N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES EXACTES
SIDI BEL ABBÈS

THESE DE DOCTORAT

Présentée par

AMINA ZEBLAH

Domaine : Mathématique informatique
Filière : Mathématique
Intitulé de la formation : Probabilités et Statistique appliquées

Intitulée

Inégalités exponentielles et théorèmes limites conditionnelles des suites de variables aléatoires, applications aux modèles statistiques

Soutenue le 10/03/2021

Devant le jury composé de :

Président :

Mr MECHAB Boubaker Professeur à l'université de S.B.A.

Directeur de thèse :

Mr BENAISSA Samir Professeur à l'université de S.B.A.

Examinateurs :

Mr CHOUAF Abdelhak M.C.A à l'université de S.B.A.

Mr BLOUHI Tayeb M.C.A à L'U.S.T.O. Oran .

Année universitaire : 2020/2021

Dedications

This thesis is dedicated to my parents,and my brothers

and my sister…

Thanks and appreciation to my uncle and my father,

AEK Zeblah…

And to all my family and best friends…

ZEBLAH AMINA

Acknowledgements

First, I want to thank ALLAH, the Almighty for giving us courage and will to accomplish this modest work.

I give my sincere thanks to the members of the jury :

*I thank my thesis supervisor Mr.*Samir Benaissa *for proposing this subject to me and for guiding and counselling me during the making of this thesis. I am grateful for his patience during all these years.*

I would also like to thank Mr. MECHAB BOUBAKER *for honoring me by presiding over the jury of this thesis. He always advised me and preciously help.*

I warmly thank Mr. CHOUAF ABDELHAK *for the trust he has shown me in being part of this jury, I would also like to thank Mr.*BLOUHI TAYEB *for the interest, he was kind enough to grant to my work by accepting to participate in the jury.*

I also wish to express my gratitude to all my teachers. I thank the members of the Stochastic Statistics and Processes Laboratory from the Djillali Liabès University of Sidi Bel Abbès.

Contents

Dedications **i**

Acknowledgements **ii**

Contents **iv**

List of Figures **v**

Publication and Communication **vi**

Abbreviation list **vii**

1 Chapter1 **4**

1.1 Time serie and basic concept on stochastic processes 6

1.1.1 Time Series and Example 6

1.2 Notions on stochastic processes 7

1.2.1 Stationarity 7

1.2.2 Autocovariance function and autocorrelation function 8

1.2.3 Linear model 9

1.2.4 AR process 9

1.2.5 Invertibility and Causality 10

1.2.6 Moments of an process AR(1) 11

1.2.7 Stationarity conditions 13

1.2.8 Moments of an AR(p)process 14

1.2.9 Condition of stationarity 14

1.3 Fitting the Data to the Model 15

1.3.1 Estimation for AR(p) process 16

1.3.2 MA process 18

1.4 The ARMA model(Mixed model) 19

1.4.1 Causality and Invertibility 20

1.4.2 Parameter Estimation 22

1.4.3 Method1(Method of Moments) . . . 23
1.4.4 Method 2(Maximum Likelihood Estimation) . . . 24
1.4.5 Method 3 (Least Squares Estimation) . . . 25
1.5 The ARIMA model . . . 26
1.6 ARCH/GARCH Models . . . 27
1.6.1 ARCH(1): Definition and Properties . . . 27
1.6.2 ARCH(q):[Definition and Properties] . . . 28
1.6.3 Generalized ARCH (GARCH) . . . 28
1.7 Simulate processes . . . 30
1.7.1 Simulations and Correlograms . . . 30
1.7.2 Simulate a MA process . . . 32
1.7.3 Simulation of an ARMA process . . . 34

2 Conditional independence, conditional central limit theorem and conditional association 37

2.1 Conditional independence of events . . . 38
2.2 Conditional independence of random variables . . . 38
2.3 Conditional Borel-Cantelli Lemma . . . 39
2.4 Conditional complete convergence . . . 40
2.5 Conditional central limit theorem . . . 40
2.6 Conditional association . . . 42

3 New tail probability inequalities for widely orthant dependent random variables sequence, application to hazard estimator 48

3.1 Introduction . . . 49
3.2 Some lemmas . . . 50
3.3 Main Results and Proofs . . . 54

4 New exponential probability inequality and complete convergence for conditional LNQD random variables sequence, application to AR(1)model general 62

4.1 Introduction . . . 63
4.2 Some lemmas . . . 64
4.3 Main Results and Proofs . . . 66
4.4 Applications to the results to AR(1) model . . . 71
4.4.1 The AR(1) model . . . 71

5 Exponential inequalities and conditional complete convergence for conditional extended acceptable random variables and application to AR(1) model 76

5.1 Introduction . . . 77
5.2 Some lemmas . . . 79
5.3 Conditionally complete convergence for conditional extended acceptable . . . 81
5.4 Applications to the results to AR(1) model . . . 82
5.4.1 The AR(1) model . . . 82

Bibliography 87

List of Figures

1.1 Yields of 70 consecutive batches from a chemical process. 6
1.2 Input Series. 7
1.3 Output Series. 7
1.4 Realizations of three autoregressive processes. 11
1.5 Conditional likelihood function of a generated ARCH(1) process with $n = 100$. The true parameter is $\alpha = 0.5$ 28
1.6 Kurtosis of a GARCH(1,1) process. The left axis shows the parameter β_1, the right α_1 29
1.7 Likelihood function of a generated GARCH(1,1) process with $n = 500$. The left axis shows the parameter β, the right α. The true parameters are $\omega = 0.1$, $\alpha = 0.1$and $\beta = 0.8$ 30
1.8 Realisation of AR(1) Model, with $\phi_1 = 0.6$ and Associated Correlogram 31
1.9 Realisation of AR(2) Model, with $\phi_1 = 0.666$, $\phi_2 = 0.333$. and Associated Correlogram 32
1.10 Realisation of MA(1) Model, with $\theta_1 = 0.6$ and Associated Correlogram 33
1.11 Realisation of MA(3) Model and Associated Correlogram 34
1.12 Realisation of an ARMA(1,1) Model, with $\alpha = 0.5$ and $\beta = -0.5$ 35
1.13 Correlogram of an ARMA(1,1) Model, with $\alpha = 0.5$ and $\beta = -0.5$ 35
1.14 Realisation of an ARMA(2,2) Model,with $\alpha_1 = 0.5, \alpha_2 = -0.25\beta_1 = 0.5\ and\ \beta_2 = -0.3$ 36
1.15 Correlogram of an ARMA(2,2) Model,with $\alpha_1 = 0.5, \alpha_2 = -0.25\beta_1 = 0.5 and \beta_2 = -0.3$ 36

Publication and Communication

Publications

Amina Zeblah, Samir Benaissa,(2020): New tail probability inequalities for widely orthant dependent random variables sequence, application to hazard estimator. International Journal of Applied Mathematics and Statistics, VL 59 NO. 1, 1-14.

International Communication:

1. Zeblah, A., Benaissa, S. "Probability tail for LNQD random variables of partial sums and application to AR(1)". International Conférence Financial Mathematics: Tools and Applications.University of Bejaia, 27-28 Octobre 2019.

2. Zeblah, A., Benaissa, S."New concentration inequality and complete convergence for recursive probability density estimator in the case of END samples". Onzième Rencontre d'Analyse Mathématiques et Applications RAMA 11. Université Djilali Liabes, Sidi Bel Abbes 21-24 novembre 2019.

National Communication:

1. Zeblah, A., Benaissa, S. "New exponenetial inequality for widely orthant dependentrandom variables application to recursive estimator". Journée Académique des Mathématiques Appliquées (JAMA19)
2. Zeblah, A., Benaissa, S. Journée Doctorales de la Facultés des Sciences Exactes (JDFSE 2019)

Abbreviation list

$ACVF$:	Auto covariance function.
ACF :	Auto correlation function.
$PACF$:	Partiel auto correlation function.
AR :	Autoregressive.
MA :	Moving average.
$ARMA$:	Autoregressive moving average.
$ARIMA$:	Autoregressive integrated moving average .
$ARCH$:	Autoregressive conditional heteroskedasticity.
$GARCH$:	Generilized autoregressive conditional heteroskedasticity .
$\mathcal{F}-WOD$:	Conditional widely orthant dependent.
$\mathcal{F}-WUOD$:	Conditional widely upper orthant dependent.
$\mathcal{F}-WLOD$:	Conditional widely lower orthant dependent.
$\mathcal{F}-END$:	Conditional extended negatively dependent.
$\mathcal{F}-NQD$:	Conditional negative quadrant dependent.
$\mathcal{F}-LEND$:	Conditional extended negatively dependent.
$\mathcal{F}-LNQD$:	Conditional linearly negative quadrant dependent.

General Introduction

The analysis of experimental data that have been observed at different points in time leads to new and unique problems in statistical modeling and inference. The obvious correlation introduced by the sampling of adjacent points in time can severely restrict the applicability of the many conventional statistical methods traditionally dependent on the assumption that these adjacent observations are independent and identically distributed. The systematic approach by which one goes about answering the mathematical and statistical questions posed by these time correlations is commonly referred to as time series analysis.

Time series modeling is a dynamic research area which has attracted attentions of researchers community over last few decades. The main aim of time series modeling is to carefully collect and rigorously study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series. This model is then used to generate future values for the series, i.e. to make forecasts. Time series forecasting thus can be termed as the act of predicting the future by understanding the past [36]. Due to the indispensable importance of time series forecasting in numerous practical fields such as business, economics, finance,science and engineering, etc. [55], [56], [43], proper care should be taken to fit an adequate model to the underlying time series. It is obvious that a successful time series forecasting depends on an appropriate model fitting. A lot of efforts have been done by researchers over many years for the development of efficient models to improve the forecasting accuracy. As a result, various important time series forecasting models have been evolved in literature.

A time series is non-deterministic in nature, i.e. we cannot predict with certainty what will occur in future. Generally a time series $\{\{X_t\}, t = 0, 1, 2, ...\}$ is assumed to follow certain probability model [13] which describes the joint distribution of the random variable X_t. The mathematical expression describing the probability structure of a time series is termed as a stochastic process [15]. Thus the sequence of observations of the series is actually a sample realization of the stochastic process that produced it.

A usual assumption is that the time series variables X_t are independent and identically distributed (i.i.d) following the normal distribution. However as mentioned in [13], an interesting point is that time series are in fact not exactly i.i.d; they follow more or less some regular pattern in long term. For example if the temperature today of a particular city is extremely high, then it can be reasonably presumed that tomorrow's temperature will also likely to be high. This is the reason

why time series forecasting using a proper technique, yields result close to the actual value.
In general models for time series data can have many forms and represent different stochastic processes. There are two widely used linear time series models in literature, viz. Autoregressive (AR) [4], [23], [15] and Moving Average (MA) [4], [15] models. Combining these two, the Autoregressive Moving Average (ARMA) [4], [23], [15] and Autoregressive Integrated Moving Average (ARIMA) [4],[15] models have been proposed in literature.The Autoregressive Fractionally Integrated Moving Average (ARFIMA) [35], [17] model generalizes ARMA and ARIMA models. For seasonal time series forecasting, a variation of ARIMA, the Seasonal Autoregressive Integrated Moving Average (SARIMA) [4], [15] model is used. ARIMA model and its different variations are based on the famous Box-Jenkins principle [4],[23], [15] and so these are also broadly known as the Box-Jenkins models.
The concept of independence for event systems or for collections of variables random concepts is among the main concepts in probability theory. There are many results for independent random variables. It can be said that such results form a nucleus of modern probability theory. Particularly in the 20th century, the emergence of stochastic models and dependent random variables was striking. The phenomena studied in physics, chemistry, biology, economics and reliability were main sources for these models. Thus, the theory of stochastic processes and fields randomness has emerged and evolved intensively. As a result, control of dependency between random variables has always been a topic of interest and concern for probabilists statisticians. Several ways of controlling this dependence have been introduced and this work concerns the notion of the association of random variables. The association and a few other concepts dependence were introduced in the 1960s. Lehman [24] introduced the concept of positive quadrant dependence between two random variables, as a result of Esary and Proschan Walkup [5] have generalized this notion and introduced the notion of association. Interest in these notions of dependence has come from models where monotonous transformations were considered.At first, this notion received little attention from the probability community. statistics, but interest has increased in recent years due to their applicability in the different engineering sciences.
There are many concepts of dependence between them. A notion of dependence is the so called extended negatively dependent (END) introduced by Liu [25], the random variables are said to be END if they are at the same time upper extended negatively dependent (UEND) and lower extended negatively dependent(LEND). The independent random variables and the NOD random variables are END, but the END random variables are much smaller than the independent random variables. Another notion of dependence is this one called widely orthant dependent (WOD) was defined by Wang, Wang and Gao [49], the random variables are said to be WOD when both are widely upper orthant dependent (WUOD) and widely lower orthant dependent (WLOD). WOD random variables are lower than NA random variables, NSD random variables, NOD random variables and END random variables.
Our aim is to give performant technical tools to Mathematicians or Statisticians which are interested in weak conditional dependence. Sometimes we will give more general results. In view of the important role played by conditioning and dependence in the models used to describe many situations in the applied sciences, the concepts and results in the aforementioned paper are extended herein to the case of randomly weighted sums of dependent random variables when a sequence of conditioning sigma-algebras is given. The dependence conditions imposed on the random variables (conditional negative quadrant dependence and conditional strong mixing) as well as the convergence results obtained are conditional relative to the conditioning sequence of

sigma-algebras.

A concrete example where conditional limit theorems are useful is the study of statistical inference for non-ergodic models as discussed in Basawa and Prakasa Rao [1] and Basawa and Scott [2]. For instance, if one wants to estimate the mean off spring θ for a Galton Watson branching process, the asymptotic properties of the maximum likelihood estimator depend on the set of non-extinction.

In the past few decades, a lot of efforts have been dedicated to prove limit theorems and statistics applications for dependent random variables, and large numbers of sharp and elegant results are available, for example, Newman [29] established the central limit theorem, Matula [28] derived the functional central limit theorem, Wang and Zhang [50] provided uniform rates of convergence in the central limit theorem, Wang et al[51] obtained some exponential inequalities, complete convergence and almost sure convergence. However, nothing is available for conditional LNQD random variables. Yuan and Wu [47] extended many results from negative association to asymptotically negative association, Yuan and Yang [44] extended many results from association conditional association, Yuan et al. [45] extended many results from negative association to conditional negative association, and these motivate our original interest in conditional LNQD.

This memoire contains five chapters, which are organized as follows: Chapter 1 gives an introduction to the basic concepts of time series modeling and the tools used on the stochastic processes, and in particular, we recall the definition of the processes AR(p), MA(q), ARMA (p,q), ARIMA(p,d,q), ARCH(q) and GARCH(p,q), etc. Chapter 2 is designed to discuss about the concept of conditional independence and conditional association for sequences of random variables. We discuss some stochastic inequalities and limit theorems for such sequences of random variables, etc. In Chapter 3 is devoted to prove new exponential inequality for a new case of dependence WOD for the distributions of sums of widely orthant dependent (WOD, in short) random variables. Using these inequality and obtain of complete convergence for kernel estimators of density and hazard functions, under some suitable conditions. This work was published in International Journal of International Journal of Applied Mathematics and Statistics. In the fourth chapter, We prove a new tail probability inequality for the distributions of sums of conditionally linearly negative quadrant dependent ($\mathcal{F}$-LNQD, in short) random variables, and obtain a result dealing with conditionally complete convergence of first-order autoregressive processes with identically distributed ($\mathcal{F}$-LNQD) innovations. The last chapter is devoted to prove new exponential inequality for the distributions of sums of conditional extended acceptable random variables. Using these inequality. The results are applied to the first-order autoregressive processes AR(1) model.

Chapter 1

Chapter1

Summary

1.1 Time serie and basic concept on stochastic processes 6

1.1.1 Time Series and Example 6

1.2 Notions on stochastic processes 7

1.2.1 Stationarity 7

1.2.2 Autocovariance function and autocorrelation function 8

1.2.3 Linear model 9

1.2.4 AR process 9

1.2.5 Invertibility and Causality 10

1.2.6 Moments of an process AR(1) 11

1.2.7 Stationarity conditions 13

1.2.8 Moments of an AR(p)process 14

1.2.9 Condition of stationarity 14

1.3 Fitting the Data to the Model 15

1.3.1 Estimation for AR(p) process 16

1.3.2 MA process 18

1.4 The ARMA model(Mixed model) 19

1.4.1 Causality and Invertibility 20

1.4.2 Parameter Estimation 22

1.4.3 Method1(Method of Moments) 23

1.4.4 Method 2(Maximum Likelihood Estimation) 24

1.4.5 Method 3 (Least Squares Estimation) 25

1.5 The ARIMA model 26

1.6 ARCH/GARCH Models 27
1.6.1 ARCH(1): Definition and Properties 27
1.6.2 ARCH(q):[Definition and Properties] 28
1.6.3 Generalized ARCH (GARCH) 28
1.7 Simulate processes 30
1.7.1 Simulations and Correlograms 30
1.7.2 Simulate a MA process 32
1.7.3 Simulation of an ARMA process 34

Introduction

In this chapter we introduce some basic ideas of time series analysis and stochastic processes. Of particular importance are the concepts of stationarity and the autocovariance and sample autocovariance functions.
The main objective of this chapter is examined, all the important destinations of the analysis of the autoregressive processes and we present to ourselves an integrated view of the explanation of the frequently used AR, MA, ARMA, ARIMA and ARCH/GARCH models after we quote properties and important results.

1.1 Time serie and basic concept on stochastic processes

1.1.1 Time Series and Example

The first definition clarifies the notion of time series analysis.

Definition 1.1.1 (Time Series) *Let* $y_t = \{..., y_{t-1}, y_t, y_{t+1}, ...\}$ *denote a sequence of random variables indexed by some time subscript* t. *Call such a sequence of random variables a time series. Some real examples of time series are (see Figure 1.1):*

– We analyze the series F data set in Box, Jenkins, and Reinsel, 1994. A plot of the 70 raw data points is shown below. shows a time series consisting of the yields from 70 consecutive batches of a chemical process.

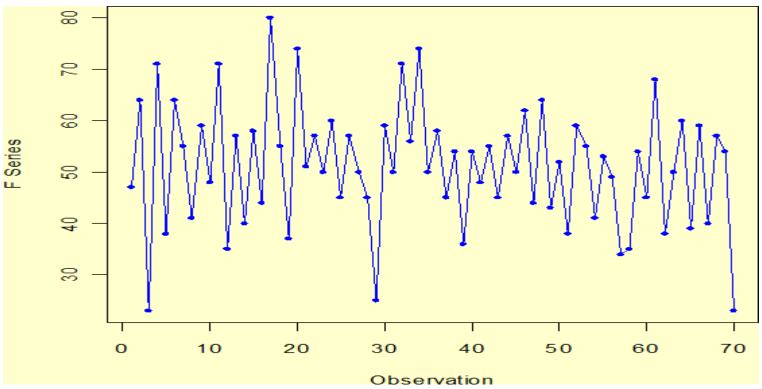

Figure 1.1: Yields of 70 consecutive batches from a chemical process.

– The gas furnace data from Box, Jenkins, and Reinsel, 1994 is used to illustrate the analysis of a bivariate time series. Inside the gas furnace, air and methane were combined in order to obtain a mixture of gases containing CO2 (carbon dioxide). The input series xt is the methane gas feedrate and the CO2 concentration is the output series y_t.
In this experiment 296 successive pairs of observations (x_t, y_t) were collected from continuous records at 9-second intervals.
The plots of the input and output series are displayed below.

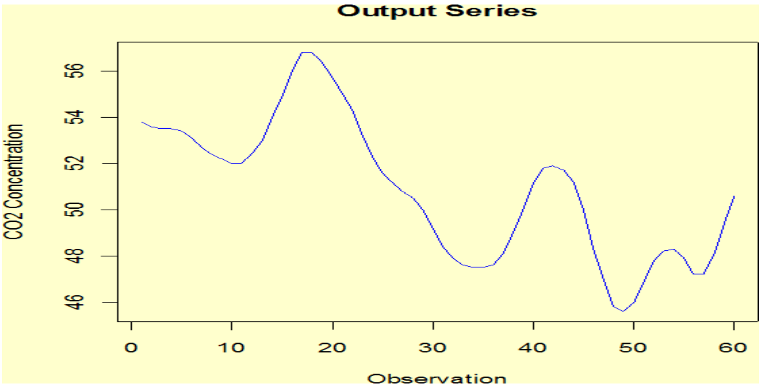

Figure 1.2: Input Series.

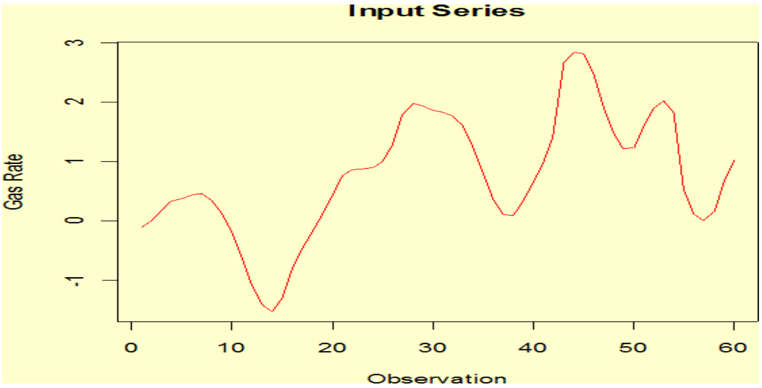

Figure 1.3: Output Series.

1.2 Notions on stochastic processes

The notion of a stochastic processes is very important both in mathematical theory and its applications in science, engineering, economics, etc. It is used to model a large number of various phenomena where the quantity of interest varies discretely or continuously through time in a non-predictable fashion.

Definition 1.2.1 *Let $\mathcal{T}$ be a subset of $[0, \infty]$. A family of random variables $\{X_t\}t \in \mathcal{T}$, indexed by $\mathcal{T}$, is called a* ***stochastic (or random) process****. When $\mathcal{T} = N(or \mathcal{T} = N_0)$, $\{X_t\}t \in \mathcal{T}$ is said to be a* ***discrete-time process****, and when $\mathcal{T} = [0, \infty]$, it is called a* ***continuous-time process****.*

1.2.1 Stationarity

Definition 1.2.2 (Joint Distribution) *The joint distribution function of $X_1, ..., X_T$ is given by*

$$F_{X_1,X_2,...,X_T}(X_1, X_2, ..., X_T) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_T)$$

Definition 1.2.3 (Strict Stationarity) *A process is said to be strictly stationary if the joint distribution of $X_1, X_2, ..., X_k$ is the same as the joint distribution of $X_{t+1}, X_{t+2}, ..., X_{t+k}$, evaluated at the same set of points $x_1, x_2, ..., x_k$, i.e.*

$$F_{X_1,X_2,...,X_k}(X_1, X_2, ..., X_k) = F_{X_{t+1},X_{t+2},...,X_{t+k}}(X_1, X_2, ..., X_k)$$

for all t and for all k.

Definition 1.2.4 (Wide Sense Stationarity) *A process is said to be second order (or wide sense) stationary if*

$$\mathbb{E}(X_t) = \mu \quad and \quad Var(X_t) = \sigma^2$$

or all t and, for all k,

$$Cov(X_t, X_{t+k}) = Cov(X_t, X_{t+1+k}) = Cov(X_t, X_{t+2+k}) + ...$$

is a function of the time lag k only and does not depend on time t.

1.2.2 Autocovariance function and autocorrelation function

the autocovariance is a function that gives the covariance of the process with itself at pairs of time points. Autocovariance is closely related to the autocorrelation of the process in question.

Definition 1.2.5 (ACF) *Similarly, we define so called autocorrelation function (ACF) as*

$$\phi(t) = \frac{\mu(T)}{\mu(0)} = Cor(X_{t+T}, X_t) \quad for\ all\ t, T.$$

Definition 1.2.6 (ACVF) *The sequence $\{\mu_k, k \in \mathbb{Z}\}$ is called the autocovariance function (ACVF),*

$$Cov(X_t, X_{t+k}) = \mu_k \quad for\ all\ t, k$$

Remark 1.2.1

1. *A strictly stationary process is weakly stationary.*
2. *If the process is Gaussian, that is $(X_{t_1}, ..., X_{t_k})$ is multivariate normal, for all $t_1, ..., t_k$, then weak stationarity implies strong stationarity.*
3. $\phi_0 = Var(X_t) > 0$, *assuming X_t is genuinely random.*
4. *By symmetry, $\phi_k = \phi_{-k}$, for all k.*

Definition 1.2.7 *Let $(\varepsilon_t)_{t\in\mathbf{Z}}$ a stochastic process, $(\varepsilon_t)_{t\in Z}$ is said to be a white noise if the following three properties are verified:*

a. $\mathbb{E}(\varepsilon_t) = 0 \quad \forall t \in \mathbf{Z}$

b. $Var(\varepsilon_t) = \sigma^2 \quad \forall t \in \mathbf{Z}$

c. $Cov(\varepsilon_t, \varepsilon_s) = \mathbb{E}(\varepsilon_t, \varepsilon_s) = 0 \quad for\ all\ t \neq s.$

Remark 1.2.2

- *If $\{\varepsilon_t, t \geq 0\}$ is a weak white noise, we will denote by $\varepsilon_t \sim WN(0, \sigma^2)$*
- *If $\{\varepsilon_t, t \geq 0\}$ is a strong white noise, we will denote by $\varepsilon_t \sim iid(0, \sigma^2)$*
- *If $\{\varepsilon_t, t \geq 0\}$ is a Gaussian white noise, we will denote by $\varepsilon_t \sim N(0, \sigma^2)$.*

1.2.3 Linear model

The linear process is a stochastic process formed by a linear combination (no necessarity finite) of strong white noises, and when they are weak the linear process is general.

Definition 1.2.8 *The time series X_t is a linear process if it has the representation*

$$X_t = \sum_{j=-\infty}^{\infty} \phi_j Z_{t-j}, \tag{1.1}$$

for all t, where $\{Z_t, WN(0,\sigma^2)\}$ and ϕ_j is a sequence of constants with $\sum_{j=-\infty}^{\infty} \phi_j < \infty$.

In terms of the backward shift operator B (1.1) can be written more compactly as

$$X_t = \phi(B)Z_t, \tag{1.2}$$

where $\phi(B) = \sum_{j=-\infty}^{\infty} \phi_j B^j$. A linear process is called a moving average or $MA(\infty)$ if $\phi_j = 0$ for all $j < 0$, i.e., if

$$X_t = \sum_{j=0}^{\infty} \phi_j Z_{t-j}$$

1.2.4 AR process

In many practical situations the value in an instant t of a time series can be written as the sum of a linear combination of previous values series and a term of white noise. Such a model is known as a process AR (AutoRegressive).

Definition 1.2.9 *An autoregressive process of order p is written as*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + Z_t, \tag{1.3}$$

where $\{Z_t\}$ is white noise, i.e., $\{Z_t\} \sim WN(0,\sigma^2)$, and $\{Z_t\}$is uncorrelated with X_s for each $s < t$.

Remark 1.2.3 *We assume(for simplicity of notation) that the mean of (X_t) is zero. If the mean is $\mathbb{E}(X_t) = \mu \neq 0$, then we replace X_t by $X_t - \mu$ to obtain*

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + ... + \phi_p(X_{t-p} - \mu) + Z_t$$

what can be written as

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + Z_t,$$

where

$$\alpha = \mu(1 - \phi_1 - ... - \phi_p)$$

Other ways of writing $AR(p)$ model use:

***Vector notation**: Denote*

$$\phi = (\phi_1, \phi_2, ..., \phi_p)^T$$

$$X_{t-1} = (X_{t-1}, X_{t-2}, ..., X_{t-p})^T$$

Then the formula (1.3) can be written as

$$X_t = \phi^T X_{t-1} + Z_t$$

***Backshift operator**: Namely, writing the model (1.3) in the form*

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - ... - \phi_p X_{t-p} = Z_t,$$

and applying $BX_t = X_{t-1}$ we get

$$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)X_t = Z_t$$

or, using the concise notation we write

$$\phi(B)X_t = Z_t, \tag{1.4}$$

*where $\phi(B)$ denotes the **autoregressive operator***

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p.$$

Then the $AR(p)$ can be viewed as a solution to the equation(1.4), i.e.,

$$X_t = \frac{1}{\phi(B)} Z_t. \tag{1.5}$$

Corollary 1.2.1 *says that an infinite combination of white noice variables is a stationary process. Here, due to the recursive of the TS we can write AR(1) in such a from. Namely*

$$\begin{aligned} X_t &= \phi X_{t-1} + Z_t \\ &= \phi(\phi X_{t-2} + Z_{t-1}) + Z_t \\ &= \phi^2 X_{t-2} + \phi Z_{t-1} + Z_t \\ &. \\ &. \\ &. \\ &= \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j Z_{t-j}. \end{aligned}$$

This can be rewritten as

$$\phi^k X_{t-k} = X_t - \sum_{j=0}^{k-1} \phi^j Z_{t-j}$$

1.2.5 Invertibility and Causality

Definition 1.2.10 (Invertibility) *A linear process $\{X_t\}$ is invertible (strictly, an invertible function of $\{W_t\}$) if there is a*

$$\pi(B) = \pi_0 + \pi_1 B + \pi_2 B + ...$$

with $\sum_{j=0}^{\infty} |\pi_j| < \infty$
and $W_t = \pi(B)X_t$.

Theoreme 1.2.1 (Causality) *A (unique) stationary solution to $\phi(B)X_t = W_t$ exists if the roots of $\phi(z)$ avoid the unit circle:*

$$|z| = 1 \Longrightarrow \phi(z) = 1 - \phi_1 z - ... - \phi_p z^p \neq 0,$$

This AR(p) process is causal iff the roots of $\phi(z)$ are outside the unit circle:

$$|z| \leq 1 \Longrightarrow \phi(z) = 1 - \phi_1 z - ... - \phi_p z^p \neq 0,$$

Example 1.2.1 (AR Processes) *Figure 1.4 displays two AR(1) processes with respective parameters $\phi_1 = -0.9$ (left) and $\phi_1 = 0.8$ (middle) as well as an AR(2) process with parameters $\phi_1 = -0.5$ and $\phi_2 = 0.3$.*

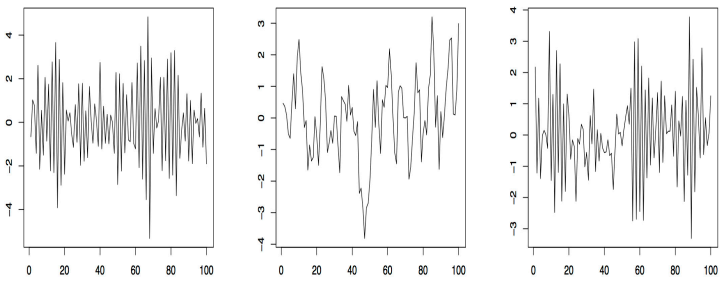

Figure 1.4: Realizations of three autoregressive processes.

1.2.6 Moments of an process AR(1)

To calculate the different moments of an AR process (1), namely its hope, variance, self-variance and self-corrosion, one will assume that white noises are independently and similarly distributed, of zero expectancy and variance σ^2 that we note ($\epsilon_i \sim iid(0, \sigma^2)$).

Expectation:

$$\mathrm{E}[X_t] = \varphi^t X_0 + c\sum_{i=0}^{t-1} \varphi^i$$

Proof (reasoning by recurrence:)

- $P(0)$ (initialization):
  $\mathrm{E}[X_0] = X_0$, because X_0 is deterministic. The expression is:

$$\varphi^0 X_0 + c\sum_{i=0}^{-1} \varphi^i = 1X_0 + 0 = X_0$$

- $P(t+1)$ (heredity):
  $\mathrm{E}[X_{t+1}] = \mathrm{E}[c + \varphi X_t + \epsilon_t]$,
  Since E is a linear operator:
  $\mathrm{E}[X_{t+1}] = c + \varphi\,\mathrm{E}[X_t]$

With the induction hypothesis:

$$\mathrm{E}[X_{t+1}] = c + \varphi(\varphi^t X_0 + c\sum_{i=0}^{t-1}\varphi^i)$$

$$\mathrm{E}[X_{t+1}] = c + \varphi^{t+1} X_0 + c\sum_{i=0}^{t-1}\varphi^{i+1},$$

By a change of variables in the sum, $i \to i-1$:

$$\mathrm{E}[X_{t+1}] = \varphi^{t+1} X_0 + c + c\sum_{i=1}^{t}\varphi^{i},$$

And, with $c = c\sum_{i=0}^{0}\varphi^i$:

$$\mathrm{E}[X_{t+1}] = \varphi^{t+1} X_0 + c\sum_{i=0}^{t}\varphi^{i}$$

Variance:

$$\mathrm{Var}[X_t] = \sum_{i=0}^{\infty}\varphi^{2i}\sigma^2$$

Proof

$$\begin{aligned}
\mathrm{Var}[X_t] &= \mathrm{E}\left[(X_t - \mathrm{E}[X_t])^2\right] \\
&= \mathrm{E}\left[\left(c\sum_{i=0}^{\infty}\varphi^i + \sum_{i=0}^{\infty}\varphi^i\varepsilon_{t-i} - c\sum_{i=0}^{\infty}\varphi^i\right)^2\right]
\end{aligned}$$

According to the results obtained in the previous page

$$\begin{aligned}
&= \mathrm{E}\left[\left(\sum_{i=0}^{\infty}\varphi^i\varepsilon_{t-i}\right)^2\right] \\
&= \mathrm{Var}\left[\sum_{i=0}^{\infty}\varphi^i\varepsilon_{t-i}\right] \quad \text{because } \mathrm{E}\left(X^2\right) = \mathrm{Var}(X) + \mathrm{E}(X)^2 \quad \text{and}
\end{aligned}$$

$$\mathrm{E}\left[\sum_{i=0}^{\infty}\varphi^i\varepsilon_{t-i}\right] = \sum_{i=0}^{\infty}\varphi^i\,\mathrm{E}[\varepsilon_{t-i}] = 0$$

and by hypothesis $\mathrm{E}[\varepsilon_t] = 0$

$$\begin{aligned}
&= \sum_{i=0}^{\infty}\mathrm{Var}\left[\varphi^i\varepsilon_{t-i}\right] \quad \text{by independence} \quad \varepsilon_t \\
&= \sum_{i=0}^{\infty}\varphi^{2i}\,\mathrm{Var}[\varepsilon_{t-i}] \quad \text{and}\, \mathrm{Var}[aX] = a^2\,\mathrm{Var}[X] \\
&= \sum_{i=0}^{\infty}\varphi^{2i}\sigma^2
\end{aligned}$$

Autocovariance: $\mathrm{Cov}[X_t, X_{t-j}] = \varphi^j\sum_{i=0}^{\infty}\varphi^{2i}\sigma^2$

Proof

$$
\begin{aligned}
\operatorname{Cov}[X_t, X_{t-j}] &= \mathrm{E}\left[(X_t - \mathrm{E}[X_t])(X_{t-j} - \mathrm{E}[X_{t-j}])\right] \\
&= \mathrm{E}\left[(\sum_{i=0}^{\infty} \varphi^i \varepsilon_{t-i})(\sum_{k=0}^{\infty} \varphi^k \varepsilon_{t-k-j})\right] \\
&= \mathrm{E}\left[\sum_{i=0}^{\infty}\sum_{k=0}^{\infty} \varphi^{i+k} \varepsilon_{t-i}\varepsilon_{t-k-j}\right] \\
&= \sum_{i=0}^{\infty} \sum_{k=0,\ k+j\neq i}^{\infty} \varphi^{i+k}\, \mathrm{E}\left[\varepsilon_{t-i}\varepsilon_{t-k-j}\right] + \sum_{k=0}^{\infty} \varphi^{2k+j}\, \mathrm{E}\left[\varepsilon_{t-k-j}^2\right] \\
&= \sum_{k=0}^{\infty} \varphi^{2k+j} \operatorname{Var}[\varepsilon_{t-k-j}] \quad \text{by the assumption of independence} \quad \varepsilon_l, \\
&\qquad\qquad \mathrm{E}\left[\varepsilon_{t-i}\varepsilon_{t-k-j}\right] = \mathrm{E}[\varepsilon_{t-i}]\,\mathrm{E}[\varepsilon_{t-k-j}] = 0, \\
&\qquad\qquad \text{and } \mathrm{E}\left[\varepsilon_{t-k-j}^2\right] = \operatorname{Var}[\varepsilon_{t-k-j}] + \mathrm{E}[\varepsilon_{t-k-j}]^2 = \operatorname{Var}[\varepsilon_{t-k-j}] \\
&= \varphi^j \sum_{i=0}^{\infty} \varphi^{2i}\sigma^2
\end{aligned}
$$

Autocorrélation

$$
\operatorname{Corr}[X_t, X_{t-j}] \equiv \frac{\operatorname{Cov}[X_t, X_{t-j}]}{\sqrt{\operatorname{Var}(X_t)\operatorname{Var}(X_{t-j})}} = \varphi^j \sqrt{\frac{1-\varphi^{2(t-j)+2}}{1-\varphi^{2t+2}}}
$$

1.2.7 Stationarity conditions

The φ setting determines whether the $AR(1)$ process is stationary or not:

$$
|\varphi| = \begin{cases} < 1 & \text{the process is stationary.} \\ = 1 & \text{Marche aléatoire : the process is therefore non-stationary.} \\ > 1 & \text{the process is explosive.} \end{cases}
$$

- $\varphi < 1$

  The following results come from the fact that if q_1 then the geometric series $\sum_{n=0}^{\infty} aq^n = \frac{a}{1-q}$.

  if $|\varphi| < 1$:

$$
\mathrm{E}[X_t] = \frac{c}{1-\varphi}
$$

$$
\operatorname{Var}[X_t] = \frac{\sigma^2}{1-\varphi^2}
$$

$$
\operatorname{Cov}[X_t, X_{t-j}] = \frac{\varphi^j}{1-\varphi^2}\sigma^2
$$

$$\mathrm{Corr}[X_t, X_{t-j}] = \varphi^j$$

We can see that the function of autocovariance decreases with a rate of $\tau - 1/\ln(\varphi)$. We see here that hope and variance are constant and that autocovariance does not depend on time: the process is therefore stationary.

- $\varphi = 1$

When $\varphi = 1$, the process is written : $X_t = c + X_{t-1} + \varepsilon_t$ and, therefore, considering the $X_t = ct + X_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i}$ contrary to before $t_0 = 0$,

if $|\varphi| = 1$:

$$\mathrm{E}[X_t] = ct + \mathrm{E}[X_0],$$
$$\mathrm{Var}[X_t] = t\sigma^2,$$

$$\mathrm{Cov}[X_t, X_{t-j}] = (t-j)\sigma^2,$$

1.2.8 Moments of an AR(p)process

In general, an AR(p) process is a process that depends on the p previous values:

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + ... + \varphi_p X_{t-p} + \varepsilon_t.$$

or ε_t is a white noise.

The different moments in a stationary process (see next section) are

$\mathrm{E}(X_t) = \frac{c}{1-\varphi_1-\varphi_2-...-\varphi_p}$
$\mathrm{Var}(X_t) = \varphi_1\gamma_1 + \varphi_2\gamma_2 + ... + \varphi_p\gamma_p + \sigma^2$
$\mathrm{Cov}(X_t, X_{t-j}) = \varphi_1\gamma_{j-1} + \varphi_2\gamma_{j-2} + ... + \varphi_p\gamma_{j-p}$
The formulas of variance and covariance correspond to the so-called Yule and walker equations (see below).

1.2.9 Condition of stationarity

Theoreme 1.2.2 *An AR process (p) is stationary if the solution module (roots) of its characteristic equation is strictly greater than* 1 *in absolute terms each time.*

The condition is frequently formulated differently, according to which the roots must be outside the unitary complex circle.

Example 1.2.2 *The polynomial of the backshift of a process* $AR(1)$ $X_t = \varphi X_{t-1} + \varepsilon_t$ *is written:* $(1 - \varphi L)X_t = \varepsilon_t$
Its resolution (replacing the Backshift operator B by the simple value x) gives $1 - \varphi x = 0 \Rightarrow x = \frac{1}{\varphi}$.
The condition that the solution is greater than 1 is equivalent to $|\frac{1}{\varphi}| > 1 \Rightarrow |\varphi| < 1$

Example 1.2.3 *The characteristic polynomial of the backshift of a process* $AR(2)$ $X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varepsilon_t$ *is written:* $(1 - \varphi_1 L - \varphi_2 L^2)X_t = \varepsilon_t$. *The resolution of the characteristic equation of the second degree* $(1 - \varphi_1 x - \varphi_2 x^2)$ *leads to the following conditions:*

- $\varphi_1 + \varphi_2 < 1$
- $\varphi_1 - \varphi_2 < 1$
- $|\varphi_2| < 1$

Example 1.2.4 *The characteristic polynomial of the backshift of a process AR(p)* $X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + ... + \varphi_p X_{t-p} + \varepsilon_t$ *is written:* $(1 - \varphi_1 L - \varphi_2 L^2 - ... - \varphi_p L^p)X_t = \varepsilon_t$.
The resolution of the characteristic equation $(1 - \varphi_1 x - \varphi_2 x^2 - ... - \varphi_p x^p)$ *leads to the following necessary (but not sufficient) conditions:*

- $\varphi_1 + \varphi_2 + ... + \varphi_p < 1$
- $|\varphi_p| < 1$

1.3 Fitting the Data to the Model

We will focus on the estimation of the autoregressive parameter of the first order with Gaussian innovations. We consider the model

$$X_k = \rho X_{k-i} + \varepsilon_k, \quad k = 1, 2, ...$$

This will be accomplished using the least squares estimation. The error term is ε_k . We want to minimize the sum of the square of errors for our observed values with respect to ρ. We take the derivative of the sum of squares to get

$$\frac{\partial}{\partial_\rho} \sum_{k=1}^{n} (X_k - \rho X_{k-1})^2 = 2 \sum_{k=1}^{n} (X_k - \rho X_{k-1})(-X_{k-1})$$

We set the derivative equal to 0 and obtain the following:

$$\begin{aligned} 2\sum_{k=1}^{n}(X_k - \rho X_{k-1})(-X_{k-1}) &= 0 \\ \sum_{k=1}^{n}(-X_k X_{k-1} + \rho X_{k-1}^2) &= 0 \\ -\sum_{k=1}^{n} X_k X_{k-1} + \rho \sum_{k=1}^{n} X_{k-1}^2 &= 0 \\ \rho \sum_{k=1}^{n} \rho X_{k-1}^2 &= \sum_{k=1}^{n} X_k X_{k-1} \end{aligned} \tag{1.6}$$

Hence, we obtain our least squares estimator for ρ :

$$\hat{\rho} = \frac{\sum_{k=1}^{n} X_k X_{k-1}}{\sum_{k=1}^{n} X_{k-1}^2} \tag{1.7}$$

1.3.1 Estimation for AR(p) process

Yule Walker equations

The Yule-Walker equations, named for Udny Yule and Gilbert Walker,[48], [52] are the following set of equations.[42]

$$\gamma_m = \sum_{k=1}^{p} \varphi_k \gamma_{m-k} + \sigma_\varepsilon^2 \delta_{m,0},$$

where $m = 0, ..., P$, yielding $p+1$ equations. Here γ_m is the autocovariance function $X_t \sigma_\varepsilon$ is the standard deviation of the input noise process, and $\delta_{m,0}$ is the kronecher delta function.
Because the last part of an individual is non-zero only if $m = 0$, the set of equations can be solved by representing the equations for $m > 0$ in matrix form, thus getting the equation

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots \\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \vdots \\ \varphi_p \end{bmatrix}$$

which can be solved for all $\{\varphi_m; m = 1, 2, \cdots, p\}$. The remaining equation for $m = 0$ is

$$\gamma_0 = \sum_{k=1}^{p} \varphi_k \gamma_{-k} + \sigma_\varepsilon^2,$$

which, once $\{\varphi_m; m = 1, 2, \cdots, p\}$ are known, can be solved for σ_ε^2.

An alternative formulation is in terms of the autocorrelation function. The AR parameters are determined by the first $p+1$ elements $\rho(\tau)$ of the autocorrelation function.The full autocorrelation function can then be derived by recursively calculating

$$\rho(\tau) = \sum_{k=1}^{p} \varphi_k \rho(k - \tau)$$

Examples for some Low-order AR(p) processes

$$\gamma_1 = \varphi_1 \gamma_0 \gamma_1 = \varphi_1 \gamma_0$$

- $p = 1$

$$\gamma_1 = \varphi_1 \gamma_0$$

Hence

$$\rho_1 = \gamma_1 / \gamma_0 = \varphi_1$$

- $p = 2$

- The-Yule Walker equations for an AR(2) process are

$$\gamma_1 = \varphi_1\gamma_0 + \varphi_2\gamma_{-1}$$

$$\gamma_2 = \varphi_1\gamma_1 + \varphi_2\gamma_0$$

- Using the first equation yields $\rho_1 = \gamma_1/\gamma_0 = \dfrac{\varphi_1}{1-\varphi_2}$

- Using the recursion formula yields $\rho_2 = \gamma_2/\gamma_0 = \dfrac{\varphi_1^2 - \varphi_2^2 + \varphi_2}{1-\varphi_2}$

Example 1.3.1 *[AR(1)]*
For an $AR(1)$ process, one has:

$$\gamma_j = \varphi\gamma_{j-1} \qquad \forall j = 1, \ldots, p$$

We notice that we quickly find, with $j = 1$, the result obtained above:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \varphi$$

$\mathrm{Var}[X_t] = \dfrac{\sigma^2}{1-\varphi^2}$ *taking the extra equation to $\gamma_0 = \varphi\gamma_1 + \sigma_\varepsilon^2$, which then becomes* $\gamma_0 = \varphi\gamma_0\varphi + \sigma_\varepsilon^2 = \varphi^2\gamma_0 + \sigma_\varepsilon^2 \Rightarrow (1-\varphi^2)\gamma_0 = \sigma^2 \Rightarrow \gamma_0 = \dfrac{\sigma^2}{1-\varphi^2}$.

Example 1.3.2 *[AR(p)]*

$$\begin{cases} \gamma_1 = \varphi_1\gamma_0 + \varphi_2\gamma_{-1} + \ldots + \varphi_p\gamma_{-(p-1)} \\ \gamma_2 = \varphi_1\gamma_1 + \varphi_2\gamma_0 + \ldots + \varphi_p\gamma_{-(p-2)} \\ \vdots \\ \gamma_p = \varphi_1\gamma_{p-1} + \varphi_2\gamma_{p-2} + \ldots + \varphi_p\gamma_0 \end{cases}$$

That can be written in matrix form:

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots \\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \vdots \end{bmatrix}$$

Proof: The defining equation of the AR process is

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t.$$

By multiplying the two members by X_{t-j} and taking hope, one obtains

$$\mathbb{E}[X_t X_{t-j}] = \mathbb{E}\left[\sum_{i=1}^{p} \varphi_i X_{t-i} X_{t-j}\right] + \mathbb{E}[\varepsilon_t X_{t-j}].$$

However, it turns out that $\mathbb{E}[X_tX_{t-j}] = \gamma_j + \mathbb{E}[X_t]\mathbb{E}[X_{t-j}]$.
In the event that X the process is considered to be zero-average $(c = 0), \mathbb{E}[X_tX_{t-j}]$ comes down to the self-correlation function. The terms of white noise are independent of each other and, moreover, X_{t-j} is independent of ε_t where j is larger than zero. For $j > 0, \mathbb{E}[\varepsilon_t X_{t-j}] = 0$. for $j = 0$,

$$\begin{aligned}
\mathbb{E}[\varepsilon_t X_t] &= \mathbb{E}\left[\varepsilon_t\left(\sum_{i=1}^{p}\varphi_i, X_{t-i} + \varepsilon_t\right)\right] \qquad (1.8)\\
&= \sum_{i=1}^{p}\varphi_i, \mathbb{E}[\varepsilon_t, X_{t-i}] + \mathbb{E}[\varepsilon_t^2] = 0 + \sigma_\varepsilon^2, \mathbb{E}[\varepsilon_t X_t]\\
&= \mathbb{E}\left[\varepsilon_t\left(\sum_{i=1}^{p}\varphi_i, X_{t-i} + \varepsilon_t\right)\right]\\
&= \sum_{i=1}^{p}\varphi_i, E[\varepsilon_t, X_{t-i}] + E[\varepsilon_t^2] = 0 + \sigma_\varepsilon^2,
\end{aligned}$$

Now we have got for $j \leq 0$,

$$\gamma_j = E\left[\sum_{i=1}^{p}\varphi_i\, X_{t-i}X_{t-j}\right] + \sigma_\varepsilon^2\delta_j.$$

On the other hand,

$$E\left[\sum_{i=1}^{p}\varphi_i\, X_{t-i}X_{t-j}\right] = \sum_{i=1}^{p}\varphi_i\, E[X_tX_{t-j+i}] = \sum_{i=1}^{p}\varphi_i\,\gamma_{j-i},$$

that gives the equations of Yule-Walker

$$\gamma_j = \sum_{i=1}^{p}\varphi_i\gamma_{j-i} + \sigma_\varepsilon^2\delta_j.$$

for $j < 0$,

$$\gamma_j = \gamma_{-j} = \sum_{i=1}^{p}\varphi_i\gamma_{|j|-i} + \sigma_\varepsilon^2\delta_j.$$

1.3.2 MA process

Definition 1.3.1 *The notation $MA(q)$ refers to the moving average model of order q:*

$$X_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$$

where μ is the mean of the series, the $\theta_1, \cdots, \theta_q$ are the parameters of the model and the $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ are white noise error terms. The value of q is called the order of the MA model. This can be equivalently written in terms of the backshift operator B as

$$X_t = \mu + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t.$$

Proposition 1.3.1 *The mean of an MA(q) process is μ.*

Proof:

$$\mathbb{E}[X_t] = \mu + \mathbb{E}[\varepsilon_t] + \theta_1\mathbb{E}[\varepsilon_{t-1}] + \cdots + \theta_q\mathbb{E}[\varepsilon_{t-q}] = \mu + 0 + \theta_1.0 + \cdots + \theta_q.0 = \mu$$

Proposition 1.3.2 *The variance of an MA(q) process is*

$$Var(X_t) = \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2)$$

Proof

$$\begin{aligned} Var(X_t) &= 0 + Var(\varepsilon_t) + \theta_1^2 Var(\varepsilon_{t-1}) + \cdots + \theta_q^2 Var(\varepsilon_{t-k}) \\ &= \sigma^2 + \theta_1^2\sigma^2 + \theta_q^2\sigma^2 \\ &= \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2) \end{aligned}$$

Proposition 1.3.3 *The autocorrelation function of an $MA(1)$ process is*

$$\rho_1 \frac{\theta_1}{1+\theta_1^2} \ \rho_h = 0 \ for \ h > 1$$

Proof

$$\begin{aligned} Cov(X_i, X_{i-h}) &= \mathbb{E}[(X_i - \mu)(X_{i-h} - \mu)] \\ &= \mathbb{E}[(\varepsilon_i + \theta_1\varepsilon_{i-1})(\varepsilon_{i-h} + \theta_1\varepsilon_{i-h-1})] \\ &= \mathbb{E}[\varepsilon_i + \varepsilon_{i-h}] + \theta_1\mathbb{E}[\varepsilon_{i-1} + \varepsilon_{i-h}] + \theta_1\mathbb{E}[\varepsilon_i + \varepsilon_{i-h-1}] + \theta_1^2\mathbb{E}[\varepsilon_{i-1} + \varepsilon_{i-h-1}] \end{aligned}$$

when $h = 1$

$$Cov(X_i, X_{i-h}) = \theta_1\mathbb{E}[\varepsilon_{i-1} + \varepsilon_{i-1}] = \theta_1\sigma^2$$

since $\mathbb{E}[\varepsilon_{i-1}] = 0. \quad When \quad h > 1$

$$Cov(X_i, X_{i-h}) = 0$$

Thus for $h = 1$, by Proppsition 2

$$\rho_1 = \frac{Cov(X_i, X_{i-h})}{Var(X_i)} = \frac{\theta_1\sigma^2}{\sigma^2(1+\theta_1^2)} = \frac{\theta_1}{1+\theta_1^2}$$

and for $h > 1$

$$\rho_1 = \frac{Cov(X_i, X_{i-h})}{Var(X_i)} = \frac{0}{Var(X_i)} = 0$$

1.4 The ARMA model(Mixed model)

In the statistical analysis of time series, autoregressive moving-average (ARMA) models provide a parsimonious description of a (weakly) stationary stochastic process in terms of two polynomials, one for the autoregression (AR) and the second for the moving average (MA). The general ARMA model was described in the 1951 thesis of Peter Whittle, Hypothesis testing in time series analysis, and it was popularized in the 1970 book by George E. P. Box and Gwilym Jenkins [4]. Given a time series of data X_t , the ARMA model is a tool for understanding and, perhaps, predicting future values in this series. The AR part involves regressing the variable on its own lagged (i.e., past) values. The MA part involves modeling the error term as a linear combination of error terms occurring contemporaneously and at various times in the past. The model is usually referred to as the ARMA(p,q) model where p is the order of the AR part and q is the order of the MA part (as defined below).

Definition 1.4.1 *The notation ARMA(p,q) refers to the model with p autoregressive terms and* q *moving-average terms. This model contains the* $AR(p)$ *and* $MA(q)$ *models,(short for ARMA (p,q))*

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{1.9}$$

where the φ_i and θ_i are the parameters of the model and the ε_i the terms of error.

- An autoregressive model $AR(p)$ is a $ARMA(p,0)$.
- A moving average model $MA(q)$ is a $ARMA(0,q)$.

1.4.1 Causality and Invertibility

While a moving average process of order (q) will always be stationary without conditions on the coefficients $(\theta_1),(...),(\theta_q)$, some deeper thoughts are required in the case of AR(p) and ARMA(p,q) processes. For simplicity, we start by investigating the autoregressive process of order one, which is given by the equations $(X_t = \phi X_{t-1} + Z_t)$ (writing $(\phi = \phi_1)$). Repeated iterations yield that

$$X_t = \phi X_{t-1} + Z_t = \phi^2 X_{t-2} + Z_t + \phi Z_{t-1} = ... = \phi^N X_{t-N} + \sum_{j=0}^{N-1} \phi^j Z_{t-j}.$$

Letting $N \to \infty$, it could now be shown that, with probability one,

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

is the weakly stationary solution to the AR(1) equations, provided that $|\phi| < 1$.These calculations would indicate moreover, that an autoregressive process of order one can be represented as linear process with coefficients $\psi_j = \phi^j$.

Definition 1.4.2 *An ARMA(p,q) process is causal if there is a sequence* $(\psi_j : j \in \mathbb{N}_0)$ *such that* $(\sum_{j=0}^{\infty} |\psi_j| < \infty)$ *and* $[X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t \in \mathbb{Z}.]$

Causality means that an ARMA time series can be represented as a linear process. It was seen earlier in this section how an AR(1) process whose coefficient satisfies the condition $(|\phi| < 1)$ can be converted into a linear process.
It was also shown that this is impossible if $(|\phi| > 1)$.
The conditions on the autoregressive parameter (ϕ) can be restated in terms of the corresponding autoregressive polynomial $(\phi(z) = 1 - \phi(z))$ as follows. It holds that

$$(|\phi| < 1) \; if \; and \; only \; if (\phi(z) \neq 0) \; for \; all (|z| \leq 1,)$$

$$(|\phi| > 1) \; if \; and \; only \; if \; (\phi(z) \neq 0) \; for \; all \; (|z| \geq 1).$$

It turns out that the characterization in terms of the zeroes of the autoregressive polynomials carries over from the AR(1) case to the general ARMA(p,q) case. Moreover, the (ψ)-weights of the resulting linear process have an easy representation in terms of the polynomials $(\phi(z))$ and $(\theta(z))$. The result is summarized in the next theorem.

Theoreme 1.4.1 *Let $(X_t : t \in \mathbb{Z})$ be an ARMA(p,q) process such that the polynomials $(\phi(z))$ and $(\theta(z))$ have no common zeroes. Then $(X_t : t \in \mathbb{Z})$ is causal if and only if $(\phi(z) \neq 0)$ for all $(z \in \mathbb{C})$ with $(|z| \leq 1)$. The coefficients $(\psi_j : j \in \mathbb{N}_0)$ are determined by the power series expansion*

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \ |z| \leq 1.$$

Definition 1.4.3 *An ARMA(p,q) process given by (1.9) is invertible if there is a sequence $(\pi_j : j \in \mathbb{N}_0)$ such that $(\sum_{j=0}^{\infty} |\pi_j| < \infty)$ and*

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \qquad t \in \mathbb{Z}.$$

Theoreme 1.4.2 *Let $(X_t : t \in \mathbb{Z})$ be an ARMA (p,q) process such that the polynomials $\phi(z)$ and $\theta(z)$ have no common zeroes. Then $(X_t : t \in \mathbb{Z})$ is invertible if and only if $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$. The coefficients $(\pi_j)_{j \in \mathbb{N}_0}$ are determined by the power series expansion*

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \ |z| \leq 1.$$

From now on it is assumed that all ARMA sequences specified in the sequel are causal and invertible unless explicitly stated otherwise. The example of this section highlights the usefulness of the established theory. It deals with parameter redundancy and the calculation of the causality and invertibility sequences $(\psi_j : j \in \mathbb{N}_0)$ and $(\pi_j : j \in \mathbb{N}_0)$.

Example 1.4.1 *Consider the ARMA equations*

$$X_t = 0.4X_{t-1} + 0.21X_{t-2} + Z_t + 0.6Z_{t-1} + 0.09Z_{t-2},$$

which seem to generate an ARMA (2,2) sequence. However, the autoregressive and moving average polynomials have a common zero:

$$\tilde{\phi}(z) = 1 - 0.4z - 0.21z^2 = (1 - 0.7z)(1 + 0.3z),$$
$$\tilde{\theta}(z) = 1 + 0.6z + 0.09z^2 = (1 + 0.3z)^2.$$

Therefore, one can reset the ARMA equations to a sequence of order $(1,1)$ and obtain $X_t = 0.7X_{t-1} + Z_t + 0.3Z_{t-1}$.

Now, the corresponding polynomials have no common roots. Note that the roots of $\phi(z) = 1 - 0.7z$ and $\theta(z) = 1 + 0.3z$ are $10/7 > 1$ and $-10/3 < -1$, respectively. Thus Theorems (1.4.1) and (1.4.2) imply that causal and invertible solutions exist. In the following, the corresponding coefficients in the expansions

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \qquad \text{and} \qquad Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \qquad t \in \mathbb{Z},$$

are calculated. Starting with the causality sequence $(\psi_j : j \in \mathbb{N}_0)$. Writing, for $|z| \leq 1$,

$$\sum_{j=0}^{\infty} \psi_j z^j = \psi(z) = \frac{\theta(z)}{\phi(z)} = \frac{1 + 0.3z}{1 - 0.7z} = (1 + 0.3z) \sum_{j=0}^{\infty} (0.7z)^j,$$

it can be obtained from a comparison of coefficients that

$\psi_0 = 1 \quad and \quad \psi_j = (0.7 + 0.3)(0.7)^{j-1} = (0.7)^{j-1}, \qquad j \in \mathbb{N}.$
Similarly one computes the invertibility coefficients $(\pi_j : j \in \mathbb{N}_0)$ *from the equation*

$$\sum_{j=0}^{\infty} \pi_j z^j = \pi(z) = \frac{\phi(z)}{\theta(z)} = \frac{1 - 0.7z}{1 + 0.3z} = (1 - 0.7z) \sum_{j=0}^{\infty} (-0.3z)^j$$

$(|z| \leq 1)$ *as*

$$\pi_0 = 1 \quad and \quad \pi_j = (-1)^j (0.3 + 0.7)(0.3)^{j-1} = (-1)^j (0.3)^{j-1}.$$

Together, the previous calculations yield to the explicit representations

$$X_t = Z_t + \sum_{j=1}^{\infty} (0.7)^{j-1} Z_{t-j} \; and \; Z_t = X_t + \sum_{j=1}^{\infty} (-1)^j (0.3)^{j-1} X_{t-j}.$$

In the remainder of this section, a general way is provided to determine the weights $(\psi_j : j \geq 1)$ for a causal ARMA (p, q) process given by $\phi(B)X_t = \theta(B)Z_t$, where $\phi(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. Since $\psi(z) = \theta(z)/\phi(z)$ for these z, the weight ψ_j can be computed by matching the corresponding coefficients in the equation $\psi(z)\phi(z) = \theta(z)$, that is,

$$(\psi_0 + \psi_1 z + \psi_2 z^2 + \ldots)(1 - \phi_1 z - \ldots - \phi_p z^p) = 1 + \theta_1 z + \ldots + \theta_q z^q.$$

Recursively solving for $\psi_0, \psi_1, \psi_2, \ldots$ gives

$$\begin{aligned} \psi_0 &= 1, \\ \psi_1 - \phi_1 \psi_0 &= \theta_1, \\ \psi_2 - \phi_1 \psi_1 - \phi_2 \psi_0 &= \theta_2, \end{aligned}$$

and so on as long as $j < \max\{p, q + 1\}$. The general solution can be stated as

$$\psi_j - \sum_{k=1}^{j} \phi_k \psi_{j-k} = \theta_j, \qquad 0 \leq j < \max\{p, q + 1\}, \tag{1.10}$$

$$\psi_j - \sum_{k=1}^{p} \phi_k \psi_{j-k} = 0, \qquad j \geq \max\{p, q + 1\}, \tag{1.11}$$

if we define $\phi_j = 0$ if $j > p$ and $\theta_j = 0$ if $j > q$. To obtain the coefficients ψ_j one therefore has to solve the homogeneous linear difference equation (1.11) subject to the initial conditions specified by (1.10).

1.4.2 Parameter Estimation

Let $(X_t : t \in \mathbb{Z})$ be a causal and invertible ARMA(p,q) process with known orders p and q, possibly with mean μ. This section is concerned with estimation procedures for the unknown parameter vector

$$\beta = (\mu, \phi_1, \ldots, \phi_p, \theta_1, ..., \theta_q, \sigma^2)^T. \tag{1.12}$$

To simplify the estimation procedure, it is assumed that the data has already been adjusted by subtraction of the mean and the discussion is therefore restricted to zero mean ARMA models.

In the following, three estimation methods are introduced. The method of moments works best in case of pure AR processes, while it does not lead to optimal estimation procedures for general ARMA processes. For the latter, more efficient estimators are provided by the maximum likelihood and least squares methods which will be discussed subsequently.

1.4.3 Method1(Method of Moments)

Since this method is only efficient in their case, the presentation here is restricted to AR(p) processes

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + Z_t, t \in \mathbb{Z},$$

where $(Z_t\colon t \in \mathbb{Z}) \sim \text{WN}(0, \sigma^2)$. The parameter vector β consequently reduces to $(\phi, \sigma^2)^T$ with $\phi = (\phi_1, \ldots, \phi_p)^T$ and can be estimated using the Yule-Walker equations

$$\Gamma_p \phi = \gamma_p \text{ and } \sigma^2 = \gamma(0) - \phi^T \gamma_p, \tag{1.13}$$

where $\Gamma_p = (\gamma(k-j))_{k,j=1,\ldots,p}$ and $\gamma_p = (\gamma(1), ..., \gamma(p))^T$. Observe that the equations are obtained by the same arguments applied to derive the Durbin-Levinson algorithm in the previous section. The method of moments suggests to replace every quantity in the Yule-Walker equations with their estimated counterparts, which yields the Yule-Walker estimators

$$\widehat{\phi} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p = \hat{R}_p^{-1} \hat{\rho}_p \tag{1.14}$$

$$\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\gamma}_p^T \hat{\Gamma}_p^{-1} \hat{\gamma}_p = \hat{\gamma}(0) \left[1 - \hat{\rho}_p^T \hat{R}_p^{-1} \hat{\rho}_p\right]. \tag{1.15}$$

There in, $\hat{R}_p = \hat{\gamma}(0)^{-1} \hat{\Gamma}_p$ and $\hat{\rho}_p = \hat{\gamma}(0)^{-1} \hat{\gamma}_p$.

Using $\hat{\gamma}(h)$ as estimator for the ACVF at lag h, a dependence on the sample size n is obtained in an implicit way. This dependence is suppressed in the notation used here. The following theorem contains the limit behavior of the Yule-Walker estimators as n tends to infinity.

Theoreme 1.4.3 *If $(X_t\colon t \in \mathbb{Z})$ is a causal AR(p) process, then*

$$\sqrt{n}(\widehat{\phi} - \phi) \xrightarrow{\mathcal{D}} N(0, \sigma^2 \Gamma_p^{-1}) \qquad \textit{and} \qquad \hat{\sigma}^2 \xrightarrow{P} \sigma^2$$

as $n \to \infty$, where $\to^P$ indicates convergence in probability.

A proof of this result is given in Section 8.10 of Brockwell and Davis (1991) [6].

Corollary 1.4.1 *If $(X_t\colon t \in \mathbb{Z})$ is a causal AR(p) process, then*

$$\sqrt{n}\widehat{\phi}_{hh} \xrightarrow{\mathcal{D}} Z \qquad (n \to \infty)$$

for all $h > p$, where Z stands for a standard normal random variable.

Example 1.4.2 (Yule-Walker estimates for AR(2) processes) *Suppose that $n = 144$ values of the autoregressive process $X_t = 1.5X_{t-1} - 0.75X_{t-2} + Z_t$ have been observed, where $(Z_t\colon t \in \mathbb{Z})$ is a sequence of independent standard normal variates.Assume further that $\hat{\gamma}(0) = 8.434, \hat{\rho}(1) = 0.834$ and $\hat{\rho}(2) = 0.476$ have been calculated from the data. The Yule-Walker* *estimators for the parameters are then given by*

$$\hat{\phi} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} = \begin{pmatrix} 1.000 & 0.834 \\ 0.834 & 1.000 \end{pmatrix}^{-1} \begin{pmatrix} 0.834 \\ 0.476 \end{pmatrix} = \begin{pmatrix} 1.439 \\ -0.725 \end{pmatrix}$$

and

$$\hat{\sigma}^2 = 8.434 \left[1 - (0.834, 0.476) \begin{pmatrix} 1.439 \\ -0.725 \end{pmatrix} \right] = 1.215.$$

To construct asymptotic confidence intervals using Theorem (1.4.3), the unknown limiting covariance matrix $\sigma^2\Gamma_p^{-1}$ needs to be estimated. This can be done using the estimator

$$\frac{\hat{\sigma}^2\hat{\Gamma}_p^{-1}}{n} = \frac{1}{144}\frac{1.215}{8.434} \begin{pmatrix} 1.000 & 0.834 \\ 0.834 & 1.000 \end{pmatrix}^{-1} = \begin{pmatrix} 0.057^2 & -0.003 \\ -0.003 & 0.057^2 \end{pmatrix}.$$

Then, the $1-\alpha$ level confidence interval for the parameters ϕ_1 and ϕ_2 are computed as
$1.439 \pm 0.057 z_{1-\alpha/2}$ *and* $-0.725 \pm 0.057 z_{1-\alpha/2}$,
respectively, where $z_{1-\alpha/2}$ is the corresponding normal quantile.

Example 1.4.3 *Consider the invertible MA(1) process $X_t = Z_t + \theta Z_{t-1}$, where $|\theta| < 1$. Using invertibility, each X_t has an infinite autoregressive representation*

$$X_t = \sum_{j=1}^{\infty} (-\theta)^j X_{t-j} + Z_t$$

that is nonlinear in the unknown parameter θ to be estimated. The method of moments is here based on solving

$$\hat{\rho}(1) = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \frac{\hat{\theta}}{1+\hat{\theta}^2}.$$

for $\hat{\theta}$. The foregoing quadratic equation has the two solutions

$$\hat{\theta} = \frac{1 \pm \sqrt{1-4\hat{\rho}(1)^2}}{2\hat{\rho}(1)},$$

of which we pick the invertible one. Note moreover, that $|\hat{\rho}(1)|$ is not necessarily less or equal to $1/2$ which is required for the existence of real solutions. (The theoretical value $|\rho(1)|$, however, is always less than $1/2$ for any MA(1) process, as an easy computation shows). Hence, θ can not always be estimated from given data samples.

1.4.4 Method 2(Maximum Likelihood Estimation)

The innovations algorithm of the previous section applied to a causal ARMA(p,q) process $(X_t \colon t \in \mathbb{Z})$ gives

$$\hat{X}_{i+1} = \sum_{j=1}^{i} \theta_{ij}(X_{i+1-j} - \hat{X}_{i+1-j}), \qquad 1 \le i < \max\{p,q\},$$

$$\hat{X}_{i+1} = \sum_{j=1}^{p} \phi_j X_{i+1-j} + \sum_{j=1}^{q} \theta_{ij}(X_{i+1-j} - \hat{X}_{i+1-j}), \qquad i \ge \max\{p,q\},$$

with prediction error

$$P_{i+1} = \sigma^2 R_{i+1}.$$

In the last expression, σ^2 has been factored out due to reasons that will become apparent from the form of the likelihood function to be discussed below. Recall that the sequence $(X_{i+1}-\hat{X}_{i+1} : i \in \mathbb{Z})$ consists of uncorrelated random variables if the parameters are known. Assuming normality for the errors, we moreover obtain even independence. This can be exploited to define the Gaussian maximum likelihood estimation(MLE) procedure. Throughout, it is assumed that $(X_t : t \in \mathbb{Z})$ has zero mean ($\mu = 0$). The parameters of interest are collected in the vectors $\beta = (\phi, \theta, \sigma^2)^T$ and $\beta' = (\phi, \theta)^T$, where $\phi = (\phi_1, ..., \phi_p)^T$ and $\theta = (\theta_1, ..., \theta_q)^T$. Assume finally that we have observed the variables $X_1, ..., X_n$. Then, the Gaussian likelihood function for the innovations is

$$L(\beta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \left(\prod_{i=1}^{n} R_i^{1/2} \right) \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{R_j} \right).$$

Taking the partial derivative of $\ln L(\beta)$ with respect to the variable σ^2 reveals that the MLE for σ^2 can be calculated from

$$\hat{\sigma}^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n}, \qquad S(\hat{\phi}, \hat{\theta}) = \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{R_j}.$$

Therein, $\hat{\phi}$ and $\hat{\theta}$ denote the MLEs of ϕ and θ obtained from minimizing the profile likelihood or reduced likelihood

$$\ell(\phi, \theta) = \ln\left(\frac{S(\phi, \theta)}{n} \right) + \frac{1}{n} \sum_{j=1}^{n} \ln(R_j).$$

Observe that the profile likelihood $\ell(\phi, \theta)$ can be computed using the innovations algorithm. The speed of these computations depends heavily on the quality of initial estimates. These are often provided by the non-optimal Yule-Walker procedure.

The limit distribution of the MLE procedure is given as the following theorem. Its proof can be found in Section 8.8 of Brockwell and Davis (1991) [6].

Theoreme 1.4.4 *Let $(X_t : t \in \mathbb{Z})$ be a causal and invertible ARMA(p,q) process defined with an iid sequence $(Z_t : t \in \mathbb{Z})$ satisfying $\mathbb{E}[Z_t] = 0$ and $\mathbb{E}[Z_t^2] = \sigma^2$. Consider the MLE $\hat{\beta}'$ of β' that is initialized with the moment estimators of Method 1.Then,*

$$\sqrt{n}(\hat{\beta}' - \beta') \xrightarrow{\mathcal{D}} N(0, \sigma^2 \Gamma_{p,q}^{-1}) \qquad (n \to \infty).$$

The result is optimal. The covariance matrix $\Gamma_{p,q}$ is in block form and can be evaluated in terms of covariances of various autoregressive processes.

1.4.5 Method 3 (Least Squares Estimation)

An alternative to the method of moments and the MLE is provided by the least squares estimation (LSE). For causal and invertible ARMA(p,q) processes, it is based on minimizing the weighted sum of squares

$$S(\phi, \theta) = \sum_{j=1}^{n} \frac{(X_j) - \hat{X}_j^{\,2}}{R_j} \tag{1.16}$$

with respect to ϕ and θ, respectively. Assuming that $\tilde{\phi}$ and $\tilde{\theta}$ denote these LSEs, the LSE for σ^2 is computed as

$$\tilde{\sigma}^2 = \frac{S(\tilde{\phi}, \tilde{\theta})}{n - p - q}.$$

The least squares procedure has the same asymptotics as the MLE.

1.5 The ARIMA model

ARIMA(p,d,q) forecasting equation: ARIMA models are, in theory, the most general class of models for forecasting a time series which can be made to be stationary by differencing (if necessary), perhaps in conjunction with nonlinear transformations such as logging or deflating (if necessary). A random variable that is a time series is stationary if its statistical properties are all constant over time. A stationary series has no trend, its variations around its mean have a constant amplitude, and it wiggles in a consistent fashion, i.e., its short-term random time patterns always look the same in a statistical sense.The latter condition means that its autocorrelations (correlations with its own prior deviations from the mean) remain constant over time, or equivalently, that its power spectrum remains constant over time. A random variable of this form can be viewed (as usual) as a combination of signal and noise, and the signal (if one is apparent) could be a pattern of fast or slow mean reversion, or sinusoidal oscillation, or rapid alternation in sign, and it could also have a seasonal component. An ARIMA model can be viewed as a filter that tries to separate the signal from the noise, and the signal is then extrapolated into the future to obtain forecasts.

The ARIMA forecasting equation for a stationary time series is a linear (i.e., regression-type) equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors. That is:
Predicted value of $Y = a$ constant and/or a weighted sum of one or more recent values of Y and/or a weighted sum of one or more recent values of the errors.
If the predictors consist only of lagged values of Y, it is a pure autoregressive (self-regressed) model, which is just a special case of a regression model and which could be fitted with standard regression software. For example, a first-order autoregressive ($AR(1)$) model for Y is a simple regression model in which the independent variable is just Y lagged by one period ($LAG(Y, 1)$ in Statgraphics or Y_LAG1 in RegressIt). If some of the predictors are lags of the errors, an ARIMA model it is a linear regression model, because there is no way to specify last period's error? as an independent variable: the errors must be computed on a period-to-period basis when the model is fitted to the data. From a technical standpoint, the problem with using lagged errors as predictors is that the model's predictions are not linear functions of the coefficients, even though they are linear functions of the past data. So, coefficients in ARIMA models that include lagged errors must be estimated by nonlinear optimization methods (hill-climbing) rather than by just solving a system of equations.
The acronym ARIMA stands for Auto-Regressive Integrated Moving Average. Lags of the stationarized series in the forecasting equation are called "autoregressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series. Random-walk and random-trend models, autoregressive models, and exponential smoothing

models are all special cases of ARIMA models.
A nonseasonal ARIMA model is classified as an "$ARIMA(p, d, q)$" model, where:

- p is the number of autoregressive terms,
- d is the number of nonseasonal differences needed for stationarity, and
- q is the number of lagged forecast errors in the prediction equation.

1.6 ARCH/GARCH Models

Autoregressive integrated moving average (ARIMA) models that allow modeling of volatility are unable to deal with volatility over time.
Monetary and financial series are characterized by volatility clustering means period of high volatility alternate with periods of low volatility. This phenomenon is called conditional heterocedents that is particularly common in stock market data, ...
From an empirical point of view, we can note the presence of nonlinear phenomena in the time series such as the presence of non-constant variability, the presence of the evolution cycle of volatility
In order to highlight the persistence of volatility, Engle [11] proposed a new model called Heteroscedastic Conditional Autoregressive (ARCH) able to capture the behavior of volatility as a function of time.
This takes into account the numeber of stylized facts that characterize the majority of financial series such as persistence, volatility clusters and leptokuretic behavior of data.
GARCH modeling has becom an essential tool in finance for analyzing and predicting volatility.
It was introduced to correct the weaknesses of the ARCH model at the estimation level with a high number of parameters.
In order to highlight the long memory of volatility Bollerslev [3] generalized the ARCH(q) model by proposing the GARCH model (p,q) which consists in adding the delayed variance in its equation.

1.6.1 ARCH(1): Definition and Properties

The ARCH model of order 1, ARCH(1), is defined as follows:

Definition 1.6.1 *The process $\varepsilon_t, t \in \mathbb{Z}$, is ARCH(1), if* $\mathbb{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$,

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 \tag{1.17}$$

with $\omega > 0, \alpha \geq 0$ and

- $\operatorname{Var}(\varepsilon_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$ and $Z_t = \varepsilon_t/\sigma_t$ is i.i.d. (strong ARCH)
- $\operatorname{Var}(\varepsilon_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$ (semi-strong ARCH),
- $\mathcal{P}(\varepsilon_t^2 \mid 1, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots) = \sigma_t^2$ (weak ARCH),

Theoreme 1.6.1 *Assume that the process ε_t is a weak ARCH(1) process with* $\operatorname{Var}(\varepsilon_t) = \sigma^2 < \infty$. *Then it follows that ε_t is white noise.*

Proof: From $\mathbb{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$ it follows that $\mathbb{E}[\varepsilon_t] = 0$ and $Cov(\varepsilon_t, \varepsilon_{t-k}) = ...\varepsilon_{t-k}\mathbb{E}(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$. □

Note that ε_t is not an independent white noise.

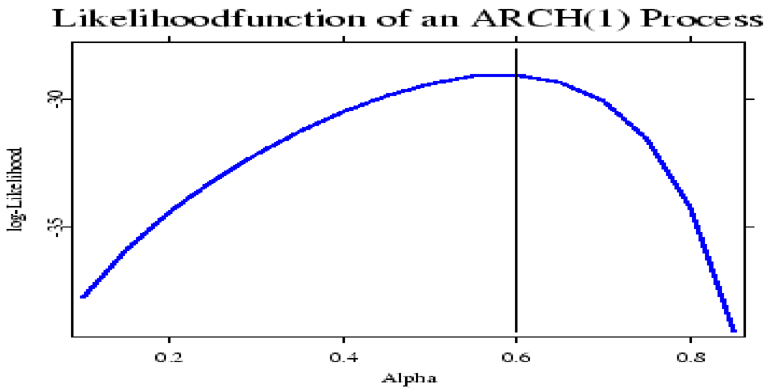

Figure 1.5: Conditional likelihood function of a generated ARCH(1) process with $n = 100$. The true parameter is $\alpha = 0.5$

1.6.2 ARCH(q):[Definition and Properties]

The definition of an ARCH(1) model will be extended for the case that $q > 1$ lags, on which the conditional variance depends.

Definition 1.6.2 *The process* (ε_t)*,* $t \in \mathbb{Z}$*, is ARCH(q), when* $\mathbb{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$*,*

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + ... + \alpha_q \varepsilon_{t-q}^2 \tag{1.18}$$

with $\omega > 0,\ \alpha_1 \geq 0, ..., \alpha_q \geq 0$ and

- $\mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$ and $Z_t = \varepsilon_t / \sigma_t$ is i.i.d. (strong ARCH)
- $\mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$ (semi-strong ARCH), or
- $\mathcal{P}(\varepsilon_t^2 \mid 1, \varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, ...) = \sigma_t^2$ (weak ARCH)

The conditional variance σ_t^2 in an ARCH(q) model is also a linear function of the q squared lags.

Theoreme 1.6.2 *Let* ε_t *be a semi-strong ARCH(q) process with* $\mathrm{Var}(\varepsilon_t) = \sigma^2 < \infty$*. Then*

$$\sigma^2 = \frac{\omega}{1 - \alpha_1 - ... - \alpha_q}$$

with $\alpha_1 + ... + \alpha_q < 1$

Proof:

If instead $\alpha_1 + ... + \alpha_q \geq 1$, then the unconditional variance does not exist and the process is not covariance-stationary.

1.6.3 Generalized ARCH (GARCH)

The ARCH(q) model can be generalized by extending it with autoregressive terms of the volatility.

Definition 1.6.3 (GARCH(p,q)) *The process* (ε_t)*,* $t \in \mathbb{Z}$*, is GARCH(p,q), if* $\mathbb{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$*,*

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \tag{1.19}$$

and

- $\mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$ and $Z_t = \varepsilon_t \sigma_t$ is i.i.d. (strong GARCH)
- $\mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$ (semi-strong GARCH), or
- $\mathcal{P}(\varepsilon_t^2 \mid 1, \varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots) = \sigma_t^2$ (weak GARCH).

The sufficient but not necessary conditions for

$$\sigma_t^2 > 0 \quad a.s., (\mathcal{P}[\sigma_t^2 > 0] = 1) \tag{1.20}$$

$\omega > 0,\ \alpha_i \geq 0,\ i = 1, ..., q$ and $\beta_j \geq 0,\ j = 1, ..., p$. In the case of the GARCH(1,2) model

$$\begin{aligned}
\sigma_t^2 &= \sigma_t^2 \\
&= \frac{\omega}{1-\beta} + \alpha_1 \sum_{j=0}^{\infty} \beta_1^j \varepsilon_{t-j-1}^2 + \alpha_2 \sum_{j=0}^{\infty} \beta_1^j \varepsilon_{t-j-2}^2 \\
&= \frac{\omega}{1-\beta} + \alpha_1 \varepsilon_{t-1}^2 + (\alpha_1 \beta_1 + \alpha_2) \sum_{j=0}^{\infty} \beta_1^j \varepsilon_{t-j-2}^2
\end{aligned}$$

with $0 \leq \beta_1 < 1.\omega > 0, \alpha_1 \geq 0$ and $\alpha_1 \beta_1 + \alpha_2 \geq 0$ are necessary assuming that the sum $\sum_{j=0}^{\infty} \beta_1^j \varepsilon_{t-j-2}^2$ converges.

Theoreme 1.6.3 *Let* ε_t *be a semi-strong GARCH(1,1) process with* $\mathrm{Var}(\varepsilon_t) = \sigma^2 < \infty$ *and* $Z_t \sim \mathrm{N}(0,1)$. *Then* $\mathbb{E}[\varepsilon_t^4] < \infty$ *holds if and only if* $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$. *The Kurtosis* $\mathrm{Kurt}(\varepsilon_t)$ *is given as*

$$\mathrm{Kurt}[\varepsilon_t] = \frac{\mathbb{E}[\varepsilon_t^4]}{\left(\mathbb{E}[\varepsilon_t^2]\right)^2} = 3 + \frac{6\alpha_1^2}{1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2}. \tag{1.21}$$

Proof

It can be proved that $\mathbb{E}[\varepsilon_t^4] = 3\mathbb{E}[(\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^2]$ and the stationarity of ε_t.

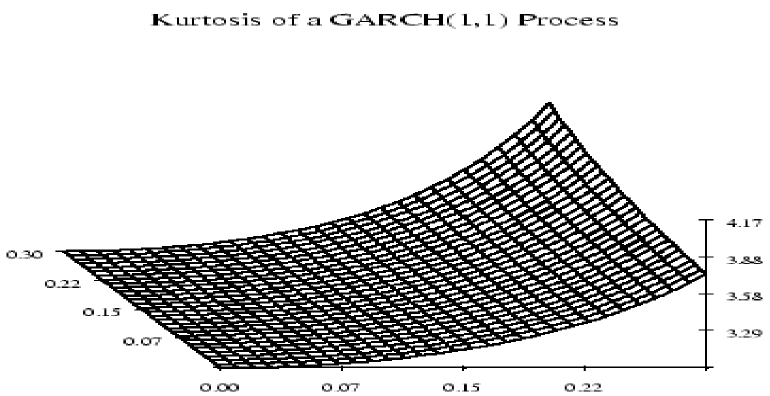

Figure 1.6: Kurtosis of a GARCH(1,1) process. The left axis shows the parameter β_1, the right α_1

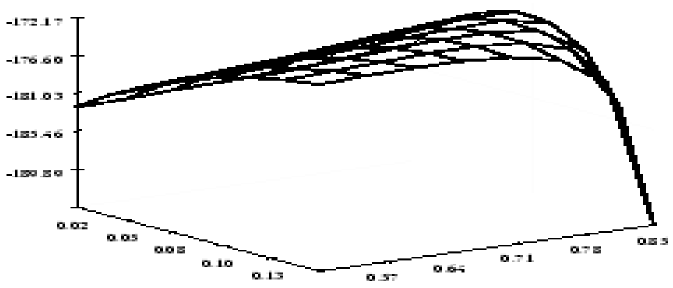

Figure 1.7: Likelihood function of a generated GARCH(1,1) process with $n = 500$. The left axis shows the parameter β, the right α. The true parameters are $\omega = 0.1$, $\alpha = 0.1$and $\beta = 0.8$

1.7 Simulate processes

1.7.1 Simulations and Correlograms

AR(1)

Let's begin with an $AR(1)$ process. This is similar to a random walk, except that ϕ_1 does not have to equal unity. Our model is going to have $\phi_1 = 0.6$. The R code for creating this simulation is given as follows:

```
x <- w <- rnorm(100)
for (t in 2:100) x[t]<- 0.6*x[t-1] + w[t]
```

Notice that our for loop is carried out from 2 to 100, not 1 to 100, as $x[t-1]$ when $t = 0$ is not indexable. Similarly for higher order AR(p) processes, t must range from p to 100 in this loop.
We can plot the realisation of this model and its associated correlogram using the layout function:

```
 layout(1:2)
plot(x, type="l")
acf(x)
pacf(x)
```

AR(2)

Let's add some more complexity to our autoregressive processes by simulating a model of order 2. In particular, we will set $\phi_1 = 0.666$, but also set $\phi_2 = 0.333$. Here's the full code to simulate and plot the realisation, as well as the correlogram for such a series:

```
x <- w <- rnorm(100)
 for (t in 3:100) x[t] <- 0.666*x[t-1] - 0.333*x[t-2] + w[t]
```

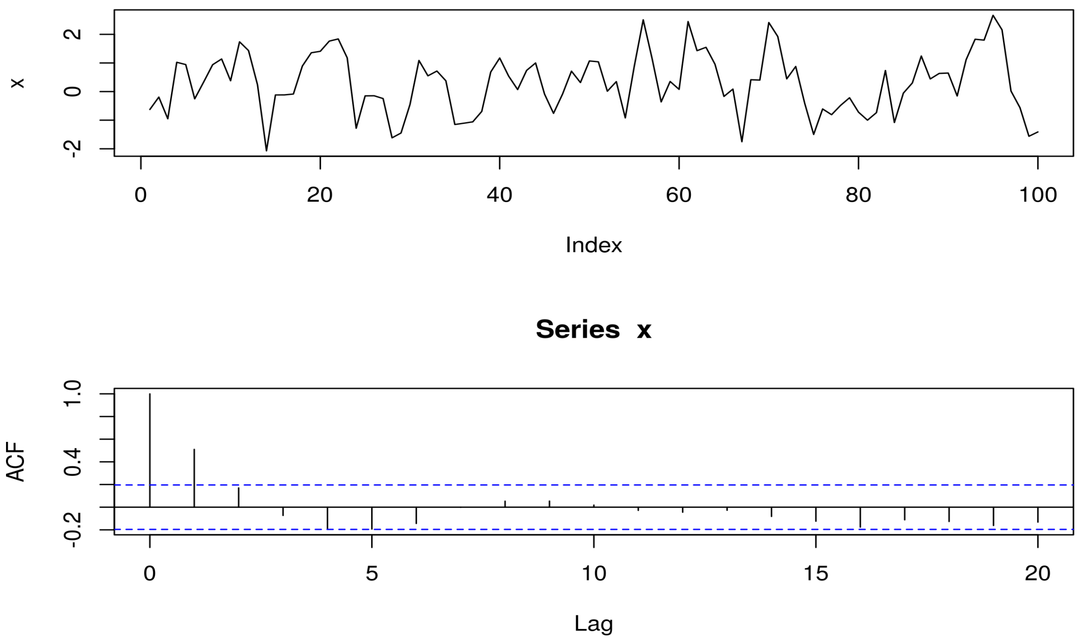

Figure 1.8: Realisation of AR(1) Model, with $\phi_1 = 0.6$ and Associated Correlogram

```
layout(1:2)
plot(x, type="l")
acf(x)
```

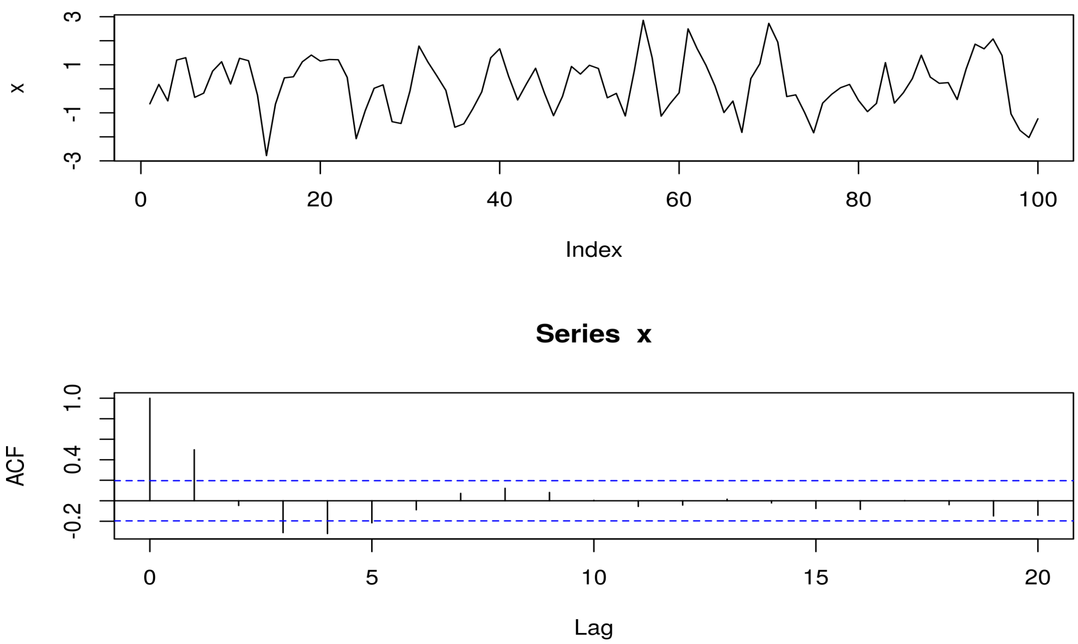

Figure 1.9: Realisation of AR(2) Model, with $\phi_1 = 0.666$, $\phi_2 = 0.333$. and Associated Correlogram

1.7.2 Simulate a MA process

MA(1)

Let's start with a MA(1) process. If we set $\theta_1 = 0.6$ we obtain the following model:

$$x_t = \epsilon_t + 0.6\epsilon_{t-1}$$

As with the AR(p) models, we can use $\mathbb{R}$ to simulate such a series and then plot the correlogram.

```
x <- w <- rnorm(100)
for (t in 2:100) x[t] <- w[t] + 0.6*w[t-1]
layout(1:2)
plot(x, type="l")
acf(x)
```

MA(3)

Let's run through the same procedure for a MA(3) process. This time we should expect significant peaks at $k \in \{1, 2, 3\}$, and insignificant peaks for $k > 3$.
We are going to use the following coefficients: $\theta_1 = 0.6, \theta_2 = 0.4 \quad and \quad \theta_3 = 0.2$. Let's simulate a MA(3) process from this model. I've increased the number of random samples to 1000 in this simulation, which makes it easier to see the true autocorrelation structure, at the expense of making the original series harder to interpret:

```
x <- w <- rnorm(1000)
```

```
for (t in 4:1000) x[t] <- w[t] + 0.6*w[t-1] + 0.4*w[t-2] + 0.3*w[t-3]
layout(1:2)
 plot(x, type="l")
acf(x)
```

The output is as follows:

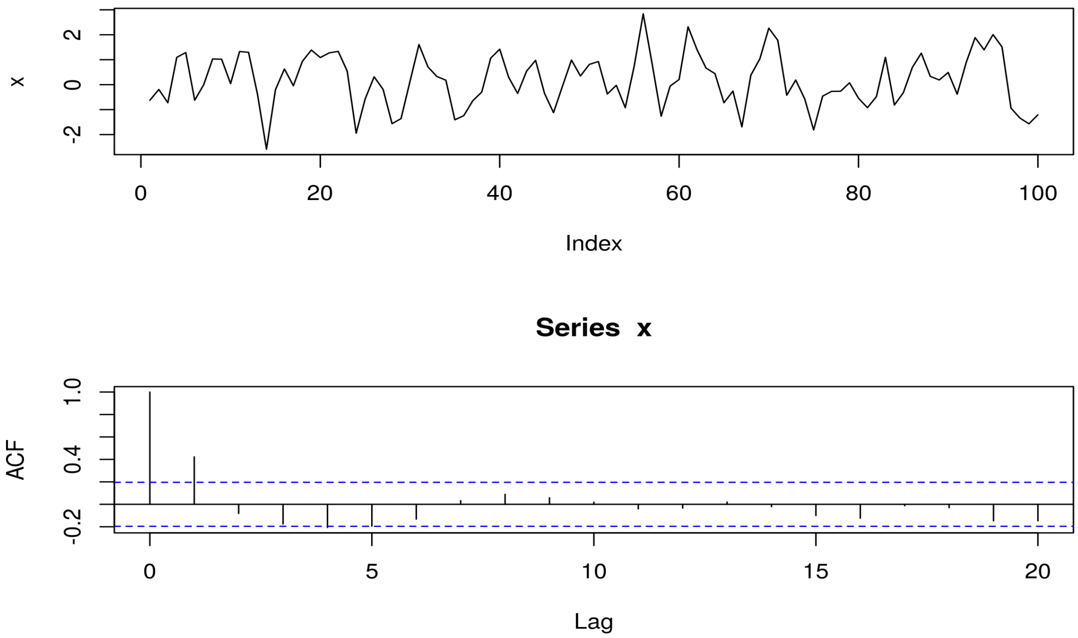

Figure 1.10: Realisation of MA(1) Model, with $\theta_1 = 0.6$ and Associated Correlogram

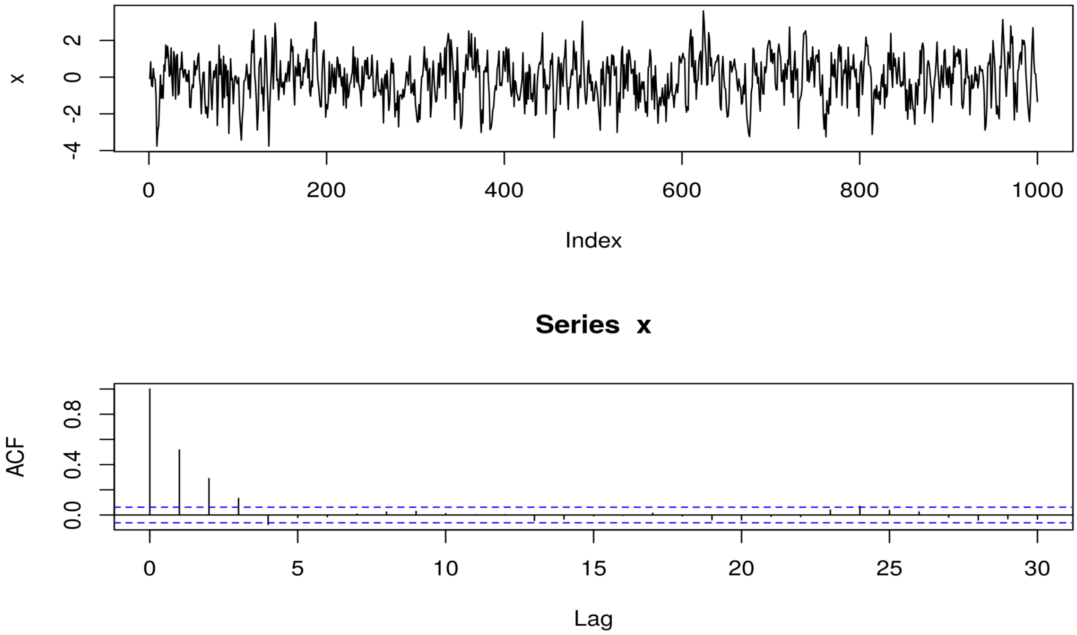

Figure 1.11: Realisation of MA(3) Model and Associated Correlogram

1.7.3 Simulation of an ARMA process

As with the autoregressive and moving average models we will now simulate various ARMA series and then attempt to fit ARMA models to these realisations. We carry this out because we want to ensure that we understand the fitting procedure, including how to calculate confidence intervals for the models, as well as ensure that the procedure does actually recover reasonable estimates for the original ARMA parameters.
In Part 1 and Part 2 we manually constructed the AR and MA series by drawing n samples from a normal distribution.
However, there is a more straightforward way to simulate AR, MA, ARMA and even ARIMA data, simply by using the $arima.sim$ method in R.
Let's start with the simplest possible non-trivial ARMA model, namely the ARMA(1,1) model. That is, an autoregressive model of order one combined with a moving average model of order one. Such a model has only two coefficients, α and β, which represent the first lags of the time series itself and the "shock" white noise terms. Such a model is given by:

$$x_t = \alpha x_t + w_t + \beta w_{t-1}$$

We need to specify the coefficients prior to simulation. Let's take $\alpha = 0.5$ and $\beta = -0.5$:

```
x <- arima.sim(n=1000, model=list(ar=0.5, ma=-0.5))
plot(x)
```

The output is as follows:
Let's also plot the correlogram:

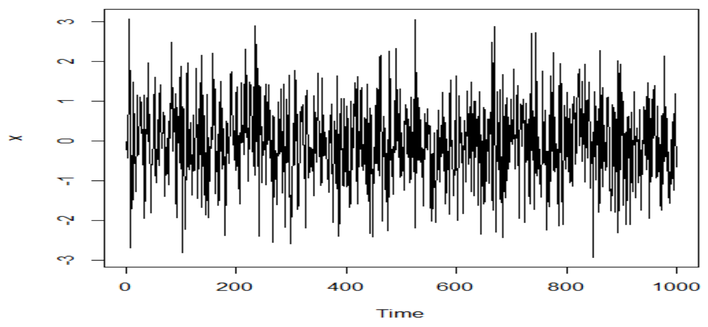

Figure 1.12: Realisation of an ARMA(1,1) Model, with $\alpha = 0.5$ and $\beta = -0.5$

```
acf(x)
```

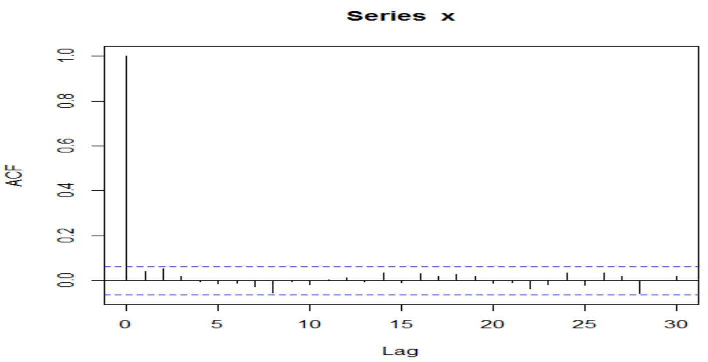

Figure 1.13: Correlogram of an ARMA(1,1) Model, with $\alpha = 0.5$ and $\beta = -0.5$

Let's now try an ARMA(2,2) model. That is, an AR(2) model combined with a MA(2) model. We need to specify four parameters for this model:$\alpha_1, \alpha_2, \beta_1$ *and* β_2. Let's take $\alpha_1 = 0.5, \alpha_2 = -0.25\beta_1 = 0.5$ *and* $\beta_2 = -0.3$:

```
x <- arima.sim(n=1000, model=list(ar=c(0.5, -0.25), ma=c(0.5, -0.3)))
 plot(x)
```

The output of our ARMA(2,2) model is as follows:

And the corresponding autocorelation:

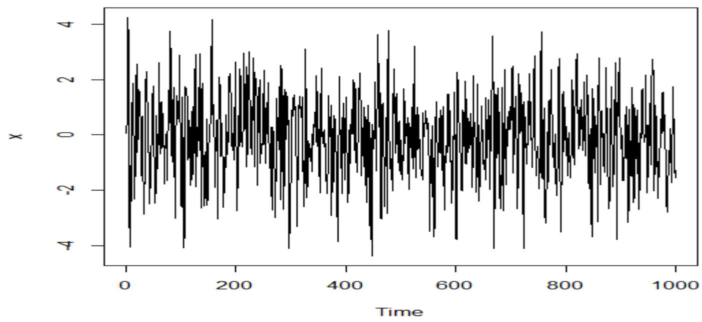

Figure 1.14: Realisation of an ARMA(2,2) Model,with $\alpha_1 = 0.5, \alpha_2 = -0.25\beta_1 = 0.5$ *and* $\beta_2 = -0.3$

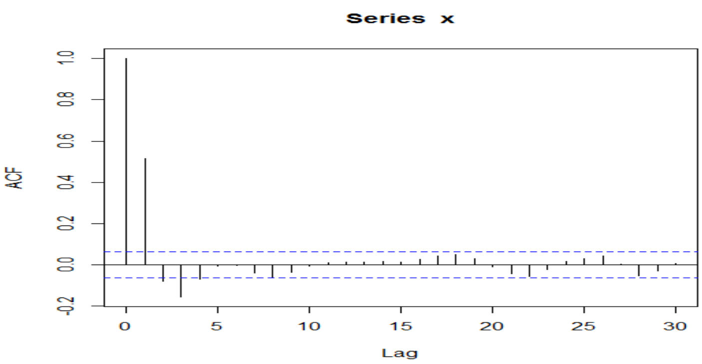

Figure 1.15: Correlogram of an ARMA(2,2) Model,with $\alpha_1 = 0.5, \alpha_2 = -0.25\beta_1 = 0.5 and \beta_2 = -0.3$

Chapter 2

Conditional independence, conditional central limit theorem and conditional association

Summary

2.1 Conditional independence of events 38
2.2 Conditional independence of random variables 38
2.3 Conditional Borel-Cantelli Lemma 39
2.4 Conditional complete convergence 40
2.5 Conditional central limit theorem 40
2.6 Conditional association 42

abstract: Our aim in this chapter is to review the concept of conditional independence and introduce the notions of conditional strong mixing and conditional association for sequences of random variables. We discuss some stochastic inequalities and limit theorems for such sequences of random variables. Earlier discussions on the topic of conditional independence can be found in [3] and more recently in [10].

2.1 Conditional independence of events

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A set of events $A_1, A_2, ..., A_n$ are said to be independent if

$$\mathbb{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k}(A_{i_j}) \tag{2.1}$$

for all $1 \leq i_1 < i_2 < ... < i_k \leq n, 2 \leq k \leq n$.

Definition 2.1.1 *The set of events $A_1, A_2, ..., A_n$ are said to be conditionally independent given an event B with $\mathbb{P}(B) > 0$ if*

$$\mathbb{P}\left(\bigcap_{j=1}^{k} A_{i_j}|B\right) = \prod_{j=1}^{k}(A_{i_j}|B) \tag{2.2}$$

for all $1 \leq i_1 < i_2 < ... < i_k \leq n, 2 \leq k \leq n$.

The following examples [10] show that the independence of events does not imply conditional independence and that the conditional independence of events does not imply their independence.

2.2 Conditional independence of random variables

We will be working on a fixed probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let $\mathcal{F}$ be a sub-σ-field such that $\mathcal{F} \subset \mathcal{A}$.

Definition 2.2.1 *Random events $A_1, A_2, ..., A_n$ are $\mathcal{F}$-independent if*

$$\bigwedge_{1\leq k\leq n} \bigwedge_{1\leq i_1<i_2<...<i_k\leq n} \mathbb{E}^{\mathcal{F}} \prod_{s=1}^{k} I_{A_{is}} = \prod_{s=1}^{k} \mathbb{E}^{\mathcal{F}} I_{A_{is}} \tag{2.3}$$

If $\mathcal{F} = (\emptyset, \Omega)$ we obtain the definition of independence of random variables events.
Note that, if $\mathcal{F} = \mathcal{A}$, then all random events are $\mathcal{A}$-independent.
A sequence of families $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n, \mathcal{A}_k \subset \mathcal{A}$ for $k = 1, ..., n$, is $\mathcal{F}$-independent, if each sequence $A_1, A_2, ..., A_n$ such that $A_i \in \mathcal{A}_i, i = 1, 2, ..., n$ is the sequence of $\mathcal{F}$-independent random events.

If $X : \Omega \to \mathbb{R}$ is a random variable, then by $\mathcal{F}_X$ we denote the smallest σ-field with respect to which the random variable X is measurable. Thus $\mathcal{F}_X$ is σ-field generated by random variable X. The random variables $X_1, X_2, ..., X_n$ are $\mathcal{F}$-independent if σ-fields $\mathcal{F}_{X_1}, \mathcal{F}_{X_2}, ..., \mathcal{F}_{X_n}$ are $\mathcal{F}$-independent.
Note that conditional independence does not not imply independence, the opposite implication is also not true, as incorrectly given in the book [16]. The broad considerations of the independence and the $\mathcal{F}$-independence are contained in [10].

2.3 Conditional Borel-Cantelli Lemma

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathcal{F}$ nonempty sub-σ-field $\mathcal{A}$. The first lemma of Borel-Cantelli states that if $\{A_n, n \geq 1\}$ is an arbitrary sequence $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then with probability one only finite number of events of the sequence $\{A_n, n \geq 1\}$ holds. The second lemma of Borel-Cantelli shows that if the sequence of independent random events such that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then probability that infinity many events of the sequence $\{A_n, n \geq 1\}$ holds equals one.

Theoreme 2.3.1 *Let $\{A_n, n \geq 1\}$ be a sequence of random eventssuch that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\sum_{n=1}^{\infty} \mathbb{P}(A_n|\mathcal{F}) < \infty$ a.s*

Proof. Obviously

$$\sum_{n=1}^{\infty} \mathbb{E}I_{A_n} = \sum_{n=1}^{\infty} \mathbb{E}(\mathbb{E}^{\mathcal{F}} I_{A_n}) = \lim_{n\to\infty} \sum_{n=1}^{\infty} \mathbb{E}(\mathbb{E}^{\mathcal{F}} I_{A_n}) = \lim_{n\to\infty} \mathbb{E}(\sum_{n=1}^{\infty} \mathbb{E}^{\mathcal{F}} I_{A_n})$$

By the Lebesgue's Monotone Convergence Theorem we have that

$$\lim_{n\to\infty} (\sum_{k=1}^{n} \mathbb{E}^{\mathcal{F}} I_{A_n}) = \mathbb{E}(\lim_{n\to\infty} \sum_{n=1}^{\infty} \mathbb{E}^{\mathcal{F}} I_{A_n}) = \mathbb{E}(\sum_{k=1}^{n} \mathbb{E}^{\mathcal{F}} I_{A_n}) = \infty$$

if

$$\mathbb{P}\left(\sum_{k=1}^{n} \mathbb{E}^{\mathcal{F}} I_{A_n}\right) = \infty > 0$$

The opposite implication is not true by Example 1 in [9]

Lemma 2.3.1 (Conditional Lemma of Borel-Cantelli I) *Let $\{A_n, n \geq 1\}$ be a sequence of random events such that $A = \{\omega : \sum_{n=1}^{\infty} \mathbb{P}(A_n|\mathcal{F}) < \infty\}, \mathbb{P}(A) < 1$, then only finitely many events from the sequence $\{A_n \bigcap A, n \geq 1\}$ hold with probability one.*

Proof. Let

$$U = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \bigcap A,$$

then

$$\mathbb{P}(U|\mathcal{F}) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} [A_k \bigcap A]|\mathcal{F}\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} [A_k \bigcap A]|\mathcal{F}\right)$$

$$= \lim_{n\to} \mathbb{E}^{\mathcal{F}} I_{\bigcup_{k=n}^{\infty} [A_k \bigcap A]} \leq \lim_{n\to\infty} \left(\sum_{k=n}^{\infty} \mathbb{E}^{\mathcal{F}} I_{[A_k \bigcap A]}\right) = 0 \quad a.s.$$

Hence $\mathbb{P}(U) = \mathbb{E}[\mathbb{P}(U)|\mathcal{F}] = 0 \quad a.s.$

Lemma 2.3.2 (Conditional Lemma of Borel-Cantelli II) *Let $\{A_n, n \geq 1\}$ be a sequence of $\mathcal{F}$-independent events and let $A = \{\omega : \sum_{n=1}^{\infty} \mathbb{P}(A_n|\mathcal{F}) = \infty\}$. Then $\mathbb{P}(\limsup A_n) = \mathbb{P}(A)$*

Proof. Let $\mathbb{E} = (\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k)^c = \bigcup_{n=n}^{\infty} \bigcap_{k=1}^{\infty} A_k^c$. Properties of conditional expectation imply

$$\begin{aligned}
\mathbb{P}(\mathbb{E}|\mathcal{F}) &= \lim_{n\to\infty} \mathbb{P}(\bigcap_{k=1}^{\infty} A_k^c|\mathcal{F}) = \lim_{n\to\infty} (\lim_{k\to\infty} \mathbb{P}(\bigcap_{i=n}^{k} A_i^c|\mathcal{F})) \\
&= \lim_{n\to\infty} (\lim_{k\to\infty} \prod_{i=n}^{k} \mathbb{P}(A_i^c|\mathcal{F})) = \lim_{n\to\infty} \lim_{k\to\infty} [\prod_{i=n}^{k}(1-\mathbb{P}(A_i|\mathcal{F}))] \\
&= \lim_{k\to\infty} [\prod_{i=n}^{\infty}(1-\mathbb{P}(A_i|\mathcal{F}))] \leq \lim_{n\to\infty} \exp(-\sum_{i=n}^{\infty}(1-\mathbb{P}(A_i|\mathcal{F}))) \quad a.s.
\end{aligned}$$

Thus for almost every $\omega \in A$ we have

$$0 \leq \mathbb{P}(\mathbb{E}|\mathcal{F})(\omega) \leq \lim_{n\to\infty} \exp(-\sum_{i=n}^{\infty} \mathbb{P}(A_i|\mathcal{F})(\omega) = 0 \quad a.s.$$

Thus

$$\mathbb{P}(\mathbb{E}) = \int_{\Omega} \mathbb{P}(\mathbb{E}|\mathcal{F})dP = \int_{A} \mathbb{P}(\mathbb{E}|\mathcal{F})dP + \int_{A^c} \mathbb{P}(\mathbb{E}|\mathcal{F})dP \leq \mathbb{P}(A^c)$$

so $\mathbb{P}(\mathbb{E}^c) \geq \mathbb{P}(A)$.
On the other hand, following the teasoning given in Lemma (2.3.1) , we state that on the set A^c only finitely many events from the sequence $\{A_n, n \geq 1\}$ hold, so $\mathbb{P}(\mathbb{E}^c) \leq \mathbb{P}(A)$.

2.4 Conditional complete convergence

Complete convergence results are well known for independent random variables(see,e.g.,Gut [8]). The classical results of Hsu, Robbins, Erdés, Baum, and Katz were extended to certain dependent sequences. We will need the following definition of conditional comlete convergence (see Christofides and Hadjikyriakou [4] for details).

Definition 2.4.1 *A sequence of random variables $\{X_n, n \in \mathbb{N}\}$ is said to converge completely given $\mathcal{F}$ to a random variable X if*

$$\sum_{i=1}^{\infty} \mathbb{P}^{\mathcal{F}}(|X_i - X| > \epsilon) < \infty \quad a.s.$$

for any $\mathcal{F}$-measurable random variable ϵ such that $\epsilon > 0 \quad a.s.$

The following set of sequence, which serves purposes of brevity, was first defined by Shen et al.[17]:

$$\mathcal{H} = \left\{ b_n : \sum_{n=1}^{\infty} h^{b_n} < \infty \quad for \quad every \quad 0 < h < 1 \right\}.$$

2.5 Conditional central limit theorem

The central boundary theorems are at the heart of every probability model, so it is not surprising that this problem was one of the first to be addressed in the literature for associated random variables. Indeed, after the first developments, mainly concerned with the dependency structure itself, the first asymptomatic result was a central theorem limit and a proven principle of invariance in Newman

and Wright [11] for associated and stationary random variables.
The following conditional version of the classical limit theorem has been stated in Prakasa Rao [14] without a proof. Although Grzenda and Zieba [7] gave a proof, it was not completly stringent.

Lemma 2.5.1 *Let X and Y be $\mathcal{F}$-identically distributed random variables, and let ξ be an $\mathcal{F}$ measurable random variable. Then for arbitrary $\mathbb{B}^2$-measurable function $f(x,y)$,*

$$\mathbb{E}^{\mathcal{F}} f(X,\xi) = \mathbb{E}^{\mathcal{F}} f(Y,\xi) \quad a.s.,$$

and similarly for any finite number of random variables.

Theoreme 2.5.1 *Let $\{X_n, n \geq 1\}$ be a sequence of $\mathcal{F}$-independent and $\mathcal{F}$-identically distributed random variables with $\sigma^2_{\mathcal{F}} = \mathbb{E}^{\mathcal{F}}(X_1 - \mathbb{E}^{\mathcal{F}} X_1)^2 < \infty \quad a.s.$*
Then

$$\mathbb{E}^{\mathcal{F}} \exp\left(\frac{S_n - \mathbb{E}^{\mathcal{F}} S_n}{\sqrt{n}\sigma_{\mathcal{F}}}\right) \to e^{-\frac{t^2}{2}} \quad a.s. \tag{2.4}$$

as $n \to \infty$ for every $t \in \mathbb{R}$. In particular,

$$\frac{S_n - \mathbb{E}^{\mathcal{F}} S_n}{\sqrt{n}\sigma_{\mathcal{F}}} \to N(0,1) \quad in\ distrribution. \tag{2.5}$$

Proof Relation (2.5) follows from 2.4 by using the dominated convergence theorem and the continuity theorem for characteristic functions. So we need only to prove (2.4). In view of the $\mathcal{F}$-independence of $X_1, X_2, ..., X_n$ and Lemma (2.5.1), we conclude that

$$\begin{aligned}
\mathbb{E}^{\mathcal{F}} \exp\left(it\frac{S_n - \mathbb{E}^{\mathbb{F}} S_n}{\sqrt{n}\sigma_{\mathcal{F}}}\right) &= \mathbb{E}^{\mathcal{F}} \left\{\prod_{k=1}^{n} \exp\left[\frac{it}{\sqrt{n}\sigma_{\mathcal{F}}}(X_k - \mathbb{E}^{\mathcal{F}} X_k)\right]\right\} \\
&= \prod_{k=1}^{n} \mathbb{E}^{\mathcal{F}} \exp\left[\frac{it}{\sqrt{n}\sigma_{\mathcal{F}}}(X_k - \mathbb{E}^{\mathcal{F}} X_k)\right]
\end{aligned}$$

In view $\mathcal{F}$-identical distribution and Lemma 2.5.1, $\mathbb{E}^{\mathcal{F}} X_1 = \mathbb{E}^{\mathcal{F}} X_2 = ... = \mathbb{E}^{\mathcal{F}} X_n \quad a.s.$ and $\mathbb{E}^{\mathcal{F}} \exp\left(\frac{itX_1}{\sqrt{n}\sigma_{\mathcal{F}}}\right) = \mathbb{E}^{\mathcal{F}} \exp\left(\frac{itX_2}{\sqrt{n}\sigma_{\mathcal{F}}}\right) = ... = \mathbb{E}^{\mathcal{F}} \exp\left(\frac{itX_n}{\sqrt{n}\sigma_{\mathcal{F}}}\right) \quad a.s.$ so that

$$\begin{aligned}
\prod_{k=1}^{n} \mathbb{E}^{\mathcal{F}} \exp\left[\frac{it}{\sqrt{n}\sigma_{\mathcal{F}}}(X_k - \mathbb{E}^{\mathcal{F}} X_k)\right] &= \prod_{k=1}^{n} \mathbb{E}^{\mathcal{F}} \left[\exp\left(\frac{itX_k}{\sqrt{n}\sigma_{\mathcal{F}}}. \exp\left(-\frac{it\mathbb{E}^{\mathcal{F}} X_k}{\sqrt{n}\sigma_{\mathcal{F}}}\right)\right)\right] \\
&= \prod_{k=1}^{n} \left[\mathbb{E}^{\mathcal{F}} \exp\left(\frac{itX_k}{\sqrt{n}\sigma_{\mathcal{F}}}. \exp\left(-\frac{it\mathbb{E}^{\mathcal{F}} X_k}{\sqrt{n}\sigma_{\mathcal{F}}}\right)\right)\right] \\
&= \left[\mathbb{E}^{\mathcal{F}} \exp\left(\frac{itX_1}{\sqrt{n}\sigma_{\mathcal{F}}}\right). \exp\left(-\frac{it\mathbb{E}^{\mathcal{F}} X_1}{\sqrt{n}\sigma_{\mathcal{F}}}\right)\right]^n \\
&= \left[\mathbb{E}^{\mathcal{F}} \exp\left(\frac{it}{\sqrt{n}} \frac{X_1 - \mathbb{E}^{\mathcal{F}} X_1}{\sigma_{\mathcal{F}}}\right)\right]^n \\
&= \left[\phi_{\mathcal{F}}\left(\frac{t}{\sqrt{n}}\right)\right]^n
\end{aligned}$$

where $\phi_{\mathcal{F}}(t)$ is the $\mathcal{F}$-characteristic function corresponding to $(X_1 - \mathbb{E}^{\mathcal{F}} X_1)/\sigma_{\mathcal{F}}$.
However,

$$\phi_{\mathcal{F}}\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} - \frac{t^2}{2n}\epsilon_2^{\mathcal{F}}\left(\frac{t}{\sqrt{n}}\right)$$

where $\epsilon_2^{\mathcal{F}}(t)$ satisfies $\epsilon_2^{\mathcal{F}}(t/n) \to 0 \quad a.s. n \to \infty$ for every $t \in \mathbb{R}$, and therefore

$$\mathbb{E}^{\mathcal{F}}\left[\exp\left(it\frac{S_n - \mathbb{E}^{\mathcal{F}} S_n}{\sqrt{n}\sigma_{\mathcal{F}}}\right)\right] = \left[1 - \frac{t^2}{2n} - \frac{t^2}{2n}\epsilon_2^{\mathcal{F}}\left(\frac{t}{\sqrt{n}}\right)\right]^n \to e^{-\frac{t^2}{2}} a.s.$$

an $n \to \infty$ for fixed $t \in \mathbb{R}$.

2.6 Conditional association

Let X and Y be random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$. Let $\mathcal{F}$-be a sub algebra of $\mathcal{A}$. We define the conditional covariance of X and Y given $\mathcal{F}$ or $\mathcal{F}$-covariance as

$$Cov^{\mathcal{F}}(X, Y) = \mathbb{E}^{\mathcal{F}}[(X - \mathbb{E}^{\mathcal{F}} X)(Y - \mathbb{E}^{\mathcal{F}} Y)]. \tag{2.6}$$

It easy to see that $\mathcal{F}$-covarinace reduces to the ordinary concept of covariance when $\mathcal{F} = \{\emptyset, \Omega\}$.

A set of random variables $\{X_k, 1 \leq k \leq n\}$ is said to be $\mathcal{F}$-associated if for any coordinatewise non-decreasing functions h, g defined on $\mathbb{R}^n$,

$$Cov^{\mathcal{F}}(h(X_1, ..., X_n), g(X_1, ..., X_n)) \geq 0 \quad a.s. \tag{2.7}$$

A sequence of random variables $X_n, n \geq 1$ is said to be $\mathcal{F}$-associated if every finite subset of the sequence $X_n, n \geq 1$ is $\mathcal{F}$-associated.

An example of a $\mathcal{F}$-associated sequence $X_n, n \geq 1$ is obtained by defining $X_n = Z + Y_n, n \geq 1$ where Z and $Y_n, n \geq 1$ are $\mathcal{F}$-independent random variables. It can be shown by standard arguments that

$$Cov^{\mathcal{F}}(X, Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H^{\mathcal{F}}(X, Y) dx dy \quad a.s.$$

where

$$H^{\mathcal{F}}(X, Y) = \mathbb{E}^{\mathcal{F}}[I_{(X \leq x, Y \leq y)}] - \mathbb{E}^{\mathcal{F}}[I(X \leq x)]\mathbb{E}^{\mathcal{F}}[I(Y \leq y)].$$

Proofs of these results can be obtained following the methods used for the study of associated random variables. As a consequence of these covariance inequalities, it should be possible to obtain a central limit theorem for conditionally associated sequences of random variables following the methods in [11]. Note that $\mathcal{F}$-association does not imply association and vice versa. For results on associated random variables, see [13] and [15].

The natural question of relation between the two concepts of association and conditional association arises. The following examples show that the association of random variables does not imply the conditional association, and that the conditional association does not imply the association.

Example 2.6.1 *Let $\Omega = \{1, 2, 3, 4\}$ and $p_i = 1/4$ be the probability assigned to the event $\{i\}$. If the events A_1 and A_2 are defined by $A_1 = \{1, 2\}$ and $A_2 = \{2, 3\}$ and the random variables X_1 and X_2 as follows $X_1 = I_{A_1}$ and $X_2 = I_{A_2}$ where I_A denotes the indicator function of an event A, then X_1 and X_2 are independent, so that the family $\{X_1, X_2\}$ is associated by Theorem 2.1 in [5]*
Let $B = \{4\}$ and let $\mathcal{F} = \{\Omega, B, B^c, \emptyset\}$ be the sub-σ-algebra generated by the event B. We will show that $\{X_1, X_2\}$ is not $\mathcal{F}$-associated. In fact, let $f(\mathcal{X}_1, \mathcal{X}_2) = \mathcal{X}_1$ and $g(\mathcal{X}_1, \mathcal{X}_2) = \mathcal{X}_2$, then functions f

and g are both coordinatewise nondecreasing.
Some simple calculations show that

$$\mathbb{E}^{\mathcal{F}}[f(X_1,X_2)] = \begin{cases} f(0,0), & \omega \in B, \\ \frac{1}{3}[f(0,1)+f(1,0)+f(1,1)], & \omega \in B^c, \end{cases} = \begin{cases} 0 & \omega \in B, \\ \frac{2}{3}, & \omega \in B^c, \end{cases}$$

$$\mathbb{E}^{\mathcal{F}}[g(X_1,X_2)] = \begin{cases} g(0,0), & \omega \in B, \\ \frac{1}{3}[g(0,1)+g(1,0)+g(1,1)], & \omega \in B^c, \end{cases} = \begin{cases} 0 & \omega \in B, \\ \frac{2}{3}, & \omega \in B^c, \end{cases}$$

and

$$\mathbb{E}^{\mathcal{F}}[f(X_1,X_2)g(X_1,X_2)] = \begin{cases} f(0,0)g(0,0), & \omega \in B, \\ \frac{1}{3}[f(0,1)g(0,1)+f(1,0)g(1,0)+f(1,1)g(1,1)], & \omega \in B^c, \end{cases}$$

$$= \begin{cases} 0, & \omega \in B, \\ \frac{1}{3}, & \omega \in B^c, \end{cases}$$

So that $\mathbb{E}^{\mathcal{F}}[f(X_1,X_2)g(X_1,X_2)] < \mathbb{E}^{\mathcal{F}}[f(X_1,X_2)].\mathbb{E}^{\mathcal{F}}[g(X_1,X_2)]$ on B^c with $\mathbb{P}(B^c)=\frac{3}{4}>0$, this indicates that X_1, X_2 is not $\mathcal{F}$-associated.

Example 2.6.2 *Let Ω and p_i be defined as in Example (2.6.1). Define the events A_1 and A_2 by $A_1=\{1,2,3\}$ and $A_2=\{3,4\}$, and the random variables X_1 and X_2 by $X_1=I_{A_1}$ and $X_2=I_{A_2}$. Let $B=\{3,4\}$ and let $\mathcal{F}=\{\Omega,B,B^c,\emptyset\}$ be the sub-σ-algebra generated by the event B. Some simple calculations show that*

$$\mathbb{E}^{\mathcal{F}}(I_{A_1}) = \begin{cases} \mathbb{E}(I_{A_1}|B), & \omega \in B, \\ \mathbb{E}(I_{A_1}|B^c), & \omega \in B^c, \end{cases} = \begin{cases} \frac{1}{2} & \omega \in B, \\ 1, & \omega \in B^c, \end{cases}$$

$$\mathbb{E}^{\mathcal{F}}(I_{A_2}) = \begin{cases} \mathbb{E}(I_{A_2}|B), & \omega \in B, \\ \mathbb{E}(I_{A_2}|B^c), & \omega \in B^c, \end{cases} = \begin{cases} 1 & \omega \in B, \\ 0, & \omega \in B^c, \end{cases}$$

and

$$\mathbb{E}^{\mathcal{F}}(I_{A_1A_2}) = \begin{cases} \mathbb{E}(I_{A_1A_2}|B), & \omega \in B, \\ \mathbb{E}(I_{A_1A_2}|B^c), & \omega \in B^c, \end{cases} = \begin{cases} \frac{1}{2} & \omega \in B, \\ 0, & \omega \in B^c, \end{cases}$$

so that $\mathbb{E}^{\mathcal{F}}I_{A_1A_2} < \mathbb{E}^{\mathcal{F}}I_{A_1}.\mathbb{E}^{\mathcal{F}}I_{A_2}$, further

$$\mathbb{E}^{\mathcal{F}}I(X_1 \le \mathcal{X}_1, X_2 \le \mathcal{X}_2) \quad a.s.$$

for all real numbers $\mathcal{X}_1$ and $\mathcal{X}_2$, this indicates that X_1 and X_2 are $\mathcal{F}$-independent, and therefore $\{X_1,X_2\}$ is $\mathcal{F}$-associated.
We claim that the family $\{X_1,X_2\}$ is not associated. For this purpose, let functions f and g be defined as in Example (2.6.1), we have

$$\mathbb{E}[f(X_1,X_2)] = \frac{1}{4}[f(0,1)+2f(1,0)+f(1,1)] = \frac{3}{4}$$

$$\mathbb{E}[g(X_1,X_2)] = \frac{1}{4}[g(0,1)+2g(1,0)+g(1,1)] = \frac{1}{2}$$

and

$$\mathbb{E}[f(X_1,X_2)g(X_1,X_2)] = \frac{1}{4}[f(0,1)g(0,1)+2f(1,0)g(1,0)+f(1,1)g(1,1)] = \frac{1}{4}$$

so that

$$\mathbb{E}[f(X_1,X_2)g(X_1,X_2)] = \mathbb{E}[f(X_1,X_2)].\mathbb{E}[g(X_1,X_2)]$$

which indicates that $\{X_1, X_2\}$ *is not associated. The association of random variables may be inherited by the conditional association, the following is such an example.*

Example 2.6.3 *Let* Ω, p_i, X_1, X_2 *be defined as in Example* (2.6.1)*. We have shown that* $\{X_1, X_2\}$ *is associated.*
Let $B = \{3,4\}$ *and let* $\mathcal{F} = \{\Omega, B, B^c, \emptyset\}$ *be the sub-*σ*-algebra generated by the event* B*. For every pair of coordinatewise nondecreasing* f *and* g *defined on* $\mathbb{R}^2$ *with* $\mathbb{E}[(X_1,X_2)]$ *finite, some simple calculations show that*

$$\mathbb{E}^{\mathcal{F}}[f(X_1,X_1)] = \left\{ \begin{array}{ll} \frac{1}{2}[f(0,0)+f(0,1)], & \omega \in B, \\ , \frac{1}{2}[f(1,0)+f(1,1)] & \omega \in B^c, \end{array} \right\},$$

$$\mathbb{E}^{\mathcal{F}}[g(X_1,X_1)] = \left\{ \begin{array}{ll} \frac{1}{2}[g(0,0)+g(0,1)], & \omega \in B, \\ \frac{1}{2}[g(1,0)+g(1,1)], & \omega \in B^c, \end{array} \right.$$

and

$$\mathbb{E}^{\mathcal{F}}[f(X_1,X_1)g(X_1,X_1)] = \left\{ \begin{array}{ll} \frac{1}{2}[f(0,0)g(0,0)+f(0,1)g(0,1)], & \omega \in B, \\ , \frac{1}{2}[f(1,0)g(1,0)+f(1,1)g(1,1)] & \omega \in B^c, \end{array} \right\},$$

It is easy to see that

$$\mathbb{E}^{\mathcal{F}}[f(X_1,X_1)g(X_1,X_1)] \geq \mathbb{E}^{\mathcal{F}}[f(X_1,X_1)]\mathbb{E}^{\mathcal{F}}[g(X_1,X_1)]$$

since $[f(0,0)-f(0,1)][g(0,0)-g(0,1)] \geq 0$ *and* $[f(1,0)-f(1,1)][g(1,0)-g(1,1)] \geq 0$*, and hence* $\{X_1, X_2\}$ *is* $\mathcal{F}$*-associated.*

Remark 2.6.1 *As it was pointed out earlier, conditional independence does not imply independence and vice versa. Hence one does have to derive limit theorems under conditioning if there is a need for such results even though the results and proofs of such results may be analogous to those under the non-conditioning set up. This was one of the reasons for developing results for conditional sequences in the earlier sections.A concrete example where conditionallimit theorems are useful is in the study of statistical inference for non-ergodic models as discussed in [1] and [2].For instance, if one wants to estimate the mean off-spring* θ *for a Galton-Watson Branching process, the asymptotic properties of the maximum likelihood estimator depend on the set of non-extinction.Conditional limit theorems under a martingale set up have been used for the study of asymptotic properties for maximum likelihood estimators of parameters in a stochastic process framework where the Fisher information need not be additive and might not increase to infinity or might even increase to infinity at an exponential rate (cf. [1]; [12];[6]).*

Bibliography

[1] Basawa, I. V., Prakasa Rao, B. L. S. (1980). Statistical inference for stochastic processes. London: Academic.

[2] Basawa, I. V., Scott, D. (1983). Asymptotic optimal inference for non-ergodic models. Lecture Notes in Statistics, Vol. **17**. New York: Springer.

[3] Chow, Y. S., Teicher, H. (1978). Probability theory: independence, interchangeability, martingales. New York: Springer.

[4] Christofides, TC., Hadjikyriakou, M., (2013). Conditional demimartingales and related results.J.Math.Anal.Appl. **398**,380-391

[5] Esary, J., Proschan, F., Walkup, D., (1967). Association of random variables with applications, Ann. Math. Statist. **38** 1466-1474.

[6] Guttorp, P. (1991). Statistical inference for branching processes. New York: Wiley.

[7] Grzenda, W., Zieba, W. (2008) Conditional central limit theorems, Int. Math. Forum **3**, no 29-32, 1521-1528.

[8] Gut, A., (2005). Probability:AGraduateCourse.Springer,Berlin

[9] Kruk, L., Zieba, W., (1993) On almost sure convergence asymptotic martingales, Annales Universitas Marie Curie-Sklodowska, **47** ,No. 8, 82-89.

[10] Majerak, D., Nowak, W., Zieba, W., (2005). Conditional strong law of large number. International Journal of Pure and Applied Mathematics, **20**, 143-157.

[11] Newman, C. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. In Y. L. Tong (Ed.), Inequalities in statistics and probability (pp. 127-140).

[12] Prakasa Rao, B. L. S. (1999b). Semimartingales and their statistical inference. Boca Raton: CRC Press

[13] Prakasa Rao, B. L. S., Dewan, I. (2001). Associated sequences and related inference problems, In C. R. Rao, D. N. Shanbag (Eds.), Handbook of statistics, 19, stochastic Processes: theory and methods (pp. 693-728). Amsterdam: North Holland.

[14] Prakasa Rao, B. L. S., (2009) Conditional independence, conditional mixing and conditional association, Ann. Inst. Statist. Math. **61**, no. 1, 441-460.

[15] Roussas, G. G. (1999). Positive and negative dependence with some statistical applications. In S. Ghosh (Ed.), Asymptotics, nonparametrics and time series . New York: Marcel Dekker, (pp. 757-788).

[16] Stoyanov, J.M. (1989) Counterexamples in probability. John Wiley and Sons, Chichester, New York, Brisbane, Torento, Singapore.

[17] Shen, A., Hu, S., Volodin, A., Wang, X., (2011). Some exponential inequalities for acceptable random variables and complete convergence.J.Inequal. Appl,2011, 142.

This article is published in the International Journal of Applied Mathematics and Statistics

Chapter 3

New tail probability inequalities for widely orthant dependent random variables sequence, application to hazard estimator

Summary

3.1 Introduction 49

3.2 Some lemmas 50

3.3 Main Results and Proofs 54

abstract: Tail probability inequalities have be an important tool in probability and statistics. Version of Bernstein type inequalities have proved for independent and for some dependence structure. We prove a new Tail probability inequality for the distributions of sums of widely orthant dependent (WOD, in short) random variables, and obtain of complete convergence for kernel estimators of density and hazard functions, under some suitable conditions.

3.1 Introduction

Hazard estimation is quite an important problem in several fields of applied statistics (medicine,econometric, seismic risk, reliability, etc.). Nonparametric estimation of the hazard function started with Watson and Leadbetter (1964a,b) who introduced the kernel estimator, and from that time on, a lot of papers on this topic have come out in the nonparametric literature. The estimation of the hazard function, in the nonparametric case, has been widely studied in the literature when the variables are of finite dimensions.

Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables with an unknown marginal probability density function $f(x)$ and distribution function $F(x)$. Assume that $K(x)$ is a known kernel function, the kernel estimate of $f(x)$ and the empirical distribution function of $F(x)$ are given by

$$\widetilde{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), \quad F_n(x) = n^{-1}\sum_{i=1}^{n}\mathbb{I}(X_i < x), \tag{3.1}$$

where $\{h_n, n \geq 1\}$ is a sequence of positive bandwidths tending to zero as $n \to +\infty$. The function $\widetilde{f}_n(.)$ is the well known kernel estimator of $f(.)$ defined by Rosenblatt[6] and Parzen[5]. $\mathbb{I}(.)$ is the indicator of the event specified in the parentheses. The hazard rate of function can be written as ration between the pdf $f(.)$ and the survivor function $S(.) = 1 - F(.)$ as follows :

$$H(x) = f(x)/S(x),$$

and it can be estimated by

$$\widetilde{H}_n(x) = \frac{\widetilde{f}_n(x)}{1 - F_n(x)}. \tag{3.2}$$

In all these works, we establish the complete convergence of hazard function estimator based on widely orthant dependent (WOD in short) random variables by using the new exponential probability.It is well known that the probability limit theorem and its applications for independent random variables have been studied by many authors, while the assumption of independence is not reasonable in real practice.If the independent case is classical in the literature, the treatment of dependent random variables is more recent.Widely orthant dependence (WOD) is one of dependence structure. Now we recall the definition of WOD. The concept of widely orthant dependent was introduced by Wang and Cheng [7] for risk model. They studied the basic renewal theorems for random walks with WOD dependent increments.

By definition, r.v.s $\{X_i, i \geq 1\}$, are said to be widely upper orthant dependent (WUOD) if for each $n \geq 1$, there exists a positive number $g_U(n)$ such that, for all $x_i \in (-\infty, +\infty), i = 1, ..., n$

$$\mathbb{P}(X_1 > x_1, X_2 > x_2, ..., X_n > x_n) \leq g_U(n)\Pi_{i=1}^{n}\mathbb{P}(X_i > x_i); \tag{3.3}$$

they are said to be widely lower orthant dependent (WLOD) if for each $n \geq 1$, there exists some finite positive number $g_L(n)$ such that, for all $x_i \in (-\infty, +\infty), i = 1, ..., n$,

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n) \leq g_L(n)\Pi_{i=1}^{n}\mathbb{P}(X_i \leq x_i); \tag{3.4}$$

and they are said to be widely orthant dependent (WOD) if they are both WUOD and WLOD. WUOD, WLOD and WOD r.v.s. are called by a joint name widely dependent r.v.s. and $g_U(n) \geq 1, g_L(n) \geq 1, n \geq 2$, are called dominating coefficients.Clearly, we have $g_u(1) = g_L(1) = 1$.

The concept of complete convergence was introduced by Hsu and Robbins[3] as follows: a sequence $X_n, n \geq 1$ of random variables converges completely to a constant ϑ if for all $\varepsilon > 0$,

$$\sum_{i=1}^{n} \mathbb{P}(|X_i - \vartheta| \geq \varepsilon) < +\infty.$$

By the Borel.Cantelli lemma, this implies that $X_n \to \vartheta$ a.s., and so complete convergence is a stronger concept than a.s. convergence.

In this paper,we attempt to establish a new probability inequality and to derive the complete convergence for the estimators of the probability density estimator (3.1) and hazard functions under strictly stationary WOD random variables. The results obtained in the paper improve and extend the corresponding ones of Li and Yang [4] for NA samples. We will also study the complete consistency for the estimator (3.1) under some mild conditions. The main results are presented in Section 2. Some lemmas are provided in Section 3. The proofs are given in Section 4. $\mathcal{C}(f)$ denotes all the continuity points of function f and $\mathcal{C}^2(f)$ stands for a point set in where the second-order derivative f'' exists and is bounded and continuous.

3.2 Some lemmas

To prove the main results of the paper, we need the following important lemmas. The first one is a basic property for WOD random variables, which was obtained by Wang et al. [4].

Lemma 3.2.1

(i) Let $\{X_n, n \geq 1\}$ be WLOD (WUOD) with dominating coefficients $g_L(n), n \geq 1(g_U(n), n \geq 1)$,

- *if $\{f_n(.), n \geq 1\}$ are nondecreasing, then $\{f_n(X_n), n \geq 1\}$ are still WLOD (WUOD) with dominating coefficients $g_L(n), n \geq 1(g_U(n), n \geq 1)$;*
- *if $\{f_n(.), n \geq 1\}$ are nonincreasing, then $\{f_n(X_n), n \geq 1\}$ are WUOD (WLOD) with dominating coefficients $g_L(n), n \geq 1(g_U(n), n \geq 1)$.*

(ii) If $\{X_n, n \geq 1\}$ are nonnegative and WUOD with dominating coefficients $g_U(n), n \geq 1$, then for each $n \geq 1$,

$$\mathbb{E}\prod_{i=1}^{n} X_i \leq g_U(n) \prod_{i=1}^{n} \mathbb{E}X_i.$$

In particular, if $\{X_n, n \geq 1\}$ are WUOD with dominating coefficients $g_U(n), n \geq 1$, then for each $n \geq 1$ and any $\lambda > 0$,

$$\mathbb{E}\exp\{\lambda \sum_{i=1}^{n} X_i\} \leq g_U(n) \prod_{i=1}^{n} \mathbb{E}\exp\{\lambda X_i\}.$$

By Lemma (3.2.1), we can get the following corollary immediately.

Corollary 3.2.1 *Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables.*

(i) If $\{f_n(.), n \geq 1\}$ are all nondecreasing (or all nonincreasing), then $\{f_n(X_n), n \geq 1\}$ are still WOD.

(ii) For each $n \geq 1$ and any $\lambda \in \mathbb{R}$,

$$\mathbb{E}\exp\{\lambda \sum_{i=1}^{n} X_i\} \leq g(n) \prod_{i=1}^{n} \mathbb{E}\exp\{\lambda X_i\}.$$

Proof. For $\lambda > 0$, it is easy to see that λX_i and $\lambda \sum_{j=i+1}^{n} X_j$ are WOD by the definition. Which implies that $\exp(\lambda X_i)$ and $\exp(\lambda \sum_{j=i+1}^{n} X_j)$ are also WOD for $i = 1, 2, \ldots, n-1$, by Lemma 3.2.1 (i). It follows from corollary (3.2.1) we obtain

$$\begin{aligned}
\mathbb{E}(\prod_{i=1}^{n} e^{\lambda X_i}) &= \mathbb{E}(\exp(\lambda X_1)\exp(\lambda \sum_{j=2}^{n} X_j)), \\
&\leq g_1(n)\mathbb{E}[\exp(\lambda X_1)]\mathbb{E}[\exp(\lambda \sum_{j=2}^{n} X_j)], \\
&= g_1(n)\mathbb{E}[\exp(\lambda X_1)]\mathbb{E}[\exp(\lambda X_2)\exp(\lambda \sum_{j=3}^{n} X_j)], \\
&\leq g_1(n)g_2(n)\mathbb{E}[\exp(\lambda X_1)]\mathbb{E}[\exp(\lambda X_2)]\mathbb{E}[\exp(\lambda \sum_{j=3}^{n} X_j)], \\
&\leq \prod_{i=1}^{n-1} g_i(n) \prod_{i=1}^{n} (\mathbb{E}e^{\lambda X_i}), \\
&= g(n) \prod_{i=1}^{n} (\mathbb{E}e^{\lambda X_i}),
\end{aligned}$$

where $g(n) = \prod_{i=1}^{n-1} g_i(n)$.

The following lemma is very useful in the proof of Lemma (3.2.3).

Lemma 3.2.2 *[8] For any $x \in \mathbb{R}$, we have*

$$\exp(x) \leq 1 + x + |x|^{1+\alpha} \exp(2|x|), \ 0 < \alpha \leq 1.$$

Lemma 3.2.3 *Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables with $\mathbb{E}X_n = 0$ for each $n \geq 1$. If there exists a sequence of positive numbers $\{a_n, n \geq 1\}$ such that $|X_i| \leq a_i$ for each $i \geq 1$, then for any $\lambda > 0$ and $0 < \alpha \leq 1$,*

$$\mathbb{E}\exp\left\{\lambda \sum_{i=1}^{n} X_i\right\} \leq g(n) \exp\left\{\lambda^{1+\alpha} \sum_{i=1}^{n} e^{2\lambda a_i} \mathbb{E}|X_i|^{1+\alpha}\right\}. \tag{3.5}$$

Proof. By lemma (3.2.2), for all $x \in \mathbb{R}$ and $0 < \alpha \leq 1$, $\exp(x) \leq 1 + x + |x|^{1+\alpha} \exp(2|x|)$. Thus, by $\mathbb{E}(X_i) = 0$ and $|X_i| \leq a_i$ for each $i \geq 1$, we have

$$\begin{aligned}
\mathbb{E}\exp(\lambda X_i) &\leq \mathbb{E}\left\{1 + \lambda X_i + |\lambda X_i|^{1+\alpha} \exp(2|\lambda X_i|)\right\} \\
&= 1 + \lambda\mathbb{E}(X_i) + \lambda^{1+\alpha}\mathbb{E}\left\{|X_i|^{1+\alpha} \exp(2|\lambda X_i|)\right\} \\
&= 1 + \lambda^{1+\alpha}\mathbb{E}\left\{|X_i|^{1+\alpha} \exp(2|\lambda X_i|)\right\} \\
&\leq 1 + \lambda^{1+\alpha}\mathbb{E}\left\{|X_i|^{1+\alpha} \exp(2\lambda a_i)\right\} \\
&= 1 + \lambda^{1+\alpha} \exp(2\lambda a_i)\mathbb{E}\left\{|X_i|^{1+\alpha}\right\} \\
&\leq \exp\left\{\lambda^{1+\alpha} \exp(2\lambda a_i)\mathbb{E}\left\{|X_i|^{1+\alpha}\right\}\right\} \qquad (3.6)
\end{aligned}$$

(using $1 + y \leq \exp(y)$ for all $y \in \mathbb{R}$) for any $\lambda > 0$. By lemma (3.2.1) and (3.6), we have can see that

$$\begin{aligned} \mathbb{E}\exp\left\{\lambda\sum_{i=1}^{n}X_i\right\} &\leq g(n)\prod_{i=1}^{n}\mathbb{E}\exp\{\lambda X_i\} &(3.7)\\ &\leq g(n)\exp\left\{\lambda^{1+\alpha}\sum_{i=1}^{n}e^{2\lambda a_i}\mathbb{E}|X_i|^{1+\alpha}\right\}. &(3.8)\end{aligned}$$

The lemma is thus proved.

Lemma 3.2.4 *Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables with $\mathbb{E}X_n = 0$ for each $n \geq 1$. If there exists a sequence of positive numbers $\{a_n, n \geq 1\}$ such that $|X_i| \leq a_i$ for each $i \geq 1$, then for any $\lambda > 0$, $0 < \alpha \leq 1$ and $\epsilon > 0$*

$$\mathbb{P}(|\sum_{i=1}^{n}X_i| \geq \epsilon) \leq 2g(n)\exp\left\{-\lambda\epsilon + \lambda^{1+\alpha}\sum_{i=1}^{n}e^{2\lambda a_i}\mathbb{E}|X_i|^{1+\alpha}\right\}. \quad (3.9)$$

Proof. By Markov's inequality and lemma (3.2.3), we can see that

$$\begin{aligned} \mathbb{P}(\sum_{i=1}^{n}X_i \geq \epsilon) &\leq \exp(-\lambda\epsilon)\mathbb{E}\exp\left\{\lambda\sum_{i=1}^{n}X_i\right\}\\ &\leq \exp(-\lambda\epsilon)g(n)\prod_{i=1}^{n}\mathbb{E}\exp\{\lambda X_i\}\\ &\leq g(n)\exp\left\{-\lambda\epsilon + \lambda^{1+\alpha}\sum_{i=1}^{n}e^{2\lambda a_i}\mathbb{E}|X_i|^{1+\alpha}\right\}. &(3.10)\end{aligned}$$

The desired result follos by remplacing X_i by $-X_i$ in (3.10). This completes the proof of the lemma.

Lemma 3.2.5 *Suppose that (H_1) holds, then for all $x \in C^2(f)$,*

$$\lim_{h\to 0}\int_{\mathbb{R}}K(u)f(x-hu)\,du = f(x).$$

Lemma 3.2.6 *Suppose that (H_1) holds, then for all $x \in \mathcal{C}^2(f)$,*

$$h_n^{-2}|\mathbb{E}\widetilde{f}_n(x) - f(x)| \leq C < +\infty.$$

Proof. By Taylor's expansion and $(H_1) - (H_2)$, for any $0 < \theta < 1$ we can see that

$$\begin{aligned} \mathbb{E}\widetilde{f}_n(x) &= \mathbb{E}(\frac{1}{nh_n}\sum_{i=1}^{n}K(\frac{x-X_i}{h_n}))\\ &= \frac{1}{nh_n}\sum_{i=1}^{n}\mathbb{E}K(\frac{x-X_i}{h_n})\\ &= \frac{1}{nh_n}\sum_{i=1}^{n}\int_{\mathbb{R}}K(u)f(x-h_nu)\,du\\ &= \frac{1}{nh_n}\sum_{i=1}^{n}\int_{\mathbb{R}}K(u)[f(x) - f^{'}(x)h_nu + \frac{f^{''}(x-\theta h_nu)}{2}(h_nu)^2]\,du\\ &= f(x) + \frac{h_n^2}{2}\int_{\mathbb{R}}K(u)f^{''}(x-\theta h_nu)u^2\,du \to f(x) + \frac{h_n^2}{2}f^{''}(x)\int_{\mathbb{R}}K(u)u^2\,du. &(3.11)\end{aligned}$$

By assumption (H_1) and (3.11), we get the result.

Lemma 3.2.7 *Let $\{X_n; n \geq 1\}$ be a sequence of WOD random variables with unknown distribution function $F(x)$ and bounded probability density function $f(x)$. Let $F_n(x)$ be the empirical distribution function. If $\widetilde{\xi_n} = \frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}}(ng(n))}{n^{\frac{\alpha}{(1+\alpha)}}} \to 0$ for some $p > 1$ and $0 < \alpha \leq 1$, then*

$$\sup_x |F_n(x) - F(x)| = O(\widetilde{\xi_n}), \ completely.$$

In particular, if $g(n) = O(n^\delta)$ for somme $\delta \geq 0$, then

$$\sup_x |F_n(x) - F(x)| = O(\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}} n}{n^{\frac{\alpha}{(1+\alpha)}}}), \ completely.$$

Proof. Let $F(x_{ni}) = i/n$ for $n \geq 3$ and $1 \leq i \leq n-1$. By lemma 2 in Yang [4] we have that

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \leq \max_{1\leq j\leq n-1} |F_n(x_{nj}) - F(x_{nj})| + 2/n. \tag{3.12}$$

Noting that $n\widetilde{\xi_n} \to \infty$, then for any positive constant B_1, we have that $2/n < B_1\widetilde{\xi_n}/2$ for all n large enough. Then if follows from 3.12 that

$$\begin{aligned} \mathbb{P}(\sup_x |F_n(x) - F(x)| \geq B_1\widetilde{\xi_n}) &\leq \mathbb{P}(\max_{1\leq j\leq n-1} |F_n(x_{nj}) - F(x_{nj})| \geq B_1\widetilde{\xi_n}/2) \\ &\leq \sum_{j=1}^{n-1} \mathbb{P}(|F_n(x_{nj}) - F(x_{nj})| \geq B_1\widetilde{\xi_n}/2). \end{aligned} \tag{3.13}$$

Let $\widetilde{Y_i} = I(X_i < x_{nj}) - \mathbb{E}I(X_i < x_{nj})$. By Lemma 3.2.1, $\{\widetilde{Y_i}; i \geq 1\}$ is still a sequence of WOD with $\mathbb{E}\widetilde{Y_i} = 0, |\widetilde{Y_i}| \leq 2$.Thus by choosing $\lambda = \left(\frac{B_1 bp}{2^{p(1+\alpha)+1}(1+\alpha)}\right)^{\frac{1}{p(1+\alpha)-1}} \left(\frac{\widetilde{\xi_n}}{n^{p-1}}\right)^{\frac{1}{p(1+\alpha)-1}}$ in Lemma (3.2.4) we have that for all n large enough,

$$\begin{aligned} \mathbb{P}(|F_n(x_{nj}) - F(x_{nj})| \geq B_1\widetilde{\xi_n}/2) &\leq \mathbb{P}(|\sum_{j=1}^{n} \widetilde{Y_i}| \geq B_1\widetilde{\xi_n}/2) \\ &\leq 2g(n) \exp\left\{-\lambda B_1\widetilde{\xi_n}/2 + \lambda^{1+\alpha} \sum_{i=1}^{n} e^{2\lambda a_i} \mathbb{E}|\widetilde{Y_i}|^{1+\alpha}\right\} \\ &\leq 2g(n) \exp\left\{-\lambda B_1\widetilde{\xi_n}/2 + \lambda^{1+\alpha} n e^{2\lambda \max_{1\leq i\leq n} a_i} \mathbb{E}|\widetilde{Y_i}|^{1+\alpha}\right\} \\ &\leq 2g(n) \exp\{-\lambda B_1\widetilde{\xi_n}/2 + \lambda^{1+\alpha} n e^{2\lambda \max_{1\leq i\leq n} a_i} \\ &\quad 2^\alpha(\mathbb{E}(I(X_i < x_{nj})^{1+\alpha}) + (\mathbb{E}I(X_i < x_{nj}))^{1+\alpha})\} \\ &\quad \text{(using the inequality } (a+b)^{\alpha+1} \leq \max(2^\alpha, 1)(a^{\alpha+1} + b^{\alpha+1})) \\ &\leq 2g(n) \exp\{-\lambda B_1\widetilde{\xi_n}/2 + \lambda^{1+\alpha} n e^{2\lambda \max_{1\leq i\leq n} a_i} \\ &\quad 2^\alpha 2\mathbb{E}(I(X_i < x_{nj})^{1+\alpha})\} \\ &\quad \text{(using the Hölder's inequality } (\mathbb{E}(Z))^{1+\alpha} \leq \mathbb{E}(Z^{1+\alpha})) \\ &\leq 2g(n) \exp\{-\lambda B_1\widetilde{\xi_n}/2 + \lambda^{1+\alpha} n e^{2\lambda \max_{1\leq i\leq n} a_i} 2^{\alpha+1}\}. \end{aligned}$$

Let $p > 1$. It is well known that

$$uv = \inf_{b>0} \left\{\frac{1}{pb} u^p + \frac{1}{q} b^{q/p} v^q\right\} \ \text{for} u > 0, v > 0 \text{ and } 1/p + 1/q = 1.$$

This yields the inequality

$$\lambda^{1+\alpha} n e^{2\lambda \max_{1\leq i\leq n} a_i} 2^{\alpha+1} \leq \frac{1}{pb}\lambda^{p(1+\alpha)} n^p 2^{p(\alpha+1)} + \frac{1}{q} b^{q/p} e^{2q\lambda \max_{1\leq i\leq n} a_i}. \tag{3.14}$$

We can thus conclude that for every $p > 1$ and $0 < \alpha \leq 1$, there for all $\lambda > 0$, such that

$$\begin{aligned}\mathbb{P}(|F_n(x_{nj}) - F(x_{nj})| \geq B_1\widetilde{\xi_n}/2) &\leq 2g(n)\exp\{-\lambda B_1\widetilde{\xi_n}/2 + \frac{1}{pb}\lambda^{p(1+\alpha)} n^p 2^{p(\alpha+1)}\} \\ &\times \exp\{\frac{1}{q} b^{q/p} e^{2q\lambda \max_{1\leq i\leq n} a_i}\} \\ &= 2g(n)\exp\{\frac{1}{q} b^{q/p} e^{2q\lambda \max_{1\leq i\leq n} a_i}\}\exp(\Phi(\lambda, n)).\end{aligned} \tag{3.15}$$

The equation $\frac{\partial \Phi(\lambda,n)}{\partial \lambda} = 0$ has the unique solution

$$\lambda = \left(\frac{B_1 bp}{2^{p(1+\alpha)+1}(1+\alpha)}\right)^{\frac{1}{p(1+\alpha)-1}} \left(\frac{\widetilde{\xi_n}}{n^{p-1}}\right)^{\frac{1}{p(1+\alpha)-1}} \tag{3.16}$$

which minimizes $\Phi(\lambda, n)$. Then from (3.15),(3.16) and taking $2q\lambda \max_{1\leq i\leq n} a_i \leq 1$ it follows that

$$\begin{aligned}\mathbb{P}(|F_n(x_{nj}) - F(x_{nj})| \geq B_1\widetilde{\xi_n}/2) &\leq 2\exp\{\frac{1}{q} b^{q/p} e\} g(n)\exp\{-\mu_1 n^{\frac{p\alpha}{p(1+\alpha)-1}} \widetilde{\xi_n}^{\frac{p(1+\alpha)}{p(1+\alpha)-1}}\} \\ &\leq 2\exp\{\frac{1}{q} b^{q/p} e\} g(n)\exp\{-\mu_1 C_0 \log(ng(n))\} \\ &= 2\exp\{\frac{1}{q} b^{q/p} e\} g(n)\frac{1}{(ng(n))^{\mu_1 C_0}}\end{aligned} \tag{3.17}$$

Where $\mu_1 = (1 - \frac{1}{p(1+\alpha)})\frac{B_1^{\frac{p(1+\alpha)}{p(1+\alpha)-1}}}{2^{\frac{2p(1+\alpha)}{p(1+\alpha)-1}}}(bp/(1+\alpha))^{\frac{1}{p(1+\alpha)-1}}$ Recall that $g(n) \geq 1$. Taking μ_1 sufficiently large such that $\mu_1 C_0 > 2$ by (3.13) and ((3.17)) we have that

$$\begin{aligned}\sum_{n=1}^{\infty}\mathbb{P}(\sup_{-\infty<x<+\infty}|F_n(x) - F(x)| \geq B_1\widetilde{\xi_n}) &\leq 2\exp\{\frac{1}{q} b^{q/p} e\}\sum_{n=1}^{\infty}\sum_{i=1}^{n-1} g(n)\frac{1}{(ng(n))^{\mu_1 C_0}} \\ &= 2\exp\{\frac{1}{q} b^{q/p} e\}\sum_{n=1}^{\infty}(n-1)g(n)\frac{1}{(ng(n))^{\mu_1 C_0}} \\ &\leq 2\exp\{\frac{1}{q} b^{q/p} e\}\sum_{n=1}^{\infty}(ng(n))\frac{1}{(ng(n))^{\mu_1 C_0}} \\ &= 2\exp\{\frac{1}{q} b^{q/p} e\}\sum_{n=1}^{\infty}\frac{1}{(ng(n))^{\mu_1 C_0 - 1}} \\ &< +\infty.\end{aligned}$$

This complete the proof of the lemma.

3.3 Main Results and Proofs

In this section, we will present the complete convergence for kernel estimators of density and hazard functions. We adopt the following assumptions.

(H_1) $\int_{-\infty}^{+\infty} K(u)\mathrm{d}u = 1, \int_{-\infty}^{+\infty} uK(u)\mathrm{d}u = 0, \int_{-\infty}^{+\infty} u^2K(u)\mathrm{d}u < \infty, K(u) \in L_1;$
(H_2) The bandwidths h_n satisfy that $h_n \downarrow 0$ and $nh_n \to \infty$ as $n \to \infty$. Now we state our main results as follows.

Theoreme 3.3.1 *Suppose that* $(H_1) - (H_2)$ *hold. Let* $\{X_n; n \geq 1\}$ *be a sequence of strictly stationary of WOD random variables with* $g(n) = O(n^\delta)$ *for some* $\delta \geq 0$. *Suppose that the kernel* $K(.)$ *is a bounded monotone density function and the bandwidth* $h_n = O(n^{-\frac{\alpha}{3(1+\alpha)}} \log^{\frac{p(1+\alpha)-1}{3p(1+\alpha)}} n)$ *for some* $p > 1$ *and* $0 < \alpha \leq 1$. *Then for any* $x \in \mathcal{C}^2(f)$.

$$|\widetilde{f}_n(x) - f(x)| = O\left(\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}} n}{n^{\frac{\alpha}{(1+\alpha)}} h_n}\right), \textit{ completely.} \tag{3.18}$$

Proof. Set $Y_i = h_n^{-1}\left[K(\frac{x-X_i}{h_n}) - \mathbb{E}K(\frac{x-X_i}{h_n})\right]$ for $1 \leq i \leq n$. Since $K(.)$ is a bounded monotone density function, then $\{Y_i, i \geq 1\}$ is still a sequence of WOD random variables. Moreover, it follows from $0 < h_n \downarrow 0$ that there exists some positive constant $\widetilde{C}$ such that $\max_{1\leq i\leq n} |Y_i| \leq \widetilde{C}h_n^{-1}$. By lemma (3.2.4) and elementary inequality $(a+b)^{1+\alpha} \leq \max(2^\alpha, 1)(a^{1+\alpha} + b^{1+\alpha})$ we have that

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}|Y_i|^{1+\alpha} &= \sum_{i=1}^n \mathbb{E}\left|h_n^{-1}\left[K\left(\frac{x-X_i}{h_n}\right) - \mathbb{E}K\left(\frac{x-X_i}{h_n}\right)\right]\right|^{1+\alpha} \\
&\leq \sum_{i=1}^n \frac{2^\alpha}{h_n^{1+\alpha}}\mathbb{E}\left[\left(K\left(\frac{x-X_i}{h_n}\right)\right)^{1+\alpha} + \left(\mathbb{E}K\left(\frac{x-X_i}{h_n}\right)\right)^{1+\alpha}\right] \\
&= \frac{2^\alpha}{h_n^{1+\alpha}}\sum_{i=1}^n \left[\mathbb{E}\left(K\left(\frac{x-X_i}{h_n}\right)\right)^{1+\alpha} + \left(\mathbb{E}K\left(\frac{x-X_i}{h_n}\right)\right)^{1+\alpha}\right] \\
&\leq \frac{2^{\alpha+1}}{h_n^{1+\alpha}}\sum_{i=1}^n \mathbb{E}\left(K\left(\frac{x-X_i}{h_n}\right)\right)^{1+\alpha} \quad \text{(by hölder's inequality)} \\
&= \frac{2^{\alpha+1}}{h_n^{1+\alpha}} n\mathbb{E}\left(K\left(\frac{x-X_i}{h_n}\right)\right)^{1+\alpha} \\
&= \frac{2^{\alpha+1}}{h_n^{1+\alpha}} n\int_{\mathbb{R}}\left(K\left(\frac{x-u}{h_n}\right)\right)^{1+\alpha} f(u)\, du \\
&= \frac{2^{\alpha+1}}{h_n^{1+\alpha}} n\int_{\mathbb{R}} K(u)f(x-h_nu)\, du \\
&\leq \widetilde{C}n\frac{2^{\alpha+1}}{h_n^{1+\alpha}}.
\end{aligned}$$

Set $\zeta_n = \frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}} n}{n^{\frac{\alpha}{(1+\alpha)}} h_n}$. Applying lemma 3.2.4 with $\lambda = \left(\frac{A_1 n\zeta_n h_n^{p(1+\alpha)}}{\widetilde{C_2}n^p p(1+\alpha)}\right)^{\frac{1}{p(1+\alpha)-1}}$, where A_1 is some positive constant, for all n large enough, then we get that

$$\begin{aligned}
\mathbb{P}(|\widetilde{f}_n(x) - \mathbb{E}\widetilde{f}_n(x)| \geq A_1\zeta_n) &\leq \mathbb{P}(|\sum_{j=1}^n Y_i| \geq A_1 n\zeta_n) \\
&\leq 2g(n)\exp\left\{-\lambda A_1 n\zeta_n + \lambda^{1+\alpha}\sum_{i=1}^n e^{2\lambda a_i}\mathbb{E}|Y_i|^{1+\alpha}\right\} \\
&\leq 2g(n)\exp\left\{-\lambda A_1 n\zeta_n + \lambda^{1+\alpha}e^{2\lambda\max_{1\leq i\leq n} a_i}\mathbb{E}|Y_i|^{1+\alpha}\right\} \\
&\leq 2g(n)\exp\{-\lambda A_1 n\zeta_n + \lambda^{1+\alpha}n2^{1+\alpha}\widetilde{C}h_n^{-(1+\alpha)}e^{2\lambda\max_{1\leq i\leq n} a_i}\}.
\end{aligned}$$

Let $p>1$. It is well known that

$$uv=\inf_{b>0}\left\{\frac{1}{pb}u^p+\frac{1}{q}b^{q/p}v^q\right\}\text{ for} u>0, v>0 \text{ and } 1/p+1/q=1.$$

This yields the inequality

$$\lambda^{1+\alpha}ne^{2\lambda\max_{1\le i\le n}a_i}2^{\alpha+1}\widetilde{C}h_n^{-(1+\alpha)}\le\frac{1}{pb}\lambda^{p(1+\alpha)}n^p2^{p(\alpha+1)}\widetilde{C}^ph_n^{-p(1+\alpha)}+\frac{1}{q}b^{q/p}e^{2q\lambda\max_{1\le i\le n}a_i}. \quad (3.19)$$

We can thus conclude that for every $p>1$ and $0<\alpha\le 1$, there for all $\lambda>0$, such that

$$\begin{aligned}\mathbb{P}(|\widetilde{f}_n(x)-\mathbb{E}\widetilde{f}_n(x)|\ge A_1\zeta_n) &\le 2g(n)\exp\{-\lambda A_1n\zeta_n+\frac{1}{pb}\lambda^{p(1+\alpha)}n^p2^{p(1+\alpha)}\widetilde{C}^ph_n^{-p(1+\alpha)}\}.\\ &\times \exp\{\frac{1}{q}b^{q/p}e^{2q\lambda\max_{1\le i\le n}a_i}\}\\ &= 2g(n)\exp\{\frac{1}{q}b^{q/p}e^{2q\lambda\max_{1\le i\le n}a_i}\}\exp(\Psi(\lambda,n)).\end{aligned} \quad (3.20)$$

The equation $\frac{\partial\Psi(\lambda,n)}{\partial\lambda}=0$ has the unique solution

$$\lambda=\left(\frac{A_1\zeta_nbp}{2^{p(1+\alpha)}(1+\alpha)n^{p-1}\widetilde{C}^ph_n^{-p(1+\alpha)}}\right)^{\frac{1}{p(1+\alpha)-1}} \quad (3.21)$$

which minimizes $\Phi(\lambda,n)$. Then from (3.20),(3.21) and noting $2q\lambda\max_{1\le i\le n}a_i\le 1$, we obtain upper bound for the tail probability as

$$\begin{aligned}\mathbb{P}(|\sum_{j=1}^nY_i|\ge A_1n\zeta_n) &\le 2g(n)\exp\{\frac{1}{q}b^{q/p}e\}\exp\left\{-\left(\frac{A_1\zeta_nbp}{2^{p(1+\alpha)}(1+\alpha)n^{p-1}\widetilde{C}^ph_n^{-p(1+\alpha)}}\right)^{\frac{1}{p(1+\alpha)-1}}\right.\\ &\left. A_1n\zeta_n(1-\frac{1}{p(1+\alpha)})\right\}\\ &\le 2g(n)\exp\{\frac{1}{q}b^{q/p}e\}\exp\{-\widetilde{M}n^{\frac{p\alpha}{p(1+\alpha)-1}}\zeta_n^{\frac{p(1+\alpha)}{p(1+\alpha)-1}}h_n^{\frac{p(1+\alpha)}{p(1+\alpha)-1}}(1-\frac{1}{p(1+\alpha)})\}\\ &\le 2M_1\exp\{\frac{1}{q}b^{q/p}e\}n^\delta\exp\{-\widetilde{M}(1-\frac{1}{p(1+\alpha)})\log n\}\\ &= 2M_1\exp\{\frac{1}{q}b^{q/p}e\}n^{\delta-\widetilde{M}(1-\frac{1}{p(1+\alpha)})}.\end{aligned}$$

Taking $\widetilde{M}=(1-\frac{1}{p(1+\alpha)})A_1\frac{p(1+\alpha)}{p(1+\alpha)-1}\frac{bp}{\widetilde{C}^p2^{p(1+\alpha)}}$ large enough such that $\delta-\widetilde{M}(1-\frac{1}{p(1+\alpha)})<-1$, then we have that

$$\sum_{j=1}^n\mathbb{P}(|\widetilde{f}_n(x)-\mathbb{E}\widetilde{f}_n(x)|\ge A_1\zeta_n)<+\infty, \quad (3.22)$$

that is

$$|\widetilde{f}_n(x)-\mathbb{E}\widetilde{f}_n(x)|=O\left(\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}}n}{n^{\frac{\alpha}{(1+\alpha)}}h_n}\right), \text{ completely.} \quad (3.23)$$

By lemma(3.2.6) that

$$h_n^{-2}|\mathbb{E}\widetilde{f}_n(x)-f(x)|\le C<+\infty.$$

Which implies that

$$|\mathbb{E}\widetilde{f}_n(x)-f(x)|=O\left(\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}}n}{n^{\frac{\alpha}{(1+\alpha)}}h_n}\right). \quad (3.24)$$

Note that

$$\begin{aligned} |\widetilde{f}_n(x) - f(x)| &= |\widetilde{f}_n(x) - \mathbb{E}\widetilde{f}_n(x) + \mathbb{E}\widetilde{f}_n(x) - f(x)| \\ &\leq |\widetilde{f}_n(x) - \mathbb{E}\widetilde{f}_n(x)| + |\mathbb{E}\widetilde{f}_n(x) - f(x)|. \end{aligned}$$

Therefore, the desired result ((3.18)) follows immediately by (3.23)-(3.24). The proof is completed.

Theoreme 3.3.2 *Suppose that* $(H_1) - (H_2)$ *hold.* $\beta_n = \left(\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}}(ng(n))}{n^{\frac{\alpha}{(1+\alpha)}} h_n}\right) \to 0$. *Let* $\{X_n; n \geq 1\}$ *be a sequence of strictly stationary of WOD random variables. Suppose that the kernel* $K(.)$ *is a bounded monotone density function and the bandwidth satisfies that* $h_n = O(n^{-\frac{\alpha}{3(1+\alpha)}} \log^{\frac{p(1+\alpha)-1}{3p(1+\alpha)}} n)$ *for some* $p > 1$ *and* $0 < \alpha \leq 1$. *Then for any* $x \in \mathcal{C}^2(f)$.

$$|\widetilde{f}_n(x) - f(x)| = O\left(\beta_n\right), \text{ completely.} \tag{3.25}$$

Proof. In view of the proof of Theorem (3.3.1), we only need to show that

$$|\widetilde{f}_n(x) - \mathbb{E}\widetilde{f}_n(x)| = O\left(\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}}(ng(n))}{n^{\frac{\alpha}{(1+\alpha)}} h_n}\right), \text{ completely.} \tag{3.26}$$

and

$$|\mathbb{E}\widetilde{f}_n(x) - f(x)| = O\left(\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}}(ng(n))}{n^{\frac{\alpha}{(1+\alpha)}} h_n}\right). \tag{3.27}$$

Noting that $g(n) \geq 1$ then (3.27) follows from (3.24) immediately and thus we only need to prove (3.26). Similar to the proof of (3.23), set $\beta_n = \left(\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}}(ng(n))}{n^{\frac{\alpha}{(1+\alpha)}} h_n}\right)$. Applying Lemma(3.2.4) with $\lambda = \left(\frac{A_2 n \beta_n h_n^{p(1+\alpha)}}{\widetilde{C}_2 n^p p(1+\alpha)}\right)^{\frac{1}{p(1+\alpha)-1}}$ to obtain that for all n large enough,

$$\begin{aligned} \mathbb{P}(|\widetilde{f}_n(x) - \mathbb{E}\widetilde{f}_n(x)| \geq A_2\beta_n) &\leq \mathbb{P}(|\sum_{j=1}^{n} Y_i| \geq A_2 n\beta_n) \\ &\leq 2g(n)\exp\left\{-\lambda A_2 n\beta_n + \lambda^{1+\alpha}\sum_{i=1}^{n} e^{2\lambda a_i}\mathbb{E}|Y_i|^{1+\alpha}\right\} \\ &\leq 2g(n)\exp\left\{-\lambda A_2 n\beta_n + \lambda^{1+\alpha} e^{2\lambda \max_{1\leq i\leq n} a_i}\mathbb{E}|Y_i|^{1+\alpha}\right\} \\ &\leq 2g(n)\exp\{-\lambda A_2 n\beta_n + \lambda^{1+\alpha} n 2^{1+\alpha}\widetilde{C} h_n^{-(1+\alpha)} e^{2\lambda \max_{1\leq i\leq n} a_i}\}. \end{aligned}$$

Let $p > 1$. It is well known that

$$uv = \inf_{b>0}\left\{\frac{1}{pb}u^p + \frac{1}{q}b^{q/p}v^q\right\} \text{ for} u > 0, v > 0 \text{ and } 1/p + 1/q = 1.$$

This yields the inequality

$$\lambda^{1+\alpha} n e^{2\lambda \max_{1\leq i\leq n} a_i} 2^{\alpha+1}\widetilde{C} h_n^{-(1+\alpha)} \leq \frac{1}{pb}\lambda^{p(1+\alpha)} n^p 2^{p(\alpha+1)}\widetilde{C}^p h_n^{-p(1+\alpha)} + \frac{1}{q}b^{q/p} e^{2q\lambda \max_{1\leq i\leq n} a_i}. \tag{3.28}$$

We can thus conclude that for every $p > 1$ and $0 < \alpha \leq 1$, there for all $\lambda > 0$, such that

$$\begin{aligned}\mathbb{P}(|\widetilde{f}_n(x) - \mathbb{E}\widetilde{f}_n(x)| \geq A_2\beta_n) &\leq 2g(n)\exp\{-\lambda A_2 n\beta_n + \frac{1}{pb}\lambda^{p(1+\alpha)}n^p 2^{p(1+\alpha)}\widetilde{C}^p h_n^{-p(1+\alpha)}\}. \\ &\times \exp\{\frac{1}{q}b^{q/p}e^{2q\lambda\max_{1\leq i\leq n}a_i}\} \\ &= 2g(n)\exp\{\frac{1}{q}b^{q/p}e^{2q\lambda\max_{1\leq i\leq n}a_i}\}\exp(\widetilde{\Psi}(\lambda,n)).\end{aligned} \tag{3.29}$$

The equation $\frac{\partial\widetilde{\Psi}(\lambda,n)}{\partial\lambda} = 0$ has the unique solution

$$\lambda = \left(\frac{A_2\beta_n bp}{2^{p(1+\alpha)}(1+\alpha)n^{p-1}\widetilde{C}^p h_n^{-p(1+\alpha)}}\right)^{\frac{1}{p(1+\alpha)-1}} \tag{3.30}$$

which minimizes $\widetilde{\Psi}(\lambda,n)$. Then from (3.29),(3.30) and noting $2q\lambda\max_{1\leq i\leq n}a_i \leq 1$, we obtain upper bound for the tail probability as

$$\begin{aligned}\mathbb{P}(|\sum_{j=1}^{n}Y_i| \geq A_2 n\beta_n) &\leq 2g(n)\exp\{\frac{1}{q}b^{q/p}e\}\exp\left\{-\left(\frac{A_2\beta_n bp}{2^{p(1+\alpha)}(1+\alpha)n^{p-1}\widetilde{C}^p h_n^{-p(1+\alpha)}}\right)^{\frac{1}{p(1+\alpha)-1}}\right. \\ &\quad \left. A_2 n\beta_n(1-\frac{1}{p(1+\alpha)})\right\} \\ &\leq 2g(n)\exp\{\frac{1}{q}b^{q/p}e\}\exp\{-\widetilde{M_1}n^{\frac{p\alpha}{p(1+\alpha)-1}}\beta_n^{\frac{p(1+\alpha)}{p(1+\alpha)-1}}h_n^{\frac{p(1+\alpha)}{p(1+\alpha)-1}}(1-\frac{1}{p(1+\alpha)})\} \\ &\leq \frac{2}{n}(ng(n))\exp\{\frac{1}{q}b^{q/p}e\}\exp\{-\widetilde{M_1}(1-\frac{1}{p(1+\alpha)})\log(ng(n))\} \\ &= 2(ng(n))\exp\{\frac{1}{q}b^{q/p}e\}(ng(n))^{-\widetilde{M_1}(1-\frac{1}{p(1+\alpha)})} \\ &= 2\exp\{\frac{1}{q}b^{q/p}e\}(ng(n))^{1-\widetilde{M_1}(1-\frac{1}{p(1+\alpha)})}.\end{aligned}$$

Taking $\widetilde{M_1} = (1-\frac{1}{p(1+\alpha)})A_2^{\frac{p(1+\alpha)}{p(1+\alpha)-1}}\frac{bp}{\widetilde{C}^p 2^{p(1+\alpha)}}$ large enough such that $1-\widetilde{M}(1-\frac{1}{p(1+\alpha)}) < -1$, then we have that

$$\sum_{j=1}^{n}\mathbb{P}(|\widetilde{f}_n(x) - \mathbb{E}\widetilde{f}_n(x)| \geq A_2\beta_n) < +\infty, \tag{3.31}$$

which is equivalent to (3.26). The proof is completed.

Furthermore, by relaxing the restriction on the bandwidth h_n, we have the following result.

Theoreme 3.3.3 *Suppose that $(H_1)-(H_2)$ hold. Let $\{X_n; n \geq 1\}$ be a sequence of strictly stationary of WOD random variables. Suppose that the kernel $K(.)$ is a bounded monotone density function and* $\left(\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}}(ng(n))}{n^{\frac{\alpha}{(1+\alpha)}}h_n}\right) \to 0$. *Then for any $x \in \mathcal{C}^2(f)$,*

$$\widetilde{f}_n(x) - f(x) \to 0, \ \textit{completely}. \tag{3.32}$$

Proof. In view of the proof theorem3.3.2, by ((3.26)) and $\left(\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}}(ng(n))}{n^{\frac{\alpha}{(1+\alpha)}}h_n}\right) \to 0$ we have that

$$\widetilde{f}_n(x) - \mathbb{E}\widetilde{f}_n(x) \to 0, \text{completely}.$$

Therefore, we only need to show that

$$|\mathbb{E}\widetilde{f}_n(x) - f(x)| \to 0 \tag{3.33}$$

without the condition $h_n = O(n^{-\frac{\alpha}{3(1+\alpha)}} \log^{\frac{p(1+\alpha)-1}{3p(1+\alpha)}} n)$ in theorem(3.3.2).Actually, by lemma(3.2.4) we have that

$$\lim_{n\to+\infty} |\mathbb{E}\widetilde{f}_n(x) - f(x)| \le C \lim_{n\to+\infty} h_n^2 = 0.$$

Consequently,(3.33) is proved and thus the proof of the theorem is completed.
Since the proofs of Theorems (3.3.4)-(4.4.1) are similar, we only present the proof of Theorem (3.3.2) as follows.

As an application of the results above, we obtain the complete convergence and the complete convergence for the hazard function estimator $\widetilde{H}_n(x)$ as follows.

Theoreme 3.3.4 *Suppose that* $(H_1) - (H_2)$ *hold. Let* $\{X_n; n \ge 1\}$ *be a sequence of strictly stationary of WOD random variables with* $g(n) = O(n^\delta)$ *for some* $\delta \ge 0$*. Suppose that the kernel* $K(.)$ *is a bounded monotone density function and the bandwidth satisfies that* $h_n = O(n^{-\frac{\alpha}{3(1+\alpha)}} \log^{\frac{p(1+\alpha)-1}{3p(1+\alpha)}} n)$ *for some* $p > 1$ *and* $0 < \alpha \le 1$*. If there exists a point* x_0 *such that* $F(x_0) < 1$*, then for any* $x \in \mathcal{C}^2(f)$ *and* $x \le x_0$*,*

$$|\widetilde{H}_n(x) - H(x)| = O(n^{-\frac{\alpha}{(1+\alpha)}} h_n^{-1} \log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}} n), \text{ completely.} \tag{3.34}$$

Proof. The proof of this theorem is based on the following decomposition:

$$\begin{aligned} |\widetilde{H}_n(x) - H(x)| &\le \left| \frac{\widetilde{f}_n(x)}{1 - F_n(x)} - \frac{f(x)}{1 - F(x)} \right| \\ &\le \left| \frac{(1 - F_n(x))(\widetilde{f}_n(x) - f(x)) + (F_n(x) - F(x))f(x)}{(1 - F_n(x))(1 - F(x))} \right| \\ &\le \frac{1}{1 - F_n(x)} |\widetilde{f}_n(x) - f(x)| + \frac{f(x)}{(1 - F_n(x))(1 - F(x))} |F_n(x) - F(x)| \end{aligned} \tag{3.35}$$

From $0 \le F(x) \le F(x_0) < 1$ for all $x \le x_0$, $\sup_x f(x) \le M < +\infty$, applying theorem 3.3.1 and taking $\widetilde{\xi_n} = \frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}} n}{n^{\frac{\alpha}{(1+\alpha)}}}$ in lemma(3.2.7), we can see that

$$|\widetilde{f}_n(x) - f(x)| = O(n^{-\frac{\alpha}{(1+\alpha)}} h_n^{-1} \log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}} n), \text{ completely.} \tag{3.36}$$

and

$$\sup_{x \le x_0} |F_n(x) - F(x)| = O(n^{-\frac{\alpha}{(1+\alpha)}} \log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}} n), \text{ completely.} \tag{3.37}$$

On the other hand, from ((3.37)) one has that $x \le x_0$ and all n large enough,

$$1 - F_n(x) \ge 1 - F(x) \ge \frac{1 - F(x_0)}{2} > 0. \tag{3.38}$$

consequently, the desired result (3.34) follows from (3.35)-(3.38). The proof is completed. Since the proof of theorems (3.3.4)-(4.4.1) are similar, we only present the proof of theorem (3.3.4) as follows.

Theoreme 3.3.5 *Suppose that* $(H_1)-(H_2)$ *hold.* $\beta_n = \left(\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}}(ng(n))}{n^{\frac{\alpha}{(1+\alpha)}}h_n}\right) \to 0$. *Let* $\{X_n; n \geq 1\}$ *be a sequence of strictly stationary of WOD random variables. Suppose that the kernel* $K(.)$ *is a bounded monotone density function and the bandwidth satisfies that* $h_n = O(n^{-\frac{\alpha}{3(1+\alpha)}}\log^{\frac{p(1+\alpha)-1}{3p(1+\alpha)}} n)$ *for some* $p > 1$ *and* $0 < \alpha \leq 1$. *Then for any* $x \in \mathcal{C}^2(f)$ *and* $x \leq x_0$,

$$|\widetilde{H}_n(x) - H(x)| = O(\beta_n), \text{ completely.} \tag{3.39}$$

Theoreme 3.3.6 *Suppose that* $(H_1)-(H_2)$ *hold. Let* $\{X_n; n \geq 1\}$ *be a sequence of strictly stationary of WOD random variables. Suppose that the kernel* $K(.)$ *is a bounded monotone density function and* $\frac{\log^{\frac{p(1+\alpha)-1}{p(1+\alpha)}}(ng(n))}{n^{\frac{\alpha}{(1+\alpha)}}h_n} \to 0$. *If there exists a point* x_0 *such that* $F(x_0) < 1$, *then for any* $x \in \mathcal{C}^2(f)$ *and* $x \leq x_0$,

$$\widetilde{H}_n(x) - H(x) \to 0, \text{ completely.} \tag{3.40}$$

Conclusion.

Our work consists in establishing some new exponential inequalities for the distribution of sums of WOD random variables. Using these inequalities, we proved the complete convergence for kernel estimators of density and hazard functions, under some suitable conditions.

Bibliography

[1] Chen, X. and Liu, F. (2017). Strong laws of large numbers for negatively dependent random variables under sublinear expectations, Communications in Statistics - Theory and Methods 46 (24) :12387-12400.

[2] He, W., Cheng D. Y. and Wang, Y. B. (2013) Asymptotic lower bounds of precise large deviations with nonnegative and dependent random variables. Stat Probab Lett 83: 331-338.

[3] Hsu, P. L. and Robbins, H. (1947). Complete convergence and the law of large numbers. Proceedings of the National Academy of Sciences of the United States of America 33: 25-31.

[4] Li, Y. and Yang, S. (2005). Strong convergence rate of recursive probability density estimator for NA sequences, Chinese J. Engineering Mathematics, 22(4): 659-665.

[5] Parzen,E. (1962). On estimation of a probability density function and mode, Ann. Math. Statist. 33: 1065-1076.

[6] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of density function, Ann. Math. Statist. 27: 832-837.

[7] Wang, K. Y., Wang, Y. B. and Gao, Q. W. (2013). Uniform asymptotics for the finite-time ruin probability of a new dependent risk model with a constant interest rate. Methodology and Computing in Applied Probability 15: 109-124.

[8] Watson, G. S. and Leadbetter, M. R. (1964a). Hazard analysis, I. Biometrika 51: 175-184.

[9] Watson, G. S. and Leadbetter, M. R. (1964b). Hazard analysis, II.Sankhya A 26: 110-116.

Chapter 4

New exponential probability inequality and complete convergence for conditional LNQD random variables sequence, application to AR(1)model general

Summary

4.1 Introduction 63

4.2 Some lemmas 64

4.3 Main Results and Proofs 66

4.4 Applications to the results to AR(1) model 71

4.4.1 The AR(1) model 71

abstract: The exponential probability inequalities have been important tools in probability and statistics. In this paper, We prove a new tail probability inequality for the distributions of sums of conditionally linearly negative quadrant dependent ($\mathcal{F}$-LNQD, in short) random variables, and obtain a result dealing with conditionally complete convergence of first-order autoregressive processes with identically distributed ($\mathcal{F}$-LNQD) innovations.

4.1 Introduction

The exponential inequality plays an important role in various proofs of limit theorems. In particular, it provides a measure of the complete convergence for partial sums.
Firstly, we will recall the definitions of conditionally negative quadrant dependent, conditionally negatively associated, and conditionally linearly negative quadrant dependent sequence.
Let $(\omega, \mathcal{A}, \mathbb{P})$ be a probability space, and all random variables in this paper are defined on it unless otherwise mentioned. Let $\mathcal{F}$ be a sub-algebra of $\mathcal{A}$, two random variables ζ_1 and ζ_2 are said to be conditionally negative quadrant dependent given $\mathcal{F}$($\mathcal{F}$-NQD, in short) if, for all $\epsilon_1, \epsilon_2 \in \mathbb{R}$

$$\mathbb{P}^{\mathcal{F}}(\zeta_1 \leq \epsilon_1, \zeta_2 \leq \epsilon_2) \leq \mathbb{P}^{\mathcal{F}}(\zeta_1 \leq \epsilon_1)\mathbb{P}^{\mathcal{F}}(\zeta_2 \leq \epsilon_2). \tag{4.1}$$

One of the many possible multivariate generalizations of conditionally negative quadrant dependence is conditionally negatively association introduced by Yuan et al.[6] .
A finite collection of random variables $\zeta_1, \zeta_2, ..., \zeta_n$ is said to be conditionally negatively associated ($\mathcal{F}$-NA, in short) if for every pair of disjoint subsets A, B of $\{1, 2, ..., n\}$

$$Cov^{\mathcal{F}}(f(\zeta_i : i \in A), g(\zeta_j : j \in B)) \leq 0,$$

whenever f and g are coordinatewise nondecreasing such that this covariance exists. An infinite sequence $\{\zeta_n, n \geq 1\}$ is $\mathcal{F}$-NA if every finite subcollection is $\mathcal{F}$-NA.
We now propose another multivariate generalization of conditionally negative quadrant dependence called conditionally linearly negative quadrant dependence, which is weaker than $\mathcal{F}$-NA property.

Definition 4.1.1 *A finite sequence of random variables $\{\zeta_n, n \geq 1\}$ is said to be conditionally linearly negative quadrant dependent given ($\mathcal{F}$-LNQD, in short) if for any disjoint subsets $A, B\mathcal{Z}$ and positive $r_j^{'}s$,*

$$\sum_{k \in A} r_k \zeta_k \text{ and } \sum_{j \in B} r_j \zeta_j \text{ are } \mathcal{F} - NQD.$$

As mentioned earlier, it can be shown that the concepts of linearly negative quadrant dependent and conditional linearly negative quadrant dependent are not equivalent. See, for example, Yuan and Xie [7], where various of counterexamples are given.

A concrete example where conditional limit theorems are useful is the study of statistical inference for non-ergodic models as discussed in Bassawa and Prakasa Rao [1] and Basawa and Scott [2]. For instance, if one wants to estimate the mean off-spring θ for a Galton-Watson branching process, the asymptotic properties of the maximum likelihood estimator depend on the set of non-extinction.

As it was pointed out earlier, the conditional LNQD property does not imply the LNQD property and the opposite implication is also not true. Hence one does have to derive limit theorems under conditioning if there is a need for such results even through the results and proofs of such results

may be analogous to those under the non-conditioning setup. This one of the reasons for developing results for sequences of $\mathcal{F}$-LNQD random variables in this paper.
As mentioned earlier, large numbers of results for LNQD random variables have been achieved. However, nothing is variable for conditional LNQD random variables. Yuan and Wu [9] extended many results from negative association to asymptotically negative association, Yuan and Yang [4] extended many results from association to conditional association, Yuan et al [6] extended many results from negative association to conditional negative association, and these motivate our original interest in conditional LNQD.

On the other hand, the concept of complet convergence of a sequence of random variables was introduced by [4]. Note that complete convergence implies almost sure convergence in view of the Borel-Cantelli lemma. Now we extend this concept a conditionally converge completely given $\mathcal{F}$ to a constant a if $\sum_{i=1}^{\infty} \mathrm{P}(|X_i - a| > \varepsilon/\mathcal{F}) < \infty$ for every $\varepsilon > 0$, and we whrite $X_n \to a$ conditionally completely given $\mathcal{F}$.

The main purpose of this paper is to establish a new probability inequality and conditional complete convergence for the $\mathcal{F} - LNQD$ random variables and to extend and improve the results of Wang et al [5].
Throughout the paper, let $S_n = \sum_{i=1}^{n} X_i$ for a sequence $\{X_n, n \geq 1\}$ of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{A}$, $\{X_n, n \geq 1\}$ will be called $\mathcal{F}$-centered if $\mathbb{E}^{\mathcal{F}} X_n = 0$ for every $n \geq 1$. Denote $B_n = \sum_{i=1}^{n} \mathbb{E}^{\mathcal{F}}|X_i|^2$ for each $1 \leq i \leq n$.

4.2 Some lemmas

Lemma 4.2.1 *Let random variables X and Y be $\mathcal{F}$-NQD. Then*
(i) $\mathbb{E}^{\mathcal{F}}(XY) \leq \mathbb{E}^{\mathcal{F}}(X)\mathbb{E}^{\mathcal{F}}(Y)$;
(ii) $\mathbb{P}^{\mathcal{F}}(X > x, Y > y) \leq \mathbb{P}^{\mathcal{F}}(X > x)\mathbb{P}^{\mathcal{F}}(Y > y)$;
(iii) *If f and g are both nondecreasing (or both nonincreasing) functions, then $f(X)$ and $g(Y)$ are $\mathcal{F}$-NQD.*

Corollary 4.2.1 *Let $\{X_n, n \geq 1\}$ be a sequence of $\mathcal{F}$-LNQD random variables and $t > 0$, then for each $n \geq 1$,*

$$\mathbb{E}^{\mathcal{F}}\left[\sum_{i=1}^{n} \exp(tX_i)\right] \leq \prod_{i=1}^{n} \mathbb{E}^{\mathcal{F}}(\exp(tX_i)) \tag{4.2}$$

Proof. For $t > 0$, it is easy to see that tX_i and $t\sum_{j=i+1}^{n} X_j$ are $\mathcal{F}$-NQD by the definition of $\mathcal{F}$-LNQD, which implies that $\exp(tX_i)$ and $\exp(t\sum_{j=i+1}^{n} X_j)$ are also $\mathcal{F}$-NQD for $i = 1, 2, ..., n-1$ by Lemma

(4.2.1)(iii). It follows from Lemma (4.2.1)(i) and induction that

$$\begin{aligned}
\mathbb{E}^{\mathcal{F}}\left[\sum_{i=1}^{n}\exp(tX_i)\right] &= \mathbb{E}^{\mathcal{F}}\left[\exp(tX_1)\exp\big(\sum_{i=2}^{n}tX_i\big)\right] \\
&\leq \mathbb{E}^{\mathcal{F}}\left[\exp(tX_1)\right]\mathbb{E}^{\mathcal{F}}\left[\exp\big(\sum_{i=2}^{n}tX_i\big)\right] \\
&= \mathbb{E}^{\mathcal{F}}\left[\exp(tX_1)\right]\mathbb{E}^{\mathcal{F}}\left[\exp(tX_2)\exp\big(\sum_{i=3}^{n}tX_i\big)\right] \\
&\leq \mathbb{E}^{\mathcal{F}}\left[\exp(tX_1)\right]\mathbb{E}^{\mathcal{F}}\left[\exp(tX_2)\right]\mathbb{E}^{\mathcal{F}}\left[\exp\big(\sum_{i=3}^{n}tX_i\big)\right] \\
&\leq \prod_{i=1}^{n}\mathbb{E}^{\mathcal{F}}(\exp(tX_i)).
\end{aligned}$$

This completes the proof of the lemma.

Lemma 4.2.2 *[3] For any $x \in \mathbb{R}$, we have*

$$\exp(x) \leq 1 + x + \frac{|x|}{2}\ln(1+|x|)\exp(2|x|).$$

Lemma 4.2.3 *Let $\{X_n, n \geq 1\}$ be a sequence of $\mathcal{F}$-LNQD random variables with $\mathbb{E}^{\mathcal{F}}(X_n) = 0$ for each $n \geq 1$. If there exists a sequence of positiv e numbers $\{c_n, n \geq 1\}$ such that $|X_i| \leq c_i$ for each $i \geq 1$, then for any $t > 0$,*

$$\mathbb{E}^{\mathcal{F}}\exp\left\{t\sum_{i=1}^{n}X_i\right\} \leq \exp\left\{\frac{t^2}{2}\sum_{i=1}^{n}e^{2tc_i}\mathbb{E}^{\mathcal{F}}|X_i|^2\right\}. \tag{4.3}$$

Proof. By lemma (4.2.2), for all $x \in \mathbb{R}$, $\exp(x) \leq 1 + x + \frac{|x|}{2}\ln(1+|x|)\exp(2|x|)$. Thus, by $\mathbb{E}^{\mathcal{F}}(X_i) = 0$ and $|X_i| \leq c_i$ for each $i \geq 1$, we have

$$\begin{aligned}
\mathbb{E}^{\mathcal{F}}\exp(tX_i) &\leq \mathbb{E}^{\mathcal{F}}\left\{1 + tX_i + \frac{t}{2}|X_i|\ln(1+|tX_i|)\exp(2|tX_i|)\right\} \\
&= 1 + t\mathbb{E}^{\mathcal{F}}(X_i) + \frac{t}{2}\mathbb{E}^{\mathcal{F}}\left\{|X_i|\ln(1+|tX_i|)\exp(2|tX_i|)\right\} \\
&= 1 + \frac{t}{2}\mathbb{E}^{\mathcal{F}}\left\{|X_i|\ln(1+|tX_i|)\exp(2|tX_i|)\right\} \\
&\leq 1 + \frac{t}{2}\mathbb{E}^{\mathcal{F}}\left\{|X_i|\ln(1+|tX_i|)\exp(2tc_i)\right\} \\
&= 1 + \frac{t}{2}\exp(2tc_i)\mathbb{E}^{\mathcal{F}}\left\{t|X_i|^2\right\} \\
&= 1 + \frac{t^2}{2}\exp(2tc_i)\mathbb{E}^{\mathcal{F}}\left\{t|X_i|^2\right\} \\
&\leq \exp\left\{\frac{t^2}{2}\exp(2tc_i)\mathbb{E}^{\mathcal{F}}\left\{|X_i|^2\right\}\right\} \text{ (using } 1+y \leq \exp(y) \text{ for all } y \in \mathbb{R})
\end{aligned} \tag{4.4}$$

for any $t > 0$. By Lemma (4.2.1) and (4.4), we have can see that

$$\mathbb{E}^{\mathcal{F}}\exp\left\{t\sum_{i=1}^{n}X_i\right\} \leq \prod_{i=1}^{n}\mathbb{E}^{\mathcal{F}}\exp\{tX_i\} \tag{4.5}$$

$$\leq \exp\left\{\frac{t^2}{2}\sum_{i=1}^{n}e^{2tc_i}\mathbb{E}^{\mathcal{F}}|X_i|^2\right\}. \tag{4.6}$$

The lemma is thus proved.

Lemma 4.2.4 *Let* $\{X_n, n \geq 1\}$ *be a sequence of* $\mathcal{F}$*-LNQD random variables with* $\mathbb{E}^{\mathcal{F}}(X_n) = 0$ *for each* $n \geq 1$*. If there exists a sequence of positive numbers* $\{c_n, n \geq 1\}$ *such that* $|X_i| \leq c_i$ *for each* $i \geq 1$*, then for any* $t > 0$ *and* $\varepsilon > 0$

$$\mathbb{P}^{\mathcal{F}}(|\sum_{i=1}^{n} X_i| \geq \varepsilon) \leq \exp\left\{-t\varepsilon + \frac{t^2}{2}\sum_{i=1}^{n} e^{2tc_i}\mathbb{E}^{\mathcal{F}}|X_i|^2\right\}. \tag{4.7}$$

Proof. By Markov's inequality and lemma (4.2.3), we can see that

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}(\sum_{i=1}^{n} X_i \geq \varepsilon) &\leq \exp(-t\varepsilon)\mathbb{E}^{\mathcal{F}} \exp\left\{t\sum_{i=1}^{n} X_i\right\} \\ &\leq \exp(t\varepsilon)\prod_{i=1}^{n}\mathbb{E}^{\mathcal{F}} \exp\{tX_i\} \\ &\leq \exp\left\{-t\varepsilon + \frac{t^2}{2}\sum_{i=1}^{n} e^{2tc_i}\mathbb{E}^{\mathcal{F}}|X_i|^2\right\}. \end{aligned} \tag{4.8}$$

The desired result follos by remplacing X_i by $-X_i$ in (4.8). This completes the proof of the lemma.

4.3 Main Results and Proofs

Theoreme 4.3.1 *Let*$\{X_n, n \geq 1\}$ *be a sequence of* $\mathcal{F}$*-LNQD random variables with* $\mathbb{E}^{\mathcal{F}}(X_i) = 0$*. If there exists a positive numbers* c *such that* $|X_i| \leq c_i, i \geq 1$*, where* $B_n = \sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}}|X_i|^2$*, then for any* $\varepsilon > 0$ *and* $n \geq 1$*, then*

$$\mathbb{P}^{\mathcal{F}}(S_n/B_n \geq \varepsilon) \leq \exp\left\{\frac{1}{q}b^{q/p}e\right\}\exp\left\{-\left(\frac{\varepsilon 2^{p-1}bp}{B_n^{p-1}}\right)^{\frac{1}{2p-1}}\varepsilon B_n\left(1-\frac{1}{p-1}\right)\right\} \tag{4.9}$$

Proof.By Markov's inequality, we have that for any $t > 0$,

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}(S_n/B_n \geq \varepsilon) &= \mathbb{P}^{\mathcal{F}}(e^{tS_n} \geq e^{t\varepsilon B_n}), \\ &\leq e^{-t\varepsilon B_n}\mathbb{E}^{\mathcal{F}}\left(\prod_{i=1}^{n} e^{tX_i}\right), \\ &\leq \exp\left\{-t\varepsilon B_n + \frac{t^2}{2}e^{2t\max_{1\leq i\leq n} c_i}B_n\right\}. \end{aligned} \tag{4.10}$$

Let $p > 1$. It is well known that

$$uv = \inf_{b>0}\left\{\frac{1}{pb}u^p + \frac{1}{q}b^{q/p}v^q\right\} \text{ for} u > 0, v > 0 \text{ and } 1/p + 1/q = 1.$$

This yields the inequality

$$\frac{t^2}{2}e^{2t\max_{1\leq i\leq n} c_i}B_n \leq \frac{1}{pb}\frac{t^{2p}}{2^p}B_n^p + \frac{1}{q}b^{q/p}e^{2tq\max_{1\leq i\leq n} c_i}. \tag{4.11}$$

We can thus conclude that for every $p > 1$, there for all $t > 0$, such that

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}(S_n/B_n \geq \varepsilon) &\leq \exp\left\{-t\varepsilon B_n + \frac{1}{pb}\frac{t^{2p}}{2^p}B_n^p\right\}\\ &\times \exp\left\{\frac{1}{q}b^{q/p}e^{2tq\max_{1\leq i\leq n}c_i}\right\}\\ &= \exp\left\{\frac{1}{q}b^{q/p}e^{2tq\max_{1\leq i\leq n}c_i}\right\}\exp(\Phi(t,n)).\end{aligned} \tag{4.12}$$

The equation $\frac{\partial\Phi(t,n)}{\partial t} = 0$ has the unique solution

$$t = \left(\frac{\varepsilon 2^{p-1}bp}{B_n^{p-1}}\right)^{\frac{1}{2p-1}} \tag{4.13}$$

which minimizes $\Phi(t,n)$. Then from (4.12),(4.13) and taking $2tq\max_{1\leq i\leq n}c_i \leq 1$ we obtain (4.9).

Theoreme 4.3.2 *Let$\{X_n, n \geq 1\}$ be a sequence of $\mathcal{F}$-LNQD random variables with $\mathbb{E}^{\mathcal{F}}(X_i) = 0$. If there exists a positive numbers c such that $|X_i| \leq c_i, i \geq 1$, then for any $\varepsilon > 0$ and $n \geq 1$,*

$$\mathbb{P}^{\mathcal{F}}(|S_n| \geq \varepsilon) \leq 2\exp\left\{\frac{1}{q}b^{q/p}e\right\}\exp\left\{-\left(\frac{\varepsilon 2^{p-1}bp}{B_n^{p-1}}\right)^{\frac{1}{2p-1}}\varepsilon\left(1-\frac{1}{p-1}\right)\right\} \tag{4.14}$$

Proof. From conditions $\mathbb{E}^{\mathcal{F}}(X_i) = 0$ and $|X_i| < c_i$ for each $i \geq 1$. By Markov's inequality and Lemma (4.2.4), Corollary (4.2.1) with the fact that $1 + x \leq e^x$, then

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}(S_n \geq \varepsilon) &= e^{-t\varepsilon}\mathbb{E}^{\mathcal{F}}(e^{tS_n}),\\ &\leq e^{-t\varepsilon}\prod_{i=1}^{n}\exp\left(\frac{t^2}{2}e^{2tc_i}\mathbb{E}^{\mathcal{F}}|X_i|^2\right),\\ &\leq \exp\left\{-t\varepsilon + \frac{t^2}{2}e^{2t\max_{1\leq i\leq n}c_i}B_n\right\}.\end{aligned} \tag{4.15}$$

Let $p > 1$. It is well known that

$$uv = \inf_{b>0}\left\{\frac{1}{pb}u^p + \frac{1}{q}b^{q/p}v^q\right\} \text{ for} u > 0, v > 0 \text{ and } 1/p + 1/q = 1.$$

This yields the inequality

$$\frac{t^2}{2}e^{2t\max_{1\leq i\leq n}c_i}B_n \leq \frac{1}{pb}\frac{t^{2p}}{2^p}B_n^p + \frac{1}{q}b^{q/p}e^{2tq\max_{1\leq i\leq n}c_i}. \tag{4.16}$$

We can thus conclude that for every $p > 1$, there for all $t > 0$, such that

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}(|S_n| \geq \varepsilon) &\leq 2\exp\left\{-t\varepsilon + \frac{1}{pb}\frac{t^{2p}}{2^p}B_n^p\right\}\\ &\times \exp\left\{\frac{1}{q}b^{q/p}e^{2tq\max_{1\leq i\leq n}c_i}\right\}\\ &= 2\exp\left\{\frac{1}{q}b^{q/p}e^{2tq\max_{1\leq i\leq n}c_i}\right\}\exp(\Phi(t,n)).\end{aligned} \tag{4.17}$$

The equation $\frac{\partial\Phi(t,n)}{\partial t}=0$ has the unique solution

$$t=\left(\frac{\varepsilon 2^{p-1}bp}{B_n^p}\right)^{\frac{1}{2p-1}} \tag{4.18}$$

which minimizes $\Phi(t,n)$. Then from (4.17),(4.18)) and taking $2tq\max_{1\leq i\leq n}c_i\leq 1$ we obtain upper bound for the tail probability as

$$\mathbb{P}^{\mathcal{F}}(|S_n|\geq\varepsilon)\leq 2\exp\{\frac{1}{q}b^{q/p}e\}\exp\left\{-\left(\frac{\varepsilon 2^{p-1}bp}{B_n^p}\right)^{\frac{1}{2p-1}}\varepsilon\left(1-\frac{1}{p-1}\right)\right\} \tag{4.19}$$

Since $\{-X_n,n\geq 1\}$ is also a sequense of $\mathcal{F}$-LNQD random variables it follows from (4.19) that

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}(S_n\leq-\varepsilon)=\mathbb{P}^{\mathcal{F}}(-S_n\geq\varepsilon) &\leq \exp\left\{\frac{1}{q}b^{q/p}e\right\}\\ &\times \exp\left\{-\left(\frac{\varepsilon 2^{p-1}bp}{B_n^p}\right)^{\frac{1}{2p-1}}\varepsilon\left(1-\frac{1}{p-1}\right)\right\}\end{aligned} \tag{4.20}$$

From (4.19) and (4.20) we obtain

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}(|S_n|\geq\varepsilon)=\mathbb{P}^{\mathcal{F}}(S_n\geq-\varepsilon)+\mathbb{P}^{\mathcal{F}}(S_n\leq\varepsilon) &\leq 2\exp\left\{\frac{1}{q}b^{q/p}e\right\}\\ &\times \exp\left\{-\left(\frac{\varepsilon 2^{p-1}bp}{B_n^p}\right)^{\frac{1}{2p-1}}\varepsilon\left(1-\frac{1}{p-1}\right)\right\}\end{aligned} \tag{4.21}$$

Theoreme 4.3.3 *Let$\{X_n,n\geq 1\}$ be a sequence of $\mathcal{F}$-LNQD random variables with mean zero and finite variances. If there exists a positive numbers c such that $|X_i|\leq c_i, i\geq 1$, where $c_n,n\geq 1$ is a sequence of positive numbers. Then for any $\varepsilon>0$ and $n\geq 1$,*

$$\mathbb{P}^{\mathcal{F}}(|S_n-\mathbb{E}^{\mathcal{F}}S_n|\geq\varepsilon)\leq 2\exp\left\{\frac{1}{q}b^{q/p}e\right\}\exp\left\{-\left(\frac{\varepsilon 2^{p-1}bp}{B_n^{p-1}}\right)^{\frac{1}{2p-1}}\varepsilon B_n\left(1-\frac{1}{p-1}\right)\right\} \tag{4.22}$$

Proof.By Markov's inequality and Lemma (4.2.3), we have that for any $t>0$,

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}(S_n-\mathbb{E}^{\mathcal{F}}S_n\geq\varepsilon) &\leq e^{-t\varepsilon}\mathbb{E}^{\mathcal{F}}[\exp(t\sum_{i=1}^{n}(X_i-\mathbb{E}^{\mathcal{F}}X_i))],\\ &\leq e^{-t\varepsilon}\mathbb{E}^{\mathcal{F}}\prod_{i=1}^{n}\left[e^{t(X_i-\mathbb{E}^{\mathcal{F}}X_i)}\right],\\ &\leq \exp\left\{-t\varepsilon+\frac{t^2}{2}e^{2t\max_{1\leq i\leq n}c_i}B_n\right\}.\end{aligned} \tag{4.23}$$

Let $p>1$. It is well known that

$$uv=\inf_{b>0}\left\{\frac{1}{pb}u^p+\frac{1}{q}b^{q/p}v^q\right\}\text{ for}u>0,v>0\text{ and }1/p+1/q=1.$$

This yields the inequality

$$\frac{t^2}{2} e^{2t\max_{1\le i\le n} c_i} B_n \le \frac{1}{pb}\frac{t^{2p}}{2^p} B_n^p + \frac{1}{q} b^{q/p} e^{2tq\max_{1\le i\le n} c_i}. \tag{4.24}$$

We can thus conclude that for every $p>1$, there for all $t>0$, such that

$$\begin{aligned} \mathbb{P}^{\mathcal{F}}(|S_n - \mathbb{E}^{\mathcal{F}} S_n| \ge \varepsilon) &\le 2\exp\left\{-t\varepsilon + \frac{1}{pb}\frac{t^{2p}}{2^p} B_n^p\right\} \\ &\times \exp\left\{\frac{1}{q} b^{q/p} e^{2tq\max_{1\le i\le n} c_i}\right\} \\ &= 2\exp\left\{\frac{1}{q} b^{q/p} e^{2tq\max_{1\le i\le n} c_i}\right\} \exp(\Phi(t,n)). \end{aligned} \tag{4.25}$$

The equation $\frac{\partial \Phi(t,n)}{\partial t} = 0$ has the unique solution

By take $t = \left(\frac{\varepsilon 2^{p-1} bp}{B_n^p}\right)^{\frac{1}{2p-1}}$. Hence it follows from (4.23) that

$$\mathbb{P}^{\mathcal{F}}(S_n - \mathbb{E}^{\mathcal{F}} S_n \ge \varepsilon) \le \exp\left\{\frac{1}{q} b^{q/p} e\right\} \exp\left\{-\left(\frac{\epsilon 2^{p-1} bp}{B_n^{p-1}}\right)^{\frac{1}{2p-1}} \varepsilon B_n \left(1 - \frac{1}{p-1}\right)\right\} \tag{4.26}$$

Let $-S_n = T_n = \sum_{i=1}^{n}(-X_n)$. Since $\{-X_n, n\ge 1\}$ is also a sequence of $\mathcal{F}$-LNQD random variables we also have

$$\begin{aligned} \mathbb{P}^{\mathcal{F}}(S_n - \mathbb{E}^{\mathcal{F}} S_n \le -\varepsilon) = \mathbb{P}^{\mathcal{F}}(T_n - \mathbb{E}^{\mathcal{F}} T_n \ge \varepsilon) &\le \exp\left\{\frac{1}{q} b^{q/p} e\right\} \\ &\times \exp\left\{-\left(\frac{\varepsilon 2^{p-1} bp}{B_n^{p-1}}\right)^{\frac{1}{2p-1}} \varepsilon B_n \left(1 - \frac{1}{p-1}\right)\right\} \end{aligned} \tag{4.27}$$

by Combining (4.26) and(4.27) we get (4.22)

Corollary 4.3.1 *Let$\{X_n, n\ge 1\}$ be a sequence of $\mathcal{F}$-LNQD random variables. Assume that there exists a positive integer n_0 such that $|X_i| \le c_n$, for each $1\le i\le n, n\ge n_0$, where $\{c_n, n\ge 1\}$ is a sequence of positive numbers. Then for any $\varepsilon > 0$*

$$\mathbb{P}^{\mathcal{F}}(|S_n - \mathbb{E}^{\mathcal{F}} S_n| \ge n\varepsilon) \le 2\exp\left\{\frac{1}{q} b^{q/p} e\right\} \exp\left\{-\left(\frac{n\varepsilon 2^{p-1} bp}{B_n^p}\right)^{\frac{1}{2p-1}} n\varepsilon \left(1 - \frac{1}{p-1}\right)\right\} \tag{4.28}$$

Theoreme 4.3.4 *Let$\{X_n, n\ge 1\}$ be a sequence of $\mathcal{F}$-LNQD random variables with $\mathbb{E}^{\mathcal{F}}(X_i) = 0$. If there exists a positive numbers c such that $|X_i| \le c_i, i\ge 1$. Then for any $r>0$*

$$n^{-r} S_n \to 0 \quad completely, \quad n\to\infty. \tag{4.29}$$

Proof. Let $B = \sum_{n=1}^{\infty} \mathbb{E}^{\mathcal{F}}(X_n)^2 \le \infty$. For any $\varepsilon > 0$, it follows from Theoreme 4.3.2 we have

$$\begin{aligned}
\sum_{n=1}^{\infty}\mathbb{P}^{\mathcal{F}}(|S_n|\geq n^r\varepsilon) &\leq 2\sum_{n=1}^{\infty}\exp\left\{\frac{1}{q}b^{q/p}e\right\}\exp\left\{-\left(\frac{n^r\varepsilon 2^{p-1}bp}{B_n^p}\right)^{\frac{1}{2p-1}}\varepsilon n^r\left(1-\frac{1}{p-1}\right)\right\}\\
&\leq 2\sum_{n=1}^{\infty}\exp\left\{\frac{1}{q}b^{q/p}e\right\}\exp\left\{-\left(\frac{\varepsilon 2^{p-1}bp}{B_n^p}\right)^{\frac{1}{2p-1}}\varepsilon\left(1-\frac{1}{p-1}\right)\right\}^{n^{\frac{2rp}{2p-1}}}\\
&\leq 2\exp\left\{\frac{1}{q}b^{q/p}e\right\}\sum_{n=1}^{\infty}[\exp(-c)]^{n^{\frac{2rp}{2p-1}}}.
\end{aligned} \tag{4.30}$$

where C is positive number not depending on n. (by the inequality $e^{-y}\leq(\frac{a}{ey})^a$), choosing $a=\frac{2p-1}{rp}$, since $a>0,x>0$. Then the right-hand side of (4.30) become

$$\begin{aligned}
\sum_{n=1}^{\infty}\mathbb{P}^{\mathcal{F}}(|S_n|\geq n^r\varepsilon) &\leq 2\exp\{\frac{1}{q}b^{q/p}e\}\sum_{n=1}^{\infty}\left(\frac{a}{ec}\right)^a\left(\frac{1}{n}\right)^{\left(\frac{2rp}{2p-1}\right)^a}\\
&\leq 2\exp\left\{\frac{1}{q}b^{q/p}e\right\}\frac{a^a}{(ec)^a}\sum_{n=1}^{\infty}\frac{1}{n^{\frac{2rpa}{2p-1}}}\\
&\leq 2\exp\left\{\frac{1}{q}b^{q/p}e\right\}\frac{a^a}{(ec)^a}\sum_{n=1}^{\infty}\frac{1}{n^2},\\
&= 2\exp\left\{\frac{1}{q}b^{q/p}e\right\}\frac{a^a}{(ec)^a}\frac{\pi^2}{6},\\
&< \infty
\end{aligned} \tag{4.31}$$

Theoreme 4.3.5 *Let$\{X_n,n\geq 1\}$ be a sequence of $\mathcal{F}$-LNQD random variables. Assume that there exists a positive integer n_0 such that $|X_i|\leq c_n$, for each $1\leq i\leq n, n\geq n_0$, where $\{c_n,n\geq 1\}$ is a sequence of positive numbers.*

$$\sum_{n=1}^{\infty}\mathbb{P}^{\mathcal{F}}\left(\frac{1}{n}|S_n-\mathbb{E}^{\mathcal{F}}S_n|\geq\varepsilon_n\right)<\infty. \tag{4.32}$$

Theoreme 4.3.6 *Let$\{X_n,n\geq 1\}$ be a sequence of $\mathcal{F}$-LNQD random variables with $\mathbb{E}^{\mathcal{F}}(X_i)=0$. If there exists a positive numbers c such that $|X_i|\leq c_i, i\geq 1$. Then for any $r>0$*

$$n^{-r}(S_n-\mathbb{E}^{\mathcal{F}}S_n)\to 0\quad completely, n\to\infty. \tag{4.33}$$

Proof. For any $\varepsilon>0$, it follows from Corollary (4.3.1) that

$$\begin{aligned}
\sum_{n=1}^{\infty}\mathbb{P}^{\mathcal{F}}(|S_n-\mathbb{E}^{\mathcal{F}}S_n|\geq n^r\varepsilon) &\leq 2\sum_{n=1}^{\infty}\exp\left\{\frac{1}{q}b^{q/p}e\right\}\exp\left\{-\left(\frac{n^r\varepsilon 2^{p-1}bp}{B_n^p}\right)^{\frac{1}{2p-1}}\varepsilon n^r\left(1-\frac{1}{p-1}\right)\right\}\\
&\leq 2\sum_{n=1}^{\infty}\left[\exp\left\{\frac{1}{q}b^{q/p}e\right\}\right]\\
&\times\left[\exp\left\{-\left(\frac{\varepsilon 2^{p-1}bp}{B_n^p}\right)^{\frac{1}{2p-1}}\varepsilon\left(1-\frac{1}{p-1}\right)\right\}\right]^{n^{\frac{2rp}{2p-1}}}
\end{aligned} \tag{4.34}$$

After this result we get (4.33).

4.4 Applications to the results to AR(1) model

The basic object of this section is applying the results to first-order autoregressive processes(AR(1)).

4.4.1 The AR(1) model

We consider an autoregressive time series of first order AR(1) defined by

$$X_{n+1} = \theta X_n + \zeta_{n+1},\ n = 1, 2, ..., \tag{4.35}$$

where $\{\zeta_n, n \geq 0\}$ is a sequence of identically distributed $\mathcal{F}$-LNQD random variables with $\zeta_0 = X_0 = 0$, $0 < \mathbb{E}^{\mathcal{F}}\zeta_k^4 < \infty$, $k = 1, 2, ...$ and where θ is a parameter with $|\theta| < 1$. Here, we can rewrite X_{n+1} in (4.35) as follows:

$$X_{n+1} = \theta^{n+1}X_0 + \theta^n\zeta_1 + \theta^{n-1}\zeta_2 + ... + \zeta_{n+1}. \tag{4.36}$$

The coefficient θ is fitted least squares, giving the estimator

$$\widehat{\theta}_n = \frac{\sum_{j=1}^{n} X_j X_{j-1}}{\sum_{j=1}^{n} X_{j-1}^2} \tag{4.37}$$

It immediately follows from (4.35) and (4.37) that

$$\widehat{\theta}_n - \theta = \frac{\sum_{j=1}^{n} \zeta_j X_{j-1}}{\sum_{j=1}^{n} X_{j-1}^2} \tag{4.38}$$

Theoreme 4.4.1 *Let the conditions of Theorem 4.3.3 be satisfied then for any* $\frac{(\mathbb{E}^{\mathcal{F}}\zeta_1^2)^{1/2}}{\rho^2} < \xi$ *positive, we have*

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}(\sqrt{n}|\widehat{\theta}_n - \theta| > \rho) \quad &\leq \quad 2\exp\left\{-\left(\frac{(\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}\zeta_1^2)n2^{p-1}bp}{B_n^{p-1}}\right)^{\frac{1}{2p-1}} (\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}\zeta_1^2)nB_n\left(1 - \frac{1}{p-1}\right)\right\} \\ &\times \quad \exp\left\{\frac{1}{q}b^{q/p}e\right\} + \exp\left\{-\frac{1}{2}n\frac{(K_1 - n\xi^2)^2}{K_2}\right\}\end{aligned} \tag{4.39}$$

where $K_1 = \mathbb{E}^{\mathcal{F}}(X_i^2) < \infty, K_2 = \mathbb{E}^{\mathcal{F}}(X_i^4) < \infty.$

Proof. Firstly, we notice that :

$$\widehat{\theta}_n - \theta = \frac{\sum_{j=1}^{n} \zeta_j X_{j-1}}{\sum_{j=1}^{n} X_{j-1}^2}.$$

It follows that

$$\mathbb{P}^{\mathcal{F}}(\sqrt{n}|\widehat{\theta}_n - \theta| > \rho) = \mathbb{P}^{\mathcal{F}}(\left|\frac{1/\sqrt{n}\sum_{j=1}^{n}\zeta_j X_{j-1}}{1/n\sum_{j=1}^{n}X_{j-1}^2}\right| > \rho)$$

By virtue of the probability properties and Hölder's inequality, we have for any ξ positive

$$\begin{aligned}
\mathbb{P}^{\mathcal{F}}(\sqrt{n}|\widehat{\theta}_n - \theta| > \rho) &\leq \mathbb{P}^{\mathcal{F}}\left(1/n\sum_{j=1}^{n}\zeta_j^2 \geq \rho^2\xi^2\right) + \mathbb{P}^{\mathcal{F}}\left(1/n^2\sum_{j=1}^{n}X_{j-1}^2 \leq \rho^2\right) \\
&= \mathbb{P}^{\mathcal{F}}\left(\sum_{j=1}^{n}\zeta_j^2 \geq (\rho^2\xi^2)n\right) + \mathbb{P}^{\mathcal{F}}\left(\sum_{j=1}^{n}X_{j-1}^2 \leq n^2\xi^2\right) \\
&= I_{1n} + I_{2n}.
\end{aligned}$$

Next we estimate I_{1n} and I_{2n}.

$$\begin{aligned}
I_{1n} &= \mathbb{P}^{\mathcal{F}}\left(\sum_{j=1}^{n}\zeta_j^2 \geq (\rho^2\xi^2)n\right) \\
&= \mathbb{P}^{\mathcal{F}}\left(\sum_{j=1}^{n}(\zeta_j^2 - \mathbb{E}^{\mathcal{F}}\zeta_j^2 + \mathbb{E}^{\mathcal{F}}\zeta_j^2) \geq (\rho^2\xi^2)n\right) \\
&= \mathbb{P}^{\mathcal{F}}\left(\sum_{j=1}^{n}(\zeta_j^2 - \mathbb{E}^{\mathcal{F}}\zeta_j^2) \geq (\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}\zeta_1^2)n\right) \\
&\leq \mathbb{P}^{\mathcal{F}}\left(\left|\sum_{j=1}^{n}(\zeta_j^2 - \mathbb{E}^{\mathcal{F}}\zeta_j^2)\right| \geq (\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}\zeta_1^2)n\right)
\end{aligned} \tag{4.40}$$

By using the Theorem 4.3.3 the right hand side of (4.40) become

$$\begin{aligned}
I_{1n} &= \mathbb{P}^{\mathcal{F}}\left(\sum_{j=1}^{n}\zeta_j^2 \geq (\rho^2\xi^2)n\right) \\
&\leq 2\exp\left\{-\left(\frac{(\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}\zeta_1^2)n2^{p-1}bp}{B_n^{p-1}}\right)^{\frac{1}{2p-1}}(\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}\zeta_1^2)nB_n\left(1 - \frac{1}{p-1}\right)\right\} \\
&\times \exp\left\{\frac{1}{q}b^{q/p}e\right\}
\end{aligned} \tag{4.41}$$

We will bound now, the second probability of the right-hand side of the expression I_{2n}.According to the Markov's inequality, it follows for any t positive

$$\begin{aligned}
I_{2n} &= \mathbb{P}^{\mathcal{F}}\left(\frac{1}{n^2}\sum_{i=1}^{n} X_{i-1}^2 \le \xi^2\right) \\
&= \mathbb{P}^{\mathcal{F}}\left(n^2\xi^2 - \sum_{i=1}^{n} X_{i-1}^2 \ge 0\right) \\
&= \mathbb{E}^{\mathcal{F}}\left(\mathbb{I}_{\{n\epsilon^2 - \sum_{i=1}^{n} X_{i-1}^2 \ge 0\}}\right) \\
&\le \mathbb{E}^{\mathcal{F}}\left(\exp t\left(n^2\xi^2 - \sum_{i=1}^{n} X_{i-1}^2\right)\right) \qquad (t>0) \\
&\le e^{tn^2\xi^2}\mathbb{E}^{\mathcal{F}}\left(\exp -t\sum_{i=1}^{n} X_{i-1}^2\right) \\
&\le e^{tn^2\xi^2}\prod_{i=1}^{n}\mathbb{E}^{\mathcal{F}}\left(\exp -tX_{i-1}^2\right).
\end{aligned}$$

Since

$$I_{2n} \le e^{tn^2\xi^2}\Pi_{i=1}^{n}\mathbb{E}^{\mathcal{F}}\left(\exp -tX_{i-1}^2\right).$$

we first claim that for $x \ge 0$

$$e^{-x} \le 1 - x + \frac{1}{2}x^2. \tag{4.42}$$

To see this let $\psi(x) = e^{-x}$ and $\phi(x) = 1 - x + \frac{1}{2}x^2$, ($\psi'(x) = -e^{-x}$) and recall that for every x

$$e^{x} \ge 1 + x \quad \forall x, \tag{4.43}$$

so that $\psi'(x) = -e^{-x} \le -1 + x = \phi'(x)$. Since $\psi(0) = 1 = \phi(0)$ this implies $\psi(x) \le \phi(x)$ for aall $x \ge 0$ and (4.42) is claimed.
From (4.42) and (4.43) it follows that for $t > 0$

$$\begin{aligned}
e^{tn\epsilon^2}\Pi_{i=1}^{n}\mathbb{E}^{\mathcal{F}}\left(\exp(-tX_{i-1}^2)\right) &\le e^{tn^2\xi^2}\left(1 - tK_1 + \frac{t^2}{2}K_2\right)^n \\
&\le e^{tn^2\xi^2}\left(\exp\left(-tK_1 + \frac{t^2}{2}K_2\right)\right)^n \\
&\le e^{tn^2\xi^2}\exp\left(-ntK_1 + \frac{t^2}{2}nK_2\right)
\end{aligned}$$

where $K_1 = \mathbb{E}^{\mathcal{F}}(X_i^2) < \infty, K_2 = \mathbb{E}^{\mathcal{F}}(X_i^4) < \infty$.
Hence

$$I_{2n} = \mathbb{P}^{\mathcal{F}}\left(\sum_{i=1}^{n} X_{i-1}^2 \le n^2\xi^2\right) \le \exp\left[t\left(n^2\xi^2 - nK_1\right) + \frac{nt^2K_2}{2}\right]. \tag{4.44}$$

With $h(t) = n^2\xi^2 - nK_1 + \frac{nt^2K_2}{2}$ and $t > 0$, the equation $h^{'}(t) = 0$ has the unique solution $t = \frac{K_1 - n\xi^2}{K_2}$ which minimize $h(t)$. Hence

$$\mathbb{P}^{\mathcal{F}}\left(\sum_{i=1}^{n} X_{i-1}^2 \le n^2\xi^2\right) \le \exp\left\{-\frac{1}{2}n\frac{(K_1 - n\xi^2)^2}{K_2}\right\} \tag{4.45}$$

Then for every $\rho > 0$, $K_1 < \infty$, $_2 < \infty$, and by the assumption

$$\begin{aligned}
\mathbb{P}^{\mathcal{F}}(\sqrt{n}|\widehat{\theta}_n - \theta| > \rho) \quad &\le \quad 2\exp\left\{-\left(\frac{(\rho^2\xi - \mathbb{E}^{\mathcal{F}}\zeta_1^2)n2^{p-1}bp}{B_n^{p-1}}\right)^{\frac{1}{2p-1}}(\rho^2\xi - \mathbb{E}^{\mathcal{F}}\zeta_1^2)nB_n\left(1 - \frac{1}{p-1}\right)\right\} \\
&\times \quad \exp\left\{\frac{1}{q}b^{q/p}e\right\} + \exp\left\{-\frac{1}{2}n\frac{(K_1 - n\xi^2)^2}{K_2}\right\}.
\end{aligned} \tag{4.46}$$

These complete the proof.

Corollary 4.4.1 *The sequence $(\widehat{\theta}_n)_{n\in\mathbb{N}}$ is completely converges to the parameter θ of autoregressive process AR(1) model. Then we have*

$$\sum_{n=1}^{+\infty} \mathbb{P}^{\mathcal{F}}(\sqrt{n}|\widehat{\theta}_n - \theta| > \rho) < +\infty. \tag{4.47}$$

Proof. By using Theorem 4.3.4 and $\mathbb{E}^{\mathcal{F}}(X_i^2) < \infty, \mathbb{E}^{\mathcal{F}}(X_i^4) < \infty$ we get the result of (4.47) immediately.

Bibliography

[1] Basawa, I.V, Paraska Rao, B.L.S (1980). Statistical Inference for Stochastic Processes.Academic press, London

[2] Basawa, I.V, Scott, D (1983). Asymptotic Optimal inference for non-ergodic models, Lecture Note in Statistics, Vol. 17. Springer, New York.

[3] Cheng, Hu. (2016). A strong law of large numbers for sub-linear expectation under a general moment condition. Statistics and Probability Letters. 119: 248-258 .

[4] Hsu, P.L., and Robbins, H. (1947). Complete convergence and the law of large numbers. Proceedings of the National Academy of Sciences, USA 33: 25-31.

[5] Wang, X.J., Hu, S.H., Yang, W.Z., Li, X.Q. (2010). Exponential inequalities and complete convergence for a LNQD sequence. J. Korean Statist. Soc. 39: 555-564.

[6] Yuan, D-M, An, J, Wu, X-S (2010). Conditional limit theorems for conditionally negatively associated random variables. Monatshefte Math. 161: 449-473.

[7] Yuan, D-M, Xie, Y (2012). Conditional limit theorems for conditionally linearly negative quadrant dependent random variables. Monatshefte Math. 166: 281-299.

[8] Yuan, D-M, Xie, Y (2009). Rosenthal type inequalities for asymptotically almost negatively associated random variables and applications, Sci. China Set. A 52: 1887-1904.

[9] Yuan, D.M., Wu, X.S. (2010). Limiting behavior of the maximum of the partial sum for asymptotically negatively associated random variables under residual Cesàro alpha-integrability assumption. J. Stat. Plan.Infer. 140: 2395-2402

[10] Yuan, D.M., Yang, Y.K. (2011). Conditional versions of limit theorems for conditionally associated random variables. J. Math. Anal. Appl. 376: 282-293

Chapter 5

Exponential inequalities and conditional complete convergence for conditional extended acceptable random variables and application to AR(1) model

Summary

5.1 Introduction **77**

5.2 Some lemmas **79**

5.3 Conditionally complete convergence for conditional extended acceptable **81**

5.4 Applications to the results to AR(1) model **82**

5.4.1 The AR(1) model 82

abstract: In this paper, we present a new exponential inequality and conditional complete convergence for conditional extended acceptable random variables and obtain a result dealing with conditional complete convergence of first-order autoregressive processes AR(1).

5.1 Introduction

Chow and Teicher [2], Majerek et al.[6], Roussas [9] and Prakasa Rao [7] studied the concept of conditionally independent random variables as well as the concept of conditional association and provided several results. They include conditional versions of generalized Borel Cantelli lemma, generalized Kolmogorov's inequality, generalized Hájek Rényi inequalities and further related results.
Prakasa Rao [7] provides counterexamples where independent random variables lose their independence under conditioning and dependent random variables become independent under conditioning.
Conditional association is defined in analogy to (unconditional) association. All random variables are defined on the probability space $(\omega, \mathcal{A}, \mathbb{P})$. Following Prakasa Rao [7] for simplicity we will use the notation $\mathbb{E}^{\mathcal{F}}(X)$ to denote $\mathbb{E}[X|\mathcal{F}]$ where $\mathcal{F}$ is sub-σ-algebra of $\mathcal{A}$. In addition, $Cov^{\mathcal{F}}(X,Y)$ denotes the conditional covariance of X and Y given $\mathcal{F}$, i.e.,

$$Cov^{\mathcal{F}}(X,Y) = \mathbb{E}^{\mathcal{F}}(XY) - \mathbb{E}^{\mathcal{F}}(X)\mathbb{E}^{\mathcal{F}}(Y).$$

Definition 5.1.1 *A finite collection of random variables $X_1, \ldots, X_n$ is said to be $\mathcal{F}$-associated if*

$$Cov^{\mathcal{F}}(f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)) \geq 0,$$

for any real-valued componentwise nondecreasing functions f, g on $\mathbb{R}^n$ such that the covariance is defined. An infinite collection is $\mathcal{F}$- associated if every finite subcollection is $\mathcal{F}$-associated. Roussas[9] introduced the concept of conditional negative association as follows.

Definition 5.1.2 *A finite collection of random variables $X_1, \ldots, X_n$ is said to be conditionally negatively associated given $\mathcal{F}$ ($\mathcal{F} - NA$) if*

$$Cov^{\mathcal{F}}(f(X_i : i \in A), g(X_j : j \in B)) \leq 0 \quad a.s.,$$

for any disjoint subsets A and B of $\{1, 2, \ldots, n\}$ and for any real-valued compenentwise nondecreasing functions f, g on $\mathbb{R}^{|A|}$ and $\mathbb{R}^{|B|}$ respectively where $|A| = card(A)$ provided that the covariance is defined. An infinite collection is conditionally negatively associated given $\mathcal{F}$ if every finite subcollection is $\mathcal{F} - NA$

Yuang et al [5] provide examples where negative association does not imply conditional negative association and vise versa.
Further, the concept of ENOD random variables was proposed by Liu [5] by extending the NOD. In this note, we combine the concept of conditioning on a σ-algebra with the concept of ENOD and define conditionally extended negative orthant dependent random variables as follows;

Definition 5.1.3 *A sequence of random variables $\{X_i, i \geq 1\}$ is said to be conditional extended negatively orthant dependent($\mathcal{F}$-ENOD) if there exists a constant $M > 0$ such that both*

$$\mathbb{P}^{\mathcal{F}}(X_1 \leq x_1, \ldots, X_n \leq x_n) \leq M \prod_{i=1}^{n} \mathbb{P}^{\mathcal{F}}(X_i \leq x_i) \tag{5.1}$$

and

$$\mathbb{P}^{\mathcal{F}}(X_1 > x_1, \ldots, X_n > x_n) > M \prod_{i=1}^{n} \mathbb{P}^{\mathcal{F}}(X_i > x_i) \tag{5.2}$$

hold for each $n = 1, 2, \ldots$ and all $x_1, \ldots, x_n$. Recall that the sequence $\{X_i, i \geq 1\}$ is said to be conditional negatively orthant dependent($\mathcal{F}$-NOD) if both 5.1 and 5.2 hold when $M = 1$; it is called conditional positively orthant dependent($\mathcal{F}$-POD) if 5.1 and 5.2 hold both in the reverse direction when $M = 1$.

In this paper, we define the concept of conditional acceptability by combining the concept of conditioning on a σ-algebra and the concept of acceptability see [4]. We therefore give the following definition.

Definition 5.1.4 *A finite family of random variables $X_1, X_2 \ldots, X_n$ is said to be $\mathcal{F}$-acceptable for $\delta > 0$ if $\mathbb{E}^{\mathcal{F}}(\exp(\delta|X_i|)) < \infty$ for all i and such that for any real $\lambda > 0$ such that $|\lambda| \leq \delta$,*

$$\mathbb{E}^{\mathcal{F}}\left(\exp\left(\lambda \sum_{i=1}^{n} X_i\right)\right) \leq \prod_{i=1}^{n} \mathbb{E}^{\mathcal{F}}(\exp(\lambda X_i)) a.s.$$

A sequence of random variables $\{X_n, n \in \mathbb{N}\}$ is $\mathcal{F}$-acceptable for $\delta > 0$ if every finit subfamily is $\mathcal{F}$-acceptable for δ.

Remark 5.1.1 *It can be easily verified that if random variables $X_1, \ldots, X_n$ are $\mathcal{F}$-acceptable, then the random variables $X_1 - \mathbb{E}^{\mathcal{F}}(X_1), X_2 - \mathbb{E}^{\mathcal{F}}(X_2), \ldots, X_n - \mathbb{E}^{\mathcal{F}}(X_n)$ are also $\mathcal{F}$-acceptable, and $-X_1, -X_2, \ldots, -X_n$ are also $\mathcal{F}$-acceptable.*

So in this work, we defined an conditional extended acceptability from the definitions of conditional acceptability and conditional extended orthant dependence as follows.

Definition 5.1.5 *A finite sequence $\{X_i, 1 \leq i \leq n\}$ of random variables is said to be conditional extended acceptable if there exists a constant $M > 0$ such that for any real λ*

$$\mathbb{E}^{\mathcal{F}}\left(\exp\left(\lambda \sum_{i=1}^{n} X_i\right)\right) \leq M \prod_{i=1}^{n} \mathbb{E}^{\mathcal{F}}(\exp(\lambda X_i)). \tag{5.3}$$

An infinite sequence $\{X_n, n \geq 1\}$ of random variables is conditional extended acceptable if every finite subcollection is conditional extended acceptable.

A sequence $\{X_i, i \geq 1\}$ of random variables is obviously acceptable if 5.3 holds when $M = 1$ and hence an acceptable sequence must be an conditinal extended acceptable sequence. In addition, 5.1 and 5.2 obviously satisfy 5.3. Therefore, the $\mathcal{F}$-ENOD random variables are conditional extended acceptable random variables.

It is know that exponential inequalities played an important role in obtaining asymptotic results for sums of independent random variables. Classical exponential inequalities were obtained, for example, by Bernstein, Hoeffding, Kolmogrov, Fuk, and Nagaev (see the monograph of Petrov [8]). The main goal of our paper is to establish a new probability inequality for conditional extended acceptable random variables. This paper is organized as follows. In section 2, we provide the establish the exponential in- equalities for sum of conditional extended acceptable random variables and in section 3, we obtain a result dealing with the conditional complete convergence for these random variables by using the exponential inequality. Finally, we obtain a result dealing with conditional complete convergence of first-order autoregressive processes AR(1).

5.2 Some lemmas

Lemma 5.2.1 *Let* $\{X_n, n \geq 1\}$ *be a sequence of* $\mathcal{F}$*-extended acceptable random variables and* $\lambda > 0$*, then for each* $n \geq 1$*,*

$$\mathbb{E}^{\mathcal{F}}\left[\sum_{i=1}^{n} \exp(\lambda X_i)\right] \leq M \prod_{i=1}^{n} \mathbb{E}^{\mathcal{F}}(\exp(\lambda X_i)) \tag{5.4}$$

Proof. For $\lambda > 0$, it is easy to see that λX_i and $\lambda \sum_{j=i+1}^{n} X_j$ are $\mathcal{F}$-extended acceptable by the definition, which implies that $\exp(\lambda X_i)$ and $\exp(\lambda \sum_{j=i+1}^{n} X_j)$ are also $\mathcal{F}$-extended acceptable for $i = 1, 2, ..., n-1$. Thus, by induction we have

$$\begin{aligned}
\mathbb{E}^{\mathcal{F}}\left[\sum_{i=1}^{n} \exp(\lambda X_i)\right] &= \mathbb{E}^{\mathcal{F}}\left[\exp(\lambda X_1)\exp(\sum_{i=2}^{n} \lambda X_i)\right] \\
&\leq M_1 \mathbb{E}^{\mathcal{F}}\left[\exp(\lambda X_1)\right] \mathbb{E}^{\mathcal{F}}\left[\exp(\sum_{i=2}^{n} \lambda X_i)\right] \\
&= M_1 \mathbb{E}^{\mathcal{F}}\left[\exp(\lambda X_1)\right] \mathbb{E}^{\mathcal{F}}\left[\exp(\lambda X_2)\exp(\sum_{i=3}^{n} \lambda X_i)\right] \\
&\leq M_1 M_2 \mathbb{E}^{\mathcal{F}}\left[\exp(\lambda X_1)\right] \mathbb{E}^{\mathcal{F}}\left[\exp(\lambda X_2)\right] \mathbb{E}^{\mathcal{F}}\left[\exp(\sum_{i=3}^{n} \lambda X_i)\right] \\
&\leq \prod_{i=1}^{n-1} M_i \prod_{i=1}^{n} \mathbb{E}^{\mathcal{F}}(\exp(\lambda X_i)).
\end{aligned}$$

where $M = \prod_{i=1}^{n-1} M_i$.

Lemma 5.2.2 *[1] For any* $x \in \mathbb{R}$*, we have*

$$\exp(x) \leq 1 + x + |x|^{1+\alpha}\exp(2|x|),\ 0 < \alpha \leq 1.$$

Lemma 5.2.3 *Let* X *be a random variable with* $\mathbb{E}^{\mathcal{F}} e^{\delta|X|} < \infty$ *for some* $\delta > 0$*. Then for any* $0 < \lambda \leq \delta/4$*,*

$$\mathbb{E}^{\mathcal{F}} e^{\lambda(X - \mathbb{E}^{\mathcal{F}} X)} \leq \exp(K\lambda^{1+\alpha}). \tag{5.5}$$

where K is defined as $K = 2^{(2\alpha+1)/2}(\mathbb{E}^{\mathcal{F}}|X|^{2(1+\alpha)})^{1/2}\mathbb{E}^{\mathcal{F}} e^{\delta|X|}$

Proof. From the inequality $\exp(x) \leq 1 + x + |x|^{1+\alpha}\exp(2|x|),\ 0 < \alpha \leq 1$ we have by the Hölder inequality, the $c_r$inequality , and the Jensen inequality that for any $0 < \lambda \leq \delta/4$

$$\begin{aligned}
\mathbb{E}^{\mathcal{F}} e^{\lambda(X-\mathbb{E}^{\mathcal{F}}X)} &\leq \mathbb{E}^{\mathcal{F}}\{1+\lambda(X-\mathbb{E}^{\mathcal{F}}X)+(|\lambda(X-\mathbb{E}^{\mathcal{F}}X)|)^{1+\alpha}e^{(2\lambda|X-\mathbb{E}^{\mathcal{F}}X|)}\} \\
&\leq 1+\lambda\mathbb{E}^{\mathcal{F}}(X-\mathbb{E}^{\mathcal{F}}X)+\lambda^{1+\alpha}\mathbb{E}^{\mathcal{F}}\left\{|X-\mathbb{E}^{\mathcal{F}}X|^{1+\alpha}e^{(2\lambda|X-\mathbb{E}^{\mathcal{F}}X|)}\right\} \\
&\leq 1+\lambda^{1+\alpha}\mathbb{E}^{\mathcal{F}}\left\{|X-\mathbb{E}^{\mathcal{F}}X|^{1+\alpha}e^{(2\lambda|X-\mathbb{E}^{\mathcal{F}}X|)}\right\} \\
&\leq 1+\lambda^{1+\alpha}(\mathbb{E}^{\mathcal{F}}|X-\mathbb{E}^{\mathcal{F}}X|^{2(1+\alpha)})^{1/2}(\mathbb{E}^{\mathcal{F}}e^{4\lambda|X-\mathbb{E}^{\mathcal{F}}X|})^{1/2} \\
&\leq 1+\lambda^{1+\alpha}\left(2^{2\alpha+1}\mathbb{E}^{\mathcal{F}}|X|^{2(1+\alpha)}+\left|\mathbb{E}^{\mathcal{F}}X\right|^{2(1+\alpha)}\right)^{1/2}\left(\mathbb{E}^{\mathcal{F}}e^{4\lambda|X|}\mathbb{E}^{\mathcal{F}}e^{4\lambda|X|}\right)^{1/2} \\
&\leq 1+2^{\frac{2\alpha+1}{2}}\lambda^{1+\alpha}\left(\mathbb{E}^{\mathcal{F}}|X|^{2(1+\alpha)}\right)^{1/2}\left(\mathbb{E}^{\mathcal{F}}e^{4\lambda|X|}\mathbb{E}^{\mathcal{F}}e^{4\lambda|X|}\right)^{1/2} \\
&\leq 1+2^{\frac{2\alpha+1}{2}}\lambda^{1+\alpha}\left(\mathbb{E}^{\mathcal{F}}|X|^{2(1+\alpha)}\right)^{1/2}\mathbb{E}^{\mathcal{F}}e^{4\lambda|X|} \\
&\leq 1+2^{\frac{2\alpha+1}{2}}\lambda^{1+\alpha}\left(\mathbb{E}^{\mathcal{F}}|X|^{2(1+\alpha)}\right)^{1/2}\mathbb{E}^{\mathcal{F}}e^{\delta|X|} \\
&= 1+\lambda^{1+\alpha}K \\
&\leq \exp(K\lambda^{1+\alpha})
\end{aligned}$$

Since $1+x\leq e^x$ for all $x\in\mathbb{R}$. Here $K=2^{(2\alpha+1)/2}(\mathbb{E}^{\mathcal{F}}|X|^{2(1+\alpha)})^{\frac{1}{2}}\mathbb{E}^{\mathcal{F}}e^{\delta|X|}$
Hence the result is proved.

Theoreme 5.2.1 *Let $\{X_n, n\geq 1\}$ be a sequence of identically distributed conditional extended acceptable random variables with $\mathbb{E}^{\mathcal{F}}e^{\delta|X_1|}<\infty$ for some $\delta>0$. Then There exists a constant $M>0$ such that for any $0<\epsilon\leq K(1+\alpha)(\delta/2)^{\alpha}$,*

$$\mathbb{P}^{\mathcal{F}}\left(\left|(S_n-\mathbb{E}^{\mathcal{F}}S_n)\right|>n\epsilon\right)\leq 2M\exp\left\{-\left(\frac{\epsilon}{K(1+\alpha)}\right)^{1/\alpha}n\epsilon\left(1-\frac{1}{(1+\alpha)}\right)\right\} \tag{5.6}$$

where $S_n=X_1+\cdots+X_n$.

Proof. Suppose that $0<\epsilon\leq K(1+\alpha)(\delta/2)^{\alpha}$ then by Markov's inequality, the definition of conditional extended acceptable random variables and Lemma (5.2.3) that for any $0<\lambda\leq\delta/4$,

$$\begin{aligned}
\mathbb{P}^{\mathcal{F}}\left(\sum_{i=1}^{n}(X_i-\mathbb{E}^{\mathcal{F}}X_i)>n\epsilon\right) &= \mathbb{P}^{\mathcal{F}}\left(\exp\left(\lambda\sum_{i=1}^{n}(X_i-\mathbb{E}^{\mathcal{F}}X_i)\right)>\exp(\lambda n\epsilon)\right) \\
&\leq \exp(\lambda n\epsilon)\mathbb{E}^{\mathcal{F}}\exp\left(\lambda\sum_{i=1}^{n}(X_i-\mathbb{E}^{\mathcal{F}}X_i)\right) \\
&\leq M\exp(\lambda n\epsilon)\prod_{i=1}^{n}\mathbb{E}^{\mathcal{F}}\exp(\lambda(X_i-\mathbb{E}^{\mathcal{F}}X_i)) \\
&\leq M\exp(\lambda n\epsilon)\prod_{i=1}^{n}\exp(K\lambda^{(1+\alpha)}) \\
&= M\exp(-\lambda n\epsilon+Kn\lambda^{(1+\alpha)})
\end{aligned}$$

Optimizing the exponent in the term of this upper bound, we find $\lambda=\left(\frac{\epsilon}{(1+\alpha)K}\right)^{1/\alpha}$. Note that $\frac{\epsilon}{(1+\alpha)k}\leq(\delta/2)^{\alpha}$, by condition $0<\epsilon\leq K(1+\alpha)(\delta/2)^{\alpha}$. Thus, we get that

$$\mathbb{P}^{\mathcal{F}}\left(\sum_{i=1}^{n}(X_i-\mathbb{E}^{\mathcal{F}}X_i)>n\epsilon\right)\leq M\exp\left\{-\left(\frac{\epsilon}{K(1+\alpha)}\right)^{1/\alpha}n\epsilon\left(1-\frac{1}{(1+\alpha)}\right)\right\} \tag{5.7}$$

Since $\{-X_n, n \geq 1\}$ are also conditional extended acceptable random variables, we can remplace X_i by $-X_i$ in the above statement. That is,

$$\mathbb{P}^{\mathcal{F}}\left(-\sum_{i=1}^{n}(X_i - \mathbb{E}^{\mathcal{F}}X_i) > n\epsilon\right) \leq M\exp\left\{-\left(\frac{\epsilon}{K(1+\alpha)}\right)^{1/\alpha} n\epsilon\left(1-\frac{1}{(1+\alpha)}\right)\right\} \tag{5.8}$$

It follows from (5.7) and (5.8) that

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}\left(\left|\sum_{i=1}^{n}(X_i - \mathbb{E}^{\mathcal{F}}X_i)\right| > n\epsilon\right) &\leq \mathbb{P}^{\mathcal{F}}\left(\sum_{i=1}^{n}(X_i - \mathbb{E}^{\mathcal{F}}X_i) > n\epsilon\right) + \mathbb{P}^{\mathcal{F}}\left(-\sum_{i=1}^{n}(X_i - \mathbb{E}^{\mathcal{F}}X_i) > n\epsilon\right) \\ &\leq 2M\exp\left\{-\left(\frac{\epsilon}{K(1+\alpha)}\right)^{1/\alpha} n\epsilon\left(1-\frac{1}{(1+\alpha)}\right)\right\}\end{aligned}$$

Theoreme 5.2.2 *Let $\{X_n, n \geq 1\}$ be a sequence of identically distributed conditional extended acceptable random variables with $\mathbb{E}^{\mathcal{F}}e^{\delta|X_1|} < \infty$ for some $\delta > 0$. Set $\epsilon_n = \left(\frac{\log n}{n}\right)^{\frac{\alpha}{1+\alpha}}$ and $K = 2^{(2\alpha+1)/2}(\mathbb{E}^{\mathcal{F}}|X|^{2(1+\alpha)})^{\frac{1}{2}}\mathbb{E}^{\mathcal{F}}e^{\delta|X|}$. Then there exists a constant $M > 0$ such that*

$$\mathbb{P}^{\mathcal{F}}\left(\left|\sum_{i=1}^{n}(S_n - \mathbb{E}^{\mathcal{F}}S_n)\right| > n\epsilon_n\right) = 2M\sum_{i=1}^{n}\exp\left\{-\widetilde{M}\log n\left(1-\frac{1}{1+\alpha}\right)\right\} \tag{5.9}$$

Proof. Let $\epsilon_n = \left(\frac{\log n}{n}\right)^{\frac{\alpha}{1+\alpha}}$ and $K = 2^{(2\alpha+1)/2}(\mathbb{E}^{\mathcal{F}}|X|^{2(1+\alpha)})^{\frac{1}{2}}\mathbb{E}^{\mathcal{F}}e^{\delta|X|}$. Then $\epsilon_n/(K(1+\alpha)(\delta/2)^{\alpha}) \leq 1$ for all large n. By choosing $\widetilde{M} = \left\{\left(\frac{1}{K(1+\alpha)}\right)^{1/\alpha}\left(1-\frac{1}{1+\alpha}\right)\right\}$.
Hence, the result follows directly from Theorem (5.2.1).

5.3 Conditionally complete convergence for conditional extended acceptable

A sequence of random variables $\{X_n, n \geq 1\}$ is said to convergence completely to a constant ϑ if for all $\varepsilon > 0$,

$$\sum_{i=1}^{\infty}\mathbb{P}(|X_n - \vartheta| \geq \varepsilon) < \infty.$$

This concept was defined by Hsu and Robbins [3] . Note that complete convergence implies almost sure convergence in view of the Borel-Cantelli lemma. Now we extend this concept (also see Definition 3 in Yaun et al.[10])

Definition 5.3.1 *A sequence of random variables $\{X_n, n \geq 1\}$ is said to conditionally convergence completely given $\mathcal{F}$ to a constant a if $\sum_{n=1}^{\infty}\mathbb{P}^{\mathcal{F}}(|X_n - \vartheta| \geq \varepsilon) < \infty$ for every $\varepsilon > 0$, and we write $X_n \to \vartheta$ conditionally completely given $\mathcal{F}$.*

Theoreme 5.3.1 *Let $\{X_n, n \geq 1\}$ be a sequence of identically distributed and conditional extended acceptable random variables with $\mathbb{E}^{\mathcal{F}}(X_1) = 0$ and $\mathbb{E}^{\mathcal{F}}e^{\delta|X_1|} < \infty$ for some $\delta > 0$. Then $n^{-1}(S_n - \mathbb{E}^{\mathcal{F}}S_n) \to 0$ completely asn $\to \infty$.*

Proof By using Theorem (5.2.1), we can be obtained the result of Theorem 5.3.1 and the proof is omitted.

Theoreme 5.3.2 *Let $\{X_n, n \geq 1\}$ be a sequence of conditional extended acceptable random variables with $\mathbb{E}^{\mathcal{F}} e^{\delta|X_1|} < \infty$ for some $\delta > 0$ and $|X_i| \leq C < \infty$ for each $i \geq 1$, where C is a positive constant. Then, for any $s > 0$*

$$n^{-s}(S_n - \mathbb{E}^{\mathcal{F}} S_n) \to 0 \quad \textit{completelyas} \quad n \to \infty \tag{5.10}$$

Proof. It follows from ((5.6)) that for any $\epsilon > 0$ and some constant $M > 0$ we obtain

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}^{\mathcal{F}}(|S_n - \mathbb{E}^{\mathcal{F}} S_n| \geq n^s \epsilon) &\leq 2M \sum_{n=1}^{\infty} \left[\exp\left\{ -\left(\frac{\epsilon}{K(1+\alpha)} \right)^{1/\alpha} n^s \epsilon \left(1 - \frac{1}{(1+\alpha)} \right) \right\} \right] \\ &\leq 2M \sum_{n=1}^{\infty} \left[\exp\left\{ -\left(\frac{\epsilon}{K(1+\alpha)} \right)^{1/\alpha} \epsilon \left(1 - \frac{1}{(1+\alpha)} \right) \right\} \right]^{n^s} \\ &\leq 2 \sum_{n=1}^{\infty} [\exp(-m)]^{n^s} < \infty\end{aligned}$$

which yields (5.10), where m is a positive number not depending on n.

5.4 Applications to the results to AR(1) model

The basic object of this section is applying the results to first-order autoregressive processes(AR(1)).

5.4.1 The AR(1) model

We consider an autoregressive time series of first order AR(1) defined by

$$X_i = \theta X_{i-1} + \xi_i, \quad i = 1, 2, \cdots, \tag{5.11}$$

where $\{\xi_i, i \geq 0\}$ is a sequence of identically distributed $\mathcal{F}$-extended acceptable random variables with $\xi_0 = X_0 = 0$, $0 < \mathbb{E}^{\mathcal{F}} \xi_k^4 < \infty$, $k = 1, 2, ...$ and where θ is a parameter with $|\theta| < 1$. Hence (5.11) as follows:

$$\xi_i = \sum_{\infty}^{j=0} \theta^j \xi_{i-j}. \tag{5.12}$$

The coefficient θ is fitted least squares, giving the estimator

$$\widehat{\theta}_n = \frac{\sum_{j=1}^{n} X_j X_{j-1}}{\sum_{j=1}^{n} X_{j-1}^2} \tag{5.13}$$

from (5.11)and (5.13) we obtain that

$$\widehat{\theta}_n - \theta = \frac{\sum_{j=1}^{n} X_j X_{j-1}}{\sum_{j=1}^{n} X_{j-1}^2} \tag{5.14}$$

Theoreme 5.4.1 *Let the conditions of Theorem (5.2.1) be satisfied then for any* $\frac{(\mathbb{E}^{\mathcal{F}}X_1^2)^{1/2}}{\rho^2} < \xi$ *positive, we have*

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}(\sqrt{n}|\widehat{\theta}_n - \theta| > \rho) &\leq 2M\exp\left\{-\left(\frac{(\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}X_1)}{K(1+\alpha)}\right)^{1/\alpha} n(\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}X_1^2)\left(1 - \frac{1}{1+\alpha}\right)\right\} \\ &+ \exp\left\{-\left(\frac{-n\epsilon^2 + \mathbb{E}^{\mathcal{F}}(X_{j-1}^2)}{K'(1+\alpha)}\right)^{1/\alpha} n^2\epsilon^2\mathbb{E}^{\mathcal{F}}(X_{j-1}^2)\left(1 - \frac{1}{1+\alpha}\right)\right\}\end{aligned}$$

Proof. Firstly, we notice that :

$$\widehat{\theta}_n - \theta = \frac{\sum_{j=1}^{n} X_j X_{j-1}}{\sum_{j=1}^{n} X_{j-1}^2}.$$

It follows that

$$\mathbb{P}^{\mathcal{F}}(\sqrt{n}|\widehat{\theta}_n - \theta| > \rho) = \mathbb{P}^{\mathcal{F}}\left(\left|\frac{1/\sqrt{n}\sum_{j=1}^{n} X_j X_{j-1}}{1/n\sum_{j=1}^{n} X_{j-1}^2}\right| > \rho\right)$$

By virtue of the probability properties and Hölder's inequality, we have for any ξ positive

$$\begin{aligned}\mathbb{P}^{\mathcal{F}}(\sqrt{n}|\widehat{\theta}_n - \theta| > \rho) &\leq \mathbb{P}^{\mathcal{F}}\left(1/n\sum_{j=1}^{n} X_j \geq \rho^2\xi^2\right) + \mathbb{P}^{\mathcal{F}}\left(1/n^2\sum_{j=1}^{n} X_{j-1}^2 \leq \rho^2\right) \\ &= \mathbb{P}^{\mathcal{F}}\left(\sum_{j=1}^{n} X_j \geq (\rho^2\xi^2)n\right) + \mathbb{P}^{\mathcal{F}}\left(\sum_{j=1}^{n} X_{j-1}^2 \leq n^2\xi^2\right) \\ &= I_{1n} + I_{2n}.\end{aligned}$$

Next we estimate I_{1n} and I_{2n}.

$$\begin{aligned}I_{1n} &= \mathbb{P}^{\mathcal{F}}\left(\sum_{j=1}^{n} X_j \geq (\rho^2\xi^2)n\right) \\ &= \mathbb{P}^{\mathcal{F}}\left(\sum_{j=1}^{n}(X_j - \mathbb{E}^{\mathcal{F}}X_j + \mathbb{E}^{\mathcal{F}}X_j) \geq (\rho^2\xi^2)n\right) \\ &= \mathbb{P}^{\mathcal{F}}\left(\sum_{j=1}^{n}(X_j - \mathbb{E}^{\mathcal{F}}X_j) \geq (\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}X_1)n\right) \\ &\leq \mathbb{P}^{\mathcal{F}}\left(\left|\sum_{j=1}^{n}(X_j - \mathbb{E}^{\mathcal{F}}X_j)\right| \geq (\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}X_1)n\right)\end{aligned} \tag{5.15}$$

By using the Theorem (5.2.1) the right hand side of (5.15) become

$$I_{n1} = \mathbb{P}^{\mathcal{F}}\left(\sum_{i=1}^{n} X_j \geq (\rho^2\xi^2)n\right) \leq M\exp\left\{-\left(\frac{(\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}X_1)}{K(1+\alpha)}\right)^{1/\alpha} n(\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}X_1)\left(1 - \frac{1}{1+\alpha}\right)\right\}$$

We will bound now, the second probability of the right-hand side of the expression I_{2n}.According to the Markov's inequality, it follows for any λ positive

$$\begin{aligned} I_{2n} &= \mathbb{P}^{\mathcal{F}}\left(\frac{1}{n^2}\sum_{i=1}^{n} X_{i-1}^2 \leq \xi^2\right) \\ &= \mathbb{P}^{\mathcal{F}}\left(n^2\xi^2 - \sum_{i=1}^{n} X_{i-1}^2 \geq 0\right) \\ &= \mathbb{E}^{\mathcal{F}}\left(\mathbb{I}_{\{n\epsilon^2 - \sum_{i=1}^{n} X_{i-1}^2 \geq 0\}}\right) \\ &\leq \mathbb{E}^{\mathcal{F}}\left(\exp\lambda\left(n^2\xi^2 - \sum_{i=1}^{n} X_{i-1}^2\right)\right) \qquad (\lambda > 0) \\ &\leq e^{\lambda n^2\xi^2}\mathbb{E}^{\mathcal{F}}\left(e^{(-\lambda\sum_{i=1}^{n} X_{i-1}^2)}\right) \\ &\leq e^{\lambda n^2\xi^2}\mathbb{E}^{\mathcal{F}}\left(\prod_{i=1}^{n} e^{-\lambda X_{i-1}^2}\right). \end{aligned}$$

By using Lemma (5.2.1) and (5.2.3) the right hand side of the expression I_{2n} become

$$\begin{aligned} I_{2n} &= Me^{tn^2\epsilon^2}\prod_{i=1}^{n}\mathbb{E}^{\mathcal{F}}\left(e^{-\lambda X_{i-1}^2}\right) \\ &\leq Me^{\lambda n^2\epsilon^2}\prod_{i=1}^{n}\mathbb{E}^{\mathcal{F}}\left(1 - \lambda X_{j-1}^2 + (|\lambda X_{j-1}^2|)^{1+\alpha}\exp(2|\lambda X_{j-1}^2|)\right) \\ &\leq Me^{\lambda n^2\epsilon^2}\left(1 - \lambda X_{j-1}^2 + (|\lambda X_{j-1}^2|)^{1+\alpha}\exp(2|\lambda X_{j-1}^2|)\right)^n \\ &\leq Me^{\lambda n^2\epsilon^2}\left(1 - \lambda\mathbb{E}^{\mathcal{F}}(X_{j-1}^2) + \lambda^{1+\alpha}\mathbb{E}^{\mathcal{F}}\left((|X_{j-1}^2|)^{1+\alpha}\right)\exp(2\lambda\mathbb{E}^{\mathcal{F}}(X_{j-1}^2))\right)^n \\ &\leq Me^{\lambda n^2\epsilon^2}\left(1 - \lambda\mathbb{E}^{\mathcal{F}}(X_{j-1}^2) + \lambda^{1+\alpha}\left(\mathbb{E}^{\mathcal{F}}(|X_{j-1}^2|)^{2(1+\alpha)}\right)^{1/2}\left(\mathbb{E}^{\mathcal{F}}e^{4\lambda(|X_{j-1}^2|)}\right)^{1/2}\right)^n \\ &\leq Me^{\lambda n^2\epsilon^2}\left(1 - \lambda\mathbb{E}^{\mathcal{F}}(X_{j-1}^2) + \lambda^{1+\alpha}\left(\mathbb{E}^{\mathcal{F}}(|X_{j-1}^2|)^{2(1+\alpha)}\right)^{1/2}\left(\mathbb{E}^{\mathcal{F}}e^{\delta(|X_{j-1}^2|)}\right)\right)^n \\ &\leq M\exp\left(\lambda n^2\epsilon^2 - n\left\{\lambda\mathbb{E}^{\mathcal{F}}(X_{j-1}^2) + \lambda^{1+\alpha}\left(\mathbb{E}^{\mathcal{F}}(|X_{j-1}^2|)^{2(1+\alpha)}\right)^{1/2}\left(\mathbb{E}^{\mathcal{F}}e^{\delta(|X_{j-1}^2|)}\right)\right\}\right) \\ &\leq M\exp\left\{-\left(\frac{-n\epsilon^2 + \mathbb{E}^{\mathcal{F}}(|X_{j-1}^2|)}{K'(1+\alpha)}\right)^{1/\alpha} n^2\epsilon^2\mathbb{E}^{\mathcal{F}}(X_{j-1}^2)\left(1 - \frac{1}{1+\alpha}\right)\right\} \end{aligned}$$

By taking $\lambda = \left(\frac{-n\epsilon^2 + \mathbb{E}^{\mathcal{F}}(|X_{j-1}^2|)}{K'(1+\alpha)}\right)^{1/\alpha}$, K' *is defined as* $k' = n(1+\alpha)\mathbb{E}^{\mathcal{F}}(X_{j-1}^2)^{1+\alpha}\mathbb{E}^{\mathcal{F}}e^{\delta X_{j-1}^2}$. Then for any $\rho > 0$

$$\begin{aligned} \mathbb{P}^{\mathcal{F}}(\sqrt{n}|\widehat{\theta}_n - \theta| > \rho) \quad \leq \quad & M\exp\left\{-\left(\frac{n(\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}X_1)}{K(1+\alpha)}\right)^{1/\alpha}(\rho^2\xi^2 - \mathbb{E}^{\mathcal{F}}X_1)\left(1 - \frac{1}{1+\alpha}\right)\right\} \\ & + \quad \exp\left\{-\left(\frac{-n\epsilon^2 + \mathbb{E}^{\mathcal{F}}(X_{j-1}^2)}{K'(1+\alpha)}\right)^{1/\alpha} n^2\epsilon^2\mathbb{E}^{\mathcal{F}}(|X_{j-1}^2|)\left(1 - \frac{1}{1+\alpha}\right)\right\} \end{aligned}$$

Corollary 5.4.1 *The sequence* $(\widehat{\theta}_n)_{n\in\mathbb{N}}$ *is completely converges to the parameter* θ *of* *autoregressive process AR(1) model. Then we have*

$$\sum_{n=1}^{+\infty} \mathbb{P}^{\mathcal{F}}(\sqrt{n}|\widehat{\theta}_n - \theta| > \rho) < +\infty. \tag{5.16}$$

Proof. By using Theorem (5.3.2) we get the result of (5.16) immediately.

Bibliography

[1] Chen, X. and Liu, F. (2017). Strong laws of large numbers for negatively dependent random variables under sublinear expectations, Communications in Statistics - Theory and Methods 46 (24) :12387-12400.

[2] Chow, Y.S, Teicher, H (1978). Probability: Independence, Interchangeability, Martingales. Springer, New York.

[3] Hsu, P. L. and Robbins, H. (1947). Complete convergence and the law of large numbers. Proceedings of the National Academy of Sciences of the United States of America 33: 25-31.

[4] Giuliano Antonini, R. Kozachenko, Y. Volodin, A (2008) Convergence of series dependentϕ-sub Gaussian random variables, J. Math. Anal. Appl. 338: 1188-1203.

[5] Liu, L (2009) Pricise large deviations for dependent random variables with heavy tails, Statist. Probab. Lett. 99: 1290-1298;

[6] Majerek, D. Nowak, W. Zieba, W (2005). Conditional strong law of large numbers, Int.J.Pure Appl.Math. 20: 143-157.

[7] Parakasa Rao, B.L.S (2009) Conditional independence, conditional mixing and conditional association, Ann. Inst. Statist. Math. 61: 441-473.

[8] Petrov, W (1995) Limit theorems of probability theory, Oxford University Press, London.

[9] Roussas, G.C (2008) On conditional independence, mixing and association, Stoch. Anal. Appl. 26: 1274-1309.

[10] Yuan, D.M, An, J, Wu, X.S (2010) Conditional limit theorems for conditionally negativelly associated random variables, Monatsh. Math. 161: 449-473.

Conclusion and Perspectives

Conclusion

We are interested in this thesis to establish a new tail probability inequality for the distributions of sums of widely orthant dependent (WOD, in short) random variables with application to hazard estimator. Then we sutdy the conditionally complete convergence of partial sums of random variables with applicatin to AR(1) generated by the errors in the conditional dependent cases (conditional linearly negative quadrant dependent, conditional extended acceptable)

Perspectives

In this section, we sketch some perspectives for possible future researches.

1. Consider the asymptotique distributions of the error density estimators in p order autoregressive models
2. It is possible to study the conditional complete convergence for $\mathcal{F}-WOD$ random variables and its applications in autoregressive AR(p) models.
3. Study the case of a process with $\phi-$ mixing functional variables.
4. study the cases of the models ARMA and GARCH.

Bibliography

[1] Basawa, I.V., Prakasa Rao, B.L.S., (1980). Statistical Inference for Stochastic Processes. Academic Press, London.

[2] Basawa, I.V., Scott, D., (1983). Asymptotic optimal inference for non-ergodic models. Lecture Notes in Statistics, vol.17. Springer, New York

[3] Bollerslev, T., (1986). Generalized Autoregressive Conditional Heterosckedasticity. J. of Econometrics: 307-327.

[4] Box, G., Jenkins, G., (1970). Time series analysis.Forecasting and Daviscontrol. HoldenDay.

[5] Bulinski, A.V., and Shanavich, E., (1998). Asymptotical behaviour for some functionals of positively and dependand random fields.

[6] Brockwell P.J and Davis R.A., (1991). Time series: theory and methods. Springer Series in Statistics. Springer Verlag, New York, Second Edition.

[7] Cheng, Hu. (2016). A strong law of large numbers for sub-linear expectation under a general moment condition. Statistics and Probability Letters. **119**: 248-258.

[8] Chen, X. and Liu, F. (2017). Strong laws of large numbers for negatively dependent random variables under sublinear expectations, Communications in Statistics - Theory and Methods **46 (24)**:12387-12400.

[9] Chow, Y. S., Teicher, H. (1978). Probability theory: independence, interchangeability, martingales. New York: Springer.

[10] Christofides, TC., Hadjikyriakou, M., (2013). Conditional demimartingales and related results.J.Math.Anal.Appl. **398**:380-391.

[11] Engle, B. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Ináation. Econometrica: 987-1008.

[12] Esary, J., Proschan, F., Walkup, D., (1967). Association of random variables with applications, Ann. Math. Statist. **38**: 1466-1474.

[13] John H. Cochrane., (1997). Time Series for Macroeconomics and Finance, Graduate School of Business, University of Chicago, spring.

[14] He, W., Cheng D. Y. and Wang, Y. B. (2013) Asymptotic lower bounds of precise large deviations with nonnegative and dependent random variables. Stat Probab Lett **83**: 331-338.

[15] Hipel, K.W., McLeod, A.I., (1994). Time Series Modelling of Water Resources and Environmental System, Amsterdam, Elsevier.

[16] Hsu, P. L. and Robbins, H. (1947). Complete convergence and the law of large numbers. Proceedings of the National Academy of Sciences of the United States of America**33**: 25-31.

[17] Galbraith, J.W., Zinde-Walsh, V., (2001). Autoregression-based estimators for ARFIMA models, CIRANO Working Papers, No: 2011s-11.

[18] Giuliano Antonini, R. Kozachenko, Y. Volodin, A (2008) Convergence of series dependent ϕ-sub Gaussian random variables, J. Math. Anal. Appl.**338**: 1188-1203.

[19] Gut, A., (2005). Probability:AGraduateCourse.Springer,Berlin.

[20] Guttorp, P. (1991). Statistical inference for branching processes. New York: Wiley.

[21] Grzenda, W., Zieba, W. (2008). Conditional central limit theorems, Int. Math. Forum **3**: no 29-32, 1521-1528.

[22] Kruk, L., Zieba, W., (1993). On almost sure convergence asymptotic martingales, Annales Universitas Marie Curie-Sklodowska, **47**: ,No. 8, 82-89.

[23] Lee, J., Univariate time series modeling and forecasting (Box-Jenkins Method), Econ **413**, lecture 4.

[24] Lehmann, E.L., (1966). Some concepte of dependance. Ann.Math.Statist.

[25] Liu, L., (2009). Precise large deviations for dependent random variables with heavy tails, Stat Probab Lett, **79**: 1290-1298.

[26] Li, Y. and Yang, S. (2005). Strong convergence rate of recursive probability density estimator for NA sequences, Chinese J. Engineering Mathematics,**22(4)**: 659-665.

[27] Majerak, D., Nowak, W., Zieba, W., (2005). Conditional strong law of large number. International Journal of Pure and Applied Mathematics, **20**, 143-157.

[28] Matula, P., (1994). Some limit theorems for negatively dependent random variables. Yokohama Math. J. **41**:163-173.

[29] Newman, C.M., (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. In: Tong, Y.L. (ed). Inequalities in Statistics and Probability. IMS Lectures Notes-Monograph Series, vol. 5, pp. 127-140. Hayward, California

[30] Park, H., (1999). Forecasting Three-Month Treasury Bills Using ARIMA and GARCH Models, Econ 930, Department of Economics, Kansas State University.

[31] Prakasa Rao, B. L. S., (2009). Conditional independence, conditional mixing and conditional association, Ann. Inst. Statist. Math. **61**: no. 1, 441-460.

[32] Prakasa Rao, B. L. S., Dewan, I. (2001). Associated sequences and related inference problems, In C. R. Rao, D. N. Shanbag (Eds.), Handbook of statistics, 19, stochastic Processes: theory and methods(pp. 693-728). Amsterdam: North Holland.

[33] Prakasa Rao, B. L. S. (1999b). Semimartingales and their statistical inference. Boca Raton: CRC Press

[34] Parzen,E. (1962). On estimation of a probability density function and mode, Ann. Math. Statist. **33**: 1065-1076.

[35] Petrov, W (1995) Limit theorems of probability theory, Oxford University Press, London.

[36] Raicharoen, T., Lursinsap, C., Sanguanbhoki, P., (2003). Application of critical support vector machine to time series prediction, Circuits and Systems,. ISCAS 03.Proceedings of the 2003 International Symposium on Volume **5**: 741-744.

[37] Roussas, G. G. (1999). Positive and negative dependence with some statistical applications. In S. Ghosh(Ed.), Asymptotics, nonparametrics and time series . New York: Marcel Dekker, (pp. 757-788).

[38] Roussas, G.C (2008) On conditional independence, mixing and association, Stoch. Anal. Appl. **26**: 1274-1309.

[39] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of density function, Ann. Math. Statist.**27**: 832-837.

[40] Shen, A., Hu, S., Volodin, A., Wang, X., (2011). Some exponential inequalities for acceptable random variables and complete convergence.J.Inequal. Appl,2011, 142.

[41] Stoyanov, J.M. (1989). Counter examples in probability. John Wiley and Sons, Chichester, New York, Brisbane, Torento, Singapore.

[42] Theodoridis, Sergios (2015-04-10). Chapter 1. Probability and Stochastic Processes. Machine Learning: A Bayesian and Optimization Perspective. Academic Press, 2015. pp. **951**. ISBN 978-0-12-801522-3.

[43] Tong, H., (1983). Threshold Models in Non-Linear Time Series Analysis, Springer-Verlag, New York.

[44] Yuan, D.M., Yang, Y.K., (2011). Conditional versions of limit theorems for conditionally associated random variables. J. Math. Anal. Appl. **376**: 282-293

[45] Yuan, D. M., An, J., Wu, X.S., (2010). Conditional limit theorems for conditionally negatively associated random variables. Monatsh Math. **161**: 449-473.

[46] Yuan, D-M, Xie, Y (2012). Conditional limit theorems for conditionally linearly negative quadrant dependent random variables. Monatshefte Math. **166**: 281-299.

[47] Yuan, D.M., Wu, X.S., (2010). Limiting behavior of the maximum of the partial sum for asymptotically negatively associated random variables under residual Cesàro alpha-integrability assumption. J. Stat. Plan.Infer. **140**: 2395-2402

[48] Yule, G., Udny (1927). On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers, Philosophical Transactions of the Royal Society of London, Ser. A, Vol.**226**: 267-298

[49] Wang, K. Y., Wang, Y. B. and Gao, Q. W. (2013). Uniform asymptotics for the finite-time ruin probability of a new dependent risk model with a constant interest rate. Methodology and Computing in Applied Probability **15**: 109-124.

[50] Wang, J.F., Zhang, L.X., (2006). A Berry?Esseen theorem for weakly negatively dependent random variables and its applications. Acta Math. Hung. **110**: 293-308

[51] Wang, X.J., Hu, S.H., Yang, W.Z., Li, X.Q., (2010). Exponential inequalities and complete convergence for a LNQD sequence. J. Korean Statist. Soc. **39**: 555-564

[52] Walker, Gilbert (1931). On Periodicity in Series of Related Terms, Proceedings of the Royal Society of London, Ser. A, Vol. **131**: 518-532.

[53] Watson, G. S. and Leadbetter, M. R. (1964a). Hazard analysis, I. Biometrika **51**: 175-184.

[54] Watson, G. S. and Leadbetter, M. R. (1964b). Hazard analysis, II.Sankhya A **26**: 110-116.

[55] Zhang, GP., (2007). A neural network ensemble method with jittered training data for time series forecasting, Information Sciences, pages, 5329-5346.

[56] Zhang, GP., (2003). Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing **50**, pages: 159-175.

الـمــلـــخص

في هذه الأطروحة، نحن مهتمون في المقام الأول بدراسة الاحتمال الأسي الشرطي والحدود الشرطية النظرية لتسلسل المتغيرات العشوائية مع تطبيقها على النموذج التراجعي الذاتي. هدفنا الرئيسي هو دراسة التقارب الكامل المشروط للمقدر لعملية الانحدار الذاتي من الدرجة الأولى في حالة اعتماد الخطأ (المرتبطة على نطاق واسع، شرط الممتدة المقبول).

Résumé

Dans cette thèse, nous sommes principalement intéressés à étudier des inégalités de probabilité conditionnelle et theorèmes limites conditionnelles de la séquence des variables aléatoires avec application au modèle autorégressif. Nous fixons comme objectif principal, l'étude de la convergence complète conditionnelle de l'estimateur d'un processus autorégressif de premier ordre dans le cas où l'erreur est dépendante (WOD , $\mathcal{F}-QLND$, conditionnelle prolongée acceptable) .

Summary

In this thesis, we are primarily interested in studying the conditional exponential probability inequalities and Conditional limit theorem of the sequence of random variables with application to the autoregressive model. We have identified the main objective, that the study of the conditional complete convergence of the estimator of first-order autoregressive process in the case where the error is dependent (WOD, $\mathcal{F}-LNQD$, Conditional étendu acceptable).