\sqrt{n}(G_n(p) - G(p)) \xrightarrow{d} N(0, \sigma^2),\]

where $\sigma^2 = \text{Var}(\theta_p)$ and

\[\theta_p = \frac{1}{p} W_1 \mathbb{1}_{W_1 < F^{-1}(p)} - F^{-1}(p) \mathbb{1}_{W_1 < F^{-1}(p)} + \frac{1}{1-p} W_1 - F^{-1}(p) \mathbb{1}_{W_1 \geq F^{-1}(p)} - 2 \frac{p - \mathbb{1}_{W_1 \geq F^{-1}(p)}}{f(F^{-1}(p))},\]

$\mathbb{1}(\cdot)$ is the indicator function.
Moreover the authors established the asymptotic behaviour of $p_n$ as follows.

**Theorem 2.1.4.** Suppose $p_0$ is the unique split point and $G'(p_0) \neq 0$. Under assumptions A1–A3, as $n \to \infty$,

$$
\sqrt{n}(p_n - p_0) \xrightarrow{d} N\left(0, \frac{\text{Var}(\theta_{p_0})}{G''(p_0)}\right),
$$

where $\theta_{p_0}$ is as defined in Theorem 2.1.3.

3. Illustration of the testing procedure

Bharath et al. (2012) demonstrated the testing procedure for the Merton-Kou model

$$
X_t = X_0 + \mu t + \sigma W_t + \sum_{k=0}^{N_t(\lambda)} J_k, \quad 0 \leq t \leq T, \tag{2.1.18}
$$

where $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_{\geq 0}$, $W_t$ is a Wiener Process, $N_t(\lambda)$ is a Poisson process, and $J_k$ are i.i.d. random variables, all random variables being independent of each others.

The authors assumed that the jump sizes are deterministic of constant size $h$. Therefore, (2.1.18) becomes

$$
X_t = X_0 + \mu t + \sigma W_t + hN_t(\lambda), \quad 0 \leq t \leq T. \tag{2.1.19}
$$

Let $\Delta_{n,i}X = X_{(i+1)\Delta_n} - X_{i\Delta_n}$ for $0 \leq i \leq n-1$, where $\Delta_n = T/n$. Since $X_t$ is a Lévy process, it has independent stationary increments and so $\Delta_{n,i}X$ are i.i.d. random variables, and we can just let $\Delta_nX \overset{d}{=} \Delta_{n,i}X$ for $0 \leq i \leq n$. Therefore since

$$
P(\mu\Delta_n + \sigma W_{\Delta_n} + hN_{\Delta_n}(\lambda) \leq x)
= \sum_{k=0}^{\infty} P(\mu\Delta_n + \sigma W_{\Delta} + hN_{\Delta_n}(\lambda) \leq x | N_{\Delta_n}(\lambda) = k) P(N_{\Delta_n}(\lambda) = k)
= \sum_{k=0}^{\infty} P(\mu\Delta_n + \sigma W_{\Delta} + hk \leq x) P(N_{\Delta_n}(\lambda) = k).
$$

That implies that $\Delta_{n,i}X$ has density

$$
f(x) = \sum_{k=0}^{\infty} \phi_k(x) \frac{e^{-\lambda \Delta_n(\lambda \Delta_n)^k}}{k!},
$$
where $\phi_k$ is the density of $N(\mu \Delta_n + hk, \sigma^2 \Delta_n^2)$. Notice that we have an infinite mixture of Gaussian densities, not mixture of finitely many densities. However,

$$f(x) = \phi_0(x)e^{-\lambda \Delta_n} + \phi_1(x)e^{-\lambda \Delta_n} + \sum_{k=2}^{\infty} \phi_k(x) \frac{e^{-\lambda \Delta_n} (\lambda \Delta_n)^k}{k!}$$

and

$$\sum_{k=2}^{\infty} \phi_k(x) \frac{e^{-\lambda \Delta_n} (\lambda \Delta_n)^k}{k!} = o(\Delta_n) \text{ as } \Delta_n \to 0.$$  

Moreover, as $\Delta_n \to 0$, $e^{-\lambda \Delta_n} = 1 - \lambda \Delta_n + o(\Delta_n)$ and $\lambda \Delta_n e^{-\lambda \Delta_n} \sim \lambda \Delta_n$. So $e^{-\lambda \Delta_n} = 1 - \lambda \Delta_n + o(\Delta_n)$ and $\lambda \Delta_n e^{-\lambda \Delta_n} = \lambda \Delta_n + o(\Delta_n)$. Hence, $f$ can be rewritten as

$$f(x) = (1 - \lambda \Delta_n) \phi_0(x) + \lambda \Delta_n \phi_1(x) + o(\Delta_n).$$

Since $\Delta_n \to 0$ as $n \to \infty$, when $n$ is large, we can instead consider the following density,

$$g(x) = (1 - \lambda \Delta_n) \phi_0(x) + \lambda \Delta_n \phi_1(x). \quad (2.1.20)$$

Therefore we have reduced the infinite mixture to a mixture of two densities and the problem of determining the presence of jumps reduced to determining if the observations come from a mixture of normal densities or not. However, note that the mixing proportion in this problem goes to 0 as $n \to \infty$. Therefore, the existing methods for classical mixtures do not work in this case.

If the underlying model $X$ is continuous, we do not have a mixture, that is, $\Delta_n, X$ will be just normally distributed. Due to the fact that the normal density is symmetric, $p = 1/2$ solves the equation $G(p) = 0$ for the cross-over function (2.1.15). Therefore the true split point, in this case is $p_0 = 1/2$.

Thus, provided we have consistency and asymptotic behaviour of the empirical split point $p_n$, the author proposed the following test statistics:

$$S_n = \sqrt{n} \left( \frac{G'(1/2)(p_n - 1/2)}{\sqrt{\text{Var}(\theta_{1/2})}} \right), \quad (2.1.21)$$

where $\theta_p$ and $G'(p)$ are as defined in the previous section. The following result is an immediate consequence of Theorem 2.1.4,

**Lemma 2.1.1.** As $n \to \infty$, $S_n \xrightarrow{d} Z$

where $Z \sim N(0,1)$.  

46
Therefore we obtained the following decision rule:

\[
\zeta(n, t, \alpha) = \begin{cases} 
X_t = X_0 + \mu t + \sigma W_t & \text{if } |S_n| \leq z_{\alpha/2} \\
X_t = X_0 + \mu t + \sigma W_t + hP_t(\lambda) & \text{otherwise}
\end{cases}
\]  

(2.1.22)

**Remark 2.1.1.** One may have concern about the separability of clusters, that is, whether we can tell there is a clear separation between clusters. It is being discussed in Bharath et al. (2012) and we shall not reproduce their arguments here.

As shown in this illustration, this test requires prior knowledge about the alternative hypothesis. That is, the distribution of the jump size. Also this test will not work if the jump size \( J_k \) has symmetric density.
2.2 Another test

2.2.1 Setting and assumptions

Suppose we have a jump diffusion process, that is, a continuous time process $X_t$ of the following form:

$$X_t = \mu t + \sigma W_t + \sum_{i=1}^{N_t} J_i, \quad 0 \leq t \leq T,$$

(2.2.1)

where $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_{>0}$, $N_t$ a Poisson process with rate $\lambda$, $J_1, J_2, J_3, \ldots$ are i.i.d. random variables, and $\{W_t\}_{t \geq 0}, \{N_t\}_{t \geq 0}, \{J_i : i \geq 1\}$ are all independent of each other.

Let $\Delta_n = T/n$ and $\Delta_{n,i}X = X_{\Delta_n} - X_{(i-1)\Delta_n}$ for $i = 1, 2, \ldots, n$. Since $X_t$ is a Lévy process, $\Delta_{n,1}X, \Delta_{n,2}X, \ldots, \Delta_{n,n}X$ are i.i.d. random variables. We want to develop a test similar to what was obtained in Aït-Sahalia and Jacod (2009).

Observe that we have a finite jump intensity $\lambda$, hence when $n$ is large, only a very small proportion of the observations $\Delta_{n,1}X, \Delta_{n,2}X, \ldots, \Delta_{n,n}X$ actually contain jumps, whereas the rest are just increments of the diffusion. So testing for the presence of jumps will be just equivalent to testing for the presence of “outliers”.

One sensible way to do this is to compare $\sum_{i=1}^{n} |\Delta_{n,i}X|^2$ to the quantity

$$S_n^{(\varepsilon)} = \sum_{k=1}^{\lfloor (1-\varepsilon)n \rfloor} Y_{kn},$$

(2.2.2)

where $Y_{in}$ is the $i$th order statistic for the sample, $|\Delta_{n,1}X|^2, \ldots, |\Delta_{n,n}X|^2$ and $\lfloor x \rfloor$ is the integer part of $x$. So $S_n^{(\varepsilon)}$ differs from $\sum_{i=1}^{n} |\Delta_{n,i}X|^2$ in that the largest $\lfloor \varepsilon n \rfloor$ summands were removed from that sum. Note that, as we have finite intensity of jumps and the proportion of the increments we remove in $S_n^{(\varepsilon)}$ is fixed, as $n \to \infty$, we will eventually have all the jumps removed, if there are any.

Suppose our null hypothesis $H_0$ is that there are no jumps, that is

$$
\begin{cases}
H_0 : & X_t = \mu t + \sigma W_t, \quad 0 \leq t \leq T, \\
H_1 : & X_t = \mu t + \sigma W_t + \sum_{i=1}^{N_t} J_i, \quad 0 \leq t \leq T.
\end{cases}
$$

(2.2.3)
Observe that, as $n \to \infty$,
\[
\begin{cases}
\sum_{i=1}^{n} |\Delta_{n,i}X|^2 \xrightarrow{P} \sigma^2 T & \text{under } H_0, \\
\sum_{i=1}^{n} |\Delta_{n,i}X|^2 \xrightarrow{P} \sigma^2 T + \sum_{k=0}^{N_T} j_k^2 & \text{under } H_1.
\end{cases}
\]

Therefore under $H_0$, we expect $S_n^{(\varepsilon)}$ to be somehow behaving like $\sum_{i=1}^{n} |\Delta_{n,i}X|^2$.

Observe that $S_n^{(\varepsilon)}$ is in fact a “trimmed sum”. Hence it is worthwhile to investigate the asymptotic properties of $S_n^{(\varepsilon)}$ so that we can formulate the Central Limit Theorem for test statistics.

2.2.2 The asymptotic behaviour of $S_n^{(\varepsilon)}$

In order to investigate the behaviour of $S_n^{(\varepsilon)}$, we need to prove results for the asymptotic behaviour of “trimmed sums”. First introduce some technical results.

Technical Results

Suppose $\xi_1, \xi_2, \ldots, \xi_n$ are i.i.d. random variables with distribution function $F$. We can assume without loss of generality that our sample is obtained as $F^{-1}(U_1), F^{-1}(U_2), \ldots, F^{-1}(U_n)$, where $U_1, U_2, \ldots, U_n$ are i.i.d. uniform random variables on $(0, 1)$.

**Definition 2.2.1.** Let $\xi$ be a random variable with distribution function $F$ and $\xi_1, \xi_2, \ldots, \xi_n$ a sample of independent copies of $\xi$. Then the Empirical Quantile Function (EQF) for the sample is the function $Q_n^*$ on $\Omega \times [0, 1]$, defined as

$$Q_n^*(\omega, s) = \inf\{t \in \mathbb{R} : F_n^*(\omega, t) \geq s\}$$

for all $n \geq 1$, where

$$F_n^*(\omega, t) = \sum_{i=1}^{n} 1\{X_i(\omega) \leq t\}$$

is the Empirical Distribution Function (EDF) for the sample $\xi_1, \xi_2, \ldots, \xi_n$.

In what follows, for the sake of simplicity, we will denote $Q_n^*(\omega, s)$ by $Q_n^*(s)$ when there is no necessity for specifying $\omega \in \Omega$. 49
Proposition 2.2.1. Let $F : \mathbb{R} \to \mathbb{R}$ be a non-decreasing càdlàg function and $A \subseteq \mathbb{R}$. Then

$$\inf_{x \in A} F(x) = F(\inf A)$$

Proof. Since $F$ is increasing, $F(\inf A) \leq F(y)$ for all $y$ in $A$ and hence $F(\inf A) \leq \inf_{x \in A} F(x)$.

To see that $F(\inf A) \geq \inf_{x \in A} F(x)$, note that, given $\varepsilon > 0$, there is a $y$ in $A$ such that $y < \inf A + \varepsilon$. By the monotonicity of $F$,

$$\inf_{x \in A} F(x) \leq F(y) \leq F(\inf A + \varepsilon)$$

Since $F$ is a càdlàg function and $\varepsilon > 0$ can be arbitrarily small, we conclude that $F(\inf A) \geq \inf_{x \in A} F(x)$.

Proposition 2.2.2. Let $\xi$ be a random variable with distribution function $F$ where $F$ is a homeomorphism and let $Q_n^\ast$ be the Empirical Quantile Function of the sample of $n$ i.i.d. copies of $\xi$. Then, as processes,

$$Q_n^\ast(\cdot) \overset{d}{=} F^{-1} \circ R_n^\ast(\cdot),$$

where $R_n^\ast$ is the EQF for a sample of $n$ i.i.d. uniform random variable.

Proof. Since $(\xi_1, \xi_2, \ldots, \xi_n) = (F^{-1}(U_1), F^{-1}(U_2), \ldots, F^{-1}(U_n))$, where $U_1, U_2, \ldots, U_n$ are i.i.d. uniform random variables, $\xi(k) = F^{-1}(U(k))$. Moreover, for $0 \leq k \leq n - 1$,

$$Q_n^\ast(s) = \xi(k+1), \quad s \in [k/n, (k + 1)/n). \quad (2.2.4)$$

Therefore proving the result.

Proposition 2.2.3. Let $\xi_1, \xi_2, \ldots, \xi_n$ be i.i.d. random variables with distribution function $F$ such that $F$ is a homeomorphism. Then for any $0 \leq \alpha \leq \beta \leq 1$,

$$\sum_{i=\lfloor \alpha n \rfloor+1}^{\lfloor \beta n \rfloor} \xi(i) \overset{d}{=} n \int_{\lfloor \alpha n \rfloor/n}^{\lfloor \beta n \rfloor/n} F^{-1} \circ R_n^\ast(s)ds, \quad (2.2.5)$$

where $\xi(j)$ is the $j$th order statistics of $X_1, X_2, \ldots, X_n$ and $R_n^\ast$ is the EQF of an i.i.d. sample of $n$ uniform random variables.

Proof. This follows immediately from the representation

$$\tilde{Q}_n^\ast(\cdot) = F^{-1} \circ R_n^\ast(\cdot),$$

as in the proof of the above proposition and the fact that, for $1 \leq k \leq n$,

$$\tilde{Q}_n^\ast(s) = \xi(k), \quad k - 1 < s \leq k.$$
Central Limit Theorems

Approximation of the trimmed sums was considered in a few literatures such as Stigler (1969), Shorack and Wellner (2009). Such an approximation can be used to deduce the asymptotic behaviour of the “trimmed” sum of i.i.d. random variables. However in the aforementioned literatures, assumptions made on the distribution function are rather strong.

In this section, we provide a central limit theorem for the “trimmed” sum of i.i.d. random variables with different assumptions on the distribution function. We will also further extend the central limit theorem to a triangular array of row-wise i.i.d. random variables.

**Theorem 2.2.1.** Let \( \xi_1, \xi_2, \xi_3, \ldots \) be i.i.d. random variables with distribution function \( F: \mathbb{R} \to [0, 1] \) such that \( F \) is continuously differentiable with non-vanishing derivative \( F' = f \) on the interval \( [F^{-1}(\alpha - \varepsilon), F^{-1}(\beta + \varepsilon)] \) for some \( \varepsilon > 0 \), where \( 0 \leq \alpha < \beta < 1 \) and \( S_{n}^{\alpha, \beta} = \sum_{i=\lfloor \alpha n \rfloor + 1}^{\lfloor \beta n \rfloor} \xi_i \). Then, as \( n \to \infty \),

\[
\frac{S_{n}^{\alpha, \beta} - n \int_{\alpha}^{\beta} F^{-1}(s)ds}{\sqrt{n} \sigma} \xrightarrow{d} Z
\]

where \( Z \sim N(0, 1) \) and

\[
\sigma^2 = \int_{\alpha}^{\beta} \int_{\alpha}^{\beta} \frac{\min\{s, t\}(1 - \max\{s, t\})}{f(F^{-1}(s))f(F^{-1}(t))}dsdt. \tag{2.2.6}
\]

**Remark 2.2.1.** It follows from the Inverse Function Theorem that \( F \) is invertible on the interval \( [F^{-1}(\alpha - \varepsilon), F^{-1}(\beta + \varepsilon)] \). Therefore \( F^{-1} \) is well defined.

**Proof.** By Proposition 2.2.3, \( S_{n}^{\alpha, \beta} \) admits the following representation:

\[
S_{n}^{\alpha, \beta} = n \int_{\alpha_n}^{\beta_n} F^{-1} \circ R_n'(s)ds,
\]

where \( \alpha_n = (\lfloor \alpha n \rfloor + 1)/n \), \( \beta_n = \lfloor \beta n \rfloor/n \) and \( R_n' \) is the empirical quantile function for a sample of \( n \) i.i.d. uniform random variables.

Let \( R \) be the quantile function of a \((0, 1)\)-uniform random variable, that is \( R(s) = s \). Since \( F' = f \) exists and is continuous, non-vanishing on \( [F^{-1}(\alpha - \varepsilon), F^{-1}(\beta + \varepsilon)] \), the derivative \( G'(s) = 1/f(G(s)) \), \( G = F^{-1} \), exists and is continuous on the interval \([\alpha - \varepsilon, \beta + \varepsilon]\). Therefore, by the Mean Value Theorem,

\[
n \int_{\alpha_n}^{\beta_n} F^{-1} \circ R_n'(s)ds = n \int_{\alpha_n}^{\beta_n} G \left( R(s) + \frac{1}{\sqrt{n}} \sqrt{n}(R_n'(s) - R(s)) \right) ds
\]

\[
= n \int_{\alpha_n}^{\beta_n} \left[ G(s) + G'(\theta_n(s)) \frac{1}{\sqrt{n}} \sqrt{n}(R_n'(s) - s) \right] ds
\]

51
where $\theta_n(s)$ is a random variable such that $|\theta(s) - s| \leq |R_n^*(s) - s|$. Note that $\theta_n$ converge uniformly to the identity map almost surely. To see this, observe that

$$|\theta_n(x) - s| \leq |R_n^*(s) - s| = |U_n(s) - s|,$$

where $U_n$ is the empirical distribution function of $n$ i.i.d. uniform random samples. By the Glivenko-Cantelli theorem,

$$\sup_{0 \leq x \leq 1} |U_n(s) - s| \to 0, \quad n \to \infty,$$

showing that $\theta_n$ converge uniformly to the identity map almost surely.

Letting $W_n = \sqrt{n}(R_n^*(s) - s)$, we see that

$$n \int_{\alpha_n}^{\beta_n} F^{-1} \circ R_n^*(s)ds = n \int_{\alpha_n}^{\beta_n} G(s) + G'(\theta_n(s)) \frac{1}{\sqrt{n}} W_n^* ds$$

After centring and normalising,

$$\frac{1}{\sqrt{n}} \left[ n \int_{\alpha_n}^{\beta_n} F^{-1} \circ R_n^*(s)ds - n \int_{\alpha_n}^{\beta_n} G(s)ds \right]$$

$$= \int_{\alpha_n}^{\beta_n} \left[ G'(\theta_n(s)) \frac{1}{\sqrt{n}} W_n^*(s) \right] ds$$

As $\max\{|\alpha_n - \alpha|, |\beta_n - \beta|\} \leq 1/n$, it is obvious that

$$\frac{1}{\sqrt{n}} \left| n \int_{\alpha_n}^{\beta_n} G(s)ds - n \int_{\alpha_n}^{\beta_n} G(s)ds \right|$$

$$\leq \sqrt{n} \sup_{\alpha - \epsilon < t < \beta + \epsilon} |G(s)| (|\alpha_n - \alpha| + |\beta_n - \beta|)$$

$$\leq \frac{2}{\sqrt{n}} \sup_{\alpha - \epsilon < t < \beta + \epsilon} |G(s)|$$

$$\to 0$$

as $n \to \infty$. Therefore we have

$$\frac{1}{\sqrt{n}} \left[ n \int_{\alpha_n}^{\beta_n} F^{-1} \circ R_n^*(s)ds - n \int_{\alpha_n}^{\beta_n} G(s)ds \right]$$

$$= \int_{\alpha_n}^{\beta_n} G'(\theta_n(s)) W_n^*(s)ds + o(1).$$

Now it remains to determine the limiting distribution of $\int_{\alpha_n}^{\beta_n} G'(\theta_n(s)) W_n^*(s)ds$. 

52
Before we proceed any further, observe that, as random elements from the probability space to the Skorokhod space $D$, $W_n^\circ$ converges weakly the Brownian bridge, $W^\circ$ (see, Theorem 1.3.17). Since the Skorokhod space $D$ is a Polish space in a sense that it can be completely metrised (see Section 1.3.4). Therefore by the Skorokhod representation theorem (see Theorem 1.3.18), we can construct a probability space of the form $(\Omega', \mathcal{F}', P')$ and random elements $\tilde{W}_n^\circ : [0, 1] \to D$, $n = 1, 2, 3, \ldots$, and $\tilde{W}^\circ : [0, 1] \to D$ such that

(i) $\tilde{W}_n^\circ$ and $W_n^\circ$ are equal in distribution for all $n = 1, 2, 3, \ldots$,

(ii) $\tilde{W}^\circ$ and $W^\circ$ are equal in distribution,

(iii) $\tilde{W}_n^\circ \xrightarrow{a.s.} \tilde{W}^\circ$, $n \to \infty$ under the metric $d_0$ which completely metrised the Skorokhod space $D$.

Therefore, we have

$$\int_{\alpha_n}^{\beta_n} G' (\theta_n (s)) W_n^\circ (s) ds \xrightarrow{d} \int_{\alpha_n}^{\beta_n} G' (\theta_n (s)) \tilde{W}_n^\circ (s) ds$$

Hence,

$$\left| \int_{\alpha_n}^{\beta_n} G' (\theta_n (s)) \tilde{W}_n^\circ (s) ds - \int_{\alpha}^{\beta} G' (s) \tilde{W}^\circ (s) ds \right|$$

$$\leq \left| \int_{\alpha}^{\beta} G' (\theta_n (s)) \tilde{W}_n^\circ (s) ds - \int_{\alpha}^{\beta} G' (s) \tilde{W}^\circ (s) ds \right|$$

$$+ \left| \int_{\alpha_n}^{\beta_n} G' (\theta_n (s)) \tilde{W}_n^\circ (s) ds - \int_{\alpha}^{\beta} G' (\theta_n (s)) \tilde{W}_n^\circ (s) ds \right|$$

$$\leq \int_{\alpha}^{\beta} \left| G' (\theta_n (s)) \left( \tilde{W}_n^\circ (s) - \tilde{W}^\circ (s) \right) \right| ds$$

$$+ \int_{\alpha}^{\beta} \left| (G' (\theta_n (s)) - G' (s)) \tilde{W}^\circ (s) \right| ds$$

$$+ \left| \int_{\alpha_n}^{\beta_n} G' (\theta_n (s)) \tilde{W}_n^\circ (s) ds - \int_{\alpha}^{\beta} G' (\theta_n (s)) \tilde{W}_n^\circ (s) ds \right|$$

$$\leq \sup_{0 \leq s \leq 1} |\tilde{W}_n^\circ (s) - \tilde{W}^\circ (s)| \int_{\alpha}^{\beta} |G' (\theta_n (s))| ds$$

$$+ \sup_{0 \leq s \leq 1} |G' (\theta_n (s)) - G' (s)| \int_{\alpha}^{\beta} |\tilde{W}^\circ (s)| ds$$

$$+ \left| \int_{\alpha_n}^{\beta_n} G' (\theta_n (s)) \tilde{W}_n^\circ (s) ds - \int_{\alpha}^{\beta} G' (\theta_n (s)) \tilde{W}_n^\circ (s) ds \right|$$

53
Since \( \widetilde{W}_n \overset{a.s.}{\longrightarrow} \widetilde{W} \) under the metric \( d_0 \) implies \( \widetilde{W}_n \overset{a.s.}{\longrightarrow} \widetilde{W} \) under the uniform metric, \( \sup_{0 \leq s \leq 1} |\tilde{W}_n^\circ(s) - \tilde{W}^\circ(s)| \overset{a.s.}{\longrightarrow} 0 \), so the first term tends to 0 almost surely as \( n \to \infty \).

Furthermore, since \( G' \) is a continuous function on the compact set \([\alpha, \beta]\), it is hence uniformly continuous on \([\alpha, \beta]\). Also, since \( \theta_n \) converge uniformly to the identity map, \( G' \circ \theta_n \) converge uniformly to \( G' \) (see Theorem A.1.2 in the Appendix). Hence the second term tends to 0 as \( n \to \infty \) as well. Finally, as the third term can be bounded by

\[
(|\alpha_n - \alpha| + |\beta_n - \beta|) \sup_{\alpha - \varepsilon \leq s \leq \beta + \varepsilon} |G'(\theta_n(s))| \sup_{\alpha - \varepsilon \leq s \leq \beta + \varepsilon} |\tilde{W}_n^\circ| \\
\leq \frac{2}{\sqrt{n}} \sup_{\alpha - \varepsilon \leq s \leq \beta + \varepsilon} |G'(\theta_n(s))| \sup_{\alpha - \varepsilon \leq s \leq \beta + \varepsilon} |\widetilde{W}_n^\circ| \\
\to 0
\]

as \( n \to \infty \). Therefore

\[
\int_\alpha^\beta G'(\theta_n(s))\tilde{W}_n^\circ(s)ds \overset{a.s.}{\longrightarrow} \int_\alpha^\beta G'(s)\widetilde{W}^\circ(s)ds, \quad n \to \infty
\]

This clearly means that, as \( n \to \infty \),

\[
\int_\alpha^\beta G'(s)W_n^\circ(s)ds \overset{d}{\longrightarrow} \int_\alpha^\beta G'(s)W^\circ(s)ds.
\]

Since \( W^\circ \) is the Brownian bridge, \( W^\circ \) is a Gaussian process. Therefore \( \int_\alpha^\beta G'(s)W^\circ(s)ds \) is normally distributed. The mean and variance can be explicitly calculated using Fubini’s theorem. The mean is equal to

\[
E \int_\alpha^\beta G'(s)W^\circ(s)ds = \int_\alpha^\beta G'(s)EW^\circ(s)ds = 0
\]
Similarly, we can compute the variance,
\[
\text{Var} \left( \int_{\alpha}^{\beta} G'(s) W^\circ(s) ds \right) = \mathbb{E} \left( \int_{\alpha}^{\beta} G'(s) W^\circ(s) ds \right)^2 = \mathbb{E} \int_{\alpha}^{\beta} G'(s) W^\circ(s) ds \int_{\alpha}^{\beta} G'(t) W^\circ(t) dt
\]
\[
= \mathbb{E} \int_{\alpha}^{\beta} \int_{\alpha}^{\beta} G'(s) G'(t) W^\circ(s) W^\circ(t) dsdt = \int_{\alpha}^{\beta} \int_{\alpha}^{\beta} G'(s) G'(t) E(W^\circ(s) W^\circ(t)) dsdt
\]
\[
= \int_{\alpha}^{\beta} \int_{\alpha}^{\beta} G'(s) G'(t) \min\{s, t\}(1 - \max\{s, t\}) dsdt
\]
since \(\text{Cov}(W^\circ(s), W^\circ(t)) = s(1 - t)\) for \(s < t\), so that
\[
\text{Cov}(W^\circ(s), W^\circ(t)) = \min\{s, t\}(1 - \max\{s, t\}).
\]
Note also that \(G'(y) = (F^{-1})'(y) = 1/F'(F^{-1}(y)) = 1/f(F^{-1}(y))\) and it is continuous on \([\alpha, \beta]\), so the above integral exists hence the variance exists.

Hence, we conclude that
\[
\int_{\alpha}^{\beta} G'(s) W^\circ(s) ds \sim N(0, \sigma^2)
\]
where
\[
\sigma^2 = \int_{\alpha}^{\beta} \int_{\alpha}^{\beta} \min\{s, t\}(1 - \max\{s, t\}) f(F^{-1}(s)) f(F^{-1}(t)) dsdt.
\]
Thus we proved that, as \(n \to \infty\),
\[
\frac{S_{n,\alpha,\beta} - n \int_{\alpha}^{\beta} F^{-1}(s) ds}{\sigma \sqrt{n}} \overset{d}{\to} Z,
\]
where \(Z \sim N(0, 1)\).

Theorem 2.2.1 gives us the Central Limit Theorem for the “trimmed” sums of i.i.d. random variables. It is natural (and in fact necessary for our purposes) to attempt to extend the Central Limit Theorem to the i.i.d.
triangular array scheme (see Definition 1.2.12), that is to the situation where we consider random variables

\[ ξ_{1,1}, ξ_{2,1}, ξ_{2,2}, ξ_{3,1}, ξ_{3,2}, ξ_{3,3}, \ldots, X_{ξ_{n,1}}, X_{ξ_{n,2}}, X_{ξ_{n,3}}, \ldots, X_{ξ_{n,n}} \]

where for each \( i \geq 1 \), the random variables \( ξ_{i,1}, ξ_{i,2}, ξ_{i,3}, \ldots \) are independent and identically distributed. Call this the row-wise i.i.d. case and denote by \( ξ_{n,(i)} \) to be the \( i \)th order statistic in the \( n \)th row.

Indeed, with some additional mild assumptions on the distributions, the desired result can be proved.

**Theorem 2.2.2.** Let \( ξ_{n,1}, ξ_{n,2}, \ldots, ξ_{n,n}, n \geq 1 \), be a triangular array of row-wise i.i.d. random variables with respective distribution functions \( F_n \), and \( S_{n,α,β} = \sum_{i=\lfloor αn \rfloor + 1}^{\lfloor βn \rfloor} ξ_{n,(i)} \) for \( 0 \leq α < β < 1 \). Suppose that, for each \( n \geq 1 \),

(i) \( F_n \) is continuously differentiable and \( F_n \) converge uniformly to a distribution function \( F \) on \([F^{-1}(α - ε), F^{-1}(β + ε)]\) for some \( ε > 0 \) and the density \( f \) of \( F \) is non-vanishing on \([F^{-1}(α - ε), F^{-1}(β + ε)]\).

(ii) The density \( f_n \) of \( F_n \) exists and is non-vanishing, continuous such that \( 1/f_n \) converge uniformly to a limit \( 1/f \) on \([F^{-1}(α), F^{-1}(β)]\) as \( n \to \infty \).

Then, as \( n \to \infty \),

\[
\frac{S_{n,α,β} - n \int_{α}^{β} F^{-1}(s)ds}{\sqrt{nσ}} \xrightarrow{d} Z,
\]

where \( Z \sim N(0,1) \) and

\[
σ^2 = \int_{α}^{β} \int_{α}^{β} \frac{\min\{s, t\}(1 - \max\{s, t\})}{f(F^{-1}(s))f(F^{-1}(t))}dndsdt.
\]

**Proof.** By Proposition 2.2.3, we can assume that

\[
S_{n,α,β} = n \int_{α_n}^{β_n} F^{-1}_n \circ R_n(s)ds,
\]
where \( \alpha_n = (\lfloor \alpha n \rfloor + 1)/n \) and \( \beta_n = \lfloor \beta n \rfloor /n \) and \( R^*_n \) is the empirical quantile function for an i.i.d. sample of \( U(0,1) \) random variables of size \( n \). Since \( F \) is differentiable for \( n \geq 1 \) and \( (F^{-1})' = 1/(f \circ F^{-1}) \), the function \( F^{-1} \) is differentiable on \([\alpha - \varepsilon, \beta + \varepsilon]\). Hence letting \( G_n = F^{-1}_n \) and expanding \( G_n \) around \( s \in [0,1] \), we obtain by the Mean Value theorem that

\[
\int_{\alpha_n}^{\beta_n} F^{-1}_n \circ R^*_n(s) ds = \int_{\alpha_n}^{\beta_n} \left[ G_n(s) + G_n'(\theta_n(s)) \frac{1}{\sqrt{n}} \sqrt{n}(R^*_n(s) - s) \right] ds,
\]

where \( \theta_n = \theta_n(s) \) is a random variable such that \( s < \theta_n(s) < s + |R^*_n(s) - s| \). Let \( W^0_n = \sqrt{n}(R^*_n(s) - s) \), as in the proof of Theorem 2.2.1, after centring and normalising,

\[
\frac{1}{\sqrt{n}} \left[ n \int_{\alpha_n}^{\beta_n} F^{-1}_n \circ R^*_n(s) ds - n \int_{\alpha_n}^{\beta_n} G_n(s) ds \right] = \int_{\alpha}^{\beta} G_n'(s) W^0_n(s) ds\]

Similarly, as in the proof on Theorem 2.2.1, for \( n \to \infty \),

\[
\frac{1}{\sqrt{n}} \left| n \int_{\alpha_n}^{\beta_n} G_n(\theta_n(s)) ds - n \int_{\alpha}^{\beta} G(s) ds \right| \leq \frac{1}{\sqrt{n}} \sup_{\alpha - \varepsilon < s < \beta + \varepsilon} |G_n(\theta_n(s)) - G(s)| \xrightarrow{a.s.} 0
\]

from the fact that \( G \) is uniformly continuous on \([\alpha - \varepsilon, \beta + \varepsilon]\) and \( \theta_n \) converge uniformly to the identity map as \( n \to \infty \), we have (See Theorem A.1.2 in the Appendix)

\[
\sup_{\alpha + \varepsilon < s < \beta + \varepsilon} |G_n(\theta_n(s)) - G(s)| \leq \sup_{\alpha + \varepsilon < s < \beta + \varepsilon} |G_n(\theta_n(s)) - G(s)|
\]

\[
\leq \sup_{\alpha + \varepsilon < s < \beta + \varepsilon} |G_n(\theta_n(s)) - G(\theta_n(s))| + \sup_{\alpha + \varepsilon < s < \beta + \varepsilon} |G(\theta_n(s)) - G(s)| \xrightarrow{a.s.} 0.
\]

Therefore, as in the proof of Theorem 2.2.1, we showed that

\[
\frac{n \int_{\alpha_n}^{\beta_n} F^{-1}_n \circ R^*_n(s) ds - n \int_{\alpha}^{\beta} G(s) ds}{\sqrt{n}} = \int_{\alpha}^{\beta} G_n'(s) W^0_n(s) ds + o(1)
\]

Following exactly the same logic as in the proof of Theorem 2.2.1, we use the fact there exists a probability space and random elements, \( \tilde{W}^0_n \), \( n \geq 1 \)
and $\widetilde{W}$ defined on that probability space such that $\widetilde{W}_n \xrightarrow{a.s.} \widetilde{W}$ as $n \to \infty$ in the Skorokhod metric and

$$\int_{\alpha}^{\beta} G'_n(s)W_n^\circ(s)\,ds \xrightarrow{d} \int_{\alpha}^{\beta} G'_n(s)\widetilde{W}_n^\circ(s)\,ds, \quad \forall n \geq 1$$

Since by our hypothesis $1/f_n$ converges uniformly to $1/f$ on $[F^{-1}(\alpha), F^{-1}(\beta)]$ and $G'(s) = 1/f(F^{-1}(s))$, we see that $G_n$ converges uniformly to $G$ on $[\alpha, \beta]$:

$$\sup_{\alpha<s<\beta} |G_n(s) - G(s)| \to 0, \quad n \to \infty$$

Therefore,

$$\left| \int_{\alpha}^{\beta} G'_n(\theta_n(s))W_n^\circ(s)\,ds - \int_{\alpha}^{\beta} G'(s)\widetilde{W}_n^\circ(s)\,ds \right|$$

$$\leq \int_{\alpha}^{\beta} \left| (G'_n(\theta_n(s)) - G'(s))\widetilde{W}_n^\circ(s) \right| \,ds$$

$$+ \int_{\alpha}^{\beta} \left| G'_n(\theta_n(s))(W_n^\circ(s) - \widetilde{W}_n^\circ(s)) \right| \,ds$$

$$\leq \sup_{\alpha<s<\beta} |G_n(\theta_n(s)) - G(s)| \int_{\alpha}^{\beta} \left| \widetilde{W}_n^\circ(s) \right| \,ds$$

$$+ \sup_{\alpha<s<\beta} \left| W_n^\circ(s) - \widetilde{W}_n^\circ(s) \right| \int_{\alpha}^{\beta} |G'_n(\theta_n(s))| \,ds$$

Since $\widetilde{W}$ and $G'$ are continuous on $[\alpha, \beta]$, $\int_{\alpha}^{\beta} \left| \widetilde{W}_n^\circ(s) \right| \,ds < \infty$ and

$$\int_{\alpha}^{\beta} |G'(\theta_n(s))|\,ds < \int_{\alpha}^{\beta} G(s)\,ds + \sup_{\alpha+\varepsilon<s<\beta+\varepsilon} |G'(\theta_n(s)) - G'(s)| < \infty,$$

so we have

$$\int_{\alpha}^{\beta} G'_n(\theta_n(s))W_n^\circ(s)\,ds \xrightarrow{a.s.} \int_{\alpha}^{\beta} G'(s)\widetilde{W}_n^\circ(s)\,ds, \quad n \to \infty.$$
Hence
\[
\int_{\alpha}^{\beta} G'_n(\theta_n(s)) W_n^\circ(s) ds \xrightarrow{d} \int_{\alpha}^{\beta} G'(s) W^\circ(s) ds \quad n \to \infty.
\]

Since \( W^\circ \) is a Brownian bridge, \( W^\circ \) is a Gaussian processes. Therefore \( \int_{\alpha}^{\beta} G'(s) W^\circ(s) ds \) is normally distributed. And the mean and variance of \( \int_{\alpha}^{\beta} G'(s) W^\circ(s) ds \) can be computed exactly in the same way as in the proof of Theorem 2.2.1. Therefore we established that
\[
\int_{\alpha}^{\beta} G'(s) W^\circ(s) ds \sim N(0, \sigma^2),
\]
where
\[
\sigma^2 = \int_{\alpha}^{\beta} \int_{\alpha}^{\beta} \min\{s, t\} (1 - \max\{s, t\}) f(F^{-1}(s)) f(F^{-1}(t)) ds dt.
\]

### 2.2.3 The test statistic

Denote \( \sum_{j=1}^{n} |\Delta_{j,n}X|^2 \) by \( S_n \) and \( S_n^{(e)} \) be as defined in (2.2.2). Let
\[
\widetilde{S}_n^{(e)} = S_n - S^{(e)}_n. \tag{2.2.7}
\]
Then \( \widetilde{S}_n^{(e)} \) is the sum of the biggest \( \epsilon \) percent of the squared increments \( |\Delta_{n,1}X|^2, \ldots, |\Delta_{n,n}X|^2 \).

Under the null hypothesis \( H_0 \), it is intuitive to expect that \( \widetilde{S}_n^{(e)} \) will be “very likely” to converge to some value less than \( \sigma^2 T \) as \( n \to \infty \). The following results show that our intuition is indeed correct.

**Proposition 2.2.4.**

\[
\frac{n}{\sigma^2 T} |\Delta_{1,n}X|^2 \xrightarrow{d} \chi^2_1,
\]
under the null hypothesis \( H_0 \), where \( \chi^2_1 \) follows the Chi-Square distribution with degree of freedom 1.

**Proof.** Denote \( \Delta_{1,n}X \) by \( \Delta_n X \). For all \( x \geq 0 \),
\[
P\left( \frac{n}{\sigma^2 T} |\Delta_{j,n}X|^2 \leq x \right) = P\left( \frac{n}{\sigma^2 T} \left| \frac{T}{n} \frac{\mu}{\sqrt{n}} \frac{\sqrt{T}}{\sigma} Z \right|^2 \leq x \right)
\]
\[
= P\left( \left| \frac{\mu}{\sigma} \sqrt{\frac{T}{n}} - Z \right|^2 \leq x \right)
\]
\[
= P\left( Z \in \left[ - \sqrt{x + a_n}, \sqrt{x + a_n} \right] \right)
\]
where \( Z \) is a standard normal random variable and \( a_n = \frac{\mu}{\sigma} \sqrt{\frac{T}{n}} \). Since \( a_n \downarrow 0 \), as \( n \to \infty \),
\[
P \left( Z \in [-\sqrt{x} + a_n, \sqrt{x} + a_n] \right) = \Phi(\sqrt{x} + a_n) - \Phi(-\sqrt{x} + a_n) \\
\to \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) \\
= P(Z^2 \leq x),
\]
where \( \Phi \) denotes the cdf of a normal random variable.

Since \( Z^2 \) follows the Chi-Square distribution with 1 degree of freedom, the result follows.

**Lemma 2.2.1.** Suppose \( \varepsilon > 0 \). Then, as \( n \to \infty \),
\[
S_n^{(\varepsilon)} \xrightarrow{P} \sigma^2 T \int_0^{1-\varepsilon} F^{-1}(s)ds
\]
under the null hypothesis \( H_0 \) where \( F \) denotes the Chi-Square distribution with 1 degree of freedom.

**Proof.** This is an immediate consequence of Proposition 2.2.4 and Theorem 2.2.2.

From the above result, we obtain the following,

**Lemma 2.2.2.** Suppose \( \varepsilon > 0 \). Then, as \( n \to \infty \),
\[
\tilde{S}_n^{(\varepsilon)} \xrightarrow{P} \sigma^2 T \int_{1-\varepsilon}^{1} F^{-1}(s)ds
\]
under the null hypothesis \( H_0 \) where \( F \) denotes the Chi-Square distribution with 1 degree of freedom.

**Proof.** It follows from Lemma 2.2.1 and the fact that \( \tilde{S}_n^{(\varepsilon)} = S_n - S_n^{(\varepsilon)} \).

As stated at the beginning of this section, we want a test that is parameter free. That is, the test does not require any prior estimation of the drift coefficient \( \mu \) and the diffusion coefficient \( \sigma \). Therefore one may not simply employ \( S_n^{(\varepsilon)} \) and \( S_n^{(\varepsilon)} \) as the test statistic since both involve parameter estimations when the actual testing procedure is carried out.

One way to avoid the parameter estimation problem is to define our test statistic \( T_n \) as
\[
T_n = \frac{\tilde{S}_n^{(\varepsilon)}}{S_n^{(\varepsilon)}}. \tag{2.2.8}
\]
The convergence in probability of the test statistics follows as an immediate consequence of Lemma 2.2.2 and Lemma 2.2.1.
Theorem 2.2.3. Fix $\varepsilon > 0$. Then, as $n \to \infty$,

$$T_n \xrightarrow{P} \frac{\nu_2}{\nu_1}$$

where

$$\nu_1 = \int_0^{1-\varepsilon} F^{-1}(s) ds, \quad \nu_2 = \int_{1-\varepsilon}^1 F^{-1}(s) ds.$$

under the null hypothesis $H_0$.

Proof. It follows from the facts that $\tilde{S}_n^{(\varepsilon_n)} \xrightarrow{P} \sigma^2 T \nu_2$ and $S_n^{(\varepsilon_n)} \xrightarrow{P} \sigma^2 T \nu_1$ as $n \to \infty$. Then, as $n \to \infty$,

$$T_n = \frac{\tilde{S}_n^{(\varepsilon_n)}}{S_n^{(\varepsilon_n)}} \xrightarrow{P} \frac{\nu_1}{\nu_2}.$$

When under the alternative hypothesis, $\tilde{S}_n^{(\varepsilon_n)}$ converge in probability to some value bigger than $\nu_2$. So our test statistic $T_n$ will be bigger compare to that under the null hypothesis. There the task of detecting jumps would be equivalent to decide whether $T_n$ is significantly larger than $\nu_2/\nu_1$.

We need the asymptotic behaviour of our test statistics $T_n$ in order to derive the rejection region for our hypothesis test. The following results allows us to have more insight into the asymptotic distribution of $T_n$.

Lemma 2.2.3. For any fixed $\varepsilon > \varepsilon' > 0$,

$$n^{1/2} \left( (\sigma^2 T)^{-1} S_n^{(\varepsilon)} - \nu, (\sigma^2 T)^{-1} (\tilde{S}_n^{(\varepsilon)} - \tilde{S}^{(\varepsilon')}) - \tilde{\nu} \right) \xrightarrow{d} (Z_1, Z_2)$$

where $(Z_1, Z_2)$ is jointly Gaussian with mean 0 and covariance matrix,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix},$$

$$f(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}, \quad F(x) = \int_0^x f(s) ds$$

$$\nu = \int_0^{1-\varepsilon} F^{-1}(s) ds, \quad \tilde{\nu} = \int_{1-\varepsilon}^{1-\varepsilon'} F^{-1}(s) ds,$$

$$\nu = \int_0^{1-\varepsilon} F^{-1}(s) ds, \quad \tilde{\nu} = \int_{1-\varepsilon}^{1-\varepsilon'} F^{-1}(s) ds.$$
\[
\rho = \frac{\int_0^{1-\varepsilon} \int_{1-\varepsilon}^{1-\varepsilon'} H(s, t) \, ds \, dt}{\left( \int_0^{1-\varepsilon} \int_0^{1-\varepsilon} H(s, t) \, ds \, dt \int_{1-\varepsilon}^{1-\varepsilon'} H(s, t) \, ds \, dt \right)^{1/2}}, \quad (2.2.10)
\]

\[
\sigma_1 = \text{Var} Z_1 = \int_0^{1-\varepsilon} \int_0^{1-\varepsilon} H(s, t) \, ds \, dt, \quad (2.2.11)
\]

\[
\sigma_2 = \text{Var} Z_2 = \int_{1-\varepsilon}^{1-\varepsilon'} \int_{1-\varepsilon}^{1-\varepsilon'} H(s, t) \, ds \, dt, \quad (2.2.12)
\]

and

\[
H(s, t) = \min\{s, t\}(1 - \max\{s, t\}) f(F^{-1}(s)) f(F^{-1}(t)).
\]

**Proof.** By Proposition 2.2.4, the distribution function of \( \frac{n}{\sigma^2 T} |\Delta_{j,n} X|^2 F_n(x) \) converge to the Chi-Square distribution \( F(x) \) for all \( x \geq 0 \). Since \( F_n \) are increasing functions and converge pointwise to \( F \), \( F_n \) converge uniformly to \( F \) on \([0, \infty)\) as \( n \to \infty \) (see Corollary A.1.1 in the Appendix). So condition \((i)\) of Theorem 2.2.2 is satisfied.

Moreover, let \( f_n = F_n' \) for \( n \geq 1 \). Observe that \( 1/f_n \) and \( 1/f \) are all increasing continuous functions such that \( 1/f_n(x) \to 1/f(x) \) for all \( x \geq 0 \). Therefore \( 1/f_n \) converge uniformly to \( 1/f \) on the interval \([0, K]\) for any \( K > 0 \) (see Corollary A.1.1 in the Appendix). So condition \((ii)\) of Theorem 2.2.2 is satisfied and hence we can apply the Central Limit Theorem (Theorem 2.2.2).

As in the proof of Theorem 2.2.2, by the Skorokhod Representation Theorem, \( n^{-1/2}(S_n^{(e)} - \nu) \) and \( n^{-1/2}((\overline{S}_n^{(e)} - \overline{S}^{(e')}) \) can be represented by using the same process \( \bar{W}_n^{(\nu)} \) such that, as \( n \to \infty \),

\[
n^{-1/2}(n(\sigma^2 T)^{-1} S_n^{(e)} - n\nu) = \int_0^{1-\varepsilon} G_n'(\theta_n(s)) \xrightarrow{a.s.} \int_0^{1-\varepsilon} G'(s)\bar{W}_n^{(\nu)}(s) \, ds
\]

and

\[
n^{-1/2}(n(\sigma^2 T)^{-1} \overline{S}_n^{(e)} - n(\sigma^2 T)^{-1} \overline{S}^{(e')}) - n\tilde{\nu}) = \int_{1-\varepsilon}^{1-\varepsilon'} G_n'(\theta_n(s)) \xrightarrow{a.s.} \int_{1-\varepsilon}^{1-\varepsilon'} G'(s)\bar{W}_n^{(\nu)}(s) \, ds,
\]

62
where \( G_n = F_n^{-1} \) for \( n \geq 1 \) and \( G = F^{-1} \). Therefore

\[
n^{1/2} \left( \sigma^2 T_n \right)^{-1} S_n - \nu, (\sigma^2 T_n)^{-1}(\tilde{S}_n - \tilde{S}(\nu)) - \tilde{v} \right) = \left( \int_{0}^{1-\varepsilon} G'(s)\tilde{W}_n(s)ds, \int_{1-\varepsilon}^{1-\varepsilon'} G'(s)\tilde{W}_n(s)ds \right)
\]

is jointly gaussian and the variances and covariance can be obtained by computing

\[
E \left( \int_{0}^{1-\varepsilon} G'(s)\tilde{W}_n(s)ds, \int_{1-\varepsilon}^{1-\varepsilon'} G'(s)\tilde{W}_n(s)ds \right)
= \int_{0}^{1-\varepsilon} \int_{1-\varepsilon}^{1-\varepsilon'} G'(t)G'(s)E \left( \tilde{W}_n(s)\tilde{W}_n(t) \right) dsdt
= \int_{0}^{1-\varepsilon} \int_{1-\varepsilon}^{1-\varepsilon'} H(s,t)dsdt
\]

and similarly the variances of \( \int_{0}^{1-\varepsilon} G'(s)\tilde{W}_n(s)ds \) and \( \int_{\varepsilon}^{1-\varepsilon'} G'(s)\tilde{W}_n(s)ds \) are computed as

\[
\int_{0}^{1-\varepsilon} \int_{0}^{1-\varepsilon} H(s,t)dsdt, \quad \int_{1-\varepsilon}^{1-\varepsilon'} \int_{1-\varepsilon}^{1-\varepsilon'} H(s,t)dsdt
\]

respectively. and hence we obtained the correlation \( \rho \) and variances \( \sigma_1 \) and \( \sigma_2 \),

\[
\rho = \frac{\int_{0}^{1-\varepsilon} \int_{1-\varepsilon}^{1-\varepsilon'} H(s,t)dsdt}{\left( \int_{0}^{1-\varepsilon} \int_{0}^{1-\varepsilon} H(s,t)dsdt \int_{1-\varepsilon}^{1-\varepsilon'} \int_{1-\varepsilon}^{1-\varepsilon'} H(s,t)dsdt \right)^{1/2}}
\]

\[
\sigma_1 = Var Z_1 = \int_{0}^{1-\varepsilon} \int_{0}^{1-\varepsilon} H(s,t)dsdt, \quad (2.2.13)
\]

\[
\sigma_2 = Var Z_2 = \int_{1-\varepsilon}^{1-\varepsilon'} \int_{1-\varepsilon}^{1-\varepsilon'} H(s,t)dsdt. \quad (2.2.14)
\]

\[\square\]

### 2.2.4 Approximating the distribution of \( T_n \)

Since as \( \varepsilon \to 0 \),

\[
\nu_2 = \int_{1-\varepsilon}^{1} F^{-1}(s)dx \to 0
\]
where $F$ is as defined in (2.2.3), we approximate $\tilde{S}_n^{(e)}$ by $\tilde{S}_n^{(e)} - \tilde{S}_n^{(e')}$ for $e' \ll e$.

Therefore, we approximate the distribution of our test statistics $T_n$ by the ratio of two correlated normal random variables without prior estimations of the drift coefficient $\mu$ and the diffusion coefficient $\sigma$. To see why this approximation does not require any parameter, suppose $(Z_1, Z_2)$ is a binormal random vector with mean 0 and covariance matrix $\Sigma$ as defined in (2.2.10). Then

$$T_n = \frac{\tilde{S}_n^{(e)}}{S_n^{(e)}} = \frac{n(\sigma^2T)^{-1}\tilde{S}_n^{(e)}}{n(\sigma^2T)^{-1}S_n^{(e)}} \approx \frac{\nu + n^{-1/2}Z_2}{\bar{\nu} + n^{-1/2}Z_1},$$

where $\nu$ and $\bar{\nu}$ are defined as in (2.2.9). And the rejection region is $\{T_n > a\}$ for some value $a$.

Since we do not have an analytic formula for (2.2.15), we need to simulate the $p^{th}$ quantiles $q_p$ for our test statistics $T_n$ and obtain the rejection region by Monte Carlo simulation. So we have the following decision rule, for any $1 < p < 1$,

$$\begin{cases} 
X \text{ is continuous in } [0,T] & \text{if } T_n \le q_p \\
X \text{ is discontinuous} & \text{if } T_n > q_p.
\end{cases}$$

(2.2.16)

Another approximation

The approximation suggested in the previous section requires Monte Carlo simulation and in practice, it can be very undesirable as the actual simulation can be very slow. In fact, we may take

$$\tilde{T}_n = \sqrt{n} \left(T_n - \frac{\nu}{\bar{\nu}}\right).$$

(2.2.17)

where $\nu, \bar{\nu}$ defined as in (2.2.9). Let

$$\xi_{1,n} = n^{1/2} \left((\sigma^2T)^{-1}S_n^{(e)} - \nu\right)$$

(2.2.18)

and

$$\xi_{2,n} = n^{1/2} \left((\sigma^2T)^{-1}(\tilde{S}_n^{(e)} - \tilde{S}_n^{(e')} - \bar{\nu}\right)$$

(2.2.19)
Hence the distribution of $\tilde{T}_n$ can be approximated by,

$$\tilde{T}_n \approx \frac{1}{\sqrt{n}} \left( \frac{\nu + n^{-1/2} \xi_{2,n} - \tilde{\nu}}{\nu + n^{-1/2} \xi_{1,n}} - \frac{\nu - (\nu + n^{-1/2} \xi_{1,n})\tilde{\nu}}{\nu + n^{-1/2} \xi_{1,n}} - \nu \right) = \frac{\nu \xi_{2,n} - \tilde{\nu} \xi_{1,n}}{\nu + n^{-1/2} \xi_{1,n}} \nu \frac{\nu Z_2 - \tilde{\nu} Z_1}{\nu^2}$$

as $n \to \infty$, where $(Z_1, Z_2)$ is a bi-variate normal vector as described in Lemma 2.2.3. Therefore we can work out the explicitly the asymptotic distribution of $\tilde{T}_n$ and hence that of $T_n$.

By Lemma 2.2.3, the distribution of $\nu^{-2}(\nu Z_2 - \tilde{\nu} Z_1)$ is computed as

$$N \left( 0, \left( \frac{\sigma_1}{\nu} \right)^2 + \left( \frac{\tilde{\nu} \sigma_2}{\nu^2} \right)^2 - \frac{2 \tilde{\nu}}{\nu^3} \rho \sigma_1 \sigma_2 \right)$$

where $\nu, \tilde{\nu}, \sigma_1, \sigma_2$ and $\rho$ are as defined in Lemma 2.2.3. Thus we can compute the quantile of (2.2.20) and get the rejection region.

This approximation gives us an explicit formula to work with, therefore no Monte Carlo simulations are required.
Chapter 3

Simulation results

This chapter is devoted to the numerical and graphical results from simulations to illustrate the convergence of our Central Limit Theorem formulated in Chapter 2, and the rejection rate of the hypothesis test employing our test statistics.

This chapter is organised as follows. In the first section we will provide numerical simulations illustrating the central limit theorem for trimmed sums. In the second section, we will employ our test statistics to conduct hypothesis tests on a diffusion model and a jump diffusion model. We will access the performance of our test by simulating their respective rejection rates.

3.1 The Central Limit Theorem

We devote this section to simulations which will provide numerical evidence to support the Central Limit Theorems formulated in Chapter 2 and examine the convergence speed.

Suppose, for $0 \leq \alpha < \beta < 1$,

$$S_{n}^{\alpha,\beta} = \sum_{k=[\alpha n]+1}^{[\beta n]} \xi_{(k)},$$

where $\xi_{(k)}$ is the $k$th ordered statistic for a sample of i.i.d. Chi-Squared random variables $\xi_1, \xi_2, \ldots, \xi_n$ with degree of freedom 1.

We will simulate $S_{n}^{\alpha,\beta}$ for different values of $n, \alpha$ and $\beta$ for illustration purposes. Then we will compare the standard normal density curve to the
histogram of the standardised trimmed sums,

\[ \frac{S_{n,\alpha}^{\alpha,\beta} - n \int_{\alpha}^{\beta} F^{-1}(s) ds}{\sigma \sqrt{n}} \]

where \( \sigma \) is as defined in (2.2.6), \( F \) is the cdf of \( X_1 \) and \( F = f' \).

Figure 3.1: The histogram of the standardised trimmed sums for different values of \( n \) with \( \alpha = 0, \beta = 0.95 \). The solid line represents the standard normal density curve.
Figure 3.2: The histogram of the standardised trimmed sums for different values of $n$ with $\alpha = 0.37$, $\beta = 0.73$. The solid line represents the standard normal density curve.

We see from the simulations that the Central Limit Theorem provides a good approximation. The convergence rate is fast as we see from the case where $n = 1000$. The histograms are very close to the standard normal density curve.

### 3.2 Testing for the presence of jumps

In this section, we will employ our test statistic $T_n$ proposed in Chapter 2 to test for the presence of jumps in the jump diffusion model as in (2.2.1)

$$X_t = \mu t + \sigma W_t + \sum_{i=1}^{N_t} J_i, \quad 0 \leq t \leq T,$$  \hspace{1cm} (3.2.1)
We will examine numerically the rejection rates under both null and alternative hypothesis. We first used the Monte Carlo simulation to obtain the rejection region, then the normal approximation (2.2.20) and they give the same rejection region, numerically.

3.2.1 Under the null hypothesis $H_0$

We simulate a diffusion and use our test statistics $T_n$ to perform a hypothesis test. As we have saw in Chapter 2, the asymptotic distribution of $T_n$ does not depend on the drift coefficient $\mu$ and the diffusion coefficient $\sigma$. Therefore we take $\mu = 0$ and $\sigma = 1$ and $T = 1$. Further, we take $\varepsilon = 0.05$ and $\varepsilon' = 1 - 10^{-5}$. We set the level of the test to be 0.05. The rejection rates of the null hypothesis $H_0$, from 500 simulations, are shown in the table below,

<table>
<thead>
<tr>
<th>$n$</th>
<th>Rejection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1 \times 10^5$</td>
<td>0.334</td>
</tr>
<tr>
<td>$5 \times 10^5$</td>
<td>0.198</td>
</tr>
<tr>
<td>$10 \times 10^5$</td>
<td>0.118</td>
</tr>
<tr>
<td>$20 \times 10^5$</td>
<td>0.066</td>
</tr>
<tr>
<td>$25 \times 10^5$</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Table 3.1: The rejection rates converge to 0.05 with as $n$ grows.

The rejection rates converge to 0.05, as expected but in a rather slow manner.

3.2.2 Under the alternative hypothesis $H_1$

Our objective is to detect jumps from discretely observed data. So our test should be able to pick up jumps even when they are very tiny that can not be detected by naked eyes.

The plots in Figure 3.3 show the sample paths from two Jump Diffusions, that is, $X$ is a process as defined in (2.2.1).
Figure 3.3: (a) $\mu = 1, \sigma = 1, J_i \sim N(0, 0.01)$. (b) $\mu = 1, \sigma = 1, J_i \sim N(0, 1)$.

As we see, jumps can be spotted by eye in (b), but not in (a).

We also expect that as $T$ grows, we will have more jumps. The longer we wait, the higher the chance we can detect the jumps. Indeed, numerical results obtained from simulation, for $\mu = 1, \sigma = 1, J_i \sim N(0, 0.01), \varepsilon = 0.05$ and $\varepsilon' = 1 - 10^{-10}$, confirm this observation.

<table>
<thead>
<tr>
<th>$n$</th>
<th>Rejection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.394</td>
</tr>
<tr>
<td>5000</td>
<td>0.600</td>
</tr>
<tr>
<td>10000</td>
<td>0.704</td>
</tr>
<tr>
<td>50000</td>
<td>0.752</td>
</tr>
<tr>
<td>100000</td>
<td>0.764</td>
</tr>
</tbody>
</table>

Table 3.2: The case $T = 1$
Table 3.3: The case $T = 5$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Rejection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.364</td>
</tr>
<tr>
<td>5000</td>
<td>0.818</td>
</tr>
<tr>
<td>10000</td>
<td>0.956</td>
</tr>
<tr>
<td>50000</td>
<td>0.990</td>
</tr>
<tr>
<td>100000</td>
<td>0.994</td>
</tr>
</tbody>
</table>

Table 3.4: The case $T = 10$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Rejection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>326</td>
</tr>
<tr>
<td>5000</td>
<td>0.886</td>
</tr>
<tr>
<td>10000</td>
<td>0.988</td>
</tr>
<tr>
<td>50000</td>
<td>1</td>
</tr>
<tr>
<td>100000</td>
<td>1</td>
</tr>
</tbody>
</table>

As expected, the numerical results confirm that, as $T$ grows, the rejection rates grows. That is, the power of the test grows. Also, the convergence rate is a lot faster compared to that in the case of approaching the limiting values for rejection rates of the null hypothesis.
3.2.3 Conclusion

The performance of our test statistic $T_n$, judging from the simulations, is quite satisfactory. As we saw in the previous section, it fulfils its role of detecting jumps with rather high efficiency, with a relatively small $n$.

One weakness of $T_n$ is the slow convergence rate of Type I error. The following heuristic argument gives us a picture behind its slow convergence rate.

For some constant $c$ and $(Z_1, Z_2)$ is jointly Gaussian with mean 0 and covariance matrix $\Sigma$ as defined in (2.2.3), as $n$ is large,

$$T_n \approx \frac{c + n^{-1/2}Z_1}{1 + n^{-1/2}Z_2} \approx (c + n^{1/2}Z_1)(1 - n^{-1/2}Z_2) \approx c + n^{-1/2}Z_1 - cn^{-1/2}Z_2.$$

So as $n$ increases 100 fold, the variability of $T_n$ will just roughly reduce 10 times. Hence we need to increase $n$ substantially to achieve convergence.

Apart from the slow convergence rate of Type I error, the test performs well in general.

It is also worth mentioning that the Central Limit Theorems for the trimmed sums provide a very good approximation with a fast convergence rate. However, the joint Gaussian property of $(S_n^{(c)}, \tilde{S}_n^{(c)})$ was not established and it is left for the future investigation.
Appendix A

This appendix is for the miscellaneous results we used in the previous chapters, but not in the Preliminary. We provided the results with proofs.

A.1 Miscellaneous results from real analysis

**Theorem A.1.1.** Let \( \{f_n\}_{n \geq 1} \) and \( f \) be a collection of bounded non-decreasing functions from \( \mathbb{R} \) to \( \mathbb{R} \) such that \( f \) is continuous on \( \mathbb{R} \) and
\[
f_n(x) \to f(x), \quad \forall x \in D \cup \{-\infty, \infty\},
\]
where \( D \) is dense in \( \mathbb{R} \). Then \( f_n \) converge to \( f \) uniformly on \( \mathbb{R} \).

**Proof.** Let us agree to denote \( f(\infty) \) to be \( \lim_{x \to \infty} f(x) \) and likewise \( f(-\infty) \) to be \( \lim_{x \to -\infty} f(x) \). Fix a \( \varepsilon > 0 \), since \( f \) is bounded continuous on \( \mathbb{R} \), there exists \( K_1, K_2 \in D \) such that
\[
f(K_1) - f(-\infty) < \varepsilon, \quad f(\infty) - f(K_2) < \varepsilon \quad (A.1.1)
\]
Since \( f \) is continuous, \( f \) is therefore uniformly continuous on \( [K_1, K_2] \). So there exists a \( \delta > 0 \) such that
\[
|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon, \quad \forall x, y \in [K_1, K_2]. \quad (A.1.2)
\]
Moreover, there exists \( a_1 = K_1 < a_2 < \ldots < a_k = K_2 \) where \( k \in \mathbb{N} \) and \( \{a_1, a_2, \ldots, a_k\} \subset D \) such that
\[
\max_{1 \leq i \leq k-1} (a_{i+1} - a_i) < \delta.
\]
Let $a_0 = -\infty$ and $a_{k+1} = \infty$. Since $f_n(x)$ converge to $f(x)$ for all $x \in D \cup \{-\infty, \infty\}$, we have

$$\max_{0 \leq i \leq k+1} |f_n(a_i) - f(a_i)| = \Delta_n \to 0, \quad n \to \infty. \quad (A.1.3)$$

Now for any $x \in \bigcup_{1 \leq i \leq k-1} [a_i, a_{i+1}]$, since $f_n$ and $f$ are non-decreasing, by (A.1.2) and (A.1.3),

$$f_n(x) - f(x) \leq f_n(a_{i+1}) - f(a_i)$$

$$\leq f(a_{i+1}) - f(a_i) + f(a_{i+1}) - f(a_i)$$

$$\leq \Delta_n + \varepsilon,$$

and likewise,

$$f_n(x) - f(x) \geq -\Delta_n + \varepsilon.$$

Therefore

$$\sup_{a_1 \leq x \leq a_k} |f_n(x) - f(x)| \leq \Delta_n + \varepsilon. \quad (A.1.4)$$

For $x < a_1$, by (A.1.1) and (A.1.4),

$$|f_n(x) - f(x)| \leq |f_n(x) - f_n(-\infty)| + |f_n(-\infty) - f(-\infty)|$$

$$+ |f(-\infty) - f(x)|$$

$$\leq |f_n(x) - f_n(-\infty)| + |f_n(-\infty) - f(-\infty)| + \varepsilon$$

$$\leq |f_n(a_1) - f(a_1)| + |f(a_1) - f(-\infty)|$$

$$+ 2|f_n(-\infty) - f(-\infty)| + \varepsilon$$

$$\leq 3\Delta_n + 2\varepsilon.$$

Likewise, for $x > a_k$,

$$|f_n(x) - f(x)| \leq 3\Delta_n + 2\varepsilon.$$

By the above arguments and (A.1.4), we obtained

$$\sup_{-\infty \leq x \leq \infty} |f_n(x) - f(x)| \leq 3\Delta_n + 2\varepsilon$$

It follows from (A.1.3) that

$$\lim_{n \to \infty} \sup_{-\infty \leq x \leq \infty} |f_n(x) - f(x)| \leq 2\varepsilon,$$

and since $\varepsilon > 0$,

$$\sup_{-\infty \leq x \leq \infty} |f_n(x) - f(x)| \to 0, \quad n \to \infty.$$

Hence proving the theorem. \qed
Corollary A.1.1. Let $X_n, n \geq 1$ and $X$ are random variables with respective distribution functions $F_n, n \geq 1$ and $F$. If $X_n \xrightarrow{d} X$ as $n \to \infty$ and $F$ is continuous, then $F_n$ converge to $F$ uniformly.

Proof. Since $X_n \xrightarrow{d} X$ as $n \to \infty$, by definition, $F_n(x) \to F(x)$ for all $x \in C(F)$ where $C(F)$ denotes the set of all continuity points of $F$. From the fact that $F$ can only have at most countably many discontinuity points, $C(F)$ is dense in the domain of $F$ which is a subset of $\mathbb{R}$. Therefore the result follows from Theorem A.1.1. $\blacksquare$

The following Corollary is an immediate consequence of Theorem A.1.1.

Corollary A.1.2. Let $K \subset \mathbb{R}$ be a compact subset, $\{f_n\}_{n \geq 1}$ be a sequence of non-decreasing functions on $K$. If $f_n$ converge uniformly to $f$ where $f$ is continuous on $K$. Then $f_n$ converge uniformly to $f$ on $K$.

Theorem A.1.2. Let $(X,d_X),(Y,d_Y),(Z,d_Z)$ be metric spaces, $g : Y \to Z$ is uniformly continuous and $\{f_n\}_{n \geq 1}$ be a sequence of functions from $X$ to $Y$ such that $f_n$ converge uniformly to $f$. Then $g \circ f_n$ converge uniformly to $g \circ f$.

Proof. Since $g$ is uniformly continuous, given $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $y_1, y_2 \in Y$,

$$d_Y(y_1, y_2) < \delta \Rightarrow d_Z(g(y_1), g(y_2)) < \varepsilon.$$

Moreover, as $f_n$ converge to $f$ uniformly, there exist a $N \in \mathbb{N}$ such that

$$\sup_{x \in X} d_Y(f_n(x), f(x)) < \delta, \quad \forall n \geq N.$$

Therefore implies, for all $x \in X$

$$d_Z(g(f_n(x)), g(f(x))) < \varepsilon.$$

Thus proving that $g \circ f_n$ converge uniformly to $g \circ f$. $\blacksquare$
A.2 Mill’s Ratio

Lemma A.2.1 (Mill’s ratio). Let $\Phi$ be the standard normal distribution function. Then
\[
\left(\frac{1}{x} - \frac{1}{x^3}\right) \phi(x) < \Phi(x) < \frac{\phi(x)}{x}, \quad \forall x > 0
\]
where $\text{phi} = \Phi'$ and $\bar{\Phi} = 1 - \Phi$.

Proof. Let us denote
\[
f(x) = \left(\frac{1}{x} - \frac{1}{x^3}\right) \phi(x), \quad g(x) = \bar{\Phi}(x), \quad h(x) = \frac{\phi(x)}{x}
\]

Since $f(x)$, $g(x)$ and $h(x)$ are differentiable and vanishing when $x \to \infty$. So,
\[
f(x) = -\int_x^\infty f'(t)dt, \quad g(x) = -\int_x^\infty g'(t)dt, \quad h(x) = -\int_x^\infty h'(t)dt.
\]
Therefore it suffices to show $h'(x) < f'(x) < g'(x)$ for $x > 0$.

Since,
\[
f'(x) = \left(-\frac{1}{x^2} + \frac{3}{x^4}\right) \phi(x) + \left(\frac{1}{x} - \frac{1}{x^3}\right) \phi'(x)
\]
\[
= \left(-\frac{1}{x^2} + \frac{3}{x^4}\right) \phi(x) + \left(\frac{1}{x} - \frac{1}{x^3}\right) x \phi(x)
\]
\[
= \left(1 - \frac{3}{x^4}\right) \phi(x)
\]
and
\[
g'(x) = -\phi(x)
\]
and
\[
h'(x) = \frac{x \phi'(x) - \phi(x)}{x^2}
\]
\[
= -x^2 \phi(x) - \phi(x)
\]
\[
= -\left(1 + \frac{1}{x^2}\right) \phi(x).
\]
Therefore we see that $h'(x) < g'(x) < f'(x)$ and thus proving the result. \qed


