Law of Large Numbers for Monotone Convolution

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Mathematics
University of Saskatchewan
Saskatoon

By

Enzo Wendler

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Department of Mathematics and Statistics
142 Mclean Hall
106 Wiggin's Road
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

Abstract

In this thesis, we use martingales to show that the dilation of a sequence of monotone convolutions $D_{\frac{1}{b_n}}(\mu_1 \triangleright \mu_2 \triangleright \cdots \triangleright \mu_n)$ is stable, where μ_j are probability distributions with the condition $\sum\limits_{n=1}^{\infty} \frac{1}{b_n}\mathrm{var}(\mu_n) < \infty$. This proves a law of large numbers for monotonically independent random variables.

Acknowledgements

First and foremost I would like to thank my adviser, JC Wang, for his support and guidance through my master's degree. I would then like to thank my committee for the comments they provided to improve my thesis. Finally I would like to thank my wife, Megan Richards, for her love and support and for proofreading my thesis.

Contents

Permission to Use **i**

Abstract **ii**

Acknowledgements **iii**

Contents **iv**

1 Introduction **1**

2 Preliminaries **7**
2.1 Types of independence 7
2.2 Analytic tools 12
2.3 Discrete Markov Chains and Martingales 18

3 Previous results **23**
3.1 Limit Theorems for free and Boolean Convolution 23
3.2 Limit Theorems for Monotone convolution 28

4 Law Of Large Numbers For Monotone Convolution **33**
4.1 Proof of Theorem 4.2 34
4.2 Future research 40

References **41**

Chapter 1

Introduction

Consider the experiment of flipping a fair coin. One would expect that if the coin is flipped many times, the difference between the number of heads and the number of tails should be relatively close to 0 when compared with the number of times the coin is flipped. This is, in fact, an example of the law of large numbers for classically independent and identically distributed random variables. The law of large numbers is perhaps one of the most well known results in classical probability theory.

There are two types of laws of large numbers, strong laws and weak laws. In this thesis, we only consider weak laws of large numbers, which means convergence of random variables in probability or, as we will see, equivalently in distribution. Normally, convergence in probability is stronger than convergence in distribution. However, in the law of large numbers the limiting random variable is a constant, and from this we can prove that convergence in distribution implies convergence in probability.

Let us recall definitions of these two types of convergence and prove that convergence to a constant in distribution implies convergence in probability.

Definition 1.1. A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ on a probability space $(\Omega, \mathcal{F}, P)$ is said to *converge* to X *in probability* if

$$\lim_{n\to\infty} P\left(|X_n - X| < \epsilon\right) = 1$$

for every $\epsilon > 0$.

Recall that the distribution μ of a random variable X on a probability space $(\Omega, \mathcal{F}, P)$ is the Borel measure defined by

$$\mu(B) = P(\{\omega : X(\omega) \in B\}),$$

for any Borel set $B \subset \mathbb{R}$.

Definition 1.2. A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ is said to *converge* to X *in distribution* if the distributions of X_n converge weakly to the distribution of X, i.e.

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(t) d\mu_n(t) = \int_{\mathbb{R}} f(t) d\mu(t)$$

for every continuous bounded function f on $\mathbb{R}$, where μ_n is the distribution of X_n and μ is the distribution of X.

Now let us show that convergence in distribution to a constant implies convergence in probability to the same constant.

Lemma 1.3. *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables defined on some probability space $(\Omega, \mathcal{F}, P)$ that converge to a constant c in distribution. Fix $\epsilon > 0$ and let $f_\epsilon : \mathbb{R} \to \mathbb{R}$ be the tent function defined by*

$$f_\epsilon(x) = \max\left(0, 1 - \frac{|x - c|}{\epsilon}\right).$$

Notice that $f_\epsilon \leq 1$ and $f_\epsilon(x) = 0$ for any $x \notin (c - \epsilon, c + \epsilon)$. Hence, for any $n \in \mathbb{N}$, one has that

$$\begin{aligned} P(|X_n - c| < \epsilon) &= \int_{\mathbb{R}} \mathbf{1}_{(c-\epsilon, c+\epsilon)}(x) d\mu_n(x) \\ &\geq \int_{\mathbb{R}} f_\epsilon(x) d\mu_n(x), \end{aligned}$$

where μ_n is the distribution of X_n. Since X_n converges to c in distribution and f_ϵ is continuous, we have that

$$\begin{aligned} \liminf_{n\to\infty} P(|X_n - c| < \epsilon) &\geq \lim_{n\to\infty} \int_{\mathbb{R}} f_\epsilon(x) d\mu_n(x) \\ &= \int_{\mathbb{R}} f_\epsilon(x) d\delta_c(x) = 1 \end{aligned}$$

Since P is a probability measure, P is bounded above by 1. Thus, we must have

$$\lim_{n\to\infty} P(|X_n - c| < \epsilon) = 1.$$

Thus, we have shown that X_n converges to c in probability.

If a sequence of probability distributions $\{\mu_n\}_{n=1}^{\infty}$ converges weakly to a point mass then any sequence of random variables distributed according to $\{\mu_n\}_{n=1}^{\infty}$ converge in distribution to a constant (and thus they must also converge in probability). Thus in the weak law of large numbers, we do not have to refer to any random variables but only their distributions.

Recall that the classical convolution $\mu * \nu$ of two probability distributions, μ and ν, is the distribution of the random variable $X + Y$, where X and Y are classically independent and distributed according to μ and ν, respectively. The convolution $\mu * \nu$ does not depend on the realization of the random variables X and Y. We now state the most general form of the weak law of large numbers for sums of classically independent random variables in terms of the convolution $*$.

Let $\{k_n\}_{n=1}^{\infty}$ be a sequence of positive integers. We say that a triangular array of probability distributions $\{\mu_{nj} : n \geq 1, 1 \leq j \leq k_n\}$ follows the *classical weak law of large numbers* if

$$\mu_{n1} * \mu_{n2} * \cdots * \mu_{nk_n}$$

converges weakly to a point mass as $n \to \infty$.

Let us show how the coin flipping example above can be put into the framework of distributions. To do so, we first need to build a probabilistic model.

Consider the set $\Omega = \{\omega = (\omega_1, \omega_2, ...) : \omega_i \in \{H, T\}\}$, i.e the set of all infinite sequences of Hs and Ts. Given a fixed $n \in \mathbb{N}$ and a fixed sequence $(a_1, a_2, ...a_n)$ where $a_i \in \{H, T\}$, we define

$$A_{(a_1,\ldots,a_n)} = \{\omega = (a_1, a_2, \ldots, a_n, \omega_1, \omega_2, \ldots) : \omega_i \in \{H, T\}\}.$$

That is, $A_{(a_1,\ldots,a_n)}$ is the subset of Ω consisting of all sequences that begin with the prescribed sequence $(a_1, a_2, \ldots, a_n)$. We define a σ-algebra $\mathcal{F}$ by

$$\mathcal{F} = \sigma\left(\{A_{(a_1,a_2,\ldots,a_n)} : n \in \mathbb{N}, a_i \in \{H, T\}\}\right).$$

From Ω and $\mathcal{F}$, we create a probability space $(\Omega, \mathcal{F}, P)$, where the probability is defined by $P(A_{(a_1,a_2,\ldots,a_n)}) = 2^{-n}$. We define a sequence of random variables $\{X_n\}_{n=1}^{\infty}$ on $(\Omega, \mathcal{F}, P)$ by

$$X_n(\omega) = \begin{cases} 1 & \text{if } \omega_n = H \\ -1 & \text{if } \omega_n = T. \end{cases}$$

That is, X_n is 1 if the n-th flip comes up heads and -1 otherwise. Then, after the first n flips, the number of heads minus the number of tails is given by

$$X_1 + X_2 + ... + X_n.$$

Since we are comparing the difference in heads and tails to the total number of times the coin is flipped, we are interested in the random variable

$$S_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Notice that $\frac{X_i}{n}$ has a equal chance of being $\frac{1}{n}$ or $\frac{-1}{n}$, i.e. it is distributed according to the measure $\mu_n = \frac{1}{2}\left(\delta_{\frac{1}{n}} + \delta_{\frac{-1}{n}}\right)$. Thus, the sum S_n is distributed according to μ_n^{*n}, where μ_n^{*n} denotes μ_n convolved with itself n times. It is well known that the sequence $\{\mu_n^{*n}\}$ converges weakly to the point mass δ_0 (see Example 2.21). Thus, the sum S_n converges to the constant 0 in distribution. Since this implies S_n converges in probability, it is extremely likely (i.e. with probability 1) that the difference in head and tails compared to the total number of times the coin is flipped is close to 0 (if the coin is flipped enough times).

The introduction of non-commutative probability theory has given rise to different types of independencies, including Boolean, free, and monotone. Each of these independencies introduces a new notion of convolution on the set of probability distributions. For each type of independence we can investigate the corresponding limit theorems. In particular, we can study laws of large numbers for different types of convolution.

As we will discuss in chapter 3, the weak law of large numbers for the free and Boolean independence has been shown in [3] for identically distributed random variables, and extended to non-identically distributed random variables in [2] and [14].

In this thesis, we focus on monotone convolution. The monotone convolution $\mu \triangleright \nu$ of two probability distributions μ and ν is the distribution of $X + Y$, where X and Y are monotonically independent random variables distributed according to μ and ν, respectively (see chapter 2 for more details). A general limit theorem for monotone convolution can be stated in the following form. Let $\{k_n\}_{n=1}^{\infty}$ be a sequence of positive integers, and let $\{\mu_{nj} : n \geq 1, 1 \leq j \leq k_n\}$ be a triangular array of probability distributions. The triangular array $\{\mu_{nj} : n \geq 1, 1 \leq j \leq k_n\}$ obeys a limit theorem if

$$\mu_{n1} \triangleright \mu_{n2} \triangleright \cdots \triangleright \mu_{nk_n}$$

converges weakly to a probability distribution. We say that the limit theorem is a weak law of large numbers if the limiting distribution is a point mass.

The central limit theorem for monotone convolution was the first limit theorem proved for monotone convolution [12]. The law of large numbers for monotonically independent and identically distributed random variables was already shown in [15]. However, the monotonically independent and non-identically distributed case had not been shown until recently in a joint work with JC Wang [16], on which this thesis is based.

One way to calculate the monotone convolution of two measures is to use the F-transform, which we will discuss in Chapter 2. In Example 2.30, we consider the measure

$$\mu = \frac{1}{2}(\delta_{-1} + \delta_1)$$

and use the F-transform to show

$$\mu \triangleright \mu = \frac{5+\sqrt{5}}{20}\left(\delta_{\gamma_1} + \delta_{-\gamma_1}\right) + \frac{5-\sqrt{5}}{20}\left(\delta_{\gamma_2} + \delta_{-\gamma_2}\right),$$

where

$$\gamma_1 = \sqrt{\frac{3+\sqrt{5}}{2}} \qquad \text{and} \qquad \gamma_2 = \sqrt{\frac{3-\sqrt{5}}{2}}.$$

In this thesis, we focus on a sequence of probability distributions $\{\mu_n\}_{n=1}^{\infty}$ and consider the convolution $\mu_1 \triangleright \mu_2 \triangleright \cdots \triangleright \mu_n$. If we use the F-transform to try to find a limiting distribution of $\mu_1 \triangleright \mu_2 \triangleright \cdots \triangleright \mu_n$, we have to consider the dynamics of the sequence of functions

$$F_{\mu_1} \circ F_{\mu_2} \circ \cdots \circ F_{\mu_n}.$$

Using complex analysis to deal with such compositions quickly becomes overwhelming. To get around this difficulty, we use Markov chains and martingales to find the limiting distribution.

The main result in this thesis provides a weak law of large numbers for non-identically distributed monotonically independent random variables with finite variances. The conditions we assume are not necessary conditions as random variables without finite variance can satisfy the weak law of large numbers.

The outline for the remainder of the thesis is as follows: In Chapter 2, we introduce basic concepts of non-commutative probability theory, and pertaining to the proof of our theorem. More precisely, in section 2.1, we introduce the different types of non-commutative independence and show some examples of independent algebras. In section 2.2 we introduce useful analytic tools for different types of convolution, such as the the Fourier transform and the F-transform, and we state some useful properties of the transforms such as Nevanlinna's form. In section 2.2, we introduce the notion of Markov chains and martingales. Chapter 3 is devoted to previous proofs in non-commutative probability theory, including the weak law of large numbers for free, Boolean and classical convolution. In Chapter 4, we prove our main result for monotonically independent (but not necessarily identically distributed) random variables.

Chapter 2

Preliminaries

We begin this chapter by stating essential definitions in non-commutative probability theory and showing some examples of different types of independence. We then go on to state same tool and theorem often used in proving limit theorems for different types of independent random variables. Finally we go on to state the definitions and some results concerning Markov chains and martingales.

2.1 Types of independence

Definition 2.1. A *non-commutative probability space* is a pair $(\mathcal{A}, \varphi)$, where $\mathcal{A}$ is a unital algebra and $\varphi : \mathcal{A} \to \mathbb{C}$ is a linear functional such that

$$\varphi(1) = 1.$$

We call a non-commutative probabilistic space, a C^**-probability space* if the algebra is a C^*-algebra and the function φ is positive.

Example 2.2. Given any classical probability space $(\Omega, \mathcal{F}, P)$, we can construct a non-commutative probability space $(\mathcal{A}, \varphi)$ by setting $\mathcal{A} = L^\infty(\Omega, \mathcal{F}, P)$ (i.e the set of all bounded random variables on $(\Omega, \mathcal{F}, P)$), where the multiplication of two elements $X, Y \in \mathcal{A}$ is defined as

$$(X \cdot Y)(\omega) = X(\omega) \cdot Y(\omega), \qquad \omega \in \Omega$$

and the linear map φ is given by

$$\varphi(X) = E[X] = \int_{\omega \in \Omega} X(\omega) P(d\omega).$$

Definition 2.3. Given a non-commutative probability space $(\mathcal{A}, \varphi)$ and two sub-algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ of $\mathcal{A}$ such that $ab = ba$ for all $a \in \mathcal{A}_1$ and $b \in \mathcal{A}_2$, we say $\mathcal{A}_1$ and $\mathcal{A}_2$ are *tensor independent* if we have

$$\varphi(ab) = \varphi(a)\varphi(b), \qquad a \in \mathcal{A}_1, b \in \mathcal{A}_2.$$

Example 2.4. Let $(\mathcal{A}, \varphi)$ be the non-commutative probability space constructed in Example 2.2. If $X, Y \in \mathcal{A}$ are classically independent random variables on $(\Omega, \mathcal{F}, P)$, then $\varphi(XY) = \varphi(X)\varphi(Y)$. Thus, while X and Y are classically independent random variables on the probability space $(\Omega, \mathcal{F}, P)$, they are tensor independent in the new framework $(\mathcal{A}, \varphi)$.

Tensor independence is only one type of independence that arises due to non-commutative probability theory. Let us now state the definition of some other types of independence.

Definition 2.5. Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. We say two unital sub-algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are freely independent if for any $n \in \mathbb{N}$, $i_1 \neq i_2 \neq \cdots \neq i_n$ with $i_k \in \{1, 2\}$,

$$\varphi(a_1 a_2 \cdots a_n) = 0,$$

for every $a_k \in \mathcal{A}_{i_k}$ such that $\varphi(a_k) = 0$.

Example 2.6. The free product is the classical example of free independence. Let us recall the free product of two groups G_1 and G_2. We define the free product of G_1 and G_2 by

$$G = \{e\} \cup \left\{g_1 g_2 \cdots g_n : n \in \mathbb{N}, i_1 \neq i_2 \neq \cdots \neq i_n, i_k \in \{1, 2\}, g_j \in G_{i_j} \setminus \{e_{i_j}\}\right\}$$

where we set the identities $e_1 = e_2 = e$. Here multiplication of two elements is given by juxtaposition and reduction to the above form by combining neighboring elements from the same group.

We now consider the algebras $\mathcal{A}_j = \mathbb{C}G_j$, $j \in \{1, 2\}$, as sub-algebras of $\mathcal{A} = \mathbb{C}G$. We define a linear functional $\varphi : \mathcal{A} \to \mathbb{C}$ by

$$\varphi(e) = 1 \quad \text{and} \quad \varphi(g) = 0 \text{ for g } \in \mathbb{C}(G \setminus \{e\}).$$

Then the sub-algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are freely independent.

Definition 2.7. Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. We say two sub-algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are *Boolean independent* if for any $n \in \mathbb{N}$, $i_1 \neq i_2 \neq \cdots \neq i_n$ with $i_k \in \{1, 2\}$,

$$\varphi(a_1 a_2 \cdots a_n) = \prod_{k=1}^{n} \varphi(a_k),$$

for every $a_k \in \mathcal{A}_{i_k}$.

Example 2.8. The standard example of Boolean independence was constructed by Bozejko [5]. Consider two Hilbert spaces H_1 and H_2 that share a common one-dimensional subspace spanned by a unit vector v (i.e. their intersection is the one dimensional $\mathbb{C}v$). Consider the decomposition of $H_j = \mathbb{C}v \oplus H_j^0, j \in \{1, 2\}$, and consider the space $\mathcal{H} = H_1^0 \oplus \mathbb{C}v \oplus H_2^0$. Let $\mathcal{B}(H_j)$ denote the set of bounded linear operators on H_j. For any bounded operator $a_1 \in \mathcal{B}(H_1)$ we define an extension $\mathbf{a}_1$ on $\mathcal{H}$ by

$$\mathbf{a}_1(h_1 \oplus cv \oplus h_2) = a_1(h_1 + cv).$$

That is, we let $\mathbf{a}_1 = a_1 \circ p_1$, where p_1 is a projection from $\mathcal{H}$ onto the subspace H_1. We similarly define an extension of any bounded operator $a_2 \in \mathcal{B}(H_2)$. Let $\mathcal{A}_j \subset \{\mathbf{a}_j : a_j \in \mathcal{B}(H_j)\}, j \in \{1, 2\}$ be sub-algebras of $\mathcal{B}(\mathcal{H})$ and let $\varphi : \mathcal{B}(\mathcal{H}) \to \mathbb{C}$ be defined by

$$\varphi(a) = \langle av, v \rangle.$$

Then the algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are Boolean independent on the non-commutative probability space $(\mathcal{B}(\mathcal{H}), \varphi)$.

Definition 2.9. Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. We say that two sub-algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are *monotonically independent* if for any $n \in \mathbb{N}$, $i_1 \neq i_2 \neq \cdots \neq i_n$, $i_k \in \{1, 2\}$,

$$\varphi(a_1 a_2 \ldots a_n) = \varphi(a_k)\varphi(a_1 a_2 \cdots a_{k-1} a_{k+1} \cdots a_n)$$

whenever $i_k = 2$, for any elements $a_k \in \mathcal{A}_{i_k}$.

Unlike other types of independence, monotone independence is not symmetric; that is, if $\mathcal{A}_1$ and $\mathcal{A}_2$ are monotonically independent, in general this does not imply that $\mathcal{A}_2$ and $\mathcal{A}_1$ are monotonically independent.

Remark 2.10. As we will show a construction of monotone convolution in Example 2.15, we will not give an example of monotone independence here.

We will now restrict ourselves to C^*-probability spaces. Given such a non-commutative probability space $(\mathcal{A}, \varphi)$, the self-adjoint elements of the algebra $\mathcal{A}$ are called random variables. We say two random variables X_1 and X_2 are monotonically (Boolean) independent if the algebras $\mathcal{A}_i = \{f(X_i) : f : \mathbb{R} \to \mathbb{R} \text{ is continuous and bounded}, f(0) = 0\}$, $i = 1, 2$, are monotonically (Boolean) independent, where the operator $f(X_i)$ is given by the functional calculus. Free independence of random variables is defined similarly. However, for two subalgebras to be freely independent, they must be unital, and thus we remove the constraint $f(0) = 0$.

Remark 2.11. Free, monotone and Boolean independence cannot be seen in classical probability theory, other than in constant random variables. We will only show this for monotone independence as the others follow from similar arguments. Assume that X and Y are monotonically independent commuting random variables. Notice that

$$\varphi(Y)\varphi(X^2)\varphi(Y) = \varphi(YX^2Y) = \varphi(X^2Y^2) = \varphi(Y^2)\varphi(X^2).$$

If X is nonzero, then dividing by $\varphi(X^2)$ gives

$$\varphi(Y^2) = \varphi(Y)^2,$$

which implies

$$\varphi((Y - \varphi(Y))^2) = 0.$$

Thus $Y = \varphi(Y)$, i.e. Y is constant.

Now we will show how the distribution of a non-commutative random variable is defined. Consider a C^*-probability space $(\mathcal{A}, \varphi)$ and a random variable $a \in \mathcal{A}$. Let $\mathcal{P}(X)$ denote the set all polynomials in X, and let $\Psi_a : \mathcal{P}(X) \to \mathbb{R}$ be defined by

$$\Psi_a(\sum_{k=1}^{n} c_k X^k) = \sum_{k=1}^{n} c_k \varphi(a^k).$$

Since polynomials are dense in the set of continuous functions there exists an extension of Ψ_a to continuous functions. By the Riesz-Markov-Kakutani Representation theorem, there exists a unique probability measure μ such that for any continuous function $f : \mathbb{R} \to \mathbb{R}$,

$$\Psi_a(f) = \int f d\mu.$$

We call the measure μ the distribution of the random variable a.

Remark 2.12. Given a non-commutative probability space $(\mathcal{A}, \varphi)$ constructed from a probability space $(\Omega, \mathcal{F}, P)$, as in Example 2.2, the distribution of a random variable $a \in \mathcal{A}$ matches the definition of the classically defined distribution of the random variable $a : \Omega \to \mathbb{R}$.

Now we can define the monotone convolution of two probability measures on $\mathbb{R}$.

Definition 2.13. The *monotone convolution* $\mu \triangleright \nu$ of two probability distributions μ and ν is the distribution of $X + Y$, where X and Y are monotonically independent random variables distributed according to μ and ν, respectively.

Remark 2.14. Definition 2.13 is a direct analogue of classical convolution. Also, note that monotone convolution is independent of the choice of random variables, see [7].

We similarly define Boolean and free convolution denoted by $\uplus, \boxplus$, respectively. Again both Boolean and free convolution are independent of the choice of random variables.

Example 2.15. Let us now consider the setting in which monotone convolution is often realized. Given a Hilbert space H and a fixed unit vector ξ, we can construct a non-commutative probability space $(\mathcal{A}, \varphi)$, where $\mathcal{A}$ is the set of all operators on H and φ is given by

$$\varphi(X) = \langle X\xi, \xi \rangle.$$

Let μ and ν be two Borel probability distributions, and consider the Hilbert space $H = L^2(\mathbb{R} \times \mathbb{R}, \mu \otimes \nu)$ and $\varphi(\cdot) = \langle \cdot \mathbf{1}, \mathbf{1} \rangle$. We define the (possibly unbounded) operators

$$Xf(x, y) = x \int_{t \in \mathbb{R}} f(x, t) d\nu(t), \qquad Yf(x, y) = yf(x, y).$$

Then X and Y are distributed according to μ and ν, respectively, and X and Y are monotonically independent, see [7]. Thus, the monotone convolution $\mu \triangleright \nu$ is the distribution of the random variable $X + Y$.

While this method works to find the monotone convolution of two measures, we rarely use the definition of monotone convolution directly. Instead, we rely on analytic tools such as the Cauchy transform.

2.2 Analytic tools

Let us begin by looking at the Fourier transform.

Definition 2.16. Given a measure μ the *Fourier transform* of μ is given by

$$\mathcal{F}\mu(t) = \int_{\mathbb{R}} e^{ixt} d\mu(x), \qquad t \in \mathbb{R}.$$

Remark 2.17. Given a Fourier transform it is possible to recover the underlying measure, i.e. $\mathcal{F}\mu = \mathcal{F}\nu$ if and only if $\mu = \nu$.

Remark 2.18. The Fourier transform of a classical convolution is the product of the individual Fourier transforms, i.e.

$$\mathcal{F}(\mu * \nu)(t) = \mathcal{F}\mu(t)\mathcal{F}\nu(t). \tag{2.1}$$

Definition 2.19. A family of probability distributions C is said to be *tight* if

$$\lim_{y\to\infty} \sup_{\mu\in C} \mu(\{t : |t| > y\}) = 0.$$

Theorem 2.20 (Lévy continuity theorem)**.** *Given probability distributions μ_n and μ such that*

$$\lim_{n\to\infty} \mathcal{F}\mu_n(t) = \mathcal{F}\mu(t), \qquad t \in \mathbb{R}$$

then the following statements are equivalent:

(1) μ_n converges weakly to μ;

(2) the sequence μ_n is tight.

See e.g. [17] for the proof.

Let us now show how the Fourier transform can be used to prove results for the classical weak law of large numbers.

Example 2.21. Let $\mu_n = \frac{1}{2}\left(\delta_{\frac{-1}{n}} + \delta_{\frac{1}{n}}\right)$, we wish to show that

$$\underbrace{\mu_n * \mu_n * \cdots * \mu_n}_{n\text{times}}$$

converges weakly to δ_0. Notice that

$$\mathcal{F}\mu_n = \frac{e^{\frac{-it}{n}} + e^{\frac{it}{n}}}{2} = \cos\left(\frac{t}{n}\right).$$

Thus, using equation (2.1), the Fourier transform of μ_n^{*n} is given by

$$\mathcal{F}\mu_n^{*n} = \cos^n\left(\frac{t}{n}\right) \xrightarrow[n\to\infty]{} 1$$

Notice that 1 is the Fourier transform of δ_0 and thus by the Lévy continuity theorem 2.20, we have that μ_n^{*n} converges weakly to δ_0.

Thus the Fourier transform can be used as a tool in showing the weak law of large numbers for classically independent random variables. It can also be used to prove other limit theorems such as the central limit theorem.

Now let us state some transforms that are useful in non-commutative probability theory. We use the notation $\mathbb{C}^+ = \{x + iy : x \in \mathbb{R}, y > 0\}$ and $\mathbb{C}^- = \{x + iy : x \in \mathbb{R}, y < 0\}$.

Definition 2.22. For a fixed probability measure μ, the Cauchy transform $G_\mu : \mathbb{C}^+ \to \mathbb{C}^-$ is a function defined by

$$G_\mu(z) = \int\limits_{t\in\mathbb{R}} \frac{1}{z-t}\mu(dt), \qquad z \in \mathbb{C}^+.$$

The Cauchy transform G_μ uniquely determines the probability measure μ. It is possible to recover the underlying measure μ using a process called the Stieltjes inversion process. Let

$$h_\epsilon(t) = -\frac{1}{\pi}\Im\big(G_\mu(t+i\epsilon)\big), \qquad \epsilon > 0,\ t \in \mathbb{R}.$$

Notice that

$$\begin{aligned}
h_\epsilon(t) &= -\frac{1}{\pi}\Im\big(G_\mu(t+i\epsilon)\big)\\
&= -\frac{1}{\pi}\Im\int_{\mathbb{R}} \frac{1}{t-s+i\epsilon}d\mu(s)\\
&= -\frac{1}{\pi}\Im\int_{\mathbb{R}} \frac{t-s-i\epsilon}{(t-s)^2+\epsilon^2}d\mu(s)\\
&= \frac{1}{\pi}\int_{\mathbb{R}} \frac{\epsilon}{(t-s)^2+\epsilon^2}d\mu(s),
\end{aligned}$$

which is the Poisson integral of the measure μ. In the weak topology,

$$d\mu(t) = \lim_{\epsilon \to 0} h_\epsilon(t) dt.$$

That is, for any bounded continuous function $f : \mathbb{R} \to \mathbb{C}$

$$\int_{\mathbb{R}} f(t) d\mu(t) = \lim_{\epsilon \to 0} \int_{\mathbb{R}} f(t) h_\epsilon(t) dt.$$

Thus, given a Cauchy transform, we are able to recover the underlying measure.

In non-commutative probability theory, we are often more interested in the F-transform, which is the reciprocal of the Cauchy transform.

Definition 2.23. For a fixed probability measure μ, the F-transform $F_\mu : \mathbb{C}^+ \to \mathbb{C}^+$ is given by

$$F_\mu(z) = \frac{1}{G_\mu(z)},$$

where G_μ is the Cauchy transform defined in Definition (2.22).

The F-transform has certain invertibility properties. More precisely, following Bercovici and Pata, for any two constants $\eta, M > 0$ we define

$$\Gamma_\eta = \{z = x + iy \in \mathbb{C}^+ : |x| < \eta y\}$$

and

$$\Gamma_{\eta,M} = \{z = x + iy \in \Gamma_\eta : y > M\}. \tag{2.2}$$

Then for every $\eta > 0$ there exists $M = M(\eta, \mu)$ such that F_μ has a left inverse F_μ^{-1} on $\Gamma_{\eta,M}$, see [3] for details.

Definition 2.24. Let μ be a probability distribution, $\eta > 0$, and let M be defined such that the F-transform of μ is invertible on $\Gamma_{\eta,M}$. The *Voiculescu transform* φ_μ of μ is defined by

$$\varphi_\mu(t) = F_\mu^{-1}(z) - z, \qquad z \in \Gamma_{\eta,M}.$$

Remark 2.25. Given any two probability distributions μ and ν, the Voiculescu transform satisfies

$$\varphi_{\mu \boxplus \nu}(z) = \varphi_\mu(z) + \varphi_\nu(z),$$

for any z in a truncated cone $\Gamma_{\eta,M}$ such that all three Voiculescu transforms are defined (see e.g. [3] for proof).

Thus, the Voiculescu transform is the free analogue of the Fourier transform (or more precisely the logarithm of the Fourier transform).

Remark 2.26. The Boolean analogue to the Fourier transform is given by

$$E_\mu(z) = z - F_\mu(z),$$

where F_μ is the F-transform of μ. That is,

$$E_{\mu \uplus \nu}(z) = E_\mu(z) + E_\nu(z).$$

Since monotone convolution is not symmetric, there cannot be any transform such that the transform of the monotone convolution is the sum of the individual transforms. However, in [7], Franz showed that the F-transform of a monotone convolution is the composition of the individual F-transforms; that is,

$$F_{\mu \triangleright \nu}(z) = F_\mu \circ F_\nu(z), \qquad z \in \mathbb{C}^+. \tag{2.3}$$

Let us now show some examples where we can use the F-transform to calculate the monotone convolution of measures. The classical example for monotone convolution is the arcsine distribution. Before showing the example, let us define the dilation of a measure.

Definition 2.27. The *dilation* of a probability measure μ by a positive real factor b is defined by

$$D_b(\mu)(B) = \mu(b^{-1}B)$$

for every Borel set $B \subset \mathbb{R}$.

Example 2.28. Let a be a positive real number and let μ be defined by $\mu(dt) = \frac{1}{\pi}\frac{1}{\sqrt{a^2-t^2}}dt$ for $t \in (-a, a)$. The F-transform of μ is given by

$$F_\mu(z) = \sqrt{z^2 - a^2}.$$

Thus,

$$F_{\mu \triangleright \mu} = \sqrt{z^2 - 2a^2},$$

which can be recognized as the F-transform of the distribution given by

$$\nu(dt) = \frac{1}{\pi}\frac{1}{\sqrt{2a^2 - t^2}}dt, \qquad t \in (-\sqrt{2}a, \sqrt{2}a).$$

Thus, since the F-transform uniquely determines the measure, we have that $\mu \triangleright \mu = \nu$. Further, notice that ν is just a dilation of μ, i.e.

$$\mu \triangleright \mu = D_{\sqrt{2}}\mu.$$

Example 2.29. Let x be any real number, and let δ_x be the Dirac delta measure. We compute the F-transform of δ_x as follows:

$$F_{\delta_x}(z) = \left(\int_{t\in\mathbb{R}} \frac{1}{z-t}\delta_x(dt)\right)^{-1} = z - x. \tag{2.4}$$

Now, (2.4) combined with (2.3) implies

$$F_{\delta_x \triangleright \mu}(z) = F_\mu(z) - x, \qquad z \in \mathbb{C}^+.$$

Example 2.30. Consider the measure

$$\mu = \frac{1}{2}\left(\delta_{-1} + \delta_1\right).$$

Let us find $\mu \triangleright \mu$. Note that

$$F_\mu(z) = 2\left(\frac{1}{z-1} + \frac{1}{z+1}\right)^{-1} = \frac{z^2-1}{z}$$

and hence

$$F_{\mu\triangleright\mu}(z) = F_\mu \circ F_\mu(z) = \frac{z^4 - 3z^2 + 1}{z^3 - z}.$$

Now we must find the measure that corresponds to this F-transform. To this end, let us first look at the Cauchy transform

$$G_{\mu\triangleright\mu}(z) = \frac{z^3 - z}{z^4 - 3z^2 + 1}.$$

Using partial fraction decomposition, we find that

$$G_{\mu\triangleright\mu}(z) = \frac{5+\sqrt{5}}{20}\left(\frac{1}{z-\gamma_1} + \frac{1}{z+\gamma_1}\right) + \frac{5-\sqrt{5}}{20}\left(\frac{1}{z-\gamma_2} + \frac{1}{z+\gamma_2}\right)$$

where

$$\gamma_1 = \sqrt{\frac{3+\sqrt{5}}{2}} \qquad \text{and} \qquad \gamma_2 = \sqrt{\frac{3-\sqrt{5}}{2}}.$$

From this, we see that

$$\mu \triangleright \mu = \frac{5+\sqrt{5}}{20}\left(\delta_{\gamma_1} + \delta_{-\gamma_1}\right) + \frac{5-\sqrt{5}}{20}\left(\delta_{\gamma_2} + \delta_{-\gamma_2}\right).$$

Thus, we see that the F-transform gives us a way to find the monotone convolution of two measures. However, it can be difficult to recover the underlying measure. Hasebe proved a general formula for the monotone convolution of two discrete distributions in [9].

Since the F-transform is an analytic self map from the upper half plane to itself, every F-transform admits a unique Nevanlinna's form. That is, the F-transform of a measure can be written as

$$F_\mu(z) = C + Dz + \int\limits_{t\in\mathbb{R}} \frac{1+tz}{t-z} d\sigma \tag{2.5}$$

where C and D are real constants and σ is a finite Borel measure.

Lemma 2.31. *If μ is a probability measure with finite variance, then Nevanlinna's form of the F-transform of μ reduces to*

$$F_\mu(z) = z - m(\mu) + \int\limits_{t\in\mathbb{R}} \frac{1}{t-z} d\sigma \tag{2.6}$$

where $m(\mu)$ is the mean of the measure μ and σ is a Borel probability measure with $\sigma(\mathbb{R}) = var(\mu)$.

Proof. Let μ be a probability measure with finite variance and consider Nevanlinna's form for F_μ. Pata showed that $var(\mu)$ is finite if and only if $var(\sigma)$ is finite in [13]. Thus we have that

$$\begin{aligned}\int_{\mathbb{R}} \frac{1+tz}{t-z} d\sigma &= \int_{\mathbb{R}} \left(\frac{1+t^2}{t-z} - t\right) d\sigma \\ &= \int_{\mathbb{R}} \frac{1}{t-z}(1+t^2) d\sigma - m(\sigma) \\ &= \int_{\mathbb{R}} \frac{1}{t-z} d\sigma' - m(\sigma)\end{aligned} \tag{2.7}$$

where $\sigma' = (1+t^2)\sigma$. Notice that since $var(\sigma)$ is finite, the measure σ' is a finite measure. Substituting (2.7) into (2.5) and absorbing $m(\sigma)$ into the constant, we obtain

$$F_\mu(z) = Dz + C + \int\limits_{t\in\mathbb{R}} \frac{1}{t-z} d\sigma'. \tag{2.8}$$

Now we show that $C = -m(\mu)$ and $D = 1$. Let m_1 and m_2 be the first two moments of μ. Notice that for $z = iy$ with large y,

$$G_\mu(z) = \int \frac{1}{z-t} d\mu \;=\; \frac{1}{z} \int \sum_{n=0}^{\infty} \frac{t^n}{z^n} d\mu \;=\; \frac{1}{z} + \frac{m_1}{z^2} + \frac{m_2}{z^3} + \mathcal{O}\left(\frac{1}{z^4}\right).$$

Thus, we have that

$$F_\mu(z) = \frac{z^3}{z^2 + m_1 z + m_2 + \mathcal{O}\left(\frac{1}{z}\right)}.$$

Using long division,

$$F_\mu(z) = z - m_1 - \frac{m_2 - m_1^2}{z} + \mathcal{O}\left(\frac{1}{z^2}\right). \tag{2.9}$$

Note the right hand side of (2.8) can be written as the power series

$$F_\mu(z) = Dz + C - \frac{\sigma'(\mathbb{R})}{z} + \mathcal{O}\left(\frac{1}{z^2}\right). \tag{2.10}$$

By comparing (2.10) and (2.9), we get our desired result. □

2.3 Discrete Markov Chains and Martingales

In this section we give a brief introduction to discrete Markov chains and martingales. A discrete Markov chain is a sequence of random variables (a discrete stochastic process) such that the next step in the Markov chain only depends on the current step and not on the past. Before giving a more rigorous definition, let us look at an example of a Markov chain in a 'toy' model.

Example 2.32. Let us pretend that the weather follows the following rules.

(1) Each day it can rain, be cloudy or be sunny.

(2) On day 1 it is sunny.

(3) If it is sunny the next day the weather will be $\begin{cases} \text{sunny} & 70\% \\ \text{rain} & 10\% \\ \text{cloudy} & 20\% \end{cases}$ of the time.

(4) If it is raining the next day the weather will be $\begin{cases} \text{sunny} & 30\% \\ \text{rain} & 40\% \\ \text{cloudy} & 30\% \end{cases}$ of the time.

(5) If it is cloudy the next day the weather will be $\begin{cases} \text{sunny} & 30\% \\ \text{rain} & 30\% \\ \text{cloudy} & 40\% \end{cases}$ of the time.

This (very bad) model of the weather would then be a Markov chain as tomorrow's weather only depends on what the weather is today but not on the weather in the past.

To give a more rigorous definition of a Markov chain, we need the notion of conditional expectation, whose existence is guaranteed by the Radon-Nikodym theorem.

Theorem 2.33 (Radon-Nikodym theorem)**.** *If μ and ν are σ-finite signed measures on $\mathbb{R}$, and v is absolutely continuous with respect to μ, then there exists $g \geq 0$ such that $\nu(E) = \int_E g d\mu$. If h is another such function then $g = h$ μ-almost everywhere.*

Let us now state the definition of conditional expectation.

Definition 2.34. Given a probability space $(\Omega, \mathcal{F}_0, P)$, a random variable X which is $\mathcal{F}_0$-measurable, and a $\sigma-$field $\mathcal{F} \subset \mathcal{F}_0$, the conditional expectation $E[X|\mathcal{F}]$ of X given $\mathcal{F}$ is a random variable Y with the following properties:

(1) Y is $\mathcal{F}$-measurable, and

(2) $\int_B X dP = \int_B Y dP, \qquad B \in \mathcal{F}$.

Remark 2.35. The Radon-Nikodym theorem can be used to show that conditional expectation exists and is a.s. unique. Let $\nu(E) = \int_E X dP$ for all $E \in \mathcal{F}$. Clearly, ν is absolutely continuous with respect to P, and thus there exists a Y such that

$$\nu(E) = \int_E Y dP, \qquad E \in \mathcal{F}.$$

Further, Y is $\mathcal{F}$-measurable and a.s. unique.

Remark 2.36. We often use $E[X_n|X_{n-1}]$ to mean $E[X_n|\sigma(X_{n-1})]$ and similarly we use $E[X_n|X_{n-1}, X_{n-2}, \cdots, X_1]$ to mean $E[X_n|\sigma(X_{n-1}, X_{n-2}, \cdots, X_1)]$.

Definition 2.37. The conditional probability of event A given event B is defined as

$$P(A|B) = E[1_A|B].$$

Now we state the definition of a discrete Markov chain.

Definition 2.38. A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ is a *Markov chain* if

$$P(X_{n+1} \in B | X_n \in B_n, \cdots, X_1 \in B_1) = P(X_{n+1} \in B | X_n \in B_n)$$

for any Borel measurable sets $B, B_1, \ldots, B_n$.

Remark 2.39. A Markov chain $\{X_n\}_{n=1}^{\infty}$ where X_n has finite expectation for all n satisfies

$$E[X_n | X_{n-1}, X_{n-2}, \ldots, X_1] = E[X_n | X_{n-1}] \qquad a.s. \qquad n > 1. \tag{2.11}$$

Note the converse is not true, i.e. a sequence that satisfies (2.11) is not necessarily a Markov chain.

Example 2.40. A more useful example of a Markov chain is a random walk. Consider a particle starting at position 0, and at each time step the particle either moves right 1 unit with probability a or left 1 unit with probability $1-a$ for some $a \in [0, 1]$. If X_n denotes the position after n steps, then the sequence $\{X_n\}_{n=1}^{\infty}$ is a Markov chain.

We also state the definition of a martingale.

Definition 2.41. A sequence $\{X_n\}_{n=1}^{\infty}$ is a *martingale* if

$$E[|X_n|] < \infty, \qquad n \geq 1,$$
$$E[X_n | X_{n-1}, \cdots, X_1] = X_{n-1}, \qquad n \geq 2.$$

A martingale can be stated more informally as a stochastic process such that, at any given step in the martingale, the expectation of the next step is the current one. The classical example of this is a 'fair' gambling game in which knowing the past does not help.

Example 2.42. Consider the 'game' of flipping a coin. Suppose if the coin comes up heads we win a dollar and if it comes up tails we lose a dollar. Let X_n be the amount of money we have after the n-th flip. Then the sequence $\{X_n\}_{n=1}^{\infty}$ is a martingale.

Note that the example above was both a Markov chain and a martingale. While the martingales we use in this thesis will also be Markov chains, this is not always the case.

One way to construct a Markov chain is with an initial distribution and a sequence of transition probabilities. Let us formally define a transition probability on Borel measurable distributions.

Definition 2.43. A transition probability is a function $p : \mathbb{R} \times \mathcal{B} \to \mathbb{R}$ such that

(1) for any $x \in \mathbb{R}$, $p(x, \cdot)$ is a Borel probability measure on $\mathbb{R}$, and

(2) for any $B \in \mathcal{B}$, $p(\cdot, B)$ is a Borel measurable function.

Remark 2.44. It is possible to generalize Definition 2.43 to any measurable space [6], but for simplicity we have chosen our measurable space to be $(\mathbb{R}, \mathcal{B})$.

One way to prove that a function is measurable is to use Dynkin's π-λ theorem. First, let us recall the definition of a π-system and a λ-system.

Definition 2.45. Given a set Ω, a collection of subsets $\mathcal{P}$ is a π-system if for all $A, B \in \mathcal{P}$, we have that $A \cap B \in \mathcal{P}$.

Definition 2.46. Given a set Ω, a collection of subsets $\mathcal{L}$ is a λ-system if it satisfies

(1) $\Omega \in \mathcal{L}$;

(2) if $A \in \mathcal{L}$, then $A^c \in \mathcal{L}$;

(3) if $\{A_n\}_{n=1}^{\infty} \subset \mathcal{L}$ where $A_{n-1} \subset A_n$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$.

Theorem 2.47 (Dynkin's π-λ Theorem)**.** *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system such that $\mathcal{P} \subset \mathcal{L}$, then*

$$\sigma(\mathcal{P}) \subset \mathcal{L}.$$

See [6] for proof.

Given a sequence of transition probabilities $\{p_n\}_{n=2}^{\infty}$ and an initial distribution μ, we can construct a Markov chain $\{X_n\}_{n=1}^{\infty}$ with joint distribution given by

$$P(X_1 \in B_1, \cdots, X_n \in B_n) = \int\limits_{t_1 \in B_1} \mu(dt_1) \int\limits_{t_2 \in B_2} p_2(t_1, dt_2) \cdots \int\limits_{t_n \in B_n} p_n(t_{n-1}, dt_n).$$

To show that such a Markov chain exists, we use Kolmogorov's extension theorem.

Theorem 2.48 (Kolmogorov's extension theorem). *Let μ_n be probability measures on $(\mathbb{R}^n, \mathcal{B}^n)$ such that for every n*

$$\mu_n(B_1 \times B_2 \times \cdots \times B_{n-1} \times \mathbb{R}) = \mu_{n-1}(B_1 \times B_2 \times \cdots \times B_{n-1})$$

for any Borel measurable sets $B_i \in \mathcal{B}$, where $1 \leq i \leq n-1$. Then there exists a unique probability measure P on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ such that

$$P(\omega_1 \in B_1, \ldots, \omega_n \in B_n) = \mu_n(B_1 \times \cdots \times B_n)$$

for any Borel measurable sets $B_i \in \mathcal{B}$, where $1 \leq i \leq n$.

See e.g. [6] for proof. We will also use Doob's L^2 martingale convergence theorem.

Theorem 2.49 (Doob's L^2 martingale convergence theorem). *Let $\{X_n\}_{n=1}^{\infty} \subset L^2(\Omega, \mathcal{F}, P)$ be a martingale. If*

$$\sup_n E(X_n^2) < \infty$$

then

$$X_n \to X \qquad \text{both in } L^2 \text{ and a.e.}$$

Again see e.g. [6] for proof.

Chapter 3

Previous results

In this chapter, we look at what has already been proved in non-commutative probability theory. We use $\mathcal{M}$ to denote the set of all Borel probability measures on $\mathbb{R}$.

3.1 Limit Theorems for free and Boolean Convolution

Let us begin by stating some definitions used by Bercovici and Pata, then go on to state some of their theorems obtained in [3].

Definition 3.1. We say $\mu, \nu \in \mathcal{M}$ are *equivalent*, and we write $\mu \sim \nu$, if there exists $a > 0$ and $b \in \mathbb{R}$ such that

$$\mu(B) = \nu(aB + b)$$

for every $B \in \mathcal{B}$. We say the equivalence is strict if $b = 0$.

Definition 3.2. We say that a measure μ is in the $*$*-partial domain of attraction* of ν, denoted $\mathcal{P}_*(\nu)$ if there exists a sequence of positive integers $k_1 < k_2 < \cdots$ and a measure $\mu_n \sim \mu$ such that

$$\nu = \underbrace{\mu_n * \mu_n * \cdots * \mu_n}_{k_n \text{ times}}.$$

For each type of independence, we define a partial domain of attraction by replacing $*$ in Definition 3.2. For example, we say $\mu \in \mathcal{P}_{\triangleright}(\nu)$ if there exists $k_1 < k_2 < \cdots$ and a measure $\mu_n \sim \mu$ such that

$$\nu = \underbrace{\mu_n \triangleright \mu_n \triangleright \cdots \triangleright \mu_n}_{k_n \text{ times}}.$$

Recall in Definition 2.16, the Fourier transform $\mathcal{F}\mu$ of μ is given by

$$\mathcal{F}\mu(t) = \int_{\mathbb{R}} e^{itx} d\mu(x), \qquad t \in \mathbb{R},$$

and that a Fourier transform uniquely determines a measure.

Theorem 3.3. *A measure ν is $*$-infinitely divisible if and only if there exists a finite positive Borel measure σ on $\mathbb{R}$ and a real number γ such that*

$$\mathcal{F}\mu(t) = exp\left[i\gamma t + \int_{\mathbb{R}} (e^i tx - 1 - itx)\frac{x^2+1}{x^2} d\sigma(x)\right], \qquad t \in \mathbb{R}.$$

Remark 3.4. We use the notation $\nu_*^{\gamma,\sigma}$ for the measure whose Fourier transform is given by

$$F\nu_*^{\gamma,\sigma}(t) = exp\left[i\gamma t + \int_{\mathbb{R}} (e^i tx - 1 - itx)\frac{x^2+1}{x^2} d\sigma(x)\right], \qquad t \in \mathbb{R}.$$

Clearly $\nu_*^{\gamma,\sigma}$ is $*$-infinitely divisible, if σ is positive.

Using the Voiculescu transform, we define a free analogue to $\nu_*^{\gamma,\sigma}$. More precisely, given a finite positive Borel measure σ on $\mathbb{R}$ and a real number γ we define $\nu_{\boxplus}^{\gamma,\sigma}$ to be the Borel measure such that the Voiculescu transform of $\nu_{\boxplus}^{\gamma,\sigma}$ is given by

$$\varphi_{\nu_{\boxplus}^{\gamma,\sigma}}(z) = \gamma + \int_{\mathbb{R}} \frac{1+tx}{z-t} d\sigma(t), \qquad z \in \mathbb{C}^+.$$

Remark 3.5. As we will show shortly, a probability measure μ is $\boxplus$-infinitely divisible if and only if $\mu = \nu_{\boxplus}^{\gamma,\sigma}$ for some real number γ and finite Borel measure σ.

We also define the Boolean analogue by

$$E_{\nu_{\uplus}^{\gamma,\sigma}}(z) = \gamma + \int_{\mathbb{R}} \frac{1+tx}{z-t} d\sigma(t).$$

Remark 3.6. Any probability distribution can be written in the form $\nu_{\uplus}^{\gamma,\sigma}$ and therefore any measure is $\uplus$-infinitely divisible.

Remark 3.7. Given a family of probability distributions C, we have

$$F_\mu(z) = z(1+\mathrm{o}(1)), \quad \text{as } |z| \to \infty, \quad z \in \Gamma_\eta,$$

for all μ in C if and only if the family C is tight (see [15]).

Another useful property of tightness is that any tight sequence of probability distributions will have a convergent subsequence.

We are now ready to state some theorems. We begin with the free analogue of the Levy continuity theorem (see Theorem 2.20) which we will state without proof.

Proposition 3.8. *Let $\{\mu_n\}_{n=1}^{\infty}$ be a sequence of probability distributions. The following are equivalent:*

(1) μ_n converges weakly to a probability measure μ;

(2) there exists $\eta, M > 0$ such that the sequence φ_{μ_n} converges uniformly on $\Gamma_{\eta,M}$ to a function R and $\varphi_{\mu_n}(z) = \mathrm{o}(|z|)$ uniformly in n as $z \to \infty, z \in \Gamma_{\eta,M}$.

If (1) and (2) hold, the function $R = \varphi_{\mu}$.

The proposition allows us to use the Voiculescu transform to prove limit theorems for free convolution. Now we state Theorem 6.3 from [3], which is a limit theorem for free, Boolean, and classical convolutions.

Theorem 3.9. *Fix a finite positive Borel measure σ on $\mathbb{R}$, a real number γ, a sequence $\mu_n \in \mathcal{M}$ and a sequence of positive integers $k_1 < k_2 < \cdots$. The following assertions are equivalent:*

(1) The sequence $\underbrace{\mu_n * \mu_n * \cdots * \mu_n}_{k_n times}$ converges weakly to $\nu_*^{\gamma,\sigma}$;

(2) The sequence $\underbrace{\mu_n \boxplus \mu_n \boxplus \cdots \boxplus \mu_n}_{k_n times}$ converges weakly to $\nu_{\boxplus}^{\gamma,\sigma}$;

(3) The sequence $\underbrace{\mu_n \uplus \mu_n \uplus \cdots \uplus \mu_n}_{k_n times}$ converges weakly to $\nu_{\uplus}^{\gamma,\sigma}$;

(4) The measures

$$d\sigma_n(x) = k_n \frac{x^2}{x^2+1} d\mu_n(x)$$

converge weakly to σ and

$$\lim_{n\to\infty} k_n \int_{\mathbb{R}} \frac{x}{x^2+1} d\mu_n(x) = \gamma.$$

We will not present a full proof of Theorem 3.9, we will just state the main idea. The method of proof is to show each statement is equivalent to statement (4). The proof that (1) is equivalent to (4) is a classical result which is contained in [8].

Proof. Let us go over the proof that (2) implies (4).

By the free analogue of the Lévy continuity theorem (i.e. Proposition 3.8), we have

$$\lim_{n\to\infty} k_n \varphi_{\mu_n} = \varphi_{\nu_{\boxplus}^{\gamma,\sigma}}.$$

Notice that μ_n weakly converges to 0. As Bercovici and Pata showed, it follows that the Voiculescu transform of μ_n can be written as

$$\varphi_{\mu_n}(z) = z^2 \left[G_{\mu_n}(z) - \frac{1}{z} \right] (1 + v_n(z)),$$

for z in some truncated cone $\Gamma_{\eta,M}$, and where $v_n(z) \to 0$ as $n \to \infty$. Then it follows that

$$k_n z^2 \left[G_{\mu_n}(z) - \frac{1}{z} \right] \xrightarrow[n\to\infty]{} \varphi_{\nu_{\boxplus}^{\gamma,\sigma}}. \tag{3.1}$$

Writing the left hand side of (3.1) in Nevanlinna's form,

$$k_n z^2 \left[G_{\mu_n}(z) - \frac{1}{z} \right] = \gamma_n + \int_{\mathbb{R}} \frac{1+tz}{z-t} d\sigma_n(t),$$

where

$$\gamma_n = k_n \int_{\mathbb{R}} \frac{t}{1+t^2} d\mu_n(t)$$

and

$$d\sigma_n(t) = \frac{k_n t^2}{1+t^2} d\mu_n(t).$$

Now it remains to be shown that γ_n converges to γ and σ_n weakly converges to σ. To do this, Bercovici and Pata showed that the sequence σ_n is tight and hence has a convergent subsequence. Using (3.1), it can be shown that the converging subsequence must converge to σ. Then, using the definition of $\nu_{\boxplus}^{\gamma,\sigma}$ and the fact that σ_n converges to σ, it can be shown that $\gamma = \lim_{n\to\infty} \gamma_n$. This shows that (2) implies (4).

To prove the converse, notice that μ_n converges weakly to δ_0, since

$$\mu_n(\{t : |t| < \epsilon\}) \le \frac{1+\epsilon^2}{\epsilon^2} \int_{\mathbb{R}} \frac{t^2}{1+t^2} d\mu_n(t) = \frac{1+\epsilon^2}{\epsilon^2} \frac{1}{k_n} \sigma_n(\mathbb{R}),$$

which converges to 0 for any $\epsilon > 0$. Hence

$$\varphi_{\mu_n}(z) = z^2 \left[G_{\mu_n}(z) - \frac{1}{z} \right] (1 + v_n(z)).$$

Using Nevanlinna's form again

$$k_n z^2\left[G_{\mu_n}(z)-\frac{1}{z}\right]=\gamma_n+\int_{\mathbb{R}}\frac{1+tz}{z-t}d\sigma_n(t), z\in\mathbb{C}^+.$$

Thus, by the definition of $\nu_{\boxplus}^{\gamma,\sigma}$, it follows that

$$k_n\varphi_{\mu_n}(z)=k_n z^2\left[G_{\mu_n}(z)-\frac{1}{z}\right](1+v_n(z))$$

converges to $\varphi_{\mu_{\boxplus}^{\gamma,\sigma}}$. So, by Proposition 3.8, all that remains to be shown is that

$$k_n(iy)^2\left[G_{\mu_n}(iy)-\frac{1}{iy}\right]=\mathrm{o}(y),$$

uniformly in n, as $y\to\infty$. Since γ_n converges to γ, we only must show that

$$\int_{\mathbb{R}}\frac{1+ity}{iy-t}d\sigma_n=\mathrm{o}(y)$$

uniformly in n, as $y\to\infty$. To this end, we notice that

$$\left|\frac{1+ity}{iy-t}\right|\leq y$$

for $y\geq 1$. Hence, for any $M>0$ and $y>1$,

$$\begin{aligned}\left|\int_{\mathbb{R}}\frac{1+ity}{iy-t}d\sigma_n(t)\right| &\leq \int_{-M}^{M}2\frac{1+|t|y}{y+|t|}d\sigma(t)+y\sigma_n(\{t:|t|\geq M\})\\ &\leq 2\frac{1+My}{y+M}++y\sigma_n(\{t:|t|\geq M\}),\end{aligned}$$

which tends to $\mathrm{o}(y)$ as the sequence σ_n is tight. The proof for the Boolean case is similar and thus we shall not discuss it here. □

Theorem 3.9 was extended to the non-identically distributed case. To be precise we need the concept of an infinitesimal array.

Definition 3.10. A triangular array $\{\mu_{nj}:1\leq n, 1\leq j<k_n\}$ is said to be *infinitesimal* if

$$\lim_{n\to\infty}\max_{1<j<k_n}\mu_{nj}(t:|t|>\epsilon\})=0.$$

Theorem 3.3 in [2] states the following

Theorem 3.11. *For an infinitesimal array $\{\mu_{nj}\}_{n,j}\subset\mathcal{M}$ and a sequence $\{c_n\}_{n=1}^{\infty}\subset\mathbb{R}$, the following statements are equivalent:*

(1) The sequence $\mu_{n1} * \mu_{n2} * \cdots * \mu_{nk_n} * \delta_{c_n}$ *converges weakly to* $\nu_*^{\gamma,\sigma}$*;*

(2) The sequence $\mu_{n1} \boxplus \mu_{n2} \boxplus \cdots \boxplus \mu_{nk_n} \boxplus \delta_{c_n}$ *converges weakly to* $\nu_\boxplus^{\gamma,\sigma}$*;*

(3) The sequence of measures

$$d\sigma_n = \sum_{j=1}^{k_n} \frac{t^2}{1+t^2} d\mu_{nj}$$

converges weakly to σ *and*

$$\gamma_n = c_n + \sum_{j=1}^{k_n} \left[\int_{|t|<1} t d\mu_{nj}(t) + \int_{\mathbb{R}} \frac{t}{1+t^2} d\mu_{nj}(t) \right]$$

converges to γ *as* $n \to \infty$*.*

Remark 3.12. As the method used in the proof of Theorem 3.11 is similar to the method used for the proof of Theorem 3.9, we will not show it here. We also remark that the Boolean analogue of Theorem 3.11 was proven in [14].

Remark 3.13. We remark that for any real constant a,

$$\delta_a = \nu_*^{a,0} = \nu_\boxplus^{a,0} = \nu_{\uplus}^{a,0},$$

where 0 is the zero measure. Thus, Theorem 3.11 gives a necessary and sufficient condition for a triangular array to follow the weak law of large numbers for free and classical convolution.

3.2 Limit Theorems for Monotone convolution

In this section, we discuss the limit theorems pertaining to monotone convolution. We discuss results obtained in [15] and then quickly mention the analogue of Theorem 3.9 obtained in [1].

One difficulty when working with monotone convolution is that there is no transform to break up a monotone convolution into a sum (or product) of individual transforms. However, as discussed in chapter 2, the F-transform of $\mu \triangleright \nu$ is the composition, $F_\mu \circ F_\nu$. In [15], the author uses this to prove the weak law of large numbers for identically distributed and monotonically independent random variables. However, let us first discuss another result for which we use the following definitions.

Definition 3.14. We say $\mu, \nu \in \mathcal{M}$ are of the same *strict type* if there exists $b > 0$ such that $D_b\mu = \nu$.

Definition 3.15. A measure $\mu \in \mathcal{M}$ is said to be *strictly stable* if given two real constants $a, b > 0$, there exists a $c > 0$ such that $D_a\mu \triangleright D_b\mu = D_c\mu$

We use the notation $\mu^{\triangleright n}$ to denote the distribution

$$\underbrace{\mu \triangleright \mu \triangleright \cdots \triangleright \mu}_{n \text{ times}}.$$

Theorem 3.16. *Given $\nu \in \mathcal{M}$ with $\nu \neq 0$, the following statements are equivalent:*

(1) for each positive integer k, the measure $\nu^{\triangleright k}$ is of the same strict type as ν;

(2) there exists $\mu \in \mathcal{M}$ and constants $b_n > 0$ such that $D_{\frac{1}{b_n}}\mu^{\triangleright n}$ weakly converges to ν;

(3) the measure ν is strictly stable.

Proof. (1) implies (2) is clear since we can find constants $b_n > 0$ such that $D_{\frac{1}{b_n}}(\nu^{\triangleright n}) = \nu$ for any integer n. The proof that (3) implies (1) follows from the definition of strictly stable. Thus all we must show is (2) implies (3). Assume (2) holds; that is, assume there exists $\mu \in \mathcal{M}$ and constants $b_n > 0$ such that $D_{\frac{1}{b_n}}(\mu^{\triangleright n})$ weakly converges to ν. We use the notation

$$\mu_n = D_{\frac{1}{b_n}}(\mu^{\triangleright n}).$$

Given a constant $a > 0$, we let $m = m(n)$ be a sequence of integers such that

$$a_n = \frac{b_m}{b_n} \to a.$$

For existence of these integers we refer the reader back to [15]. Then we note that

$$D_{\frac{b_{n+m}}{b_n}}(\mu_{n+m}) = \mu_n \triangleright D_{a_n}(\mu_m).$$

The left hand side converges weakly to $D_c(\nu)$, where $c = \lim_{n\to\infty} \frac{b_{n+m}}{b_n}$ and the right hand side weakly converges to $\nu \triangleright D_a(\nu)$. However, since the limit is unique, we have found a value c such that

$$D_c(\nu) = \nu \triangleright D_a(\nu).$$

Now notice that, given constants $a, b > 0$, we can find c such that

$$\begin{aligned} D_a(\nu) \triangleright D_b(\nu) &= D_a\left(\nu \triangleright D_{\frac{b}{a}}(\nu)\right) \\ &= D_a D_{c_1}(\nu) \\ &= D_c(\nu). \end{aligned}$$

Thus, we have proved (2) implies (3). □

Let us now look at the weak law of large numbers for monotonically independent and identically distributed random variables that was proved in [15].

Theorem 3.17. *Let $a \in \mathbb{R}$. The sequence*

$$\mu_n = \underbrace{D_{\frac{1}{b_n}}(\mu) \triangleright D_{\frac{1}{b_n}}(\mu) \triangleright \cdots \triangleright D_{\frac{1}{b_n}}(\mu)}_{ntimes}$$

weakly converges to δ_a if and only if

$$\lim_{n\to\infty} \int_{\mathbb{R}} \frac{nb_n t}{b_n^2 + t^2} d\mu(t) = a \quad \textit{and} \quad \lim_{n\to\infty} \int_{\mathbb{R}} \frac{nt^2}{b_n^2 + t^2} d\mu(t) = 0.$$

Proof. Suppose μ_n weakly converges to δ_a. Thus $\{\mu_n\}_{n=1}^{\infty}$ is a tight sequence. Thus, there exists a truncated cone $\Gamma_{\eta,M}$ such that

$$\left| \frac{1}{b_j} F_\mu^{\circ j}(b_j z) - z \right| = |F_{\mu_j}(z) - z| \leq |z|, \qquad z \in \Gamma_{\eta,M}.$$

In particular, we can let $z = iy$, for any $y > M$. Using the monotonic property of b_n, we see that

$$|F_\mu^{\circ j}(ib_n y) - ib_n y| \leq b_n y, \qquad 0 \leq j \leq n-1, \quad y > M.$$

We can write F_μ in Nevanlinna's form so that $F_\mu(z) = z - \gamma + A(z)$, where

$$A(z) = \int_{\mathbb{R}} \frac{1+tz}{t-z} d\sigma(t), \qquad z \in \mathbb{C}^+.$$

One can show that

$$\left|A(F_\mu^{\circ j}(ib_n y) - A(ib_n y)\right| \leq 2\Im A(F_\mu^{\circ j}(ib_n y)), \qquad 0 \leq j \leq n-1.$$

Thus, we have that

$$\frac{1}{b_n}\sum_{j=0}^{n-1}\left|A(F_\mu^{\circ j}(ib_ny) - A(ib_ny)\right| \leq \frac{2}{b_n}\Im\sum_{j=0}^{n-1}A(F_\mu^{\circ j}(ib_ny)).$$

Further, notice that

$$\begin{aligned}\frac{1}{b_n}\sum_{j=0}^{n-1}A(F_\mu^{\circ j}(b_nz)) &= \frac{1}{b_n}\sum_{j=0}^{n-1}\left(F_\mu^{\circ j+1}(b_nz) - \gamma - F_\mu^{\circ j}(b_nz)\right)\\ &= \frac{1}{b_n}\left(F_\mu^{\circ n}(b_nz) - b_nz - n\gamma\right)\\ &= F_{\mu_n}(z) - z - \frac{n}{b_n}\gamma.\end{aligned}$$

Thus

$$\left|\frac{1}{b_n}\sum_{j=0}^{n-1}A(F_\mu^{\circ j}(ib_ny)) - A(ib_ny)\right| \leq \frac{1}{b_n}\sum_{j=0}^{n-1}\left|A(F_\mu^{\circ j}(ib_ny)) - A(ib_ny)\right| \leq 2\Im\left(F_{\mu_n}(z) - z\right),$$

which tends to 0 as $n \to \infty$. So we have that

$$\lim_{n\to\infty} F_{\mu_n}(z) - z + \frac{n}{b_n}(\gamma + A(b_nz)) = 0.$$

But, since μ_n converges weakly to δ_a, we have that $\lim_{n\to\infty} F_{\mu_n}(z) - z = a$. So

$$\lim_{n\to\infty}\frac{n}{b_n}(\gamma + A(b_nz)) = -a. \tag{3.2}$$

Now note that

$$n\left(F_{D_{\frac{1}{b_n}}\mu}(z) - z\right) = \frac{n}{b_n}F_\mu(b_nz) - nz = \frac{n}{b_n}\left(\gamma + A(b_nz)\right) \xrightarrow[n\to\infty]{} -a. \tag{3.3}$$

Also note that since $D_{\frac{1}{b_n}}\mu$ weakly converges to δ_0, we have

$$F_{D_{\frac{1}{b_n}}\mu}(z) - z = \left[\int_{\mathbb{R}}\frac{tz}{t - b_nz}d\mu(t)\right](1 + \epsilon_n(z)), \tag{3.4}$$

where $\epsilon_n(z) \to 0$ as $n \to \infty$. Plugging $z = i$ into equation (3.4) and taking $n \to \infty$, we get the first implication (from equation (3.3)).

As for the converse, we just give an overview of the proof. See [15] for details. We use the notation $F_n = F_{D_{\frac{1}{b_n}}\mu}$. The idea is to consider a particular set U and show that for any $z \in U$, we can bound $|F_n(z) - z|$ by a distance of $\frac{c}{n}$, where c is a constant that depends on a and μ. Then we consider a subset $U_0 \subset U$ such that for $z \in U_0$, we can show that $F_n^{\circ j}(z) \in U$ for $1 \leq j \leq n$. We use this to show that $F_n^{\circ n}(z)$ converges to $z - a$ as $n \to \infty$, and the result follows. □

The main result in [1] is the monotone analogue to Theorem 3.9. Using previous results, the author defines a measure $\nu_{\triangleright}^{\gamma,\sigma}$. Unlike the Boolean, free and classical analogues, $\nu_{\triangleright}^{\gamma,\sigma}$ cannot be written in terms of a simple formula. The paper goes on to prove that

$$\mu_n \triangleright \mu_n \triangleright \cdots \triangleright \mu_n \quad \text{converges weakly to } \nu_{\triangleright}^{\gamma,\sigma}$$

if and only if

$$\mu_n * \mu_n * \cdots * \mu_n \quad \text{converges weakly to } \nu_{*}^{\gamma,\sigma}.$$

The proof relies on the Chernoff product formula, which we will not discuss here. For more details, see the original paper [1].

The results stated above show necessary and sufficient conditions for a triangular array of distribution to follow the weak law of large numbers for classical, free and Boolean convolution. However, for monotonic convolution, the results are only for the identically distributed case. In the next chapter, we provide a condition that guarantees that a triangular arrow follows the weak law of large numbers for monotone convolution. However, this condition is not a necessary condition as we will assume finite variance, which by the previous results in not a necessary condition in the identically distributed case (and hence cannot be for the non-identically distributed case).

Chapter 4

Law Of Large Numbers For Monotone Convolution

In this chapter we will state our main result and the proof.

Definition 4.1. A sequence of distributions $\{\mu_n\}_{n=1}^{\infty}$ is said to be *stable* if there exists real numbers $\{a_n\}_{n=1}^{\infty}$ such that

$$\lim_{n\to\infty} \mu_n(\{t : |a_n - t| > \epsilon\}) = 0,$$

for any $\epsilon > 0$.

Theorem 4.2. *Let $\{\mu_n\}_{n=1}^{\infty}$ be a sequence of probability distributions with finite variances, and let $\{b_n\}_{n=1}^{\infty}$ be an increasing sequence of positive real numbers such that $b_n \to \infty$. Further, suppose that*

$$\sum_{n=1}^{\infty} \frac{var(\mu_n)}{b_n^2} < \infty. \tag{4.1}$$

Then the sequence

$$D_{\frac{1}{b_n}}(\mu_1 \rhd \mu_2 \rhd \cdots \rhd \mu_n)$$

is stable with asymptotic constants

$$a_n = \frac{1}{b_n}\sum_{k=1}^{n} m(\mu_k). \tag{4.2}$$

Before going on to the proof, let us make a few remarks. First, we would like to remark that condition 4.1 is not a necessary condition. Currently, there does not exist a theorem that provides necessary and sufficient conditions for a non-identically distributed but monotonically independent sequence of random variables. The proof for the weak law of large numbers for free (or Boolean) relies heavily on the Voiculescu transform (or E-transform). Trying to use the F-transform for the monotone weak law of large numbers is much more difficult as it relies on the composition of functions (rather than addition of functions in the free and Boolean case). Thus, the proof we present here relies on martingales.

The proof of Theorem 4.2 presented here is not unique; in particular, there exists a direct proof that relies on Chebyshev's inequality. However, the proof in this thesis shows how the concept of martingales can be applied to monotone convolution, which suggests that Markov chains and martingales can be used to prove other limit theorems for monotone convolution.

Let us look at some consequences of Theorem 4.2. First, Theorem 4.2 provides a law of large numbers for monotone convolution. To see this, let $\{\mu_n\}_{n=1}^{\infty}$ be a sequence of probability measures satisfying condition (4.1), and suppose that the asymptotic constants a_n in (4.2) converge to a real number a. Then the triangular array defined by

$$\mu_{nj} = D_{\frac{1}{b_n}}\mu_j, \qquad 1 \leq n, \quad 1 \leq j \leq n,$$

follows the weak law of large numbers for monotone convolution.

In addition, note that Theorem 4.2 implies that if $\{X_n\}_{n=1}^{\infty}$ is a monotonically independent sequence of identically distributed random variables with finite variances, then the average

$$\frac{X_1 + X_2 + \cdots + X_n}{n}$$

converges in distribution to a point mass. This is the result obtained from Theorem 3.17 with the extra condition of finite variance.

4.1 Proof of Theorem 4.2

In this section we prove Theorem 4.2, which is the main result of this thesis.

Proof of Theorem 4.2. Let us consider a sequence of distributions $\{\mu_n\}_{n=1}^{\infty}$ as in Theorem 4.2 . Let $\mathcal{B}$ denote the Borel σ-field, and consider the function $p_n : \mathbb{R} \times \mathcal{B} \to \mathbb{R}$ defined by

$$p_n(x, B) = \delta_x \rhd \mu_n(B).$$

We will now show that p_n is a transition probability. It is clear that for any x, $p_n(x, \cdot) = \delta_x \triangleright \mu_n$ is a probability measure. So all that remains to be shown is that for any $B \in \mathcal{B}$ the function $p_n(\cdot, B)$ is Borel measurable. To do this we will use Dynkin's $\pi - \lambda$ theorem.

Let n be a fixed positive integer, and let T be the operator defined by

$$Tf(x) = \int_{t \in \mathbb{R}} f(t) p_n(x, dt)$$

for any bounded Borel measurable function f. Also, let $\mathcal{H}$ denote the set of all Borel measurable functions f such that Tf is also Borel measurable. Note that any bounded continuous function $f : \mathbb{R} \to \mathbb{R}$ is in $\mathcal{H}$ since the map $x \to \delta_x \triangleright \mu$ is weakly continuous. Now consider the set $\mathcal{L} = \{B \in \mathcal{B} : I_B \in \mathcal{H}\}$. Notice that for any $B \in \mathcal{L}$, we have that $TI_B = p_n(\cdot, B)$ is Borel measurable. Now we will show that $\mathcal{L} = \mathcal{B}$. To do this, consider the set of all open intervals

$$\mathcal{P} = \{(a, b) | a, b \in \mathbb{R} \text{ and } a < b\} \cup \{\emptyset\}.$$

This set is clearly a π-system (i.e. closed under finite intersection), and it is well known that the σ-algebra generated by $\mathcal{P}$ is the Borel σ-algebra. Given any interval $E \in \mathcal{P}$, we can find a sequence of continuous functions $f_n \in \mathcal{H}$ such that $f_n(x) \geq f_{n-1}(x) \geq 0$ and $\lim_{n\to\infty} f_n(x) = I_E(x)$ for all $x \in \mathbb{R}$. Then the monotone convergence theorem (see e.g. [6]) states that $I_E \in \mathcal{H}$. Thus we have that $\mathcal{P} \subset \mathcal{L}$.

The constant function $1_\mathbb{R}$ is Borel measurable and $T1_\mathbb{R} = 1_\mathbb{R}$. Thus, we see that $\mathbb{R} \in \mathcal{L}$. Also, since the difference of two measurable function is measurable, for any $A \in \mathcal{L}$ we must have $A^c \in \mathcal{L}$. Finally, we have that if $\{A_n\}_{n=1}^\infty \subset \mathcal{L}$ where $A_n \subset A_{n+1}$, then $\{I_{A_n}\}_{n=1}^\infty \subset \mathcal{H}$ is an increasing sequence. It follows from the monotone convergence theorem that the limit 1_A must belong to $\mathcal{H}$, and thus the limit $A = \bigcup_n A_n \in \mathcal{L}$. Therefore, we have that $\mathcal{L}$ is a λ-system. So, by Dynkin's $\pi - \lambda$ theorem (Theorem 2.47),

$$\mathcal{B} = \sigma(\mathcal{P}) \subset \mathcal{L}.$$

But clearly $\mathcal{L} \subset \mathcal{B}$, and thus we have proven $\mathcal{L} = \mathcal{B}$. Hence, by our previous remark, $p_n(\cdot, B)$ is a Borel measurable function and thus a transition probability.

We can construct a Markov chain $\{X_n\}_{n=1}^\infty$ with the initial distribution μ_1 and transition probabilities $\{p_n\}_{n=2}^\infty$ by using Kolomogorov's extension theorem.

The joint distribution of the Markov chain is given by

$$Pr(X_1 \in B_1, \cdots, X_n \in B_n) = \int\limits_{t_1 \in B_1} \int\limits_{t_2 \in B_2} \cdots \int\limits_{t_n \in B_n} p_n(t_{n-1}, dt_n) \cdots p_3(t_2, dt_3) p_2(t_1, dt_2) d\mu_1(t_1).$$

Let us show that this does indeed satisfy the consistency condition for Kolmogorov's extension theorem (Theorem 2.48. Notice that

$$\begin{aligned}
&Pr(X_1 \in B_1, \ldots, X_n \in B_n, X_{n+1} \in \mathbb{R}) \\
&= \int\limits_{t_1 \in B_1} \cdots \int\limits_{t_n \in B_n} \int\limits_{t_{n+1} \in \mathbb{R}} p_{n+1}(t_n, dt_{n+1}) p_n(t_{n-1}, dt_n) \cdots p_2(t_1, dt_2) d\mu_1(t_1) \\
&= \int\limits_{t_1 \in B_1} \cdots \int\limits_{t_n \in B_n} \delta_{t_n} \triangleright \mu_{n+1}(\mathbb{R}) p_n(t_{n-1}, dt) \cdots p_2(t_1, dt_2) d\mu_1(t_1) \\
&= Pr(X_1 \in B_1, \ldots, X_n \in B_n).
\end{aligned}$$

To find the distribution of X_n first notice that

$$\begin{aligned}
G_{\mu \triangleright \nu}(z) &= G_\mu \left(F_\nu(z) \right) \\
&= \int_{x \in \mathbb{R}} \frac{1}{F_\nu(z) - x} d\mu(x) \\
&= \int_{x \in \mathbb{R}} G_{\delta_x \triangleright \nu}(z) d\mu(x) \\
&= \int_{x \in \mathbb{R}} \int_{t \in \mathbb{R}} \frac{1}{z - t} \, d\delta_x \triangleright \nu(t) d\mu(x) \\
&= \int_{t \in \mathbb{R}} \frac{1}{z - t} \int_{x \in \mathbb{R}} d\delta_x \triangleright \nu(t) \mu(dx).
\end{aligned}$$

Since the last line is a Cauchy transform, and since the Cauchy transform of a measure is unique, we have the following equality:

$$\mu \triangleright \nu(B) = \int_{x \in \mathbb{R}} d\delta_x \triangleright \nu(B) \mu(dx) \qquad \text{for any B } \in \mathcal{B}.$$

Recalling our definition of p_n, we see that

$$\begin{aligned}
&Pr(X_1 \in \mathbb{R}, \dots, X_{n-1} \in \mathbb{R}, X_n \in B_n) \\
&= \int\limits_{t_n \in B_n} \int\limits_{t_{n-1} \in \mathbb{R}} p_n(t_{n-1}, dt_n) \cdots \int\limits_{t_2 \in \mathbb{R}} p_3(t_2, dt_3) \int\limits_{t_1 \in \mathbb{R}} p_2(t_1, dt_2) d\mu_1(t_1) \\
&= \int\limits_{t_n \in B_n} \int\limits_{t_{n-1} \in \mathbb{R}} p_n(t_{n-1}, dt_n) \cdots \int\limits_{t_2 \in \mathbb{R}} p_3(t_2, dt_3) d\mu_1 \triangleright \mu_2(t_2) \\
&\quad \vdots \\
&= \mu_1 \triangleright \mu_2 \triangleright \cdots \triangleright \mu_n(B_n).
\end{aligned}$$

Thus, we see that the distribution of X_n is exactly the measure $\mu_1 \triangleright \mu_2 \cdots \triangleright \mu_n$.

While the Markov chain $\{X_n\}_{n=1}^{\infty}$ is not a martingale, it is very close to being one. Let us calculate the conditional expectation. Note

$$E[X_n | X_{n-1}] = \int\limits_{t \in \mathbb{R}} t p_n(x, dt) \Big|_{x = X_n} = m(\delta_x \triangleright \mu_n) \Big|_{x = X_n}. \tag{4.3}$$

To calculate the mean of the measure $\delta_x \triangleright \mu_n$, we will look at the F-transform. Recall that

$$F_{\delta_x \triangleright \mu_n}(z) = F_{\mu_n}(z) - x, \qquad z \in \mathbb{C}^+.$$

Writing both sides of this equation in Nevanlinna's form, we have that

$$z - m(\delta_x \triangleright \mu) + \int_{t \in \mathbb{R}} \frac{1}{t - z} d\sigma(t) = z - m(\mu) - x + \int_{t \in \mathbb{R}} \frac{1}{t - z} d\sigma'(t).$$

Since Nevanlinna's form is unique, we get the following two identities:

$$m(\delta_x \triangleright \mu_n) = m(\mu_n) + x, \qquad n \geq 1, \tag{4.4}$$

$$\sigma = \sigma'. \tag{4.5}$$

Combining (4.3) and (4.4), we have that

$$E[X_n | X_{n-1}] = X_{n-1} + m(\mu_n). \tag{4.6}$$

We construct a martingale $\{Y_n\}_{n=1}^{\infty}$ by shifting X_n; that is, we define

$$Y_n = X_n - \sum_{k=1}^{n} m(\mu_k), \qquad n \geq 1.$$

Notice the sigma field generated by $X_n, X_{n-1}, \ldots, X_1$ is the same as the sigma field generated by $Y_n, Y_{n-1}, \ldots, Y_1$. Thus, using (4.6) and (2.11), we have

$$\begin{aligned} E[Y_{n+1}|Y_n, Y_{n-1}, \ldots, Y_1] &= E[X_{n+1} - \sum_{k=1}^{n+1} m(\mu_k)|X_n, X_{n-1}, \ldots, X_1] \\ &= X_n + m(\mu_{n+1}) - \sum_{k=1}^{n+1} m(\mu_k) \\ &= Y_n. \end{aligned}$$

So indeed $\{Y_n\}_{n=1}^{\infty}$ is a martingale. Now, let us look at the martingale difference defined by

$$\begin{aligned} &Z_1 = Y_1; \\ &Z_n = Y_n - Y_{n-1}, \qquad n \geq 2. \end{aligned}$$

Again we have that the algebra generated from $Z_1, Z_2, \ldots, Z_n$ is the same as the algebra generated from $X_1, X_2, \ldots, X_n$. Thus,

$$E[Z_{n+1}|Z_1, Z_2, \ldots, Z_n] = 0, \qquad n \geq 1. \tag{4.7}$$

Next, let us consider the martingale given by

$$S_n = \sum_{k=1}^{n} \frac{1}{b_k} Z_k, \qquad n \geq 1.$$

This is indeed a martingale, since the sigma algebra generated by $S_1, \ldots, S_n$ is the same as the algebra generated by $Z_1, \ldots, Z_n$. Hence, using (4.7), we have that

$$E[S_{n+1}|S_n, \cdots, S_1] = E\left[S_n + \frac{1}{b_{n+1}} Z_{n+1}|Z_1, \cdots, Z_n\right] = S_n, \qquad n \geq 1.$$

For a fixed n, we will now calculate the expectation of S_n^2. To do so, notice that

$$E[X_n^2|X_{n-1}] = \int_{t \in \mathbb{R}} t^2 p_n(x, dt)\bigg|_{x=X_n} = m_2(\delta_x \triangleright \mu_n)\bigg|_{x=X_n}. \tag{4.8}$$

Equation (4.5) implies that $var(\mu_n) = var(\delta_x \triangleright \mu_n)$. Combining this with equations (4.8) and (4.3) implies

$$\begin{aligned} E[X_n^2|X_{n-1}] &= var(\delta_x \triangleright \mu_n) + m(\delta_x \triangleright \mu_n)^2 \\ &= m_2(\mu_n) - m(\mu_n)^2 + (m(\mu_n)^2 + X_{n-1})^2 \\ &= X_{n-1}^2 + 2m(\mu_n)X_{n-1} + m_2(\mu_n). \end{aligned}$$

Thus, using the definition of Z_n and (4.6), we have

$$\begin{aligned}
E[Z_n^2|Z_{n-1}, Z_{n-2}, \ldots, Z_1] &= E[(X_n - X_{n-1} - m(\mu_n))^2|X_{n-1}, X_{n-2}, \ldots, X_1] \\
&= E[X_n^2|X_{n-1}] - 2(X_{n-1} + \mu_n)E[X_n|X_{n-1}] + (X_{n-1} + \mu_n)^2 \\
&= X_{n-1}^2 + 2m(\mu_n)X_{n-1} + m_2(\mu_n) - (X_{n-1} + \mu_n)^2 \\
&= var(\mu_n).
\end{aligned}$$

Using the conditional expectation (4.7), we see that the conditional expectation of S_n^2 is given by

$$\begin{aligned}
E[S_n^2|Z_{n-1}, \ldots, Z_1] &= E\big[(\frac{1}{b_n}Z_n + S_{n-1})^2|Z_{n-1}, \ldots, Z_1\big] \\
&= \frac{1}{b_n^2}E[Z_n^2] + \frac{2}{b_n}S_{n-1}E[Z_n|Z_{n-1}, \ldots, Z_1] + S_{n-1}^2 \\
&= \frac{1}{b_n^2}var(\mu_n) + S_{n-1}^2.
\end{aligned}$$

Thus, we have that

$$\begin{aligned}
E[S_n^2] &= E\Big[E[S_n^2|Z_{n-1}, \ldots, Z_1]\Big] \\
&= \frac{1}{b_n^2}var(\mu_n) + E[S_{n-1}^2].
\end{aligned}$$

An induction argument shows

$$E[S_n^2] = \sum_{k=1}^{n} \frac{1}{b_k^2} var(\mu_k),$$

which is finite for all n by the hypothesis. Thus, the martingale convergence theorem (Theorem 2.49 together with Kronecker's lemma (see e.g. **??**) implies that

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^{n} Z_k = 0 \qquad a.s.$$

But notice that

$$\sum_{k=1}^{n} Z_k = Y_n = X_n - \sum_{k=1}^{n} m(\mu_k).$$

Thus, if we define $a_n = \frac{1}{b_n}\sum_{k=1}^{n} m(\mu_k)$, we have that

$$\lim_{n\to\infty} \Pr\left(\left|\frac{1}{b_n}X_n - a_n\right| > \epsilon\right) = 0$$

for any $\epsilon > 0$, which gives the desired result. □

Thus, we have found a sufficient condition for a triangular array to follow the weak law of large numbers. another consequence of our result is given a sequence of monotonically independent random variables, $\{X_n\}_{n=1}^{\infty}$ with bounded variance then the sum

$$\frac{X_1 + X_2 + \cdots + X_n}{n}$$

converges (in distribution) to a constant.

4.2 Future research

We would like to find answers to the general limit problem for monotone convolution. That is, we would like to find necessary and sufficient conditions for an infinitesimal triangular array $\{\mu_{nj} : 1 \leq n, 1 \leq j \leq k_n\}$ such that $\mu_{n1} \triangleright \mu_{n2} \triangleright \cdots \triangleright \mu_{nk_n}$ has a limiting distribution ν. Further, we would also like to know what properties the limiting distribution ν must posses.

The weak law of large numbers we proved in this thesis gives a partial answer to the first question, when we require the limiting distribution to be a point mass. However, even in this case our result is only a partial result, since, for example, distributions without finite variance can obey the weak large numbers.

In the classical case, it has been shown that the limiting distribution must be $*$-infinitely divisible. That is, if ν is the limiting distribution, then for each k there must exist ν_k such that

$$\nu = \underbrace{\nu_k * \nu_k * \cdots \nu_k}_{\text{k times}}.$$

Similarly, a limiting distribution for free convolution was shown to be $\boxplus$-infinitely divisible. However, as of yet, there are no results showing that the limiting distribution for monotone convolutions must be $\triangleright$-infinitely divisible.

References

[1] M. Anshelevich and J. D. Williams, *Limit theorems for monotonic convolution and the Chernoff product formula*, Int. Math. Res. Not. (2013). doi:10.1093/imrn/rnt018.

[2] H. Bercovici and J.-C. Wang, *The asymptotic behavior for free additive convolution*, Oper. Matrices **2** (2008), no. 1, 115-124.

[3] H. Bercovici, V. Pata, with an appendix by P. Biane, *Stable laws and domains of attraction in free probability theory*, Annals of Mathematics, **149** (1999), 10231060.

[4] H. Bercovici, D. Voiculescu *Lvy-Hinin type theorems for multiplicative and additive free convolution*, Pacific Journal of Mathematics **153** (1992), no. 2, 217–248.

[5] Bozejko, M., *Positive definite functions on the free group and the non-commutative Riesz product*, Boll. Un. Mat. Ital., 5A (6): 1321 (1986).

[6] R. Durrett, *Probability: Theory and Examples*, Wadsworth Inc., Belmont California, 1991.

[7] U. Franz, *Monotone and Boolean convolutions for non-compactly supported probability measures*, Indiana Univ. Math. J. **58** (2009), no. 3, 1151-1185.

[8] B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sums of independent random variables*, Addison-Wesley Publishing Company, Cambridge, 1954.

[9] T. Hasebe, *On monotone convolution and monotone infinite divisibility* Dim. Anal. Quantum Probab. Relat. Topics **13** (2010), no. 1, 111-131.

[10] Isak Wulff Mottelson, *Introduction to non-commutative probability* (2012).

[11] N. Muraki, *The five independences as natural products*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **6** (2003), no. 3, 337-371.

[12] N. Muraki, *Monotonic independence, monotonic central limit theorem and monotonic law of small numbers*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **4** (2001), no. 1, 39-58.

[13] V. Pata, *The Central Limit Theorem for Free Additive Convolution* J. of Funct. Anal. **140**, Issue 2, (1996), 359380.

[14] J.-C. Wang, *Limit laws for Boolean convolutions*, Pacific J. Math.237, No. 2 (2008),349371.

[15] J.-C. Wang, *Strict limit types for monotone convolution*, J. Funct. Anal. **262** (2012), no. 1, 35-58.

[16] J.-C. Wang, Enzo Wendler, *Law of large numbers for monotone convolution* Probability and Mathematical Statistics Vol. 33, Fasc. 2 (2013), pp. 225231.

[17] Williams, D. *Probability with Martingales* Cambridge University Press, (1991), ISBN 0-521-40605-6.