A Measure Theory and Integration Approach to Probability Theory, and Applications in Financial Markets (Black-Scholes model), and Actuarial Mathematics (Ruin Probability)

A Thesis Presented to the Department of Pure and Applied Mathematics, African University of Science and Technology.

By

Nwonu Blessing Uchenna

In Partial Fulfillment of the Requirements for the Award of Master of Science Degree in Pure and Applied Mathematics

Supervisor:
Prof. Gane Samb Lo

June 2021

Approval

Title of thesis

A Measure Theory and Integration Approach to Probability Theory, and Applications in Financial Markets (Black-Scholes model), and Actuarial Mathematics (Ruin Probability)

By

Nwonu Blessing Uchenna

A THESIS PRESENTED TO THE DEPARTMENT OF PURE AND APPLIED MATHEMATICS

RECOMMENDED: ==

Supervisor, Prof. Gane Samb Lo

2021/08/01

==

Head Department of Pure and Applied Mathematics

APPROVED: ==

Vice president, Prof. C.E. Chidume

Dedication

I dedicate this project to God Almighty my creator, source of inspiration, and understanding. I also dedicate this work to my parents, Mr & Mrs Emmanuel Nwonu, who has encouraged me all the way, and whose encouragement has made sure that I give it all it takes to finish that which I have started. To my lovely siblings, Precious, Ezekiel, Emmanuel, Favour, and Somto, who have been affected in every way possible by this quest, Thank you. My love for you all can never be quantified. God bless you.

Acknowledgments and Thanks

First and foremost, I would like to express my sincere gratitude to my esteemed supervisor, Prof. Gane Samb Lo, for his invaluable supervision, support and tutelage during the course of my Masters study and research, I would like to thank him for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

My gratitude extends to Prof. C.E Chidume for the funding opportunity to undertake my studies at the Department of Pure and Applied Mathematics, Africa University of Science and Technology. Additionally, I would love to express gratitude to Dr. Usman Abdulmalik for his treasured support throughout the course of my study in AUST, which was really influential in shaping my understanding of mathematical analysis. I also thank Prof. Yakubu Ochefu for his mentorship, and Prof. Joseph Ushie for his unwavering support.

Finally, I would like to thank my course mates and colleagues for a cherished time spent together in AUST, and in social settings. My appreciation goes out to my family, friends, and in particular, my uncle, Mr. Emeka Nwachukwu, for their encouragement and support all through my studies.

Contents

Introduction

The contents of this dissertation fall within the general area of Measure Theory and Integration (MTI), an area whose theoretical applications are numerous, diversified and almost everywhere present in all branches of mathematics. Its application in Probability Theory and their statistical applications seems to be more powerful with real-life applications. Everywhere the notion of random or chance is, MTI can help through probability theory, and later Statistics if estimations have to be dealt with. This implies that the core probability Theory is already encapsulated in the MTI course. At some point, probability theory is an oriented and restricted view of MTI.

The aim of the dissertation is to reformulate concepts of MTI in way that can be interpreted as calculus of chances around the following important parts:

- Probabilistic terminology of MTI concepts
- Use of the measure-images as probability laws and their properties
- Use of integrals as the mathematical expectations
- Use of product-measure as way of defining the probabilistic notions of independence and its characterizations
- Use of the Radon-Nikodym theorem to have a powerful foundation of the probabilistic density functions.
- Combination of the Radon-Nikodym theorem and the series as integral to characterize discrete probability laws
- Combination of the Radon-Nikodym theorem and the Lebesgue integral in $\mathbb{R}^k$ to characterize absolutely continuous probability laws

The exercise above will lead to two important points that build up in the core probability theory

CPT1 High level in dealing with Probability laws in specific spaces, in particular in $\mathbb{R}^k$, $k \geq 1$

CPT2 Characterizing usual Probability laws in $\mathbb{R}^k$.

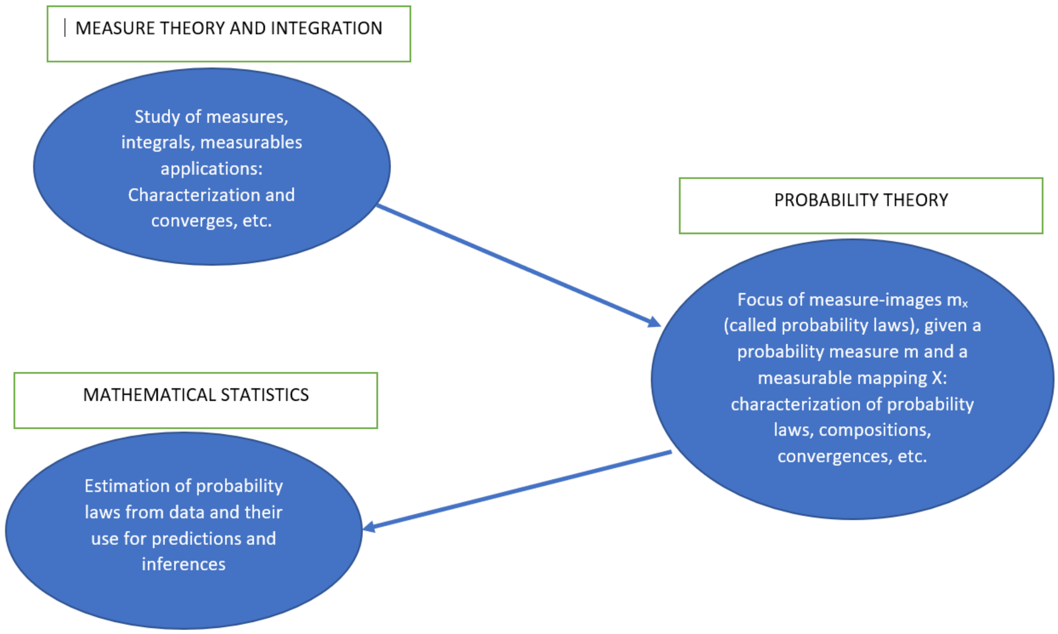

FIGURE 1. Measure theory and integration, probability theory, mathematical statistics

It would be possible to go further to cover advanced notions of

CPT3 Mathematical expectations and conditional probability laws (Chapter 8 in Lo (2018))

CPT4 Arbitrary product of probability spaces (Chapter 9 in Lo (2018)),

which provide the largest foundation to deal with probabilistic aspects in probability laws. According the quote (Gane Samb Lo):

> The essence of Probability Theory resides in probability laws, anything else is a matter of technicalities based on other branches of Mathematics,

achieving steps [CPT3] and [CPT4] is the final step to be able to be ready to cover all aspects of probability theory, mathematical statistics. Fig. 1 summarizes the main relations between these three areas.

But, due to time limitations, we are restricting ourselves on covering Stages [CPT1] and [CPT2] and to give applications on Financial markets and in Actuarial Sciences.

We organize the rest of the dissertation as follows. We cover the core part of probability theory in four chapters.

In Chapter 1, we proceed to a specialization of some contents of Measure Theory and Integration to normed measures. The probabilistic terminology is introduced. The main results in MTI are rephrased within Probability theory.

In Chapter 2, the main subject of Probability is described as the probability law of random variables. Properties and characterizations of probability laws are stated in general spaces. The important notion of independence, which is based on tensor products of measures and Fubini's Theorem, is introduced and characterized. In the specific case on $\Omega = \mathbb{R}^k$, particular characterizations of probability laws are described, among them: moments, cumulative distribution functions, probability distribution functions, characteristic and moment functions.

In Chapter 3, we proceed to a wide study of particular and usual probability laws on $\mathbb{R}$. The chapter stands as dictionary of usual laws. The cumulative distribution function, probability distribution function, the characteristic and moment functions are provided for each case.

In Chapter 4, we show that the covered part of probability laws is already enough to solve important real-life problems, for example in Financial Markets and in Actuarial Sciences.

In Financial Markets, we show that if the returns are independent and follow log-normal laws, we are able to derive the so-called Black-Sholes model which is used to determine the right price of an asset at time zero to avoid a loss at the striking time.

In Actuarial Sciences, the ruin problem is estimated. If the claims of the clients are exponential and the counting process of the claims times is Poisson, we are able to compute the exact ruin probability.

CHAPTER 1

From measure Theory and Integration to Probability Theory

We are going to summarize in Chapters 1 and 2 the contents of the same chapters in Lo (2018) without complete proofs.

1. Probabilistic Terminology

1.1. 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) $\mathbb{P} \geq 0$*.*

(b) $0 \leq \mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1$*,* $\forall A \in \mathcal{A}$*.*

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

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

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 experiment 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

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

1.2. 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

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

(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

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

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

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

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.

1.3. Random variables.

A random variable is a measurable function,

$$X : \quad (\Omega, \mathcal{A}, \mathbb{P}) \quad \to \quad (E, \mathcal{B}) \tag{1.1}$$

from a set of possible outcomes Ω, to a measurable space E. The probability that X takes on a value in a measurable set $S \subset E$ is written as

$$\mathbb{P}(X \in S) = \mathbb{P}(\{\omega : X(\omega) \in S\})$$

Probability law.

Suppose that we have a probability measure $\mathbb{P}$ on the measurable space $(\Omega, \mathcal{A})$ in Formula 1.1. We have the following definition.

DEFINITION 2. *The Probability law of the random variable X in Formula 1.1 is the image-measure of $\mathbb{P}$ by X, denoted as $\mathbb{P}_X$, which is a probability measure on E given by*

$$\begin{array}{llll}\mathbb{P}_X: & \mathcal{B} & \rightarrow & \mathbb{R}. \\ & B & \hookrightarrow & \mathbb{P}_X(B) = \mathbb{P}(X \in B)\end{array}$$

Classification of random variables.

Although the space E in Formula (1.1) 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.

1.4. Mathematical Expectation.

It is very important to note that, at the basic level, the mathematical expectation and later the conditional mathematical expectation, are 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 :

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

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

For an arbitrary random variable as defined in Formula (1.1) and for any real-valued measurable mapping

$$h : \quad (E, \mathcal{B}) \quad \rightarrow \quad (\overline{\mathbb{R}}, \mathcal{B}_\infty(\overline{\mathbb{R}}), \tag{1.2}$$

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, that we may compute the mathematical expectation of $h(X)$, if it exists, by

$$\mathbb{E}(h(X)) = \int_\Omega h \ d\mathbb{P}_X = \int_\Omega h(x) \ d\mathbb{P}_X(x). \tag{1.3}$$

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

$$\mathbb{E}(X) = \int_\mathbb{R} x \ d\mathbb{P}_X(x). \tag{1.4}$$

(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

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

1.5. Convergences of real-valued random variables.

(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

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

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

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

(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 $\varepsilon > 0$,

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

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

(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 $k \rightarrow +\infty$.

2. Independence

2.1. Independence of random variables.

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

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

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

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

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

$$\mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_n} \left(\prod_{1\leq i\leq n} A_i \right) = \prod_{1\leq i\leq n} \mathbb{P}_{X_i}(A_i). \tag{2.1}$$

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 later probability measure is called the joint probability measure.

By the λ-π Lemma, these two probability measures are equal whenever they agree on the semi-algebra $\mathcal{S}$.

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

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, ..., X_n \in B_n) = \prod_{1 \leq i \leq n} \mathbb{P}_{X_i}(B_i). \tag{2.2}$$

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 $A \in \mathcal{B}$ and B$\in \mathcal{G}$,

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

Formula (2.2) may be rephrased by means of measurable functions in place of measurable subsets. We have

THEOREM 1. *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*

$$\mathbb{E}\left(\prod_{1 \leq i \leq n} h_i(X_i) \right) = \prod_{1 \leq i \leq n} \mathbb{E}(h_i(X_i)). \tag{2.3}$$

2.2. Independence of events.

Independence of events is obtained from independence of random variables.

(a) Simple case of two events.

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*

$$\mathbb{P} =(AB) = \mathbb{P}(A) \times \mathbb{P}(B). \tag{2.4}$$

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

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,*

$$\mathbb{P}\left(\bigcap_{1\leq j\leq s} A_{i_j}\right) = \prod_{1\leq j\leq s} \mathbb{P}(A_{i_j}). \tag{2.5}$$

(c) An interesting remark.

A useful by-product of Formula (2.3) 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 may 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 2.3 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

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

2.3. 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 $h_i : F_i \rightarrow \mathbb{R}$, $1 \leq i \leq n$, 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

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

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

PROPOSITION 1. *Measurable transformations of independent random variables are independent*

2.4. Family of independent random variables. .

Consider a family of random variables

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

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.

CHAPTER 2

Finite dimensional Probability laws

1. Moments of Real Random Variables

(a) Definition of the moments.

Let X and Y be two *rrv*'s, let X_1, X_2, ... and Y_1, Y_2, $\cdots$ be finite sequences of *rrv*'s, and let α_1, α_2, ... and β_1, β_2, $\cdots$ are finite sequences of real numbers.

We 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 $m_1(X) = \mathbb{E}X$ exists and is finite.

(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

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

Note. $\mathbb{E}h(X)$ exists and is finite if and only if $\mathbb{E}|h(X)|$ is finite.

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

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

(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 $X = \mathbb{E}(X)$ $a.s.$

(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

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

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

(P6) If $X_1, ..., X_k$ 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.
(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) Let $\lambda > 0, \mathbb{V}ar(\lambda X) = \mathbb{E}((\lambda X)^2) - \mathbb{E}(\lambda X)^2 = \lambda^2(\mathbb{E}(X^2) - \mathbb{E}(X)^2) = \lambda^2\mathbb{V}ar(X)$

(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 2.3 implies that : $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. 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 $|\rho_{XY}| \leq 1$. $\square$

2. Cumulative distribution functions

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

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

$$\begin{array}{llll} \mathbb{F}_{\mathbb{X}} : & \mathbb{R} & \rightarrow & \mathbb{R} \\ & x & \hookrightarrow & F_X(x) = \mathbb{P}_X(]-\infty, x]) \end{array}$$

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

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

We 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

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

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

It has the following 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 $F(+\infty) = \lim_{x \to +\infty} F(x) = 1$.

(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 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 $\mathbb{R}$ and the class of ***cfd****'s $F_{\mathbb{P}}$ on $\mathbb{R}$ according to 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 real-valued random variables X and Y which have the same distribution function have the same probability law.

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

Given a (*cdf*), F, we can 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 of 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-Stieltjes measure $\mathbb{P}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R})$ defined by

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

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 (LS11), 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*.

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

Important Remarks. 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 $(Y_1, Y_2, \cdots, X_d)^t$ be another d-random vector and $(Z_1, Z_2, \cdots, Z_r)^t$ and $(T_1, T_2, \cdots, T_s)^t$ be two other random vectors of dimensions $r \geq 1$ and $s \geq 1$, all of them being defined on $(\Omega, \mathcal{A}, \mathbb{P})$.

(a) Variance-covariance and Covariance Matrices.

(a1) 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 another notation

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

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

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

(a2) Properties.

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, we have

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

(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 the computations work 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 (P3), we have

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

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

$$c_{ij} = \biggl((X_i - \mathbb{E}(X_i))(Z_j - \mathbb{E}(Z_j))\biggr),$$

$(i, j) \in \{1, ..., d\} \times \{1, ..., r\}$. 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 the i-th line of AC are $\{A_iC^1, A_iC^2, ..., A_iC^p\}$ 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

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

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

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

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

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

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

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

(b) Cumulative Distribution Functions.

(b1) Notion of Volume of cuboids by F.

Rule of forming $\Delta F(a,b)$.
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

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

or

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

where $s(\varepsilon)=\sum_{i=1}^{k}\varepsilon_i$

(b2) Cumulative Distribution Functions.

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 following 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)

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

and Condition(3-ii)

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

As we did in one dimension, we have :

Definition. A function $F : \mathbb{R}^d \rightarrow [0, 1]$ 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.

(b4) Joint *cdf*'s and marginal *cdf*'s.

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

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

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 $X^{(2)} = (X_{r+1}, ..., X_d)^t$ 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

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

THEOREM 2. *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) 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

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

(b5) 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 $(\overline{\mathbb{R}}^d, \mathcal{B}_\infty(\overline{\mathbb{R}}^d)$ given 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 :

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

4. 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 $\mathcal{V}_X = \{x^{(j)}, j \in J\}$, $\emptyset \neq J \subset \mathbb{N}$.

We already know from Measure Theory that X is measurable 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 $h : \mathcal{B}_\infty(\overline{\mathbb{R}}^d) \to \mathcal{B}_\infty(\overline{\mathbb{R}})$ with respect to the probability law $\mathbb{P}_X$ is performed through the *pdf* f_X in the Discrete Integral Formula

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

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

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

pdf's with respect to counting measures are referred to as *mass pdf's* or 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

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

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

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

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 Theorem for integrating a measurable function $h : \mathcal{B}_\infty(\overline{\mathbb{R}}^d) \to \mathcal{B}_\infty(\overline{\mathbb{R}})$ through the formula

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

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.

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 unique 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.

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

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x_1} d\lambda(t_1) \int_{-\infty}^{x_2} \\ & .... \int_{-\infty}^{x_{d-1}} d\lambda(t_{d-1}) \int_{-\infty}^{x_d} f_X(t_1, ..., t_d) d\lambda(t_d) \end{aligned}$$

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

(b4) Criterion for Absolute Continuity from the Cdf.

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

$$F_X(x) = \int_{-\infty}^{x_1} d\lambda(t_1) \int_{-\infty}^{x_2} d\lambda(t_2)...d\lambda(t_{d-1}) \int_{-\infty}^{x_k} f_X(t_1, ..., t_d)\ d\lambda(t_d).\ (AC01)$$

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

$$f_X(x_1, x_2, ..., x_k) = \frac{\partial^k F_X(x_1, x_2, ..., x_k)}{\partial x_1 \partial x_2 ... \partial x_k},\ \lambda_d - a.e.\ (AC02)$$

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

(b5) Marginal Probability Density functions.

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

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

For the extension to higher dimensions, 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 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

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

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

$$f_{X^{(2)}}(x) = \int_{\mathbb{R}^r} f_{(X_1,...,X_d)}(x_1, ..., x_d)\ d\biggl(\otimes_{1\leq i\leq r}\, m_i \biggr)(x_1, ..., x_r).$$

THEOREM 3. *Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ be a random vector. We 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, then 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 :*

$$\forall (x_1, ..., x_d) \in \mathbb{R}^d,\ f_{(X_1,...,X_d)}(x_1, ..., x_d) = \prod_{j=1}^{d} f_{X_i}(x_i),\ m.a.e.\ (DLM01)$$

5. Characteristic functions

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* $X = (X_1, \cdots, X_d)^t : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto \mathbb{R}^d$, *the function*

$$u = (u_1, \cdots, u_d)^t \mapsto \phi_X(u) = \mathbb{E} \exp\left(i\left(\sum_{j=1}^{d} X_j u_j\right)\right) = \mathbb{E}\left(\prod_{j=1}^{d} \exp\left(X_j u_j\right)\right),$$

written in a more compact form as

$$u \mapsto \phi_X(u) = \mathbb{E}(e^{i<X,u>}),$$

is called the characteristic function of the random vector 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

$$\mathbb{E}(e^{i<X,u>}) = \mathbb{E}(\cos < X, u >) + i\ \mathbb{E}(\sin < X, u >),$$

or

$$\mathbb{E}(e^{i<X,u>}) = \mathbb{E}\cos\left(\sum_{j=1}^{d} X_j u_j\right) + i\ \mathbb{E}\sin\left(\sum_{j=1}^{d} X_j u_j\right),$$

which is defined since the integrands for the real and imaginary parts are bounded.

(b) Moment Generating Function (*mgf*).

The following function

$$u \mapsto \varphi_X(u) = \mathbb{E}(e^{<X,u>}),\ u \in \mathbb{R}^d, \tag{5.1}$$

when defined on a domain D of $\mathbb{R}^d$ containing the null vector as an interior point, is called the first moment generating function (*mfg*) of X.

We may find the characteristic function by using the moment generating functions as follows :

$$\Phi_X(u) = \varphi_X(iu), u \in \mathbb{R}^d.$$

Main properties of the characteristic function.

THEOREM 4. *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. Then 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

$$\phi_{(X,Y)}(u,v) = \phi_X(u) \times \phi_Y(v). \tag{5.2}$$

Proof of Theorem 4.

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

$$< \left(\begin{array}{c} X \\ Y \end{array}\right), \left(\begin{array}{c} u \\ v \end{array}\right) >=< X, u > + < Y, v > .$$

Then

$$\begin{aligned} \phi_{(X,Y)}(u,v) &= E\left(\exp < \left(\begin{array}{c} X \\ Y \end{array}\right), \left(\begin{array}{c} u \\ v \end{array}\right) >\right) \\ &= \mathbb{E}\left(e^{<X,u>+<Y,v>}\right) = E(e^{<X,u>})\mathbb{E}(e^{<Y,v>}) \\ &= \phi_X(u) \times \phi_Y(v). \end{aligned}$$

Characterization of a probability law on $\mathbb{R}^d$ by its characteristic function.

THEOREM 5. *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 their probability laws coincide on $\mathcal{B}(\mathbb{R}^d)$, that is*

$$\Phi_X = \Phi_Y \Leftrightarrow \mathbb{P}_X = \mathbb{P}_Y.$$

A characterization of independence.

THEOREM 6. *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*

$$\phi_{(X,Y)}(u,v) = \phi_X(u) \times \phi_Y(v) \tag{5.3}$$

CHAPTER 3

A review of usual probabilty laws on $\mathbb{R}$

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 6 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 4 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 $\mathcal{V}_X = \{0, 1, ..., min(n, M)\}$:

$$\mathbb{P}(X = k) = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n}, \ k = 0, ..., \min(n, M).$$

Characteristic function of no use.

Mathematical expectation and variance :

$$\mathbb{E}(X) = rM/n, \ \text{and} \ V(X) = rM(n - M)(n - r)/\{n^2(n - 1)\}.$$

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

$$\mathbb{P}(X \in \mathcal{V}_X) = 1.$$

(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 on $\mathbb{R}_+$.

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 :

$$\mathbb{E}(X) = a/b, Var(X) = a/b^2.$$

(4) 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 :

$$\mathbb{E}(X) = a/(a+b) \text{ and } Var(X) = ab/[(a+b)^2(a+b+1)].$$

(5) 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 $\mathcal{V}_X = \mathbb{R}$:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-(x-m)^2/(2\sigma^2)), x \in \mathbb{R}.$$

Moment generating function :

$$\Phi_Y(u) = e^{um+\frac{\sigma^2 u^2}{2}}.$$

Moments of order $k \geq 1$.

$$\mathbb{X}\left(\frac{X-m}{\sigma^2}\right) = \frac{2^k\ k!}{(2k)}.$$

Mathematical expectation and variance :

$$\mathbb{E}(X) = m,\ Var(X) = \sigma^2.$$

(6) Log-Normal Random Variable with parameters $m \in \mathbb{R}$ and σ^2.

DEFINITION 7. *A random variable X is said to have a lognormal distribution with parameters m and σ^2 if $\log X$ has a normal distribution with parameters m and σ^2. We say $X \sim \mathcal{LN}(m, \sigma^2)$, and is supported by the positive real line $\mathcal{V}_X = \mathbb{R}_+$.*

Absolutely continuous probability density function on $\mathcal{V}_X = \mathbb{R}$:

Let $X \sim \mathcal{LN}(m, \sigma^2)$, and let $Y = \log X$. Then $Y \sim \mathcal{N}(m, \sigma^2)$, and,

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-(y-m)^2/(2\sigma^2)), y \in \mathbb{R}.$$

Now,

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(\log X \leq \log x) \\ &= \mathbb{P}(Y \leq \log x) \\ &= F_Y(\log x) \end{aligned}$$

$$\begin{aligned}
\Rightarrow F_X(x) &= F_Y(\log x)\\
\Rightarrow F'_X(x) &= \frac{1}{x}F'_Y(\log x)\\
\Rightarrow f_X(x) &= \frac{1}{x}f_Y(\log x)\\
&= \frac{1}{x}\frac{1}{\sigma\sqrt{2\pi}}\exp(-(\log x - m)^2/(2\sigma^2))\\
&= \frac{1}{\sigma\sqrt{2\pi}}x^{-1}\exp(-(\log x - m)^2/(2\sigma^2))\\
\Rightarrow f_X(x) &= \frac{1}{\sigma\sqrt{2\pi}}x^{-1}\exp(-(\log x - m)^2/(2\sigma^2)).
\end{aligned}$$

Moments of order $t \geq 1$:

Let $X \sim \mathcal{LN}(m, \sigma^2)$, and let $Y = \log X$ i.e $X = e^Y$, and $X^t = e^{tY}$. We have that,

$$\Phi_Y(u) = e^{um+\frac{\sigma^2 u^2}{2}}.$$

$X^t = e^{tY}$ implies:

$$\begin{aligned}
\mathbb{E}(X^t) &= \mathbb{E}(e^{tY})\\
&= \mathbb{E}(e^{tY})\\
&= \Phi_Y(t)\\
&= e^{tm+\frac{\sigma^2 t^2}{2}}.
\end{aligned}$$

This implies that $\mathbb{E}(X^t) = e^{tm+\frac{\sigma^2 t^2}{2}}$, for $t \in \mathbb{R}$.

Mathematical expectation and variance :

$$\mathbb{E}(X) = e^{m+\frac{\sigma^2}{2}}$$

$$
\begin{aligned}
Var(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
&= e^{2m+\frac{4\sigma^2}{2}} - e^{2m+\sigma^2} \\
&= e^{2m+\sigma^2}(e^{\sigma^2} - 1)
\end{aligned}
$$

CHAPTER 4

Applications of simple probability laws in Financial Markets and Actuarial sciences

1. Pricing in financial Markets using lognormal laws

In this section, we consider the seminal result of Black and Scholes (1973) that allows to set the pricing of a stock in a financial market. The original paper of these authors was derived in a very descriptive method.

Following Turner (2010) for example, we may place the problem of the two authors in very clear probabilistic approach, set correct assumptions of the market (assumption of no free risk money for instance), on the probability laws and the independence of the data and recover that powerful formula.

By doing this, the result could be adapted to to different probability laws and eventually to certain dependence structures of the data.

1.1. Recalls on European Financial Markets. **Definition 1:** Options are derivative contracts that gives the holder the right but not the obligation to buy or sell an underlying asset at a pre determined price before the option expires. This right comes with a premium.

The two most common types of options are calls and put options.

Definition 2: A call option is a contract between two parties to exchange stock at a strike price by a predetermined date. One part, the buyer of the Call, has the right but not an obligation to buy the stock at the strike price by the future date, while the other party, the seller has the obligation to sell the stock to the buyer at the strike price if the buyer exercises the option.

The more the underline asset is percieved to appreciate the higher the premium demanded by the market for the call opton.

Definition 3: A put option is a contract between two parties to exchange stock at a strike price by a predetermined date. One part, the buyer of the Puts, has the right but not an obligation to sell the stock at the strike price by the future date, while the other party, the seller has the obligation to buy the stock from the buyer at the strike price if the buyer exercises the option.

The higher the percieved risk, the higher the premium demanded by the market for the put option.

Definition 4: A European option is simply an option that can only be exercised at the expiry of the option, which is specified in the contract.

Definition 5: The forward price of a stock is the current price of the stock S_0 plus an expected return, which will exactly offset the cost of holding the stock over a period of time t.

$$F = S_0 e^{rt}$$

where,
F = Forward price
S_0 = The Underlying asset current price
t = The delivery date in years
r = risk free interest rate that applies to the life of the forward contract

1.2. Free risk markets. The notion of *fairness* is important in any financial frame as in gambling, financial markets, etc. The rule should be that no-one can force a gain on his own, The market should be led by pure random behavior. In fair gambling, the portfolio should follow a martingale in order for the players and the casino owner have the same chance of winning.

In a financial market, *fairness* is described by the risk neutral notion. According to that principle, no one can can gain money mechanically, without facing a risk. The money obtained without risk is called *free money*. Let us give an example of free market.

Suppose that our market include two places : Abuja and Dakar, that is we may buy and sell in both places. Suppose some asset A is sold at $100naira$ and $141CFA$ in Abuja and Dakar's Market respectively, and at the same time $1Naira$ is exchanged with $1.31FCA$. We may borrow $1000Naira$, get ten assets A in Abuja, and convert our portfolio to $1310 CFA$, and sell our 10 assets A at 1410. Without any risk, we gain 100 CFA as free money. If free-money can be obtained, we say that there is an arbitrage opportunity.

In Financial markets, we may be interested by two activities:

(1) Modeling the value of the stock S_t at a time $t > 0$, specially at the striking time T. This activity is called *hedging*.

(2) Determining the price of the option at time 0, $C(S, 0)$.

Now, let us suppose that by placing S_0 at the bank at rate $r > 0$ and the same amount at risk. The amount placed at fixed rate at the bank is non-random and is

$$V_t = S_0 e^{rt}$$

The risky placement provide a random wealth S_t at time t. Finding or estimating the probability law of S_t or of the return $r_t = S_t/S_{t-1}$ or the log-return

$$\log(S_t/S_{t-1})$$

is important. Later, to derive the original Black and Sholes formula, we will use the log-normal model for r_t.

In a *fair market*, with no arbitrage opportunities, the two parts of the portfolios S_t and V_t (with the same amount placed) should behave such that $S_t - V_t$ may be positive (i.e. the risky placement is more advantageous that the non-risky one) or $S_t - V_t$ is negative (i.e. the risky placement is less advantageous that the non-risky one). But because of the fairness, those gains and loss should neutralize in mean, that is

$$\mathbb{E}S_t = V_t. \tag{1.1}$$

statistically, we should be able to observe

$$\frac{1}{t}\sum_{j=1}^{t}(S_t - V_t) \approx 0$$

for large values of t.

The same analysis can be done in the reverse case. For a no risky placement, if we get to have V_T a time T we have to place $V_T e^{-rT}$ at time 0. The same fairness principle says that the discount process $S_T V_T e^{-rT}$ should be equal, in average, to the price P_0 at which the asset should be sold, that is

$$P_0 = \mathbb{E}(S_T V_T e^{-rT}). \tag{1.2}$$

Definition 6: A universe is said to be risk neutral if for all asset A, and time period t, the value of an asset at time $t = 0$ $C(A, 0)$ is the expected value of the asset at time t discounted to the present value using the risk free rate.

$$C(A, 0) = \mathbb{E}(C(A, t))e^{-rt}$$

r is the continuously compounded risk free interest rate.

Definition 7: An In The Money (ITM) option is one with a strike price that has already been surpassed by the current stock price. An Out of The Money (OTM) option is one that has a strike price that the underlying security has yet to reach, meaning the option has no intrinsic value.

Definition 8: An option premium is the current market price of an option contract. It is thus the income received by the seller (writer) of an option contract to another party. It is composed of two parts, The intrinsic value and the time value (extrinsic value).

Total value of an option = Intrinsic value of the option + Time premium of the option

Factors determining the value of an option

1. Intrinsic Value of an option: The intrinsic value of an option is the value the option would have if exercised immediately. Basically, the intrinsic value is the amount by which the strike price of an option is profitable or in-the-money as compared to the stock's price in the market.

Intrinsic value = max{Strike price - current price, 0}

2. Time value of an option (Extrinsic Value): Time value refers to the portion of an option's premium that is attributable to the amount of time remaining until the expiration of the option contract. For example, when a stock is selling for $60 a share, its call option with exercise price, $55 is selling for $8. Then the intrinsic value of the call is $5 and the time value $3. For another option priced at $3 with stock price $79 and exercise price $80, the intrinsic value is zero, and hence the time premium is $3.

3. Volatility: Volatility often refers to the amount of uncertainty or risk related to the size of changes in a security's value. A higher volatility means that a security's value can potentially be spread out over a larger range of values. This means that the price of the security can change dramatically over a short time period in either direction. A lower volatility means that a security's value does not fluctuate dramatically, and tends to be more steady. Typically, stocks with high volatility have a higher probability for the option to be profitable or in-the-money by expiry. As a result, the time value as a component of the option's premium,is typically higher to compensate for the increased chance that the stock's price could move beyond the strike price and expire in-the-money. For stocks that are not expected to move much, the option's time value will be relatively low.

1.3. The Black and Sholes Pricing Formula.

LEMMA 1. *Let S_0 be the initial value of the stock price, S_t be the price at time t, and denote by σ the annual volatility in the percent change in the stock price, i.e., the standard deviation of the percent change in the price over one year. Finally, assume S_t is a log-normally distributed random variable, i.e. $\log S_t/S_0$ is normally distributed with mean μ and variance v, and let the mean of the log-normal distribution be located at the forward price of the stock. Then,*

$$v(t) = \sigma^2 t \tag{1.3}$$

and,

$$\mu(t) = \left(r - \frac{\sigma^2}{2}\right) t \tag{1.4}$$

Proof

We prove equation 1.3 by the using mathematical induction. At $t = 1$, $v(1) = \sigma^2$, That is $\mathbb{V}ar(\ln \frac{S_1}{S_0}) = \sigma^2(1)$ (After one year $\ln \frac{S_1}{S_0}$ has variance σ^2. This also implies that $\mathbb{V}ar(\ln \frac{S_t}{S_{t-1}}) = \sigma^2 \ \forall t \geq 1$). If we assume that $\mathbb{V}ar(\ln \frac{S_{t-1}}{S_0}) = \sigma^2(t-1)$, i.e., After $t-1$ years, $\ln \frac{S_{t-1}}{S_0}$ will have variance $\sigma^2(t-1)$, then after t years,

$$\begin{aligned}
\mathbb{V}ar\left(\ln \frac{S_t}{S_0}\right) &= \mathbb{V}ar\left(\ln \frac{S_t S_{t-1}}{S_0 S_{t-1}}\right) \\
&= \mathbb{V}ar\left(\ln \frac{S_t S_{t-1}}{S_{t-1} S_0}\right) \\
&= \mathbb{V}ar\left(\ln \frac{S_t}{S_{t-1}} + \ln \frac{S_{t-1}}{S_0}\right) \\
&= \mathbb{V}ar\left(\ln \frac{S_t}{S_{t-1}}\right) + \mathbb{V}ar\left(\ln \frac{S_{t-1}}{S_0}\right) () \\
&= \sigma^2 + \sigma^2(t-1) \\
&= \sigma^2 t
\end{aligned}$$

Therefore $\ln \frac{S_t}{S_0}$ will have variance $\sigma^2 t$

Important remark: For simplicity of notation, In the rest of this document, we will denote $v(t)$ and $\mu(t)$ as just v and μ respectively.

Equation 1.4 results from the following:

$$\begin{aligned}
F_{S_t}(x) &= \mathbb{P}(S_t \leq x) \\
&= \mathbb{P}\left(\frac{S_t}{S_0} \leq \frac{x}{S_0}\right) \\
&= \mathbb{P}\left(\ln\frac{S_t}{S_0} \leq \ln\frac{x}{S_0}\right) \\
&= \mathbb{P}\left(X_t \leq \ln\frac{x}{S_0}\right), (X_t = \ln\frac{S_t}{S_0}) \\
&= F_{X_t}\left(\ln\frac{x}{S_0}\right)
\end{aligned}$$

This implies $F_{S_t}(x) = F_{X_t}(\ln\frac{x}{S_0})$, and $F'_{S_t}(x) = F'_{X_t}(\ln\frac{x}{S_0})\frac{1}{x}$. Then we have that $f_{S_t}(x) = \frac{1}{x} f_{X_t}(\ln\frac{x}{S_0}) = \frac{1}{\sqrt{2\pi v}x} e^{-\frac{\left(\ln\frac{x}{S_0}-\mu\right)^2}{2v}}$

Hence,

$$f_{S_t}(x) = \frac{1}{\sqrt{2\pi v}x} e^{-\frac{\left(\ln\frac{x}{S_0}-\mu\right)^2}{2v}}$$

Now,

$$\begin{aligned}
\mathbb{E}(S_t) &= \int_0^\infty x f_{S_t}(x)dx \\
&= \int_0^\infty x\frac{1}{\sqrt{2\pi v}x} e^{-\frac{\left(\ln\frac{x}{S_0}-\mu\right)^2}{2v}} dx \\
&= \frac{1}{\sqrt{2\pi v}} \int_0^\infty e^{-\frac{\left(\ln\frac{x}{S_0}-\mu\right)^2}{2v}} dx
\end{aligned}$$

Let $z = \frac{\ln\frac{x}{S_0}-\mu}{\sqrt{v}}$, then $dz = \frac{dx}{x\sqrt{v}}$, with $x = S_0 e^{z\sqrt{v}+\mu}$. So that,

$$\begin{aligned}
\mathbb{E}(S_t) &= \frac{1}{\sqrt{2\pi v}} \int_{-\infty}^{\infty} S_0 e^{z\sqrt{v}+\mu} \sqrt{v} e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} S_0 e^{z\sqrt{v}+\mu} e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} S_0 e^{z\sqrt{v}+\mu-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} S_0 e^{\frac{-1}{2}(z^2-2z\sqrt{v}+v)+\frac{v}{2}+\mu} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} S_0 e^{\frac{-1}{2}(z-\sqrt{v})^2+\frac{v}{2}+\mu} dz \\
&= \frac{S_0 e^{\frac{v}{2}+\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-1}{2}(z-\sqrt{v})^2} dz
\end{aligned}$$

Let $x = z - \sqrt{v}$, then $dx = dz$, and we see that,

$$\begin{aligned}
\mathbb{E}(S_t) &= \frac{S_0 e^{\frac{v}{2}+\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-1}{2}(x)^2} dx \qquad (1.5) \\
&= \frac{S_0 e^{\frac{v}{2}+\mu}}{\sqrt{2\pi}} \sqrt{2\pi} \\
&= S_0 e^{\frac{v}{2}+\mu} \qquad (1.6)
\end{aligned}$$

So, we have

$$\mathbb{E}(S_t) = S_0 e^{\frac{v}{2}+\mu}.$$

But by assumption, $\mathbb{E}(S_t) = S_0 e^{rt}$, where r is the risk free rate. By using Formula (1.1), we get, $S_0 e^{\frac{v}{2}+\mu} = S_0 e^{rt}$ and, $\mu = \left(rt - \frac{v}{2}\right)$. Hence, $\mu(t) = \left(r - \frac{\sigma^2}{2}\right) t$ since $v(t) = \sigma^2 t$.

THEOREM 7. ***(Black Scholes)*** *In a risk-neutral universe with an initial stock price S_0 and a log-normally distributed stock price S_t, as in the Lemma 1, at time t, the value C of a European call option at time $t = 0$ with strike K, and expiration time T, and r being the continuously compounded risk-free rate is:*

$$C = S_0 \mathbb{N}\left(\frac{rT + \frac{\sigma^2 T}{2} + \ln\frac{S_0}{K}}{\sigma\sqrt{T}}\right) - Ke^{rT}\mathbb{N}\left(\frac{rT - \frac{\sigma^2 T}{2} + \ln\frac{S_0}{K}}{\sigma\sqrt{T}}\right) \tag{1.7}$$

where $\mathbb{N}$ is the cumulative distribution function of the standard normal variable.

Proof. We have $C(S,T) = max(S_T - K, 0)$. By assumption,

$$\begin{aligned} C = C(S,0) &= \mathbb{E}(C(S,T))e^{rT} \\ &= \mathbb{E}(max(S_T - K, 0))e^{rT} \\ &= e^{rT}\int_K^\infty (x-K) f_{S_t}(x)dx \\ &= e^{rT}\left(\int_K^\infty x f_{S_t}(x)dx - \int_K^\infty K f_{S_t}(x)dx\right) \\ &= e^{rT}\left(\int_K^\infty \frac{x}{\sqrt{2\pi v}x} e^{-\frac{\left(\ln\frac{x}{S_0} - \mu\right)^2}{2v}} dx - \int_K^\infty \frac{K}{\sqrt{2\pi v}x} e^{-\frac{\left(\ln\frac{x}{S_0} - \mu\right)^2}{2v}} dx\right) \end{aligned}$$

So,

$$C = e^{rT}\left(\int_K^\infty x\frac{1}{\sqrt{2\pi v}x} e^{-\frac{\left(\ln\frac{x}{S_0} - \mu\right)^2}{2v}} dx - \int_K^\infty K\frac{1}{\sqrt{2\pi v}x} e^{-\frac{\left(\ln\frac{x}{S_0} - \mu\right)^2}{2v}}\right). \tag{1.8}$$

Evaluating the first integral in equation 1.8, we have:

$$\int_K^\infty x\frac{1}{\sqrt{2\pi v}x} e^{-\frac{\left(\ln\frac{x}{S_0} - \mu\right)^2}{2v}} dx = \int_K^\infty x\frac{1}{\sqrt{2\pi T}\sigma x} e^{-\frac{\left(\ln\frac{x}{S_0} - \mu\right)^2}{2\sigma^2 T}} dx.$$

Let $z = \frac{\ln\frac{x}{S_0} - \mu}{\sigma\sqrt{T}}$, then $dz = \frac{dx}{x\sigma\sqrt{T}}$, with $x = S_0 e^{z\sigma\sqrt{T} + \mu}$. and let $A_0 = \frac{\ln\frac{K}{S_0} - \mu}{\sigma\sqrt{T}}$
So that,

$$\begin{aligned}
\int_K^\infty \frac{1}{\sqrt{2\pi T}\sigma} e^{-\frac{\left(\ln \frac{x}{S_0} - \mu\right)^2}{2\sigma^2 T}} dx &= \frac{1}{\sigma\sqrt{2\pi T}} \int_A^\infty S_0 e^{z\sigma\sqrt{T}+\mu} \sigma\sqrt{T} e^{\frac{-z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_A^\infty S_0 e^{z\sigma\sqrt{T}+\mu} e^{\frac{-z^2}{2}} dz \\
&= \frac{S_0}{\sqrt{2\pi}} \int_A^\infty e^{z\sigma\sqrt{T}+\mu+\frac{-z^2}{2}} dz \\
&= \frac{S_0}{\sqrt{2\pi}} \int_A^\infty e^{\frac{-1}{2}(z-\sigma\sqrt{T})^2+\frac{\sigma^2 T}{2}+\mu} dz \\
&= \frac{S_0}{\sqrt{2\pi}} e^{\frac{\sigma^2 T}{2}+\mu} \int_A^\infty e^{\frac{-1}{2}(z-\sigma\sqrt{T})^2} dz
\end{aligned}$$

Let $x = z - \sigma\sqrt{T}$, then $dx = dz$. Let $A_1 = A_0 - \sigma\sqrt{T} = \frac{\ln \frac{K}{S_0} - rT - \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}}$ and $d_1 = (-A_1) = frac \ln \frac{S_0}{K} + rT + \frac{\sigma^2 T}{2} \sigma\sqrt{T}$ and we see that,

$$\begin{aligned}
\frac{S_0}{\sqrt{2\pi}} e^{\frac{\sigma^2 T}{2}+\mu} \int_{A_0}^\infty e^{\frac{-1}{2}(z-\sigma\sqrt{T})^2} dz &= \frac{S_0}{\sqrt{2\pi}} e^{\frac{\sigma^2 T}{2}+\mu} \int_{A_1}^\infty e^{\frac{-x^2}{2}} dx \\
&= S_0 e^{\frac{\sigma^2 T}{2}+\mu} \int_{A_1}^\infty \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \\
&= S_0 e^{\frac{\sigma^2 T}{2}+\mu} (1 - \mathbb{N}(A_1)) \\
&= S_0 e^{rT} \mathbb{N}(d_1) \qquad (1.9)
\end{aligned}$$

And Evaluating the Second integral in equation 1.8, we have:

$$\int_K^\infty K \frac{1}{\sqrt{2\pi v}x} e^{-\frac{\left(\ln \frac{x}{S_0} - \mu\right)^2}{2v}} = K \int_K^\infty \frac{1}{\sqrt{2\pi T}\sigma x} e^{-\frac{\left(\ln \frac{x}{S_0} - \mu\right)^2}{2\sigma^2 T}}$$

Let $z = \frac{\ln \frac{x}{S_0} - \mu}{\sigma\sqrt{T}}$, then $dz = \frac{dx}{x\sigma\sqrt{T}}$. Let $A_0 = \frac{\ln \frac{K}{S_0} - \mu}{\sigma\sqrt{T}}$, and $d_2 = -A = \left(\frac{rT - \frac{\sigma^2 T}{2} + \ln \frac{S_0}{K}}{\sigma\sqrt{T}}\right)$ So that,

$$
\begin{aligned}
K\int_K^\infty \frac{1}{\sqrt{2\pi T}\sigma x} e^{-\frac{\left(\ln\frac{x}{S_0}-\mu\right)^2}{2\sigma^2 T}}dx &= \int_{A_0}^\infty \frac{K}{\sigma\sqrt{2\pi T}}\sigma\sqrt{T}e^{\frac{-z^2}{2}}dz \\
&= K\int_{A_0}^\infty \frac{e^{\frac{-z^2}{2}}}{\sqrt{2\pi}}dz \\
&= K\mathbb{N}(-A_0) \\
&= K\mathbb{N}(d_2) \qquad (1.10)
\end{aligned}
$$

From 1.9 and 1.10, we have:

$$
\begin{aligned}
C &= e^{-rT}(S_0e^{rT}(\mathbb{N}(d_1)) - K\mathbb{N}(d_2)) \\
&= e^{-rT}S_0e^{-rT}(\mathbb{N}(d_1)) - e^{-rT}K\mathbb{N}(d_2) \\
&= S_0\mathbb{N}(d_1) - Ke^{-rT}\mathbb{N}(d_2) \\
&= S_0\mathbb{N}\left(\frac{rT+\frac{\sigma^2T}{2}+\ln\frac{S_0}{K}}{\sigma\sqrt{T}}\right) - Ke^{-rT}\mathbb{N}\left(\frac{rT-\frac{\sigma^2T}{2}+\ln\frac{S_0}{K}}{\sigma\sqrt{T}}\right)
\end{aligned}
$$

2. Estimating Ruin probability in Actuarial Sciences using exponential laws

In this section, we follow ideas in Paulsen (1998), Paulsen (2008) but essentially Grandell (1991).

2.1. Introduction. Let us begin by some notations. Customers signing insurance policies with an insurer are insured against a precise undesired event (sickness, cars accidents, etc.). For each specified event, there is a product. Although a company offers a number of product, we suppose for simplicity's sake that only one product is running. Let us define the following objects.

The original investment of the insurer is denoted by $u \geq 0$.

The costumer pays a premium $c > 0$ at the signature of the prime.

The claim X, is the amount (in Naira) to be paid by the insurer if the undesired event occurs. It is a random variable of *cdf* F.
The random number of occurrences of claims payments up to time $t \geq 0$ is denoted by $N(t) = N_t$.

At time $t \geq 0$, the claims to be paid are denoted : $X_1, \cdots, X_{N(t)}$.

The times at which the claims are to be paid are denoted : $Z_1 < Z_2$, so that

$$N(t) = \sum_{j \geq 0} 1_{Z_j \leq t}, \ t \geq 0.$$

Given these notation, the financial balance of the company at time $t > 0$ is given by

$$Y_t = Y(t) = u + ct - \sum_{j=1}^{N(t)} X_j, \ t \geq 0, \tag{2.1}$$

In this thesis, the Risk process is given by

$$Y_t^* = Y^*(t) = ct - \sum_{j=1}^{N(t)} X_j, \ t \geq 0, \tag{2.2}$$

and the balance is given by

$$Y_t = u + Y_t^* \; t \geq 0, \tag{2.3}$$

We have the following definitions.

Definition 1. Insolvency/ruin is a state of financial distress in which a person or business is unable to pay their debts. We have ruin if and only if the financial balance of the company is negative at some time. i.e., if:

$$\exists t > 0, \; u + Y_t^* < 0.$$

Definition 2. The ruin time $T_{ruin}(u)$ is defined by

$$T_{ruin}(u) = \begin{cases} +\infty & \text{if} \quad : \forall t \geq 0, \; u + Y^*(T_u) \geq 0 \\ \inf\{t \geq 0, \; t \geq 0, u + Y^*(T_u) < 0\} & \text{if not} \end{cases}.$$

Note: $T_{ruin}(u) = \infty$ iff the set $\{t \geq 0, \; t \geq 0, \; u + Y^*(T_u) < 0\} = \emptyset$, and we set that $inf\emptyset = +\infty$.

Definition 3. The probability that the insurer's surplus level eventually falls below zero (making the firm bankrupt) is called the probability of ultimate ruin, $P_{ruin}(u)$, and it is defined by:

$$P_{ruin}(u) = \mathbb{P}(\exists t > 0, \; u + Y_t^* < 0),$$

or

$$P_{ruin}(u) = \mathbb{P}(T_{ruin} < +\infty).$$

Our main task is to compute or approximate the probability of ruin, $P_{ruin}(u)$. We will proceed to our analysis by using the following hypotheses.

(HP1) All the random objects are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

(HP2) The sequence of claims $(X_n \geq 0)_{n \geq 0}$ is independent from the counting process $(N(t))_{t \geq 0}$, based on the arrival times $(Z_n \geq 0)_{n \geq 0}$.

(HP3) $(X_n \geq 0)_{n\geq 0}$ is a sequence of independent and identically distributed non-negative random variables with *cdf* F and *mgf* φ and finite mathematical expectation $\mu > 0$.

(HP4) $(Z_n \geq 0)_{n\geq 0}$ are the arrival times of a Poisson process of intensity $\lambda > 0$, that is the inter-arrival times $(Z_n - Z_{n-1} \geq 0)_{n\geq 1}$ are independent exponential $\mathcal{E}(\lambda)$ random variables, and consequently, $(N(t))_{t\geq 0}$ is the counting process of that Poisson process.
(HP4) The safety load

$$\rho = \frac{c}{\lambda\mu} - 1 \geq 0.$$

These hypothesis are classical in Acutarial Science. Now we are going to determine or approximate the ruin probability.

2.2. Integral Formula and Applications. Sometimes, computing the non-ruin probability

$$P_{noruin}(u) = 1 - P_{ruin}(u) = \mathbb{P}(T_{ruin}(u) = +\infty)$$

is easier to compute. So we will use below.

Here are our main result.

THEOREM 8. *Suppose that the hypothesis (HP1)-(HP4) holds and that the claim follow a cdf F. Then the non-ruin probability $\Phi(u)$, given (T_1, X_1), is determinied as follows:*

$$\Phi(u) = \Phi(0) + \frac{\lambda}{c}\int_0^u \Phi(u-x)\,[1-F(x)]\,dx. \tag{2.4}$$

Proof of Theorem 8.

We let T_1 be the time of the first claim. Given $(T_1, X_1) = (t, x)$, the process starts again from $T_1 = t$. Then the initial capital which we know to be u, will then be $cT_1 - X_1$. By conditioning on (T_1, X_1), we get

$$\Phi(u) = \mathbb{E}(Probability\ of\ nonruin/(T_1, X_1)) = \int \Phi(u + ct - x)\ dF_{(T_1,X_1)}.$$

We recall that T_1 and X_1 are independent by assumption. So we have

$$dF_{(T_1,X_1)}(t,x) = dF_{T_1}(t)dF_{X_1)}(x).$$

We also recall that T_1 follows an exponential law $\mathcal{E}(\lambda)$, that is

$$dF_{T_1}(t) = \lambda e^{-\lambda t}\ dt,\ \ t \geq 0.$$

We finally have, by using Fubini's Theorem

$$\Phi(u) = \int\int \lambda e^{-\lambda t} 1_{t\geq 0} \Phi(u + ct - x)\ dF_{X_1)}(x)dt.$$

We have ruin at the first claim if $u + ct - x \geq 0$, that is if $u + ct \leq x$. Since we are studying the non-ruin event, we surely have $u + ct \geq x$. This combined with the fact that X_1 is a non-negative random variable, give the following bounds in Equation (2.5) which becomes

$$\Phi(u) = \int_0^{+\infty} \int_0^{u+ct} \lambda e^{-\lambda t} \Phi(u + ct - x) dF_{X_1)}(x)dt.$$

where $F = F_{X_1}$ is the common *cdf* of the claims X_i. By Applying Fubini's Theorem, we repeatedly integrate and get

$$\Phi(u) = \int_0^{+\infty} \lambda e^{-\lambda t} \left(\int_0^{u+ct} \Phi(u + ct - x) dF(x) \right) dt.$$

The change of variable $z = u + ct$ implies $dt = dz/c$ leads to

$$\Phi(u) = \frac{\lambda}{c} e^{\lambda u/c} \int_u^{\infty} e^{-\lambda z/c} \int_0^z \Phi(z - x) dF(x) dz.$$

we now differentiate $\Phi(u)$ using the quotient rule we have

(2.5)

$$\Phi'(u) = \frac{\lambda}{c}\cdot\frac{\lambda}{c}e^{\lambda u/c}\int_u^\infty e^{-\lambda z/c}\int_0^z \Phi(z-x)dF(x)dz - \frac{\lambda}{c}e^{\lambda u/c}\cdot e^{-\lambda u/c}\int_0^u \Phi(u-x)dF(x).$$

By formula (2.5)we have

$$\Phi'(u) = \frac{\lambda}{c}\cdot\Phi(u) - \frac{\lambda}{c}\int_0^u \Phi(u-x)dF(x). \tag{2.6}$$

We replace dF(x)=-d(1-F(x)) and integrating by parts we have

$$\Phi'(u) = \frac{\lambda}{c}\cdot\Phi(u) + \frac{\lambda}{c}\int_0^u \Phi(u-x)d(1-F(x)).$$

$$\Phi'(u) = \frac{\lambda}{c}\cdot\Phi(u) + \frac{\lambda}{c}\left[\Phi(0)(1-F(u)) - \Phi(u)\right] + \lambda/c\int_0^u \Phi'(u-x)(1-F(x))dx.$$

$$\Phi'(u) = \frac{\lambda}{c}\cdot\Phi(0)(1-F(u)) + \frac{\lambda}{c}\int_0^u \Phi'(u-x)(1-F(x))dx.$$

Integrating over the limits (0,t) gives

$$\Phi(t) - \Phi(0) = \frac{\lambda}{c}\Phi(0)\int_0^t (1-F(u))du + \frac{\lambda}{c}\int_0^t\int_0^u \Phi'(u-x)(1-F(x))dxdu.$$

$$\Phi(t) - \Phi(0) = \frac{\lambda}{c}\Phi(0)\int_0^t (1-F(u))du + \frac{\lambda}{c}\int_0^t (1-F(x))\int_x^t \Phi'(u-x)(1-F(x))dxdu.$$

$$= \frac{\lambda}{c}\Phi(0)\int_0^t (1-F(u))du + \frac{\lambda}{c}\int_0^t (1-F(x))\left[\Phi(t-x) - \Phi(0)\right]dx.$$

$$= \frac{\lambda}{c}\int_0^t (1-F(x))\Phi(t-x)dx$$

Finally we can write

$$\Phi(u) = \Phi(0) + \frac{\lambda}{c}\int_0^u \Phi(u-x)\left[1-F(x)\right]dx \tag{2.7}$$

The proof is finished. ■

Solving that integral equation is rarely easily. In general, the ruin probabability is approximated or simply bounded. But in the case where F follow an exponential law, an explicit formula can be obtained.

2.3. Application to exponential laws $\mathcal{E}(\lambda)$.

THEOREM 9. *Suppose that the hypothesis (HP1)-(HP4) holds and that the claim follow a exponential law of parameter $\mathcal{E}(\lambda)$. Then the non-ruin probability $\Phi(u)$ is explicitely given by:*

$$\Phi(u) = 1 - \frac{1}{1+\rho} e^{-\frac{\rho u}{\mu(1+\rho)}}. \tag{2.8}$$

Proof of Theorem9. We use the equation

$$\Phi'(u) = \frac{\lambda}{c} \cdot \Phi(u) - \frac{\lambda}{c} \int_0^u \Phi(u-x) dF(x).$$

Since the claims are exponentially distributed, we can replace

$$dF(x) = \frac{1}{\mu} e^{-x/\mu} dx$$

to obtain

$$\Phi'(u) = \frac{\lambda}{c} \cdot \Phi(u) - \frac{\lambda}{c\mu} \int_0^u \Phi(u-x) e^{-x/\mu} dx.$$

Change of variables we let $z = u - x$, $dx = -dz$ gives

$$\Phi'(u) = \frac{\lambda}{c} \cdot \Phi(u) + \frac{\lambda}{c\mu} \int_u^0 \Phi(z) e^{-(u-z)/\mu} dz.$$

$$\Phi'(u) = \frac{\lambda}{c} \cdot \Phi(u) - \frac{\lambda}{c\mu} e^{(-u)/(\mu)} \int_0^u \Phi(z) e^{-(z)/(\mu)} dz.$$

Differentiating by u leads to

$$\Phi''(u) = \frac{\lambda}{c} \cdot \Phi'(u) - \left[\frac{-\lambda}{c\mu^2} e^{\frac{-u}{\mu}} \int_0^u \Phi(z) e^{\frac{z}{\mu}} dz + \frac{-\lambda}{c\mu} e^{\frac{-u}{\mu}} \cdot \Phi(u) e^{\frac{u}{\mu}} \right]$$

Simplifying further, gives

$$\Phi''(u) = \left(\frac{\lambda}{c} - \frac{1}{\mu}\right) \cdot \Phi'(u) = -\frac{\rho}{\mu(1+\rho)}\Phi'(u).$$

From here we have

$((\ln \Phi(u)))' = \frac{\Phi''(u)}{\Phi'(u)} = -\frac{\rho}{\mu(1+\rho)}$ thus

$$(\ln \Phi'(u)) = -\frac{\rho u}{\mu(1+\rho)} + C_1$$

From which

$$\Phi'(u) = C_2 e^{-\frac{\rho u}{\mu(1+\rho)}}$$

This leads to

$$\Phi(u) = C_3 e^{-\frac{\rho u}{\mu(1+\rho)}} + C_4$$

The constants C_3 and C_4 are defined by conditions $\Phi(\infty) = 1$ (giving $C_4 = 1$) and $\Phi(0) = 1 - 1/(1+\rho)$ (giving $C_3 = -1/(1+\rho)$ Therefore,the non ruin probability for exponentially distribution claim is.

$$\Phi(u) = 1 - \frac{1}{1+\rho} e^{-\frac{\rho u}{\mu(1+\rho)}}.$$

Conclusion

3. Summary of the dissertation

In this dissertations, we succeeded in adapting the concepts of MTI to cover the core part of Probability theory as followw.

(1) Probabilistic terminology of MTI concepts
(2) Use of the measure-images as probability laws and their properties
(3) Use of integrals as the mathematical expectations
(4) Use of product-measure as way of defining the probabilistic notions of independence and its characterizations
(5) Use of the Radon-Nikodym theorem to have a powerful foundation of the probabilistic density functions.
(6) Combination of the Radon-Nikodym theorem and the series as integral to characterize discrete probability laws
(7) Combination of the Radon-Nikodym theorem and the Lebesgue integral in $\mathbb{R}^k$ to characterize absolutely continuous probability laws

From this, we ensure the two first pillars in Probability theory

CPT1 High level in dealing with Probability laws in specific spaces, in particular in $\mathbb{R}^k$, $k \geq 1$
CPT2 Characterizing and knowing usual Probability laws in $\mathbb{R}^k$.

Finally, we showed that important problems in Financial mathematics can be solved with the mastering of probability theory and their applications.

Bibliography

Lo, G.S.(2016). A Course on Elementary Probability Theory. SPAS Editions. Saint-Louis, Calgary, Abuja. Doi : 10.16929/sbs/2016.0003.

Lo, G. S. (2017) Measure Theory and Integration By and For the Learner. SPAS Books Series. Saint-Louis, Senegal - Calgary, Canada.

Lo, G.S.(2018). *Mathematical Foundation of Probability Theory*. SPAS Books Series. Saint-Louis, Senegal - Calgary, Canada. Doi : http://dx.doi.org/10.16929/sbs/2016.0008. Arxiv : arxiv.org/pdf/1808.01713

Lo G.S., K. T. A. Ngom M. and Diallo M.(2018). Weak Convergence (IIA) - Functional and Random Aspects of the Univariate Extreme Value Theory. Arxiv : 1810.01625

Lo G.S.(2018) A flexible application of the Stone-Weierstrass Theorem in General and Aoolication in Probability Theory. *Journal of Mathematical Facts and Short Papers*. To appear.

Barndorff-Nielsen O. and C. Halgreen (1977). Infinite Divisibility of the Hyperbolic and Generalized Inverse Gaussian Distribution. Z. Wahrs. verw. Gebiete. Vol. 38, 309-311

Grosswald E.(1976) The Student t-distribution on any degree is infinitely divisible. Z. Wahrs. verw. Gebiete. Vol. 36, 103-109

Feller W. (1966) An introduction to Probability Theory, Vol. II. Wiley, New-York.

McNeil A.J, Frey R. and Embrechts P.(2005) Quantitative Risk Management. Princeton University Press.

Black F. and Sholes M.(1973) The pricing of options and corporate liabilities, *J. of Economy*, 81, pp.635-654

Merton, R.C.(1976) Option pricing when underlying stock returns are discontinuous, *J. Financial Economics*, 3, pp.125-144

Grandel J.(1991). *Aspects of Risk Theory*. Springer.

Paulsen, J. (1998). Ruin theory with compounding assets - a survey. *Insurance, Mathematics & Economics*, Vol. 22, pp. 3–16. MR1625827

Paulsen J.(2008). Ruin models with investment income. **Probability Surveys**. Vol. 5, pp. 416–434. DOI: 10.1214/08-PS134

Turner E.(2010) The Black-Scholes Model and extensions. University of Chicago. See Website: Semantic scholar, visited 2021/05/31 (https://www.semanticscholar.org/paper/THE-BLACK-SCHOLES-MODEL-AND-EXTENSIONS-Turner/f6a7be0984eefee223bba734bf615cdbbddeffe5)

Black F. and Scholes M.(1973) *The Journal of Political Economy*, 81 (3) pp. 637-654