Gane Samb LO

Mathematical Foundations of Probability Theory

Statistics and Probability African Society (SPAS) Books Series.
Saint-Louis, Calgary, Alberta. 2018.

DOI : http://dx.doi.org/10.16929/sbs/2016.0008

ISBN 978-2-9559183-8-8

SPAS TEXTBOOKS SERIES

List of published books

List of published or scheduled books in English

□ Weak Convergence (IA) - Sequences of Random Vectors. Gane Samb LO, Modou NGOM and Tchilabalo A. KPANZOU. 2016.
Doi : 10.16929/sbs/2016.0001. ISBN 978-2-9559183-1-9

□ A Course on Elementary Probability Theory. Gane Samb LO. 2017.
Doi : 10.16929/sbs/2016.0003. ISBN 978-2-9559183-3-3

□ Measure Theory and Integration By and For the Learner. Gane Samb LO.
Doi : http://dx.doi.org/10.16929/sbs/2016.0005. ISBN 978-2-9559183-5-7

Library of Congress Cataloging-in-Publication Data

Gane Samb LO, 1958-

Mathematical Foundations of Probability Theory.

SPAS Books Series, 2018.

DOI : 10.16929/sbs/2016.0008

ISBN 978-2-9559183-8-8

Author : Gane Samb LO

Emails:
gane-samb.lo@ugb.edu.sn, ganesamblo@ganesamblo.net.

Url's:
www.ganesamblo@ganesamblo.net
www.statpas.net/cva.php?email.ganesamblo@yahoo.com.

Affiliations.
Main affiliation : University Gaston Berger, UGB, SENEGAL.
African University of Sciences and Technology, AUST, ABuja, Nigeria.
Affiliated as a researcher to : LSTA, Pierre et Marie Curie University, Paris VI, France.

Teaches or has taught at the graduate level in the following universities:
Saint-Louis, Senegal (UGB)
Banjul, Gambia (TUG)
Bamako, Mali (USTTB)
Ouagadougou - Burkina Faso (UJK)
African Institute of Mathematical Sciences, Mbour, SENEGAL, AIMS.
Franceville, Gabon

Dedicatory.

To my first and tender assistants, my daughters Fatim Zahrà Lo and Maryam Majigun Azrà Lo

Acknowledgment of Funding.

The author acknowledges continuous support of the World Bank Excellence Center in Mathematics, Computer Sciences and Intelligence Technology, CEA-MITIC. His research projects in 2014, 2015 and 2016 are funded by the University of Gaston Berger in different forms and by CEA-MITIC.

Mathematical Foundations of Probability Theory

Abstract. (**English**) In the footsteps of the book *Measure Theory and Integration By and For the Learner* of our series in Probability Theory and Statistics, we intended to devote a special volume of the very probabilistic aspects of the first cited theory. The book might have assigned the title : From Measure Theory and Integration to Probability Theory. The fundamental aspects of Probability Theory, as described by the keywords and phrases below, are presented, not from experiences as in the book *A Course on Elementary Probability Theory*, but from a pure mathematical view based on Measure Theory. Such an approach places Probability Theory in its natural frame of Functional Analysis and constitutes a firm preparation to the study of Random Analysis and Stochastic processes. At the same time, it offers a solid basis towards Mathematical Statistics Theory. The book will be continuously updated and improved on a yearly basis.
(**Français**)

Keywords. Measure Theory and Integration; Probabilistic Terminology of Measure Theory and Applications; Probability Theory Axiomatic; Fundamental Properties of Probability Measures; Probability Laws of Random Vectors; Usual Probability Laws review; Gaussian Vectors; Probability Inequalities; Almost sure and in Probability Convergences; Weak convergences; Convergence in Lp; Kolmogorov Theory on sequences of independent real-valued random variables; Central Limit Theorem, Laws of Large Numbers, Berry-Essen Approximation, Law of the iterated logarithm for real valued independent random variables; Existence Theorem of Kolmogorov and Skorohod for Stochastic processes; Conditional Expectations; First examples of stochastic process : Brownian and Poisson Processes.

AMS 2010 Classification Subjects : 60-01; 60-02; 60-03;G0Axx; 62GXX.

Contents

General Preface

This textbook is one of the elements of a series whose ambition is to cover a broad part of Probability Theory and Statistics. These textbooks are intended to help learners and readers, of all levels, to train themselves.

As well, they may constitute helpful documents for professors and teachers for both courses and exercises. For more ambitious people, they are only starting points towards more advanced and personalized books. So, these textbooks are kindly put at the disposal of professors and learners.

Our textbooks are classified into categories.

A series of introductory books for beginners. Books of this series are usually destined to students of first year in universities and to any individual wishing to have an initiation on the subject. They do not require advanced mathematics. Books on elementary probability theory (See Lo (2017a), for instance) and descriptive statistics are to be put in that category. Books of that kind are usually introductions to more advanced and mathematical versions of the same theory. Books of the first kind also prepare the applications of those of the second.

A series of books oriented to applications. Students or researchers in very related disciplines such as Health studies, Hydrology, Finance, Economics, etc. may be in need of Probability Theory or Statistics. They are not interested in these disciplines by themselves. Rather, they need to apply their findings as tools to solve their specific problems. So, adapted books on Probability Theory and Statistics may be composed to focus on the applications of such fields. A perfect example concerns the need of mathematical statistics for economists who do not necessarily have a good background in Measure Theory.

A series of specialized books on Probability theory and Statistics of high level. This series begins with a book on Measure

Theory, a book on its probability theory version, and an introductory book on topology. On that basis, we will have, as much as possible, a coherent presentation of branches of Probability theory and Statistics. We will try to have a self-contained approach, as much as possible, so that anything we need will be in the series.

Finally, a series of **research monographs** closes this architecture. This architecture should be so diversified and deep that the readers of monograph booklets will find all needed theories and inputs in it.

We conclude by saying that, with only an undergraduate level, the reader will open the door of anything in Probability theory and statistics with **Measure Theory and integration**. Once this course validated, eventually combined with two solid courses on topology and functional analysis, he will have all the means to get specialized in any branch in these disciplines.

Our collaborators and former students are invited to make live this trend and to develop it so that the center of Saint-Louis becomes or continues to be a re-known mathematical school, especially in Probability Theory and Statistics.

Introduction

Mathematical Foundation of Probability Theory.

In the introduction to the book *Measure Theory and Integration By and For The Learner*, we said :

Undoubtedly, Measure Theory and Integration is one of the most important part of Modern Analysis, with Topology and Functional Analysis for example. Indeed, Modern mathematics is based on functional analysis, which is a combination of the Theory of Measure and Integration, and Topology.

The application of mathematics is very pronounced in many fields, such as finance (through stochastic calculus), mathematical economics (through stochastic calculus), econometrics [which is a contextualization of statistical regression to economic problems], physic statistics. Probability Theory and Statistics has become an important tool for the analysis of biological phenomena and genetics modeling.

This quotation already stressed the important role played by Probability Theory in the application of Measure Theory. So, Probability Theory seems to be one of the most celebrated extensions of Measure Theory and Integration when it comes to apply it to real life problems.

Probability Theory itself may be presented as the result of modeling of stochastic phenomena based on random experiences. This way is illustrated in the element of this series : *A Course on Elementary Probability Theory.*

But for theoretical purposes, it may be presented as a mathematical theory, mainly based on Measure Theory and Integration, Topology and Functional Analysis. This leads to impressive tools that reveal themselves very powerful in dealing real-life problems.

In this book, we tried to give the most common elements of the Theory as direct rephrasing and adaptation of results Measure Theory according to the following scenario.

Chapter 1 is devoted to a complete rephrasing of the Measure Theory and Integration Terminology to that of Probability Theorem, moving from a general measures to normed measures called Probability Measures.

Chapters 2, 3 and deal with a simple fact in Measure Theory and Integration, namely the image-measure, which becomes the most important notion in Probability Theory and called under the name of Probability Laws. Chapter 2 includes a wide range of characterizations for Probability Laws of Random vectors we might need in research problems in Probability Theory and Mathematical Statistics. In particular, the concept of independence is visited from various angles, which leads to a significant number of important characterizations of it. In Chapter 3, usual and important probability Laws are given and reviewed in this chapter in connection with the their generations described made in Lo (2017a). Finally Chapter 3 presents the so important Gaussian random vectors.

Chapter 5 is concerned with the theory of convergence of sequences of (real-valued, mainly) random variables. The three types of Convergence : Almost-sure, in Probability and in L^p.

It is important to notice the the book *Weak Convergence (IA). Sequences of random vectors* (See **?**) has its place exactly here, within the global frame of the series. Due to its importance and its size, we preferred to devote a booklet of medium size (about two hundred pages) to an introduction to weak convergence.

Because of the importance of Inequalities in Probability Theory, we devote Chapter 6 to them. This chapter will continuously updated and augmented on a yearly basis.

In Chapter 7, we presented the main results of the study of sequence of independent random variables which occupied the researchers in a great part of the 19th century. The laws that were studied are until now the most important ones of the theory, although they are exented to the non-independent cases nowadays. But there is no way to join the current studies if the classical main tools and proofs are not mastered.

We introduce to the Kolmogorov Strong Law of Large numbers, the Central Limit Theorem, the Berry-Essen Approximation and the Law of the Iterated Logarithm.

Chapter 8 uses the Radon-Nikodym Theorem to found the important notion of Mathematical Expectation which is the main tool form moving to independent to dependent data.

Finally, Chapter 9 presents the Fundamental Theorem of Kolmogorov which is considered as the foundation of Modern Probability Theory. Versions of the Theorem are given, among them, the Skorohod Theorem. This chapter is the bridge with the course on Stochastic processes.

The place of the book within the series.

While the book *A Course on Elementary Probability Theory* may read at any level, the current one should no be read before the full exposition of Measure Theory and Integration (Lo (2017b) or a similar book). Indeed, the latter book is cited in any couple of pages. The demonstrations in that book are quoted in the current one. Without assuming those demonstration, this textbook would have a very much greater number of pages.

Reading the textbook **?** is recommended after Chapter 5 of the current book.

Now, this book combined with Lo *et al.* (2016) open the doors of many other projects of textbooks, among whom we cite :

(a) Asymptotics of Sequences of Random Vectors

(b) Stochastic Processes

(c) Mathematical Statistics

(d) Random Measures

(e) Times Series

(f) etc.

Consequently, the series will expand to those areas.

CHAPTER 1

An update of the Terminology from Measure Theory to Probability Theory

1. Introduction

This course of Probability Theory is the natural continuation of the one on Measure Theory and Integration. Its constitutes the very minimal basis for a fundamental course which enables to prepare for more advanced courses on Probability Theory and Statistics, like Stochastic Processes, Stochastic Calculus or to prepare specialized Mathematical statistics, etc.

The book *A Course on Elementary Probability Theory* (Lo (2017a)) of this series concentrated on discrete probability measures and focused on random experiences, urn models, generation of random variables and associated computations. The reader will not find such results here. We recommend him to go back to this book or to similar ones which directly deal with Probability Theory related to real experiences. This textbook treats the mathematical aspects of Probability Theory, as a branch of Measure Theory and Integration as exposed in Lo (2017b), where the Measure Theory terminology can be found.

This course begins with new expressions and names of concepts introduced in Measure Theory and Integration. Next, a specific orientation will be taken to present the base of modern Probability Theory.

2. Probabilistic Terminology

2.1. Probability space.

A probability space is a measure space $(\Omega, \mathcal{A}, m)$ where the measure assigns the unity value to the whole space Ω, that is,

$$m(\Omega) = 1.$$

Such a measure is called a probability measure. Probability measures are generally denoted in blackboard font : $\mathbb{P}$, $\mathbb{Q}$, etc.

We begin with this definition :

Definition 1. *Let* $(\Omega, \mathcal{A})$ *be a measurable space. The mapping*

$$\begin{array}{rccc} \mathbb{P}: & \mathcal{A} & \rightarrow & \mathbb{R} \\ & A & \hookrightarrow & \mathbb{P}(A) \end{array}$$

is a probability measure if and only if $\mathbb{P}$ *is a measure and* $\mathbb{P}(\Omega) = 1$, *that is :*

(a) $0 \leq \mathbb{P} \leq \mathbb{P}(\Omega) = 1$.

(b) For any countable collection of measurable sets $\{A_n, n \geq 0\} \subset \mathcal{A}$, *pairwise disjoints, we have*

$$\mathbb{P}(\sum_{n \geq 0} A_n) = \sum_{n \geq 0} \mathbb{P}(A_n).$$

We adopt a special terminology in Probability Theory.

(1) The whole space Ω is called *universe*.

(2) Measurable sets are called *events*. Singletons are elementary events whenever they are measurable.

Example. Let us consider a random experience in which we toss two dies and get the outcomes as the ordered pairs (i, j), where i and j are respectively the number of the first and next the second face of the two dies which come out. Here, the universe is $\Omega = \{1, ..., 6\}^2$. An ordered pair $\{(i, j)\}$ is an elementary event. As an other example, the event : *the sum of the faces is less or equal to 3* is exactly

$$A = \{(1, 1), (1, 2), (2, 1)\}.$$

(3) *Contrary event.* Since $\mathbb{P}(\Omega) = 1$, the probability of the complement of an event A, also called the contrary event to A and denoted $\overline{A}$, is computed as

$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A).$$

The previous facts form simple transitions from Measure Theory and Integration terminology to that of Probability Theory. We are going to continue to do the same in more elaborated transitions in the rest of that chapter.

2.2. Properties of a Probability measure.

Probability measures inherit all the properties of a measure.

(P1) A probability measure is sub-additive, that is, for any countable collection of events $\{A_n, n \geq 0\} \subset \mathcal{A}$, we have

$$\mathbb{P}(\bigcup_{n\geq 0} A_n) \leq \sum_{n\geq 0} \mathbb{P}(A_n).$$

(P2) A probability measure $\mathbb{P}$ is non-decreasing, that is, for any ordered pair of events $(A, B) \in \mathcal{A}^2$ such that $A \subset B$, we have

$$\mathbb{P}(A) \leq \mathbb{P}(B)$$

and more generally for any ordered pair of events $(A, B) \in \mathcal{A}^2$, we have

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

(P3) A probability measure $\mathbb{P}$ is continuous below, that is, for any non-decreasing sequence of events $(A_n)_{n\geq 0} \subset \mathcal{A}$, we have

$$\mathbb{P}(\bigcup_{n\geq 0} A_n) = \lim_{n\to+\infty} \mathbb{P}(A_n),$$

and is continuous above, that is, for any non-increasing sequence of events $(A_n)_{n\geq 0} \subset \mathcal{A}$, we have

$$\mathbb{P}(\bigcap_{n\geq 0} A_n) = \lim_{n\to+\infty} \mathbb{P}(A_n)$$

The *continuity above* in Measure Theory requires that the values of the measures of the A_n's be finite for at least one integer $n \geq 0$. Here, we do not have to worry about this, since all $\mathbb{P}(A_n)$'s are bounded by one.

2.3. Random variables.

Measurable mappings are called random variables. Hence, a mapping

$$X : \ (\Omega, \mathcal{A}) \ \rightarrow \ (E, \mathcal{B}) \tag{2.1}$$

is a random variable, with respect to the σ-algebras $\mathcal{A}$ and $\mathcal{B}$ if and only if it is measurable with respect to the same σ-algebras.

Probability law.

There is not a more important phrase in Probability Theory that *Probability law*. I dare say that the essence of probability Theory is finding probability laws of random phenomena by intellectual means and the essence of Statistical theory is the same but by means of inference from observations or data.

Suppose that we have a probability measure $\mathbb{P}$ on the measurable space $(\Omega, \mathcal{A})$ in Formula 2.1. We have the following definition.

Definition 2. *The Probability law of the random variable X in Formula 2.1 is the image-measure of $\mathbb{P}$ by X, denoted as $\mathbb{P}_X$, which is a probability measure on E given by*

$$\mathcal{B} \ni \mathcal{B} \mapsto \mathbb{P}_X(B) = \mathbb{P}(X \in B). \ \diamondsuit$$

Such a simple object holds everything in Probability Theory.

Classification of random variables.

Although the space E in Formula (2.1) is arbitrary, the following cases are usually and commonly studied :

(a) If E is $\overline{\mathbb{R}}$, endowed with the usual Borel σ-algebra, the random variable is called a *real random variables* (*rrv*).

(b) If E is $\overline{\mathbb{R}}^d$ ($d \in \mathbb{N}^*$), endowed with the usual Borel σ-algebra, X is called a d-random vector or a random vector of dimension d, denoted $X = (X_1, X_2, ..., X_d)^t$, where X^t stands for the transpose of X.

(c) More generally if E is of the form $\mathbb{R}^T$, where T is a non-empty set, finite or countable infinite or non-countable infinite, X is simply called a stochastic process. The σ-algebra on $\mathbb{R}^T$, which is considered as the collections of mapping from T to $\mathbb{R}$ is constructed by using the fundamental theorem of Kolmogorov, which generally is stated in the first chapter of a course on Stochastic processes, and which is extensively stated in Chapter 9.

In the special case where $T = \mathbb{N}$, X is a sequence of real random variables $X = \{X_1, X_2, ...\}$.

(d) If E is some metric space (S, d) endowed with the Borel σ-algebra denoted as $\mathcal{B}(S)$, the term *random variable* is simply used although some authors prefer using *random element*.

2.4. Mathematical Expectation.

It is very important to notice that, at the basic level, the mathematical expectation, and later the conditional mathematical expectation, is defined for a real random variable.

(a) Mathematical expectation of rrvs's.

Let $X \; : \; (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B}_\infty(\overline{\mathbb{R}})$ be a *real random variable*. Its mathematical expectation with respect to the probability measure $\mathbb{P}$ or its $\mathbb{P}$-mathematical expectation, denoted by $\mathbb{E}_{\mathbb{P}}(X)$ is simply its integral with respect to

$$\mathbb{P}$$

whenever it exists and we denote :

$$\mathbb{E}_{\mathbb{P}}(X) = \int_\Omega X \; d\mathbb{P}.$$

The full notation $\mathbb{E}_{\mathbb{P}}$ of the mathematical expectation reminds us to which probability measure the mathematical expectation is relative to. In many examples, it may be clear that all the mathematical expectations are relative to only one probability measure so that we may drop the subscript and only write

$$\mathbb{E}(X) = \int_\Omega X \; d\mathbb{P}.$$

Also, the parentheses may also be removed and we write $\mathbb{E}X$.

(b) Mathematical expectation of a function of an arbitrary random variable.

For an arbitrary random variable as defined in Formula (2.1) and for any real-valued measurable mapping

$$h: \ (E, \mathcal{B}) \ \rightarrow \ (\overline{\mathbb{R}}, \mathcal{B}_\infty(\overline{\mathbb{R}}), \tag{2.2}$$

the composite mapping

$$h(X) = h \circ X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \overline{\mathbb{R}}$$

is a real random variable. We may define the mathematical expectation of $h(X)$ with respect to $\mathbb{P}$ by

$$\mathbb{E}h(X) = \int_\Omega h(X)\ d\mathbb{P},$$

whenever the integral exists.

(c) Use of the probability law for computing the mathematical expectation.

We already know from the properties of image-measures (See page Lo (2017b), Doc 04-01 Point (V)), that we may compute the mathematical expectation of $h(X)$, if it exists, by

$$\mathbb{E}(h(X)) = \int_\Omega h\ d\mathbb{P}_X = \int_\Omega h(x)\ d\mathbb{P}_X(x). \tag{2.3}$$

If X is itself a real random variable, its expectation, if it exists, is

$$\mathbb{E}(X) = \int_\mathbb{R} x\ d\mathbb{P}_X(x). \tag{2.4}$$

(d) **Mathematical expectation of a vector**.

The notion of mathematical expectation may be extended to random vectors by considering the vector of the mathematical expectations of the coordinates. Let us consider the random vector X such that

$X^t = (X_1, X_2, ..., X_d)$. The Mathematical vector expectation $\mathbb{E}(X)$ is defined by

$$(\mathbb{E}(X))^t = (\mathbb{E}X_1, \mathbb{E}X_2, ..., \mathbb{E}X_d).$$

A similar extension can be operated for random matrices.

(e) Properties of the Mathematical expectation.

As an integral of real-valued measurable application, the mathematical expectation inherits all the properties of integrals we already had in Measure Theory. Here, we have to add that : *constant real random variables and bounded random variables have finite expectations*. The most important legacy to highlight is the following.

THEOREM 1. *On the class of all random variables with finite mathematical expectation denoted $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, the mathematical expectation operator :*

(a) is linear, that is for all $(\alpha, \beta) \in \mathbb{R}^2$, for all $(X, Y) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$,

$$\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y),$$

(b) is non-negative, that for all non-negative X random variable, we have $\mathbb{E}(X) \geq 0$

(c) and satisfies for all non-negative X random variable : $\mathbb{E}(X) = 0$ if and only if $X = 0$, $\mathbb{P}$-a.e.

(c) Besides, we have for all real-valued random variables X and Y defined on $(\Omega, \mathcal{A})$,

$$\left(|X| \leq Y,\ Y \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})\right) \Rightarrow X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P}).$$

and

$$\left|\int X \, d\mathbb{P}\right| \leq \int |X| \, d\mathbb{P} \leq \int Y \, d\mathbb{P}.$$

The first formula in the following Lemma is often used to computing the mathematical expectation of non-negative real-valued randoms variables. We generalize with respect to the counting measure. For example, this will render much comprehensible the proof the Kolmogorov Theorem 17 (Chapter 7, page 228) on strong laws of large numbers.

Let us define, for a real-valued random variable X, its lower endpoint $lep(X)$ and the upper endpoint $uep(X)$ respectively by

$$lep(X) = \inf\{t \in \mathbb{R},\ \mathbb{P}(X \leq t) > 0\},\ uep(X) = \sup\{t \in \mathbb{R},\ \mathbb{P}(X \leq t) < 1\}.$$

This means that $\mathbb{P}(X \leq t) = 0$ for all $t > uep(X)$ and similarly, we have $\mathbb{P}(X \leq t) = 0$ for all $t \leq lep(X)$. Actually, we have $uep(X) = \|X\|_\infty$ in the L^∞ space. The values space of X becomes $\mathcal{V}_X = [lep(X),\ uep(X)]$.

We have :

Proposition 1. *Let X be any real-valued and non negative random variable, we have*

$$\mathbb{E}(X) = \int_0^{uep(X)} \mathbb{P}(X > t)dt, \quad (CF)$$

and

$$1 + \mathbb{E}\left([X]_+\right) = \sum_{n \in [0,\ [uep(X)]_+]} \mathbb{P}(X \geq n), \quad (DF1)$$

where $[x]_+$ (resp. $[x]^+$) stands for the greatest (resp. smallest) integer less or equal (resp. greater or equal) to $x \in \overline{\mathbb{R}}$. Also, for any a.s finite real-valued random variable, we have

$$-1 + \sum_{n \in [0,\ [uep(X)]_+]} \mathbb{P}(|X| \geq n) \leq \mathbb{E}|X| \leq \sum_{n \in [0,\ [uep(X)]_+]} \mathbb{P}(|X| \geq n). \ (DF2)$$

Since $\mathbb{P}(X > t) = 0$ for all $t > uep(X)$, extending the integration domain to $+\infty$ does not effect the value of the integral.

Proof.

Proof of (CF). The function $t \mapsto \mathbb{P}(X > t)$ is bounded and has at most a countable number of discontinuity. So its improper Riemann integral is a Lebesgue's one and we may apply the Tonelli's Theorem (See Chapter 8, Doc 07-01 in Lo (2017b)) at Line (L13) below as follows :

$$\begin{aligned}
\int_0^{uep(X)} \mathbb{P}(X>t)dt &= \int_0^{uep(X)} \mathbb{P}(X>t)d\lambda(t) \\
&= \int_{(t\in]0,uep(X)])} \left(\int_{\mathcal{V}_X} 1_{(x>t)}\ d\mathbb{P}_X(x)\right) d\lambda(t) \\
&= \int_{(t\in]0,uep(X)])} 1_{(x>t)}\ d\mathbb{P}_X(x)d\lambda(t) \\
&= \int_{\mathcal{V}_X} \left(\int_{(t\in]0,uep(X)])} 1_{(x>t)}\right) d\mathbb{P}_X(x) \quad (L13) \\
&= \int_{\mathcal{V}_X} \left(\int_0^{\min(x,uep(X)} d\lambda(t)\right) d\mathbb{P}_X(x) \\
&= \int_{\mathcal{V}_X} \min(x, uep(X) d\mathbb{P}_X(x) \\
&= \mathbb{E}(min(X, uep(X))) = \mathbb{E}(X),
\end{aligned}$$

since $X \leq uep(X)$ $a.s.$

Proof of (DF1). We use the counting measure ν on $\mathbb{N}$ and say

$$\begin{aligned}
\sum_{n\geq 0,\ n\leq [uep(X)]^+} \mathbb{P}(X\geq n) &= \int_{[0,\ [uep(X)]^+]} \mathbb{P}(X\geq n)\ d\nu(n) \\
&= \int_{[0,\ [uep(X)]^+]} \left(\int_{\mathcal{V}_X} 1_{(x\geq n)}\ d\mathbb{P}_X(x)\right) d\nu(n) \quad (L22) \\
&= {}_1\int_{\mathcal{V}_X} \left(\int_{[0,\ [uep(X)]^+]} 1_{(x\geq n)} d\nu(n)\right)\ d\mathbb{P}_X(x) \quad (L23) \\
&= \int_{\mathcal{V}_X} \nu([0, \max(x, [uep(X)]^+)])\ d\mathbb{P}_X(x)\ (L24) \\
&= \int_0^{+\infty} \left(\left[\max(x, [uep(X)]^+)\right]_+ + 1\right)\ d\mathbb{P}_X(x) \\
&= \mathbb{E}\left(\left[\max(X, [uep(X)]^+)\right]_+\right) + 1.
\end{aligned}$$

We conclude that

$$\sum_{n\geq 0,\ n\leq [uep(X)]^+} \mathbb{P}(X \geq n) \quad = \quad \mathbb{E}[X]_+ + 1.$$

Proof of (DF2). The left-hand inequality is derived from (DF1) when applied to non-negative random variabe $|X|$. To establish the right-hand inequality, suppose that X is non-negative. Let us denote $A_n = (X \geq n)$, $n \geq 0$ with $A_0 = \Omega$ clearly. We have for any $n \geq 1$, $A_{n-1} \setminus A_n) = (n-1 \leq X < n)$. If $uep(X) = +\infty$, the sets $]n-1, n]$, $n \geq 1$, form a partition of $\mathbb{R}_+$. If $uep(X)$ if finite, the sets $]n-1, n]$, $1 \leq n \leq N = [uep(X)]_+ + 1$ for a partition of $[0, N]$ which covers X *a.s.*. So we have

$$\sum_{1\leq n<N+1} \biggr(A_{n-1} \setminus A_n\biggr) = \Omega.$$

By the Monotone Convergence Theorem when $N = +\infty$, but by finite additivity for N finite, we have

$$\begin{aligned}
\mathbb{E}(X) \quad &= \quad \mathbb{E}\biggr(X \sum_{1\leq n<N+1} 1_{A_{n-1}\setminus A_n}\biggr) \quad (L31) \\
&= \quad \sum_{1\leq n<N+1} \mathbb{E}\biggr(X1_{A_{n-1}\setminus A_n}\biggr) \\
&= \quad \sum_{1\leq n<N+1} \mathbb{E}\biggr(X1_{(n-1\geq |X|<n)}\biggr) \\
&\leq \quad \sum_{1\leq n<N+1} n\mathbb{E}\biggr(1_{(n-1\geq X<n)}\biggr) \\
&\leq \quad \sum_{1\leq n<N+1} n\biggr(\mathbb{P}(A_{n-1}) - \mathbb{P}(A_n)\biggr) \quad (L35)
\end{aligned}$$

Suppose that N is infinite. By developing the last line, we have for

$$\begin{aligned}
\sum_{1\leq n\leq k+1} n(\mathbb{P}(A_{n-1}) - \mathbb{P}(A_n)) \quad &= \quad \sum_{0\leq n\leq k} \mathbb{P}(A_n) - (k+1)\mathbb{P}(A_{k+1}) \\
&\leq \quad \sum_{0\leq n\leq k} \mathbb{P}(A_n). \quad (L42)
\end{aligned}$$

By letting $k \to +\infty$ in Line (42) and by combining the results with Lines (L31) and (L36), we get the inequality.

If N is finite, the last line is exactly

$$\sum_{0 \leq n \leq N} \mathbb{P}(A_n) - N\mathbb{P}(A_N \leq \sum_{0 \leq n \leq N} \mathbb{P}(A_n),$$

and hence, is less or equal to $\sum_{0 \leq n \leq [uep(X)]_+} \mathbb{P}(A_n)$ and we conclude that

$$\mathbb{E}(X) \leq \sum_{0 \leq n \leq [uep(X)]_+} \mathbb{P}(A_n).$$

To get the right-hand inequality in (DF2), we just apply the last formula to $|X|$. ■

An easy example. Suppose that X is Bernoulli random variable with $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p$, $0 < p < 1$. We have $\mathbb{E}(X) = p$, $uep(X) = 1$, $[uep(X)]_+ = 1$, $\mathbb{P}(A_0) = 1$, $\mathbb{P}(A_1) = p$ and we exactly have

$$\sum_{0 \leq n \leq [uep(X)]_+} \mathbb{P}(A_n) = 1 + p = \mathbb{E}(X) + 1.$$

2.5. Almost-sure events.

In Measure Theory, we have studied null-sets and the notion of almost-everywhere *(a.e)* properties. In the context of probability theory, for any null-set N, we have

$$\mathbb{P}(\overline{N}) = 1 - \mathbb{P}(N) = 1.$$

So, the complement of any null-set is an almost-sure (*a.s.*) event. Then, a random property $\mathbb{P}$ holds *a.s.* if and only if

$$\mathbb{P}(\{\omega \in \Omega,\ \mathbb{P}(\omega) \text{ true }\}) = 1.$$

An almost-everywhere (a.e.) property is simply called an almost-sure (a.s.) property. Let us recall some properties of *a.e.* properties.

(P1) If A is an *a.e.* event and if the event B is contained A, then B is an *a.s.* event.

(P2) A countable union of *a.s.* events is an *a.s.* event.

(P3) If each assertion of a countable family of assertions holds *a.s.*, then all the assertions of the family hold simultaneously *a.s.*

2.6. Convergences of real-valued random variables.

We may also rephrase the convergence results in Measure Theory as follows.

(a) Almost-sure convergence.

Let X and X_n, $n \geq 0$, be random variables defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The sequence $(X_n)_{n\geq 0}$ converges to X almost-surely, denoted

$$X_n \to X, \ a.s.,$$

as $n \to +\infty$ if and only if $(X_n)_{n\geq 0}$ converges to X *a.e.*, that is

$$\mathbb{P}(\{\omega \in \Omega, \ X_n(\omega) \to X(\omega)\}) = 1.$$

(b) Convergence in Probability.

The convergence in measure becomes the convergence in Probability. Let X be an *a.s.*-finite real random variable and $(X_n)_{n\geq 0}$ be a sequence of *a.e.*-finite real random variables. We say that $(X_n)_{n\geq 0}$ converges to X in probability, denoted

$$X_n \xrightarrow{\mathbb{P}} X,$$

if and only if, for any $\varepsilon > 0$,

$$\mathbb{R}(|X_n - X| > \varepsilon) \to 0 \ as \ n \to +\infty.$$

We remind that the *a.s.* limit and the limit in probability are *a.s.* unique.

We re-conduct the comparison result from Measure Theory.

(c) Comparison between these two Convergence types.

Let X be an *a.s.*-finite real random variable and $(X_n)_{n\geq 0}$ be a sequence of *a.e.* finite real random variables. Then we have the following implications, where all the unspecified limits are done as $n \rightarrow +\infty$.

(1) If $X_n \rightarrow X$, *a.s.*, then $X_n \xrightarrow{\mathbb{P}} X$.

(2) If $X_n \xrightarrow{\mathbb{P}} X$, then there exists a sub-sequence $(X_{n_k})_{k\geq 0}$ of $(X_n)_{n\geq 0}$ such that $X_{n_k} \rightarrow X$, *a.s.*, as $k \rightarrow +\infty$.

Later, we will complete the comparison theorem by adding the convergence in the space

$$L^p = \{X \in L^0,\ \mathbb{E}\,|X|^p < \infty\},\ p \geq 1.$$

The weak convergence also will be quoted here while its study is done in Lo *et al.* (2016).

3. Independence

The notion of independence is extremely important in Probability Theory and its applications. The main reason that the theory, in its earlier stages, has been hugely developed in the frame of independent random variables. Besides, a considerable number methods of handling dependent random variables are still generalizations of techniques used in the independence frame. In some dependence studies, it is possible to express the dependence from known functions of independent objects. In others, approximations based on how the dependence is near the independence are used.

So, mastering the independence notion and related techniques is very important. In the elementary book (Lo (2017a)), we introduced the independence of events in the following way :

Definition. Let A_1, A_2, . . . , A_n be events in a probability space $(\Omega, \mathbb{P}(\Omega), \mathbb{P})$. We have the following definitions.

(A) The events A_1, A_2, . . . , A_{n-1} and A_n are pairwise independent if and only if

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\ \mathbb{P}(A_j), \text{ for all } 1 \leq i \neq j \leq n.$$

(B) The events A_1, A_2, . . ., A_{n-1} and A_n are mutually independent if and only if for any subset $\{i_1, i_2, ..., i_k\}$ of $\{1, 2, ..., n\}$, with $2 \leq k \leq n$, we have

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}) = \mathbb{P}(A_{i_1})\ \mathbb{P}(A_{i_2})...\mathbb{P}(A_{i_k}).$$

(C) Finally, the events A_1, A_2, . . ., A_{n-1} and A_n fulfills the global factorization formula if and only if

$$\mathbb{P}(A_1 \cap A_2 \cap ... \cap A_n) = \mathbb{P}(A_1)\ \mathbb{P}(A_2)...\mathbb{P}(A_n).$$

We showed with examples that none two definitions from the three definitions (A), (B) and (C) are equivalence. It is important to know that, without any further specification, *independence* refers to Definition (B).

Measure Theory and Integration (MTI) *gives the nicest and most perfect way* to deal with the notion of independence and, by the way, with the notion of dependence with copulas.

3.1. Independence of random variables.

Let $X_1, ..., X_n$ be n random variables defined on the same probability space

$$X_i \quad (\Omega, \mathcal{A}, \mathbb{P}) \ \mapsto \ (E_i, \mathcal{B}_i).$$

Let $(X_1, ..., X_n)$ be a the n-tuple defined by

$$(X_1, ..., X_n)^t : \ (\Omega, \mathcal{A}) \ \mapsto \ (E, \mathcal{B})$$

where $E = \Pi_{1 \leq i \leq n} E_i$ is the product space of the E_i's endowed with the product σ-algebra, $\mathcal{B} = \otimes_{1 \leq i \leq n} \mathcal{B}_i$. On each $(E_i, \mathcal{B}_i)$, we have the probability law $\mathbb{P}_{X_i}$ of X_i.

Each of the $\mathbb{P}_{X_i}$'s is called a marginal probability law of $(X_1, ..., X_n)^t$.

On $(E, \mathcal{B})$, we have the following product probability measure

$$\mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_n},$$

characterized on the semi-algebra

$$S = \{\Pi_{1 \leq i \leq n}\ A_i, A_i \in \mathcal{B}_i\}$$

of measurable rectangles by

$$\mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_n} \left(\prod_{1 \leq i \leq n} A_i) = \prod_{1 \leq i \leq n} \mathbb{P}_{X_i}(A_i \right). \tag{3.1}$$

Now, we have two probability measures

$$\mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_n}$$

that is the product probability measure of the marginal probability measures and the probability law

$$\mathbb{P}_{(X_1,...,X_n)}(B) = \mathbb{P}((X_1, ..., X_n) \in B).$$

of the n-tuple $(X_1, ..., X_n)$ on $(E, \mathcal{B})$, with is the image-measure of $\mathbb{P}$ by $(X_1, ..., X_n)$. The latter probability measure is called the joint probability measure.

By the λ-π Lemma (See Lo (2017b), Exercise 11 of Doc 04-02, Part VI, page 228), these two probability measures are equal whenever they agree on the semi-algebra $\mathcal{S}$.

Now, we may give the most general definition of the independence of random variables :

DEFINITION 3. *The random variables $X_1, ...,$ and X_n are independent if and only if the joint probability law $\mathbb{P}_{(X_1,...,X_n)}$ of the vector $(X_1, ..., X_n)$ is the product measure of its marginal probability laws $\mathbb{P}_{X_i}$, that is :*

For any $B_i \in \mathcal{B}_i$, $1 \leq i \leq n$,

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, ..., X_n \in B_n) = \prod_{1 \leq i \leq n} \mathbb{P}_{X_i}(B_i). \tag{3.2}$$

For an ordered pair of random variables, the two random variables

$$X : \ (\Omega, \mathcal{A}) \ \mapsto \ (E, \mathcal{B})$$

and

$$Y \ \ (\Omega, \mathcal{A}) \ \mapsto \ (F, \mathcal{G})$$

are independent if and only if $A \in \mathcal{B}$ et B$\in \mathcal{G}$,

$$\mathbb{P}(X \in B, Y \in G) = \mathbb{P}(X \in A) \times \mathbb{P}(Y \in B).$$

Important Remark. The independence is defined for random variables defined on the same probability space. The space in which they take values may differ.

Formula (3.2) may be rephrased by means of measurable functions in place of measurable subsets. We have

THEOREM 2. *The random variables X_1, ..., and X_n are independent if and only if, for all non-negative and measurable real-valued functions $h_i : (E_i, \mathcal{B}_i) \mapsto \mathbb{R}$, we have*

$$\mathbb{E}\left(\prod_{1 \leq i \leq n} h_i(X_i)\right) = \prod_{1 \leq i \leq n} \mathbb{E}(h_i(X_i)). \tag{3.3}$$

Proof.

We have to show the equivalence between Formulas (3.2) and (3.3). Let us begin to suppose that Formula (3.3) holds. Let us prove Formula (3.2). Let $A_i \in \mathcal{B}$ and set $h_i = 1_{A_i}$. Each h_i is non-negative and measurable. Further

$$h_i(X_i) = 1_{A_i}(X) = 1_{(X_i \in A_i)}.$$

and then

$$\mathbb{E}(h_i(X_i)) = \mathbb{E}(1_{(X_i \in A_i)}) = \mathbb{P}(X_i \in A_i). \tag{3.4}$$

As well, we have

$$\prod_{1 \leq i \leq n} h_i(X_i) = \prod_{1 \leq i \leq n} 1_{(X_i \in A_i)} = 1_{(X_1 \in A_1, \dots, X_n \in A_n)}$$

and then

$$\begin{aligned} \mathbb{E}(\prod_{1 \leq i \leq n} h_i(X_i)) &= \mathbb{E}(1_{(X_1 \in A_1, \dots, X_n \in A_n)}) \\ &= \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n). \end{aligned} \tag{3.5}$$

By putting together (3.4) and (3.5), we get (3.2).

Now, assume that (3.2) holds. Let $h_i : (E_i, \mathcal{B}_i) \mapsto \mathbb{R}$ be measurable functions. Set

$$\mathbb{E}\left(\prod_{1 \leq i \leq n} h_i(X_i)\right) = \mathbb{E}(h(X_1, ..., X_n)),$$

where $h(x_1, ..., x_n) = h_1(x_1)h_2(x_2)...h_n(x_n)$. The equality between the joint probability law and the product margin probability measures leads to

$$\mathbb{E}(h(X_1, ..., X_n)) = \int h(x_1, ..., x_n) \ d\mathbb{P}_{(X_1,...,X_n)}(h(x_1, ..., x_n)$$

$$= \int h(x_1, ..., x_n) \ d\left\{\mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_n}\right\}(x_1, ..., x_n).$$

From there, we apply Fubini's theorem,

$$\begin{aligned}
& \mathbb{E}(h(X_1, ..., X_n)) \\
= \ & \int_{\Omega_1} d\mathbb{P}_{X_1}(x_1) \int_{\Omega_2} d\mathbb{P}_{X_2}(x_2) \int ... \\
... \ & d\mathbb{P}_{X_{n-1}}(x_{n-1}) \int h(x_1, ..., x_n) \ d\mathbb{P}_{X_n}(x_1) \\
= \ & \int_{\Omega_1} d\mathbb{P}_{X_1}(x_1) \int_{\Omega_2} d\mathbb{P}_{X_2}(x_2) \int ...d\mathbb{P}_{X_{n-1}}(x_{n-1}) \int h_1(x_1)h_2(x_2)...h_n(x_n) \ d\mathbb{P}_{X_n}(x_1) \\
= \ & \int_{\Omega_1} h_1(x_1) \ d\mathbb{P}_{X_1}(x_1) \int_{\Omega_2} h_2(x_2) \ d\mathbb{P}_{X_2}(x_2)... \int_{\Omega_n} .h_n(x_n) \ d\mathbb{P}_{X_n}(x_1) \\
= \ & \prod_{1 \leq i \leq n} \int_{\Omega_i} h_i(x_i) \ d\mathbb{P}_{X_i}(x_i) = \prod_{1 \leq i \leq n} \int_{\Omega_i} h_i(X_i) \ d\mathbb{P} \\
= \ & \prod_{1 \leq i \leq n} \mathbb{E}(h_i(X_i)). \ \blacksquare
\end{aligned}$$

This demonstration says that we have independence if and only if Formula (3.3) holds for all measurable functions $h_i : (E_i, \mathcal{B}_i) \mapsto \mathbb{R}$, $\mathbb{P}_{X_i}$-integrable or simply for all measurable and bounded functions $h_i : (E_i, \mathcal{B}_i) \mapsto \mathbb{R}$ or for all non-negative measurable functions h_i, $i \in \{1, ..., n\}$.

Let us come back to independence of events.

3.2. Independence of events.

Independence of events is obtained from independence of random variables.

(a) Simple case of two events.

We say that two events $A \in \mathcal{A}$ and $B \in \mathcal{A}$ are independent if and only if the random variables 1_A and 1_B are independent, that is, for all $h_i : \mathbb{R} \rightarrow \mathbb{R}$ (i=1,2) non-negative and measurable

$$\mathbb{E}h_1(1_A)h_2(1_B) = \mathbb{E}h_1(1_A)\mathbb{E}h_2(1_B) \tag{3.6}$$

As a direct consequence, we have for $h_i(x) = x$, that Formula 3.6 implies that

$$\mathbb{E}(1_A 1_B) = \mathbb{E}(1_{AB}) = \mathbb{E}(1_A)\mathbb{E}(1_B),$$

that is

$$\mathbb{P}(AB) = \mathbb{P}(A) \times \mathbb{P}(B). \tag{3.7}$$

Now, we are going to prove that (3.7), in its turn, implies (3.6). First, let us show that (3.7) implies

$$\mathbb{P}(A^c B) = \mathbb{P}(A^c) \times \mathbb{P}(B), \tag{3.8}$$

$$\mathbb{P}(AB^c) = \mathbb{P}(A) \times \mathbb{P}(B^c) \tag{3.9}$$

and

$$\mathbb{P}(A^c B^c) = \mathbb{P}(A^c) \times \mathbb{P}(B^c). \tag{3.10}$$

Assume that (3.7). Since,

$$B = AB + A^c B,$$

we have

$$\mathbb{P}(B) = \mathbb{P}(AB) + \mathbb{P}(A^c B) = \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A^c B).$$

Then

$$\mathbb{P}(A^c B) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(A^c)\mathbb{P}(B).$$

Hence (3.8) holds. And (3.9) is derived in the same manner by exchanging the role of A and B. Now, to prove (3.10), remark that

$$A^cB^c = (A \cup B)^c = (AB^c + A^cB + AB)^c.$$

Then, we get

$$\begin{aligned}
\mathbb{P}(A^cB^c) &= 1 - \mathbb{P}(AB^c) - \mathbb{P}(A^cB) - \mathbb{P}(AB) \\
&= 1 - \mathbb{P}(A)\mathbb{P}(B^c) - \mathbb{P}(A^c)\mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \\
&= 1 - \mathbb{P}(A) - \mathbb{P}(A^c)\mathbb{P}(B) \\
&= 1 - \mathbb{P}(A) - \mathbb{P}(A^c)(1 - \mathbb{P}(B^c)) \\
&= 1 - \mathbb{P}(A) - \mathbb{P}(A^c) + \mathbb{P}(A^c)\mathbb{P}(B^c) \\
&= \mathbb{P}(A^c)\mathbb{P}(B^c).
\end{aligned}$$

Hence (3.10) holds.

Finally, let us show that Formula (3.7) ensures Formula (3.6). Consider two non-negative and measurable mappings $h_i : \mathbb{R} \to \mathbb{R}$, (i=1,2). We have

$$h_1(1_A) = h_1(1)1_A + h_1(0)1_{A^c}$$

and

$$h_2(1_B) = h_2(1)1_B + h_2(0)1_{B^c}.$$

As well, we have

$$\begin{aligned}
h_1(1_A)h_2(1_B) &= h_1(1)h_2(1)1_{AB} + h_1(1)h_2(0)1_{AB^c} \\
&+ h_1(0)h_2(1)1_{A^cB} + h_1(0)h_2(0)1_{A^cB^c}.
\end{aligned}$$

Then, we have

$$\mathbb{E}(h_1(1_A)) = h_1(1)\mathbb{P}(A) + h_1(0)\mathbb{P}(A^c)$$

and

$$\mathbb{E}(h_2(1_B)) = h_2(1)\mathbb{P}(B) + h_2(0)\mathbb{P}(B^c).$$

We also have

$$\mathbb{E}h_1(1_A)h_2(1_B) = h_1(1)h_2(1)\mathbb{P}(A)\mathbb{P}(B) + h_1(1)h_2(0)\mathbb{P}(A)\mathbb{P}(B^c)$$

$$+h_1(0)h_2(1)\mathbb{P}(A^c)\mathbb{P}(B) + h_1(0)h_2(0)\mathbb{P}(A^c)\mathbb{P}(B^c).$$

By comparing the three last formulas, we indeed obtain that

$$\mathbb{E}h_1(1_A)h_2(1_B) = \mathbb{E}h_1(1_A)\mathbb{E}h_2(1_B).$$

The previous developments lead to the definition (and theorem).

DEFINITION 4. *(Definition-Theorem). The events A and B are independent if and only if 1_A and 1_B are independent if and only if*

$$\mathbb{P} =(AB) = \mathbb{P}(A) \times \mathbb{P}(B). \tag{3.11}$$

(b) Case of an arbitrary finite number $k \geq 2$ of events.

Let us extend this definition to an arbitrary number k of events and compare it with the definition (B) in the preliminary remarks of this section.

Let A_i, $1 \leq i \leq k$, be k events and $h_i : \mathbb{R} \rightarrow \mathbb{R}$, $(i = 1, ..., k)$, be k non-negative and measurable mappings. The events A_i are independent if and only if the mappings 1_{A_i}, $1 \leq i \leq k$, are independent if and only if for all measurable finite mappings h_i, $1 \leq i \leq k$, we have

$$\mathbb{E}(\prod_{1\leq i\leq k} h_i(1_{A_i})) = \prod_{1\leq i\leq k} \mathbb{E}(h_i(1_{A_i})). \tag{3.12}$$

Let us put for each s-tuple of non-negative integers $1 \leq i_1 \leq i_2 \leq ... \leq i_s \leq k$, $1 \leq s \leq k$,

$$h_{i_j}(x) = x, \ j = 1, ..., s$$

and

$$h_i(x) = 1 \ for \ i \notin \{i_1, i_2, ..., i_s\}.$$

Hence, by Formula (3.12), we get for any subset $\{i_{1,}i_{2,}..., i_k\}$ of $\{1, 2, ..., n\}$, with $2 \leq k \leq n$

$$\mathbb{P}(\bigcap_{1\leq j\leq s} A_{i_j}) = \prod_{1\leq j\leq s} \mathbb{P}(A_{i_j}). \tag{3.13}$$

This is Definition (B) in the preliminary remarks of this section. By the way, it is also a generalization of Formula (3.7) for two events ensembles. We may, here again, use straightforward computations similar to those done for the case $k \geq 2$, to show that Formula 3.13 also implies Formula 3.12. This leads to the definition below.

Definition 5. *(Definition-Theorem) The events A_i, $1 \leq i \leq k$, are independent if and only if the mappings 1_{A_i} are independent if and only if for each s-tuple $1 \leq i_1 \leq i_2 \leq ... \leq i_s \leq k$, of non-negative integers,*

$$\mathbb{P}\left(\bigcap_{1\leq j\leq s} A_{i_j}\right) = \prod_{1\leq j\leq s} \mathbb{P}(A_{i_j}). \tag{3.14}$$

(c) An interesting remark.

A useful by-product of Formula (3.12) is that if $\{A_i,\ 1 \leq i \leq n\}$, is a collection of independent events, then any elements of any collection of events $\{B_i,\ 1 \leq i \leq n\}$, with $B_i = A_i$ or $B_i = A_i^c$, are also independent.

To see this, it is enough to establish Formula (B). But for any $\{i_1, i_2, ..., i_k\}$ of $\{1, 2, ..., n\}$, with $2 \leq k \leq n$, we make take $h_{i_j}(x) = x$ if $B_{i_j} = A_{i_j}$ or $h_{i_j}(x) = 1 - x$ if $B_{i_j} = A_{i_j}^c$ for $j = 1, ..., k$ and $h_i(x) = 1$ for $i \notin \{i_1, ..., i_k\}$ in Formula 3.12 and use the independence of the A_i's.

We get, for $\{i_1, i_2, ..., i_k\} \subset \{1, 2, ..., n\}$, with $2 \leq k \leq n$, that

$$\mathbb{P}(\bigcap_{1\leq j\leq s} B_{i_j}) = \prod_{1\leq j\leq s} \mathbb{P}(B_{i_j}). \ \square$$

3.3. Transformation of independent random variables.

Consider the independent random variables

$$X_i : \ (\Omega, \mathcal{A}, \mathbb{P}) \ \mapsto \ (E_i, \mathcal{B}_i),$$

$i = 1, ..., n$ and $g_i : (E_i, \mathcal{B}_i) \mapsto (F_i, F_i)$, n measurable mappings.

Then, the random variables $g_i(X_i)$ are also independent.

Indeed, if $h_i : F_i \to \mathbb{R}$, $1 \leq i \leq n$, are measurable and bounded real-valued mappings, then the $h_i(g_i)$ are also real-valued bounded and measurable mappings. Hence, the $h_i(g_i(X_i))$'s are $\mathbb{P}$-integrable. By independence of the X_i, we get

$$\mathbb{E}(\prod_{1\leq i\leq n} h_i \circ g_i(X_i)) = \prod_{1\leq i\leq n} \mathbb{E}(h_i \circ g_i(X_i)),$$

and this proves the independence of the $h_i \circ g_i(X_i)$'s. We have the proposition :

PROPOSITION 2. *Measurable transformations of independent random variables are independent*

3.4. Family of independent random variables. .

Consider a family of random variables

$$X_t \quad (\Omega, \mathcal{A}, \mathbb{P}) \;\mapsto\; (E_t, \mathcal{B}_t),\ (t \in T).$$

This family $\{X_t, t \in T\}$ may be finite, infinite and countable or infinite and non countable. It is said that the random variables of this family are independent if and only the random variables in any finite sub-family of the family are independent, that is, for any subfamily $\{t_1, t_2, ..., t_p\} \subset T$, $2 \leq p < +\infty$, the mappings $X_{t_1}, X_{t_2}, ..., X_{t_p}$ are independent.

The coherence of this definition will be a consequence of the Kolmgorov Theorem.

4. Pointcarré and Bonferroni Formulas

Poincarré or Inclusion-exclusion Formula.

In Lo (2017b), we already proved these following formulas for subsets A_1, ..., A_n of Ω, $n \geq 2$:

$$1_{\bigcup_{(1\leq j\leq j} A_j)}$$

$$= \sum_{1\leq j\leq n} 1_{A_j} + \sum_{r=2}^{n} (-1)^{r+1} \sum_{1\leq i_1<...<t_r\leq n} 1_{A_{i_1}...A_{i_r}} \tag{4.1}$$

and

(4.2)

$$Card(\bigcup_{1\leq j\leq n} A_j) = \sum_{1\leq j\leq n} Card(A_j) + \sum_{r=2}^{n}(-1)^{r+1} \sum_{1\leq i_1<...<t_r\leq n} Card(A_{i_1}...A_{ir}).$$

In the cited book, Formula (4.1) is proved and very similar techniques may be repeated to have Formula (4.2). The same techniques also lead the formula

$$\mathbb{P}(\bigcup_{1\leq j\leq n} A_j) = \sum_{1\leq j\leq n} \mathbb{P}(A_j) + \sum_{r=2}^{n}(-1)^{r+1} \sum_{1\leq i_1<...<t_r\leq n} \mathbb{P}(A_{i_1}...A_{ir}), \tag{4.3}$$

if A_1, ..., A_n are events.

These three formula are different versions of the Pointcarré's Formula, also called Inclusion-Exclusion Formula.

Bonferroni's Inequality.

Let A_1, ..., A_n be measurable subsets of Ω, $n \geq 2$. Define

$$\begin{aligned}
\alpha_0 &= \sum_{1\leq j\leq n} \mathbb{P}(A_j) \\
\alpha_1 &= \alpha_0 - \sum_{1\leq i_1<t_2\leq n} \mathbb{P}(A_{i_1}A_{i_2}) \\
\alpha_2 &= \alpha_1 + \sum_{1\leq i_1<...<t_3\leq n} \mathbb{P}(A_{i_1}...A_{i_3}) \\
\cdots &= \cdots \\
\alpha_r &= \alpha_{r-1} + (-1)^{r+1} \sum_{1\leq i_1<...<t_r\leq n} \mathbb{P}(A_{i_1}...A_{i_r}) \\
\cdots &= \cdots \\
\alpha_r &= \alpha_{r-1} + (-1)^{n+1}\mathbb{P}(A_1A_2A_3...A_n)
\end{aligned}$$

Let $p = n \text{ mod } 2$, that is $n = 2p+1+h$, $h \in \{0,1\}$. We have the Bonferroni's inequalities : if n is odd,

$$\alpha_{2k+1} \leq \mathbb{P}\left(\bigcup_{1\leq j\leq n} A_n\right) \leq \alpha_{2k}, \ k = 0, ..., p \ (BF1)$$

and if n is even,

$$\alpha_{2k+1} \leq \mathbb{P}\biggl(\bigcup_{1\leq j\leq n} A_j\biggr) \leq \alpha_{2k}, \ k=0,...,p-1. \ \ (BF2)$$

We may easily extend this formula to cardinalities in the following way. Suppose the A_i's are finite subsets of Ω and one of them at least is non-empty. Denote by M the cardinality of $\Omega_0 = \bigcup_{1\leq j\leq n} A_n$. Hence

$$\mathbb{P}(\Omega_0) \ni A \mapsto \mathbb{P}(A) = \frac{1}{M} Card(A),$$

is a probability measure and the Bonferroni inequalities hold. By multiplying the formulas by M, we get

$$\beta_{2k+1} \leq Card\biggl(\bigcup_{1\leq j\leq n} A_n\biggr) \leq \alpha_{2k}, \ p=0,1,...$$

where the sequence $(\beta_s)_{0\leq s\leq n}$ is defined sequentially by

$$\beta_0 = \sum_{1\leq j\leq n} Card(A_j)$$

and for $r>0$,

$$\beta_r = \alpha_{r-1} + (-1)^{r+1} \sum_{1\leq i_1<...<t_r\leq n} Card(A_{i_1}...A_{i_r}).$$

The extension has been made in the case where one of A_i's is non-empty. To finish, we remark that all inequalities hold as equalities of null terms if all the sets A_i's are empty.

Remark also that for $0<s\leq n$, we have

$$\alpha_s = \sum_{1\leq j\leq n} \mathbb{P}(A_j) + \sum_{r=2}^{s} (-1)^{r+1} \sum_{1\leq i_1<...<t_r\leq n} \mathbb{P}(A_{i_1}...A_{ir})$$

and

$$\beta_s = \sum_{1\leq j\leq n} Card(A_j) + \sum_{r=2}^{s} (-1)^{r+1} \sum_{1\leq i_1<...<t_r\leq n} Card(A_{i_1}...A_{ir}).$$

CHAPTER 2

Random Variables in $\mathbb{R}^d$, $d \geq 1$

This chapter will focus on the basic important results of Probability Theory concerning random vectors. Most of the properties exposed here and relative to discrete *real* random variables are already given and proved in the textbook Lo (2017a) of this series. The new features are the extensions of those results to vectors and the treatment of the whole thing as applications of the contents of Measure Theory and integration.

Three important results of Measure Theory and Integration, namely L^p spaces, Lebesgue-stieljes measures and Radon-Nokodym's Theorem are extensively used.

First, we will begin with specific results for *real* random variables.

1. A review of Important Results for Real Random variables

First, let us recall inequalities already established in Measure Theory. Next, we will introduce the new and important Jensen's one and give some of its applications.

Remarkable inequalities.

The first three inequalities are results of Measure Theory and Integration (See Chapter 10 in Lo (2017b)).

(1) Hölder Inequality. Let $p > 1$ and $q > 1$ be two conjugated positive rel numbers, that is, $1/p + 1/q = 1$ and let

$$X, Y : \ (\Omega, \mathcal{A}, \mathbb{P}) \ \mapsto \ \mathbb{R}$$

be two random variables $X \in L^p$ and $Y \in L^q$. Then XY is integrable and we have

$$|\mathbb{E}(XY)| \leq \|X\|_p \times \|Y\|_q ,$$

where for each $p \geq 1$, $\|X\|_p = (\inf |X|^p)^{1/p}$.

(2) Cauchy-Schwartz's Inequality. For $p = q = 2$, the Hölder inequality becomes the Cauchy-Schwartz one :

$$|\mathbb{E}(XY)| \leq \|X\|_2 \times \|Y\|_2 .$$

(3) Minskowski's Inequality. Let $p \geq 1$ (including $p = +\infty$). If X and Y are in L^p, then we have

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p .$$

(4) C_p Inequality. Let $p \in [1, +\infty[$. If X and Y are in L^p, then for $C_p = 2^{p-1}$, we have

$$\|X + Y\|_p^p \leq C_p(\|X\|_p^p + \|Y\|_p^p).$$

(5) Jensen's Inequality.

(a) Statement and proof of the inequality.

PROPOSITION 3. *(Jensen's inequality). Let ϕ be a convex function defined from a closed interval I of $\mathbb{R}$ to $\mathbb{R}$. Let X be a rrv with values in I such that $\mathbb{E}(X)$ is finite. Then $\mathbb{E}(X) \in I$ and*

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)).$$

Proof. Here, our proof mainly follows the lines of the one in Parthasarathy (2005).

Suppose that the hypotheses hold with $0 \in I$ and $\phi(0) = 0$. That $\mathbb{E}(X) \in I$ is obvious. First, let us assume that I is a compact interval, that is, $I = [a, b]$, with a and b finite and $a < b$. A convex function has left-hand and right-hand derivatives and then, is continuous (See Exercise 6 on Doc 03-09 of Chapter 4, page 191). Thus, ϕ is uniformly continuous on I. For $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|x - y| \leq \delta \Rightarrow |\phi(x) - \phi(y)| \leq \varepsilon. \tag{1.1}$$

We may cover I with a finite number of disjoint intervals E_j $(1 \leq j \leq k)$, of diameters not greater than δ. By using the Choice's Axiom, let us pick one x_j in each E_j. Let μ be a une probability measure on I. We have

$$\begin{aligned}\left|\int_I \phi(x)\ d\mu - \sum_{1\leq j\leq k} \phi(x_j)\right| &= \left|\sum_{1\leq j\leq k} \int_{E_j} \phi(x)\ d\mu - \sum_{1\leq j\leq k} \phi(x_j)\ \mu(E_j)\right| \quad (J02)\\ &= \left|\sum_{1\leq j\leq k} \int_{E_j} \phi(x)\ d\mu - \sum_{1\leq j\leq k} \int_{E_j} \phi(x_j)\ d\mu\right|\\ &\leq \sum_{1\leq j\leq k} \int_{E_j} |\phi(x) - \phi(x_j)|\ \ d\mu \leq \sum_{1\leq j\leq k} \varepsilon\mu(E_j) \leq \varepsilon.\end{aligned}$$

We also have

$$\begin{aligned}\left|\int_I x\ d\mu - \sum_{1\leq j\leq k} x_j\ \mu(E_j)\right| &= \left|\sum_{1\leq j\leq k} \int_{E_j} x\ d\mu - \sum_{1\leq j\leq k} x_j\ \mu(E_j)\right|\\ &= \left|\sum_{1\leq j\leq k} \int_{E_j} x\ d\mu - \sum_{1\leq j\leq k} \int_{E_j} x_j\right|\\ &\leq \sum_{1\leq j\leq k} \int_{E_j} |x - x_j|\, d\mu\\ &\leq \sum_{1\leq j\leq k} \delta\ \mu(E_j) \leq \delta.\end{aligned}$$

Then, by uniform continuity, we get

$$\left|\phi\left(\int_I x d\mu\right) - \phi\left(\sum_{1\leq j\leq k} x_j \mu(E_j)\right)\right| \leq \varepsilon. \tag{1.2}$$

By applying the convexity of ϕ, we have

$$\phi\left(\int_I x\ d\mu\right) \leq \varepsilon + \phi\left(\sum_{1\leq j\leq k} x_j\ \mu(E_j)\right) \leq \varepsilon + \sum_{1\leq j\leq k} \phi(x_j)\ \mu(E_j).$$

By applying Formula (J02) to last term of the right-hand side, we have

$$\phi‘ \left(\int_I x \ d\mu \right) \leq 2\varepsilon + \int_I \phi(x) \ d\mu,$$

for any $\varepsilon > 0$. This implies

$$\phi \left(\int_I x \ d\mu \right) \leq \int_I \phi(x) \ d\mu. \tag{1.3}$$

Now let I be arbitrary and μ be a probability measure on $\mathbb{R}$. Put, for each $n \geq 1$, $I_n = [a_n, b_n]$ with $(a_n, b_n) \to (-\infty, +\infty)$ as $n \to \infty$ and $\mu(I_n) > 0$ for large values of n. Let us consider the probability measures μ_n on I_n defined by

$$\mu_n(A) = \mu(A)/\mu(I_n), \ A \subset I_n.$$

Let us apply the inequality (1.3) to have

$$\phi \left(\int_{I_n} x \ d\mu_n \right) \leq \int_{I_n} \phi(x) \ d\mu_n.$$

But, by the Monotone Convergence Theorem, we get

$$\int x \ d\mu = \lim_{n \uparrow \infty} \int_{I_n} x \ d\mu = \lim_{n \uparrow \infty} \mu(I_n) \int_{I_n} x \ d\mu_n$$

and

$$\lim_{n \uparrow \infty} \mu(I_n) \int_{I_n} \phi(x) \ d\mu_n = \int \phi(x) \ d\mu.$$

By using the continuity of ϕ, and the the Monotone Convergence Theorem, and the fact that $\int x d\mu$ exists, we conclude by

$$\begin{aligned}
\phi\left(\int_I x d\mu\right) &= \lim_{n\to\infty} \phi(\int_{I_n} x\ d\mu) \\
&= \lim_{n\to\infty} \phi(\mu(I_n) \int_{I_n} x\ d\mu_n) \\
&= \lim_{n\to\infty} \phi(\mu(I_n) \int_{I_n} x\ d\mu_n + (1-\mu(I_n)) \times 0) \\
&\leq \lim_{n\to\infty} \mu(I_n)\ \phi\left(\int_{I_n} x\ d\mu_n\right) + (1-\mu(I_n))\ \phi(0)\ (By\ convexity) \\
&\leq \lim_{n\to\infty} \mu(I_n)\ \phi\left(\int_{I_n} x d\mu_n\right) \leq \lim_{n\to\infty} \mu(I_n)\ \phi\left(\int_{I_n} x d\mu_n\right) \\
&= \lim_{n\to\infty} \mu(I_n) \int_{I_n} \phi(x)\ d\mu_n\ (Since\ \phi(0)=0) \\
&= \int \phi(x)\ d\mu.\ (J03)
\end{aligned}$$

The proof above is valid for any probability measure on $\mathbb{R}$. Since X is integrable, X is *a.e.* finite and hence the support of $\mathbb{P}_X$ is a subset of $\mathbb{R}$. Hence, by applying (J3) to $\mathbb{P}_X$, we have the Jensen's inequality with the restrictions $0 \in I$, $\phi(0) = 0$. We remove them as follows :

If $0 \notin I$, we may enlarge I to contains 0 without any change of the inequality. If $\phi(0) \neq 0$, we may still apply the inequality to the convex function $\psi(x) = \phi(x) - \phi(0)$ which satisfies $\psi(0) = 0$ and get the result.

(b) Some applications of the Jensen's Inequality.

The following stunning results on L^p hold when the measure is a probability measure. They do not hold in general.

(b1) Ordering the spaces L^p.

Let $1 < p < q$, p finite but $q \in [1, +\infty]$. Let $X \in L^q$. Then $X \in L^p$ and

$$\|X\|_p \leq \|X\|_q\,.$$

For $q = +\infty$, the inequality holds for any finite measure.

Proof. We consider two cases.

Case q finite. Set $g_1(x) = x^p$, $g_2(x) = x^q$. Then the function $g_2 \circ g_1^{-1}(x) = x^{q/p}$ is convex on $(0, +\infty)$ since its second derivative is non-negative on $(0, +\infty)$. Let us set $X = g_1^{-1}(Y)$. In order to stay on $(0, +\infty)$, put $Z = |X|$ and take $Z = g_1^{-1}(Y)$, $Y \in (0, +\infty)$. The application of Jensen's Inequality leads to

$$g_2 \circ g_1^{-1}(\mathbb{E}(Y)) \leq \mathbb{E}(g_2 \circ g_1^{-1}(Y)).$$

Then we have

$$g_1^{-1}(\mathbb{E}(Y)) \leq g_2^{-1}(\mathbb{E}(g_2 \circ g_1^{-1}(Y)),$$

that is

$$g_1^{-1}(\mathbb{E}(g_1(Z)) \leq g_2^{-1}(\mathbb{E}(g_2(Z)).$$

This is exactly :

$$\|X\|_p \leq \|X\|_q \,.$$

Case $q = +\infty$. By definition, $X \in L^\infty$ means that the set

$$\{M \in [0, +\infty[, \; |X| \leq M, \; \mathbb{P} - a.e\}$$

*is not empty and the infimum of that set is $\|X\|_\infty$. But for any $0 \leq M < +\infty$ such that $|X| \leq M$, $\mathbb{P}$-$a.e.$ By taking the power and integrating, we get that

$$\left(\int |X|^p \; d\mathbb{P} \right)^{1/p} \leq M.$$

By taking the minimum of those values M, we get $\|X\|_p \leq \|X\|_\infty$.

Conclusion. If we have two real and finite numbers p and q such that $1 \leq p \leq q$, we have the following ordering for L^p spaces associated to a probability measure :

$$L^\infty \subset L^q \subset L^p \subset L^1.$$

(b2) Limit of the sequence of L^p-norm.

We have

$$\|X\|_p \nearrow \|X\|_\infty \text{ as } p \nearrow +\infty. \ (LN)$$

Proof. If $\|X\|_{p_0} = +\infty$ for some $p_0 \geq 1$, the results of Point (b2) above imply that $\|X\|_\infty = +\infty$ and $\|X\|_p = +\infty$ for all $p \geq p_0$ and the Formula (LN) holds.

Now suppose that $\|X\|_p < +\infty$ for all $p \geq 1$. By definition, $\|X\|_\infty = +\infty$ if the set

$$\{M \in [0, +\infty[,\ |X| \leq M,\ \mathbb{P} - a.e\}$$

is empty and is its infimum in the either case. In both cases, we have $\mathbb{P}(|X| > c) > 0$ for all $c < \|X\|_\infty$ (as a consequence of the infimum). We get the following inequalities, which first exploit the relation : $|X| \leq \|X\|_\infty$, *a.e.*. Taking the powers in that inequality and integrating yield, for $c < \|X\|_\infty$,

$$\|X\|_\infty \geq \biggl(\int |X|^p \, d\mathbb{P} \biggr)^{1/p} \geq \biggl(\int_{(|X| \geq c)} |X|^p \, d\mathbb{P} \biggr)^{1/p} \geq c \biggl(\mathbb{P}(|X| \geq c) \biggr)^{1/p}.$$

By letting first $p \to +\infty$, we get

$$c \leq \liminf_{p \to +\infty} \|X\|_p \leq \limsup_{p \to +\infty} \|X\|_p \leq \|X\|_\infty.$$

By finally letting $c \nearrow \|X\|_\infty$, we get the desired result.

2. Moments of Real Random Variables

(a) Definition of the moments.

The moments play a significant role in Probability Theory and in Statistical estimation. In the sequel, X and Y are two *rrv*'s, X_1, X_2, ... and Y_1, Y_2 are finite sequences of *rrv*'s, α_1, α_2, ... and β_1, β_2 are finite sequences of real numbers.

Let us define the following parameters, whenever the concerned expressions make sense.

(a1) Non centered moments of order $k \geq 1$:

$$m_k(X) = E\,|X|^k\,,$$

which always exists as the integral of a non-negative random variable.

(a2) Centered Moment of order $k \geq 1$.

$$\mu_k(X) = E\,|X - m_1|^k\,,$$

which is defined if If $m_1(X) = \mathbb{E}X$ exists and is finite.

(b) Focus on the centered moment of order 2.

(b1) Definition.

If $\mathbb{E}X$ exists and is finite, the centered moment of second order

$$\mu_2(X) = \mathbb{E}\biggl(X - \mathbb{E}(X)\biggr)^2,$$

is called the variance of X. Throughout the textbook, we will use the notations

$$\mu_2(X) =: \mathbb{V}ar(X) =: \sigma_X^2.$$

The number $\sigma_X = \mathbb{V}ar(X)^{1/2}$ is called the standard deviation of X.

(b2) Covariance between X and Y.

If $\mathbb{E}X$ and $\mathbb{E}Y$ exist and are finite, we may define the covariance between X and Y by

$$\mathbb{C}ov(X,Y) = \mathbb{E}\biggl((X - \mathbb{E}(X))((\mathbb{Y} - \mathbb{E}(Y))\biggr).$$

Warning. It is important to know that the expectation operator is used in the Measure Theory and Integration frame, that is, $\mathbb{E}h(X)$ exists and is finite if and only if $\mathbb{E}|h(X)|$ is finite. Later, when using Radon-Nikodym derivatives and replacing Lebesgue integrals by Riemann integrals, one should always remember this fact.

Warning. From now on, we implicitly assume the existence and the finiteness of the first moments of the concerned real random variables when using the variance or the covariance.

(b3) Expansions of the variance and covariance.

By expanding the formulas of the variance and the covariance and by using the linearity of the integral, we get, whenever the expressions make sense, that

$$\mathbb{V}ar(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

(In other words, the variance is the difference between the non centered moment of order 2 and the square of the expectation), and

$$\mathbb{C}ov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

(b4) Two basic inequalities based on the expectation and the variance.

The two first moments, when they exist, are basic tools in Statistical estimation. In turn, two famous inequalities are based on them. The first is the :

Markov's inequality : For any random variable X, we have for any $\lambda > 0$

$$\mathbb{P}(|X| > \lambda) \leq \frac{\mathbb{E}\,|X|}{\lambda}.$$

(See Exercise 6 in Doc 05-02 in Chapter 6 in Lo (2017b)). Next we have the :

Tchebychev's inequality : If $X - \mathbb{E}(X)$ is defined *a.e.*, then for any $\lambda > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| > \lambda) \leq \frac{\mathbb{V}ar(X)}{\lambda^2}.$$

This inequality is derived by applying the Markov's inequality to $|X - \mathbb{E}(X)|$ and by remarking that $(|X - \mathbb{E}(X)| > \lambda) = ((X - \mathbb{E}(X))^2 > \lambda^2)$, for any $\lambda > 0$.

(c) Remarkable properties on variances and covariances.

Whenever the expressions make sense, we have the following properties.

(P1) $\mathbb{V}ar(X) = 0$ if and only if $X = (X)$ $a.s.$

(P2) For all $\lambda > 0$, $\mathbb{V}ar(\lambda X) = \lambda^2 \mathbb{V}ar(X)$

(P3) We have

$$\mathbb{V}ar(\sum_{1 \leq i \leq k} \alpha_i X_i) = \sum_{1 \leq i \leq k} \mathbb{V}ar(X_i)\alpha_i^2 + 2\sum_{i<j} \mathbb{C}ov(X_i, Y_j)\alpha_i \alpha_j.$$

(P4) We also have

$$\mathbb{C}ov\Big(\sum_{1 \leq i \leq k} \alpha_i X_i), \sum_{1 \leq i \leq \ell} \beta_i Y_i)) = \sum_{1 \leq i \leq k} \sum_{1 \leq j \leq \ell} \mathbb{C}ov(X_i, Y_j)\alpha_i \beta_j.$$

(P5)] If X and Y are independent, then $\mathbb{C}ov(X, Y) = 0$.

(P6) Si $X_1, ..., X_k$ are pairwise independent, then

$$\mathbb{V}ar(\sum_{1 \leq i \leq k} \alpha_i X_i) = \sum_{1 \leq i \leq k} \mathbb{V}ar(X_i)\alpha_i^2.$$

(P7) If none of σ_X and σ_Y is null, then the coefficient

$$\rho_{XY} = \frac{\mathbb{C}ov(X, Y)}{\sigma_X \sigma_Y},$$

is called the linear correlation coefficient between X and Y and satisfies

$$|\rho_{XY}| \leq 1.$$

Proofs or comments. Most of these formulas are proved in the textbook Lo (2017a) of this series. Nevertheless we are going to make comments of the proofs at the light of Measure Theory and Integration and prove some of them.

(P1) We suppose that $\mathbb{E}(X)$ exists and is finite. We have $Y = (X - \mathbb{E}(X))^2 \geq 0$ and $\mathbb{V}ar(X) = \mathbb{E}Y$. Hence, $\mathbb{V}ar(X) = 0$ if and only if

$Y = 0$ *a.e.* $\square$

(P2) This is a direct application of the linearity of the integral as recalled in Theorem 1. $\square$

(P3) This formula uses (P2) and the following the identity :

$$\left(\sum_{1\leq i\leq k} a_i\right)^2 = \sum_{1\leq i\leq k} a_i^2 + 2\sum_{i<j} a_i a_j,$$

where a_i, $1 \leq i \leq k$, are real and finite numbers. Developing the variance and applying this alongside the linearity of the mathematical expectation together lead to the result. $\square$

(P4) This formula uses the following identity

$$\left(\sum_{1\leq i\leq k} a_i\right)\left(\sum_{1\leq i\leq \ell} bi\right) = \sum_{1\leq i\leq k}\sum_{1\leq j\leq \ell} a_i b_j,$$

where the a_i, $1 \leq i \leq k$, and the b_i, $1 \leq i \leq \ell$, are real and finite numbers. By developing the covariance and applying this alongside the linearity of the mathematical expectation lead to the result. $\square$

(P5) Suppose that X and Y are independent. Since X and Y are *real* random variables, Theorem 3.3 implies that : $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ by. Hence, by Point (b3) above, we get

$$\mathbb{C}ov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) = 0.\square$$

(P6) If the X_i's are pairwise independent, the covariances in the formula in (P3) vanish and we have the desired result. $\square$

(P7) By applying the Cauchy-Schwartz inequality to $X - \mathbb{E}(X)$ and to $Y - \mathbb{E}(Y)$, that is the Hőlder inequality for $p = q = 2$, we get

$$|\mathbb{C}ov(X,Y)| \leq \sigma_X \sigma_Y.$$

If none of σ_X and σ_Y is zero, we get $|\rho_{XY}| \leq 1$. $\square$

3. Cumulative distribution functions

An important question in Probability Theory is to have parameters or functions which characterize probability laws. In Mathematical Statistics, these characteristics may be used in statistical tests. For example, if X is a real value random variable having finite moments of all orders, that is : for all $k \geq 1$, $\mathbb{E}|X|^k < +\infty$. Does the sequence $(\mathbb{E}|X|^k)_{k\geq 1}$ characterize the probability law $\mathbb{P}_X$? This problem, named after the moment problem, will be addressed in a coming book.

The first determining function comes from the Lebesgue-Stieljes measure studied in Chapter 11 in Lo (2017b). We will use the results of that chapter without any further recall.

(a) The cumulative distribution function of a real-random variable.

Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto \mathbb{R}$ be a random real-valued random variable. Its probability law $\mathbb{P}_X$ satisfies :

$$\forall x \in \mathbb{R},\ \mathbb{P}_X(]-\infty, x]) < +\infty.$$

Hence, the function

$$\mathbb{R} \ni x \mapsto F_X(x) = \mathbb{P}_X(]-\infty, x]),$$

is a distribution function and $\mathbb{P}_X$ is the unique probability-measure such that

$$\forall (a,b) \in \mathbb{R}^2 \text{ such that } a \leq b,\ \mathbb{P}_X(]a,b]) = F_X(b) - F_X(a).$$

Before we go further, let us give a more convenient form of F_X by writing for any $x \in \mathbb{R}$,

$$\begin{aligned} F_X(x) &= \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X^{-1}(]-\infty, x])) \\ &= \mathbb{P}(\{\omega \in \Omega,\ X(\omega) \leq x\}) = \mathbb{P}(X \leq x). \end{aligned}$$

Now, we may summarize the results of the Lebesgue-Stieljes measure in the context of probability Theory.

Definition. For any real-valued random variable $X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto \mathbb{R}$, the function defined by

$$\mathbb{R} \ni \mapsto F_X(x) = \mathbb{P}(X \leq x),$$

is called the cumulative distribution (cdf) function of X.

It has the two sets of important properties.

(b) Properties of F_X.

(1) It assigns non-negative lengths to intervals, that is

$$\forall (a, b) \in \mathbb{R}^2 \text{ such that } a \leq b, \ \Delta_{a,b}F = F_X(b) - F_X(a) \geq 0.$$

(2) It is right-continuous at any point $t \in \mathbb{R}$.

(3) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$ and $F(+\infty) = \lim_{x \to +\infty} F(x) = 1$.

Warning. Point (1) means, in the case of one-dimension, that F_X is non-decreasing. *So, it happens that the two notions of non-negativity of lengths by F and non-drecreasingness of F coincide in dimension one.* However, we will see that this is not the case in higher dimensions, and that non-decreasingness is not enough to have a *cdf*.

(c) Characterization.

The *cdf* is a characteristic function of the probability law of a random variable with values in $\mathbb{R}$ from the following fact, as seen Chapter 11 in Lo (2017b) of this series :

*There exists a one-to-one correspondence between the class of Probability Lebesgue-Stieljes measures $\mathbb{P}_F$ on $\mathbb{R}$ and the class of **cfd**'s $F_{\mathbb{P}}$ on $\mathbb{R}$ according the relations*

$$\left(\forall x \in \mathbb{R}, \ F_{\mathbb{P}}(x) = \mathbb{P}(]-\infty, x])\right), \ \left(\forall (a,b) \in (\mathbb{R}), \ a \leq b, \ \mathbb{P}_F(]a,b]) = \Delta_{a,b}F\right)$$

The *cdf* is a characteristic function of the probability law of random variables. The means that two random real variable X and Y with the same distribution function have the same probability law.

(d) How Can we Define a Random Variable Associated to a *Cdf*.

Let us transform the properties in Point (b) into a definition.

(d1) Definition. A function $F : \mathbb{R} \rightarrow [0,1]$ is *cdf* if and only if conditions (1), (2) and (3) of Point (b) above are fulfilled.

Once we know that F is (*cdf*), can you produce a random variable $X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto \mathbb{R}$ such that $F_X = X$? meaning : can we construct a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding a random variable such that for all $x \in \mathbb{R}$, $F(x) = \mathbb{P}(X \leq x)$?

This is the simplest form the Kolmogorov construction. A solution is the following.

(d2) A Simple form of Kolmogorov construction.

Since F is a *cdf*, we may define the Lebesgue-Stieljes measure $\mathbb{P}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R})$ defined by

$$\mathbb{P}(]y,x]) = \Delta_{y,x} F = F(x) - F(y), \quad -\infty < y < x < +\infty. \ (LS11)$$

By Conditions (3) in the definition of a *cdf* in Point (b) above, $\mathbb{P}$ is normed and hence, is a probability measure. By letting $y \downarrow -\infty$ in (LS1), we get

$$\forall x \in \mathbb{R}), \ F(x) = \mathbb{P}(]-\infty, x]). \ \ (LS12)$$

Now take $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ and let $X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto \mathbb{R}$ be the identity function

$$\forall \omega \in \Omega, \ X(\omega) = \omega.$$

It is clear that X is a random variable and we have for $x \in \mathbb{R}$, we have

$$\begin{aligned} F_X(x) &= \mathbb{P}(\{\omega \in \mathbb{R}, \ X(\omega) \leq x\}) \\ &= \mathbb{P}(\{\omega \in \mathbb{R}, \ \omega \leq x\}) \\ &= \mathbb{P}(]-\infty, x) = F(x), \end{aligned}$$

where we used (LS12). We conclude the X admits F as a *cdf*.

Warning. This construction may be very abstract at a first reading. If you feel confused with it, we may skip it and wait a further reading to catch it.

(e) Decomposition of *cdf* in discrete and continuous parts.

Let F be a *cdf* on $\mathbb{R}$ and let us denote by $\mathbb{P}_F$ the associated Lebesgue-measure. We already know from Measure Theory that : $x \in \mathbb{R}$ is a continuity point of F if and only if

$$\mathbb{P}_F(\{x\}) = F(x) - F(x-0) = 0, \quad (CC)$$

where for each $x \in \mathbb{R}$

$$F(x+) \equiv F(x+0) = \lim_{h \searrow 0} F(x+h)$$

and

$$F(x-) \equiv F(x-0) = \lim_{h \searrow 0} F(x+h)$$

are the right-limit hand and the the left-limit hand of $F(\circ)$ at x, whenever they exist. In the present case, they do because of the monotonicity of F.

So, a *cdf* is continuous if and only if Formula (CC) holds for each $x \in \mathbb{R}$. In the general case, we are able to decompose the *cdf* into two non-negative distributions functions F_c and F_d, where F_c is continuous and F_d is discrete in a sense we will define. As a reminder, a distribution function (*df*) on $\mathbb{R}$ is a function satisfying only Conditions (1) and (2) in Point (b) above.

Let us define a discrete *df* F_d on $\mathbb{R}$ as a function such that there exists a countable number of distinct real numbers $\mathcal{D} = \{x_j, \ j \in J\}$, $J \subset \mathbb{N}$ and a family of finite and positive real numbers $(p_j)_{j \in J}$ such that

$$\forall x \in \mathbb{R}, F_d(x) = \sum_{x_j \leq x, j \in J} p_j < +\infty. \ (DDF1)$$

Let ν be the discrete measure $\mathcal{D}$ defined by

$$\forall (y,x) \in \mathbb{R}^2 \text{such that } y \leq x, \nu(]y,x]) = \sum_{x_j \in]y,x], j \in J} p_j < +\infty. \ (DDF2)$$

By combining (DDF1) and (DDF2), we see thatF_d a discrete *df* is a *df* of a counting measure which is finite on bounded above intervals. It follows that for each $j \in J$,

$$\nu(\{x_j\}) = F_d(x_j) - F_d(x_j - 0) = p_j > 0.$$

This implies that a discrete *df* is never continuous at all points. We still may call $\mathcal{D}$ the support of F_d by extension of the support of ν.

We know that F, as a non-decreasing function, has at most a countable number of discontinuity points. Let us denote the set of those discontinuity points by $\mathcal{D} = \{x_j,\ j \in J\}$, $J \subset \mathbb{N}$ and put $p_j = F(x_j) - F(x_j - 0) > 0$. Going Back to Measure Theory (see Solution of Exercise 1, Doc 03-06, Chapter 4 in Lo (2017b) of this series), we have that

$$\forall (y,x) \in \mathbb{R}^2 \text{such that } y \leq x, \sum_{x_j \in]y,x], j \in J} F(x_j) - F(x_j - 0) \leq F(x) - F(y).$$

By letting $y \downarrow -\infty$, we have

$$\forall x \in \mathbb{R}, F_d(x) = \sum_{x_j \leq x, j \in J} p_j \leq F(x) < +\infty.$$

Besides, the set discontinuity points of F_d is $\mathcal{D}$ since discontinuity points x of F_d must satisfy $\nu(\{x\}) > 0$.

Next, let us define $F_c = F - F_d$. It is clear that F_c is right-continuous and non-negative. Let us prove that F_c is continuous. By the developments above, F_c is continuous outside $\mathcal{D}$ and for each $j \in J$, we have

$$\begin{aligned} F_c(x_j) - F_c(x_j - 0) &= \Big(F(x_j) - F(x_j - 0)\Big) - \Big(F_d(x_j) - F_d(x_j - 0)\Big) \\ &= p_j - p_j = 0. \end{aligned}$$

It remains to show F_c is a *df* by establishing that : it assigns to intervals non-negative lengths. For each $x \in \mathbb{R}$, $h > 0$, we have

$$\Delta_{x,x+h}F_c = \left(F(x+h) - F(x)\right) - \left(F_d(x+h) - F_d(x)\right).$$

But, by definition $F_d(x+h) - F_d(x)$ is the sum of the jumps of F at discontinuity points in $]x,\ x+h]$. We already know (otherwise, get help from a simple drawing) that this sum of jumps is less than $F(x+h+0) - F(x)$ which is $F(x+h) - F(x)$ by right-continuity of F. Hence for all $x \in \mathbb{R}$, for all $h > 0$, $\Delta_{x,x+h}F_c \geq 0$. In total, F is a *df*.

We get the desired decomposition : $F = F_c + F_d$. Suppose that we have another alike decomposition $F = F_c^* + F_d^*$. Since the functions are bounded, we get $F_c - F_c^* = F_d^* - F_d$. Let us denote by $\mathcal{D}$ and $\mathcal{D}^*$ and by p_x and p_x^* the supports and the discontinuity jumps (at x) of F_d and F_d^* respectively.

If the supports are not equal, thus for $x \in \mathcal{D}\Delta\mathcal{D}^*$, $F_d^* - F_d$ is discontinuous at x.

If $\mathcal{D} = \mathcal{D}^*$ and $p_x \neq p_x^*$, the discontinuity jump of $F_d^* - F_d$ at x is $p_x^* - p_x > 0$.

Since none of the two last conclusions is acceptable, we get that the equation $F_c - F_c^* = F_d^* - F_d$ implies that F_d^* and F_d have the same support and the same discontinuity jumps, and hence are equal and then so are F_c^* and F_c.

We get the following important result.

Proposition. A *cdf* F is decomposable into the addition of two non-negative distribution functions (*df*) F_c and F_d, where F_c is continuous and F_d is discrete and $F_c(-\infty) = F_d(-\infty) = 0$. The decomposition is unique.

Warning. Such a result is still true for a *df* but the condition $F_c(-\infty) = F_d(-\infty) = 0$ is not necessarily true.

NB. We did not yet treat the probability density existence and its use for real random variables. This will be done in the next section which is concerned with random vectors.

4. Random variables on $\overline{\mathbb{R}}^d$ or Random Vectors

(a) Introduction.

Random vectors are generalizations of real random variables. A random vector of dimension $d \geq 1$ is a random variable

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto (\overline{\mathbb{R}}^d, \mathcal{B}_\infty(\overline{\mathbb{R}}^d).$$

with values in $\overline{\mathbb{R}}^d$.

Important Remarks. In general, it is possible to have $\left(\overline{\mathbb{R}}^d, \mathcal{B}_\infty\left(\overline{\mathbb{R}}^d\right)\right)$ as the set of values of random vectors, especially when we are concerned with general probability laws. But, the most common tools which are used for the study of random vectors such as the cumulative random vectors, the characteristic functions, the absolute probability density function are used for finite component random vectors with values in $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d\right)$.

Throughout this section, we use a random vector with d components as follows. Let

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_{d-1} \\ X_d \end{bmatrix}$$

From Measure Theory, we know that X is a random variable if and only if each X_i, $1 \leq i \leq d$, is a real random variable.

If $d = 1$, the random vector becomes a real random variable, abbreviated (rrv).

Notation. To save space, we will rather use the transpose operator and write $X^t = (X_1, X_2, \cdots, X_d)$ or $X = (X_1, X_2, \cdots, X_d)^t$. Let $(Y_1, Y_2, \cdots, X_d)^t$ another be d-random vector and two other random vectors $(Z_1, Z_2, \cdots, Z_r)^t$ and $(T_1, T_2, \cdots, T_s)^t$ of dimensions $r \geq 1$ and $s \geq 1$, all of them being defined on $(\Omega, \mathcal{A}, \mathbb{P})$.

Matrix Notation. To prepare computations on the matrices, let us denote any real matrix A of $r \geq 1$ lines and $s \geq 1$ columns in the form $A = (a_{ij})_{1 \leq i \leq r,\ 1 \leq j \leq s}$, where the lowercase letter a is used to denote the elements of the matrix whose name is the uppercase letter A. As well, we will use the notation $(A)_{ij} = a_{ij},\ 1 \leq i \leq r,\ 1 \leq j \leq s$.

A matrix of r lines and s columns is called a $(r \times s)$-matrix, a square matrix with r lines and r columns is a r-matrix and a vector of r components is a d-vector.

The s columns of a matrix A are elements of $\mathbb{R}^r$ and are denoted by A^1, A^2, ..., A^s. The r lines of the matrix A are $(1 \times s)$-matrices denoted A_1, ...,A_r, that is A_1^t, ...,A_r^t belong to $\mathbb{R}^s$.

So, for $1 \leq j \leq s$, $A^j = (a_{1j}, a_{2j}, \cdots, a_{rj})^t$ and for $1 \leq i \leq r$, $A_i = (a_{i1}, a_{i2}, \cdots, a_{is})$.

We also have $A = [A^1, A^2, ..., A^d]$ and $A^t = [A_1^t, A_2^t, ..., A_p^t]$.

Introduce the scalar product in $\mathbb{R}^s$ in the following way. Let x and y be two elements $\mathbb{R}^s$ with $x^t = (x_1, \cdots, x_s)$ and $y^t = (y_1, \cdots, y_s)$.

We define the scalar product $< x, y >$ of x and y as the matrix product of the $(1 \times s)$-matrix x^t by the $(s \times 1)$-matrix y which results in the real number

$$< x, y > = \ x^t y = \sum_{i=1}^{s} x_i y_i.$$

With the above notation, the matrix operations may be written in the following way.

If (1) *Sum of matrices of same dimensions.* If $A = (a_{ij})_{1 \leq i \leq r,\ 1 \leq j \leq s}$ and $B = (b_{ij})_{1 \leq i \leq r,\ 1 \leq j \leq s}$ are two $(r \times s)$-matrix, then $A + B$ is the $(r \times s)$-matrix : $A+B = (a_{ij}+b_{ij})_{1 \leq i \leq r,\ 1 \leq j \leq s}$, that is : $(A+B)_{ij} = (A)_{ij}+(B)_{ij}$ for $1 \leq i \leq r,\ 1 \leq j \leq s$.

If (2) *Multiplication by a scalar.* If λ is a real number and if $A = (a_{ij})_{1 \leq i \leq r,\ 1 \leq j \leq s}$, then λA is the $(r \times s)$-matrix : $\lambda A = (\lambda a_{ij})_{1 \leq i \leq r,\ 1 \leq j \leq s}$, that is :$(\lambda A)_{ij} = \lambda (A)_{ij}$ for $1 \leq i \leq r,\ 1 \leq j \leq s$.

If (3) *Product of Matrices.* If $A = (a_{ij})_{1\leq i\leq r,\ 1\leq j\leq s}$ and $B = (b_{ij})_{1\leq i\leq s,\ 1\leq j\leq q}$ such that the number of columns of A (the first matrix) is equal to the number of lines of B (the second of the second matrix), the product matrix AB is a (r,q)-matrix defined by

$$AB = (A_i B^j)_{1\leq i\leq r,\ 1\leq j\leq q},\ (PM1)$$

that is, for $1 \leq i \leq r,\ 1 \leq j \leq q$,

$$(AB)_{ij} = (A_i B^j) = \sum_{k=1}^{d} a_{ik} b_{kj}.\ (MP2)$$

(b) Variance-covariance and Covariance Matrices.

(b1) Definition of Variance-covariance and Covariance Matrices.

We suppose that the components of our random vectors have finite second moments. We may define

(i) **the mathematical expectation vector** $\mathbb{E}(X)$ **of** $X \in \mathbb{R}^d$ by the vector

$$\mathbb{E}(X)^t = (\mathbb{E}(X_1), \mathbb{E}(X_2), \cdots, \mathbb{E}(X_d)),$$

(ii) **the covariance matrix** $\mathbb{C}ov(X,Y)$ **between** $X \in \mathbb{R}^d$ **and** $Z \in \mathbb{R}^r$ by the $(d \times r)$-matrix

$$\mathbb{C}ov(X,Y) = \Sigma_{XY} = \mathbb{E}\biggl((X - \mathbb{E}(X))(Z - \mathbb{E}(Z))^t\biggr),$$

in an other notation

$$\mathbb{C}ov(X,Y) = \Sigma_{XY} = \biggl(\mathbb{E}(X_i - \mathbb{E}(X_i))(Z_j - \mathbb{E}(Z_j))\biggr)_{1\leq i\leq d,\ 1\leq j\leq r},$$

(iii) **the variance-covariance matrix** $\mathbb{V}ar(Y)$ **of** $X \in \mathbb{R}^d$ by the $(d \times d)$-matrix

$$\begin{aligned}\mathbb{V}ar(X) &= \Sigma_X = \mathbb{E}\biggl((X-\mathbb{E}(X))(X-\mathbb{E}(X))^t\biggr)\\ &= \biggl(\mathbb{E}(X_i-\mathbb{E}(X_i))\mathbb{E}(X_j-\mathbb{E}(X_j))\biggr)_{1\leq j\leq d,\ 1\leq j\leq d}. \square\end{aligned}$$

Let us explain more the second definition. The matrix

$$(X-\mathbb{E}(X))(Z-\mathbb{E}(Z))^t$$

is the product of the $(d\times 1)$-matrix $X-\mathbb{E}(X)$, with

$$(X-\mathbb{E}(X))^t = (X_1-\mathbb{E}(X_1), X_2-\mathbb{E}(X_2), \cdots, X_d-\mathbb{E}(X_d)),$$

by the $(1\times r)$-matrix with

$$(Z-\mathbb{E}(Z))^t = (Z_1-\mathbb{E}(Z_1), Z_2-\mathbb{E}(Z_2), \cdots, X_r-\mathbb{E}(Z_r)).$$

The (ij)-element of the product matrix, for $1\leq i\leq d,\ 1\leq j\leq r$, is

$$(X_i-\mathbb{E}(X_i))(Z_j-\mathbb{E}(Z_j)).$$

By taking the mathematical expectations of those elements, we get the matrix of covariances

$$\mathbb{C}ov(X,Y) = \biggl(\mathbb{C}ov(X_i,Y_j)\biggr)_{1\leq i\leq d,\ 1\leq j\leq r}$$

For $X=Z$ (and then $d=r$), we have

$$\mathbb{V}ar(X) \equiv \mathbb{C}ov(X,X) = \biggl(\mathbb{C}ov(X_i,X_j)\biggr)_{1\leq i\leq d,\ 1\leq j\leq d}.$$

We have the following properties.

(b2) Properties.

Before we state the properties, it is useful to recall that linear mappings from $\mathbb{R}^d$ to $\mathbb{R}^r$ are of the form

$$\mathbb{R}^d \ni X \mapsto AX \in \mathbb{R}^r,$$

where A is a $(r \times d)$-matrix of real scalars. Such mappings are continuous (uniformly continuous, actually) and then measurable with respect to the usual σ-algebras on $\mathbb{R}^d$ and $\mathbb{R}^p$.

Here are the main properties of the defined parameters.

(P1) For any $\lambda \in \mathbb{R}$,

$$\mathbb{E}(\lambda X) = \lambda \mathbb{E}(X).$$

(P2) For two random vectors X and Y of the same dimension d, $AX \in \mathbb{R}^p$ and

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

(P3) For any $(p \times d)$-matrix A and any d-random vector X,

$$\mathbb{E}(AX) = A\mathbb{E}(X) \in \mathbb{R}^p.$$

(P4) For any d-random vector X and any s-random vector Z,

$$\mathbb{C}ov(X,Z) = \mathbb{C}ov(Z,X)^t.$$

(P5) For any $(p \times d)$-matrix A, any $(q \times s)$-matrix B, any d-random vector X and any s-random vector Z,

$$\mathbb{C}ov(AX, BZ) = A\mathbb{C}ov(X,Z)B^t,$$

which is a (p,q)-matrix.

Proofs.

We are just going to give the proof of (P3) and (P4) to show how work the computations here.

Proof of (P3). The i-th element of the column vector AX of $\mathbb{R}^p$, for $1 \leq i \leq p$, is

$$(AX)_i = A_i X = \sum_{1 \leq j \leq d} a_{ij} X_j$$

and its real mathematical expectation, is

$$\mathbb{E}(AX)_i = \sum_{1 \leq j \leq d} a_{ij}\mathbb{E}(X_j).$$

But the right-hand member is, for $1 \leq i \leq p$, the i-th element of the column vector $A\mathbb{E}(X)$. Since $\mathbb{E}(AX)$ and $A\mathbb{E}(X)$ have the same components, we get

$$\mathbb{E}(AX) = A\mathbb{E}(X). \square$$

Proof of (P5). We have

$$\mathbb{C}ov(AX, BZ) = \mathbb{P}\left((AX - \mathbb{E}(AX))(BZ - \mathbb{E}(BZ))^t\right). \ (COV1)$$

By (P5), we have

$$\begin{aligned}(AX - \mathbb{E}(AX))(BX - \mathbb{E}(BZ)) &= A(X - \mathbb{E}(X))(B(Z - \mathbb{E}(Z)))^t \\ &= A\Bigg((X - \mathbb{E}(X))(Z - \mathbb{E}(Z))^t\Bigg)B^t.\end{aligned}$$

Let us denote $C = (X - \mathbb{E}(X))(Z - \mathbb{E}(Z))^t$. We already know that

$$c_{ij} = \Bigg((X_i - \mathbb{E}(X_i))(Z_j - \mathbb{E}(Z_j))\Bigg), \ (i, j) \in \{1, ..., d\}^2.$$

Let us fix $(i, j) \in \{1, ..., p\} \times \{1, ..., q\}$. The ij-element of the (p, q)-matrix ACB^t is

$$(ACB^t)_{ij} = (AC)_i(B^t)^j.$$

But the elements of i-th line of AC are $\{A_iC^1, A_2C^2, ..., A_iC^p\}$ and the column $(B^t)^j$ contains the elements of the j-th line of B, that is b_{j1}, b_{j2}, ...,b_{js}. We get

$$
\begin{aligned}
(ACB^t)_{ij} &= \sum_{1 \leq k\ leqs} \biggl((AC)_i\biggr)_k \biggl((B^t)^j\biggr)_k \\
&= \sum_{1 \leq k\ \leq s} \biggl(A_i C^k\biggr) b_{jk} \\
&= \sum_{1 \leq k\ \leq s} \sum_{1 \leq p\ \leq p} a_{ih}(X_h - \mathbb{E}(X_h))(Z_k - \mathbb{E}(Z_k)) b_{jk}. \ (COV2)
\end{aligned}
$$

Hence, by applying Formula (COV1), the ij-element of $\mathbb{C}ov(AX, BZ)$ is

$$
\mathbb{E}\biggl((ACB^t)_{ij}\biggr) = \sum_{1 \leq k\ \leq s} \sum_{1 \leq p\ \leq p} a_{ih} \mathbb{C}ov(X_h, Z_k) b_{jk}. \ (COV3)
$$

Actually we have proved that for any $(p \times d)$-matrix A, for any $(d \times s)$-matrix and for any $(q \times)$-matrix, the ij-element of ACB^t is given by

$$
\sum_{1 \leq k\ \leq s} \sum_{1 \leq p\ \leq p} a_{ih} c_{hk} b_{jk}. \ (ACBT)
$$

When applying this to Formula (COV2), we surely have that

$$
\mathbb{C}ov(AX, BZ) = A \mathbb{C}ov(X, Z) B^t.
$$

(b3) Focus on the Variance-covariance matrix.

(P6) Let A be a $(p \times d)$-matrix and X be a d-random vector. We have

$$
\Sigma_{AX} = A \Sigma_X A^t.
$$

(P7) The Variance-covariance matrix Σ_X of $X \in \mathbb{R}^d$ is a positive matrix as a quadratic form, that is :

$$
\forall u \in \mathbb{R}^d, \ u \Sigma_X u^t \geq 0.
$$

If Σ_X is invertible, then it is definite-positive that is

$$
\forall u \in \mathbb{R}^d \setminus \{0\}^d, \ u \Sigma_X u^t > 0.
$$

(P8) Σ_X is symmetrical and there exists an orthogonal d-matrix T such that $T\Sigma_X T^t$ is a diagonal matrix

$$T\sigma_X T^t = diag(\delta_1, \delta_2, ..., \delta_d),$$

with non-negative eigen-values $\delta_j \geq 0$, $1 \leq j \leq d$. Besides we have the following facts :

(P8a) The columns of T are eigen-vectors of Σ_X respectively associated the eigen-values $\delta_j \geq 0$, $1 \leq j \leq d$ respectively.

(P8b) The columns, as well as the lines, of T form an orthonormal basis of $\mathbb{R}^d$.

(P8c) $T^{-1} = T$.

(P8d) The number of positive eigen-values is the rank of Σ_X and Σ_X is invertible if and only all the eigen-values are positive.

(P8e) The determinant of Σ_X is given by

$$|\Sigma_X| \equiv det(\Sigma_X) = \prod_{j=1}^{d} \delta_j.$$

Proofs. Property (P6) is a consequence of (P5) for $A = B$. Formula (P8) and its elements are simple reminders of Linear algebra and diagonalization of symmetrical matrices. The needed reminders are gathered in Subsection 2 in Section 10. The only point to show is (P7). And by (P5), we have for any $u \in \mathbb{R}^d$

$$\begin{aligned} u\Sigma_X u^t &= u\mathbb{E}\biggl((X - \mathbb{E}(X))(X - \mathbb{E}(X))^t\biggr)u^t \\ &= \mathbb{E}\biggl(u(X - \mathbb{E}(X))(X - \mathbb{E}(X))^t u^t\biggr). \end{aligned}$$

But $u(X - \mathbb{E}(X))(X - \mathbb{E}(X))^t u^t = \biggl(u(X - \mathbb{E}(X)\biggr)\biggl(u(X - \mathbb{E}(X))\biggr)^t$. Since $u(X - \mathbb{E}(X))$ is d-vector, we have

$$\left(u(X-\mathbb{E}(X)\right)\left(u(X-\mathbb{E}(X))\right)^t = \|u(X-\mathbb{E}(X)\|^2 \geq 0.$$

Hence $u\sigma_X u^t \geq 0$. ■.

(c) Cumulative Distribution Functions.

The presentation of *cdf*'s on $\mathbb{R}^d$ follows that lines we already used for *cdf*'s on $\mathbb{R}$. But the notations are heavier.

Let us recall the notion of volume we already introduced in Chapter 11 in Lo (2017b).

(c1) Notion of Volume of cuboids by F.

Simple case. Let us begin by the case $d = 2$. Consider a rectangle

$$]a,b] =]a_1,b_1] \times]a_2,b_2] = \prod_{i=1}^{2}]a_i,b_i],$$

for $a = (a_1,a_2) \leq b = (b_1,b_2)$ meaning $a_i \leq b_i$, $1 \leq i \leq 2$. The volume of $]a,b]$ by F is denoted

$$\Delta F(a,b) = F(b_1,b_2) - F(b_1,a_2) - F(a_1,b_2) + F(a_1,a_2).$$

Remark. In the sequel we will use both notations $\Delta_{a,b}F$ and $\Delta F(a,b)$ equivalently. The function $\Delta F(a,b)$ is obtained according to the following rule :

Rule of forming $\Delta F(a,b)$. First consider $F(b_1,b_2)$ the value of the distribution function at the right endpoint $b = (b_1,b_2)$ of the interval $]a,b]$. Next proceed to the replacements of each b_i by a_i by replacing exactly one of them, next two of them etc., and add each value of F at the formed points, with a sign *plus* $(+)$ if the number of replacements is even and with a sign *minus* $(-)$ if the number of replacements is odd.

We also may use a compact formula. Let $\varepsilon = (\varepsilon_1, \varepsilon_2) \in \{0,1\}^2$. We have four elements in $\{0,1\}^2$: $(0,0)$, $(1,0)$, $(0,1)$, $(1,1)$. Consider a particular $\varepsilon_i = 0$ or 1, we have

$$b_i + \varepsilon_i(a_i - b_i) = \begin{cases} b_i & if \quad \varepsilon_i = 0 \\ a_i & if \quad \varepsilon_i = 1 \end{cases}.$$

So, in

$$F(b_1 + \varepsilon_1(a_1 - b_1), b_1 + \varepsilon_2(a_2 - b_2)),$$

the number of replacements of the b_i's by the corresponding a_i is the number of the coordinates of $\varepsilon = (\varepsilon_1, \varepsilon_2)$ which are equal to the unity 1. Clearly, the number of replacements is

$$s(\varepsilon) = \varepsilon_1 + \varepsilon_2 = \sum_{i=1}^{2} \varepsilon_i$$

We way rephrase the Rule of forming $\Delta F(a, b)$ into this formula

$$\Delta F(a, b) = \sum_{\varepsilon=(\varepsilon_1,\varepsilon_2)\in\{0,1\}} (-1)^{s(\varepsilon)} F(b_1 + \varepsilon_1(a_1 - b_1), b_1 + \varepsilon_2(a_2 - b_2)).$$

We may be more compact by defining the product of vectors as the vector of the products of coordinates as

$$(x, y) * (X, Y) = (x_1 X_1, ..., y_d Y_d), \ d = 2.$$

The formula becomes

$$\Delta F(a, b) = \sum_{\varepsilon\in\{0,1\}} (-1)^{s(\varepsilon)} F(b + \varepsilon * (a - b)).$$

Once the procedure is understood for $d = 2$, we may proceed to the general case.

General case, $d \geq 1$.

Let $a = (a_1, ..., a_d) \leq b = (b_1, ..., b_d)$ two points of $\mathbb{R}^d$. The volume of the cuboid

$$]a, b] = \prod_{i=1}^{d}]a_i, b_i],$$

by F, is defined by

$$\Delta F(a,b) = \sum_{\varepsilon=(\varepsilon_1,\ldots,\varepsilon_d)\in\{0,1\}^k} (-1)^{s(\varepsilon)} F(b_1+\varepsilon_1(a_1-b_1), \ldots, b_d+\varepsilon_k(a_d-b_d))$$

or

$$\Delta F(a,b) = \sum_{\varepsilon\in\{0,1\}^d} (-1)^{s(\varepsilon)} F(b+\varepsilon * (a-b)).$$

Similarly to the case $d = 2$, we have the

General rule of forming $\Delta F(a,b)$. $\Delta F(a,b)$ in formed as follows. First consider $F(b_1, b_2, ..., b_d)$ the value of F at right endpoint $b = (b_1, b_2, ..., b_d)$ of the interval $]a,b]$. Next proceed to the replacement of each b_i by a_i by replacing exactly one of them, next two of them etc., and add the each value of F at these points with a sign plus $(+)$ if the number of replacements is even and with a sign minus $(-)$ if the number of replacements is odd.

(c2) Cumulative Distribution Function.

In this part, we study finite components vectors.

Definition. For any real-valued random variable $X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto \mathbb{R}^d$, the function defined by

$$\mathbb{R}^d \ni x \mapsto F_X(x) = \mathbb{P}(X \leq x),$$

where $x^t = (x_1, ..., x_d)$ and

$$F_X(x_1,...,x_d) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, ..., X_d \leq x_d) = \mathbb{P}_X\left(\prod_{i=1}^{d}]-\infty, x_i]\right).$$

is called the cumulative distribution (cdf) function of X.

It has the two sets of important properties.

Properties of F_X.

(1) It assigns non-negative volumes to cuboids, that is

$$\forall (a,b) \in (\mathbb{R}^d)^2 \text{ such that } a \leq b, \ \Delta_{a,b}F \geq 0.$$

(2) It is right-continuous at any point $t \in \mathbb{R}^d$, that is,

$$F_X(t^{(n)}) \downarrow F_m(t)$$

as

$$(t^{(n)} \downarrow t) \Leftrightarrow (\forall 1 \leq i \leq d, \ t_i^{(n)} \downarrow t_i).$$

(3) F_X satisfies the limit conditions :

Condition (3-i)

$$\lim_{\exists i, 1 \leq i \leq k, t_i \to -\infty} F_X(t_1, ..., t_k) = 0$$

and Condition(3-ii)

$$\lim_{\forall i, 1 \leq i \leq k, t_i \to +\infty} F_X(t_1, ..., t_k) = 1.$$

As we did in one dimension, we have :

Definition. A function $F : \mathbb{R}^d \to [0,1]$ is **cdf** if and only if Conditions (1), (2) and (3) above hold.

(c3) Characterization.

The *cdf* is a characteristic function of the probability law of random variables of $\mathbb{R}^d$ from the following fact, as seen in Chapter 11 in Lo (2017b) of this series :

*There exists a one-to-one correspondence between the class of Probability Lebesgue-Stieljes measures $\mathbb{P}_F$ on $\mathbb{R}^d$ and the class of **cfd**'s $F_{\mathbb{P}}$ on $\mathbb{R}^d$ according the relations*

$$\forall x \in \mathbb{R}^d, \quad F_{\mathbb{P}}(x) = \mathbb{P}(]-\infty, x])$$

and

$$\forall (a,b) \in \left(\mathbb{R}^d\right), \ a \leq b, \quad \mathbb{P}_F(]a,b]) = \Delta_{a,b}F.$$

This implies that two d-random vectors X and Y having the same distribution function have the same probability law.

(c4) Joint *cdf*'s and marginal *cdf*'s.

Let us begin by the sample case where $d = 2$. Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^2$ be a random coupe, with $X^t = (X_1, X_2)$. We have,

$$\begin{aligned}(X_1 \leq x) &= X_1^{-1}(]-\infty, x]) = X_1^{-1}(]-\infty, x]) \cap X_1^{-1}(]-\infty, +\infty]) \\ &= \lim_{y \uparrow +\infty} X_1^{-1}(]-\infty, x]) \cap X_2^{-1}(]-\infty, y]) = \lim_{y \uparrow +\infty} (X_1 \leq x, X_2 \leq y)\end{aligned}$$

and by applying the Monotone Convergence Theorem, we have

$$\begin{aligned}\forall x \in \mathbb{R},\ F_{X_1}(x) &= \mathbb{P}(X_1 \leq x) \\ &= \lim_{y \uparrow +\infty} \mathbb{P}(X_1 \leq x, X_2 \leq y) = \lim_{y \uparrow +\infty} F_{(X_1,X_2)}(x, y).\end{aligned}$$

We write, for each $x \in \mathbb{R}$,

$$F_{X_1}(x) = F_{(X_1,X_2)}(x, +\infty).$$

The same thing could be done for the X_2. We may now introduce the following terminology.

Definition. $F_{(X_1,X_2)}$ is called the joint *cdf* of the ordered pair (X_1, X_2). F_{X_1} and F_{X_2} are called the marginal *cfd*'s of the couple. The marginal *cdf*'s may be computed directly but they also may be derived from the joint *cdf* by

$$F_{X_1}(x_1) = F_{(X_1,X_2)}(x_1, +\infty) \text{ and } F_{X_2}(x_2) = F_{(X_1,X_2)}(+\infty, x_2),\ (x_1, x_2) \in \mathbb{R}^2.$$

The extension to higher dimensions is straightforward. Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ be a random vector with $X^t = (X_1, ..., X_d)$.

(i) Each marginal *cdf* F_{X_i}, $1 \leq i \leq d$, is obtained from the joint *cdf* $F_X =: F_{(X_1,...,X_d)}$ by

$$F_{X_i}(x_i) = F_{(X_1,...,X_d)}\left(+\infty, ..., +\infty, \underbrace{x_i}_{i-th\ argument}, +\infty, ..., +\infty\right),\ x_i \in \mathbb{R},$$

or

$$F_{X_i}(x_i) = \lim_{(\forall j \in \{1,...,d\} \setminus \{i\},\ x_j \uparrow +\infty)} F_{(X_1,...,X_d)}(x_1, ..., x_d), \quad x_i \in \mathbb{R}.$$

(ii) Let $(X_{i_1}, ..., X_{i_r})^t$ be a sub-vector of X with $1 \leq r < d$, $1 \leq i_1 < i_2 < ... < i_r)$. Denote $I = \{i_1, ..., i_r\}$, the marginal *cfd* of $(X_{i_1}, ..., X_{i_r})$ is given by

$$F_{(X_{i_1},...,X_{i_r})}(x_{i_1}, ..., x_{i_r}) = \lim_{\forall j \in \{1,...,d\} \setminus I,\ x_j \uparrow +\infty} F_{(X_1,...,X_d)}(x_1, ..., x_d),\ (x_{i_1}, ..., x_{i_r}) \in \mathbb{R}^r.$$

(iii) Let $X^{(1)} = (X_1, ..., X_r)^t$ and $X^{(2)} = (X_{r+1}, ..., X_b)^t$ be two sub-vectors which partition X into consecutive blocs. The marginal *cdf*'s of $X^{(1)}$ and $X^{(2)}$ are respectively given by

$$F_{X^{(1)}}(x) = F_{(X_1,...,X_d)}\left(x, \underbrace{+\infty, ..., +\infty}_{(d-r) \text{ times}}\right),\ x \in \mathbb{R}^r$$

and

$$F_{X^{(2)}}(y) = F_{(X_1,...,X_d)}\left(\underbrace{+\infty, ..., +\infty}_{r \text{ times}}, y\right),\ y \in \mathbb{R}^{d-r}.$$

After this series of notation, we have this important theorem concerning a new characterization of the independence.

THEOREM 3. *Let* $X\ :\ (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ *be a random vector. Let us adopt the notation above. The following equivalences hold.*

(i) The margins X_i, $1 \leq i \leq d$ *are independent if and only if the joint cdf of* X *is factorized in the following way :*

$$\forall (x_1, ..., x_d) \in \mathbb{R}^d,\ F_{(X_1,...,X_d)}(x_1, ..., x_d) = \prod_{j=1}^{d} F_{X_i}(x_i).\ (FLM01)$$

(i) The two marginal vectors $X^{(1)}$ *and* $X^{(2)}$ *are independent if and only if the joint cdf of* X *is factorized in the following way : for* $(x^{(1)}, x^{2)}) \in \mathbb{R}^d$*, we have*

$$F_{(X_1,...,X_d)}(x^{(1)}, x^{2)}) = F_{X^{(1)}}(x^{(1)})F_{X^{(2)}}(x^{(2)}). \quad (FLM02)$$

Proof. This important characterization follows as a simple result of Measure Theory and Integration. The proof of the two points are very similar. So, we only give the proof of the first one.

Suppose that the components of X are independent. By Theorem 9 in Section 3 in Chapter 1, we have for any $(x_1, ..., x_d) \in \mathbb{R}^d$,

$$\begin{aligned} F_{(X_1,...,X_d)}(x_1, ..., x_d) &= \mathbb{P}(X_1 \leq x_1, ..., X_d \leq x_d) = \mathbb{E}\biggl(\prod_{j=1}^{d} 1_{]-\infty,x_j]}(X_j)\biggr) \\ &= \prod_{j=1}^{d} \mathbb{E}\biggl(1_{]-\infty,x_j]}(X_j)\biggr) = \prod_{j=1}^{d} F_{X_j}(x_j). \end{aligned}$$

Conversely, if Formula (FLM01) holds, the Factorization Formula (FACT02) in Part (10.03) in Doc 10-01 in Chapter 11 in Lo (2017b) of this series, we have : for any $a = (a_1, ..., a_k) \leq b = (b_1, ..., b_k)$,

$$\Delta_{a,b}F_X = \prod_{1 \leq i \leq k} (F_{X_i}(b_i) - F_{X_i}(a_i)).$$

By using the Lebesgue-Stieljes measures and exploiting the product measure properties, we have for any $(a, b) \in \mathbb{R}^2$, $a \leq b$,

$$\mathbb{P}_X(]a, b]) = \prod_{1 \leq i \leq k} \mathbb{P}_{X_i}(]a_i, b_i]) = \biggl(\otimes_{j=1}^{d} \mathbb{P}_{X_i} \biggr)(]a, b]).$$

So the probability measures $\mathbb{P}_X$ and $\otimes_{j=1}^{d}\mathbb{P}_{X_i}$ coincide on the π-system of rectangles of the form $]a, b]$ which generates $\mathcal{B}(\mathbb{R}^d)$. Hence they simply coincide. Thus the components of X are independent.

One handles the second point similarly by using Formula (FACT05) in the referred book at the same part, in the same document and the same section.

(c5) How Can we Define a Random Variable Associated to a *Cdf*.

As on $\mathbb{R}$, the Kolmogorov construction on $\mathbb{R}^d$, $d \geq 2$, is easy to perform.

For any *cdf* F on $\mathbb{R}^d$, we may define the Lebesgue-Stieljes measure $\mathbb{P}$ on $(\overline{\mathbb{R}}^d, \mathcal{B}_\infty(\overline{\mathbb{R}}^d)$ defined by

$$\mathbb{P}(]y,x]) = \Delta_{y,x}F, \ \ (y,x) \in (\mathbb{R}^d)^2, \ y \leq x. \ (LS21)$$

Now take $\Omega = \mathbb{R}^d$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^d)$ and let $X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto \mathbb{R}^d$ be the identity function

$$\forall \omega \in \Omega, \ X(\omega) = \omega.$$

Thus we have :

$$\forall x \in \mathbb{R}^d, \ F(x) = \mathbb{P}(]-\infty, x]). \ (LS22)$$

Particular case. In may situations, the above construction is stated as following : let $n \geq 1$ and F_1, F_2, ..., F_n be n *cdf*'s respectively defined on $\mathbb{R}^{d_i}$, $d_i \geq 1$. Can we construct a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding n independent random vectors X_1, ..., X_n such that for any $1 \leq i \leq n$, $F_{X_i} = F_i$.

The answer is yes. It suffices to apply the current result to the *cdf* F defined on $\mathbb{R}^d$, with $d = d_1 + ... + d_n$ and defined as follows :

$$\forall (x_1, ..., x_n)^t \prod_{j=1}^{d} \mathbb{R}^{d_j}, \ F(x_1, ..., x_n) = \prod_{j=}^{d} F_j(x_j).$$

Using Formula (FACT05) in Part (10.03) in Doc 10-01 in Chapter 11 in Lo (2017b) of this series, we see that F is a *cdf*. We may consider the identity function on $\mathbb{R}^d$ as above, form X_1 by taking the first d_1 components, X_2 by the next d_2 components, ..., X_d by the last d_n components. These subvectors are independent and respectively have the *cdf*'s F_1, ..., F_n.

5. Probability Laws and Probability Density Functions of Random vectors

Throughout this section we deal with random vectors, like the d-random vector ($d \geq 1$)

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto (\overline{\mathbb{R}}^d, \mathcal{B}_\infty(\overline{\mathbb{R}}^d),$$

with $X^t = (X_1, X_2, \cdots, X_d)$.

A- Classification of Random vectors.

(a) Discrete Probability Laws.

Definition. The random variable X is said to be discrete if it takes at most a countable number of values in $\overline{\mathbb{R}}$ denoted $\mathcal{V}_X = \{x^{(j)}, j \in J\}$, $\emptyset \neq J \subset \mathbb{N}$.

NB. On $\mathbb{R}$, we denote the values taken by such a random variable by sub-scripted sequences x_j, $j \in J$. In $\mathbb{R}^d$, $d \geq 2$, we use super-scripted sequences in the form $x^{(j)}$, $j \in J$, to avoid confusions with notation of components or powers.

Next, we give a set of facts from which we will make a conclusion on how to work with such random variables.

We already know from Measure Theory that X is measurable (Se Chapter 4, Doc 08-03, Criterion 4) if and only if

$$\forall j \in J, (X = x^{(j)}) \in \mathcal{A}.$$

Besides, we have for any $B \in \mathcal{B}_\infty(\overline{\mathbb{R}}^d)$,

$$(X \in B) = \sum_{j \in J, x^{(j)} \in B} (X = x^{(j)}). \ (DD01)$$

Now, we clearly have

$$\sum_{j \in J} \mathbb{P}(X = x^{(j)}) = 1. \ (DD02)$$

From (DD01), the probability law $\mathbb{P}_X$ of X is given by

$$\mathbb{P}_X(B) = \sum_{j \in J, x^{(j)} \in B} \mathbb{P}(X = x^{(j)}), \ (DD03)$$

for any $B \in \mathcal{B}_\infty(\overline{\mathbb{R}}^d)$. Let us denote the function defined on $\mathcal{V}_X$ by

$$\mathcal{V}_X \in x \mapsto f_X(x) = \mathbb{P}_X(\{x\}) = \mathbb{P}(X = x^{(j)}).$$

Next, let us consider the counting measure ν on $\mathbb{R}^d$ with support $\mathcal{V}_X$. Formulas (DD02) and (DD03) imply that

$$\int f_X \ d\nu = 1. \ (RD01)$$

and for any $B \in \mathcal{B}_\infty(\overline{\mathbb{R}}^d)$, we have

$$\int_B d\mathbb{P}_X = \int_B f_X \ d\nu. \ (RD02)$$

We conclude that f_X is the Radon-Nikodym derivative of $\mathbb{P}_X$ with respect to the σ-finite measure ν. Formula (RD02) may be written in the form

$$\int h d\mathbb{P}_X = \int h f_X \ d\nu. \ (RD03)$$

where $h = 1_B$. By using the four steps method of the integral construction, Formula (RD03) becomes valid whenever $\mathbb{E}h(X) = \int h \ d\mathbb{P}_X$ make senses.

We may conclude as follows.

Discrete Probability Laws.

If X is discrete, that is, it takes a countable number of values in $\overline{\mathbb{R}}^d$ denoted $\mathcal{V}_X = \{x^{(j)}, j \in J\}$, its probability law $\mathbb{P}_X$ is also said to be discrete. It has a probability density function *pdf* with respect to the counting measure on $\mathbb{R}^d$ supported by $\mathcal{V}_X$ and defined by

$$f_X(x) = \mathbb{P}(X = x), \ x \in \overline{\mathbb{R}}^d,$$

which satisfies

$$f_X(x^{(j)}) = \mathbb{P}(X = x^{(j)}) \text{ for } j \in J \text{ and } f_X(x) = 0 \text{ for } x \notin \mathcal{V}_X.$$

As a general rule, integrating any measurable function $h : \mathcal{B}_\infty(\overline{\mathbb{R}}^d) \rightarrow \mathcal{B}_\infty(\overline{\mathbb{R}})$ with respect to the probability law $\mathbb{P}_X$ is performed through the *pdf* f_X in the Discrete Integral Formula

$$\mathbb{E}h(X) = \int h f_X d\nu = \sum_{j \in J} h(x_j) f_X(x^{(j)}). \ \ (DIF1)$$

which becomes for $h = 1_B$, $B \in \mathcal{B}_\infty(\overline{\mathbb{R}}^d)$,

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \sum_{j \in J, x^{(j)}} f_X(x^{(j)}).(DIF2)$$

Some authors name *pdf*'s with respect to counting measures as *mass pdf's*. For theoretical purposes, they are Radon-Nikodym derivatives.

(b) Absolutely Continuous Probability Laws.

(b1) Lebesgue Measure on $\mathbb{R}^d$.

We already have on $\overline{\mathbb{R}}^d$ the σ-finite Lebesgue measures λ_d, which is the unique measure defined by the values

$$\lambda_d\biggl(\prod_{i=1} 2]a_i, b_i]\biggr) = \prod_{i=1}^{d} (b_i - a_i), \ \ (LM01)$$

for any points $a = (a_1, ..., a_d)^t \leq b = (b_1, ..., b_d)^t$ of $\mathbb{R}^d$. This formula also implies

$$\lambda_d\biggl(\prod_{i=1}^{d}]a_i, b_i]\biggr) = \prod_{i=1}^{d} \lambda_1(]a_i, b_i]), \ \ (LM02)$$

Let us make some Measure Theory reminders. Formula (LM02) ensures that λ_d is the product measure of the Lebesgue measure $\lambda_1 = \lambda$, that is

$$\lambda_d = \lambda^{\otimes d}.$$

Hence, we may use Fubini's Theorm for integrating a measurable function function $h : \mathcal{B}_\infty(\overline{\mathbb{R}}^d) \rightarrow \mathcal{B}_\infty(\overline{\mathbb{R}})$ through the formula

$$\int h d\lambda_d = \int d\lambda(x_1) \int .... \int d\lambda(x_{d-1}) \int h(x_1, ..., x_d) f_X(x_1, ..., x_d) d\lambda(x_d),$$

when applicable (for example, when h is non-negative or h is integrable).

(b2) Definition.

The probability Law $\mathbb{P}_X$ is said to be absolutely continuous if it is continuous with respect to λ_d. By extension, the random variable itself is said to be absolutely continuous.

NB. It is important to notice that the phrase **absolutely continuous** is specifically related to the continuity with respect to Lebesgue measure.

In the rest of this Point (b), we suppose that X is absolutely continuous.

(b3) Absolutely Continuous *pdf*'s.

By Radon-Nikodym's Theorem, there exists a Radon-Nikodym derivative denoted f_X such that for any $B \in \mathcal{B}_\infty(\overline{\mathbb{R}}^d)$,

$$\int_B d\mathbb{P}_X = \int_B f_X \ d\lambda_d.$$

The function f_X satisfies

$$f_X \geq 0 \text{ and } \int_{\mathbb{R}} f_X \ d\lambda_d = 1.$$

Such a function is called a *pdf* with respect to the Lebesgue measure. Finally, we may conclude as follows.

As a general rule, integrating any measurable function $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ with respect to the probability law $\mathbb{P}_X$, which is absolutely continuous, is performed through the *pdf* f_X with the Absolute Continuity Integral Formula

$$\mathbb{E}h(X) = \int h f_X \ d\lambda_d. \ (ACIF)$$

Since λ_d is the product of the Lebesgue measure on $\mathbb{R}$ d times, we may use Fubini's Theorem when applicable to have

$$\mathbb{E}h(X) = \int d\lambda_1(x_1) \int .... \int d\lambda_1(x_{d-1}) \int h(x_1, ..., x_d) f_X(x_1, ..., x_d) d\lambda_1(x_d).$$

In particular, the *cdf* of X becomes

$$\begin{aligned} F_X(x) &= \int_{-\infty} x_d d\lambda(x_1) \int_{-\infty} x_2 \\ &.... \int_{-\infty} x_{d-1} d\lambda(x_{d-1}) \int_{-\infty} x_d h(x_1, ..., x_d) f_X(x_1, ..., x_d) d\lambda(x_d) \end{aligned}$$

for any $x = (x_1, ..., x_d)^t \in \mathbb{R}^d$.

(b4) Criterion for Absolute Continuity from the Cdf.

In practical computations, a great deal of Lebesgue integrals on $\mathbb{R}$ are Riemann integrals. Even integrals with respect to the multidimensional Lebesgue Measure can be multiple Riemann ones. But we have to be careful for each specific case (See Points (b5) and (b6) below).

Let be given the *dcf* F_X of a random vector, the absolute continuity of X would give for any $x \in \mathbb{R}^d$

$$F_X(x) = \int_{-\infty}^{x_1} d\lambda(x_1) \int_{-\infty}^{x_2} d\lambda(x_2) ... d\lambda(x_{d-1}) \int_{-\infty}^{x_k} f_X(x_1, ..., x_d) \ d\lambda(x_d). \ (AC01)$$

If f_X is locally bounded and locally Riemann integrable (LLBRI), we have

$$f_X(x_1, x_2, ..., x_k) = \frac{\partial^k F_X(x_1, x_2, ..., x_k)}{\partial x_1 \partial x_2 ... \partial x_x}, \ (AC02)$$

λ_d-*a.e.*. (See Points (b5) and (b6) below for a more detailed explanation of *LLBRI* functions and for a proof).

From a computational point of view, the above Formula quickly helps to find the *pdf*, if it exists.

(b5) Cautions to be taken when replacing Lebesgue integral by Riemann ones.

Let us consider that we are on $\mathbb{R}$ and let X be a real random variable with an absolutely *pdf* f. For any measurable function h from $\mathbb{R}$ to $\mathbb{R}$, the expectation

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)f(x)\ d\lambda(x),\ (EL)$$

is defined with respect to the Lebesgue measure. It happens that for computation such an integral, we lean to use the improper Riemann integral

$$\mathbb{E}(h(X)) = \int_{-\infty}^{+\infty} h(x)f(x)\ dx.\ (ER)$$

Although this works for a lot of cases, we cannot use the just mentioned formula without a minimum of care, since in Riemann integration we may have that $\int_{\mathbb{R}} h(x)f(x)\ dx$ is finite and $\int_{\mathbb{R}} |f(x)|\ dx$ infinite, a situation that cannot occur with Lebesgue integration.

We may use the results of Doc 06-07 in Chapter 7 in Lo (2017b) of this series to recommend the following general rule that we will follow in this book.

Let us suppose that the function hf is *LLBRI* (implying that hf is λ-*a.e.* continuous on $\mathbb{R}$). We have :

(a) If $\mathbb{E}(h(X))$ exists and is finite, then Formula (ER) holds as an improper Riemann integral (as an application of the Dominated Convergence Theorem), that is

$$\mathbb{E}(h(X)) = \lim_{n\to+\infty} \int_{a_n}^{b_n} h(x)f(x)\ dx,\ (ER02)$$

for any sequence $(a_n, b_n)_{n\geq 0}$ such that $(a_n, b_n) \to (-\infty,\ +\infty)$ as $n \to +\infty$. In such a case, we may chose a particular alike sequence to compute $\mathbb{E}(h(X))$.

To check whether $\mathbb{E}(h(X))$ is finite, we may directly use Riemann integrals (which are based on the Monotone Convergence Theorem)

$$\mathbb{E}(h^-(X)) = \int_{-\infty}^{+\infty} (hf)^+(x)f(x)\ dx \quad (ENPa)$$

and

$$\mathbb{E}(h^-(X)) = \int_{-\infty}^{+\infty} (hf)^-(x)f(x)\ dx, (ENPb)$$

and apply the classical Riemann integrability criteria.

(b) If the Riemann improper integral of $|hf|$ exists and is finite, then the Lebesgue integral of hf exists (by using the Monotone Convergence Theorem on the positive and negative parts) and Formula (ER) holds.

(c) Even if $\mathbb{E}(h(X))$ exists and is infinite, Formula (ER) still holds, by using the Monotone Convergence Theorem on the positive and negative parts and exploiting Formula (ENP).

Finally, such results are easily extended in dimension $d \geq 2$, because of the Fubini's integration formula.

(b6) Back to Formula (AC01).

Dimension one. If f is *LLBRI*, we surely have that f is λ-$a.e.$ continuous and we may treat the integrals $\int_{-\infty}^{x} f_X(t)\ d\lambda(t)$ as a Riemann ones. By the known results for indefinite Riemann integrals, we have

$$\left(\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^{x} f_X(t)\ d\lambda(t)\right) \Leftrightarrow \frac{dF_X}{dx} = f_X\ \lambda - a.e..$$

Remark that the constant resulting in the solution of the differential equation in the right-hand assertion is zero because of $F_X(-\infty) = 0$.

Dimension $d \geq 2$. Let $d = 2$ for example. Let f_X be *LLBRI*. By Fubini's theorem,

$$\forall (x,y) \in \mathbb{R}^2, F_X(x,y) = \int_{-\infty}^{x} d\lambda(s) \biggl(\int_{-\infty}^{y} f_X(s,t)\ d\lambda(t) \biggr).$$

The function,

$$t \rightarrow \int_{-\infty}^{y} f_X(s,t)\ d\lambda(t)$$

is bounded (by the unity) and continuous. By, returning back to Riemann integrals, we have

$$\forall (x,y) \in \mathbb{R}^2, \quad \frac{\partial F_X(x,y)}{\partial x} = \int_{-\infty}^{y} f_X(x,t)\ d\lambda(t).$$

By applying the results for dimension one to the partial function $f_X(x,t)$, for x fixed, which is (LBLI), we get

$$\forall (x,y) \in \mathbb{R}^2, \frac{\partial^2 F_X(x,y)}{\partial y \partial x} = \int_{-\infty}^{y} f_X(x,t)\ \lambda - a.e.$$

The order of derivation may be inverted as in the Fubini's Theorem.

The general case $d \geq 2$ is handled by induction. ■

(c) General case.

Let us be cautious! Later, we will be concerned by practical computations and applications of this theory. We will mostly deal with discrete or absolutely continuous random variables. But, we should be aware that these kind of probability laws form only a small part of all the possibilities, as we are going to see it.

By the Lebesgue Decomposition Theorem (Doc 08-01, Part III, Point (05-06), Chapter 9), there exists a unique decomposition of $\mathbb{P}_X$ into an absolutely continuous measure ϕ_{ac}, associated to a non-negative Radon-Nikodym $f_{r,X}$ and λ_d-singular measure ϕ_s, that is, for any $B \in \mathcal{B}_\infty(\overline{\mathbb{R}}^d)$,

$$\mathbb{P}_X(B) = \int_B f_{r,X}\ d\lambda_d + \phi_s(B).$$

The λ-singularity of ϕ_s means that there exists a λ_d-null set N such that for all $B \in \mathcal{B}_\infty(\overline{\mathbb{R}}^d)$,

$$\phi_s(B) = \phi_s(B \cap N)$$

Suppose that none of ϕ_s and ϕ_s is the null measure. If N is countable that is N may be written as $N = \{x^{(j)}, j \in J\}$, $\emptyset \neq J \subset \mathbb{N}$, the measure is discrete and for any $B \in \mathcal{B}_\infty(\overline{\mathbb{R}}^d)$, we have

$$\phi_s(B) = \sum_{j \in J} f_{d,X}(x^{(j)}),$$

where

$$f_{d,X}(x^{(j)}) = \phi_s(x^{(j)}), \quad j \in J.$$

We have $0 < a = \phi_{ac}(\overline{\mathbb{R}}^d)$, $b = \phi_s(\overline{\mathbb{R}}^d) \leq 1$, and $a + b = 1$. Let us denoting by ν the counting measure with support N. Then $f_X^{(1)} = f_{r,X}/a$ is an absolutely continuous *pdf* and $f_X^{(2)} = f_{d,X}/b$ is a discrete *pdf* and we have for all $B \in \mathcal{B}_\infty(\overline{\mathbb{R}}^d)$,

$$\mathbb{P}_X(B) = a \int_B f_X^{(1)} \, d\lambda_d + (a-1) \int_B f_X^{(2)} \, d\nu.$$

Hence, $\mathbb{P}_X$ is mixture of two probability laws, the first being absolutely continuous and the second being discrete.

We may be more precise in dimension one.

More detailed decomposition on $\mathbb{R}$. We already saw that a real *cdf* Fmay be decomposed into two *df*'s :

$$F = F_c + F_d,$$

where F_c is continuous and F_d is discrete. Surely, the Lebesgue-Stieljes measure associated with F_d, denoted by ϕ_d, is discrete. The Lebesgue-Stieljes measure associated with F_c, denoted by ϕ_c, may decomposed as above into

$$\phi_c = \phi_{ac} + \phi_s$$

where ϕ_{ac} is absolutely continuous and ϕ_s is singular. Since $F_c(-\infty) = F_d(-\infty) = 0$, we may go back to the *df*'s to have :

$$F = F_{ac} + F_c + F_d,$$

where F_{ac} is *df* of measure absolutely continuous, F_d is a discrete *df* and F_c is a continuous and, unless it is equal to the null measure, is neither discrete nor absolutely continuous.

This fact is obvious since F_c is continuous and cannot be discrete. Also, it is singular and cannot be absolutely continuous.

We have the following conclusion.

Position of any probability law with respect to the Lebesgue measure. Any probability law is a mixture of an absolutely continuous probability measure $\mathbb{P}_{ac,X}$, associated to a *pdf* $f_{ac,X}$, a discrete distribution probability measure $\mathbb{P}_{d,X}$, which is a λ-singular measure $\mathbb{P}_{d,X}$ which has a countable strict support $\mathcal{V}_{d,X}$ and of a λ-singular probability measure $\mathbb{P}_{c,X}$ which has a non-countable λ-null set support, respectively associated to $p_1 \geq 0$, $p_2 \geq 0$ and $p_2 \geq 0$, with $p_1 + p_2 + p_3 = 1$, such that

$$\mathbb{P}_X = p_1\mathbb{P}_{ac,X} + p_2\mathbb{P}_{d,X} + p_3\mathbb{P}_{c,X}.$$

The probability measures are respectively associated to the *df*'s F_{ac}, F_d, F_c so that we have

$$F_X = F_{ac} + F_d + F_c,$$

$$\frac{dF_{ac}(x)}{dx} = f_{ac,X}, \ \lambda - a.e,$$

$\mathcal{V}_{d,X}$ is the set of discontinuity points of F, and F_c is continuous but not λ-$a.e.$ differentiable.

By *strict countable support of* $\mathbb{P}_{d,X}$, we mean a support such that for any point x in, we have $\mathbb{P}_{d,X}(\{x\}) > 0$.

Warning. If the decomposition has more that one term, the corresponding functions among F_{ac}, F_d and F_c are not *cdf*'s but only *df*'s.

(b7) Marginal Probability Density functions.

Let us begin, as usual, by the simple case where $d = 2$. Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^2$ be a random couple, with $X^t = (X_1, X_2)$. Let us suppose that X has a *pdf* $f_{(X_1,X_1)}$ with respect to a σ-finite product measure $m = m_1 \otimes m_2$ on $\mathbb{R}^2$. Let us show that each X_i, $i \in \{1, 2\}$, has a *pdf* with respect to m. We have, for any Borel set B,

$$\begin{aligned} \mathbb{P}(X_1 \in B) &= \mathbb{P}((X_1, X_2) \in B \times \mathbb{R}) \\ &= \int 1_{B\times\mathbb{R}} f_{(X_1,X_2)}(x, y) \ dm(x, y) \\ &= \int_B \left(\int_{\mathbb{R}} f_{(X_1,X_2)}(x, y) \ dm_2(y) \right) dm_1(x). \end{aligned}$$

By definition, the function

$$f_{X_1}(x) = \int_{\mathbb{R}} f_{(X_1,X_1)}(x, y) \ dm_2(y), \ m - a.e \ in \ x \in \mathbb{R},$$

is the *pdf* of X with respect of m_1, named as the *marginal pdf* of X_1. We could do the same for X_2. We may conclude as follows.

Definition. Suppose that the random order pair $X^t = (X_1, X_2)$ has a *pdf* $f_{(X_1,X_1)}$ with respect to a σ-finite product measure $m = m_1 \otimes m_2$ on $\mathbb{R}^2$. Then each X_i, $i \in \{1, 2\}$, has the marginal *pdf*'s f_{X_i} with respect to m_i, and

$$f_{X_1}(x) = \int_{\mathbb{R}} f_{(X_1,X_2)}(x, y) \ dm_2(y), m_1 - a.e. \in \ x \in \mathbb{R}$$

and

$$f_{X_2}(x) = \int_{\mathbb{R}} f_{(X_1,X_2)}(x, y) \ dm_1(x), \ m_2 - a.e. \in x \in \mathbb{R}.$$

The extension to higher dimensions is straightforward. Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ be a random vector with $X^t = (X_1, ..., X_d)$. Suppose that X has a *pdf* $f_{(X_1,...,X_d)}$ with respect to a σ-finite product measure $m = \otimes_{j=1}^d m_j$.

(i) Then each X_j, $j \in \{1, d\}$, has the marginal *pdf*'s f_{X_j} with respect to m_j given m_i-*a.e.*, for $x \in \mathbb{R}$, by

$$f_{X_j}(x) = \int_{\mathbb{R}^{d-1}} f_{(X_1,...,X_d)}(x_1, ..., x_d)\, d\biggl(\otimes_{i\leq i\leq d,\ i\neq j} m_i\biggr)(x_1, ..., x_{j-1}, x_{j+1}, ..., x_d).$$

(ii) Let $(X_{i_1}, ..., X_{i_r})^t$ be a sub-vector of X with $1 \leq r < d$, $1 \leq i_1 < i_2 < ... < i_r)$. Denote $I = \{i_1, ..., i_r\}$, the marginal *pdf* of $(X_{i_1}, ..., X_{i_r})$ with respect to $m = \otimes_{i=1}^{r} m_{i_j}$ is given for $(x_1, ..., x_r) \in \mathbb{R}^r$ by

$$f_{(X_{i_1},...,X_{i_r})}(x_1, ..., x_r) =$$
$$\int_{\mathbb{R}^{d-r}} f_{(X_1,...,X_d)}(x_1, ..., x_d)\, d\biggl(\ \otimes_{1\ i\leq d,\ i\notin I}\ m_i\biggr)(x_j, \mathrm{J} \in \{1, ..., n\} \setminus I).$$

Let $X^{(1)} = (X_1, ..., X_r)^t$ and $X^{(2)} = (X_{r+1}, ..., X_b)^t$ be two sub-vectors which partition X into two consecutive blocs. Then $X^{(1)}$ and $X^{(2)}$ have the *pdf* $f_{X^{(1)}}$ and $f_{X^{(2)}}$ with respect to $\otimes_{j=1}^{r} m_j$ and $m = \otimes_{j=r+1}^{d} m_j$ respectively, and given for $x \in \mathbb{R}^r$ by

$$f_{X^{(1)}}(x) = \int_{\mathbb{R}^{d-r}} f_{(X_1,...,X_d)}(x_1, ..., x_d)\, d\biggl(\ \otimes_{r+1\leq i\leq d}\ m_i\biggr)(x_{r+1}, ..., x_d),$$

and for $x \in \mathbb{R}^{d-r}$ by

$$f_{X^{(2)}}(x) = \int_{\mathbb{R}^{r}} f_{(X_1,...,X_d)}(x_1, ..., x_d)\, d\biggl(\ \otimes_{1\leq i\leq r}\ m_i\biggr)(x_1, ..., x_r).$$

After this series of notations, we have this important theorem for characterizing the independence.

THEOREM 4. *Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ be a random vector. Let us adopt the notation above. Suppose that we are given a σ-finite product measure $m = \otimes_{j=1}^{d} m_j$ on $\mathbb{R}^d$, and X has a pdf f_X with respect to m. We have the following facts.*

(i) The margins X_i, $1 \leq i \leq d$ are independent if and only if the joint pdf of X is factorized in the following way :

$$\forall (x_1, ..., x_d) \in \mathbb{R}^d,\ f_{(X_1,...,X_d)}(x_1, ..., x_d) = \prod_{j=1}^{d} f_{X_i}(x_i),\ m.a.e.\ (DLM01)$$

(i) The two marginal vectors $X^{(1)}$ and $X^{(2)}$ are independent if and only if the joint pdf of X is factorized in the following way : for all $(x^{(1)}, x^{2)}) \in \mathbb{R}^d$,

$$f_{(X_1,...,X_d)}(x^{(1)}, x^{2)}) = f_{X^{(1)}}(x^{(1)})f_{X^{(2)}}(x^{(2)}), \ m.a.e. \ (FLM02)$$

Proof. It will be enough to prove the first point, the proof of the second being very similar. Suppose that the X_i are independent. It follows that for any Borel rectangle $B = B_1 \times ... \times B_d$, we have

$$\begin{aligned}
\mathbb{P}(X \in B) &= \int_B f_X(x) \ dm(x) \\
&= \mathbb{P}(X_1 \in B_1, ..., X_d \in B_d) \\
&= \prod_{j=1}^{d} \mathbb{P}(X_j \in B_j) \\
&= \prod_{j=1}^{d} \int_{B_j} f_{X_j}(x_i) \ dm_j(x_j) \\
&= \int_{B_1 \times ... \times B_d} \left(\prod_{j=1}^{d} f_{X_j}(x_i) \right) dm(x)
\end{aligned}$$

Thus the two finite measures

$$B \mapsto \int_B f_X(x) \ dm(x) \text{ and } B \mapsto \int_B \left(\prod_{j=1}^{d} f_{X_j}(x_i) \right) dm(x)$$

coincide on a π-system generating the whole σ-algebra. Thus, they coincide. Finally, we get two finite indefinite integrals with respect to the same σ-finite measure m. By the Radon-Nikodym Theorem, the two Radon-Nikodym derivatives are equal m-*a.e.*

Suppose now that Formula (DLM01) holds. Thanks to Fubini's Theorem, we readily get the factorization of the joint *cdf* and get the independence through Theorem 3.

6. Characteristic functions

After the *cdf*'s, are going to see a second kind of characterization function for probability laws.

I - Definition and first properties.

It is important to say that, in this section, we only deal with finite components random vectors with values in spaces $\mathbb{R}^d$, $d \geq 1$, endowed with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}$ which is the product σ-algebra of $\mathcal{B}(\mathbb{R})$ d times.

(a) Characteristic function.

DEFINITION 6. *For any random variable $X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto \mathbb{R}^d$, the function*

$$u \mapsto \phi_X(u) = \mathbb{E}(e^{i<X,u>}),$$

is called the characteristic function of X. Here, i is the complex number with positive imaginary part such that $i^2 = -1$.

This function always exists since we interpret the integral in the following way

$$\mathbb{E}(e^{i<X,u>}) = \mathbb{E}(\cos < X, u >) + i \ \mathbb{E}(\sin < X, u >),$$

which is defined since the integrated real and imaginary parts are bounded.

The role played by the characteristic function in Probability Theory may also be played by a few number of functions called moment generating functions. These functions do not always exist, and if they do, they may be defined only on a part of $\mathbb{R}^d$. The most used of them is defined as follows.

(a) Moment Generated Function (*mgf*).

The following function

$$u \mapsto \varphi_X(u) = \mathbb{E}(e^{<X,u>}), \ u \in \mathbb{R}^d,$$

when defined on a domain D of $\mathbb{R}^d$ containing the null vector as an interior point, is called the moment generating function (*mfg*) of X.

If φ_X exists on some domain D to which zero is interior, we will prefer it to $\Phi_X(u)$, to avoid to use the *complex* number i involved in Φ_X. Non-mathematician users of Probability Theory would like this.

Besides, we may find the characteristic function by using the moment generating functions as follows :

$$\Phi_X(u) = \varphi_X(iu), u \in \mathbb{R}^d.$$

The characteristic function has these two immediate properties.

PROPOSITION 4. *For all* $u \in \mathbb{R}^d$,

$$\|\phi_X(u)\| \leq 1 = \|\phi_X(0)\|.$$

Besides $\phi_X(u)$ *is uniformly continuous at any point* $u \in \mathbb{R}^d$.

This proposition is easy to prove. In particular, the second point is an immediate application to the Dominated Convergence Theorem.

Here are the :

II - Main properties of the characteristic function.

THEOREM 5. *We have the following facts.*

(a) Let X be a random variable with value in $\mathbb{R}^d$, A a $(k \times d)$-matrix of real scalars, B a vector of $\mathbb{R}^k$. Then the characteristic function of $Y = AX + B \in \mathbb{R}^k$ is given,

$$\mathbb{R}^k \ni u \mapsto \phi_Y(u) = e^{<B,u>}\phi_X(A^t u), \ u \in \mathbb{R}^k.$$

(b) Let X and Y be two independent random variables with values in $\mathbb{R}^d$, defined on the same probability space. The for any $u \in \mathbb{R}^d$, we have

$$\phi_{X+Y}(u) = \phi_X(u) \ \times \phi_Y(u).$$

(c) Let X and Y be two random variables respectively with values in $\mathbb{R}^d$ and in $\mathbb{R}^k$ and defined on the same probability measure. If the random variables X and Y are independent, then for any $u \in \mathbb{R}^d$ and for $v \in \mathbb{R}^k$, we have

$$\phi_{(X,Y)}(u, v) = \phi_X(u) \times \phi_Y(v). \tag{6.1}$$

Let us make some remarks before we give the proof of the theorem. In Part A, Section 3, Chapter 6, the characterization (c) was stated and quoted as (CI4), and admitted without proof. Here, the proof will be based on a characterization of product measure.

Point (c) provides a characterization of the independence between X and Y. But the decomposition in Point (b) is not enough to ensure the independence. You may consult counter-examples book of Stayonov (1987) or the monograph Lo (2017a) of this series, Part A, Section 3, Chapter 6, where is reported a counter-example from Stayonov (1987).

Proof of Theorem 5.

Point (a). By definition, we have $< AX+B, u >= {}^t(AX+B)u = {}^tX(A^Tu) + B^Tu$. Hence,

$$\begin{aligned}\phi_{AX+B}(u) &= \mathbb{E}(e^{{}^tX(A^tu)+B^tu}) = e^{<B,u>} \times \mathbb{E}(e^{<X,A^tu)}) \\ &= e^{<B,u>}\phi_X(A^tu).\end{aligned}$$

Point (b). Let X and Y be independent. We may form $X+Y$ since they both have their values in $\mathbb{R}^d$, and they are defined on the same probability space. We have for any $u \in \mathbb{R}^d$,

$$\phi_{X+Y}(u) = \mathbb{E}\left(e^{<X+Y,u>}\right) = \mathbb{E}\left(e^{<X,u>}e^{<Y,u>}\right) = \mathbb{E}\left(e^{<X,u)}\right)\times\mathbb{E}\left(e^{<Y,u)}\right).$$

Point (c). Let X and Y be two independent random variables with values in $\mathbb{R}^d$ and $\mathbb{R}^k$. Let u and v be two respectively elements of $\mathbb{R}^d$ and $\mathbb{R}^k$. We have

$$< \begin{pmatrix} X \\ Y \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} >=< X, u > + < Y, v > .$$

Then

$$\begin{aligned}\phi_{(X,Y)}(u,v) = E\left(\exp < \begin{pmatrix} X \\ Y \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} >\right) \\ = \mathbb{E}\left(e^{<X,u>+<Y,v>}\right) = E(e^{<X,u>})\mathbb{E}(e^{<Y,v>}) \\ = \phi_X(u) \times \phi_Y(v).\end{aligned}$$

The proof is over. $\blacksquare$.

Now, we want to move to next very important other characterization. When $d = 1$, we have an explicit inversion formula which expresses the *cdf* of a probability law on $\mathbb{R}^d$ by means of its characteristic function. The characterization of a probability law on $\mathbb{R}$ by its characteristic function follows from this inversion formula.

But when $d > 1$, things are more complicated and we may need a non-standard version of the Theorem of Stone-Weierstrass Theorem. In that case a more general characterization of probability measures in metric spaces may be useful. So we begin with general characterizations.

III - Characterization of a probability law on a metric space.

Let us suppose that we are working on a metric space (E, ρ) endowed with the metric ρ. We are going to use the class $C_b(E)$ of real-valued continuous and bounded functions defined on E. Let us begin by reminding that, by the λ-π Lemma (See Lo (2017b), Exercise 11 of Doc 04-02, Part VI, page 228), the class of open sets $\mathcal{O}$ is a determining class of probability measures since it is a π-system, containing E and generating $\mathcal{B}(E)$, that is, for two probability measures $\mathbb{P}_j$ ($j \in \{1, 2\}$) on $(E, \mathcal{B}(E))$, we have

$$(\mathbb{P}_1 = \mathbb{P}_2) \Leftrightarrow (\forall G \in \mathcal{O},\ \mathbb{P}_1(G) = \mathbb{P}_2(G)). \tag{6.2}$$

Actually, this characterization can be extended to integrals of $f \in C_b(E)$. For this, we need the following tool.

LEMMA 1. *Let G be a non-empty open in E. There exists a non-decreasing sequence of functions $(f_m)_{m\geq 1}$ such that :*

(1) for each $m \geq 1$, f_m is a Lipschitz function of coefficient m and

$$0 \leq f_m \leq 1_G.$$

and $f_m = 0$ on ∂G and

(2) we have

$$f_m \uparrow 1_G,\ \text{as } m \uparrow +\infty.$$

The proof is given in the Appendix Chapter 10 in Lemma 15 (page 330).

This lemma may be used to get the following characterization : for two probability measures $\mathbb{P}_j$ $(j \in \{1,2\})$ on $(E, \mathcal{B}(E))$, we have

$$(\mathbb{P}_1 = \mathbb{P}_2) \Leftrightarrow \left(\forall f \in C_b(E),\ \int f\ d\mathbb{P}_1 = \int f\ d\mathbb{P}_2\right). \tag{6.3}$$

To establish this, we only need to show the indirect implication. Suppose that right-hand assertion holds. For any $G \in \mathcal{O}$, we consider the the sequence $(f_m)_{n\geq 1}$ in Lemma 1 and we have

$$\forall m \geq 1,\ \int f_m\ d\mathbb{P}_1 = \int f_m\ d\mathbb{P}_2.$$

By letting $m \uparrow +\infty$ and by applying the Monotone Convergence Theorem, we get $\mathbb{P}_1(G) = \mathbb{P}_2(G)$. Since this holds for any $G \in \mathcal{O}$, we get $\mathbb{P}_1 = \mathbb{P}_2$ by Formula .

IV - Characterization of a probability law on $\mathbb{R}^d$ by its characteristic function.

We are going to prove that characteristic functions also determine probability laws on $\mathbb{R}^d$.

Theorem 6. *Let X and Y be two random variables with values in $\mathbb{R}^d$. Their characteristic functions coincide on $\mathbb{R}^d$ if and only if do their probability laws on $\mathcal{B}(\mathbb{R}^d)$, that is*

$$\Phi_X = \Phi_Y \Leftrightarrow \mathbb{P}_X = \mathbb{P}_Y.$$

Proof. We are going to use an approximation based on a version of the theorem of Stone-Weierstrass. Let us begin by reminding that the class of intervals of $\mathbb{R}^d$

$$\mathcal{I}_d = \{]a,b[= \prod_{j=1}^{d}]a_j, b_j[,\ a \leq b, (a,b) \in \left(\mathbb{R}^d\right)^2\}$$

is a π-system, contains $E = \mathbb{R}^d$ and generates $\mathcal{B}\left(\mathbb{R}^d\right)$. By the the λ-π Lemma, it constitutes a determining class for probability measures.

Fix $G =]a,b[$ with $a_j < b_j$, for all $1 \leq j \leq d$. For any $j \in \{1, \cdots, d\}$ and consider the sequence $(f_{j,m})_{m\geq 0} \subset C_b(\mathbb{R}^d)$ constructed for $G_j =]a_j, b_j[$ in Lemma 1. The numbers $f_{j,m}(a_j)$ and $f_{j,m}(a_j)$ are zero. So the functions

$$f_m(x) = \prod_{j=1}^{d} f_{j,m}(x_j),\ x = (x_1, ..., x_d)^t \in \mathbb{R}^d,\ m \geq 1,$$

vanish on the border ∂G of G since

$$\partial G = \{x \in G,\ \exists j \in \{1, \cdots, d\},\ x_j = a_j \text{ or } x_j = b_j\}$$

It becomes clear that for any probability measure $\mathbb{L}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, we have

$$\forall]a,b[\in \mathcal{I}_d,\ f_m \uparrow 1_{]a,b[} \text{ and } \int f_m \, d\mathbb{L} \uparrow \mathbb{L}(]a,b[),\ \text{as } m \uparrow +\infty.$$

We may seize the opportunity to state a new characterization of probability measures of $\mathbb{R}^d$. Let $C_{b,0}(\mathbb{R}^d)$ be the class of functions f for which there exists $]a,b[\in \mathcal{I}_d$ such that $0 \leq f \leq 1$ and $f = 0$ outside $]a,b[$. We get that :

For two probability measures $\mathbb{P}_j$ $(j \in \{1,2\})$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$:

$$(\mathbb{P}_1 = \mathbb{P}_2) \Leftrightarrow \left(\forall f \in C_{b,0}(\mathbb{R}^d),\ \int f \, d\mathbb{P}_1 = \int f \, d\mathbb{P}_2\right). \tag{6.4}$$

Now fix $f \in C_{b,0}(\mathbb{R}^d)$ associated with $[a,b]$. Let $\varepsilon \in]0,1[$. Fix $r > 0$ and $K_r = [-r,r]^d$. We choose r such that

$$-r \leq \min(a_1, ..., a_d) \text{ and } r \geq \max(b_1, ..., b_d) \tag{6.5}$$

and

$$\mathbb{P}_X(K_r^c) + \mathbb{P}_Y(K_r^c) \leq \frac{\varepsilon}{2(2+\varepsilon)}. \tag{6.6}$$

Now consider the class $\mathcal{H}$ of finite linear combinations of functions of the form

$$\prod_{j=1}^{d} \exp\left(in_j \pi x_j / r\right), \tag{6.7}$$

where $n_j \in \mathbb{Z}$ is a constant and i is the normed complex of angle $\pi/2$ and let $\mathcal{H}_r$ be the class of the restrictions h_r of elements $h \in \mathcal{H}$ on $K_r = [-r,r]^d$.

It is clear that $\mathcal{H}_r$ is a sub-algebra of $C_b(K_r)$ with the following properties.

(a) for each $h \in \mathcal{H}$, the uniform norm of h on $\mathbb{R}^d$ is equal to the uniform norm of h on K_r, that is

$$\|f\|_\infty = \sup_{x\in\mathbb{R}^d} |h(x)| = \sup_{x\in K_r} |h(x)| = \|f\|_{K_r}.$$

This comes from that remark that h is a finite linear combination of functions of the form in Formula 6.7 above and each factor $\exp(in_j \pi x_j/r)$ is a $2r$-periodic function.

(b) $\mathcal{H}_r$ separates the points of $K_r \setminus \partial K_r$ and separates points of $K_r \setminus \partial K_r$ from points of ∂K_r. Indeed, if x and y are two points in K_r, at the exception where both of them are edge points of K_r of the form

$$(x, y) \in \{(s_1, ..., s_d) \in K_r,\ \forall j \in \{1, ..., d\},\ s_j = r \textit{ or } s_j = r\}^2,$$

there exists $j_0 \in \{1, \cdots, d\}$ such that $0 < |x_{j_0} - y_{j_0}| < 2r$ that is $|(x_{j_0} - y_{j_0})/r| < 2$ and the function

$$h_r(x) = \exp(i\pi x_{j_0}/r)$$

separates x and y since $h_r(x) = h_r(y)$ would imply $\exp(i\pi(x_{j_0} - x_{j_0})/r) = 1$, which in term would imply $x_{j_0} - x_{j_0} = 2\ell r$, $\ell \in \mathbb{Z}$. The only possible value of ℓ would be zero and this is impossible since $x_{j_0} - y_{j_0} \neq 0$.

(c) For all the points in $t \in \partial K_r$, the function $g(t) \equiv 0 \in \mathcal{H}_r$ converges to $f(t) = 0$.

(d) $\mathcal{H}_r$ contains all the constant functions.

We may then apply Corollary 2 in Lo (2018b) (Corollary 4 in the appendix, page 330) to get that : there exists $h_r \in \mathcal{H}_r$ such that

$$\|f - h_r\|_{K_r} \leq \varepsilon/4. \tag{6.8}$$

and by Point (a) above (using also that the norm of $f \in C_{b,0}$ less or equal to 1), we have

$$\|h\|_\infty = \|h_r\|_{K_r]} \leq \|f\|_\infty + \varepsilon/4 \leq 1 + \varepsilon. \tag{6.9}$$

Now, y the assumption of equality of the characteristic functions, we have

$$\mathbb{E}(h(X)) = \mathbb{E}(h(Y)).$$

We have have

$$\begin{aligned}
\mathbb{E}(f(X)) - \mathbb{E}(f(Y)) &= \left(\int f\ d\mathbb{P}_X - \int h\ d\mathbb{P}_X\right) + \left(\int h\ d\mathbb{P}_X - \int h\ d\mathbb{P}_Y\right) \\
&+ \left(\int h\ d\mathbb{P}_Y - \int f\ d\mathbb{P}_Y\right) \\
&= \left(\int f\ d\mathbb{P}_X - \int h\ d\mathbb{P}_X\right) + \left(\int h\ d\mathbb{P}_Y - \int f\ d\mathbb{P}_Y\right).
\end{aligned}$$

The first term satisfies

$$\begin{aligned}
(6.10)\, \mathbb{E}\left|\int f\ d\mathbb{P}_X - \int h_r\ d\mathbb{P}_X\right| &\leq \int_{K_r]} |f - h_r|\ d\mathbb{P}_X \\
&+ \int_{K_r^c} |f - h|\ d\mathbb{P}_X \\
&\leq \varepsilon/4 + (\|f\| + \|h\|)\mathbb{P}_X(K_r^c), \\
&\leq \varepsilon/4 + (2+\varepsilon)\mathbb{P}_X(K_r^c),
\end{aligned}$$

where we used Formulas 6.8 and 6.9.

By treating the second term in the same manner, we also get

$$\mathbb{E}\left|\int f\ d\mathbb{P}_Y - \int h d\mathbb{P}_Y\right| \leq \varepsilon/4 + (2+\varepsilon)\mathbb{P}_Y(K_r^c). \tag{6.11}$$

By putting together Formulas (6.10) and (6.11) and by remembering Formulas (6.5) and (6.6), we get

$$|\mathbb{E}(f(X)) - \mathbb{E}(f(Y))| \leq \varepsilon/2 + (2+\varepsilon)(\mathbb{P}_X(K_r^c) + \mathbb{P}_Y(K_r^c)) \leq \varepsilon.$$

for any $\varepsilon \in]0,1[$. So, for all $f \in C_{b,0}(\mathbb{R}^d)$,

$$\int f d\mathbb{P}_X = \int f d\mathbb{P}_Y.$$

We close the proof by applying Formula (6.4) above.

V - Inversion Formula on $\mathbb{R}$ and applications.

Here, we consider the characteristic function of a Lebesgue-Stieljes measures on $\mathbb{R}$, not necessarily a probability measure. After the proof of the following proposition, we will get another characterization of probability laws by characteristic functions by means of *cdf*'s. Let us begin to state the

PROPOSITION 5. *Let F be an arbitrary distribution function. Let*

$$\Phi(x) = \int exp(itx)\ d\lambda_F(x),\ \ x \in \mathbb{R},$$

where λ_F denotes the Lebesgue-Stieljes measure associated with F. Set for two reals numbers a and b such that $a < b$,

$$J_U =: J_U(a,b) = \frac{1}{2\pi}\int_{-U}^{U} \frac{e^{-iau} - e^{-ibu}}{iu} \Phi_X(u)\ du. \tag{6.12}$$

(a) Then, we have, as $U \to +\infty$, J_U converges to

$$F(b-) - F(a) + \frac{1}{2}\bigg(F_X(a) - F(a-) + F(b) - F(b-) \bigg)). \tag{6.13}$$

(b) If a and b are continuity points of F, then

$$F(b) - F(a) = \lim_{U \to +\infty} J_U. \tag{6.14}$$

If F is absolutely continuous, that is there exists a measurable λ-a.e. finite function f such that for $x \in \mathbb{R}$,

$$F(x) = \int_{-\infty}^{x} f(t) d(x), \tag{6.15}$$

then, we have λ-a.e.,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ixu}\Phi(u)\ du.$$

Proof. Recall Dirichlet's Formula

$$\begin{aligned}\int_{-\infty}^{0} \frac{\sin x}{x} dx &= \int_{0}^{+\infty} \frac{\sin x}{x} dx \\ &= \lim_{b \to +\infty} \int_{0}^{b} \frac{\sin x}{x} dx = \pi/2,\end{aligned}$$

which can be proved, for example, using complex integration based on residues. We deduce from it that the numbers

$$\int_{a}^{b} \frac{\sin x}{x} dx = \int_{a}^{0} \frac{\sin x}{x} dx + \int_{0}^{b} \frac{\sin x}{x} dx, \ a \leq 0 \leq b$$

are uniformly bounded in a and b, say by M. By using Fubini's theorem, we have

$$\begin{aligned}J_U &= \frac{1}{2\pi} \int_{-U}^{U} \frac{e^{-iau} - e^{-ibu}}{iu} \left(\int e^{iux} d\mathbb{P}_X(x) \right) \ du \\ &= \int d\mathbb{P}_X(dx) \times \frac{1}{2\pi} \int_{-U}^{U} \frac{e^{-i(a-x)u} - e^{-i(b-x)u}}{iu} du \\ &= \int J(U,x) \ d\mathbb{P}_X(x),\end{aligned}$$

where

$$\begin{aligned}J(U,x) &= \frac{1}{2\pi} \int_{-U}^{U} \frac{e^{-i(a-x)u} - e^{-i(b-x)u}}{iu} du \\ &= \frac{1}{2\pi i} \int_{-U}^{U} \frac{\cos(u(a-x)) - \cos(u(b-x))}{u} du \\ &+ \frac{1}{2\pi} \int_{-U}^{U} \frac{\sin(u(b-x)) - \sin(u(a-x))}{u} du.\end{aligned}$$

But, we also have

$$\int_{-U}^{U} \frac{\cos(u(a-x)) - \cos(u(b-x))}{u} du = 0.$$

Since the integrated functions are odd and the integration is operated on a symmetrical compact interval with respect to zero. We get

$$\begin{aligned} J(U,x) &= \frac{1}{2\pi}\int_{-U}^{U}\frac{\sin(u(b-x))-\sin(u(a-x))}{u}du \\ &= \frac{1}{2\pi}\int_{-U(b-x)}^{U(b-x)}\frac{\sin v}{v}dv - \frac{1}{2\pi}\int_{-U(a-x)}^{U(a-x)}\frac{\sin v}{v}dv. \end{aligned}$$

Thus, $J(U,x)$ uniformly bounded bounded by M/π. Next by considering the position of x with respect of the interval (a,b) and by handling accordingly the signs of $(b-x)$ and $(a-x)$, we easily arrive at the following set of implications :

$$(x < a \text{ } or \text{ } x > b) \Rightarrow J(U,x) \to 0 \text{ as } U \to +\infty,$$

$$(x = a \text{ } or \text{ } x = b) \Rightarrow J(U,x) \to 1/2 \text{ as } U \to +\infty$$

$$(a < x < b) \Rightarrow J(U,x) \to 1 \text{ as } U \to +\infty.$$

Then

$$J(U,x) \to 1_{]a,b[} + \frac{1}{2}1_{\{a\}} + \frac{1}{2}1_{\{b\}}.$$

From there, we apply the Fatou-Lebesgue Theorem to get

$$\begin{aligned} J_U &\to \int \left(1_{]a,b[} + \frac{1}{2}1_{\{a\}} + \frac{1}{2}1_{\{b\}}\right) d\mathbb{P}_X(x) \\ &= F(b-) - F(a) + \frac{1}{2}\Big(F(a) - F(a-) + F(b) - F(b-)\Big). \end{aligned}$$

This proves Point (a). □

Point (b) If a and b are continuity points of F, the limit in (6.13) reduces to $F(b) - F(a)$. □

Point (c) Now, from (6.14), we deduce that F is continuous and next, the derivative of F at x is $f(x)$ when f is continuous. But a measurable function that is integrable is λ-$a.e.$ continuous. So,

$$\frac{dF(x)}{dx} = f(x), \ \lambda - a.e.$$

Also, by (6.14), we have for all $h > 0$,

$$\frac{F(a+h)-F(a)}{h} = \lim_{U \to +\infty} \frac{1}{2\pi} \int_{-U}^{U} \frac{e^{-iau} - e^{-i(a+h)u}}{ihu} \Phi_X(u)\, du. \tag{6.16}$$

Then for any $a \in \mathbb{R}$,

$$\begin{aligned}
f(a) &= \lim_{h \to 0} \frac{F(a+h)-F(b)}{h} \\
&= \frac{1}{2\pi} \lim_{h \to 0} \lim_{U \to +\infty} \int_{-U}^{U} \frac{e^{-iau} - e^{-i(a+h)u}}{ihu} \Phi(u)\, du \\
&= \frac{1}{2\pi} \lim_{U \to +\infty} \lim_{h \to 0} \int_{-U}^{U} \frac{e^{-iau} - e^{-i(a+h)u}}{ihu} \Phi(u)\, du \\
&= \frac{1}{2\pi} \lim_{U \to +\infty} \int_{-U}^{U} \lim_{h \to 0} \frac{e^{-iau} - e^{-i(a+h)u}}{ihu} \Phi(u)\, du;
\end{aligned}$$

where the exchange between integration and differentiation in the last line is allowed by the use the Fatou-Lebesgue theorem based on the fact that the integrated function is bounded by the unity which is integrable on $(-U, U)$, U fixed.

So, we arrive at

$$\begin{aligned}
f(a) &= \frac{1}{2\pi} \lim_{U \to +\infty} \int_{-U}^{U} \lim_{h \to 0} \frac{e^{-iau} - e^{-i(a+h)u}}{ihu} \Phi(u)\, du \\
&= \frac{1}{2\pi} \lim_{U \to +\infty} \int_{-U}^{U} e^{-iau} \Phi(u)\, du \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iau} \Phi(u)\, du.
\end{aligned}$$

λ-a.e. ■

Application.

Now, let us use this to prove Theorem 6 for $k = 1$.

Let X and Y be two *rrv*'s with equal characteristic functions. By (6.14), their distribution functions F_X and F_Y are equal on the set $D_{X,Y}$ of continuity points of both F_X and F_Y. The complement of that set is at most countable. So, for $x \in D_{X,Y}$ fixed, we may find a sequence of numbers $(x_n)_{n\geq 0}$ such that

$$(x_n)_{n\geq 0} \subset D, \text{ such that } x_n \rightarrow x \text{ as } n \uparrow +\infty.$$

So, we will have for any $n \geq 0$

$$F_X(x) - F_X(a_n) = F_Y(x) - F_Y(a_n).$$

By letting $n \rightarrow +\infty$, we get for all $x \in D_{X,Y}$

$$F_X(x) = F_Y(x).$$

For any $x \in \mathbb{R}$, we also can can find monotone sequence $(x_n)_{n\geq 0}$ such that

$$(x_n)_{n\geq 0} \subset D, \text{ such that } x_n \downarrow x \text{ as } n \uparrow +\infty.$$

By right-continuity at x of F_X and F_Y, we have

$$F_X(x) = \lim_{n\uparrow+\infty} F_X(x_n) = \lim_{n\uparrow+\infty} F_Y(x_n) = F_Y(x_n).$$

Conclusion $F_X = F_X$. Thus by the first characterization, X and Y have the same probability law.■

IV - A characterization of independence.

We are going to see that Point (c) of Theorem 5 is a rule for independence because of Theorem 6. We have

THEOREM 7. *Let X and Y be two random variables respectively with values in $\mathbb{R}^d$ and in $\mathbb{R}^k$ and defined on the same probability measure. The random variables X and Y are independent if and only if for any $u \in \mathbb{R}^d$ and for $v \in \mathbb{R}^k$, we have*

$$\phi_{(X,Y)}(u,v) = \phi_X(u) \times \phi_Y(v) \tag{6.17}$$

Proof. We need only to prove that (6.17) implies independence of X and Y. Suppose that (6.17) holds. It is clear that the left-hand member of (6.17) is the characteristic function of the product measure $\mathbb{P}_X \otimes \mathbb{P}_Y$.

Since the characteristic functions of the probability laws $\mathbb{P}_{(X,Y)}$ and $\mathbb{P}_X \otimes \mathbb{P}_Y$ coincide, we get

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y,$$

which is the definition of the independence between X and Y. ■

V - Characteristic functions and moments for *rrv*.

We are going to see how to find the moments from the characteristic function in the following. Let us write

$$\Phi_X(u) = \int e^{iux} d\mathbb{P}_X(x), \ u \in \mathbb{R}.$$

The function

$$g(u,x) = \cos(ux) + i\sin(ux) = e^{iux}$$

is differentiable with respect to u and its derivative is

$$g'(u,x) = ix(\cos(ux) + i\sin ux)) = ixe^{iux}.$$

It is bounded by $Y(x) = |x|$. The integral of this function $Y(x)$ is the mathematical expectation of X, that is,

$$\int Y(x) d\mathbb{P}_X(x) = \int |x| \, d\mathbb{P}_X(X) = E\,|X|\,.$$

Suppose that the mathematical expectation is finite. Then, by the Dominated Convergence Theorem (See Point 06.14 in Doc 06.14, Chapter 7, in Lo (2017b) of this series), we may exchange integration and differentiation. The method may be repeated by a second differentiation and so forth. We conclude this quick discussion in

PROPOSITION 6. *If* $\mathbb{E}(X)$ *exists and is finite, then the function* $u \mapsto \phi_X(u)$ *is differentiable and we have*

$$\phi'_X(u) = \int ixe^{iux} d\mathbb{P}_X(x).$$

And we have

$$i \times \mathbb{E}(X) = \phi'_X(0).$$

More generally, if for $k \geq 1$, $\mathbb{E}\,|X|^k$ exists and is finite, then the function $u \mapsto \phi_X(u)$ is differentiable k times with

$$\phi_X^{(k)}(u) = i^k \int x^k e^{iux} d\mathbb{P}_X(x)$$

and

$$\mathbb{E}X^k = -i^k \quad \phi_X^{(k)}(0).$$

7. Convolution, Change of variables and other properties

I - Convolution product of probability density functions on $\mathbb{R}$.

Let X and Y be two real-valued random variables which are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and mutually independent. Set $Z = X + Y$. By definition, the probability law of Z is called the convolution product of the probability laws of $\mathbb{P}_X$ and of $\mathbb{P}_Y$, denoted as

$$\mathbb{P}_Z = \mathbb{P}_X * \mathbb{P}_Y. \tag{7.1}$$

Now, suppose that X and Y have probability density functions f_X and f_Y with respect to the Lebesgue measure λ. Then Z has an absolutely probability density function f_Z denoted as

$$f_{X+Y} = f_X * f_Y.$$

We have the following

Proposition 7. *Let X and Y be to real-valued and independent random variables, defined on the same probability measure $(\Omega, \mathcal{A}, \mathbb{P})$, and admitting the probability density functions f_X and f_Y with respect to a σ-finite product measure $\nu = \nu_1 \otimes \nu_2$. Then Z has a* pdf *f_Z which has the two following to expressions :*

$$f_X * f_Y(z) = \int_{\mathbb{R}} f_X(z-x)\ f_Y(x)\ d\lambda(x).$$

Proof. Assume the hypotheses of the proposition hold. Let us use the joint probability law of (X, Y) to have

$$F_Z(z) = \mathbb{P}(X + Y \leq x) = \int_{(x+y \leq z)} d\mathbb{P}_{(X,Y)}(x, y).$$

Since X and Y are independent, we have

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y.$$

We may apply Fubini's Theorem to get

$$\begin{aligned} F_Z(z) &= \int (x + y) d\mathbb{P}_{(X,Y)}(x, y) = \int d\mathbb{P}_X(x) \int_{y \leq z-x} d\mathbb{P}_Y(y) \\ &= \int f_X(x) d\nu_1(y) \left(\int_{y \leq z-x} f_Y(y) \ d\nu_2(y) \right). \end{aligned}$$

We recall the the Lebesgue measure is invariant by translation. Let us make the change variable $u = y + x$, to have

$$\begin{aligned} F_Z(z) &= \int f_X(x) \ dx \left(\int_{u \leq z} f_Y(u - x) \ du \right) \\ &= \int f_X(x) \ dx \ \left(\int_{-\infty}^{z} f_Y(u - x) \ du \right). \end{aligned}$$

Let us use again the Fubini's Theorem to get

$$F_Z(z) = \int \int_{-\infty}^{z} f_Y(u - x) \ f_X(x) \ dx \ dy = \int_{-\infty}^{z} \left(\int f_Y(u - x) \ f_X(x) \ dx \right) \ du.$$

Taking the differentiation with respect to z, we get

$$f_Z(z) = \int f_Y(z - x) \ f_X(x) \ dx.$$

For such a formula for discrete random variables, the reader is referred Lo (2017a) of this series, Formula (3.24), Part D, Section 3, Chapter 6.

II - Change of Variable by Diffeomorphisms and Introduction to the Gauss Random variables.

(a) Recall of the Change of Variable Formula for Riemann Integrals on $\mathbb{R}^d$ (See Valiron (1946), page 275, for double integration).

Suppose we have the following Riemann integral on $\mathbb{R}^d$,

$$I = \int_D f(x_1, x_2, \cdots, x_d)\ dx_1 dx_2 \cdots dx_d,$$

where D is a domain of $\mathbb{R}^d$. We will write for short with $x^t = (x_1, x_2, \cdots, x_d)$,

$$I = \int_D f(x)\ dx.$$

Suppose that we have a diffeomorphism h from an other domain Δ of $\mathbb{R}^k$ to D. This means that the function

$$h : \Delta \mapsto D$$

(a) is a bijection (one-to-one mapping).

(b) h and its inverse function $g = h^{-1}$ have continuous partial derivatives (meaning that they are both of class C^1).

Let us write h as :

$$D \ni x = h(y) \longleftrightarrow y \in \Delta.$$

The components of h are denoted by h_i :

$$x_i = h_i(y) = h_i(y_1, ..., y_d).$$

The d-square matrix of elements

$$\frac{\partial x_i}{\partial y_j} = \frac{h_i(y_1, ..., y_d)}{\partial y_j}$$

written also as

$$M(h) = \left[\left(\frac{\partial x_i}{\partial y_j} \right)_{ij} \right].$$

is called the Jacobian matrix of the transformation. The absolute value of its determinant is called the **Jacobian coefficient** of the change of variable. We way write it as

$$J(h,y) = \det\left(\left[\left(\frac{\partial x_i}{\partial y_j}\right)_{ij}\right]\right).$$

The change of variable formula is the following

$$I = \int_{\Delta} f(h(y)) \ |J(h,y)| \ dy.$$

We replace x by $h(y)$, the domain D by Δ, but we multiply the integrated function by the Jacobian coefficient (depending on y).

(b) An example leading to the Gaussian probability Law.

Let us give a classical example. Suppose we want to compute

$$I = \int_{[0,+\infty[\times[0,+\infty[} e^{-(x^2+y^2)} dx \ dy.$$

Let us the polar coordinates of (x,y) in $(\mathbb{R}_+)^2$:

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}$$

with

$$(x,y) \in D = [0,+\infty[\times[0,+\infty[\longleftrightarrow (r,\theta) \in [0,+\infty[\times[0,\pi/2].$$

The Jacobian coefficient of the transformation is

$$J(r,\theta) = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

We apply the change of variable formula to have

$$I = \int_{[0,+\infty[\times[0,\pi/2]} re^{-r^2} dr \ d\theta = \int_{[0,\pi/2]} d\theta \int_{[0,+\infty[} re^{-r^2} dr = \frac{\pi}{4}.$$

By the Fubini's Formula, we have

$$I = \int_{[0,+\infty[} e^{-x^2} dx \ \int_{[0,+\infty[} e^{-y^2} dy = (\int_0^{+\infty} e^{-u^2} du)^2.$$

Then, we have

$$\int_0^{+\infty} e^{-u^2} du = \frac{\sqrt{\pi}}{2}.$$

Finally, by a new change of variable, where we take the evenness of the function $u \mapsto exp(-u^2/2)$, leads to

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-u^2/2} du = 1.$$

This is a probability density function. Compare this with the lengthy proof in Section 5, Chapter 7, in Lo (2017a) of this series.

Let us apply this Formula to finding new density functions.

(c) Finding a probability density function by change of variables.

Let X be a random variable in $\mathbb{R}^d$ of probability density function f_X with respect to the Lebesgue measure on $\mathbb{R}^k$, still denoted by $\lambda_k(x) = dx$. Suppose that D is the support of X. Let

$$h : \Delta \mapsto D$$

be a diffeomorphism and

$$Y = h^{-1}(X)$$

be another random vector. Then, the probability density function of Y exists and is given by

$$f_Y(y) = f_X(h(y)) \ |J(h)| \ 1_\Delta(y). \ (CVF)$$

This follows from an immediate application of the variable change formula. Let B be a borel set of $\mathbb{R}^d$, we have

$$\int_{x \in h(B)} f_X(x) \ dx = \int_{h^{-1}(x) \in B} f_X(x) \ dx.$$

Let us apply the variable change formula as follows :

$$\begin{aligned}\mathbb{P}(Y \in B) &= \int_{y \in B} f_X(h(y))\ 1_\Delta(y)\,|J(h,y)|\,dy \\ &= \int_B \{f_X(h(y))\ 1_\Delta(y)\,|J(h,y)|\}\,dy.\end{aligned}$$

We deduce from this that

$$f_Y(y) = f_X(h(y))\ 1_\Delta(y)\,|J(h,y)|$$

is the probability density function of Y.

In Mathematical Statistics, this tool is extensively used, especially for Gaussian random variables.

(d) Important example.

This example is important for two reasons. First, we will have to apply many of the techniques used in the this chapter and secondly, the object of the example is the starting point of the study of stable laws.

Let us consider two independent $\mathcal{E}(\lambda)$-random variables X_1 and X_2, $\lambda > 0$ on a same probability space (such a construction is achieved through the Kolmogorov construction method) and let us set $X_s = X_1 - X_2$. The *pdf* of X_s is the convolution product of f_{X_1} and f_{-X_2}. The *pdf* f_{-X_0} is

$$f_{-X_2}(y) = \lambda \exp(\lambda y),\ y \leq 0.$$

So, we have for all $x \in \mathbb{R}$,

$$\begin{aligned}f_{X_s}(x) &= (f_{X_1} * f_{-X_2})(x) \\ &= \int f_{X_1}(x-y) f_{-X_2}(y)\ dy \\ &= \lambda^2 \int \Big(\exp(-\lambda(x-y))1_{(x-y\geq 0)}\Big)\Big(\exp(\lambda y)1_{(y\leq 0)}\Big)\ dy \\ &= \lambda^2 \int \Big(\exp(-\lambda(x-y))1_{(y\leq x)}\Big)\Big(\exp(\lambda y)1_{(y\leq 0)}\Big)\ dy.\end{aligned}$$

If $x \leq 0$, we have

$$\begin{aligned}
f_{X_s}(x) &= \lambda^2 \int_{-\infty} x \exp(-\lambda(x-y)) \exp(\lambda y)\ dy \\
&= \lambda^2 \exp(-\lambda x) \int_{-\infty} x \exp(2\lambda y)\ dy \\
&= \lambda^2 \exp(-\lambda x) \left[\frac{e^{2\lambda y}}{2\lambda}\right]_{-\infty}^{x} \\
&= \frac{\lambda}{2} \exp(\lambda x).
\end{aligned}$$

If $x \geq 0$, we have

$$\begin{aligned}
f_{X_s}(x) &= \lambda^2 \int_{-\infty} 0 \exp(-\lambda(x-y)) \exp(\lambda y)\ dy \\
&= \lambda^2 \exp(-\lambda x) \int_{-\infty}^{0} \exp(2\lambda y)\ dy \\
&= \frac{\lambda}{2} \exp(-\lambda x).
\end{aligned}$$

In total, we have

$$f_{X_s}(x) = \frac{\lambda}{2} \exp(-\lambda |x|),\ x \in \mathbb{R}. \tag{7.2}$$

Next, let us see an interesting application of the inversion formula. The characteristic function of X_s is

$$\begin{aligned}
\Phi_{X_s}(u) &= \Phi_{X_1 - X_2}(u) = \Phi_{X_1 - X_2}(u) \\
&= \Phi_{X_1}(u)\Phi_{X_2}(-u) \\
&= \frac{1}{1 - it/\lambda} \frac{1}{1 + it/\lambda}
\end{aligned}$$

which leads to

$$\Phi_{X_s}(u) = \frac{\lambda^2}{\lambda^2 + u^2},\ u \in \mathbb{R}. \tag{7.3}$$

Now let us apply the inversion formula to this characteristic function. We have λ-a.e. for all $x \in \mathbb{R}$

$$\begin{aligned}\frac{\lambda}{2}\exp(-\lambda|x|) &= \frac{1}{2\pi}\int e^{-iux}\Phi_{X_s}(u)\ du\\ &= \frac{1}{2\pi}\int e^{-iux}\frac{\lambda^2}{\lambda^2+u^2}\ du,\end{aligned}$$

and by dividing both members by by $(\lambda/2)$ we get

$$\exp(-\lambda|x|) = \int e^{-iux}\frac{\lambda}{\pi(\lambda^2+u^2)}\ du, \tag{7.4}$$

and by replacing x by $-x$, we conclude that we have λ-a.e. for all $x \in \mathbb{R}$,

$$\int e^{iux}\frac{\lambda}{\pi(\lambda^2+u^2)}\ du = \exp(-\lambda|x|).$$

It happens that

$$f_{C(0,\lambda)} = \frac{\lambda}{\pi(\lambda^2+u^2)},\ x \in \mathbb{R},$$

is the *pdf* of a Cauchy random variable of parameters 0 and $\lambda > 0$ (see Chapter 3, Section 2, 115). We just found the characteristic of a Cauchy random variable, which is not easy to find by direct methods.

8. Copulas

The lines below should form a part of Section 8 which was devoted to *cdf*'s. But, nowadays, the notion of copula is central in Statistics theory, although copulas are simply particular *cdf*'s in Probability. So we think that introducing to copulas in a section might serve for references.

A very recurrent source on copulas is Nelsen (2006). However, the lines below will use the note of Lo (2018).

Definition A copula on $\mathbb{R}^d$ is a *cdf* C whose marginal *cdf*'s defined by, for $1 \leq i \leq d$,

$$\mathbb{R} \ni s \mapsto C_i(s) = C\left(+\infty, ..., +\infty, \underbrace{s}_{i-th\ argument}, +\infty, ..., +\infty\right),$$

are all equal to the $(0,1)$-uniform *cdf* which in turn is defined by

$$x \mapsto x1_{[0,1[} + 1_{[1,+\infty[},$$

and we may also write, for all $s \in [0,1]$,

$$C_i(s) = C\left(1, ..., 1, \underbrace{s}_{i-th\ argument}, 1, ..., 1\right) = s. \tag{8.1}$$

The copula became very popular with following the important theorem of Sklar (1959)

THEOREM 8. *For any cdf F on $\mathbb{R}^d$, $d \geq 1$, there exists a copula C on $\mathbb{R}^d$ such that*

$$\forall x \in \mathbb{R}^d,\ F(x) = C(F_1(x), ..., F_d(x)). \tag{8.2}$$

This theorem is now among the most important tools in Statistics since it allows to study the dependence between the components of a random vector through the copula, meaning that the intrinsic dependence does not depend on the margins.

We are going to provide a recent proof due to Lo (2018). Fortunately, the tools we need are available in the current series, in particular in Lo *et al.* (2016).

Proof of Sklar (1959)'s Theorem.

(A) - Complements. We first need some complements to the properties of the generalized inverse function given in Lo *et al.* (2016). Let us begin by defining generalized functions. Let $[a,b]$ and $[c,d]$ be non-empty intervals of $\mathbb{R}$ and let $G : [a,b] \mapsto [c,d]$ be a non-decreasing mapping such that

$$\begin{aligned} c &= \inf_{x\in[a,b]} G(x), \quad (L11) \\ d &= \sup_{x\in[a,b]} G(x). \quad (L12) \end{aligned}$$

Since G is a mapping, this ensures that

$$\begin{aligned} a &= \inf\{x \in \mathbb{R},\ G(x) > c\}, \quad (L13) \\ b &= \sup\{x \in \mathbb{R}, G(x) < d\}. \quad (L14) \end{aligned}$$

If $x = a$ or $x = b$ is infinite, the value of G at that point is meant as a limit. If $[a, b]$ is bounded above or below in $\mathbb{R}$, G is extensible on $\mathbb{R}$ by taking $G(x) = G(a+)$ for $x \leq a$ and $G(x) = G(b-0)$ for $x \geq b$. As a general rule, we may consider G simply as defined on $\mathbb{R}$. In that case, $a = lep(G)$ and $b = uep(G)$ are called *lower end-point* and *upper end-point* of G.

The generalized inverse function of G is given by

$$\forall u \in [lep(G), uep(G)],\ G^{-1}(u) = \inf\{x \in \mathbb{R},\ G(x) \geq u\}.$$

The properties of G^{-1} have been thoroughly studied, in particular in Billinsgley (1968), Resnick (1987). The results we need in this paper are gathered and proved in wcrv or in Lo *et al.* (2016b) (Chapter 4, Section 1) and reminded as below.

Lemma 2. *Let G be a non-decreasing right-continuous function with the notation above. Then G^{-1} is left-continuous and we have*

$$\forall u \in [c, d],\ G(G^{-1}(u)) \geq u\ (A) \text{ and } \forall x \in [a, b],\ G^{-1}(G(x)) \leq x\ (B)$$

and

$$\forall x \in [lep(G), uep(G)],\ G^{-1}(G(x) + 0) = x. \tag{8.3}$$

Proof. The proof of Formulas (A) and (B) are well-known and can be found in the cited books above. Let us prove Formula (8.3) for any $x \in [a, b]$.

On one side, we start by the remark that $G^{-1}(G(x) + 0)$ is the limit of $G^{-1}(G(x) + h)$ as $h \searrow 0$. But for any $h > 0$, $G^{-1}(G(x) + h)$ is the infimum of the set of $y \in [a, b]$ such that $G(y) \geq G(x) + h$. Any these y satisfies $y \geq x$. Hence $G^{-1}(G(x) + 0) \geq x$.

On the other side $G(x+h) \searrow G(x)$ by right-continuity of G, and by the existence of the right-hand limit of the non-decreasing function $G^{-1}(\circ)$,

$G^{-1}(G(x+h)) \searrow G^{-1}(G(x)+0)$. Since $G^{-1}(G(x+h)) \leq x+h$ by Formula (B), we get that $G^{-1}(G(x)+0) \leq x$ as $h \searrow 0$. The proof is complete. □

(B) - Proof of Sklar's Theorem. Define for $s=(s_1, s_2, \cdots, s_d) \in [0,1]^d$,

$$C(s) = F(F_1^{-1}(s_1+0), F_2^{-1}(s_2+0), \cdots, F_d^{-1}(s_d+0)). \tag{8.4}$$

It is immediate that C assigns non-negative volumes to cuboids of $[0,1]^d$, since according to Condition (DF2), Formula (**??**) for C derives from the same for F where the arguments are the form $F_i^{-1}(\circ+0)$, $1 \leq i \leq d$.

Also C is right-continuous since F is right-continuous as well as each $F_i^{-1}(\circ+0)$, $1 \leq i \leq d$. By passing, this explains why we took the right-limits because the $F_i^{-1}(\circ)$'s are left-continuous.

Finally, by combining Formulas (8.3) and (8.4), we get the conclusion of Sklar in Formula (8.2). The proof is finished. □

9. Conclusion

(A) Back to independence of Random vectors.

Because of the importance of the notion of independence and since several characterizations of the independence are scattered this chapter and in Chapter 1, we think that a summary on this point may be useful to to reader.

(1) The most general definition of a finite family of random variables is given in Definition 3 (page 21). This definition covers all type of random variables and uses the finite product measure. Random variables of an infinite family are independent if and only if the elements each finite sub(family are independence.

In this general case, Theorem (page 101) gives a general characterization.

(2) When we have a random real-valued vector in $\mathbb{R}^d$, $d \leq 1$, the independence of the coordinates and the independence of sub-vectors are characterized :

(2a) in Theorem 3 (page 61), using the cumulative distribution functions,

(2b) in Theorem 7 (page 89), using the characteristic functions,

(2c) in Theorem 4 (page 75), using the probability density functions with respect to the measure.

(B) General advices to determine probability laws.

Now, we have the means to characterize the usual probability laws by their distribution functions or their characteristic functions. Its is also important to know the parameters of the usual laws. In the next two chapters, we will be dealing with them. Estimating these from data is one of the most important purposes of Statistics.

In trying to find the probability laws, the following ideas may be useful.

(A) Using the convolution product to find the probability law of the sum of two independent real-value random variables.

(B) Using the product of characteristic function to find the probability law of the sum of two independent random variables of equal dimension.

(C) Finding the distribution function of the studied random variable and differentiate it if possible, and try to identify a known probability law.

(D) Directly finding the characteristic function of the studied random variable and trying to identify a known probability law.

(E) Using the Change of Variable Formula to derive *pdf*'s if applicable.

(F) In particular, the following easy *stuff* may be useful :

A useful stuff. Suppose that two random elements X and Y, are defined on the same probability space and take their values in the same measure space $(E, \mathcal{B}, \nu)$, which is endowed with a measure ν. Suppose that X and Y have *pdf*'s f_X and f_Y with respect to ν and that these two *pdf*'s a common support $\mathcal{V}$ and have a common variable part, meaning that there exist a non-negative function $h : E \rightarrow \mathbb{R}$ and constants $C_1 > 0$ and $C_2 > 0$ such that

$$\forall x \in E,\ f_X(x) = C_1 h(x) \text{ and } f_Y(x) = C_2 h(x).$$

Then $f_X = f_Y$, ν-*a.s.* and $C_1 = C_2$. $\diamondsuit$

The proof is obvious since

$$1 = \int_{\mathcal{V}} f_X \ d\nu = C_1 \int_{\mathcal{V}} h \ d\nu = \int_{\mathcal{V}} f_X \ d\nu = C_2 \int_{\mathcal{V}} h \ d\nu.$$

which leads to

$$C_1 = C_2 = 1/\left(\int_{\mathcal{V}} h \ d\nu\right).$$

Despite its simplicity, this *stuff* is often used and allows to get remarkable Analysis formulas, some of them being extremely difficult, even impossible, to establish by other methods.

CHAPTER 3

Usual Probability Laws

We begin to focus on real random variables. Later, we will focus on Random vectors in Chapter 4.

Actually, the researchers have discovered a huge number of probability laws. A number of dictionaries of probability laws exist (See for example, Kotz *et al.* (199), which is composed of 13 volumes at least). Meanwhile, people are still continuing to propose new probability laws and their properties (see Okorie *et al.* (2017) for a recent example).

This chapter is just a quick introduction to this wide area. A short list among the most common laws is given. Some others concern new important probability laws (Skewed normal, hyperbolic, etc.).

I - Review of usual probability law on $\mathbb{R}$.

We begin with discrete random variables. For such random variables, the discrete integration formula is used to find the parameters and the characteristic functions. This has already been done in the monograph of Lo (2017a). We will not repeat the computations here.

1. Discrete probability laws

For each random variable X, the values set or support $\mathcal{V}_X$, the probability density function with respect to the appropriate counting measure, the characteristic function and/or the moment generating function and the moments are given.

(1) Constant random variable $X = a$, a.s, $a \in \mathbb{R}$.

X takes only one value, the value a.

Discrete probability density function on $\mathcal{V}_X = \{a\}$:

$$\mathcal{V}_X = \{a\} \text{ and } \mathbb{P}(X = a) = 1.$$

Distribution function :

$$F_X(x) = 1_{[a,+\infty[},\ x \in \mathbb{R}.$$

Characteristic function :

$$\Phi_X(u) = e^{iau},\ t \in \mathbb{R}.$$

Moment generating function :

$$\varphi_X(u) = e^{au},\ t \in \mathbb{R}.$$

Moments of order $k \geq 1$

$$\mathbb{E}X^k = a^k, \mathbb{E}(X - a)^k = 0.$$

A useful remark. A constant random variable is independent from any other random variable defined on the same probability space. Indeed let $X = a$ and Y be another any other random variable defined on the same probability space. The joint characteristic function of (X, Y) is given by

$$\begin{aligned}\Phi_{(X,Y)}(u, v) &= \mathbb{E}\exp(iXu + iYv) = \mathbb{E}\biggl(\exp(iau)\exp(iYv)\biggr)\\ &= \exp(iau)\mathbb{E}\exp(iYv) = \Phi_X(u)\Phi_Y(v),\end{aligned}$$

for any $(u, v) \in \mathbb{R}^2$. By Theorem 7 in Chapter 2, X and Y are independent.

(2) Uniform Random variable on $\{1, 2, ..., n\}$, $n \geq 1$.

$X \sim \mathcal{U}(1, 2, ..., n)$ takes each value in $\{1, 2, ..., n\}$ with the same probability.

Discrete probability density function on $\mathcal{V}_X = \{1, 2, ..., n\}$:

$$\mathbb{P}(X = k) = 1/n,\ k \in \{1, ..., n\}$$

Distribution function :

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 1, \\ \frac{i-1}{n} & \text{if} \quad \frac{i-1}{n} \leq x < \frac{i}{n},\ 1 \leq i \leq n, \\ 1 & \text{if} \quad x \geq n. \end{cases}$$

Characteristic function :

$$\Phi_X(u) = \frac{1}{n}\sum_{j=1}^{n} e^{iju},\ u \in \mathbb{R}.$$

Moments of order $k \geq 1$:

$$\mathbb{E}X^k = \frac{1}{n}\sum_{j=1}^{n} j^k.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = \frac{n+1}{2},\ \mathbb{V}ar(X) = \frac{(n-1)(n+1)(4n+3)}{12}.$$

(3) Bernoulli Random Variable with parameter $0 < p < 1$.

$X \sim \mathcal{B}(p)$ takes two values : 1 (Success) and 0 (failure).

Discrete probability density function on $\mathcal{V}_X = \{0, 1\}$:

$$\mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0).$$

Distribution function :

$$F(x) = 0 \times 1_{]-\infty,0[} + p \times 1_{[0,1[} + 1_{[1,+\infty[},\ x \in \mathbb{R}.$$

Characteristic function :

$$\Phi_X(u) = q + pe^{iu},\ u \in \mathbb{R}.$$

Moments of order $k \geq 1$:

$$\mathbb{E}X^k = p.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = p, \ \mathbb{V}ar(X) = pq.$$

(4) Binomial random variable with parameters $0 < p < 1$ **and** $n \geq 1$.

$X \sim \mathcal{B}(n,p)$ takes its values in $\{0, 1, ..., n\}$.

Discrete probability density function on $\mathcal{V}_X = \{0, 1, ..., n\}$:

$$\mathbb{P}(X = k) = C_n^k \ p^k (1-p)^{n-k}, \ k = 0, ..., n.$$

Characteristic function. Since X is the sum of n independent Bernoulli $\mathcal{B}(p)$ random variables, Point (b) and Theorem 5 and the value of the characteristic function of a Bernoulli random variable, yield

$$\Phi_X(u) = (q + pe^{iu})^n, \ u \in \mathbb{R}.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = np, \text{ and } Var(X) = np(1-p).$$

The above parameters are computed by still using the decomposition of Binomial random variable by into a sum of independent Bernoulli random variables.

(5) Geometric Random Variable with parameter $0 < p < 1$.

$X \sim \mathcal{G}(p)$ takes its values in $\mathbb{N}$.

Discrete probability density function on $\mathcal{V}_X = \mathbb{N}$:

$$\mathbb{P}(X = k) = p(1-p)^k, \ k \in \mathbb{N}.$$

Characteristic function :

$$\Phi_X(u) = p/(1 - qe^{iu}), \ u \in \mathbb{R}.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = q/p, \ Var(X) = q/p^2.$$

(6) Negative Binomial Random Variable with parameters $r \geq 1$ **and** $0 < p < 1$.

$X_r \sim \mathcal{B}N(r,p)$ takes the values in $\{r, r+1, ...\}$.

Discrete probability density function on $\mathcal{V}_X = \{r, r+1, ...\}$:

$$\mathbb{P}(X = k) = C_{k-1}^{r-1} p^k (1-p)^{r-k}, \quad k \geq r.$$

Characteristic function. Since X_r is the sum of r independent Geometric $\mathcal{G}(p)$ random variables, Theorem and the value of the characteristic function of a Bernoulli random variable, yield

$$\Phi_X(u) = \left\{pe^{iu}/(1-qe^{iu}\right\}^r, \ u < -\log(1-p).$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = rq/p, \ Var(X) = rq/p^2.$$

(7) Poisson Random variable of parameter $\lambda > 0$.

$X \sim \mathcal{P}(\lambda)$ takes its values in $\mathbb{N}$.

Discrete probability density function on $\mathcal{V}_X = \mathbb{N}$:

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \ k \geq 0.$$

Characteristic function

$$\Phi_X(u) = \exp(\lambda(e^{iu} - 1)), \ u \in \mathbb{R}.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = Var(X) = \lambda.$$

(8) Hyper-geometric Random Variable.

$X \sim \mathcal{H}(N, \theta, n)$ or $H(N, M, n)$, $1 \leq n \leq N$, $0 < \theta < 1$, $\theta = M/N$, takes its values in $\{0, 1, ..., min(n, M)\}$.

Discrete probability density function on $\mathcal{V}_X = \{0, 1, ..., min(n, M)\}$:

$$(X = k) = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n}, \; k = 0, ..., \min(n, M).$$

Characteristic function of no use.

Mathematical expectation and variance :

$$\mathbb{E}(X) = rM/n, \text{ and } V(X) = rM(n-M)(n-r)/\{n^2(n-1)\}.$$

(9) Logarithmic Random Variable.

$X \sim Log(p)$ takes its values in $\{1, 2, ...\}$.

Discrete probability density function on $\mathcal{V}_X = \{1, 2, ...\}$:

$$\mathbb{P}(X = k) = -qk/(k \log p), \; k \geq 1.$$

Characteristic function :

$$\Phi_X(u) = \log(1 - qe^{iu})/\log(p), \; u \in \mathbb{R}.$$

Moment Generating function :

$$\Phi_X(u) = \log(1 - qe^{u})/\log(p), \; u < -\log(1-p).$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = -q/(p\log(p)), V(X) = -q(q + \log(p))/(p\log(p)).$$

2. Absolutely Continuous Probability Laws

For each random variable X, the support $\mathcal{V}_X$, the probability density function with respect to the Lebesgue measure, the characteristic function and/or the moment generating function, the moments are given. By definition, the support $\mathcal{V}_X$ of X is given by

$$\mathcal{V}_X = \overline{\{x \in \mathbb{R}, f_X(x) \neq 0\}}$$

We also have

$$\mathbb{P}(X \in \mathcal{V}_X) = 1.$$

For any real-valued random variable, we may define

$$lep(F) = \inf\{x, F(x) > 0\}$$

and

$$uep(F) = \sup\{x, F(x) < 1\}.$$

where $lep(F)$ and $uep(F)$ respectively stand for *lower end-point of* F and *upper end-point of* F. As a result we have

$$X \in [lep(F), uep(F)], \ \ a.e.$$

The first examples given without computations are done in Lo (2017b).

(1) Continuous uniform Random variable on a bounded compact set.

Let a and b be two real numbers such that $a < b$. $X \sim \mathcal{U}(a, b)$.

Domain : $\mathcal{V}_X = [a, b]$.

Absolutely continuous probability density function on $\mathcal{V}_X = [a, b]$:

$$f_X(x) = \frac{1}{b-a} 1_{[a,b]}(x), \ \ x \in \mathbb{R}.$$

Distribution function :

$$F_X(x) = \begin{cases} 1 \ \ if \ \ x \geq b, \\ (x-a)/(b-a) \ \ if \ \ a \leq x \leq b, \\ 0 \ \ if \ \ x \leq a. \end{cases}$$

Characteristic function :

$$\Phi_X(u) = \frac{e^{ibu} - e^{iau}}{iu(b-a)}, u \in \mathbb{R}.$$

Moments of order $k \geq 1$:

$$\mathbb{E}X^k = \frac{b^{k+1} - b^{k+1}}{(k+1)(b-a)}.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = (a+b)/2, \ et \ Var(X) = (b-a)^2/12.$$

(2) Exponential Random Variable of parameter $b > 0$.

$X \sim \mathcal{E}(b)$ is supported on $\mathbb{R}_+$.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}_+$:

$$f_X(x) = be^{-bx}1_{(x \geq 0)}.$$

Distribution function :

$$F_X(x) = (1 - e^{-bx})1_{(x \geq 0)}.$$

Characteristic function :

$$\Phi_X(u) = (1 - iu/b)^{-1},$$

Moment Generating Function :

$$\phi_X(u) = (1 - u/b)^{-1}, \ u < b.$$

Moments of order $k \geq 1$

$$\mathbb{E}(X^k) = \frac{k!}{\lambda^k}.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = 1/\lambda, Var(X) = 1/\lambda^2.$$

(3) Gamma Random variable with Parameter $a > 0$ and $b > 0$.

$X \sim \gamma(a, b)$ is defined ob $\mathbb{R}_+$.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}_+$:

$$f_X(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} 1_{(x \geq 0)}$$

with

$$\Gamma(a) = \int_0^\infty x^{a-1}\ e^{-x}\ dx.$$

Characteristic function :

$$\Phi_X(u) = (1 - iu/b)^{-a}.$$

Moments of order $k \geq 1$:

$$\mathbb{E}(X^k) = \frac{1}{b^k} \prod_{j=0}^{k} a + j.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = a/b, Var(X) = a/b^2.$$

Be careful. Some authors, many of them in North America, take $\gamma(a, 1/b)$ as the *gamma law*. If you read somewhere that $\mathbb{E}(X) = ab$ for $X \sim \gamma(a, b)$, be aware that in our definition we have $X \sim \gamma(a, 1/b)$.

(4) Symmetrized Exponential random variable with $\lambda > 0$.

$X \sim \mathcal{E}_s(\lambda)$ in defined on $\mathbb{R}$.

From the non-negative random variable X , it is always possible to define a symmetrized random variable X_s by considering two independent $\mathcal{E}(\lambda)$-random variables X_1 and X_2 on a same probability space (such a construction is achieved through the Kolmogorov construction method) and by setting $X_s = X_1 - X_2$. Another way to define it is to have an $\mathcal{E}(\lambda)$-random variable X and a $(0,1)$-uniform random variable U independent of X and to set $X_s = -X1_{(U \leq 0.5)} + X1_{(U>0.5)}$. We are going to use the first method. It is clear that X_s is a symmetric random variable. Further if X admits an absolutely continuous *pdf*, X_s has the *pdf*

$$f_{X_x}(x) = \frac{1}{2} f_X(|x|), \ x \in \mathbb{R}.$$

By applying this to the exponential random variable, a Symmetrized Exponential random following $X \sim \mathcal{E}_s(\lambda)$ has the following *pdf*.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}$:

$$f_X(x) = \frac{\lambda}{2} \exp(-\lambda |x|), \ x \in \mathbb{R}.$$

Distribution function :

$$F_X(x) = \frac{1}{2} e^{\lambda x} 1_{(x<0)} + \left(1 - \frac{1}{2} e^{-\lambda x}\right) 1_{(x \geq 0)}, \ x \in \mathbb{R}.$$

Characteristic function : (See Formula 7, Chapter 2, page 98)

$$\Phi_X(u) = \exp(-\lambda |u|), \ u \in \mathbb{R}.$$

Mathematical expectation and variance :

$$\mathbb{E}X_s = 0 \ \ and \ \ \mathbb{V}ar(X_s) = \frac{2}{\lambda^2}.$$

To justify the variance, we may remark that $X_s = (X_1 - 1/\lambda) - (X_2 - 1/\lambda)$, that is, $X_s = (X_1 - \mathbb{E}(X_1)) - (X_2 - \mathbb{E}(X_1 = 2))$ and exploit that X_s is the difference between two independent and centered random variables.

Remark. For $\gamma = 1$, this law holds the name of Laplace random variable.

(5) Beta Random variables of parameter $a > 0$ and $b > 0$.

$X \sim B(a,b)$ is defined on $(0,1)$.

Absolutely continuous probability density function on $\mathcal{V}_X$:

$$f_X(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} 1_{(0,1)}(x),$$

where

$$B(a,b) = \int_1^1 x^{a-1}(1-x)^{b-1}\ dx.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = a/(a+b) \text{ and } Var(X) = ab/[(a+b)^2(a+b+1)].$$

(6) Pareto Random Variable of parameter $a > 0$.

$X \sim Par(a, \alpha)$, with parameters $\alpha > 0$ and $a \geq 0$, is supported by $]a, +\infty]$.

Absolutely continuous probability density function on $\mathcal{V}_X =]a, +\infty[$:

$$f_X(x) = \alpha a^{\alpha} x^{-\alpha-1} 1_{(x>a)}.$$

(7) Cauchy random variable with $\lambda > 0$ and $a \in \mathbb{R}$.

$X \sim C(a, \lambda)$ in defined on $\mathbb{R}$.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}$:

$$f_X(x) = \frac{\lambda}{\pi(\lambda^2 + (x-a)^2)},\ x \in \mathbb{R}.$$

Distribution function :

$$F_X(x) = \frac{1}{\pi}\left(\arctan\left(\frac{x-a}{\lambda}\right) - \frac{\pi}{2}\right),\ \in \mathbb{R}.$$

Characteristic function :

$$\Phi_X(u) = \exp(iua - \lambda|u|),\ u \in \mathbb{R}.$$

The mathematical expectation does not exist.

Proof. We have to prove that for $a = 0$ and $\lambda = 1$, Formula 2 gives a *pdf*. Indeed, using that the primitive of $(1 + x^2)$ is $\arctan x$, the inverse of the tangent function $\tan x$, we have

$$\int_{-\infty}^{+\infty} \frac{dx}{\pi(1+x^2)} = \frac{1}{\pi}\left[\arctan x\right]_{-\infty}^{+\infty} = 1.$$

Next, setting $X = \lambda Z + a$, where Z follows a $C(0, 1)$ law leads to the general case in 2 by differentiating $F_Z(x) = F_X((x - a)/\lambda)$, $x \in \mathbb{R}$.

The expression of the characteristic function of a *rrv* Z following a standard Cauchy law is given by Formula 7.4 (Chapter 2, page 98). By the transform $X = \lambda Z + a$, we have the general characteristic function of a Cauchy distribution.

Finally, we have for $a = 0$ and $\lambda = 1$,

$$\mathbb{E}(X^+) = \int_0^{+\infty} \frac{x}{\pi(1+x^2)} dx = +\infty,$$

and

$$\mathbb{E}(X^-) = \int_{-\infty}^{0} \frac{x}{\pi(1+x^2)} dx = -\infty,$$

and then $\mathbb{E}(X)$ is not defined. Concerning that point, we recommend to go back to the remark concerning the caution to take while using the improper Riemann integration at the place of the Lebesgue integral (See Point (b5) in Section 5 in Chapter 2, page 69).

(8) Logistic Random Variable with parameters $a \in \mathbb{R}$ and $b > 0$.

$X \sim \ell(a, b)$ is supported by the whole real line.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}$:

$$f_X(x) = b^{-1}e^{-(x-a)/b}/(1 + e^{-(x-a)/b}),\ x \in \mathbb{R}.$$

Characteristic function :

$$\Phi_X(u) = e^{iau}\pi b\ cosec(i\pi bu).$$

Mathematical expectation and variance :

$$E(X) = a; V(X) = b^2\pi^2/3.$$

(9) Weibull Random Variable with parameters $a > 0$ **and** $b > 0$.

$X \sim W(a, b)$ is supported by $\mathbb{R}_+$.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}_+$:

$$f_X(x) = ab\ x^{b-1}\exp(-ax^{-b})1_{(x>0)}.$$

Characteristic function :

$$\Phi_X(u) = a^{-iu/b}\Gamma(1 + iu/b),\ u \in \mathbb{R}.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = (1/a)^{1/b}\Gamma(1 + 1/b); V(X) = a^{-2/b}(\Gamma(1 + 2/b) - \Gamma(1 + 1/b)).$$

(10) Gumbel Random Variable $a \in \mathbb{R}$ **and** $b > 0$.

$X \sim Gu(a, b)$ is supported by the whole line $\mathbb{R}$.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}$:

$$f_X(x) = (u/b)e^{-u},\ with\ u = e^{-(x-a)/b}.$$

Characteristic function :

$$\Phi_X(u) = e^{iua}\Gamma(1 - ibu),\ u \in \mathbb{R}.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = a + \gamma b.$$

where $\gamma =$ is the Euler's number and

$$Var(X) = \pi^2 b^2/2.$$

(11) Double-exponential Random Variable with parameter $b > 0$.

See Point (4) above.

$X \sim \mathcal{E}_d(b)$ is defined on the whole real line.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}$:

$$f_X(x) = \frac{b}{2}\exp(-b\,|x|), x \in \mathbb{R}.$$

Characteristic function :

$$\Phi_X(u) = (1 + (u/b)^2)^{-1}, \quad u \in \mathbb{R}.$$

Moments of order $k \geq 1$:

Mathematical expectation and variance :

$$\mathbb{E}(X) = 0, \text{ and } Var(X) = 2b^{-2}.$$

(12) Gaussian Random Variable with parameters $m \in \mathbb{R}$ **and** $\sigma > 0$. $X \sim \mathcal{N}(m, \sigma^2)$ is supported by the whole real line $\mathcal{V}_X = \mathbb{R}$.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}$:

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\exp(-(x-m)^2/\sigma^2), x \in \mathbb{R}.$$

Characteristic function :

$$\Phi_X(u) = e^{-um}\exp(-\sigma^2 u^2/2).$$

Moments of order $k \geq 1$.

$$\mathbb{X}\left(\frac{X-m}{\sigma^2}\right) = \frac{2^k\ k!}{(2k)}.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = m,\, Var(X) = \sigma^2.$$

Because of its importance in the history of Probability Theory, as explained by its name of *normal* probability law, we will devote a special study to it in Chapter 4.

(13) Chi-square Probability law of parameters $d \geq 1$.

$X \sim \chi_d^2$ is supported by $\mathcal{V}_X = \mathbb{R}_+$.

Definition A Chi-square Probability law of $d \geq 1$ degrees of freedom is simply a Gamma law of parameters $a = d/2$ and $b = 1/2$, that is

$$\chi_d^2 = \gamma(d/2, 1/2).$$

By reporting the results of γ-laws, we have the following facts.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}_+$:

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}_+$:

$$f_X(x) = \frac{1}{2^{d/2}\Gamma(d/2)} x^{\frac{d}{2}-1} e^{-\frac{x}{2}} 1_{(x\geq 0)}.$$

Characteristic function :

$$\Phi_X(u) = (1 - i2u)^{-d/2}.$$

Moments of order $k \geq 1$.

$$\mathbb{E}(X^k) = \frac{1}{2^{-k}} \prod_{j=0}^{k} \left(\frac{d}{2} + j\right).$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = d,\ Var(X) = 2d.$$

Important properties. Chi-square distributions are generated from Gaussian random variables as follows.

Fact 1. If Z follows a standard Gaussian probability law, Z^2 follows a Chi-square law of one degree of freedom :

$$Z \sim \mathcal{N}(0,1) \Rightarrow Z^2 \sim \chi_1^2.$$

Proof. Suppose that $Z \sim \mathcal{N}(0,1)$ and put $X = Z^2$. It is clear that the domain of Y is $\mathcal{V}_X = \mathbb{R}_+$. For any $y \geq 0$,

$$\begin{aligned} F_X(x) &= \mathbb{P}(Z^2 \leq x) \\ &= \mathbb{P}(|Z| \leq \sqrt{x}) \\ &= \mathbb{P}(Z \in]-\infty, \sqrt{x}] \setminus]-\infty, \sqrt{x}[) \\ &= \mathbb{P}(Z \leq \sqrt{x}) - \mathbb{P}(Z \leq -\sqrt{x}). \end{aligned}$$

Remind that Z has an even absolutely continuous *pdf* f_Z. This implies that $\mathbb{P}(Z = t) = 0$ for any $t \in \mathbb{R}$ and we get for any $y \geq 0$,

$$F_Y(x) = F_Z(\sqrt{x}) - F_Z(-\sqrt{x}).$$

By differentiating by x, we get the absolutely continuous *pdf* of X for any any $x \in \mathcal{V}_X$

$$f_X(x) = \frac{1}{2\sqrt{x}}\Big(f_Z(\sqrt{x}) + f_Z(-\sqrt{x})\Big) = \frac{1}{\sqrt{x}} f_Z(\sqrt{x}),$$

which leads to

$$f_X(x) = \frac{\left(\frac{1}{2}\right)^{1/2}}{\sqrt{\pi}} x^{1-1/2} \exp(-x/2), \quad x \in \mathbb{R}_+.$$

By comparing with the absolutely continuous *pdf* $f_{\chi^1_2}$ of a Chi-square probability law, we see that $f_{\chi^1_2}$ and f_X are two absolutely continuous **pdf**'s with the same support $\mathcal{V}$ and a common variable part

$$h(x) = x^{1-1/2} \exp(-x/2), \quad x \in \mathcal{V}.$$

By the Easy Stuff remark in Section 9 in Chapter Section 9, it follows that they are equal and by the way, we get the stunning equality

$$\Gamma(1/2) = \sqrt{\pi}.$$

Fact 2. Let $d \geq 2$. The convolution product of d Chi-square law of one degree of freedom is a Chi-square law probability of d degrees of freedom. In particular, if X_1, ..., X_d are d independent real-valued random variables, defined on the same probability space, identically following a Chi-square law of one degree of freedom, we have

$$X_1^2 + \cdots + X_d^2 = \sum_{1 \leq i \leq d} X_i^2 \sim \chi_d^2.$$

Indeed, if X_1, ..., X_d are independent and identically follow a Chi-square law of one degree of freedom, the probability law $X_1^2 + \cdots + X_d^2$ is characterized by its characteristic function

$$\varphi_{X_1^2+\cdots+X_d^2}(t) = \prod_{j=1}^{d} = \varphi_{X_j}(t) = (1-2it)^{d/2},$$

which establishes that $X_1^2 + \cdots + X_d^2$ follows a χ_d^2 law.

(14) Around the Normal Variance Mixture class of random variables

We are introducing some facts on this class of random variables which are important tools in financial data statistical studies. We only provide some of their simple features, not dwelling in their deep relations. It is expected to treat these random variables in completion of Chapter later.

(a). A normal variance mixture is defined as follows :

$$X = \mu + \sigma\sqrt{W}Z, \ (NMV)$$

where Z is a standard random variable, W is a positive random variable defined on the same space as Z and independent of Z, μ is a real number (the mean of X) and σ is a positive random variable. Hence we have

$$\mathbb{E}(X) = \mu + \sigma\mathbb{E}(\sqrt{W})\mathbb{E}(Z) = \mu$$

and

$$\mathbb{V}(X) = \sigma^2 \mathbb{E}(W^2)\mathbb{E}(Z^2) = \sigma^2 \mathbb{E}(W^2).$$

The following distributions of W are generally used.

(b). The inverse gamma law $W = 1/Y \sim Ig(\alpha, \beta)$, where $Y \sim \gamma(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}_+$:

$$f_W(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp(-\beta/x), \; x > 0.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = \frac{\beta}{\alpha - 1} \; for \; \alpha > 1; \mathbb{V}ar(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} \; for \; \alpha > 2.$$

(c). The Generalized Inverse Gaussian (GIG) law : $W \sim Gig(a, b, c)$, $(a, b, c) \in \mathbb{R}^3_+$.

Parameters domains :

$$\begin{array}{llll} b > 0 & and \;\; c \geq 0 & if & a < 0 \\ b > 0 & and \;\; c > 0 & if & a = 0 \\ b \geq 0 & and \;\; c > 0 & if & a > 0. \end{array}$$

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}_+$:

$$f_W(x) = \frac{b^{-a}(bc)^a}{2K_a\left((bc)^{1/2}\right)} x^{a-1} e^{-(cx+b/x)/2}, \; x \in \mathbb{R}.$$

$$K_a\left((bc)^{1/2}\right) = \frac{b^{-a}(bc)^a}{2} \int_0^{+\infty} x^{a-1} e^{-(cx+b/x)/2} dx.$$

This function, called a *modified Bessel function*, is not directly defined in a simple argument, but on a composite argument $(bc)^{1/2}$ and one should pay a particular attention to the simultaneous domain of the parameters (a, b, c).

Mathematical expectation and variance :

$$\mathbb{E}(X) = \frac{\beta}{\alpha - 1} \ for \ \alpha > 1; \mathbb{V}ar(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} \ for \ \alpha > 2.$$

(d). Student distribution of $\nu \geq 1$ degrees of freedom : $X \sim t(\nu)$.

If in Formula (NMV), we take W as inverse Gamma random variable $Ig(\nu/2, \nu/2)$, where $\nu \geq 1$ is an integer, the *pdf* of X becomes :

for $x \in \mathcal{V}_X = \mathbb{R}$:

$$f(x) = \frac{\Gamma((\nu+1)/2)}{\sigma\Gamma(\nu/2)(\nu\pi)^{1/2}} \left(1 + \frac{(x-\mu)^2/2}{\nu}\right)^{-(\nu+1)/2}.$$

(e). Symmetric Generalized Hyperbolic distribution : $X \sim SGH(\mu, a, b, c)$.

Parameters : $\mu \in \mathbb{R}$, a, b and c given in the Gig law presentation.

If, in Formula (NMV), we take W as the generalized inverse Gaussian random variable $Gig(a, b, c)$, the *pdf* of X becomes :

for $x \in \mathcal{V}_X = \mathbb{R}$:

$$f(x) = \frac{(ab)^{-a/2}c^{1/2}}{\sigma(2\pi)^{1/2}K_a\left((bc)^{1/2}\right)} \frac{K_{a-1/2}\left(\left(b + c(x-mu)^2/\sigma\right)^{1/2}\right)}{\left(b + c(x-mu)^2/\sigma\right)^{1/4-a/2}}.$$

(f). Generalized Hyperbolic distribution : $X \sim GH(\mu, a, b, c)$.

The latter probability law is a particular case of the following model :

$$X = \mu + \gamma W + \sigma\sqrt{W}Z, \ (GNMV)$$

for $\gamma = 0$. If γ is an arbitrary real number and we take W as a generalized inverse Gaussian random variable $Gig(a, b, c)$, the *pdf* of X is :

for $x \in \mathcal{V}_X = \mathbb{R}$:

$$f(x) = c\frac{\exp(\gamma(x-\mu)/\sigma)K_{a-1/2}\left((b+\sigma^{-1}(x-mu)^2(c+\gamma^2/\sigma))^{1/2}\right)}{(b+\sigma^{-1}(x-mu)^2(c+\gamma^2/\sigma))^{1/4-a/2}},$$

where

$$c = \frac{(ab)^{-a/2}c^c(c+\gamma^2/\sigma)^{1/2-a}}{\sigma(2\pi)^{1/2}K_a\left((bc)^{1/2}\right)}.$$

Comments. This part (14) was only an introduction to an interesting modern and broad topic in Statistical studies in Finance. The multivariate version has also been developed.

(15) Probabiliy Laws of the Gaussian sample.

In Mathematical Statistics, the study Gaussian samples holds a special place, at least at the beginning of the exposure of the theory. The following probability laws play the major roles.

(a) The Chi-square probability law of $n \geq 1$ degrees of freedom.

$X \sim \chi^n_2$.

This law has been introduced in Point (13) above.

(b) The Student probability law of $n \geq 1$ degrees of freedom.

$X \sim t(n)$ is defined on the whole real line.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}$:

$$f_X(x) = \frac{\Gamma((n+1)/2)}{(n\pi)^{1/2}\Gamma(n/2)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}$$

Characteristic function. No explicit form.

Moments of order $k \geq 1$:

Mathematical expectation and variance :

$$\mathbb{E}(X) = 0, \text{ and } Var(X) = \frac{n}{n-2}), \ n \geq 3.$$

(c) The Fisher probability law of degrees of of freedom $n \geq 1$ *and* $m \geq 1$.

$X \sim F(n,m))$ is defined on the positive real line.

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}_+$:

$$f_X(x) = \frac{n^{n/2}m^{m/2}\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \frac{x^{n/2-1}}{(m+nx)^{(n+m)/2}}.$$

Characteristic function. No explicit form.

Mathematical expectation and variance :

$$\mathbb{E}(X) = \frac{m}{m-2}, \ m \geq 3 \text{ and } Var(X) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}, \ m \geq 5.$$

We take this opportunity to propose an exercise which illustrate the change of variable formula given in page 92 and which allows to find the just given laws.

EXERCISE 1. *Let (X,Y) be a 2-random vector with pdf $f_{(X,Y)}$, on its support D with respect to the Lebesgue measure on $\mathbb{R}^2$. Consider the following transform*

$$(x,y) \mapsto h(x,y) = (x, x+y) \in \Delta,$$

that is :

$$\begin{cases} U = X \\ V = X + Y \end{cases}$$

(a) Find the law of (U,V) and their marginals law.

(b) Precise the pdf of V if X and Y are independent.

(c) Application : Let $X \sim \gamma(\alpha, b)$ and $Y \sim \gamma(\beta, b)$. Show that $V = X + Y \sim \gamma(\alpha+\beta, b)$.

EXERCISE 2. *Let Let (X,Y) be a 2-random vector with pdf $f_{(X,Y)}$ with respect to the Lebesgue measure on $\mathbb{R}^2$. Consider the following transfrom*

$$(x,y) \mapsto h(x,y) = (x/y, y),$$

that is

$$(X,Y) \mapsto (U,V) = (X/Y, Y).$$

(a) Apply the general change of variable formula to write the pdf of $(X/Y, Y)$ and deduce the marginal pdf of $U = X/Y$.

(b) Precise it for X and Y independent.

(c) In what follows, X and Y are independent. Precise the pdf of U when X and Y are both standard Gaussian random variables. Identify the found probability law.

(d) Let X be a standard Gaussian random variable and $Y = Z^{1/2}$ the square-root of a χ^2_n random variable with $n \geq 1$. Precise the pdf of $U = X/\sqrt{Z}$. Begin to give the pdf of Y by using its cdf.

Deduce from this the probability law of

$$t(n) = \frac{\sqrt{n}X}{Y} \equiv \frac{\mathcal{N}(0,1)}{\sqrt{\chi_2^n/n}}$$

where the term after the sign $\equiv$ is a rephrasing of a ratio of two independent random variables : a $\mathcal{N}(0,1)$ random variable by the square root of a chi-square random variable divided by its number of freedom degrees.

Conclude that a t(n)-random variable has the same law as the ratio of two independent random variables : a $\mathcal{N}(0,1)$ random variable by by square root of a chi-square random variable divided by its number of freedom degrees.

(e) Let X and Y be two independent random variable following Chi-square laws of respective number of freedom degrees $n \geq 1$ and $m \geq 1$. Precise the pdf of $U = X/Y$.

Deduce from this the probability law of

$$F_{n,m} = \left(\frac{m}{n}\right)\frac{X}{Y} \equiv \frac{\chi_2^n/n}{\chi_2^m/m},$$

where the term after $\equiv$ is a rephrasing of a ratio of two independent random variables : a Chi-square random variable of number of freedom degrees $n \geq 1$ by a Chi-square random variable of number of freedom degrees $m \geq 1$.

Conclude that a Fisher random variable with numbers of freedom degrees $n \geq 1$ and $m \geq 1$ has the same probability law as a ratio of two independent random variables : a Chi-square random variable of number of freedom degrees $n \geq 1$ by a Chi-square random variable of number of freedom degrees $m \geq 1$.

Solutions of Exercise 1. We have the transformation :

$$\left\{\begin{array}{c} u = x \\ v = x + y \end{array}\right. \Leftrightarrow \left\{\begin{array}{c} x = u \\ y = -u + v \end{array}\right.$$

The Jacobian matrix is :

$$J_{(u,v)} = \left(\begin{array}{cc} 1 & 0 \\ -1 & 1 \end{array}\right)$$

with determinantdet $\left(J_{(u,v)}\right) = 1$. By the Change of variable formula, we have

$$f_{(U,V)}(u,v) = f_{(X,Y)}(u, -u+v)\, 1_{\Delta}(u,v).$$

The marginal laws are

$$f_U(u) = \int_{D_v} f_{(U,V)}(u,v)\,dv?\ U \in \mathcal{V}_U$$

and

$$\begin{aligned} f_V(v) &= \int_{D_u} f_{(U,V)}(u,v)\,du \\ &= \int_{D_U} f_{(X,Y)}(u,-u+v)\,\mathbb{I}_{D(U,V)}(u,v)\,du. \end{aligned}$$

Question (b) We have

$$f_V(v) = f_{X+Y}(v) = \int f_X(u)\,f_Y(v-u)\,du = \int f_Y(u)\,f_X(v-u)\,du,$$

which is the convolution product between X and Y.

Question (c). We recall that

$$f_X(x) = \frac{b^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-bx} \mathbb{I}_{\mathbb{R}_+}(x)$$

and

$$f_Y(x) = \frac{b^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-bx} \mathbb{I}_{\mathbb{R}_+}(x).$$

We have

$$\begin{aligned} f_{X+Y}(v) &= \frac{b^\alpha}{\Gamma(\alpha)} \frac{b^\beta}{\Gamma(\beta)} \int_0^v u^{\alpha-1} e^{-bu} (v-u)^{\beta-1} e^{-b(v-u)} du \\ &= \frac{b^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \times v^{\alpha+\beta-2} e^{-bv} \int_0^v \left(\frac{u}{v}\right)^{\alpha-1} \left(1-\frac{u}{v}\right)^{\beta-1} du. \end{aligned}$$

By taking the further change of variables $x = u/v$, we get

$$f_{X+Y}(v) = \left[\frac{b^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 (x)^{\alpha-1}(1-x)^{\beta-1}\,dx\right] \times v^{\alpha+\beta-2} e^{-bv} \mathbb{I}_{\mathbb{R}_+}(v).$$

Since $X+Y$ has the same domain and the same variable part of a $\gamma(\alpha+\beta, b)$, they have the same constant and then we have $X+Y \sim$

$\gamma(\alpha+\beta,b)$ and

$$\frac{b^{\alpha+\beta}}{\Gamma\left(\alpha+\beta\right)}=\frac{b^{\alpha+\beta}}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}\times\beta\left(\alpha,\beta\right)$$

with

$$B\left(\alpha,\beta\right)=\int_0^1\left(x\right)^{\alpha-1}\left(1-x\right)^{\beta-1}dx=\frac{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}{\Gamma\left(\alpha+\beta\right)}.$$

□

Solutions of Exercise 2.

Question (a). We have the transformation :

$$\left\{\begin{array}{l} u=\frac{x}{y} \\ v=y \end{array}\right. \Rightarrow \left(\left\{\begin{array}{l} x=uv \\ y=v \end{array}\right.\right)$$

The the Jacobian matrix is

$$J_{(u,v)}=\begin{pmatrix} v & u \\ 1 & 0 \end{pmatrix}$$

with determinant $\det\left(J_{(u,v)}\right)=v$. The *pdf* of (U,V) becomes

$$f_{\left(\frac{X}{Y},Y\right)}\left(u,v\right)=f_{(X,Y)}\left(uv,v\right)|v|\,\mathbb{I}_{D\left(\frac{X}{Y},Y\right)}\left(u,v\right),$$

and the marginal law of $V=X/Y$ is

$$\begin{aligned} f_{\frac{X}{Y}}\left(u\right) &= \int_{D_v} f_{(U,V)}\left(u,v\right)dv \\ &= \int_{D_v} f_{(X,Y)}\left(uv,v\right)|v|\,\mathbb{I}_{D\left(\frac{X}{Y},Y\right)}\left(u,v\right)dv \qquad (1). \end{aligned}$$

Question (b) If X and Y are independent, we have

$$\begin{aligned} f_{\frac{X}{Y}}\left(u\right) &= \int_{D_v} f_{(U,V)}\left(u,v\right)dv \\ &= \int_{(uv\in D_X,v\in D_Y)} f_X\left(uv\right)f_Y\left(v\right)|v|\,dv. \end{aligned}$$

Question (c) Now if X and Y are standard Gaussian random variables, we get

$$\begin{aligned} f_{\frac{X}{Y}}(u) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2v^2}e^{-\frac{1}{2}v^2}\,|v|\,dv \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(u^2+1\right)v^2}\,|v|\,dv \\ &= \frac{1}{\pi}\int_{0}^{\infty} ve^{-\frac{1}{2}\left(u^2+1\right)v^2}dv \\ &= \frac{1}{\pi}\left[-\frac{e^{-\frac{1}{2}\left(u^2+1\right)v^2}}{u^2+1}\right]_0^{\infty} \\ &= \frac{1}{\pi\left(u^2+1\right)}, \end{aligned}$$

which is the standard Cauchy probability law.

Question (d). The *pdf* of Y is easily derived from the relation

$$\forall y \geq 0,\ \ F_Y(y) = F_Z(y^2)$$

which, after differentiation, gives

$$f_Y\left(y\right) = 2yf_{Y_1}\left(y^2\right) = \frac{2\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}y^{n-1}e^{-\frac{1}{2}y^2}1_{\mathbb{R}_+}\left(y\right).$$

From there, we have

$$\begin{aligned} f_{\frac{X}{Y}}(u) &= \int f_X\left(uv\right)f_Y\left(v\right)|v|\,dv \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}\int_0^{\infty} v^n e^{-\frac{1}{2}u^2v^2}e^{-\frac{v^2}{2}}dv. \end{aligned}$$

Let us set

$$\begin{aligned} A &= \int_0^{\infty} v^n e^{-\frac{1}{2}u^2v^2}e^{-\frac{v^2}{2}}dv \\ &= \int_0^{\infty} v^n e^{-\frac{v^2}{2}\left(u^2+1\right)}dv\ , \end{aligned}$$

and make the change of variable $t = \frac{v^2}{2}\left(u^2+1\right)$. Then we have

$$v = \sqrt{\frac{2t}{u^2+1}} = \sqrt{\frac{2}{u^2+1}} t^{\frac{1}{2}}$$

and

$$\begin{aligned}
dv &= \frac{1}{2}\sqrt{\frac{2}{u^2+1}} t^{-\frac{1}{2}} dt && (2.1)\\
&= \frac{1}{2}\sqrt{\frac{2}{u^2+1}} \times \frac{1}{v} \times \sqrt{\frac{2}{u^2+1}} dt && (2.2)\\
&= \frac{dt}{u^2+1} \times \frac{(u^2+1)^{\frac{1}{2}}}{\sqrt{2}} t^{-\frac{1}{2}}. && (2.3)
\end{aligned}$$

Next, we have

$$\begin{aligned}
A &= \left(\frac{2}{u^2+1}\right)^{\frac{n}{2}} \frac{1}{\left[2\left(u^2+1\right)\right]^{\frac{1}{2}}} \int_0^{\infty} t^{\frac{n+1}{2}-1} e^{-t} dt \\
&= \frac{2^{\frac{n}{2}}}{\sqrt{2}\left(u^2+1\right)^{\frac{n}{2}+\frac{1}{2}}} \Gamma\left(\frac{n+1}{2}\right),
\end{aligned}$$

which leads to

$$\begin{aligned}
f_{\frac{X}{Y}}(u) &= \frac{1}{\sqrt{2\pi}} \frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} \times \frac{2^{\frac{n}{2}}}{\sqrt{2}\left(u^2+1\right)^{\frac{n}{2}+\frac{1}{2}}} \Gamma\left(\frac{n+1}{2}\right) \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)} \left(u^2+1\right)^{-\frac{n+1}{2}}
\end{aligned}$$

Finally, by taking $W = \sqrt{n}U = \sqrt{n}X/Y$, the pdf of W is

$$f_{\frac{\sqrt{n}X}{Y}}(u) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{u^2}{n}\right)^{(n+1)/2}.$$

Question (e). Using the right expressions for the Chi-square random variable leads to

$$\begin{aligned}
f_{\frac{X}{Y}}(u) &= \int f_X(uv)\, f_Y(v)\, |v|\, dv \\
&= \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}\left(\frac{1}{2}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\int_0^\infty (uv)^{\frac{n}{2}-1}\, e^{-\frac{1}{2}uv} v^{\frac{m}{2}-1} e^{-\frac{1}{2}v} dv \\
&= \frac{u^{\frac{n}{2}-1}}{2^{\frac{n+m}{2}}\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\int_0^\infty v^{\frac{n+m}{2}-1} e^{-\frac{1}{2}v(u+1)} dv \\
&= \frac{u^{\frac{n}{2}-1}}{2^{\frac{n+m}{2}}\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \times \frac{\Gamma\left(\frac{n+m}{2}\right)}{(u+1)^{\frac{n+m}{2}}} \\
&= \frac{\Gamma\left(\frac{n+m}{2}\right)}{2^{\frac{n+m}{2}}\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \times u^{\frac{n}{2}-1} \times (u+1)^{-\left(\frac{n+m}{2}\right)}.
\end{aligned}$$

Now using the general rule $f_{aZ}(t) = |a|^{-1} f_X(t/a)$ gives the *pdf*

$$f_X(u) = \frac{n^{n/2} m^{m/2} \Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \frac{u^{n/2-1}}{(m+nux)^{(n+m)/2}} 1_{(u \geq 0)},$$

which is the *pdf* of a Fisher $F_{n,m}$ random variable.

CHAPTER 4

An Introduction to Gauss Random Measures

This chapter focuses of Gaussian probability measures on $\mathbb{R}$ first and next on $\mathbb{R}^d$, $d \geq 2$ exclusively. This is explained by the role of such probability laws in the history of Probability Theory and its presence in a great variety of sub-fields of Mathematics and in a considerable number of Science domains. Knowing that law and its fundamental properties is mandatory.

1. Gauss Probability Laws on $\mathbb{R}$

(A) Standard Gauss Gauss Probability Law.

We already encounter the function

$$f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2), \ x \in \mathbb{R}. \tag{1.1}$$

and we proved that $\int_{\mathbb{R}} f_{0,1}(x) \ dx = 1$. We remark that f is locally bounded and locally Riemann integrable (LLBRI). So, we may equivalently consider the Riemann integral of $f_{0,1}$ or its Lebesgue integral. Without express notification, we will use Riemann integrals as long as we stay in the case where these Riemann integrals are Lebesgue's one.

DEFINITION 7. *A random variable $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ is said to follow a standard normal or standard Gaussian probability law, or in other words : X is a standard normal or standard Gaussian random variable if and only if $f_{0,1}$ is the pdf of X, that is the Radon-Nikodym of $\mathbb{P}_X$ with respect to the Lebesgue measure. Its strict support is the whole real line $\mathbb{R}$.*

The historical derivation of such *pdf* in the earlier Wworks of *de* Moivre, Laplace and Gauss (1732 - 1801) is stated in Loève (1997) and in Lo (2017b) of this series.

The main properties of a standard normal random variable are the following.

THEOREM 9. *If X is a standard normal random variable, then :*

(1) $\mathbb{E}(X) = 0$ *and* $\mathbb{V}ar(X) = 1$.

(2) X has finite moments of all orders and :

$$\mathbb{E}X^{2k+1} = 0 \text{ and } \mathbb{E}X^{2k} = \frac{(2k)!}{2^k k!}, \quad k \geq 1,$$

and, in particular, its kurtosis parameter K_X satisfies

$$K_X = \frac{\mathbb{E}(X^4)}{\mathbb{E}(X^2)^2} = 3.$$

(3) Its mgf is

$$\varphi(u) = \exp(u^2/2), \; u \in \mathbb{R}$$

and its characteristic function is

$$\Phi(u) = \exp(-u^2/2), \; u \in \mathbb{R}.$$

(5) Its cdf

$$G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-u^2/2) \; du, \; u \in \mathbb{R}$$

admits the approximation, for $x > 1$,

$$C \left\{ \frac{1}{x} - \frac{1}{x^2} \right\} e^{-x^2/2} \leq 1 - G(x) \leq \frac{Ce^{-x^2/2}}{x},$$

where $C = 1/\sqrt{2\pi}$.

(6) The quantile function $G^{-1}(1-s)$ is expanded as $s \downarrow 0$, according to

$$\phi^{-1}(1-s)$$
$$= \left\{ (2\log(1/s))^{1/2} - \frac{\log 4\pi + \log\log(1/s)}{2(2\log(1/s))^{1/2}} + O((\log\log(1/s)^2(\log 1/s)^{-1/2})) \right\}.$$

and the derivative of $G^{-1}(1-s)$ is, as $s \downarrow 0$,

$$\left(G^{-1}(1-s)\right)' = (2\log(1/s))^{1/2} - \frac{\log 4\pi + \log\log(1/s)}{2(2\log(1/s))^{1/2}}$$
$$+O((\log\log(1/s)^2(\log 1/s)^{-1/2})).$$

(7) The following property holds. For each $x \in \mathbb{R}$,

$$\lim_{n\rightarrow+\infty} G\left((2\log n)^{1/2}x+(2\log n)^{1/2}-\frac{\log 4\pi + \log\log n}{2(2\log n)^{1/2}}\right)^n = \exp(-e^{-x}).$$

For right now, we are only concerned with the three first Points. The other points are related to the tail $1-G$ of a normal law. We will deal with this in the monograph devoted to extreme value theory.

Proof of Theorem.

Points (1) and (2). Let $k \geq 0$. By using Formula (ACIF) (See Section 5 in Chapter 2, page 67), we have

$$\mathbb{E}X^{2k+1} = \int_{\mathbb{R}} x^{2k+1} f_{0,1}(x)\ d\lambda(x).$$

The function $|x|^{2k+1} f_{0,1}(x)$ is locally bounded and locally Riemann integrable. So, we may use the recommendations in Point (b) in Section 5 in Chapter 2 to get

$$\int_{\mathbb{R}} x^{2k+1} f_{0,1}(x)\ d\lambda(x) = \lim_{n\rightarrow+\infty} \int_{[-n,n]} x^{2k+1} f_{0,1}(x)\ dx.$$

Now, Riemann integration techniques ensure that, for each $n \geq 1$, $\int_{-n}^{n} x^{2k+1} f_{0,1}(x)\ dx = 0$ since the continuous function $x^{2k+1} f_{0,1}(x)$ is odd on the symmetrical interval $[-n, n]$ with respect to zero. By putting together all the previous facts, we have

$$\mathbb{E}X^{2k+1} = 0.$$

For even order moments, we denote $I_k = \mathbb{E}X^{2k}$, $k \geq 0$. We have $I_0 = 1$. For $k \geq 1$, let us use Riemann integrals and integrations by parties. We have

$$\begin{aligned}
I_k &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^{2k}\exp(-x^2/2)\ dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^{2k-1}\Big(x\exp(-x^2/2)\Big)\ dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^{2k-1}d\Big(-\exp(-x^2/2)\Big) \\
&= \frac{1}{\sqrt{2\pi}}\Big[-\exp(-x^2/2)\Big]_{-\infty}^{+\infty} + (2k-1)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^{2k-2}\exp(-x^2/2)\ dx \\
&= (2k-1)I_{k-1}.
\end{aligned}$$

We get by induction that

$$I_k = (2k-1)I_{k-1} = (2k-1)(2k-3)I_{k-3} = \cdots = (2k-1)(2k-3)(2k-5)\cdots 3I_0.$$

Hence

$$I_k = (2k-1)(2k-3)(2k-5)\cdots 3.$$

By multiplying I_k by the even numbers $(2k)(2k-2)\cdots 2 = 2^k k!$ and dividing it as well, we get the results.

(3) By still using Riemann integrals and using Formula (AC01) (See page 68), we have for all $u \in \mathbb{R}$

$$\begin{aligned}
\varphi(u) &= \mathbb{E}(e^{tX}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{ux}\exp(-x^2/2)\ dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \exp\left(\frac{1}{2}\Big(x^2-2tu\Big)\right)\ dx
\end{aligned}$$

By using $x^2 - 2tu = (u-t)^2 - u^2$, we get

$$\varphi(u) = \exp(u^2/2)\left(\int_{-\infty}^{+\infty}\exp(-(x-u)^2/2)\ dx\right)$$

By using the change of variable $y = u-1$, we get that integral between the parentheses is one, and the proof is finished. □

(B) Real Gauss Probability Laws.

Now given a standard random variable Z, m a real number and $\sigma > 0$, the random variable

$$X = \sigma Z + m,$$

has the *cdf*, for $x \in \mathbb{R}$,

$$F_X(x) = \mathbb{P}(\sigma Z + m \leq x) = \mathbb{P}\left(Z \leq \frac{x-m}{\sigma}\right) = F_Z\left(\frac{x-m}{\sigma}\right)$$

which, by differentiating the extreme members, leads to

$$f_{m,\sigma}(x) = \frac{dF_X(x)}{dx} = \frac{1}{\sigma}\frac{dF_Z((x-m)/\sigma)}{dx},$$

for $x \in \mathbb{R}$. Since the functions $f_{m,\sigma}$ and $f_{0,1}$ are bounded and continuous, we may apply the recommendations of Point (b) [Section 5, Chapter 2] to conclude that

$$f_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2}\right), \ x \in \mathbb{R}. \ \ (RG)$$

is the absolute *pdf* of X. By using the properties of expectations and variances and properties characteristic functions, we have :

$$\mathbb{E}X = m \text{ and } \mathbb{V}ar(X) = \sigma^2,$$

$$\varphi_X(u) = \exp(mu + \sigma^2 u^2/2), \ u \in \mathbb{R},$$

and

$$\Phi_X(u) = \exp(imi - \sigma^2 u^2/2), \ u \in \mathbb{R}.$$

Before we conclude, we see that if $\sigma = 0$, $X = m$ and its mgf is $\exp(mu)$, which is of the form $\exp(mu + \sigma^2 u^2/2)$ for $\sigma = 0$. We may conclude as follows.

Definition - Proposition (DEF01).

A real random variable is said to follow a Gaussian or normal probability law, denoted $X \sim \mathcal{N}(m, \sigma^2)$, if and only if its *mgf* is given by

$$\varphi_X(u) = \exp(mu + \sigma^2 u^2/2), \ u \in \mathbb{R}, \ (RGM)$$

or, if and only of its, characteristic function given by

$$\Phi_X(u) = \exp(imu - \sigma^2 u^2/2), \ u \in \mathbb{R}. \ (RGC)$$

If X is not degenerate, that is $\sigma > 0$, its absolutely continuous *pdf* is

$$f_X = \frac{1}{\sigma\sqrt{2\pi}} \exp\biggl(- \frac{(x-m)^2}{2} \biggr), \ x \in \mathbb{R}. \ (RGD)$$

Its first parameters are

$$m = \mathbb{E}X \text{ and } \sigma^2 = \mathbb{E}X^2.$$

$\Diamond$.

(D) Some immediate properties.

(D1) Finite linear combination of independent real Gaussian randoms.

Any linear combination of a finite number $d \geq 2$ of independent random variables $X_1, \cdots, X_d$ with coefficient $\delta_1, \cdots, \delta_d$ follows a normal law. Precisely, if the X_i's are independent and $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$, $1 \leq i \leq d$, if we denote $m^t = (m_1, ..., m_d)$, $\delta^t = (\delta_1, \cdots, \delta_d)$ and $\Sigma = diag(\sigma_1, \cdots, \sigma_d)$, we have

$$\sum_{1 \leq j \leq d} \delta_j X_j \sim \mathcal{N}(m^t\delta, \delta^t \Sigma \delta).$$

To see this, put

$$Y = \sum_{1 \leq j \leq d} \delta_j X_j$$

By the factorization property formula, we have for any $u \in \mathbb{R}^d$,

$$\begin{aligned}
\Phi_Y(u) &= \mathbb{E}\exp\biggl(\sum_{1\leq j\leq d} u\delta_j X_j\biggr) \\
&= \mathbb{E}\exp\biggl(\prod_{1\leq j\leq d} u\delta_j X_j\biggr) \\
&= \prod_{1\leq j\leq d} \mathbb{E}\exp(u\delta_j X_j) \\
&= \prod_{1\leq j\leq d} \exp(im_j\delta_j u - \delta_j^2\sigma_j^2 u^2/2) \\
&= \exp\biggl(i\langle m,\delta\rangle u - \biggl(\delta^t\Sigma\delta\biggr)\frac{u^2}{2}\biggr).
\end{aligned}$$

From there, we may conclude that Y follows a real random vector with the given parameters.

(D2) Towards Gaussian Random Vectors.

Let us remain in the frame of the previous point (D1). Let X be the vector defined by $X^t = (X_1, \cdots, X_d)$ with independent real Gaussian Random variables with the given parameters. We have for any $u \in \mathbb{R}^d$ with $u^t = (u_1, ..., u_d)$,

$$\begin{aligned}
\Phi_X(u) &= \mathbb{E}\exp i\langle u, Z\rangle \\
&= \mathbb{E}\exp\biggl(\sum_{1\leq j\leq d} u_j X_j\biggr) \\
&= \prod_{1\leq j\leq d} \mathbb{E}\exp(uu_j X_j) \\
&= \prod_{1\leq j\leq d} \exp(im_j u_j - \sigma_j^2 u^2/2) \\
&= \exp\left(im^t u - \frac{u^t\Sigma u}{2}\right).
\end{aligned}$$

By also using the same techniques for the *mgf*, we get that for any $u \in \mathbb{R}^d$ with $u^t = (u_1, ..., u_d)$,

$$\varphi_X(u) = \exp\left(\langle m, u\rangle + \frac{u^t\Sigma u}{2}\right).$$

A random vector X whose components are independent and satisfy $X_i \sim \mathcal{N}(m_j, \sigma_j^2)$, $1 \leq j \leq d$ has the *mgf*

$$\varphi_X(u) = \exp\left(\langle m, u\rangle + \frac{u^t \Sigma u}{2}\right). \ (RV01)$$

for any for any $u \in \mathbb{R}^d$, where $m^t = (m_1, ..., m_d)$ and $\Sigma = diag(\sigma_1, \cdots, \sigma_d)$. Besides, we have

$$\mathbb{E}X = m \text{ and } \mathbb{V}ar(X) = \Sigma.$$

This offers us a good transition to the introduction of Gaussian random vectors.

2. Gauss Probability Law on $\mathbb{R}^d$, Random Vectors

(A) Introduction and immediate properties.

In general, the study of random vectors relies so much on quadratic forms and orthogonal matrices topic. We advice the reader to read at least the definitions, theorems and propositions on the aforementioned topic in Section 2 in the Appendix Chapter 10. Each time a property on orthogonal matrices is quoted, it is supposed to be found in the appendix in the aforementioned section.

The above formula (RV01) gave us a lead to the notion of Gaussian random vectors. we have :

Definition 8. *A random variable $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$, $d \geq 1$, is said to follow a d-multivariate Gaussian probability law, or in other words : X is a d-Gaussian random vector if and only its* mgf *is defined by*

$$\varphi_X(u) = \exp\left(\langle m, u\rangle + \frac{u^t \Sigma u}{2}\right), \ u \in \mathbb{R}^2, \ (RV02)$$

where m is a d-vectors or real numbers and Σ is a symmetrical and semi-positive d-matrix or real numbers, and we write $X \sim \mathcal{N}_d(m, \Sigma)$.

By comparing with Formula (RV01), we immediately have :

PROPOSITION 8. *A random vectors with real-valued independent Gaussian components is a Gaussian vector.*

We also have the following properties.

PROPOSITION 9. *X admits the mgf in Formula (REV02), then we have*

$$\mathbb{E}(X) = m \text{ and } \mathbb{V}ar(X) = \Sigma.$$

Proof. We are going to construct a random vector Y which has the *mgf*

$$\exp\left(\langle m, u\rangle + \frac{u^t \Sigma u}{2}\right), \ u \in \mathbb{R}^d. \ (RV03)$$

and next use the characterization of the probability law by the *mgf*. Since Σ is symmetrical and semi-positive, we may find an orthogonal d-matrix T such that

$$T\Sigma T^t = diag(\delta_1, \cdots, \delta_d),$$

where $\delta_1, \cdots, \delta_d$ are non-negative real numbers. By the Kolmogorov Theorem as applied in Point (c5) in Section 5.2 in Chapter 2, we may find a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding a d-random vector Z whose components are centered independent real-valued Gaussian random vectors with respective variances δ_j, $1 \leq j \leq d$. Its follows that Z is Gaussian and hence, by Formula (REV01), we have for $D = diag(\delta_1, \cdots, \delta_d)$, for $u \in \mathbb{R}^d$,

$$\varphi_Z(u) = \exp\left(\frac{u^t D u}{2}\right).$$

Now let us set $Y = m + T^t Z$. We have

$$\begin{aligned}
\varphi_Y(u) &= \mathbb{E}\exp\left((m + T^t Z)^t u\right) = \exp(m^t u)\mathbb{E}\exp\left(Z^t(Tu)\right) \\
&= \exp(m^t u)\mathbb{E}\exp\left(\langle Z, Tu\rangle\right) \\
&= \exp(\langle m, u\rangle)\exp\left(\frac{u^t T^t D T u}{2}\right)
\end{aligned}$$

But, by the properties of orthogonal matrices, we have $T\Sigma T^t = D$, which implies that $T\Sigma T^t DT = \Sigma$. Thus we have

$$\varphi_Y(u) = \exp\left(\langle m, u\rangle + \frac{u^t\Sigma u}{2}\right),\ u \in \mathbb{R}^d.$$

This a direct proof, based on the Kolmogorov construction, that the function in Formula (REV03) is characteristic. An other method would rely on the Bochner Theorem we do not mention here. At the end, we have that $Y \sim \mathcal{N}_d(m, \Sigma)$. By using the properties and expectation vectors and variance-covariance properties seen in Chapter 2, we have

$$\mathbb{E}Y = \mathbb{E}(m + T^t Z) = \mathbb{E}(m) + T^t\mathbb{E}(Z) = m$$

and, since the constant vector m is independent from $T^t Z$,

$$\mathbb{V}ar(Y) = \mathbb{V}ar(m + T^Z) = \mathbb{V}ar(T^Z) = T^t DT = \Sigma.$$

We conclude as follows : for any random variable characterized by its *mdf* given in Formula (REV02), its expectation vector and its variance-covariance matrix are given as above. ■

Important Remark. In the notation $X \sim \mathcal{N}_d(m, \Sigma)$, m and Σ are the respective expectation vector and the variance-covariance matrix of X.

Let us now study other important properties of Gaussian vectors.

(B) - Linear transforms of Gaussian Vectors.

Proposition 10. *The following assertions hold.*

(a) Any finite-dimension linear transform of a Gaussian random vector is a Gaussian random vector.

(b) Any linear combination of the components of a Gaussian random vector is a real Gaussian random variable.

(c) If a random vector $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$, $d \geq 1$, follows a $\mathcal{N}_d(m, \Sigma)$ probability law and if A is a $(k \times d)$-matrix and B a k-vector, $k \geq 1$,

then $AX+B$ follows a $\mathcal{N}_k(Am+B, A\Sigma A^t)$ probability law.

Proof. It is enough to prove Point (c). Suppose that the assumption of that point hold. Thus $Y = AX$ is k-random vector. By Point (a) of Theorem 5 in Section 6 in Chapter 2, we have

$$\Phi_{AX+B}(v) = \exp(B^t v)\Phi_X(A^t v)$$

and combining this with Formula (RV02) gives, for any $v \in \mathbb{R}^k$,

$$\begin{aligned}
\varphi_{AX+B}(v) &= \exp(B^t v)\Phi_X(A^t v)\\
&= \exp(B^t v)\exp\left(m^t(A^t v) + \frac{v^t A\Sigma A^t v}{2}\right)\\
&= \exp\left((B+Am)^t v + \frac{(A^t v)^t A\Sigma(A^t v)}{2}\right)\\
&= \exp\left(\langle B+Am, v\rangle + \frac{v^t\,(A\Sigma A)\,v)}{2}\right).\ \square.
\end{aligned}$$

This proves (c) which is a more precise form of (a). Point (c) is only an application of Point (c) to a $(d\times 1)$-matrix A.

Point (c) provides a new definition of Gaussian vectors **given we already have the definition of a real-valued Gaussian random variable**. We have :

Definition - Proposition.

(a) (DEF01) Any d-random vector, $d \geq 2$, is Gaussian if an only if any linear combination of its components is a real-valued Gaussian random variable.

(b) (DEF02) *Given we already have the definition of a real-valued Gaussian random variable*, a d-random vector, $d \geq 2$, is Gaussian if any linear combination of its components is a real-valued Gaussian random variable. $\Diamond$

Proof of Point (a). Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$, $d \geq 2$, be a random vector such that any linear combination of its components is a real-valued

Gaussian random variable.

First, each component is Gaussian and hence is square integrable and next, by Cauchy-Schwartz inequality, any product of two components is integrable. Hence the expectation vector m and the variance-covariance matrix Σ of X have finite elements. Next, the characteristic function of X is satisfies, for any $u \in \mathbb{R}^d$.

$$\Phi_X(u) = \mathbb{E}\exp\biggl(iu^tX\biggr) = \Phi_{u^tX}(1), \ (RV04)$$

where Φ_{u^tX} is the characteristic of $u^tX = u_1X_1 + \cdots + u_dX_d$ which is supposed to be a real-valued normal random variable with parameters

$$\mathbb{E}(u^tX) = \sum_{1\leq j\leq d} u_jX_j = \langle u, m\rangle$$

and

$$\begin{aligned}\mathbb{V}ar(u^tX) &= \mathbb{V}ar(\sum_{1\leq j\leq d} u_jX_j) \\ &= \sum_{1\leq j\leq d}\sum_{1\leq j\leq d} \mathbb{C}ov(X_i, X_j)u_iu_j \\ &= u^t\Sigma u.\end{aligned}$$

Now using the characteristic function of a $\mathcal{N}(\langle u, m\rangle, u^t\Sigma u)$ allows to conclude. $\square$.

Some consequences.

(a) A sub-vector of a Gaussian Vector is a Gaussian vector since it is a projection, then a finite-dimensional linear transform, of the vector.

(b) As particular cases of Point (a), components of a Gaussian vector are Gaussian.

(c) A vector whose components are **independent** and Gaussian is Gaussian.

(c) **But, in general**, a vector whose components are Gaussian is not necessarily Gaussian. Here is a general, using Sklar (1959)'s Theorem, to construct counter-examples. As stated in Section 8 of Chapter 2, for any random vector X of dimension $d \geq 1$, the *cdf* F_X of X satisfies

$$\forall x \in \mathbb{R}^d,\ F_X(x) = C(F_{X,1}(x), ..., F_{X,d}(x)),$$

where C is a copula and $F_{X,j}$ stand for the individual marginal *cdf*'s and the copula is unique if the marginal *cdf*'s are continuous. By choosing the $F_{X,j}$ as *cdf*'s of Gaussian random variables X_j, the vector $X = (X_1, ..., X_d)^t$ has Gaussian components. But not any copula C makes F_X a *cdf* of Gaussian vector.

For example, for $d = 2$, by taking the least copula $C(u, v) = max(u + v - 1, 0)$, $(u, v) \in [0, 1]^2$, Φ the *cdf* of a $\mathcal{N}(0, 1)$ random variable, a random vector $(X, Y)^t$ associated with the *cfd*

$$\forall x = (x, y)^t \in \mathbb{R}^2,\ F(x) = \max(\Phi(x) + \Phi(y) - 1, 0),$$

is not Gaussian but has Gaussian components.

(C) - Uncorrelated and Gaussian Component.

Let us begin to resume the result of this part by saying this : For a Gaussian vector, uncorrelation and independence of its sub-vectors are the same. Precisely we have :

PROPOSITION 11. *Let $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^r$ and $Z : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^s$, $r \geq 1$, $s \geq b$, be a two random vectors such that $X^t = (Y^t, Z^t)$ is a d-Gaussian vector, $d = r + s$. Suppose that Y and Z are uncorrelated, that is, their covariance matrices are null matrices*

$$\mathbb{C}ov(Y, Z)\mathbb{C}ov(Z, Y)^t = \biggl(\mathbb{C}ov(Y_i, Z_j\biggr)_{1 \leq i \leq r,\ 1 \leq j \leq s} = 0,$$

that is also

$$\forall (u, j) \in \{1, ..., r\} \times \{1, ..., s\}, \quad \mathbb{C}ov(Y_i, Z_j) = 0.$$

Then Z and Y and independent.

Proof.

Since X is Gaussian, its sub-vectors Z and Z are Gaussian and have *mgf* functions

$$\mathbb{R}^r \ni v \mapsto \varphi_Y(v) = \exp\left(m_Y^t v + v^t \Sigma_Y v\right) \quad (RV05a)$$

and

$$\mathbb{R}^s \ni w \mapsto \varphi_Z(w) = \exp\left(m_Z^t w + w^t \Sigma_Y w\right), \ (RV05b)$$

where m_Y and Σ_Y (resp. m_Z and Σ_Z) are the expectation vector and the variance-covariance matrix of Y (resp. Z). The components of X are $X_i = Y_i$ for $1 \leq i \leq r$ and $X_i = Z_i$ for $r+1 \leq i \ d$. Suppose that Y and X are uncorrelated. Denote also by m_X and Σ_X the expectation vector and the variance-covariance matrix of X.

Thus for any $v \in \mathbb{R}^r$, $w \in \mathbb{R}^r$, we have by denoting $u^t = (v^t, w^t)$, $u \in \mathbb{R}^d$,

$$\begin{aligned} u^t \Sigma_X u &= \sum_{1 \leq i \leq d, \ 1 \leq j \leq d} \mathbb{C}ov(X_i, X_j) \\ &= \sum_{1 \leq i \leq r, \ 1 \leq j \leq r} \mathbb{C}ov(X_i, X_j) + \sum_{r+1 \leq i \leq r, \ r+1 \leq j \leq d} \quad (L2) \\ &+ \sum_{1 \leq i \leq r, \ r+1 \leq j \leq d} \mathbb{C}ov(X_i, X_j) + \sum_{r+1 \leq i \leq d, \ 1 \leq j \leq s} \quad (L3) \end{aligned}$$

The covariances of Line (L3) are covariance between a component of Y and another of Z and by hypothesis, the summation in that line is zero. In the first term of Line (L2), the covariances are those between components of Y and the second term contains those of components of Z. We get

$$u^t \Sigma_X u = v^t \Sigma_Y v + w^t \Sigma_Y w$$

with the same notation, we have $u^t m_X = v^t m_Y + v^t m_Z$ and, by taking Formula (RV05) into account, we arrive at

$$\begin{aligned}\varphi_X(u) &= \varphi_{(Y,Z)}(u,w) = \exp\left(u^t m_X + \frac{u^t \Sigma_X u}{2}\right)\\ &= \exp\left(v^t m_Y + v^t m_Z + \frac{v^t \Sigma_Y v + w^t \Sigma_Y w}{2}\right)\\ &= \varphi_Y(v)\varphi_Z(w).\end{aligned}$$

We finally have for any $v \in \mathbb{R}^r$, $w \in \mathbb{R}^r$,

$$\varphi_{(Y,Z)}(u,w) = \varphi_Y(v)\varphi_Z(w).$$

By Theorem 7 in Section 6 in Chapter 2, we conclude that Y and Z are independent.

WARNING Gaussian Random vectors do not have the exclusivity of such a property. To make it simple, this property holds for a random ordered pair (X,Y) if for example, for any $(u,v) \in \mathbb{R}^2$,

$$\left|\Phi_{(X,Y)}(u,v) - \Phi_X(u)\Phi_Y(v)\right| \leq h_{X,Y}(n,u),$$

where $h_{X,Y}$ is a function satisfying $h_{X,Y}(0,0) = 0$.

Example : Associated random variables. A finite family of d real random variables $X_j : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$, $1 \leq j \leq d$ is said to be associated if and only for any pair(f,g) of bounded real-valued and measurable functions functions defined on $\mathbb{R}^d$ both coordinate-wisely non-decreasing, we have

$$\mathbb{C}ov\left(f(X_1,...,X_d)g(X_1,...,X_d)\right) \geq 0.$$

Let us denote $X^t = (X_1,...,X_d)$ and let Σ_X be the variance-covariance matrix of X. If that sequence is associated, Newman and Wright (1981) Theorem states that for any $u \in \mathbb{R}^d$,

$$\left|\Phi_{(X_1,...,X_d)}(u) - \prod_{1 \leq j \leq d} \Phi_{X_j}(u_j)\right| \leq \frac{1}{2} \sum_{1 \leq i \leq d,\ 1 \leq j \leq d} |u_i u_j| \mathbb{C}ov(X_i, X_j).$$

It is useful to know that the covariances $\mathbb{C}ov(X_i, X_j)$ are non-negative for associated variables. Thus, associated and uncorrelated variables are independent.

(D) - Density probability function of Gaussian Vectors with a positive variance-covariance matrix.

Probability laws of Gaussian vectors with non-singular variance-covariance matrix may be characterized by their absolute *pdf*. We have the following :

Proposition - Definition (DEF03).

(a) Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$, $d \geq 1$, be Gaussian random vector of expectation vector m and variance-covariance matrix Σ. If Σ is invertible, then X has the *pdf*

$$\frac{det(\Sigma)^{-1/2}}{(2\pi)^{d/2}} \exp\biggl(- \frac{(x-m)^t \Sigma^{-1} (x-m)}{2} \biggr), \ x \in \mathbb{R}^d. \ x \in \mathbb{R}^d. \ (RVD)$$

(b) (DEF03) A random vector $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$, $d \geq 1$, whose variance-covariance is invertible is a Gaussian vector if and only if it admits the absolute density probability *pdf* (RVD) above.

Proof. We use the same techniques as in the proof of Proposition 9 and based on the Kolmogorov construction of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding a d-random vector Z whose components are centered independent real-valued Gaussian random vectors having as variances the eigen-values δ_j, $1 \leq j \leq d$ of Σ. Set $D = diag(\delta_1, ..., \delta_d)$. All those eigen-value δ_j, $1 \leq j \leq d$, are positive and

$$det(\Sigma) = \prod_{1 \leq j \leq d} \delta_j.$$

We may use the *pdf*'s of each Z_j and make profit of their independence to get the *pdf* of Z, which is for any $x \in \mathbb{R}^d$,

$$f_Z(z) = \prod_{1 \leq j \leq d} f_{Z_i}(z_i) = \prod_{1 \leq j \leq d} \frac{1}{(2\pi\delta_j)} \exp\biggl(- \sum_{1 \leq j \leq d} \frac{z_i^2}{2\delta_j} \biggr),$$

which yields

$$f_Z(z) = \frac{det(\Sigma)^{-1/2}}{(2\pi)^{d/2}} \exp\biggl(- \frac{z^t D^{-1} z}{2} \biggr), \ z \in \mathbb{R}^d.$$

Now, let T be an orthogonal matrix such that $T\Sigma T^t = D$. Set $Y = T(Z+m)$, that is : $Z = T^t Y - m$, is a diffeomorphism which preserves the whole domain $\mathbb{R}^d$ of Z and the Jacobian coefficient $J(y)$ is the determinant of T^t which is ± 1. The change of variable formula (CVF) in Section 7 in Chapter 2 leads to

$$f_Y(y) = f_Y(T(z-m) = \frac{det(\Sigma)^{-1/2}}{(2\pi)^{d/2}} \exp\biggl(- \frac{(y-m)T^t D^{-1} T(y-m)}{2} \biggr),$$

where $y \in \mathbb{R}^d$. Since $T^t D^{-1} T = \Sigma^{-1}$, we conclude that

$$f_Y(y) = \frac{det(\Sigma)^{-1/2}}{(2\pi)^{d/2}} \exp\biggl(- \frac{(y-m)\Sigma^{-1}(y-m)}{2} \biggr), \ y \in \mathbb{R}^d.$$

By combining this with Proposition 9, we conclude that the non-negative function given in formula (RVD) is an absolute *pdf* and is the *pdf* of any random vector with the *mgf* given in Formula (RV02) for a non-singular matrix Σ. ■

Different definitions. We provided three definitions (DEF01), (DEF02) and (DEF03) for Gaussian vectors. The first which is based of the characteristic function or the *mgf* is the most general. The second suppose we already have the definition a real Gaussian random variable. The last assumes that the variance-covariance is invertible.

Remark. Another way to proceed for the last proof is to directly show that the function given Formula (RVD) is a *pdf* and to compute its *mgf* by using the orthogonal transform of Σ. By trying to do so, Formula (UID) in Section 2 in Chapter 2 may be useful.

(E) Quadratic forms of Gaussian Vectors.

Let $X \sim \mathcal{N}_d(m, \Sigma)$, $d \geq 1$, be a d-dimensional Random Vector. We have the following sample result.

PROPOSITION 12. *If Σ is invertible, then the quadratic form $(X-m)^t\Sigma^{-1}(X-m)$ follows a Chi-square probability law of d degrees of freedom, that is*

$$(X-m)^t\Sigma^{-1}(X-m) \sim \chi^2_d.$$

Proof. Suppose that $X \sim \mathcal{N}_d(m,\Sigma)$ and Σ is invertible. Let T be an orthogonal matrix such that

$$T\Sigma T^t = D = diag(\delta_1, ..., \delta_d)$$

which entails

$$\Sigma^{-1} = T^t D^{-1} T.$$

Set $Y = T(X-m)$. Thus Y is a Gaussian vector. Its variance-covariance matrix is $\Sigma_Y = T\Sigma T^t = D$. Hence the components Y_1, ..., Y_d are Gaussian and not correlated. Hence they are independent. By Fact 2 in Point (11) on the Chi-square probability law in Section 2 in Chapter 2, we have

$$Q = \sum_{1\leq j\leq d} \frac{Y_j^2}{\delta_j} \sim \chi^2_d.$$

Since $D^{-1} = diag(1/\delta_1, ..., 1/\delta_d)$, we have

$$Q = Y^t D^{1-} Y = (X-m)^t T^t D^{1-} T(X-m) = (X-m)^t\Sigma^{-1}(X-m) \sim \chi^2_d. \square$$

CHAPTER 5

Introduction to Convergences of Random Variables

1. Introduction

The convergence of random variables, extended by the convergence of their probability laws, is a wide field with quite a few number of sub-fields. In Statistical terms, any kind of convergence theory of sequences of random variables is classified in the asymptotic methods area.

We are going to introduce some specific types of convergence.

Let $(X_n)_{n\geq 0}$ be a sequence of random elements with values in a Borel space $(E, \mathcal{B})$, where $\mathcal{B}$ is the σ-algebra generated by the class of open set $\mathcal{O}$, such that each X_n, $n \geq 0$, is defined on some probability space $(\Omega_n, \mathcal{A}_n, \mathbb{P}^{(n)})$.

Let also $X_\infty : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}^{(\infty)}) \rightarrow (E, \mathcal{O})$ be some random element.

Notation. We will simply write $X = X_\infty$ if no confusion is possible.

Regularity Condition. At least we suppose that the topological space $(E, \mathcal{O})$ is separated ensuring that limits are unique and for sequences of any random elements $X, X_n : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E, \mathcal{O})$, $n \geq 0$, we have

$$(X_n \rightarrow X) \in \mathcal{A}.$$

Now let us present some the following definitions for convergence of random variables after the

Warning : In this textbook, only the convergences (A), (B) , (F) and (G) will be addressed, and they will studied on $E = \mathbb{R}^d$.

(A) Almost-sure Convergence. *Suppose that X_∞ and all the elements of the sequence $(X_n)_{n\geq 0}$ are defined on the same probability space*

$(\Omega, \mathcal{A}, \mathbb{P})$.

The sequence $(X_n)_{n\geq 0}$ converges almost-surely to X_∞ and we denote

$$X_n \rightarrow X_\infty, \ a.s. \ as \ n \rightarrow +\infty,$$

if and only

$$\mathbb{P}(X_n \nrightarrow X_\infty) = 0. \ (ASC)$$

(B) Convergence in Probability. *Suppose that X_∞ and all the elements of the sequence $(X_n)_{n\geq 0}$ are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and E is a normed real linear space and its norm is denoted by $\|.\|$.*

The sequence $(X_n)_{n\geq 0}$ converges in probability to X_∞ and we denote

$$X_n \stackrel{\mathbb{P}}{\longrightarrow} X_\infty, \ as \ n \rightarrow +\infty,$$

if and only for any $\varepsilon > 0$

$$\lim_{n\rightarrow+\infty} \mathbb{P}(\|X_n - X_\infty\| > \varepsilon) = 0. \ (CP)$$

(C) General Convergence in Probability. *Suppose that X_∞ and all the elements of the sequence $(X_n)_{n\geq 0}$ are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*

The sequence $(X_n)_{n\geq 0}$ generally converges in probability to X_∞ and we denote

$$X_n \stackrel{\mathbb{P}(g)}{\longrightarrow} X_\infty, \ as \ n \rightarrow +\infty,$$

if and only for any open set $G \in E$,

$$\lim_{n\rightarrow+\infty} \mathbb{P}(X \in G, \ X_n \notin G) = 0. \ (GCP)$$

(D) Complete Convergence. *Suppose that X_∞ and all the elements of the sequence $(X_n)_{n\geq 0}$ are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and E is a normed real linear space and its norm is denoted* *by* $\|.\|$.

The sequence $(X_n)_{n\geq 0}$ completely converges to X_∞ and we denote

$$X_n \stackrel{c.c}{\longrightarrow} X_\infty,\ as\ n \rightarrow +\infty,$$

if and only for any $\varepsilon > 0$,

$$\sum_{n}^{+\infty} \mathbb{P}(\|X_n - X_\infty\| > \varepsilon) < +\infty. \ (CC)$$

(E) Convergence in p-th moment, $p > 0$. *Suppose that* X_∞ *and all the elements of the sequence* $(X_n)_{n\geq 0}$ *are defined on the same probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *and* E *is a normed real linear space and its norm is denoted by* $\|.\|$. Let $r > 0$.

The sequence $(X_n)_{n\geq 0}$ converges to X_∞ in the r-th moment and we denote

$$X_\infty \stackrel{m^r}{\longrightarrow} X_\infty,\ as\ n \rightarrow +\infty,$$

if and only for

$$\lim_{n\rightarrow+\infty} \mathbb{E}\|X_n - X_\infty\|^r = 0. \ \ (MR)$$

(F) Convergence in moment L^p**,** $p \geq 1$. *Suppose that* X_∞ *and all the elements of the sequence* $(X_n)_{n\geq 0}$ *are* ***real-valued*** *mappings defined on the same probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *and belong all to* $L^p(\Omega, \mathcal{A}, \mathbb{P})$.

The sequence $(X_n)_{n\geq 0}$ convergences to X_∞ in L^p and we denote

$$X_\infty \stackrel{L^p}{\longrightarrow} X_\infty,\ as\ n \rightarrow +\infty,$$

if and only if

$$\lim_{n\rightarrow+\infty} \mathbb{E}\|X_n - X_\infty\|_p = 0. \ \ (CLP)$$

Important Remark. It is of the greatest importance to notice that all the previous limits, the random variables X_∞ and X_n, $n \geq 0$, are defined on the same **probability space**. This will not be the case in the next definition. Each random element may be defined on its own probability space. We will come back to this remark after the definition.

(G) Weak Convergence in a metric space. *Suppose E is a metric space (E, d) endowed with the metric Borel σ-algebra.* Denote by $\mathcal{C}_b(E)$ the class of all real-valued, bounded and continuous functions defined on E. Define the probability laws :

$$\mathbb{P}_\infty = \mathbb{P}^{(\infty)} X_\infty^{-1}, \ \ \mathbb{P}_n = \mathbb{P}^{(n)} X_n^{-1}, \ n \geq 0.$$

The sequence $(X_n)_{n\geq 0}$ weakly convergences to X_∞ and we denote

$$X_\infty \rightsquigarrow X_\infty, \ as \ n \to +\infty,$$

if and only for $f \in \mathcal{C}_b(E)$

$$\lim_{n\to+\infty} \int_E f \ d\mathbb{P}_n = \int_E f \ d\mathbb{P}_\infty. \ \ (WC)$$

Remark. We effectively see that only the probability laws of X_n, $n \geq 0$ and X_∞ are concerned in Formula (WC), at the exclusion of the paths $\{X_\infty(\omega), \ \omega \in \Omega_\infty\}$ and $\{X_n(\omega), \ \omega \in \Omega_n\}$, $n \geq 0$. In general, a type of convergence which ignores the domain of elements of the sequence whose limit is considered, is called *weak* or *vague*.

As announced earlier, we are going to study convergences type (A), (B) , (F) and (G) for sequences of random vectors in $\mathbb{R}^d$, $d \geq 1$. At this step, the three remarks are should be made.

(a) Convergence (A) and (B) are already treated in the Measure Theory and Integration book. We will give easy extensions only.

(b) Convergence (G) is treated in a separate monograph. At this step of this course of probability theory, the weak convergence theory for random vectors may be entirely treated. This is what we did in Lo *et al.* (2016), as an element of the current series. The reason we expose that theory in an independent textbook us that we want it to be a first part of the exposition of Weak convergence embracing the most general spaces, including, stochastic processes.

The reader is free to read it as soon as he has completed the chapters 1 to 4 of this textbook. But, for coherence's sake, we will give the needed reminders to have a comprehensive comparison between the different kinds of convergence.

(c) Space L^p. Convergence in L^p is simply a convergence in a normed space L^p. We already know for the Measure Theory and Integration book that this space is a Banach one.

(d) Convergences of real sequences. When dealing with random vectors, a minimum prerequisite is to master the convergence theory for non-random sequences of real number. This is why we always include a related appendix in our monographs dealing with it. In this book, the reminder is exposed in Section 3 in the Appendix chapter 10.

After the previous remarks, we see that this chapter is rather a review one with some additional points. In particular, the equi-continuity notion will be introduced for the comparison between the convergence in measure and the L^p-convergence.

Part A : Convergences of real-valued random variables. .

2. Almost-sure Convergence, Convergence in probability

As recalled previously, such convergences have been studied in probability and Integration [Chapter 7 in Lo (2017b)]. We are just going to report the results.

(a) Almost-everywhere convergence.

A sequence of random variables $X_n)_{n\geq 1}$ defined from $(\Omega, \mathcal{A}, \mathbb{P})$ to $\overline{\mathbb{R}}$ converges almost-surely to a random variable $X : (\Omega, \mathcal{A}, m) \longmapsto \overline{\mathbb{R}}$ and we denote

$$X_n \longrightarrow f, \ a.s.,$$

if and only if

$$\mathbb{P}(X_n \ \nrightarrow \ X) = 0.$$

If the elements of the sequences X_n are finie *a.s.*, we have :

Characterization. A sequence of *a.s.* finite random variables $(X_n)_{n\geq 1}$ defined from $(\Omega, \mathcal{A}, \mathbb{P})$ to $\overline{\mathbb{R}}$ converges almost-surely to a random variables $X : (\Omega, \mathcal{A}, m) \longmapsto \overline{\mathbb{R}}$ if and only if

$$\mathbb{P}\left(\bigcap_{k\geq 1}\bigcap_{N\geq 1}\bigcup_{n\geq N}(|X_n - f| < 1/k)\right) = 0,$$

if and only if, for any $k \geq 1$

$$\mathbb{P}\left(\bigcup_{N\geq 1}\bigcap_{n\geq N}(|X_n - f| < 1/k)\right) = 0$$

if and only if, for any $\varepsilon > 0$

$$\mathbb{P}\left(\bigcup_{N\geq 1}\bigcap_{n\geq N}(|X_n - f| \geq \varepsilon)\right) = 0.$$

(b) Convergence in Probability.

A sequence of **a.s. finite** random variables $(X_n)_{n\geq 1}$ defined from $(\Omega, \mathcal{A}, \mathbb{P})$ to $\mathbb{R}$ converges in Probability with respect to the probability measure $\mathbb{P}$ to an *a.s* finite random variable $X : (\Omega, \mathcal{A}, m) \longmapsto \mathbb{R}$, denoted

$$X_n \rightarrow_{\mathbb{P}} X$$

if and only for any $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0 \text{ as } n \longrightarrow +\infty.$$

Remark. The convergence in probability is only possible if the limit f and the X_n's are **a.s.** since we need to get the differences $X_n - X$. The *a.s.* finiteness justifies this.

NB. It is important to notice that the inequality in $(|X_n - X| > \varepsilon)$ may be strict or not.

(c) - Properties of the *a.s.* convergence.

(c1) The *a.s.* limit is *a.s.* unique.

(c2) We have the following operations on *a.s* limits :

Let $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ be sequences of *a.s.* finite functions. Let a and b be finite real numbers. Suppose that $X_n \rightarrow X$ *a.s.* and $Y_n \rightarrow Y$ *a.s.*. Let $H(x, y)$ a continuous function of $(x, y) \in D$, where D is an open set of $\mathbb{R}^2$. We have :

(1) $aX_n + bY_n \rightarrow aX + Yg$ *a.s.*

(2) $X_nY_n \rightarrow XY$ *a.s*

(3) If $\mathbb{P}(Y = 0) = 0$ (that is Y is *a.s* nonzero), then

$$X_n/Y_n \rightarrow X/Y, \ a.s.$$

(4) If $(X_n, Y_n)_{n\geq 1} \subset D$ *a.s.* and $(X, Y) \in D$ *a.s.*, then

$$H(X_n, Y_n) \rightarrow H(X, Y), \ a.s.$$

(d)- *a.s.* Cauchy sequences. If we deal with *a.s.* finite functions, it is possible to consider Cauchy Theory on sequences of them. And we have the following definition and characterizations.

Definition. A sequence $(X_n)_{n\geq 1}$ of *a.s.* finite functions is an $\mathbb{P}$-*a.s.* Cauchy sequence if and only if

$$\mathbb{P}\biggl(X_p - Y_q \nrightarrow 0,\ as\ (p,q) \to (+\infty, +\infty)\biggr) = 0,$$

that is, the ω for which the real sequence $(X_n(\omega))_{n\geq 0}$ is a Cauchy sequence on $\mathbb{R}$ form an *a.s.* event.

Other expressions. A sequence $(X_n)_{n\geq 1}$ of *a.s.* finite functions is an $\mathbb{P}$-*a.s.* Cauchy sequence :

if and only if for any $k \geq 1$,

$$\mathbb{P}\left(\bigcap_{n\geq 1}\bigcup_{p\geq n}\bigcup_{q\geq n}(|f_p - f_q| > 1/k)\right) = 0$$

if and only if for any $k \geq 1$,

$$\mathbb{P}\left(\bigcap_{n\geq 1}\bigcup_{p\geq 0}(|f_{p+n} - X_n| > 1/k)\right) = 0$$

if and only if for any $\varepsilon > 0$,

$$\mathbb{P}\left(\bigcap_{n\geq 1}\bigcup_{p\geq n}\bigcup_{q\geq n}(|f_p - f_q| > \varepsilon)\right) = 0$$

if and only if for any $\varepsilon > 0$,

$$\mathbb{P}\left(\bigcap_{n\geq 1}\bigcup_{p\geq 0}(|X_{p+n} - X_n| > \varepsilon)\right) = 0$$

Property. Let $(X_n)_{n\geq 1}$ be a sequence of *a.s.* finite functions.

$(X_n)_{n\geq 1}$ is an $\mathbb{P}$-*a.s.* Cauchy sequence if and only if $(X_n)_{n\geq 1}$ converges *a.s.* to an *a.s.* finite function.

(e) - Properties of the convergence in probability.

(e1) The limit in probability is *a.s.* unique.

(e2) Operation on limits in Probability.

The operations of limits in probability are not simple as those for *a.s.* limits. The secret is that such operations are related to weak convergence. The concepts of tightness or boundedness are needed to handle this. But we still have some general laws and complete results on operations on constant and non-random limits.

Let $X_n \to_{\mathbb{P}} X$ and $Y_n \to Y$, $a \in \mathbb{R}$. We have :

(1) In the general case where X and Y are random and *a.s.* finite, we have :

(1a) $X_n + Y_n \to_{\mathbb{P}} X + Y$.

(2b) $aX_n \to_{\mathbb{P}} aX$

(2) - Finite and constant limits in probability.

Let $X = A$ and $Y = B$ be constant and non-random. we have

(2a) $aX_n + bY_n \to_{\mathbb{P}} aA + bB$.

(2b) $X_nY_n \to_{\mathbb{P}} AB$.

(3c) If $B \neq 0$, then

$$X_n/Y_n \to_{\mathbb{P}} A/B.$$

(3d) If $(X_n, Y_n)_{n\geq 1} \subset D$ *a.s.* and $(A, B) \in D$, then

$$H(X_n, Y_n) \to_{\mathbb{P}} H(A, B).$$

(f) - Cauchy sequence in probability or mutually convergence in probability.

Here again, we deal with *a.s.* finite random variables and consider a Cauchy Theory on sequences of them. And we have the following definition and characterizations.

Definition. A sequence $(X_n)_{n\geq 1}$ of *a.s.* random variables is a Cauchy sequence in probability if and only if, for any ε,

$$\mathbb{P}(|X_p - X_q| > \varepsilon) \rightarrow 0 \ as \ (p, q) \rightarrow (+\infty, +\infty).$$

Properties. Let $(X_n)_{n\geq 1}$ be a sequence of *a.s.* random variables. We have :

P1 $(X_n)_{n\geq 1}$ is a Cauchy sequence in probability if and only if $(X_n)_{n\geq 1}$ converges in probability to an *a.s.* random variable.

P2 If $(X_n)_{n\geq 1}$ is a Cauchy sequence in probability, then $(X_n)_{n\geq 1}$ possesses a subsequence $(X_{n_k})_{k\geq 1}$ and an *a.s.* random variable such that f such that

$$X_{n_k} \rightarrow X \ a.s. \ as \ k \rightarrow +\infty,$$

and

$$X_n \rightarrow_{\mathbb{P}} X \ as \ n \rightarrow +\infty.$$

(g) - Comparison between *a.e.* convergence and convergence in probability.

(1). If $X_n \rightarrow X$ *a.s.*, then $X_n \rightarrow_{\mathbb{P}} f$.

The reverse implication is not true. It is only true for a sub-sequence as follows.

(2). Let $X_n \rightarrow_{\mathbb{P}} X$. Then, there exists a sub-sequence $(X_{n_k})_{k\geq 1}$ of $(X_n)_{n\geq 1}$ converging *a.s* to X.

Terminology. Probability Theory results concerning a *a.s.* limit is qualified as *strong*. Since such results imply versions with limits in probability which are called *weak*.

3. Convergence in L^p

We already know that $L^p(\Omega, \mathcal{A}, \mathbb{P})$ is a Banach space, with for $X \in L^p$,

$$\|X\|_p = \left(\mathbb{E}(|X|^p)\right)^{1/p},\ p \in [1, \infty[$$

and

$$\|X\|_\infty = \inf\{M > 0,\ |X| \leq M,\ \text{a.s.}\},\ p = +\infty.$$

In this section, we are going to compare L^p convergence and the *a.s.* convergence or the convergence in probability.

We restrict ourselves to the case where p is finite.

(a) Immediate implications.

We have the following facts.

PROPOSITION 13. *$(X_n)_n \subset L^p$ and $X \in L^p$ and let $X_n \stackrel{L^p}{\longrightarrow} X$. Then :*

(i) $X_n \longrightarrow_{\mathbb{P}} X$

and

(ii) $\|X_n\|_p \to \|X\|_p$,

meaning that : the convergence in L^p implies the convergence in probability and the convergence of p-th absolute moments.

Proof. $(X_n)_n \subset L^p$ and $X \in L^p$ and let $X_n \stackrel{L^p}{\longrightarrow} X$.

Proof of Point (i). For any $\varepsilon > 0$ and by the Markov inequality, we have

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X|^p > \varepsilon^p) \leq \frac{\|X_n - X\|_p^p}{\varepsilon} \to 0.$$

Thus the convergence in L^p implies the convergence in probability.

Proof of Point (ii). This is immediate from the second triangle inequality

$$\left| \|X_n\|_p - \|X\|_p \right| \leq \|X_n - X\|_p \to 0. \square$$

On can the question : does one of Points (i) and (ii) implies the convergence in L^p? We need the concepts of continuity of a sequence of real random variables. Most of the materials used below comes from Loève (1997).

(b) Continuity of a sequence of random variables.

We have already seen the notion of continuity for a real-valued σ-additive application defined on the σ-algebra $\mathcal{A}$ with respect to the probability measure $\mathbb{P}$ pertaining to the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, which holds whenever as follows :

$$\forall A \in \mathcal{A},\ \mathbb{P}(A) = 0 \Rightarrow \phi(A).$$

Such a definition may be extended to the situation where we replace $\mathbb{P}(A) = 0$ by a limit of the form :

$$\phi(A) \to 0 \ as\ \mathbb{P}(A) \to 0, \quad (AC01)$$

which may be discretized in the form :

$$\biggl((A_p)_{p \geq 0} \subset \mathcal{A} \ and\ \mathbb{P}(A_p) \to 0 \biggr) \Rightarrow \biggl(\phi(A_p) \to 0 \biggr), \quad (AC02)$$

where the limits are meant as $p \to +\infty$.

Let $\phi = i = \phi_X$ be an indefinite integral associated to the absolute value of random variable X, that is

$$\phi_X(A) = \int_A |X| \, d\mathbb{P},\ A \in \mathcal{B}(\mathbb{R}).$$

We denote $B(X, c) = (|X| > c)$ for any $c > 0$ and introduce the condition

$$\lim_{c\uparrow+\infty} \phi_X(B(X,c)) = 0,$$

that is

$$\lim_{c\uparrow+\infty} \int_{(|X|>c)} |X| \, d\mathbb{P} = 0. \quad (CI)$$

Let us introduce the following :

Definitions.

(a) A random variable $X \in \overline{\mathbb{R}}$ is $\mathbb{P}$-absolutely continuous if and only if Formula (AC01) holds.

(b) A random variable $X \in \overline{\mathbb{R}}$ is $\mathbb{P}$-continuously integrable if and only if Formula (CI01) holds. $\Diamond$

We have the following first result.

PROPOSITION 14. *If X is integrable, then it is $\mathbb{P}$-absolutely continuous and $\mathbb{P}$-continuously integrable.*

Proof. Let X be integrable. Now, since $(|X| > c) \downarrow (|X| = +\infty)$ as $c \uparrow +\infty$, we get by the monotone convergence theorem (Do not forget that any limit is achieved through a discretized form)

$$\lim_{c\uparrow+\infty} \int_{(|X|>c)} X \, d\mathbb{P} = \int_{(|X|=+\infty)} X \, d\mathbb{P}.$$

Since X in integrable, it is *a.s.* finite, that is $\mathbb{P}(|X| = +\infty) = 0$, which leads to $\int_{(|X|=+\infty)} X \, d\mathcal{P}$ since the indefinite integral of the integrable random variable X is continuous with respect to $\mathcal{P}$. Hence X is $\mathbb{P}$-continuous integrable.

Now, suppose that $(A_p)_{p\geq 0} \subset \mathcal{A}$ and $\mathbb{P}(A_p) \to 0$ as $p \to +\infty$. We have for any $c > 0$, $p \geq 0$,

$$\begin{aligned}\int_{A_p} |X| \, d\mathbb{P} &= \int_{A_p \cap B(X,c)} |X| \, d\mathbb{P} + \int_{A_p \cap B(X,c)^c} |X| \, d\mathbb{P} \\ &\leq \int_{B(X,c)} |X| \, d\mathbb{P} + c\mathbb{P}(A_p).\end{aligned}$$

By letting $p \to +\infty$ first and next $c \uparrow +\infty$, we get Formula (AC02). Hence X is $\mathbb{P}$-absolutely continuous. $\square$

Now we may extend the definitions above to a sequence of integrable random variables by requiring that Formulas (AC02), page **??** or (CI01), page **??**, to hold uniformly. This gives :

Definitions.

(a) A sequence of integrable random variables $(X_n)_{n\geq 0} \subset L^1$, is $\mathbb{P}$-uniformly and absolutely continuous (*uac*) if and only if

$$\lim_{\mathbb{P}(A)\to 0} \sup_{n\geq 0} \int_A |X_n| \, \mathbb{P} = 0, \quad (UAC1)$$

which is equivalent to

$$\forall \varepsilon > 0, \exists \eta > 0, \, \forall A \in \mathcal{A}, \mathbb{P}(A) < \eta \Rightarrow \forall n \geq 0, \, \int_A |X_n| \, d\mathbb{P} < \varepsilon. \, (UAC2)$$

(b) A sequence of integrable random variables $(X_n)_{n\geq 0} \subset L^1$, is $\mathbb{P}$-uniformly continuously integrable (*uci*) if and only if

$$\lim_{c\uparrow+\infty} \sup_{n\geq 0}, \int_{(|X_n|>c)} |X_n| \, \mathbb{P}. \quad (UCI)$$

. $\Diamond$

Example. As in Billinsgley (1968), let us consider a sequence of random variables $(X_n)_{n\geq 0} \subset L^{1+r}$, $r > 0$ such that

$$\sup_{n\geq 0} \mathbb{E}|X_n|^{1+r} = C < +\infty.$$

Such a sequence is $\mathbb{P}$-*uci* since for all $c > 0$,

$$\int_{(|X_n|>c)} |X_n| \ \mathbb{P} = \int_{(|X_n|>c)} \frac{|X_n|^{1+r}}{|X_n|^r} \ \mathbb{P} \leq c^{-r}C,$$

and next

$$\sup_{n\geq 0} \int_{(|X_n|>c)} |X_n| \ \mathbb{P} \leq c^{-r}C \rightarrow 0 \ as \ c \uparrow +\infty.$$

Unlike the situation where we had only one integrable random variable, the two notions of $\mathbb{P}$-*uac* and $\mathbb{P}$-*uci* do not coincide for sequences. We have :

Proposition 15. *A sequence of integrable random variables* $(X_n)_{n\geq 0} \subset L^1$

(i) is $\mathbb{P}$*-uci*

if and only if

(ii) it is $\mathbb{P}$*-uci and the sequence of integrals* $(\mathbb{E}|X_n|)_{n\geq 0}$ *is bounded.* $\diamondsuit$

Proof. Let us consider a sequence of integrable random variables $(X_n)_{n\geq 0} \subset L^1$.

Let us suppose that is $\mathbb{P}$-*uci*. Hence by definition, by the classical results of limits in $\mathbb{R}$, where

$$\sup_{c>0} \sup_{n\geq 0} \int_{(|X_n|>c)} |X_n| \ \mathbb{P} = C < +\infty,$$

next, for any $n \geq 0$, for any $c_0 > 0$,

$$\mathbb{E}|X_n| = \int_{(|X_n|>c)} |X_n| \ \mathbb{P} + \int_{(|X_n|\leq c)} |X_n| \ \mathbb{P} \leq C + c_0(|X_n| \leq c) \leq C + c_0,$$

and thus

$$\sup_{n\geq 0} \mathbb{E}|X_n| \leq C + c_0 < +\infty.$$

Besides, if we are given $(A_p)_{p\geq 0} \subset \mathcal{A}$ and $\mathbb{P}(A_p) \to 0$ as $p \to +\infty$, we have for any $c > 0$, $p \geq 0$,

$$\begin{aligned}
\int_{A_p} |X_n| \, d\mathbb{P} &= \int_{A_p \cap B(X_n,c)} |X_n| \, d\mathbb{P} + \int_{A_p \cap B(X_n,c)^c} |X_n| \, d\mathbb{P} \\
&\leq \int_{B(X_n,c)} |X_n| \, d\mathbb{P} + c\mathbb{P}(A_p) \\
&\leq \int_{B(X_n,c)} |X_n| \, d\mathbb{P} + c\mathbb{P}(A_p) \\
&\leq \sup_{n\geq 0} \int_{B(X_n,c)} |X_n| \, d\mathbb{P} + c\mathbb{P}(A_p).
\end{aligned}$$

By letting $p \to +\infty$ first and next $c \uparrow +\infty$, we get Formula (UCA). Hence the sequence $\mathbb{P}$-*uac*.

Suppose now that the sequence is $\mathbb{P}$-*uac* and the sequence of integrals $(\mathbb{E}|X_n|)_{n\geq 0}$ is bounded. Put

$$\sup_{n\geq 0} \mathbb{E}|X_n| = C < +\infty.$$

By the Markov inequality, we have

$$\sup_{n\geq 0} \mathbb{P}(|X_n| > c) \leq \sup_{n\geq 0} \frac{\mathbb{E}(|X_n|}{c} \leq Cc^{-1}. \ \ (MK)$$

Let us apply Formula (UAC2). Let $\varepsilon > 0$ and let $\eta > 0$ such that

$$\mathbb{P}(A) < \eta \Rightarrow \forall n \geq 0, \ \int_A |X_n| \, d\mathbb{P} < \varepsilon. \ \ (MK1)$$

Let $c_0 > 0$ such that $Cc_0^{-1} < \eta/2$. By Formula (MK) above we have for all $c \leq c_0$, for all $n \geq 0$, $\mathbb{P}(|X_n| > c) \leq \eta/2 < \eta$, and by Formula (MK1),

$$\forall n \geq 0, \ \int_{|X_n|>c} |X_n| \, d\mathbb{P} < \varepsilon$$

that is

$$\forall c \leq c_0, \ \ \sup_{n\geq 0} \int_{|X_n|>c} |X_n| \, d\mathbb{P} \leq \varepsilon.$$

This means that the sequence is $\mathbb{P}$-*uci*. ■

Now, we are able to give the converse of Proposition 13.

THEOREM 10. *Let $(X_n)_{n\geq 0} \subset L^p$ be e sequence of elements of L^p, and X some random variable $X \in \overline{\mathbb{R}}$. We have :*

(a) If $X_n \xrightarrow{L^p} X$, then $X \in L^p$.

(b) If If $X_n \xrightarrow{L^p} X$, then If $X_n \longrightarrow_{\mathbb{P}} X$.

(c) Suppose that $X_n \longrightarrow_{\mathbb{P}} X$ and one of the three conditions holds.

(c1) The sequence $(|X_n|^p)_{n\geq 0}$ is $\mathbb{P}$-uniformly and absolutely integrable.

(c2) The sequence $(|X_n - X|^p)_{n\geq 0}$ is $\mathbb{P}$-uniformly and absolutely integrable.

(c3) The sequence $(|X_n|^p)_{n\geq 0}$ is $\mathbb{P}$-uniformly and continuously integrable.

(c4) $\|X_n\|_p \to \|X\|_p < +\infty$.

Then $X_n \xrightarrow{L^p} X$.

(All the limits are meant when $n \to +\infty$).

Proof of Theorem 10.

Proof of (a). First remark that the random variables $|X_n - X|$ are *a.s.* defined since the X_n's are *a.s.* finite. Next, by Minkowski's inequality, for any $n \geq 0$.

$$\|X\|_p \leq \|X_n\|_p + \|X_n - X\|_p$$

By $X_n \xrightarrow{L^p} X$, there exists $n_0 \geq 0$ such that $\|X_n - X\|_p \leq 1$ and thus, $\|X\|_p \leq 1 + \|X_{n_0}\|_p < +\infty$.

Proof of (b). It is done in Proposition 13.

Proof of (c). Suppose that $X_n \longrightarrow_{\mathbb{P}} X$.

Let (c1) hold. Let $\varepsilon > 0$. By Point (f) in Section 2, the sequence $(X_n)_{n\geq 0}$ is of Cauchy in probability. Hence for $B_{r,s}(\varepsilon) = (|X_r - X_s| > (\varepsilon/2)^{1/p})$, we have

$$\mathbb{P}(B_{r,s}(\varepsilon)) \rightarrow 0 \ as \ (r,s) \rightarrow (+\infty, +\infty). \ \ (LP1)$$

Since (c1) holds, we use Formula (UAC2) to find a value $\eta > 0$ such that, for $C_p = 2p^{p-1}$,

$$\mathbb{P}(A) < \eta \Rightarrow \forall n \geq 0, \ \int_A |X_n|^p \ d\mathbb{P} < (\varepsilon/2C_p). \ \ (LP2)$$

From Formula (LP1), we can find an integer r_0 such that for any $r \geq r_0$, for any $s \geq 0$,

$$\mathbb{P}(B_{r,r+s}(\varepsilon)) < \eta.$$

Now, based on the previous facts and the C_p inequality, we have for all $r \geq r_0$ and $s \geq 0$,

$$\begin{aligned}\int |X_r - X_{r+s}|^p \ d\mathbb{P} &\leq \int_{B_{r,r+s}(\varepsilon)} |X_r - X_{r+s}|^p \ d\mathbb{P} \\ &+ \int_{B_{r,s}(\varepsilon)^c} |X_r - X_{r+s}|^p \ d\mathbb{P} \ \ (L1) \\ &\leq C_p\left(\int_{B_{r,r+s}(\delta)} |X_r|^p + \int_{B_{r,r+s}(\delta)} |X_{r+s}|^p\right) + \varepsilon/2 \\ &\leq \varepsilon/2 + \varepsilon/2 = \varepsilon.\end{aligned}$$

This implies that the sequence $(X_n)_{n\geq 0}$ us a Cauchy sequence in L^p and since L^p is a Banach space, it converges in L^p to Y. By Point (b), we also have that X_n converges in Probability to Y. Thus $X = Y$ *a.s.*. Finally $\|X - n - Y\|_p = \|X - n - X\|_p \rightarrow 0$ and X_n converges to X in L^p.

Let (c2) hold. The same method may used again. The form of $B_{r,s}(\varepsilon)$ does not change since X is dropped in the difference. When concluding in Line (L1) in the last group of formulas, we use

$$\int_{B_{r,r+s}(\varepsilon)} |X_r - X_{r+s}|^p \leq \int_{B_{r,r+s}(\varepsilon)} |X_r - X|^p \, d\mathbb{P} + \int_{B_{r,r+s}(\varepsilon)} |X_{r+s} - X|^p \, d\mathbb{P},$$

and the conclusion is made similarly.

Let (c3) hold. By Proposition 15, (c1) holds and we have the results.

Let (c4) hold. We are going to use the Young version of the Dominated Convergence Theorem [YCDT] (See Lo (2017b), Chapter 7, Doc 06-02, Point (06.07c)). We have

$$x|X_n - X|^p \leq C_p(|X_n|^p + |X|^p) = Y_n.$$

Hence $|X_n - X|^p$ converges to zero in probability and is bounded, term by term, by a sequence $(Y_n)_{n\geq 0}$ of non-negative and integrable random variables such that :

(i) Y_n converges to $Y = 2C_p|X|^p$

and

(ii) $\int Y_n \ \mathbb{P}$ convergences to $\int Y \ \mathbb{P}$.

By the YDCT, we get the conclusion, that is $\|X_n - X\|_p^p = \int |X_n - X|^p \, d\mathbb{P} \rightarrow 0$ as $n \rightarrow 0$.

We still have to expose a simple review of weak convergence on $\mathbb{R}$. But we prefer stating it, for once, on $\mathbb{R}^d$ in the next part.

Part B : Convergence of random vectors. .

4. A simple review on weak convergence

A general introduction of the theory of weak convergence is to be found in Lo *et al.* (2016). The main fruits of that theory on $\mathbb{R}^d$ are summarized below.

First of all, it is interesting that characteristic elements of probability laws on $\mathbb{R}^d$ (*cdf*'s, *pdf*'s, *mgf*'s, characteristic functions, etc.) still play the major roles in weak convergence.

The main criteria for weak convergence are stated here :

THEOREM 11. *(A particular version of Portmanteau Theorem) Let d be a positive integer. The sequence of random vectors $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}^{(n)}) \mapsto (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, ≥ 1, weakly converges to the random vector $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ if and only if one of these assertions holds.*

(i) For any real-valued continuous and bounded function f defined on $\mathbb{R}^d$,

$$\lim_{n\to+\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X).$$

(ii) For any open set G in $\mathbb{R}^d$,

$$\liminf_{n\to+\infty} \mathbb{P}_n(X_n \in G) \geq \mathbb{P}_\infty(X \in G).$$

(iii) For any closed set F of $\mathbb{R}^d$, we have

$$\limsup_{n\to+\infty} \mathbb{P}_n(X_n \in F) \leq \mathbb{P}_\infty(X \in F).$$

(iv) For any Borel set B of $\mathbb{R}^d$ that is $\mathbb{P}_X$-continuous, that is $\mathbb{P}_\infty(X \in \partial B) = 0$, we have

$$\lim_{n\to+\infty} \mathbb{P}_n(X_n \in B) = \mathbb{P}_X(B) = \mathbb{P}_\infty(X \in B).$$

(v) For any continuity point $t = (t_1, t_2, ..., t_d)$ of F_X, we have,

$$F_{X_n}(t) \to F_X(t) \text{ as } n \to +\infty.$$

where for each $n \geq 1$, F_{X_n} is the distribution function of X_n and F_X that of X.

(vi) For any point $u = (u_1, u_2, ..., u_d) \in \mathbb{R}^k$,

$$\Phi_{X_n}(u) \mapsto \Phi_X(u) \text{ as } n \to +\infty,$$

where for each $n \geq 1$, Φ_{X_n} is the characteristic function of X_n and Φ_X is that of X.

***(c)** If the moment functions φ_{X_n} exist on B_n, $n \geq 1$ and φ_X exists on B, where the B_n and B are neighborhoods of 0 and $B\cap_{n\geq 1}$, and if for any $x \in B$,*

$$\Psi_{X_n}(x) \to \Psi_X(x) \text{ as } n \to +\infty,$$

then X_n weakly converges to X.

The characteristic function as a tool of weak convergence is also used through the following criteria.

Wold Criterion. The sequence $\{X_n, \ \ n \geq 1\} \subset \mathbb{R}^d$ weakly converges to $X \in \mathbb{R}^d$, as $n \to +\infty$ if and only if for any $a \in \mathbb{R}^d$, the sequence $\{< a, X_n >, \ \ n \geq 1\} \subset \mathbb{R}$ weakly converges to $X \in \mathbb{R}$ as $n \to +\infty$.

We also have :

The Continuous mapping Theorem. Assume that the sequence $\{X_n, \ \ n \geq 1\} \subset$ weakly converges to $X \in \mathbb{R}^d$, as $n \to +\infty$. Let $k \geq 1$ and let $f : \mathbb{R}^d \to \mathbb{R}^d$ be a continuous function. Then $\{f(X_n), \ \ n \geq 1\} \subset \mathbb{R}^k$ weakly converges to $f(X) \in \mathbb{R}^k$.

The *pdf*'s may be used in the following way.

PROPOSITION 16. *These two assertions hold.*

(A) Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}^{(n)}) \mapsto (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be random vectors and $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ another random vector, all of them absolutely continuous with respect to the Lebesgue measure denoted as λ_d. Denote f_{X_n} the probability density function of X_n, $n \geq 1$ and by f_X the probability density function of X. Suppose that we have

$$f_{X_n} \to f_X,\ \lambda_k - a.e.,\ \text{as } n \to +\infty.$$

Then X_n weakly converges to X as $n \to +\infty$.

(B) Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}^{(n)}) \mapsto (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be discrete random vectors and $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ another discrete random vector. For each n, define D_n the countable support of X_n, that

$$\mathbb{P}^{(n)}(X_n \in D_n) = 1 \text{ and for each } x \in D_n,\ \mathbb{P}^{(n)}(X_n = x) \neq 0,$$

and D_∞ the countable support of X. Set $D = D_\infty \cup (\cup_{n\geq 1} D_n)$ and denote by ν as the counting measure on D. Then the probability densities of the X_n and of X with respect to ν are defined on D by

$$f_{X_n}(x) = \mathbb{P}^{(n)}(X_n = x),\ n \geq 1,\ f_X(x) = \mathbb{P}^{(\infty)}(X = x),\ x \in D.$$

If

$$(\forall x \in D), f_{X_n}(x) \to f_X(x),$$

then X_n weakly converges to X.

In summary, the weak convergence in $\mathbb{R}^d$ holds when the distribution functions, the characteristic functions, the moment functions (if they exist) or the probability density functions (if they exist) with respect to the same measure ν, point-wisely converge to the distribution function, or to the characteristic function or to moment function (if it exists), or to the probability density function (if it exists) with respect to ν of a probability measure in $\mathbb{R}^d$. In the case of point-wise convergence of the distribution functions, only matters the convergence for continuity points of the limiting distribution functions.

In Chapter 1 in Lo *et al.* (2016), a number of direct applications are given and a review of some classical weak convergence results are stated.

5. Convergence in Probability and *a.s.* convergence on $\mathbb{R}^d$

Let us denote by $\|.\|$ one of the three equivalent usual norms on $\mathbb{R}^d$. Because of the continuity of the norm, $\|X\|$ becomes a real-valued random variable for any random vector. From this simple remark, we may extend the *a.s.* convergence and the convergence in probability on $\mathbb{R}^d$ in the following way.

Definitions.

Let X and $(X_n)_{n\geq 0}$ be, respectively, a random vector and a sequence of random vectors defined on the same on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values $\overline{\mathbb{R}}^d$. Let us denote by $X_n^{(j)}$ the j-th component of X_n for each $1 \leq j \leq d$, $n \geq 1$.

(a) The $(X_n)_{n\geq 0}$ converges *a.s.* to X as $n \rightarrow +\infty$ if and only if, each sequence of components $(X_n^{(j)})_{n\geq 0}$ converges to X_j as $n \rightarrow +\infty$.

(b) Let X and the elements of sequences $(X_n)_{n\geq 0}$ have *a.s.*-finite components. Then $(X_n)_{n\geq 0}$ converges *a.s.* to X if and only if

$$\|X_n - X\| \rightarrow 0, \text{ a.s. } as\ n \rightarrow +\infty.$$

(c) Let X and the elements of sequences $(X_n)_{n\geq 0}$ have *a.s.*-finite components. Then $(X_n)_{n\geq 0}$ converges to X in probability if and only if

$$\|X_n - X\| \rightarrow_{\mathbb{P}} 0, \quad as\ n \rightarrow +\infty.$$

$\diamondsuit$

For the coherence of the definition, we have to prove the equivalence between Points (a) and (b) above in the case where the random vectors have *a.s.* finite components. This is let as an easy exercise.

We have the following properties.

PROPOSITION 17. *Let X, Y, $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ be, respectively, two random vectors and two sequences of random vectors defined on the same on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values $\overline{\mathbb{R}}^d$. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ be a continuous function. Finally let a and b tow real numbers. The limits in the proposition are meant as $n \rightarrow +\infty$.*

(1) If $X_n \to X$, a.s, then $X_n \to_{\mathbb{P}} X$.

(2) Let $X_n \to X$, a.s and $Y_n \to Y$, a.s. Then, we have

(2a) $aX_n + bY_n \to aX + bY$, a.s.

and

(2b) $f(X_n) \to f(X)$, a.s.

(3) Let $X_n \to_{\mathbb{P}} X$, and $Y_n \to_{\mathbb{P}} Y$. Then, we have

(3a) $aX_n + bY_n \to_{\mathbb{P}} aX + bY$.

and, if $X = A$ is a non-random constant vector, we have

(3b) $f(X_n) \to_{\mathbb{P}} f(A)$.

But in general, if f is a Lipschitz function, we have

(3c) $f(X) \to_{\mathbb{P}} f(X)$.

Proofs. By going back to the original versions on $\mathbb{R}$ for *a.s.* and convergence in probability, all these results become easy to prove except Points (3b) and (3c). But a proof of Point (2v) is given in the proof of Lemma 8 in Lo *et al.* (2016) of this series. Point (3c) is proved as follows.

Let $X_n \to_{\mathbb{P}} X$ and let f be a Lipschitz function associated to a coefficient $\rho > 0$, that is

$$\forall (x, y) \in \mathbb{R}^d, \ \|f(x) - f(y)\| \leq \rho \|x - y\|.$$

Hence for any $\varepsilon > 0$,

$$(\|f(X_n) - f(X)\| > \varepsilon) \subset (\|f(X_n) - f(X)\| > \varepsilon/\rho)$$

and hence

$$\mathbb{P}(\|f(X_n) - f(X)\| > \varepsilon) \subset \mathbb{P}(\|f(X_n) - f(X)\| > \varepsilon/\rho) \to 0.$$

Thus $\|f(X_n) - f(X)\| \to_{\mathbb{P}} 0$. □

Immediate implications. Since projections are Lipschitz functions, we get the if $X_n \to_{\mathbb{P}} X$, then we also get the convergence in probability component-wise, that is : each sequence of components $(X_n^{(j)})_{n\geq 0}$ converges to X_j in probability. Conversely, the convergence in probability implies the convergence if probability of the vectors. Indeed, take for example

$$\|x\| = \max_{1\leq i \ \ d} |x_i|, \ x = (x_1, ..., x_d)^t \in \mathbb{R}^d.$$

We have, for each $n \geq 0$,

$$(\|X_n - X\| > \varepsilon) \subset \bigcup_{1\leq i \ \ d} (\|X_n^{(j)} - X_j\| > \varepsilon),$$

which leads, for each $n \geq 0$, to

$$\mathbb{P}(\|X_n - X\| > \varepsilon) \leq \sum_{1\leq i \ \ d} \mathbb{P}(\|X_n^{(j)} - X_j\| > \varepsilon).$$

Since d is fixed, the conclusion is obvious.

6. Comparison between convergence in probability and weak convergence

This section is reduced to the statements of results concerning the comparison between the weak convergence and the convergence in probability.

We remember that in the definition of weak convergence, the elements of the sequence $(X_n)_{n\geq 0}$ may have their own probability spaces. So, in general, the comparison with convergence in probability does not make sense unless we are in the particular case where all the elements of the sequence $(X_n)_{n\geq 0}$ and the limit random variable X are defined on the same probability space.

Before, we state the results, let us give this definition.

Definition. Let $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ be two random vectors and two sequences of random vectors defined on the same on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values $\overline{\mathbb{R}}^d$.

They are equivalent in probability if and only if :

$$\|X_n - Y_n\| \rightarrow_{\mathbb{P}} 0, \ as \ n \rightarrow +\infty.$$

They are *a.s.* equivalent with respect to their *a.s.* convergence or divergence if and only if

$$\|X_n - Y_n\| \rightarrow 0 \ a.s., \ as \ n \rightarrow +\infty. \ \Diamond$$

We have :

PROPOSITION 18. *Let X, Y, $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ be, respectively, two random vectors and two sequences of random vectors defined on the same on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values $\overline{\mathbb{R}}^d$. We have :*

(a) The convergence in probability implies the weak convergence, that is :

If $X_n \rightarrow_{\mathbb{P}} X$, that d $Y_n \rightsquigarrow c$, then $(X_n, Y_n) \rightsquigarrow (X, c)$.

(b) The weak convergence and convergence in probability to a constant are equivalent, that is :

$X_n \longrightarrow_{\mathbb{P}} c$ as $n \rightarrow +\infty$ if and only if $X_n \rightsquigarrow c$ as $n \rightarrow +\infty$.

(c) Two equivalent sequences in probability weakly converge to the same limit if one of them does.

(d) (Slutsky's Theorem) If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$, then $(X_n, Y_n) \rightsquigarrow (X, c)$.

(e) (Coordinate-wise convergence in probability) $X_n \longrightarrow_{\mathbb{P}} X$ and $Y_n \longrightarrow_{\mathbb{P}} Y$ if and only if $(X_n, Y_n) \longrightarrow_{\mathbb{P}} (X, Y)$.

The proofs of all these facts are given in Lo *et al.* (2016) of this series.

A comment. We know that the result (e) does not holds in general for the weak convergence. This means that the convergence in probability implies the weak convergence but not the contrary.

The *a.s.* equivalence takes a special shape for partial sums. Let us consider a sequence $(X_n)_{n\geq 0}$ of real random variables defined on the same on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For a sequence of positive numbers $(c_n)_{n\geq 1}$, let us consider the truncated random variables $X_n^{(t)} = X_n 1_{|X_n|\leq c_n}$, $n \geq 1$, that is each $X_n^{(t)}$, $n \geq 1$, remains unchanged for $|X_n| \leq c_n$ but vanishes otherwise. Let us form the the partial sums $S_n = \sum_{1\leq k\leq n} X_n$ and $S_n^{(t)} = \sum_{1\leq k\leq n} X_n^{(t)}$, $n \geq 1$.

Let $(b_n)_{n\geq 1}$ be a sequence of real numbers converging to $+\infty$. We are going to see that the *a.s.* equivalence between $(S_n^{(t)}/b_n)_{n\geq 1}$ and $(S_n/b_n)_{n\geq 1}$ is controlled by the series

$$\sum_{n\geq 0} \mathbb{P}(|X_n| > c_n).$$

Indeed, since the event $(X_k \neq X_k^{(t)})$ occurs only if $(|X_k| \geq c_k)$, we have

$$\mathbb{P}(X_n \neq X_n^{(t)},\ i.o) = \lim_{n\uparrow+\infty} \mathbb{P}\left(\bigcup_{k\geq n} (X_k \neq X_k^{(c)})\right) \leq \lim_{n\uparrow+\infty} \sum_{k\geq n} \mathbb{P}(|X_k| \geq c).$$

So if the series $\sum_{n\geq 0} \mathbb{P}(|X_n| > c_n)$ is convergent, we have $\mathbb{P}(X_n \neq X_n^{(t)},\ i.o)$ which implies that there exists a null-set Ω_0^c such that for any $\omega \in \Omega_0$, we can find $N(\omega)$ such that for any $n \geq N(x)$, $X_n = X_n^{(t)}$ so that for $n > N(\omega)$,

$$\left|\frac{S_n - S_n^{(t)}}{b_n}\right| = \frac{1}{b_n}|S_N(\omega)^{(t)} - S_N(\omega)| \to 0,\ as\ n \to +\infty.$$

This proves the claim. □

CHAPTER 6

Inequalities in Probability Theory

Here, we are going to gather a number of some inequalities we may encounter and use in Probability Theory. Some of them are already known from the first chapters.

The reader may skip this chapter and comes back to it only when using, later, an inequality which is is stated here and especially when he/she wants to see the proof.

Unless an express specification is given, the random variables X, Y, X_i, Y_i, $i \geq 1$, which used below, are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Readers who want to read this chapter in the first place will need an earlier introduction to the notion of conditional expectation right now, instead of waiting Chapter 8 where this notion is studied.

1. Conditional Mathematical Expectation

We are going to use the Radon-Nikodym Theorem as stated in Doc 08-01 in Chapter 9 in Lo (2017b).

Let be given a sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$, Y a measurable mapping from $(\Omega, \mathcal{A})$ to a measurable space $(E, \mathcal{F})$ and finally a measurable mapping h from $(E, \mathcal{F})$ to $\overline{\mathbb{R}}$, endowed with the usual σ-algebra $\mathcal{B}_\infty(\overline{\mathbb{R}})$. We always suppose that $h(Y)$ is defined, and quasi-integrable, that is : $\mathbb{E}(h(Y)^+)$ or $\mathcal{E}(h(Y)^-)$ is finite.

Now the mapping

$$\mathcal{B} \ni B \mapsto \phi_{\mathcal{B}}(B) = \int_B h(Y) \, d\mathbb{P}$$

is σ-additive and is continuous with respect to $\mathbb{P}$. By the Radon-Nikodym Theorem as recalled earlier, $\phi_{\mathcal{B}}$ possesses a Radon-Nikodym derivative with respect to $\mathbb{P}$, we denoted as

$$\frac{d\phi_{\mathcal{B}}}{d\mathbb{P}} =: \mathbb{E}(h(Y)/\mathcal{B}).$$

By the properties of Radon-Nikodym derivatives (please, visit again the aforementioned source if needed), we may define.

Definition If the mathematical of expectation $h(Y)$ exists, the conditional mathematical expectation of $h(Y)$ denoted as

$$\mathbb{E}(h(Y)/\mathcal{B}),$$

is the $\mathbb{P}$.**a.s** unique real-valued and $\mathcal{B}$-measurable random variable such that

$$\forall B \in \mathcal{B}, \quad \int_B h(Y)\ d\mathbb{P} = \int_B \mathbb{E}(h(Y)/\mathcal{B})\ d\mathbb{P}. \ (CE01)$$

Moreover, $\mathbb{E}(h(Y)/\mathcal{B})$ is *a.s.* finite if $h(Y)$ is integrable. $\Diamond$

Extension of the Definition. By putting $Z = 1_B$ in (CE01), and by using the classical three steps method of Measure Theory and Integration, we easily get that when $\mathbb{E}|h(Y)| < +\infty$, Formula (CE01) is equivalent to any one of the two following others :

(a) For any non-negative and $\mathcal{B}$-measurable random variable Z,

$$\int Zh(Y)\ d\mathbb{P} = \int Z\mathbb{E}(h(Y)/\mathcal{B})\ d\mathbb{P}. \ (CE02) \ \Diamond$$

(b) For any $\mathcal{B}$-measurable and integrable random variable Z,

$$\int Zh(Y)\ d\mathbb{P} = \int Z\mathbb{E}(h(Y)/\mathcal{B})\ d\mathbb{P}. \ (CE02) \ \Diamond$$

In the extent of this chapter, we will directly utilize (CE02) as a definition each time we need it.

The following exercises will be proved in Chapter 8 as properties of the mathematical expectation.

Exercise. Show that the following properties.

(1) If X is a real-valued and quasi-integrable random variable, then

$$\mathbb{E}\biggl(\mathbb{E}(X)/\mathcal{B})\biggr) = \mathbb{E}(X).$$

(2) If X is a real-valued and quasi-integrable random variable $\mathcal{B}$-measurable, then

$$\mathbb{E}(X)/\mathcal{B}) = X. \ a.s.$$

(3) Let X be a real-valued and quasi-integrable random variable independent of $\mathcal{B}$ in the following sense : for all real-valued and quasi-integrable random variable Z $\mathcal{B}$-measurable,

$$\mathbb{E}(ZX) = \mathbb{E}(Z)\mathbb{E}(X).$$

Then, we have

$$\mathbb{E}(X)/\mathcal{B}) = \mathbb{E}(X)). \ a.s.$$

(4) If X is a real-valued and quasi-integrable random variable independent $\mathcal{B}$ and if Z is a real-valued, quasi-integrable and $\mathcal{B}$-measurable random variable, we have

$$\mathbb{E}(ZX)/\mathcal{B}) = Z\mathbb{E}(X)/\mathcal{B}).$$

(5) If X and Y are real-valued random variables both non-negative or both integrable, then

$$\mathbb{E}\biggl((X+Y)/\mathcal{B}\biggr) = \mathbb{E}(X/\mathcal{B}) + \mathbb{E}(Y/\mathcal{B}). \ \Diamond$$

2. Recall of already known inequalities

.

(1) Inequality of Markov. If $X \geq 0$, then for all $x > 0$,

$$\mathbb{P}(X \geq x) \leq \frac{1}{x}. \ \Diamond$$

(2) Inequality of Chebychev. If $\mathbb{E}(X)$ exists and $X - \mathbb{E}(X)$ is *a.s.* defined, then for all $x > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| > x) \leq \frac{\mathbb{E}(X - \mathbb{E}(X))^2}{x^2}. \ \Diamond$$

These two inequalities are particular forms of the following one.

(3) Basic Inequality. (As in Loève (1997)) Let X be any real-valued random variable and g be a non-null, non-decreasing and non-negative mapping from $\mathbb{R}$ to $\mathbb{R}$. Then for any $a \in \mathbb{R}$, we have

$$\frac{\mathbb{E}g(X) - g(a)}{\|g(X = \|_\infty} \leq \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}g(X)}{g(a)}. \quad (BI01)$$

If, in addition, g is even or if g satisfies

$$\forall a \geq 0, \ g(\max(-a, a)) \geq g(a), \quad (AI01)$$

$$\frac{\mathbb{E}g(X) - g(a)}{\|g(X = \|_\infty} \leq \mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}g(X)}{g(a)}. \quad (BI02) \ \Diamond$$

Proof of Formula (BI01). The mathematical expectation $\mathbb{E}(g(X))$ exists since g is of constant sign. By using the same method of establishing the Markov inequality, we have

$$\mathbb{E}g(X) \geq \int_{X \geq a} d\mathbb{P} \geq g(a)\mathbb{P}(X \geq a), \quad (BI03)$$

where we used the non-decreasingness of g. So, we get the right-hand inequality of Formula (BI01) even if $g(a) = 0$. We also gave

$$\int_{X \geq a} d\mathbb{P} \leq \|g(X = \|_\infty \mathbb{P}(|X| \geq a)$$

and

$$\int_{X < a} d\mathbb{P} \leq g(a)\mathbb{P}(|X| < a) \leq g(a)$$

and by these formulas,

$$\begin{aligned} \mathbb{E}g(X) &= \int_{X \geq a} g(X) \ d\mathbb{P} + \int_{X < a} g(X) \ d\mathbb{E} \\ &= \|g(X = \|_\infty \mathbb{P}(|X| \geq a) + g(a), \end{aligned}$$

that is

$$\mathbb{E}g(X) \leq \|g(X = \|_\infty \mathbb{P}(|X| \geq a) + g(a),$$

which gives the left-and inequality in Formula (BI01) even if $\|g(X = \|_\infty = +\infty$ (it cannot be zero by assumption). □

Proof of Formula (BI02). Since $\mathbb{P}(X \leq a) \leq \mathbb{P}(|X- \leq a)$, we pnly have to justify the right-hand inequality of (BI02). But we may use the simple remark that $X \geq \max(-a, a)$ on $(|X| \geq a)$ to modify (BI03) as follows

$$\mathbb{E}g(X) \geq \int_{|X|\geq a} d\mathbb{P} \geq g(\max(-a, a))\mathbb{P}(|X| \geq a), \quad (BI04)$$

So, using Assumption (AI01) - which holds if g is even - allows to conclude. □

(4) Hölder Inequality. Let $p > 1$ and $q > 1$ be two conjugated positive rel numbers, that is, $1/p + 1/q = 1$ and let

$$X, Y : \ (\Omega, \mathcal{A}, \mathbb{P}) \ \mapsto \ \mathbb{R} \ ,$$

be two random variables $X \in L^p$ and $Y \in L^q$. Then XY is integrable and we have

$$|\mathbb{E}(XY)| \leq \|X\|_p \times \|Y\|_q \,,$$

where for each $p \geq 1$, $\|X\|_p = (\mathbb{E}(|X|^p)^{1/p}$.

(5) Cauchy-Schwartz's Inequality. For $p = q = 2$, the Hölder inequality becomes the Cauchy-Schwartz one :

$$|\mathbb{E}(XY)| \leq \|X\|_2 \times \|Y\|_2 \,.$$

(6) Minskowski's Inequality. Let $p \geq 1$ (including $p = +\infty$). If X and Y are in L^p, then we have

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p \,.$$

(7) C_p Inequality. Let $p \in [1, +\infty[$. If X and Y are in L^p, then for $C_p = 2^{p-1}$, we have

$$\|X + Y\|_p^p \leq C_p(\|X\|_p^p + \|Y\|_p^p).$$

(8) Ordering the spaces $L^p, \ p \geq 1$.

Let $1 < p < q$, p finite but $q \in [1, +\infty]$. Let $X \in L^q$. Then $X \in L^p$ and

$$\|X\|_p \leq \|X\|_q \leq \|X\|_{+\infty}.$$

(9) Jensen's Inequality.

Let ϕ be a convex function defined from a closed interval I of $\mathbb{R}$ to $\mathbb{R}$. Let X be a *rrv* with values in I such that $\mathbb{E}(X)$ is finite. Then $\mathbb{E}(X) \in I$ and

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)).$$

(10) Inequality for two convex functions a random variable. Let g_i, $i \in \{1,2\}$ be two finite real-valued convex and increasing functions (then invertible function as increasing and continuous functions) such that g_2 is convex in g_1 meaning that $g_2 g_1^{-1}$ is convex. For any real-valued random variable X such that X and $g_1(X)$ are integrable, we have

$$g_1^{-1}(\mathbb{E}(g_1(Z)) \leq g_2^{-1}(\mathbb{E}(g_2(Z)).$$

(11) Bonferroni's Inequality.

Let A_1, ..., A_n be measurable subsets of Ω, $n \geq 2$. Define

$$\begin{aligned}
\alpha_0 &= \sum_{1 \leq j \leq n} \mathbb{P}(A_j) \\
\alpha_1 &= \alpha_0 - \sum_{1 \leq i_1 < t_2 \leq n} \mathbb{P}(A_{i_1} A_{i_2}) \\
\alpha_2 &= \alpha_1 + \sum_{1 \leq i_1 < \ldots < t_3 \leq n} \mathbb{P}(A_{i_1} \ldots A_{i_3}) \\
\cdots &= \cdots \\
\alpha_r &= \alpha_{r-1} + (-1)^{r+1} \sum_{1 \leq i_1 < \ldots < t_r \leq n} \mathbb{P}(A_{i_1} \ldots A_{i_r}) \\
\cdots &= \cdots \\
\alpha_r &= \alpha_{r-1} + (-1)^{n+1} \mathbb{P}(A_1 A_2 A_3 \ldots A_n).
\end{aligned}$$

Let $p = n \bmod 2$, that is $n = 2p + 1 + h$, $h \in \{0, 1\}$. We have the Bonferroni's inequalities : if n is odd,

$$\alpha_{2k+1} \leq \mathbb{P}\left(\bigcup_{1\leq j\leq n} A_n\right) \leq \alpha_{2k},\ k = 0, ..., p\ (BF1)$$

and if n is even,

$$\alpha_{2k+1} \leq \mathbb{P}\left(\bigcup_{1\leq j\leq n} A_j\right) \leq \alpha_{2k},\ k = 0, ..., p-1.\ \ (BF2)$$

3. Series of Inequalities

(12) Order relations for conditional expectations. Let X and Y be two real-valued random variables such that $X \leq Y$. Let $\mathcal{B}$ be a σ-sub-algebra of $\mathcal{A}$. Then, whenever the expressions in the two sides make sense and are finite, we have

$$\mathbb{E}(X/\mathcal{B}) \leq \mathbb{E}(Y/\mathcal{B})\ a.s.\ (CE03)$$

Besides, the conditional expectation is a contracting operator in the following sense : for any real-valued and quasi-integrable random variable X, we have

$$|\mathbb{E}(X/\mathcal{B})| \leq \mathbb{E}(|X|/\mathcal{B}).\ \ (CE04)\ \Diamond$$

Proof. Suppose that all the assumptions hold. We have for all $B \in \mathcal{B}$,

$$\int_B \mathbb{E}(X/\mathcal{B})\ d\mathbb{P} = \int_B X\ d\mathbb{P} \leq \int_B Y = \int_B \mathbb{E}(Y/\mathcal{B})\ d\mathbb{P}$$

Take an arbitrary $\varepsilon > 0$ and set $B(\varepsilon) = (\mathbb{E}(X/\mathcal{B} > \mathbb{E}(Y/\mathcal{B}) + \varepsilon)$. It it sure that $B_0 \in \mathcal{B}$ and we have

$$\begin{aligned}\int_{B(\varepsilon)} \mathbb{E}(X/\mathcal{B})\ d\mathbb{P} &\geq \left(\int_{B(\varepsilon)} \mathbb{E}(Y/\mathcal{B}) + \varepsilon\right) d\mathbb{P} \\ &\geq \left(\int_{B(\varepsilon)} \mathbb{E}(Y/\mathcal{B})\ d\mathbb{P}\right) + \varepsilon\mathbb{P}(B(\varepsilon)).\end{aligned}$$

The two last formulas cannot hold together unless $\mathbb{P}(B(\varepsilon)) = 0$ for all $\varepsilon > 0$. By the Monotone convergence Theorem, we get that $\mathbb{P}(\mathbb{E}(X/\mathcal{B} > \mathbb{E}(Y/\mathcal{B}) = 0$, which proves Inequality (CE03). To prove Inequality (CE04), we apply (CE03) and Point (4) in the exercise in Section 1 to $X \leq |X| = X^+ + X^-$ and to $-X \leq |X| = X^+ + X^-$, we get

$$\begin{aligned} |\mathbb{E}(X/\mathcal{B})| &= \max(-\mathbb{E}(X/\mathcal{B}), \mathbb{E}(X/\mathcal{B}) \\ &= \max(\mathbb{E}(-X/\mathcal{B}), \mathbb{E}(X/\mathcal{B}) \leq \mathbb{E}(|X|/\mathcal{B}). \end{aligned}$$

(13) Jensen's Inequality for Conditional Mathematical Expectations. Let $\mathcal{B}$ be a σ-sub-algebra of $\mathcal{A}$. Let ϕ be a convex function defined from a closed interval I of $\mathbb{R}$ to $\mathbb{R}$. Let X be a *rrv* with values in I such that $\mathbb{E}(X)$ is finite. Then $\mathbb{E}(X) \in I$ and

$$\phi(\mathbb{E}(X/\mathcal{B})) \leq \mathbb{E}(\phi(X)/\mathcal{B}).$$

Proof. It will be given on Chapter 8, Theorem 23 (See page 290) □

(14) Kolmogorov's Theorem for sums independent random variables.

Let X_1, ..., X_n be independent centered and square integrable random variables. We denote $\mathbb{V}ar(X_i) = \sigma_i^2$, $1 \leq i \leq n$. Let c be a non-random number (possibly infinite) satisfying

$$\sup_{1 \leq k \leq n} |X_k| \leq c \ a.s.$$

Denote the partial sums by

$$S_0 = 0, \ S_k = \sum_{i=1}^{k} X_i, \ k \geq 1 \text{ and } s_0 = 0, \ s_k^2 = \sum_{i=1}^{k} \sigma_i^2.$$

We have the double inequality, for any ε

$$1 - \frac{(\varepsilon + c)^2}{s_n^2} \leq \mathbb{P}(max(|S_1|, |S_2|, ..., |S_n|) \geq \varepsilon) \leq \varepsilon^{-2} s_n^2. \ (KM01)$$

Proof. We follow the proof in Loève (1997). Let $\varepsilon > 0$ and put

$$A_0 = \Omega, \ A_1 = (|S_1| < \varepsilon), \ A_k = (|S_1| < \varepsilon, ..., |S_k| < \varepsilon), \ k \geq 2.$$

We easily see that the sequence $(A_k)_{1\leq k\leq n}$ is non-increasing and we have

$$B_2 = A_1 \setminus A_2 = (|S_1| < \varepsilon, |S_2| \geq \varepsilon),$$

$$B_k = A_{k-1} \setminus A_k = (|S_1| < \varepsilon, ..., |S_{k-1}| < \varepsilon, |S_k| \geq \varepsilon), \ \ k \geq 3.$$

We also have

$$A_n^c = \sum_{1\leq k\leq n} B_k.$$

To see this quickly, say that $A_n^c = \cup_{1\leq k\leq n} C_k$, where $C_k = (|S_k| \geq \varepsilon)$. We are now accustomed to how rendering a union into a sum of sets since the course of Measure Theory and Integration by taking $D_1 = C_1$, $D_2 = C_1^c \cap C_2$, $D_k = C_1^c \cap ... \cap C_{k-1}^c C_k$, $k \geq 3$ to have

$$\bigcup_{1\leq k\leq n} C_k = \sum_{1\leq k\leq n} D_k.$$

We have just to check that the D_k's are exactly the B_k's. In the coming developments, we repeatedly use the fact that an indication function is equal to any of its positive power. Now, for any $1 \leq k \leq n$, we may see that $S_k 1_{B_k}$ is independent of $S_n - S_k$ (even when k=n with $S_n - S_k = 0$). Reminding that the S_k's are centered, we have

$$\begin{aligned}
\int_{B_k} S_n^2 \, d\mathbb{P} &= \mathbb{E}(S_n 1_{B_k})^2 \\
&= \mathbb{E}\biggl(S_k 1_{B_k} + (S_n - S_k)1_{B_k}\biggr) \\
&= \mathbb{E}(S_k 1_{B_k})^2 + \mathbb{E}((S_n - S_k)1_{B_k})^2 + 2\mathbb{E}((S_k 1_{B_k})(S_n - S_k)) \ \ (L02) \\
&= \mathbb{E}(S_k 1_{B_k})^2 + \mathbb{E}((S_n - S_k)1_{B_k})^2 \ \ (L03) \\
&\geq \mathbb{E}(S_k 1_{B_k})^2 \geq \varepsilon^2 \mathbb{P}.(B_k).
\end{aligned}$$

Line (L3) derives from Line (L2) by the fact that $(S_k 1_{B_k})$ and $(S_n - S_k)$ are independent and $S_n - S_k$ is centered. Hence, we get for each $1 \leq k \leq n$,

$$\int_{B_k} S_n^2 \, d\mathbb{P} \geq \varepsilon^2 \mathbb{P}(B_k).$$

By summing both sides over $k \in \{1, ..., k\}$ and by using the decomposition of A_n^c into the B_k's, we get

$$\int_{A_n^c} S_n^2 \, d\mathbb{P} \geq \varepsilon^2 \mathbb{P}(A_n).$$

which, by the simple remark that

$$\sum_{i=1}^{k} \sigma_i^2 = s_n^2 = \int S_n^2 \, d\mathbb{P}$$

leads to

$$\sum_{i=1}^{k} \sigma_i^2 \geq \varepsilon^2 \mathbb{P}(A_n),$$

which is the right-side inequality in Formula (KM01).

To prove the left-side inequality, let us start by remarking that for $2 \leq k \leq n$,

$$S_k 1_{A_{k-1}} = S_{k-1} 1_{A_{k-1}} + X_k 1_{A_{k-1}} = S_k 1_{A_k} + S_k 1_{B_k}.$$

Now, on one side, we have

$$\begin{aligned}
\mathbb{E}(S_{k-1} 1_{A_{k-1}} + X_k 1_{A_{k-1}})^2 &= \mathbb{E}(S_{k-1} 1_{A_{k-1}})^2 \\
&+ \mathbb{E}(X_k 1_{A_{k-1}})^2 + 2\mathbb{E}((S_{k-1} 1_{A_{k-1}}) X_k) \quad (L11) \\
&= \mathbb{E}(S_{k-1} 1_{A_{k-1}})^2 + \mathbb{E}(X_k 1_{A_{k-1}})^2 \quad (L12) \\
&= \mathbb{E}(S_{k-1} 1_{A_{k-1}})^2 + \sigma_k^2 \mathbb{P}(A_{k-1}). \quad (L13)
\end{aligned}$$

Line (L12) derives from Line (L11) since $S_{k-1} 1_{A_{k-1}}$ and X_k are independent and X_k is centered. Line (L13) derives from Line (L12) since X_k^2 is independent of $1_{A_{k-1}}$.

On the other side, we have

$$\begin{aligned}
\mathbb{E}(S_k 1_{A_k} + S_k 1_{B_k})^2 &= \mathbb{E}(S_k 1_{A_k})^2 + \mathbb{E}(S_k 1_{B_k})^2 + 2\mathbb{E}((S_k S_k)(1_{A_k} 1_{B_k})) \\
&= \mathbb{E}(S_k 1_{A_k})^2 + \mathbb{E}(S_k 1_{B_k})^2,
\end{aligned}$$

since the sets A_k and B_k are disjoint [recall that $B_k = A_{k-1} \setminus A_k = A_{k-1} \cap A_k^c \subset A_k^c$].

We get for $2 \leq k \leq n$,

$$\mathbb{E}(S_{k-1}1_{A_{k-1}})^2 + \sigma_k^2 \mathbb{P}(A_{k-1}) = \mathbb{E}(S_k 1_{A_k})^2 + \mathbb{E}(S_k 1_{B_k})^2. \quad (KM02)$$

But the expression $S_k 1_{B_k}$, which is used in last term in the right-hand member in Formula (KM02) is bounded as follows

$$|S_k 1_{B_k}| \leq |S_{k-1} 1_{B_k}| + |X_k 1_{B_k}| \leq (\varepsilon + c) 1_{B_k}.$$

Hence the last term in the right-hand member in Formula (KM02) itself is bounded as follows

$$\mathbb{E}(S_k 1_{B_k})^2 \leq (\varepsilon + c)^2 \mathbb{P}(B_k).$$

Further, we may bound below the last term in the left-hand member in Formula (KM02) by $s_k^2 \mathbb{P}(A_n)$, to get for $2 \leq k \leq n$

$$\mathbb{E}(S_{k-1}1_{A_{k-1}})^2 + \sigma_k^2 \mathbb{P}(A_n) \leq \mathbb{E}(S_k 1_{A_k})^2 + (\varepsilon + c)^2 \mathbb{P}(B_k). \quad (KM03)$$

Now, we may sum over $k \in \{2, ..., n\}$ in both sides to get in the left-hand side

$$\sum_{k=1}^{n-1} \mathbb{E}(S_k 1_{A_k})^2 + \sum_{k=2}^{n} \sigma_k^2 \mathbb{P}(A_n) \quad (KM03a)$$

and in the right-hand side, by rigorously handling the ranges of summation and by using the decomposition of A_n's into the B_k's, we have

$$\begin{aligned}
& \sum_{k=1}^{n} \mathbb{E}(S_k 1_{A_k})^2 - \mathbb{E}(S_1 1_{A_1})^2 + (\varepsilon + c)^2 \mathbb{P}(A_n^c \setminus B_1) \\
= & \sum_{k=1}^{n} \mathbb{E}(S_k 1_{A_k})^2 - \mathbb{E}(S_1 1_{A_1})^2 + (\varepsilon + c)^2 (\mathbb{P}(A_n^c) - \mathbb{P}(B_1)) \\
\leq & \sum_{k=1}^{n} \mathbb{E}(S_k 1_{A_k})^2 - \mathbb{E}(S_1 1_{A_1})^2 + (\varepsilon + c)^2 \mathbb{P}(A_n^c) - (\varepsilon + c)^2 \mathbb{P}(B_1). \quad (KM03b))
\end{aligned}$$

By moving the first term in (KM03a) to the right-hand member in (KM03) and by moving the terms in (KM03b) which are preceded by a minus sign to the left-hand member in (KM03) and by reminding that $B_1 = A^c$ and $S_1 = X_1$, we get

$$a + \sum_{k=2}^{n} \sigma_k^2 \mathbb{P}(A_n) \leq \mathbb{E}(S_n 1_{A_n})^2 + (\varepsilon + c)^2 \mathbb{P}(A_n^c), \quad (KM03c)$$

where

$$a = \mathbb{E}(X_1 1_{A_1})^2 + (\varepsilon + c)^2 \mathbb{P}(A_1^c).$$

But, since $|X_1| \leq c$, we have

$$\begin{aligned} \sigma_1^2 &= \mathbb{E}(X_1^2) = \mathbb{E}(X_1 1_{A_1})^2 + \mathbb{E}(X_1 1_{A_1^c})^2 \\ &\leq \mathbb{E}(X_1 1_{A_1})^2 + c^2 \mathbb{P}(A_1^c) \\ &\leq \mathbb{E}(X_1 1_{A_1})^2 + (\varepsilon + c)^2 \mathbb{P}(A_1^c) = a. \end{aligned}$$

Since $s_1^2 \mathbb{P}(A_n) \leq s_1^2 \leq a$, we may bound below a by $s_1^2 \mathbb{P}(A_n)$ in Formula (KM03c) to set

$$\begin{aligned} \left(\sum_{k=1}^{n} \sigma_k^2 \right) \mathbb{P}(A_n) &\leq \mathbb{E}(S_n 1_{A_n})^2 + (\varepsilon + c)^2 \mathbb{P}(A_n^c) \\ &\leq \varepsilon^2 \mathbb{P}(A_n) + (\varepsilon + c)^2 \mathbb{P}(A_n^c) \\ &\leq (\varepsilon + c)^2 \mathbb{P}(A_n) + (\varepsilon + c)^2 \mathbb{P}(A_n^c) = (\varepsilon + c)^2, \end{aligned}$$

which implies

$$\left(\sum_{k=1}^{n} \sigma_k^2 \right) (1 - \mathbb{P}(A_n^c)) \leq (\varepsilon + c)^2,$$

and hence

$$\mathbb{P}(A_n^c) \geq 1 - \frac{(\varepsilon + c)^2}{\sum_{k=1}^{n} \sigma_k^2},$$

which is the first inequality in Formula (KM01). The proof is complete now. ■

(15) Maximal inequality for sub-martingales.

Let X_1, ..., X_n be rel-valued integrable random variables. Let us consider the following sub-σ-algebras : for $1 \leq k \leq n$,

$$\mathcal{B}_k = \sigma(\{X_j^{-1}(B),\ 1 \leq j \leq k,\ B \in \mathcal{B}_\infty(\overline{\mathbb{R}})\}).$$

In clear, each $\mathcal{B}_k$ is the smallest σ-algebra rendering measurable the mapping X_j, $1 \leq j \leq k$. It is also clear that $(\mathcal{B}_k)_{1\leq k\leq n}$ is an non-decreasing sequence of sub-σ-algebras of $\mathcal{A}$.

Definition. The sequence $(X_k)_{1\leq k\leq n}$ is a martingale if and only if

$$\forall 1 \leq k_1 \leq k_2 \leq n,\ \forall A \in \mathcal{B}_{k_1}),\ \int_A X_{k_2}\ d\mathbb{P} = \int_A X_{k_1}\ d\mathbb{P},$$

and is a sub-martingale if and only if

$$\forall 1 \leq k_1 \leq k_2 \leq n,\ \forall A \in \mathcal{B}_{k_1}),\ \int_A X_{k_2}\ d\mathbb{P} \geq \int_A X_{k_1}\ d\mathbb{P}.\ \Diamond.$$

Let us adopt the notations given in Inequality (11).

If $(X_k)_{1\leq k\leq n}$ is a sub-martingale, we have

$$\mathbb{P}(\max(X_1, X_2, ..., X_n) \leq \varepsilon) \leq \varepsilon^{-1}\mathbb{E}(X_n).(IM01)$$

Proof. It is clear that

$$C = (\max(X_1, X_2, ..., X_n) \geq \varepsilon) = \bigcup_{1\leq k\leq n} (X_j \geq \varepsilon) = \sum_{1\leq k\leq n} C_k,$$

with

$$C_1 = (X_1 \geq \varepsilon),\ C_2 = (X_1 < \varepsilon, X_2 \geq \varepsilon),\ C_k = (X_1 < \varepsilon, ..., X_{k-1} < \varepsilon, X_k \geq \varepsilon),\ k \geq 3.$$

We remark that $C_k \in \mathcal{B}_\ell$, for all $1 \leq k \leq n$, $k \leq \ell$. We have

$$\begin{aligned}
\mathbb{E}(X_n) &= \int X_n \, d\mathbb{P} \geq \int_X X_n \, d\mathbb{P} \quad (L51) \\
&= \sum_{1 \leq k \leq n} \int_{C_k} X_n \, d\mathbb{P} \\
&\geq \sum_{1 \leq k \leq n} \int_{C_k} X_k \, d\mathbb{P} \quad (L53) \\
&\geq \sum_{1 \leq k \leq n} \varepsilon \mathbb{P}(C_k) \quad (L54) \\
&= \varepsilon \sum_{1 \leq k \leq n} \mathbb{P}(C_k) = \varepsilon \mathbb{P}(C). \quad (L55)
\end{aligned}$$

In Line (53), we applied the definition of a sub-martingale. In Line (L54), we applied that $X_k \geq \varepsilon$ on C_k. Finally, the combination of Lines (L51) and (L55) gives

$$\mathbb{P}(C) \leq \varepsilon^{-1} \mathbb{E}(X_n),$$

which is Formula (MT01). ■

(16) - Kolmogorov's Exponential bounds.

Let us fix an integer n such that $n \geq 1$. Suppose that we have n independent and centered random variables on the same probability space, as previously, which is *a.s.* bounded. As usual S_n is the partial sum at time n with variance s_n^2. We fix n such that $s_n > 0$. Define

$$c = \max_{1 \leq k \leq n} \frac{X_k}{s_n} < +\infty.$$

The following double inequality which is proved in Point (A1) in Chapter 10, Section 5 (page 360) will be instrumental in our proofs :

$$\forall t \in \mathbb{R}_+, \ e^{t(1-t)} \leq 1 + t \leq e^t. \ (EB1)$$

Now, let us begin by the following Lemma, which is part, of the body of exponential bounds.

LEMMA 3. *Let X be a centered random variable which is bounded, in absolute value, by $c < +\infty$. Let us denote $\mathbb{E}X^2 = \sigma^2$. Then for any $t > 0$ such that $tc \leq 1$, we have*

$$\mathbb{E}e^{tX} < \exp\left(\frac{t^2\sigma^2}{2}\left(1+\frac{tc}{2}\right)\right) \ (EB2)$$

and

$$\mathbb{E}e^{tX} > \exp\left(\frac{t^2\sigma^2}{2}\left(1-tc\right)\right). \quad (EB3)$$

Proof of Lemma 3. We begin to remark that the *mgf* $t \mapsto \mathbb{E}e^{tX}$ admits an infinite expansion on the whole real line of the form

$$\mathbb{E}e^{tX} = 1 + \frac{t^2}{2!}\mathbb{E}X^2 + \frac{t^3}{3!}\mathbb{E}X^3 + \dots.$$

For $t > 0$ and $tc \leq 1$, we have $\mathbb{E}X^{2+\ell} \leq \sigma^2 c^\ell$ for $\ell > 0$. Hence

$$\begin{aligned}\mathbb{E}e^{tX} &= 1 + \frac{t^2\sigma^2}{2}\left(1 + 2\left(\frac{tc}{3!} + \frac{(tc)^2}{4!} + \dots\right)\right)\\ &= 1 + \frac{t^2\sigma^2}{2}\left(\sum_{k\geq 3}\left(\frac{(tc)^{k-2}}{k!}\right)\right).\end{aligned}$$

Hence, by using the left inequality in Formula (EB1), we have

$$\begin{aligned}&\mathbb{E}e^{tX} - \left(1 + \frac{(t^2\sigma^2}{2}\left(1+\frac{tc}{2}\right)\right)\\ \leq\ & t^2\sigma^2\left(2\sum_{k\geq 3}\left(\frac{(tc)^{k-3}}{k!} - \frac{tc}{2}\right)\right.\\ =\ & 2t^2(tc)\sigma^2\left(\sum_{k\geq 3}\left(\frac{(tc)^{k-3}}{k!} - \frac{1}{4}\right)\right.\\ \leq\ & 2t^2(tc)\sigma^2\left(\sum_{k\geq 3}\left(\frac{1}{k!} - \frac{1}{4}\right)\right. \ (L23)\\ \leq\ & 2t^2(tc)\sigma^2(e - 7/4) \leq 0,\end{aligned}$$

where we used $tc \leq 1$ in Line (L23). Hence

$$\mathbb{E}e^{tX} \leq \left(1 + \frac{(t^2\sigma^2}{2}\left(1+\frac{tc}{2}\right) \leq \exp\left(\frac{t^2\sigma^2}{2}\left(1+\frac{tc}{2}\right)\right),$$

which proves Formula (EB2).

To prove the left-hand inequality, we remark that $\mathbb{E}X^{2+\ell} \geq \sigma^2(-c)^\ell$ for $\ell > 0$, we also have

$$\begin{aligned}\mathbb{E}e^{tX} &\geq 1 + \frac{t^2\sigma^2}{2}\left(1 + 2\left(\frac{-tc}{3!} + \frac{(-tc)^2}{4!} + \dots\right)\right) \\ &\geq 1 + \frac{t^2\sigma^2}{2}\left(1 - 2\left(\frac{tc}{3!} - \frac{(tc)^2}{4!} + \dots\right)\right)\end{aligned}$$

The same method, word by word, leads to

$$\mathbb{E}e^{tX} \geq -2t^2(tc)\sigma^2(e - 7/4) \geq 0$$

and next, by using the right inequality in Formula (EB1), we get

$$\mathbb{E}e^{tX} \geq \left(1 + \frac{(t^2\sigma^2}{2}\left(1-\frac{tc}{2}\right) \geq \exp\left(\frac{(t^2\sigma^2}{2}\left(1-tc\right)\right),$$

which establishes Formula (EB3). □

Here the first result concerning the exponential bounds.

Theorem 12. *Let us use the same notations as in Lemma 3. Then the assertions below hold true, for any $\varepsilon > 0$, for any $n \geq 1$.*

(i) for $c\varepsilon \leq 1$,

$$\mathbb{P}\left(S_n > \varepsilon s_n)\right) < \exp\left(-\frac{\varepsilon^2}{2}\left(1 - \frac{\varepsilon c}{2}\right)\right).$$

(ii) and for $c\varepsilon > 1$,

$$\mathbb{P}\left(S_n > \varepsilon s_n)\right) < \exp\left(-\frac{\varepsilon^2}{4c}\right).$$

Proof of Theorem 12. To make the notation shorter, we put $S = S_n$ and $s_n = s$ and some times $S^* = S/s$. Now let us apply Formulas (EB2) and (EB3) in Lemma 3 in the following way : for $t > 0$ and $tc \leq 1$, and since

$$\mathbb{E}\exp(tS^*) = \prod_{1\leq k\leq n} \mathbb{E}\exp(tX_k/s),$$

we have

$$\prod_{1\leq k\leq n} \exp\biggl(\frac{(t^2\sigma_k^2}{2s^2}\,(1-tc)\biggr) < \mathbb{E}\exp(tS^*) \prod_{1\leq k\leq n} \exp\biggl(\frac{(t^2\sigma_k^2}{2s^2}\,\biggl(1+\frac{tc}{2}\biggr)\biggr)$$

This obviously leads to

$$\exp\biggl(\frac{(t^2}{2}\,(1-tc)\biggr) < \mathbb{E}\exp(tS^*)$$
$$\exp\biggl(\frac{(t^2}{2}\,\biggl(1+\frac{tc}{2}\biggr)\biggr). \quad (DE)$$

From this, we are able to handle both Points (i) or (ii).

For (i), we may apply the the Markov inequality and left-hand inequality in Formula (DE) above to $t > 0$, ε 0 such that $c\varepsilon \leq 1$ and $tc \leq 1$, to get

$$\begin{aligned} \mathbb{P}\,(S^* > \varepsilon) &= \mathbb{P}\,(\exp(tS^*) > \exp(t\varepsilon)) \\ &\leq \exp(-t\varepsilon)\mathbb{E}\exp(tS^*) \\ &\leq \exp\biggl(-t\varepsilon\frac{(t^2}{2}\,\biggl(1+\frac{tc}{2}\biggr)\biggr). \ (L23) \end{aligned}$$

We point out that the condition $tc \leq 1$ intervenes only in the conclusion in Line (L23). Taking $t = \varepsilon$ in in Line (L23) (which is possible since both conditions $tc \leq 1$ and $\varepsilon c \leq 1$ hold) leads to

$$\begin{aligned} \mathbb{P}\,(S^* > \varepsilon) &< \exp\biggl(-\varepsilon^2 + \frac{\varepsilon^2}{2}\,\biggl(1+\frac{\varepsilon c}{2}\biggr)\biggr) \\ &= \exp\biggl(-\frac{\varepsilon^2}{2}\,\biggl(1-\frac{\varepsilon c}{2}\biggr)\biggr), \end{aligned}$$

which is the announced result for Point (i).

To prove Point (ii), let $c\varepsilon > 1$, we use the value $t = 1/c$ (here again, the condition $tc \leq 1$ holds) to get

$$\begin{aligned}\mathbb{P}\left(S^* > \varepsilon\right) &< \exp\biggl(-\frac{\varepsilon}{c} + \frac{1}{2c^2}\biggl(1 + \frac{1}{2}\biggr)\biggr) \\ &= \exp\biggl(-\frac{\varepsilon}{4c}\biggr),\end{aligned}$$

which is the announced result for Point (ii). □

Finally, the coming exponential bound is very important when dealing with the *Law of iterated logarithm* (LIL). We have :

THEOREM 13. *Let us use the same notation as in Lemma 3.*

Let us fix $0 < \alpha < 1/4$, *we set* $\beta = 2\sqrt{\alpha}$ *and*

$$\gamma = \frac{1 + 2\alpha + \beta^2/2}{(1-\beta)^2} - 1 > 0.$$

Then there exists $t(\alpha)$ *large enough such that for* $c(\alpha)$ *small enough, that is* $c(\alpha) < \alpha/t$ *and* $8c(\alpha)t(\alpha) \leq 1$ *such that for* $\varepsilon = t(\alpha)(1-2\sqrt{(\alpha)})$ *we have*

$$\mathbb{P}\biggl(S_n > \varepsilon s_n)\biggr) < \exp\biggl(-\frac{\varepsilon^2}{4c}(1+\gamma)\biggr).\Diamond$$

Proof of Theorem 13. The proof is so really technical that some authors like Gutt (2005) omitted and explained : *this one is no pleasure to prove* it. He referred to Stout (1974).

Here, we will follow the lines of the proof in Loève (1997). However, the presentation and the ordering of the arguments have been significantly improved.

From Formula (DE), we may fix $0 < \alpha = t_0 c < 1$ so that for all $t \leq t_0$, we have

$$\mathbb{E}\exp(tS^*) > \exp\biggl(\frac{t^2}{2}(1-\alpha)\biggr).\ (EB4)$$

The principle of all the proof is to fix first $t > 0$, as large as necessary, and to choose c so that the desired conclusions hold. Then, let us choose α such that $2\sqrt{\alpha} < 1$. Put

$$(i) \quad \beta = 2\sqrt{\alpha},$$

We have

$$(ii) \quad \gamma = \frac{1 + 2\alpha + \beta^2/2}{(1-\beta)^2} = \frac{(1+\beta)^2 + 2\beta + 1}{2(1-\beta)} - 1 > 0.$$

The positivity of γ is clear since $0 < \beta < 1$ and $\gamma > 0$. We first choose $0 < \alpha < 1/4$ which guarantees that $1 + \beta < 2$. Formula (EB4) shows that when α is fixed, the following conditions make sense : For t large enough, we have

$$(iii) \quad \left(8t^2 \exp\left(-\frac{\alpha t^2}{4})\right)\right) < 1/4, \quad (ii) \quad \mathbb{E}e^{tS^*} > 8,$$

$$(iv) \quad \frac{1}{4}\mathbb{E}e^{tS^*} > 2 \quad and \quad (v) \quad \frac{1}{4t^2}\exp\left(\frac{t^2}{2}\alpha\right) > 1.$$

We choose a value $t > 0$ satisfying points (iii), (iv), (v). Next we suppose that

$$(vi) \quad c < \alpha/t \quad (vii) \quad 8tc \leq 1 \quad and \ (viii) \quad c \leq 4t/(1-\beta).$$

Once these conditions are set, we may proceed to the proof. First of all, Formula (EB4) is justified by condition (vi). Put $q(x) = \mathbb{P}(S^* > x)$. By Formula (CF), in Chapter 1, page 14, we have, for $t >$ and $Z = \exp(tS^*)$,

$$\mathbb{E}(Z) = \int_0^{+\infty} \mathbb{P}(Z > t)dt, \ (EB5)$$

We get

$$\begin{aligned}\mathbb{E}(\exp(tS^*) &= \int_0^{+\infty} \mathbb{P}(\exp(tS^* > y)dy \\ &= \int_0^{+\infty} \mathbb{P}\left(S^* > \frac{\log y}{t}\right) dy \\ &= \int_0^{+\infty} q\left(\frac{\log y}{t}\right) dy \\ &= t\int_{-\infty}^{+\infty} e^{tx} q(x)dx \ (L44)\end{aligned}$$

Now we split the integral in Line (L44) above by decomposing the integration domain $I =]-\infty, +\infty[$ using the intervals $I_1 =]-\infty, 0]$, $I_2 =]0, (1-\beta)]$, $I_3 =]t(1-\beta), t(1+\beta)]$, $I_4 =]t(1+\beta), 8t]$, and $I_5 =]8t, +\infty[$, that is

$$I = I_1 + I_2 + I_3 + I_4 + I_5.$$

Let us name the integrals over I_i by J_i, $i \in \{1, ..., 5\}$, respectively

Let us begin by J_5. Let $s \in I_5$. We have for $0 \leq xc \geq 1$, by Formula (DE) and by Condition (vii)

$$q(x) < \exp\left(-\frac{x}{4c}\right) < \exp(-2tx),$$

where we use $1/(4c) = 2t/(8tc) \geq 2t$. If $xc < 1$, we apply again Formula (DE) and use $1 - xc/2 \geq 1/2$ in the middle member, to get

$$q(x) < \exp\biggl(-\frac{x^2}{2}\left(1-\frac{xc}{2}\right)\biggr) < \exp\left(-\frac{x}{4c}\right) < \exp(-2tx)$$

We get

$$J_5 = t\int_{8t}^{ta_n} e^{tx} q(x)dx \leq \int_{8t}^{+\infty} e^{tx}\exp(-2tx)dx <\leq \int_0^{+\infty} e^{-tx}dx = 1.$$

Since q is a bounded by one, we have

$$J_1 = J_5 = t\int_{-\infty}^{0} e^{tx} q(x)dx = t\int_0^{+\infty} e^{-tx} q(x)dx \leq t\int_0^{+\infty} e^{-tx}dx = 1.$$

Now we handle J_2 and J_4 by using a maximization argument. On I_4 and I_4, we have $x \leq 0$ and $xc \leq 8tc < 1$. From Point (ii) of Theorem 12, and by using again $xc \leq 8tc$ in the second inequality below, we arrive at

$$e^{tx}q(x) < \exp\biggl(tx - \frac{x^2}{2}\left(1 - \frac{xc}{2}\right)\biggr) \leq \exp\biggl(tx - \frac{x^2}{2}(1-4tc)\biggr) \equiv g(x),$$

where we remind that $4tc < 1/4$. On $\mathbb{R}_+$, $g'(x) = t - x(1-4tc)$ and thus g attains its maximum at $x_0 = t/(1-4tc)$.

Where lies x_0? $x_0 > t(1-\beta)$ is equivalent to $-\beta/(1-\beta) < 4tc$ which is true. As well $x_0 \leq t(1+\beta)$ is equivalent to Condition (viii). Thus $x_0 \in J_3$. Hence on $I_2 =]0, (1-\beta)]$, g is non-decreasing and thus, for $x \in I_2$,

$$\begin{aligned}
g(x) &= g(t(1-\beta)) = t^2(1-\beta) - \frac{t^2(1-\beta)^2}{2}(1-4tc)\\
&= (1-\beta)\biggl(t^2 - \frac{t^2}{2}(1-\beta) + \frac{t^2}{2}8tc\frac{1-\beta}{2}\biggr)\\
&= \frac{t^2}{2}(1-\beta)\biggl(2 - (1-\beta) + \frac{1-\beta}{2}\biggr) \quad (\textit{we used Condition (iii)})\\
&= \frac{t^2}{2}(1-\beta)\biggl((1+\beta) + \frac{(1-\beta)}{2}\biggr)\\
&= \frac{t^2}{2}\biggl(1-\beta^2 + \frac{(1-\beta)^2}{2}\biggr)\\
&= \frac{t^2}{2}\biggl((1-\beta^2/2) - \frac{1}{2}(1+2\beta)\biggr).
\end{aligned}$$

It follows that

$$\begin{aligned}
J_2 &= t\int_0^{t(1-\beta)} e^{g(x)}dx \leq t\int_0^{(1+\beta)} e^{g(t(1-\beta))}dx\\
&\leq t^2(1-\beta)\exp\biggl((1-\beta^2/2) - \frac{1}{2}(1+2\beta)\biggr).
\end{aligned}$$

As well on $I_4 =]t(1+\beta), 8t]$ or on $I_4^* =]t(1+\beta), ta_n]$, g is non-decreasing and we have for $x \in I_4^* \cup I_4$,

$$\begin{aligned}
g(x) &= g(t(1+\beta)) = t^2(1+\beta) - \frac{t^2(1+\beta)^2}{2}(1-4tc) \\
&= (1+\beta)\left(t^2 - \frac{t^2}{2}(1+\beta) + \frac{t^2}{2}8tc\frac{1+\beta}{2}\right) \\
&= \frac{t^2}{2}(1+\beta)\left(2-(1+\beta)+\frac{1+\beta}{2}\right) \quad (\textit{we used Condition (iii)}) \\
&= \frac{t^2}{2}(1+\beta)\left((1-\beta)+\frac{(1+\beta)}{2}\right) \\
&= \frac{t^2}{2}\left(1-\beta^2+\frac{(1+\beta)^2}{2}\right) \\
&= \frac{t^2}{2}\left((1-\beta^2/2)-\frac{2\beta^2-2\beta+1}{2}\right),
\end{aligned}$$

since the polynomial $2\beta^2-2\beta+1$ has a negative discriminant and thus, is constantly positive. It follows that

$$\begin{aligned}
\max(J_4, J_4^*) &\leq t\int_{t(1+\beta)}^{8t} e^{g(x)}dx \\
&\leq t\int_0^{(1+\beta)} e^{g(t(1+\beta))}dx \\
&\leq 7 \quad t^2\exp\left((1-\beta^2/2)-\frac{1}{2}(1+2\beta)\right).
\end{aligned}$$

So, we have

$$\max(J_4, J_4^*) \quad < \quad 7t^2\exp\left(\left(1-\beta^2/2\right)\right).$$

Now we remind that α is fixed and $\alpha = \beta^2/4$ and hence

$$1-\beta^2/2 = (1-\alpha)+\alpha/2.$$

Hence Inequality (EB4) gives

$$\mathbb{E}e^{tX} > \exp\left(\frac{t^2}{2}(1-\alpha)\right) = \exp\left(\frac{t^2}{2}(1-\beta^2/2))\right)\exp\left(\frac{\alpha t^2}{4})\right)$$

and hence

$$\exp\left(\frac{\alpha t^2}{4}(1-\beta^2/2))\right) < \mathbb{E}e^{tS^*}\left(\frac{\alpha t^2}{4}(1-\beta^2/2))\right),$$

which, by using Conditions (iiia) and (iiib), leads to

$$J_2 + J_4 < \left(8t^2 \exp\left(\frac{\alpha t^2}{4})\right)\right)\mathbb{E}e^{tS^*}$$

and by Condition (iv), we get

$$J_1 + J_5) < 2 < \frac{1}{4}\mathbb{E}e^{tS^*} \ and \ J_2 + J_4, J_2 + J_4^*) < \frac{1}{4}\mathbb{E}e^{tS^*}.$$

Since $\mathbb{E}e^{tS^*} = J_1 + J_2 + J_3 + J_4 + J_5$, it follows

$$J_3 = t\int_{t(1-\beta)}^{t(1+\beta)} e^{tx}q(x)dx > \frac{1}{2}\mathbb{E}e^{tS^*}.$$

Now using the bound of $\mathbb{E}e^{tS^*}$ as in Formula (S) and using the non-increasingness of q and the non-decreasingness of $x \mapsto e^{tx}$ for $t > 0$, leads to

$$\frac{1}{2}\exp\left(\frac{t^2}{2}(1-\alpha)\right) < tq(t(1-\beta))\int_{t(1-\beta)}^{t(1+\beta)} e^{t^2(1+\beta)}dx,$$

that is

$$\frac{1}{2}\exp\left(\frac{t^2}{2}(1-\alpha)\right) < 2t^2q(t(1-\beta))e^{t^2(1+\beta)}$$

and next

$$q(t(1-\beta)) \geq \left[\frac{1}{4t^2}\exp\left(\frac{t^2}{2}\alpha\right)\right]\left[\exp\left(-t^2(1+\beta)+\frac{t^2}{2}(1-\alpha)-\frac{t^2}{2}\alpha\right)\right], \ (EB6)$$

with

$$\begin{aligned}
& \exp\left(\frac{t^2}{2}\alpha\right)\exp\Big(-t^2(1+\beta)\frac{t^2}{2}(1-\alpha)-\frac{t^2}{2}\alpha\Big) \\
= & \exp\left(\frac{t^2}{2}(1-\alpha-\alpha-2(1+\beta)\right) \\
= & \exp\left(\frac{t^2}{2}(1-2\alpha-2(1+\beta)\right) \\
= & \exp\left(-\frac{t^2}{2}(1+2\alpha+2\beta)\right) \ (EB7)
\end{aligned}$$

We take $t = \varepsilon/(1-\beta)$. The quantity between the big brackets is bounded below by one in virtue of Condition (v). From this the combination of (EB6) and (EB7) gives

$$q(\varepsilon) > \exp\left(-\frac{\varepsilon^2}{2}\frac{1+2\alpha+2\beta}{(1-\beta^2)}\right) = \exp\left(-\frac{\varepsilon^2}{2}\frac{1+2\beta+\beta^2/2}{(1-\beta^2)}\right)$$

and finally, by Condition (ii), we get

$$q(\varepsilon) > \exp\left(-\frac{\varepsilon^2}{2}(1+\gamma)\right),$$

which was the target. □

(17) - Billingsley's Inequality (See Billinsgley (1968), page 69).

Let $(X_n)_{n\geq 0}$ be a sequence of square integrable and centered real-valued random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We have for any $\varepsilon > \sqrt{2}$,

$$\mathbb{P}\left(\max_{1\leq k\leq n} S_n \geq \varepsilon\right) \leq 2\mathbb{P}\left(S_n \geq \varepsilon - \sqrt{2\mathbb{V}ar(S_n)}\right).$$

where, as usual, S_n, $n \geq 1$, are the partial sums of the studied sequence.

Proof. Put $s_k^2 = \mathbb{V}ar(S_k)$ for $k \geq 1$. As usual,

$$\begin{aligned}
A &= \left(\max_{1\leq k\leq n} S_n \geq \varepsilon s_n\right)\\
&= \sum_{1\leq j\leq n} (S_1 < \varepsilon s_n, \cdots, S_{j-1} < \varepsilon s_n, S_j \geq \varepsilon s_n)\\
&\equiv \sum_{1\leq j\leq n} A_j.
\end{aligned}$$

Now we have

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A \cap (S_n \geq (\varepsilon - \sqrt{2})s_n) + \mathbb{P}(A \cap (S_n < (\varepsilon - \sqrt{2})s_n)\\
&\leq \mathbb{P}(S_n \geq (\varepsilon - \sqrt{2})s_n) + \sum_{1\leq j\leq n} \mathbb{P}(A_j \cap (S_n < (\varepsilon - \sqrt{2})s_n)\\
&= \mathbb{P}(S_n \geq (\varepsilon - \sqrt{2})s_n) + \sum_{1\leq j\leq n-1} \mathbb{P}(A_j \cap (S_n < (\varepsilon - \sqrt{2})s_n),
\end{aligned}$$

since $\mathbb{P}(A_n \cap (S_n < (\varepsilon - \sqrt{2})s_n) = \emptyset$. We also have for each $1 \leq j < n$,

$$(S_j \geq \varepsilon_n)\, and\, (S_n \leq \varepsilon - \sqrt{2})s_n) \Rightarrow (S_n - S_j \geq \sqrt{2}s_n) \Rightarrow (|S_n - S_j| \geq \sqrt{2}s_n.)$$

Since we still have that $S_n - S_j = X_{j+1} + \cdots + X_{j+1}$, for $1 \leq j < n$, is independent of A_j, we get

$$\mathbb{P}(A_j \cap (S_n < (\varepsilon - \sqrt{2})s_n) \leq \mathbb{P}(A_j)\mathbb{P}(|S_n - S_j| \geq s_n\sqrt{2}),\ 1 \leq j < n.$$

Now using the Tchebychev inequality, we get

$$\begin{aligned}
\mathbb{P}(A) &\leq \mathbb{P}(S_n \geq (\varepsilon - \sqrt{2})s_n) + \sum_{1\leq j\leq n-1} \frac{s_n^2 - s_j^2}{2s_n^2})\\
&\leq \mathbb{P}(S_n \geq (\varepsilon - \sqrt{2})s_n) + \sum_{1\leq j\leq n-1} \frac{1}{2}\mathbb{P}(A_j)\\
&\leq \mathbb{P}(S_n \geq (\varepsilon - \sqrt{2})s_n) + \frac{1}{2}\sum_{1\leq j\leq n} \mathbb{P}(A_j)\\
&= \mathbb{P}(S_n \geq (\varepsilon - \sqrt{2})s_n) + \frac{1}{2}\mathbb{P}(A)
\end{aligned}$$

which leads to the desired result. □

18 - Etemadi's Inequality. Let $X_1, \cdots, \leq X_n$ be n independent real-valued random variables such that the partial sums $S_k = X_1 + ... + X_k$, $1 \leq k \leq n$, are definied. Then for any $\alpha \geq 0$, we have

$$\mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq 3\alpha\right) \leq 3 \max_{1\leq k\leq n} \mathbb{P}\left(|S_k| \geq \alpha\right). \diamond$$

proof. The formula is obvious for $n = 1$. Let $n \geq 2$. As usual, denote $B_1 = (|X_1| \leq 3\alpha)$, $B_k = (|S_1| < 3\alpha, \cdots, |S_{k-1}| < 3\alpha, |S_k| \geq\geq 3\alpha)$, $k \geq 2$. By decomposing $(\max_{1\leq j\leq n} |S_j| \geq 3\alpha)$ over the partition

$$(|S_n| \geq \alpha) + (|S_n| < \alpha) = \Omega,$$

we have

$$(\max_{1\leq j\leq n} |S_j| \geq 3\alpha) \subset (|S_n| \geq \alpha) \cup (|S_n| < \alpha, \max_{1\leq j\leq n} |S_j| \geq 3\alpha)$$

And by the principle of the construction of the B_j,

$$(\max_{1\leq j\leq n} |S_j| \geq 3\alpha) = \sum_{1\leq j\ \ n} B_j$$

and hence

$$(\max_{1\leq j\leq n} |S_j| \geq 3\alpha) \subset (|S_n| \geq \alpha) \cup \sum_{1\leq j\ \ n-1} (|S_n| < \alpha 3\alpha) \cap B_j$$

where the summation is restricted to $j \in \{1, ..., n-1\}$ since the event $(|S_n| < \alpha 3\alpha) \cap B_n$ is empty. Further, on $(|S_n| < \alpha) \cup B_j$, we have $(|S_n| < \alpha)$ and $(|S_j| < 3\alpha)$ and the second triangle inequality $|S_n - S_j| \geq |S_j| - |S_n| \geq 3\alpha - \alpha = 2\alpha$, that is

$$(|S_n| < \alpha) \cap B_j \subset B_j \cap (|S_n| < \alpha) \cap (|S_n - S_j| \geq 2\alpha) \subset B_j \cap (|S_n - S_j| \geq 2\alpha).$$

Now, we remind that and B_j and $S_n - S_j$ are independent. Translating all this into probabilities gives

$$\begin{aligned}\mathbb{P}(\max_{1\leq j\leq n} |S_j| \geq 3\alpha) &\leq \mathbb{P}(|S_n| \geq \alpha) + \sum_{1\leq j\ \ n-1} \mathbb{P}(B_j)\mathbb{P}(|S_n - S_j| \geq 2\alpha) \\ &\leq \mathbb{P}(|S_n| \geq \alpha) + \sum_{1\leq j\ \ n-1} \mathbb{P}(B_j)\Big(\mathbb{P}(|S_n| \geq \alpha) + \mathbb{P}(|S_j| \geq \alpha)\Big).\end{aligned}$$

But $(|S_n| \geq 2\alpha)$, $(|S_n| \geq \alpha)$ and $(|S_j| \geq 2\alpha)$ are subsets of

$$(\max_{1\leq j\leq n} |S_j| \geq \alpha)$$

and hence, we may conclude that

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq j\leq n} |S_j| \geq 3\alpha\right) &\leq \mathbb{P}\left(\max_{1\leq j\leq n} |S_j| \geq \alpha\right)\left(1+2\sum_{1\leq j\leq n}\mathbb{P}(B_j)\right)\\
&\leq \mathbb{P}\left(\max_{1\leq j\leq n} |S_j| \geq 3\alpha\right)\left(1+2\mathbb{P}\left(\sum_{1\leq jn} B_j\right)\right)\\
&\leq 3\mathbb{P}(\max_{1\leq j\leq n} |S_j| \geq 3\alpha). \ \square
\end{aligned}$$

CHAPTER 7

Introduction to Classical Asymptotic Theorems of Independent Random variables

1. Easy Introduction

We are going to quickly discover three classical types of well-known convergences which are related to sequences of independent random variables. In the sequel :

$(X_n)_{n\geq 0}$ is a sequence of centered real-valued random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. If the expectations $\mu_n = \mathbb{E}X_n$'s exist, we usually center the X_n's at their expectations by taking $X_n - \mu_n$ in order to have centered random variables. If the variances exist, we denote $\sigma_n^2 = \mathbb{V}ar(X_n)$ and

$$s_0^2 = 0,\ s_1^2 = \sigma_1^2,\ s_n^2 = \sigma_1^2 + ... + \sigma_n^2,\ n \geq 2.$$

The laws we will deal with in this chapter are related to the partial sums

$$S_0 = 0,\ S_n = X_1,\ S_n = X_1 + ... + X_n.$$

(a) Discovering the simplest Weak Law of Large Numbers (WLLN).

Suppose that the random variables X_n are independent and are identically distributed (*iid*) and have the common mathematical expectation μ. We are going to find the limit in probability of the sequence

$$\overline{X}_n = \frac{S_n}{n},\ \neq 0.$$

By Proposition 18 in Section 6 in Chapter 5, a non-random weak limit is also a limit in probability and vice-versa. So we may directly try to show that $\overline{X}_n$ converges to a non-random limit (which is supposed to be μ). To do this, we have many choices through the Portmanteau

Theorem 11 in Section 4 in Chapter 5. Let us use the characteristic function tool $\Phi_{X_j} = \Phi$ for all $j \geq 1$. Since we have, by Proposition 6 in Section 6 in Chapter 2,

$$\Phi'(0) = i\mu \ \ and \ \ \Phi(0) = 1,$$

(where i is the normed pure complex number with a positive angle), we may use a one order Taylor expansion of Φ at zero to have

$$\Phi(u) = 1 + i\mu u + O(u^2), \ as \ u \rightarrow 0. \ (EX)$$

By the properties of the characteristic function and by taking into account the fact that the variables are *iid*, we have

$$\Phi_{S_n/n}(u) = \Phi_{X_1+\ldots+X_n}(u/n) = \Phi(u/n)^n, \ u \in \mathbb{R}.$$

Now, for u fixed, we have $u/n \rightarrow 0$ as $n \rightarrow \infty$, and we may apply Formula (EX) to have, as $n \rightarrow +\infty$,

$$\Phi_{S_n/n}(u) = \exp\biggl(n \log(1 + i\mu u/n + O(n^{-2}))\biggr) \rightarrow \exp(i\mu u) = \Phi_\mu(u).$$

Here, we skipped the computations that lead to $n\log(1 + i\mu u/n + O(n^{-2}) \rightarrow i\mu u$. In previous books asLo (2017a) and Lo *et al.* (2016), such techniques based on expansions of the logarithm function have been given in details.

We just show that $S_n/n \rightsquigarrow \mu$, hence $S_n/n \rightarrow_{\mathbb{P}} \mu$. This gives us the first law.

THEOREM 14. *(Kintchine) If $(X_n)_{n\geq 0}$ is a sequence of independent and are identically distributed (iid) random variables with a finite common mathematical expectation μ, we have the following Weak Law of Large Numbers (WLLN) :*

$$S_n/n \rightarrow_{\mathbb{P}} \mu, \ as \ n \rightarrow +\infty.$$

(b) Discovering the Strong Law of Large Numbers (SLLN).

Before we proceed further, let us state a result of measure theory and integration (See Lo (2017b), given in Exercise 3 in Doc 04-05, and its solution in Doc 04-08) in Chapter 5 in Lo (2017b) in the following famous lemma.

LEMMA 4. ***(Borel-Cantelli Lemma)*** *Let* $(A_n)_{n\geq 0} \subset \mathcal{A}$.

(i) If the series $\sum_{n\geq 0} \mathbb{P}(A_n) < +\infty$ *is convergent, then*

$$\mathbb{P}\left(\limsup_{n\to+\infty} A_n\right) = 0.$$

(ii) If the events A_n *are independent and if the series diverges, that is*

$\sum_{n\geq 0} \mathbb{P}(A_n) = \infty$, *then*

$$\mathbb{P}\left(\limsup_{n\to+\infty} A_n\right) = 1.$$

This lemma is the classical basis of the simple *SLLN*. But before we continue, let us give the following consequence.

COROLLARY 1. *Let* $(X_n)_{n\geq 0}$ *be a sequence of independent a.e. finite real-valued random variables such that* $X_n \to 0$ *a.s as* $n \to +\infty$. *Then for any finite real number* $c > 0$,

$$\sum_{n\geq 0} \mathbb{P}(|X_n| \geq c) < +\infty.$$

Proof. Given the assumptions of the corollary, the events A_n's are independent. By the Borel-Cantelli Lemma, $\sum_{n\geq 0} \mathbb{P}(|X_n| \leq c) = +\infty$ would imply $\mathbb{P}(|X_n| > c, i.o) = 1$ and hence $(X_n \to 0)$ *a.e.* would be false. The proof is complete with this last remark. □

Let us expose the simple strong law of large number.

THEOREM 15. *(Simple Strong Law of Large Numbers) Let* $(X_n)_{n\geq 0}$ *be a sequence of independent centered and square integrable random variables with variance one, that is* $\mathbb{E}X_n^2 = 1$ *for all* $n \geq 1$. *Then*

$$\frac{1}{n} \sum_{1\leq k\leq n} X_k \to 0 \ a.s. \ as \ n \to +\infty.$$

* **Proof**. Suppose that the assumption of the theorem hold. We are going to use the perfect square method. Put

$$Y_n = S_{n^2}/n^2,\ n \geq 1,$$

that is, we only consider the elements of the sequence $(S_k/k)_{k\geq 1}$ corresponding to a square index $k = n^2$. Remark that $\mathbb{V}ar(Y_n) = n^{-2}$, $n \geq 2$. Fix $0 < \beta < 1/2$. By Chebychev's inequality, we have

$$\mathbb{P}(|Y_n| > n^{-\beta}) \leq n^{2(1-\beta)}$$

and thus,

$$\sum_n \mathbb{P}(|Y_n| > n^{-\beta}) \leq \sum n^{2(1-\beta)} < \infty.$$

By Borel-Cantelli's Lemma, we conclude that

$$\mathbb{P}(\liminf_n (|Y_n| \leq n^{-\beta}) = 1.$$

Let us remind that

$$\Omega_0 = \liminf_n (|Y_n| \leq n^{-\beta}) = \bigcup_{n\geq 0} \bigcap_{r\geq n} (|Y_r| \leq r^{-\beta}).$$

Hence, for all $\omega \in \Omega_0$, there exists $n(\omega) \geq 0$ such that for any $r \geq n$,

$$|Y_r| \leq r^{-\beta}.$$

* By the sandwich's rule, we conclude that, for any $\omega \in \Omega_0$, we have

$$Y_m(\omega) \to 0.$$

This means that

$$\Omega_0 \subset (Y_n \to 0).$$

We conclude that $\mathbb{P}(Y_n \to 0) = 1$ and hence $Y_n \to 0$, a.s..

To extend this result to the whole sequence, we use the decomposition of $\mathbb{N}$ by segments with perfect squares bounds. We have

$$\forall (n \geq 0), \exists m \geq 0, k(n) = m^2 \leq n \leq (\sqrt{k(n)} + 1)^2.$$

We have

$$\mathbb{E}(\frac{1}{n}(S_n - S_{k(n)}) = 0$$

and

$$\mathbb{V}ar(\frac{1}{n}(S_n - S_{k(n)}) = \frac{1}{n^2}E\sum_{i=k(n)+1}^{n} X_i^2 \leq \frac{1}{n^2}(2\sqrt{k(n)}+1) \leq \frac{3\sqrt{n}}{2} = 3n^{-3/2}.$$

Hence,

$$\sum_n \mathbb{P}(\left|\frac{1}{n}(S_n - S_{k(n)})\right| > n^{-\beta}) \leq 3\sum n^{-(\frac{3}{2}-2\beta)} < \infty$$

whenever $\beta < 3/4$. We conclude as previously that

$$\frac{1}{n}(S_n - S_{k(n)}) \rightarrow 0, \ a.s.$$

Finally we have

$$\frac{S_n}{n} = \frac{S_n - S_{k(n)}}{n} + \frac{k(n)}{k(n)} \times \frac{S_{k(n)}}{n} \rightarrow 0 \ a.s.,$$

since

$$1 \leq \frac{n}{k(n)} < 1 + \frac{2}{\sqrt{k(n)}} + \frac{1}{k(n)}$$

and

$$\frac{k(n)}{n} \rightarrow 1.$$

We just finished to prove that

$$\frac{S_n}{n} \rightarrow 0 \ a.s. \ \blacksquare$$

In a more general case of random variables with common variance, we may center and normalize them to be able to use the result above as in

COROLLARY 2. *Let $(X_n)_{n\geq 0}$ be a sequence of independent and square integrable random variables with equal variance $\sigma^2 > 0$,that is $\mathbb{V}ar(X_n) = \sigma^2$ for all $n \geq 1$. Then*

$$\frac{1}{n\sigma} \sum_{1\leq k\leq n} (X_k - \mathbb{E}(X_k)) \to 0 \ a.s. \ as \ n \to +\infty.$$

* We may also derive the

PROPOSITION 19. *(Kolmogorov) If $(X_n)_{n\geq 0}$ is a sequence of independent random variables with mathematical expectations μ_n and variances $0 < \sigma_n^2 < +\infty$, we have*

$$\frac{1}{n} \sum_{1\leq j\leq n} \frac{X_j - \mu_j}{\sigma_j} \to 0, \ a.s. \ as \ \to 0.$$

If the expectations are zero's that is $\mu_n = 0$, $n \geq 0$ and if the variances are equal, that $\sigma_n^2 = \sigma^2$, $n \geq 0$, we have the simple SLLN :

$$\frac{S_n}{n\sigma} \to 0, \ a.s. \ as \ \to 0.$$

(c) Discovering the Central limit Theorem.

The Central Limit Theorem in Probability Theory turns around finding conditions under which the sequence of partials sums S_n, $n \geq 1$, when appropriately centered and normalized, weakly converges to some random variable. Generally, the probability law of the limiting random variable is Gaussian.

Actually, we already encountered the *CTL* in our series, through Theorem 4 in Chapter 7 in Lo (2017a) in the following way.

If the X_n's are *iid* according to a Bernoulli probability law $\mathcal{B}(p)$, $0 < p < 1$, S_n follows a Biniomial laws of parameters $p = 1 - q$ and $n \geq 1$ and we have

$$Z_n = \frac{S_n - \mathbb{E}S_n}{\mathbb{V}ar(S_n)^{1/2}} = \frac{S_n - npq}{\sqrt{npq}}, \ n \geq 1.$$

The invoked theorem (in Lo (2017a)) states that, as $n \to +\infty$,

$$\forall x \in \mathbb{R}, \ F_{S_n}(x) \to N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-t^2/2) \ d\lambda(t).$$

At the light of the Portmanteau Theorem 11 in Section 4 in Chapter 5, we have

PROPOSITION 20. *Let $(X_n)_{n\geq 1}$, be a sequence of independent random variables identically distributed as a Bernoulli probability law $\mathcal{B}(p)$, $0 < p < 1$. Then we have the following Central limit Theorem (CLT)*

$$\frac{S_n - npq}{\sqrt{npq}} \rightsquigarrow \mathcal{N}(0,1), \text{ as } n \to +\infty.$$

We are going to see that result is a particular case the following one.

PROPOSITION 21. *(CLT for an iid sequence with finite variance). Let $(X_n)_{n\geq 1}$, be a sequence of centered and iid random variables with common finite variance $\sigma^2 > 0$. Then, we have the following CLT*

$$\frac{S_n}{s_n} \rightsquigarrow \mathcal{N}(0,1), \text{ as } n \to +\infty. \ (CLTG)$$

If the common expectation is μ, we may write

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1), \text{ as } n \to +\infty.$$

Proof. The Portmanteau Theorem 11 in Section 4 in Chapter 5 offers us a wide set of tools for establishing weak laws. In on dimensional problems, the characteristic method is the favored one. Here, we have $\Phi_{X_j} = \Phi$ for all $i \geq 1$. Let us give the proof for $\sigma = 1$. By Proposition 6 in Section 6 in Chapter 2, we have

$$\Phi(0) = 1, \quad \Phi'(0) = 0 \text{ and } \Phi''(0) = -1.$$

Let us use two-order Taylor expansion of Φ in the neighborhood of 0 to have :

$$\Phi(u) = 1 - u^2/2 + O(u^2), \text{ as } u \to 0. \ (EX2)$$

By the properties of the characteristic function and by taking into account that the variables are iid, we have

$$\Phi_{S_n/\sqrt{n}}(u) = \Phi_{X_1+\ldots+X_n}(u/\sqrt{n}) = \Phi(u/\sqrt{n})^n, \ u \in \mathbb{R}.$$

Now for u fixed, we have $u/\sqrt{n} \to 0$ as $n \to \infty$, and we may apply Formula (EX2) to have, for $n \to +\infty$,

$$\Phi_{S_n/\sqrt{n}}(u) = \exp\biggl(n\log(1 - u^2/(2n) + O(n^{-3/2}))\biggr) \to \exp(-u^2/2),$$

where again we skipped details on the expansions of the logarithm function. So we have just proved that

$$S_n/\sqrt{n} \rightsquigarrow \mathcal{N}(0,1).$$

If the common expectation is μ, we may transform the sequence to $((X_n - \mu)/\sigma)_{n\geq 1}$, which is an *iid* sequence of centered random variables with variance one. By applying the result above, we get

$$\frac{S_n}{\sigma\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1).$$

We finish the proof by noticing that : $s_n^2 = n\sigma^2$, $n \geq 1$. □

(d) A remark leading the Berry-Essen Bounds.

Once we have a *CLT* in the form of Formula (CLTG), the Portmanteau theorem implies that for any fixed $x \in \mathbb{R}$

$$\left|\mathbb{P}\left(\frac{S_n}{s_n} \leq x\right) - N(x)\right| \to 0 \ as \ n \to +\infty.$$

Actually, the formula above holds uniformly (See Fact 4 in Chapter 4 in Lo *et al.* (2016)), that is

$$B_n = \sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{S_n}{s_n} \leq x\right) - N(x)\right| \to 0 \ as \ n \to +\infty.$$

A Berry-Bound is any bound of B_n. We will see later in this chapter a Berry-Essen bound for sequence of independent random variables with third finite moments.

Conclusion.

Through Theorem 14 and Propositions 19 and 21, we discovered simple forms of three of the most important asymptotic laws in Probability

Theory.

Establishing *WLLN*'s, *SLLN*'s, *CLT*'s, Berrey-Essen bounds, etc. is still a wide and important part in Probability Theory research under a variety of dependence type and in abstract spaces.

For example, the extensions of such results to set-valued random variables constitute an active research field.

The results in this section are meaningful and are indeed applied. But we will give important more general cases in next sections. The coming results represent advanced forms for sequence of independent random variables.

2. Tail events and Kolmogorov's zero-one law and strong laws of Large Numbers

This chapter will be an opportunity to revise generated σ-algebras and to deepen our knowledge on independence.

(A) Introduction and statement of the zero-one law.

At the beginning, let $X = (X_t)_{t\in T}$ be an non-empty of mappings from $(\Omega, \mathcal{A})$ to some measure spaces $(F_t, \mathcal{F}_t)$. The σ-algebra on Ω generated by this family is

$$\mathcal{A}_X = \sigma\{X_{t_i}^{-1}(B_{t_i}),\ B_{t_i} \in \mathcal{F}_{s_i}, (t_1, ..., t_p) \in T^p,\ p \geq 1\}.$$

It is left as an exercise to check that $\mathcal{A}_X$ is also generated by the class of finite intersections of the form

$$\mathcal{C}_X = \{\bigcap_{1\leq k\leq n} X_{t_i}^{-1}(B_{t_i}),\ B_{t_i} \in \mathcal{F}_{s_i}, (t_1, ..., t_p) \in T^p,\ p \geq 1\}, \quad (P01)$$

which is a π-system.

Coming to our topic on the zero-one law, we already saw from the Borel-Cantelli Lemma 4 that : for a sequence of independent events $(A_n)_{n\geq 0} \subset \mathcal{A}$ on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $A_n \to A$ as $n \to +\infty$, then $\mathbb{P}(A) \in \{0, 1\}$.

We are going to see that this is a more general law called the Kolmogorov zero-one law. Let $(X_n)_{n\geq 1}$ be a sequence of measurable mappings from $(\Omega, \mathcal{A})$ to some measure spaces $(F_i, \mathcal{F}_i)$, $i \geq 1$. For each $n \geq 0$, the smallest σ-algebra on Ω rendering measurable all the mapping X_k, $k \geq n$, with respect to $\mathcal{F}$ is

$$\mathcal{A}_{tail,n} = \sigma\{X_k^{-1}(B_k),\ B_k \in \mathcal{F}_k,\ k \geq n\}.$$

It is usually denoted as $\mathcal{A}_{tail,n} = \sigma(X_k,\ k \geq n)$ and quoted as the σ-algebra generated by the mappings X_k, $k \geq n$.

Definition. The tail σ-algebra generated by the sequence $(X_n)_{n\geq 0}$, relatively to $\mathcal{F}$, is the intersection

$$\mathcal{A}_{tail} = \bigcap_{n\geq 0} \mathcal{A}_{tail,n}.$$

The elements of $\mathcal{A}_{tail}$ are, by definition, the *tail events* with respect to the sequence $(X_n)_{n\geq 0}$. $\diamondsuit$

Let us give an example. Let $B_k \in \mathcal{F}_k$, for $k \geq 1$. We have that

$$\liminf_{n\rightarrow+\infty}(X_n \in B_k) \in \mathcal{A}_{tail} \text{ and } \limsup_{n\rightarrow+\infty}(X_n \in B_k) \in \mathcal{A}_{tail}.$$

Here is why. Because of the increasingness of the $(\cap_{p\geq n} A_p)_{n\geq 0}$ (in n), we have for any fixed $n_0 \geq 0$,

$$\liminf_{n\rightarrow+\infty}(X_n \in B) = \bigcup_{n\geq 1}\bigcap_{p\geq n}(X_p \in B) = \bigcup_{n\geq n_0}\bigcap_{p\geq n}(X_p \in B)$$

and

$$\left\{\bigcup_{k\geq n}(X_n \in B),\ n \geq n_0\right\} \subset \mathcal{A}_{tail,n_0},$$

and then $\liminf_{n\rightarrow+\infty}(X_n \in B) \in \mathcal{A}_{tail,n_0}$ for all $n_0 \geq 0$ and is in $\mathcal{A}_{tail}$. To get the same conclusion for the superior limit, we applied that conclusion to its complement.

Let us prove a useful result before we proceed further.

The zero-one Law. If the sequence elements of the sequence $(A_n)_{n\geq 0}$ are mutually independent, then any tail A event with respect to that sequence is such that $\mathbb{P}(A) \in \{0, 1\}$, that is the tail σ-algebra is behaves as the trivial σ-algebra. $\diamondsuit$

Before we give the proof, let us get more acquainted with independent σ-algebras.

(B) Independence of σ-algebras.

Definition. Two non-empty sub-classes $\mathcal{C}_1$ and $\mathcal{C}_2$ of $\mathcal{A}$ are *mutually independent* if and only if : for any subsets $\{A_1, ..., A_{\ell_1}\} \subset \mathcal{C}_1$

and $\{B_1, ..., B_{\ell_2}\} \subset \mathcal{C}_2$, for any non-negative, real-valued and measurable functions h_i, $i \in \{1, 2\}$, defined on a domain containing $\{0, 1\}$, $h_1(1_{A_1}, ..., 1_{A_{\ell_1}})$ and $h_2(1_{B_1}, ..., 1_{B_{\ell_2}})$ are independent. $\Diamond$

For easy notation, let us denote by $\mathcal{I}(\mathcal{C}_1)$ the class of all elements of the form $h_1(1_{A_1}, ..., 1_{A_{\ell_1}})$ as described above. $\mathcal{I}(\mathcal{C}_2)$ is defined similarly.

Example. Let $\mathcal{F}_1 = \{X_t, t \in T\}$ a non-empty family of measurable mappings from $(\Omega, \mathcal{A})$ to some measure space $(F, \mathcal{F})$ and $\mathcal{F}_2 = \{Y_s, s \in S\}$ a non-empty family of measurable mappings from $(\Omega, \mathcal{A})$ to some measure space $(G, \mathcal{G})$. Suppose that any finite pairs of sub-families $(X_{t_j})_{1 \leq j \leq p}$ $(p \geq 1)$and $(Y_{s_j})_{1 \leq j \leq q}$ $(q \geq 1)$ of $\mathcal{F}_1$ and $\mathcal{F}_2$ respectively, the random vectors $(X_{t_1}, ..., X_{t_p})^t$ and $(Y_{s_1}, ..., Y_{s_q})^t$ are independent, that is

$$\mathbb{P}_{(X_{t_1},...,X_{t_p},Y_{s_1},...,Y_{s_q})} = \mathbb{P}_{(X_{t_1},...,X_{t_p})} \otimes \mathbb{P}_{(Y_{s_1},...,Y_{s_q})}. \quad (DE01)$$

The classes

$$\mathcal{C}_X = \{X_t^{-1}(B),\ B \in \mathcal{F},\ t \in T\} \text{ and } \mathcal{C}_Y = \{Y_s^{-1}(C),\ C \in \mathcal{G},\ s \in S\}$$

are independent. To see that, we consider two finite subsets of $\mathcal{C}_X$ and $\mathcal{C}_Y$ of the forms

$$(X_{t_1}^{-1}(B_1), ..., X_{t_p}^{-1}(B_p)) \text{ and } (Y_{s_1}^{-1}(C_1), ..., S_{t_q}^{-1}(C_q)).$$

where the $(B_j)_{1 \leq j \leq p} \subset \mathcal{F}$ and $(C_j)_{1 \leq j \leq q} \subset \mathcal{G}$ and, accordingly, two real-valued and measurable functions h_1 and h_1 of their indicators functions as

$$H_1 = h_1(1_{B_1}(X_{t_1}), ..., 1_{B_p}(X_{t_p}))) \text{ and } H_2 = h_1(1_{C_1}(Y_{s_1}), ..., 1_{C_q}(Y_{s_p}))).$$

So, the functions H_1 and H_2 are independent because of Formula (DE). $\Diamond$

For now, we need the two results in the next proposition.

Proposition 22. *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two mutually independent and non-empty π-sub-classes of $\mathcal{A}$. Consider the generated σ-algebra $\mathcal{A}_i = \sigma(\mathcal{C}_i)$, $i \in \{1, 2\}$.*

(1) Then for any $(A, B) \in \mathcal{A}_1 \times \mathcal{A}_2$, A and B are independent.

(2) $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent.

(3) For any non-negative and real-valued function Z_i, $i \in \{1, 2\}$, such that each Z_i is $\mathcal{C}_i$-measurable, we have

$$\mathbb{E}(Z_1 Z_2) = \mathbb{E}(Z_1)\mathbb{E}(Z_2). \ \Diamond$$

proof. We easily see that each $\mathcal{A}_i$, $i \in \{1, 2\}$, is also generated by the class of finite intersections of sets which are either elements of $\mathcal{C}_i$ or complements of elements of $\mathcal{A}_i$, denoted

$$\tilde{\mathcal{C}}_1 = \{ \bigcap_{1 \leq k \leq n} A_i, \ A_i \in \mathcal{C}_i \text{ or } A_i^c \in \mathcal{C}_i, \ 1 \leq i \leq p, \ p \geq 1\}.$$

Also, for example, we already learned in Chapter 1 (Subsection 3 in Section 3.2) how to choose h_1 such that $\mathbb{E}h_1(1_{A_1}, ..., 1_{A_{\ell_1}})$ be of the form

$$\mathbb{P}(A'_1, ..., A'_i, ..., A'_{\ell_1}), \quad (IN02)$$

where $A'_i = A_i$ or $A'_i = A_i^c$, $1 \leq i \leq \ell_1$. In general, for any element $Z_1 \in \mathcal{I}(\mathcal{C}_1)$ of the form $h_1(1_{A_1}, ..., 1_{A_{\ell_1}})$ with $\{A_1, ..., A_{\ell_1}\} \subset \mathcal{C}_1$, ℓ_1, we have

$$h_1(1_{A_1}, ..., 1_{A_{\ell_1}}) = \sum_{\varepsilon \in D_{e \ll 1}} h(\varepsilon) 1_{\prod_{1 \leq i \leq \ell_1} A_i^{(\varepsilon_i)}}, \quad (IN02)$$

and Z_1 is simply a finite linear combination of elements of $\mathcal{A}_1$. So the independence between $\mathcal{C}_1$ and $\mathcal{C}_2$ is that of $\mathcal{A}_1$ and $\mathcal{A}_2$ since the factorization is preserved by finite liner combinations.

After these preliminary considerations, we going to prove a first step.

Step 1. We prove that for any $A \in \mathcal{A}_1$, A is independent from $\mathcal{C}_2$.

To see this, define

$$\mathcal{A}_{0,1} = \{A \in \mathcal{A}_1, \forall Z \in \mathcal{I}(\mathcal{C}_2),\ 1_A \text{ independent of } Z\}.$$

By the assumption we have that $\mathcal{C}_1 \subset \mathcal{A}_{0,1}$. Let us quickly prove that $\mathcal{A}_{0,2}$ is a σ-algebra. For sure, $\Omega \in \mathcal{A}_{0,2}$. If $A \in \mathcal{A}_{0,1}$, 1_{A^c} is a measurable function of 1_A, and by this is still in $\mathcal{A}_{0,1}$.

Let $(A_1, A_2) \in \mathcal{A}_{0,1}$, $A_2 \subset A_1$. For any non-negative and measurable functions $h(1_{A_2 \setminus A_1})$ and $\ell(Z)$ of $1_{A_2 \setminus A_1}$ and $Z \in \mathcal{I}(\mathcal{C}_2)$, we have

$$\begin{aligned} h(1_{A_2 \setminus A_1}) &= h(0)(1_{(A_2 \setminus A_1)^c}) + h(1)(1_{A_2} - 1_{A_1}) \\ &= h(0)(1_{A_2^c} + 1_{A_1}) + h(1)(1_{A_2} - 1_{A_1}) \end{aligned}$$

Hence, by multiplying $h(1_{A_2 \setminus A_1})$ by $\ell(Z)$ and by taking the expectations, we will be able to factorize $\mathbb{E}(\ell(Z)1_B)$ for $B \in \{A_1, A_2, A_1^c\}$ in all the terms of the products and, by this, we get

$$\mathbb{E}(h(1_{A_2 \setminus A_1})\ell(Z)) = \mathbb{E}(h(1_{A_2 \setminus A_1}))\mathbb{E}(\ell(Z)).$$

We get that $A_2 \setminus A_1 \in \mathcal{A}_{0,1}$.

Finally, let $(A_k)_{leqk \geq 0} \in \mathcal{A}_{0,1}$ be a sequence of pairwise disjoint elements of $\mathcal{A}_{0,1}$. We define

$$B_n = \bigcup_{1 \leq k \leq n} A_k = \sum_{k \geq 0} A_k \text{ and } B_n = \bigcup_{1 \leq k \leq n} A_k = \sum_{1 \leq k \leq n} A_k,\ n \geq 0.$$

For any non-negative and measurable functions $h(1_{B_n})$ and $\ell(Z)$ of 1_{B_n} and $Z \in \mathcal{I}(\mathcal{C}_2)$, we have

$$h(1_{B_n} = h(0)\left(1 - \sum_{1 \leq k \leq n} 1_{A_k}\right) + h(1) \sum_{1 \leq k \leq n} 1_{A_k}.$$

Here again, by multiplying $h(1_{B_n})$ by $\ell(Z)$ and by taking the expectations, we will be able to factorize any $\mathbb{E}(\ell(Z)1_B)$ for $B \in \{B_k,\ 1 \leq k \leq n\}$ in all the terms of the product and, by this, we get also that :

$$\forall n \geq 0,\ \mathbb{E}(h(1_{B_n})\ell(Z)) = \mathbb{E}(h(1_{B_n}))\mathbb{E}(\ell(Z)).$$

Next by letting $n \uparrow +\infty$, we get by the Monotone Convergence Theorem that

$$\mathbb{E}(h(1_B)\ell(Z)) = \mathbb{E}(h(1_B))\mathbb{E}(\ell(Z)),$$

any non-negative and measurable functions $h(1_{B_n})$ and $\ell(Z)$ of 1_{B_n} and of $Z \in \mathcal{I}(\mathcal{C}_2)$. This proves that $B \in \mathcal{A}_{0,1}$. In summary $\mathcal{A}_{0,1}$ is a Dynkin system containing the π-system. So by the $\lambda - \pi$-Lemma (See Lo (2017b), Doc 04-02, Chapter 5), it contains $\mathcal{A}_1$. We conclude that $\mathcal{A}_1 = \mathcal{A}_{0,1}$ and we get that :

Any element of $\mathcal{A}_1$ is independent of $\mathcal{I}(\mathcal{C}_2)$.

Step 2. For any $Z_1 \in \mathcal{I}(\mathcal{A}_1)$ of the form $h_1(1_{A_1}, ..., 1_{A_{\ell_1}})$ with $\{A_1, ..., A_{\ell_1}\} \subset \mathcal{A}_1$, ℓ_1, we have Z_1 independent of $\mathcal{C}_2$.

This is an easy consequence of Formula (IN01) and the previous result.

Final Step 3. Put

$$\mathcal{A}_{0,2} = \{B \in \mathcal{A}_2, \forall Z \in \mathcal{I}(\mathcal{A}_1),\ 1_B \text{ independent of } Z\}.$$

By the previous steps, $\mathcal{A}_{0,2}$ includes $\mathcal{C}_2$. We use the same techniques as in Step 1 to prove that $\mathcal{A}_{0,2}$ is a Dynkin-system and get that $\mathcal{A}_{0,2} = \mathcal{A}_2$ by the classical methods. Next, we proceed to the same extension as in Step 2 to conclude that any elements of $\mathcal{I}(\mathcal{A}_1)$ is independent of any other element of $\mathcal{I}(\mathcal{A}_2)$. ■

Now, we may go back the proof of the Kolmogorov law.

(C) Proof of the zero-one law. Define the σ-algebras

$$\mathcal{A}_{part,\ n} = \sigma(\{X_k^{-1}(B),\ B \in \mathcal{F},\ 0 \leq k \leq n\}).$$

If A is a tail event, hence for each $n \geq 1$, $A \in \mathcal{A}_{part,\ n}$. Hence, by the principle underlying Formula (P01) at the beginning at the section and by Proposition above, we get that A is independent to any $\mathcal{A}_{part,\ n}$, $n \geq 1$. Since these latter sub-classes are π-system (being σ-algebras), A is also independent of

$$\sigma\Bigg(\mathcal{A}_{part,\ n}\Bigg) = \sigma(\{X_k^{-1}(F),\ F \in \mathcal{F},\ k \geq 0\}) = \mathcal{A}_{part,\ 0}.$$

Since $A \in \mathcal{A}_{part,\ 0}$, we get that A is independent to itself, that is $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$. The equation $\mathbb{P}(A)^2 = P(A)$ has only two solutions 0 or 1 in $[0,1]$. $\square$

(D) Limits Laws for independent random variables.

We are going to derive series of a three interesting asymptotic laws from the Kolmogorov Inequality (Inequality 14 in Chapter 6), the last of them being the celebrated Three-series law of Kolmogorov.

Let X_1, X_2, ... be independent centered and square integrable random variables. We denote $\mathbb{V}ar(X_i) = \sigma_i^2$, $1 \leq i \leq n$. Define

$$C_\infty = \inf\{C > 0, \forall k \geq 0,\ |X_k| \leq C\ \ a.s.\}.$$

Define the partial sums by

$$S_0 = 0,\ S_k = \sum_{i=1}^{k} X_i,\ k \geq 1 \text{ and } s_0 = 0,\ s_k^2 = \sum_{i=1}^{k} \sigma_i^2.$$

We have :

Proposition 23. *The following statements hold.*

(1) If s_n^2 converges $\sigma^2 \in \mathbb{R}$ as $n \rightarrow +\infty$, then $(S_n)_{n\geq 0}$ converges a.s to a a.s. finite (possibly constant) random variable.

(2) If the sequence $(X_n)_{n\geq 0}$ is uniformly bounded, that is C_∞ is finite, s_n^2 converges in $\mathbb{R}$ as $n \rightarrow +\infty$ if and only if $(S_n)_{n\geq 0}$ converges a.s.

More precisely, if s_n^2 diverges as $n \rightarrow +\infty$ and if C_∞ is finite, then $(S_n)_{n\geq 0}$ diverges on any measurable subset of Ω with a positive probability, that is $(S_n)_{n\geq 0}$ non-where converges.

Proof. Since the random variables are centered and independent, we have for any $0 \leq k \leq n$,

$$\mathbb{V}(S_n - S_k) = \sum_{k<j\leq n} \mathbb{V}(X_j) = s_n^2 - s_k^2.$$

Let us apply the right-hand Inequality 14 in Chapter 6, to get for any $\varepsilon > 0$, for any $0 \leq k \geq n$,

$$\mathbb{P}(\max(|S_{k+1}-S_k|,...,|S_n-S_k|)\geq\varepsilon)\leq\frac{|s_n^2-s_k^2|}{\varepsilon^2},$$

in other words, for any $k\geq 0$, $n\geq 0$, for any $\varepsilon>0$

$$\mathbb{P}\left(\bigcup_{1\leq j\leq n}(|S_{k+j}-S_k|\geq\varepsilon)\right)\leq\frac{s_n^2-s_k^2}{\varepsilon^2}.$$

Let us suppose that s_n^2 converges in $\mathbb{R}$ as $n\to+\infty$, that is $(s_n^2)_{n\geq 0}$ is a Cauchy sequence. By applying the Monotone convergence Theorem, we have for any $k\geq 0$ for any $\varepsilon>0$,

$$\mathbb{P}\left(\bigcup_{n\geq k}(|S_n-S_k|\geq\varepsilon)=0\right).$$

and next any $\varepsilon>0$

$$\mathbb{P}(\bigcap_{k\geq 0}\left(\bigcup_{n\geq k}(|S_n-S_k|\geq\varepsilon)=0\right).$$

To conclude, set

$$\Omega_p=\bigcap_{k\geq 0}\left(\bigcup_{n\geq k}(|S_n-S_k|\geq 1/p\right),\ p\geq 1 \text{ and } \Omega_\infty=\bigcup_{p\geq 1}\Omega_p.$$

We still have $\mathbb{P}(\Omega_\infty)=0$ and for any $\omega\in\Omega_\infty$, for any $p\geq 1$, $\exists k_0\geq 0$, for all $n\geq k_0$,

$$|S_n-S_k|(\omega)<1/p.$$

We conclude that $(S_n)_{n\geq 0}$ is Cauchy on Ω_∞ and then converges on Ω_∞, and simply converges **a.s.**

It remains to prove that if C_∞ is finite and if $s_n^2\to+\infty$, $(S_n)_{n\geq 0}$ diverges *a.s.*. By Inequality 14 in Chapter 6, we also have for any $\varepsilon>0$, for any $0\leq k\leq n$,

$$\mathbb{P}\left(\bigcup_{1\leq j\leq n}\left(|S_{k+j}-S_k|\geq\varepsilon\right)\right)\geq 1-\frac{(\varepsilon+C_\infty)^2}{s_n^2-s_k^2}.$$

For k fixed and $n\to+\infty$, we get for any $\varepsilon>0$,

$$\mathbb{P}(\bigcup_{j\geq 0}(|S_{k+j}-S_k|\geq \varepsilon)=1.$$

and next, for any $\varepsilon>0$,

$$\mathbb{P}(\bigcap_{k\geq 0}\bigcup_{j\geq 0}(|S_{k+j}-S_k|\geq \varepsilon)=1.$$

Denote

$$\Omega_0=\biggl(\bigcap_{k\geq 0}\bigcup_{j\geq 0}(|S_{k+j}-S_k|\geq 1\biggr).$$

It is clear that $\mathbb{P}(\Omega_0)=1$ and $(S_n)_{n\geq 0}$ is not Cauchy on Ω_0. This proves the two last statements of the proposition. □

PROPOSITION 24. *Suppose that the assumptions in Proposition 24 hold, except we assume that the X_k's are not necessarily centered. Then, if $(S_n)_{n\geq 0}$ converges a.s as $n\to+\infty$ and the sequence $(X_n)_{n\geq 0}$ is uniformly bounded (that is $C_\infty<+\infty$), then the two sequences $(s_n^2)_{n\geq 0})$ and $\left(\sum_{1\leq k\leq n}\mathbb{E}X_k\right)_{n\geq 0}$ both converge to finite numbers as $n\to+\infty$.*

Proof. It uses the Kolmogorov construction of probability spaces. At this stage, we know this result only in finite distribution (See Chapter 2, Section 5.2, Point (c5)). Here, we anticipate and use Theorem 27 (see page 314) in Chapter 9, and say:

There exists a probability space holding independent random variables X_k, X_k', $k\geq 0$ such that for each k, $X_k=_d X_k'$.

Let us suppose that the assumptions hold and let us define the symmetrized sequence $X_k^{(s)}=X_k-X_k'$, $k\geq 0$. Then the sequence $(X_k^{(s)})_{k\geq 0}$ is centered and uniformly bounded by $2C_\infty$. Now, if $\left(\sum_{1\leq k\leq n}X_k\right)_{n\geq 0}$ converges *a.s.*, so does $\left(\sum_{1\leq k\leq n}X_k'\right)_{n\geq 0}$ by the equality in law. Hence $\left(\sum_{1\leq k\leq n}X_k^{(s)}\right)_{n\geq 0}$ also converges. Next, by applying Point (2) of Proposition 24, the sequence $\left(\sum_{1\leq k\leq n}\mathbb{V}(X_k^{(s)})\right)_{n\geq 0}$ converges. Since $\sum_{1\leq k\leq n}\mathbb{V}(X_k^{(s)})=2s_n^2$, we have the first conclusion.

It remains to prove that $\left(\sum_{1\leq k\leq n} \mathbb{E}X_k\right)_{n\geq 0}$ converges. But we have for all $n \geq 0$,

$$\sum_{1\leq k\leq n} \mathbb{E}(X_k) = \sum_{1\leq k\leq n} X_k - \sum_{1\leq k\leq n} (X_k - \mathbb{E}(X_k))$$

From this and from that assumption that $(S_n)_{n\geq 0}$ converges $a.s$, we may apply Point (1) of Proposition 24 to see that the second series in the right-hand of the formula above converges and get our last conclusion. □

Remark. To fully understand this proof, the reader should seriously know the Kolmogorov construction Theorem and its consequences. For example, because of the independence, the vectors $(X_0, ..., X_k)$ and $(X'_1, ..., X'_k)$ have the same law of $k \geq 0$ and by this, the sequences $(X_k)_{k\geq 0}$ and $(X'_k)_{k\geq 0}$ have the same law as stochastic processes. So the *a.s.* depends only on the probability law of $(X_k)_{k\geq 0}$, the proved results remain valid on any other probability space for a sequence of the same probability law. We advice the reader to come back to this proof after reading Chapter 9.

Before we continue, let us denote for any real-valued random variable X and a real-number $c > 0$, the truncation of X at c by

$$X^{(c)} = X1_{(|X|\leq c)}$$

which is bounded by c.

PROPOSITION 25. *Suppose that the X_n's are are square integrable, centered and independent. If s_n^2 converges $\sigma^2 \in \mathbb{R}$ as $n \to +\infty$, then $(S_n)_{n\geq 0}$ converges a.s to a a.s. finite . The series $(S_n)_{n\geq 0}$ converges a.s. if and only any of the three series below converges :*

$$(i)\forall c \in \mathbb{R}_+\backslash\{0\},\ \sum_{k\geq 0} \mathbb{P}(|X_k| \geq c),\quad (ii)\ \sum_{k\geq 0} \mathbb{V}ar\left(X_k^{(c)}\right) \quad \text{and } (iii) \sum_{k\geq 0} \mathbb{E}(X_k^{(c)})$$

Proof. Suppose the three Conditions (i), (ii) and (iii) hold. From (ii), $\left(\sum_{1\leq k\leq n}\left(X_k^{(c)} - \mathbb{E}(X_k^{(c)})\right)\right)_{n\geq 0}$ converges. This combined with Condition (iii) implies that $\left(\sum_{1\leq k\leq n} X_k^{(c)}\right)_{n\geq 0}$ converges *a.s.*, based on the remark that $X_k^{(c)} = \left(X_k^{(c)} - \mathbb{E}(X_k^{(c)})\right) + \left(\mathbb{E}(X_k^{(c)})\right)$ for all $k \geq 0$.

Next for all $c > 0$, the event $(X_k \neq X_k^{(c)})$ occurs only if $(|X_k| \geq c)$ and hence

$$\mathbb{P}(X_n \neq X_n^{(c)},\ i.o) = \lim_{n\uparrow+\infty} \mathbb{P}\left(\bigcup_{k\geq n}(X_k \neq X_k^{(c)})\right) \leq \lim_{n\uparrow+\infty} \sum_{k\geq n} \mathbb{P}(|X_k| \geq c).$$

Hence, by Condition (i), $\mathbb{P}(X_n \neq X_n^{(c)},\ i.o) = 0$ and next, the series $\sum_{k\geq 0} X_k$ and $\sum_{k\geq 0} X_k^{(c)}$ converge or diverge *a.s.* simultaneously. We get that $\sum_{k\geq 0} X_k$ converges *a.s.*.

Conversely, if $(S_n)_{n\geq 0}$ converges *a.s.*, it follows that $(X_n)_{n\geq 0}$ converges *a.s.* to zero, by Corollary 1 in Section 1 below, Condition (i) holds. The latter, by the *a.s.* equivalence between $\sum_{k\geq 0} X_k$ and $\sum_{k\geq 0} X_k^{(c)}$, ensures that $\sum_{k\geq 0} X_k^{(c)}$ converges, which by Proposition 24, yields Conditions (ii) and (iii).

It remains to prove that none of the three conditions cannot fail, whenever the series $\sum_{k\geq 0} X_k$ converges *a.s.*. First, by Corollary 1 in Section 1, Condition (i) cannot fail and hence the *a.s.* convergence of $\sum_{k\geq 0} X_k^{(c)}$ also cannot fail and this bears Conditions (ii) and (ii). □

Now let us close this introduction to these following important Kolmogorov's Theorems.

(E) Strong Law of Large numbers of Limits Laws of Kolmororov.

Before we state the Kolmogorov laws, we state the following :

Kronecker Lemma. If $(b_n)_{n\geq 0}$ is an increasing sequence of positive numbers and $(x_n)_{n\geq 0}$ is a sequence of finite real numbers such that $\left(\sum_{1\leq k\leq n} x_k\right)_{n\geq 0}$ converges to a finite real number s, then

$$\frac{\sum_{1\leq k\leq n} b_k x_k}{b_n} \to 0 \ as\ n \to \infty. \ \Diamond$$

This Lemma is proved in the Appendix, where it is derived from the Toeplitz Lemma.

Let us begin by

(E-a) The Strong Law of Large Numbers for Square integrable and independent random variables.

THEOREM 16. *Let $(X_n)_{n\geq 0}$ be a sequence of square integrable and independent random variables and let $(b_n)_{n\geq 1}$ be an increasing sequence of finite real numbers. If*

$$\sum_{n\geq 0} \frac{\mathbb{V}(X_n)}{b_n^2} < +\infty, \quad (CK01)$$

then we have the following SLLN

$$\frac{S_n - \mathbb{E}(S_n)}{b_n} \to 0 \text{ as } n \to +\infty. \quad (SK01)$$

Proof. The proof comes as the conclusion of the previous developments. Suppose that assumptions of the theorem hold and Condition (CK01) is true. By Proposition 23, we have

$$\sum_{n\geq 0} \frac{X_n - \mathbb{E}(X_n)}{b_n} < +\infty, \ a.s. \quad (SK02)$$

Applying the Kronecker Lemma with $x_k = X_k - \mathbb{E}(X_k)$ and the same sequence $(b_n)_{n\geq 0}$ leads to (SK01). □

Example. If the X_n's have the same variance σ^2, we may take $b_n = n$, $n \geq 1$ and see that Condition (SK01) is verified since

$$\sum_{n\geq 1} \frac{\mathbb{V}(X_n)}{b_n^2} = \sigma^2 \sum_{n\geq 1} \frac{1}{n^2} < +\infty,$$

and next,

$$\frac{S_n - \mathbb{E}(S_n)}{n} \to 0 \text{ as } n \to +\infty. \quad (SK03)$$

We find again the simple SLLN as in Corollary 2 above.

Now, what happens if the first moments of the X_n exist but we do not have information about the second moments of the X_n's. We already saw in Kintchine's Theorem 14 that we have a WLLN if the X_n has the same Law. Here again, the Kolmogorov theory goes far and establishes the SLLN even if the common second moment is infinite. We are going to see this in the next part.

(E-b) The Strong Law of Large Numbers for independent and identically random variables with finite mean.

We will need the following simplified Toeplitz lemma which is proved in the Appendix in its integrability.

Simple Toeplitz Lemma. Suppose that $k(n) = n$ for all $n \geq 1$. Let $(c_k)_{k\geq 0}$ be sequence such that the sequence $(b_n)_{n\geq 0} = (\sum_{1\leq k\leq n} |c_k|_{n\geq 0}$ is non-decreasing and $b_n \rightarrow \infty$. If $x_n \rightarrow x \in \mathbb{R}$ as $n \rightarrow +\infty$, then

$$\frac{1}{b_n} \sum_{1\leq k\leq n} c_k x_k \rightarrow x \quad as \quad n \rightarrow +\infty. \ \Diamond$$

THEOREM 17. *Let* $(X_n)_{n\geq 0}$ *be a sequence of independent and identically distributed random variables having the same law as* X*. Then*

(a) $\mathbb{E}|X| < +\infty$

if and only if

(b) Sn/n *converges* ***a.s.*** *to a finite number* c*, which is necessarily* $\mathbb{E}(X)$.

Proof. Set $A_n = (|X| \geq n)$, $n \geq 0$ with $A_0 = \Omega$ clearly.

Now suppose that Point (b) holds. We have to prove that of $\mathbb{E}|X|$ is finite. If X is bounded, there is nothing to prove. If not, the upper endpoint of $uep(X)$ is infinite and Formula (DF3) in Proposition 1 (See Chapter 1, page 14)

$$-1 + \sum_{n\in[0,\ [uep(X)]^+]} \mathbb{P}(|X| \geq n) \leq \mathbb{E}|X| \leq \sum_{n\in[0,\ [uep(X)]^+]} \mathbb{P}(|X| \geq n)$$

becomes

$$-1 + \sum_{n \geq 0} \mathbb{P}(|X| \geq n) \leq \mathbb{E}|X| \leq \sum_{n \neq 0} \mathbb{P}(|X| \geq n). \quad (DF4)$$

Then we have

$$\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} = \frac{S_n}{n} - \frac{n-1}{n}\frac{S_{n-1}}{n-1} \to c - c = 0 \ a.s.$$

By the Borel-Cantelli Corollary 1, the serie $\sum_{n\geq 0} \mathbb{P}(|X_n/c| \geq 1)$ is convergent, that $\sum_{n\geq 0} \mathbb{P}(A_n)$ is finite and by Formula (DF4), $\mathbb{E}|X|$ is finite.

Now suppose that $\mathbb{E}|X|$ is finite. If X is bounded, we are in the case of the last example above with $\mathbb{V}ar(X) = \sigma^2$ is stationary and we have that S_n/n converges to $\mathbb{E}(X)$ $a.s.$. If not, we use the truncated random variables $X_k^{(t)} = X_k 1_{(|X_k|<k)}$, $k \geq 1$ and let $S_0^{(t)} = 0$, $S_1^{(t)} = X_1^{(t)}$, $S_n^{(t)} = X_1^{(t)} + ... + X_n^{(t)}$, $n \geq 2$.

We already explained in page 225 that $S_n^{(t)}/n$ and S_n/n have the same $a.s.$ limit or diverge $a.s.$ together whenever $\sum_{n\geq 1} \mathbb{P}(|X_n| \geq n)$. But since $\mathbb{E}|X|$ is finite, the series $\sum_{n\geq 1} \mathbb{P}(|X_n| \geq n)$ converges by Formula (DF4). Hence we only have to prove that $S_n^{(t)}/n \to \mathbb{E}(X)$ $a.s.$. Now, by the Dominated Convergence Theorem, we have

$$\mathbb{E}(X_n^{(t)} = \int X 1_{(|X|<n)} \, d\mathbb{P} \to \mathbb{E}X \ as \ n \to +\infty.$$

By applying the simple Toeplitz Lemma with $c_k = 1$ and $x_k = \mathbb{E}(X_n^{(t)}$, we get

$$\frac{\mathbb{E}(S_n^{(t)}}{n}) \to \mathbb{E}X.$$

So, our task is to prove that

$$\frac{S_n^{(t)} - \mathbb{E}(S_n^{(t)}}{n}) \to 0 \ a.s.$$

But this derives form Theorem 16 whenever we have

$$\sum_{n\geq 1} \frac{\mathbb{V}ar(X_n^{(t)})}{n^2} < +\infty.$$

But we have

$$\begin{aligned}
\sum_{n\geq 1} \frac{\mathbb{V}ar(X_n^{(t)})}{n^2} &= \sum_{n\geq 1} \frac{\mathbb{E}(X_n^{(t)})^2 - (\mathbb{E}(X_n^{(t)}))^2}{n^2} \\
&\leq \sum_{n\geq 1} \frac{\mathbb{E}(X_n^{(t)})^2}{n^2} \\
&= \sum_{n\geq 1} \mathbb{E}\left(\frac{X^2}{n^2} 1_{(|X|<n)}\right).
\end{aligned}$$

Next, define $B_m = (m-1 \leq |X| < m)$, $m \geq 1$. For $m \geq 1$ fixed, we have

$$\frac{X^2}{n^2} 1_{(|X|<n)\cap B_m} = \frac{X^2}{n^2} 1_{\emptyset} = 0 \ for \ n < m,$$

and for $n \geq m$

$$\frac{X^2}{n^2} 1_{(|X|<n)\cap B_m} = \frac{X^2}{n^2} 1_{B_m} \leq \frac{m^2}{n^2} 1_{B_m},$$

so that

$$\begin{aligned}
\sum_{n\geq 1} \frac{X^2}{n^2} 1_{(|X|<n)\cap B_m} &= \sum_{n\geq m} \frac{X^2}{n^2} 1_{(|X|<n)\cap B_m} \\
&\leq \left(m^2 \sum_{n\geq m} \frac{1}{n^2}\right) 1_{B_m}.
\end{aligned}$$

By comparing the series of the form $\sum_{n\geq m} f(n)$ with the integral $\int_{x\geq m} f(x)dx$ for a non-decreasing and continuous function $f(x) = x^{-2}$, we have

$$\sum_{n\geq m} \frac{1}{n^2} \leq \int_m^{+\infty} x^{-2} \ dx = 1/m.$$

Hence, we have

$$\sum_{n\geq 1} \frac{X^2}{n^2} 1_{(|X|<n)\cap B_m} \leq m 1_{B_m} = (1+(m-1))1_{B_m} \leq (1+|X|)1_{B_m}.$$

Since, we obviously have $\sum_{m\geq 1} B_m = \Omega$, we may sum over m in the previous formula to have

$$\sum_{n\geq 1} \frac{X^2}{n^2} 1_{(|X|<n)} \leq (1+|X|).$$

We arrive at

$$\sum_{n\geq 1} \frac{\mathbb{V}ar(X_n^{(t)})}{n^2} \leq \mathbb{E} \sum_{n\geq 1} \frac{X^2}{n^2} 1_{(|X|<n)} \leq (1+\mathbb{E}|X|) < +\infty.$$

We reached the desired condition which allows to conclude the proof. ■

This nice theory of Kolmogorov opens the wide field of SLLN's. The first step for the generalization will be the Hájèk-Rényi approach we will see soon in special monograph reserved to limits laws for sequences of random variables of arbitrary probability laws.

3. Convergence of Partial sums of independent Gaussian random variables

Let us give the following interesting equivalences between different types of convergences for partial sums of independent Gaussian real-valued random variables.

we have

THEOREM 18. *Let $(X_n)_{n\geq 1}$ be sequence of independent and centered Gaussian real-valued random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let us define their partial sums $S_n = \sum_{1\leq k\leq n} X_n$, $n \geq 1$ and $s_n^2 = \sum_{1\leq k\leq n} \mathcal{V}ar(X_n)$, $n \geq 1$. The the following convergences are equivalent, as $n \rightarrow +\infty$,*

(1) $(S_n)_{n\geq 1}$ converges a.s. to an a.s. finite random variable Z.

(2) $(S_n)_{n\geq 1}$ converges in probability to an a.s. finite random variable Z.

(3) $(S_n)_{n\geq 1}$ weakly converges to an a.s. finite random variable Z.

(4) $(s_n)_{n\geq 1}$ converges in $\mathbb{R}$.

(5) $(S_n)_{n\geq 1}$ converges in L^2.

Proof. The proof is based on the compa rison between type of convergences in Chapter 5. The implication (1) → (2). Next (2) → (3) by Point (a) of Proposition 18 in Section 5. Further (3) that for each $t \in \mathbb{R}$,

$$\mathbb{E}\exp(itS_n) \rightarrow \mathbb{E}\exp(X_{(p,\ +\infty)}).$$

Since the X_n are independent, we have $S_n \sim \mathcal{N}(0, s_n^2)$. Hence for all $t \in \mathbb{R}$,

$$\exp(-ts_n^2/2) \rightarrow \mathbb{E}\exp(X_{(p,\ +\infty)}),$$

This is possible only if s_n^2 converge in $\mathbb{R}$, where we took into account the fact that Z is *a.s.* finite. Now, by Proposition 23 (in Section 2 in Section 5, page 216) (4) implies (1). By this circular argument, the assertions (1) to (4) are equivalent.

Let us handle Assertion (5). Suppose (5) holds with $s_n^2 \to s^2$. Let us denote $S = \sum_{n \geq 1} X_n$. We have for all $n \geq 1$

$$\mathbb{E}(S - S_n)^{=} \|S - S_n\|_2^2 = (s^2 - s_n^2) \to 0.$$

So (5) implies (5). Finally, suppose that (2) holds. We have that by Point (3) that S_n converge to S in probability and s_n^2 converge to s^2, and next

$$\|S_n\|_2^2 = s_n^2 \to s^2 = \|S\|_2^2.$$

Thus by Point(c4) of Theorem 10 (in Section 2 in Section 5, page 167), (2) implies (5). The proof is complete now.

4. The Lindenberg-Lyapounov-Levy-Feller Central Limit Theorem

We do not treat the Central limit Theorem on $\mathbb{R}^d$, $d \geq 2$, which is addressed in Lo *et al.* (2016) in its simplest form.

We already described the *CLT* question on $\mathbb{R}$ with *iid* sequences. The current section will give the most finest results for independent random variables. Researchers are trying to export the Linderberg-Levy-Feller Central Limit Theorem to abstract spaces under dependance conditions. In that generalization process, mastering the techniques which are used in the independence case significantly help.

Let us begin by the key result of Lyapounov.

(A) Lyapounov Theorem.

THEOREM 19. *Let X_1, X_2, ... a sequence of real and independent random variables centered at expectations, with finite $(n+\delta)-$moment, $\delta > 0$. Put for each $n \geq 1$, $S_n = X_1 + ... + X_n$ and $s_n^2 = \mathbb{E}X_1^2 + \mathbb{E}X_2^2 + ... + \mathbb{E}X_n^2$. We denote $\sigma_k^2 = \mathbb{E}X_k^2, k \geq 1$ and F_k denotes the probability distribution function of X_k. Suppose that*

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left|X_k\right|^{2+\delta} \to 0 \text{ as } n \to \infty. \tag{4.1}$$

Then, we have as $n \to +\infty$,

$$S_n/s_n \rightsquigarrow N(0,1).$$

Proof of Theorem 19. According to Lemma 3 below, if (4.1) holds for $\delta > 1$, then it holds for $\delta = 1$. So it is enough to prove the theorem for $0 < \delta \leq 1$. By lemma 4 below, the assumption (4.1) implies $s_n \to +\infty$ and

$$\max_{1 \leq k \leq n} \left(\frac{\sigma_k}{s_n}\right)^{2+\delta} \leq \max_{1 \leq k \leq n} \frac{\mathbb{E}\left|X_k\right|^{2+\delta}}{s_n^{2+\delta}} \leq \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left|X_k\right|^{2+\delta} =: A_n(\delta) \to 0.$$

Let us use the expansion of the characteristic functions

$$f_k(u) = \int e^{iux} dF_k(x)$$

at the order two to get for each $k, 1 \leq k \leq n$ as given in Lemma 1 below

$$f_k(u/s_n) = 1 - \frac{u^2}{2} \cdot \frac{\sigma_k^2}{s_n^2} + \theta_{nk} \frac{|u|^{2+\delta} \, \mathbb{E} \, |X_k|^{2+\delta}}{s_n^{2+\delta}}. \tag{4.2}$$

Now the characteristic function of S_n/s_n is, for $u \in R$,

$$f_{S_n/s_n}(u) = \prod_{k=1}^{n} f_k(u/s_n)$$

S that is

$$\log f_{S_n/s_n}(u) = \sum_{k=1}^{n} \log f_k(u/s_n).$$

Now, we use the uniform expansion of $\log(1+u)$ at the neighborhood at 1, that is

$$\sup_{|u| \leq z} \left| \frac{\log(1+u)}{u} \right| = \varepsilon(z) \to 0. \tag{4.3}$$

For each k in (4.2), we have

$$f_k(u/s_n) = 1 - u_{kn} \tag{4.4}$$

with the uniform bound

$$\begin{aligned} |u_{kn}| &\leq \sum_{j=1}^{n} \frac{|u|^2}{2} \cdot \frac{\sigma_k^2}{s_n^2} + \frac{|u|^{2+\delta} \, \mathbb{E} \, |X_k|^{2+\delta}}{s_n^{2+\delta}} \\ &= \frac{|u|^2}{2} \cdot \max_{1 \leq k \leq n} \left(\frac{\sigma_k^2}{s_n} \right) + \frac{|u|^{2+\delta} \sum_{j=1}^{n} \mathbb{E} \, |X_k|^{2+\delta}}{s_n^{2+\delta}} = u_n. \end{aligned}$$

By applying (4.3) to (4.4), we get

$$\log f_k(u/s_n) = -u_{kn} + \theta_n u_{kn} \varepsilon(u_n)$$

and next

$$\begin{aligned}\log f_{S_n/s_n}(u) &= \sum_{k=1}^{n} \log f_k(u/s_n) \\ &= -\frac{u^2}{2} + |u|^{2+\delta}\, \theta_n A_n(\delta) + (\frac{u^2}{2} + |u|^{2+\delta}\, \theta_n A_n(\delta))\varepsilon(u_n) \\ &\rightarrow -u^2/2.\end{aligned}$$

We get for u fixed,

$$f_{S_n/s_n}(u) \rightarrow \exp(-u^2/2).$$

This completes the proof. □

An expression of Lyapounov Theorem using triangular arrays.

Since the proof is based on the distribution of $\{X_k, 1 \leq k \leq n\}$ for each $n \geq 1$, it may be extended to triangular array to the following corollary.

COROLLARY 3. *Consider the triangular array* $\{X_{nk}, 1 \leq k \leq n, n \geq 1\}$*. Put for each* $n \geq 1$, $S_{nn} = X_{n1} + ... + X_{nn}$ *and* $s_{nn}^2 = \mathbb{E}X_{n1}^2 + \mathbb{E}X_{n2}^2 + ... + \mathbb{E}X_{nn}^2$*. Suppose that for each* $n \geq 1$*, the random variables* $X_{nk}, 1 \leq k \leq n,$ *are centered and independent such that*

$$\frac{1}{s_{nn}^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left|X_{nk}\right|^{2+\delta} \rightarrow 0 \text{ as } n \rightarrow \infty. \tag{4.5}$$

Then

$$S_{nn}/s_{nn} \rightsquigarrow N(0,1).$$

Now, we are able to prove the Lyapounov-Feller-Levy Theorem (see Lecam for an important historical note with the contribution of each author in this final result).

(B) The General Central Limit Theorem on $\mathbb{R}$.

THEOREM 20. *Let X_1, X_2, ... a sequence of real and independent random variables centered at expectations, with finite $(n+\delta)-$moment, $\delta > 0$. Put for each $n \geq 1$, $S_n = X_1 + ... + X_n$ and $s_n^2 = \mathbb{E}X_1^2 + \mathbb{E}X_2^2 + ... + \mathbb{E}X_n^2$. We denote $\sigma_k^2 = \mathbb{E}X_k^2, k \geq 1$ and F_k denotes the probability distribution function of X_k. We have the equivalence between*

$$\max_{1\leq k\leq n}\left(\frac{\sigma_k^2}{s_n}\right) \to 0 \quad \text{and} \quad S_n/s_n \rightsquigarrow N(0,1) \tag{4.6}$$

and

$$\forall \varepsilon > 0, \;\; \frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x|\geq\varepsilon s_n} x^2 dF_k(x) \to 0 \text{ as } n \to \infty. \tag{4.7}$$

Proof of Theorem 20.

The proof follows the lines of the proof in Loève (1997). But they are extended by more details and adapted and changed in some parts. Much details were omitted. We get them back for making the proof understandable for students who just finished the measure and probability course.

Before we begin, let us establish an important property of

$$g_n(\varepsilon) = \frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x|\geq\varepsilon s_n} x^2 dF_k(x),$$

when (4.7) holds. Suppose that this latter holds. We want to show that there exists a sequence $\varepsilon_n \to 0$ such that $\varepsilon_n^{-2} g_n(\varepsilon_n) \to 0$ (this implying also that $\varepsilon_n^{-1} g_n(\varepsilon_n) = o(\varepsilon_n) \to 0$ and that $g_n(\varepsilon_n) = o(\varepsilon_n^2) \to 0$). To this end, let $k \geq 1$ fixed. Since $g_n(1/k) \to 0$ as $n \to \infty$, we have $0 \leq g_n(1/k) \leq k^{-3}$ for n large enough.

We will get what we want from an induction on this property. Fix $k = 1$ and denote n_1 an integer such that $0 \leq g_n(1) \leq 1^{-3}$ for $n \geq n_1$. Now we apply the same property on the sequence $\{g_n(\circ), n_1 + 1\}$ with $k = 2$. We find a $n_2 > n_1$ such that $0 \leq g_n(1/2) \leq 2^{-3}$ for $n \geq n_2$. Next we apply the same property on the sequence $\{g_n(\circ), n_2 + 1\}$ with

$k=2$. We find a $n_3 > n_2$ such that $0 \leq g_n(1/3) \leq 3^{-3}$ for $n \geq n_3$. Finally, an infinite sequence of integers $n_1 < n_2 < .. < n_k < n_{k+1} < ..$ such that for each $k \geq 1$, one has $0 \leq g_n(1/k) \leq k^{-3}$ for $n \geq n_k$. Put

$$\varepsilon_n = 1/k \text{ on } n_k \leq n < n_{k+1}.$$

We surely have $\varepsilon_n \to 0$ and $\varepsilon_n^{-2} g_n(\varepsilon_n)$. This is clear from

$$\begin{cases} \varepsilon_n = 1/k & \text{on} \quad n_k \leq n < n_{k+1} \\ \varepsilon_n^{-2} g_n(\varepsilon_n) = k^2(1/k^3) \leq (1/k) & \text{on} \quad n_k \leq n < n_{k+1}. \end{cases}$$

Now we arge going to use

$$\varepsilon_n \to 0 \text{ and } \varepsilon_n^{-2} g_n(\varepsilon_n) \to 0. \tag{4.8}$$

Proof of (4.7) $\Longrightarrow$(4.6). Suppose (4.7) holds. Thus there exists a sequence $(\varepsilon_n)_{n\geq 0}$ of positive numbers such that (4.8) prevails. First, we see that, for each $j, 1 \leq j \leq n$,

$$\begin{aligned} \frac{\sigma_j^2}{s_n^2} &= \frac{1}{s_n^2} \int x^2 dF_j = \frac{1}{s_n^2} \left\{ \int_{|x| \leq \varepsilon_n s_n} x^2 dF_j + \int_{|x| \leq \varepsilon_n s_n} x^2 dF_j \right\} \\ &\leq \frac{1}{s_n^2} \int_{|x| \leq \varepsilon_n s_n} x^2 dF_k + \varepsilon_n^2 \\ &\leq \frac{1}{s_n^2} \sum_{k=1}^{k} \int_{|x| \leq \varepsilon_n s_n} x^2 dF_k = g(\varepsilon_n) + \varepsilon_n^2. \end{aligned}$$

It follows that

$$\max_{1 \leq j \leq n} \frac{\sigma_j^2}{s_n^2} \leq g(\varepsilon_n) + \varepsilon_n^2 \to 0.$$

Its remains to prove that $S_n/s_n \to \mathbb{N}(0,1)$. To this end we are going to use this array of truncated random variables $\{X_{nk}, 1 \leq k \leq n, n \geq 1\}$ defined as follows. For each fixed $n \geq 1$, we set

$$X_{nk} = \begin{cases} X_k & \text{if} \quad |X_k| \leq \varepsilon_n s_n \\ 0 & \text{if} \quad |X_k| > \varepsilon_n s_n \end{cases}, 1 \leq k \leq n.$$

Now, we consider summands S_{nn} as defined in Corollary 3. Weremark that for any $\eta > 0$,

$$P\left(\left| \frac{S_{nn}}{s_n} - \frac{S_n}{s_n} \right| > \eta \right) \leq P\left(\frac{S_{nn}}{s_n} \neq \frac{S_n}{s_n} \right) = P\left(\frac{S_{nn}}{s_n} \neq \frac{S_n}{s_n} \right)$$

and also,

$$\begin{aligned}\left(\frac{S_{nn}}{s_n} \neq \frac{S_n}{s_n}\right) &= ((\exists 1 \leq k \leq n), X_{nk} \neq X_k) \\ &= (\exists (1 \leq k \leq n), |X_k| > \varepsilon_n s_n) = \bigcup_{k=1} (|X_k| > \varepsilon_n s_n).\end{aligned}$$

We get

$$\begin{aligned}P\left(\left|\frac{S_{nn}}{s_n} - \frac{S_n}{s_n}\right| > \eta\right) &\leq \sum_{k=1}^{n} P(|X_k| > \varepsilon_n s_n) \\ &\leq \sum_{k=1}^{n} \int_{|x| \leq \varepsilon_n s_n} dF_k = \sum_{k=1}^{n} \int_{|x| \leq \varepsilon_n s_n} \left\{\frac{1}{x^2}\right\} x^2 dF_k \\ &\leq \left\{\frac{1}{(\varepsilon_n s_n)^2}\right\} \sum_{k=1}^{n} \int_{|x| \leq \varepsilon_n s_n} x^2 dF_k \\ &\leq \frac{1}{\varepsilon_n^2} g_n(\varepsilon_n) \to 0.\end{aligned}$$

Thus S_{nn}/s_n and S_n/s_n are equivalent in probability. This implies that they have the same limit law or do not have a limit law together. So to prove that S_n/s_n has a limit law, we may prove that S_{nn}/s_n has a limit law. Next by Slutsky lemma, it will suffice to establish the limiting law of S_{nn}/s_{nn} whenever $s_{nn}/s_n \to 1$. We focus on this. We begin to remark that, since $\mathbb{E}(X_k) = 0$, we have the decomposition

$$0 = \mathbb{E}(X_k) = \int x dF_k = \int_{|x| \leq \varepsilon_n s_n} x dF_k + \int_{|x| > \varepsilon_n s_n} x dF_k$$

to get that

$$\left|\int_{|x| \leq \varepsilon_n s_n} x dF_k\right| = \left|\int_{|x| > \varepsilon_n s_n} x dF_k\right|.$$

We remark also that

$$\begin{aligned}\mathbb{E}(X_{nk}) &= \int_{|X_k| \leq \varepsilon_n s_n} X_{nk} d\mathbb{P} + \int_{|X_k| > \varepsilon_n s_n} X_{nk} d\mathbb{P} \\ &= \int_{|X_k| \leq \varepsilon_n s_n} X_k d\mathbb{P} + \int_{|X_k| > \varepsilon_n s_n} 0 \; d\mathbb{P}.\end{aligned} \tag{4.9}$$

Combining all what precedes leads to

$$\begin{aligned}
|E(X_{nk})| &= \left|\int_{|X_k|\leq \varepsilon_n s_n} X_k dP\right| \\
&= \left|\int_{|x|>\varepsilon_n s_n} x dF_k\right| = \left|\int_{|x|>\varepsilon_n s_n} \left\{\frac{1}{x}\right\} x^2 dF_k\right| \\
&\leq \left|\int_{|x|>\varepsilon_n s_n} \frac{1}{|x|} x^2 dF_k\right| \leq \frac{1}{\varepsilon_n s_n}\left|\int_{|x|>\varepsilon_n s_n} x^2 dF_k\right|.
\end{aligned}$$

Therefore,

$$\frac{1}{s_n}\sum_{k=1}^{n}|E(X_{nk})| \leq \varepsilon_n^{-1} g(\varepsilon_n) \to 0. \tag{4.10}$$

Based on this, let us evaluate s_{nn}/s_n. Notice that for each fixed $n \geq 1$, the X_{nk} are still independent. The technique used in 4.9 may be summarized as follows : any any measurable function $g(\circ)$ such that $g(0) = 0$,

$$Eg(X_{nk}) = \int_{|X_k|\leq \varepsilon_n s_n} g(X_{nk}) dP + \int_{|X_k|>\varepsilon_n s_n} g(0) dP = \int_{|X_k|\leq \varepsilon_n s_n} g(X_{nk}) dP.$$

By putting these remarks together, we obtain

$$\begin{aligned}
1-\frac{s_{nn}^2}{s_n^2} &= \frac{s_n^2-s_{nn}^2}{s_{n^2}} \\
&= \frac{1}{s_{n^2}}\left\{\sum_{k=1}^n EX_k^2-\sum_{k=1}^n E(X_{nk}-E(X_{nk}))^2\right\} \\
&= \frac{1}{s_{n^2}}\left\{\sum_{k=1}^n EX_k^2-\left(\sum_{k=1}^n E(X_{nk}^2)-E(X_{nk})^2\right)\right\} \\
&= \frac{1}{s_{n^2}}\left\{\sum_{k=1}^n EX_k^2-\sum_{k=1}^n EX_{nk}^2+\sum_{k=1}^n (EX_{nk})^2\right\}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{s_{n^2}}\left\{\sum_{k=1}^n \int X_k^2 dP-\sum_{k=1}^n \int_{|X_k|\leq\varepsilon_n s_n} X_k^2 dP+\sum_{k=1}^n (EX_{nk})^2\right\} \\
&= \frac{1}{s_{n^2}}\left\{\sum_{k=1}^n \int_{|X_k|>\varepsilon_n s_n} X_k^2 dP+\sum_{k=1}^n (EX_{nk})^2\right\} \\
&\leq \frac{1}{s_{n^2}}\left\{\sum_{k=1}^n \int_{|X_k|>\varepsilon_n s_n} X_k^2 dP+\sum_{k=1}^n (E\left|X_{nk}\right|)^2\right\}
\end{aligned}$$

Finally, we use the simple inequality of real numbers $(\sum |a_i|)^2 = \sum |a_i|^2 + \sum_{i\neq j}^2 |a_i|\,|a_j| \geq \sum |a_i|^2$ and conclude from the last inequality that

$$\begin{aligned}
\left|1-\frac{s_{nn}^2}{s_n^2}\right| &\leq \frac{1}{s_{n^2}}\left\{\sum_{k=1}^n \int_{|X_k|>\varepsilon_n s_n} X_k^2 dP+\sum_{k=1}^n (E\left|X_{nk}\right|)^2\right\} \\
&\leq \frac{1}{s_{n^2}}\left\{\sum_{k=1}^n \int_{|X_k|>\varepsilon_n s_n} X_k^2 dP+\left(\sum_{k=1}^n E\left|X_{nk}\right|\right)^2\right\} \\
&= g(\varepsilon_n)+\left(\frac{1}{s_n}\sum_{k=1}^n E\left|X_{nk}\right|\right).
\end{aligned}$$

By (4.10) above, we arrive at

$$\left|1-\frac{s_{nn}^2}{s_n^2}\right| \leq g(\varepsilon_n)+\varepsilon_n^{-1}g(\varepsilon_n)\to 0.$$

It comes that $s_{nn}/s_n \to 1$. Finally, the proof of this part will be derived from S_{nn}/s_{nn}. We center the X_{nk} at their expectations. To prove that the sequence of the new summands T_{nn}/s_{nn} converges to $\mathcal{N}(0,1)$, we use Corollary 1 by checking the Lyapounov's condition

$$\frac{1}{s_{nn}^3}\sum_{k=1}^{n}\mathbb{E}\left|X_{nk}-EX_{nk}\right|^3 \to 0 \text{ as } n\to\infty.$$

By

$$\begin{aligned}\frac{1}{s_{nn}^3}\sum_{k=1}^{n}\mathbb{E}\left|X_{nk}-EX_{nk}\right|^3 &= \frac{1}{s_{nn}^3}\sum_{k=1}^{n}\mathbb{E}\left|X_{nk}-EX_{nk}\right|\times\mathbb{E}\left|X_{nk}-EX_{nk}\right|^2\\ &\leq \frac{2}{s_{nn}^3}\sum_{k=1}^{n}\mathbb{E}\left|X_{nk}\right|\times\mathbb{E}\left|X_{nk}-EX_{nk}\right|^2.\end{aligned}$$

Take $g(\cdot)=|\cdot|$ in (4) to see again that

$$\mathbb{E}\left|X_{nk}\right| = \int_{|X_k|\leq\varepsilon_n s_n}\left|X_{nk}\right| dP \leq \varepsilon_n s_n.$$

The last two formula yied

$$\begin{aligned}\frac{1}{s_{nn}^3}\sum_{k=1}^{n}\mathbb{E}\left|X_{nk}-EX_{nk}\right|^3 &\leq \frac{2\varepsilon_n s_n}{s_{nn}^3}\sum_{k=1}^{n}\mathbb{E}\left|X_{nk}-EX_{nk}\right|^2\\ &= \frac{2\varepsilon_n s_n s_{nn}^2}{s_{nn}} = \varepsilon_n\frac{2s_n}{s_{nn}} \to 0.\end{aligned}$$

It comes by Corrolary 3 that

$$\frac{T_{nn}}{s_{nn}} = \frac{S_{nn}-\sum_{k=1}^{n}E(X_{nk})}{s_{nn}} \to N(0,1).$$

Since $s_{nn}/s_n \to 1$ and by 4.10

$$\left|\frac{\sum_{k=1}^{n}E(X_{nk})}{s_{nn}}\right| \leq \frac{s_n}{s_{nn}}\left\{\frac{1}{s_n}\sum_{k=1}^{n}E\left|X_{nk}\right|)\right\} \to 0.$$

We conclude that S_{nn}/s_{nn} converges to $\mathcal{N}(0,1)$.

Proof of : (4.6)$\Longrightarrow$ (4.7). The convergence to $\mathcal{N}(0,1)$ implies that for any fixed $t \in \mathcal{R}$, we have

$$\prod_{k=1}^{n} f_k(u/s_n) \rightarrow \exp(-u^2/2). \tag{4.11}$$

We are going to use uniform expansions of $\log(1+z)$. We have

$$\lim_{z\rightarrow 0} \left|\frac{\log(1+z)-z}{z^2}\right| = \frac{1}{2}$$

this implies

$$\sup_{z\leq u} \left|\frac{\log(1+z)-z}{z^2}\right| = \varepsilon(u) \rightarrow 1/2 \text{ as } u \rightarrow 0. \tag{4.12}$$

Now, use the expansion

$$f_k(u/s_n) = 1 + \theta_k \frac{u^2\sigma_k^2}{2s_n^2}.$$

This implies that

$$\max_{1\leq k\leq n} \leq |f_k(u/s_n)-1| \leq \frac{u^2}{2} \max_{1\leq k\leq n} \frac{\sigma_k^2}{s_n^2} = u_n \rightarrow 0. \tag{4.13}$$

and next

$$|f_k(u/s_n)-1| = \theta_k^2 \frac{u^2\sigma_k^2}{2s_n^2} \times \frac{u^2\sigma_k^2}{2s_n^2} \leq \left[\frac{u^4}{4} \max_{1\leq k\leq n} \frac{\sigma_k^2}{s_n^2}\right] \times \frac{\sigma_k^2}{s_n^2}.$$

This latter implies

$$\sum_{k=1}^{n} |f_k(u/s_n)-1| \leq \left[\frac{u^4}{4} \max_{1\leq k\leq n} \frac{\sigma_k^2}{s_n^2}\right] = B_n(u) \rightarrow 0.$$

By (4.13), we see that $\log f_k(u/s_n)$ is uniformly defined in $1 \leq k \leq n$ for n large enough and (4.11) becomes

$$\sum_{k=1}^{n} \log f_k(u/s_n) \rightarrow -u^2/2,$$

that is

$$\frac{u^2}{2} + \sum_{k=1}^{n} \log f_k(u/s_n) \rightarrow 0.$$

Now using the uniform bound of $|f_k(u/s_n) - 1|$ by u_n to get

$$\log(f_k(u/s_n)) = f_k(u/s_n) - 1 + (f_k(u/s_n) - 1)^2 \varepsilon(u_n)$$

and then

$$\begin{aligned}\frac{u^2}{2} + \sum_{k=1}^{n} \log f_k(u/s_n) &= \frac{u^2}{2} + \sum_{k=1}^{n} f_k(u/s_n) - 1 + (f_k(u/s_n) - 1)^2 \varepsilon(u_n) \\ &= \left\{ \frac{u^2}{2} - \sum_{k=1}^{n} 1 - f_k(u/s_n) \right\} \\ &+ \left\{ \sum_{k=1}^{n} (f_k(u/s_n) - 1)^2 \right\} \varepsilon(u_n),\end{aligned}$$

with

$$\left| \left\{ \sum_{k=1}^{n} (f_k(u/s_n) - 1)^2 \right\} \varepsilon(u_n) \right| \leq B_n(u) \left| \varepsilon(u_n) \right| = o(1).$$

We arrive at

$$\frac{u^2}{2} = \sum_{k=1}^{n} 1 - f_k(u/s_n) + o(1).$$

If we take the real parts, we have for any fixed $\varepsilon > 0$,

$$\begin{aligned}\frac{u^2}{2} &= \sum_{k=1}^{n} \int (1 - \cos \frac{ux}{s_n}) dF_k(x) + o(1) \\ &= \sum_{k=1}^{n} \int_{|x|<\varepsilon s_n} (1 - \cos \frac{ux}{s_n}) dF_k(x) \\ &+ \sum_{k=1}^{n} \int_{|x|\geq\varepsilon s_n} (1 - \cos \frac{ux}{s_n}) dF_k(x) + o(1),\end{aligned}$$

that is

$$\frac{u^2}{2} - \sum_{k=1}^{n} \int_{|x|<\varepsilon s_n} (1 - \cos \frac{ux}{s_n}) dF_k(x) = \sum_{k=1}^{n} \int_{|x|\geq\varepsilon s_n} (1 - \cos \frac{ux}{s_n}) dF_k(x) + o(1),$$

We have by Fact 2 below that $\sqrt{2(1-\cos a)} \leq 2\,|a/2|^{\delta}$ for all $\delta, 0 < \delta \leq 1$. Apply this for $\delta = 1$ to have

$$\begin{aligned}
&\sum_{k=1}^{n} \int_{|x|<\varepsilon s_n} (1-\cos\frac{ux}{s_n}) dF_k(x) \\
\leq\ & \frac{u^2}{2s_n^2} \sum_{k=1}^{n} \int_{|x|<\varepsilon s_n} x^2 dF_k(x) \\
=\ & \frac{u^2}{2s_n^2} (\sum_{k=1}^{n} \int x^2 dF_k(x) - \sum_{k=1}^{n} \int_{|x|\geq\varepsilon s_n} x^2 dF_k(x)) \\
=\ & \frac{u^2}{2s_n^2} (s_n^2 - \sum_{k=1}^{n} \int_{|x|\geq\varepsilon s_n} x^2 dF_k(x)) = \frac{u^2}{2}(1-g_n(\varepsilon)).
\end{aligned}$$

On the other hand

$$\begin{aligned}
\sum_{k=1}^{n} \int_{|x|\geq\varepsilon s_n} (1-\cos\frac{ux}{s_n}) dF_k(x) &\leq 2\sum_{k=1}^{n} \int_{|x|\geq\varepsilon s_n} dF_k(x) \\
&= 2\sum_{k=1}^{n} \int_{|x|\geq\varepsilon s_n} \left\{\frac{1}{x^2}\right\} x^2 dF_k(x) \\
&\leq \frac{2}{\varepsilon^2 s_n^2} \sum_{k=1}^{n} \int_{|x|\geq\varepsilon s_n} x^2 dF_k(x) \leq \frac{2}{\varepsilon^2}.
\end{aligned}$$

By putting all this together, we have

$$\frac{u^2}{2} \leq \frac{u^2}{2}(1-g_n(\varepsilon)) + \frac{2}{\varepsilon^2} + o(1)$$

which leads

$$\frac{u^2}{2} g_n(\varepsilon) \leq \frac{2}{\varepsilon^2} + o(1)$$

which in turns implies

$$g_n(\varepsilon) \leq \frac{2}{u^2}(\frac{2}{\varepsilon^2} + o(1)).$$

By letting first $n \to +\infty$ and secondly $u \to 0$, we get

$$g_n(\varepsilon) \to 0.$$

This concludes the proof. $\square$

(C) APPENDIX : TOOLS, FACTS AND LEMMAS

1 - A useful development for the characteristic function.

Consider the characteristic function associated with the real probability distribution function F that is

$$\mathbb{R} \ni x \hookrightarrow f(x) = \int e^{itx} dF(x)$$

Suppose that the n^{th} moment exists, that is

$$m_n = \int x^n dF(x).$$

In the following, we also denote

$$\mu_n = \int |x|^n \, dF(x)$$

LEMMA 5. *Let $0 < \delta \leq 1$. If μ_{n+2} is finite, then we have the following expansion*

(4.14)

$$f(u) = 1 + \sum_{k=1}^{n} \frac{(iu)^k mk}{k!} + \theta 2^{1-\delta} \mu^{n+\delta} \frac{|u|^{n+\delta}}{(1+\delta)(2+\delta)...(n+\delta)}, |\theta| \leq 1.$$

Proof of Lemma 5. By using the Lebesgue Dominated Theorem, we get the f is n-times differentiable and the k-th derivative is

$$f^{(k)}(0) = i^k m_k = \int x^k dF(x), 1 \leq k \leq n. \tag{4.15}$$

We may use the Taylor-Mac-Laurin formula expansion

$$f(u) = 1 + \sum_{k=1}^{n-1} \frac{(iu)^k mk}{k!} + \int_0^u \frac{(u-x)^{n-1}}{n!} f^{(n)}(x) dx.$$

We are going to handle $\rho_n(u) = \int_0^u \frac{x^{n-1}}{n!} f^{(n)}(x) dx$. Let us make the change variable $t = x/u$ and use 4.15 to get

$$\begin{aligned}
\rho_n(u) &= u^n \int_0^1 \frac{(1-t)^{n-1}}{n!} f^{(n)}(tu)dt \\
&= (iu)^n \int_0^1 \int \frac{(1-t)^{n-1}}{(n-1)!} x^n e^{itux} dF(x)dt \\
&= (iu)^n \int_0^1 \int \frac{(1-t)^{n-1}}{(n-1)!} x^n (e^{itux} - 1 + 1) dF(x)dt \\
&= (iu)^n \int_0^1 \int \frac{(1-t)^{n-1}}{(n-1)!} x^n dF(x)dt \\
&+ (iu)^n \int_0^1 \int \frac{(1-t)^{n-1}}{(n-1)!} x^n (e^{itux} - 1) dF(x)dt
\end{aligned}$$

The first term is

$$\rho_n(1,u) = (iu)^n \int_0^1 \frac{(1-t)^{n-1}}{(n-1)!} dt \int x^n dF(x) = (iu)^n m_n \left[-\frac{(1-t)^n}{n!} \right]_{t=0}^{t=1} = \frac{(iu)^n m_n}{n!}.$$

To handle the second term, we remark that,

$$\left| e^{ia} - 1 \right| = \sqrt{2 - 1 - \cos(a)} = 2 \left| \sin(a/2) \right|.$$

Let $0 < \delta \leq 1$. If $|a/2| \geq 1$, we have

$$\left| e^{ia} - 1 \right| = 2 \left| \sin(a/2) \right| \leq |a| \leq 2 \left| a/2 \right|^{\delta}$$

by the decreasingness in δ of the function $|a/2|^{\delta}$. If $|a/2| \leq 1$, we get by Fact 1 below that also $|e^{ia} - 1| = 2 \left| \sin(a/2) \right| \leq 2 \left| a/2 \right|^{\delta}$. We have for all $a \in R$, for all $0 \leq \delta \leq 1$,

$$\left| e^{ia} - 1 \right| \leq 2 \left| a/2 \right|^{\delta}$$

Applying this to (4.16) yields

$$\begin{aligned}
|\rho_n(2,u)| &\leq |u| \int_0^1 \int \frac{(1-t)^{n-1}}{(n-1)!} |x|^n \left| e^{itux} - 1 \right| dF(x)dt \\
&\leq 2^{1-\delta} |u|^{n+\delta} \int_0^1 \frac{(1-t)^{n-1} t^{\delta}}{(n-1)!} dt \int |x|^{n+\delta} dF(x) \\
&\leq 2^{1-\delta} |u|^{n+\delta} \mu^{n+\delta} \int_0^1 \frac{(1-t)^{n-1} t^{\delta}}{(n-1)!} dt.
\end{aligned}$$

Since by Fact 2 below,

$$\int_0^1 \frac{(1-t)^{n-1}t^\delta}{(n-1)!}dt = \frac{1}{(1+\delta)(2+\delta)...(n+\delta)},$$

we get

$$\rho_n(2,u) = \theta 2^{1-\delta}\mu^{n+\delta}\frac{|u|^{n+\delta}}{(1+\delta)(2+\delta)...(n+\delta)},$$

with $|\theta| \leq 1$. By getting together all these pieces, we get (4.14). This concludes the proof of Lemma 5.

FACT 1. For any $a \in \mathbb{R}$,

$$\left|e^{ia}-1\right| = \sqrt{2(1-\cos a)} \leq 2\left|\sin(a/2)\right| \leq 2\left|a/2\right|^\delta.$$

This is easy for $|a/2| > 1$. Indeed for $\delta > 0, |a/2|^\delta > 0$ and

$$2\left|\sin(a/2)\right| \leq 2 \leq 2\left|a/2\right|^\delta$$

Now for $|a/2| > 1$, we have the expansion

$$\begin{aligned} 2(1-\cos a) &= a^2 - \sum_{k=2}^{\infty}(-1)^2\frac{a^{2k}}{(2k)!} = x^2 - 2\sum_{k\geq 2, k\ even}^{\infty}\frac{a^{2k}}{(2k)!} - \frac{a^{2(k+1)}}{(2(k+1))!} \\ &= a^2 - 2x^{2(k+1)}\sum_{k\geq 2, k\ even}^{\infty}\frac{1}{(2k)!}\left\{\frac{1}{a^2} - \frac{1}{(2k+1)((2k+2)...(2k+k)}\right\}. \end{aligned}$$

For each $k \geq 2$, for $|a/2| < 1$,

$$\left\{\frac{1}{a^2} - \frac{1}{(2k+1)((2k+2)...(2k+k)}\right\} \geq \left\{\frac{1}{4} - \frac{1}{(2k+1)((2k+2)...(2k+k)}\right\} \geq 0.$$

Hence

$$2(1-\cos a) \leq a^2.$$

But for $|a/2|$, the function $\delta \hookrightarrow |a/2|^\delta$ is non-increasing $\delta, 0 \leq \delta \leq 1$. Then

$$\sqrt{2(1-\cos a)} \leq |a| = 2\left|a/2\right|^1 \leq 2\left|a/2\right|^\delta.$$

FACT 2. For any $1 < \delta \leq 1$, for any $n \geq 1$

$$\int_0^1 \frac{(1-t)^{n-1}t^{\delta}}{(n-1)!}dt = \frac{1}{(1+\delta)(2+\delta)...(n+\delta)}.$$

Proof. By integrating by parts, we get

$$\int_0^1 \frac{(1-t)^{n-1}t^{\delta}}{(n-1)!}dt = \frac{1}{\delta+1}\left[\frac{(1-t)^{n-1}t^{\delta}}{(n-1)!}\right]_{t=0}^{t=1} + \frac{1}{\delta+1}\int_0^1 \frac{(1-t)^{n-2}t^{\delta+1}}{(n-2)!}dt,$$

that is

$$\int_0^1 \frac{(1-t)^{n-1}t^{\delta}}{(n-1)!}dt = \frac{1}{\delta+1}\int_0^1 \frac{(1-t)^{n-2}t^{\delta+1}}{(n-2)!}dt.$$

From there, we easily get by induction that, for $1 \leq \ell \leq n-1$,

$$\int_0^1 \frac{(1-t)^{n-1}t^{\delta}}{(n-1)!}dt = \frac{1}{(\delta+1)(\delta+2)...(\delta+\ell)}\int_0^1 \frac{(1-t)^{n-\ell-1}t^{\delta+\ell}}{(n-2)!}dt.$$

For $\ell = n-1$, we have

$$\begin{aligned}\int_0^1 \frac{(1-t)^{n-1}t^{\delta}}{(n-1)!}dt &= \frac{1}{(\delta+1)(\delta+2)...(\delta+n-1)}\int_0^1 t^{\delta+n-1}dt \\ &= \frac{1}{(\delta+1)(\delta+2)...(\delta+n)}.\end{aligned}$$

This finishes the proof. □

LEMMA 6. *Let Y a random variable with r_0-th finite moment, $r_0 > 0$. Then the function $g(x) = \log \mathbb{E}|Y|^x$, $0 \leq x \leq r_0$, is convex.*

Proof of Lemma 6. Let $0 \leq r_1 < r_2 \leq r_0$. Use the Cauchy-Scharwz inequality to $|Y|^{(r_1+r_2)/2}$ and $|Y|^{(r_2-r_1)/2}$ to have

$$(E|Y|^{r_1})^2 \leq E|Y|^{(r_1+r_2)} \times E|Y|^{(r_2-r_1)}$$

which implies

$$2\log E|Y|^{r_1} \leq \log \mathbb{E}|Y|^{(r_1+r_2)} + \log \mathbb{E}|Y|^{(r_2-r_1)}$$

that is, since g is continuous,

$$g(r_1) \leq \frac{1}{2}(g(r_1 + r_2) + g(r_2 - r_1)). \tag{4.16}$$

Now, set $x = r_1 + r_2$ and $y = r_2 - r_1$ and (4.16) becomes

$$g(\frac{x+y}{2}) \leq \frac{1}{2}(g(x) + g(y)) \tag{4.17}$$

for $0 \leq x \leq r_0$. Now, the Dominated Convergence Theorem, the fonction $g(\cdot)$ is continuous. So (4.17) implies the convexity of $g(\cdot)$. □

LEMMA 7. *Let X_1, X_2, ... a sequence of real and independent random variables centered at expectations, with finite $(n+\delta)-$moment, $\delta > 0$. Put for each $n \geq 1$, $S_n = X_1 + ... + X_n$ et $s_n^2 = \mathbb{E}X_1^2 + \mathbb{E}X_2^2 + ... + \mathbb{E}X_n^2$. We denote $\sigma_k^2 = \mathbb{E}X_k^2, k \geq 1$ et F_k denotes the probability distribution function of X_k. If $\delta > 1$, then any fixed $n \geq 1$,*

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left|X_k\right|^{2+\delta} \leq \left(\frac{1}{s_n^3} \sum_{k=1}^{n} \mathbb{E}\left|X_k\right|^3\right)^{(\delta-2)/\delta}. \tag{4.18}$$

Proof of Lemma 7. Let $n \geq 1$ be fixed. Let $(\pi_1, ..., \pi_n)$ following a multinomial law of n issues having all the probability n or occuring but only on repeatition. This means that only one of the $\pi_k's$ is one, the remaining being zero. Sey

$$Y = \sum_{k=1}^{n} \pi_k X_k.$$

The meaning of this expression is the following :

$$Y = X_k \text{ on } (\pi_k = 1).$$

So we have, for $r \geq 0$.

$$\left|Y\right|^r = \sum_{k=1}^{n} \pi_k \left|X_k\right|^r.$$

Hence

$$\begin{aligned} E\left|Y\right|^r &= E\sum_{j=1}^{n}\pi_j\left|X_j\right|^r = \sum_{k=1}^{n}P(\pi_k=1)E\left(\sum_{j=1}^{n}\pi_j\left|X_j\right|^r \quad \mid \pi_k=1\right) \\ &= \frac{1}{n}\sum_{k=1}^{n}E\left|X_k\right|^r. \end{aligned}$$

Use now the convexity of $g(r)=\log E\left|Y\right|^r$ for $\delta>1$ like that :

$$\frac{\delta-1}{\delta}\times 2+\frac{1}{\delta}\times(2+\delta)=3$$

and convexity implies

$$g(\frac{\delta-1}{\delta}\times 2+\frac{1}{\delta}\times(2+\delta))\leq\frac{\delta-1}{\delta}g(2)+\frac{1}{\delta}g(2+\delta)).$$

This implies

$$\delta\log E\left|Y\right|^3\leq(\delta-1)\log E\left|Y\right|^2+\log E\left|Y\right|^{2+\delta}$$

and by taking exponentials, we get

$$\left(E\left|Y\right|^3\right)^{\delta}\leq\left(E\left|Y\right|^2\right)^{\delta-1}E\left|Y\right|^{2+\delta}\Longrightarrow E\left|Y\right|^3\leq\left(E\left|Y\right|^2\right)^{\delta-1}\left(E\left|Y\right|^{2+\delta}\right)^{1/\delta}.$$

Replacing by the values of $E\left|Y\right|^r$, we get

$$\frac{1}{n}\sum_{k=1}^{n}E\left|X_k\right|^3\leq\frac{1}{n^{(\delta-1)/\delta}}s_n^{2(\delta-1)/\delta}\left(\frac{1}{n}\sum_{k=1}^{n}E\left|X_k\right|^{2+\delta}\right)^{1/\delta}.$$

From there, easy computations lead to

$$\frac{1}{s_n^3}\sum_{k=1}^{n}E\left|X_k\right|^3\leq\left(\frac{1}{s_n^{2+\delta}}\sum_{k=1}^{n}E\left|X_k\right|^{2+\delta}\right)^{1/\delta}.$$

LEMMA 4. Let $\delta>0$ and let X be a real random variable such that $\left|X\right|^{2+\delta}$ is integrable. Then

$$\left(EX^2\right)^{(2+\delta)/2}\leq E\left|X\right|^{2+\delta}.$$

PROOF. Use Lemma 2 and the convexity of $g(x)=\log E\left|X\right|^x, 0<x\leq 2+\delta$ to the convex combination

$$2=\frac{2}{2+\delta}\times(2+\delta)+\frac{2}{2+\delta}\times 0$$

to get

$$g(2) \leq \frac{2}{2+\delta} g(2+\delta) + \frac{2}{2+\delta} g(0).$$

Since $g(0) = 0$, we have

$$\log E\,|X|^2 \leq \frac{2}{2+\delta} \log E\,|X|^{2+\delta},$$

which gives the desired results upon taking the exponentials.

5. Berry-Essen approximation

Once the central theorem holds, the convergence of the distribution functions of S_n/s_n, denoted F_n, $n \geq 1$, to that of a standard Gaussian random variable denoted by G holds uniformly, by a known result of weak convergence (See Lo *et al.* (2016), chapter 4, Fact 5), that is

$$\sup_{x \in \mathbb{R}} |F_n(x) - G(x)| \rightarrow, \text{ as } n \rightarrow +\infty.$$

The Berry-Essen inequality is the most important result on the rate of convergence of F_n to G. Here is a classical form of it.

5.1. Statement of the Berry-Essen Inequality.

THEOREM 21. *(Berry-Essen) Let $X_1, X_2, ...$ be independent random variables with zero mean and with partial sums $\{S_n, n \geq 1\}$. Suppose that $\gamma_k^3 = \mathbb{E}|X_k|^3 < +\infty$ for all $k \geq 1$, and set $\sigma_k^2 = Var(X_k)$, $s_n^2 = \sum_{1 \leq j \leq k} \sigma_j^2$ and $\beta_n^3 = \sum_{1 \leq j \leq k} \gamma_j^3$. Then*

$$\sup_{x \in R} \left| \mathbb{P}\left(\frac{S_n}{s_n} \leq x\right) - \mathbb{P}(N(0,1) \leq x) \right| \leq C \frac{\beta_n^3}{s_n^3}.$$

Remarks This result may be extended to some dependent data. Generally, one seeks to get a Berry-Essen type results each time a Central limit Theorem is obtained.

The value of C may be of interest and one seeks to have it the lowest possible. In the proof below, C will be equal to 36.

PROOF The proof is very technical. But, it is important to do it at least one time, since, it may give ideas when no longer prevails the independence.

The proof itself depends on two interesting lemmas. We suggest to the reader who wants to develop an expertise in this field, to do the following.

1) The reader who wishes to master this very technical proof is recommended to read the statement and the proof of the Essen Lemma 10. This lemma gives the important formula (5.4). It is based on the inversion formula that expresses the density probability function with respect to the characteristic function. It also uses a characterization the supremum of bounded and right-continuous with left-limits (rcll)

of real-valued functions vanishing at $\pm\infty$ given in Lemma 13.

2) Next, read the statement of Lemma 12 which gives the approximation of the characteristic function of S_n/s_n to that of a standard normal random variable which is $exp(-t^2/2)$. The proof of this Lemma uses a special expansion of the characteristic function in the neighborhood of zero given in Lemma 11.

From these two points, the proof of the Theorem of Berry-Essen comes out naturally in the following lines by plugging the results of Lemma 12 in the formula (5.4) of Lemma 10. And we say :

By Lemma 5.4,

$$\sup_x \left|F_{S_n/s_n}(x) - G(x)\right| \leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\psi_{S_n/s_n}(t) - \exp(-t^2/2)}{t}\right| dt + 24A/(\pi T), \tag{5.1}$$

where A is an upper bound of the derivative the standard gaussian distribution function G whose infimum is $1/\sqrt{2\pi}$. Take $A = 1/\sqrt{2\pi}$ and $T = T_n = s_n^3/(4\beta_n^3)$. We use Formula (5.1) and the following inequality

$$\left\|\psi_{S_n/s_n}(t) - \exp(-t^2/2)\right\| \leq 16\exp(-t^2/2)\frac{\beta_n^3\,|t|^3}{s_n^3}.$$

to grap on

$$\begin{aligned}
\sup_x \left|F_{S_n/s_n}(x) - G(x)\right| &\leq \frac{16}{\pi}\frac{\beta_n^3}{s_n^3}\int_{-s_n^3/(4\beta_n^3)}^{s_n^3/(4\beta_n^3)} t^2\exp(-t^2/2)dt + \frac{24(1/)}{\pi s_n^3/(4\beta_n^3)} \\
&\leq \frac{16}{\pi}\frac{\beta_n^3}{s_n^3}\int_{s_n^3/(4\beta_n^3)}^{s_n^3/(4\beta_n^3)} t^2\exp(-t^2/2)dt + \frac{96\beta_n^3}{\pi\sqrt{2\pi}s_n^3}.
\end{aligned}$$

The integral $\int_{s_n^3/(4\beta_n^3)}^{s_n^3/(4\beta_n^3)} t^2\exp(-t^2/2)dt$ is bounded by

$$\int_{s_n^3/(4\beta_n^3)}^{s_n^3/(4\beta_n^3)} t^2\exp(-t^2/2)dt = \frac{3}{2}\sqrt{2\pi}.$$

We get

$$\begin{aligned}\sup_x \left|F_{S_n/s_n}(x) - G(x)\right| &\leq \frac{24\sqrt{3}}{\sqrt{\pi}}\frac{\beta_n^3}{s_n^3} + \frac{96\beta_n^3}{\pi\sqrt{2\pi}s_n^3}\\ &\leq \left(\frac{24\sqrt{3}}{\sqrt{\pi}} + \frac{96}{\pi\sqrt{2\pi}}\right)\frac{\beta_n^3}{s_n^3}\\ &\leq 36\frac{\beta_n^3}{s_n^3}.\end{aligned}$$

This concludes the proof.

5.2. Tools, Facts and Lemmas.

LEMMA 8. *Define the triangle probability density function pdf, with parameter T as following*

$$f_{tri}(x) = \frac{1}{T}(1 - \frac{|x|}{T})1_{(|x|\leq T)}. \tag{5.2}$$

(i) Then its characteristic function is

$$\Phi_{tri(T)}(t) = \sin^2(tT/2)/(tT)^2.$$

(ii) The function

$$g(x) = \frac{1-\cos xT}{\pi x^2 T}, x \in \mathbb{R},$$

defines a density distribution function and its characteristic function is $1 - |t|/T$ *.*

Proof. We have

$$\begin{aligned}\Phi_{tri(T)}(t) &= \frac{1}{T}\int_{-T}^{T} e^{itx}(1 - \frac{|x|}{T})dx\\ &= \frac{1}{T}\left\{\int_{-T}^{0} e^{itx}(1+\frac{x}{T})dx + \int_0^T e^{itx}(1-\frac{x}{T})dx\right\}.\end{aligned}$$

Next, we have

$$\int_{-T}^{0} e^{itx}(1+\frac{x}{T})dx = \left[\frac{e^{itx}}{it}\right]_{-T}^{0} + \frac{1}{T}\int_{-T}^{0} xe^{itx}dx.$$

By integrating by parts, we get

$$\begin{aligned}\int_{-T}^{0} e^{itx}(1+\frac{x}{T})dx &= \left[\frac{e^{itx}}{it}\right]_{-T}^{0} + \frac{1}{T}\left[\frac{xe^{itx}}{it}\right]_{-T}^{0} + \frac{1}{itT}\int_{-T}^{0} e^{itx}dx \\ &= \left[\frac{e^{itx}}{it}\right]_{-T}^{0} + \frac{1}{T}\left[\frac{xe^{itx}}{it}\right]_{-T}^{0} - \frac{1}{itT}\left[\frac{e^{itx}}{it}\right]_{-T}^{0} \\ &= \frac{1}{it}(1-e^{-itT}) + \frac{1}{it}e^{-itT} + \frac{1}{t^2T}(1-e^{-itT}).\end{aligned}$$

Likewise, we get

$$\begin{aligned}\int_{0}^{T} e^{itx}(1-\frac{x}{T})dx &= \left[\frac{e^{itx}}{it}\right]_{0}^{T} - \frac{1}{T}\int_{0}^{T} xe^{itx}dx. \\ &= \left[\frac{e^{itx}}{it}\right]_{0}^{T} - \frac{1}{T}\left[\frac{xe^{itx}}{it}\right]_{0}^{T} + \frac{1}{itT}\int_{0}^{T} e^{itx}dx \\ &= \left[\frac{e^{itx}}{it}\right]_{0}^{T} - \frac{1}{T}\left[\frac{xe^{itx}}{it}\right]_{0}^{T} + \frac{1}{itT}\left[\frac{e^{itx}}{it}\right]_{0}^{T} \\ &= \frac{1}{it}\left(e^{itT}-1) - \frac{1}{it}e^{itT} - \frac{1}{t^2T}(e^{itT}-1\right).\end{aligned}$$

By putting all this together, and by adding term by term, we get

$$\begin{aligned}\Phi_{tri(T)}(t) &= \frac{1}{T}\left\{\frac{2\sin tT}{t} - \frac{2\sin tT}{t} - \frac{2\cos tT - 2}{t^2T}\right\} \\ &= \frac{2(1-\cos tT)}{t^2T^2} \\ &= \frac{\sin^2 tT/2}{t^2T^2}.\end{aligned}$$

Remark that $\Phi_{tri(T)}(t)$ is well defined for $t = 0$. From now, we may use the inversion theorem for an absolutely continuous distribution function :

$$f_{tri(T)}(t) = \frac{1}{2\pi}\int e^{-itx}\Phi_{tri(T)}(x)dx.$$

Then for $|t| \leq T$,

$$\frac{1}{T}\left(1-\frac{|t|}{T}\right) = \frac{1}{2\pi}\int e^{-itx}\frac{\sin^2 xT/2}{x^2T^2}dx,$$

which gives

$$\begin{aligned}1-\frac{|t|}{T} &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\frac{\sin^2 xT/2}{2\pi x^2T^2}dx\\ &= \frac{1}{\pi}\int_{\mathbb{R}} e^{-itx}\frac{1-\cos xT}{x^2T^2}dx\\ &= \int_{\mathbb{R}} e^{itx}\frac{1-\cos xT}{\pi x^2T^2}dx.\end{aligned}$$

Taking $t=0$ in that formula proves that

$$\frac{1-\cos xT}{\pi x^2T^2}, x\in\mathbb{R}$$

is a density probability on $\mathbb{R}$, and its characteristic function is $1-|t|/T$. This gives

LEMMA 9. *The following function*

$$\frac{1-\cos xT}{\pi x^2T^2}, x\in\mathbb{R}$$

is a probability density function with characteristic function $1-|t|/T$.

The following lemma uses the inverse formulas in Proposition 5 (see section Part V, Section 6, Chapter 2).

LEMMA 10. *Let U and V be two random variables, and suppose that*

$$\sup_{x\in\mathbb{R}} F_V'(x)\leq A. \tag{5.3}$$

Then

$$\sup_x |F_U(x)-F_V(x)| \leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\psi_U(t)-\psi_V(t)}{t}\right|dt + 24A/(\pi T). \tag{5.4}$$

Proof. Suppose that

$$\int_{-T}^{T}\left|\frac{\psi_U(t)-\psi_V(t)}{t}\right|dt<+\infty,$$

for $T>0$, otherwise (5.4) is obvious. We may consider, by using Kolmogorov Theorem, that we are on a probability space holding the ordered pair (U,V) and an absolutely continuous random variable Z_T with characteristic function $(1-|t|/T)1_{(|t|\leq T)}$ as allowed by Lemma 9 such that Z_T is independent from U and V. Given the *cdf* $F_{(U,V)}$ of (U,V) and the *cdf* F_{Z_T}, the *cdf* of (U,V,Z_T) is given by

$$F_{(U,V,Z_T)}(u,v,z) = F_{(U,V)}(u,v) \times F_{Z_T}(z), \quad (u,v,z) \in \mathbb{R}^3.$$

The probability space by using the Lebesgue-Stieljes measure of $F_{(U,V,Z_T)}$ following Point (c5), Section (page 62).

Now, we recall the convolution formula on U and Z_T :

$$F_{U+Z_T}(x) = \int F_U(x-y) f_{Z_T}(y)\ dy,\ x \in \mathbb{R}.$$

Define F_{U+Z_T} likewise. Set

$$\Delta(x) = F_U(x) - F_V(x),\ x \in \mathbb{R}.$$

and

$$\Delta_T(x) = \int \Delta(x-y) f_{Z_T}(y) dy = F_{U+Z_T}(x) - F_{V+Z_T}(x). \tag{5.5}$$

We remark that for any fixed t, $\psi_{Z_T}(t) = (1 - |t|/T)1_{(|t| \leq T)} \rightarrow 1$, which is the characteristic function of 0. Then Z_T weakly converges to 0, that is equivalent to $Z_T \rightarrow_P 0$. Using results of weak theory (see for example Chapter 5, Subsection 3.2.3, Proposition 21 in Lo *et al.* (2016)) implies that $U + Z_T \rightsquigarrow U$ and $V + Z_T \rightsquigarrow V$. By returning back to the distribution functions that are continuous, we have from (5.5)

$$\text{For any x, } \Delta_T(x) \rightarrow \Delta(x) \text{ as } T \rightarrow \infty.$$

By applying Proposition 5 (see section Part V, Section 6, Chapter 2), we have for continuity points x and b of both F_{U+Z_T} and F_{V+Z_T}, with $b < x$,

$$\begin{aligned}
& (F_{U+Z_T}(x) - F_{U+Z_T}(b) - ((F_{U+Z_T}(x) - F_{U+Z_T}(b)) \\
= & \lim_{U \rightarrow +\infty} \frac{1}{2\pi} \int_{-U}^{U} \frac{e^{-ixt} - e^{-ibt}}{it} \left(\psi_U(t) - \psi_V(t)\right) \psi_{Z_T}(t) dt \\
= & \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ixt} - e^{-ibt}}{it} \left(\psi_U(t) - \psi_V(t)\right) \psi_{Z_T}(t) dt, \\
= & \frac{1}{2\pi} \int_{-T}^{T} \left(\psi_U(t) - \psi_V(t)\right) \psi_{Z_T}(t) \left(- \int_x^b e^{-itv} dv \right) dt,
\end{aligned}$$

since $\psi_{Z_T}(t)$ vanishes outside $[-T,T]$. By letting $b \downarrow -\infty$ over the set of continuity points of both F_{U+Z_T} and F_{V+Z_T}, and by using the Fatou-Lebesgue convergence theorem at right, we get

$$\begin{aligned} & F_{U+Z_T}(x) - F_{U+Z_T}(x) \\ = \ & \frac{1}{2\pi}\int_{-T}^{T} \left(\psi_U(t) - \psi_V(t)\right)\psi_{Z_T}(t)\left(-\int_{x}^{-\infty} e^{-itv}dv\right)dt \\ = \ & \frac{1}{2\pi}\int_{-T}^{T} -e^{-itx}\left(\psi_U(t) - \psi_V(t)\right)\psi_{Z_T}(t)\left(-\int_{x}^{-\infty} e^{-itv}dv\right)dt \end{aligned}$$

which gives, for any continuity point x of both F_{U+Z_T} and F_{V+Z_T}. By taking the supremum of those continuity point x of both F_{U+Z_T} and F_{V+Z_T} (which amounts to taking the supremum over $\mathbb{R}$ by right-continuity), we finally get

$$\|\Delta_T\|_{+\infty} \leq \frac{1}{2\pi}\int_{-T}^{T}\left|\frac{\psi_U(t) - \psi_V(t)}{t}\right| dt.$$

Since we want to prove Formula (5.4), the last formula says it will be enough to prove

$$\|\Delta\|_\infty \leq 2\,\|\Delta_T\|_\infty + 24A/(\pi T) \tag{5.6}$$

We remark that Δ is bounded and is right-continuous with left-limits at each point of $\mathbb{R}$ and $\Delta(+\infty) = \Delta(-\infty) = 0$. Then by Lemma 13 below, there exists a $x_0 \in R$ such that $\|\Delta\|_\infty = |\Delta(x_0)|$ or $\|\Delta\|_\infty = |f(x_0-)|$. We continue with the case where $\|\Delta\|_\infty = |\Delta(x_0)| = \Delta(x_0)$. Handling the other cases is similar. We have for any $s > 0$:

$$\Delta(x_0+s) - \Delta(x_0) = \{F_U(x_0+s) - F_U(x_0)\} - \{F_V(x_0+s) - F_V(x_0)\}$$

and, by (5.3),

$$F_V(x_0+s) - F_V(x_0) = \int_{x_0}^{x_0+s} F_V'(t)dt \leq As.$$

Next

$$\Delta(x_0+s) - \Delta(x_0) = \{F_U(x_0+s) - F_U(x_0)\} - As \geq -As$$

since $\{F_U(x_0+s) - F_U(x_0)\} \geq 0$ (F_U increasing). This gives for any $s \geq 0$

$$\Delta(x_0+s) \geq \|\Delta\|_\infty - As.$$

By applying this to $s = \|\Delta\|_\infty /(2A)) + y$ for $|y| \leq \|\Delta\|_\infty /(2A)$, we get

$$\Delta\left(x_0 + \frac{\|\Delta\|_\infty}{2A} + y\right) \geq \frac{\|\Delta\|_\infty}{2} - Ay \tag{5.7}$$

We going to apply this to Δ_T, while reminding the definition, to see that

$$\begin{aligned}
\Delta_T\left(x_0 + \frac{\|\Delta\|_\infty}{2A}\right) &= \int \Delta\left(x_0 + \frac{\|\Delta\|_\infty}{2A} - y\right) f_{Z_T}(y)dy \\
&= \int_{\left\{|y|\leq \|\Delta\|_\infty/(2A)\right\}} \Delta\left(x_0 + \frac{\|\Delta\|_\infty}{2A} - y\right) f_{Z_T}(y)dy \\
&+ \int_{\left\{|y|> \|\Delta\|_\infty/(2A)\right\}} \Delta\left(x_0 + \frac{\|\Delta\|_\infty}{2A} - y\right) f_{Z_T}(y)dy.
\end{aligned}$$

On one hand, by (5.7), we have

$$\begin{aligned}
&\int_{\{|y|\leq \|\Delta\|/(2A)\}} \Delta(x_0 + \frac{\|\Delta\|_\infty}{2A} - y) f_{Z_T}(y)dy \\
\geq\ & \int_{\left\{|y|\leq \|\Delta\|_\infty/(2A)\right\}} \left(\frac{\|\Delta\|_\infty}{2} - Ay\right) f_{Z_T}(y)dy \\
\geq\ & \frac{\|\Delta\|_\infty}{2} \int_{\{|y|\leq \|\Delta\|/(2A)\}} f_{Z_T}(y)dy \\
=\ & \frac{\|\Delta\|_\infty}{2} P\left(|Z_T| \leq \frac{\|\Delta\|_\infty}{2A}\right) \\
=\ & \frac{\|\Delta\|_\infty}{2} \left\{1 - P\left(|Z_T| > \frac{\|\Delta\|_\infty}{2A}\right)\right\}
\end{aligned}$$

and for the other term, we use the following trivial inequality

$$\Delta(\cdot) \geq -\sup_{x\in R} |\Delta(x)| = -\|\Delta\|_\infty$$

to have

$$\begin{aligned}
&\int_{\{|y|> \|\Delta\|/(2A)\}} \Delta\left(x_0 + \frac{\|\Delta\|_\infty}{2A} - y\right) f_{Z_T}(y)dy \\
\geq\ & -\|\Delta\|_\infty \int_{\{|y|> \|\Delta\|/(2A)\}} f_{Z_T}(y)dy \\
\geq\ & -\Delta \|\Delta\|_\infty \int_{\{|y|> \|\Delta\|/(2A)\}} f_{Z_T}(y)dy \\
=\ & -\Delta \|\Delta\|_\infty P\left(|Z_T| > \frac{\|\Delta\|_\infty}{2A}\right).
\end{aligned}$$

The two last formulas lead to

$$\Delta_T\left(x_0+\frac{\|\Delta\|_\infty}{2A}\right)\geq\frac{\|\Delta\|_\infty}{2}\left\{1-3P\left(|Z_T|>\frac{\|\Delta\|_\infty}{2A}\right)\right\}$$

and next

$$\|\Delta_T\|_\infty\geq\frac{\|\Delta\|_\infty}{2}\left\{1-3P\left(|Z_T|>\frac{\|\Delta\|_\infty}{2A}\right)\right\}. \tag{5.8}$$

In this last step, we have

$$\begin{aligned}P\left(|Z_T|>\frac{\|\Delta\|_\infty}{2A}\right) &= \int_{\left\{|y|>\|\Delta\|_\infty/(2A)\right\}}\frac{1-\cos yT}{\pi Ty^2}dy\\ &= \int_{\Delta\|\Delta\|_\infty/(2A)}^{+\infty}\frac{1-\cos yT}{\pi Ty^2}dy\\ &= \frac{1}{\pi}\int_{\Delta\|\Delta\|_\infty T/(4A)}^{+\infty}\frac{1-\cos 2y}{y^2}dy\\ &= \frac{2}{\pi}\int_{\Delta\|\Delta\|_\infty T/(4A)}^{+\infty}\frac{\sin^2 y}{y^2}dy\\ &\leq \frac{2}{\pi}\int_{\|\Delta\|_\infty T/(4A)}^{+\infty}\frac{1}{y^2}dy\\ &= \frac{8A}{\pi T\|\Delta\|_\infty}.\end{aligned}$$

This and (5.8) yield

$$\begin{aligned}2\|\Delta_T\|_\infty &\geq \left\{\|\Delta\|_\infty-3\|\Delta\|_\infty P\left(|Z_T|>\frac{\|\Delta\|_\infty}{2A}\right)\right\}\\ &\geq \|\Delta\|_\infty-\frac{24A}{T\pi},\end{aligned}$$

which implies

$$\|\Delta\|_\infty\leq 2\|\Delta_T\|_\infty+\frac{24A}{T\pi}.$$

This was the target, that is Formula (5.6), which is enough to have the final result (5.4). ■

Technical Lemmas used by the proof.

LEMMA 11. . *Let X be a real random variable with $n+1$ finite moments. Then for any $t \in \mathbb{R}$,*

$$\left|\psi_X(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!}\mathbb{E}\,|X|^k\right| \leq \min\left(\frac{2\,|t|^n}{n!}\mathbb{E}\,|X|^n\,,\, \frac{|t|^{n+1}}{(n+1)!}\mathbb{E}\,|X|^{n+1}\right).$$

Proof. We may use the Taylor-Mac-Laurin expansion formula,

$$f(y) = \sum_{k=0}^{n} \frac{y^k}{k!} f^{(k)}(0) + \int_0^y \frac{(y-x)^n}{n!} f^{(n)}(x)dx$$

for $f(y) = e^{iy}$. We have $f^{(k)}(y) = i^k f(y)$ and then

$$e^{iy} = \sum_{k=0}^{n} \frac{(iy)^k}{k!} + i^{n+1}\int_0^y e^{ix}\frac{(y-x)^n}{n!}dx. \tag{5.9}$$

Then

$$\left|e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}\right| \leq \frac{1}{n!}\int_0^{|y|} |y-x|^n\,dx \leq \frac{|y|^{n+1}}{(n+1)!}.$$

We apply (5.9) for $n-1$, that is

$$e^{iy} = \sum_{k=0}^{n-1} \frac{(iy)^k}{k!} + i^n\int_0^y e^{ix}\frac{(y-x)^{n-1}}{(n-1)!}dx$$

and we use the decomposition $e^{ix} = 1 + (e^{ix} - 1)$ to get

$$e^{iy} = \sum_{k=0}^{n-1} \frac{(iy)^k}{k!} + i^n\int_0^y \frac{(y-x)^{n-1}}{(n-1)!}dx + i^n\int_0^y (e^{ix}-1)\frac{(y-x)^{n-1}}{(n-1)!}dx.$$

We have

$$i^n\int_0^y \frac{(y-x)^{n-1}}{(n-1)!}dx = (iy)^n,$$

which leads to

$$e^{iy} = \sum_{k=0}^{n} \frac{(iy)^k}{k!} + i^n\int_0^y (e^{ix}-1)\frac{(y-x)^{n-1}}{(n-1)!}dx$$

and next, since $|(e^{ix}-1)| \leq 2$,

$$\left|e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}\right| \leq \frac{2}{(n-1)!} \int_0^{|y|} |y-x|^n \, dx \leq \frac{2\,|y|^n}{n!}.$$

We then get

$$\left|e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}\right| \leq \min\left(\frac{2\,|y|^n}{n!}, \frac{|y|^{n+1}}{(n+1)!}\right).$$

We apply this to a random real variable X with enough finite moments to get

$$\left|\mathbb{E}e^{itX} - \mathbb{E}\sum_{k=0}^{n} \frac{(itX)^k}{k!}\right| \leq \mathbb{E}\left|e^{itX} - \sum_{k=0}^{n} \frac{(itX)^k}{k!}\right| \leq \mathbb{E}\min\left(\frac{2\,|tX|^n}{n!}, \frac{|tX|^{n+1}}{(n+1)!}\right),$$

and then

$$\left|\psi_X(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!}\mathbb{E}\,|X|^k\right| \leq \min\left(\frac{2\,|t|^n}{n!}\mathbb{E}\,|X|^n, \frac{|t|^{n+1}}{(n+1)!}\mathbb{E}\,|X|^{n+1}\right)$$

LEMMA 12. *With the notations and assumptions of the Theorem, we have :*

Part 1.

$$\left|\exp(itS_n/s_n) - \exp(-t^2/2)\right| \leq 0.4466464\frac{\beta_n^3}{s_n^3}\,|t|^3 \exp(-\,|t|^2/2) \text{ for } |t| \leq s_n/(2\beta_n).$$

Part 2.

$$\left|\exp(itS_n/s_n) - \exp(-t^2/2)\right| \leq 16\frac{\beta_n^3}{s_n^3}\,|t|^3 \exp(-\,|t|^2/3) \text{ for } |t| \leq s_n^3/(4\beta_n^3).$$

Proof.

Proof of Part 1. Let us prove that

$$\left|\exp(itS_n/s_n) - \exp(-t^2/2)\right| \leq 0.5\frac{\beta_n^3}{s_n^3}\,|t|^3 \exp(-\,|t|^3/3) \text{ for } |t| \leq s_n/(2\beta_n)$$

To this end, we use the following expansion

$$\exp(itX_k/s_n) = 1 - \frac{t^2\sigma_k^2}{2s_n^2} + \theta\frac{|t|^3\,\gamma_k^3}{6s_n^3}, \quad |\theta| \leq 1.$$

For $|t| \leq s_n/(2\beta_n)$,

$$r_{1,k} = \left|\theta \frac{|t|^3 \gamma_k^3}{6s_n^3}\right| \leq \frac{|t|^3 \gamma_k^3}{6s_n^3}.$$

Next

$$\begin{aligned}\log \exp(itX_k/s_n) &= \log\left(1 - \frac{t^2\sigma_k^2}{2s_n^2} + r_{1,k}\right)\\ &= -\frac{t^2\sigma_k^2}{2s_n^2} + r_{1,k} + r_{2,k},\end{aligned}$$

with, after having used the c_r-inequality,

$$\begin{aligned} r_{2,k} &\leq \left|-\frac{t^2\sigma_k^2}{2s_n^2} + r_1\right|^2 \leq 2\left|-\frac{t^2\sigma_k^2}{2s_n^2}\right|^2 + 2\left|r_1\right|^2 \\ &\leq \frac{1}{2}\left(\frac{|t|^3\sigma_k^3}{s_n^3}\right)\left(\frac{|t|\,\sigma_k}{s_n}\right) + \frac{1}{18}\left(\frac{|t|^3\gamma_k^3}{s_n^3}\right)\left\{\frac{|t|^3\gamma_k^3}{s_n^3}\right\}.\end{aligned}$$

Now $|t| \leq s_n/(2\beta_n)$ implies

$$\left(\frac{|t|\,\sigma_k}{s_n}\right) \leq \left(\frac{\sigma_k}{s_n} \times \frac{s_n}{2\beta_n}\right) = \frac{1}{2}\frac{\sigma_k}{\beta_n} = \frac{1}{2}\left(\frac{\sigma_k^3}{\beta_n^3}\right)^{1/3} \leq \frac{1}{2}\left(\frac{\sigma_k^3}{\gamma_n^3}\right)^{1/3} = \frac{1}{2}\frac{\sigma_k}{\gamma_k} \leq \frac{1}{2},$$

by Lyapounov's inequality, that is for $1 \leq p \leq q$,

$$(E\left|X_k\right|^p)^{1/p} \leq (E\left|X_k\right|^q)^{1/q}$$

and next

$$\left\{\frac{|t|^3\gamma_k^3}{s_n^3}\right\} \leq \left\{\frac{\gamma_k^3}{s_n^3} \times \frac{s_n^3}{8\beta_n^3}\right\} \leq \frac{1}{8}.$$

We arrive, after applying again Lyapounov's inequality, at

$$r_{2,k} \leq \frac{1}{2}\left(\frac{|t|^3\sigma_k^3}{s_n^3}\right) \times \frac{1}{2} + \frac{1}{18}\left(\frac{|t|^3\gamma_k^3}{s_n^3}\right)\frac{1}{8} = \frac{37}{144}\frac{|t|^3\gamma_k^3}{s_n^3}.$$

Next

$$\begin{aligned}
\log \exp(itS_n/s_n) &= \sum_{k=1}^{n} \log \exp(itX_k/s_n) \\
&= -t^2/2 + r_n \\
&= -t^2/2 + \sum_{k=1}^{n} \frac{|t|^3 \gamma_k^3}{6s_n^3} + \frac{37}{144}\frac{|t|^3 \gamma_k^3}{s_n^3} \\
&= -t^2/2 + \frac{61}{144}\frac{|t|^3 \beta_n^3}{s_n^3} \\
&\leq -t^2/2 + \frac{61}{144} \times \frac{1}{8},
\end{aligned}$$

where we used $|t| \leq s_n/(2\beta_n)$ at the last step. We already have

$$\exp(itS_n/s_n) = \exp(-t^2/2 + r_n)$$

so that

$$\left|\exp(itS_n/s_n) - \exp(-t^2/2)\right| = \exp(-t^2/2)\left\|e^{r_n} - 1\right\|.$$

We use the formula $\|e^z - 1\| \leq \|z\| \, e^{\|z\|}$ to see that

$$\begin{aligned}
\left|\exp(itS_n/s_n) - \exp(-t^2/2)\right| &= \exp(-t^2/2)\left\|e^{r_n} - 1\right\| \\
&\leq \exp(-t^2/2)\left\|r_n\right\| e^{\|r_n\|} \\
&\leq \exp(-t^2/2)\frac{61}{144}\frac{|t|^3 \beta_n^3}{s_n^3} e^{61/(8*144)} \\
&\leq 0.4466464 \times \exp(-t^2/2)\frac{|t|^3 \beta_n^3}{s_n^3} \\
&\leq 16 \times \exp(-t^2/2)\frac{|t|^3 \beta_n^3}{s_n^3} \\
&\leq 16 \times \exp(-t^2/2)\frac{|t|^3 \beta_n^3}{s_n^3},
\end{aligned}$$

since $\exp(-t^2/2) \leq \exp(-t^2/3)$.

Proof of Part 2. This is proved as follows. If $s_n^3/(4\beta_n^3) \leq s_n/(2\beta_n)$, Part 2 is implied by Part 1. Then, we only need to prove Part 2 in the case

$$s_n/(2\beta_n) < s_n^3/(4\beta_n^3),$$

and only for t satisfying

$$s_n/(2\beta_n) < t \leq s_n^3/(4\beta_n^3).$$

Let us proceed by considering the symmetrized form of X_k, denoted by X_k^s, and defined by

$$X_k^s = X_k - X_k',$$

where X_k' is a random variable with the same law than X_k and independent of X_k. Then, obviously, $\mathbb{E}X_k^s = \mathbb{E}X_k - \mathbb{E}X_k' = \mathbb{E}X_k - \mathbb{E}X_k = 0$ and

$$\mathbb{V}ar(X_k^s) = \mathbb{V}ar(X_k) + \mathbb{V}ar(X_k') + 2\ \mathbb{C}ov(X_k, X_k') = \sigma_k^2 + \sigma_k^2 + 0 = 2\sigma_k^2$$

and finally, by the C_r-inequality

$$\mathbb{E}\left|X_k^s\right|^r = \mathbb{E}\left|X_k - X_k\right|^r \leq c_r(\mathbb{E}\left|X_k\right|^r + \mathbb{E}\left|X_k'\right|^r),$$

with $c_r = 2^{r-1}$, $r \geq 1$. Apply it to $r = 3$ to get

$$\mathbb{E}\left|X_k^s\right|^3 \leq 4(\mathbb{E}\left|X_k\right|^3 + \mathbb{E}\left|X_k'\right|^3) = 8\gamma_k^3.$$

Now, we remark that we have for any real random variable

$$\psi_X(t) = \int \cos(tx)\ d\mathbb{P}_X(x) + i \int \sin(tx)\ d\mathbb{P}_X(x).$$

and

$$\psi_{-X}(t) = x\psi_{-X}(t) = \int \cos(tx)\ d\mathbb{P}_X(x) - i \int \sin(tx)\ d\mathbb{P}_X(x) = \overline{\psi_X(t)},$$

where $\overline{\psi_X(t)}$ is the conjugate of $\psi_X(t)$. Next, from this and by independence, we have

$$\begin{aligned} \psi_{X_k^s}(t) &= \psi_{X_k - X_k'}(t) = \psi_{X_k}(t)\psi_{-X_k'}(t) = \psi_{X_k}(t)\psi_{-X_k}(t) \\ &= \psi_{X_k}(t)\overline{\psi_{X_k}(t)} = \left\|\psi_{X_k}(t)\right\|^2, \end{aligned}$$

where, here, $\|\circ\|$ denotes the norm in the complex space. Next, we apply Lemma 11 to X_k^s at the order $n = 2$ to get

$$\left|\psi_{X_k^s}(t) - (1 - 2\sigma_k^2 t^2)\right| \leq \frac{8\left|t\right|^3}{6}\gamma_k^3 = \frac{4\left|t\right|^3}{3}\gamma_k^3.$$

The triangle inequality leads to

$$\psi_{X_k^s}(t) \leq 1 - 2\sigma_k^2 t^2 + \frac{4\left|t\right|^3}{3}\gamma_k^3,$$

which gives ($\bowtie$)

$$\psi_{X_k^s/s_n}(t) \leq \exp\{-2\frac{\sigma_k^2 t^2}{s_n^2} + \frac{4\,|t|^3}{3s_n^3}\gamma_k^3\}.$$

Denote also $S_n^s = X_1^s + ... + X_n^s$. Then, by reminding that $\psi_{X_k^s}(t)$ is real and non-negative and that $\left\|\psi_{X_k^s/s_n}(t)\right\|$ is an absolute value, we have

$$\begin{aligned}\psi_{S_n^s/s_n}(t) &= \prod_{k=1}^{n}\psi_{X_k^s}(t) \leq \prod_{k=1}^{n}\exp\{-\frac{\sigma_k^2 t^2}{s_n^2} + \frac{4\,|t|^3}{3s_n^3}\gamma_k^3\}\\ &= \exp\{-\sum_{k=1}^{n}\frac{\sigma_k^2 t^2}{s_n^2} + \sum_{k=1}^{n}\frac{4\,|t|^3}{3s_n^3}\gamma_k^3\}\\ &= \exp\{-t^2 + \frac{4\,|t|^3\,\beta_n^3}{3s_n^3}\}.\end{aligned}$$

Now, for $|t| \leq s_n^3/(4\beta_n^3)$, it comes that

$$\begin{aligned}\psi_{S_n^s/s_n}(t) &\leq \exp\{-t^2 + \frac{4\,|t|^3\,\beta_n^3}{3s_n^3}\}\\ &\leq \exp\{-t^2 + \frac{t^2}{3}\} = \exp\{-2t^2/3\}.\end{aligned}$$

Since $\psi_{S_n^s/s_n}(t) = \left\|\psi_{S_n/s_n}(t)\right\|^2$, we have

$$\left\|\psi_{S_n/s_n}(t)\right\| \leq \exp\{-t^2/3\}.$$

Now, since, $s_n/(2\beta_n) < |t|$,

$$\begin{aligned}1 &\leq \frac{2\beta_n\,|t|}{s_n} = \frac{2\beta_n\,|t|^3}{s_n} \times \left(\frac{1}{|t|^2}\right)\\ &\leq \frac{2\beta_n\,|t|^3}{s_n} \times \left(\frac{4\beta_n^2}{s_n^2}\right) = \frac{8\beta_n^3\,|t|^3}{s_n^3}.\end{aligned}$$

To conclude, we say that

$$\begin{aligned}\left\|\psi_{S_n/s_n}(t) - \exp(-t^2/2)\right\| &\leq \exp(-t^2/2) + \left\|\psi_{S_n/s_n}(t)\right\|\\ &\leq \exp(-t^2/2) + \exp\{-t^2/3\}\\ &\leq 2\exp(-t^2/2).\end{aligned}$$

We conclude by using the following stuff :

$$\begin{aligned}\left\|\psi_{S_n/s_n}(t)-\exp(-t^2/2)\right\| &\leq 2\exp(-t^2/2)\times(1)\\ &\leq 2\exp(-t^2/2)\frac{8\beta_n^3\,|t|^3}{s_n^3}\\ &= 16\exp(-t^2/2)\frac{\beta_n^3\,|t|^3}{s_n^3}.\end{aligned}$$

The following lemma on elementary real analysis has been used in the proof the Essen Lemma 10.

Lemma 13. *Let f be a bounded and non-constant right-continuous mapping from $\mathbb{R}$ to $\mathbb{R}$ with left-limits at each point of $\mathbb{R}$, such that*

$$\lim_{x\to-\infty} f(x)=0 \text{ and } \lim_{x\to+\infty} f(x)=0.$$

Then there exists some $x_0\in\mathbb{R}$ such that

$$0<c=\sup_{x\in\mathbb{R}}|f(x)|=|f(x_0)| \text{ or } c=|f(x_0-)|$$

where $f(x-)$ stands for the left-limit of f at x.

Proof. Let $c=\sup_{x\in\mathbb{R}}|f(x)|$. The number c is strictly positive, otherwise f would be constant and equal to zero, which would be contrary to the assumption. Now since $\lim_{x\to-\infty} f(x)=0$ and $\lim_{x\to+\infty} f(x)=0$, we can find $A>0$ such that

$$\forall x, (|x|>A)\Longrightarrow(|f(x)|<c/2).$$

So, we get

$$c=\sup_{x\in[-A,A]}|f(x)|.$$

We remark that c is finite since f is bounded. Now consider a sequence $(x_n)_{n\geq 0}\subset[-A,A]$ such that $|f(x_n)|\to c$. Since $(x_n)_{n\geq 0}\subset[-A,A]$, by the Bolzano-Weierstrass property, there exists a subsequence $(x_{n(k)})_{k\geq 0}\subset(x_n)_{n\geq 0}$ converging to some $x_0\in[-A,A]$. Consider

$$I(\ell)=\{k\geq 1, x_{n(k)}\geq x_0\} \text{ and } I(r)=\{k\geq 1, x_{n(k)}<x_0\}.$$

One of these two set is infinite. If $I(\ell)$ is infinite, we can find a subsequence $\left(x_{n(k_j)}\right)_{j\geq 1}$ such that $x_{n(k_j)}\geq x_0$ for any $j\geq 1$ and $x_{n(k_j)}\to x_0$

as $j \to \infty$. Then by right-continuity, $|f(x_{n(k_j)})| \to |f(x_0)|$ and as a sub-sequence of $|f(x_n)|$ which converges to c, we also have $|f(x_{n(k_j)})| \to c$ as $j \to \infty$. Then

$$c = |f(x_0)|$$

If $I(r)$ is infinite, we can find a sub-sequence $\left(x_{n(k_j)}\right)_{j \geq 1}$ such that $x_{n(k_j)} < x_0$ for any $j \geq 1$ and $x_{n(k_j)} \to x_0$ as $j \to \infty$. Then by the existence of the left-limit of f at x_0, $|f(x_{n(k_j)})| \to |f(x_0-)|$ and as a sub-sequence of $|f(x_n)|$ which converges to c, we also have $|f(x_{n(k_j)})| \to c$ as $j \to \infty$. Then

$$c = |f(x_0-)| .$$

LEMMA 14. *We have the following inequality, for any complex number z*

$$\|e^z - 1\| \leq \|z\| \, e^{\|z\|}$$

6. Law of the Iterated Logarithm

The Law of the Iterated Logarithm, abbreviated *LIL* is one of the classical results in Probability Theory. As usual, it was discovered for a sequence of *iid* real-valued random variables. From a quick tour of the introduction on the question in Loève (1997), in Gutt (2005) and in Feller (1968b), we may say that the *LIL* goes back to Kintchine, and to Levy in the binary case and finally to Kolmogorov and to Cantelli in the general case for independent random variables. Other important contributors in the stationary case are Hartman & Wintner, and Strassen. Here, we present the Kolmogorov Theorem as cited by Loève (1997).

Throughout this section, the iterated logarithm function $\log(\log(x))$, $x > e$, is denoted by $\log_2(x)$.

Let us ive the statement of *LIL* law, by using the notation introduced above. A sequence of square integrable and centered real-valued random variables $(X_n)_{n\geq 0}$ defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ satisfies the *LIL* if we have

$$\limsup_{n\rightarrow+\infty} \frac{S_n}{\sqrt{2s_n log_2 s_n^2}} = 1, \ a.s.$$

If the $(-X_n)_{n\geq 1}$ also satisfies the *LIL*, we also have

$$\liminf_{n\rightarrow+\infty} \frac{S_n}{\sqrt{2s_n log_2 s_n^2}} = -1, \ a.s.$$

The two conditions which required in the independent scheme to have the *LIL* are :

$$s_n \rightarrow +\infty \ as \ n \rightarrow +\infty, \ \ (C1).$$

and

$$|X_n/s_n| = o((log_2 s_n^2)^{-1}), \ as \ n \rightarrow +\infty. \ (C2)$$

The conditions (C2) is used to ensure the following one :

(C3) : For all $\mathbb{R}_+ \setminus \{0\} \ni c > 1$, there exists a sub-sequence $(s_{n_k})_{k\geq 1}$ of $(s_n)_{n\geq 1}$ such that

$$s_{n_{k+1}}/s_{n_k} \sim c \text{ as } k \to +\infty,$$

which is ensured if $s_{n_k} \sim \beta c^k$, where $\beta > 0$ is a real constant.

But it is important that the proof below is based only (C1) and (C3). In the *iid*, we have $s_n = \sigma\sqrt{n}$, $n \geq 1$. For any $c > 0$, we may take $n_k = \sigma[c^k]$, $k \geq 1$ to have (C2).

Let us state the Kolmogorov Theorem.

THEOREM 22. *Let $(X_n)_{n\geq 0}$ be a sequence of square integrable and centered real-valued random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that Condition (C1) and (C3) hold. Then the sequence satisfies the LIL, that is*

$$\limsup_{n\to+\infty} \frac{S_n}{\sqrt{2s_n log_2 s_n^2}} = 1, \ a.s.$$

and, by replacing X_n by $-X_n$ (which replacement does not change the variances), we have

$$\liminf_{n\to+\infty} \frac{S_n}{\sqrt{2s_n log s_n}} = -1, \ a.s.$$

Proof. Let $\delta > 0$. By applying (C3), let $(s_{n_k})_{k\geq 1}$ be a sub-sequence of $(s_n)_{n\geq 1}$ such that $s_{n_k} \sim c^k$, as $k \to +\infty$, with $1 < c < 1+\delta$ so that

$$2\delta' = 1 - \frac{1+\delta}{c} > 0 \ \ (S1)$$

and

$$(s_{n_k}(2\log_2 s_{n_k})^{1/2})/(s_{n_{k-1}}(2\log_2 s_{n_{k-1}})^{1/2}) \to c. \ \ (S2)$$

Now, since the the following class of integers intervals

$$\{[1, n_1[, [n_{k-1}, n_k[, k \geq 1\}$$

is a partition of $\mathbb{N} \setminus \{0\}$, we have for all each $n \geq 1$, there exists a unique $k \geq 1$ such that $n \geq [n_{k-1}, n_k[$ and so

$$S_n > (1+\delta)s_n(2\log_2 s_n)^{1/2},$$

implies that

$$S^*_{n_k} = \sup_{\ell \leq n_k} S_\ell \geq S_n > (1{+}\delta)s_n(2\log_2 s_n^2)^{1/2} \geq (1{+}\delta)s_{n_{k-1}}(2\log_2 s_{n_{k-1}})^{1/2},$$

which by (S1) and (S2), implies, for large values of k, that

$$S^*_{n_k} > (1+\delta'_1)^2 s_{n_k}(2\log_2 s_{n_k})^{1/2}.$$

When put together, these formulas above prove that

$$\left(S_n > (1{+}\delta)s_n(2\log_2 s_n)^{1/2},\ i.o.\right) \subset \left(S^*_{n_k} > (1{+}\delta')s_{n_k}(2\log_2 s^2_{n_k})^{1/2},\ i.o.\right)$$

Hence by Inequality (12) in Chapter 6 (See page 202), we have

$$\begin{aligned}\mathbb{P}\biggl(S^*_{n_k} \quad &> \quad (1+\delta'_2)s_{n_k}(2\log_2 s_{n_k})^{1/2}\biggr) \\ &\leq 2 \quad \mathbb{P}\left(S^*_{n_k} > \left(1+\delta' - \frac{\sqrt{2}}{(2\log_2 s_{n_{k-1}})^{1/2}}\right) s_{n_k}(2\log_2 s^2_{n_k})^{1/2}\right).\end{aligned}$$

So, for any $0 < \delta'' < \delta'$, we have for large values of k,

$$\mathbb{P}\left(S^*_{n_k} > (1{+}\delta')(2\log_2 s^2_{n_k})^{1/2}\right) \leq 2\mathbb{P}\left(S^*_n > (1{+}\delta'')s_{n_k}(2\log_2 s^2_{n_{k-1}})^{1/2}\right).$$

At this step, let us apply the exponential inequality, Statement (i) in Theorem 13, to have, with $\varepsilon_{n_k} = (1+\delta'')(2\log_2 s_{n_k})^{1/2}$ and c_{n_k}. Since, by assumption, $c_{n_k}\varepsilon_{n_k} \to 0$ and for k large to ensure based on $c_{n_k}\varepsilon_{n_k} < 1$, the last formula yields

$$\begin{aligned}
\mathbb{P}\biggl(S^*_{n_k}/s_{n_k} > \varepsilon_{n_k}) &< \exp(-\frac{\varepsilon^2_{n_k}}{2}(1+\varepsilon_{n_k}c_{n_k}/2)(2\log_2 s_{n_{k-1}})^{1/2}\biggr)\\
&\leq \exp\biggl(-\frac{\varepsilon^2_{n_k}}{2}\biggr)\\
&= \exp\biggl(-\frac{(1+\delta'')^2(2\log_2 s_{n_k}}{2})\biggr)\\
&\leq \exp\biggl(-(1+\delta'')\log_2 s^2_{n_k}\biggr)\\
&= \frac{1}{(1+\delta)\log s^2_{n_k}} \sim \frac{1}{\biggl(2k\log(1+\delta'')\biggr)^{(1+\delta'')}}.
\end{aligned}$$

Since the last term in the group of formulas above is the general term of a converging series, we also see that the series of general term

$$\mathbb{P}\biggl(S_n > (1+\delta)s_n(2\log_2 s^2_n)^{1/2},\ i.o.\biggr)$$

also converges. By Point (i) of Borel Cantelli's Lemma 4, we have

$$\mathbb{P}\biggl(S^*_{n_k} > (1+\delta'')s_{n_k}(2\log_2 s^2_{n_k})^{1/2},\ i.o\ (in\ k)\ \biggr) = 0,$$

and by the bounds and inclusions that are proved above, we have

$$\mathbb{P}\biggl(S_n > (1+\delta)s_n(2\log_2 s^2_n)^{1/2},\ i.o.\biggr) = 0,$$

that is, for any arbitrary $\delta > 0$, we have

$$\limsup_{n\to+\infty} \frac{S_n}{\sqrt{2s^2_n log_2 s^2_n}} \leq 1+\delta,\ a.s.$$

which proves that

$$\limsup_{n\to+\infty} \frac{S_n}{\sqrt{2s_n log_2 s^2_n}} \leq 1.\ a.s.$$

To prove that this superior limit is also greater than one, we just remark that the first part of the proof applied to the $-X_n$'s with $\delta = 1$ leads to

$$\mathbb{P}(\{\omega \in \Omega,\ \exists N(\omega),\ \forall n \geq N,\ -S_n \leq 2(2s_n^2 log_2 s_n^2)^{1/2})\}) = 1$$

Set

$$\Omega_1^* = \{\omega \in \Omega,\ \exists N(\omega),\ \forall n \geq N,\ S_n \leq 2(2s_n^2 log_2 s_n^2)^{1/2})\}.$$

Now, let $\delta > 0$ and let $c > 1$ such that

$$\frac{\delta'}{2} = 1 - \left((1-\delta)\left(1 - \frac{1}{c^2}\right)^{1/2} - \frac{2}{c}\right) > 0$$

so that

$$\left((1-\delta)\left(1 - \frac{1}{c^2}\right)^{1/2} - \frac{2}{c}\right) > 1 - \delta'.\ (K1)$$

Now we select, By Formula (C3), a sub-sequence $(s_{n_k})_{k\geq 1}$ of $(s_n)_{n\geq 1}$ such that $s_{n_k} \sim c^k$, as $k \to +\infty$. Next, we wish to apply Theorem 13 to the non-overlapping spacings $Y_{n_k} = S_{n_k} - S_{n_{k-1}}$'s of variances $u_{n_k} = s_n^2 - s_{n-1}^2$. We immediately check that

$$u_{n_k} = s_n^2 - s_{n-1}^2 \sim \left(1 - \frac{1}{c^2}\right)\ and\ v_{n_k}^2 = (2\log_2 u_{n_k}^2) \sim (2\log_2 u_k^2).\ (K2)$$

We have to remark that we still have, as $k + \infty$,

$$\begin{aligned} b_{n_k} &= \frac{1}{u_{n_k}} \max_{n_{k-1} < n \leq n_k} |X_{n_k}| \\ &= \max_{n_{k-1} < n < n_k} |Y_{n_k}| \\ &\leq \frac{1}{u_{n_k}} \max_{1 \leq n_k} |X_{n_k}| \\ &= O\Big((2\log_2 s_{n_k}^2)^{-1/2})\Big) \to 0. \end{aligned}$$

because of the first part of Formula (K2) above. Now, we are in the position to re-conduct the same method to the Y_{n_k}. So for $0 < \delta$,

$\varepsilon_{n_k} = (1-\delta)v_{n_k} \to +\infty$. We take $\gamma = (1-\delta)^{-1}$. Applying Theorem 13 with that value of γ for large values of k so that we have b_{n_k} is small enough and $(1-\delta)(2\log_2 u_{n_k}^2)^{1/2}$ is large enough, leads to

$$\begin{aligned}
\mathbb{P}\biggl(Y_{n_k}/u_{n_k} > (1+\delta)(2\log_2 u_{n_k}^2)^{1/2}\biggr) &> \exp\left(-\frac{\varepsilon_{n_k}^2}{2}\right)\\
&> \exp\left((1-\delta)\log_2(u_{n_k}^2)\right)\\
&= \frac{1}{(2k\log u_{n_k})^(1-\delta)}\\
&\sim \frac{1}{\biggl(2k\log(1+\delta)\biggr)^{(1-\delta)}}.
\end{aligned}$$

Since the series of general term

$$\mathbb{P}\biggl(Y_{n_k}/u_{n_k} > (1-\delta_1)(2\log_2 u_{n_k}^2)^{1/2}\biggr)$$

diverges and the Y_{n_k} are independent, we have by Point (i) of Borel-Cantelli Lemma 4.

$$\mathbb{P}\biggl(Y_{n_k}/u_{n_k} > (1-\delta)(2\log_2 u_{n_k}^2)^{1/2},\ i.o.\biggr) = 1.$$

Set

$$\Omega_1^* = (Y_{n_k}/u_{n_k} > (1-\delta)(2\log_2 u_{n_k}^2)^{1/2},\ i.o.\biggr).$$

On $\Omega^* = \Omega_1^* \cap \Omega_2^*$, we have

$$\begin{aligned}
S_{n_k} &= S_{n_k} - S_{n_{k-1}} + S_{n_{k-1}} > (1-\delta)u_k(2\log_2 u_k^2)^{1/2} + S_{n_{k-1}},\ i.o\\
&\Rightarrow (S_{n_k} > (1-\delta)s_{n_k}(2\log_2 s_{s_k}^2)^{1/2} - 2s_{n_{k-1}}(2\log_2 s_{n_{k-1}}^2)^{1/2}),\ i.o\\
&\Rightarrow (S_{n_k} > (1-\delta)(s_{n_k}(2\log_2 s_{n_k}^2)^{1/2})\\
&\times \left(\frac{u_k(2\log_2 u_k^2)^{1/2}}{s_{n_k}(2\log_2 s_{n_k}^2)^{1/2})} - \frac{s_{n_{k-1}}(2\log_2 s_{n_{k-1}}^2)^{1/2})}{s_{n_k}(2\log_2 s_{n_k}^2)^{1/2}}\right),\ i.o.\quad (L45)
\end{aligned}$$

But, as $k \to +\infty$

$$\left(\frac{u_k(2\log_2 u_k^2)^{1/2}}{s_{n_k}(2\log_2 s_{n_k}^2)^{1/2})} - \frac{2s_{n_{k-1}}(2\log_2 s_{n_{k-1}}^2)^{1/2})}{s_{n_k}(2\log_2 s_{n_k}^2)^{1/2}}\right)$$

converges to

$$\left((1-\delta)\left(1-\frac{1}{c^2}\right)^{1/2} - \frac{2}{c}\right) > 1-\delta'.$$

We conclude that Line (L41) above that

$$\Omega^* \subset (S_{n_k} > (1-\delta)s_{n_k}(2\log_2 s_{n_k}^2)^{1/2},\ i.o).$$

Since $\mathbb{P}(\Omega^*) = 1$, we get that for any $\delta > 0$, we have

$$\mathbb{P}((S_{n_k} > (1-\delta)s_{n_k}(2\log_2 s_{n_k}^2)^{1/2},\ i.o) = 1.$$

The proof of the theorem is now complete. □

CHAPTER 8

Conditional Expectation

1. Introduction and definition

We already saw in Chapter 7 the key role played independence in Probability Theory. But a very great part, even the greatest part, among studies in Probability Theory rely on some kind on dependence rather that on independence. However, the notion of independence, in most situations, is used as a theoretical modeling tool or as an approximation method. Actually, many methods which are used to handle dependence are based transformation of independent objects or based on some *nearness* measure from the independence frame. So, the better one masters methods based on independence, the better one understands methods for dependence studies.

However there is a universal tool to directly handle dependence, precisely the *Conditional Mathematical Expectation* tool. This chapter which is devoted to it, is the door for the study of arbitrary sequences or family of random objects.

The most general way to deal and to introduce to this tool relies on the Radon-Nikodym Theorem as stated in Doc 08-01 in Chapter 9 in Lo (2017b). We already spoke a little on it in the first lines in Chapter 6.

Definition 9. *Let X be real-valued random variable $(\Omega, \mathcal{A}, \mathbb{P})$ which is quasi-integrable, that is $\int X^+ d\mathbb{P} < \infty$ for example. Let $\mathcal{B}$ be a σ-sub-algèbra of $\mathcal{A}$, meaning that $\mathcal{B}$ is a σ-algebra of subsets of Ω and $\mathcal{B} \subset \mathcal{A}$. The mapping*

$$\begin{array}{rccc} \phi : & \mathcal{B} & \longrightarrow & \overline{\mathbb{R}} \\ & A & \hookrightarrow & \phi(B) = \int_B X d\mathbb{P} \end{array}$$

is well-defined and continuous with respect to $\mathbb{P}$ in the following sense

$$(\forall B \in \mathcal{B}),\ (\mathbb{P}(B) = 0 \Longrightarrow \phi(B) = 0).$$

By Radon-Nikodym's Theorem (See Doc 08-01 in Chapter 9 in Lo (2017b)), there exists a random variable, uniquely defined a.s.,

$$Z \; : \; (\Omega, \mathcal{B}) \mapsto \overline{\mathbb{R}}$$

which is $\mathcal{B}$ -measurable such that

$$(\forall B \in \mathcal{B}), \left(\int_B X \; d\mathbb{P} = \int_B Z \; d\mathbb{P} \right). \tag{1.1}$$

This random variable Z, is defined as the mathematical expectation of X with respect to $\mathcal{B}$ and denoted by by

$$Z = \mathbb{E}(X/\mathcal{B}) = \mathbb{E}^{\mathcal{B}}(X) \; a.s.,$$

** and the mathematical expectation is uniquely $\mathcal{B}$-almost surely.*

Let Y be a measurable mapping $Y(\Omega, \mathcal{A}) \to (G, \mathcal{D})$, where $(G, \mathcal{D})$ is an arbitrary measurable space. As previously explained in the first lines in Section 2 in Chapter 6;

$$\mathcal{B}_Y = \{Y^{-1}(H), H \in \mathcal{D}\}$$

is the σ-algebra generated by Y, the smallest one rendering Y measurable. It is a σ-sub-algebra of $\mathcal{A}$. The mathematical expectation with respect to Y, denoted by $\mathbb{E}(X/Y)$, is the mathematical expectation with respect to $\mathcal{B}_Y$ that is

$$\mathbb{E}(X/Y) = \mathbb{E}(X/\mathcal{B}_Y).$$

Extension. Later we will define the mathematical expectation with respect to a family a measurable mappings similarly to the one with respect to one mapping Y as in the definition.

This definition is one of the most general ones. In the special case where we work with square integrable real-valued random variables, a specific definition based on the orthogonal projection on the closed linear space $H = L^2(\Omega, \mathcal{B})$ of $\mathcal{B}$-measurable random variables is possible. And the mathematical expectation of X is its orthogonal projection on H. We will see this in Section 1. But as we will see it, even in the general case, the mathematical expectation is still a linear projection L^1 as explained in Remark 1 below.

2. The operator of the mathematical expectation

We already knew in Doc 08-01 in Chapter 9 in Lo (2017b), that $\mathbb{E}(X/\mathcal{B})$ is *a.s.* finite (in the frame of Probability Theory) and is integrable whenever X is. In stating the properties below, we fix $\mathbb{B}$ and we do not need write the mention of *with respect to* $\mathcal{B}$.

We have

PROPOSITION 26. *Considered as an operator from* $L^1(\Omega, A, \mathbb{P})$ *to* $L^1(\Omega, \mathcal{B}, \mathbb{P})$*, the mathematical expectation mapping*

$$\begin{array}{cccc} L & L^1(\Omega, A, \mathbb{P}) & \longrightarrow & L^1(\Omega, \mathcal{B}, \mathbb{P}) \\ & X & \hookrightarrow & \mathbb{E}(X/\mathcal{B}) \end{array}$$

is linear and satisfies

$$L^2 = L \text{ and } \|L\| = 1.$$

It is non-negative in the following sense

$$X \geq 0 \Longrightarrow L(X) \geq 0.$$

The operator L *is non-decreasing in the sense that, for* $(X, Y) \in (L^1)^2$,

$$X \leq Y \ a.s. \ \Rightarrow L(X) \leq L(Y), \ a.s.$$

Proof. Let X and Y be two integrable random variables, defined both on $(\Omega, \mathcal{A})$, α and β two real numbers. Then $\alpha X + \beta Y$ and $\alpha\mathbb{E}(X/\mathcal{B}) + \beta\mathbb{E}(Y/\mathcal{B})$ are *a.s.* defined. Then for any $B \in \mathcal{B}$, we have

$$\int_B (\alpha X + \beta Y) d\mathbb{P} = \alpha \int_B X d\mathbb{P} + \beta \int_B Y d\mathbb{P},$$

which, by the the definition of the mathematical expectation, implies

$$\alpha \int_B \mathbb{E}(X/\mathcal{B}) d\mathbb{P} + \beta \int_B \mathbb{E}(Y/\mathcal{B}) d\mathbb{P} = \int_B \{\alpha\mathbb{E}(X/\mathcal{B}) + \beta\mathbb{E}(Y/\mathcal{B})\} d\mathbb{P}.$$

Since $\alpha\mathbb{E}(X/\mathcal{B}) + \beta\mathbb{E}(Y/\mathcal{B})$ is $\mathcal{B}$-measurable, the equality entails that

$$\mathbb{E}((\alpha X + \beta Y)/\mathcal{B}) = \alpha\mathbb{E}(X/\mathcal{B}) + \mathbb{E}(Y/\mathcal{B}), \ a.s.$$

Next, as an immediate consequence of the definition, $\mathbb{E}(X/\mathcal{B}) = X$ *a.s.* whenever X is $\mathcal{B}$-measurable. Since $L(X) = \mathbb{E}(X/\mathcal{B})$ is $\mathcal{B}$-measurable, it comes that

$$L^2(X) = \mathbb{E}(L(X)/\mathcal{B}) = L(X),$$

which implies that $L^2 = L$. The non-negativity comes from Radon-Nikodym's Theorem which says the if X is non-negative, the mapping

$$\mathcal{B} \ni B \longmapsto \phi(B) = \int_B X d\mathbb{P} = \int_B Z \ d\mathbb{P}$$

is non-negative. Hence the its Radon-Nikodym derivative Z is non-negative. Here is an easy proof. Indeed $B_a = (Z < a) \in \mathcal{B}$, $a < 0$. Fix $a < 0$. If B_a is not a null-set, we would have

$$\phi(B) = \int_{B_a} X \ dP = \int_{B_a} Z \ dP \leq a\mathbb{P}(B_a) < 0.$$

This is impossible since we have $\phi(B_a) \geq 0$. So for $k \geq 1$, all the events $B_{-1/k}$ are null-sets. Since

$$(Z \leq 0) = \bigcap_{k \geq 1} B_{-1/k},$$

$(Z \leq 0)$ is a null-event and thus, $Z \geq 0$ *a.s.*.

The non-decreasingness is immediate from the combination of the linearity and the non-negativity.

Let us determine the norm of L defined by

$$\|L\| = \sup\{\|L(X)\|_1 / \|X\|_1, \ X \in L^1, \ \|X\|_1 \neq 0\}.$$

First let us show that L is contracting, that is

$$|\mathbb{E}(X/\mathcal{B})| \leq \mathbb{E}(|X|/\mathcal{B}).$$

We have $X \leq |X|$ and $-X \leq |X|$. We have $\mathbb{E}(X/\mathcal{B}) \leq \mathbb{E}(|X|/\mathcal{B})$ and $-\mathbb{E}(X/\mathcal{B}) = \mathbb{E}(-X/\mathcal{B}) \leq \mathbb{E}(|X|/\mathcal{B}) \geq 0$.

We conclude that

$$|\mathbb{E}(X/\mathcal{B})| \leq \mathbb{E}(|X|/\mathcal{B}).$$

By applying the definition of the mathematical expectation, we have

$$\|\mathbb{E}(X/\mathcal{B})\|_1 = \mathbb{E}|\mathbb{E}(X/\mathcal{B})| \leq \mathbb{E}\mathbb{E}(|X|/\mathcal{B}) = \mathbb{E}(|X|) = \|X\|_1 .$$

Hence we have

$$\|L(X)\|_1 \leq \|X\|_1 .$$

Next we have

$$\|L\| \leq 1.$$

But if $\|X\|_1 \neq 0$ and X is $\mathcal{B}$-measurable, we have $L(X) = X$ and for such random variables, we have$\|L(X)\|_1 / \|X\|_1 = 1$. We conclude that

$$\|L\| = 1.$$

The proof of the proposition is complete. ■

Remark 1. *A linear operator L such that $L^2 = L$ is called a projection. Hence the mathematical operator is a projection of L^1 to the sub-space of $\mathcal{B}$-measurable functions.*

3. Other Important Properties

Proposition 27. *We have the following facts.*

(1) If X is $\mathcal{B}$-measurable, then $\mathbb{E}(X/\mathcal{B}) = X$ a.s.

(2) The mathematical conditional expectation is anon-negative linear and non-decreasing operator.

(3) The mathematical conditional expectation is a contracting operator, that is, whenever the expressions make sense,

$$|\mathbb{E}(X/\mathcal{B})| \leq \mathbb{E}(|X|/\mathcal{B}).$$

(4) Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two σ-sub-algebras of $\mathcal{A}$ with $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \mathcal{A}$. We have

$$\mathbb{E}(\mathbb{E}(X/\mathcal{B}_2)/\mathcal{B}_1) = \mathbb{E}(X/\mathcal{B}_1) \tag{3.1}$$

and

$$\mathbb{E}(\mathbb{E}(X/\mathcal{B}_1)/\mathcal{B}_2) = \mathbb{E}(X/\mathcal{B}_1). \tag{3.2}$$

(5) Let X be a random variable independent of $\mathcal{B}$ in the following sense : for any mapping $\mathcal{B}$-measurable mapping $Z : \Omega \longmapsto \mathbb{R}$ and for any measurable application $h : \mathbb{R} \longmapsto \mathbb{R}$,

$$\mathbb{E}(Z \times \ h(X)) = \mathbb{E}(Z) \times \ \mathbb{E}(h(X)).$$

Then, if $\mathbb{E}(X)$ exists, we have

$$\mathbb{E}(X/\mathcal{B}) = \mathbb{E}(X).$$

(6) (Monotone Convergence Theorem for Mathematical expectation). Let $(X_n)_{n\geq 0}$ be a non-decreasing sequence of integrable random variables which are all non-negative or all integrable. Then we have

$$\mathbb{E}(\lim_{n\to\infty} X_n/\mathcal{B}) = \lim_{n\to\infty} \mathbb{E}(X_n/\mathcal{B})$$

(7) (Fatou-Lebesgue Theorems). Let $(X_n)_{n\geq 0}$ be quasi-integrable real-valued random variables which is a.s. bounded below by an integrable random variable, then

$$\mathbb{E}(\liminf_{n\to\infty} X_n/\mathcal{B}) \leq \liminf_{n\to\infty} \mathbb{E}(X_n/\mathcal{B}).$$

If the sequence is a.s. bounded above by an integrable random variable, then

$$\mathbb{E}(\limsup_{n\to\infty} X_n/\mathcal{B}) \geq \limsup_{n\to\infty} \mathbb{E}(X_n/\mathcal{B}).$$

If the sequence is uniformly a.s. bounded by an integrable random variable Z and converges a.s. to X, then

$$\lim_{n\to\infty} \mathbb{E}(X_n/\mathcal{B}) = \mathbb{E}(X/\mathcal{B})$$

and

$$\mathbb{E}(|X|/\mathcal{B}) \leq \mathbb{E}(|Z|/\mathcal{B}).$$

(8) Let X be a quasi-integrable random variable. Let Z be $\mathcal{B}$-measurable and non-negative or integrable. Then we have

$$\mathbb{E}(ZX/\mathcal{B}) = Z \times \mathbb{E}(X/\mathcal{B}).$$

Proof.

Points from (1) to (3) are already proved in the first proposition.

Proof of Point (4). First we know that $\mathbb{E}(X/\mathcal{B}_1)$ is $\mathcal{B}_1$-measurable and thus $\mathcal{B}_2$-measurable. By Point (1), we have

$$\mathbb{E}(\mathbb{E}(X/\mathcal{B}_1)/\mathcal{B}_2) = \mathbb{E}(X/\mathcal{B}_1).$$

Formula (3.1) is proved. Next, for any $B \in \mathcal{B}_1 \subset \mathcal{B}_2$, we have

$$\int_B X d\mathbb{P} = \int_B \mathbb{E}(X/\mathcal{B}_2) d\mathbb{P},$$

since B is also in $\mathcal{B}_2$. Now we apply the definition of the mathematical expectation with respect to $\mathcal{B}_1$ in the right-hand member to have

$$\int_B X d\mathbb{P} = \int_B \mathbb{E}(X/\mathcal{B}_2) d\mathbb{P} = \int_B \mathbb{E}(\mathbb{E}(X/\mathcal{B}_2)/\mathcal{B}_1) d\mathbb{P}.$$

Since $\mathbb{E}(\mathbb{E}(X/\mathcal{B}_2)/\mathcal{B}_1)$ is $\mathcal{B}_1$-measurable, we conclude that $\mathbb{E}(X/\mathcal{B}_1) = \mathbb{E}(\mathbb{E}(X/\mathcal{B}_2)/\mathcal{B}_1)$ *a.s.* Thus, we reach Formula (3.2).

Proof of (5). It is clear that the constant mapping $\omega \longmapsto \mathbb{E}(X)$ is $\mathcal{B}$-measurable. Hence for any $B \in \mathcal{B}$,

$$\int_B X \ d\mathbb{P} = \mathbb{E}(1_B \times \ X) = \mathbb{E}(1_B) \times \mathbb{E}(X) = \mathbb{E}(X) \int_B d\mathbb{P} = \int_B \mathbb{E}(X) d\mathbb{P},$$

which proves that

$$\mathbb{E}(X/\mathcal{B}) = \mathbb{E}(X), \ a.s.$$

Proof of Point (6). Let $(X_n)_{n\geq 0}$ be a non-decreasing sequences of random variables such that $\mathbb{E}(X_n^+) < \infty$ pour tout $n \geq 0$. For any $B \in \mathcal{B}$, we have

$$\int_B X_n d\mathbb{P} = \int_B \mathbb{E}(X_n/\mathcal{B}) d\mathbb{P}.$$

Since the sequences $(X_n)_{n\geq 0}$ and $(\mathbb{E}(X_n/\mathcal{B}))_{n\geq 0}$ are non-decreasing of integrable random variable, we may apply the Monotone Convergence Theorem to get

$$\int_B \lim_{n\to\infty} X_n d\mathbb{P} = \lim_{n\to\infty} \int_B X_n d\mathbb{P} = \lim_{n\to\infty} \int_B \mathbb{E}(X_n/\mathcal{B}) d\mathbb{P} = \int_B \lim_{n\to\infty} \mathbb{E}(X_n/\mathcal{B}) d\mathbb{P}.$$

Since $\lim_{n\to\infty} \mathbb{E}(X_n/\mathcal{B})d\mathbb{P}$ is $\mathcal{B}$-measurable, we have

$$\mathbb{E}(\lim_{n\to\infty} X_n/\mathcal{B}) = \lim_{n\to\infty} \mathbb{E}(X_n/\mathcal{B}).$$

Proof of Point (7). Based on the Monotone convergence Theorem for conditional expectation in the previous Point (6), the Fatou-Lebesgue Theorem and the Lebesgue Dominated Theorem are proved as in the unconditional case, as done in Chapter 6 in Lo (2017b). ■

Proof of Point (8). Let Z be a $\mathcal{B}$-measurable random variable non-negative or integrable. Thus $Z\mathbb{E}(X/\mathcal{B})$ is $\mathcal{B}$-measurable. Suppose that $Z = 1_C$, $C \in \mathcal{B}$, that is Z is $\mathcal{B}$-measurable indicator function. We have for any $B \in \mathcal{B}$,

$$\begin{aligned} \int_B Z\ X\ d\mathbb{P} &= \int_B 1_C\ X\ d\mathbb{P} = \int_{BC} X\ d\mathbb{P} \\ &= \int_{BC} \mathbb{E}(X/\mathcal{B}) d\mathbb{P} = \int_B 1_C \mathbb{E}(X/\mathcal{B}) d\mathbb{P}. \end{aligned}$$

Thus, we have

$$\int_B Z\ X\ d\mathbb{P} = \int_B Z\ \mathbb{E}(X/\mathcal{B}) d\mathbb{P}.$$

Since $Z \times \mathbb{E}(X/\mathcal{B})$ is $\mathcal{B}-$measurable, we get

$$\mathbb{E}(ZX/\mathcal{B}) = Z\mathbb{E}(X/\mathcal{B}).$$

To finish the proof, we follow the famous three steps method by extending the last formula to elementary functions based on $\mathcal{B}$-measurable sets, next to non-negative random variables using Point (6) above and finally to an arbitrary random variable Z using the additivity of both the expectation and the conditional expectation.

* We may and do have the same theory by using non-negative $\mathcal{B}$-measurable random variables in place of the elements of $\mathcal{B}$ in the definition of the mathematical expectation.

4. Generalization of the definition

Let us define by $\mathbb{E}(X/\mathcal{B})$ a $\mathcal{B}$-measurable random variable such that for all $\mathcal{B}$-measurable mapping $h : (\Omega, \mathcal{B}) \longmapsto (\mathbb{R}, \mathcal{B})$, we have

$$\int h\ X\ d\mathbb{P} = \int h\mathbb{E}(X/\mathcal{B})\ d\mathbb{P} \tag{4.1}$$

We are going to quickly show that the definitions based on Formulas (4.1) and (1.1) respectively are the same.

Before we do it, let us just say that Formula (4.1) usually offers a more comfortable handling of the mathematical expectation.

Proof of the equivalence between Formulas (4.1) and (1.1). The implication (4.1) $\Longrightarrow$ (1.1) by taking $h = 1_B$ for $B \in \mathcal{B}$. To prove the converse implication, we use the classical three steps methods. Suppose that Formula (1.1) holds.

Step 1. If $h = 1_B$ for $B \in \mathcal{B}$, Formula (4.1) is obvious.

Step 2. h is an elementary function of the form

$$h = \sum_{i=1}^{p} \alpha_i 1_{B_i},$$

where $B_i \in \mathcal{B}$ et $\alpha_i \in \mathbb{R}$. By using the linearity, we have

$$\begin{aligned}
\int h\ X\ d\mathbb{P} &= \int \left(\sum_{i=1}^{p} \alpha_i 1_{B_i}\right) X\ d\mathbb{P} \\
&= \sum_{i=1}^{p} \alpha_i \left(\int 1_{B_i} X\ d\mathbb{P}\right) \\
&= \sum_{i=1}^{p} \alpha_i \left(\int 1_{B_i} \mathbb{E}(X/\mathcal{B})\ d\mathbb{P}\right) \\
&= \int \left(\sum_{i=1}^{p} \alpha_i 1_{B_i}\right) \mathbb{E}(X/\mathcal{B}) d\mathbb{P} = \int h\mathbb{E}(X/\mathcal{B})\ d\mathbb{P}.
\end{aligned}$$

Step 3. h is $\mathcal{B}$-measurable and non-negative There exists a sequence of elementary $(h_n)_{n\geq 0}$ based on elements of $\mathcal{B}$ such that $h_n \nearrow h$ and thus,

$$h_n X^+ \nearrow hX^+ \text{ and } h_n\mathbb{E}(X^+/\mathcal{B}) \nearrow h\mathbb{E}(X^+/\mathcal{B}).$$

By the monotone convergence Theorem, we have

$$\int h_n X^+ d\mathbb{P} = \int h_n \mathbb{E}(X^+/\mathcal{B}) d\mathbb{P} \nearrow \int h X^+ d\mathbb{P} = \int h \mathbb{E}(X^+/\mathcal{B}) d\mathbb{P}.$$

We similarly get

$$\int h\ X^-\ d\mathbb{P} = \int h \mathbb{E}(X^-/\mathcal{B}) d\mathbb{P}.$$

Thus by quasi-integrability, we have

$$\begin{aligned}\int h\ X\ d\mathbb{P} &= \int h\mathbb{E}(X^+/\mathcal{B})d\mathbb{P} - \int h\ X^-\ d\mathbb{P} \\ &= \int h\mathbb{E}(X^+/\mathcal{B})d\mathbb{P} - \int h\ \mathbb{E}(X^-/\mathcal{B})d\mathbb{P} = \int h\mathbb{E}(X/\mathcal{B})\ d\mathbb{P}.\end{aligned}$$

The proof is over. □

With the second definition, some properties are easier to prove as the following one.

Proposition 28. *if Z is $\mathcal{B}$-measurable either non-negative or integrable, we have for any quasi-integrable random variable,*

$$\mathbb{E}(Z \times\ X/\mathcal{B}) = Z \times\ \mathbb{E}(X/\mathcal{B}).$$

Proof. Let h be any non-negative and real valued $\mathcal{B}$)-measurable function. We have

$$\int h\{ZX\} d\mathbb{P} = \int \{hZ\} X d\mathbb{P} = \int \{hZ\} \mathbb{E}(X/\mathcal{B}) d\mathbb{P} = \int h \left\{Z\mathbb{E}(X/\mathcal{B})\right\} d\mathbb{P}.$$

Since $Z\mathbb{E}(X/\mathcal{B})$ is $\mathcal{B}$-measurable, we get that $\mathbb{E}(ZX/\mathcal{B}) = Z\mathbb{E}(X/\mathcal{B})$ *a.s.* □

5. Mathematical expectation with respect to a random variable

Let us consider that $\mathcal{B}$ is generated by a measurable mapping $Y\ : (\Omega, , \mathcal{A}) \to (G, \mathcal{D})$, where $(G, \mathcal{D})$ is an arbitrary leasure space, that is

$$\mathcal{B} = \mathcal{B}_Y = \{Y^{-1}(H), H \in \mathcal{D}\}$$

Par definition, we denote

$$\mathbb{E}(X/\mathcal{B}_Y) = \mathbb{E}(X/Y).$$

Let us prove that any real valued and $\mathcal{B}_Y$-measurable mapping h is of the form $g(Y)$, where g is a measurable mapping defined on $(G, \mathcal{D})$ and takes its values in $\overline{\mathbb{R}}$.

To see this, let us use again the four steps method. In the first step, let us suppose is an indicator function of an element of $\mathcal{B}_Y$. So there exists $C \in \mathcal{D}$, such that

$$h = 1_{Y^{-1}(C)} = 1_C(Y)$$

Clearly $g = 1_C$ is a real-valued measurable mapping defined on G such that $h = g(Y)$.

In a second step, let h be a of the form

$$h = \sum_{i=1}^{p} \alpha_i 1_{Y^{-1}}(B_i) = \sum_{i=1}^{p} \alpha_i 1_{B_i}(Y) = \left(\sum_{i=1}^{p} \alpha_i 1_{B_i}\right)(Y) = g(Y),$$

where g is clearly $\mathcal{D}$-measurable. We easily move to non-negative $\mathcal{B}_Y$ -measurable function by Point (6) of Proposition ... and the classical fact that any measurable and non-negative function is a non-decreasing limit of a sequence of non-negative elementary function. The extension to an arbitrary quasi-integrable $\mathcal{B}$-measurable function is done by using the positive and negative parts.

In summary, whenever it exists, $\mathbb{E}(X/Y)$ is has the form

$$\mathbb{E}(X/Y) = g(Y),$$

g is a measurable mapping defined on $(G, \mathcal{D})$ and takes its values in $\overline{\mathbb{R}}$.

This function g is also called the **regression function** of X in Y denoted as

$$E(X/Y = y) = g(y).$$

It is very interesting to see a discrete version of that formula, which is very commonly used. Suppose that Y takes a countable number of values denoted by $(y_j)_{j \in J}$, $j \subset \mathbb{N}$. We recall that we have

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X/Y)),$$

and next by using the regression function, we have

$$\begin{aligned} \mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X/Y)) &= \mathbb{E}g(Y) \\ &= \sum_{j \in J} g(y_j)\mathbb{P}(Y = y_j) = \sum_{j \in J} \mathbb{E}(X/Y = y_j)\mathbb{P}(Y = y_j). \end{aligned} \tag{5.1}$$

This gives

$$\mathbb{E}(X) = \sum_{j \in J} \mathbb{E}(X/Y = y_j)\mathbb{P}(Y = y_j).$$

If X itself is discrete and takes the values $(x_i)_{i \in I}$, we have

$$\mathbb{E}(X) = \sum_{j \in J} x_i \mathbb{P}(X = x_i/Y = y_j)\mathbb{P}(Y = y_j). \tag{5.2}$$

Let us study the Jensen's inequality for the mathematical expectation. We keep the same notations. For some details on convex function in our series, on may consult Exercise 6 and its solution in Doc 03.09 in Chapter 4 in Lo (2017b).

6. Jensen's Inequality for Mathematical Expectation

THEOREM 23. *Let X be random variable supported by an interval I on which is defined a real-valued convex function ϕ. Suppose that X and $\phi(X)$ are integrable. Then for any σ-sub-algebra $\mathcal{B}$ of $\mathcal{A}$, we have*

$$\phi(\mathbb{E}(X/\mathcal{B}) \leq \mathbb{E}(\phi(X)/\mathcal{B}).$$

Proof. Let us follows Chung (1974) in the first proof therein. Let us proceed by step.

Step 1. Let us suppose that X takes a finite number of distinct values (x_j), $j \in J$, J finite. Let us denote $B_j = (X = x_j)$, $j \in J$ so that

$$X = \sum_{j \in J} x_j 1_{B_j}, \ \textit{and} \ \sum_{j \in J} 1_{B_j} = 1_\Omega = 1. \ (F1)$$

and hence

$$\phi(X) = \sum_{j \in J} \phi(x_j) 1_{B_j}.$$

By the linearity of the mathematical expectation, we have

$$\mathbb{E}(\phi(X)/\mathcal{B}) = \sum_{j\in J} \phi(x_j)\mathbb{E}(1_{B_j}/\mathcal{B}). \quad (F2)$$

But the real numbers $\mathbb{E}(1_{B_j}/\mathcal{B})$ add up to one since, because of Formula (F1), we get

$$\sum_{j\in J} \mathbb{E}(1_{B_j}/\mathcal{B}) = \mathbb{E}\left(\sum_{j\in J} 1_{B_j}/\mathcal{B}\right) = \mathbb{E}(1_\Omega/\mathcal{B}) = 1.$$

Hence by the convexity of ϕ, the right-hand member of Formula (F2) satisfies

$$\sum_{j\in J} \phi(x_j)\mathbb{E}(1_{B_j}/\mathcal{B}) \geq \phi\left(\left(\sum_{j\in J} x_j \mathbb{E}(1_{B_j}\right)/\mathcal{B}\right),$$

and, surely, the right-hand member is

$$\phi\left(\mathbb{E}\left(\sum_{j\in J} x_j 1_{B_j}\right)/\mathcal{B}\right) = \phi(\mathbb{E}(X/\mathbb{B}). \quad (F3)$$

By comparing the left-hand term of Formula (F2) and the right-hand term of Formula (F3), we get the desired result.

Step 2. For a *a.s.* finite general random variable, we already know from Measure Theory and Integration that X is limit of a sequence elementary functions $(X_p)_{p\geq 1}$ with $|X_p| \leq |X|$ for all $p \geq 1$.

If X is bounded *a.s*, say $|X| \leq A < +\infty$ *.a.s*, then by the continuity ϕ, we have

$$\max\left(|\phi(X)|, \sup_{p\geq 1} |\phi(X_p)|\right) \leq \|\phi\|_{[-A,A]} < +\infty.$$

By applying the result of Step 1, we have for all $p \geq 1$

$$\phi(\mathbb{E}(X_p/\mathcal{B}) \leq \mathbb{E}(\phi(X_p)/\mathcal{B}).$$

By applying the Dominated Convergence Theorem in both sides, we get the desired result.

Step 3. Now suppose that X is not bounded above. By Proposition 17.6 in Choquet (1966), each point of $(a, \phi(a))$ of the Graph Γ of the convex function ϕ has a supporting line, that is a straight line which passes through $(a, \phi(a))$ and is below Γ. A quick drawing may help to catch the meaning of this. For each $n \geq 1$, consider a supporting line at the point $(n, \phi(n))$ with equation $f_n(x) = A_n x + B_n$.

If X is not bounded below, we consider, for each $n \geq 1$, a supporting line at the point $(-n, \phi(-n))$ with equation $g_(x) = C_n x + D_n$.

We may have X bounded below and not bounded above, X not bounded below and bounded above or X neither bounded below and nor bounded above. In all these situations, we will have similar way to handle the situation. Let us take the last case. We define

$$\phi_n = g_n 1_{]-\infty,-n[} + \phi 1_{[-n,n]} + f_n 1_{]n,+\infty[}$$

We may check quickly that each ϕ_n is convex, $\phi_n \leq \phi$ and $\phi_n \uparrow \phi$ as $n \uparrow +\infty$. By denoting, for each $n \geq 1$,

$$E_n = \|\phi\|_{[-n,n]}$$

and $a_n = |A_n| + |C_n| + |E_n|$ and $b_n = |B_n| + |D_n|$, we have for all $n \geq 1$, for all $x \in \mathbb{R}$

$$|\phi_n(x)| \leq a_n |x| + b_n,$$

and next for all $n \geq 1$, for all $p \geq 1$

$$|\phi_n(X_p)| \leq a_n |X_p| + b_n \leq a_n |X| + b_n, \ (F4)$$

Since for each $n \geq 1$, ϕ_n is convex, the result of Step 1 gives for all $p \geq 1$,

$$\phi_n(\mathbb{E}(X_p/\mathcal{B}) \leq \mathbb{E}(\phi_n(X_p)/\mathcal{B}).$$

By fixing $n \geq 1$, by letting $p \to +\infty$ and by applying the Dominated Convergence Theorem in both sides on the account of Formula (F4), we get

$$\phi_n(\mathbb{E}(X/\mathcal{B}) \leq \mathbb{E}(\phi_n(X)/\mathcal{B}).$$

By letting $n \uparrow +\infty$, and by applying the Monotone convergence Theorem of the integrable functions in the right-hand member, we get the general conclusion. ■

7. The Mathematical Expectation as an Orthogonal Projection in L^2

Let us suppose that $X \in E = L^2(\Omega, \mathcal{A}, \mathbb{P})$. For any σ-sub-algebra $\mathcal{B}$ of $\mathcal{A}$, let us consider $H = L^2(\Omega, \mathcal{B}, \mathbb{P})$ the square integrable and real-valued $\mathcal{B}$-measurable functions. At least $1 = 1_\Omega$ and $0 = 1_\emptyset$ are elements of H.

We already know that $L^2(\Omega, \mathcal{A}, \mathbb{P})$ is a Hilbert space endowed with inner product

$$L^2(\Omega, \mathcal{A}, \mathbb{P})^2 \ni (X, Y) \mapsto \langle X, Y \rangle = \mathbb{E}(XY).$$

We have the following projection theorem in Hilbert spaces (See for example Theorem 6.26 in Chidume (2014), page 109).

Proposition 29. *Suppose that E is a Hilbert space and H a closed sub-linear space. Fix $x \in E$. We have the following facts.*

(1) There exists a unique element $p_H(x) \in H$ such that

$$d(x, H) = \inf\{\|x - h\|,\ h \in H\} = \|x = p_H(x)\|.$$

(2) $p_H(x)$ is also the unique element of H such that $x - p_H(x)$ is orthogonal all elements of H.

We are going to apply it in order to characterize $\mathbb{E}(X/\mathcal{B})$. We have

Theorem 24. *For any σ-sub-algebra $\mathcal{B}$ of $\mathcal{A}$, $H(\mathcal{B}) = L^2(\Omega, \mathcal{B}, \mathbb{P})$ is a closed linear space, and for any $X \in E = L^2(\Omega, \mathcal{A}, \mathbb{P})$,*

$$\mathbb{E}(X/\mathcal{B}) = p_{H(\mathcal{B})}(X).\ a.s.$$

Proof. Let us begin to show that $H(\mathcal{B})$ is closed. Let Z a limit of a sequence of elements of $H(\mathcal{B})$. Since E is a Hilbert space, we know that Z is still in E, thus is square integrable. By Theorem 10 in Chapter 5, page 167, the concerned sequence converges to Z in probability and next, by the relation between weak and strong limits seen in the same chapter, a sub-sequence of the sequence converges *a.s.* to Z. Finally Z being an *a.s.* limit of a sequence $\mathcal{B}$-measurable functions is $\mathcal{B}$-measurable. In total $Z \in H(\mathcal{B})$. thus $H(\mathcal{B})$ is closed in E.

Now for any $X \in E$, Point (2) of Proposition 29 characterizes $Z = p_{H(\mathcal{B})}(X)$. Thus for any $B \in \mathcal{B}$, $h = 1_B \in H(\mathcal{B})$, $\langle X - Z, h\rangle = 0$, that is $\langle X, Z\rangle = \langle X, h\rangle$

$$\int_B X \; d\mathbb{P} = \int_B Z \; d\mathbb{P}.$$

We conclude that $Z = \mathbb{E}(X/\mathcal{B})$. □

8. Useful Techniques

In a great number of situations, we need to compute the mathematical expectation of a real-valued function $h(X)$ of X and we have to use a conditioning based on another random variables Y where of course X and Y are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$, even if they may have their values in different measure spaces $(E_1, \mathcal{G}_{E_1})$ and $(E_2, \mathcal{G}_{E_2})$. Suppose $h : (E_1, \mathcal{G}_{E_1}) \to \mathbb{R}$ is measurable and that h is non-negative or $h(X)$ is integrable. We already know that there exists a measurable function $g : (E_2, \mathcal{G}_{E_2}) \to \mathbb{R}$ such that

$$\mathbb{E}(h(X)/Y) = g(Y) \quad (CG01)$$

and by this, we have

$$\mathbb{E}(h(X)) = \mathbb{E}(g(Y)). \quad (CG02)$$

Now suppose that Y is continuous with respect to a measure ν that is given on the measure space $(E_2, \mathcal{G}_{E_2})$. By reminding that $g(y) = \mathbb{E}(h(X)/(Y = y))$ for $y \in E_2$, we have from formula (CG02)

$$\begin{aligned}\mathbb{E}(h(X)) &= \mathbb{E}(g(Y)) \\ &= \int_{E_2} g(y) \; d\mu(y) \\ &= \int_{E_2} \mathbb{E}(h(X)/(Y = y)) \; d\nu(y)\end{aligned}$$

which is, in short,

$$\mathbb{E}(h(X)) = \int_{E_2} g(y)\ d\nu(y). \quad (CG03)$$

This result takes the following particular forms.

(I) Conditioning by a discrete random variable.

Suppose that Y is discrete, that is the values set of Y is countable and is written as $D = \{y_j, \in J\}$, $J \subset \mathbb{N}$. Hence the probability law of Y is continuous with respect to the counting measure ν with support D. Thus by the Discrete Integration Formula (DIF1) (see page 66) applied to Formula (CG03), we get

$$\mathbb{E}(h(X)) = \sum_{j \in J} \mathbb{E}(h(X)/(Y = y_j))\mathbb{P}(X = j_j). \quad (CD)$$

(II) Conditioning by an absolutely continuous real random vector.

Suppose that $E_2 = \mathbb{R}^r$, $r \geq 1$. If Y is continuous with respect to the Lebesgue measure λ_r, we get the formula

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}^r} \mathbb{E}(h(X)/(Y = y))\ f_Y(y)\ d\lambda_r(y). \quad (CC01)$$

(III) Conditional probability density function.

On top of the assumptions in Part (II) above, let us suppose also that $E_2 = \mathbb{R}^s$, $s \geq 1$, with $d = r + s$. Let us suppose that $Z = (X^t, Y^t)^t$ has a *pdf* $f_Z \equiv f_{(X,Y)}$ with respect to the Lebesgue measure on $\mathbb{R}^d$. Thus the marginal *pdf*'s of X and Y are defined by

$$f_X(x) = \int_{\mathbb{R}^r} f_{(X,Y)}(x, y)\ d\lambda_r(y), \quad x \in \mathbb{R}^s$$

and

$$f_Y(y) = \int_{\mathbb{R}^s} f_{(X,Y)}(x, y)\ d\lambda_s(x), \quad y \in \mathbb{R}^r.$$

Let us define

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)}, \quad f_Y(y) > 0,$$

as the conditional *pdf* of X given $Y = y$. The justification of such a definition relies in the fact that replacing $\mathbb{E}(h(X)/(Y=y))$

$$\int_{\mathbb{R}^s} h(x) f_{X|Y=y}(x) \ d\lambda_s(x)$$

and using Tonelli's Theorem (when h in non-negative) or Fubini's Theorem (when $h(X)$ is integrable) leads to

$$\begin{aligned}
& \int_{\mathbb{R}^r} \left(\int_{\mathbb{R}^s} h(x) f_{X|Y=y}(x) \ d\lambda_s(x) \right) \ f_Y(y) \ d\lambda_t(y) \\
= & \int_{\mathbb{R}^r} \left(\int_{\mathbb{R}^s} h(x) \frac{f_{(X,Y)}(x,y)}{f_Y(y)} \ d\lambda_s(x) \right) \ f_Y(y) \ d\lambda_t(y) \\
= & \int_{\mathbb{R}^s} h(x) \left(\int_{\mathbb{R}^s} \frac{f_{(X,Y)}(x,y)}{f_Y(y)} \ d\lambda_t(y) \right) \ d\lambda_s(x) \\
= & \int_{\mathbb{R}^s} h(x) f_X(x) d\lambda_s(x) \\
= & \ \mathbb{E}(h(X)).
\end{aligned}$$

This leads to the frequent use the following formula

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}^r} \left(\int_{\mathbb{R}^s} h(x) f_{X|Y=y}(x) \ d\lambda_s(x) \right) \ f_Y(y) \ d\lambda_r(y). \quad (CC02)$$

CHAPTER 9

Probability Laws of family of Random Variables

1. Introduction

We have already studied the finite product measure in Chapter 8 in Lo (2017b) for σ-finite measures defined on arbitrary measure spaces. In this chapter we give the Theorem of Kolmogorov which establishes the arbitrary product of probability measures, but in special measure spaces.

This theorem of Kolmogorov is the foundation of the modern theory of probability. There is nothing above it, in term of probability laws. On this basis, the modern theory of random analysis, which extends Real Analysis (paths smoothness, differentiability, integration according to different types, etc) is built on.

We recommend the reader, especially the beginner, to read it as many as possible and to often and repeatedly come back to it in order to see its deepness and to understand its consequences.

Among special spaces on which the construction is made, we count Polish spaces. A Polish space is a complete and separable metric space (E, d) like $(\mathbb{R}^s, \rho)$, $s \geq 1$, where ρ is one of its three classical metrics. An interesting remark is that the finite product of Polish spaces is a Polish space. The finite Borel product σ of Polish spaces is generated by the product of open balls.

Here, the level of abstraction is moderately high. Once again, we recommend the beginner to go slow and to give himself the needed time to understand the definitions and the notation. This chapter may be considered as a continuation of Chapter 8 in Lo (2017b).

We already encountered this Kolmogorov construction in finite dimensions in Chapter 2 in pages 44 and 62. The results in this chapter will be the most general extension of this kind of result.

In the first section, we state and prove the existence of the product probability measure. Next, we will see how to state a number of particular forms involving Lebesgue-Stieljes measures.

2. Arbitrary Product Probability Measure

Let $(E_t, \mathcal{B}_t, \mathbb{P}_t)$, $t \in T \neq \emptyset$, be a family of probability spaces. We define the product space by

$$E = \prod_{t \in T} E_t.$$

If T is finite, even countable, we may use the classical notation : $T = \{t_j, j \geq 0\}$. It make senses to speak about the first factor E_{t_1}, the second E_{t_2}, etc. The elements of

$$E = \prod_{j \geq 0} E_{t_j},$$

may be denoted by $x = (x_{t_1}, x_{t_2}, ...)$ as an ordered set.

But, the index set T may arbitrary and uncountable. For example T be may a set of functions. If the functions are real-valued, T is uncountable and has a partial order. Sometimes we may not have an order at all. So the general appropriate way to study E seems to consider E as a set of functions. Thus, an element x of E, written as

$$x = (x_t)_{t \in T} = (x(t), t \in T),$$

is perceived as a function x which corresponds to each $t \in T$ a value $x(t) = x_t \in E_t$.

Let us begin to introduce the projections and give relevant notation. We denote by $\mathcal{P}_f$ the class of finite and non-empty subsets of T. Given an element $S = \{s_1, ..., s_k\}$ of $\mathcal{P}_f$, we may write in any order of the subscripts. This leads to the class of ordered and non-empty subsets denoted by $\mathcal{P}_{of}$. Elements of $\mathcal{P}_{of}$ are written as k-tuples $S = (s_1, ..., s_k)$, $k \geq 1$. For any $S = (s_1, ..., s_k) \in \mathcal{P}_{of}$, we have the finite product space

$$E_S = \prod_{j=1}^{k} E_{s_j}$$

which is endowed with the finite product σ-algebra

$$\mathcal{B}_S = \bigotimes_{j=1}^{k} \mathcal{B}_{s_j}.$$

The projection of this space E_S is defined by

$$\begin{array}{rccc} \Pi_S : & (E, \mathcal{B}) & \longrightarrow & (E_S, \mathcal{B}_S) \\ & x = (x_t)_{t \in T} & \longmapsto & \Pi_S(x) = (x_{s_1}, ..., x_{s_k}) \end{array} . \tag{2.1}$$

We name S as the index support of the projection.

Our first objective is to define a σ-algebra on E, which renders measurable all the projections on finite sub-products spaces.

2.1. The Product σ-algebra on the product space. Let us begin by the definition

DEFINITION 10. *The product σ-algebra on E, denoted by $\mathcal{B}$, is the smallest σ-algebra rendering measurable all the projections of finite index support.*

We already know that such a σ-algebra exists. Compared to the finite product σ-algebra, there is nothing new yet.

In the sequel, we have to change the order of elements of $V \in \mathcal{P}_{of}$. So the following recall may be useful. Indeed, by permuting the elements of $V = (v_1, ..., v_k) \in \mathcal{P}_{of}$, $k \geq 1$, by means of a permutation s of $\{1, 2, ..., n\}$, the correspondence

$$\begin{array}{ccc} (E, \mathcal{B}) & \longrightarrow & (E_{s(S)}, \mathcal{B}_{s(S)}) \\ (x_{v_1}, ..., x_{v_k}) & \longmapsto & (x_{s(v_1)}, ..., x_{s(v_k)}) \end{array} . \tag{2.2}$$

is a one-to-one mapping. Also, in parallel of the notation of E_V, we may and do adopt the following notation

$$x_V = (x_{s(v_1)}, ..., x_{s(v_k)}).$$

As well, the space $\mathcal{S}_V$ denote the class of measurable rectangles in E_V.

Now, as in the finite product case, we have to see how to generate $\mathcal{B}$ by what should correspond to the class of measurable rectangles.

Here, we use the phrasing in Loève (1997) of measurable cylinders. Let $S = (s_1, ..., s_k) \in \mathcal{P}_{of}$, $k \geq 1$. A finite measurable rectangle in E_S is generally denoted by

$$A_S = \prod_{j=1}^{k} A_{s_j}, \ (A_{s_j} \in \mathcal{B}_{s_j}, 1 \leq j \leq k) \tag{2.3}$$

It is clear that $\Pi_S^{-1}(A_S)$ is the set of all $x = (x_t)_{t \in T}$ such that

$$x_{s_j} \in A_{s_j}, \ 1 \leq j \leq k.$$

We write the above fact as

$$\Pi_S^{-1}(A_S) = A_S \times \prod_{t \notin S} E_t. \ \ (FP)$$

DEFINITION 11. *The class $\mathcal{S}$ of measurable cylinders of E is the class of subsets of E which are of the form $\Pi_S^{-1}(A_S)$, $S \in \mathcal{P}_{of}$.*

In other words, a measurable cylinder of E is a product of measurable subsets $A_t \in \mathcal{B}_t$ of the form

$$\prod_{t \in T} A_t, \ \ \ (SP01)$$

such that at most a finite number of the A_t, $t \in T$, are non-empty. If $V = \{v_1, ..., v_k\} \subset T$ is such that $A_t = \emptyset$ for $t \notin V$, then the product

$$A_V = \prod_{v \in V} A_v \ \ \ (SP02)$$

is called a finite support of the cylinder and the cylinder is written as

$$c(A_V) = A_V \times \prod_{t \notin V} E_t. \ \ (SP03)$$

Remarks. The following remarks are important.

(1) In the definition of the support of the cylinder in Formula (SP02), the order of V is not relevant in the writing of the cylinder $c(A_V)$, but it really counts in the writing of the support A_V.

(2) A support is not unique. For example if one of E_v, $v \in V$, is equal to E_v, may may drop it from the support. As well, we may add to V any other $w \notin V$ such that $A_w = E_w$: we may drop full spaces from the support and add full spaces to it.

(3) Formula (SP03) means that $x = (x_t)_{t \in T}$ is in the cylinder only if $x_v \in E_v$, $v \in V$, and we do not care about where are the x_t, $t \notin V$. The only knowledge about them is that they remain in their full space E_t, $t \notin V$.

(4) The notation $c(A_V)$ introduced in Formula (SP03) stands for cylinder of support A_V where V is non-empty set of T.

(5) For the sake of shorter notation, we may write the formula in (SP03) in the form

$$c(A_V) = A_V \times E'_S \ \ where \ \ E'_S = \prod_{t \in V} E_t$$

The coming concept of coherence, which is so important to the theory of Kolmogorov, depends on the understanding of the remarks above and the next remark.

(6) Common index support of two cylinders. Consider two cylinders

$$c(A_V) = A_V \times \prod_{t \notin V} E_t \quad \text{and} \quad c(B_W) = B_W \times \prod_{t \notin W} E_t. \tag{2.4}$$

of respective supports $V = (v_1, ..., v_q) \in \mathcal{P}_{of}$ and $W = (w_1, ..., w_p) \in\in \mathcal{P}_{of}$, $p \geq 1$, $q \geq 1$.

We want to find a common support for both $c(A_V)$ and $c(A_W)$. We proceed as follows. Let us form an ordered set U by selecting first all the elements of V in the ascendent order of the subscripts. Next we complete by adding the elements of W which were not already in U, still in the ascendent order of the subscripts. At the arrival, the elements of W, corresponding to the common elements of V and W if they exists, may not be present if U in the original order of the their subscripts in W. Rather, they are present in U in the subscripts order of some permutation $s(W)$ of W.

Example. Suppose that $V = (v_1, v_2, v_3, v_4, v_5)(1, 4, 7, 2, 5)$ and $W = (w_1, w_2, w_3, w_4) = (5, 2, 10, 8)$. We have

$$U = (1, 4, 7, 2, 5, 10, 8)$$

* So the elements of W are given in order in (w_2, w_1, w_3, w_4) which is $s(W)$ with $s(1) = 2$, $s(2) = 2$, $s(3) = 3$ and $s(4) = 4$.

But we already saw that the order of the subscripts of V, or W or U does not alter the cylinders $c(A_V)$, $c(B_W)$, $c(A_U)$ or $c(B_U)$. We have

$$c(A_V) = A_U \times E'_U \quad \text{and} \quad c(B_W) = B_U \times E'_U. \tag{2.5}$$

Actually, we formed A_U (resp. A_U) by adding full spaces E_t to the support A_V (resp. B_W) for $t \in U \setminus V$) (resp. for $t \in U \setminus W$) .

We say that we have written $c(A_V)$ and $c(B_W)$ with a common index support U. This consideration will be back soon.

(5) In the definition of a cylinder, the finite support, say A_V, is a measurable rectangle. But in general, A_V may be a measurable subset of E_V, that is $A_V \in \mathcal{B}_V$ and we still have

$$\Pi^{-1}(A_V) = A_V \times E'_V,$$

which is to be interpreted as

$$x \in A_V \times E'_V \Leftrightarrow x_V \in A_V.$$

Now, we are ready to go further and to give important properties of $\mathcal{S}$.

PROPOSITION 30. *$\mathcal{S}$ is a semi-algebra.*

Proof. (i) Let us see that $E \in \mathcal{S}$. If we need to prove it, we consider a point $t_0 \in T$, put $A_{t_0} = E_{t_0}$ and get that

$$E = A_{t_0} \times \prod_{t \neq t_0} E_t = c\left(A_{\{t_0\}}\right).$$

(ii) Next, by the definition of a cylinder of support $S \in \mathcal{P}_{of}$, checking that x belongs to $c(A_S)$ or not depends only of $x_S \in A_S$ or not. So we have

$$c(A_S)^c = A_S^c \times \prod_{t \notin S} E_t, \tag{2.6}$$

Next, let us check that the complement of any element of $\mathcal{S}$ is a finite sum of elements of S. We already knew that the class of measurable rectangles in E_V is a semi-algebra, so that A_S^c is a finite sum of elements of measurable rectangles $A_S^{(j)}$, $1 \leq j \leq r$, $r \geq 1$, of E_S. And it becomes obvious that

$$\begin{aligned}
c(A_S)^c &= A_S^c \times \prod_{t \notin S} E_t \\
&= \left(\sum_{1 \leq j \leq r} A_S^{(j)} \right) \times \prod_{t \notin S} E_t \\
&= \sum_{1 \leq j \leq r} \left(A_S^{(j)} \times \prod_{t \notin S} E_t \right) \\
&= \sum_{1 \leq j \leq r} c(A_S^{(j)}).
\end{aligned}$$

Our checking is successful.

(iii) Finally, let us check that $\mathcal{S}$ is stable under finite intersection. To do so, let us consider two cylinders

$$c(A_V) = A_V \times \prod_{t \notin V} E_t \quad \text{and} \quad c(B_W) = B_W \times \prod_{t \notin W} E_t, \tag{2.7}$$

and next their expressions using a common index support as explained earlier, we have

$$c(A_V) = A_U \times E'_U \quad \text{and} \quad c(B_W) = B_U \times E'_U. \tag{2.8}$$

It becomes clear that we have

$$c(A_V) \cap c(B_W) = c(A_V) = \left(A_U \cap B_U \right) \times E'_U, \tag{2.9}$$

which is element of $\mathcal{S}$ since

$$A_U \cap B_U = \prod_{t \in U} A_t \cap B_t.$$

* We also have

THEOREM 25. *The σ-algebra on E generated by the projections, denoted $\mathcal{B}$ in Definition 10, is also generated by the class of cylinders of finite support $\mathcal{S}$, called the product σ-algebra and denoted as*

$$\mathcal{B} = \bigotimes_{t \in T} \mathcal{B}_t.$$

Proof. Let us denote by $\mathcal{B}$ the σ-algebra on E generated by the projections with finite support and by $\mathcal{B}_0$ the one generated by $\mathcal{S}$.

(1) Let us prove that $\mathcal{B} \subset \mathcal{B}_0$. Let us fix $V \in \mathcal{P}_{of}$. For any measurable rectangle A_V which in E_V, that is $A_V \in \sigma(\mathcal{S}_V)$, we already now, since Formula (FP), that

$$\pi^{-1}(A_V) = A_V \times E'_V$$

and next

$$\pi_V^{-1}(A_V) = A_V \times E'_V \in \mathcal{S} \subset \mathcal{B}_0.$$

So each projection π_V^{-1} of finite support is $\mathcal{B}_0$-measurable. We conclude $\mathcal{B} \subset \mathcal{B}_0$ by the definition of $\mathcal{B}$.

2) Let us prove that $\mathcal{B}_0 \subset \mathcal{B}$. It is enough to prove that $\mathcal{S} \subset \mathcal{B}$. But any element A of (S) can be written as

$$A = \prod_{t \in V} A_t \times \prod_{t \notin V} E_t =: A_V \times E'_V, \ A_t \in \mathcal{B}_t$$

which is

$$A = \Pi_V^{-1}(A_V),$$

and then to $\mathcal{B}$, since $A_V \in \mathcal{B}_V$ and Π_V is $\mathcal{B}$-measurable. $\square$

3. Stochastic Process, Measurability for a family of Random Variables

I - General case : family of random variable.

Now we have the product space

$$E = \prod_{t \in T} E_t,$$

endowed with the product σ-algebra

$$\mathcal{B} = \bigotimes_{t \in T} \mathcal{B}_t.$$

We may study the measurability of mappings $X \ : \ (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{B})$. According to the notation above, we denote

$$\forall \omega \in \Omega, \ X(\omega) = (X_t(\omega))_{t \in T}.$$

For all $t \in T$, the mapping $\omega \longmapsto X_t(\omega)$ taking its values in E_t is called the t-th component or margin.We immediately have that for each $t \in T$,

$$X_t = \Pi_t \circ X.$$

It become clear that if X is measurable, thus each margin X_t, $t \in T$, is also measurable. Actually, this is a characterization of the measurability of such mappings.

Proposition 31. *A mapping $X \ (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$ is measurable if and only if each margin X_t, $t \in T$, is measurable. Indeed, we have*

Proof. We only need to prove the implication that if all the margins X_t, $t \in T$, are measurable, then X is. Suppose that all the margins X_t, $t \in T$, are measurable. It will be enough to show that $X^{-1}(cl(A_V))$ is measurable whenever $cl(A_V) \in \mathcal{S}$. By using the notation above, we have

$$\begin{aligned} \omega \in X^{-1}(c(A_V)) \quad & \leftrightarrow \quad X(\omega) \in c(A_V) \\ & \leftrightarrow \quad (\forall v \in V, X_v(\omega) \in A_v) \\ & \leftrightarrow \quad (\forall v \in V, \omega \in X_v^{-1}(A_v)) \\ & \leftrightarrow \quad \omega \in \bigcap_{v \in A} X_v^{-1}(A_v), \end{aligned}$$

which, by the measurability of the X_v's, gives

$$X^{-1}(c(A_V)) = \bigcap_{v \in A} X_v^{-1}(A_v) \in \mathcal{A}. \square$$

II - Stochastic Processes.

Let consider the special case where all the E_t are equal to one space E_0 on which is defined a σ-algebra $\mathcal{B}_0$. The product space is denoted by

$$E = E_0^T$$

and is the class of all mappings defined from T to E_0. As in the general context, elements of E are denoted $x = (x_t)_{t \in T}$, where for all $t \in T$, $x_t \in E_0$. The product σ-algebra is denoted by

$$\mathcal{B} = \mathcal{B}_0^{\otimes T}.$$

We have the general terminology :

(1) A measurable application $X \ : \ (\Omega, \mathcal{A}) \rightarrow \left(E_0^T, \mathcal{B}_0^{\otimes T}\right)$ is called a **stochastic process**.

(2) E_0 is called the states space of the stochastic process.

(3) T is called the time space in a broad sense.

(4) If $T = \{1\}$ is a singleton, the stochastic process is called a simple random variable.

(5) If $T = \{1, ..., k\}$ is finite with $2 \leq k \in \mathbb{N}$, the stochastic process is called a random vector.

(6) If $T = \mathbb{N}$, the stochastic process is a sequence of random variables.

(7) If $T = \mathbb{Z}$, the stochastic process is called a time series.

(8) If $T = \mathbb{R}_+$, the terminology of time space is meant in the real-life case.

(9) If T in endowed with a partial order, we generally speak of a random field or random net.

(10) For any $\omega \in \Omega$, the mapping

$$T \ni t \mapsto X_t(\omega).$$

is called a path of the stochastic process on E_0.

4. Probability Laws of Families of random Variables

(I) The concept of Coherence.

Consider a probability measure $\mathbb{P}$ on the product measure space $(E, \mathcal{B})$. The image measure on a sub-product $(E_S, \mathcal{B}_S)$ by the projection Π_S, where $S = (s_1, ..., s_k) \in \mathcal{P}_{of}$ is

$$\mathbb{P}_S = \mathbb{P}\Pi_S^{-1},$$

We recall that for any $B_S \in \mathcal{B}_S$, we have

$$\mathbb{P}_S(B_S) = \mathbb{P}(\Pi_S^{-1}(B_S)).$$

We get the family of probability measures

$$\{\mathbb{P}_S, S \in \mathcal{P}_{of}(T)\},$$

which we called *the family of marginal probability measures with finite index support*. By a language abuse, we also use the phrase of family of *finite-dimensional* marginal probability measures of $\mathcal{P}$.

We are going to discover some important relations between the *finite-dimensional* marginal probability measures. But we should also keep in mind that, for any $S \in \mathcal{P}_{of}(T)$, a probability measure $\mathbb{P}_S$ on $(E_S, \mathcal{B}_S)$ is characterized by its values on $\mathcal{S}_S$, which the class of measurable rectangles on E_S.

First, let us consider $(S_1, S_2) \in \mathcal{P}_{of}(T)^2$, such that one of them is a permutation of the other, that is $S_1 = s(S_2)$, where $S_1 = (s_1, ..., s_k)$, and s is a permutation $\{1, 2, ..., k\}$. Consider any $A_{S_1} \in \mathcal{S}_{S_1}$. We have

$$s(A_{S_1}) = s(A_{s_1} \times ... \times A_{s_k}) = A_{s(s_1)} \times ... \times A_{s(s_k)} = A_{S_2}.$$

Furthermore, the projection on E_{S_2} is the composition of the projection on E_{S_1} and the permutation of that projection by s, which gives

$$\Pi_{S_2} = s \circ \Pi_{S_1},$$

from which, by the characterization of a finite product probability by its values on the measurable rectangles, we have

$$\Pi_{S_2}^{-1} = \Pi_{S_1}^{-1} \circ s^{-1},$$

and similarly,

$$\Pi_{S_1}^{-1} = \Pi_{S_2}^{-1} \circ s.$$

Hence, for any $B_{S_1} \in \mathcal{B}_{S_1}$, we have

$$\mathbb{P}\Pi_{S_1}^{-1}(B_{S_1}) = \mathbb{P}\Pi_{S_2}^{-1} \circ s(B_{S_1}),$$

which leads to

$$\mathbb{P}_{S_1}(B_{S_1}) = \mathbb{P}_{S_2}(s(B_{S_1})),$$

and

$$\mathbb{P}_{S_1}(\cdot) = \mathbb{P}_{s(S_1)}(s(\cdot)).$$

* We already reached a first coherence (or consistency) condition. Let us discover a second one. Let $U = (u_1, .., u_r) \subset S = (s_1, ..., s_k)$, where the inclusion holds with the preservation of the ascendent order of the subscripts. Then the projection on E_U is obtained by the projection on E on E_S first, and next by the projection of E_S on E_U denoted $\Pi_{S,U}$. Accordingly to the notation above, we have for any $B_U \subset E_U$,

$$\Pi_{S,U}^{-1}(B_U) = B_U \times E'_{S \setminus U},$$

which is interpreted as

$$x_S = (x_U,\ x_{S\setminus U}) \in \Pi_{S,U}^{-1}(B_U) \Leftrightarrow x_U \in B_U.$$

Going back to the considerations which were made above about the projection on S, we have

$$\Pi_U = \Pi_{S,U} \circ \Pi_S,$$

and next,

$$\mathbb{P}\Pi_U^{-1} = \mathbb{P}\Pi_S^{-1}\Pi_{S,U}^{-1},$$

and finally,

$$\mathbb{P}_U = \mathbb{P}_S\Pi_{S,U}^{-1}.$$

We get a second relation between the marginal probability measures. Based on the previous developments, we may define

Definition 12. *A family of finite-dimensional probability measures* $\{\mathbb{P}_S, S \in \mathcal{P}_{of}(T)\}$ *is said to be coherent if and only if we have the following two conditions, called coherence coherent or consistency conditions :*

(CH1a) For any ordered and finite subsets U *and* S *of* T *such that* U *is subset of* S *with the preservation of the ascendent ordering of the subscripts of* U *in* S*, we have*

$$\mathbb{P}_U = \mathbb{P}_S \Pi_{S,U}^{-1}.$$

(CH2) For any ordered and finite subset S *of* T *and for any permutation* s *of* E_S*, for any* $B_S \in \mathcal{B}_S$*,*

$$\mathbb{P}_S(B_S) = \mathbb{P}_{s(S)}(s(B_S)).$$

Important Remarks.

(a) The condition (CH2) is useless when T is endowed with a total ordering. In that case, we may and do write the finite subsets of T always in that total order.

(b) The main coherence condition (CH1a) may have different equivalent forms.

(b1) First we may write (CH1a) when S has only one point more than U. From the new condition, we have the general one by simple induction.

(b2) We may also consider V and W two finite ordered subsets of T such that $U = V \cap W$ is not empty, and as usual, we suppose that $V \cap W$ is in V and in W with the same ascendent order of the subscripts. Condition (CH1a) gives

$$\mathbb{P}_S \Pi_{V,U}^{-1} = \mathbb{P}_S \Pi_{W,U}^{-1}$$

and this, in turn, implies (CH1a) for $U = V \subset W = S$. So we have the following new coherence condition :

(CH1b) For any $U = (u_1, .., u_r) \in \mathcal{P}_{of}(T)$ and $S = (u_1, ..., u_r, u_{r+1})$ with $u_{r+1} \notin U$,

$$\mathbb{P}_U = \mathbb{P}_S \Pi_{S,U}^{-1}$$

which is equivalent to saying that for $B \in \mathcal{B}_{(u_1,...,u_r)}$,

$$\mathbb{P}_{(u_1,...,u_r)}(B) = \mathbb{P}_{(u_1,...,u_r,u_{r+1})}(B \times E_{u_{r+1}}).$$

(CH1c) For any two finite and ordered subsets of V and W of T such that $U = V \cap W$ is not empty and is in V and in W with the same ascendent order of the subscripts, we have

$$\mathbb{P}_S \Pi_{V,U}^{-1} = \mathbb{P}_S \Pi_{W,U}^{-1}.$$

(II) Towards the construction of a probability law of a coherent family of marginal probability.

In this part, we try to solve the following problems.

(i) Given a coherent (or consistence) family of real-valued, non-negative, normed and additive mappings $\mathbb{L}_V$, $V \in \mathcal{P}_{of}(T)$ defined on $\mathcal{B}_V$, and denoted

$$\mathcal{F} = \{\mathbb{L}_V,\ V \in \mathcal{P}_{of}(T)\},$$

does it exists a real-valued, normed and non-negative and additive mapping $\mathbb{L}$ on $\mathcal{B}$ such that the elements of $\mathcal{F}$ are the finite-dimensional margins of $\mathbb{L}$, that is for any $V \in \mathcal{P}_{of}(T)$,

$$\mathbb{L}_V = \mathbb{L}\Pi_V^{-1}?$$

(ii) Given a coherent family of finite-dimensional probability measures $\mathbb{P}_V$, $S \in \mathcal{P}_{of}(T)$, defined on $\mathcal{B}_V$ and denoted

$$\mathcal{F} = \{\mathbb{P}_V,\ V \in \mathcal{P}_{of}(T)\},$$

does it exists a probability measure $\mathbb{P}$ on $\mathcal{B}$ such that the elements of $\mathcal{F}$ are the finite-dimensional marginal probability measures of $\mathbb{P}$, that is for any ,

$$\mathbb{P}_V = \mathbb{P}\Pi_V^{-1}?$$

Of course, if Problem (ii) is solved, Problem (i) is also solved, by taking $\mathbb{L} = \mathbb{P}$. On the other side, the solution of Problem (i) is the first step to the solution of Problem (ii).

We are going to see that Problem (i) has a solution with no supplementary conditions. We have

THEOREM 26. *Given a coherent family $\mathcal{F} = \{\mathbb{L}_V, V \in \mathcal{V}_{of}(T)\}$ of non-negative, normed and additive applications, as described above, it exists* ***a normed and non-negative and additive application*** $\mathbb{L}$ *on $\mathcal{B}$ such that the elements of $\mathcal{F}$ are the finite-dimensional margins of $\mathbb{L}$, that is for any $V \in \mathcal{P}_{of}(T)$,*

$$\mathbb{L}_V = \mathbb{L}\Pi_V^{-1}.$$

Proof. We adopt the notation introduced before to go faster. Let us suppose we are given a coherent family of $\{\mathbb{L}_V,\ S \in \mathcal{P}_{of}(T)\}$. Let us define $\mathcal{S} \subset \mathcal{P}(E)$ the class of cylinders of finite support, the following mapping

$$A_V \times E'_V \mapsto \mathbb{L}(A_V \times E'_V) = \mathbb{L}_V(A_V) \tag{4.1}$$

for all $V \in \mathcal{P}_{of}(T)$ and $A_V \in \mathcal{S}_V$, or in an other notation

$$\Pi_V^{-1}(A_V) \mapsto \mathbb{L}(\Pi_V^{-1}(A_V)) = \mathbb{L}_V(A_V) \tag{4.2}$$

The first thing to do is to show that $\mathbb{L}$ is well-defined. Indeed, the support of $A = A_V \times E'_V \in \mathcal{S}$ (with $A_V \in \mathcal{S}_V$) is not unique. Let us consider an other represent of A : $A = A_W \times E'_W,\ A_W \in \mathcal{S}_W$). If $U = V \cap W$ is empty, it means that all the factors of A are full spaces and and so $A = E$ and, and since all the $\mathbb{L}_t$'s are normed, we have

$$\mathbb{L}_V(A_V) = \mathbb{L}_W(A_W) = 1 = \mathbb{L}(A).$$

If U is not empty, what ever how it is ordered, it is present in V according to a certain order corresponding to a permutation of r of it. Also, there exists a permutation s of E_W such that $r(U)$ is in W with the preservation of the ascendent order of the script. So we may denote $r(U) = (u_1, ..., u_k)$, $V = (v_1, ..., v_p)$, $s(W) = (w_1, ..., w_q)$, $p \geq k$, $q \geq k$. We have

$$A_V = A_{r(U)} \times E_{V \setminus r(U)} \quad and \quad A_{s(W)} = A_{r(U)} \times E_{s(W) \setminus r(U)}$$

and it is clear that

$$\begin{aligned}
\mathbb{L}_V(A_V) &= \mathbb{L}_V(A_{r(U)} \times E_{V \setminus r(U)}) \quad (L11) \\
&= \mathbb{L}_V\left(\Pi^{-1}_{V,r(U)}(A_{r(U)})\right) \quad (L12) \\
&= \mathbb{L}_{r(U)}(A_{r(U)}) \quad (L13) \\
&= \mathbb{L}_{r(U)}(r(A_U)) \quad (L14) \\
&= \mathbb{L}_U(A_U) \quad (L15)
\end{aligned}$$

In Lines (L11)-(L13), we used the coherence condition (CH1a) while (CH2) was used in Lines (L14) and (L15).

At the arrival, using any writing of $A \in \mathcal{S}$ leads to the same value. Then the mapping $\mathbb{L}$ is well-defined and normed.

In the next step, we have to show that $\mathbb{L}$ is additive of $\mathcal{S}$. For this, let us consider an element of $\mathcal{S}$ that is split into two disjoint elements of $\mathcal{S}$. Suppose

$$A = B + C$$

with

$$A = A_U \times E'_U, \quad B = A_V \times E'_V \quad and \quad C = C_W \times E'_W.$$

Let consider $Z = U \cup V \cup W$ given in some order of the subscripts. There exist permutations s, r and p of E_U, E_V and E_W respectively such that $r(U)$, $s(V)$ and $p(W)$ are given in Z with the preservation of the ascendent order of the subscripts and we have :

$$A = \left(A_{r(U)} \times E_{Z \setminus r(U)}\right) \times E'_Z \equiv A^*_Z \times E'_Z,$$

$$B = \left(B_{s(V)} \times E_{Z \setminus s(V)}\right) \times E'_Z \equiv B^*_Z \times E'_Z,$$

and

$$C = \left(C_{p(Z)} \times E_{Z \setminus p(W)}\right) \times E'_Z \equiv C^*_Z \times E'_Z,$$

with

$$A^*_Z \times E'_Z = \left(B^*_Z \times E'_Z\right) + \left(C^*_Z \times E'_Z\right)$$

This is possible only if we have

$$A^*_Z = B^*_Z + C^*_Z,$$

with

$$\mathbb{L}_Z(A^*_Z) = \mathbb{L}_U(A_U), \quad \mathbb{L}_Z(B^*_Z) = \mathbb{L}_V(B_V), \quad \textit{and} \quad \mathbb{L}_Z(C^*_Z) = \mathbb{L}_W(C_Z)$$

Using the coherence conditions, we have

$$\begin{aligned}
\mathbb{L}(B) &= \mathbb{L}_Z(B^*_Z)) \quad (L31) \\
&= \mathbb{L}_Z\left(\Pi^{-1}_{r(V),Z}\left(B_{r(V)}\right)\right) \quad (L32) \\
&= \mathbb{L}_{r(V)}\left(B_{r(V)}\right) \quad (L33) \\
&= \mathbb{L}_V(B_V).
\end{aligned}$$

By doing the same for A and C, we have

$$\begin{aligned}
\mathbb{L}(A) = \mathbb{L}_U(A_U) = \mathbb{L}_Z(A^*_Z)), \\
\mathbb{L}(B) = \mathbb{L}_V(B_V) = \mathbb{L}_Z(B^*_Z)), \\
\mathbb{L}(C) = \mathbb{L}_W(C_W) = \mathbb{L}_Z(C^*_Z))
\end{aligned}$$

Bu using the additivity of $\mathbb{L}_Z$, we conclude that

$$\mathbb{L}(A) = \mathbb{L}(B) + \mathbb{L}(C).$$

The mapping $\mathbb{L}$ is normed and additive on the semi-algebra. From Measure Theory and Integration (See Doc 04-02, Exercise 15, in Lo (2017b)), $\mathbb{L}$ is automatically extended to a normed and additive mapping on the algebra $\mathcal{C}$ generated by $\mathcal{S}$.

Now, we face Problem (ii). Surely, the assumptions and the solution Problem (i) ensure that there exists a normed and additive mapping $\mathbb{L}$ whose margins are the $\mathbb{P}_V$, $V \in \mathcal{P}_{of}(T)$. Let us call it $\mathbb{P}$. All we have to do is to get an extension of $\mathbb{P}$ to $\sigma(\mathcal{C}) = \mathcal{B}$.

A way to do it is to use Caratheodory's Theorem (See **Doc 04-03 in Lo (2017b)** for a general revision). But, unfortunately, we need special spaces. Suppose that each (E_t, d_t), $t \in T$, is Polish space, that is a metric separable and complete separable space. The following facts are known in Topology. For $V \in \mathcal{P}_{of}(T)$, the space E_V is also a Polish space. In such spaces, the extension of $\mathbb{P}$ to probability measure is possible. The proof heavily depends on topological notions, among them a characterization of compact sets.

We give the proof in the last section as an Appendix. In the body of the text, we focus on probability theory notions. However, we strongly recommend the learners to read the proof in small groups. We have the following Theorem.

THEOREM 27. *(Fundamental Theorem of Kolmogorov) Let us suppose that each (E_t, d_t), $t \in T$ is Polish Space. For any $V \in \mathcal{P}_{of}(T)$, E_V is endowed with the Borel σ-algebra associated with the product metric of the metrics of its factors.*

For $T \neq \emptyset$, given a coherent family $\mathcal{F} = \{\mathbb{P}_V,\ V \in \mathcal{P}_{of}(T)\}$ of finite-dimensional probability measures, there exists a unique probability measure on $\mathcal{B}$ such that the elements of $\mathcal{F}$ are the finite-dimensional margins of $\mathbb{L}$, that is for any $V \in \mathcal{P}_{of}(T)$,

$$\mathbb{P}_V = \mathbb{P}\Pi_V^{-1}.$$

Now, we are going to derive different versions of that important basis of Probability Theory and provide applications and examples.

To begin, let us see how to get the most general forms the Kolmogorov construction in finite dimensions (See Chapter 2 in pages 44 and 62). Let us repeat a terminology we already encountered. For any mapping

$$X \ :\ (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E, \mathcal{B})$$

we have $X(\omega) = (X_t(\omega))_{t \in T}$. For any $V = (v_1, ..., v_k) \in \mathcal{P}_{of}(T)$, $k \geq 1$,

$$X_V \equiv (X_{v_1}, ..., X_{v_k}) = \Pi_V(X),$$

is called a finite-dimensional (and ordered) margin of X. We have

THEOREM 28. *Let us suppose that each* (E_t, d_t), $t \in T$ *is a Polish Space. For any* $V \in \mathcal{P}_{of}(T)$, E_S *is endowed with the Borel* σ*-algebra associated with the product metric of the metrics of its factors.*

For $T \neq \emptyset$, *given a coherent family* $\mathcal{F} = \{\mathbb{P}_V,\ V \in \mathcal{P}_{of}(T)\}$ *of probability measures, there exists a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *and a measurable mapping* X *defined on* Ω *with values in* $(E, \mathcal{B})$ *such that for any* $V \in \mathcal{P}_{of}(T)$, $k \geq 1$,

$$\mathbb{P}_V = \mathbb{P}_{X_V} = \mathbb{P}X_V^{-1}.$$

In other words, there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ *holding a measurable mapping* X *with values in* $(E, \mathcal{B})$ *such that the finite-dimensional marginal probability measures* $\mathbb{P}_V$, $V \in \mathcal{P}_{of}(T)$ *are the probability laws of the finite-dimensional (and ordered) margins* X_V *of* X.

Furthermore, the probability laws of the finite-dimensional (and ordered) margins X_V *of* X *determine the probability law of* X.

Proof. We apply Theorem 27 above to get the probability measure on $\mathbb{P}$ on $(E, \mathcal{B})$ whose finite-dimensional marginal probabilities are the $\mathbb{P}_V$, $V \in \mathcal{P}_{of}(T)$. Now we take

$$(\Omega, \mathcal{A}, \mathbb{P}) = (E, \mathcal{B}, \mathbb{P})$$

* and set X as the identity mapping. We have

$$\begin{aligned} \mathbb{P}_{X_V} &= \mathbb{P}X_V^{-1} = \mathbb{P}\Pi_V(X)^{-1} \\ &= \mathbb{P}\Pi_V^{-1}X^{-1} \\ &= \mathbb{P}_V X^{-1}. \end{aligned}$$

Now for any $B_V \in \mathcal{B}_V$,

$$\mathbb{P}_{X_V}(B_V) = \mathbb{P}_V(\{\omega \in \Omega, X(\omega) = \omega \in B_V\}) = \mathbb{P}_V(B_V).$$

We get the desired result : $\mathbb{P}_{X_V} = \mathbb{P}_V$, $V \in \mathcal{P}_{of}(T)$. To finish, the probability law of X is given by

$$\mathbb{P}_X(B) = \mathbb{P}X^{-1}(B),\ B \in \mathcal{B}.$$

By the uniqueness of the Caratheodory's extension for a σ-additive and proper mapping of from an algebra to the σ-algebra generated, the probability measure $\mathbb{P}X^{-1}$ on $\mathcal{B}$ is characterized by its values on $\mathcal{S}$. But an element of $\mathcal{S}$ is of the form

$$B = B_V \times E'_V, B_V \in \mathcal{B},\ V \in \mathcal{P}_{of}(T).$$

We have have

$$X \in B = B_V \times E'_V \Leftrightarrow X_V \in B_V$$

that is

$$X^{-1}(B) = X_V^{-1}(B_V).$$

Hence, by applying $\mathbb{P}$ at both sides, we get

$$\mathbb{P}_X(B_V \times E'_V) = \mathbb{P}_{X_V}(B_V).$$

* Since the values of $\mathbb{P}_X$ are functions only of the values of the probability laws of the finite-dimensional (and ordered) margins X_V of X, these latter finally determine $\mathbb{P}_X$. $\square$

We are continuing to see developments of the Kolmogorov Theorem in special sections.

5. Skorohod's Construction of real vector-valued stochastic processes

(I) - The General Theorem.

Let T be an non-empty index set. For each $t \in T$, let be given $E_t = \mathbb{R}^{d(t)}$, where $d(t)$ is positive integer number. Let us consider a family of probability distribution functions described as follows : for $V = (v_1, ..., v_k) \in \mathcal{P}_{of}(T)$, $k \geq 1$, we set $d(V) = d(v_1) + ... + d(v_k)$. The probability distribution function associated to V is defined for $x_{v_j} \in \mathbb{R}^{d(v_j)}$, $1 \leq j \leq k$, by

$$\mathbb{R}^{d(V)} \ni (x_{v_1}, ..., x_{v_k}) \mapsto F_V(x_{v_1}, ..., x_{v_k}).$$

The family of $\{F_V,\ V = (v_1, ..., v_k) \in \mathcal{P}_{of}(T)\}$ is coherent if and only if :

(CHS1), for $V \in \mathcal{P}_{of}(T)$, for any $(x_{v_1}, ..., x_{v_k}) \in \mathbb{R}^{d(V)}$, for any permutation of E_V,

$$F_V(x_{v_1},...,x_{v_k}) = F_{s(V)}(s(x_{v_1},...,x_{v_k}))$$

and

(CHS2) for $V \in \mathcal{P}_{of}(T)$, for any $(x_{v_1},...,x_{v_k}) \in \mathbb{R}^{d(V)}$, for any $u \in T \backslash V$,

$$F_V(x_{v_1},...,x_{v_k}) = \lim_{u \uparrow \{+\infty\}^{d(u)}} F_V(x_{v_1},...,x_{v_k},u).$$

In all this part, by writing $(x_{v_1},...,x_{v_k}) \in \mathbb{R}^{d(V)}$, we also mean that $x_{v_j} \in \mathbb{R}^{d(v_j)}$, for all $j \in \{1,...,k\}$.

Here is the Skorohod Theorem as follows.

THEOREM 29. *Given a coherent family of probability distribution functions $\{F_V,\ V = (v_1,...,v_k) \in \mathcal{P}_{of}(T)\}$, there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding a measurable mapping X with values in $(E, \mathcal{B})$ such that each finite-dimensional marginal probability distribution function F_{X_V}, $V \in \mathcal{P}_{of}(T)$ is F_V, that is for of $V = (v_1,...,v_k)$, $k \geq 1$, for $(x_{v_1},...,x_{v_k}) \in \mathbb{R}^{d(V)}$,*

$$F_V(x_{v_1},...,x_{v_k}) = \mathbb{P}(X_{v_1} \leq x_{v_1},...,X_{v_k} \leq x_{v_k}).$$

Proof. The proof results from the application of The Kolmogorov Theorem and a smart use of the Lebesgue-Stieljes measures. We remind first that for any $\ell \geq 1$, a finite measure on $\mathbb{R}^\ell$ is characterized by its values of the elements of the form

$$]-\infty, a] = \prod_{1 \leq j \leq \ell}]-\infty, a_j],\ a = (a_1,...,a_\ell)$$

which form a π-system denoted $\mathcal{D}_\ell$, which in turn, generates $\mathcal{B}(\mathbb{R}^\ell)$. Consider the unique Lebesgue-Stieljes probability measure $\mathbb{P}_V$ on $\mathbb{R}^{d(V)}$ associated with F_V, $V = (v_1,...,v_k) \in \mathcal{P}_{of}(T)\}$, $k \geq 1$. By keeping the previous notation, we have that for any $(x_{v_1},...,x_{v_k}) \in \mathbb{R}^{d(V)}$ and for any permutation of E_V

$$
\begin{aligned}
\mathbb{P}_V(v) &= F_V(x_{v_1}, ..., x_{v_k}) \\
&= F_{s(V)}(s(x_{v_1}, ..., x_{v_k})) \\
&= \mathbb{P}_{s(V)}\left(s\left(\prod_{1\leq j\leq k}]-\infty, x_{v_j}]\right)\right) \\
&= \mathbb{P}_{s(V)}s^{-1}\left(\prod_{1\leq j\leq k}]-\infty, x_{v_j}]\right).
\end{aligned}
$$

Since the probability measures $\mathbb{P}_V$ and $\mathbb{P}_{s(V)}s^{-1}$ coincide on $\mathcal{D}_{d(V)}$, they are equal and the first coherence condition is proved. To prove the second, we have

$$
\begin{aligned}
\mathbb{P}_V\left(\prod_{1\leq j\leq k}]-\infty, x_{v_j}]\right) &= F_V(x_{v_1}, ..., x_{v_k}) \\
&= \lim_{\mathbb{R}^{d(u)}\uparrow\{+\infty\}^{d(u)}} F_{V\cup\{u\}}(s(x_{v_1}, ..., x_{v_k}), u) \\
&= \lim_{\mathbb{R}^{d(u)}\uparrow\{+\infty\}^{d(u)}} \mathbb{P}_{V\cup\{u\}}\left(\prod_{1\leq j\leq k}]-\infty, x_{v_j}]\times]-\infty, u]\right) \\
&= \mathbb{P}_{V\cup\{u\}}\left(\prod_{1\leq j\leq k}]-\infty, x_{v_j}]\times E_u\right).
\end{aligned}
$$

But the two probability measures on $\mathbb{R}^{d(V)}$: $\mathbb{P}_V(B)$ and $\mathbb{P}_{V\cup\{u\}}(B \times E_u)$, $B \in \mathcal{B}_V$, coincide on $\mathcal{D}_{d(V)}$. Hence for any $B \in \mathcal{B}_V$, we have

$$
\mathbb{P}_V(B) = \mathbb{P}_{V\cup\{u\}}(B \times E_u).
$$

Thus, the coherence condition (CH1b) holds. Finally there exists a probability measure $\mathbb{P}$ on $(E, \mathcal{B})$ whose finite-dimensional marginal probability measures are the elements of $\{\mathbb{P}_V,\ V \in \mathcal{P}_{of}\}$. Let us take

$$
(\Omega, \mathcal{A}, \mathbb{P}) = (E, \mathcal{B}, \mathbb{P})
$$

and set X as the identity mapping. We have, for any $(x_{v_1}, ..., x_{v_k}) \in \mathbb{R}^{d(V)}$,

$$\begin{aligned}
\mathbb{P}(X_{v_1} \leq x_{v_1}, ..., X_{v_k} \leq x_{v_k}) &= \mathbb{P}(\{\omega \in E, \omega_{v_1} \leq x_{v_1}, ..., \omega_{v_k} \leq x_{v_k}\}) \\
&= \mathbb{P}(\prod_{1 \leq j \leq k}]-\infty, x_{v_j}] \times E'_V) \\
&= \mathbb{P}_V \left(\prod_{1 \leq j \leq k}]-\infty, x_{v_j}] \right) \\
&= F_V(x_{v_1}, ..., X_{v_k}).
\end{aligned}$$

The proof is finished. □

Other forms of the Skorohod Theorem using densities of probability.

Suppose we have the similar following situation as earlier. For each $t \in T$, let be given $E_t = E_0^{d(t)}$, where $d(t)$ is positive integer number. Let ν be a σ-finite measure on E_0. On each $E_V = E_0^V$, $V \in \mathcal{P}_{of}(T)$, we have a the finite product probability :

$$\nu_V = \nu^{\otimes d(v_1)} \nu \otimes^{\otimes d(v_2)} \otimes \cdots \otimes \nu^{\otimes d(v_k)} = \nu^{\otimes d(V)}.$$

* Now a family of marginal probability density functions (*pdf*) $\{f_V,\ V = (v_1, ..., v_k) \in \mathcal{P}_{of}(T)\}$, each f_V is *pdf* with respect to $\nu \otimes d(V)$ on $E_0^{d(V)}$, is said to be coherent if the two conditions hold :

(CHSD1) For $V \in \mathcal{P}_{of}(T)$, for any $(x_{v_1}, ..., x_{v_k}) \in E_0^{d(V)}$, for any permutation of E_V, we have

$$f_V(x_{v_1}, ..., x_{v_k}) = f_{s(V)}(s(x_{v_1}, ..., x_{v_k}))$$

and,

(CHSD2) for $V \in \mathcal{P}_{of}(T)$, for any $(x_{v_1}, ..., x_{v_k}) \in \mathbb{R}^{d(V)}$, for any $u \in T \setminus V$,

$$f_V(x_{v_1}, ..., x_{v_k}) = \int_{E_{v_u}} f_V(x_{v_1}, ..., x_{v_k}, u) d\nu^{\otimes d(v_u)}(u),$$

meaning that $f_V(x_{v_1}, ..., x_{v_k})$ is a marginal *pdf* of $f_V(x_{v_1}, ..., x_{v_k}, u)$.

Let us consider the finite distribution probability measure on E_V defined by

$$E_V(=E_0^{d(V)}) \ni B \mapsto \mathbb{P}_V(B) = \int_{E_V} f_V(x_{v_1},...,x_{v_k})d\nu^{\otimes d(V)}(x_{v_1},...,x_{v_k}),$$

It is easy to see that (CHSD1) and (CHSD2) both ensure that the $\mathbb{P}_V$ form a coherent family of finite dimensional probability measures. We apply theorem to conclude that :

For any coherent family of marginal probability density functions (*pdf*) $\{f_V,\ V=(v_1,...,v_k) \in \mathcal{P}_{of}(T)\}$, each f_V is *pdf* with respect $\nu^{\otimes d(V)}$ on $E_0^{d(V)}$, there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding a measurable mapping X with values in $(E, \mathcal{B})$ such that each finite-dimensional *pdf* f_{X_V}, $V \in \mathcal{P}_{of}(T)$ is F_V, that is for of $V=(v_1,...,v_k)$, $k \geq 1$, for $(x_{v_1},...,x_{v_k}) \in \mathbb{R}^{d(V)}$,

$$d\mathbb{P}(X_{v_1},\cdots,X_{v_k}) = f_V d\nu^{\otimes d(V)}.$$

In general, this is used in the context of $\mathbb{R}^d$, $d \geq 1$, with ν being the Lebesgue measure or a counting measure on $\mathbb{R}$. But it goes far beyond as a general law.

6. Examples

To make it simple, let $T=\mathbb{R}_+$ or $T=\mathbb{N}$. So we do not need to care about the first coherence condition since we have a natural order. Let be given $E_t = E_0^{d(t)}$, where E_0 is a polish space and $d(t)$ is positive integer number.

Problem 1. Given a family of Probability measures $\mathbb{P}_t$ on each E_t, of dimension $d(t)$. Does-it exist a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding a stochastic process $(X_t)_{t\in T}$ with independent margins such that each margin X_t follows the probability law $\mathbb{P}_t$.

Solution. We can easily that the family of finite dimensional probability measure,

$$\mathbb{P}_{(t_1,t_2,...,t_k} = \bigotimes_{j=1}^{k} \mathbb{P}_{t_j};$$

for $t_1 < ... < t_k$, defined by, for any $B_j \in \mathcal{B}(E_t)$, $j \in \{1, ..., k\}$,

$$\mathbb{P}_{(t_1,t_2,...,t_k}\left(\prod_{1\leq j\leq k} B_j\right) = \prod_{j=1}^{k} \mathbb{P}_{t_j}(B_j).$$

is coherent since for $t_{j+1} > t_j$,

$$\begin{aligned}\mathbb{P}_{(t_1<t_2<...<t_k}\left(\prod_{1\leq j\leq k} B_j \times E_{j+1}\right) &= \prod_{j=1}^{k} \mathbb{P}_{t_j}(B_j) \times \mathbb{P}_{t_{j+1}}(E_{j+1}) \\ &= \prod_{j=1}^{k} \mathbb{P}_{t_j}(B_j).\end{aligned}$$

Thus, the answer is positive.

Problem 2. Many techniques are based on the symmetrization method as in the proof of Proposition 24 (See page 224). We need to have two sequences $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ on the same probability space and having their values on $\mathbb{R}^d$ such that $X_n =_d Y_n$ for each $n \geq 1$. Is it possible?

Here is the statement of the problem for independent margins.

Given a family of Probability measures $\mathbb{P}_t$ on each E_t, of dimension $d(t)$. Does-it exist a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding a stochastic process $(X_t)_{t\in T}$ with independent margins such that :

(a) $X_t \in E_t^2$, that is $X_t = (X_t^{(1)}, X_t^{(2)})^t$

(b) For each $t \in T$, for each $i \in \{1, 2\}$, $\mathbb{P}_{X_t^{(i)}} = \mathbb{P}_t$.

If this problem is solved and if E_0 is a linear space, we may form the symmetrized form $X^{(s)} = X_t^{(1)} - X_t^{(2)}$ with $X_t^{(1)} =_d X_t^{(2)}$.

Solution. Let us apply the solution of Problem 1 for the case where $E_0 = \mathbb{R}$ in the context of independent margins. We notice that nothing is said about the dependence between $X_t^{(1)}$ and $X_t^{(2)}$. So we may take, for any $t \in T$, an arbitrary probability distribution function F_t on $\mathbb{R}^{2d(t)}$ such that the margins

$$F_t(\underbrace{x_1, \cdots, x_{d(t)}}_{d(t) \text{ times}}, \underbrace{+\infty, \cdots, +\infty}_{d(t) \text{ times}})$$

and

$$F_t(\underbrace{+\infty, \cdots, +\infty}_{d(t) \text{ times}}, \underbrace{x_{d(t)+1}, \cdots, x_{2d(t)}}_{d(t) \text{ times}}),$$

are equal both the probability distribution function of $\mathbb{P}_t$. This is possible by the use of copulas. With such a frame, we apply again the Skorohod Theorem to get our solution.

Problem 3. Existence of the Poisson Process. Given $\theta > 0$, by the solution of Problem 1, there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding a sequence independent random variables identically distributed as the standard exponential law $\mathcal{E}$ denoted X_1, X_2, etc.

Let us call them the independent and exponential inter-arrival times.

Let us define the arrival times Z_0, $Z_j = X_1 + \cdots + X_j$, $j \geq 1$, so that we have

$$Z_0 < Z_1 < \cdots Z_j \cdots$$

If we suppose that the Z_j are the arrival times of clients at a desk (say a bank desk) and $Z_0 = 0$ is the opening time of the desk, we may wish to know the probability law of the number at arrived clients at a time $t > 0$,

$$N(]0,t]) = N_t = \sum_{j \geq 1} 1_{(Z_j \leq t)}.$$

Here, we say that we have a standard Poisson Process (SPP) of intensity θ. Sometimes, authors mean $(N_t)_{t \geq 0}$ which is the counting function of the SPP, others mean $(X_n)_{n \geq 1}$ which is the sequence of arrival times or $(Z_n)_{n \geq 1}$ which is the sequence of inter-arrival times.

Problem 4. Existence of Brownian Movement by the exercise. Let $0 = t_0 < t_1 < ... < t_n$ be n real numbers and consider $Y_1, Y_2, ..., Y_n$, n non-centered Gaussian with respective variances $t_1, t_2 - t_1, ..., t_n - t_{n-1}$. Set

$$X = (X_1, X_2, ..., X_n) = (Y_1, Y_1 + Y_2, ..., Y_1 + Y_2 + \cdots + Y_n).$$

a) Find the density of X.

b) Give the distribution function of X.

c) Now, consider that family of distribution functions indexed by the ordered and finite subsets of $\mathbb{R}_+$: for any $(x_1, x_2, ..., x_k) \in \mathbb{R}^k$, $k \geq 1$,

$$F_{(t_1,t_2,...<t_k)}(x_1, x_2, ..., x_k) \qquad (BR01)$$
$$= \int_{-\infty}^{x_1} dy_1 \int_{-\infty}^{x_2} dy_2 \cdots \int_{-\infty}^{x_k} \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \exp\biggl(-\frac{1}{2} \frac{(y_i - y_{i-1)}^2}{ti - t{-}1} \biggr)\ dy_k,$$

where $(t_1 < t_2, ... < t_k)$ is an ordered and finite subset of $\mathbb{R}_+$, with $y_0 = t_0 = 0$.

c1) Say on the basis of questions (a) and (b), why do we have, for all $(t_1 < t_2, ... < t_k)$, $k \geq 2$; for all $(x_1, x_2, ..., x_{k-1}) \in \mathbb{R}^{k-1}$

$$\lim_{x_k \uparrow \infty} F(t_1, t_2, ..., t_k)(x_1, x_2, ..., x_k) = F_{(t_1,t_2,...,t_{k-1})}(x_1, x_2, ..., x_{k-1})$$

c2) Show this property directly from the definition (BR01).

c3) Conclude by the Kolmogorov-Skorohod, that there is a stochastic process $(\Omega, \mathcal{A}, \mathbb{P}, (B_t)_{t \in \mathbb{R}_+})$ for which we have

$$F_{(B(t_1),B(t_2),...,B(t_k))}(x_1, x_2, ..., x_k) = F_{(t_1,t_2,...<t_k)}(x_1, x_2, ..., x_k)$$

for any finite and ordered subset $(t_1, t_2, ... < t_k)$ of $\mathbb{R}_+$.

Alternatively, use the system of *pdf*'s indexed by the ordered and finite subsets of $\mathbb{R}_+$: for any $(x_1, x_2, ..., x_k) \in \mathbb{R}^k$, $k \geq 1$,

$$f_{(t_1,t_2,...<t_k)}(x_1, x_2, ..., x_k) \ (BR02)$$
$$= \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \exp\biggl(-\frac{1}{2} \frac{(y_i - y_{i-1)}^2}{ti - t{-}1} \biggr)$$

where $(t_1 < t_2, ... < t_k)$ is an ordered and finite subset of $\mathbb{R}_+$, with $y_0 = t_0 = 0$, and the coherence condition (CHSD2) (page 319) to justify the existence of such a process.

d) Such a stochastic process $(B_t)_{t \in \mathbb{R}_+}$ is called *Brownian motion* in Probability Theory and *Wiener Process* in Statistics.

Show or state the following facts.

d1) Its finite distributions are non-centered Gaussian vectors.

d2) For all $0 \leq s < t$, $B_t - B_s$ and B_s are independent.

d3) For all $t \geq 0$,

$$B_t - B_s \sim B(t-s) \sim \mathcal{N}(0, t-s).$$

d4) $\Gamma(s,t) = \mathbb{C}ov(B_t, B_s) = \min(s,t), \ (t,s) \in \mathbb{R}_+^2$.

(e) A stochastic process $(X_t)_{t \geq 0}$ is said to be Gaussian if and only if its finite margins are Gaussian vectors.

Show the following points :

(e1) Show that the probability law of a Gaussian Process is entirely determined by its mean function

$$m(t) = \mathbb{E}(X_t), \ t \in \mathbb{R}_+.$$

and by its variance-covariance function

$$\Gamma(s,t) = \mathbb{C}ov(X_t, X_s), \ (t,s) \in \mathbb{R}_+^2.$$

(e2) Deduce from this that the probability law of the Brownian Process is entirely the variance-covariance function

$$\Gamma(s,t) = \mathbb{C}ov(X_t, X_s) = \min(s,t), \ (t,s) \in \mathbb{R}_+^2.$$

7. Caratheodory's Extension and Proof of the Fundamental Theorem of Kolmogorov

Our departure point is the end of the Proof of Theorem 26. The construction mapping $\mathbb{L}$; we denote now as $\mathbb{P}$ is additive and normed on the algebra $\mathcal{C} = a(\mathcal{S})$ generated by $\mathcal{S}$. By Carathéodory Theorem (Doc 04-03 in Lo (2017b)), $\mathbb{P}$ is uniquely extensible to a probability measure whenever it is continuous at $\emptyset$, that is, as $n \to \infty$,

$$\left(\mathcal{C} \ni A_n \downarrow \emptyset\right) \implies \left(\mathbb{P}(A_n) \downarrow 0\right).$$

Actually, we are going to use an *ab contrario* reason. Suppose that there exists a non-increasing sequence $(A_n)_{n\geq 0} \subset \mathcal{C}$ and $\mathbb{P}(A_n)$ does not converges to zero. Since the sequence $(\mathbb{P}(A_n))_{n\geq 1}$ is non-increasing, its non-convergence to zero is equivalent to

$$\exists \varepsilon > 0), \ (\forall n \geq 1, \mathbb{P}(A_n) > \varepsilon).$$

At the beginning let us remark that $\mathcal{C} = a(\mathcal{S})$ is formed by the finite sum of elements of $\mathcal{S}$, we rely on the above considerations on finite sums of elements of $\mathcal{S}$, and easily get that any element of $\mathcal{C}$, and then any A_n is of the form

$$A_n = B_{V_n} \times E'_{V_n}, \ n \geq 1,$$

where $B_{V_n} \in \mathcal{B}_{V_n}$, and $V_n \in \mathcal{P}_{of}(T)$. Hence the whole sequence does involve only a countable spaces E_t, $t \in T_0$, where

$$T_0 = \bigcup_{j=1}^{\infty} V_n$$

and denote, accordingly,

$$E_{T_0} = \prod_{t \in T_0} E_t.$$

So we may ignore all the other factors in E.

$$A_n = B_{V_n} \times \prod_{t \notin T_0 \setminus V_n} E_t.$$

As well, we may use the natural order of integers and only consider supports index sets of the form $V_n = (1, ..., m(n))$, $n \geq 1$. Now we are going to use the following key topological property : in a Polish space, for any Borel set B, for any finite measure μ, for any $\varepsilon > 0$, there exists a compact set $K(\varepsilon)$ such

$$\mu(B \setminus K(\varepsilon)) < \varepsilon.$$

Then for any $\varepsilon > 0$, for any $n \geq 1$, there exists a compact set $B'_{V_n} \subset B_{V_n}$ such that

$$P_{V_n}(B_{V_n} - B'_{V_n}) \leq \varepsilon 2^{-(n+1)}.$$

Let us denote

$$A'_n = B'_{V_n} \times E'_{T_0 \setminus V_n}$$

Hence for each $n \geq 1$,

$$P(A_n - A'_n) = P_{V_n}(B_{V_n} - B'_{V_n}) < \varepsilon 2^{-(n+1)}.$$

Let us set

$$C_n = A'_1 \cap ... \cap A'_n.$$

We have for each $n \geq 1$,

$$\begin{aligned} P(A_n - C_n) &= P(A_n \cap (\bigcup_{j=1}^{n} (A'_j)^c)) \leq \sum_{j=1}^{n} P(A_n \cap (A'_j)^c)) \\ &\leq \sum_{j=1}^{n} P(A_n - A'_j) \\ &\leq \sum_{j=1}^{n} P(A_j - A'_j) \\ &\leq \sum_{j=1}^{n} \varepsilon 2^{-(j+1)} < \varepsilon/2. \end{aligned}$$

But $C_n \subset A'_n \subset A_n$, we have

$$P(C_n) = P(A_n) - P(A_n - C_n) > P(A_n) - \varepsilon/2 > \varepsilon/2.$$

We conclude that for all $n \geq 1$, C_n is non-empty. So, by the axiom of Choice, we may choose, $n \geq 1$, $x^{(n)} = (x_1^{(n)},\ x_2^{(n)}, ...) \in E_{T_0}$ such that $x^{(n)}$. By the non-decreasingness of the sequence $(C_n)_{n\geq 1}$, the sequence $(x^{(n)})_{n\geq 1}$ is in C_1, which we recall is such that

$$C_1 \subset A_1' = B_{V_1}' \times E_{T_0 \setminus V_1}'$$

$$\forall n \geq 1,\ (x_1^{(n)}, x_2^{(n)}, ..., x_{m_1}^{(n)})) \in E_{V_1}'.$$

Since B_{V_1}' est compact, there exists a sub-sequence $(x^{(n_{1,k})})$ of $(x^{(n)})$ such that

$$(x_1^{(n_{1,k})}, x_2^{(n_{1,k})}, ..., x_{m_1}^{(n_{1,k})})) \rightarrow (x_1^*, ..., x_{m(1)}^*) \in B_{V_1}' \subset B_{V_1}.$$

But the sub-sequence Now by the nature $(x^{(n_{1,k})}) \in C_2$ whenever $n_{1,k} \geq 2$ (which happens for from some value $k_2 > 0$ since the sequence $(n_{1,k})_{k \geq 1}$ is an increasing sequence of non-negative integers). We thus have

$$\forall k > k_2 \geq 1,\ (x_1^{(n_{1,k})}, x_2^{(n_{1,k})}, ..., x_{m_1}^{(n_{1,k})})) \in E_{V_2}'.$$

We conclude similarly that there exists a sub-sequence $(x^{(n_{2,k})})_{k \geq 1}$ of $(x^{(n_{1,k})})_{k \geq 1}$ such that

$$(x_1^{(n_{2,k})}, x_2^{(n_{2,k})}, ..., x_{m_1}^{(n_{2,k})})) \rightarrow (x_1^*, ..., x_{m(2)}^*) \in B_{V_2}.$$

It is important the for a common factor j between A_{V_1} and A_{V_2}, the limit x_j^* remains unchanged as the limit of a sub-sequence of converging sequence. We may go so-forth and consider the diagonal sub-sequence

$$(x^{(n_{k,k})})_{k \geq 1}.$$

We have that for each $n \geq 1$, there exists $K(n) > 0$ such that $(x^{(n_{k,k})})_{k \geq K(n)} \subset C_n$. Hence, for each $n \geq 1$,

$$(x_1^*, ..., x_{m(n)}^*) \in B_{V_n}.$$

So by denoting

$$x^* = (x_1^*, x_2^*, ...),$$

we get that x^* belongs to each A_n, $n \geq 0$. Hence A is not empty.

CHAPTER 10

Appendix

1. Some Elements of Topology

.

I - Stone-Weierstrass Theorem.

Here are two forms of Stone-Weierstrass Theorem. The second is more general and is the one we use in this text.

PROPOSITION 32. *Let (S, d) be a compact metric space and H a non-void subclass of the class $\mathcal{C}(S, \mathbb{R})$ of all real-valued continuous functions defined on S. Suppose that H satisfies the following conditions.*

(i) H is lattice, that is, for any couple (f, g) of elements of H, $f \wedge g$ et $f \vee g$ are in H

(ii) For any couple (x, y) of elements of S and for any couple (a, b) of real numbers such that $a = b$ if $x = y$, there exists a couple (h, k) of elements of H such that

$$h(x) = a \text{ and } k(y) = b.$$

Then H is dense in $\mathcal{C}(S, \mathbb{R})$ endowed with the uniform topology, that is each continuous function from S to $\mathbb{R}$ is the uniform limit of a sequence of elements in H.

THEOREM 30. *Let (S, d) be a compact metric space and H a non-void subclass of the class $\mathcal{C}(S, \mathbb{C})$ of all real-valued continuous functions defined on S. Suppose that H satisfies the following conditions.*

(i) H contains all the constant functions.

(ii) For all $(h, k) \in H^2$, $h + k \in H, h \times k \in H, \overline{u} \in H$.

(iii) H separates the points of S, i.e., for two distinct elements of S, x and y, that is $x \neq y$, there exists $h \in H$ such that

$$h(x) \neq h(y).$$

Then H is dense in $\mathcal{C}(S, \mathbb{C})$ endowed with the uniform topology, that is each continuous function from S to $\mathbb{C}$ is the uniform limit of a sequence of elements in H.

Remark.

If we work in $\mathbb{R}$, the condition on the conjugates - $\overline{u} \in H$ - becomes needless.

But here, these two classical versions do not apply. We use the following extension.

COROLLARY 4. *Let K be a non-singleton compact space and $\mathcal{A}$ be a non-empty sub-algebra of $C(K, \mathbb{C})$. Let $f \in C(K, \mathbb{C})$. Suppose that there exists $K_0 \subset K$ such that $k \setminus K_0$ has at least two elements and f is constant on K_0. Suppose that the following assumption hold.*
(1) $\mathcal{A}$ separates the points of $K \setminus K_0$ and separates any point of K_0 from any point of $K \setminus K_0$.

(2) A contains all the constant functions.

(3) For all $f \in \mathcal{A}$, its conjugate function $\bar{f} = \mathcal{R}(f) - i\mathcal{I}\Updownarrow(f) \in \mathcal{A}$,

Then

$$f \in \overline{\mathcal{A}}.$$

A proof if it available in Lo (2018b).

II- Approximations of indicator functions of open sets by Lipschitz function.

We have the

LEMMA 15. *Let (S, d) be an arbitrary metric space and G be an open set of S. Then there exists a non-decreasing sequence $(f_k)_{k \geq 1}$ of non-negative real-valued and Lipschitz functions defined on S converging to 1_G.*

Proof. Let G be an open set of S. For any integer number $k \geq 1$, set the function $f_k(x) = \min(kd(x, G^c), 1)$, $x \in S$. We may see that for any $k \geq 1$, f_m has values in $[0, 1]$, and is bounded. Since G^c is closed, we have

$$d(x, G^c) = \begin{cases} > 0 \text{ if } x \in G \\ 0 \text{ if } x \in G^c \end{cases}.$$

Let us show that f_k is a Lipschitz function. Let us handle $|f_k(x) - f_k(y)|$ through three cases.

Case 1. $(x, y) \in (G^c)^2$. Then

$$|f_k(x) - f_k(y)| = 0 \leq k\ d(x, y).$$

Case 2. $x \in G$ and $y \in G^c$ (including also the case where the roles of x and y are switched). We have

$$|f_k(x) - f_k(y)| = |\min(kd(x, G^c), 1)| \leq k\ d(x, G^c) \leq k\ d(x, y),$$

by the very definition of $d(x, G^c) = \inf\{d(x, z), \text{z} \in G^c\}$.

Case 3. $(x, y) \in G^2$. We use Lemma 16 in this section, to get

$$|f_k(x) - f_k(y)| = |\min(kd(x, G^c), 1) - \min(kd(y, G^c), 1)| \leq |kd(x, G^c) - kd(y, G^c)|,$$
$$\leq k\ d(x, y)$$

by the second triangle inequality. Then f_k is a Lipschitz function with coefficient k. Now, let us show that

$$f_k \uparrow 1_G \text{ as k} \uparrow \infty.$$

Indeed, if $x \in G^c$, we obviously have $f_k(x) = 0 \uparrow 0 = 1_G(x)$. If $x \in G$, that $d(x, G^c) > 0$ and $kd(x, G^c) \uparrow \infty$ as $k \uparrow \infty$. Then for k large enough,

$$f_k(x) = 1 \uparrow 1_G(x) = 1 \text{ as} k \uparrow \infty. \tag{1.1}$$

■.

III -Lipschitz property of finite maximum or minimum.

We have the

LEMMA 16. *For any real numbers x, y, X, and Y,*

$$|\min(x, y) - \min(X, Y)| \leq |x - X| + |y - Y|. \tag{1.2}$$

Proof. Let us have a look at the four possibles case.

Case 1 : $min(x, y) = x$ and $min(X, Y) = X$. We have

$$|\min(x, y) - \min(X, Y)| \leq |x - X|$$

Case 2 : $min(x, y) = x$ and $min(X, Y) = Y$. If $x \leq Y$, we have $Y \geq X$, we have

$$0 \leq \min(X, Y) - \min(x, y) = Y - x \leq X - x$$

If $x > Y$, we have $X \geq Y$, we have

$$0 \leq \min(x, y) - \min(X, Y) = x - Y \leq y - Y$$

Case 3 : $min(x, y) = y$ and $min(X, Y) = Y$. We have

$$|\min(x, y) - \min(X, Y)| \leq |y - Y|$$

Case 4 : $min(x, y) = y$ and $min(X, Y) = X$. This case id handled as for Case 2 by permuting the roles of (x, y) and (X, Y).

2. Orthogonal Matrices, Diagonalization of Real Symmetrical Matrices and Quadratic forms

I - Orthogonal matrices.

We begin by this result.

Proposition 33. *For any square d-matrix T, we have the equivalence between the following assertions.*

(1) T is invertible and the inverse matrix T^{-1} of T is its transpose matrix, that is

$$TT^t = T^tT = I_d,$$

where I_d is the identical matrix of dimension d.

(2) T is an isometry, that is T preserves the norm : For any $x \in \mathbb{R}^d$

$$\|Tx\| = \|x\|.$$

(3) The columns $\left(T^{(1)}, T^{(2)}, .., T^{(d)}\right)$ form an orthonormal basis of $\mathbb{R}^d$.

(4) The transposes of the lines $\left(T_1^t, ..., T_d^t\right)$ form an orthonormal basis of $\mathbb{R}^d$.

Besides, if T is orthogonal, its transpose is also orthogonal and satisfies

$$det(T) = \pm 1.$$

Before we give the proof, we provide the definition of an orthogonal matrix.

Definition 13.
A square d-matrix is orthogonal if and only if one of the equivalent assertions of Proposition 33 holds.

Now we may concentrate of the

Proof of Proposition 33.

Recall that, in finite dimension linear theory, the d-matrix B is the inverse of the d-matrix A if and only if $AB = I_d$ if and only if $BA = I_d$. (See the reminder at the end of the proof).

Let us show the following implications or equivalences.

(i) (1) $\Leftrightarrow$ (3). By definition, for any $(i, j) \in \{1, ..., d\}^d$,

$$\left(T^t T\right)_{ij} = \left(T^t\right)_i T^{(j)} = \left(T^{(j)}\right)^t T^{(j)} = \langle T^{(i)}, T^{(j)}\rangle. \ (I01)$$

and

$$\left(T T^t\right)_{ij} = T_i (T^t)_j = T_i (T_j)^t = \left(T_i^t\right)^t (T_j)^t = \langle T_i^t, T_j^t\rangle. \ (I02)$$

By Formula (I01), we have the equivalence between $T^t T = I_d$ and (3), and thus, (1) and (3) are equivalent.

(i) : (1) $\Leftrightarrow$ (4). The same conclusion is immediate by using Formuka (I02) instead of Formula (I01).

(ii) (3) $\Longleftrightarrow$ (2). We have for all $x \in \mathbb{R}^d$,

$$\begin{aligned} \|Tx\|^2 &= \langle Tx, Tx\rangle =^t (Tx)\,(Tx) =^t x^t T T x \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \left({}^t T T\right)_{ij} x_i x_j = \sum_{i=1}^{n}\sum_{j=1}^{n} \langle T^{(i)}, T^{(j)}\rangle x_i x_j. \ (IS01) \end{aligned}$$

Hence, (3) implies that for all $x \in \mathbb{R}^d$,

$$\|Tx\|^2 = \sum_{i=1}^{n} x_i^2 = \ \|x\|^2 ,$$

which is the definition of an isometry.

(iii) (2) $\Longleftrightarrow$ (3). Let us suppose that (2) holds.

To show that each $T^{(i_0)}$, for a fixed $i_0 \in \{1, ..., d\}$ is normed, we apply (ISO) to the vector x whose coordinates are zero except $x_{i_0} = 1$. We surely have $\|x\| = 1$ and all the terms of

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\langle T^{(i)}, T^{(j)}\rangle x_i x_j$$

are zero except for $i = j = i_0$, and the summation reduces to $\langle T^{(i_0)}, T^{(i_0)}\rangle x_{i_0}^2 = \langle T^{(i_0)}, T^{(i_0)}\rangle$. Equating the summation with $\|x\|^2$ gives that

$$\langle T^{(i_0)}, T^{(i_0)}\rangle = 1.$$

So, $T^{(i_0)}$ is normed.

To show that two different columns $T^{(i_0)}$ and $T^{(j_0)}$, for a fixed ordered pair $(i_0, j_0) \in \{1, ..., d\}^2$, are orthogonal, we apply (ISO) the vector x whose coordinates are zero except $x_{i_0} = 1$ and $x_{j_0} = 1$. We surely have $\|x\|^2 = 2$ and all the terms of

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\langle T^{(i)}, T^{(j)}\rangle x_i x_j$$

are zero except for $i = j = i_0$, $i = j = j_0$ and $(i, j) = (i_0, j_0)$, and the summation reduces to

$$\langle T^{(i_0)}, T^{(i_0)}\rangle x_{i_0}^2 + \langle T^{(j_0)}, T^{(j_0)}\rangle x_{i_0}^2 + 2\langle T^{(i_0)}, T^{(j_0)}\rangle x_{i_0} x_{j_0}^2$$

By equating with the summation with $\|x\|^2$, we get

$$2 = 2 + 2\langle T^{(i_0)}, T^{(j_0)}\rangle.$$

This implies that $\langle T^{(i_0)}, T^{(j_0)}\rangle = 0$.

We conclude that (2) holds whenever (3) does.

We obtained the following equivalences

$$\begin{array}{ccc} (1) & \Leftrightarrow & (3) \\ \Updownarrow & & \Updownarrow \\ (4) & & (2), \end{array}$$

from which we deive the equivalence between the four assertions.

It remains the two last points. That the transpose of T is orthogonal with T, is a direct consequence of the equivalence between assertions (3) and (4). Since a square matrix and its transpose have the same determinant and since $TT^t = I_d$, we get that $1 = det(TT^t) = det(T)det(T^t) = det(T)^2$. $\blacksquare$.

A useful reminder.

In finite dimension linear theory, the d-matrix B is the inverse of the d-matrix A if and only if $AB = I_d$ if and only if $BA = I_d$. But in an arbitrary algebraic structure $(E, \star)$ endowed with an internal operation $\star$ having a unit element e, that is an element of e satisfying $x \star e = e \star x = x$ for all $x \in E$, an inverse y of x should should fulfills : $x \star y = y \star x = e$. The definition may be restricted to $x \star y = e$ or to $y \star x = e$ if the operation e is commutative.

In the case of d-matrices, the operation is not commutative. So using only one of the two conditions $AB = I_d$ and $BA = I_d$ to define the inverse of a matrix A is an important result of linear algebra in finite dimensions.

II - Diagonalization of symmetrical matrices.

Statement and proof.

We have the important of theorem.

THEOREM 31. *For any real and symmetrical d-matrix A, there exists an orthogonal d-matrix T such that TAT^t is a diagonal matrix $diag(\delta_1, ..., \delta_d)$, that is*

$$TAT^t = diag(\delta_1, ..., \delta_d),$$

where δ_i, $1 \leq i \leq d$, are finite real numbers.

Remark. In other words, any real and symmetrical d-matrix A admits d real eigen-values (not necessarily distinct) δ_i, $1 \leq i \leq d$ and the passage matrix may be chosen to be an orthogonal matrix.

Proof. Let us suppose that A is symmetrical, which means that for any $u \in \mathbb{R}^d$, we have

$$\langle Au, v\rangle = \langle u, Av\rangle. \ (S)$$

.

In a first step, let us borrow tools from Analysis. The linear application $\mathbb{R}^d \ni u \to Au$ is continuous so that

$$\|A\| = \sup_{u\in\mathbb{R}^d,\ \|u\|\leq 1} \|Au\| = \sup_{u\in\mathbb{R}^d,\ \|X\|u1} \|Au\| < +\infty.$$

Since the closed ball is closed a compact set in $\mathbb{R}^d$, there exists, at least, u_0 such that $\|u_0\| = 1$ and

$$\sup_{\|u\|\leq 1} \|Au\| = \|Au_0\|.$$

In a second step, let us assume that A has two eigen-vectors u and v associated to two distinct real eigen-valeues μ and λ. By Formula (S) above, we have

$$\begin{aligned} &\mu\langle u, v\rangle = \lambda\langle u, v\rangle \\ \Rightarrow\ &(\mu - \lambda)\langle u, v\rangle = 0. \end{aligned}$$

We get that u and v are orthogonal. We get the rule : two eigen-vectors of a symmetric square matrix which are associated to two distinct real eigen-values are orthogonal.

In a third step, let us show that if a linear sub-space F in invariant by A, that is for all $u \in F$, $Au \in F$ (denoted $AF \subset F$), then the orthogonal $F^{\perp}$ of F is also invariant by A. Indeed, if F is A-invariant and $v \in F^{\perp}$, we have

$$\forall u \in F,\ \langle Av, u\rangle = \langle v, Au\rangle = 0;$$

since $Au \in F$.

Finally, in the last and fourth step, we have for $u \in \mathbb{R}^d$, such that $\|Au\| = 1$, by applying the Cauchy-Schwartz Inequality

$$\|Au\| =^2= \langle Au, Au\rangle = \langle u, A^2u\rangle \leq \|A^2u\|. \ (S1)$$

The equality is reached for some u_1 (with $\|u_1\| = 1$) only if u_1 and A^2u_1 are linearly dependent, that is exists λ such that

$$A^2u_1 = \lambda u_1,$$

meaning that u_1 is an eigen-vector of A^2 where, by taking the norms, we have

$$\lambda = \|A^2u_1\|.$$

Now, we have all the tools to solve the problem by induction. By definition, we have

$$\|A\|^2 = \|Au_0\|^2 \leq \|A^2u_0\| \leq \|A\|\|Au_0\| \leq \|A\|\|A\|\|u_0\|,$$

and hence Formula (S1) becomes an equality for u_0. The conclusion of the fourth step says that u_0 is an eigen-vector of A^2 associated to $\lambda = \|A\|^2$. For $\lambda = \mu^2$, this leads to $A^2u_0 = u^2u_0$, that is

$$(A - \mu I_d)(A + \mu I_d)u_0 = 0.$$

Now, either $(A + \mu I_d)u_0 = 0$ and u_0 is an eigen-vector of A associated to $-\mu$, or $v_0 = (A+\mu I_d)u_0 \neq 0$ and v_0 is an eigen-vector of A associated to μ. In both case, the eigen-value is $\pm\|A\|$.

We proved that A has at least on real eigen-vector we denote by e_1 associated to $\lambda_1 = \pm\|A\|$. In a next step, let us denote $F_1 = Lin(\{e_1\})$ and $G_2 = F_1^{\perp}$. It is clear that F_1 is invariant by A, so is G_2. We consider the restriction of A on G_2. We also have that A_2 symmetrical and clearly $\|A_2\| \leq \|A\|$. We find an eigen-vector e_2 of A_2, thus of A, associated with $\lambda_2 = \pm\|A_2\|$ and $|\lambda_1| \geq |\lambda_2|$ and e_1 and e_2 are orthogonal. We do the same for $F_2 = Lin(e_1, e_2)$, $G_3 = F_2^{\perp}$ and A_3 the restriction of A_2 (and hence of A) on G_3. We will find an eigen-vector e_3 of A_2, thus of A, associated with $\lambda_3 = \pm\|A_3\|$ and $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$ with $\{e_1, e_2, e_3\}$ orthonormal. We proceed similarly to get exactly d normed eigen-vectors orthogonal associated to a decreasing sequence of eigen-values in absolute values. ■

(b) Some consequences.

(b1) Determinant.

If $TAT^t = diag(\delta_1, ..., \delta_d)$, where T is orthogonal, we have

$$det(TAT^t) = det(T)det(T^t)det(A) = det(T)^2 det(A) = det(A)$$

and next

$$det(A) = det\biggl(diag(\delta_1, ..., \delta_d)\biggr)$$

which leads to

$$det(A) = \prod_1^d \delta_j.$$

(b2) A useful identity.

If $TAT^t = diag(\delta_1, ..., \delta_d)$, where T is orthogonal, we have, for any $(i, j) \in \{1, ..., d\}^2$,

$$\sum_{j=1}^{d} \delta_h \biggl(T^{(h)}(T^{(h)})^t\biggr)_{ij} = a_{ij}. \ (UID)$$

Proof. Let us denote $D = diag(\delta_1, ..., \delta_d)$ and suppose that $TAT^t = D$, T being orthogonal. We get A=T^tDT. Hence, for any $(i, j) \in \{1, ..., d\}^2$, we have

$$(A)_{ij} = \biggl(T^t D\biggr)_i T^{(j)}.$$

But the h elements, $1 \leq h \leq d$, of the line $\biggl(T^t A\biggr)_i$ are $\biggl(T^t\biggr)_i D^{(h))}$, which are

$$\sum_{1 \leq r \leq d} t_{ti} \delta_h \delta_{rh} = \delta_h t_{hi}.$$

Thus, we have

$$(A)_{ij} = \left(T^t D\right)_i T^{(j)} = \sum_{1\leq h \leq d} \delta_h t_{hi} t_{hj} = \sum_{1\leq h \leq d} \delta_h \left(T^{(h)} (T^{(h)}) T\right)_{ij}.$$

III - Elements from Bi-linear Forms and Quadratic Forms Theory.

Before we begin, let us remind Formula (ACBT), seen in the proof of (P5) in Points (b)-(b2) in Section 5.2 in Chapter 2 : for any $(p \times d)$-matrix A, any $(d \times s)$-matrix C and any $(q \times s)$-matrix B, the ij-element of ACB^t is given by

$$\sum_{1\leq k \ \leq s} \sum_{1\leq p \ \leq p} a_{ih} c_{hk} b_{jk}.$$

Let us apply this to vectors $u = A^t \in \mathbb{R}^d$, $v = B^t \in \mathbb{R}^k$ and to a matrix $(d \times k)$-matrix C. The unique element of the (1×1)-matrix $u^t C v$ is

$$u^t C v = \sum_{1\leq i \ \leq d} \sum_{1 \leq j \leq k} c_{ij} u_i v_j.$$

This formula plays a key role in bi-linear forms studies in finite dimensions.

If d-matrix C is diagonal, that is $C_{ij} = 0$ for $i \neq j$, we have

$$u^t C v = \sum_{1\leq i \ \leq d} \delta_j v_j^2. \quad (C0)$$

where $\delta_j = c_{jj}$, $j \in \{1, ..., d\}$. We may use the Kronecker's symbol defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}.$$

to get the following notation of a diagonal matrix. A d-diagonal matrix D whose diagonal elements are denoted by δ_j, $j \in \{1, ..., d\}$, respectively, may be written as follows :

$$D = diag(\delta_1, ..., \delta_2) = \left(\delta_i \delta_{ij}\right)_{1 \leq i,j \leq d}.$$

(a) - Bi-linear Forms.

By definition, a function

$$f \ : \ \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}$$

is bi-linear if and only if :

(i) for any fixed $u \in \mathbb{R}^d$, the partial application $v \mapsto f(u, v)$ is linear

and

(ii) for any fixed $v \in \mathbb{R}^k$, the partial application $u \mapsto f(u, v)$ is linear.

The link with matrices theory is the following. Let $(e_1, e_2, ...e_n)$ be an orthonormal basis of $\mathbb{R}^d$ and $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_k)$ an orthonormal basis of $\mathbb{R}^k$. Let us define the $(d \times k)$-matrix A by

$$a_{ij} = f(e_i, \varepsilon_j)$$

and denote the coordinates of $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^k$ in those bases by

$$u = \sum_{i=1}^{n} u_i e_i \text{ and } v = \sum_{j=1}^{m} v_j \varepsilon_j.$$

We have the following expression of the bi-linear form

$$f(u, v) = u^t A v.$$

The proof is the following :

$$\begin{aligned} f(u,v) &= f\left(\sum_{i=1}^{n} u_i e_i, \sum_{j=1}^{m} v_j \varepsilon_j\right) \\ &= \sum_{i=1}^{n} u_i f\left(e_i, \sum_{j=1}^{m} v_j \varepsilon_j\right) \\ &= \sum_{i=1}^{n} u_i \sum_{j=1}^{m} v_j f(e_i, \varepsilon_j) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} u_i a_{ij} v_j \end{aligned}$$

Thus, we may conclude with the help of Formula (uTCv) above.

(b) - Quadratic forms.

For any bi-linear form $f\ :\ (\mathbb{R}^d)^2 \rightarrow \mathbb{R}$, the mapping

$$x \ni \mathbb{R}^d \mapsto Q_f(x) = f(x, x)$$

is called the quadratic form associated with f.

The quadratic form is said to be **semi-positive** if and only if $Q_f(x) \geq 0$, for all $x \in \mathbb{R}^d$.

It is said to be **positive** if and only if $Q_f(x) > 0x$, for all $0 \neq x \in \mathbb{R}^d$.

We already know that f may be represented by a d-matrix A and thus, Q_f may be represented as

$$Q_f(u) = u^t A u,\ \ u \in \mathbb{R}^d.$$

But, since $Q_f(u)^t = u^t A^t u = Q_f(u)$, we also have for all $u \in \mathbb{R}^d$ that

$$Q_f(u) = u^t \frac{A + A^t}{2} u,\ \ u \in \mathbb{R}^d.$$

The matrix $B = (A + A^t)/1$ is symmetrical and we have

$$Q_f(u) = u^t B u,\ \ u \in \mathbb{R}^d,$$

which leads to the :

PROPOSITION 34. *Any quadratic form Q on $\mathbb{R}^d$ is of the form.*

$$Q(u) = u^t B u,\ \ u \in \mathbb{R}^d,$$

where B is a d-symmetrical form.

(c) - Canonical reduction of a Quadratic forms.

Reducing a quadratic form Q on $\mathbb{R}^d$ to a canonical form consists in finding an invertible linear change of variable $v = Tu = (v_1, ..., v_d)^d$ such that $Q_0(v) = Q(Tv)$ is of the form

$$Q_0(v) = \sum_{1 \leq j \leq d} \delta_j v_j^2.$$

This may be achieved in finite dimension in the following ways. Let B be a symmetrical matrix associated to the quadratic form B. According to Part II of this section, we can find an orthogonal matrix T such that TBT^t is a diagonal matrix $D = diag(\delta_1, ..., \delta_d)$. For $v = Tu = (v_1, ..., v_d)^d$, we have

$$Q(vT) = v^t(T^t AT)v = v^t Dv = \sum_{1 \leq i \leq d,\ 1 \leq j \leq d} d_{ij} v_i v_j = \sum_{1 \leq j \leq d} \delta_j v_j^2.$$

This leads to the

PROPOSITION 35. *Any quadratic form Q on $\mathbb{R}^d$ the form.*

$$Q(u) = u^t Bu,\ u \in \mathbb{R}^d,$$

where B is a d-symmetrical matrix, may be reduced to the canonical for

$$Q(u) = \sum_{1 \leq j \leq d} \delta_j v_j^2, \quad (CF)$$

where $v = Tu$ and the columns of T^t form an orthonormal basis of $\mathbb{R}^d$ and are eigen-vector of B respectively associated to the eigen-values δ_j, $1 \leq j \leq d$.

Consequences. From the canonical form (CF), we may draw the straightforward following facts based on the facts that T is invertible and its determinant is ± 1. Hence each element $u \in \mathbb{R}^d$ is of the from $u = Tv$. Hence, Formula (FC) holds for all $u \in \mathbb{R}^d$ with $u = Tv$.

(1) If all the eigen-values are non-negative, the quadratic form Q is semi-positive.

(2) If all the eigen-values are positive, the quadratic form Q is semi-positive.

(3) If the quadratic form Q is semi-positive and B is invertible or has a non-zero determinant, that it is positive.

Before we close the current section, let us remind that a canonical form as in (CF) is not unique. But the three numbers of positive terms $\eta_{(+)}$, of negative terms (η_-) and zero terms (η_0) are unique. The triplet $(\eta_{(-)}, \eta_{(0)}, \eta_{(+)})$ is called the signature of the quadratic form.

====

3. What should not be ignored on limits in $\overline{\mathbb{R}}$ - Exercises with Solutions

Definition $\ell \in \overline{\mathbb{R}}$ is an accumulation point of a sequence $(x_n)_{n\geq 0}$ of real numbers finite or infinite, in $\overline{\mathbb{R}}$, if and only if there exists a sub-sequence $(x_{n(k)})_{k\geq 0}$ of $(x_n)_{n\geq 0}$ such that $x_{n(k)}$ converges to ℓ, as $k \to +\infty$.

Exercise 1.

Set $y_n = \inf_{p\geq n} x_p$ and $z_n = \sup_{p\geq n} x_p$ for all $n \geq 0$. Show that :

(1) $\forall n \geq 0, y_n \leq x_n \leq z_n$.

(2) Justify the existence of the limit of y_n called limit inferior of the sequence $(x_n)_{n\geq 0}$, denoted by $\liminf x_n$ or $\underline{\lim}\ x_n$, and that it is equal to the following

$$\underline{\lim}\ x_n = \liminf x_n = \sup_{n\geq 0} \inf_{p\geq n} x_p.$$

(3) Justify the existence of the limit of z_n called limit superior of the sequence $(x_n)_{n\geq 0}$ denoted by $\limsup x_n$ or $\overline{\lim}\ x_n$, and that it is equal

$$\overline{\lim}\ x_n = \limsup x_n = \inf_{n\geq 0} \sup_{p\geq n} x_p x_p.$$

(4) Establish that

$$-\liminf x_n = \limsup(-x_n) \quad \text{and} \quad -\limsup x_n = \liminf(-x_n).$$

(5) Show that the limit superior is sub-additive and the limit inferior is super-additive, i.e. : for two sequences $(s_n)_{n\geq 0}$ and $(t_n)_{n\geq 0}$

$$\limsup(s_n + t_n) \leq \limsup s_n + \limsup t_n$$

and

$$\liminf(s_n + t_n) \geq \liminf s_n + \liminf t_n.$$

(6) Deduce from (1) that if

$$\liminf x_n = \limsup x_n,$$

then $(x_n)_{n\geq 0}$ has a limit and

$$\lim x_n = \liminf x_n = \limsup x_n$$

Exercise 2. Accumulation points of $(x_n)_{n\geq 0}$.

(a) Show that if ℓ_1=$\liminf x_n$ and $\ell_2 = \limsup x_n$ are accumulation points of $(x_n)_{n\geq 0}$. Show one case and deduce the second one and by using Point (3) of Exercise 1.

(b) Show that ℓ_1 is the smallest accumulation point of $(x_n)_{n\geq 0}$ and ℓ_2 is the biggest. (Similarly, show one case and deduce the second one and by using Point (3) of Exercise 1).

(c) Deduce from (a) that if $(x_n)_{n\geq 0}$ has a limit ℓ, then it is equal to the unique accumulation point and so,

$$\ell = \overline{\lim}\ x_n = \limsup x_n = \inf_{n\geq 0} \sup_{p\geq n} x_p.$$

(d) Combine this result with Point **(6)** of Exercise 1 to show that a sequence $(x_n)_{n\geq 0}$ of $\overline{\mathbb{R}}$ has a limit ℓ in $\overline{\mathbb{R}}$ if and only if $\liminf x_n = \limsup x_n$ and then

$$\ell = \lim x_n = \liminf x_n = \limsup x_n.$$

Exercise 3. Let $(x_n)_{n\geq 0}$ be a non-decreasing sequence of $\overline{\mathbb{R}}$. Study its limit superior and its limit inferior and deduce that

$$\lim x_n = \sup_{n\geq 0} x_n.$$

Deduce that for a non-increasing sequence $(x_n)_{n\geq 0}$ of $\overline{\mathbb{R}}$,

$$\lim x_n = \inf_{n\geq 0} x_n.$$

Exercise 4. (Convergence criteria)

Prohorov Criterion Let $(x_n)_{n\geq 0}$ be a sequence of $\overline{\mathbb{R}}$ and a real number $\ell \in \overline{\mathbb{R}}$ such that: Every subsequence of $(x_n)_{n\geq 0}$ also has a subsequence (that is a subssubsequence of $(x_n)_{n\geq 0}$) that converges to ℓ. Then, the limit of $(x_n)_{n\geq 0}$ exists and is equal ℓ.

Upcrossing or Downcrossing Criterion.

Let $(x_n)_{n\geq 0}$ be a sequence in $\overline{\mathbb{R}}$ and two real numbers a and b such that $a < b$. We define

$$\nu_1 = \begin{cases} \inf & \{n \geq 0, x_n < a\} \\ +\infty & \text{if } (\forall n \geq 0, x_n \geq a) \end{cases}.$$

If ν_1 is finite, let

$$\nu_2 = \begin{cases} \inf & \{n > \nu_1, x_n > b\} \\ +\infty & \text{if } (n > \nu_1, x_n \leq b) \end{cases}.$$

.

As long as the $\nu_j's$ are finite, we can define for $\nu_{2k-2} (k \geq 2)$

$$\nu_{2k-1} = \begin{cases} \inf & \{n > \nu_{2k-2}, x_n < a\} \\ +\infty & \text{if } (\forall n > \nu_{2k-2}, x_n \geq a) \end{cases}.$$

and for ν_{2k-1} finite,

$$\nu_{2k} = \begin{cases} \inf & \{n > \nu_{2k-1}, x_n > b\} \\ +\infty & \text{if } (n > \nu_{2k-1}, x_n \leq b) \end{cases}.$$

We stop once one ν_j is $+\infty$. If ν_{2j} is finite, then

$$x_{\nu_{2j}} - x_{\nu_{2j-1}} > b - a.$$

We then say : by that moving from $x_{\nu_{2j-1}}$ to $x_{\nu_{2j}}$, we have accomplished a crossing (toward the up) of the segment $[a, b]$ called *up-crossings*. Similarly, if one ν_{2j+1} is finite, then the segment $[x_{\nu_{2j}}, x_{\nu_{2j+1}}]$ is a crossing downward (down-crossing) of the segment $[a, b]$. Let

$D(a, b) =$ number of up-crossings of the sequence of the segment $[a, b]$.

(a) What is the value of $D(a,b)$ if ν_{2k} is finite and ν_{2k+1} infinite.

(b) What is the value of $D(a,b)$ if ν_{2k+1} is finite and ν_{2k+2} infinite.

(c) What is the value of $D(a,b)$ if all the $\nu_j's$ are finite.

(d) Show that $(x_n)_{n\geq 0}$ has a limit iff for all $a<b$, $D(a,b)<\infty$.

(e) Show that $(x_n)_{n\geq 0}$ has a limit iff for all $a<b$, $(a,b)\in\mathbb{Q}^2$, $D(a,b)<\infty$.

Exercise 5. (Cauchy Criterion). Let $(x_n)_{n\geq 0}$ $\mathbb{R}$ be a sequence of (**real numbers**).

(a) Show that if $(x_n)_{n\geq 0}$ is Cauchy, then it has a unique accumulation point $\ell\in\mathbb{R}$ which is its limit.

(b) Show that if a sequence $(x_n)_{n\geq 0}\subset\mathbb{R}$ converges to $\ell\in\mathbb{R}$, then, it is Cauchy.

(c) Deduce the Cauchy criterion for sequences of real numbers.

SOLUTIONS

Exercise 1.

Question (1). It is obvious that :

$$\inf_{p\geq n} x_p \leq x_n \leq \sup_{p\geq n} x_p,$$

since x_n is an element of $\{x_n, x_{n+1}, ...\}$ on which we take the supremum or the infimum.

Question (2). Let $y_n = \inf_{p\geq 0} x_p = \inf_{p\geq n} A_n$, where $A_n = \{x_n, x_{n+1}, ...\}$ is a non-increasing sequence of sets : $\forall n \geq 0$,

$$A_{n+1} \subset A_n.$$

So the infimum on A_n increases. If y_n increases in $\overline{\mathbb{R}}$, its limit is its upper bound, finite or infinite. So

$$y_n \nearrow \underline{\lim}\ x_n,$$

is a finite or infinite number.

Question (3). We also show that $z_n = \sup A_n$ decreases and $z_n \downarrow \overline{\lim}$ x_n.

Question (4) . We recall that

$$-\sup\{x, x \in A\} = \inf\{-x, x \in A\},$$

which we write

$$-\sup A = \inf(-A).$$

Thus,

$$-z_n = -\sup A_n = \inf(-A_n) = \inf\{-x_p, p \geq n\}.$$

The right hand term tends to $-\overline{\lim}\ x_n$ and the left hand to $\underline{\lim}(-x_n)$ and so

$$-\overline{\lim}\ x_n = \underline{\lim}\ (-x_n).$$

Similarly, we show:

$$-\underline{\lim}\ (x_n) = \overline{\lim}\ (-x_n).$$

Question (5). These properties come from the formulas, where $A \subseteq \mathbb{R}, B \subseteq \mathbb{R}$:

$$\sup\{x + y, A \subseteq \mathbb{R}, B \subseteq \mathbb{R}\} \leq \sup A + \sup B.$$

In fact :

$$\forall x \in \mathbb{R}, x \leq \sup A$$

and

$$\forall y \in \mathbb{R}, y \leq \sup B.$$

Thus

$$x + y \leq \sup A + \sup B,$$

where

$$\sup_{x \in A, y \in B} x + y \leq \sup A + \sup B.$$

Similarly,

$$\inf(A + B \geq \inf A + \inf B.$$

In fact :

$$\forall (x, y) \in A \times B, x \geq \inf A \text{ and } y \geq \inf B.$$

Thus

$$x + y \geq \inf A + \inf B,$$

and so

$$\inf_{x \in A, y \in B}(x + y) \geq \inf A + \inf B$$

Application.

$$\sup_{p \geq n} (x_p + y_p) \leq \sup_{p \geq n} x_p + \sup_{p \geq n} y_p.$$

All these sequences are non-increasing. By taking the infimum, we obtain the limits superior :

$$\overline{\lim}\ (x_n + y_n) \leq \overline{\lim}\ x_n + \overline{\lim}\ x_n.$$

Question (6). Set

$$\underline{\lim}\ x_n = \overline{\lim}\ x_n.$$

Since :

$$\forall x \geq 1,\ y_n \leq x_n \leq z_n,$$

$$y_n \to \underline{\lim}\ x_n$$

and

$$z_n \to \overline{\lim}\ x_n,$$

we apply the Sandwich Theorem to conclude that the limit of x_n exists and :

$$\lim\ x_n = \underline{\lim}\ x_n = \overline{\lim}\ x_n.$$

Exercice 2.

Question (a).

Thanks to Question (4) of Exercise 1, it suffices to show this property for one of the limits. Consider the limit superior and the three cases:

The case of a finite limit superior :

$$\underline{\lim} x_n = \ell \text{ finite.}$$

By definition,

$$z_n = \sup_{p \geq n} x_p \downarrow \ell.$$

So:

$$\forall \varepsilon > 0, \exists (N(\varepsilon) \geq 1), \forall p \geq N(\varepsilon), \ell - \varepsilon < x_p \leq \ell + \varepsilon.$$

Take less than that:

$$\forall \varepsilon > 0, \exists n_\varepsilon \geq 1 : \ell - \varepsilon < x_{n_\varepsilon} \leq \ell + \varepsilon.$$

We shall construct a sub-sequence converging to ℓ.

Let $\varepsilon = 1$:

$$\exists N_1 : \ell - 1 < x_{N_1} = \sup_{p \geq n} x_p \leq \ell + 1.$$

But if

$$z_{N_1} = \sup_{p \geq n} x_p > \ell - 1, \tag{3.1}$$

there surely exists an $n_1 \geq N_1$ such that

$$x_{n_1} > \ell - 1.$$

If not, we would have

$$(\forall p \geq N_1, x_p \leq \ell - 1\ ) \Longrightarrow \sup\{x_p, p \geq N_1\} = z_{N_1} \geq \ell - 1,$$

which is contradictory with (3.1). So, there exists $n_1 \geq N_1$ such that

$$\ell - 1 < x_{n_1} \leq \sup_{p \geq N_1} x_p \leq \ell - 1.$$

i.e.

$$\ell - 1 < x_{n_1} \leq \ell + 1.$$

We move to step $\varepsilon = \frac{1}{2}$ and we consider the sequence$(z_n)_{n\geq n_1}$ whose limit remains ℓ. So, there exists $N_2 > n_1$:

$$\ell - \frac{1}{2} < z_{N_2} \leq \ell - \frac{1}{2}.$$

We deduce like previously that $n_2 \geq N_2$ such that

$$\ell - \frac{1}{2} < x_{n_2} \leq \ell + \frac{1}{2}$$

with $n_2 \geq N_1 > n_1$.

Next, we set $\varepsilon = 1/3$, there will exist $N_3 > n_2$ such that

$$\ell - \frac{1}{3} < z_{N_3} \leq \ell - \frac{1}{3}$$

and we could find an $n_3 \geq N_3$ such that

$$\ell - \frac{1}{3} < x_{n_3} \leq \ell - \frac{1}{3}.$$

Step by step, we deduce the existence of $x_{n_1}, x_{n_2}, x_{n_3}, ..., x_{n_k}, ...$ with $n_1 < n_2 < n_3 < ... < n_k < n_{k+1} < ...$ such that

$$\forall k \geq 1, \ell - \frac{1}{k} < x_{n_k} \leq \ell - \frac{1}{k},$$

i.e.

$$|\ell - x_{n_k}| \leq \frac{1}{k},$$

which will imply:

$$x_{n_k} \to \ell$$

Conclusion : $(x_{n_k})_{k\geq 1}$ is very well a subsequence since $n_k < n_{k+1}$ for all $k \geq 1$ and it converges to ℓ, which is then an accumulation point.

Case of the limit superior equal $+\infty$:

$$\overline{\lim}\, x_n = +\infty.$$

Since $z_n \uparrow +\infty$, we have : $\forall k \geq 1, \exists N_k \geq 1,$

$$z_{N_k} \geq k + 1.$$

For $k = 1$, let $z_{N_1} = \inf_{p\geq N_1} x_p \geq 1 + 1 = 2$. So there exists

$$n_1 \geq N_1$$

such that :

$$x_{n_1} \geq 1.$$

For $k = 2$, consider the sequence $(z_n)_{n \geq n_1+1}$. We find in the same manner

$$n_2 \geq n_1 + 1$$

and

$$x_{n_2} \geq 2.$$

Step by step, we find for all $k \geq 3$, an $n_k \geq n_{k-1} + 1$ such that

$$x_{n_k} \geq k,$$

which leads to $x_{n_k} \to +\infty$ as $k \to +\infty$.

Case of the limit superior equal $-\infty$:

$$\overline{\lim} x_n = -\infty.$$

This implies : $\forall k \geq 1, \exists N_k \geq 1$, such that

$$z_{n_k} \leq -k.$$

For $k = 1$, there exists n_1 such that

$$z_{n_1} \leq -1.$$

But

$$x_{n_1} \leq z_{n_1} \leq -1.$$

Let $k = 2$. Consider $(z_n)_{n \geq n_1+1} \downarrow -\infty$. There will exist $n_2 \geq n_1 + 1$:

$$x_{n_2} \leq z_{n_2} \leq -2$$

Step by step, we find $n_{k1} < n_{k+1}$ in such a way that $x_{n_k} < -k$ for all k bigger than 1. So

$$x_{n_k} \to +\infty$$

Question (b).

Let ℓ be an accumulation point of $(x_n)_{n \geq 1}$, the limit of one of its sub-sequences $(x_{n_k})_{k \geq 1}$. We have

$$y_{n_k} = \inf_{p \geq n_k} x_p \leq x_{n_k} \leq \sup_{p \geq n_k} x_p = z_{n_k}.$$

The left hand side term is a sub-sequence of (y_n) tending to the limit inferior and the right hand side is a sub-sequence of (z_n) tending to the limit superior. So we will have:

$$\underline{\lim}\ x_n \leq \ell \leq \overline{\lim}\ x_n,$$

which shows that $\underline{\lim}\ x_n$ is the smallest accumulation point and $\overline{\lim}\ x_n$ is the largest.

Question (c). If the sequence $(x_n)_{n\geq 1}$ has a limit ℓ, it is the limit of all its sub-sequences, so subsequences tending to the limits superior and inferior. Which answers question (b).

Question (d). We answer this question by combining point (d) of this exercise and Point 6) of the Exercise 1.

Exercise 3. Let $(x_n)_{n\geq 0}$ be a non-decreasing sequence, we have:

$$z_n = \sup_{p\geq n} x_p = \sup_{p\geq 0} x_p, \forall n \geq 0.$$

Why? Because by increasingness,

$$\{x_p, p \geq 0\} = \{x_p, 0 \leq p \leq n-1\} \cup \{x_p, p \geq n\}.$$

Since all the elements of $\{x_p, 0 \leq p \leq n-1\}$ are smaller than than those of $\{x_p, p \geq n\}$, the supremum is achieved on $\{x_p, p \geq n\}$ and so

$$\ell = \sup_{p\geq 0} x_p = \sup_{p\geq n} x_p = z_n.$$

Thus

$$z_n = \ell \to \ell.$$

We also have $y_n = \inf\{x_p, 0 \leq p \leq n\} = x_n$, which is a non-decreasing sequence and so converges to $\ell = \sup_{p\geq 0} x_p$.

Exercise 4.

Let $\ell \in \overline{\mathbb{R}}$ having the indicated property. Let ℓ' be a given accumulation point.

$$(x_{n_k})_{k\geq 1} \subseteq (x_n)_{n\geq 0} \text{ such that } x_{n_K} \to \ell'.$$

By hypothesis this sub-sequence (x_{n_K}) has in turn a sub-sub-sequence $\left(x_{n_{(k(p))}}\right)_{p\geq 1}$ such that $x_{n_{(k(p))}} \to \ell$ as $p \to +\infty$.

But as a sub-sequence of $\left(x_{n_{(k)}}\right)$,

$$x_{n_{(k(\ell))}} \to \ell'.$$

Thus

$$\ell = \ell'.$$

Applying that to the limit superior and limit inferior, we have:

$$\overline{\lim}\, x_n = \underline{\lim}\, x_n = \ell.$$

And so $\lim x_n$ exists and equals ℓ.

Exercise 5.

Question (a). If ν_{2k} is finite and if ν_{2k+1} is infinite, then there ate exactly k up-crossings : $[x_{\nu_{2j-1}}, x_{\nu_{2j}}]$, $j = 1, ..., k$, that is, we have $D(a,b) = k$.

Question (b). If ν_{2k+1} is finite and ν_{2k+2} is infinite, then there are exactly k up-crossings: $[x_{\nu_{2j-1}}, x_{\nu_{2j}}]$, $j = 1, ..., k$, that is we have $D(a,b) = k$.

Question (c). If all the $\nu_j's$ are finite, then there are an infinite number of up-crossings : $[x_{\nu_{2j-1}}, x_{\nu_{2j}}]$, $j \geq 1k$: $D(a,b) = +\infty$.

Question (d). Suppose that there exist $a < b$ rationals such that $D(a,b) = +\infty$. Then all the $\nu_j's$ are finite. The subsequence $x_{\nu_{2j-1}}$ is strictly below a. So its limit inferior is below a. This limit inferior is an accumulation point of the sequence $(x_n)_{n\geq 1}$, so is more than $\underline{\lim}\, x_n$, which is below a.

Similarly, the subsequence $x_{\nu_{2j}}$ is strictly below b. So the limit superior is above a. This limit superior is an accumulation point of the sequence $(x_n)_{n\geq 1}$, so it is below $\overline{\lim}\, x_n$, which is directly above b. This leads to :

$$\underline{\lim}\, x_n \leq a < b \leq \overline{\lim}\, x_n.$$

That implies that the limit of (x_n) does not exist. In contrary, we just proved that the limit of (x_n) exists, meanwhile for all the real numbers a and b such that $a < b$, $D(a,b)$ is finite.

Now, suppose that the limit of (x_n) does not exist. Then,

$$\underline{\lim}\, x_n < \overline{\lim}\, x_n.$$

We can then find two rationals a and b such that $a < b$ and a number ϵ such that $0 < \epsilon$, such that

$$\underline{\lim}\, x_n < a - \epsilon < a < b < b + \epsilon < \overline{\lim}\, x_n.$$

If $\underline{\lim}\, x_n < a - \epsilon$, we can return to Question **(a)** of Exercise 2 and construct a sub-sequence of (x_n) which tends to $\underline{\lim}\, x_n$ while remaining below $a - \epsilon$. Similarly, if $b + \epsilon < \overline{\lim}\, x_n$, we can create a sub-sequence of (x_n) which tends to $\overline{\lim}\, x_n$ while staying above $b + \epsilon$. It is evident with these two sequences that we could define with these two sequences all ν_j finite and so $D(a, b) = +\infty$.

We have just shown by contradiction that if all the $D(a, b)$ are finite for all rationals a and b such that $a < b$, then, the limit of $(x_n)_{n \geq 0}$ exists.

Exercise 5. Cauchy criterion in $\mathbb{R}$.

Suppose that the sequence is Cauchy, *i.e.*,

$$\lim_{(p,q)\to(+\infty,+\infty)} (x_p - x_q) = 0.$$

Then let $x_{n_{k,1}}$ and $x_{n_{k,2}}$ be two sub-sequences converging respectively to $\ell_1 = \underline{\lim}\, x_n$ and $\ell_2 = \overline{\lim}\, x_n$. So

$$\lim_{(p,q)\to(+\infty,+\infty)} (x_{n_{p,1}} - x_{n_{q,2}}) = 0.$$

, By first letting $p \to +\infty$, we have

$$\lim_{q\to+\infty} \ell_1 - x_{n_{q,2}} = 0,$$

which shows that ℓ_1 is finite, else $\ell_1 - x_{n_{q,2}}$ would remain infinite and would not tend to 0. By interchanging the roles of p and q, we also have that ℓ_2 is finite.

Finally, by letting $q \to +\infty$, in the last equation, we obtain

$$\ell_1 = \underline{\lim}\, x_n = \overline{\lim}\, x_n = \ell_2.$$

which proves the existence of the finite limit of the sequence (x_n).

Now suppose that the finite limit ℓ of (x_n) exists. Then

$$\lim_{(p,q)\to(+\infty,+\infty)} (x_p - x_q) = \ell - \ell = 0,$$

which shows that the sequence is Cauchy.

Improper Riemann integral of an odd function on $\mathbb{R}$. Consider

$$f(x) = \frac{1}{\pi(1+x^2)}, \ x \in \mathbb{R}.$$

We have

$$\int_{-\infty}^{+\infty} \frac{1}{\pi(1+x^2)} \ dx = \int_{-\infty} +\infty d(\tan x) = \Big[\tan x \Big]_{-\infty}^{+\infty} = \pi.$$

Hence f is a *pdf*. Let X be a random variable associated to the *pdf* f. Set $g(x) = xf(x)$, $x \in \mathbb{R}$. Since g^+ and g^- are non-negative and locally bounded and Riemann integrable, we have

$$\int_{\mathbb{R}} g^+(x) d\lambda(x) = \int_{\mathbb{R}} g^-(x) d = +\infty \text{ and } \int_{\mathbb{R}} g^+(x) d\lambda(x) = \int_{\mathbb{R}} g^+(x) d = +\infty,$$

by using for example the D'Alembert criterion. Hence $\mathbb{E}(X)$ does not exist.

4. Important Lemmas when dealing with limits on limits in $\overline{\mathbb{R}}$

(1) - Cesaro generalized Limit.

The following result is often quoted as the Cesaro lemma.

LEMMA 17. *Let $(x_n)_{n\geq 1} \subset \mathbb{R}$ be a sequence of finite real numbers converging to $x \in \mathbb{R}$, then sequence of arithmetic means*

$$y_n = \frac{x_1 + ... + x_n}{n}, \ n \geq 1$$

S also converges to x.

Proof. Suppose that $(x_n)_{n\geq 1} \subset \mathbb{R}$ converge to $x \in \mathbb{R}$ as $n \to +\infty$. Fix $\varepsilon > 0$. Thus, there exists $N \geq 1$ such that $|x_n - x| < \varepsilon$ for all $n \geq N$. Now, for any $n \geq N$, we have

$$\begin{aligned}
|y_n - x| &= \left|\frac{x_1 + ... + x_n}{n} - \frac{y + ... + y}{n}\right| \\
&= \left|\frac{(x_1 - x) + ... + (x_n - x)}{n}\right| \\
&= \left|\frac{(x_1 - x) + ... + (x_N - x)}{n}\right| + \left|\frac{(x_{N+1} - x) + ... + (x_n - x)}{n}\right| \\
&= A_N + \left|\frac{(x_{N+1} - x) + ... + (x_n - x)}{n}\right|.
\end{aligned}$$

with $A_N = |(x_1 - x) + ... + (x_N - x)|$, which is constant with N. Hence for for any $n \geq N$, we have

$$\begin{aligned}
|y_n - x| < &\leq A_N + \frac{1}{n}(|x_{N+1} - x| + ... + |x_n - x|) \\
&\leq \frac{A_N}{n} + \frac{(n-N)}{n}\varepsilon
\end{aligned}$$

we conclude that, for all $\varepsilon > 0$,

$$\limsup_{n \to +\infty} |y_n - x| \leq \varepsilon,$$

and this achieves the proof. □

Remark. The limit of sequence of arithmetic means $(x_1 + ... + x_n)/n$, $n \geq 1$, may exists and that of $(x_n)_{n\geq 1}$ does not. In that sense the limit of the arithmetic means, whenever it exists, is called the *Cesaro generalized limit* of the sequence of $(x_n)_{n\geq 1}$.

(2) - Toeplitz Lemma. Let $(a_{n,k})_{(n\geq 1,\ 1\leq k\leq k(n))}$ be an array of real numbers such that

(i) For any fixed $k \geq 1$, $a_{n,k} \to 0$ as $n \to +\infty$,

(ii) there exists a finite real number c such that $\sup_{n\geq 1} \sum_{1\leq h\leq k(n)} |a_{n,k}| \leq c$.

Let $(x_n)_{n\geq 1}$ be a sequence of real number and define $y_n = \sum_{1\leq h\leq k(n)} x_k a_{n,k}$ and $b_n = \sup_{1\leq h\leq k(n)} a_{n,k}$, $n \geq 1$. We have the following facts.

(1) If $x_n \rightarrow 0$ as $n \rightarrow +\infty$, then $y_n \rightarrow 0$ as $n \rightarrow +\infty$.
(2) If $b_n \rightarrow 1$ and $x_n \rightarrow x \in \mathbb{R}$ as $n \rightarrow +\infty$, then $y_n \rightarrow x$ as $n \rightarrow +\infty$.

(3) Suppose that $k(n) = n$ for all $n \geq 1$. Let $(c_k)_{k\geq 0}$ be sequence such that the sequence $(b_n)_{n\geq 0} = (\sum_{1\leq k\leq n} |c_k|_{n\geq 0}$ is non-decreasing and $b_n \rightarrow \infty$. If $x_n \rightarrow x \in \mathbb{R}$ as $n \rightarrow +\infty$, then

$$\frac{1}{b_n} \sum_{1\leq k\leq n} c_k x_k \rightarrow x \quad as \quad n \rightarrow +\infty. \ \Diamond$$

Proof. All the convergence below are meant as $n \rightarrow +\infty$.

Proof of (1). Since x_n converges to 0, we can find for any fixed $\eta > 0$ a number $k_0 = k_0(\eta) > 0$ such that for any $k \geq k_0$, $|x_k| \leq \eta/c$ and (by this), we have for any $n \geq 0$

$$\begin{aligned} |y_n| &\leq \max(\sum_{1\leq h\leq k_0} |x_k||a_{n,k}|, \sum_{1\leq h\leq k_0} |x_k||a_{n,k}| + \sum_{k_0\leq h\leq k(n)} |x_k||a_{n,k}|) \ (L1) \\ &\leq \sum_{1\leq h\leq k_0} |x_k||a_{n,k}| + (\eta/c) \sum_{1\leq h\leq k(n)} |a_{n,k}| \\ &\leq \sum_{1\leq h\leq k_0} |x_k||a_{n,k}| + \eta, \end{aligned}$$

in short

$$|y_n| \leq \sum_{1\leq h\leq k_0} |x_k||a_{n,k}| + \eta.$$

In the Line (L1) above, it is not sure that $k(n)$ might exceed k_0, so we bound by the first argument of the *max* if $k(n) \leq k_0$. The the last equation, we let n go to infinity to have, for all $\eta > 0$,

$$\limsup_{n\rightarrow +\infty} |y_n| \leq \eta,$$

since that finite number of k_0 sequences $|x_k||a_{n,k}|$ (in n) converge to zero, which implies that y_n converges to zero.

Proof of 2. We have

$$y_n = x \sum_{1 \le h \le k_0} a_{n,k} + \sum_{1 \le h \le k_0} a_{n,k}(x_k - x)$$

which implies

$$|y_n - x| \le |x| |\sum_{1 \le h \le k_0} a_{n,k} - 1| + c|x_k - x|,$$

* which by the assumptions lead to $y_n \to 0$.

Proof of (3). By setting $a_{n,k} = c_k/b_n$, $1 \le kn$, we inherit the assumption is the former points with $c = 1$ and we may conclude by applying Point (2). □

(3) - Kronecker Lemma. If $(b_n)_{n \ge 0}$ is an increasing sequence of positive numbers and $(x_n)_{n \ge 0}$ is a sequence of finite real numbers such that $\left(\sum_{1 \le k \le n} x_k\right)_{n \ge 0}$ converges to a finite real number s, then

$$\frac{\sum_{1 \le k \le n} b_k x_k}{b_n} \to 0 \ as \ n \to \infty. \ \Diamond$$

Proof. Set $b_0 = 0$, $a_k = b_{k+1} - b_k$, $k \ge 0$, $s_1 = 0$, $s_{n+1} = x_1 + .. + s_n$, $n \ge 2$. We have

$$\begin{aligned} \frac{\sum_{1 \le k \le n} b_k x_k}{b_n} &= \frac{\sum_{1 \le k \le n} b_k (s_{k+1} - s_k)}{b_n} \\ &= s_{n+1} - \frac{1}{b_n} \sum_{1 \le k \le n} b_k s_k. \quad (L2) \end{aligned}$$

To see how to get Line, we just have to develop the summation and to make the needed factorizations as in

$$b_n(s_{n+1} - s_n) + b_{n-1}(s_n - s_{n-1}) + b_{n-2}(s_{n-1} - s_{n-2}) + \cdots b_3(s_4 - s_3) + b_2(s_3 - s_2) + b_1(s_2 -$$
$$b_n s_{n+1} - s_n)(b_n - b_{n-1}) - s_{n-1}(s_{n-1} - b_{n-2}) + \cdots s_2(b_2 - b_1) + s_1 b_1.$$

From Line (L2), we may apply Point (3) of the Toeplitz's Lemma above, since $b_n = a_1 + ... + a_n$, to conclude that the expression in Line (L2) converges to zero as $n \to +\infty$. □

5. Miscellaneous Results and facts

A - Technical formulas.

A1. We have for all $t > 0$, we have

$$\forall t \in \mathbb{R}_+,\ e^{t(1-t)} \leq 1 + t \leq e^t.\ \Diamond$$

Proof. Put $g(t) = (1+t) - e^t\ t \geq 0$. It is clear that $g'(t) = 1 - e^t$ is non-positive and hence g is non-decreasing on $\mathbb{R}_+$, and thus : for any $t \in \mathbb{R}_+$, $g(t) \leq g(0)$, which leads to the right-hand. To deal with the left-hand one, we put $g(t) = e^{t(t-1)} - t(t-1))$, $t \in \mathbb{R}$. The first two derivatives of g are

$$g'(t) = (-2t+1)e^{t(t-1)} - 1 \ \ and \ \ g''(t) = (4t^2 - 4t - 1)e^{t(t-1)}, \quad t \in \mathbb{R}.$$

The zeros of $4t^2 - 4t - 1)e^{t(t-1)}$ are $t_1 = (1-\sqrt{2})/2$ and $t_2 = (1+\sqrt{2})/2$. Since $t_1 \leq 0$, g'' is negative on $[0, t_2]$ vanishes on t_2 and positive on $]t_2,\ +\infty[$. This means that g' which vanishes at 0 and tends to -1 at $+\infty$, decreases on $[0, t_2]$, reaches its minimum at value at t_2 and increases to -1 on $]t_2,\ +\infty[$. So we have proved that g' is non-positive on $t \in \mathbb{R}_{+1}$. We conclude that for any $t \in \mathbb{R}_+$, $g(t) \leq g(0)$, which is the right-hand member of the inequality.

Acknowledgement. This proof is due to Cherif Mamadou Moctar Traor, University of Bamako, Mali.

6. Quick and powerfull algorithms for Gaussian probabilities

Visual Basic™ codes to compute $F(z) = \mathbb{P}(\mathcal{N}(0,1) \leq z)$.

```
Function ProbaNormale(z As Double) As Double
Dim a1 As Double, a2 As Double, a3 As Double, A4 As Double
Dim A5 As Double, w As Double, W1 As Double, P0 As Double

a1 = 0.31938153
a2 = -0.356563782
a3 = 1.781477937
A4 = -1.821255978
A5 = 1.330274429
```

```
W1 = Abs(z)
w = 1 / (1 + 0.2316419 * W1)
W1 = 0.39894228 * Exp(-0.5 * W1 * W1)
P0 = (a3 + w * (A4 + A5 * w))
P0 = w * (a1 + w * (a2 + w * P0))

P0 = W1 * P0

If z <= 0 Then
        P0 = 1 - P0
End If

ProbaNormale = 1 - P0
End Function
```

Quantile Function.

The quantile function or inverse function of $F(z) = \mathbb{P}(\mathcal{N}(0,1) \leq z)$ is computed by :

```
Public Function inverseLoiNormal(z As Double) As Double
Dim a1 As Double, a2 As Double, a3 As Double, A4 As Double, A5 As Dou
Dim A6 As Double
Dim W1 As Double, w As Double, W2 As Double, Q As Double

a1 = 2.515517: a2 = 0.802853: a3 = 0.010328
A4 = 1.432788: A5 = 0.189269: A6 = 0.001308

        If z <= 0 Then
                inverseLoiNormal = -4
                Exit Function
        ElseIf z >= 1 Then
    inverseLoiNormal = 4
    Exit Function
        End If

Q = 0.5 - Abs(z - 0.5)
w = Sqr(-2 * Log(Q))
W1 = a1 + w * (a2 + a3 * w): W2 = 1 + w * (A4 + w * (A5 + A6 * w))
```

```
inverseLoiNormal = (w - W1 / W2) * Sgn(z - 0.5)
End Function
```

Bibliography

Chidume C. 2014). Applicable Functional Analysis. Ibadan University Press, Nigeria. ISBN : 978-978-4856-31-5

Gustave Choquet (1966). *Topology : translated by Amiel Feinstein*. Academic Press. New-York.

Kai Lai Chung (1974). *A Course in Probability Theory*. Academic Press. New-York.

Patrick Billingsley (1995). *Probability and Measure*. Wiley. Third Edition.

Billingsley, P.(1968). *Convergence of Probability measures*. John Wiley, New-York.

Vladmit I. Bogachev (2007a). *Measure Theory I*. Springer.

Vladmir I. Bogachev (2007b). *Measure Theory II*. Springer.

Buchwalter Henri (1991). *Le calcul Intégral*. Ellipse. Paris.

Feller W.(1968) *An introduction to Probability Theory and its Applications. Volume I*. Third Editions. John Wiley & Sons Inc., New-York.

Feller W.(1968) *An introduction to Probability Theory and its Applications. Volume II*. Third Editions. John Wiley & Sons Inc., New-York.

Dominique Foata and Aimé Fuchs (1998). *Calcul de probabilités*. Dunod. Paris. Second Edition.

André Gramain (1994). *Intégration*. Collections Méthodes. Hermann.

Jean Genet (1976). *Mesure et Intégration : Théorie élémentaire*. Vuibert. Paris.

(2005). Allan Gutt (2005). *Probability Theory : a graduate course*. Springer.

Paul Richard Halmos*Measure Theory*. Springer.

Alaoui Aziz El Kacimi.(1999). *Eléments d'intégration et d'analyse fonctionnelle*. Ellipse. Paris.

Samuel Kotz, Cambell B. Read, N. Balakrishna, Brani Vidakovic (1999) *Encyclopedia of Statistical Sciences*. Wiley. ISBN: 9780471667193. DOI: 10.1002/0471667196.

Lo, G. S. (2017) A Course on Elementary Probability Theory. Gane Samb LO. SPAS Books Series. Saint-Louis, Senegal - Calgary, Canada. Doi : 10.16929/sbs/2016.0003. ISBN 978-2-9559183-3-3. ArXiv:1703.09648

Lo, G. S. (2017) Measure Theory and Integration By and For the Learner. SPAS Books Series. Saint-Louis, Senegal - Calgary, Canada. Doi : http://dx.doi.org/10.16929/sbs/2016.0005, ISBN : 978-2-9559183-5-7. Arxiv :

Lo, G.S.(2016). Weak Convergence (IA). Sequences of random vectors. SPAS Books Series. Saint-Louis, Senegal - Calgary, Canada. Doi : 10.16929/sbs/2016.0001. Arxiv : 1610.05415. ISBN : 978-2-9559183-1-9

Lo, G.S.(2016). Convergence vague (IA). Suites de vecteurs aléatoires. SPAS Books Series. Saint-Louis, SENEGAL - CANADA, Canada. Arxiv : 1611.03575.

LO G.S. (2018). A simple proof of the theorem of Sklar. *Journal of Mathematical Facts and Short papers* . Volume 1(1), pp 19-23 Doi : http://dx.doi.org/10.16929/jmfsp/19.003

Lo G.S.(2018). Journal of Mathematical Facts and Short Papers. Vol. 1 (1), 2017, pages 33-38. Doi :
exible application of the Stone-Weierstrass Theorem in General and Application in Probability Theory. 6

Michel Loève (1997). *Probability Theory I*. Springer Verlag. Fourth Edition.

Idika Eke OKORIE,Anthony Chukwudi AKPANTA,Johnson OHAKWE,David Chidi CHIKEZIE, (2017). The Kumaraswamy G Exponentiated Gumbel Type-2 Distribution. *Afrika Statistika*, Volume 12(3), pp 1367-1396 Doi : http://dx.doi.org/10.16929/as/2017.1367.109 (*projecteuclid.org*)

Charles-Michel Marle (1974). *Mesure et probabilités*. Hermann. Paris.

Michel Métivier. *Notions Pondamentales de Probabilités*. Dunod Université.

Nelsen, R.B. (2006). *An introduction to copula*. Springer-Verlag, New-York.

Newman C.M and Wright A.L. (1981). An invariance principle for certain dependent sequences. *Ann. probab.* **9**(4), 671-675.

Kalyanapuram Rangachari Parthasarathy (2005). *Introduction to Probability and Measure*. Hindustan Book Agency. India.

Resnick, S.I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New-York.

Daniel Revuz (1994). *Mesure et Intégration*. Hermann. Paris.

Stanislaw Sacks (1933). *Theory of the integral : A translation by of the original book published by Monografie Matematycze, Volume II, Varsaw 1933*. Hafner Publishing Company. New-York.

Sklar A.(2006) Fonctions de rpartition n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris. Vol. 8, pp 229-231

Jordan Stoyanov (1987). *Counterexamples in Probability*. Wiley.

Jun Shao (1987). *Mathematical Statistics*. Springer. Second Edition.

Taylor J.C.(1997) *Introduction to Measure and Probability*. Springer.

Valiron, G.(1946). *Théorie des Functions*. Masson. Paris.

Tómács, T. and Líbor, Z.(2006). A Hájék - Rényi type inequality and its applications. *Ann. Math. Inform.* **33**, 141 - 149. (MR2385473)

Hájék, J. and Rényi, A.(1955). Generalization of an inequality of Kolmogorov. *Acta Math. Acad. Sci. Hungar.* **6/3** - 4, 281 - 283. (MR0076207)

Stout, W.F.(1974). *Almost Sure Convergence*. Academic Press, Cambridge, MA (1974)