Sequential Change Detection

By : Wan Yijing
Supervisor: Professor Keiji Nagai

Student ID number   10LB903
Department of Economics
Graduate School of International Social Sciences
YOKOHAMA National University
Jan 24, 2014
## Contents

Abstract............................................................................................................................. 1  
Chapter 1 Introduction...................................................................................................... 1  
Chapter 2 Mathematical Prerequisites............................................................................... 4  
  2.1 Some Definitions ........................................................................................................ 4  
    2.1.1 Probability Measures ....................................................................................... 4  
    2.1.2 Measures .......................................................................................................... 5  
    2.1.3 Random Variable and Expectation .................................................................. 6  
    2.1.4 Independence ................................................................................................... 7  
    2.1.5 Conditional Expectation .................................................................................. 7  
    2.1.6 Stochastic Process ........................................................................................... 8  
    2.1.7 Convergence .................................................................................................... 9  
      2.1.7.1 Weak Convergence ............................................................................... 9  
      2.1.7.2 Convergence of Random Variables .................................................... 10  
      2.1.7.3 Fundamental Theorems ...................................................................... 12  
    2.1.8 Uniform Integrability ..................................................................................... 14  
    2.1.9 The Law of Large Numbers ........................................................................... 15  
      2.1.9.1 Weak Law of Large Numbers ............................................................. 15  
      2.1.9.2 Strong Law of Larger Numbers .......................................................... 16  
      2.1.9.3 Differences between the Weak Law and the Strong Law ................... 16  
      2.1.9.5 Borel's Law of Large Numbers ........................................................... 17  
    2.1.10 The Central Limit Theorem ........................................................................... 23  
  2.2 Martingales ............................................................................................................... 18  
    2.2.1 Definition and Main Properties ..................................................................... 18  
    2.2.2 Submartingales and Supermartingales........................................................... 20  
    2.2.3 Space of Martingales ..................................................................................... 21  
    2.2.4 Stopping Times .............................................................................................. 21  
    2.2.5 Local Martingales .......................................................................................... 22  
    2.2.6 Martingale Convergence Theorem ............................................................... 23  
    2.2.7 Martingale Central Limit Theorem ................................................................ 23  
  2.3 Markov Process ........................................................................................................ 24  
  2.4 Brownian Motion...................................................................................................... 24
5.3.2 The expectation of stopping time $H_0$ .......................................................... 68
5.3.3 The expectation of stopping time $H_1$ .......................................................... 69
5.3.3 The error of the first kind ($\alpha$) and the error of the second kind($\beta$) .......... 69

Chapter 6 CUSUM Test (Cumulative Sum Test) .......................................................... 72
6.1 CUSUM Procedures .............................................................................................. 72
6.2 Cusum test of AR(1) process ................................................................................. 74
   6.2.1 The structural change ................................................................................ 74
   6.2.2 Stopping Time ......................................................................................... 75
   6.2.3 Change at the beginning ......................................................................... 77
   6.2.4 Change at the time $k$ $(k < \infty)$ ......................................................... 78
   6.2.5 No change occurs .................................................................................... 79

Chapter 7 Numerical Studies ......................................................................................... 80
7.1 Sequential Unit Root Test ................................................................................. 80
7.2 SPRT ................................................................................................................. 80
7.3 Cusum Test ....................................................................................................... 81
7.4 Comparison of the Test .................................................................................. 81
   7.4.1 Shiryayev-Roberts Rule ................................................................. 81
   7.4.2 Comparison of Three Tests ............................................................... 82

Chapter 8 Conclusion .............................................................................................. 83

References .............................................................................................................. 84

TABLE 1 SEQUENTIAL UNIT ROOT ....................................................................... 85
TABLE 2 SPRT 1 .................................................................................................. 85
TABLE 3 SPRT 2 .................................................................................................. 86
TABLE 4 CUSUM TEST 1 ................................................................................. 86
TABLE 5 CUSUM TEST 2 ................................................................................. 87
TABLE 6 COMPARISON OF THREE TESTS 1 .................................................. 87
Abstract

We consider SPRT and CUSUM Test under sequent sampling for an autoregressive (AR) process with dependent disturbances to deal with the change point. Motivated by the stopping rules of Wald’s SPRT and Page’s CUSUM test, we use time change and DDS (Dambis and Dubins-Shward) Brownian motion to obtain the mean of stopping time and the false alarm. And we also implement numerical computations to proof our conclusion.

Key words: autoregressive process, sequential unit root test, SPRT (sequential probability ratio test), stopping time, change point, CUSUM test, and false alarm

Chapter 1 Introduction

The first person that wanted to deal with the change point problem formally is due to Shewhart (1931). He was primarily interested in detecting a shift in the mean of a normal distribution. His research is based on a series of independent variables with known baseline distribution. He proposed that an observation will exceed the (known) baseline mean by more than three standard deviations when the alarm occurs the first time. His proposition is classical; ARL’s (average run lengths, until a false alarm, and from change to its detection) are the operating characteristics of this method. The method is known to be very good in detecting a large change quickly.

During the next decade, the fact that the Shewhart procedure does not enable information to accumulate brought about ad hoc attempts to correct for this (such as ”warning lines” and ”action lines”, where too many proximate observations in exceedence of a warning line would also be cause for alarm). The perspective of these methods is also classical.

The first person that uses Bayesian consideration to deal with the problem is due to Girschick and Rubin (1952). They assumed the observations to be independent and $f_0$ and $f_1$ to be known. They posited a geometric prior on the changepoint and a gain (or loss) function for each observation. Their objective was to maximize expected gain per observation. Their solution calls for raising an alarm whenever the posterior probability of a change having taken place is large enough. This procedure is a precursor of the Shiryaev-Roberts procedure. With the same probabilistic structure, Shiryaev (1963,
1978) considered minimizing $E(N - ν|N \geq ν)$ when each post-change observation costs $c > 0$ units and the penalty for a false alarm is 1 unit. His solution is the same.

The next development was classical: Page (1954) proposed the Cusum scheme, which in essence is a repeated sequence of SPRT’s defined by $f_0$ and $f_1$ that calls for an alarm the first time that a SPRT exits on the side of $f_1$. The ARL to false alarm and the average delay to detection are the relevant operating characteristics. The observations are assumed to be independent, $f_0$ is assumed known, and although the post-change distribution need not be known to implement the procedure, it is necessary to represent it by a (fixed) density $f_1$ in order to spell out the SPRT. The definition was ad hoc. Lorden (1971) proved that the minimum over all stopping times $T$ with ARL to false alarm $\geq B$ of $\sup_{1 \leq k < \infty} \text{ess sup} E[ν] = k(T - k|X_1, \ldots, X_{k-1})$ is $(I + (O(1)) \times (\log B)/I$, where $I$ is the Kullback-Leibler information number (of the post-change density vs. the pre-change density) and $O(1) \to 0$ as $B \to \infty$, and showed that the Cusum scheme achieves this asymptotic lower limit. Moustakides (1986) went the last leg and proved that (when $f_1$ is the true post-change density) the appropriate Cusum scheme is strictly optimal in minimizing $\sup_{1 \leq k < \infty} \text{ess sup} E[ν] = k(T - k|X_1, \ldots, X_{k-1})$ over all stopping times $T$ with ARL to false alarm $\geq B$. In this context, Ritov (1990) proved the optimality of the Cusum in a game-theoretic setup, with Nature and the statistician being opposing players. Beibel (1996) proved the same in a Brownian motion context.

Roberts (1959) proposed the exponentially weighted moving average (EWMA) method. His approach is classical (motivated by time series). Srivastava and Wu (1993) found this method to be inferior to others.

In a Brownian motion context of detecting a shift in mean, Shiryaev (1961, 1963) considered the problem of detecting an object with the aim of minimizing expected delay (from change to detection), asymptotically after a long run of false alarms raised by successive application of a stopping time $N$, under the constraint that the ARL to false alarm (in a single application of $N$) be $\geq B$. He found that the optimal procedure is analogous to that of Girschick and Rubin’s when the parameter of the geometric prior tends to zero. (A discrete time analog of Shiryaev’s result was derived by Pollak and Tartakovsky, 2009.) Independently, Roberts (1966) was the first to consider this limit in the context of reducing expected delay to detection (of a single application of the stopping time) subject to a lower bound on the ARL to false alarm. Roberts studied the
procedure by simulation, comparing it to other procedures (Cusum, EWMA and others), and found it to be good. The procedure is now known as the Shiryaev-Roberts procedure.

Pollak (1985) also considered the changepoint problem in a classical framework. The conditions considered are that the observations are independent, with known $f_0$ and $f_1$, and the goal is to minimize $\sup_{1 \leq k < \infty} E[\nu] = k(T - k|N \geq k)$ subject to ARL to false alarm $\geq B$. He found that that a method based on starting the Shiryaev-Roberts procedure at a random value is optimal to within an additive $O(1)$ term, with $O(1) \to 0$ as $B \to \infty$. The method of proof is Bayesian; he took Shiryaev’s (1978) solution of the Bayesian problem a step further, showing that the aforementioned procedure is a limit of Bayes rules. The question of whether this procedure is strictly optimal was open until very recently; Polunchenko and Tartakovsky (2009) produced a counterexample.

In this paper, we use SPRT (sequential probability ratio test) and Cusum Test (cumulative sum test) to deal with the change point problem under sequential sampling for an autoregressive process with dependent disturbances. Unlike the former methods is that we use DDS (Dambis and Dubins-Sharz) Brownian motion to obtain the joint limiting distribution of the stopping time and the false alarm.

The chapter 2 is to generalize the theorem that was used in this paper. This chapter can let readers to understand our theory better. In the chapter 3, we describe the three processes of the AR (1) model. The chapter 4, 5 and 6 is the very important chapter of this paper, and the proposition that we want to explain is in these chapters. In the chapter 4, motivated by Lai and Siegmund (1983)’s sequential estimation for AR coefficient, we formulate a sequential unit root test by using a stopping time based on the observed Fisher information. Using time change and a DDS Brownian motion, we obtain the joint limiting distribution of the stopping time and the sequential least squares estimate of the AR coefficient under null and the local alternatives. In the Chapter 5, we use SPRT and DDS Brownian motion to obtain what we want to obtain. The CUSUM test is used in the chapter 6, and by DDS Brownian motion we want to get the false alarm that we want to know. To proof our result, we implement numerical computation in the chapter 7. And we let SPRT and CUSUM test to compare with the Shirayayev-Roberts’s rule. In the final chapter, we summarize the main points that we describe in this paper. Generally speaking, in this paper, we just want to find a more accurate method to detect change point by SPRT and CUSUM test under AR(1) model.
2.1 Some Definitions

In this chapter, we try to give all the important definitions and theorems for further use. Maybe we will lose some proofs for the theorems, the reader can refer to the books of John Wiley and Sons (1995), Billingsley, P. (1999) and David Williams (1991).

2.1.1 Probability Measures

Let \( \Omega \) be an arbitrary space or set of points \( \omega \). In probability theory \( \Omega \) consists of all the possible results or outcomes \( \omega \) of an experiment or observation. Viewed probabilistically, subset of \( \Omega \) is an event and an element \( \omega \) of \( \Omega \) is a sample point.

A probability measure \( P \) on a measurable space \( (\Omega, \mathcal{F}) \) is a map from \( \mathcal{F} \) to \([0,1] \) such that:

- \( P(\Omega) = 1 \),
- \( P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \) for any countable family of disjoint sets \( A_i \in \mathcal{F} \), i.e., such that \( A_i \cap A_j = \emptyset \) for \( i \neq j \).

A set function is a real-valued function defined on some class of subsets of \( \Omega \). A set function \( P \) on a field \( \mathcal{F} \) is a probability measure if it satisfies these conditions:

(i) \( 0 \leq P(A) \leq 1 \) for \( A \in \mathcal{F} \);
(ii) \( P(\emptyset) = 0, P(\Omega) = 1 \);
(iii) if \( A_1, A_2, \ldots \) is a disjoint sequence of \( \mathcal{F} \)-sets and if \( \bigcup_{k=1}^{\infty} A_k \in \mathcal{F} \), then

\[
P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k). \tag{2.1}
\]

The condition imposed on the set function \( P \) by (iii) is called countable additivity. Since \( \mathcal{F} \) is a field but perhaps not a \( \sigma \)-field, it is necessary in (iii) to assume that \( \bigcup_{k=1}^{\infty} A_k \) lies on \( \mathcal{F} \). If \( A_1, \ldots, A_n \) are disjoint \( \mathcal{F} \)-sets, then \( \bigcup_{k=1}^{\infty} A_k \) is also in \( \mathcal{F} \) and (2.1) with \( A_{n+1} = A_{n+2} = \cdots = \emptyset \) gives

\[
P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k). \tag{2.2}
\]

The condition that (2.4) holds for disjoint \( \mathcal{F} \)-sets is additivity; it is a consequence of countable additivity.

We will give a theorem about probability measure that we will frequently use on the further chapter.
Theorem 2.1.1 Let $P$ be probability measure on a field $\mathcal{F}$.

(i) Continuity from below: If $A_n$ and $A$ lie in $\mathcal{F}$ and $A_n \uparrow A$, then $P(A_n) \uparrow P(A)$.

(ii) Continuity from above: If $A_n$ and $A$ lie in $\mathcal{F}$ and $A_n \downarrow A$, then $P(A_n) \downarrow P(A)$.

(iii) Countable subadditivity: If $A, B,$ and $\bigcup_{k=1}^{\infty} A_k$ lie in $\mathcal{F}$ (the $A_k$ need not be disjoint), then

$$P(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} P(A_k). \quad (2.3)$$

2.1.2 Measures

Given a space $\Omega$, a $\sigma$-algebra on $\Omega$ is a class $\mathcal{F}$ of subsets of $\Omega$, such that $\mathcal{F}$ is closed under complements and countable intersection (hence under countable union) and $\emptyset \in \mathcal{F}$ (hence, $\Omega \in \mathcal{F}$). For given class $\mathcal{C}$ of subsets of $\Omega$, we denote by $\sigma(\mathcal{C})$ the smallest $\sigma$-algebra which contains $\mathcal{C}$ (i.e., the intersection of all the $\sigma$-algebra containing $\mathcal{C}$).

A measurable space $(\Omega, \mathcal{F})$ is a space $\Omega$ endowed with a $\sigma$-algebra $\mathcal{F}$. A measurable map $X$ from $(\Omega, \mathcal{F})$ to another measurable space $(E, \mathcal{E})$ is a map from $\Omega$ to $E$ such that, for any $B \in \mathcal{E}$, the set

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$$

belongs to $\mathcal{F}$.

A real-valued random variable (r.v.) on $(\Omega, \mathcal{F})$ is a measurable map from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$ where $\mathcal{B}$ is the Borel $\sigma$-algebra, i.e., the smallest $\sigma$-algebra that contains the intervals.

Let $X$ be a real-valued random variable on a measurable space $(\Omega, \mathcal{F})$. The $\sigma$-algebra generated by $X$, denoted $\sigma(X)$, is $\sigma(X) := \{X^{-1}(B) ; B \in \mathcal{B}\}$. Doob’s theorem asserts that any $\sigma(X)$-measurable real-valued r.v. can be written as $h(X)$ where $h$ is a Borel function. The set of bounded Borel functions on measurable space $(E, \mathcal{E})$. If $\mathcal{H}$ is $\sigma$-algebra on $\Omega$, we shall make the slight abuse of notation by writing $X \in \mathcal{H}$ for $X$ a bounded r.v. in $\mathcal{H}$.

Let $(X_i, i \in I)$ be a set of random variables. There exists a unique r.v. with values in $\mathbb{R}$, denoted $\text{esssup}_i X_i$ (essential supremum of the family $(X_i; i \in I)$) such that, for any r.v. $Y$,

$$X_i \leq Y_i \text{ a.s. \forall } i \in I \Leftrightarrow \text{esssup}_i X_i \leq Y.$$

If the family is countable, $\text{esssup}_i X_i = \sup_i X_i$. In the case where the set $I$ is not
countable, the map \( \text{sup}_t X_t \) (pointwise supremum) may not be a random variable.

A set function \( \mu \) on a field \( \mathcal{F} \) in \( \Omega \) is a measure if it satisfies these conditions:

(i) \( \mu(A) \in [0, \infty] \) for \( A \in \mathcal{F} \);

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

(iii) If \( A_1, A_2, \ldots \) is a disjoint sequence of \( \mathcal{F} \)-sets and if

\[
\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k). \tag{2.4}
\]

The measure \( \mu \) is finite or infinite as \( \mu(\Omega) < \infty \) or \( \mu(\Omega) = \infty \); it is a probability measure if \( \mu(\Omega) = 1 \). If \( \Omega = A_1 \cup A_2 \cup \cdots \) for some finite or countable sequence of \( \mathcal{F} \)-sets satisfying \( \mu(A_k) < \infty \), then \( \mu \) is \( \sigma \)-finite.

2.1.3 Random Variable and Expectation

Let \( (\Omega, \mathcal{F}, P) \) be an arbitrary probability space, and let \( X \) be a real-valued function on \( \Omega \); \( X \) is a simple random variable if it has finite range and if

\[
[\omega : X(\omega) = x] \in \mathcal{F}
\]

for each real \( x \). And its expectation

\[
E[X] = E\left[ \sum_i x_i 1_{A_i} \right] = \sum_i x_i P(A_i) \tag{2.5}
\]

The expected value of \( X \) is the integral of \( X \) with respect to the measure \( P \):

\[
E[X] = \int X dP = \int_{\Omega} X(\omega) P(d\omega). \tag{2.6}
\]

If a real-valued random variable \( X \) defined on the space \( (\Omega, \mathcal{F}, P) \) is the probability measure \( P_X \) on \( (\mathbb{R}, \mathcal{B}) \) defined by

\[
\forall A \in \mathcal{B}, P_X(A) = P(X \in A).
\]

It is the image on \( (\mathbb{R}, \mathcal{B}) \) of \( P \) by the map \( \omega \to X(\omega) \). This definition extends to an \( \mathbb{R}^n \) random variable, and, more generally, to an \( E \) random variable (a measurable map from \( (\Omega, \mathcal{F}) \) to \( (E, \mathcal{E}) \)).

The cumulative distribution function of random variable \( X \) is the right-continuous function \( F \) defined as \( F(x) = P(X \leq x) \), then the expectation of a positive random variable \( Z \) is defined as

\[
E[Z] = \int Z dP = \int_{\mathbb{R}^+} X dP_Z(x), \tag{2.7}
\]

and, if \( E(|X|) < \infty \), then \( E(X) = E(X^+) - E(X^-) \). The random variable \( X \) is said to be \( P \)-integrable if \( E(|X|) < \infty \). The integral \( \int_A X dP \) over a set \( A \) is defined, ad before, as \( E[I_A X] \).
2.1.4 Independence

A family of random variables \((X_i, i \in I)\), defined on the space \((\Omega, \mathcal{F}, P)\), is said to be independent if, for any \(n\) distinct indices \((i_1, i_2, \ldots, i_n)\) with \(i_k \in I\) and for any \((A_1, \ldots, A_n)\) where \(A_k \in \mathcal{B}\),

\[
P\left( \bigcap_{k=1}^{n} \{X_{i_k} \in A_k\} \right) = \prod_{k=1}^{n} P\left( X_{i_k} \in A_k \right). \tag{2.8}
\]

Suppose that \(X\) and \(Y\) are independent. If they are also simple, then

\[
E[XY]=E[X]E[Y].
\]

Define \(X_1, \ldots, X_n\) are independent and integrable, then the product \(X_1, \ldots, X_n\) is also integrable and

\[
E[X_1, \ldots, X_n] = E[X_1] \cdots E[X_n]. \tag{2.9}
\]

Suppose that \(\mathcal{G}_1\) and \(\mathcal{G}_2\) are independent \(\sigma\)-fields, \(A\) lies in \(\mathcal{G}_1\), \(X_1\) is measurable \(\mathcal{G}_1\) and \(X_2\) is measurable \(\mathcal{G}_2\). Then \(I_A X_1\) and \(X_2\) are independent, so that (2.9) gives

\[
\int_A X_1 X_2 dP = \int_A X_1 dP E[X_2]. \tag{2.10}
\]

if the random variables are integrable. IN particular,

\[
\int_A X_2 dP = P(A) E[X_2]. \tag{2.11}
\]

A classical application of the monotone class theorem is that, if the random variable \((X_i, i \in I)\) are independent, then, with the same notation as above, for any bounded Borel function \(f_k\),

\[
E\left[ \prod_{k=1}^{n} f_k(X_{i_k}) \right] = \prod_{k=1}^{n} E[f_k(X_{i_k})]. \tag{2.12}
\]

2.1.5 Conditional Expectation

Let \(X\) be an integrable random variable on the space \((\Omega, \mathcal{F}, P)\) and \(\mathcal{H}\) a \(\sigma\)-algebra contained in \(\mathcal{F}\), i.e., \(\mathcal{H} \subseteq \mathcal{F}\). The conditional expectations of \(X\) given \(\mathcal{H}\) have two properties:

(i) \(E[X|\mathcal{H}]\) is measurable \(\mathcal{H}\) and integrable.

(ii) \(E[X|\mathcal{H}]\) satisfies the functional equation

\[
\int_{\mathcal{H}} E[X|\mathcal{H}] dP = \int_{\mathcal{H}} X dP, \quad H \in \mathcal{H}. \tag{2.13}
\]

The conditional expectation is denoted \(E[X|\mathcal{H}]\) and the following properties hold:

- If \(X\) is \(\mathcal{H}\)-measurable, \(E[X|\mathcal{H}]=X\), a.s.
- \(E[E[X|\mathcal{H}]] = E[X]\).
- If \(X \geq 0\), then \(E[X|\mathcal{H}] \geq 0\), a.s.
- Linearity: If \(Y\) is an integrable random variable and \(a, b \in \mathbb{R}\),
\[ E[(aX + bY)|\mathcal{H}] = aE[X|\mathcal{H}] + bE[Y|\mathcal{H}], \text{ a.s.} \]

- If \( \mathcal{G} \) is another \( \sigma \)-algebra and \( \mathcal{G} \subset \mathcal{H} \), then
  \[ E[E[X|\mathcal{G}]|\mathcal{H}] = E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{G}], \text{ a.s.} \]
- If \( Y \) is \( \mathcal{H} \)-measurable and \( XY \) is integrable, \( E[XY|\mathcal{H}] = YE[X|\mathcal{H}] \), a.s.
- Jensen’s inequality: If \( f \) is a convex function such that \( f(X) \) is integrable, \( E[f(X)|\mathcal{H}] \geq f(E[X|\mathcal{H}]), \text{ a.s.} \)

In the particular case where \( \mathcal{H} \) is the \( \sigma \)-algebra generated by a random variable \( Y \), then \( E[X|\sigma(Y)] \), which is usually denoted by \( E[X|Y] \), is \( \sigma(Y) \)-measurable, hence there exists a Borel function \( \varphi \) such that \( E[X|Y] = \varphi(Y) \). The notation \( E[X|Y=y] \) is often used for \( \varphi(y) \).

If \( X \) is an \( R^p \)-valued random variable, and \( Y \) an \( R^p \)-valued random variable, there exists a family of measures \( \mu(dx, y) \) such that, for any bound Borel function \( h \)
\[
E[h(X)|Y = y] = \int h(x)\mu(dx, y). \tag{2.14}
\]

If \( (X, Y) \) are independent random variable, and \( h \) is a bounded Borel function, the \( E[h(X,Y)|Y]=\psi(Y) \), where \( \psi(Y)=E[h(X, y)] \), i.e., the conditional law of \( X \) given \( Y=y \) does not depend on \( y \).

Note that, if \( X \) is square integrable, then \( E[X|\mathcal{H}] \) may be defined as the projection of \( X \) on the space \( L^2(\Omega, \mathcal{H}) \) of \( \mathcal{H} \)-measurable square integrable random variables. The conditional variance of a square integrable random variable \( X \) is
\[
\text{var}(X|\mathcal{H}) = E[X^2|\mathcal{H}] - (E[X|\mathcal{H}])^2. \tag{2.15}
\]

### 2.1.6 Stochastic Process

**Definition 2.1.6.1** A continuous time process \( X \) on \((\Omega, \mathcal{F}, P)\) is a family of random variables \((X_t, t \geq 0)\), such that the map \((\omega, t) \mapsto X_t(\omega)\) is \( \mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+) \) measurable.

**Definition 2.1.6.2** A process \( X \) is increasing if \( X_0 = 0 \), \( X \) is right-continuous, and \( X_s \leq X_t, \text{ a.s. for } s \leq t \).

**Definition 2.1.6.3** Let \((\Omega, \mathcal{F}, F, \mathbb{P})\) be a filtered probability space. The process \( X \) is \( F \)-adapted if for any \( t \geq 0 \), the random variable \( X_t \) is \( \mathcal{F}_t \)-measurable.
The natural filtration $F^X$ of a stochastic process $X$ is the smallest filtration $F$ which satisfies the usual hypotheses and such that $X$ is $F$-adapted. We shall write in short $\mathcal{F}_t^X = \sigma(X_s, s \leq t)$.

Let $G=(g_t, t \geq 0)$ be another filtration on $\Omega$. If $G$ is larger than $F$, and if $X$ is an $F$-adapted process, it is also $G$-adapted.

**Definition 2.1.6.4** A real-valued process $X$ is progressively measurable with respect to a given filtration $F=(\mathcal{F}_t, t \geq 0)$, if, for every $t$, the map $(\omega, s) \to X_s(\omega)$ from $\Omega \times [0, t]$ into $\mathbb{R}$ is $\mathcal{F}_t \times \mathcal{B}([0, t])$-measurable.

If $X$ is progressively measurable, then
\[
E \left[ \int_0^\infty X_t \, dt \right] = \int_0^\infty E[X_t] \, dt,
\]
where the existence of one of these expressions implies the existence of the other.

**2.1.7 Convergence**

Many of the best-know theorems in probability have to do with the asymptotic behavior of distributions. This section covers both general methods for deriving such theorems and specific applications.

**2.1.7.1 Weak Convergence**

If $F_n$ and $F$ are distribution functions, then by definition, $F_n$ converges weakly to $F$, written $F_n \Rightarrow F$, if
\[
\lim_n F_n(x) = F(x)
\]
for each $x$ at which $F$ is continuous.

If $\mu_n$ and $\mu$ are the probability measures in $(R^1, \mathcal{B}^1)$ corresponding to $F_n$ and $F$, then $F_n \Rightarrow F$ if and only if
\[
\lim_n \mu_n(A) = \mu(A)
\]
for every $A$ of the form $A = (-\infty, x]$ for which $\mu\{x\} = 0$.

Distribution functions $F$ and $G$ are of the same type if there exist constants $a$ and $b$, $a > 0$, such that $F(ax+b)=G(x)$ for all $x$. 


Theorem 2.1.7.1 Suppose that \( F_n(u_n + v_n) \Rightarrow F(x) \) and \( F_n(a_n + b_n) \Rightarrow G(x) \), where \( u_n > 0, a_n > 0, \) and \( F \) and \( G \) are nondegenerate. Then there exist \( a \) and \( b, a > 0, \) such that \( a_n \xi_n \to a, (b_n - v_n) \xi_n \to b, \) and \( F(ax + b) = G(x) \).

Lemma 1. If \( F_n \Rightarrow F, a_n \to a, \) and \( b_n \to b, \) then
\[
F_n(a_nx_n + b_n) \Rightarrow F(ax + b).
\]

Lemma 2. If \( F_n \Rightarrow F \) and \( a_n \to \infty, \) then
\[
\lim_{n} F_n(a_nx) = 1 \quad \text{for} \quad x > 0; \\
\lim_{n} F_n(a_nx) = 0 \quad \text{for} \quad x < 0.
\]

Lemma 3. If \( F_n \Rightarrow F \) and \( b_n \) is unbounded, then
\( F_n(x + b_n) \) cannot converge weakly.

Lemma 4. If \( F_n(x) \Rightarrow F(x) \) and \( F_n(a_nx + b_n) \Rightarrow G(x) \), where \( F \) and \( G \) are monodegenerate, then
\[
0 < \inf_n a_n \leq \sup_n a_n < \infty, \sup_n |b_n| < \infty.
\]

Lemma 5. If \( F(x) = F(ax + b) \) for all \( x \) and \( F \) is nondegenerate, then
\( a = 1 \) and \( b = 0. \)

2.1.7.2 Convergence of Random Variables

Convergence in Distribution
Let \( X_n \) and \( X \) be random variables with respective distribution functions \( F_n \) and \( F. \) If \( F_n \Rightarrow F, \) then \( X_n \) is said to converge in distribution or in law to \( X, \) written \( X_n \Rightarrow X. \)

Because of the defining conditions (2.16) and (2.17), \( X_n \Rightarrow X \) if and only if
\[
\lim_{n} P[X_n \leq x] = P[X \leq x]
\]
for every \( x \) such that \( P[X = x] = 0. \)

Convergence in Probability
Suppose that \( X_1, X_2, \ldots \) are random variables all defined on the same probability space \( (\Omega, \mathcal{F}, P). \) If \( X_n \to X \) with probability 1, then \( P[|X_n - X| \geq \epsilon \ i.o.] = 0 \) for \( \epsilon > 0, \) and hence
\[
\lim_{n} P[|X_n - X| > \epsilon] = 0.
\]
Thus there is convergence in probability \( X_n \to_p X. \)

Suppose that (2.19) holds for each positive \( \epsilon. \) Now \( P[X \leq x - \epsilon] - P[|X_n - x| \geq \epsilon] \to 0 \) as \( n \to \infty, \) hence
\[
\lim_{n} P[X \leq x - \epsilon] = P[X \leq x].
\]
\[ X \geq \epsilon \] \leq P[X_n \leq x] \leq P[X \leq x + \epsilon] + P[|X_n - x| \geq \epsilon] \; \text{letting } n \text{ tend to } \infty \text{ and then letting } \epsilon \text{ tend to } 0 \text{ shows that} \lim_{n \to \infty} P[X_n \leq x] \leq \limsup P[X \leq x] \leq P[X \leq x]. \text{ Thus } P[X_n < x] \to P[X \leq x] \text{ if } P[X=x]=0, \text{ and so } X_n \Rightarrow X:

**Theorem 2.1.7.2** Suppose that \( X_n \) and \( X \) are random variables on the same probability space. If \( X_n \to X \) with probability 1, then \( X_n \to_p X \). If \( X_n \to_p X \), then \( X_n \Rightarrow X \).

**Theorem 2.1.7.3** If \( X_n \Rightarrow X \) and \( X_n - Y_n \Rightarrow 0 \), then \( Y_n \Rightarrow X \).

**Theorem 2.1.7.4** If for each \( u \), \( X_n^{(u)} \Rightarrow X^{(u)} \) as \( n \to \infty \), if \( X^{(u)} \Rightarrow X \) as \( u \to \infty \), and if

\[
\lim_{u} \limsup_{n} P \left[ \left| X_n^{(u)} - Y_n \right| \geq \epsilon \right] = 0
\]

(2.21)

for positive \( \epsilon \), then \( Y_n \Rightarrow X \).

**Almost Surely Convergence**

This is the type of stochastic convergence that is most similar to pointwise convergence known from elementary real analysis.

\[ P(\lim_{n \to \infty} X_n = X) = 1. \] (2.22)

This means that the values of \( X_n \) approach the value of \( X \), in the sense that events for which \( X_n \) does not converge to \( X \) have probability 0. Using the probability space \((\Omega, \mathcal{F}, P)\) and the concept of the random variable as a function from \( \Omega \) to \( \mathbb{R} \), this is equivalent to the statement

\[ P(\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)) = 1. \] (2.23)

Another, equivalent, way of defining almost sure convergence is as follows:

\[ \lim_{n \to \infty} P(\omega \in \Omega : \sup_{m \geq n} |X_m(\omega) - X(\omega)| \geq \epsilon = 0, \text{ for all } \epsilon > 0 \] (2.24)

Almost sure convergence is often denoted by adding the letters \( a.s. \) over an arrow indicating convergence: \( X_n \xrightarrow{a.s.} X \).

For generic random elements \( \{X_n\} \) on a metric space \((S, d)\), convergence almost surely is defined similarly:

\[ P \left( \omega \in \Omega : d(X_n(\omega), X(\omega)) \xrightarrow{n \to \infty} 0 \right) = 1. \] (2.25)
Sure convergence

To say that the sequence of random variables \( X_n \) defined over the same probability space (i.e., a random process) converges surely or everywhere or pointwise towards \( X \) means

\[
\lim_{n \to \infty} X_n(\omega) = X(\omega), \forall \omega \in \Omega,
\]  

(2.26)

where \( \Omega \) is the sample space of the underlying probability space over which the random variables are defined. This is the notion of pointwise convergence of sequence functions extended to sequence of random variables. (Note that random variables themselves are functions).

\[
\{\omega \in \Omega | \lim_{n \to \infty} X_n(\omega) = X(\omega)\} = \Omega.
\]  

(2.27)

Sure convergence of a random variable implies all the other kinds of convergence stated above, but there is no payoff in probability theory by using sure convergence compared to using almost sure convergence. The difference between the two only exists on sets with probability zero. This is why the concept of sure convergence of random variables is very rarely used.

Convergence In Mean

We say that the sequence \( X_n \) converges in the \( r \)-th mean (or in the \( L^r \)-norm) towards \( X \), for some \( r \geq 1 \), if \( r \)-th absolute moments of \( X_n \) and \( X \) exist, and

\[
\lim_{n \to \infty} E(|X_n - X|^r) = 0
\]  

(2.28)

where the operator \( E \) denotes the expected value. Convergence in \( r \)-th mean tells us that the expectation of the \( r \)-th power of the difference between \( X_n \) and \( X \) converges to zero. This type of convergence is often denoted by adding the letter \( L^r \) over an arrow indicating convergence: \( X_n \to_{L^r} X \).

2.1.7.3 Fundamental Theorems

Theorem 2.1.7.5 Suppose that \( \mu_n \) and \( \mu \) are probability measures on \( (\mathbb{R}^1, \mathcal{B}^1) \) and \( \mu_n \Rightarrow \mu \). There exist random variables \( Y_n \) and \( Y \) on a common probability space \( (\Omega, \mathcal{F}, P) \) such that \( Y_n \) has distribution \( \mu_n \), \( Y \) has distribution \( \mu \), and \( Y_n(\omega) \to Y(\omega) \) for each \( \omega \).
Theorem 2.1.7.6 Suppose that \( h: R^1 \rightarrow R^1 \) is measurable and that the set \( D_h \) of its discontinuities is measurable. If \( \mu_n \Rightarrow \mu \) and \( \mu(D_h) = 0 \), then \( \mu_n h^{-1} \Rightarrow \mu h^{-1} \). \( \mu h^{-1} \) has value \( \mu(h^{-1}A) \) at \( A \).

Corollary 1. If \( X_n \Rightarrow X \) and \( P[X \in D_h] = 0 \), then \( h(X_n) \Rightarrow h(X) \).

Take \( X \equiv a \):

Corollary 2. If \( X_n \Rightarrow a \) and \( h \) is continuous at \( a \), then \( h(X_n) \Rightarrow h(a) \).

Theorem 2.1.7.7 The following three conditions are equivalent.

(i) \( \mu_n \Rightarrow \mu \);
(ii) \( \int f \, d\mu_n \rightarrow \int f \, d\mu \) for every bounded, continuous real function \( f \);
(iii) \( \mu_n(A) \rightarrow \mu(A) \) for every \( \mu \)-continuity set \( A \).

Helly’s Theorem
Some frequently used results in analysis is the Helly theorem:

Theorem 2.1.7.8 For every sequence \( \{F_n\} \) of distribution functions there exists a subsequence \( \{F_{n_k}\} \) and a nondecreasing, right-continuous function \( F \) such that \( \lim_{k} F_{n_k}(x) = F(x) \) at continuity points \( x \) of \( F \).

Theorem 2.1.7.9 Tightness is a necessary and sufficient condition that for every subsequence \( \{\mu_{n_k}\} \) there exist a further subsequence \( \{\mu_{n_k(j)}\} \) and a probability measure \( \mu \) such that \( \mu_{n_k(j)} \Rightarrow \mu, j \rightarrow \infty \).

Corollary. If \( \{\mu_n\} \) is a tight sequence of probability measures, and if each subsequence that converges weakly at all converges weakly to the probability measure \( \mu \), then \( \mu_n \Rightarrow \mu \).

Integration to the Limit
Theorem 2.1.7.10 If \( X_n \Rightarrow X \), then \( E[|X|] \leq \liminf_{n} E[|X|] \).

Theorem 2.1.7.11 If \( X_n \Rightarrow X \) and the \( X_n \) are uniformly integrable, then \( X \) is integrable and \( E[X_n] \rightarrow E[X] \).

Corollary. Let \( r \) be a positive integer. If \( X_n \Rightarrow X \) and \( \sup_{n} E[|X|^r + \varepsilon] < \infty \), where \( \varepsilon > 0 \), then \( E[|X|^r] < \infty \) and \( E[X_n^r] \rightarrow E[X^r] \).
Lebesgue's Convergence Theorem

Theorem 2.1.7.12 (Dominated Convergence Theorem)
Suppose \( f_n : \mathbb{R} \to [-\infty, \infty] \) are (Lebesgue) measurable functions such that the pointwise \( \lim f(x) = \lim_{n \to \infty} f_n(x) \) exists. Assume there is an integrable \( g : \mathbb{R} \to [0, \infty] \) with \( |f_n(x)| \leq g(x) \) for each \( x \in \mathbb{R} \). Then \( f \) is integrable as is \( f_n \) for each \( n \), and
\[
\lim_{n \to \infty} \int f_n \, d\mu = \int f \, d\mu. \tag{2.29}
\]

2.1.8 Uniform Integrability

If \( f \) is integrable, then \( |f| I_{\{|f| \geq \alpha\}} \) goes to 0 almost everywhere as \( \alpha \to \infty \) and is dominated by \( |f| \), and hence
\[
\lim_{\alpha \to \infty} \int |f| I_{\{|f| \geq \alpha\}} \, d\mu = 0. \tag{2.30}
\]
A sequence \( \{f_n\} \) is uniformly integrable if (2.21) holds uniformly in \( n \):
\[
\lim_{\alpha \to \infty} \sup_n \int |f| I_{\{|f| \geq \alpha\}} \, d\mu = 0. \tag{2.31}
\]
If (2.30) holds and \( \mu(\Omega) < \infty \), and if \( \alpha \) is large enough that the supremum in (2.31) is less than 1, then
\[
\int |f_n| \, d\mu \leq \alpha \mu(\Omega) + 1, \tag{2.32}
\]
and hence then \( f_n \) are integrable. On the other hand, (2.30) always holds if the \( f_n \) are uniformly bounded, but the \( f_n \) need not be in that case be integrable if \( \mu(\Omega) = \infty \). For this reason the concept of uniform integrability is interesting only for \( \mu \) finite.

If \( h \) is the maximum of \(|f|\) and \(|g|\), then
\[
\int |f_n + g| I_{\{|f| \geq 2\alpha\}} \, d\mu \leq 2 \int_{h \geq 2\alpha} h \, d\mu + 2(\int_{|g| \leq \alpha} |g| \, d\mu + \int_{|f| \leq \alpha} |f| \, d\mu).
\]
Therefore, if \( \{f_n\} \) and \( \{g_n\} \) are uniformly integrable, so is \( \{f_n + g_n\} \).

Theorem 2.1.8.1 Suppose that \( \mu(\Omega) < \infty \) and \( f_n \to f \) almost everywhere.
(i) If the \( f_n \) are uniformly integrable, then \( f \) is integrable and \( \int f_n \, d\mu \to \int f \, d\mu \).
(ii) If \( f \) and the \( f_n \) are nonnegative and integrable, then (2.32) implies that the \( f_n \) are uniformly integrable.

Corollary. Suppose that \( \mu(\Omega) < \infty \). If \( f \) and \( f_n \) are integrable, and if \( f_n \to f \) almost everywhere, then these conditions are equivalent:
(i) \( f_n \) are uniformly integrable;
(ii) \( \int |f - f_n| d\mu \to 0; \)
(iii) \( \int |f_n| d\mu \to \int |f| d\mu. \)

2.1.9 The Law of Large Numbers

In probability theory, the law of large numbers (LLN) is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

Two different versions of the Law of Large Numbers are described below; they are called the Strong Law of Large Numbers, and the Weak Law of Large Numbers. Both versions of the law state that the sample average \( \bar{X}_n \) and \( \bar{X}_n = (X_1 + \cdots + X_n)/n \), converges to the expected value \( \bar{X}_n \to \mu \) for \( n \to \infty \), where \( X_1, X_2, \ldots \) is an infinite sequence of i.i.d. integrable random variables with expected value \( E(X_1) = E(X_2) = \ldots = \mu \). Integrability means that \( E(|X_j|) < \infty \) for \( j=1,2,\ldots \). An assumption of finite variance \( Var(X_1) = Var(X_2) = \cdots = \sigma^2 < \infty \) is not necessary. Large or infinite variance will make the convergence slower, but the LLN holds anyway. This assumption is often used because it makes the proofs easier and shorter.

The difference between the strong and the weak version is concerned with the mode of convergence being asserted. For interpretation of these modes, see Convergence of random variables.

2.1.9.1 Weak Law of Large Numbers

The weak law of large numbers (also called Khintchine's law) states that the sample average converges in probability towards the expected value
\[
\bar{X}_n \xrightarrow{p} \mu, \text{ when } n \to \infty.
\]
That is to say that for any positive number \( \varepsilon \),
\[
\lim_{n \to \infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0. \tag{2.33}
\]
Interpreting this result, the weak law essentially states that for any nonzero margin specified, no matter how small, with a sufficiently large sample there will be a very high probability that the average of the observations will be close to the expected value; that is, within the margin.
Convergence in probability is also called weak convergence of random variables. This version is called the weak law because random variables may converge weakly (in probability) as above without converging strongly (almost surely) as below.

### 2.1.9.2 Strong Law of Larger Numbers

The strong law of large numbers states that the sample average converges almost surely to the expected value

\[ \bar{X}_n \xrightarrow{a.s.} \mu, \text{ when } n \to \infty. \]

That is,

\[ P(\lim_{n \to \infty} \bar{X}_n = u) = 1. \]  \hspace{1cm} (2.34)

The strong law of large numbers can itself be seen as a special case of the pointwise ergodic theorem. Moreover, if the summands are independent but not identically distributed, then

\[ \bar{X}_n - E[\bar{X}_n] \xrightarrow{a.s.} 0 \]

provided that each \( X_k \) has a finite second moment and

\[ \sum_{k=1}^{\infty} \frac{1}{k^2} \text{Var}[X_k] < \infty. \]  \hspace{1cm} (2.35)

This statement is known as Kolmogorov's strong law.

### 2.1.9.3 Differences between the Weak Law and the Strong Law

The weak law states that for a specified large \( n \), the average \( \bar{X}_n \) is likely to be near \( \mu \). Thus, it leaves open the possibility that \( |\bar{X}_n - \mu| > \varepsilon \) happens an infinite number of times, although at infrequent intervals. The strong law shows that this almost surely will not occur. In particular, it implies that with probability 1, we have that for any \( \varepsilon > 0 \) the inequality \( |\bar{X}_n - \mu| < \varepsilon \) holds for all large enough \( n \).

### 2.1.9.4 Uniform Law of Large Numbers

Suppose \( f(x,0) \) is some function defined for \( 0 \in \Theta \), and continuous in \( 0 \). Then for any fixed \( 0 \), the sequence \{\( f(X_1,0), f(X_2,0), \ldots \)\} will be a sequence of independent and identically distributed random variables, such that the sample mean of this sequence converges in probability to \( E[f(X,0)] \). This is the pointwise (in \( 0 \)) convergence.
The uniform law of large numbers states the conditions under which the convergence happens uniformly in $\theta$. If\[10][11]

1. $\Theta$ is compact,
2. $f(x, \theta)$ is continuous at each $\theta \in \Theta$ for almost all $x$’s, and measurable function of $x$ at each $\theta$.
3. There exists a dominating function $d(x)$ such that $E[d(X)] < \infty$, and

\[\left| f(x, \theta) \right| \leq dx \quad \text{for all } \theta \in \Theta.\]

Then $E[f(X, \theta)]$ is continuous in $\theta$, and

\[
\sup \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i, \theta) - E[f(X, \theta)] \right| \xrightarrow{a.s.} 0.
\] (2.36)

### 2.1.9.5 Borel's Law of Large Numbers

Borel's law of large numbers, named after Émile Borel, states that if an experiment is repeated a large number of times, independently under identical conditions, then the proportion of times that any specified event occurs approximately equals the probability of the event's occurrence on any particular trial; the larger the number of repetitions, the better the approximation tends to be. More precisely, if $E$ denotes the event in question, $p$ its probability of occurrence, and $N_n(E)$ the number of times $E$ occurs in the first $n$ trials, then with probability one,

\[
\frac{N_n(E)}{n} \rightarrow P \text{ as } n \rightarrow \infty.
\] (2.37)

**Chebyshev's Lemma.**

*Let $X$ be a random variable with finite expected value $\mu$ and finite non-zero variance $\sigma^2$. Then for any real number $k > 0$,

\[
P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.
\] (2.38)*

This theorem makes rigorous the intuitive notion of probability as the long-run relative frequency of an event's occurrence. It is a special case of any of several more general laws of large numbers in probability theory.

### 2.1.10 The Central Limit Theorem

The central limit theorem (CLT) says roughly that the sum of many independent random
variables will be approximately normally distributed if each summand has high probability of being small.

**Theorem 2.1.10.1 (Lindeberg-Lévy’s Theorem)**
Suppose that \( \{X_n\} \) is an independent sequence of random variables having the same distribution with mean \( c \) and finite positive variance \( \sigma^2 \). If \( S_n = X_1 + \cdots + X_n \), then
\[
\frac{S_n - nc}{\sigma \sqrt{n}} \Rightarrow N.
\] (2.39)

**Lemma 1.** Let \( z, \cdots, z_m \) and \( w_1, \cdots, w_m \) be complex numbers of modulus at most 1; then
\[
|z_1 \cdots z_m - w_1 \cdots w_m| \leq \sum_{k=1}^m |z_k - w_k|.
\] (2.40)

**Theorem 2.1.10.2 (Lindeberg and Lyapounov Theorems)**
Suppose that for each \( n \) the sequence \( X_{n1}, \cdots, X_{nr_n} \) is independent and satisfies
\[
E[X_{nk}] = 0, \sigma^2_{nk} = E[X_{nk}^2], S_n^2 = \sum_{k=1}^{r_n} \sigma^2_{nk}.
\] (2.41)

If the assumption of Lindeberg condition:
\[
\lim_{n \to \infty} \sum_{k=1}^{r_n} \frac{1}{\sigma^2_{nk}} \int_{|X_{nk}| \geq \epsilon \sqrt{n}} X_{nk}^2 dP = 0
\] (2.42)
holds for all positive \( \epsilon \), then
\[
\frac{S_n}{s_n} \Rightarrow N.
\]
This theorem contains the preceding one: Suppose that \( X_{nk} = X_k \) and the same distribution with mean 0 and variance \( \sigma^2 \). Then (2.35) reduces to
\[
\lim_{n \to \infty} \frac{1}{\sigma^2} \int_{|X_1| \geq \epsilon \sqrt{n}} X_1^2 dP = 0,
\] (2.43)
which holds because \( ||X_1| \geq \epsilon \sqrt{n} \downarrow \emptyset \) as \( n \uparrow \infty \).

### 2.2 Martingales

#### 2.2.1 Definition and Main Properties

**Definition 2.2.1.1** Let \( X_1, X_2, \cdots \) be a sequence of random variables on a probability space \( (\Omega, \mathcal{F}, P) \), and let \( \mathcal{F}_1, \mathcal{F}_2, \cdots \) be a sequence of \( \sigma \)-fields in \( \mathcal{F} \). The sequence \( \{(X_n, \mathcal{F}_n) : n = 1, 2, \cdots \} \) is a martingale if these four conditions hold:
(i) \( \mathcal{F}_n \subset \mathcal{F}_{n+1} \);
(ii) \( X_n \) is measurable \( \mathcal{F}_n \);
(iii) \( E[|X_n|] < \infty \);
(iv) with probability 1,

\[
E[X_{n+1}|\mathcal{F}_n] = X_n
\]  
(2.44)

Alternatively, the sequence \( X_1, X_2, \cdots \) is said to be a martingale relative to the \( \sigma \)-fields in \( \mathcal{F}_1, \mathcal{F}_2, \cdots \). Condition (i) is expressed by saying the \( \mathcal{F}_n \) form a filtration and condition (ii) by saying the \( X_n \) are adapted to the filtration.

The sequence \( X_1, X_2, \cdots \) is defined to be a martingale if it is a martingale relative to some sequence \( \mathcal{F}_1, \mathcal{F}_2, \cdots \). In this case, the \( \sigma \)-fields \( \mathcal{G}_n = \sigma(X_1, \cdots, X_n) \) always work: Obviously, \( \mathcal{G}_n \subset \mathcal{G}_{n+1} \) and \( X_n \) is measurable \( \mathcal{G}_n \), and if (2.44) holds, then

\[
E[X_{n+1}|\mathcal{G}_n] = E[E[X_{n+1}|\mathcal{F}]|\mathcal{G}_n] = E[X_n|\mathcal{G}_n] = X_n
\]

by chapter 2.1.5. For these special \( \sigma \)-fields \( \mathcal{G}_n \), (2.44) reduces to

\[
E[X_{n+1}|\mathcal{G}_1, \cdots, X_n] = X_n
\]  
(2.45)

Since \( \sigma(X_1, \cdots, X_n) \subset \mathcal{F}_n \) if and only if \( X_n \) is measurable \( \mathcal{F}_n \) for each \( n \), the \( \sigma(X_1, \cdots, X_n) \) are the smallest \( \sigma \)-fields with respect to which the \( X_n \) are a martingale.

The essential condition is embodied in (2.44) and in its specialization (2.45). Condition (iii) is of course needed to ensure that \( E[X_{n+1}|\mathcal{F}_n] \) exists. Conditional (iv) says that \( X_n \) is version of \( E[X_{n+1}|\mathcal{F}_n] \); since \( X_n \) is

\[
\int_A X_{n+1}dP = \int_A X_n dP \quad A \in \mathcal{F}_n
\]  
(2.46)

Since \( \mathcal{F}_t \) are nested, \( A \in \mathcal{F}_t \) implies that

\[
\int_A X_{n+1}dP = \int_A X_{n+1}dP = \cdots = \int_A X_{n+k}dP
\]

Therefore, \( X_n \), being measurable \( \mathcal{F}_n \), is a version of

\[
E[X_{n+k}|\mathcal{F}_n] = X_n
\]  
(2.47)

with probability 1 for \( k \geq 1 \). Note that for \( A = \Omega \), (2.46) gives

\[
E[X_1] = E[X_2] = \cdots
\]  
(2.48)

The defining conditions for a martingale can also be given in terms of the differences

\[
\Delta = X_n - X_{n-1}
\]  
(2.49)

(\( \Delta_1 = X_1 \)). By linearity, (2.44) is the same thing as

\[
E[\Delta_{n+1}|\mathcal{F}_n] = E[(X_n - X_{n-1})|\mathcal{F}_n] = 0
\]  
(2.50)

Since \( X_k = \Delta_1 + \cdots + \Delta_k \) and \( \Delta_k = X_k - X_{k-1} \), the sets \( X_1, \cdots, X_n \) and \( \Delta_1, \cdots, \Delta_n \) generate the same \( \sigma \)-field:

\[
\sigma(X_1, \cdots, X_n) = \sigma(\Delta_1, \cdots, \Delta_n)
\]  
(2.51)
**Definition 2.2.1.2** A process $X$ is of the class (D), if the family of random variables $(X_t, \tau \text{ finite stopping time})$ is uniformly integrable.

### 2.2.2 Submartingales and Supermartingales

Random variables $X_n$ are a submartingale (resp. a supermartingale) relative to $\sigma$-fields $\mathcal{F}_n$ if (i), (ii), and (iii) of the definition above hold and if this condition holds in place of (iv):

(iv') with probability 1,

\[
E[X_{n+1}|\mathcal{F}_n] \geq X_n \quad (2.52)
\]

\[
E[X_{n+1}|\mathcal{F}_n] \leq X_n. \quad (2.52')
\]

As before, the $X_n$ are a submartingale (resp. a supermartingale) if they are a submartingale (resp. a supermartingale) with respect to some sequence $\mathcal{F}_n$, and the special sequence $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ works if any dos. The requirement (2.46) is the same thing as

\[
\int_A X_{n+1} dP \geq \int_A X_n dP \quad A \in \mathcal{F}_n \quad (2.53)
\]

\[
\text{(resp. } \int_A X_{n+1} dP \leq \int_A X_n dP \quad A \in \mathcal{F}_n). \quad (2.53')
\]

This extends inductively (basic on (2.47)), and so

\[
E[X_{n+k}|\mathcal{F}_n] \geq X_n \quad (2.54)
\]

\[
\text{(resp. } E[X_{n+k}|\mathcal{F}_n] \leq X_n) \quad (2.54')
\]

for $k \geq 1$. Taking expected values in (2.52) gives

\[
E[X_1] \leq E[X_2] \leq \cdots. \quad (2.55)
\]

\[
\text{(resp. } E[X_1] \geq E[X_2] \geq \cdots). \quad (2.55')
\]

**Theorem 2.2.2.1 (Doob-Meyer Decomposition Theorem)**

The process $(X_t; t \geq 0)$ is a sub-martingale (resp. a super-martingale) of class (D) if and only if $X_t = M_t + A_t$ (resp. $X_t = M_t - A_t$) where $M$ is a uniformly integrable martingale and $A$ is an increasing predictable process with $E(A_\infty) < \infty$.

**Theorem 2.2.2.2** If $X_1, \cdots, X_n$ is a sub-martingale (resp. a supermartingales), then for $a>0$,

\[
P[\max_{i \leq n} X_i \geq a] \leq \frac{1}{a} E[|X_n|]. \quad (2.56)
\]
\[
P[\max_{i\leq n} X_i \geq a] \leq \frac{1}{a} E[|X_n|].
\] (2.56')

2.2.3 Space of Martingales

We denote by \( H^2 \) (resp. \( H^2[0,T] \)) the subset of square integrable martingales (resp. define on \([0,T]\)), i.e. martingales such that \( \sup_t E[M_t^2] < \infty \) (resp. \( \sup_{t\leq T} E[M_t^2] < \infty \)). From Jensen’s inequality, if \( M \) is a square integrable martingale, \( M^2 \) is a sub-martingale. It follows that the martingale \( M \) is square integrable on \([0,T]\) if and only if \( E[M_T^2] < \infty \).

If \( M \in H^2 \), the process \( M \) is uniformly integrable and \( M_t = E[M_\infty |\mathcal{F}_t] \). From Fatou’s lemma, the random variable \( M_\infty \) is square integrable and
\[
E[M_\infty^2] = \lim_{t\to\infty} E[M_t^2] = \sup_t E[M_t^2].
\] (2.57)

From \( M_t^2 \leq E[M_\infty^2 |\mathcal{F}_t] \), it follows that \( (M_t^2, t \geq 0) \) is uniformly integrable.

Doob’s inequality states that, if \( M \in H^2 \), then \( E[\sup_t |M_t^2|] \leq 4E[M_\infty^2] \). Hence, \( E[\sup_t M_t^2] < \infty \) is equivalent to \( \sup_t E[M_t^2] < \infty \). More generally, if \( M \) is a martingale or a positive sub-martingale, and \( p>1 \),
\[
\|\sup_{t\leq T}|M_t||\|_p \leq \frac{p}{p-1}\sup_{t\leq T} \||M_t||\|_p.
\] (2.58)

Obviously, the Brownian motion does not belong to \( H^2 \), however, it belongs to \( H^2([0,T]) \) for any \( T \).

We denote by \( H^1 \) the set of martingales \( M \) such that \( E[\sup_t |M_t|] < \infty \). More generally, the space of martingales such that \( M^* = \sup_t |M_t| \) is in \( L^p \) is denoted by \( H^p \). For \( p > 1 \), one has the equivalence
\[
M^* \in L^p \iff M_\infty \in L^p.
\]

Thus the space \( H^p \) for \( p > 1 \) may be identified with \( L^p(\mathcal{F}_\infty) \). Note that \( \sup_t E[|M_t|] \leq E[\sup_t |M_t|] \), hence any element of \( H^1 \) is \( L^1 \) bounded, but the converse if not true.

2.2.4 Stopping Times

Let \( \tau \) be random variable taking as values positive integers or the special value \( \infty \). It is a stopping time with respect to \( \{\mathcal{F}_n\} \) if \( [\tau = k] \in \mathcal{F}_k \) for each finite \( k \), or, equivalently, if \( [\tau \leq k] \in \mathcal{F}_k \) for each finite \( k \).

**Definitions 2.2.4.1**

\[
\mathcal{F}_\tau = \{ A \in \mathcal{F}_\infty : A \cap [\tau \leq k] \in \mathcal{F}_k, 1 \leq k \leq \infty \}.
\] (2.59)

21
This is a σ-field, and the definition is unchanged if \([\tau \leq k]\) replaced by \([\tau = k]\) on the right. Since clearly \([\tau = j]\) ∈ \(\mathcal{F}_\tau\) for finite \(j\), \(\tau\) is measurable \(\mathcal{F}_\tau\).

Suppose now that \(\tau_1\) and \(\tau_2\) are two stopping times and \(\tau_1 \leq \tau_2\). If \(A \in \mathcal{F}_{\tau_1}\), then \(A \cap [\tau_1 \leq k] \in \mathcal{F}_k\) and hence \(A \cap [\tau_2 \leq k] = A \cap [\tau_1 \leq k] \cap [\tau_2 \leq k] \in \mathcal{F}_k; \mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}\).

**Theorem 2.2.4.1** If \(X_1, \ldots, X_n\) is a sub-martingale with respect to \(\mathcal{F}_1, \ldots, \mathcal{F}_n\) and \(\tau_1, \tau_2\) are stopping times satisfying \(1 \leq \tau_1 \leq \tau_2 \leq n\), then \(X_{\tau_1}, X_{\tau_2}\) is a sub-martingale with respect to \(\mathcal{F}_{\tau_1}, \mathcal{F}_{\tau_2}\).

**Definition 2.2.4.2** A stopping time \(\tau\) is predictable if there exists an increasing sequence \((\tau_n)\) of stopping times such that almost surely

1. \(\lim_n \tau_n = \tau\),
2. \(\tau_n < \tau\) for every \(n\) on the set \(\{\tau > 0\}\).

A stopping time \(\tau\) is totally inaccessible if \(P(\tau = 0 < \infty) = 0\) for any predictable stopping time \(\theta\) (or, equivalently, if for any increasing sequence of stopping time \((\tau_n, n \geq 0)\), \(P\left(\lim_n \tau_n = \tau \cap A\right) = 0\) where \(A = \cap_n \{\tau_n < \tau\}\)).

If \(X\) is an \(\mathcal{F}\)-adapted process and \(\tau\) a stopping time, the \((\mathcal{F}\)-adapted) process \(X^\tau\) where \(X^\tau := X_{\tau+\tau}\) is called the process \(X\) stopped at \(\tau\).

**Theorem 2.2.4.2** (Doob’s Optional Sampling Theorem)
If \(M\) is a uniformly integrable martingale and \(\theta, \tau\) are two stopping times with \(\theta \leq \tau\), then

\[
M_\theta = E[M_\tau | \mathcal{F}_\theta] = E[M_\infty | \mathcal{F}_\theta], \text{ a.s.} \tag{2.60}
\]

If \(M\) is a positive super-martingale and \(\theta, \tau\) a pair of stopping times with \(\theta \leq \tau\), then

\[
E[M_\tau | \mathcal{F}_\theta] \leq M_\theta. \tag{2.61}
\]

2.2.5 Local Martingales

**Definition 2.2.5.1** An adapted, right-continuous process \(M\) is an \(\mathcal{F}\)-local martingale if there exists a sequence of stopping times \((\tau_n)\) such that:

- The sequence \(\tau_n\) is increasing and \(\lim_n \tau_n = \infty\), a.s.
- For every \(n\), the stopped process \(M^{\tau_n} I_{[\tau_n > 0]}\) is an \(\mathcal{F}\)-martingale.

A sequence of stopping times such that the two previous conditions hold is called a
localizing or reducing sequence. If \( M \) is a local martingale, it is always possible to choose the localizing sequence \( \left\{ \bigcap g_{n20} \cap g_{n30} \cap g_{n48} \cap \bigcap g_{n86} \cap g_{n34} \cap g_{n88} \cap g_{n46} \cap g_{n30} \cap g_{n32} \cap \bigcap g_{n83} \right\} \) such that each martingale \( \left\{ X_t \right\} \) is uniformly integrable.

Doob-Meyer decomposition can be extended to general sub-martingales:

**Proposition 2.2.5.1** A process \( X \) is a sub-martingale (resp. a super-martingale) if and only if \( X_t = M_t + A_t \) (resp. \( X_t = M_t - A_t \)) where \( M \) is a local martingale and \( A \) an increasing predictable process.

**Proposition 2.2.5.2** A continuous local martingale of local finite variation is a constant.

### 2.2.6 Martingale Convergence Theorem

**Theorem 2.2.6.2** Let \( X_1, X_2, \cdots \) be a submartingale. If \( K = \sup_n E[|X_n|] < \infty \), then \( X_n \to X \) with probability 1, where \( X \) is a random variable satisfying \( E[|X|] \leq K \).

**Lemma.** If \( Z \) is integrable and \( \mathcal{F}_n \) are arbitrary \( \sigma \)-fields, then the random variable \( E[Z|\mathcal{F}_n] \) are uniformly integrable.

**Theorem 2.2.6.1** If \( \mathcal{F}_n \uparrow \mathcal{F}_\infty \) and \( Z \) is integrable, then

\[
E[Z|\mathcal{F}_n] \to E[Z|\mathcal{F}_\infty]
\]

with probability 1.

### 2.2.7 Martingale Central Limit Theorem

Suppose \( X_1, X_2, \cdots \) is a martingale relative to \( \mathcal{F}_1, \mathcal{F}_2, \cdots \), and put \( Y_n = X_n - X_{n-1} \) (\( Y_1 = X_1 \)), so that

\[
E[Y_n|\mathcal{F}_{n-1}] = 0. \tag{2.62}
\]

Suppose for simplicity the \( Y_n \) are bounded, and define

\[
\sigma^2_n = E[Y^2_n|\mathcal{F}_{n-1}] \tag{2.63}
\]

(take \( \mathcal{F}_0 = \{ \emptyset, \Omega \} \)). Consider the stopping times

\[
\theta_t = \min \{ n: \sum_{k=1}^n \sigma^2_k \geq t \}. \tag{2.64}
\]

**Theorem 2.2.7.1** Suppose the \( Y_n = X_n - X_{n-1} \) are uniformly bounded and satisfy (2.62), an assume that \( \sum \sigma^2_n = \infty \) with probability 1. Then \( X_{\theta_t}/\sqrt{t} \to N \). (Here \( N \) is the normal distribution)
Theorem 2.2.7.2 Suppose that
\[
\sum_{k=1}^{\infty} \sigma_{nk}^2 \rightarrow_p \sigma^2,
\]
where \( \sigma \) is a positive constant, and that
\[
\sum_{k=1}^{\infty} E[Y_{nk}^2I_{\{|Y_{nk}|>\epsilon\}}] \rightarrow 0
\]
for each \( \epsilon \). Then \( \sum_{k=1}^{\infty} Y_{nk} \Rightarrow \sigma N \).

2.3 Markov Process

Let \((\Omega, \mathcal{F})\) be a measure space of elementary outcomes, let \(F = \{\mathcal{F}_t\}, t \in Z\), be a nondecreasing family of sub-\(\sigma\)-algebras \(\mathcal{F}\), and let \((E, \mathcal{B})\) be a state space. Further, let \((X_t, \mathcal{F}_t)\), \(t \in Z\), be a random process with values in \((E, \mathcal{B})\) and let probability measure \(P_x\) be given on a \(\sigma\)-algebra \(\mathcal{F}\) for each \(x \in E\).

Definition 2.3.1 The system \(X = (x_t, \mathcal{F}^s_t, P_{x,s}), s, t \in Z, x \in E_\Delta\), is said to be a (nonhomogeneous, terminating) Markov process with state space \((E, \mathcal{B})\) with an adjoined point \(\Delta\) if the following conditions are satisfied:

1. For each \(A \in \mathcal{F}^s\) and \(s \in Z\), \(P_{x,s}(A)\) is a \(\mathcal{B}\)-measurable function for \(x\);
2. For all \(x \in E_\Delta, B \in \mathcal{B}, 0 \leq s \leq t \leq u, \)
\[
P_{x,s}(x_u \in B|\mathcal{F}^s_t) = P_{t,x_t}(x_u \in B) \quad (P_{s,x} - a.s.) ;
\]
3. \(P_{s,x}(x_s = x) = 1, P_{s,x}(x_t \in \Delta) = 1, x \in E_\Delta, s \leq t; \)

The function \(P(s, x; t, B) = P_{s,x}(x_t \in B)\) is said to be the transition function of the Markov process \(X\).

Let \(X_0, X_1, X_2, \ldots\) be a sequence of random variables whose ranges are contained in \(S\). The sequence is a Markov process if
\[
P[X_{n+1} = j|X_0 = i_0, \ldots, X_n = i_n] = P[X_{n+1} = j|X_n = i] = p_{i,n,j}
\]
for every \(n\) and every sequence \(i_0, \ldots, i_n\) in \(S\) for which \(P[X_0 = i_0, \ldots, X_n = i_n] > 0\). The set \(S\) is the state space or phase space of the process, and the \(p_{ij}\) are the transition probabilities.

2.4 Brownian Motion

2.4.1 Introduction

Brownian movement is the name given to irregular movement of pollen, suspended in water, observed by the botanist Robert Brown in 1828. This random movement, now
attributed to the buffeting of the pollen by water molecules, results in dispersal or diffusion of the pollen in the water. The range of application of Brownian motion as defined here goes far beyond a study of microscopic particles in suspension and includes modeling of stock price, of thermal noise in electrical circuits, of certain limiting behavior in queuing and inventory systems, and of random perturbations in a variety of other physical, biological, economic, and management systems.

**Definition 2.4.1.1** Brownian motion is a continuous adapted process \( B = \{ B_t, \mathcal{F}_t; 0 \leq t < \infty \} \), defined on some probability space \( (\Omega, \mathcal{F}, P) \), with the properties that \( B_0 = 0 \) a.s. and for \( 0 \leq s < t \), the increment \( B_t - B_s \) is independent of \( \mathcal{F}_s \) and is normally distributed with mean zero and variance \( t-s \).

**Definition 2.4.1.2** A collection \( \mathcal{D} \) of subsets of a set \( \Omega \) is called a Dynkin system if

1. \( \Omega \in \mathcal{D} \),
2. \( A, B \in \mathcal{D} \) and \( B \subseteq A \) imply \( A \setminus B \in \mathcal{D} \),
3. \( \{ A_n \}_{n=1}^{\infty} \subseteq \mathcal{D} \) and \( A_1 \subseteq A_2 \subseteq \cdots \) imply \( \bigcup_{n=1}^{\infty} A_n \in \mathcal{D} \).

**Theorem 2.4.1.1** (Dynkin System Theorem) Let \( \mathcal{C} \) be a collection of subsets of \( \Omega \) which is closed under pairwise intersection. If \( \mathcal{D} \) is a Dynkin system containing \( \mathcal{C} \), then \( \mathcal{D} \) also contains the \( \sigma \)-field \( \sigma(\mathcal{C}) \) generated by \( \mathcal{C} \).

### 2.4.2 Some Theorems of Brownian Motion

#### 2.4.2.1 The Consistency Theorem

Let \( \mathbb{R}^{[0,\infty)} \) denote the set of all real-valued functions on \([0,\infty)\). An \( n \)-dimensional cylinder set in \( \mathbb{R}^{[0,\infty)} \) is a set of the form

\[
C = \{ \omega \in \mathbb{R}^{[0,\infty)}; (\omega(t_1), \ldots, \omega(t_n)) \in A \},
\]

where \( t_i \in [0,\infty), i = 1, \ldots, n \), and \( A \in \mathcal{B}(\mathbb{R}^n) \). Let \( \mathcal{C} \) denote the field of all cylinder sets in \( \mathbb{R}^{[0,\infty)} \), and let \( \mathcal{B}(\mathbb{R}^{[0,\infty)}) \) denote the smallest \( \sigma \)-field containing \( \mathcal{C} \).

**Definition 2.4.2.1** Let \( T \) be the set of finite sequences \( \bar{t} = (t_1, \ldots, t_n) \) of distinct, nonnegative numbers, where the length \( n \) of these sequences ranges over the set of
positive integers. Suppose that for each $\bar{t}$ of length $n$, we have a probability measure $Q_{\bar{t}}$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Then the collection $\{Q_{\bar{t}}\}_{\bar{t} \in T}$ is called a family of finite-dimensional distributions. This family is said to be consistent provided that the following two conditions are satisfied:

(I) if $\bar{s} = (t_i, \ldots, t_n)$ is a permutation of $\bar{t} = (t_1, \ldots, t_n)$ then for any $A_i \in \mathcal{B}(\mathbb{R}^n), i = 1, \ldots, n$, we have

$$Q_{\bar{t}}(A_1 \times A_2 \times \cdots \times A_n) = Q_{\bar{s}}(A_{i_1} \times A_{i_2} \times \cdots \times A_{i_n});$$

(II) if $\bar{t} = (t_1, \ldots, t_n)$ with $n \geq 1, \bar{s} = (t_1, \ldots, t_{n-1})$ and $A \in \mathcal{B}(\mathbb{R}^{n-1})$, then

$$Q_{\bar{t}}(A \times \mathbb{R}) = Q_{\bar{s}}(A).$$

If we have a probability measure $P$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, then we can define a family of finite-dimensional distributions by

$$Q_{\bar{t}}(A) = P[\omega \in \mathbb{R}^{[0,\infty)}; \omega(t_1), \ldots, \omega(t_n) \in A],$$

where $A \in \mathcal{B}(\mathbb{R}^n)$ and $\bar{t} = (t_1, \ldots, t_n) \in T$.

**Theorem 2.4.2.1 (Daniell(1918), Kolmogorov(1933)).** Let $\{Q_{\bar{t}}\}$ be a consistent family of finite-dimensional distributions. Then there is a probability measure $P$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, such that (2.69) holds for every $\bar{t} \in T$.

**Corollary** There is a probability measure $P$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, under which the coordinate mapping process

$$B_t(\omega) = \omega(t); \quad \omega \in \mathbb{R}^{[0,\infty)}, t \geq 0,$$

has stationary, independent increments. An increment $B_t - B_s$, where $0 \leq s < t$, is normally distributed with mean zero and variance $t-s$.

**2.4.2.2 The Kolmogorov-Chentsov Theorem**

**Theorem 2.4.2.2** Suppose that a process $X = \{X_t; 0 \leq t \leq T\}$ on a probability space $(\Omega, \mathcal{F}, P)$ satisfies the condition

$$E|X_t - X_s|^\alpha \leq C|t - s|^{\alpha + \beta}, 0 \leq s, t \leq T,$$

(2.70)

for some positive constants $\alpha, \beta$ and $C$. Then there exists a continuous modification $\hat{X} = \{\hat{X}_t; 0 \leq t \leq T\}$ of $X$, which is locally Holder-continuous with exponent $\gamma$ for every $\gamma \in (0, \beta/\alpha)$, i.e.,
\[
P \left[ \omega; \sup_{0 < t - s < h(\omega)} \frac{\tilde{X}_t(\omega) - \tilde{X}_s(\omega)}{|t - s|^\alpha} \leq \delta \right] = 1,
\]
where \( h(\omega) \) is an a.s. positive random variable and \( \delta > 0 \) is an appropriate constant.

**Corollary** There is a probability measure \( P \) on \((\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))\), and a stochastic process \( W = \{W, \mathcal{F}_t^W; t \geq 0\} \) on the same space, such that under \( P, W \) is a Brownian motion.

### 2.4.2.3 Dambis / Dubins-Schwarz Theorem

**Definition 2.4.2.2** The probability measure \( P \) on \((\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))\), under which the coordinate mapping process \( W_t(\omega) \equiv \omega(t), 0 \leq t < \infty \), is a standard, one-dimensional Brownian motion, is called Wiener measure.

**Theorem 2.4.2.3** The only continuous local martingale \( (M_t)_{t \geq 0} \) such that \( (M_t^2 - t, t \geq 0) \) is also a local martingale is Brownian motion.

The Dambis / Dubins-Schwarz theorem (1965) gives an important representation of a generic continuous local \((\mathcal{F}_t)\) martingale \( (M_t)_{t \geq 0} \), as a time change of a Brownian motion; precisely:

\[
M_t = B_{\langle M \rangle_t}, t \geq 0,
\]
where \( (\beta_\mu, \mu \geq 0) \) is a \((\mathcal{F}_\mu)\) Brownian motion \([\tau_\mu = \inf\{t: \langle M \rangle_t > u\}]\).

To any continuous local martingale \( (M_t)_{t \geq 0} \), one may associate its quadratic variation process \( \langle \langle M \rangle_t \rangle, t \geq 0 \) which may be defined by

\[
\langle \langle M \rangle_t \rangle = P - \lim_{n \to \infty} \sum_{t_1^{(n)} < \cdots < t_p^{(n)} \leq t} (M_{t_{i+1}^{(n)}} - M_{t_{i}^{(n)}})^2,
\]
where \( \sup_{t \geq 0} (t_{i+1}^{(n)} - t_i^{(n)}) \to 0 \).

It is not difficult to show that

\[
M_t^2 - \langle \langle M \rangle_t \rangle, t \geq 0,
\]
is a local martingale, and since \( \langle \langle M \rangle_t \rangle \) is increasing and continuous, the property (2.73) characterizes \( \langle \langle M \rangle_t \rangle \). Note that this discussion is closely related to Theorem 1.2, where in greater generality, the existence and uniqueness of \( \langle \langle M \rangle_t \rangle \) is asserted.

**Theorem 2.4.2.4** (Dambis / Dubins-Schwarz (1965)) If \( (M_t, t \geq 0) \) is a continuous local \( (\mathcal{F}_t) \) martingale such that \( \langle \langle M \rangle_{\infty} \rangle = \infty \), then the process:
\[ B_u = M_{\tau_u}, \quad \tau_u = \inf \{ t; < M >_t > u \} \]
is a \((\mathcal{F}_{\tau_u})\) martingales.

### 2.4.2.4 Other Theorems

**Theorem 2.4.2.5** If \((M_t, t \geq 0)\) is a local martingale with respect to the filtration \((B_t, t \geq 0)\) generated by an \(n\)-dimensional Brownian motion \((B_t, t \geq 0)\), it may be represented as:

\[ M_t = c + \int_0^t m_s \, dB_s, \tag{2.74} \]

where \((m_s)_{s \geq 0}\) is a \((\mathcal{B}_s)\) predictable process, taking values in \(\mathbb{R}^n\), and \(x.y\) denotes the Euclidean scalar product between \(x\) and \(y\). For (2.74) to make sense, it is necessary that \(m\) satisfies: \(\int_0^t |m_s|^2 \, ds < \infty\), for every \(t\).

### 2.4.3 The Space \(\mathcal{C}(0, \infty)\), Weak Convergence, and the Wiener Measure

Suppose \(X\) is a random variable on a probability space \((\Omega, \mathcal{F}, P)\) with values in measurable space \((S, \mathcal{B}(S))\), i.e., the function \(X: \Omega \rightarrow S\) is \(\mathcal{F}/\mathcal{B}(S)\)-measurable, then \(X\) induces a probability measure \(PX^{-1}\) on \((S, \mathcal{B}(S))\) by

\[ PX^{-1}(B) = P\{\omega \in \Omega; X(\omega) \in B\}, B \in \mathcal{B}(S). \tag{2.75} \]

An important special case of (2.75) occurs when \(\{X = \{X_t; 0 \leq t < \infty\}\}\) is a continuous stochastic process \((\Omega, \mathcal{F}, P)\). Such an \(X\) can be regarded as a random variable on \((\Omega, \mathcal{F}, P)\) with values in \((\mathcal{C}[0, \infty), \mathcal{B}(\mathcal{C}[0, \infty)))\), and \(PX^{-1}\) is called the law of \(X\).

#### 2.4.3.1 Weak Convergence

**Definition 2.4.3.1** Let \((S, \rho)\) be a metric space with Borel \(\sigma\)-field \(\mathcal{B}(S)\). Let \(\{P_n\}_{n=1}^\infty\) be a sequence of probability measures on \((S, \mathcal{B}(S))\), and let \(P\) be another measure on this space. We say that \(\{P_n\}_{n=1}^\infty\) converges weakly to \(P\) and write \(P_n \rightharpoonup P\), if and only if

\[ \lim_{n \to \infty} \int_S f(s) \, dP_n(s) = \int_S f(s) \, dP(S) \]

for every bounded, continuous real-valued function \(f\) on \(S\).

**Definition 2.4.3.2** Let \((\Omega_n, \mathcal{F}_n, P_n)_{n=1}^\infty\) be a sequence of probability spaces, and on each of them consider a random variable \(X_n\) with values in the metric space \((S, \rho)\).
Let \((\Omega, \mathcal{F}, P)\) be another probability space, on which a random variable \(X\) with values in \((S, \rho)\) is given. We say that \(\{X_n\}_{n=1}^{\infty}\) converges to \(X\) in distribution, and write \(X_n \xrightarrow{D} X\), if the sequence of measure \(\{P_nX_n^{-1}\}_{n=1}^{\infty}\) converges weakly to the measure \(PX^{-1}\).

Equivalently, \(X_n \xrightarrow{D} X\) if and only if
\[
\lim_{n \to \infty} E_n[f(X_n)] = E[f(X)]
\]
for every bounded, continuous real-valued function \(f\) on \(S\), where \(E_n\) and \(E\) denote expectations with respect to \(P_n\) and \(P\), respectively.

### 2.4.3.2 Tightness

**Definition 2.4.3.3** Let \((S, \rho)\) be metric space and let \(\Pi\) be a family of probability measures on \((S, \mathcal{B}(S))\). We say that \(\Pi\) is relatively compact if every sequence of elements of \(\Pi\) contains a weakly convergent subsequence. We say that \(\Pi\) is tight if for every \(\varepsilon > 0\), there exists a compact set \(K \subseteq S\) such that \(P(K) \geq 1 - \varepsilon\), for \(P \in \Pi\).

If \(\{X_\alpha\}_{\alpha \in A}\) is a family of random variables, each one defined on probability space \((\Omega_\alpha, \mathcal{F}_\alpha, P_\alpha)\) and taking values in \(S\). The family is relatively compact or tight if the family of induced measures \(\{P_\alpha X_\alpha^{-1}\}_{\alpha \in A}\) has appropriate property.

**Theorem 2.4.3.1** (Prohorov (1956)) Let \(\Pi\) be a family of probability measures on a complete, separable metric space \(S\). This family is relatively compact if and only if it is tight.

**Theorem 2.4.3.2** A set \(A \subseteq \mathcal{C}[0, \infty)\) has compact closure if and only if the following two conditions hold:
\[
\sup_{\omega \in A} |\omega(0)| < \infty,
\]
\[
\limsup_{\delta \downarrow 0} m^T(\omega, \delta) = 0 \quad \text{for every } T > 0.
\]

**Theorem 2.4.3.3** A sequence \(\{P_n\}_{n=1}^{\infty}\) of probability measures on \((\mathcal{C}[0, \infty), \mathcal{B}(\mathcal{C}[0, \infty)))\) is tight if and only if
\[
\limsup_{n \to \infty} P_n[\omega; |\omega(0)| > \lambda] = 0,
\]
\[
\lambda^{1/\alpha} \quad \text{for } \lambda \geq 1.
\]
\[
\lim_{\delta \to 0} \sup_{n \geq 1} P_n [\omega; m^T (\omega, \delta) > \varepsilon] = 0; \forall T > 0, \varepsilon > 0.
\]

2.4.3.3 Convergence of Finite-Dimensional Distributions

Suppose that \(X\) is a continuous process on some \((\Omega, \mathcal{F}, P)\). For each \(\omega\), the function \(t \mapsto X_t(\omega)\) is a member of \(C[0, \infty)\), which we denote by \(X(\omega)\). Since \(\mathcal{B}(C[0, \infty))\) is generated by the one-dimensional cylinder sets and \(X_t(\cdot)\) is \(\mathcal{F}\)-measurable for each fixed \(t\), the random functional \(X: \Omega \to C[0, \infty)\) is \(\mathcal{F}/\mathcal{B}(C[0, \infty))\)-measurable. Thus, if \(\{X^{(n)}\}_{n=1}^{\infty}\) is a sequence of continuous processes (with each \(X^{(n)}\) defined on a perhaps distinct probability space \((\Omega_n, \mathcal{F}_n, P_n)\), we can ask whether \(X^{(n)} \to_d X\) in the sense of Definition 2.4.3.2. We can also ask whether the finite-dimensional distributions of \(\{X^{(n)}\}_{n=1}^{\infty}\) converge to those of \(X\), i.e., whether

\[
\left( X^{(n)}_{t_1}, X^{(n)}_{t_2}, \ldots, X^{(n)}_{t_d} \right) \overset{d}{\to} \left( X_{t_1}, X_{t_2}, \ldots, X_{t_d} \right).
\]

**Theorem 2.4.3.4** Let \(\{X^{(n)}\}_{n=1}^{\infty}\) be a tight sequence of continuous processes with the property that, whenever \(0 \leq t_1 < \cdots < \infty\), then the sequence of random vectors \(\left\{ \left( X^{(n)}_{t_1}, X^{(n)}_{t_2}, \ldots, X^{(n)}_{t_d} \right) \right\}_{n=1}^{\infty}\) converges in distribution. Let \(P_n\) be the measure induced on \((C[0, \infty), \mathcal{B}(C[0, \infty)))\) by \(X^{(n)}\). Then \(\{P_n\}_{n=1}^{\infty}\) converges weakly to a measure \(P\), under which the coordinate mapping process \(W_t(\omega) \triangleq \omega(t)\) on \(C[0, \infty)\) satisfies

\[
\left( X^{(n)}_{t_1}, X^{(n)}_{t_2}, \ldots, X^{(n)}_{t_d} \right) \overset{d}{\to} \left( W_{t_1}, W_{t_2}, \ldots, W_{t_d} \right), \quad 0 \leq t_1 < \cdots < \infty, d \geq 1.
\]

2.4.3.4 The Invariance Principle and the Wiener Measure

Suppose a sequence \(\{\varepsilon_j\}_{j=1}^{\infty}\) of independent, identically distributed random variables with mean zero and variance \(\sigma^2\), \(0 < \sigma^2 < \infty\), as well as the sequence of partial sums \(S_0 = 0, S_k = \sum_{j=1}^{k} \varepsilon_j, k \geq 1\). A continuous-time process \(Y = \{Y_t; t \geq 0\}\) can be obtained from the sequence \(\{S_k\}_{k=0}^{\infty}\) by linear interpolation; i.e.,

\[
Y_t = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor) \varepsilon_{\lfloor t \rfloor + 1}, t > 0,
\]

where \(\lfloor t \rfloor\) denotes the greatest integer less than or equal to \(t\). Scaling appropriately both time and space, we obtain from \(Y\) a sequence of processes \(\{X^{(n)}\}\):
\[ X_t^{(n)} = \frac{1}{\sigma \sqrt{n}} Y_{nt}, \quad t \geq 0. \] (2.77)

**Theorem 2.4.3.5** With a sequence of processes \( \{X^{(n)}\} \) defined by (2.77) and \( 0 \leq t_1 < \cdots < t_d < \infty \), we have

\[
\left( X_{t_1}^{(n)}, \ldots, X_{t_d}^{(n)} \right) \xrightarrow{d} \left( B_{t_1}, \ldots, B_{t_d} \right) \quad \text{as } n \to \infty,
\]

where \( \{B_t, \mathcal{F}_t^B; t \geq 0\} \) is a standard, one-dimensional Brownian motion.

**Lemma 1** Set \( S_k = \sum_{j=1}^k \varepsilon_j \), where \( \{\varepsilon_j\}_{j=1}^{\infty} \) of independent, identically distributed random variables with mean zero and finite variance \( \sigma^2 > 0 \). Then, for any \( \varepsilon > 0 \),

\[
\lim_{\delta \downarrow 0} \lim_{n \to \infty} \frac{1}{\delta} P \left[ \max_{1 \leq i \leq n} |S_i| > \varepsilon \sigma \sqrt{n} \right] = 0.
\]

**Lemma 2** Under the assumptions of Lemma 1, we have for any \( T > 0 \),

\[
\lim_{\delta \downarrow 0} \lim_{n \to \infty} \frac{1}{\delta} P \left[ \max_{1 \leq i \leq n} |S_{i+k} - S_k| > \varepsilon \sigma \sqrt{n} \right] = 0.
\]

If \( B \) is a Brownian motion and \( 0 \leq t_1 < \cdots < t_n < \infty \), then the increments \( \{B_{t_j} - B_{t_{j-1}}\}_{j=1}^n \) are independent and the distribution of \( B_{t_j} - B_{t_{j-1}} \) depends on \( t_j \) and \( t_{j-1} \) only through the difference \( t_j - t_{j-1} \); i.e., it is normal with mean zero and variance \( t_j - t_{j-1} \). We say that the process \( B \) has stationary, independent increments. It is easily verified that \( B \) is a square integrable martingale and

\[
<B>_t = t, \quad t \geq 0.
\] (2.78)

**Theorem 2.4.3.6 (The Invariance Principle of Donsker(1951))** Let \( (\Omega, \mathcal{F}, P) \) be a probability space on which is given a sequence \( \{\varepsilon_j\}_{j=1}^{\infty} \) of independent, identically distributed random variables with mean zero and finite variance \( \sigma^2 > 0 \). Define \( X^{(n)} = \{X^{(n)}; t \geq 0\} \) by (2.70), and let \( P_n \) be the measure induced by \( X^{(n)} \) on \( (C[0, \infty), \mathcal{B}(C[0, \infty])) \). Then \( \{P_n\}_{n=1}^{\infty} \) converges weakly to a measure \( P_* \), under which the coordinate mapping process \( W_t(\omega) \triangleq \omega(t) \) on \( C[0, \infty) \) is a standard, one-dimensional Brownian motion.

**Definition 2.4.3.4** The probability measure \( P \) on \( (C[0, \infty), \mathcal{B}(C[0, \infty])) \), under which the coordinate mapping process \( W_t(\omega) \triangleq \omega(t), 0 \leq t < \infty, \) is standard,
one-dimensional Brownian motion, is called Wiener measure.

2.4.4 The Markov Property

2.4.4.1 Markov Processes and Markov Families

Definition 2.4.4.1 Let $d$ be a positive integer and $\mu$ a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. An adapted, $d$-dimensional process $X = \{X_t, \mathcal{F}_t; t \geq 0\}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a Markov process with initial distribution $\mu$ if

(i) $\mathbb{P}[X_0 \in \Gamma] = \mu(\Gamma), \forall \Gamma \in \mathcal{B}(\mathbb{R}^d)$;

(ii) for $s, t \geq 0$ and $\Gamma \in \mathcal{B}(\mathbb{R}^d)$,

$$
\mathbb{P}^\mu[X_{t+s} \in \Gamma | \mathcal{F}_t] = \mathbb{P}^\mu[X_{t+s} \in \Gamma | X_t], \quad p^\mu - a.s.
$$

Definition 2.4.4.2 Let $d$ be a positive integer. A $d$-dimensional Markov family is an adapted process $X = \{X_t, \mathcal{F}_t; t \geq 0\}$ on some $(\Omega, \mathcal{F})$, together with a family of probability measures $\{\mathbb{P}^x\}_{x \in \mathbb{R}^d}$ on $(\Omega, \mathcal{F})$, such that

(i) for each $F \in \mathcal{F}$, the mapping $x \mapsto \mathbb{P}^x(F)$ is universally measurable;

(ii) $\mathbb{P}^x[X_0 = x] = 1, \forall x \in \mathbb{R}^d$;

(iii) for $x \in \mathbb{R}^d, s, t \geq 0$ and $\Gamma \in \mathcal{B}(\mathbb{R}^d)$

$$
\mathbb{P}^x[X_{t+s} \in \Gamma | \mathcal{F}_t] = \mathbb{P}^x[X_{t+s} \in \Gamma | X_t], \quad p^x - a.s.
$$

(iv) for $x \in \mathbb{R}^d, s, t \geq 0$ and $\Gamma \in \mathcal{B}(\mathbb{R}^d)$,

$$
\mathbb{P}^x[X_{t+s} \in \Gamma | X_s = y] = \mathbb{P}^x[X_{t+s} \in \Gamma], \quad p^x X_s^{-1} - a.e. y
$$

in the notation of (2.75).

Theorem 2.4.4.1 A one-dimensional Brownian motion is a Markov process. A $d$-dimensional Brownian family is a Markov family.

Given an adapted process $X = \{X_t, \mathcal{F}_t; t \geq 0\}$ and a family of probability measures $\{\mathbb{P}^x\}_{x \in \mathbb{R}^d}$ on $(\Omega, \mathcal{F})$, such that condition (I) of Definition 2.4.4.2 is satisfied, we can define a collection of operators $\{U_t\}_{t \geq 0}$ which map bounded, Borel-measurable, real-valued functions on $\mathbb{R}^d$ into bounded, universally measurable, real-valued functions on the same space. These are given by

$$
(U_t f)(x) \triangleq E^x f(X_t).
$$

(2.79)
In the case where \( f \) is the indicator of \( \Gamma \in \mathcal{B}(\mathbb{R}^d) \), we have \( E^x f(X_t) = P^x[X_t \in \Gamma] \), and the universal measurable of \( U_t f \) follows directly from Definition 2.4.4.2 (I); for an arbitrary, Borel-measurable function \( f \), the universal measurability of \( U_t f \) is then a consequence of the bounded convergence theorem.

**Proposition 2.4.4.1** Conditions (III) and (IV) of Definition 2.4.4.2 can be replaced by:

\[(V) \quad \text{For } x \in \mathbb{R}^d, t \geq 0 \text{ and } \Gamma \in \mathcal{B}(\mathbb{R}^d),
\]

\[ P^x[X_{s+t} \in \Gamma|\mathcal{F}_s] = (U_{t\Gamma})(X_s), \quad P^x - a.s. \]

**Proposition 2.4.4.2** For a Markov family \( X, (\Omega, \mathcal{F}), \{P^x\}_{x \in \mathbb{R}^d} \), we have:

\[(III') \quad \text{for } x \in \mathbb{R}^d, s \geq 0 \text{ and } F \in \mathcal{B}(\mathbb{R}^d)^{[0,\infty)},
\]

\[ P^x[X_{s+} \in F|\mathcal{F}_s] = P^x[X_{s-} \in F|X_s], \quad P^x - a.s. \]

\[(IV') \quad \text{for } x \in \mathbb{R}^d, s \geq 0 \text{ and } F \in \mathcal{B}(\mathbb{R}^d)^{[0,\infty)},
\]

\[ P^x[X_{s+} \in F|X_s = y] = P^y[X \in F], \quad P^xX_s^{-1} - a.e.y. \]

Suppose a process \( X = \{X_t, \mathcal{F}_t; t \geq 0\} \) on a measurable space \((\Omega, \mathcal{F})\), that one can construct a family of so-called shift operators \( \theta_s: \Omega \to \Omega, s \geq 0 \), such that each \( \theta_s \) is \( \mathcal{F}/\mathcal{F}\)-measurable and

\[ X_{s+t}(\omega) = X_t(\theta_s \omega); \forall \omega \in \Omega, s, t \geq 0. \quad (2.80) \]

Thus, (III’) and (IV’) can be reformulated as follows: for every \( F \in \mathcal{F}_x^{X} \) and \( s \geq 0 \),

\[(III'') \quad P^x[\theta_s^{-1}F|\mathcal{F}_s] = P^x[\theta_s^{-1}F|X_s], \quad P^x - a.s \]

\[(IV'') \quad P^x[\theta_s^{-1}F|X_s = y] = P^y[F], \quad P^x - a.s \]

In a manner analogous to what was achieved in Proposition 2.4.4.1, we can capture both (III‘’) and (IV‘’) in the requirement that for every \( F \in \mathcal{F}_x^{X} \) and \( s \geq 0 \),

\[(V'') \quad P^x[\theta_s^{-1}F|\mathcal{F}_s] = P^x(X_s(F)), \quad P^x - a.s \]

**Theorem 2.4.4.2** Let \( X = \{X_t, \mathcal{F}_t; t \geq 0\} \) be an adapted process on a measurable space \((\Omega, \mathcal{F})\), let \( \{P^x\}_{x \in \mathbb{R}^d} \) be a family of probability measures on \((\Omega, \mathcal{F})\), and let \( \{\theta_s\}_{s \geq 0} \) be a family of \( \mathcal{F}/\mathcal{F}\)-measurable shift-operators satisfying (2.80). Then \( X, (\Omega, \mathcal{F}), \{P^x\}_{x \in \mathbb{R}^d} \) is a Markov family if and only if (I), (II) and (V‘’) hold.

**Definition 2.4.4.3** Let \( B = \{B_t, \mathcal{F}_t; t \geq 0\}, (\Omega, \mathcal{F}), \{P^x\}_{x \in \mathbb{R}^d} \) be a \( d \)-dimensional Brownian family. If \( \mu \in \mathbb{R}^d \) and \( \sigma \in L(\mathbb{R}^d, \mathbb{R}^d) \) are constant and \( \sigma \) is nonsingular;
then with $Y_t = \mu t + \sigma B_t$, we say $Y = \{Y_t, \mathcal{F}_t; t \geq 0\}$. $(\Omega, \mathcal{F}), \{P^{\sigma^{-1}x}\}_{x \in \mathbb{R}^d}$ is a de-dimensional Brownian family with drift $\mu$ and dispersion coefficient $\sigma$.

### 2.4.4.2 The Strong Markov Property

Part of the appeal of Brownian motion lies in the fact that the distribution of certain of its functionals can be obtained in closed form. Perhaps the most fundamental of these functionals is the passage time $T_b$ to a level $b \in \mathbb{R}$, defined by

$$T_b(\omega) = \inf \{t \geq 0; B_t(\omega) = b\} \quad (2.81)$$

We call that a passage time for a continuous process is a stopping time.

Let $\{B_t, \mathcal{F}_t; 0 \leq t < \infty\}$ be a standard, one-dimensional Brownian motion on $(\Omega, \mathcal{F}, P^0)$. For $b>0$, we have

$$P^0[T_b < t] = P^0[T_b < t, B_t > b] + P^0[T_b < t, B_t < b].$$

$$P^0[T_b < t, B_t > b] = P^0[B_t > b] \quad \text{and} \quad P^0[T_b < t, B_t < b] = P^0[B_t < b]$$

which then yields

$$P^0[T_b < t] = 2P^0[B_t > b] = \frac{1}{\sqrt{\pi}} \int_{b^2/2}^{\infty} e^{-x^2/2} dx. \quad (2.82)$$

Differentiating with respect to $t$, we obtain the density of the passage time

$$P^0[T_b \in dt] = \frac{|b|}{\sqrt{2\pi t^3}} e^{-b^2/2t} dt; \quad t > 0 \quad (2.83)$$

The preceding reasoning is based on the assumption that Brownian motion “start afresh” (in the terminology of Itô & Mckean (1974)) at the stopping time $T_b$, i.e., that the process $\{B_{t+T_b} - B_{T_b}; 0 \leq t < \infty\}$ is Brownian motion, independent of the $\sigma$-field $\mathcal{F}_{T_b}$. The fact that this “starting afresh” actually takes place at stopping times such that $T_b$ is a consequence of the strong Markov property for Brownian motion.

### Definition 2.4.4.4

Let $d$ be a positive integer and $\mu$ a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. A progressively measurable, $d$-dimensional process $X = \{X_t, \mathcal{F}_t; t \geq 0\}$ on some $(\Omega, \mathcal{F}, P^\mu)$ is said to be a strong Markov process with initial distribution $\mu$ if

(I) $P^\mu[X_0 \in \Gamma] = \mu(\Gamma), \forall \Gamma \in \mathcal{B}(\mathbb{R}^d);$

(II) for any optional time $S$ of $\{\mathcal{F}_t\}, t \geq 0$ and $\Gamma \in \mathcal{B}(\mathbb{R}^d)$,

$$P^\mu[X_{S+} \in \Gamma | \mathcal{F}_{S+}] = P^\mu[X_{S+} \in \Gamma | X_S], \quad P^\mu - a.s. \text{ on } \{S < \infty\}.\]
a progressively measurable process $X = \{X_t, F_t; t \geq 0\}$ on some $(\Omega, \mathcal{F})$, together with a family of probability measure $\{P^x\}_{x \in \mathbb{R}^d}$ on $(\Omega, \mathcal{F})$, such that:

(I) for each $F \in \mathcal{F}$, the mapping $x \mapsto P^x(F)$ is universally measurable;

(II) $P^x[X_0 = x] = 1, \forall x \in \mathbb{R}^d$;

(III) for $x \in \mathbb{R}^d, t \geq 0, \Gamma \in \mathcal{B}(\mathbb{R}^d)$, and any optional time $S$ of $\{F_t\}$,

$$P^\mu[X_{S+t}\in \Gamma|\mathcal{F}_{S+}] = P^\mu[X_{S+t}\in \Gamma|X_S], \mu\text{-a.s. on } S < \infty;$$

(IV) for $x \in \mathbb{R}^d, t \geq 0, \Gamma \in \mathcal{B}(\mathbb{R}^d)$, and any optional time $S$ of $\{F_t\}$,

$$P^x[X_{S+t}\in \Gamma|X_S = y] = P^\mu[X_t\in \Gamma], P^xX^{-1}_S\text{-a.e.y.}$$

**Proposition 2.4.4.3** For a strong Markov family $X = \{X_t, F_t; t \geq 0\}$, $(\Omega, \mathcal{F})$, $\{P^x\}_{x \in \mathbb{R}^d}$, we have

(III') for $x \in \mathbb{R}^d, t \geq 0, F \in \mathcal{B}(\mathbb{R}^d)^0, \infty$, and any optional time $S$ of $\{F_t\}$,

$$P^\mu[X_{S+t}\in F|\mathcal{F}_{S+}] = P^\mu[X_{S+t}\in F|X_S], \mu\text{-a.s. on } S < \infty;$$

(IV') for $x \in \mathbb{R}^d, t \geq 0, F \in \mathcal{B}(\mathbb{R}^d)^0, \infty$, and any optional time $S$ of $\{F_t\}$,

$$P^x[X_{S+t}\in F|X_S = y] = P^\mu[X_t\in \Gamma], P^xX^{-1}_S\text{-a.e.y.}$$

**Proposition 2.4.4.4** Let $X = \{X_t, F_t; t \geq 0\}$ be a progressively measurable process on $(\Omega, \mathcal{F})$. $\{P^x\}_{x \in \mathbb{R}^d}$ is a strong Markov if and only if for any $\{F_t\}$-optional time $S$ and $t \geq 0$, one of the following holds:

(V) for any $\Gamma \in \mathcal{B}(\mathbb{R}^d)$,

$$P^x[X_{S+t}\in \Gamma|\mathcal{F}_{S+}] = (U_t\Gamma)(X_S), P^\mu\text{-a.s. on } S < \infty$$

(V') for any bounded, continuous $f: \mathbb{R}^d \to \mathbb{R}$,

$$E^x[f(X_{S+t})|\mathcal{F}_{S+}] = (U_tf)(X_S), P^\mu\text{-a.s. on } S < \infty.$$

**Remark** If $X = \{X_t, F_t; t \geq 0\}$, $(\Omega, \mathcal{F})$, $\{P^x\}_{x \in \mathbb{R}^d}$ is a strong Markov family and $\mu$ is a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, we can define a probability measure $P^\mu$ (on $\mathcal{B}(\mathbb{R}^d)$) by $P^\mu(F) = \int_{\mathbb{R}^d} P^x(F)\mu(dx)$ for every $F \in \mathcal{F}$, and then $X$ on $(\Omega, \mathcal{F}, P^\mu)$ is a strong Markov process with initial distribution $\mu$. Condition (II) of Definition 2.4.4.4 can be verified upon writing condition (V) in integrated form:

$$\int_F(U_t\Gamma)(X_S)dP^x = P^x[X_{S+t}\in \Gamma, F]; F \in \mathcal{F}_{S+},$$

and then integrating both sides with respect to $\mu$. Similarly if $X, (\Omega, \mathcal{F}), \{P^x\}_{x \in \mathbb{R}^d}$ is a Markov family, then $X$ on $(\Omega, \mathcal{F}, P^\mu)$ is a Markov process with initial distribution $\mu$. 35
Conditions (V) and (V’) are statements about the conditional distribution of $X$ at a single time $S+t$ after the optional time $S$. If there are shift operators $\{\theta_s\}_{s \geq 0}$ satisfying (2.80), then for any random time $S$ we can define the random shift $\theta_S: \{S < \infty\} \to \Omega$ by

$$ (\theta_S \omega)(t) = (\theta_t \omega)(t) \text{ on } \{S = s\}. $$

In other words, $\theta_S$ is defined so that whenever $S(\omega) < \infty$, then

$$ X_{S(\omega)+t}(\omega) = X_{t}(\theta_S(\omega)). $$

In particular, we have $\{X_{S+t} \in E\} = \theta^{-1}_S\{X \in E\}$, and (III’) and (IV’) are, respectively, equivalent to the statements: for every $x \in \mathcal{F}_X^\omega$, and any optional time $S$ of $\{\mathcal{F}_t\}$, (III’’)

$$ P^x[\theta^{-1}_S F]\mathcal{F}_{S+t} = P^x[\theta^{-1}_S F]\mathcal{F}_s, \text{ } P^x\text{-a.s. on } \{S < \infty\}; $$

(IV’’)

$$ P^x[\theta^{-1}_S F]X_s = y = P^y[F], \text{ } P^xX_S^{-1}\text{-a.s.} $$

Both (III’’) and (IV’’) can be captured by the single condition:

(V’’)

$$ x \in \mathbb{R}^d, F \in \mathcal{F}_X^\omega, \text{ and any optional time } S \text{ of } \{\mathcal{F}_t\}, \quad \quad \quad P^x[\theta^{-1}_S F]\mathcal{F}_{S+t} = P^xS[F], \text{ } P^x\text{-a.s. on } \{S < \infty\}. $$

**Theorem 2.4.4.3** Let $X = \{X_t, \mathcal{F}_t; t \geq 0\}$ be a progressively measurable process on $(\Omega, \mathcal{F})$, let $\{P^x\}_{x \in \mathbb{R}^d}$ be a family of probability measures on $(\Omega, \mathcal{F})$, and let $\{\theta_s\}_{s \geq 0}$ be a family of $\mathcal{F}/\mathcal{F}$-measurable shift operators satisfying (2.80). Then $X, (\Omega, \mathcal{F})$, $\{P^x\}_{x \in \mathbb{R}^d}$ is a strong Markov family if and only if (I), (II) and (V’’) hold.

### 2.4.4.3 The Strong Markov Property for Brownian Motion

**Definition 2.4.4.6** Let $X$ be a random variable on a probability space $(\Omega, \mathcal{F}, P)$ taking values in complete, separable metric space $(S, \mathcal{B}(S))$. Let $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$. A regular conditional probability of $X$ given $\mathcal{G}$ is a function $Q: \Omega \times \mathcal{B}(S) \to [0,1]$ such that

(I) for each $\omega \in \Omega$, $Q(\omega; \cdot)$ is a probability measure on $(S, \mathcal{B}(S))$.

(II) for each $E \in \mathcal{B}(S)$, the mapping $\omega \mapsto Q(\omega; E)$ is $\mathcal{G}$-measurable, and

(III) for each $E \in \mathcal{B}(S)$, $P[X \in E|\mathcal{G}] = Q(\omega; E), P$-a.e.$\omega$.

**Lemma 1.** Let $X$ be a $d$-dimensional random vector on $(\Omega, \mathcal{F}, P)$. Suppose $\mathcal{G}$ is a sub-$\sigma$-field of $\mathcal{F}$ and suppose that for each $\omega \in \Omega$, there is a function $\varphi(\omega; \cdot): \mathbb{R}^d \to \mathbb{C}$ such that for each $\mu \in \mathbb{R}^d$, 

36
\[ \varphi(\omega; \mu) = E\left[e^{i(\mu,x)}\big| \mathcal{G}\right](\omega), \quad p - a.e. \omega. \]

If, for each \( \omega, \varphi(\omega; \cdot) \) is the characteristic function of some probability measure \( P^\omega \) on \((\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))\), i.e.,
\[ \varphi(\omega; \mu) = \int_{\mathbb{R}^d} e^{i(\mu,x)} P^\omega(dx), \]
then for each \( \Gamma \in \mathcal{B}(\mathbb{R}^d) \), we have
\[ P[X \in \Gamma| \mathcal{G}](\omega) = P^\omega(\Gamma) \quad p - a.e. \omega. \]

**Lemma 2.** For each \( X \in \mathbb{R}^d \), the process \( \{R_t, \mathcal{F}_t; t \geq 0\} \) and \( \{I_t, \mathcal{F}_t; t \geq 0\} \) are martingales on \((\Omega, \mathcal{F}, P)\).

**Theorem 2.4.4.7** A d-dimensional Brownian family is a strong Markov family. A d-dimensional Brownian motion is a strong Markov process.

**Theorem 2.4.4.8** Let \( S \) be an a.s. finite optional time of the filtration \( \{\mathcal{F}_t\} \) for the d-dimensional Brownian motion \( B = \{B_t, \mathcal{F}_t; t \geq 0\} \). Then with \( W_t \triangleq B_{S+t} - B_{S} \), the process \( W = \{W_t, \mathcal{F}_t^W; t \geq 0\} \) is a d-dimensional Brownian motion, independent of \( \mathcal{F}_S \).

**Proposition 2.4.4.5** Let \( X = \{X_t, \mathcal{F}_t; t \geq 0\}, (\Omega, \mathcal{F}), \{P^x\}_{x \in \mathbb{R}^d} \) be a strong Markov family, and the process \( X \) be right-continuous. Let \( S \) be an optional time of \( \{\mathcal{F}_t\} \) and \( T \) an \( \mathcal{F}_S \)-measurable random time satisfying \( T(\omega) \geq S(\omega) \) for all \( \omega \in \Omega \). Then, for any \( x \in \mathbb{R}^d \) an any bounded, continuous \( f: \mathbb{R}^d \rightarrow \mathbb{R} \),
\[ E^x[f(X_T)|\mathcal{F}_S^+](\omega) = \left(U_{T(\omega)-S(\omega)}f\right)(X_{S(\omega)}(\omega)), \quad P^x - a.e. \omega \in \{T < \infty\}. \quad (2.84) \]

**Proposition 2.4.4.6** Let \( \{B_t, \mathcal{F}_t; t \geq 0\} \) be a standard, one-dimensional Brownian motion, and for \( b \neq 0 \), let \( T \) be the first passage time to \( b \) as in (2.81). Then \( T_b \) has the density given by (2.83).

**2.4.5 Computations Based on Passage Times**

Let \( \{W_t, \mathcal{F}_t; 0 \leq t < \infty\}, (\Omega, \mathcal{F}), \{P^x\}_{x \in \mathbb{R}} \) will be a one-dimensional Brownian family. We recall from (2.81) the passage times
\[ T_b(\omega) = \inf\{t \geq 0; B_t(\omega) = b\}; \ b \in \mathbb{R}, \]
and define the running maximum (or maximum-to-date)
\[ M_t = \max_{0 \leq s \leq t} W_s. \quad (2.85) \]

**Proposition 2.4.5.1** We have for \( t > 0 \) and \( a \leq b, b > 0 \):
\[ P^0[W_t \in da, M_t \in db] = \frac{2(b-a)^2}{\sqrt{2\pi t^3}} \exp\{-\frac{(2b-a)^2}{2t}\}dadb \quad (2.86) \]

**Remark** Because of for \( t > 0 \),
\[ P^0[M_t \in db] = P^0[[W_t] \in db] = P^0[M_t - W_t \in db] = \frac{2}{\sqrt{2\pi t^3}} e^{\frac{-b^2}{2t}} db; \ b > 0, \]
we see that
\[ P^0[T_b \leq t] = P^0[M_t \geq b] = \frac{2}{\sqrt{2\pi t^3}} \int_b^\infty e^{-x^2/2} dx; \ b > 0 \quad (2.87) \]

By differentiation, we recover the passage time density (2.83). For future reference, we note that this density has Laplace transform
\[ E^0[e^{-\alpha T_b}] = e^{-b\sqrt{2\alpha}}; \ b > 0, \alpha > 0. \quad (2.88) \]

**Proposition 2.4.5.2** The process of passage times \( T = \{T_a, \mathcal{F}_{T_a}; 0 \leq a < \infty\} \) has the property that, under \( P^0 \) and for \( 0 \leq a < b \), the increment \( T_b - T_a \) is independent of \( \mathcal{F}_{T_a} \) and has the density
\[ P^0[T_b - T_a \in dt] = \frac{b-a}{\sqrt{2\pi t^3}} e^{-(b-a)^2/2t} dt; \ 0 < t < \infty. \quad (2.89) \]

In particular,
\[ E^0\left[e^{-\alpha(T_b-T_a)} \mid \mathcal{F}_{T_a}\right] = e^{-(b-a)^2\sqrt{2\alpha}}; \ \alpha > 0. \quad (2.90) \]

**Proposition 2.3.5.3** Choose \( 0 < x < a \). Then for \( t > 0, 0 < y < a \):
\[ P^x[W_t \in dy, T_0 \wedge T_a > t] = \sum_{n=-\infty}^{\infty} p_-\left(t; x, y + 2na\right). \quad (2.91) \]

**Remark** For \( t > 0, 0 \leq y < a \):
\[ P^0[T_0 \in dt, T_0 < T_a] = \frac{1}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} (2na + x) \exp\left\{-\frac{(2na + x)^2}{2t}\right\} dt, \]
\[ P^0[T_0 \in dt, T_a < T_0] = \frac{1}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} (2na + a - x) \exp\left\{-\frac{(2na + a - x)^2}{2t}\right\} dt, \]
Indeed, one can use the identity (2.88) to compute the Laplace transforms of the
right-hand sides; then the above formula are seen to be equivalent to

\[ E^x[e^{-\alpha T_0} I_{\{T_0 < T_\alpha\}}] = \frac{\sinh((\alpha - \chi)\sqrt{2}\alpha)}{\sinh(\alpha\sqrt{2}\alpha)}, \quad (2.92) \]

\[ E^x[e^{-\alpha T_\alpha} I_{\{\alpha < T_0\}}] = \frac{\sinh(\chi\sqrt{2}\alpha)}{\sinh(\alpha\sqrt{2}\alpha)}. \quad (2.93) \]

By adding (2.92) and (2.93) we obtain

\[ E^x[e^{-\alpha(T_0 \wedge T_\alpha)}] = \frac{\cosh((\alpha - \alpha/2)\sqrt{2}\alpha)}{\sinh(\alpha\sqrt{2}\alpha/2)}. \quad (2.94) \]

### 2.5 Itô’s Formula

Itô’s formula is the fundamental theorem of stochastic calculus, just as one speaks of the fundamental theorem of ordinary integral/differential calculus.

Theorem 2.5 Let \( X_t \) be a \( n \)-dimensional continuous \( (\mathcal{F}_t) \) semimartingale.

Then, for every \( f \in C^2(\mathbb{R}^n) \), we have:

\[ f(X_t) = f(X_0) + \int_0^t \nabla f(X_s) dX_s + \frac{1}{2} \int_0^t \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} f(X_s) d < X^i, X^j >_s. \]

### 2.6 Ornstein-Uhlenbeck Processes

**Proposition 2.6** Let \( k, \theta \) and \( \sigma \) be bounded Borel functions, and \( W \) a Brownian motion. The solution of

\[ dr_t = k(t)(\theta(t) - r_t) dt + \sigma(t) dW_t \quad (2.95) \]

is

\[ r_t = e^{-K(t)}(r_0 + \int_0^t e^{K(s)} \theta(s) ds + \int_0^t e^{K(s)} \sigma(s) dW_s) \quad (2.96) \]

where \( K(t) = \int_0^t k(s) ds \). The process \( (r_t, t \geq 0) \) is a Gaussian process with mean

\[ E[r_t] = e^{-K(t)}(r_0 + \int_0^t e^{K(s)} \theta(s) ds) \quad (2.97) \]

and covariance

\[ e^{-K(t) - K(s)} \int_0^{t \wedge s} e^{2K(u)} \sigma^2(u) du. \quad (2.98) \]

**Definition 2.6** An Ornstein-Uhlenbeck (OU) process driven by a Brownian motion follows the dynamics \( dr_t = -kr_t dt + \sigma dW_t \).
2.7 Bessel Process

Let $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ be an $n$-dimensional Brownian motion where $n \in N$ and define the process $R$ as $R_t = ||\beta_t||$, i.e., $R_t^2 = \sum_{i=1}^{n}(\beta_i)^2 (t)$. Itô’s formula leads to $dR_t^2 = \sum_{i=1}^{n} 2\beta_i(t) \, d\beta_i(t) + ndt$.

Note that for any $t > 0$, $P(R_t = 0) = 0$, hence the process $W$ defined as

$$dW_t = \frac{1}{R_t} \sum_{i=1}^{n} \beta_i(t) \, d\beta_i(t) \quad (2.99)$$

is a real-valued Brownian motion and the process $R$ satisfies

$$d(R_t^2) = 2R_t \, dW_t + ndt. \quad (2.100)$$

Hence, setting $\rho_t = R_t^2$

$$d\rho_t = 2\sqrt{\rho_t} \, dW_t + ndt \quad (2.101)$$

and, using Itô’s formula, which may be easily justified for $n > 1$, one obtains

$$dR_t = dW_t + \frac{n-1}{2} \frac{dt}{R_t} \quad (2.102)$$

We shall say that $R$ is a Bessel process (BES) of dimension $n$, and $\rho$ is a squared Bessel process (BESQ) of dimension $n$.

**Definitions 2.7** For every $\delta \geq 0$ and $x \geq 0$, the unique strong solution of the equation

$$Z_t = x + 2 \int_{0}^{t} \sqrt{Z_s} \, d\beta_s + \delta t \quad (2.103)$$

is called the square of $\delta$-dimensional Bessel process started at $x$ and is denoted by $BESQ^\delta(x)$. The number $\delta$ is the dimension of $BESQ^\delta$.

**Corollary.** For $\delta > 0$, the semi-group of $BESQ^\delta$ has a density in $y$ equal to

$$q_t^\delta(x,y) = \frac{1}{2} \left( \frac{y}{x} \right)^{\frac{\delta}{2}} \exp \left( -\frac{x+y}{2t} \right) I_v \left( \frac{\sqrt{xy}}{t} \right), \quad t > 0, x > 0$$

where $v$ is the index corresponding to $\delta$ and $I_v$ is the Bessel function of index $v$.

For $x = 0$, this density becomes

$$q_t^\delta(0,y) = (2t)^{-\frac{\delta}{2}} \Gamma(\delta/2) y^{\frac{\delta}{2} - 1} \exp \left( -\frac{y}{2t} \right).$$

The density of the semi-group is also obtained from that of $BESQ^\delta$ by a straightforward change of variable (Let $x' = x^2, y' = y^2$) and is found equal, for $\delta > 0$, to
\[ p_t^\delta(x, y) = t^{-1} \left( \frac{y}{x} \right)^v \exp \left( -\frac{x^2 + y^2}{2t} \right) I_v \left( \frac{xy}{t} \right), \text{ for } t > 0, x > 0 \]  \hfill (2.104)

and

\[ p_t^\delta(0, y) = 2^{-v} t^{-(v+1)} \Gamma(v + 1)^{-1} y^{2v+1} \exp \left( -\frac{y^2}{2t} \right). \]  \hfill (2.105)
Chapter 3 First-order Autoregression

Suppose the observe time series \( \{x_i\} \) is generated from the following AR(1) process, given \( x_0 \),

\[
x_i = \theta x_{i-1} + \varepsilon_i
\]

where \( \{\varepsilon_i\} \) is independent and identically distributed disturbances which is satisfying \( \text{E}(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2 \).

\[
x_t = \theta^t x_0 + \theta^{t-1} \varepsilon_1 + \theta^{t-2} \varepsilon_2 + \cdots + \varepsilon_t
\]

\( \bar{\beta}, \bar{\sigma} \) is the estimate of \( \beta, \sigma \), and

\[
\bar{\theta}_T = \frac{1}{T} \sum_{i=1}^{T} \frac{x_i x_{i-1}}{x_{i-1}^2}, \quad \bar{\sigma}_T^2 = \frac{1}{T} \sum_{i=1}^{T} (x_i - \bar{\theta}_T x_{i-1})^2.
\]

We consider \( \{x_i\} \) is a unit root process, when the absolute value of coefficient \( \theta \) is equal to 1. And consider it is stationary, when the absolute value of coefficient \( \theta \) is less than 1. And consider it is explosive process, when the absolute value of coefficient \( \theta \) is more than 1. But here we just consider situation about the coefficient is nonnegative.

3.1 Stationary AR Process

3.1.1 Stationary and Ergodic Theory

If a random variable \( X \) is indexed to time, usually denoted by \( t \), the observations \( \{X_t, t \in T\} \) is called a time series, where \( T \) is a time index set. For a time series \( \{X_t\}_{t=-\infty}^{\infty} \), we need to model the dependence over infinite number of random variables. The autocovariance and autocorrelation functions provide us a tool for this purpose.

**Definition 3.1.1.1 (Autocovariance Function).** The autocovariance function of a time series \( \{X_t\} \) with \( \text{Var}(X_t) < \infty \) is defined by

\[
\gamma_X(s, t) = \text{Cov}(X_s, X_t) = E[(X_s - EX_s)(X_t - EX_t)]. \tag{3.1}
\]

With autocovariance functions, we can define the covariance stationary, or weak stationary. In the literature, usually stationary means weak stationary, unless otherwise specified.

**Definition 3.1.1.2 (Stationary or Weak Stationary)** The time series \( \{X_t, t \in \mathbb{Z}\} \)
where $Z$ is the integer set) is said to be stationary if

(I) $E[X_t^2] < \infty \ \forall t \in Z$;

(II) $E[X_t] = \mu \ \forall t \in Z$;

(III) $\gamma_X(s, t) = \gamma_X(s + h, t + h) \ \forall s, t, h \in Z$.

In other words, a stationary time series $\{X_t\}$ must have three features: finite variation, constant first moment, and that the second moment $\gamma_X(s, t)$ only depends on $(t - s)$ and not depends on $s$ or $t$. In light of the last point, we can rewrite the autocovariance function of a stationary process as

$$\gamma_X(h) = \text{Cov}(X_t, X_{t+h}) \ \text{for} \ t, h \in Z.$$ 

Also, when $X_t$ is stationary, we must have

$$\gamma_X(h) = \gamma_X(-h).$$

When $h = 0$, $\gamma_X(0) = \text{Cov}(X_t, X_t)$ is the variance of $X_t$, so the autocorrelation function for a stationary time series $\{X_t\}$ is defined to be

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}.$$

**Definition 3.1.1.3 (Strict Stationary)** The time series $\{X_t, t \in Z\}$ is said to be strict stationary if the joint distribution of $(X_{t_1}, \cdots, X_{t_k})$ is the same as that of $(X_{t_1+h}, \cdots, X_{t_k+h})$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathbf{E} = (\mathbf{E}_1, \mathbf{E}_2, \cdots)$ a sequence of random variables or, as we say, a random sequence. Let $\theta_k \mathbf{E}$ denotes the sequence $(\mathbf{E}_{k+1}, \mathbf{E}_{k+2}, \cdots)$.

**Definition 3.1.1.4** A random sequence $\mathbf{E}$ is stationary (in the strict sense) if probability distributions of $\theta_k \mathbf{E}$ and $\mathbf{E}$ are the same for every $k \geq 1$.

$$P((\mathbf{E}_1, \mathbf{E}_2, \cdots) \in B) = P((\mathbf{E}_{k+1}, \mathbf{E}_{k+2}, \cdots) \in B) \quad B \in \mathcal{B}(\mathbb{R}^\infty).$$

If $\mathbf{E} = (\mathbf{E}_1, \mathbf{E}_2, \cdots)$ is a sequence of independent identically distribution random variables with $E[\mathbf{E}_1] < \infty$ and $E[\mathbf{E}_1] = m$, the law of the large numbers tells us that, with probability 1,

$$\frac{\mathbf{E}_1 + \cdots + \mathbf{E}_n}{n} \to m, \quad n \to \infty.$$
Definition 3.1.1.5 A transformation $T$ of $\Omega$ into $\Omega$ is measurable if, for every $A \in \mathcal{F}$,
$$T^{-1}A = \{ \omega : T\omega \in A \} \in \mathcal{F}.$$  

Definition 3.1.1.6 A measurable transformation $T$ is a measure-preserving transformation if, for every $A \in \mathcal{F}$,
$$P(T^{-1}A) = P(A).$$  

Theorem 3.1.1.1 Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $T$ be a measure-preserving transformation, and let $A \in \mathcal{F}$. Then, for almost every point $\omega \in A$, we have $T^n \omega \in A$ for infinitely many $n \geq 1$.

Definition 3.1.1.6 A set $A \in \mathcal{F}$ is invariant if $T^{-1}A = A$. A set $A \in \mathcal{F}$ is almost invariant if $A$ and $T^{-1}A$ differ only by set of measure zero, i.e. $P(A \Delta T^{-1}A) = 0$.

Definition 3.1.1.7 A measure-preserving transformation $T$ is ergodic (or metrically transitive) if every invariant set $A$ has measure either zero or one.

Theorem 3.1.1.2 Let $T$ be a measure-preserving transformation. Then the following conditions are equivalent:

(I) $T$ is ergodic

(II) Every almost invariant random variable is (P-a.s.) constant;

(III) Every invariant random variable is (P-a.s.) constant.

Definition 3.1.1.8 A measure-preserving transformation is mixing (or has the mixing property) if, for all $A$ and $B \in \mathcal{F}$,
$$\lim_{n \to \infty} P(A \cap T^{-n}B) = P(A)P(B).$$  

Theorem 3.1.1.3 Every mixing transformation $T$ is ergodic.

Theorem 3.1.1.4 (Birkhoff and Khinchin) Let $T$ be a measure-preserving transformation and $\mathcal{E} = \mathcal{E}(\omega)$ a random variable with $E|\mathcal{E}| < \infty$. Then (P-a.s.)
$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{E}(T^k \omega) = E[\mathcal{E}|\mathcal{F}].$$
If also $T$ is ergodic then (P-a.s.)

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{E}(T^k \omega) = E[\mathcal{E}].$$

**Lemma (Maximal Ergodic Theorem)** Let $T$ be a measure-preserving transformation, let $\mathcal{E}$ be a random variable with $E|\mathcal{E}| < \infty$, and let

$$S_k(\omega) = \mathcal{E}(\omega) + \mathcal{E}(T\omega) + \cdots + \mathcal{E}(T^{k-1}\omega),$$

$$M_k(\omega) = \max\{0, S_1(\omega), \ldots, S_1(\omega)\}.$$

Then

$$E[\mathcal{E}(\omega)1_{\{M_n > 0\}}(\omega)] \geq 0$$

for every $n \geq 1$.

**Corollary.** A measure-preserving transformation $T$ is ergic if and only if, for all $A$ and $B \in \mathcal{F}$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} P(A \cap T^{-n}B) = P(A)P(B).$$

**Theorem 3.1.1.5** Let $T$ be a measure-preserving transformation and $\mathcal{E} = \mathcal{E}(\omega)$ a random variable with $E|\mathcal{E}| < \infty$. Then (P-a.s.)

$$E[\lim_{n} \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{E}(T^k \omega) - E[\mathcal{E}|\mathcal{F}] \to 0.$$ 

If also $T$ is ergodic then (P-a.s.)

$$E[\lim_{n} \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{E}(T^k \omega) - E[\mathcal{E}] \to 0.$$ 

**Definition 3.1.1.9** A set $A \in \mathcal{F}$ is invariant with respect to the sequence $\mathcal{E}$ if there is a set $B \in \mathcal{B}(R\infty)$ such that for $n \geq 1$

$$A = \{\omega: (\mathcal{E}_n, \mathcal{E}_{n+1}, \ldots) \in B\}.$$ 

**Definition 3.1.1.10** A stationary sequence $\mathcal{E}$ is ergodic if the measure of every invariant set is either 0 or 1.
Theorem 3.1.1.6 (Ergodic Theorem) Let $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2, \cdots)$ be a stationary (strict sense) random sequence with $E|\mathcal{E}_1| < \infty$. Then (P-a.s., and in the mean)

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{E}_k(\omega) = E[\mathcal{E}_1|\mathcal{F}_c].
$$

If $\mathcal{E}$ is also an ergodic sequence, then (P-a.s., and in the mean)

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{E}_k(\omega) = E[\mathcal{E}_1].
$$

3.1.2 Stationary AR process

Suppose the observe time series $\{x_t\}$ is generated from the following AR(1) process, given $x_0$,

$$
x_t = \theta x_{t-1} + \varepsilon_t
$$

Here process $\{\varepsilon_t\}$ is independent and identically distributed disturbances which is satisfying normal distribution with mean zero and variance $\sigma^2$.

If $\theta < 1, x_0 = 0$, we can get

$$
x_t = \sum_{i=0}^{t} \theta^i \varepsilon_{t-i}
$$

and we also get it’s expectation and variance

$$
\mu = E(x_t) = 0,
$$

$$
\text{Var}(x_t) = \sum_{i=0}^{t} \theta^{2i} \sigma^2 \to \frac{\sigma^2}{1-\theta^2} < \infty.
$$

By (3.1) and (3.3), we can get the autocovariance

$$
\gamma_X(j) = \text{Cov}(X_t, X_{t-j}) = E[(X_t - \mu)(X_{t-j} - \mu)]
$$

$$
= \frac{\theta^j \sigma^2}{1-\theta^2} = \gamma_{-j}.
$$

(3.3), (3.4), (3.5) is satisfied the conditions (I)-(III) of Definition 3.1.1.2. When $\beta < 1$, the AR(1) is stationary process. $\{\varepsilon_t\}$ is ergodic, and $\{X_t\}$ is also ergodic.

By M.C.L.T which is given in chapter (2.2.7), we can get

$$
\sqrt{n}(\bar{X} - \mu) \to N(0, \sigma^2).
$$

46
and
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{i-1} \varepsilon_i \rightarrow N \left( 0, \frac{\sigma^2}{1-\beta^2} \right). \tag{3.7}
\]

Use the ordinary least square to get the estimate
\[
\bar{\theta}_n = \sum_{i=1}^{n} \frac{x_i x_{i-1}}{x_{i-1}^2}, \quad \bar{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{\theta}_n x_{i-1})^2,
\]
so that with (3.7) we get
\[
\sqrt{n}(\bar{\theta} - \theta) = \frac{1}{\sqrt{n} \sum_{i=1}^{n} x_{i-1}^2} \frac{d}{\sum_{i=1}^{n} x_{i-1}^2} \rightarrow N(0,1-\theta^2). \tag{3.8}
\]

By (3.4) and (3.8), we can consider
\[
t = \frac{\bar{\theta} - \theta}{\bar{\sigma}_n / \sqrt{\sum_{i=1}^{n} x_{i-1}^2}} \rightarrow N(0,1) \tag{3.9}
\]

### 3.2 Unit Root Process

**Definition 3.2.1 (Random walk)** The process \( \{X_t\} \) is a random walk if it satisfies

(I) \( X_t = X_{t-1} + \varepsilon_t \) for all \( t = 1, 2, \cdots \);

(II) \( X_0 = 0 \);

(III) \( \{\varepsilon_t\} \) is iid such that \( E(\varepsilon_t) = 0, 0 < E(\varepsilon_t^2) < \infty \).

Random walk is not a stationary process.

**Definition 3.2.2 (Unit root process)** The process \( \{X_t\} \) is a unit root process if it satisfies \( (1-L)X_t = U_t \) where \( \{U_t\} \) is a mean zero covariance stationary process with short memory.

Suppose the observe time series \( \{x_i\} \) is generated from the following AR(1) process, given \( x_0 \),
\[
X_i = \beta X_{i-1} + \varepsilon_i
\]
Here process \( \{\varepsilon_t\} \) is independent and identically distributed disturbances which is satisfying normal distribution with mean zero and variance \( \sigma^2 \).

If \( \beta = 1, x_0 = 0 \), the process \( \{X_t\} \) will change to be
\[
X_t = \sum_{i=0}^{t} \varepsilon_i. \tag{3.10}
\]

Let process \( \{Y_n\} \) satisfy with
\[ Y_{nt} = \frac{X_t}{\sqrt{n}} = \frac{\varepsilon_1}{\sqrt{n}} + \ldots + \frac{\varepsilon_{[nt]}}{\sqrt{n}} \]

By **Theorem 2.4.1**, process \( \{X_t\} \) is a standard one-dimensional Brownian motion. So
\[ X_t \to W_t \quad (3.11) \]

Phillips (1987) showed that the sample moments of \( \{x_i\} \) converge to random functions of Brownian motion:
\[
\begin{align*}
& n^{-2} \sum_{i=1}^{n} x_{i-1} \overset{d}{\to} \sigma \int_0^1 W_t \, dt \\
& n^{-2} \sum_{i=1}^{n} x_{i-1}^2 \overset{d}{\to} \sigma^2 \int_0^1 W_t^2 \, dt \\
& n^{-1} \sum_{i=1}^{n} x_{i-1} \epsilon_i \overset{d}{\to} \sigma^2 \int_0^1 W_t^2 \, dW(t),
\end{align*}
\]

where \( W_t \) denotes a standard Brownian motion (Wiener process) defined on the unit interval.

We know that the F-test statistic is
\[ F = T(\bar{\theta}_n - 1) = \frac{n^{-1} \sum_{i=1}^{n} x_{i-1} \epsilon_i}{n^{-2} \sum_{i=1}^{n} x_{i-1}^2 - 1}. \quad (3.12) \]

Using the above results Phillips showed that under the unit root \( \theta = 1 \)
\[ F = n(\bar{\theta}_n - 1) = \frac{n^{-1} \sum_{i=1}^{n} x_{i-1} \epsilon_i}{n^{-2} \sum_{i=1}^{n} x_{i-1}^2} \overset{d}{\to} \frac{\int_0^1 W_t \, dt \, dW(t)}{\int_0^1 W_t^2 \, dt} \quad (3.13) \]

So we can get the t-test statistic
\[ t = \frac{(\bar{\theta}_n - 1)}{\sigma / \sqrt{\sum_{i=1}^{n} x_{i-1}^2}} \overset{d}{\to} \frac{\int_0^1 W_t \, dt}{\sqrt{\int_0^1 W_t^2 \, dt}} \quad (3.14) \]

The above yield some surprising results:

- \( \bar{\theta} \) is super-consistent; that is, \( \bar{\theta} \overset{p}{\to} \theta \) at rate \( T \) instead of the usual rate \( n^{1/2} \)
- \( \bar{\theta} \) is not asymptotically normally distributed and \( t \) is not asymptotically standard normal.
- The limiting distribution of \( t \) is called the Dickey-Fuller (DF) distribution and does not have a closed form representation. Consequently, quantiles of the distribution must be computed by numerical approximation or by simulation.
- Since the normalized bias \( n(\bar{\theta}_n - 1) \) has a well defined limiting distribution that does not depend on nuisance parameters it can also be used as a test statistic for \( \theta = 1 \).
### 3.3 Explosive Process

Suppose the observe time series \( \{x_i\} \) is generated from the following AR(1) process, given \( x_0 \),

\[
X_i = \theta X_{i-1} + \varepsilon_i
\]

Here process \( \{\varepsilon_i\} \) is independent and identically distributed disturbances which is satisfying normal distribution with mean zero and variance \( \sigma^2 \).

If \( \theta > 1 \) and \( x_0 = 0 \), the variance

\[
Var(x_t) = \sigma^2 (1 + \theta^2 + \ldots + \theta^{2(t-1)}) = \frac{\sigma^2}{(\theta^2 - 1)\theta^t - 1}
\]

and the variance increases exponentially as \( t \) increases. \( \lim_{t \to \infty} Var(x_t) = \infty \). The process is said to be “explosive”.

To further understand the importance of the first few \( \varepsilon_i \) in the model with \( \theta > 1 \), we observe that

\[
X_t = \sum_{i=0}^{t-1} \theta^i \varepsilon_{t-i} = \sum_{i=1}^{t} \theta^i \varepsilon_i = \theta^t \sum_{i=1}^{t} \theta^{-i} \varepsilon_i, \quad (3.15)
\]

Letting \( Y_t = \sum_{i=1}^{t} \theta^{-i} \varepsilon_i \), we have

\[
Y_t \to Y \quad \text{a.s.}
\]

at \( t \) increases, where \( Y = \sum_{i=1}^{\infty} \theta^{-i} \varepsilon_i \). Thus for large \( t \), the behavior of \( X_t \) is essentially that of an exponential multiple of the random variable \( Y \). We may write \( X_t \) as

\[
X_t = \theta^t Y_t, \quad t = 0, 1, 2, \ldots \quad (3.16)
\]

Now \( Y_t \) converges \( a.s. \) as \( t \to \infty \) to random variable \( Y = \sum_{i=1}^{\infty} \theta^{-i} \varepsilon_i \). Then \( \theta^{-t} X_t \to Y \) \( a.s. \) as \( t \) increases. Note that

\[
Y - Y_t = \sum_{i=t+1}^{\infty} \theta^{-i} \varepsilon_i = \theta^{-(t+1)} \varepsilon_{t+1} + \theta^{-(t+2)} \varepsilon_{t+2} + \ldots \quad (3.17)
\]

The error in the least squares estimator of \( \theta \) is

\[
\hat{\theta} - \theta = \frac{\sum_{i=1}^{n} x_{i-1} \varepsilon_i}{\sum_{i=1}^{n} x_{i-1}^2} = \frac{\theta^{-2n} \sum_{i=1}^{n} x_{i-1} \varepsilon_i}{\theta^{-2n} \sum_{i=1}^{n} x_{i-1}^2} \quad (3.18)
\]

The denominator is equal to

\[
\theta^{-2n} \sum_{i=1}^{n} x_{i-1}^2
\]

\[
= \sum_{i=1}^{n} \theta^{-2(n-i+1)} Y^2 + 2 \sum_{i=1}^{n} \theta^{2(i-n-1)} Y (Y_{i-1} - Y) + \sum_{i=1}^{n} \theta^{2(i-n-1)} (Y_{i-1} - Y)^2.
\]

Let \( Y - Y_{i-1} = \sum_{j=i}^{\infty} \theta^{-j} \varepsilon_j \) and \( Z_i = \sum_{k=0}^{\infty} \theta^{-k} \varepsilon_{i+k} = \theta^i (Y - Y_{i-1}) \), we can get

\[
\sum_{i=1}^{n} \theta^{2(i-n-1)} Y (Y_{i-1} - Y) = - \sum_{i=1}^{n} \theta^{-2(n-i+1)} Y \theta^{-i} Z_i
\]

\[
= - \theta^{-(n+2)} \sum_{i=1}^{n} \theta^{-(n-i)} Y Z_i \quad (3.19)
\]

and
\[
\sum_{i=1}^{n} \theta^{2(i-n-1)} (Y_{i-1} - Y)^2 \\
= \sum_{i=1}^{n} \theta^{2(i-n-1)} (\theta^{-1}Z_i)^2 \\
= \sum_{i=1}^{n} \theta^{-2(n+1)} Z_i^2
\]  
(3.20)

Combining with (3.19), (3.20) and
\[
\sum_{i=1}^{n} \theta^{-2(n-i+1)} Y^2 = (\theta^{-2n} + \theta^{-2(n-1)} + \ldots + \theta^{-2}) Y^2 = \sum_{k=0}^{n-1} \theta^{-2(k+1)} Y^2,
\]  
(3.21)

the denominator is equal to
\[
\theta^{-2n} \sum_{i=1}^{n} X_i^{-2} \\
= -2\theta^{-(n+2)} \sum_{i=1}^{n} \theta^{-(n-i)} Y Z_i + \sum_{i=1}^{n} \theta^{-2(n+1)} Z_i^2 + \sum_{k=0}^{n-1} \theta^{-2(k+1)} Y^2
\]  
(3.22)

The numerator of (3.18) will be
\[
\theta^{-2n} \sum_{i=1}^{n} X_i^{-1} \varepsilon_i \\
= \theta^{-n} \sum_{i=1}^{n} \theta^{-(n-i+1)} Y_{i-1} \varepsilon_i \\
= \theta^{-n} \sum_{i=1}^{n} \theta^{-(n-i+1)} [Y - (Y - Y_{i-1})] \varepsilon_i \\
= \theta^{-n} \sum_{i=1}^{n} \theta^{-(n-i+1)} Y \varepsilon_i - \theta^{-n} \sum_{i=1}^{n} \theta^{-(n+1)} Z_i \varepsilon_i.
\]  
(3.23)

Because of \( Z_i = \sum_{k=0}^{\infty} \theta^{-k} \varepsilon_{i+k} \) is a ergodic stationary process, with the Ergodic Theorem (Theorem 3.1.1.6) we can get \( E[Z] = 0, Var[Z] = E[Z^2] = (1 - \theta^{-2})^{-1} \).

And
\[
\left| \theta^{-(n+2)} \sum_{i=1}^{n} \theta^{-(n-i)} Y Z_i \right| \leq n|\theta|^{-(n+2)} \frac{1}{n} \sum_{i=1}^{n} |YZ_i| \rightarrow 0 \text{ a.s.}
\]

so that
\[
-2\theta^{-(n+2)} \sum_{i=1}^{n} \theta^{-(n-i)} Y Z_i \rightarrow 0 \text{ a.s.}, n \rightarrow \infty.
\]  
(3.24)

By Ergodic Theorem, we also can obtain \( \frac{1}{n} \sum_{i=1}^{n} Z_i \varepsilon_i \rightarrow E[Z_1 \varepsilon_1] \text{ a.s.} \), so \( \frac{1}{n} \sum_{i=1}^{n} Z_i \varepsilon_i \) almost surely convergence to 0 and
\[
\sum_{i=1}^{n} \theta^{-(n+1)} Z_i \varepsilon_i = n\theta^{-(n+1)} \frac{1}{n} \sum_{i=1}^{n} Z_i \varepsilon_i \rightarrow 0 \text{ a.s.}, n \rightarrow \infty.
\]  
(3.25)

By Ergodic Theorem, we can get \( \frac{1}{n} \sum_{i=1}^{n} Z_i^2 \rightarrow 0 \text{ a.s.} \), so
\[
\sum_{i=1}^{n} \theta^{-2(n+1)} Z_i^2 = n\theta^{-2(n+1)} \left( \frac{1}{n} \sum_{i=1}^{n} Z_i^2 \right) \rightarrow 0 \text{ a.s.}
\]  
(3.26)

In addition,
\[
\sum_{k=0}^{n-1} \theta^{-2(k+1)} Y^2 \rightarrow (\theta^2 - 1)^{-1} Y^2, \ n \rightarrow \infty.
\]  
(3.27)

Combining with (3.22)-(3.27), the t-test statistics is equal to
\[ t = \frac{\bar{\theta}_n - \theta}{\sigma_n / \sqrt{\sum_{i=1}^{n} x_i^2}} = \frac{\theta - n \sum_{i=1}^{n} x_{i-1} \varepsilon_i}{\sigma_n / \sqrt{\sum_{i=1}^{n} x_i^2}} = \frac{\theta - n \sum_{i=1}^{n} x_{i-1} \varepsilon_i}{\sigma_n / \sqrt{\sum_{i=1}^{n} x_i^2}} = \frac{\theta - n \sum_{i=1}^{n} x_{i-1} \varepsilon_i}{\sigma_n \sqrt{\sum_{i=1}^{n} x_i^2}} \sim \frac{\sum_{i=1}^{n} \theta^{-(n-i+1)} \varepsilon_i}{\sigma_n \sqrt{\sum_{i=1}^{n} x_i^2}} . \]

And

\[ t \sim \frac{\sum_{i=1}^{n} \theta^{-(n-i+1)} \varepsilon_i}{\sigma / \sqrt{(\theta^2 - 1) \varepsilon^2}} = \frac{\sigma^2}{\theta^2 - 1} \sum_{i=1}^{n} \theta^{-(n-i+1)} \varepsilon_i. \quad (3.28) \]

By Ergodic theorem,

\[ E \left[ \frac{\sigma^2}{\theta^2 - 1} \sum_{i=1}^{n} \theta^{-(n-i+1)} \varepsilon_i \right] = 0, \quad (3.29) \]

\[ \text{VAR} \left( \frac{\sigma^2}{\theta^2 - 1} \sum_{i=1}^{n} \theta^{-(n-i+1)} \varepsilon_i \right) \rightarrow \frac{\sigma^2}{\theta^2 - 1} \left( \frac{\theta^2 - 1}{\sigma^2} \right) = 1. \quad (3.30) \]

Therefore, by (3.28)-(3.30), we will get the t-test statistics converges in distribution to normal distribution with mean zero and variance one. It means that

\[ t = \frac{\bar{\theta}_n - \theta}{\sigma_n / \sqrt{\sum_{i=1}^{n} x_i^2}} \rightarrow N(0,1) \quad (3.31) \]

### 3.4 Ornstein-Uhlenbeck(OU) Process Representation of AR(1)

Suppose the observe time series \( \{x_i\} \) is generated from the following AR(1) process, given \( x_0 = 0, \)

\[ X_i = \theta X_{i-1} + \varepsilon_i \]

Here process \( \{\varepsilon_i\} \) is independent and identically distributed disturbances which is satisfying normal distribution with mean zero and variance \( \sigma^2. \)

So we can get

\[ x_n = \varepsilon_t + \theta \varepsilon_{n-1} + \theta^2 \varepsilon_{n-2} + \cdots + \theta^{n-1} \varepsilon_1 + \theta^n x_0 \]

Let both sides subtract \( \beta^n x_0, \) so we can get

\[ x_n - \theta^n x_0 = \sum_{k=0}^{n-1} \theta^k \varepsilon_{n-k} \]

Here we consider \( S_n = \sum_{i=0}^{n} \varepsilon_i, \) so we can get

\[ x_n - \theta^n x_0 = \sum_{k=0}^{n-1} \theta^k (S_{n-k} - S_{n-k-1}) \]

Because of
As we know $x_0 = 0$, so that we can get
\[ x_n = \beta \theta^{-1} S_n + (\theta - 1) \sum_{i=0}^{n-1} \theta^{n-i-1} S_i. \tag{3.32} \]
Here we consider $n = [mt]$, $i = [ms]$, $Y_t = x_{[mt]} / \sigma \sqrt{m}$, $\theta = 1 - \delta / m$, and divide the result by $\sigma \sqrt{m}$, then we get
\[ Y_t = \frac{x_{[mt]}}{\sigma \sqrt{m}} = (1 - \delta / m)^{-1} \frac{S_{[mt]}}{\sigma \sqrt{m}} - \frac{\delta}{m} \sum_{i=0}^{[mt]-1} (1 - \delta / m)^{[mt]-[ms]-1} \frac{S_{[ms]}}{\sigma \sqrt{m}}. \]

Because of $n - i - 1 = [mt] - [ms] - 1 \rightarrow m(t - s)$ for $m \rightarrow \infty$.

By Theorem 2.4.1,
\[ \frac{S_{[ms]}}{\sigma \sqrt{m}} = \frac{\sum_{i=0}^{[ms]} \varepsilon_i}{\sigma \sqrt{m}} \rightarrow W_s. \]

By $\lim_{n \rightarrow \infty} (1 + a / n)^n = e^a$,
\[ \left[ 1 + \frac{(-\delta)}{m} \right]^m \rightarrow e^{-\delta}. \]

We can get
\[ Y_t \rightarrow_D W_t - \delta \int_0^t e^{-\delta(t-s)} W_s ds, \tag{3.33} \]
where $\rightarrow_D$ indicates convergence for every $\omega$ in the sense of $D[0,\infty)$ as $m \rightarrow \infty$. Actually the convergence in $D[0,\infty)$ is equivalent to the uniform convergence on each finite interval when the limit function is continuous. Then we multiply both side by $e^{\delta t}$, we will get
\[ Y_t e^{\delta t} = e^{\delta t} W_t - \delta \int_0^t e^{-\delta s} W_s ds. \]

We substitute $Z_t = Y_t e^{\delta t}, e^{-\delta t} = E_t$, so
\[ Y_t = E_t Z_t = f(E_t, Z_t). \]

By derivation principle of composite function, we get
\[
\frac{\partial f(E_t, Z_t)}{\partial E_t} = Z_t, \quad \frac{\partial f(E_t, Z_t)}{\partial Z_t} = E_t.
\]
\[
\frac{\partial^2 f(E_t, Z_t)}{\partial Z_t^2} = 0, \quad \frac{\partial^2 f(E_t, Z_t)}{\partial E_t^2} = 0, \quad \frac{\partial^2 f(E_t, Z_t)}{\partial Z \partial M} = 0.
\]

Derivative for \( Z_t \) and \( E_t \),
\[
dZ_t = d(Y_t e^{\delta t}) = d\left(e^{\delta t}W_t - \delta \int_0^t e^{-\delta s}W_s ds\right)
\]
\[
= -\delta e^{\delta t}W_t dt + \delta e^{\delta t}W_t dt + e^{\delta t}dW_t = e^{\delta t}dW_t,
\]
\[
dE_t = (-\delta)e^{-\delta t}dt
\]  \hspace{1cm} \text{(3.34)}

According to Itô’s process, we will know
\[
dY_t
\]
\[
= \frac{\partial f(E_t, Z_t)}{\partial Z_t} dZ_t + \frac{\partial f(E_t, Z_t)}{\partial E_t} dE_t + \frac{1}{2} \frac{\partial^2 f(E_t, Z_t)}{\partial Z_t^2} d < Z > + \frac{1}{2} \frac{\partial^2 f(E_t, Z_t)}{\partial E_t^2} d < E > + \frac{1}{2} \frac{\partial^2 f(E_t, Z_t)}{\partial Z \partial M} d < M, Z >. \hspace{1cm} \text{(3.36)}
\]

By (3.21)-(3.23), we can obtain
\[
dY_t = Z_t dE_t + E_t dZ_t
\]
\[
= Y_t e^{\delta t}(-\delta)e^{-\delta t}dt + e^{-\delta t}e^{\delta t}dW_t
\]
\[
= (-\delta)Y_t dt + dW_t.
\]

By Definition 2.6, we know it’s an Ornstein-Uhlenbeck Process.
Chapter 4 Sequential Unit Root Test

In statistics, a unit root test tests whether a time series variable is non-stationary using an autoregressive model. A well-known test that is valid in large samples is the augmented Dickey–Fuller test. The optimal finite sample tests for a unit root in autoregressive models were developed by John Denis Sargan and Alok Bhargava. Another test is the Phillips–Perron test. These tests use the existence of a unit root as the null hypothesis.

Sequential analysis or sequential hypothesis testing is statistical analysis where the sample size is not fixed in advance. Instead data are evaluated as they are collected, and further sampling is stopped in accordance with a pre-defined stopping rule as soon as significant results are observed. Thus a conclusion may sometimes be reached at a much earlier stage than would be possible with more classical hypothesis testing or estimation, at consequently lower financial and/or human cost.

In this chapter, we use sequential unit root test that combine the sequential test and unit root test to test AR(1) process. Consider a scalar AR(1) process,

$$X_i = \theta X_{i-1} + \epsilon_i$$

where \( \epsilon_i \)'s are disturbances that satisfy \( \epsilon_i \sim iid N(0, \sigma^2) \). When \( \beta = \pm 1 \), the series is said to be a unit root process.

4.1 The stopping Time

The testing hypotheses we consider are,

$$H_0: \theta = \theta_0 = 1 \quad \text{vs} \quad H_1: \theta = \theta_1 = 1 - c/m$$

where \( c \) is a non-zero constant which can be either negative or positive. \( c/m \) localizes the alternative hypothesis. For historical samples, we set local alternatives or contiguous alternatives such that they approach to the null as the sample size increases. In sequential inferences, we consider the asymptotic where \( m \), the degree of accuracy of the procedure, increases. Local alternatives in the present problem are defined such that they approach to the null as \( m \) increases. An advantage of the local alternative is that the local asymptotic normality (LAN) holds for the Fisher-information-based stopping time.

Suppose the disturbance variance \( \sigma^2 \) is known for simplicity. We define the stopping rule as Lai and Siegmund (1983)

$$T_m = \inf \{ t > 0 | \sum_{i=1}^t \frac{x_i^2}{\sigma^2} \geq m^2 s \}$$

(4.1)
in which as for \( m \) we utilize the unit of standard deviation rather than that of Fisher information unlike Lai and Siegmund's definition and \( s \) is a constant. Employing this stopping time, they estimate the parameter by

\[
\text{\( \tilde{\theta}_{T_m} = \sum_{i=1}^{T_m} \frac{x_i x_{i-1}}{x_{i-1}^2}, \)}
\] (4.2)

To motivate the stopping time (4.1), the likelihood function based on the sample is

\[
L_t(\theta | x_1, \ldots, x_t) = \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^t \prod_{i=1}^{t} \exp \left( -\frac{(x_i - \theta x_{i-1})^2}{2\sigma^2} \right).
\] (4.3)

\[
I_t = -\frac{\partial^2}{\partial \theta^2} \log L_t(\theta | x_1, \ldots, x_t) = \sum_{i=1}^{t} \frac{x_i - \theta x_{i-1}}{\sigma^2}
\] (4.4)

is so-called the “observe” Fisher information. \( m \) controls for the accuracy of the estimation in the sense that the variance of the estimate, \( I_t^{-1} \), is guaranteed to be smaller than \( m^{-2} \). There exists a trade-off between the accuracy of estimation and the cost of sampling. If we set \( m \) large, \( T_m \) will tend to be also large by construction, which will yield a more accurate estimate. If we set \( m \) small, sampling will stop relatively earlier, but the accuracy will be lower. Note that \( T_m \) is also a statistic depending on the observations.

Lai and Siegmund (1983) prove the asymptotic normality of \( \tilde{\theta}_{T_m} \) as \( m \to \infty \), uniformly on in \( \theta \) for \( \theta \in [-1, 1] \),

\[
\sqrt{T_m} (\tilde{\theta}_{T_m} - \theta) \Rightarrow N(0,1),
\] (4.5)

where \( \Rightarrow \) indicates weak convergence (We proof it in the chapter 3). To test the existence of unit root, we can determine the critical value for the test statistic \( \sqrt{T_m} (\tilde{\theta}_{T_m} - 1) \) simply from the standard normal table. They also derive the stochastic limit of the stopping time,

\[
P \left( \lim_{n \to \infty} \frac{T_m}{m^2} = 1 - \theta^2 \right) = 1
\] (4.6)

for non-local \( \theta \in (-1, 1) \). Obviously, the mean sample size \( E[T_m] \) and the power of the test statistic depend on \( m \).

And the likelihood ratio test at the Fisher-information based time \( T_m \) defined in (4.1) is LAN;

\[
A_{T_m}(\theta_m | x_1, \ldots, x_t) = \frac{L_t(\theta_1 | x_1, \ldots, x_t)}{L_t(\theta_0 | x_1, \ldots, x_t)} \equiv \exp \left\{ \frac{c}{m \sigma^2} \sum_{i=1}^{T_m} (x_i - x_{i-1})x_{i-1} - \frac{c^2}{2m^2 \sigma^2} \sum_{i=1}^{T_m} x_{i-1}^2 \right\}
\]
under the null as $m \to \infty$ where $Z \sim N(0,1)$. It is an immediate consequence from Lai and Siegmund (1983).

Therefore, the test statistics $m(\hat{\theta}_{T_m} - 1)$ is asymptotically $N(c,1)$ by Le Cam's third lemma (Le Cam (1986)) under the alternative. Hence, given the Fisher information based sample size $T_m$, the likelihood ratio test is asymptotically most powerful for each $c$ by Neyman-Peason's lemma. Thus, the equivalent test using $m(\hat{\theta}_{T_m} - 1)$ becomes asymptotically uniformly most powerful (AUMP) test. However, the AUMP property depends on the particular stopping time $T_m$ and there might exists another pair of a test statistic and a stopping time which provides more power and smaller expected sample size for some $c$. We would not be able to obtain the statistical properties of the stopping time from Le Cam's lemmas, that is an important issue especially when sampling cost is high.

By (4.1), we can obtain that

$$
\frac{T_m}{m} = \frac{1}{m} \inf \left\{ t > 0 \bigg| \sum_{i=1}^{t} \frac{x_{i-1}^2}{(\sigma m)^2} \geq s \right\},
$$

and we consider $t = [mu]$, so that

$$
\frac{T_m}{m} = \frac{1}{m} \inf \left\{ [mu] > 0 \bigg| \sum_{i=1}^{[mu]} \frac{x_{i-1}^2}{(\sigma m)^2} \cdot \frac{1}{m} \geq s \right\}. \tag{4.7}
$$

Put the initial value be $x_0 = 0$, and

$$X_m(t) = \frac{x_{[mu]}}{\sqrt{m\sigma}}.
$$

Let $X_t$ be an Ornstein-Uhlenbeck process that satisfies

$$dX_t = -cX_t dt + dW_t \tag{4.8}
$$

where $W_t$ is the standard Brownian motion, that is the limit of $(m^2\sigma)^{-1} \sum_{i=1}^{[mt]} \epsilon_i$. We shall prove in the appendix that

$$X_m(t) \Rightarrow X_t,$$

$$\frac{T_m}{m} \Rightarrow U_s \equiv \inf \{ u \geq 0 \big| \int_0^u X_t^2 dt = s \}, \tag{4.9}
$$

in the sense of $D[0, \infty)$. The first approximation was originally obtained by Bobkoski
(1983) for $D[0, \infty)$ under the null.

### 4.2 The expectation of stopping time under $H_0$

Under $H_0$, we consider $\theta = 1$, i.e., $c=0$. By Theorem 2.4.1, we know that

$$
\sum_{i=1}^{[\mu]} \frac{x_{i-1}^2}{(\sigma \sqrt{m})^2} \cdot \frac{1}{m} \Rightarrow \int_0^u W_t^2 \, dt = s. \tag{4.10}
$$

Combine with (4.7) and (4.8), we can obtain

$$
\frac{t_m}{m} \Rightarrow U_s \equiv \text{inf}\{u \geq 0| \int_0^u W_t^2 \, dt = s\}, \tag{4.11}
$$

where $m \rightarrow \infty$.

As we know, the disturbance $\{\varepsilon_i\}$ satisfy $\varepsilon_i \sim i.i.d. N(0, \sigma^2)$ and we consider $S_n = \sum_{i=1}^{n} \varepsilon_i$. Furthermore, the invariance principle on $D[0,\infty)$ holds, that is, letting $W_t^m = \frac{S_{[mt]}}{\sigma \sqrt{m}}$ and $W$ be a standard Brownian motion, $W_t^m \Rightarrow W$ as $m \rightarrow \infty$ in the Skorohod topology of $D[0,\infty)$.

We appeal to Skorohod representation theorem on $D[0, \infty)$, which we proofed in section 3.4

$$
x_n = \theta^{-1} S_n + (\theta - 1) \sum_{i=0}^{n-1} \theta^{n-i-1} \varepsilon_i. \tag{4.12}
$$

Substituting $n=\lfloor \mu \rfloor$, $i = \lfloor m s \rfloor$, $X_t = \frac{x_{[ms]}}{\sigma \sqrt{m}}$, $\beta = 1$, and divide the result by $\sigma \sqrt{m}$, then we get

$$
X_m(t) = \frac{\sum_{i=1}^{[\mu]} \varepsilon_i}{\sigma \sqrt{m}} \Rightarrow \frac{S_{[\mu]} D}{\sigma \sqrt{m}} \bigg| X_u \equiv W_u \tag{4.13}
$$

where $\Rightarrow$ indicates convergence for every $\omega$ in the sense of $D[0, \infty)$ as $m \rightarrow \infty$. Actually the convergence in $D[0, \infty)$ is equivalent to the uniform convergence on each finite interval when the limit function is continuous. And it is proofed in the subsection 3.4.

We can obtain

$$
I_m(t) = \frac{1}{\sigma^2 m^2} \sum_{i=1}^{[\mu]} \frac{x_{i-1}^2}{\sigma \sqrt{m}} \stackrel{D}{\rightarrow} \int_0^u W_t^2 \, dt, \tag{4.14}
$$

$$
J_m(t) = \frac{1}{\sigma^2 m} \sum_{i=1}^{[\mu]} \frac{x_{i-1}(x_i - x_{i-1})}{\sigma \sqrt{m}} = \sum_{i=1}^{[\mu]} \frac{x_{i-1}}{\sigma \sqrt{m}} \cdot \frac{\varepsilon_i}{\sigma \sqrt{m}} \Rightarrow \int_0^u W_t \, dW_t \tag{4.15}
$$

by (4.13) and obtain the test statistic
\[ m(\tilde{\theta}_m - 1) = \frac{m \sum_{i=1}^{m} x_{i-1} x_{i-1}}{\sum_{i=1}^{m} x_i^2} \rightarrow \int_0^{U_s} W_t \, dW_t / \int_0^{U_s} W_t^2 \, dt \]  

(4.16)

Combining with (4.11) and (4.16),

\[ m(\tilde{\theta}_m - 1) \rightarrow \int_0^{U_s} W_t \, dW_t / s \]  

(4.17)

Then we use a DDS Brownian motion to deal with the problem. The DDS theorem is play a very important role not just in this chapter but also in this paper. Suppose that \{\mathcal{F}_t\} is the complete right continuous filtration generated by the Brownian \{W_t\}. Let the continuous martingale \( M_t = \int_0^t W_s \, dW_s \) and its quadratic variation process \( \langle M \rangle_t = \int_0^t W_s^2 \, ds \). Consider stopping times \( U_s = \inf\{t \geq 0; \langle M \rangle_t \geq s\} \) defined by \( \langle M \rangle \) and stopped martingale

\[ B_s = M_{U_s} \]  

(4.18)

and let \( g_s = \mathcal{F}_{U_s} \) where \( \mathcal{F}_{U_s} = \{ A \in \mathcal{F}_\infty : A \cap \{ U_s \leq t \} \in \mathcal{F}_t \ \text{for all} \ t \} \). Since \( \langle M \rangle_\infty = \infty \ \text{a.s.} \), \( B_s \) becomes the DDS Brownian motion with respect to \( g_s \). See Revuz and Yor (1999). Therefore,

\[ M_{U_s} = \int_0^{U_s} W_t \, dW_t = B_s \]  

(4.19)

Combined by (4.17) and (4.19), we will have

\[ (m(\tilde{\theta}_m - 1), \frac{\tau_m}{m}) \Rightarrow (B_s / s, U_s). \]  

(4.20)

Here a simple application of the inverse function theorem gives \( dU_s / ds = 1 / X_{U_s}^2 \).

Ito’s formula and plugging \( \mu = U_s \) yield

\[ X_{U_s}^2 = 2 \int_0^{U_s} W_t \, dW_t + U_s = 2B_s + U_s. \]  

(4.21)

Thus

\[ \frac{dU_s}{ds} = \frac{1}{2cs + 2B_s + U_s}. \]

Put \( \rho_s \equiv X_{U_s}^2 / 2 = (2B_s + U_s) / 2 \), then we have

\[ \frac{dU_s}{ds} = \frac{1}{2\rho_s} \]  

(4.22)

\[ d\rho_s = d(B_s) + \frac{1}{2} dU_s = \frac{1}{4\rho_s} ds + dB_s. \]  

(4.23)

This indicates that \( \rho_s \) is the Bessel process of dimension 3/2 without drift and a initial value 0. See Revuz and Yor (1999) and Linetsky (2004) for the details of Bessel process.

Integrating both sides of (4.24),
\[ \rho_s = B_s + 1/2 \int_0^t \frac{1}{2\rho_s} \, ds \]  
(4.24)

We will have

\[ U_s = \int_0^t \frac{1}{2\rho_s} \, ds \]  
(4.25)

by (4.22) and (4.24).

Therefore, we can obtain that

\[ E_0 \left[ \frac{T_m}{m} \right] \Rightarrow E_0[U_s] = 2E_0[\rho_s] + 2E_0[B_1] = 2E_0[\rho_s]. \]  
(4.26)

Combining with (2.105) and (4.26), we can obtain

\[ E_0[U_s] = 2E_0[\rho_s] = 2^{-v+1} t^{-(v+1)} \Gamma(v + 1)^{-1} \int_0^\infty y \cdot y^{2v+1} \exp \left( -\frac{y^2}{2t} \right) dy \]

\[ = 2^{-v+1} t^{-(v+1)} \Gamma(v + 1)^{-1} \int_0^\infty y^{2(v+1)} \exp \left( -\frac{y^2}{2t} \right) dy. \]  
(4.27)

Here we use the definition of Gamma distribution,

\[ \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx. \]  
(4.28)

Let \( x = \frac{y^2}{\beta} \), and we can get \( dx = \frac{2y}{\beta} \, dy \). By (4.28),

\[ \frac{\Gamma(\alpha)}{\beta^\alpha} = \int_0^\infty y^{2\alpha-1} e^{-\frac{y^2}{\beta}} \, dy. \]  
(4.29)

Let \( \alpha = v + \frac{3}{2}, \beta = 2t \). (4.29) change to be

\[ \frac{\Gamma(v+3/2)}{2} (2s)^{v+3/2} \int_0^\infty y^{2(v+3/2)-1} e^{-\frac{y^2}{2t}} \, dy = \int_0^\infty y^{2(v+1)} \exp \left( -\frac{y^2}{2t} \right) dy \]  
(4.30)

By (4.26), (4.27) and (4.30), we will have

\[ E_0 \left[ \frac{T_m}{m} \right] \Rightarrow \frac{1}{2} E_0[\rho_1] = 2^{-v+1} \Gamma(v + 1)^{-1} \frac{\Gamma(v + 3/2)}{2} t^{v+3/2} \]

\[ E_0 \left[ \frac{T_m}{m} \right] \Rightarrow \frac{\Gamma(v+3/2)}{\Gamma(v+1)} 2^{3/2} \sqrt{s}. \]  
(4.31)

Because of \( \delta = 3/2, v = \delta/2 - 1 = -1/4 \).

\[ E_0 \left[ \frac{T_m}{m} \right] \Rightarrow \frac{\Gamma(5/4)}{\Gamma(3/4)} 2^{3/2} = 2.0921 \sqrt{s}. \]  
(4.32)

For the alternative hypothesis, it becomes possible to obtain the asymptotic
expectation of the stopping time by utilizing the transition density of the Bessel diffusion in Linetsky (2004). We will describe on the next section.

4.3 The expectation of stopping time under $H_1$

Here we consider the testing hypotheses the same as above that

$$H_0: \theta = \theta_0 = 1 \quad vs \quad H_1: \theta = \theta_1 = 1 - c/m$$

where $c$ is a constant. The reason is that it is not easy to detect the null of $\theta_0 = 1$ against the alternative of so-called near-unit-root, such like $\theta_1 = 0.99$. And another reason is that it allows a diffusion approximation under the alternative as well as the null. Therefore, $c/m$ localizes the alternative hypothesis. In sequential inferences, we consider the asymptotics where $m$, the degree of accuracy of the procedure, increases.

Local alternatives in the present problem are defined such that they approach to the null as the degree of accuracy $m$ increases.

Then we use the same method as section 4.2. First, we appeal to Skorohod representation theorem on $D[0, \infty)$. Substituting $n=\lfloor mu \rfloor$, $i=\lfloor ms \rfloor$, $X_t = x_{\lfloor mu \rfloor}/\sigma\sqrt{m}$, $\beta = 1 - \delta/m$, and divide the result by $\sigma\sqrt{m}$, then we get

$$X_m(t) = \frac{x_{\lfloor mu \rfloor}}{\sigma\sqrt{m}} = (1 - c/m)^{-1} \frac{S_{\lfloor mu \rfloor}}{\sigma\sqrt{m}} - \frac{c}{m} \sum_{i=0}^{\lfloor mu \rfloor - 1} (1 - c/m)^{\lfloor mu \rfloor - \lfloor ms \rfloor - 1} \frac{S_{\lfloor ms \rfloor}}{\sigma\sqrt{m}}.$$  

By (4.12), we can get

$$X_m(t) \xrightarrow{D} X_u \equiv W_u - ce^{-cu} \int_0^u e^{cs} W_s ds$$  

(4.33)

where $\xrightarrow{D}$ indicates convergence for every $\omega$ in the sense of $D[0, \infty)$ as $m \to \infty$. Actually the convergence in $D[0, \infty)$ is equivalent to the uniform convergence on each finite interval when the limit function is continuous. And it is proofed in the subsection 3.4.

We consider

$$I_m(t) = \frac{1}{\sigma^2 m^2} \sum_{i=1}^{\lfloor mu \rfloor} x_{i-1}^2 \xrightarrow{D} \int_0^u X_t^2 dt,$$  

(4.34)

$$J_m(t) = \frac{1}{\sigma^2 m} \sum_{i=1}^{\lfloor mu \rfloor} x_i - x_{i-1} = \sum_{i=1}^{\lfloor mu \rfloor} \frac{x_i - x_{i-1}}{\sigma\sqrt{m}}$$

\xrightarrow{D} \int_0^u X_t dX_t  

(4.35)
Because of (4.9), (4.34) and (4.35), we can obtain the test statistic
\[ m(\bar{\theta}_{tm} - 1) = \frac{m \sum_{i=1}^{m} x_i - (x_i - x_{i-1})}{\sum_{i=1}^{m} x_i^2} \rightarrow \int_0^{U_s} X_t dX_t / s, \quad (4.36) \]
for every \( \omega \) as \( m \to \infty \).

Second, we use a DDS Brownian motion to deal with the problem. The DDS theorem is play a very important role not just in this chapter but also in this paper. Suppose that \( \{\mathcal{F}_t\} \) is the complete right continuous filtration generated by the Brownian \( \{W_t\} \). Let the continuous martingale \( M_t = \int_0^t X_s dW_s \) and its quadratic variation process \( <M>_t = \int_0^t X_s^2 ds \). Consider stopping times \( U_s = \inf\{t \geq 0; <M>_t \geq s\} \) defined by \( <M> \) and stopped martingale
\[ B_s = M_{U_s}, \quad (4.37) \]
and let \( g_s = \mathcal{F}_{U_s} \) where \( \mathcal{F}_{U_s} = \{A \in \mathcal{F}_\infty: A \cap \{U_s \leq t\} \in \mathcal{F}_t \text{ for all } t\} \). Since \( <M> \to \infty \text{ a.s.} \), \( B_s \) becomes the DDS Brownian motion with respect to \( g_s \). Here \( U_1 = U \). And (4.8) gives
\[ \int_0^U X_t dX_t = -c \int_0^U X_t^2 dt + \int_0^U X_t dW_t = -cs + \int_0^U X_t dW_t. \quad (4.38) \]
Because of \( \int_0^{U_s} X_t dW_t = M_{U_s} = B_s \), combining with (4.36), we will have
\[ (m(\bar{\theta}_{tm} - 1), \frac{\tau_m}{m}) \Rightarrow (-c + B_s/s, U_s). \quad (4.39) \]
And a simple application of the inverse function theorem gives \( dU_s / ds = 1/X_{U_s}^2 \). Ito’s formula and plugging \( U = U_s \) yield
\[ X_{U_s}^2 = -2c \int_0^{U_s} X_t^2 dt + 2 \int_0^{U_s} X_t dW_t + U_s = -2cs + 2B_s + U_s. \quad (4.40) \]
Thus
\[ \frac{dU_s}{ds} = \frac{1}{-2cs + 2B_s + U_s}. \]
Put \( \rho_s \equiv X_{U_s}^2/2 = (-2cs + 2B_s + U_s)/2 \), then we have
\[ d\rho_s = d(-cs) + d(B_s) + \frac{1}{2} du_s = (-c + \frac{1}{4\rho_s}) ds + dB_s. \quad (4.41) \]
This indicates that \( \rho_s \) is the Bessel process of dimension 3/2 with a drift \(-c\) and a initial value 0. Therefore, we obtain that
\[ E_1 \left[ \frac{\tau_m}{m} \right] \Rightarrow E_1[U_s] = 2E_1[\rho_s - c] + 2E_0[B_s] = 2E_0[\rho_s - c] \quad (4.42) \]
Let \( f_0(z, u) \) be the joint density of the stochastic limit (4.31) of the test statistics and the stopping time under the null obtained in (4.9). Then let the joint density function
under the local alternative to be \( f_c(z, u) \). To obtain the asymptotic joint density, we appeal to a multivariate Girsanov Theorem. Note that both the coefficient estimator and the stopping time are driven by the same DDS Brownian motion. Putting \( Y_t = B_t - ct \), we have

\[
\frac{d}{d\tau} \left( \frac{X_t}{Y_t} \right) = \left( \begin{array}{c} -c + \frac{1}{4X_t} \\ -c \end{array} \right) dt + \left( \begin{array}{c} 1 \\ 1 \end{array} \right) dB_t.
\]

(4.43)

\( B_t \) is a Brownian motion with respect to, say \( \mathcal{P} \). Since

\[
\left( \begin{array}{c} -c \\ -c \end{array} \right) = -c \left( \begin{array}{c} 1 \\ 1 \end{array} \right),
\]

\((X_t^c, Y_t')\) satisfies

\[
\frac{d}{d\tau} \left( \frac{X_t^c}{Y_t'} \right) = \left( \begin{array}{c} -c + \frac{1}{4X_t^c} \\ -c \end{array} \right) dt + \left( \begin{array}{c} 1 \\ 1 \end{array} \right) d\tilde{B}_t,
\]

where \( \tilde{B}_t = \int_0^t (-c)ds + B_t = -ct + B_t \) is a Brownian motion with respect to \( \mathcal{Q} \) defined by:

\[
d\mathcal{Q} = M_t d\mathcal{P},
\]

\[
M_t = \exp \left\{ -\int_0^t (-c)dB_s - \frac{1}{2} \int_0^t (-c)^2 ds \right\} = \exp \left( cB_t - \frac{1}{2} c^2 t \right) = \exp \left( c\tilde{B}_t + \frac{1}{2} c^2 t \right)
\]

This yields

\[
f_c(z, u) = \exp (-cz - \frac{1}{2} c^2) f_0(z, u).
\]

(4.45)

Hence we obtain the log likelihood ratio:

\[
\log \frac{f_c(z, u)}{f_0(z, u)} = -cz - \frac{1}{2} c^2
\]

(4.46)

which implies the LAN property and asymptotic sufficiency of the sequential unit root test. This indicates that the stopping time does not carry any additional information in testing the null of unit root to the information carried by the AR(1) coefficient estimate. This is a natural result in the case of normal disturbances, but it is also true for non-normal cases.

We note that this likelihood ratio is not exactly the likelihood ratio in the ordinary sense because it does not present the likelihood of the observations themselves, but only their functions, namely \( \tilde{\theta}_{T_m} \) and \( T_m \). A test based on this likelihood ratio may be a reasonable approach especially when we do not know the distribution of the disturbances as we cannot write down the ordinary likelihood.
Chapter 5 Sequential Probability Ratio Test, SPRT

The sequential probability ratio test (SPRT) is a specific sequential hypothesis test, developed by Abraham Wald. Neyman and Pearson’s 1933 result inspired Wald to reformulate it as a sequential analysis problem. The Neyman-Pearson lemma, by contrast, offers a rule of thumb for when all the data is collected (and its likelihood ratio known).

While originally developed for use in quality control studies in the realm of manufacturing, SPRT has been formulated for use in the computerized testing of human examinees as a termination criterion.

Here, we first stop the sampling at the time $T_m$, which was defined on the chapter 4. Then, we stop the sampling, which was changed to be smaller by the first step, at the time $\tau_m^{(a,b)}$ by sequential probability ratio test.

5.1 Wald's Sequential Probability Ratio

Wald's Sequential Probability Ratio Test (SPRT) is currently the only Bayesian Statistical Procedure in SISA (although one might argue that 'sample size' is also Bayesian). What is required in Bayesian statistics is quite a detailed description of the expectations of the outcome under the model prior to executing the data collection. In Wald's SPRT, if certain conditions are met during the data collection decisions are taken with regard to continuing the data collection and the interpretation of the gathered data.

Wald's procedure is particularly relevant if the data is collected sequentially. Sequential Analysis is different from Classical Hypothesis Testing were the number of cases tested or collected is fixed at the beginning of the experiment. In Classical Hypothesis Testing the data collection is executed without analysis and consideration of the data. After all data is collected the analysis is done and conclusions are drawn. However, in Sequential Analysis every case is analyzed directly after being collected, the data collected up to that moment is then compared with certain threshold values, incorporating the new information obtained from the freshly collected case. This approach allows one to draw conclusions during the data collection, and a final conclusion can possibly be reached at a much earlier stage, as is the case in Classical Hypothesis Testing. The advantages of Sequential Analysis are easy to see. As data
collection can be terminated after fewer cases and decisions taken earlier, the savings in terms of human life and misery, and financial savings, might be considerable.

To make this plan work in the program implemented in SISA it is required that one stipulate both the level above which one would describe the situation as 'more' (whatever) and the level below one would describe the situation as 'less'. One has to define these values prior to the data collection. In the program implemented here this is only possible for percentages. Thus, one has to give a percentage for the number positive on the total number above one would declare the situation as improved (or more), and a percentage below one declares the situation as deteriorated (or less). The program echoes for each additional case collected the number positive on the total number that defines the threshold value for the sample, which is collected up to that moment. If a higher number positive is found in the data, the data collection should be terminated and the conclusion should be drawn that the study shows divergence between the hypothesis and the collected data. Similarly for the lower value, for each additional case a number is given on the total number collected up to that moment, if values below this are found the data collection should be stopped because the situation has deteriorated (or gotten less).

Two additional values are required, statistical power \((1 - \beta)\) and \(\alpha\) level. Power is the chance that, if a difference exists in the real world, one also observes a difference in the data. In the program the power level is set at 80%. There is an 80% chance to discover a really existing difference in the sample. Alpha is the chance that one would terminate sampling because one would think one has discovered a difference, while in fact this difference does not exist. Alpha is set at 5%, which means that 5%, or one in twenty, of the experiments would have been incorrectly stopped because the data signaling that 'something' existed, while in fact it did not. Filling in different values can change power level and alpha level. However, the values pre-given are the widely accepted 'usual' values. Only change these values if you have a good reason.

5.2 Relationship between the thresholds and error probability
As in classical hypothesis testing, SPRT starts with a pair of hypotheses, say \(H_0\) and \(H_1\) for the null hypothesis and alternative hypothesis respectively. We consider

\[
\text{Reject } H_0 \quad \text{if } L_n \geq B
\]
Accept $H_0$ if $L_n \leq A$

The next step is calculate the cumulative sum of the log-likelihood ratio $L_i$,

$$L_i = \prod_{i=1}^{n} \frac{f_i(x_i)}{f_0(x_i)}, \quad n = 1, 2, \ldots \quad (5.1)$$

The goal of the SPRT is to decide which hypothesis is correct as soon as possible (i.e., for the smallest value of $n$). To do this the SPRT requires two thresholds, $0 < A < B < \infty$ (usually $A < 1 < B$). The SPRT “stops” as soon as $\log L_n \geq \log B = b$, and we then decide $H_1$ is correct, or when $\log L_n \leq \log A = a$, and we then decide $H_0$ is correct. The key is to set the thresholds so that we are guaranteed a certain levels of error. Making $b$ larger and $a$ smaller yields a test that will tend to stop later and produce more accurate decisions. We will try to set the thresholds to provide desired probabilities of detection $P_D$ (which is the power $1 - \beta$ that we said above) and false-alarm $P_{FA}$ (which is the $\alpha$ level that we said above). We can express $P_D$ as follows. To simplify the notation, let $x := (x_1, \cdots, x_n)$ and write $f_j(x) := \sum_{i=1}^{n} f(x_i), j = 0, 1$. $P_D$ can be written in terms of the decision set $R_1$ ($R_1 := \{L_n \geq a\}$) as follows:

$$\beta = P_D = \int_{R_1} f_1(x) dx = \int_{R_1} \frac{f_1(x)}{f_0(x)} f_0(x) dx = \int_{R_1} L_n(x) f_0(x) dx = a L_n(x)$$

$$\log A_n(x) = \log \frac{1 - \beta}{\alpha} \geq b$$

where we use the fact that $\log A_n \geq b$ on the set $R_1$. Similarly,

$$1 - \alpha = 1 - P_{FA} = 1 - \int_{R_1} f_0(x) dx = \int_{R_0} \frac{f_0(x)}{f_1(x)} f_1(x) dx = \int_{R_0} L_n^{-1}(x) f_1(x) dx$$

$$\log L_n^{-1}(x) = \frac{1 - \alpha}{\beta} \geq a^{-1}$$

These expressions give us bounds on the thresholds necessary to achieve $P_D(\beta)$ and $P_{FA}(\alpha)$:

$$a \leq \log \frac{\beta}{1 - \alpha}, \quad b \geq \log \frac{1 - \beta}{\alpha} \quad (5.2)$$
Let us err on the side of conservatism and set \( a = \log \frac{\beta}{1-\alpha} \) and \( b = \log \frac{1-\beta}{1-\alpha} \). These thresholds guarantee that error probabilities of the test will be at least as small as specified by choice of \( \alpha \) and \( \beta \), but they could be too conservative.

So we will get the stopping rule that is a simple threshold scheme:

- \( a < \log L_n < b \): continue monitoring, i.e., no change occurs (critical inequality)
- \( \log L_n \geq b \): Accept \( H_1 \)
- \( \log L_n \leq a \): Accept \( H_0 \)

where \( a \) and \( b \) depend on the desired probability of the error of the first kind and the error of the second kind (\( \alpha \) and \( \beta \)). They may be chosen as follows:

\[
a \cong \log \frac{\beta}{1-\alpha}, \quad \quad b \cong \log \frac{1-\beta}{1-\alpha} \tag{5.3}
\]

\[
\alpha \cong \frac{1-A}{B-A} = \frac{1-e^a}{e^b-e^a}, \quad \beta \cong A \left( \frac{B-1}{B-A} \right) = \frac{e^a (1-e^b)}{e^b-e^a}
\]

In other words, \( \alpha \) and \( \beta \) must be decided beforehand in order to set the thresholds appropriately. The numerical value will depend on the application. The reason for using approximation signs are that, in the discrete case, the signal may cross the threshold between samples. Thus, depending on the penalty of making an error and the sampling frequency, one might set the thresholds more aggressively. Of course, the exact bounds may be used in the continuous case.

### 5.3 SPRT of AR(1)

In the case of a simple AR(1) model, we consider

\[
X_i = \theta X_{i-1} + \varepsilon_i \quad i = 1, 2, \ldots
\]

and \( X_0 = 0, \varepsilon_i \sim_{i.i.d.} N(0, \sigma^2) \).

#### 5.3.1 Stopping Time

The testing hypotheses we consider are,

\[
H_0: \theta = 1 - \frac{\delta}{m} \quad V.S. \quad H_1: \theta = 1
\]

where \( \delta \) is a positive constant. \( \delta/m \) localizes the alternative hypothesis.

Let the likelihood ratio of AR(1) process to be \( L_n, L_0 = 1, \).
Then we use a DDS Brownian motion to let problem to be easy. Suppose that

\[ f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i | x_{i-1}). \]

According to Markov Chain, we consider

\[ L_n = \prod_{i=1}^{n} \exp \left\{ \frac{(x_i - x_{i-1} + (1-\theta)x_{i-1})^2}{2\sigma^2} - \frac{(x_i - x_{i-1})^2}{2\sigma^2} \right\} \]

\[ = \prod_{i=1}^{n} \exp \left\{ \frac{(1-\theta)^2}{2\sigma^2} \sum_{i=1}^{n} x_{i-1} (x_i - x_{i-1}) + \frac{(1-\theta)^2}{2\sigma^2} \sum_{i=1}^{n} x_{i-1}^2 \right\} \]

\[ = \exp \left\{ \frac{\delta}{m\sigma^2} \sum_{i=1}^{n} x_{i-1} (x_i - x_{i-1}) + \frac{\delta^2}{2(m\sigma^2)} \sum_{i=1}^{n} x_{i-1}^2 \right\} \]  \hspace{1cm} (5.4)

Here we let \( n=[mt], i=[ms], X_t = x_{[mt]} / \sigma \sqrt{m} \), so we can get

\[ l_n \to \delta \int_0^t X_s \, dX_s + \frac{\delta^2}{2} \int_0^t X_s^2 \, ds, \quad m \to \infty. \]  \hspace{1cm} (5.5)

And the stopping time

\[ \tau_{a,b}^{(m)} = \inf \{ n \geq 1: L_n \leq e^a, \text{or} \ L_n \geq e^b \} \]

\[ = \inf \{ n \geq 1: \log L_n \notin (a, b) \} \]

\[ = \inf \{ n \geq 1: \left[ \frac{\delta}{m\sigma^2} \sum_{i=1}^{n} x_{i-1} (x_i - x_{i-1}) + \frac{\delta^2}{2(m\sigma^2)} \sum_{i=1}^{n} x_{i-1}^2 \right] \notin (a, b) \} \]  \hspace{1cm} (5.6)

will change to be

\[ \frac{\tau_{a,b}^{(m)}}{m} \to \min \{ t \geq 1: \delta \int_0^t X_s \, dX_s + \frac{\delta^2}{2} \int_0^t X_s^2 \, ds \notin (a, b) \}, \quad m \to \infty \]  \hspace{1cm} (5.7)

Then we use a DDS Brownian motion to let problem to be easy. Suppose that \( \{ \mathcal{F}_t \} \)

is the complete right continuous filtration generated by the Brownian \( \{ \mathcal{W}_t \} \). Let the continuous martingale \( M_t = \int_0^t X_v \, dW_v \text{ and its quadratic variation process } \langle M \rangle_t = \int_0^t X_v^2 \, dv \). Consider stopping times \( U_v = \inf \{ t \geq 0: \langle M \rangle_t \geq v \} \text{ defined by } \langle M \rangle \text{ and stopped martingale} \]

\[ L_n = \frac{\tilde{f}_{H_1}(x_1, x_2, \ldots, x_n)}{\tilde{f}_{H_0}(x_1, x_2, \ldots, x_n)}, \quad n \geq 1 \]

\[ = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ \frac{(x_i - x_{i-1})^2}{2\sigma^2} \right\} \]

\[ = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ \frac{(x_i - \theta x_{i-1})^2}{2\sigma^2} \right\} \]
\[ B_v = M_{U_v} \]  
(5.10)

and let \( g_v = \mathcal{F}_{U_v} \) where \( \mathcal{F}_{U_v} = \{ A \in \mathcal{F}_\infty : A \cap \{ U_v \leq t \} \in \mathcal{F}_t \text{ for all } t \} \). Since \( < M >_\infty = \infty \) a.s., \( B_v \) becomes the DDS Brownian motion with respect to \( g_s \). Here

\[ U_v \equiv \inf\{ u \geq 0 \mid \int_0^u X_s^2 \, ds = v \} \]  
(5.11)

So we can obtain

\[ B_v = M_{U_v} = \int_0^{U_v} X_v \, dW_v. \]  
(5.12)

\[ \int_0^{U_v} X_v^2 \, dt = v \]  
(5.13)

(4.8) and (5.6) gives

\[ \int_0^{U_v} X_t \, dX_t = -\delta \int_0^{U_v} X_v^2 \, dt + \int_0^{U_v} X_t \, dW_t = -\delta v + B_v. \]  
(5.14)

5.3.2 The expectation of stopping time \( H_0 \)

Under \( H_0 \), there is no change. We consider \( \theta = 1 - \frac{\delta}{m} \). By chapter 3.4, we know that

\[ dX_t = -\delta X_t \, dt + dW_t. \]

So that we can obtain

\[ \int X_t \, dX_t = \int X_t (-\delta X_t) \, dt + \int X_t \, dW_t \]

\[ = -\delta \int X_t^2 \, dt + \int X_t \, dW_t \]

\[ \int X_t \, dX_t + \frac{\delta}{2} \int X_t^2 \, dt = -\frac{\delta}{2} \int X_t^2 \, dt + \int X_t \, dW_t \]  
(5.15)

With (5.12), (5.13), (5.14) and (5.15), we can obtain

\[ \frac{\tau_{a,b}^{(m)}}{m} \Rightarrow U_{a,b} = \inf\left\{ U_v \geq 0 : \left( \delta \int_0^{U_v} X_u \, dX_u + \frac{\delta^2}{2} \int_0^{U_v} X_u^2 \, du \right) \notin (a, b) \right\} \]

will change to be

\[ \frac{\tau_{a,b}^{(m)}}{m} \Rightarrow U_{a,b} = \min\left\{ v \geq 0 : (B_v - \frac{\delta}{2} v) \notin \left( \frac{a}{\delta}, \frac{b}{\delta} \right) \right\} \]  
(5.16)

So we can get the stopping time under \( H_0 \),

\[ U_{a,b}^* = \min\left\{ u \geq 0 : (B_u - uu) \notin \left( \frac{a}{\delta}, \frac{b}{\delta} \right) \right\}, \text{ here } u = \frac{\delta}{2}. \]  
(5.17)
5.3.3 The expectation of stopping time $H_1$

Under $H_1$, it means that change occurred and the $\delta = 0$. Therefore, we will have

$$x_n = x_{n-1} + \varepsilon_n = \sum_{i=1}^{n} \varepsilon_i$$

$$X_t = \frac{x_t}{\sigma \sqrt{m}} = \frac{\sum_{i=1}^{[mt]} \varepsilon_i}{\sigma \sqrt{m}} = W_t.$$

And (5.9) will transfer to

$$\frac{\tau_{a,b}}{m} \Rightarrow \inf \left\{ t \geq 0 : (\int_0^t W_u \, dW_u + \frac{\delta}{2} \int_0^t W_u^2 \, du) \notin \left( \frac{a}{\delta}, \frac{b}{\delta} \right) \right\}$$

(5.18)

by Theorem 2.4.1, in which we know that

$$\sum_{i=1}^{[mt]} \frac{x^2_{t-1}}{(\sigma \sqrt{m})^2} \cdot \frac{1}{m} \Rightarrow \int_0^u W_s^2 \, ds = 1,$$

and the fist stopping time will transfer to

$$\frac{\tau_{a,b}}{m} \Rightarrow U_v \equiv \inf \{ u \geq 0 \mid \int_0^u W_s^2 \, ds = v \}. \quad (5.19)$$

And $M_t = \int_0^t X_s \, dW_s$ change to be $M_t = \int_0^t W_s \, dW_s$ and $< M >_t = \int_0^t W_s^2 \, ds$.

(5.14) and (5.9) will be

$$\int_0^{U_v} X_t \, dX_t = -\delta \int_0^{U_v} X_t^2 \, dt + \int_0^{U_v} X_t \, dW_t = \int_0^{U_v} W_t \, dW_t = B_v,$$

$$(5.20)$$

$$\frac{\tau_{a,b}^{(m)}}{m} \Rightarrow U^{+}_{a,b} = \min \left\{ u \geq 0 : \left( B_v + uu \right) \notin \left( \frac{a}{\delta}, \frac{b}{\delta} \right), \text{ here } u = \frac{\delta}{2} \right\} \quad (5.21)$$

5.3.3 The error of the first kind ($\alpha$) and the error of the second kind($\beta$)

Combining with (5.14) and (5.21), we will obtain

$$U_{a,b} = \begin{cases} U_{(a,b)} = \min \{ v \geq 0 : (B_v - uu) \notin (c_0, c_1) \}, & \text{when } H_0 \text{ is true} \\ U^{+}_{(a,b)} = \min \{ s \geq 0 : (B_v + uu) \notin (c_0, c_1) \}, & \text{when } H_1 \text{ is true} \end{cases} \quad (5.22)$$

Here $c_0 = \frac{a}{\delta}, c_1 = \frac{b}{\delta}, u = \frac{\delta}{2}$.

We assume a parameter $V$ and

$$V = \begin{cases} V^+(C) = V(u, C) = \min \{ v : B_v + uu = C \} \\ V^-(C) = V(-u, C) = \min \{ v : B_v - uu = C \} \end{cases} \quad (5.23)$$

And we can get the relationship between $V$ and $U$,

$$\begin{cases} U^{+}_{(a,b)} = V^+_c \land V^{+}_c \\ U^{-}_{(a,b)} = V^-_c \land V^{-}_c \end{cases} \quad (5.24)$$
By chapter 5.3.2, we can get the error of the first kind ($\alpha$) and the error of the second kind($\beta$) and

$$
\alpha = P^- = P_{H_0}\{V_{c_0}^- > V_{c_1}^-\},
$$
$$
\beta = P_{H_1}\{V_{c_0}^+ < V_{c_1}^+\} = 1 - P_{H_0}\{V_{c_0}^+ < V_{c_1}^+\} = P^+.
$$

Here we first want to proof that

$$
P\{V(u, C) < V(u, D)\} = \frac{\sinh(Dx|\mu|)e^{Cu}}{\sinh ((D-C)x|\mu|)}
$$

and

$$
\sinh(x) = \frac{e^x - e^{-x}}{2}.
$$

As we know, $B_t$ is a Brownian motion that start at the time 0. And we consider

$$
\tilde{B}_t = B_t - ut
$$

and we can obtain

$$
Z_t = \exp (uB_t - 1/2 u^2 t)
$$

by Girsanov Theorem. We consider stopping time

$$
T_a = \inf\{t: \tilde{B}_t + ut = a\}
$$

and

$$
P^{(u)}(A) = E[I_A Z_t].
$$

Therefore, we will have

$$
P^{(u)}(T_a < T_b < t) = E[I_{(T_a<T_b<t)}Z_t]
$$
$$
= E[I_{(T_a<T_b<t)}E[Z_t|\mathcal{F}_{T_a \wedge T_b}]]
$$
$$
= E[I_{(T_a<T_b<t)}Z_{T_a}]
$$

by Martingale property. Combined with (5.37), (5.38) and (5.39), (5.41) can change to

$$
P^{(u)}(T_a < T_b) = E[I_{(T_a<T_b)}Z_{T_a}] = E[I_{(T_a<T_b)} \exp(ua - 1/2 u^2 T_a)]
$$
$$
= E[I_{(T_a<T_b)} \exp(- 1/2 u^2 T_a)] e^{ua}
$$

Then we want to let the Brownian motion that starts at the time 0 to change to starts at the time $-a$. Therefore (5.42) transfer to

$$
P^{(u)}(T_a < T_b) = E[I_{(T_a<T_b)} \exp(- 1/2 u^2 T_a)] e^{ua}
$$
$$
= E^{-a}[I_{(T_0<T_{b-a})}e^{-1/2u^2T_0}] e^{ua}
$$

Combining with (2.92) in the chapter 2 and (5.43), we will have

$$
P^{(u)}(T_a < T_b) = e^{ua} \frac{\sinh ((b-a-(-a))\sqrt{2\cdot1/2u^2})}{\sinh ((b-a)\sqrt{2\cdot1/2u^2})} = \frac{\sinh (b|\mu|)e^{ua}}{\sinh ((b-a)|\mu|)}
$$

It is the same as the (5.25).
By (5.22),
\[ U_{(a,b)}^- = \min\{v \geq 0: (B_v - u v) \notin (c_0, c_1)\} = \min\{v \geq 0: (-B_v + u v) \notin (-c_1, -c_0)\} , \]
here \( u = \frac{\delta_0}{2}, c_0 = \frac{a}{\delta_0}, c_1 = \frac{b}{\delta_0}. \) So
\[ U_{(a,b)}^- = V^{r+}(u, C') = \min\{v \geq 0: (\frac{-B_v}{z} + \frac{\delta_0}{z} u) \notin (-\frac{\beta}{\delta_0}, -\frac{\alpha}{\delta_0})\} \] (5.35)

Combined with (5.25) the error of the first kind will change to be
\[ \alpha = P \left\{ V^{r+} < V^{r+} \right\} = \frac{\sinh\left(\frac{-a}{\delta_0} + \frac{b}{\delta_0} \right) \cdot \exp\left(\frac{-b}{\delta_0} \right)}{\sinh\left(\frac{b-a}{\delta_0} \right)} = \frac{\sinh\left(\frac{-a}{z} \right) \cdot e^{-\frac{b}{z}}}{\sinh\left(\frac{b-a}{z} \right)} \] (5.36)
and we also can get the error of the second kind,
\[ \beta = P_{H_1} \left\{ V^{s+}_a < V^{s+}_b \right\} \frac{\sinh\left(\frac{b}{\delta_0} \right) \cdot \exp\left(\frac{a}{\delta_0} \right)}{\sinh\left(\frac{b-a}{\delta_0} \right)} = \frac{\sinh\left(\frac{b}{z} \right) \cdot e^\frac{a}{z}}{\sinh\left(\frac{b-a}{z} \right)} \] (5.37)

Combining with (5.46) and (5.47), we will have
\[ a = \log \frac{\beta}{1-\alpha}, \quad b = \log \frac{1-\beta}{\alpha} \] (5.38)

It is the same as (5.3).
Chapter 6 CUSUM Test (Cumulative Sum Test)

A few years later, George Alfred Barnard developed a visualization method, the V-Test. The cumulative sum (CUSUM) test was proposed by Page (1954) as a means to detect sequentially changes in distributions of discrete-time random processes. Lorden (1971) introduced a min-max criterion for the change detection problem, and established the asymptotic optimality of the CUSUM test under his proposed performance measure. Moustakides (1986) proved optimality, under Lorden’s criterion, for the i.i.d. case and for known distributions before and after the change. Ritov (1990) demonstrated a Bayesian optimality property of CUSUM, based on which he also provided an alternative proof for optimality in Lorden’s sense. Finally, optimum CUSUM procedures were proposed by Poor (1998) for exponentially penalized detection delays.

In continuous time, the optimality of CUSUM has been established for Brownian motion with constant drift by Beibel (1996), in the Bayesian setting of Ritov (1990), that yielded also optimality in Lorden’s sense, and by Shiryaev (1996). These results should be compared to the significantly richer and more general ones available for the other popular sequential test, the sequential probability ratio test (SPRT). In continuous time, the SPRT was shown to be optimum in Wald’s sense (Wald (1947)) for Brownian motion with constant drift by Shiryaev (1978), page 180. However, when one replaces in Wald’s criterion the expected delay by the Kullback-Leibler (K-L) divergence, then Liptser and Shiryaev (1978), page 224, demonstrated the optimality of the SPRT for Ito’s processes. Subsequently, this result was extended by Yashin (1983), Irle (1984) and Galtchouk (2001) to more general continuous time processes.

In this chapter, we will use the Cusum Test to analyze the first-order autoregression, and we want to detect the time on which the process is from stationary to unstationary. We call it change point. And we also want to find the false alarm about this test. In section 1 of this chapter, we introduce the classical method of Cusum Procedures. And in section from 2 to 4, we will explain introduce our methods which was used to solve the problem of AR(1) models.

6.1 CUSUM Procedures

Imagine a process which produces a potentially infinite sequence of observation $x_1, x_2, \ldots$. Initially the process is “in control” in the sense that an observer is satisfied to
record the x’s without taking any action. At some unknown time v the process changes and becomes “out of control”. The observer would like to infer from the x’s that this change has taken place and take appropriate action “as soon as possible” after time v.

To give this problem a simple, precise formulation, assume that the x_i are independent random variables and that for some v ≥ 1, x_1, x_2, …, x_v−1 have the probability density function f_0, whereas x_v, x_v+1, …, have the probability density function f_1.

Let P_v denote probability when the change from f_0 to f_1 occurs at the v-th observation, v = 1, 2, …; and let P_0 denote probability when there is no change i.e. v = ∞, so x_1, x_2, … are independent and identically distributed with probability density function f_0.. We seek a stopping rule T which makes the P_v distributions of (T − v)^+ stochastically small, v ≥ 1, subject to the constraint that the P_0 distribution of T stochastically large. A simple formal requirement is to minimize.

\[ \sup_{v \geq 1} E_v [T - v + 1|T \geq v] \]

subject to

\[ E_0[T] \geq B \]

for some given (large) constant B.

An ad hoc proposal to solve this problem approximately is the following. Suppose that x_1, …, x_n have been observed. Consider for 1 ≤ v ≤ n the hypotheses H_v that x_1, …, x_v−1 are distributed according to f_0 and x_v, …, x_n according to f_1, and H_0 the hypothesis of no change. The log likelihood ratio for testing H_v against H_0 is

\[ \sum_{k=v}^{n} \log \{f_1(x_k)/f_0(x_k)\}. \]

For testing the composite hypothesis that at least one of the H_v hold (1 ≤ v ≤ n) against H_0, the log likelihood ration statistic is

\[ \max_{0 ≤ k ≤ n} (S_n - S_k) = S_n - \min_{0 ≤ k ≤ n} S_k, \]

where S_n = \sum_{j=v}^{n} \log \{f_1(x_j)/f_0(x_j)\}. An intuitively appealing stopping rule based on (5.3) is

\[ T = \inf \{n: S_n - \min_{0 ≤ k ≤ n} S_k ≥ b\}. \]

Note that S_n - \min_{0 ≤ j ≤ n} S_k measures the current height of the random walk S_j, j = 0, 1, 2, … above its minimum value. Whenever the random walk establishes a new minimum, i.e. S_n = \min_{0 ≤ k ≤ n} S_k, the process forgets its past and starts over in sense that for all j ≥ 0, S_{n+j} - \min_{0 ≤ k ≤ n} S_k = S_{n+j} - S_n - \min_{0 ≤ k ≤ n} (S_{n+k} - S_n).

This renewal property has several important consequences. First, it implies that for T defined by (5.4), \( \sup_{v \geq 1} E_v [T - v + 1|T \geq v] \leq E_1[T], \) because at all times after v-1
the process (5.3) must be at least as large as if there had been a renewal at \( v-1 \). Hence to evaluate (5.1) and (5.2) one must calculate \( E_v[T] \) for \( v = 0, 1 \), and for each of these “extreme” case \( x_1, x_2, \cdots \) are identically distributed. Second, it means that \( T \) can be defined in terms of a sequence of sequential probability ratio test follows. Let

\[
N_1 = \inf\{n: S_n \notin (0, b)\}. \tag{6.5}
\]

If \( S_n \geq b \), then \( T = N_1 \). Otherwise \( S_{N_1} = \min_{0 \leq k \leq N_1} S_k \) and we define

\[
N_1 = \inf\{n: n \geq 1, S_{N_1+n} - S_{N_1} \notin (0, b)\}.
\]

If \( S_{N_1+n} - S_{N_1} \geq b \), then \( T = N_1 + N_2 \). Otherwise \( S_{N_1+N_2} \leq S_{N_1} \), and \( S_{N_1+N_2} = \min_{0 \leq k \leq N_1+N_2} S_k \). In general let

\[
N_k = \inf\{n: n \geq 1, S_{N_1+\cdots+N_{k-1}+n} - S_{N_1+\cdots+N_{k-1}} \notin (0, b)\}. \tag{6.6}
\]

It is easy to see that

\[
T = N_1 + \cdots + N_M \tag{6.7}
\]

where

\[
M = \inf \{k: S_{N_1+\cdots+N_k} - S_{N_1+\cdots+N_{k-1}} \geq b\}. \tag{6.8}
\]

**6.2 Cusum test of AR(1) process**

In the case of a simple AR(1) process, we consider

\[
X_i = \theta X_{i-1} + \epsilon_i \quad i = 1, 2, \cdots
\]

and \( X_0 = 0, \epsilon_i \sim i.i.d. N(0, \sigma^2) \).

**6.2.1 The structural change**

We consider the autoregressive coefficient \( \beta \) is to be equal or more than 1 at the moment of \( \tau^* \). So we will get

1. \( 0 < \theta_1 = \theta_2 = \cdots = \theta_{\tau^*-1} < 1 \) is a stationary process.
2. \( 1 \leq \theta_{\tau^*} = \theta_{\tau^*+1} = \cdots \) is a unit root process or explosive process.

And we consider

\[
H_0: \theta = \theta_0 < 1 \quad \text{vs} \quad H_1: \theta = \theta_1 \geq 1
\]

When the null hypotheses is true, there is no change and \( T=\infty \). Otherwise, when alternative hypotheses is true, the structure was changed and \( T<\infty \).

We consider autoregressive coefficient \( \theta \) has three cases:

(a) The autoregressive coefficient \( \theta \) before the change or after the change is both
known in the whole process. Obviously, it is impossible.

(b) The autoregressive coefficient $\theta$ of the stationary process is known, but we do not know the autoregressive coefficient $\beta$, when the change happened.

(c) The autoregressive coefficient $\theta$ before or after the change is unknown in the whole process. Obviously, it is impossible. We consider it is the most difficult problem.

6.2.2 Stopping Time

In this paper, we consider the structural change is just from stationary to unit root process. So if we assume the process change at the time $t = k$, the AR(1) process will transform to two process

$$X_t = \begin{cases} \theta X_{t-1} + \epsilon_t & t = 1, 2, \cdots, k - 1 \\ X_{t-1} + \epsilon_t & t = k, k + 1, \cdots \end{cases}$$  \hspace{1cm} (6.9)$$

where $\epsilon_t \sim N(0, \sigma^2)$. And the testing hypotheses we consider are

$$H_0: \theta = 1 - \delta/m \hspace{1cm} vs \hspace{1cm} H_1: \theta = 1$$  \hspace{1cm} (6.10)$$

Because of $X_t = \theta X_{t-1} + \epsilon_t = (1 - \delta/m)X_{t-1} + \epsilon_t$, we can obtain that

$$X_t - X_{t-1} = -\frac{\delta}{m}X_{t-1} + \epsilon_t.$$  

and then divide the both side by $\sigma\sqrt{m}$, we will have

$$\frac{X_t - X_{t-1}}{\sigma\sqrt{m}} = -\frac{\delta}{m}\frac{X_{t-1}}{\sigma\sqrt{m}} + \frac{\epsilon_t}{\sigma\sqrt{m}}$$

Herr we consider $Y_t = X_t/\sigma\sqrt{m}$, follow the chapter 3.4, we can get

$$dY_t = -\delta Y_t dt + dW_t.$$  

It is Ornstein-Uhlenbeck process.

The time $t = 0, 1, \cdots$ transform to be $= 0, \frac{1}{m}, \frac{2}{m}, \cdots, \frac{k-1}{m}, \frac{k}{m}, \cdots$. The change point $k$ transform to $t^*$, and $t^* = \frac{k}{m}$. So we can get

$$dY_t = -\delta \mathbb{1}_{\{t < t^*\}} Y_t dt + dW_t $$  \hspace{1cm} (6.11)$$

In the case of a simple AR(1) model, we consider

$$X_t = \begin{cases} \theta X_{i-1} + \epsilon_i & i = 1, 2, \cdots, k - 1 \\ X_{i-1} + \epsilon_i & i = k, \cdots \end{cases}$$

and $X_0 = 0$, $\epsilon_i \sim i.i.d. N(0, \sigma^2)$. So we can get
\[ \varepsilon \sim f(\varepsilon) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{\varepsilon^2}{2\sigma^2} \right). \]

As we know,
\[ \varepsilon_i = \begin{cases} X_i - \theta X_{i-1} & i = 1, 2, \ldots, k - 1 \\ X_i - X_{i-1} & i = k, \ldots \end{cases}, \]
we will get
\[ X_i | X_{i-1} = x_{i-1} \sim f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{(X_i - \theta X_{i-1})^2}{2\sigma^2} \right), & i = 1, 2, \ldots, k \\ \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{(X_i - X_{i-1})^2}{2\sigma^2} \right), & i = k + 1, \ldots \end{cases}. \]

The testing hypotheses we consider are,

- \( H_\infty: \theta_1 = \theta_2 = \cdots = 1 - \frac{\delta}{m} \) i.e. no change occurs

- \( H_k: \theta_1 = \cdots = \theta_{k-1} = 1 - \frac{\delta}{m}, \theta_k = \theta_{k+1} = \cdots = 1 \) i.e. change occurs at time \( k \).

Here we consider the likelihood ratio \( L_0 = 1 \), and
\[ L(k; x_1, x_2, \ldots, x_n) = \frac{f_{H_k}}{f_{H_\infty}}, n \geq 1. \]

According to Markov Chain, we consider
\[ f(x_1, x_2, \ldots, x_n | x_0) = \prod_{i=1}^{n} f(x_i | x_{i-1}). \]

We can get
\[ L(k; x_1, x_2, \ldots, x_n) = \frac{f_{H_k}}{f_{H_\infty}} = \frac{\prod_{i=1}^{k-1} \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{(x_i - \theta x_{i-1})^2}{2\sigma^2} \right) \prod_{i=k}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{(x_i - x_{i-1})^2}{2\sigma^2} \right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{(x_i - \theta x_{i-1})^2}{2\sigma^2} \right)} = \frac{\prod_{i=k}^{n} \exp \left( -\frac{(x_i - x_{i-1})^2}{2\sigma^2} \right)}{\prod_{i=k}^{n} \exp \left( -\frac{(x_i - \theta x_{i-1})^2}{2\sigma^2} \right)} = \prod_{i=k}^{n} \exp \left\{ -\frac{(x_i - x_{i-1})^2}{2\sigma^2} + \frac{(x_i - x_{i-1} + (1-\theta) x_{i-1})^2}{2\sigma^2} \right\}. \]
\[ = \prod_{i=k}^{n} \exp \left\{ \frac{(1-\theta)x_{t-1}(x_i - x_{i-1}) + (1-\theta)x_i}{2\sigma^2} \right\} \]

\[ = \exp \left\{ \frac{(1-\theta)}{\sigma^2} \sum_{i=k}^{n} x_{t-1} (x_i - x_{i-1}) + \frac{(1-\theta)^2}{2\sigma^2} \sum_{i=k}^{n} x_i^2 \right\} \]  \hspace{1cm} (6.12)

The stopping time \( T_m \) and

\[ T_m = \inf \{ n \geq 1: max R(k; x_1, x_2, \ldots x_n) \geq c \} \]

\[ = \inf \left\{ n \geq 1: \max \left\{ \frac{(1-\theta)}{\sigma^2} \sum_{i=k}^{n} x_{t-1} (x_i - x_{i-1}) + \frac{(1-\theta)^2}{2\sigma^2} \sum_{i=k}^{n} x_i^2 \right\} \geq c \right\} \] \hspace{1cm} (6.13)

Substituting \( n = [mt], \theta = 1 - \frac{\delta}{m}, \) we can get

\[ T_m = \inf \left\{ mt \geq 1: \max \left\{ \frac{\delta}{\sigma^2 m} \sum_{i=k}^{[mt]} x_{t-1} (x_i - x_{i-1}) + \frac{\delta^2}{2\sigma^2 m^2} \sum_{i=k}^{[mt]} x_i^2 \right\} \geq c \right\} \] \hspace{1cm} (6.14)

and dividing the result by \( m \), the stopping time will transform to

\[ \tau = \frac{T_m}{m} = \inf \left\{ t \geq 0: \max \left\{ \frac{\delta}{\sigma^2 m} \sum_{i=k}^{[mt]} x_{t-1} (x_i - x_{i-1}) + \frac{\delta^2}{2\sigma^2 m^2} \sum_{i=k}^{[mt]} x_i^2 \right\} \geq c \right\} \] \hspace{1cm} (6.15)

By chapter 3.2.1, we will get the stopping time \( \frac{T_m}{m} \) and

\[ \tau = \frac{T_m}{m} \Rightarrow U_c = \inf \left\{ t \geq 0: \max \left\{ \delta \int_{t}^{t'} X_s dX_s + \frac{\delta^2}{2} \int_{t}^{t'} X_s^2 ds \right\} \geq c \right\} \] \hspace{1cm} (6.16)

6.2.3 Change at the beginning

If the process change at the time \( k=0, \)

\[ X_i = X_{i-1} + \varepsilon_i \]

\[ = \varepsilon_1 + \cdots + \varepsilon_n, \quad i = 1, 2, \ldots. \]

And

\[ \frac{X_{[mt]}}{\sqrt{m\sigma}} = \frac{\sum_{i=1}^{[mt]} \varepsilon_i}{\sqrt{m\sigma}} \rightarrow B_t \sim N(0, t), \]

here \( B_t \) is standard Brownian motion. The stopping time \( \tau \) will change to be

\[ \tau \Rightarrow U_c = \inf \left\{ t \geq 0: \max \left\{ \delta \int_{0}^{t} B_s dB_s + \frac{\delta^2}{2} \int_{0}^{t} B_s^2 ds \right\} \geq c \right\} \] \hspace{1cm} (6.17)

Suppose that \( \{ \mathcal{F}_t \} \) is the complete right continuous filtration generated by Brownian \( \{ B_t \} \). Put the continuous martingale \( M_t = \int_{0}^{t} X_s dB_s = \int_{0}^{t} B_s dB_s \), and its quadratic
variation process $< M >_t = \int_0^t X_s^2 \, ds = \int_0^t B_s^2 \, ds$. Consider stopping time $U_s = \inf\{ t \geq 0: < M >_t \geq s \} = \inf\{ t \geq 0: \int_0^t X_s^2 \, ds = \int_0^t B_s^2 \, ds \geq s \}$ define by $< M >$ and stopped martingale

$$B_s = M_{U_s}$$

and let $G_s = \mathcal{F}_{U_s}$ where $\mathcal{F}_{U_s} = \{ A \in \mathcal{F}_\infty: A \cap \{ U_s \leq t \} \in \mathcal{F}_t \text{ for all } t \}$. Since $< M >_\infty$ a.s., $B_s$ become the DDS Brownian motion with respect to $G_s$. See Revuz and Yor (1999). Hence, DDS implies

$$\int_0^{U_s} B_s \, dB_s = B_s. \quad (6.18)$$

Then the stopping time will change to be

$$\tau \Rightarrow U_c = \inf\{ t \geq 0: \max \left\{ \delta \int_0^{U_s} B_s \, dB_s + \frac{\delta^2}{2} \int_0^{U_s} B_s^2 \, ds \right\} \geq c \}$$

$$= \inf\{ t \geq 0: \max_{t' \leq t} \left\{ \delta M_{U_s} + \frac{\delta^2}{2} < M >_{U_s} \right\} \geq c \}. \quad (6.19)$$

We consider $\tau = < M >^{-1}_\nu = U_\nu$, and $\tau$ is

$$\tau \Rightarrow U_c = \inf\{ s \geq 0: \max_{s' \leq s} \left\{ \delta \left( B_s + \frac{\delta}{2} s \right) \right\} \geq c \}. \quad (6.20)$$

### 6.2.4 Change at the time $k$ ($k < \infty$)

Under $H_k$, the change occurs at the time $k$. Before the change occurs, $\theta = 1 - \frac{\delta}{m}$. And by subsection 3.4, we will get

$$dX_t = -\delta X_t \, dt + dW_t, \quad t < k.$$ And

$$\frac{X_{[mt]}}{\sqrt{m}\sigma} = \frac{\sum_{i=1}^{[mt]} \epsilon_i}{\sqrt{m}\sigma} \rightarrow B_t \sim N(0, t), \quad t > k,$$

here $B_t$ is standard Brownian motion. So

$$dX_t = dB_t$$

Suppose that $\{ \mathcal{F}_t \}$ is the complete right continuous filtration generated by Brownian $\{ W_t \}$. Put the continuous martingale $M_t = \int_0^t X_s \, dB_s = \int_0^t B_s \, dB_s$, and its quadratic variation process $< M >_t = \int_0^t X_s^2 \, ds = \int_0^t B_s^2 \, ds$. Consider stopping time $U_s = \inf\{ t \geq 0: < M >_t \geq s \} = \inf\{ t \geq 0: \int_0^t X_s^2 \, ds = \int_0^t B_s^2 \, ds \geq s \}$ define by $< M >$ and stopped martingale
\[ B_s = M_{U_s} \]

and let \( G_s = \mathcal{F}_{U_s} \) where \( \mathcal{F}_{U_s} = \{ A \in \mathcal{F}_\infty : A \cap \{ U_t \leq t \} \in \mathcal{F}_t \) for all \( t \). Since \( < M > \) a.s., \( B_s \) become the DDS Brownian motion with respect to \( G_s \). See Revuz and Yor (1999). Hence, DDS implies

\[ \int_0^{U_s} B_s dB_s = B_s. \]

The stopping time \( \tau \) will change to be

\[ \tau \Rightarrow U_c = \inf \left\{ t \geq 0 : \max \left\{ \delta \int_t^{t'} B_s \, dB_s + \frac{\delta^2}{2} \int_t^{t'} B_s^2 \, ds \right\} \geq c \right\} \]

\[ = \inf \left\{ t \geq 0 : \max_{t'} \left\{ \delta (M_t - M_{t'}) + \frac{\delta^2}{2} (< M >_t - < M >_{t'}) \right\} \geq c \right\} \quad (6.21) \]

\[ \tau \Rightarrow U_c = \inf \left\{ s \geq 0 : \max_{s' \leq s} \left\{ \delta \left( B_s + \frac{\delta}{2} s \right) - \delta \left( B_{s'} + \frac{\delta}{2} s' \right) \right\} \geq c \right\} \quad (6.22) \]

### 6.2.5 No change occurs

Under \( H_k \), the change occurs at the time \( k \). Before the change occurs, \( \theta = 1 - \frac{\delta}{m} \). And by subsection 3.4, we will get

\[ dX_t = -\delta X_t \, dt + dW_t. \]

Therefore, we can obtain that

\[ \int X_t \, dX_t + \frac{\delta}{2} \int X_t^2 \, dt = -\frac{\delta}{2} \int X_t^2 \, dt + \int X_t \, dW_t \]

Then combined with (6.19), the stopping time will change to be

\[ \tau \Rightarrow U_c = \inf \left\{ t \geq 0 : \max_{t'} \left\{ \delta \int_t^t X_s \, dW_s - \frac{\delta^2}{2} \int_t^t X_s^2 \, ds \right\} \geq c \right\}. \quad (6.23) \]

And combining with the DDS theorem, we will have

\[ \tau \Rightarrow U_c = \inf \left\{ s \geq 0 : \max_{s' \leq s} \left\{ \delta \left( B_s - \frac{\delta}{2} s \right) - \delta \left( B_{s'} - \frac{\delta}{2} s' \right) \right\} \geq c \right\}. \quad (6.25) \]
Chapter 7 Numerical Studies

We numerically evaluate the performance of the proposed test, namely the size, power, the stopping time and false alarm. In the SPRT, we need to test size, power and mean of the stopping time. But in the CUSUM test, it is different. We just need to test the mean of the stopping time and the false alarm.

To test what we want, we consider AR(1) model

\[ X_i = \theta X_{i-1} + \varepsilon_i, \quad \varepsilon_i \sim \text{iid} N(0, \sigma^2) \]

and the \( \theta_0 = 1 - \delta/m \) under \( H_0 \), \( \theta_1 = 1 \) under \( H_1 \). We pick \( \theta \) for several pre-fixed values to detect the relationship between \( \theta \) and the statistics that we want to test.

7.1 Sequential Unit Root Test

Table 1 presents the relationship between the \( \theta_1 \) and simulated reject probability. We pre-fixed the mean of the stopping time and \( s=1 \), so we can obtain \( m \) by \( m = E[T_{m1}]/2.0921 \). We will detect no matter with the change of the \( m \), the simulated rejection probability (RP) is always close to the normal size or limiting power. And the simulated mean of the stopping time is also close to the pre-fixed values. And following with the \( \theta_1 \) to be smaller, the simulated mean of the stopping time depart more from the pre-fixed values under the same \( m \).

Table 2 presents the same conclusion as table 1. Comparing with table 1 and table 2, we will detect that when \( s \) is smaller, the limiting power is getting bigger under the mean of the stopping time is the same pre-fixed value.

7.2 SPRT

Table 3 presents the relationship between the first (second) kind of errors and the size (power), and the relationship between \( \theta \) and the mean of stopping time under \( H_0 \) and \( H_1 \). We pick \( \theta_0 \) for several pre-fixed values and let \( \theta_1 = 1 \), \( \alpha = 0.05 \) and \( \beta = 0.25 \). By chapter 4, we know that size and power as closer to \( \alpha \) and \( (1 - \beta) \) as better. By Table 1 we can obtain that the size is to be big and deviate \( \alpha \) if the \( \theta_0 \) is getting bigger. Simultaneously, the power is getting smaller and close to \( (1 - \beta) \). And \( E_0(T) \) and \( E_1(T) \) is getting bigger. It is means that the stop occurs late. Because \( \theta_0 \) is close to the \( \theta_1 = 1 \), it is hard to be detect.

Table 4 presents if \( \theta_0, \theta_1 \) is fixed, following with the change of \( \alpha \) and \( (1 - \beta) \), what
will occur to the size, power and the mean of stopping time under $H_0$ and $H_1$. By table 2, we pre-fixed $\theta_0=0.95$, $\theta_1=1$. If we set $\alpha=0.05$ and pick $(1-\beta)$ for several pre-fixed values, and let $(1-\beta)$ is getting bigger. Then we can obtain that the power, and the mean of the stopping time is getting bigger. But the size is smaller and deviate the $\alpha$. Contrarily, if we fixed $(1-\beta)=0.75$ and pick $\alpha$ for several pre-fixed values, and let $\alpha$ is getting smaller. Then we can obtain that the size, the mean of the stopping time is getting bigger and the power is deviate the $(1-\beta)$ and getting smaller.

### 7.3 Cusum Test

Table 5 presents the relationship between cp (change point) and false alarm, $\text{E}[T - \text{cp}|T \geq \text{cp}]$. We fixed $\theta_0=0.98$, $\theta_1=1$ and boundary $c=1$. Following the cp to getting bigger, the false alarm is bigger and the $\text{E}[T - \text{cp}|T \geq \text{cp}]$ is smaller. It means that if the change occurs later then the stopping time occurs earlier.

Table 6 presents the relationship between $\theta_0$ and the mean, variance of stopping time, false alarm, $\text{E}[T - \text{cp}|T \geq \text{cp}]$. We fix $c=1$, $\text{cp}=10$. Following with $\theta_0$ getting bigger, the $\text{E}_{\text{cp}}[T]$, $\text{Var}(T)$ and $\text{E}[T - \text{cp}|T \geq \text{cp}]$ is bigger. But the false alarm is smaller.

### 7.4 Comparison of the Test

We want to compare the test by mean of stopping time under $H_0$ and $H_1$. Here we introduce the Shiryaev-Oberts test to compare with the SPRT and CUSUM test.

#### 7.4.1 Shiryayev-Roberts Rule

The basic Shiryaev-Roberts statistic and stopping time are respectively

$$ R_n = \sum_{k=1}^{n} \prod_{i=k}^{n} \frac{f_i(x_i|x_{i-1})}{f_0(x_i|x_{i-1})} $$

(6.1)

and

$$ T_C = \min \{n | R_n \geq C'\} $$

(6.2)

where C is a lower bound on the acceptable rate of stopping time. A generalization of this procedure is to define

$$ R_n = \sum_{k=1}^{n} \frac{f_{T=k}(x_1,\cdots,x_n|x_0)}{f_{T=\infty}(x_1,\cdots,x_n|x_0)} $$
and

\[ T_c = \min \{ n | R_n \geq C' \} \]

where \( T = \infty \) means there is no change and \( T = k \) means that change occurs at the \( k \)th time. \( f_{T=\infty} \) is the joint density of the observations when no change ever takes place and \( f_{T=k} \) is a joint density of the observations when \( T = k \) and the first \( k - 1 \) observations are distributed as they would be under the regime dictated by \( f_{T=\infty} \). The observations need not be independent.

By AR(1) model, we know that

\[ f(x_1, x_2, \ldots, x_n) = f_{T=K} \]

so that

\[ R_n = \sum_{k=1}^{n} \exp \left\{ \frac{(1-\theta)\sum_{i=k}^{n} (x_i - x_{i-1})}{\sigma^2} + \frac{(1-\theta)^2}{2\sigma^2} \sum_{t=k}^{n} x_i^2 \right\} \]

and

\[ T_c = \min \{ n | \sum_{k=1}^{n} \exp \left\{ \frac{(1-\theta)\sum_{i=k}^{n} (x_i - x_{i-1})}{\sigma^2} + \frac{(1-\theta)^2}{2\sigma^2} \sum_{t=k}^{n} x_i^2 \right\} \geq C' \} \]

\[ (6.3) \]

\[ (6.4) \]

\[ (6.5) \]

### 7.4.2 Comparison of Three Tests

Table 7 presents that which test has the smaller and bigger mean of stopping time. By table 5, we fix \( \alpha, \beta \) of SPRT to be 0.05, 0.25 and \( c \) of CUSUM test is equal to 1. Set the \( C' \) of Shiryayev-Roberts test is equal to 70. We can detect that all mean of stopping time of these tests are getting bigger following with the \( \theta_0 \) getting bigger. And on the situation of the same \( \theta_0 \), the \( E_0(T) \) and \( E_1(T) \) of CUSUM test is the biggest and \( E_0(T) \) and \( E_1(T) \) the Shiryayev-Roberts method is the smallest. It means that the CUSUM test is the latest to stop and the Shiryayev-Roberts method is the earliest to stop.

82
Chapter 8 Conclusion

In this paper, we newly develop sequential detection procedures for testing local-to-unity hypotheses. We consider diffusion approximations and derive the asymptotic results for OC's by using the Dambis-Dubins&Schwartz Brownian motion. Especially the Bessel processes with non-integer dimensions play an important role to represent the limit of the stopping times. The LAN property is shown for the Fisher-information-based stopping times, which does not hold for non-sequential cases. And we also use the DDS Brownian motion to prove that the error probability is the same as the conclusion of Wald' SPRT. And in this paper, we also want to find a method to get the change point under CUSUM test. But unfortunately, we failed. So we just can implement numerical computation in the chapter 7.

So the subject which we will continue to study is to find a way to get the false alarm rate of SPRT and the change point of CUSUM test.
References
3. David Siegmund, *Sequential Analysis*, Springer-Verlag
7. Keiji Nagai; Yoshihiko Nishiyama and KOhtaro Hitomi, Sequential Unit Root Test, unpublished
11. 笠原浩司「微分積分学」、サイエンス社
12. 伊藤清 「確率過程」、岩波書店
13. 伊藤清 「確率論」、岩波書店
Table 1 Sequential Unit Root

<table>
<thead>
<tr>
<th>s=1</th>
<th>E(Tm) under H0</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>m=E(Tm)/2.0921</td>
<td>9.5597</td>
<td>23.8994</td>
<td>47.7988</td>
<td>95.5977</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β=1</th>
<th>normal size</th>
<th>0.05</th>
<th>0.05</th>
<th>0.05</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulated RP</td>
<td>0.0504</td>
<td>0.0503</td>
<td>0.0491</td>
<td>0.0513</td>
<td></td>
</tr>
<tr>
<td>Simulated Tm</td>
<td>20.55</td>
<td>50.71</td>
<td>99.88</td>
<td>201.54</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β=0.99</th>
<th>limiting power</th>
<th>0.0607</th>
<th>0.0799</th>
<th>0.1216</th>
<th>0.2454</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulated RP</td>
<td>0.0627</td>
<td>0.0798</td>
<td>0.1173</td>
<td>0.2451</td>
<td></td>
</tr>
<tr>
<td>Simulated Tm (Tm)</td>
<td>20.9(20.82)</td>
<td>54.93(55.4)</td>
<td>121.46(122.87)</td>
<td>291.65(301.44)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β=0.95</th>
<th>limiting power</th>
<th>0.1216</th>
<th>0.3263</th>
<th>0.7718</th>
<th>0.9991</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulated RP</td>
<td>0.1247</td>
<td>0.3251</td>
<td>0.7742</td>
<td>0.9992</td>
<td></td>
</tr>
<tr>
<td>Simulated Tm (Tm)</td>
<td>24.21(24.57)</td>
<td>81.08(83.22)</td>
<td>251.01(258.27)</td>
<td>918.33(943.9)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β=0.9</th>
<th>limiting power</th>
<th>0.2454</th>
<th>0.7718</th>
<th>0.9991</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulated RP</td>
<td>0.2431</td>
<td>0.7662</td>
<td>0.999</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Simulated Tm (Tm)</td>
<td>28.57(30.15)</td>
<td>122(129.22)</td>
<td>445.97(471.95)</td>
<td>1750.14(1827.79)</td>
<td></td>
</tr>
</tbody>
</table>

(calculated value)

Table 2 SPRT 1

<table>
<thead>
<tr>
<th>m=500, n=2000, α=0.05, β=0.25, θ₁=1</th>
<th>θ₀</th>
<th>size</th>
<th>power</th>
<th>E₀[T]</th>
<th>E₁[T]</th>
<th>E₁[T]/E₀[T]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>0.1155</td>
<td>0.7325</td>
<td>115.9925</td>
<td>105.5955</td>
<td>0.910364894</td>
<td></td>
</tr>
<tr>
<td>0.98</td>
<td>0.0465</td>
<td>0.7805</td>
<td>87.5565</td>
<td>75.7135</td>
<td>0.864738769</td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>0.039</td>
<td>0.7805</td>
<td>49.8175</td>
<td>46.431</td>
<td>0.93202188</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.036</td>
<td>0.778</td>
<td>33.498</td>
<td>29.1885</td>
<td>0.871350528</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.0315</td>
<td>0.781</td>
<td>20.178</td>
<td>17.748</td>
<td>0.879571811</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.0375</td>
<td>0.805</td>
<td>15.046</td>
<td>13.555</td>
<td>0.900903895</td>
<td></td>
</tr>
</tbody>
</table>
Table 3 SPRT 2
m=500, n=2000, \( \theta_0=0.95, \theta_1=1 \)

\[
\begin{array}{|c|c|c|c|c|}
\hline
\alpha=0.05 & 1-\beta & size & E0[T] & power & E1[T] \\
\hline
0.8 & 0.0325 & 63.486 & 0.8205 & 51.4915 \\
0.75 & 0.0395 & 54.1255 & 0.7835 & 47.1095 \\
0.7 & 0.0395 & 42.4145 & 0.757 & 43.351 \\
0.65 & 0.0455 & 37.6105 & 0.6985 & 35.5595 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|c|c|}
\hline
1-\beta=0.75 & \alpha & size & E0[T] & power & E1[T] \\
\hline
0.045 & 0.032 & 55.999 & 0.784 & 47.693 \\
0.05 & 0.0395 & 54.1255 & 0.7835 & 47.1095 \\
0.055 & 0.043 & 52.9195 & 0.782 & 44.632 \\
0.06 & 0.048 & 49.8955 & 0.7685 & 44.337 \\
\hline
\end{array}
\]

Table 4  CUSUM Test 1
m=500, n=2000, c=1, \( \theta_0=0.98, \theta_1=1 \)

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
cp & maxtime & Ecp(T) & FalseAlarm & E[T-cp|T\geq cp] & Var(T) \\
\hline
1 & 490 & 71.037 & 0 & 70.037 & 3185.022 \\
5 & 439 & 68.8315 & 0 & 63.8315 & 2925.829 \\
10 & 422 & 73.455 & 7 & 63.67787 & 3358.367 \\
15 & 408 & 72.4545 & 52 & 58.98819 & 2862.584 \\
20 & 460 & 71.064 & 131 & 54.64312 & 2919.519 \\
nochange & - & 70.1155 & 0 & - & 3011.452 \\
\hline
\end{array}
\]
Table 5 CUSUM TEST 2

\[m=500, n=2000, c=1, cp=10, \beta_1=1\]

| \(\beta_0\) | maxtime | numfailse | Ecp(T) | FalseAlarm | \(E[T-cp|T>=cp]\) | Var(T) |
|-----------|---------|-----------|--------|------------|----------------|--------|
| 0.99      | 501     | 155       | 212.5645 | 1          | 202.5645       | 18906.44 |
| 0.98      | 501     | 20        | 120.8755 | 1          | 110.931        | 9081.272 |
| 0.95      | 442     | 0         | 54.742   | 7          | 44.89915       | 1750.46 |
| 0.9       | 175     | 0         | 31.9685  | 71         | 22.77709       | 475.3292 |
| 0.8       | 96      | 0         | 20.483   | 205        | 20.483         | 140.8401 |
| 0.7       | 79      | 0         | 16.8255  | 273        | 7.904459       | 76.89049 |

Table 6 Comparison of Three Tests 1

\[m=500, n=2000, \alpha=0.05, \beta=0.25, \theta_1=1, c=1, C'=70\]

<table>
<thead>
<tr>
<th>(\theta_0)</th>
<th>S-R (E_0[T])</th>
<th>SPRT (E_0[T])</th>
<th>(E_1[T])</th>
<th>Cusum (E_1[T])</th>
<th>(E_0[T])</th>
<th>(E_1[T])</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>61.14</td>
<td>61.087</td>
<td>115.9925</td>
<td>105.5955</td>
<td>214.242</td>
<td>218.186</td>
</tr>
<tr>
<td>0.98</td>
<td>48.847</td>
<td>49.5245</td>
<td>87.5565</td>
<td>75.7135</td>
<td>121.785</td>
<td>122.054</td>
</tr>
<tr>
<td>0.95</td>
<td>31.2165</td>
<td>30.945</td>
<td>49.8175</td>
<td>46.431</td>
<td>51.943</td>
<td>53.578</td>
</tr>
<tr>
<td>0.9</td>
<td>17.745</td>
<td>17.928</td>
<td>33.498</td>
<td>29.1885</td>
<td>29.361</td>
<td>28.601</td>
</tr>
<tr>
<td>0.8</td>
<td>10.681</td>
<td>8.275</td>
<td>20.178</td>
<td>17.748</td>
<td>16.7555</td>
<td>17.4765</td>
</tr>
<tr>
<td>0.7</td>
<td>6.111</td>
<td>6.0845</td>
<td>15.046</td>
<td>13.555</td>
<td>12.6095</td>
<td>12.178</td>
</tr>
</tbody>
</table>