Limit Theorems in Free Probability Theory

by

Vladislav Kargin

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Mathematics
New York University
May 2008

Gerard Ben Arous

to my mother, Valentina Alexandrovna Guseva

Acknowledgements

I am very grateful to my adviser, Professor Gerard Ben Arous. His confidence that there must be free Poisson-Dirichlet process and free extremes were instrumental in my search for these mathematical objects.

Without encouragement by Professor Raghu Varadhan, this dissertation would hardly exist.

I am thankful to Professor Charles Newman for stimulating discussions of products of random matrices.

I am grateful to all people at the Courant Institute of Mathematical Sciences who jointly create a stimulating environment of scientific endeavor.

Diana Bloom helped me with editing many mathematical texts.

Last but not least, I am grateful to my wife Yuliana who convinced me that attempting a Ph.D. in mathematics is a reasonable idea.

Abstract

The first part of this dissertation is devoted to the systematic exposition of the fundamentals of free probability theory.

The second part studies various aspects of the free Central Limit Theorem. The particular contributions of this part are as follows: (a) The free Central Limit Theorem holds for non-commutative random variables which satisfy a condition weaker than freeness. (b) The speed of convergence in the free Central Limit Theorem is the same as in the classical case, which is shown by an analogue of the Berry-Esseen inequality. (c) An estimate on the support of free additive convolutions is established.

The third part investigates products of free operators. In particular, it studies the growth in norm of products of free operators and gives an infinite-dimensional analogue of the Furstenberg-Kesten theorem about products of random matrices. This part also introduces the Lyapunov exponents of products of free operators and expresses them in terms of Voiculescu's S-transform. Finally, it gives the necessary and sufficient conditions for products of free unitary operators to converge in distribution to the uniform law on the unit disc.

The fourth part of the dissertation introduces the concept of the free point process and proves a theorem about the convergence of this process to the free Poisson measure. The free point processes are used to define free extremes, which extend the concept of the free maximum introduced earlier by Ben Arous and Voiculescu. A theorem about the convergence of free extremes is proven, which is similar to the corresponding theorem in the classical theory but results in a different set of limit laws.

Contents

List of Figures

1 Introduction

Free probability theory arose from the study of the following problems:

A) Consider the free group with k generators, $F(k)$, and the corresponding group algebra with complex coefficients. The left action of this group algebra on $L^2(F(k))$ (i.e., the Hilbert space of square-summable functions on $F(k)$) makes this algebra an algebra of operators. The closure of this algebra in the weak operator topology is a von Neumann algebra of type II_1, which we will denote as $\mathcal{N}(F(k))$. Many natural operators in this algebra are non-compact. Since they cannot be approximated in norm by finite-dimensional operators, it is hard to study their spectral properties by the usual approximation techniques. Can we find an efficient algorithm to compute the spectral properties of these operators?

B) Suppose we approximate a self-adjoint operator A in $\mathcal{N}(F(k))$ by finite matrices M_n where by approximation we mean that the spectral distributions of M_n approach the spectral distribution of A as $n \to \infty$. Can we measure how the quality of approximation improves as $n \to \infty$?

C) Von Neumann algebras of type II_1 can be thought of as generalizations of finite mass measure spaces. In the classical case, we can upgrade measure theory with the concept of independence and in this way obtain probability theory. Is there an independence concept suitable for von Neumann algebras? Recall that all measure spaces without atoms are isomorphic to each other. On the other hand, there is a rich theory of how to classify measure-preserving transformations in measure spaces. Consider two von Neumann algebras: $\mathcal{N}(F(k))$ and $\mathcal{N}(F(l))$, with $k \neq l$. They can be thought of as two non-commutative measure spaces. Are they isomorphic? What about transformations in a non-commutative measure space? What are conditions under which they are isomorphic?

D) Let G_q be a homogeneous tree of degree q, that is, a tree in which every vertex is an end point of exactly q edges. Assume that these edges are labeled by integers from 1 to q, and let p_i be non-negative real numbers such that $\sum_{i=1}^{q} p_i = 1$. Then we can define a random walk on the tree by the rule that a particle at a given vertex travels along the edge i with probability p_i. The transition matrix of this random walk defines an operator on $l^2(G_q)$, which can be thought of as a non-homogeneous Laplace operator. What is the spectral measure of this operator?

The tree G_q is the Cayley graph of the free product $Z_2 * Z_2 * \dots * Z_2$ with q elements in the product. Similar questions can also be asked for random walks on Cayley graphs of more general free product groups.

E) Suppose $\{X_n\}$ is a sequence of random n-by-n Hermitian matrices. Suppose that their probability distributions are invariant under unitary transformations and that the empirical distribution of the eigenvalues of X_n converges to a probability measure μ as n grows to infinity. Let $\{Y_n\}$ be a similar sequence of independent matrices with the empirical distribution of eigenvalues converging to a probability measure ν. Is it true that the empirical distribution of eigenvalues of $X_n + Y_n$ converges to a limit? How can we compute the limit using only μ and ν?

From this list of problems it appears that there is a need to study infinite-dimensional objects that are, in a sense, limits of independent random matrices of growing dimension.

As a response to this need, a new field emerged on the border between operator algebra and probability theories. The field was christened free probability theory by its creator, Dan Voiculescu. It developed into a complex theory, which in many respects parallels the usual probability theory. Sums and products of freely independent operators correspond to certain convolutions of their spectral measures, and free probability studies the properties of these convolutions. The theory includes analogues of characteristic functions, the Central Limit Theorem, the Law of Large Numbers, and many other concepts from classical probability theory. However, the limiting laws are different and their proofs proceed along quite different lines.

This theory has interesting connections with the theory of random matrices and is used by engineers because it significantly simplifies many calculations associated with random matrices. See Edelman and Rao (2005) for a review of applications of random matrices and free probability in numerical analysis. Free probability is also useful in statistics; see, for example, Rao et al. (2008). In another direction, an interesting application of this theory to the theory of representations was discovered by Biane (1998).

Another beautiful part of free probability theory is the theory of free entropy. In free probability, infinite-dimensional operators can be approximated by finite dimensional matrices where approximation is meant in the sense of convergence of their spectral distributions. In many respects this is similar to a theory of approximation of continuous probability measures by measures supported on finite sets. In both classical and free cases, there is a natural quantity that measures the quality of approximation and which is called entropy. What is especially surprising is that free entropy is closely related to the concept of free independence. For example, free entropy is addititive with respect to joining several freely independent variables in one vector.

The theory of free entropy is useful in the study of free operator algebras because it provides a new way to approximate infinite-dimensional operators and study the quality of these approximations. These methods have led to breakthrough results in operator theory (see Voiculescu (1990) and Haagerup and Thorbjornsen (2005)).

This text will be devoted mostly to an account of free probability theory from the point of view of its similarity to classical probability theory. Almost no attention will be paid to the theory of free entropy. Instead, the focus is on limit theorems for sums and products of free random operators. The text will present some new results and also give quantitative versions of some of the known limit theorems (that is, versions that provide quantitative bounds on the speed of convergence).

Another focal point of this dissertation is the study of free point processes and free extremes. These concepts are new and have a potential to explain why many results in free probability theory are strikingly similar to the corresponding results in classical probability theory.

In the following subsections of the introduction we briefly outline the history of the subject and indicate which results in this dissertation are new.

1.1 Historical remarks

Free probability theory was invented by Dan Voiculescu in the early 1980s when he researched von Neumann algebras of type II_1. The main motivation was to study the properties of free products of these algebras. Voiculescu formulated an axiomatic definition of what it means for two operators to be free. He pointed out the analogy to the concept of independence in classical probability theory and suggested calling free operators free random variables.

Even in the earliest of his papers, Voiculescu (1983), some fundamental results were established. In particular, it was proved that the sums of free random variables converge in distribution to the semicircle random law. As a next step, Voiculescu developed an analytic method for computation of moments of the sum and of the product of two free random variables (1986 and 1987)[1]. An application of these methods to questions in operator algebra theory was given in Voiculescu (1990). Extremely fruitful for further progress of the theory was the realization that free probability is connected with random matrix theory. In Voiculescu (1991) it was proved that the

[1] At about the same time similar formulas for additive free convolutions were independently developed by researchers who studied random walks on free products of discrete groups; see, e.g., McLaughlin (1986).

Wigner random matrices of increasing dimension become asymptotically free and this can be interpreted as the proper explanation for why the Wigner semicircle law holds for the spectral distribution of large random matrices. An offspring of this realization was a definition of free entropy in Voiculescu (1993) and Voiculescu (1994), which explains how well infinite-dimensional operators in type II_1 algebras can be approximated by finite matrices. This was followed by a series of breakthrough results for von Neumann algebras (Voiculescu (1996a)).

Addition of free self-adjoint random variables induces a convolution of probability measures which is quite different from the standard convolution. This new convolution, named the free additive convolution, became an object of study by Bercovici, Biane, Maassen, Pata, Speicher, Voiculescu himself, and others. It was found that this concept is analogous in many respects to the usual convolution of probability measures and closely related to certain classical problems of complex analysis. In particular, the concept of free additive convolution was extended to unbounded probability measures, and many properties of this convolution were investigated. (see Maassen (1992), Bercovici and Voiculescu (1992), Bercovici and Voiculescu (1993), Bercovici and Pata (1996), Bercovici et al. (1999), Belinschi and Bercovici (2004), Barndorff-Nielsen and Thorbjornsen (2005), and Ben Arous and Voiculescu (2006)).

These problems were also investigated for the free multiplicative convolution that arises from products of free random variables. In particular, Bercovici and Voiculescu (1992) classified the free infinitely-divisible laws for the measures on the real half-line and on the unit circle. The progress here, however, has been less significant than for free additive convolution.

In another development, Speicher (1990) investigated the relation of free additive and multiplicative convolutions with combinatorics. He introduced a concept of free cumulants and related this concept to a theory of non-crossing partitions. One of the successes of this method was a proof of a certain free analogue of the multivariate CLT. Using Speicher's techniques and following some early contributions by Voiculescu, Biane developed a theory of free stochastic processes.

The relationship between free probability theory and random matrices was also actively investigated. For example, Ben Arous and Guionnet (1997) related free entropy to the large deviation property for random matrices.

1.2 Summary of original contributions

Free multiplicative convolutions

While sums of free random variables and the corresponding limit theorems have been thoroughly studied, the multiplication of free random variables has been less researched. Let $X_1, \ldots, X_n$ be free and identically distributed operators in a von Neumann algebra with trace (expectation) E. What are the properties of $\Pi_n = X_n \ldots X_1$ for large n?

First, I have proved that

$$\lim_{n\to\infty} n^{-1} \log \|\Pi_n\| = \frac{1}{2} \log \left(E\left(X_1^* X_1\right)\right).$$

This result is in agreement with a previous result by Cohen and Newman on the norm of products of i.i.d. $N \times N$ random matrices, in the situation when $N \to \infty$.

Next, assume that $E\left(X_1^* X_1\right) = 1$. In this case, I have proved that

$$\lim_{n\to\infty} \sup \frac{1}{\sqrt{n}} \|\Pi_n\| \leq c\sqrt{v},$$

where $v = E\left(\left(X_1^* X_1\right)^2\right) - 1$ and c is a constant.

In order to understand the behavior of the singular values of the product Π_n in the bulk, I have defined the Lyapunov exponents for products of free self-adjoint random variables. To understand why this concept is helpful, consider the finite-dimensional situation. The sum of the logarithms of the k largest singular values of an operator A can be computed as follows:

$$\log \lambda_1 + \log \lambda_2 + ... + \log \lambda_k = \log \sup_{v_1,\ldots,v_k} vol\left(Av_1, Av_2, ...Av_k\right),$$

where $v_1, ..., v_k$ are orthonormal and $vol\left(Av_1, Av_2, ...Av_k\right)$ denote the volume of the parallelepiped spanned by vectors $Av_1, ..., Av_k$. This suggests that we consider the following limit:

$$\lim_{n\to\infty} \frac{1}{n} \log \sup_{v_1,\ldots,v_k} vol\left(\Pi_n v_1, \Pi_n v_2, ...\Pi_n v_k\right).$$

In the theory of products of random matrices, it is proved that this limit exists under a certain assumption on the distribution of matrix entries. Moreover the supremum can be removed: under a mild assumption, the limit is the same for arbitrary choice of the orthonormal vectors $v_1, ..., v_k$ with probability 1.

In the infinite-dimensional case, we can define an analogous expression:

$$\lim_{n\to\infty}\frac{1}{n}\log\sup_{P_t}\det\left(\Pi_n P_t\right), \tag{1}$$

where P_t is a t-dimensional subspace. Here $\det$ denotes a modification (Lück's determinant) of the infinite-dimensional Fuglede-Kadison determinant and it allows us to compute how a given operator changes the "volume element" of an infinite-dimensional subspace. This limit, if it exists, contains all the information about the asymptotic behavior of the singular values of Π_n.

By analogy to the finite-dimensional situation, I have studied the case when the $\sup$ is removed and instead P_t is assumed to be free of all of X_i. So, I have defined the *integrated Lyapunov exponent function* as follows:

$$F\left(t\right)=\lim_{n\to\infty}\frac{1}{n}\log\det\left(\Pi_n P_t\right). \tag{2}$$

I have proved that this is a consistent definition and that the integrated Lyapunov exponent exists for bounded free X_i. In addition, I have derived an explicit formula which relates Lyapunov exponents to Voiculescu's S-transform:

$$F'\left(t\right)=-\frac{1}{2}\log\left(S_{X^*X}\left(-t\right)\right), \tag{3}$$

where S_{X^*X} is the S-transform of X^*X.

This formula allows me to infer a number of results about the Lyapunov exponents, in particular, a formula for the largest Lyapunov exponent and the additivity property of the Lyapunov exponents with respect to the operator product.

An example with a particular choice of X_i recovers the "triangle law" discovered earlier by C. M. Newman in his work on Lyapunov exponents of large random matrices.

The next natural step would be to prove that the limit in (1) exists and coincides with the limit in (2). This would follow from a free probability version of the Oseledec theorem.

These results about Lyapunov exponents of infinite-dimensional operators can be considered a generalization of some of the results of Furstenberg, Kesten, and Oseledec regarding products of random matrices. The usual technique based on Kingman's sub-additive ergodic theorem does not work here because free operators do not form an ergodic stochastic process. Instead, we have to use directly the definition of freeness of operators.

In a somewhat different project, I have studied products of free unitary operators. For this problem, I have derived a necessary and sufficient condition for the convergence of the spectral probability distribution of the product to the uniform distribution on the circle. The necessary condition for convergence is that the product of expectations converges to zero. This condition fails to be sufficient only if all of the following statements hold: (i) exactly one of the operators has zero expectation, (ii) this operator is not uniformly distributed, (iii) the product of the expectations of the remaining operators does not converge to zero.

Free additive convolutions

Recently, there has been great progress in the theory of free additive convolutions. (See, for example, papers by Bercovici, Biane, Maasen, Pata, Speicher, and Voiculescu.) In particular, a theory of free infinitely-divisible and stable laws has been developed. Also, free versions of the Law of Large Numbers ("LLN") and the Central Limit Theorem ("CLT") have been derived.

I have extended the free CLT to the situation when the operators are not free, but "almost" free. In particular, I have devised an example of the situation when operators are not free but the free CLT is still valid for their sequence. This example takes a free group with an infinite number of generators and adds certain relations. Then it uses the method of short cancellations from combinatorial group theory to infer a weakened version of freeness. Finally, the proof uses the Lindeberg approach to the classic CLT to infer the free CLT from this weakened version of freeness.

My other work in this area is focused mostly on making the available results more quantitative. In particular, it was known that the support of normalized free additive convolutions converges to the interval $(-2, 2)$. I have shown that for large n, the support of the n-time free convolution is in $\left(-2 - cn^{-1/2}, 2 + cn^{-1/2}\right)$ and that this rate is optimal.

I have also derived a free version of the Berry-Esseen estimate for the speed of convergence in the CLT. The rate obtained in this result is $n^{-1/2}$, the same as in the classical case. An example shows that this rate cannot be improved without further conditions.This result has been obtained independently and by a different method than a similar result in Chistyakov and Gotze (2006), that also derived a free version of the Berry-Esseen inequality.

Free point processes and free extremes

This part of the dissertation is joint work with Gerard Ben Arous, which was inspired by the previous work of Ben Arous and Voiculescu (2006).

Let X_i be a sequence of identically distributed free variables with the spectral measure μ. Ben Arous and Voiculescu introduced an operation of maximum which takes any n-tuple of self-adjoint operators to another operator (possibly different from each of the original operators). If $X^{(n)} = \max_{1\leq i\leq n}(X_i)$, then a natural question is whether the sequence of the spectral distributions of $\left(X^{(n)} - a_n\right)/b_n$ converges for a certain choice of constants a_n and b_n. Ben Arous and Voiculescu proved that there are only three possible limit laws, which are different from the classical limit laws. Surprisingly, the domains of attraction for these free limit laws are the same as the domains of attraction for the three classical laws. In our work, we have explained this puzzling fact.

In the classical case the convergence of extremes is closely related to the convergence of point processes

$$N_n = \sum \delta_{(X_i - a_n)/b_n}$$

to a Poisson random measure. We have introduced a free probability analogue to the concept of a point process. Namely, a free point process M_n is a linear functional on the space of bounded measurable functions, defined by the formula:

$$\langle M_n, f\rangle = \sum_{i=1}^{n} f(X_{i,n}),$$

where $X_{i,n}$ is an array of free random variables. In application to the theory of free extremes, we use the array $X_{i,n} = (X_i - a_n)/b_n$, where X_i is a sequence of free, identically distributed variables with the spectral probability measure μ.

We have also introduced a concept of weak convergence of a free point process and proved that the free point process corresponding to the measure μ converges if and only if the classical free point process converges. Moreover, we have proved that it converges to an object which was discovered by Voiculescu (1998) and extensively studied by Barndorff-Nielsen and Thorbjornsen (2005), who called it the free Poisson random measure.

Both the condition that ensures the convergence of a free point process and the intensity of the resulting free Poisson measure are exactly the same as in the classical case. It is this fact that is at the root of the phenomenon that the domains of convergence for free and classical extremal convolutions are the same but the limit laws are different. Indeed, while the classical and free point processes associated with a measure μ converge to similar objects under similar conditions, the limiting extremal laws are built in a different way from the classical and free limit Poisson measures.

We have also applied free point processes to further develop the theory of free extremes. In the classical case the k-th order extremal distribution $F^{(k)}(x)$ can be defined as the probability that the corresponding random point process has no more than k points in the interval $[t, \infty)$. This can be codified as the following formula:

$$F^{(k)}(x) = E\left\{\mathbf{1}_{[0,k]}\left(\left\langle N_n, \mathbf{1}_{[t,\infty)}\right\rangle\right)\right\},$$

where $\left\langle N_n, \mathbf{1}_{[t,\infty)}\right\rangle =: \sum_{i=1}^{n} \mathbf{1}_{[t,\infty)}(X_{i,n})$.

This definition has a straightforward generalization to the free case:

$$F_f^{(k)}(x) = E\left\{\mathbf{1}_{[0,k]}\left(\left\langle M_n, \mathbf{1}_{[t,\infty]}\right\rangle\right)\right\}.$$

We call these distribution functions *the k-th order free extremal convolution.* Moreover, it turns out that it is possible to define in a natural way an operator that has $F_f^{(k)}(x)$ as its spectral probability distribution. We call this operator the *k-th order free extreme*. In particular, the 0-th order free extreme convolution corresponds to the free extremal convolution of Ben Arous and Voiculescu.

Using the limit theorem for free point processes, it is possible to prove a limit theorem for k-th order free extremal convolutions. We derive the explicit formulas for the limit laws. The particular case of the 0-th order convolutions corresponds to the limit law derived in Ben Arous and Voiculescu.

The rest of the dissertation is organized as follows. Part I explains the fundamentals of free probability theory. In Sections 2 – 5 we give the basic definitions and examples. In particular, we define non-commutative probability spaces, free independence, and free additive and multiplicative convolutions of probability measures. One of the main tasks of free probability theory is the study of the properties of these convolutions. As an initial step in this direction, we prove Voiculescu's addition and multiplication theorems in Section 6.

Our main tools in the study of free convolutions are analytical properties of the Cauchy transform and related functions. We collect them in Section 7 of Part I and show applications in Section 8. This section discusses measures that are infinitely-divisible with respect to free convolution.

Part II is devoted to the free Central Limit Theorem for additive free convolutions. We give the original proof by Voiculescu in Section 10 and certain extensions in Sections 11, 12, and 13.

Part III is devoted to limits of products of free random variables and multiplicative convolutions. In Section 14, we find the growth rate of the norm of the products.

The results from this section are made more precise in Section 15. In Section 16 we introduce and study the properties of Lyapunov exponents of a sequence of free random variables. And in Section 17 we prove a limit theorem for products of free unitary operators.

Part IV is devoted to the convergence of free point processes and free extremes.

The original contributions are in Sections 11 – 19.

Part I

Fundamentals of Free Probability

What is a probability space? Formally, it is a collection of random variables with an expected value functional. However, the most important block in building a probability theory is an appropriate concept of independence of random variables. In the following sections we will introduce the definitions of these basic concepts for free probability theory. Here is a brief and informal overview.

We will define non-commutative random variables as linear operators on a Hilbert space and the expected value functional as a linear functional on these operators which is similar to the trace functional on matrices.

What should be our notion of non-commutative independence? Consider operators given by two different generators in a representation of an (algebraically) free group. These operators are our model of non-commutative independence. We will say that two operators are freely independent if expectations of their products behave similarly to expectation of products in this model situation. The most important result here is that it is possible to compute the distribution of a sum and a product of two freely independent self-adjoint random variables. This can be done with the help of certain analytic transforms similar to the Fourier transform in the classical case.

2 Non-Commutative Probability Space

Space and Expectation

A *non-commutative probability space* $(\mathcal{A}, E)$ is a C^*-algebra $\mathcal{A}$ and a linear functional E defined on this algebra. An algebra $\mathcal{A}$ is a C^*-algebra if 1) it is an algebra over complex numbers; 2) it is closed under an involution $*$; that is, if $X \in \mathcal{A}$, then $X^* \in \mathcal{A}$ and $(X^*)^* = X$; 3) it has a norm with the following properties: $\|XY\| \leq \|X\| \|Y\|$ and $\|X^*X\| = \|X\|^2$, and 4) it is closed relative to convergence in norm. The algebra $\mathcal{A}$ is unital if it contains the identity element I. If the algebra is closed in the weak topology, then it is called a W^*- or a von Neumann algebra.

Intuitively, C^* algebras are non-commutative generalizations of the algebra of bounded continuous functions on a compact topological space. And von Neumann algebras are generalizations of the algebra of measurable functions.

The linear functional E is called an *expectation*[2]. We will use the notation E to emphasize the analogy with the expectation in classical probability theory. An expectation is assumed to have the following properties:

1. $E(I) = 1$ (If the algebra is not unital, then we require that $\lim_{n\to\infty} E(I_n) = 1$, where I_n is any approximate identity.)

2. $E(X^*) = \overline{E(X)}$, and

3. $X > 0$ implies that $E(X) > 0$.

If an additional property is satisfied that says that $X \geq 0$ and $E(X) = 0$ imply $X = 0$, then the expectation is called *faithful*. If E is continuous with respect to weak convergence of operators than it is called *normal*. (A theorem from operator algebra theory says that the expectation is always continuous with respect to norm convergence.) Finally, if $E(XY) = E(YX)$, then the expectation is called *tracial*, or simply a *trace*. Many of the results of non-commutative probability theory hold without assuming any of these additional properties of the expectation.

Let us give some simple examples of non-commutative probability spaces.

Example 1

A usual probability space $(\Omega, \mathfrak{A}, \mu)$ can be considered as a non-commutative probability space. Let $\mathcal{A}$ be the algebra of all measurable functions on Ω, and let E be the usual integral with respect to the measure μ:

$$E(f) = \int_\Omega f(\omega)\, d\mu(\omega).$$

Then $(\mathcal{A}, E)$ is a non-commutative probability space, although the adjective non-commutative is not very appropriate, since $\mathcal{A}$ is a commutative algebra.

Example 2

Consider an algebra of n by n matrices: $\mathcal{A} = M_n(\mathbb{C})$. Define E as the usual trace normalized by n^{-1}:

$$E(X) = n^{-1}\,\mathrm{Tr}(X).$$

Then $(\mathcal{A}, E)$ is a finite-dimensional non-commutative probability space.

[2] In operator algebra theory it is usually called a *state*. It is curious that Segal introduced the term "state" with the following comment: "[W]e use the term 'state' to mean ... – more commonly, this is called an expectation value in a state..." (See Segal (1947).)

Example 3

Consider all trace-class operators acting on a complex separable Hilbert space H. (An operator X is trace-class if eigenvalues of $|X|$ form a summable sequence.) These operators form an algebra $\mathcal{A}$ and we can take the trace as the functional E. Then $(\mathcal{A}, E)$ is almost a non-commutative probability space. Unfortunately $\mathcal{A}$ is not norm closed, so this algebra is not a C^*-algebra and does not qualify under our definition as a non-commutative probability space.

Example 4

Now let us consider the algebra of all bounded linear operators acting on H. Fix a trace-class operator Γ and define the expectation as $E(X) =: \mathrm{tr}(\Gamma X)$. Then this algebra is a non-commutative probability space. However, the expectation is not tracial.

Example 5

Let $\mathcal{A}$ be a von Neumann algebra of type II_1 and E be the trace of this algebra. Then $(\mathcal{A}, E)$ is a non-commutative probability space. In this example, the expectation is tracial, and the algebra is closed not only in norm but also in the weak topology.

One useful technique to obtain a new non-commutative probability space is by building a matrix from the elements of another non-commutative probability space. So if $\mathcal{A}$ is a non-commutative probability space with the expectation E then we can construct $M_n(\mathcal{A})$ as $\mathcal{A} \otimes M_n(C)$ and take $E \otimes \mathrm{tr}_n$ as the expectation in this new non-commutative probability space. If E is tracial then $E \otimes \mathrm{tr}_n$ is also tracial.

Variables, Moments, and Measures

We will call elements from an algebra $\mathcal{A}$ (non-commutative) *random variables*. In the usual probability theory, a random variable can be characterized by its distribution function, moments, or characteristic function. In non-commutative probability theory, the easiest and the most general way to characterize random variables is through their moments. The *joint moment map* of random variables $A_1, ..., A_n$ is the linear map from non-commutative polynomials with complex coefficients to complex numbers, induced by taking the expectation:

$$m_{A_1,..A_n}(P) = E(P(A_1, ..., A_n)).$$

The *joint moments of degree* k are expectations of the monomials that have degree k.

Two collections of variables, say $(X_1, ..., X_n)$ and $(Y_1, ..., Y_n)$, are called *equivalent* if they have the same joint moment maps. In this case we will write $(X_1, ..., X_n) \sim (Y_1, ..., Y_n)$. These variables are *star-equivalent* if $(X_1, X_1^*, ..., X_n, X_n^*)$ and $(Y_1, Y_1^* ..., Y_n, Y_n^*)$ are equivalent. Then we will write $(X_1, ..., X_n) \approx (Y_1, ..., Y_n)$. For self-adjoint random variables X_i and Y_i these concepts coincide.

Note that equivalent random variables can come from essentially different operator algebras. If random variables are from the same algebra we can calculate their sum and product. The equivalence relation is invariant relative to these operations.

Proposition 6 *If $X_1, Y_1 \in \mathcal{A}$ and $X_2, Y_2 \in \mathcal{B}$ and $(X_1, Y_1) \sim (X_2, Y_2)$, then i) $X_1 + Y_1 \sim X_2 + Y_2$, and ii) $X_1 Y_1 \sim X_2 Y_2$.*

Proof: Both claims follow from expansion of the expressions for the moments. For example,

$$\begin{aligned} E\left((X_1 + Y_1)^k\right) &= \sum E\left(X_1^{i_1} Y_1^{j_1} ... X_1^{i_k} Y_1^{j_k}\right) \\ &= \sum E\left(X_2^{i_1} Y_2^{j_1} ... X_2^{i_k} Y_2^{j_k}\right) \\ &= E\left((X_2 + Y_2)^k\right). \end{aligned}$$

QED.

Clearly, this result can be generalized to a larger number of variables.

We also want to define convergence in distribution of vectors of random variables. Suppose we have a sequence of vectors of random variables: $X^{(i)} = \left(X_1^{(i)}, ..., X_n^{(i)}\right)$, where $X_k^{(i)} \in (\mathcal{A}_i, E_i)$. Suppose also that we have a vector of variables $x = (x_1, ..., x_n)$, where $x_k \in (\mathcal{A}, E)$. Then we will say that $X^{(i)}$ *converges in distribution* to x if for every $k > 0$ the joint moments of $X^{(i)}$ of degree less than k converge to the corresponding joint moments of x of degree less than k.

For some applications it is important to generalize the concept of freeness to operator-valued random variables. We give here a sketch of the generalization. A *non-commutative $\mathcal{B}$-valued probability space* $(\mathcal{A}, \mathcal{B}, E)$ is a C^*-algebra $\mathcal{A}$, its C^*-subalgebra $\mathcal{B}$, and a linear functional E, which is defined on the algebra $\mathcal{A}$ and takes values in the sub-algebra $\mathcal{B}$. It is assumed that E is a *conditional expectation*. That is, E maps positive definite operators to positive definite operators, and if $B \in \mathcal{B}$ then $E(B) = B$.

Here is a typical example. Suppose that $(\mathcal{A}, E)$ is a usual non-commutative probability space. Then we can define a non-commutative probability space with values in the algebra of n-by-n matrices M_n We will denote this algebra as $(M_n \otimes \mathcal{A}, E_n)$. These are matrices with elements of the algebra $\mathcal{A}$ as matrix entries. We define the M_n-valued expectation E_n component-wize. That is, the ij element of $E_n(A)$ is $E(A_{ij})$.

Now let us consider the question whether we can define a distribution function for a non-commutative random variable. Let X be a self-adjoint random variable (i.e., a self-adjoint operator from an algebra A). We can write X as an integral over a resolution of identity:

$$X = \int_{-\infty}^{\infty} \lambda dP_X(\lambda),$$

where $P_X(\lambda)$ is an increasing family of commuting projections. Then we can define the *spectral probability measure* of an interval $(a, b]$ as follows:

$$\mu_X\{(a, b]\} = E[P_X(b) - P_X(a)].$$

Then we can extend this measure to all measurable subsets. We will call μ_X the *spectral probability measure* of operator X, or simply its *spectral measure*.

Alternatively we can define $f_X(t)$ as $E(\exp(itX))$ and then prove by the Bochner theorem that f_X is a characteristic function of a probability measure.

For Example 1, the spectral measure of a real-valued random variable f coincides with the usual distribution measure of this random variable:

$$\mu_f\{(a, b]\} = \mu\{\omega : f(\omega) \in (a, b]\}.$$

For Example 2, the spectral measure of an n-by-n Hermitian matrix X is supported on the set of its eigenvalues. Each eigenvalue λ_i has the mass m/n, where m is the multiplicity of this eigenvalue.

This concept can be generalized to the case of unitary operators, in which case we will have measures defined on the unit circle. We can write a spectral representation for every unitary operator:

$$X = \int_{-\pi}^{\pi} e^{i\theta} dP_X(\theta).$$

Then if $(a, b] \subset (-\pi, \pi]$, we define the measure μ_X by the same formula as before:

$$\mu_X\{(a, b]\} = E[P_X(b) - P_X(a)].$$

In this case it is natural to interpret this measure as a measure on the unit circle instead of a measure on the interval $(-\pi, \pi]$.

3 Free Independence

3.1 Definition and properties

A natural requirement for independence of random variables A and B is that

$$E(P(A)Q(B)) = E(P(A))E(Q(B))$$

for arbitrary polynomials P and Q. But what about $E(ABAB)$, for example? Should it be $E(A^2)E(B^2)$ as in the case when A and B are commutative? Or perhaps it should be $E(A)^2 E(B)^2$, as if the first and the second occurence of variables A and B were completely independent each from the other?

Inspired by examples that arise in the theory of free group algebras, Dan Voiculescu suggested a particular concept of independence, which proved to be especially fruitful. He called this concept *freeness*. We define it for subalgebras of a given algebra $\mathcal{A}$.

Let $\mathcal{A}_1, ..., \mathcal{A}_n$ be sub-algebras of algebra $\mathcal{A}$, and let $\overline{A}_i$ denote an arbitrary element of algebra $\mathcal{A}_i$.

Definition 7 *Sub-algebras $\mathcal{A}_1, ..., \mathcal{A}_n$ (and their elements) are said to be* free*, if the following condition holds:*
For every sequence $\overline{A}_{i_1}...\overline{A}_{i_m}$, if $E\left(\overline{A}_{i_s}\right) = 0$ and $i_{s+1} \neq i_s$ for every s, then

$$E\left(\overline{A}_{i_1}...\overline{A}_{i_m}\right) = 0.$$

Variables X and Y are called free if the sub-algebras generated by $\{I, X, X^*\}$ and $\{I, Y, Y^*\}$ are free.

Remark: The definition of freeness can be generalized to the case of $\mathcal{B}$-valued probability spaces in a straightforward way, if by E we understand the conditional expectation $E : \mathcal{A} \to \mathcal{B}$. In this case we typically assume that $\mathcal{B} \subset \mathcal{A}_i \subset \mathcal{A}$, and in this case the definition is literally the same.

An important property of the concept of freeness is that it allows to compute all the joint moments of a set of free random variables in terms of the moments of individual variables.

Theorem 8 *Let $\mathcal{A}_1, ..., \mathcal{A}_m$ be free sub-algebras of $\mathcal{A}$, and let $A_1, ..., A_n$ be a sequence of random variables, $A_k \in \mathcal{A}_{i(k)}$, such that $i(k) \neq i(k+1)$. Then*

$$E(A_1...A_n) = \sum_{r=1}^{n} \sum_{1 \leq k_1 < ... < k_r \leq n} (-1)^{r-1} E(A_{k_1}) ... E(A_{k_r}) E\left(A_1 ... \widehat{A}_{k_1} ... \widehat{A}_{k_r} ... A_n\right), \tag{4}$$

where ^ denotes terms that are omitted.
Conversely, if this equality holds for every sequence of elements $A_1...A_n$ from the sub-algebras $\mathcal{A}_1, ..., \mathcal{A}_m$, then these sub-algebras are free.

Remark: Note that on the right-hand side the expectations are taken from products that have no more than $n-1$ terms. So a recursive application of this formula reduces computation of a joint moment to computation of a polynomial in the moments of the individual variables.

Proof: This formula is simply an expansion of the following relation:

$$E\left[(A_1 - E(A_1) I) ... (A_n - E(A_n) I)\right] = 0, \tag{5}$$

which holds by the definition of the free relation. Conversely, if formula (5) holds for any $A_1, ..., A_n$ then the algebras $\mathcal{A}_1, ..., \mathcal{A}_m$ are by definition free. QED.

So, for the example that started this section, it is easy to calculate:

$$E(ABAB) = E\left(A^2\right) E\left(B^2\right) - \left(E\left(B^2\right) - E(B)^2\right)\left(E\left(A^2\right) - E(A)^2\right).$$

So if we use σ_X^2 to denote the variance, i.e., the centered second moment, $\sigma_X^2 =: E(X^2) - E(X)^2$, then we can write:

$$\frac{E(ABAB)}{E(A^2B^2)} = 1 - \frac{\sigma_A^2 \sigma_B^2}{E(A^2) E(B^2)}.$$

Since this ratio is a measure of how non-commutativity affects calculation of moments, we can see that the effect of non-commutativity is larger if both variables have large relative variance, that is, if $\sigma_A^2 / E(A^2)$ and $\sigma_B^2 / E(B^2)$ are both close to 1.

For $\mathcal{B}$-valued expectations the formula in the previous theorem does not hold because the scalars from $\mathcal{B}$ do not commute with operators from $\mathcal{A}$. However what is true is that we still can compute the joint moments from individual moments. To convince the reader, we show this calculation for the expectation $E(A_1A_2A_1A_2)$, where $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$, and it is assumed that $\mathcal{A}_1$ and $\mathcal{A}_2$ are free. Let $E(A_1) = B_1$ and $E(A_2) = B_2$. Then

$$E\left((A_1 - B_1)(A_2 - B_2)(A_1 - B_1)(A_2 - B_2)\right) = 0$$

by definition of freeness. On the other hand, we can write the expression on the left-hand side as

$$\begin{aligned} & E\left(A_1 A_2 A_1 A_2\right) - E\left(B_1 A_2 A_1 A_2\right) - E\left(A_1 B_2 A_1 A_2\right) - ... \\ = \; & E\left(A_1 A_2 A_1 A_2\right) - E\left(A_2' A_1 A_2\right) - E\left(A_1'' A_2\right) - ..., \end{aligned}$$

where $A_2' = B_1 A_2 \in \mathcal{A}_2$ and $A_1'' = A_1 B_2 A_1 \in \mathcal{A}_1$. From this expression it is clear that $E\left(A_1 A_2 A_1 A_2\right)$ can be expressed in terms of a sum of expectations of monomials that have a length of no more than three. Then it is clear that we can apply induction.

We collect here some basic facts about freeness.

Proposition 9 *If $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_m$ are free, then $\mathcal{A}_2, ..., \mathcal{A}_m$ are free.*

Proof: Evident from definition. QED

Proposition 10 *If $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_m$ are free and $\mathcal{B} \subset \mathcal{A}_1$, then $\mathcal{B}, \mathcal{A}_2, ..., \mathcal{A}_m$ are free.*

Proof: Any variable from $\mathcal{B}$ is also a variable from $\mathcal{A}_1$ and consequently the relation in the definition of free independence holds. QED.

Proposition 11 *Let $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_m$ be free and $\mathcal{B}$ be an algebra generated by $\mathcal{A}_1, ..., \mathcal{A}_{k-1}$. Then $\mathcal{B}, \mathcal{A}_k, ..., \mathcal{A}_m$ are free.*

Proof: Any element from $\mathcal{B}$ is a polynomial of elements from $\mathcal{A}_1, ..., \mathcal{A}_{k-1}$. Without loss of generality we can choose these elements to have zero expectation. Write, for example,

$$B = \sum_{I=(i_1,...,i_n)} \alpha_I A_{i_1} A_{i_2} ... A_{i_n},$$

where A_{i_t} belongs to one of $\mathcal{A}_1, ..., \mathcal{A}_{k-1}$, $i_t \neq i_{t+1}$, and all of A_{i_t} have zero expectations. Since $\mathcal{A}_1, ..., \mathcal{A}_{k-1}$ are free, the fact that $E(B) = 0$ implies that the constant in this sum is zero. Therefore, each product like $B A_k A_{k+1} B ... A_s$, such that operators A_k, ... , A_s are taken from the algebras $\mathcal{A}_k, ..., \mathcal{A}_m$ and all of A_k, ... , A_s have zero expectation, can be expanded to a sum that has the following form:

$$\sum_{I=(i_1,...,i_n)} \alpha_I A_{i_1} A_{i_2} ... A_{i_n}$$

where consecutive A_{i_t} are from different algebras and where all of them have zero expectations. Since the algebras $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_m$ are assumed to be free, the expectation of this sum is zero and consequently we can conclude that $\mathcal{B}$ and $\mathcal{A}_k, ..., \mathcal{A}_m$ are free. QED.

We have the following analogue of Proposition 6 for free random variables:

Proposition 12 *Suppose that $A_1, A_2 \in \mathcal{A}$, and $B_1, B_2 \in \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are free sub-algebras. If $A_1 \sim A_2$ and $B_1 \sim B_2$, then i) $(A_1, B_1) \sim (A_2, B_2)$; ii) $A_1 + B_1 \sim A_2 + B_2$, and iii) $A_1 B_1 \sim A_2 B_2$.*

Proof:

i) By Theorem 8 each joint moment of A_i and B_i can be reduced to a polynomial in moments of A_i and moments of B_i. Let $[I, J]$ denote a sequence of indices $(i_1, j_1, ..., i_n, j_n)$ with elements which are non-negative integers, and let $m_{A_1,B_1}[I, J]$ denote the joint moment that corresponds to this sequence. That is, let

$$m_{A_1,B_1}[I,J] =: E\left(A_1^{i_1} B_1^{j_1} ... A_1^{i_n} B_1^{j_n}\right).$$

Let also $m_X(k) =: E\left(X^k\right)$. Then we can write

$$\begin{aligned} m_{A_1,B_1}[I,J] &= P_{[I,J]}(m_{A_1}(k_1), m_{B_1}(k_2), ..., m_{A_1}(k_{n-1}), m_{B_1}(k_n)) \\ &= P_{[I,J]}(m_{A_2}(k_1), m_{B_2}(k_2), ..., m_{A_2}(k_{n-1}), m_{B_2}(k_n)) \\ &= m_{A_2,B_2}[I,J], \end{aligned}$$

where $P_{[I,J]}$ denotes the polynomial that computes this joint moment of free A and B in terms of their individual moments and where the second line holds by the assumption that $A_1 \sim A_2$ and $B_1 \sim B_2$.

ii) and iii) follow from i) and Proposition 6. QED.

Let us now introduce a useful class of free variables. We will say that a unitary operator U is *Haar-distributed* if its spectral distribution is the uniform distribution on the unit circle.

Haar-distributed unitaries are very useful because we can use them to build collections of free self-adjoint random variables with prescribed spectral distributions. All we need is one self-adjoint variable, X, with a given spectral distribution and a sequence of Haar-distributed unitaries that are free from each other and from the variable X.

Proposition 13 *Suppose that i) the expectation is tracial, ii) $X, U_1, ..., U_n$ are free, iii) X is self-adjoint, and iv) U_1, ..., U_n are unitary and Haar-distributed. Suppose also that $h_1(x)$, ..., $h_n(x)$ are real-valued Borel-measurable functions of real argument. Then the variables $X_i = U_i^* h_i(X) U_i$ are free.*

Proof: Note that $f(X) = U_i^* f(h_i(X)) U_i$. So we need to prove that

$$E\left(U_{i(1)}^* f_1\left(h_{i(1)}(X)\right) U_{i(1)} ... U_{i(s)}^* f_s\left(h_{i(s)}(X)\right) U_{i(s)}\right) = 0 \tag{6}$$

if $E\left(U^*_{i(k)} f_1\left(h_{i(k)}\left(X_{i(k)}\right)\right) U_{i(k)}\right) = 0$ and $i(k+1) \neq i(k)$ for each $k.$

Note that $E(U_i) = E(U_i^*) = 0$ by the assumption that the U_i are Haar-distributed. Also, since expectation is assumed tracial, we have

$$\begin{aligned} 0 &= E\left(U^*_{i(k)} f_1\left(h_{i(k)}(X)\right) U_{i(k)}\right) = E\left(f_1\left(h_{i(k)}(X)\right) U_{i(k)} U^*_{i(k)}\right) \\ &= E\left(f_1\left(h_{i(k)}(X)\right)\right). \end{aligned}$$

Therefore,

$$E\left(f_1\left(h_{i(k)}(X)\right)\right) = 0,$$

and this implies (6) because X and all U_i are assumed free. QED.

How can we construct free sub-algebras? Examples show that certain natural constructions do not work. For example, let $\mathcal{B}_1$ and $\mathcal{B}_2$ be free sub-algebras of $\mathcal{A}$. Then we can build matrix algebras $M_n(\mathcal{A})$ and $M_n(\mathcal{B}_i)$ with natural inclusions $M_n(\mathcal{B}_i) \subset M_n(\mathcal{A})$. However, $M_n(\mathcal{B}_1)$ and $M_n(\mathcal{B}_2)$ are not necessarily free sub-algebras of $M_n(\mathcal{A})$. For example,

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} = \begin{pmatrix} A_{11} & -A_{12} \\ A_{21} & -A_{22} \end{pmatrix}.$$

Suppose that the expectation of the operator A on the left-hand side is 0, i.e., $E(A_{11}) + E(A_{22}) = 0$. This does not imply that the trace of the operator on the right-hand side is zero, i.e., in general $E(A_{11}) - E(A_{22}) \neq 0$.

Subalgebras $M_n(\mathcal{B}_1)$ and $M_n(\mathcal{B}_2)$ are $M_n(\mathbb{C})$-free with respect to the conditional expectation $E \otimes I_n$ but sometimes we want more. The next section gives a method to construct free subalgebras from two algebras $\mathcal{A}_1$ and $\mathcal{A}_2$.

3.2 Free products of probability spaces

If we have two non-commutative probability spaces, $(\mathcal{A}_1, E_1)$ and $(\mathcal{A}_2, E_2)$, then we can define their *free product* $(\mathcal{A}_1 * \mathcal{A}_2, E_1 * E_2)$. The algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ can be identified with two free subalgebras of $\mathcal{A}_1 * \mathcal{A}_2$ and both E_1 and E_2 are then restrictions of $E_1 * E_2$ to the corresponding subalgebra. This free product of algebras was first defined in Avitzour (1982).

Suppose A_1 and A_2 are two unital $*$-algebras. The algebra $\mathcal{A}_1 * \mathcal{A}_2$ is constructed as follows. Let S be the set of all sequences $a_1 a_2 ... a_n$, where $a_k \in \mathcal{A}_{i(k)}$ and $i(k) \neq i(k+1)$. Let S also contain the identity element. Let $L(S)$ be the algebra of all finite linear combinations of elements of S. We can easily define $*$ operation and multiplication on basis elements and then extend these operations to $L(S)$ by linearity. In particular, $(a_1...a_n)^* = a_n^*...a_1^*$ and $(a_1...a_n)(b_1...b_n) = R^\infty(a_1....a_n b_1...b_n)$, where R^∞ is the reduction operator. It is clear how to define R^∞. First we define one-step reduction R which leaves a sequence unchanged if no two neighboring elements are from the same algebra, or replace it with a reduced sequence if there are two neighboring elements from the same algebra. The reduced sequence is obtained in two steps. First, we have all two neigboring elements from the same algebra replaced with their product, and second we remove all identity elements from the sequence unless the sequence consists only of identity elements in which case the sequence is replaces with identity. The repetition of the one-step reduction converges for any initial sequence and the result is called complete reduction and denoted $R^\infty(.)$.

Theorem 14 *With the two operations defined above $L(S)$ is an algebra, closed with respect to $*$-operation.*

We call this $*$-algebra $A_1 * A_2$ and define the expectation $E_1 * E_2$ on elements of $A_1 * A_2$ by the following construction. Let $a_1...a_n$ is a reduced representation of a monomial element from $A_1 * A_2$. Clearly, it is enough to define $E_1 * E_2$ on monomials and extend it then to linear combinations of monomials by linearity. Since there is no linear dependence relations among monomials this definition is goint to be consistent. On monomials we define $E_1 * E_2$ using formula from Theorem 8. Namely, define $E_1 * E_2(I) = 1$ and let

$$E_1 * E_2(a_1...a_n) = \sum_{r=1}^{n} \sum_{1\le k_1<...<k_r\le n} (-1)^{r-1} E(a_{k_1})...E(a_{k_r}) E(a_1...\widehat{a}_{k_1}...\widehat{a}_{k_r}...a_n). \tag{7}$$

At this stage the only reduction that we allow on the right is that any two neighboring elements from the same algebra are replaced with their product. This defines $E_1 * E_2$ recursively for all sequences $a_1...a_n$ where a_i are from alternating algebras ($A_{k(i)} \neq A_{k(i+1)}$) provided that we have not identified the units, that is, that the sequences like $a_1 I_{A_1} a_3$ with $a_1, a_3 \in A_2$ are considered different from $(a_1 a_3)$. To show that the definition makes sense for algebras with identified units, it is sufficient to prove that

$$E_1 * E_2(a_1...a_{k-1} I_{A_1} a_{k+1}...a_n) = E_1 * E_2(a_1...(a_{k-1}a_{k+1})...a_n). \tag{8}$$

This is subject of the following lemma.

Lemma 15 *Relation 8 is true.*

From this lemma it follows that $E_1 * E_2$ is a well defined linear functional on $A_1 * A_2$. By definition $E_1 * E_2 (I) = 1$. The question is whether $E_1 * E_2$ is positive and what other properties of E_1 and E_2 are preserved under taking the free products. In answering these questions it is useful to have the following decomposition of $A_1 * A_2$.

Proposition 16 *Let A_i^0 be the elements of A_i that has zero expectation. Then as a linear space $A_1 * A_2$ has the following representation:*

$$\begin{aligned} A_1 * A_2 &= C + A_1^0 + A_2^0 + A_1^0 \otimes A_2^0 + A_2^0 \otimes A_1^0 + ... \\ &= C + \bigoplus_{n=1}^{\infty} \bigoplus_{k_1,...,k_n} \bigotimes_{i=1}^{n} A_{k_i}. \end{aligned}$$

Theorem 17 *Linear functional $E_1 * E_2$ is positive on the algebra $A_1 * A_2$.*

This theorem implies that $E_1 * E_2$ is a state. Therefore, we can define a norm on $A_1 * A_2$. We just use the GNS construction to represent $A_1 * A_2$ and then take the usual operator norm as the definition of the norm in the algebra. The completion of $A_1 * A_2$ with respect to this norm is the C^*-algebra $\mathcal{A}_1 * \mathcal{A}_2$, and the expectation $E_1 * E_2$ can be extended to the whole of the algebra $\mathcal{A}_1 * \mathcal{A}_2$. Further, the completion of this algebra of linear operators with respect to weak topology gives the W^*-algebra $\mathcal{A}_1 * \mathcal{A}_2$.

Another interesting property of the free product is as follows:

Theorem 18 *If E_1 and E_2 are traces then $E_1 * E_2$ is a trace.*

3.3 Circular and semicircular systems

A special place in free probability theory belongs to so-called circular and semicircular systems. They have a role similar to the role of independent multivariate Gaussian variables in classical probability theory. A vector of operators $(X_1, ..., X_n)$ forms a *semicircular system* if the variables are free and self-adjoint, and if each of them has the *semicircle distribution*, i.e., if the distribution function for the probability spectral measure associated with X_i is given by the formula

$$\mathcal{F}_{X_i}(t) = \frac{1}{2\pi} \int_{-\infty}^{t} \sqrt{4 - \xi^2} \chi_{[-2,2]}(\xi) \, d\xi.$$

It is easy to compute the moments of this distribution. First, we can check that the total mass is 1 by using the substitution $\xi = \sin \varphi$. Then, to compute moments, we can use integration by parts. The odd moment are evidently zero by symmetry. For the even, note that

$$\begin{aligned} m_{2k} &=: \frac{1}{2\pi}\int_{-2}^{2} t^{2k}\sqrt{4-t^2}dt = \frac{1}{2\pi}\int_{-2}^{2}\sqrt{4-t^2}\frac{d\left(t^{2k+1}\right)}{2k+1} \\ &= \frac{1}{2\pi}\frac{1}{2k+1}\int_{-2}^{2}\frac{t^{2k+2}}{\sqrt{4-t^2}}dt. \end{aligned} \tag{9}$$

Next, note the following identity:

$$\frac{4x^{2k}-x^{2k+2}}{\sqrt{4-x^2}} = x^{2k}\sqrt{4-x^2}.$$

Let us integrate from -2 to 2 and apply (9). Then we get:

$$4\left(2k-1\right)m_{2k-2} - \left(2k+1\right)m_{2k} = m_{2k},$$

and, consequently, we have the following recursion

$$m_{2k} = \frac{2\left(2k-1\right)}{k+1}m_{2k-1}$$

with the initial condition $m_0 = 1$. It is easy to check that the solution is

$$m_{2k} = \frac{1}{k+1}\binom{2k}{k}.$$

Further, we can define the moment-generating funciton:

$$G_X\left(z\right) =: \frac{1}{z} + \sum_{n=1}^{\infty}\frac{E\left(X^n\right)}{z^{n+1}} = \frac{1}{z} + \sum_{k=1}^{\infty}\frac{1}{k+1}\binom{2k}{k}\frac{1}{z^{2k+1}}. \tag{10}$$

To give an analytic formula for this function we look at the function: $f\left(u\right) = \sqrt{1-4u^2}$. For small u, we can develop $f\left(u\right)$ in power series:

$$\begin{aligned} f\left(u\right) &= 1 - \frac{\frac{1}{2}}{1!}4u^2 + \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)}{2!}\left(4u^2\right)^2 - \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\left(\frac{1}{2}-2\right)}{3!}\left(4u^2\right)^3 + ... \\ &= 1 - 2u^2 - \sum_{k=2}^{\infty}\frac{2^k\left(2-1\right)\left(2\times 2-1\right)...\left(2\left(k-1\right)-1\right)}{k!}u^{2k} \\ &= 1 - 2u^2 - 2\sum_{k=2}^{\infty}\frac{1}{k}\frac{\left(2-1\right)2(2\times 2-1)\left(2\times 2\right)...\left(2\left(k-1\right)-1\right)\left(2\left(k-1\right)\right)}{\left(k-1\right)!\left(k-1\right)!}u^{2k} \\ &= 1 - 2u^2 - 2\sum_{k=1}^{\infty}\frac{1}{k+1}\binom{2k}{k}u^{2k+2}. \end{aligned}$$

Therefore,

$$\frac{1-f(u)}{2u}=u+\sum_{k=1}^{\infty}\frac{1}{k+1}\binom{2k}{k}u^{2k+1}.$$

Comparing this with (10), we conclude that

$$G_X(z)=\frac{1-\sqrt{1-4z^{-2}}}{2z^{-1}}=\frac{z-\sqrt{z^2-4}}{2}.$$

Another useful concept is that of a circular system. A vector $(X_1,...,X_n)$ forms a *circular system* if algebras generated by $\{X_i, X_i^*\}$ are free, and each of the vectors $(Y_1,...,Y_n)$ and $(Z_1,...,Z_n)$ forms a semicircular system, where $Y_k=\frac{1}{\sqrt{2}}(X_k+X_k^*)$ and $Z_k=\frac{1}{\sqrt{2}}(X_k-iX_k^*)$.

3.4 Asymptotic and approximate freeness

In finite-dimensional algebras the operators cannot be free, unless they are multiples of identity. The reason is that the concept of free independence imposes infinitely many conditions that cannot be satisfied by a finite number of entries of finite-dimensional matrices.

Consider two real symmetric matrices of order two. Suppose that they both have zero traces. We can choose the basis in such a way that one of them is diagonal. Then we can write these matrices as follows:

$$A=\begin{pmatrix}\lambda & 0\\ 0 & -\lambda\end{pmatrix}\text{ and }B=\begin{pmatrix}x & y\\ y & -x\end{pmatrix}.$$

Suppose also that $\lambda\neq 0$. Then if we impose the condition that $\operatorname{tr}AB=0$, then it must be true that $x=0$. If we further require that $\operatorname{tr}(ABAB)=0$, then we can infer that $y=0$ and therefore $B=0$.

However, as a substitute of true freeness we can define concepts of asymptotic and approximate freeness for finite-dimensional matrices.

Suppose that $(\mathcal{A}_i, E_i)$ is a sequence of non-commutative probability spaces and X_i and Y_i are random variables from $\mathcal{A}_i$. Suppose that X_i and Y_i converge in distribution to operators x and y, respectively, which belong to probability spaces $(\mathcal{A}_x, E_x)$ and $(\mathcal{A}_y, E_y)$, respectively. Consider the free product $(A_x * A_y, E_x * E_y)$, and let x and y be free operators in this product.

Definition 19 *The sequences X_i and Y_i are called* asymptotically free *if (X_i, Y_i) converge in distribution to (x,y). In more detail, we require that for any ε and any* *sequence* $(n_1, ..., n_k)$*, there exists such* i_0 *that for* $i \geq i_0$*, the following inequality holds:*

$$\left|E_i\left(X_i^{n_1} Y_i^{n_2} ... X_i^{n_{k-1}} Y_i^{n_k}\right) - E\left(x^{n_1} y^{n_2} ... x^{n_{k-1}} y^{n_k}\right)\right| \leq \varepsilon.$$

Two sequences of subalgebras, $\mathcal{A}_i^{(1)}$ *and* $\mathcal{A}_i^{(2)}$*, are called* asymptotically free *if they are generated by asymptotically free operators* X_i *and* Y_i*.*

At the cost of more complicated notation, this definiton can be generalized to the case of more than two variables and to the case of subsets of variables $\Omega_r^{(i)}$ where each $\Omega_r^{(i)} \subset \mathcal{A}_i$.

Intuitively, what this definition aims to capture is the notion that as the index of $\mathcal{A}_i$ grows, the joint moments of X_i and Y_i converge to the joint moments that these variables would have if they were free. Typically, we will apply this definition in cases when $\mathcal{A}_i$ are algebras of random matrices of increasing dimension.

Now let us describe the concept of approximate freeness. Let X_i denote finite-dimensional operators, or in other words, k-by-k matrices. More generally, let Ω_i denote sets of such finite dimensional operators and let $X = (X_1, ..., X_s)$ denote an s-component vector of operators from Ω_i, that is, $X_i \in \Omega_i$ for each $i = 1, ..., s$.

Recall that for every non-commutative polynomial P in s variables and any set of s free random variables $Y_1, ..., Y_n$, we can calculate $E\left[P\left(Y\right)\right]$ as a polynomial of individual moments of Y_i. We can formally write this as

$$E\left[P\left(Y\right)\right] = f_P\left(m(Y_1), ..., m\left(Y_s\right)\right),$$

where $m\left(Y\right)$ denotes the moment sequence of variable Y.

Definition 20 *Sets of operators* Ω_i *with* $i = 1, ..., s$ *are called* (N, ε)-approximately free*, if for any non-commutative polynomial* P *in* s *variables and of a degree that does not exceed* N*, it is true that*

$$\left|E\left(P\left(X\right)\right) - f_P\left(m\left(X_1\right), ..., m\left(X_s\right)\right)\right| \leq \varepsilon$$

for every $X = (X_1, ..., X_s)$ *such that* $X_i \in \Omega_i$*.*

Approximate freeness is a tool to establish asymptotic freeness. Suppose we can find a sequence of matrices (X_i, Y_i) which are (N_i, ε_i)-approximately free. If $N_i \to \infty$ and $\varepsilon_i \to 0$ as $i \to \infty$, then we conclude that sequences of X_i and Y_i are asymptotically free.

Suppose that $\mathcal{A}$ is the algebra of unitary n-by-n matrices. It is intuitively clear that if N and ε are fixed, then the set of matrix pairs (X, Y) which are (N, ε)-approximately free becomes in some sense larger as n grows. Indeed, it becomes easier to satisfy the fixed number of conditions in Definition 20. For example, if μ is the Haar measure on $\mathcal{A} \times \mathcal{A}$, normalized to have the unit total mass, then we can expect that the mass of the set of (X, Y), such that X and Y are (N, ε)-approximately free, approaches 1 as n grows to infinity.

3.5 Various constructions of freely independent variables

In this section we study how to construct new free random variables from already existing ones. These results are from Nica and Speicher (1996) and Voiculescu (1996b), and we formulate them without proof.

Theorem 21 *Let $P_1, \ldots, P_n$ be a family of projections, which are orthogonal to each other (i.e., $P_i^* = P_i$, $P_i P_j = \delta_{ij} P_i$). Let c be the standard circular variable and suppose that $\{P_1, \ldots, P_n\}$ and $\{c, c^*\}$ are free. Then (1) variables $c^* P_1 c, \ldots, c^* P_n c$ are free, and (2) each of $c^* P_i c$ is a free Poisson variable with the parameter $\lambda_i = E(P_i)$.*

The statements of this theorem are valid if instead of the circular c we use the standard semicircular variable s, or the standard quartercircular variable b (i.e. $b = \sqrt{s^* s}$).

Can we use other variables instead of c, b, s, for example, the standard free Poisson $m = s^* s$? Or, the free Poisson with parameter $\lambda > 1$? Or, more generally, any random variable x, which does not have an atom at zero? This is not clear at this moment.

Here is another useful result

Theorem 22 *Let $X_1, ..., X_n, P$ be free self-adjoint random variables in non-commutative probability space $(\mathcal{A}, E)$, and suppose that P is a projection. Then $PX_1P, \ldots, PX_nP$ are free in $\left(P\mathcal{A}P, E(P)^{-1} E(\cdot)\right)$.*

Remark: this theorem generalizes a similar results for the case when we use a free unitary U instead of the projection P. Again, one immediate question is whether this result holds for other classes of operators beside unitaries and projections.

4 Example I: Group Algebras of Free Groups

Consider a countable group G with the counting measure μ. Let $H = l^2(G, \mu)$ be the Hilbert space of square-summable functions on G. Define the left action of G on H by $(L_g f)(h) = f(gh)$ and let $\mathcal{A}$ be the algebra of all finite linear combinations of L_g. We can close this algebra with respect to the operator norm topology and then we get a C^*-algebra, $C^*(G)$, or we can close $\mathcal{A}$ with respect to weak topology and then we get a von Neumann algebra, $\mathcal{N}(G)$. Define the expectation functional on algebra $\mathcal{A}$ as $E(L) = \langle \delta_e, L\delta_e \rangle$, where δ_e is the characteristic function of the set that consists of the unit of G. In other words, if $L = \sum_{g \in G} a_g L_g$, then $E(L) = a_e$.

This expectation is faithful, continuous in both strong and weak topology, and tracial. It can be used to define the scalar product on algebra $\mathcal{A}$ by the formula $(L_1, L_2) =: E(L_1^* L_2) = \langle L_1 \delta_e, L_2 \delta_e \rangle$. If we complete $\mathcal{A}$ with respect to this scalar product, then we get a Hilbert space (or "Hilbert algebra"), which we denote as $H^*(G)$. If C^*-algebras generalize algebras of continuous functions on a compact topological space and W^*-algebras generalize bounded measurable functions, then Hilbert algebras generalize L^2-summable functions.

The operators L_g form a complete orthonormal basis in the Hilbert algebra $H^*(G)$. In particular, we can represent each element from this Hilbert algebra as an infinite series $\sum_g x_g L_g$ with square-summable coefficients x_g.

Example 23 *Von Neumann algebra of a free group*

Suppose F_k is a free group with k generators. Then $\mathcal{N}(F_k)$ is a *von Neumann algebra of the free group*. It is still an open question whether these algebras are isomorphic for different k.

Suppose G is a free product of groups G_1 and G_2, and $\mathcal{A}$ is an algebra generated by G. Consider subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$ generated by elements L_g, with g from correspondingly G_1 and G_2.

Theorem 24 *Subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are free.*

Proof: Indeed, consider a product $Z = X_1 Y_1 ... X_n Y_n$ where $X_i \in \mathcal{A}_1$ and $Y_i \in \mathcal{A}_2$. Suppose that $E(X_k) = 0$ and $E(Y_k) = 0$ for every $k = 1, ..., n$. The condition $E(X_k) = 0$ means that $X_k = \sum x_g^{(k)} L_g$ where $g \in G_1$ and $g \neq e$. Similarly, $E(Y_k) = 0$ means that $Y_k = \sum y_g^{(k)} L_g$ where $g \in G_2$ and $g \neq e$. This implies that $Z = \sum z_g^{(k)} L_g$ where $g \in G$ and $g \neq e$. This claim holds because the subgroups

G_1 and G_2 are free. Therefore, $E(Z) = 0$. Since this holds for any product $Z = X_1Y_1...X_nY_n$, the freeness condition is verified. QED.

In particular, let G be a free group generated by elements $g_1, ..., g_n$, and G_i be subgroups generated by elements g_i, respectively. Let $\mathcal{A}$ and $\mathcal{A}_i$ be group algebras of G and G_i, respectively. Then subalgebras $\mathcal{A}_i$ are free.

Example 25

Consider the random variable $X = L_{g^{-1}} + L_g$. This variable is self-adjoint and we can calculate $E\left(X^k\right)$ as 0 if k is odd and as $\binom{k}{k/2}$ if k is even. Indeed consider a random walk on integers that starts at time $t = 0$ from $x = 0$ and at each time step can either go up by 1 or down by -1. Then $E\left(X^k\right)$ is equal to the number of paths that at time $t = k$ end at $x = 0$. Clearly, the number of such paths is zero if k is odd and it is equal to the number of ways we can choose the k time steps, at which the random walk goes up, i.e., $\binom{k}{k/2}$.

If we define the moment-generating function of X is as follows:

$$G_X(z) =: \frac{1}{z} + \sum_{k=1}^{\infty} \frac{E\left(X^k\right)}{z^{k+1}},$$

then it is clear that for $X = L_{g^{-1}} + L_g$, we have

$$G_X(z) = \frac{1}{z} + \sum_{k=1}^{\infty} \binom{2k}{k} \frac{1}{z^{2k+1}}.$$

Can we find a probability distribution that corresponds to these moments? Consider a probability measure with the following distribution function:

$$F(x) = \frac{1}{\pi} \int_{-\infty}^{x} \frac{\chi_{[-2,2]}(t)}{\sqrt{4-t^2}} dt.$$

Using substitution $t = 2\sin\varphi$, it is easy to check that

$$\frac{1}{\pi} \int_{-2}^{2} \frac{dt}{\sqrt{4-t^2}} = 1,$$

so $F(x)$ is a valid probability distribution function.

Next, note that

$$\begin{aligned} \frac{1}{\pi} \int_{-2}^{2} \frac{\left(4t^{2k-2} - t^{2k}\right)}{\sqrt{4-t^2}} dt &= \frac{1}{\pi} \int_{-2}^{2} t^{2k-2} \sqrt{4-t^2} dt \\ &= \frac{1}{\pi} \frac{1}{2k-1} \int_{-2}^{2} \frac{t^{2k}}{\sqrt{4-t^2}} dt, \end{aligned}$$

where, in order to get the second equality, we integrated by parts. Consequently, if

$$m_k =: \frac{1}{\pi} \int_{-2}^{2} \frac{t^k}{\sqrt{4-t^2}} dt,$$

then $m_k = 0$ for odd k and for even k we get the following recursion:

$$\left(1 + \frac{1}{2k-1}\right) m_{2k} = 4 m_{2k-2},$$

or

$$m_{2k} = \frac{2\,(2k-1)}{k} m_{2k-1}.$$

with the initial condition $m_0 = 1$.

It is easy to check that the recursion is satisfied by $m_{2k} = \binom{2k}{k}$. Therefore, this probability distribution has the desired moments.

The distribution function can be computed explicitly as

$$F\,(x) = \frac{1}{\pi} \arcsin\left(\frac{x}{2}\right) + \frac{1}{2}.$$

For this reason, this distribution is called the *arcsine law*.

Example 26

Consider the random variable $Y = L_g$. This variable is unitary and $E\left(Y^k\right) = 0$ for every integer $k > 0$. Since Y is unitary, we can conclude that is ia a Haar-distributed unitary. Its moment-generating function is simply $G_Y\,(z) = z^{-1}$.

One difficulty with free group algebras is that it is difficult to construct a variable with a given sequence of moments or with a given spectral distribution. This difficulty is partially resolved in the example that we discuss in the next section.

5 Example II: Algebras of Creation and Annihilation Operators in Fock Space

5.1 Operators in Fock space as non-commutative random variables

Definition 27 *Let H be a separable complex Hilbert space and fix a vector $\xi \in H$. The Fock space $T(H)$ is the following Hilbert space:*

$$T(H) = \mathbb{C}\xi + H + H \otimes H + H \otimes H \otimes H + \ldots$$

The fixed vector ξ is often called the *vacuum vector*. If e_i is a basis of H then elements ξ and $e_{i_1} \otimes ... \otimes e_{i_n}$ form a basis of $T(H)$. We will write the basis elements as $e_{i_1} e_{i_2} ... e_{i_n}$ to lighten notation. In this basis the scalar product in $T(H)$ is determined by linearity and the following rules: $\langle \xi, \xi \rangle = 1$, $\langle \xi, e_{i_1} ... e_{i_n} \rangle = 0$, $\langle e_{i_1} ... e_{i_n}, e_{j_1} ... e_{j_m} \rangle = 0$ if $n \neq m$, and $\langle e_{i_1} ... e_{i_n}, e_{j_1} ... e_{j_n} \rangle = \delta_{i_1 j_1} ... \delta_{i_n j_n}$. In other words, ξ and $e_{i_1} ... e_{i_n}$ form an orthonormal basis of $T(H)$.

Let us fix a basis of H. For each vector e_k in this basis we define an operator a_k acting on $T(H)$, namely, $a_k(\xi) = e_k$, $a_k(e_{i_1} e_{i_2} ... e_{i_n}) = e_k e_{i_1} e_{i_2} ... e_{i_n}$. This operator is called a *(left) creation operator*. Its adjoint is called a *(left) annihilation operator*: $a_k^*(\xi) = 0$, $a_k(e_i) = \delta_{ki} \xi$, and $a_k^*(e_{i_1} e_{i_2} ... e_{i_n}) = \delta_{k i_1} e_{i_2} e_{i_2} ... e_{i_n}$ for $n \geq 2$.

The terminology came from physics where the Fock space is used in quantum models of light propagation (see, e.g. Klauder- .... (...)).

Let $\mathcal{A}$ be an algebra of all polynomials of the operators a_k and a_k^*, and $\widetilde{\mathcal{A}}$ is its closure in the weak topology. Then $\widetilde{\mathcal{A}}$ is a W^* non-commutative probability space with the expectation given by $E(X) = \langle \xi, X\xi \rangle$.

This expectation is not tracial: $E(a_k^* a_k) = 1$ but $E(a_k a_k^*) = 0$.

We will consider random variables of the following form:

$$X = \sum_{i=1}^{n} x_{-i} (a_k^*)^i + \sum_{i=0}^{\infty} x_i (a_k)^i,$$

where x_i denotes a summable sequence $(x_0, x_1, ...)$. We will call them *Toeplitz random variables*. because they have some similarities to Toeplitz matrices.

5.2 Free independence of Toeplitz random variables

Theorem 28 *If $k \neq l$ then a_k and a_l are free.*

Proof: Without loss of generality, let $k = 1$ and $l = 2$. Consider polynomials of a_1 and its adjoint a_1^*, and of a_2 and its adjoint a_2^*. Expanded, they have the following form:

$$P_r = \sum_{k=1}^{n} x_{-k}^{(r)} (a_1^*)^k + x_0^{(r)} + \sum_{k=1}^{m} x_k^{(r)} (a_1)^k$$

and

$$Q_r = \sum_{k=1}^{n} y_{-k}^{(r)} (a_2^*)^k + y_0^{(r)} + \sum_{k=1}^{m} y_k^{(r)} (a_2)^k.$$

If $E(P_r) = 0$, then $x_0^{(r)} = 0$. Similarly, if $E(Q_r) = 0$, then $y_0^{(r)} = 0$. Assume that $x_0^{(r)} = y_0^{(r)} = 0$. Then it is easy to see by induction that $Q_1 P_1 Q_2 P_2 ... Q_n P_n \xi$ is the sum of terms that have the form $e_2^k f$, where $k \geq 1$, and $P_1 Q_2 P_2 ... Q_n P_n \xi$ is the sum of terms that have the form $e_1^k f$, where $k \geq 1$. Indeed, suppose we have already proved this for $P_1 Q_2 P_2 ... Q_n P_n \xi$ and know that all terms of this product start with $e_1^k f$, where $k \geq 1$. Then the terms of Q_1 that have the form $\sum_{k=1}^{n} y_{-k}^{(1)} (a_2^*)^k$ will produce zero when they multiply terms of the form $e_1^k f$, and the terms of Q_1 that have the form $\sum_{k=1}^{m} y_k^{(1)} (a_2)^k$ will produce the terms of the form $e_2^k f$ with $k \geq 1$.

Consequently, the product $Q_1 P_1 Q_2 P_2 ... Q_n P_n \xi$ has a constant term equal to zero and therefore:

$$\langle \xi, Q_1 P_1 Q_2 P_2 ... Q_n P_n \xi \rangle = 0.$$

By a similar argument, we can write

$$\langle \xi, P_1 Q_2 P_2 ... Q_n P_n \xi \rangle = 0$$

and similar identities for products that end in Q_n. This implies the free independence of X and Y. QED.

Let $\mathcal{A}_k$ denote a subalgebra of $\mathcal{A}$, generated by a_k and a_k^* only.

Corollary 29 *If $k \neq l$, then the subalgebras $\mathcal{A}_k$ and $\mathcal{A}_l$ are free.*

5.3 Representability by Toeplitz random variables

Toeplitz random variables are useful because it is relatively easy to construct a Toeplitz variable with a given moment sequence. We will say that operator X *represents* operator Y if $X \sim Y$, that is, if X and Y have the same moment sequence: $E(X^k) = E(Y^k)$ for all $k \geq 1$. We will also say that X *represents the moment sequence* $\{m_n\}$ if $E(X^n) = m_n$ for each n.

We know that if A and B are free, then moments of $A + B$ are determined by moments of A and B. In this case, if $\widetilde{A}$ and $\widetilde{B}$ represent A and B, respectively, and $\widetilde{A}$ and $\widetilde{B}$ are free, then $\widetilde{A} + \widetilde{B}$ represents $A + B$.

Lemma 30 *For any number sequence $m_1, ..., m_n, ...$, there is a unique number sequence, x_i, such that the operator $X =: a + \sum_{i=0}^{\infty} x_i (a^*)^i$ represents $\{m_n\}$, i.e., $E(X^n) = m_n$ for each n. In particular for any operator Y, there is a unique operator $X = a + \sum_{i=0}^{\infty} x_i (a^*)^i$ that represents Y.*

Proof: Let

$$X = a + \sum_{k=0}^{m} x_k \left(a^*\right)^k ,$$

and for consistency define $x_{-1} = 1$. Consider the expansion of $X^n \xi$. It consists of terms of the form

$$\varepsilon \left(i_1, .., i_n\right) x_{i_1} x_{i_2} ... x_{i_n} e^{-i_1 - ... - i_n},$$

where $\varepsilon \left(i_1, .., i_n\right)$ is either 0 or 1 depending on the sequence $i_1, ..., i_n$, and we use the notational conventions $e^0 =: \xi$ and $e^k =: 0$ if $k < 0$. Here x_{i_n} denotes the coefficient before $\left(a^*\right)^{i_n}$ (or a, if $i_n = -1$) in the first copy of X that operated on ξ; $x_{i_{n-1}}$ denotes the coefficient before $\left(a^*\right)^{i_{n-1}}$ in the copy of X that operated on $X\xi$ after that, and so on. The coefficient x_{i_1} comes from the copy of X that operated on $X^{n-1}\xi$.

If we look at the consequitive sums i_1, $i_1 + i_2$, ..., $i_1 + i_2 + ... + i_n$, then we can note that $\varepsilon \left(i_1, .., i_n\right) = 0$ whenever a sum from this sequence is positive. This positivity means that more annihilation than creation operators acted till this moment and the effect of this sequence of operators on the vacuum vector must be zero.

A particular example of this situation arises if a specific term in the expansion has x_{i_k}, with $i_k \geq n$ as one of its elements. Then $-(i_1 + ... + i_n) < 0$ because $i_s \geq -1$ for all other $s \neq k$. Therefore, $e^{-(i_1 + ... + i_n)} = 0$ and the scalar product of this term and ξ is zero.

Next consider the situation when a particular term has x_{i_k}, with $i_k = n - 1$. Then the only possibility that $i_1 + ... + i_n \leq 0$ is that $i_s = -1$ for all $s \neq k$. Moreover, in this case $\varepsilon \left(i_1, .., i_n\right)$ is not zero if and only if $k = 1$. In other words, if $i_k = n - 1$ for some i_k then k must equal 1 (that is, the coefficient x_{n-1} must come from the copy of X that operated last) and then we must have $i_2 = i_3 = ... = i_n = -1$. Only in this case we have $i_1 + ... + i_r \leq 0$ for all $r \leq n$ and $\varepsilon \left(i_1, .., i_n\right) = 1$. In this case, the term that we are considering must be $x_{n-1} x_{-1} ... x_{-1} x_{-1} \xi = x_{n-1}\xi$ and the scalar product of this term with ξ is x_{n-1}.

The remaining terms have all $i_k < n - 1$. Therefore, they will not have x_{n-1} as the element of the multiple $x_{i_1} x_{i_2} ... x_{i_n}$. Hence, we can conclude that

$$EX^n = \langle \xi, X^n \xi \rangle = x_{n-1} + P_n \left(x_0, ..., x_{n-2}\right),$$

where P_n is a polynomial. Therefore, we can proceed inductively. Coefficient x_0 is uniquely determined by the moment $m_1 = EX$. Coefficient x_1 is uniquely determined by the moment $m_2 = E\left(X^2\right)$ and by the coefficient x_0. Coefficient x_2 is uniquely determined by the moment $m_3 = E(X^3)$ and by a polynomial of coefficients x_0 and x_1, and so on. Coefficient x_{n-1} is uniquely determined by the moment $m_n = E(X^n)$ and a polynomial of coefficients x_0, x_1, ..., and x_{n-2}. QED.

Warning: Note that an X that represents Y is not necessarily strongly equivalent to Y. That is, in general, even if $E\left(X^k\right) = E\left(Y^k\right)$ for all k, this does not imply that $E\left((X^*)^k\right) = E\left((Y^*)^k\right)$.

Example 31

Consider $X = a + a^*$. This variable is self-adjoint. To compute $E\left(X^k\right)$, note that the constant in the expansion of $(a + a^*)^k$ can be expressed in terms of the number of paths in a certain random walk. Namely, consider the random walk of a particle on the lattice of integers. At time $t = 0$ the particle starts at $x = 0$, and at each later time step it can either go up by 1 (if the creation operator acts), or down by 1 (if the annihilation operator acts). We are interested in the number of such paths that end at time $t = k$ at $x = 0$. and that are always greater than or equal to zero. This is a classical combinatorial problem and the answer can be found in Feller (...). Clearly, if k is odd, then the number of paths is 0, and it turns out that if k is even, then it equals $(k/2+1)^{-1}\binom{k}{k/2}$.

For convenience of the reader we repeat here the argument. Let us consider an equivalent problem: We are looking for a number of paths such that $x(t=0) = 1$, $x(t=k) = 1$, and that $x(t) \geq 1$ for all $t : 0 \leq t \leq k$. Then we first note that the total number of paths from $(t = 0,\ x = 1)$ to $(t = k,\ x = 1)$ is 0 if k is odd and $\binom{k}{k/2}$ if k is even. On the other hand the number of paths that go from $(t = 0,\ x = 1)$ to $(t = k,\ x = 1)$ and touch or cross the line $x = 0$ equals the total number of paths that go from $(t = 0,\ x = -1)$ to $(t = k,\ x = 1)$. This is simply reflection principle. Therefore, we can compute this number as 0 for odd k and $\binom{k}{k/2-1}$ for even k. Indeed if q is the number of down movements and p is the number of up movements, then $p - q = 2$, and $p + q = k$. Therefore, $q = k/2 - 1$, and the number of ways to choose these down movements is $\binom{k}{k/2-1}$.

Hence if $k = 2n$ then the number of paths that go from$(t = 0,\ x = 1)$ to $(t = k,\ x = 1)$ and do not drop below $x = 1$ equals

$$\binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n}$$

Therefore the moment-generating function for this example is

$$G_X(z) = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{1}{n+1} \binom{2n}{n} \frac{1}{z^{2n+1}} \tag{11}$$

Note that this function is different from the moment-generating function of the random variable in Example 25. Therefore, the variables $a + a^*$ and $g + g^{-1}$ are not equivalent, or in other words, $a + a^*$ does not represent $g + g^{-1}$.

On the other hand, comparing (11) with formula (10) on page 23, we can conclude that $a + a^*$ has the semicircle distibution.

Example 32

Consider $Y = a$. Then Y is an isometry. It is not unitary because it is not invertible. Clearly $E\left(Y^k\right) = 0$ for every $k > 0$, and the moment-generating function of Y^k is $G_Y(z) = z^{-1}$. Comparing this with Example 26, we note that a represents the Haar-unitary random variable g.

5.4 Additivity of Toeplitz random variables

Let a Toeplitz variable, X, be associated with a vector $x = (x_1, ..., x_n, ..)$ and some creation operator a_2:

$$X = a_2 + \sum_{k=0}^{m} x_k \left(a_2^*\right)^k .$$

Similarly, let Y belong to the same Fock space and be associated with a vector y and a different creation operator a_3:

$$Y = a_3 + \sum_{k=0}^{m} y_k \left(a_3^*\right)^k$$

Finally, let Z be a Toeplitz variable (possibly from a different Fock space) associated with the vector $x + y$:

$$Z = a_1 + \sum_{k=0}^{m} (x_k + y_k) \left(a_1^*\right)^k$$

Theorem 33 *Z represents $X + Y$, i.e. $Z \sim X + Y$.*

Proof: Let us write:

$$\begin{aligned} X+Y &\equiv (a_2+a_3)+\sum_{i=0}^{m} x_i \left(a_2^*\right)^i+\sum_{i=0}^{m} y_i \left(a_3^*\right)^i , \\ Z &\equiv a_1+\sum_{i=0}^{m} x_i \left(a_1^*\right)^i+\sum_{i=0}^{m} y_i \left(a_1^*\right)^i . \end{aligned}$$

Let us consider (a_2+a_3) as a single symbol. Then there is an evident correspondence between elements in these sums. This correspondence is as follows

$$\begin{aligned} (a_2+a_3) &\rightarrow a_1, \\ x_i \left(a_2^*\right)^i &\rightarrow x_i \left(a_1^*\right)^i , \text{ and} \\ y_i \left(a_3^*\right)^i &\rightarrow y_i \left(a_1^*\right)^i . \end{aligned}$$

If we write an expansion of $(X+Y)^n$ in terms of (a_2+a_3) , a_2^*, and a_3^*, then we can use this correspondence to write an expansion of Z^n in terms of a_1 and a_1^*. For example, $(a_2+a_3) x_3 \left(a_2^*\right)^3 (a_2+a_3) y_7 \left(a_3^*\right)^7$ corresponds to $a_1 x_3 \left(a_1^*\right)^3 a_1 y_7 \left(a_1^*\right)^7$. Therefore we need only to prove the equality of the expectations of these products, e.g.,

$$\left\langle \xi, (a_2+a_3) x_3 \left(a_2^*\right)^3 (a_2+a_3) y_7 \left(a_3^*\right)^7 \xi \right\rangle = \left\langle \xi, a_1 x_3 \left(a_1^*\right)^3 a_1 y_7 \left(a_1^*\right)^7 \xi \right\rangle .$$

Note that the following identities hold:

$$\begin{aligned} a_2^* (a_2+a_3) &= a_1^* a_1 = I, \text{ and} \\ a_3^* (a_2+a_3) &= a_1^* a_1 = I, \end{aligned}$$

Therefore, whenever a_2+a_3 is on the right of either a_2^* or a_3^*, we can cancel it out as well as the corresponding pair of a_1 and a_1^*. From this it follows that we need only to prove the equality of the following expectations, where a_2+a_3 is on the left of all a_2^* and a_3^*, and the corresponding terms a_1 are on the right of all a_1^*:

$$\left\langle \xi, (a_2+a_3)^n \left(a_2^*\right)^{k_1} \left(a_3^*\right)^{k_2} ...\xi \right\rangle = \left\langle \xi, (a_1)^n \left(a_1^*\right)^{k_1} \left(a_1^*\right)^{k_2} ...\xi \right\rangle .$$

However, it is evident that both are 1 if and only if $k_1 = k_2 = ... = n = 0$, and 0 otherwise. QED.

6 Addition and Multiplication Theorems

6.1 Addition

6.1.1 Motivation

We know from Theorem 8 that we can calculate every joint moment of free random variables A and B from their individual moments. Consequently, we can express $E\left(A+B\right)^k$ as a polynomial of EA^i and EB^j for $i,j\leq k$. This method, however, is not efficient and does not provide much insight. It is natural to seek a more efficient algorithm for computation of $E\left(A+B\right)^k$.

We are interested in $E\left(A+B\right)^k$ for several reasons. First, let G be a countable group. Consider an operator

$$A=\sum_{g\in G}a_gL_g,$$

where L_g are right shift operators as in Section 4, and let us impose an additional restriction that a_g are non-negative and that $\sum a_g=1$. Then we can interpret a_g as probabilities and A as a random walk on the group G. Then $E\left(A^k\right)$ has a natural interpretation as a probability of the return to the identity element after k steps. Now, suppose that G is a free group with two generators, g and h, and that we have two probability distributions, μ and ν, which assign probabilities to powers of g and h, respectively. That is, $\mu\left(g^k\right)=a_k$ and $\nu\left(h^l\right)=b_l$, where k and l are arbitrary integers, and a_k and b_k are positive numbers such that $\sum a_k=1$ and $\sum b_l=1$.

We define a random walk on the free group G by the following process. At each moment of time we throw a die and decide whether we use a power of g or a power of h. If we decide to use a power of g, then we use g^k with probability a_k and if we decided to use a power of h then we use h^l with probability b_l. For this random walk, what is the probability of return to the unit element after k steps? The answer depends on our ability to calculate the following quantity:

$$E\left(\frac{A+B}{2}\right)^k,$$

where

$$A=\sum_{k\in Z}a_kL_{g^k},\text{ and }B=\sum_{l\in Z}b_lL_{h^l}.$$

In free probability, operators A and B are prototypical free random variables and we arrive at the calculation of the k-th moment of $(A+B)$.

Second, suppose that free operators A and B are self-adjoint and that μ_A are μ_B are their spectral distributions. Then $A+B$ is also self-adjoint and its distribution μ_{A+B} depends only on μ_A and μ_B. We will call this distribution the *additive free convolution* of μ_A and μ_B and denote it as $\mu_A \boxplus \mu_B$. It is natural to ask about properties of this new operation on probability measures. Note that the moments of $\mu_A \boxplus \mu_B$ are given by $E\,(A+B)^k$.

For comparison, consider the case when G is the *commutative* free group generated by g and h. Then $A = \sum_k a_k L_{g^k}$ is a sum of commuting unitary operators. Therefore, the spectral measure of operator $A = \sum_k a_k L_{g^k}$ is well defined and it is the image of the uniform measure on the unit circle under the map $e^{i\theta} \to f_A\left(e^{i\theta}\right)$, where $f_A(z)$ is the symbol of operator A :

$$f_A(z) = \sum_k a_k z^k.$$

Let μ_A be the resulting measure on the complex plane. Similarly, define μ_B for $B = \sum_k b_k L_{h^k}$. Then A and B have the same set of eigenspaces, and therefore the spectral distibution of the sum is well defined and is easy to compute. The result is simply the additive convolution of the measures μ_A and μ_B :

$$\mu_{A+B}(dw) = \int_{z\in\mathbb{C}} \mu_A(dw - z)\,\mu_B(dz).$$

In the non-commutative case we can define the free additive convolution.

Definition 34 *Let μ and ν be the spectral probability measures of free self-adjoint random variables A and B, respectively. Then the spectral probability measure of $A+B$ is called the* free additive convolution *of measures μ and ν, and denoted $\mu \boxplus \nu$.*

In the case of the non-commutative free group of two generators, even if the operators A and B are self-adjoint they do not have the same set of eigenspaces and it is difficult to compute the spectral distribution of the sum from the spectral distributions of the summands. It is amazing that there exists an analytical way to perform this computation. Let us describe this procedure.

6.1.2 Cauchy transform and K-function

Definition 35 *We call the* moment-generating function *of a random variable X, or the* Cauchy transform *of X, the expectation of the resolvent of X:*

$$G_X(z) = E\left[(z-X)^{-1}\right].$$

In the case that we consider most often, X is a bounded operator. Therefore, $G_X(z)$ is defined in the area $\{z : |z| > \|X\|\}$. The Cauchy transform is holomorphic in this area and maps ∞ to 0. Moreover, for all sufficiently large z, the Cauchy transform is univalent and we can define the functional inverse of $G_X(z)$.

Definition 36 *We call the* K-function *of* X *the functional inverse of the Cauchy transform of* X*:*

$$G_X(K_X(z)) = K_X(G_X(z)) = z.$$

For bounded random variables, the function $K_X(z)$ is well-defined in a neighborhood of 0 and has a simple pole at 0.

It is easy to compute the expansion of the Cauchy transform at $z = \infty$:

$$G_X(z) = \frac{1}{z} + \sum_{k=1}^{\infty} \frac{E\left(X^k\right)}{z^{k+1}}.$$

Here are the first terms of the Laurent series for the K-function of X:

$$K_X(z) = \frac{1}{z} + E(X) + \left[E\left(X^2\right) - E(X)^2\right] z + ...$$

If $K_X(z) = z^{-1} + \sum_{k=0}^{\infty} c_k z^k$, then it is easy to see that c_k can be expressed as polynomials of $E(X), ..., E\left(X^{k+1}\right)$. Later, we will give an analytic expression for these coefficients.

Example 37 *Zero and scalar operators*

If a random variable $X = 0$, then its K-function is simply z^{-1}. More generally, if $X = cI$, where I is the identity operator, then

$$G_{cI}(z) = \frac{1}{z - c},$$

and

$$K_{cI}(z) = \frac{1}{z} + c.$$

Example 38 *Semicircle distribution*

For a self-adjoint random variable X that has the semicircle distibution as its spectral probability distribution (see definition of the semicircle distribution at page 22) it is easy to compute its K-function:

$$K_{SC}(z) = z^{-1} + z.$$

Example 39 *Marchenko-Pastur distribution*

One other distribution plays important role in free probability theory. It has the following K-function:

$$K_{MP}(u) = \frac{1}{u} + \frac{\lambda}{1-u}.$$

This distibution is called the *Marchenko-Pastur distribution* because this distribution was discovered by Marcenko and Pastur (1967). They discovered this distribution as a limit eigenvalue distribution for a so-called Wishart matrices. Consider a rectangular n-by-m matrix X with independent Gaussian entries that have zero expectation and variance equal to $1/n$. Consider matrix $Y = X'X$. This matrix is called the Wishart matrix with parameters (n, m). Suppose that n and m grow to infinity but in such a way that n/m approaches a limit $\lambda > 0$. Marchenko and Pastur discovered that the distibutions of eigenvalues of this matrix converges to a distribution that depends on the parameter λ. This distribution is called the Marchenko-Pastur distribution.

Let us compute the shape of this distribution. Inverting the K-function we get:

$$G_{MP}(z) = \frac{1 - \lambda + z - \sqrt{(1 - \lambda + z)^2 - 4z}}{2z}.$$

It follows that the continous part of this distribution is concentrated on the interval

$$\left[\left(1 - \sqrt{\lambda}\right)^2, \left(1 + \sqrt{\lambda}\right)^2\right]$$

and the density is

$$f_{MP}(x) = \frac{\sqrt{4x - (1 - \lambda + x)^2}}{2\pi x}.$$

In addition, if $\lambda < 1$, then there is also an atom at zero with the probability weight $1 - \lambda$.

This distribution is also called the free *Poisson distribution* because it can be obtained as a limit of free additive convolutions of Bernoulli distributions, and in the classical case a similar sequence of convolutions would converge to the Poisson random variable.

There are two other functions directly related to the K-function, which appear often in the literature.

Definition 40 *The function $R_X(z) = K_X(z) - z^{-1}$ is called the R-*transform *of the random variable X.*

The funciton $R_X(z)$ is holomorphic around $z = 0$. Its usefulness stems from the fact that $R_{X+Y}(z) = R_X(z) + R_Y(z)$. We will prove this fact in the next section.

Definition 41 *Let $F_X(z) = 1/G_X(z)$ and $F_X(z)^{(-1)}$ is the functional inverse of $F_X(z)$ in an open set of the upper half-plane, which includes infinity. Then $\varphi_X(z) = F_X(z)^{(-1)} - z$ is called the* Voiculescu transform *of the random variable X.*

The definition of Voiculescu transform is especially useful when the random variable X is not bounded. In the case when X is bounded, $\varphi_X(z) = R_X(z^{-1}) = K_X(z^{-1}) - z$. Again the main property of this function is that $\varphi_{X+Y}(z) = \varphi_X(z) + \varphi_Y(z)$.

Note that if X is self-adjoint and μ is its spectral probability measure, then the K-function, R-transform, and Voiculescu transform depend only on μ. We will say that these functions are the K-function, R-transform, and Voiculescu transform of the measure μ.

6.1.3 Addition formula

Theorem 42 (Voiculescu's Addition Formula)
Let Y_1 and Y_2 be two free bounded non-commutative random variables with the Cauchy transforms $G_{Y_1}(z)$ and $G_{Y_2}(z)$, and let $K_{Y_1}(z)$ and $K_{Y_2}(z)$ be two corresponding K-functions. If $Y_3 = Y_1 + Y_2$ then the K-function of Y_3 can be computed as

$$K_{Y_3}(z) = K_{Y_1}(z) + K_{Y_2}(z) - \frac{1}{z}.$$

This theorem plays a central role in the theory of additive free convolutions. Initially, the theorem was proved by Voiculescu using the Helton-Howe formula for traces of commutators of Toeplitz operators. Then it was simplified by Haagerup, who avoided using the Helton-Howe formula. Both Voiculescu and Haagerup worked with bounded random variables. Later, Maassen generalized the Voiculescu addition formula to the case of unbounded operators with finite variance, and Bercovici and Voiculescu further generalized these concepts to the general case of unbounded random variables. This development was especially important as it allowed the transfer of a large body of classical results about infintitely-divisible measures to the case of additive free convolutions.

Here we will prove the theorem only for bounded random variables.

6.1.4 Proof #1

Consider the following random variable:

$$X = a + \sum_{k=0}^{\infty} x_k \left(a^*\right)^k,$$

where a and a^* are the creation and annihilation operators acting on the Fock space. Define the *symbol* of X as

$$K_X(z) = z^{-1} + \sum_{k=0}^{\infty} x_k z^k.$$

Lemma 43 *The symbol $K_X(z)$ is the K-function of X in the sense of Definition 36.*

Proof: Since $G_X(z) = E\left[(z-X)^{-1}\right]$, our task is to show that

$$E\left[(K(z)-X)^{-1}\right] = z.$$

We start with computing how $K(z) - X$ acts on

$$w_z =: \xi + \sum_{n=1}^{\infty} z^n e^{\otimes n}.$$

(The series is well defined for $|z| \leq 1$.) Note that

$$aw_z = e + \sum_{n=1}^{\infty} z^n e^{\otimes n} = (w_z - \xi)/z,$$

and

$$a^* w_z = z\xi + \sum_{n=2}^{\infty} z^n e^{\otimes (n-1)} = z w_z.$$

Therefore,

$$Xw_z = \frac{w_z - \xi}{z} + \left(\sum_{k=0}^{\infty} x_k z^k\right) w_z = K(z)\, w_z - \frac{\xi}{z}.$$

In other words,

$$(K(z) - X)\, w_z = \xi/z.$$

Now, recall that $K(z)$ has a pole at $z=1$. That means that for all z in a sufficiently small circle around 0, we have $|K(z)| \geq \|X\|$. Consequently, operator $K(z)-X$ is invertible, and we can write:

$$(K(z)-X)^{-1}\xi = zw_z.$$

Hence,

$$\begin{aligned} E\left[(K(z)-X)^{-1}\right] &= \left\langle \xi, (K(z)-X)^{-1}\xi \right\rangle \\ &= \langle \xi, zw_z \rangle = z. \end{aligned}$$

QED.

Proof of Voiculescu's addition theorem:

Let A and B be arbitrary free random variables with the Cauchy transforms $G_A(z)$ and $G_B(z)$. Let the corresponding K-functions be $K_A(z)$ and $K_B(z)$ and define the Toeplitz variables X and Y as the variables that have these functions as their symbols. Then by Lemma 43, X and Y have the same K-functions as A and B. Therefore they have the same Cauchy transforms and the same moments, and therefore X and Y represent A and B, respectively. Consequently $X+Y$ represents $A+B$. In particular, $X+Y$ has the same K-function as $A+B$. But by Theorem 33, $X+Y$ is equivalent to the Toeplitz variable Z that has the symbol $K_Z(z) = K_A(z) + K_B(z) - z^{-1}$. Hence, both $X+Y$ and $A+B$ have a K-function equal to $K_Z(z) = K_A(z) + K_B(z) - z^{-1}$. QED.

Alternatively, we could avoid using Theorem 33 and prove the following lemma directly.

Lemma 44 *Let $X = a_1 + \sum_{k=0}^{\infty} x_k (a_1^*)^k$ and $Y = a_2 + \sum_{k=0}^{\infty} y_k (a_2^*)^k$. Then*

$$G_{X+Y}\left(K_X(z) + K_Y(z) - \frac{1}{z}\right) = z.$$

Proof: We need to prove that

$$E\left[\left(K_X(z) + K_Y(z) - \frac{1}{z} - X - Y\right)^{-1}\right] = z.$$

First, we investigate how $K_X(z) + K_Y(z) - z^{-1} - X - Y$ acts on

$$\rho_z =: \xi + \sum_{n=1}^{\infty} z^n (e_1 + e_2)^n,$$

where $(e_1 + e_2)^n$ is a short notation for the sum of tensor products of e_1 and e_2 obtained from expansion of the tensor product $(e_1 + e_2)^{\otimes n}$. (The series for ρ_z is well-defined for $|z| < 1/2$.)

First, note that

$$(a_1 + a_2)\,\rho_z = \sum_{n=0}^{\infty} z^n\,(e_1 + e_2)^{n+1} = \frac{\rho_z - \xi}{z},$$

and

$$(a_1^*)^k\,\rho_z = (a_2^*)^k\,\rho_z = z^k\xi + \sum_{n=1}^{\infty} z^{n+k}\,(e_1 + e_2)^n = z^k\rho_z.$$

This implies that

$$(X + Y)\,\rho_z = \frac{\rho_z - \xi}{z} + \left(K_X - \frac{1}{z}\right)\rho_z + \left(K_Y - \frac{1}{z}\right)\rho_z,$$

and therefore that

$$\left(K_X(z) + K_Y(z) - \frac{1}{z} - X - Y\right)\rho_z = \frac{\xi}{z}.$$

If z is sufficiently small, such that $|K_X(z) + K_Y(z) - z^{-1}| > \|X\| + \|Y\|$, we can invert the operator on the left of the previous equality and get:

$$\left(K_X(z) + K_Y(z) - \frac{1}{z} - X - Y\right)^{-1}\xi = z\rho_z.$$

Therefore,

$$\begin{aligned} E\left(K_X(z) + K_Y(z) - \frac{1}{z} - X - Y\right)^{-1} &= \left\langle \xi, \left(K_X(z) + K_Y(z) - \frac{1}{z} - X - Y\right)^{-1}\xi \right\rangle \\ &= \langle \xi, z\rho_z \rangle = z. \end{aligned}$$

QED.

6.1.5 Proof #2

This proof was found by researchers who studied random walks on free products of discrete groups (see: Figa-Talamanca and Steger (1994)). It was found at about the same time as Voiculescu's proof but is less well known. Recently this proof was revived by Lehner (1999) The advantage of this proof is that it does not require the machinery of the operators acting on Fock space.

Let X and Y are two free operators. Let $R_X(z) = (z - X)^{-1}$ and $R_Y(z) = (z - Y)^{-1}$, i.e., $R_X(z)$ and $R_Y(z)$ are resolvents of operators X and Y, respectively. The the Cauchy transforms are simply expectations of these resolvents: $G_X(z) = E(R_X(z))$ and $G_Y(z) = E(R_Y(z))$. We are interested in computing $G_{X+Y}(z)$, so we start with a calculation of $R_{X+Y}(z)$, that is, of $(z - (X+Y))^{-1}$. It is based on the following Proposition:

Proposition 45 *Let* $(I - X)^{-1} = I + S_X$, $(I - Y)^{-1} = I + S_Y$, *and* $(I - (X+Y))^{-1} = I + S_{X+Y}$. *Then*

$$S_{X+Y} = \sum_{n=1}^{\infty} \sum\nolimits_{i_1 \neq i_2 \neq \ldots \neq i_n} S_{i_1} S_{i_2} \ldots S_{i_n}, \tag{12}$$

where i_k *take values* X *or* Y.

Proof: First, we claim that

$$I + \sum_{n=1}^{\infty} \sum\nolimits_{i_1 \neq i_2 \neq \ldots \neq i_n} S_{i_1} S_{i_2} \ldots S_{i_n} = \left(I - \frac{S_X}{I + S_X} - \frac{S_Y}{I + S_Y} \right)^{-1}. \tag{13}$$

Indeed, we can split the epxreission on the left-hand side of (13) into 4 parts:

$$\begin{aligned}
\langle 1 \rangle &= I + S_X S_Y + S_X S_Y S_X S_Y + \ldots \\
&= \frac{I}{I - S_X S_Y}; \\
\langle 2 \rangle &= S_X + S_X S_Y S_X + S_X S_Y S_X S_Y S_X + \ldots \\
&= \frac{I}{I - S_X S_Y} S_X; \\
\langle 3 \rangle &= S_Y S_X + S_Y S_X S_Y S_X + S_Y S_X S_Y S_X S_Y S_X + \ldots \\
&= S_Y \frac{I}{I - S_X S_Y} S_X, \text{ and} \\
\langle 4 \rangle &= S_Y + S_Y S_X S_Y + S_Y S_X S_Y S_X S_Y + \ldots \\
&= S_Y \frac{I}{I - S_X S_Y}.
\end{aligned}$$

In words, $\langle 1 \rangle$ are those terms that start with S_X and end with S_Y (and also I), $\langle 2 \rangle$ are those that start with S_X and end with S_X, $\langle 3 \rangle$ are those that start with S_Y and end with S_X, and $\langle 4 \rangle$ are those that start with S_Y and end with S_Y.

Now we can compute:

$$\begin{aligned}
\langle 1\rangle + \langle 2\rangle &= \frac{I}{I - S_X S_Y}\left(I + S_X\right); \\
\langle 3\rangle + \langle 4\rangle &= S_Y \frac{I}{I - S_X S_Y}\left(I + S_X\right), \text{ and} \\
\langle 1\rangle + \langle 2\rangle + \langle 3\rangle + \langle 4\rangle &= \left(I + S_Y\right)\frac{I}{I - S_X S_Y}\left(I + S_X\right). \qquad (14)
\end{aligned}$$

On the other hand, we can compute

$$\begin{aligned}
I - \frac{S_X}{I + S_X} - \frac{S_Y}{I + S_Y} &= \left(I + S_X\right)^{-1}\left(I + S_X\right)\left(I + S_Y\right)\left(I + S_Y\right)^{-1} \\
&\quad - \left(I + S_X\right)^{-1} S_X \left(I + S_Y\right)\left(I + S_Y\right)^{-1} \\
&\quad - \left(I + S_X\right)^{-1}\left(I + S_X\right) S_Y \left(I + S_Y\right)^{-1} \\
&= \left(I + S_X\right)^{-1}\left(I - S_X S_Y\right)\left(I + S_Y\right)^{-1}.
\end{aligned}$$

Consequently,

$$\left(I - \frac{S_X}{I + S_X} - \frac{S_Y}{I + S_Y}\right)^{-1} = \left(I + S_Y\right)\frac{I}{I - S_X S_Y}\left(I + S_X\right) \qquad (15)$$

Comparing (14) and (15), we conclude that (13) holds.

Note that by assumption $S_X\left(I + S_X\right)^{-1} = X$ and $S_Y\left(I + S_Y\right)^{-1} = Y$. Therefore,

$$\left(I - \frac{S_X}{I + S_X} - \frac{S_Y}{I + S_Y}\right)^{-1} = \left(I - (X + Y)\right)^{-1} = I + S_{X+Y}.$$

Hence,

$$S_{X+Y} = \sum_{n=1}^{\infty} \sum\nolimits_{i_1 \neq i_2 \neq \ldots \neq i_n} S_{i_1} S_{i_2} \ldots S_{i_n}.$$

QED.

To use the property from the definition of free probabilities we would like S_X and S_Y to have zero expectation. For this purpose it is useful to reformulate the previous proposition in a slightly different form, which is less beautiful but easier to apply in our case

Proposition 46 *Let two functions $f_X(z)$ and $f_Y(z)$ are given and suppose that $f_X(z) = z^{-1} + O(1)$ and $f_Y(z) = z^{-1} + O(1)$ for small z. Suppose also that*

$$\left(f_X(z) - X\right)^{-1} = z\left(I + S_X(z)\right),$$

$$\left(f_Y(z) - Y\right)^{-1} = z\left(I + S_Y(z)\right),$$

and

$$\left(-z^{-1} + f_X(z) + f_Y(z) - (X+Y)\right)^{-1} = z\left(I + S_{X+Y}(z)\right)$$

for some operator-valued functions $S_X(z)$, $S_Y(z)$, *and* $S_{X+Y}(z)$. *Then*

$$S_{X+Y}(z) = \sum_{n=1}^{\infty} \sum\nolimits_{i_1 \neq i_2 \neq \ldots \neq i_n} S_{i_1}(z) S_{i_2}(z) \ldots S_{i_n}(z), \tag{16}$$

where i_k *take values* X *or* Y.

Proof: We can still apply formula (13) and write

$$I + \sum_{n=1}^{\infty} \sum\nolimits_{i_1 \neq i_2 \neq \ldots \neq i_n} S_{i_1}(z) S_{i_2}(z) \ldots S_{i_n}(z) = \left(I - \frac{S_X(z)}{I + S_X(z)} - \frac{S_Y(z)}{I + S_Y(z)}\right)^{-1}. \tag{17}$$

Next, we calculate $S_X(z)\left(I + S_X(z)\right)^{-1} = 1 - z f_X(z) + zX$ and $S_Y(z)\left(I + S_Y(z)\right)^{-1} = 1 - z f_Y(z) + zY$. Therefore,

$$\left(I - \frac{S_X(z)}{I + S_X(z)} - \frac{S_Y(z)}{I + S_Y(z)}\right)^{-1} = \frac{1}{z}\left(-\frac{1}{z} + f_X(z) + f_Y(z) - (X+Y)\right)^{-1}.$$

In combination with (17), this gives:

$$\frac{1}{-\frac{1}{z} + f_X(z) + f_Y(z) - (X+Y)} = z\left(I + \sum_{n=1}^{\infty} \sum\nolimits_{i_1 \neq i_2 \neq \ldots \neq i_n} S_{i_1}(z) S_{i_2}(z) \ldots S_{i_n}(z)\right). \tag{18}$$

QED.

Now we can prove the addition formula. Indeed, take $f_X(z) = K_X(z)$, that is, the functional inverse of the Cauchy transform, and define $S_X(z)$ as in the previous proposition:

$$z\left(I + S_X(z)\right) = \frac{I}{K_X(z) - X}.$$

This is possible because $K_X(z) = z^{-1} + O(1)$ near zero. Then, taking the expectation on both sides and using the definition of the Cauchy transform, we can write:

$$E\left[z\left(I + S_X(z)\right)\right] = E\frac{I}{K_X(z) - X} = G_X\left(K_X(z)\right) = z.$$

It follows that $E\left(S_X(z)\right) = 0$ for every z, for which $S_X(z)$ is defined. Similarly, $E\left(S_Y(z)\right) = 0$. Also $S_X(z)$ and $S_Y(z)$ are free.

Then we take expectation on both sides of (18) and use the main property of free variables to obtain:

$$E\frac{1}{-\frac{1}{z}+K_X(z)+K_Y(z)-(X+Y)}=z.$$

It follows that the functional inverse of the Cauchy transform for $X+Y$ is equal to $-z^{-1}+K_X(z)+K_Y(z)$. That is,

$$K_{X+Y}(z)=-z^{-1}+K_X(z)+K_Y(z).$$

QED.

6.2 Multiplication

If X and Y are free, we can consider their product XY. The moments of XY are determined uniquely by moments of X and Y. Below we will see that there is an efficient way to calculate them. Before this we want to address a question if we can define a concept of free multiplicative convolution similar to the concept of free additive convolution. Let X and Y be two free self-adjoint random variables with the spectral probability measures μ and ν. The difficulty is that even though X and Y are self-adjoint their product is in general not self-adjoint and so we can define its spectral probability measure. Moreover, apparently there is no guarantee that the moments of the product XY correspond to moments of a probability measure on the real line.

However, at least in the case of non-commutative probability space with tracial expectation, the following results holds.

Theorem 47 *Let X and Y be positive self-adjoint variables in a non-commutative probability space with tracial expectation. Then the k-th moments of random variables XY, YX, $X^{1/2}YX^{1/2}$, $Y^{1/2}XY^{1/2}$ are the same:*

$$E\left((XY)^k\right)=E\left((YX)^k\right)=E\left(\left(X^{1/2}YX^{1/2}\right)^k\right)=E\left(\left(Y^{1/2}XY^{1/2}\right)^k\right)$$

for any integer $k\geq 0$.

The proof is obvious from the definition of the tracial expectation. Note that for this result we do not even need freeness of variables X and Y.

Since $X^{1/2}YX^{1/2}$ is self-adjoint, this theorem shows that in non-commutative probability spaces with tracial expectation, moments of the product XY equals moments of a probability distribution on the real line. This probability distribution equals the spectral probability distribution of both $X^{1/2}YX^{1/2}$ and $Y^{1/2}XY^{1/2}$. In view of this we introduce the following definition.

Definition 48 *Let μ and ν be the spectral probability measures of free positive self-adjoint random variables A and B, respectively. Then the spectral probability measure of $A^{1/2}BA^{1/2}$ is called the* free additive convolution *of measures μ and ν, and denoted $\mu \boxtimes \nu$.*

For calculation of moments of the product XY, we introduce the *S-transform*. . It was also invented by Voiculescu. Here is how it is defined.

Let A be bounded operator in a non-commutative probability space. Define

$$\psi_A(z) = E\left(\frac{1}{1 - zA}\right) - 1 = \sum_{k=1}^{\infty} E\left(A^k\right) z^k. \tag{19}$$

If $E(A) \neq 0$, then in a sufficiently small neigborhood of 0, an inverse of $\psi_A(z)$ is defined, which we denote as $\psi_A^{-1}(z)$. The *S-transform* is defined as

$$S_A(z) = \left(1 + \frac{1}{z}\right) \psi_A^{-1}(z). \tag{20}$$

In other words, from (19) and (20) the defining functional relation for $S(z)$ is as follows:

$$E\left(\frac{1}{1 - \frac{z}{1+z} S_A(z) A}\right) = 1 + z. \tag{21}$$

Now, let us write out several first terms in the power expansions for $\psi(z)$, $\psi^{-1}(z)$, and $S(z)$. Assume for simplicity that $E(A) = 1$.

$$\begin{aligned}
\psi(z) &= z + m_2 z^2 + m_3 z^3 + \ldots, \\
\psi^{-1}(z) &= z - m_2 z^2 - \left(m_3 - 2m_2^2\right) z^3 + \ldots, \\
S(z) &= 1 + (1 - m_2)z + \left(2m_2^2 - m_2 - m_3\right) z^2 + \ldots
\end{aligned}$$

The main theorem regarding the multiplication of free random variables was proved by Voiculescu. The original proof by Voiculescu (1987) was very complicated. We give here Haagerup's simplified version (1997).

Theorem 49 (Voiculescu's multiplication formula) *Suppose X and Y are bounded free random variables. Suppose also that $E(X) \neq 0$ and $E(Y) \neq 0$. Then*

$$S_{XY}(z) = S_X(z) S_Y(z).$$

Proof: Consider the following random variables: $X = (1 + a_1) f(a_1^*)$ and $Y = (1 + a_2) g(a_2^*)$, where a_1 and a_2 are two creation operators, a_1^* and a_2^* are corresponding annihilation operators, and $f(z)$ and $g(z)$ are two functions analytical near $z = 0$ and such that $f(0) \neq 0$ and $g(0) \neq 0$. The variables X and Y are clearly free. The claim is that:

1. Every variable A with $E(A) \neq 0$ can be represented as $(1 + a) f(a^*)$, where $f(a^*)$ is an appropriate function; and

2. $$S_X = \frac{1}{f(z)},\ S_Y = \frac{1}{g(z)},\ \text{and}\ S_{XY} = \frac{1}{f(z)\, g(z)}.$$

For the first claim, recall that by Lemma 30, p. 31, every bounded random variable A can be represented by a Toeplitz random variable of the following form: $a + g(a^*)$, where $g(z)$ is a function analytic in a neighborhood of $z = 0$. Then we can define

$$f(z) = \frac{1 + zg(z)}{1 + z}.$$

Note that this implies that: 1) $f(z)$ is analytic in a neigborhood of $z = 0$ and the constant term in its Taylor expansion equals 1, and 2)

$$g(z) = \frac{(1 + z) f(z) - 1}{z}.$$

Then we can write:

$$\begin{aligned}
(1 + a) f(a^*) &= f(a^*) + a + \left.\frac{f(z) - 1}{z}\right|_{z=a^*} \\
&= a + \left.\frac{zf(z) + 1f(z) - 1}{z}\right|_{z=a^*} \\
&= a + g(a^*).
\end{aligned}$$

Therefore, every random variable A can be represented by a Toeplitz random variable of the form $(1 + a) f(a^*)$.

Now let us turn to the second claim. To calculate $S_X(z)$ (and prove that $S_X(z) = 1/f(z)$), we will aim to find such a state ω_z that for some function $h(z)$, the operator $h(z) - X$ annihilates ω_z, i.e., $(h(z) - X)\omega_z = r(z)\xi$, where ξ is the vacuum vector and $r(z)$ is some other function. Then $\omega_z / r(z) = (h(z) - X)^{-1}\xi$ and we can calculate $E\left((h(z) - X)^{-1}\right)$ as $\langle \xi, \omega_z \rangle / r(z)$. This clearly will allow us to compute $S_X(z)$.

Thus, we are looking for a quasi-eigenstate ω_z that has the following defining property:

$$X\omega_z = h(z)\omega_z - r(z)\xi.$$

It turns out that $\omega_z = \xi + \sum_{n=1}^{\infty} z^n e_1^n$ is exactly what we need. (We can also write $\omega_z = (1 - za_1)^{-1}\xi$.) First, it is easy to see that the following two formulas hold:

$$\begin{aligned} a_1\omega_z &= \frac{1}{z}(\omega_z - \xi), \text{ and} \\ a_1^*\omega_z &= z\omega_z. \end{aligned}$$

The second formula implies that $f(a_1^*)\omega_z = f(z)\omega_z$. Therefore,

$$\begin{aligned} X\omega_z &= (1 + a_1) f(z)\omega_z \\ &= f(z)\left(\omega_z + \frac{1}{z}(\omega_z - \xi)\right) \\ &= f(z)\left(\frac{1+z}{z}\omega_z - \frac{1}{z}\xi\right), \end{aligned} \tag{22}$$

which has the desired form.

Therefore,

$$\left(f(z)\frac{1+z}{z} - X\right)\omega_z = \frac{f(z)}{z}\xi.$$

Since $f(0) \neq 0$, the operator on the right-hand side is invertible for all sufficiently small z and we have

$$\begin{aligned} \omega_z &= \left(f(z)\frac{1+z}{z} - X\right)^{-1}\frac{f(z)}{z}\xi \\ &= \left(1 - \frac{z}{(1+z)f(z)}X\right)^{-1}\frac{1}{1+z}\xi. \end{aligned}$$

Therefore,

$$E\left(1 - \frac{z}{(1+z)f(z)}X\right)^{-1} = \langle \xi, (1+z)\omega_z \rangle = 1 + z.$$

Comparison with (21) shows that $S_X(z) = 1/f(z)$. The proof for $S_Y(z)$ is similar.

A harder problem is to find a quasi-eigenstate for $XY = (1+a_1) f(a_1^*)(1+a_2) g(a_2^*)$. It turns out that an appropriate state is

$$\begin{aligned} \sigma_z &= (1 - z(a_1 + a_2 + a_1 a_2))^{-1} \xi \\ &= \xi + \sum_{n=1}^{\infty} z^n (e_1 + e_2 + e_1 e_2)^n . \end{aligned}$$

First, we have

$$a_2^* \sigma_z = \sum_{n=1}^{\infty} z^n (e_1 + e_2 + e_1 e_2)^{n-1} = z\sigma_z,$$

and therefore

$$g(a_2^*)\sigma_z = g(z)\sigma_z,$$

and

$$(1 + a_2) g(a_2^*)\sigma_z = g(z)(1 + a_2)\sigma_z.$$

Next, we note that

$$\begin{aligned} a_1^* \sigma_z &= (1 + e_2) \sum_{n=1}^{\infty} z^n (e_1 + e_2 + e_1 e_2)^{n-1} \\ &= z(1 + a_2)\sigma_z. \end{aligned}$$

This implies that

$$a_1^*(1 + a_2)\sigma_z = a_1^* \sigma_z = z(1 + a_2)\sigma_z,$$

and therefore

$$f(a_1^*)(1 + a_2)\sigma_z = f(z)(1 + a_2)\sigma_z.$$

Altogether we get

$$\begin{aligned} XY\sigma_z &= (1 + a_1) f(a_1^*)(1 + a_2) f(a_2^*)\sigma_z \\ &= f(z) g(z)(1 + a_1)(1 + a_2)\sigma_z. \end{aligned}$$

However,

$$\begin{aligned} (1 + a_1)(1 + a_2)\sigma_z &= \sigma_z + \sum_{n=1}^{\infty} z^n (e_1 + e_2 + e_1 e_2)^{n+1} \\ &= \sigma_z + \frac{\sigma_z - \xi}{z} \\ &= \frac{z+1}{z}\sigma_z - \frac{1}{z}\xi. \end{aligned}$$

Therefore,

$$XY\sigma_z = f(z)\, g(z) \left(\frac{z+1}{z}\sigma_a - \frac{1}{z}\xi \right).$$

This equation has exactly the same form as equation (22). Therefore we can repeat the arguments that we made after equation (22) and conclude that $S_{XY} = 1/\left(f(z)\, g(z)\right)$. This implies that $S_{XY} = S_X S_Y$. QED.

Here is an application of the multiplication formula. For a probability distribution on the real line, μ, we intoduce the following notation:

$$\begin{aligned} E(\mu) &= \int t d\mu(t), \text{ and} \\ \operatorname{Var}(\mu) &= \int t^2 d\mu(t) - (E\mu)^2. \end{aligned}$$

Theorem 50 *Suppose that μ_i are probability distributions such that $E(\mu_i) = 1$. Let*

$$\mu^{(n)} = \mu_1 \boxtimes ... \boxtimes \mu_n$$

Then $\operatorname{Var}\left(\mu^{(n)}\right) = \sum_{i=1}^{n} \operatorname{Var}(\mu_i)$.

Proof: Let X_i be self-adjoint variables with distributions μ_i. Then $EX_i^2 = m_2^{(i)} > 1$. We can write

$$S_{X_i}(z) = 1 + (1 - m_2^{(i)})z + ...$$

Then

$$S_{\Pi_n}(z) = \prod_{i=1}^{n} S_{X_i}(z) = 1 + \sum_{i=1}^{n} \left(1 - m_2^{(i)}\right) z + ...$$

From this we can conclude that $E(\Pi_n^2) = 1 - \sum_{i=1}^{n} \left(1 - m_2^{(i)}\right)$, or $E(\Pi_n^2) - 1 = \sum_{i=1}^{n} \left(m_2^{(i)} - 1\right)$. In other words, $\operatorname{Var}\left(\mu^{(n)}\right) = \sum_{i=1}^{n} \operatorname{Var}(\mu_i)$. QED.

This interesting observation does not have an analogue in the case of multiplication of classical random variables.

7 Analytical Properties of Cauchy's Transforms and Their Functional Inverses

The main tools for investigating free convolutions of probability measures are Cauchy transforms and their functional inverses. It is important to know answers to the following questions:

1) Are Cauchy transforms in one-to-one correspondence with probability measures?

2) How can we compute a probability measure from its Cauchy transform?

3) How are properties of Cauchy transforms related to properties of corresponding probability measures?

4) How is the convergence of Cauchy transforms related to the convergence of probability measures?

3) Which analytical properties distinguish Cauchy transforms among all other analytic functions?

We can ask similar questions about functional inverses of Cauchy transforms. In addition, we have a very important question about the relation of properties of the Cauchy transform to properties of its functional inverse. In this section we compile answers to these questions.

We will call a function *holomorphic* at a point z if it can be represented by a convergent power series in a sufficiently small disc with the center at z. We call the function *holomorphic in an open domain*, D, if it is holomorphic at every point of the domain. Here D may include $\{\infty\}$, in which case it is a part of the extended complex plane $\mathbb{C} \cup \{\infty\}$ with the topology induced by the stereographic projection of the Riemann sphere on the extended complex plane.

The integral representation of the Cauchy transform shows that the Cauchy transform of every probability measure, $G(z)$, is a holomorphic function in

$$\mathbb{C}^{+} = \{z \in \mathbb{C} | \operatorname{Im} z > 0\}$$

and

$$\mathbb{C}^{-} = \{z \in \mathbb{C} | \operatorname{Im} z < 0\}.$$

If in addition the measure is assumed to be supported on interval $[-L, L]$, then the Cauchy transform is holomorphic in the area $\Omega : |z| > L$, where it can be represented by a convergent power series of z^{-1} :

$$\widetilde{G}(z) = \frac{1}{z} + \frac{m_1}{z^2} + \frac{m_2}{z^3} + ... \tag{23}$$

Here m_k denote the moments of the measure μ:

$$m_k = \int_{-\infty}^{\infty} t^k \mu(dt).$$

In particular, $G(z)$ is holomorphic at $\{\infty\}$.

In general, for an unbounded probability measure, series (23) is not convergent. In this case, the main tool for the study of properties of the Cauchy transform is the so-called Nevanlinna representation. We repeat the statement and refer for a proof to the book by Akhieser (1961).

7.1 Properties of Cauchy transforms

Representation formulas and characterization

The basis for the analysis of the Cauchy transform of a non-necessarily bounded probability measure is the Nevanlinna representation theorem. The theorem characterizes the analytic functions that map the upper half-plane to itself. The formula is the natural outgrowth of the remarkable Schwarz formula that represents a function analytic in a neighborhood of a disc through the values of its real part on the boundary of the disc.

Theorem 51 (Schwarz' formula) *Suppose $f(z)$ is a function analytic in the disc $|z| < 1$. Then there exists a real C such that for any z in the disc $|z| < R < 1$, the following formula holds:*

$$f(z) = iC + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{Re^{i\theta} + z}{Re^{i\theta} - z} u(R, \theta)\, d\theta,$$

where $u(R, \theta) = \operatorname{Re} f\left(Re^{i\theta}\right)$.

The proof is very ingenious. First, the Cauchy formula is adjusted by adding a (non-analytic) function so that the kernel in this formula is real on the circle with radius R. By taking the real part, it follows that the real part of $f(z)$ inside this circle equals to the integral of the real part.of $f(z)$ on the circle against the kernel. Then we find an analytic function that has this kernel as its real part on the circle. If we substitute this function as the new kernel we obtain the representation of $f(z)$ as the integral of $\operatorname{Re} f(z)$ against this new kernel. We wil skip the details.

From the Schwarz formula it is easy to get the following theorem:

Theorem 52 (Herglotz' Representation) *Let $f(z)$ be a function, analytic inside the unit disc and taking values in the upper half-plane. Then $f(z)$ has a unique representation of the form*

$$f(z) = C + \frac{i}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} d\sigma(\theta), \tag{24}$$

where $\sigma(\theta)$ is a non-decreasing real-valued function of finite variation and C is a real constant. Conversely, if (24) holds then $f(z)$ is an analytic map of the unit disc to the upper half-plane.

And next a change of variables gives the characterization of functions that are analytic in the upper half-plane and map the upper half-plane to itself. Akhieser calls these functions the Nevanlinna class.

Theorem 53 (Nevanlinna's Representation) *Function $f(z)$ is an analytic map of the upper half-plane to itself if and only if it has a unique representation of the form*

$$f(z) = \mu z + \nu + \int_{-\infty}^{\infty} \frac{1 + uz}{u - z} d\tau(u). \tag{25}$$

where $\tau(u)$ is a non-decreasing function of finite variation, μ and ν are real and $\mu \geq 0$.

An important property of the Cauchy transforms that follows from the Nevanlinna representation is as follows.

Theorem 54 *The following statements are equivalent.*
i) A function, $G(z)$, is the Cauchy transform of a probability measure on R;
ii) $G(z)$ is a holomorphic function mapping $\mathbb{C}^{+}$ (the open upper half-plane) to $\mathbb{C}^{-}$ (the open lower half-plane) and

$$\lim_{y \to \infty} iyG(iy) = 1; \tag{26}$$

iii) $G(z)$ is a holomorphic function mapping $\mathbb{C}^{+}$ to $\mathbb{C}^{-}$ and

$$\lim_{z \to \infty,\, z \in \Gamma_{\alpha}} zG(z) = 1, \tag{27}$$

where $\Gamma_{\alpha} = \{z \in C^{+} | \operatorname{Re} z < \alpha \operatorname{Im} z\}$.

Remark: The notation $z \to \infty$, $z \in \Gamma_{\alpha}$ means that z approaches ∞ along any sequence of values of z that belong to Γ_{α}. We will say that *z approaches ∞ in the set Γ_{α}*.

For proof, see Proposition 5.1 in Bercovici and Voiculescu (1993).

The Perron-Stieltjes inversion formula

Theorem 55 *Measure $\mu(B)$ can be recovered from its Cauchy transform $G(z)$ by the formula*

$$\mu(B) = -\frac{1}{\pi} \lim_{\varepsilon \downarrow 0} \int_B \operatorname{Im} G(x + i\varepsilon)\, dx,$$

which is valid for all Borel B such that $\mu(\partial B) = 0$.

For proof of this theorem, see Akhieser (1961). A simple concequence of the Perron-Stieltjes inversion formula is the following result, which we will need later.

Lemma 56 *Suppose that*
1) $G(z)$ is the Cauchy transform of a compactly supported probability distribution, μ, and
2) $G(z)$ is holomorphic at every $z \in \mathbb{R}$, $|z| > L$.
Then the support of μ lies entirely in the interval $[-L, L]$.

Proof: From assumption 1) we infer that in some neighborhood of infinity $G(z)$ can be represented by the convergent power series (23) and that $G(z)$ is also holomorphic everywhere in $\mathbb{C}^+$ and $\mathbb{C}^-$. Therefore, using assumption 2) we can conclude that $G(z)$ is holomorphic everywhere in the area $|z| > L$, including the point at infinity. It follows that the power series (23) converges everywhere in the area $|z| > L$. Since this power series has real coefficients we can conclude that $G(z)$ is real for $z \in R$, $|z| > L$. Also, since $G(z)$ is holomorphic and therefore continuous in $|z| > L$, we can conclude that $\lim_{\varepsilon \downarrow 0} \operatorname{Im} G(z + i\varepsilon) = 0$. Then the Stieltjes inversion formula implies that $\mu([a, b]) = 0$ for each pair of a and b which belong to $|x| > L$ and such that $\mu(a) = 0$ and $\mu(b) = 0$. It remains to prove that this impies that $\mu\{|x| > L\} = 0$.

For this purpose, note that the set of points $x \in \mathbb{R}$, for which $\mu(x) > 0$ is at most countable. Indeed, let S be the set of all x for which $\mu(x) > 0$. We can divide this set into a countable collection of disjoint subsets S_k, where k are all positive integers and $S_k = \left\{x | k^{-1} \geq \mu\{x\} > (k+1)^{-1}\right\}$. Clearly, every S_k is either empty or contains a finite number of points x. Otherwise, we could take an infinite countable sequence of $x_{i,k} \in S_k$, and we would get (by countable additivity and monotonicity of μ) that $\mu(S_k) \geq \sum_i \mu(x_{i,k}) = +\infty$. By the monotonicity of μ we would further get $\mu(\mathbb{R}) = +\infty$, which would contradict the assumption that μ is a probability measure. Therefore, S is a countable union of finite subsets S_k and hence countable.

From the countability of S we conclude that the set of points x, for which $\mu(x) = 0$ (i.e., S^c), is dense in the set $|x| > L$. Indeed, take an arbitrary non-empty interval (α, β). Then $(\alpha, \beta) \cap S^c \neq \emptyset$, since otherwise $(\alpha, \beta) \subset S$ and therefore S would be uncountable. Hence S^c is countable. This fact and the countable additivity of μ implies that $\mu(\{|x| > L\}) = 0$. Indeed, using the denseness of S^c we can cover the set $\{|x| > L\}$ by a countable union of disjoint intervals $[a, b]$, where $\mu(a) = 0$ and $\mu(b) = 0$. For each of these intervals, $\mu([a, b]) = 0$, and therefore countable additivity implies that $\mu(\{|x| > L\}) = 0$. Consequently, μ is supported on a set that lies entirely in $[-L, L]$. QED.

Convergence of Cauchy transforms and weak convergence of probability measures

Theorem 57 *If $\mu_n \to \mu$ weakly, then $G_{\mu_n}(z) \to G_\mu(z)$ uniformly on compact subsets.*

Proof Write

$$\left|G_{\mu_n}(z) - G_\mu(z)\right| = \left|\int \frac{1}{z-t} d\mu_n(t) - \int \frac{1}{z-t} d\mu(t)\right|.$$

Since $(z-t)^{-1}$ is bounded and continuous in t for every $z \in \mathbb{C}^+$, we can conclude that this difference converges to zero for every $z \in \mathbb{C}^+$. Moreover, the family $f_z(t) = (z-t)^{-1}$ is equicontinuous for z in a compact subset of $\mathbb{C}^+$. This implies that the convergence of the difference to zero is uniform in z in a compact subset of $\mathbb{C}^+$. QED.

Usually we are interested in the opposite direction. For this, we cite the following theorems that relate the closeness of two Cauchy transforms with the closeness of the corresponding probability distributions. These theorems were proved by Bai (1993). The first is for arbitrary probability distributions on $\mathbb{R}$, and the second is for the distributions on $\mathbb{R}$ that have compact support.

Theorem 58 ((Bai (1993))) *Consider probability measures with distribution functions $\mathcal{F}$ and $\mathcal{G}$ and let their Cauchy transforms be $G_{\mathcal{F}}(z)$ and $G_{\mathcal{G}}(z)$, respectively. Let $z = x + iy$. Then*

$$\begin{aligned}\sup_x |\mathcal{F}(x) - \mathcal{G}(x)| \quad \leq \quad & \frac{1}{\pi(2\gamma - 1)}\left[\int_{-\infty}^{\infty} |G_{\mathcal{F}}(x+iy) - G_{\mathcal{G}}(x+iy)|\, dx\right. \\ & \left. + \frac{1}{y}\sup_x \int_{|u|\leq 2yc} |\mathcal{G}(x+u) - \mathcal{G}(x)|\, du\right],\end{aligned}$$

where c and γ are related by the following equality

$$\gamma = \frac{1}{\pi}\int_{|x|<c} \frac{1}{1+x^2} dx > \frac{1}{2}.$$

Theorem 59 (Bai (1993)) *Let probability measures with distribution functions $\mathcal{F}$ and $\mathcal{G}$ be supported on a finite interval $[-B, B]$ and let their Cauchy transforms be $G_{\mathcal{F}}(z)$ and $G_{\mathcal{G}}(z)$, respectively. Let $z = u + iv$. Then*

$$\begin{aligned} \sup_x |\mathcal{F}(x) - \mathcal{G}(x)| \quad &\leq \quad \frac{1}{\pi(1-\kappa)(2\gamma-1)} \left[\int_{-A}^{A} |G_{\mathcal{F}}(z) - G_{\mathcal{G}}(z)| \, du \right. \\ &\left. + \frac{1}{v} \sup_x \int_{|y| \leq 2vc} |\mathcal{G}(x+y) - \mathcal{G}(x)| \, dy \right], \end{aligned} \tag{28}$$

where $A > B$, $\kappa = \frac{4B}{\pi(A-B)(2\gamma-1)} < 1$, $\gamma > 1/2$, and c and γ are related by the following equality

$$\gamma = \frac{1}{\pi} \int_{|u|<c} \frac{1}{1+u^2} du.$$

7.2 Lagrange's formulas for the functional inverse

Here we list and prove very useful results about functional inverses of holomorphic functions. By function holomorphic in a domain D we mean function which is bounded and differentiable in D. These formulas are originally due to Lagrange (see "Nouvelle Methode pour Resoudre les Equations Litterales par le Moyen des Series" (1770) on pp. 5-73 in Lagrange (1869)) and the most typical inversion formula is as follows:

Lemma 60 (Lagrange's inversion formula)
Suppose f is a function of a complex variable, which is holomorphic in a neighborhood of $z_0 = 0$ and has the Taylor expansion

$$f(z) = a_1 z + \sum_{k=2}^{\infty} a_k z^k,$$

with $a_1 \neq 0$ and converging for all sufficiently small z. Then the functional inverse of $f(z)$ is well defined in a neighborhood of 0 and the Taylor series of the inverse is given by the following formula:

$$f^{-1}(u) = \frac{u}{a_1} + \sum_{k=2}^{\infty} \left[\frac{1}{k} \operatorname*{res}_{z=0} \frac{1}{f(z)^k} \right] u^k,$$

where $\mathrm{res}_{z=z_0}$ denotes the Cauchy residual at point z_0. Alternatively, we can write:

$$f^{-1}(u) = \frac{u}{a_1} + \sum_{k=2}^{\infty} \left[\frac{1}{2\pi i k} \oint_{\gamma} \frac{dz}{f(z)^k} \right] u^k,$$

where γ is such a circle around 0, where f has only one zero.

For the modern proof see Markushevich (1977), Theorems II.3.2 and II.3.3, or Whittaker and Watson (1927), Section 7.32. We also need the following modification of the Lagrange formula, which says how to invert a function in a neigborhood of infinity.

Lemma 61 *Suppose G is a function of a complex variable, which is holomorphic in a neighborhood of $z_0 = \infty$ and has the expansion*

$$G(z) = a_0 + \frac{a_1}{z} + \frac{a_2}{z^2} + ...,$$

converging for all sufficiently large z, where $a_1 \neq 0$. Define $g(z) = G(1/z)$. Then the functional inverse of $G(z)$ is well defined for large z. The inverse is meromorphic in a neighborhood of 0 and Laurent's series of the inverse is given by the following formula:

$$G^{-1}(w) = \frac{a_1}{w - a_0} + \frac{a_2}{a_1} - \sum_{n=1}^{\infty} \left[\frac{1}{n} \frac{1}{2\pi i} \oint_{\partial\gamma} \frac{dz}{z^2 \left(g(z) - a_0\right)^n} \right] (w - a_0)^n,$$

where γ is a closed disc around 0 in which $g(z)$ has only one zero.

Proof: Let γ be a disc around 0 in which $g(z)$ has only one zero. This disk exists because $g(0) = 0$, and $g(z)$ is analytical in a neighborhood of 0 and has a non-zero derivative at 0. Let

$$r_w = \frac{1}{2} \inf_{z \in \partial\gamma} |g(z)|.$$

Then $r_w > 0$ by our assumption on γ. We can apply Rouche's theorem and conclude that the equation $g(z) - w = 0$ has only one solution inside γ if $|w - a_0| \leq r_w$. Let us fix such a w that $|w - a_0| \leq r_w$. Inside γ, the function

$$\frac{g'(z)}{z\left(g(z) - w\right)}$$

has a pole at $z = 1/G^{-1}(w)$ with the residual $G^{-1}(w)$ and a pole at $z = 0$ with the residual $a_1/\left(a_0 - w\right)$. Consequently, we can write:

$$G^{-1}(w) = \frac{1}{2\pi i} \oint_{\partial\gamma} \frac{g'(z)dz}{z\left(g(z) - w\right)} + \frac{a_1}{w - a_0}.$$

The integral can be re-written as follows:

$$
\begin{aligned}
\oint_{\partial\gamma} \frac{g'(z)dz}{z\left(g(z)-a_0-(w-a_0)\right)} &= \oint_{\partial\gamma} \frac{g'(z)}{z\left(g(z)-a_0\right)} \frac{1}{1-\frac{w-a_0}{g(z)-a_0}} dz \\
&= \sum_{n=0}^{\infty} \oint_{\partial\gamma} \frac{g'(z)dz}{z\left(g(z)-a_0\right)^{n+1}} \left(w-a_0\right)^n .
\end{aligned}
$$

For $n = 0$ we calculate

$$
\frac{1}{2\pi i} \oint_{\partial\gamma} \frac{g'(z)dz}{z\left(g(z)-a_0\right)} = \frac{a_2}{a_1},
$$

Indeed, the only pole of the integrand is at $z = 0$ and it has order two. The corresponding residual can be computed from the series expansion for $g(z)$:

$$
\begin{aligned}
\operatorname{res}_{z=0} \frac{g'(z)dz}{z\left(g(z)-a_0\right)} &= \left.\frac{d}{dz} \frac{z^2\left(a_1+2a_2z+...\right)}{z\left(a_1z+a_2z^2+..\right)}\right|_{z=0} \\
&= \left.\frac{d}{dz} \frac{1+\left(2a_2/a_1\right)z+...}{1+\left(a_2/a_1\right)z+..}\right|_{z=0} = \frac{a_2}{a_1}.
\end{aligned}
$$

For $n > 0$ we integrate by parts:

$$
\frac{1}{2\pi i} \oint_{\partial\gamma} \frac{g'(z)dz}{z\left(g(z)-a_0\right)^{n+1}} = -\frac{1}{2\pi i} \frac{1}{n} \oint_{\partial\gamma} \frac{dz}{z^2\left(g(z)-a_0\right)^n}.
$$

QED.

Most often we will use this Lemma to invert the Cauchy transform of a probability distribution and so we formulate a Corollary:

Corollary 62 *Suppose G is a function of a complex variable, which is holomorphic in a neighborhood of $z_0 = \infty$ and has the expansion*

$$
G(z) = \frac{1}{z} + \frac{a_1}{z^2} + ...,
$$

converging for all sufficiently large z. Define $g(z) = G(1/z)$. Then the functional inverse of $G(z)$ is well defined in a neighborhood of 0 and Laurent's series of the inverse is given by the following formula:

$$
G^{-1}(w) = \frac{1}{w} + a_1 - \sum_{n=1}^{\infty} \left[\frac{1}{n} \frac{1}{2\pi i} \oint_{\partial\gamma} \frac{dz}{z^2 g(z)^n}\right] w^n,
$$

where γ is a closed disc around 0 in which $g(z)$ has only one zero.

Here is another modification of the Lagrange inversion formula. This time it says how to invert a function near the point where it has a simple pole.

Lemma 63 *Suppose f is a function of a complex variable, which is meromorphic in a neighborhood of $z_0 = a$ and has the expansion*

$$f(z) = \frac{c_{-1}}{z-a} + c_0 + c_1 (z-a) + ...,$$

converging for all z sufficiently close to a.. Then the functional inverse of $f(z)$ is well defined in a neighborhood of ∞ and Laurent's series of the inverse is given by the following formula:

$$f^{-1}(u) = a + \sum_{n=1}^{\infty} \left[\frac{1}{n} \frac{1}{2\pi i} \oint_{\partial\gamma} f(z)^n \, dz \right] u^n,$$

where γ is a closed disc around a in which $1/f(z)$ has only one zero (at $z = a$).

Proof: The proof is similar to the proof of the previous Lagrange formulas. First, let $g(z) = 1/f(z)$ and $w = 1/u$. Then $g(z)$ maps $z = a$ to $w = 0$. Note that for all sufficiently small w

$$\frac{zg'(z)}{g(z) - w}$$

has the only pole at $z = g^{-1}(w)$ and the residual is $g^{-1}(w)$. Then we can write:

$$\begin{aligned} g^{-1}(w) &= \frac{1}{2\pi i} \int_{\partial\gamma} \frac{zg'(z)}{g(z) - w} dz \\ &= \frac{1}{2\pi i} \int_{\partial\gamma} \left[\sum_{k=0}^{\infty} \frac{zg'(z)}{g(z)^{k+1}} w^k \right] dz \\ &= a + \sum_{k=1}^{\infty} \left[\frac{1}{2\pi i} \frac{1}{k} \int_{\partial\gamma} \frac{dz}{g(z)^k} \right] w^k, \end{aligned}$$

where we used integration by parts.

Therefore,

$$f^{-1}(u) = a + \sum_{k=1}^{\infty} \left[\frac{1}{2\pi i} \frac{1}{k} \int_{\partial\gamma} f(z)^k \, dz \right] \frac{1}{u^k}.$$

QED.

The Lagrange inversion formula can be illustrated by the following application, in which we estimate the power series coefficients of a K-function.

Lemma 64 *Suppose that the measure μ is supported on interval $[-L, L]$, and $K(z)$ denotes the functional inverse of its Cauchy transform $G(z)$. Then the Laurent series of $K(z)$ converge in the area $\Omega = \left\{z : 0 < |z| < (4L)^{-1}\right\}$. Write these series as*

$$K(z) = \frac{1}{z} + \sum_{k=0}^{\infty} b_k z^k.$$

Then

$$|b_k| \leq \frac{2L}{k}(4L)^k.$$

Proof: Let us apply Lemma 61 to $G(z)$ with circle γ having radius $(2L)^{-1}$. We need to check that $g(z) =: G(1/z)$ has only one zero inside this circle. It holds because

$$g(z) = z\left(1 + a_2 z^2 + a_3 z^3 + \ldots\right),$$

and inside $|z| \leq (2L)^{-1}$ we can estimate:

$$\left|a_2 z^2 + a_3 z^3 + \ldots\right| \leq L^2\left(\frac{1}{2L}\right)^2 + L^3\left(\frac{1}{2L}\right)^3 + \ldots = \frac{1}{2}, \tag{29}$$

and an application of Rouche's theorem shows that $g(z)$ has only one zero inside this circle.

Another consequence of the estimate (29) is that on the circle $|z| = (2L)^{-1}$

$$|g(z)| \geq |z|/2 = 1/(4L).$$

By Lemma 61 the coefficients in the series for the inverse of $G(z)$ are

$$b_k = \frac{1}{2\pi i k}\oint_{\partial\gamma} \frac{dz}{z^2 g(z)^k},$$

and we can estimate them as

$$|b_k| \leq \frac{2L}{k}(4L)^k. \tag{30}$$

This implies that the radius of convergence of power series for $K(z)$ is at least $(4L)^{-1}$. QED.

Corollary 65 *Suppose that the measure μ is supported on interval $[-L, L]$ and b_0 and b_1 denote its first and second moments. Then for all z, such that $|z| \leq (2L)^{-1}$, the following inequality holds:*

$$\left|K(z) - \frac{1}{z} - b_0 - b_1 z\right| \leq 8Lz^2.$$

Finally, we can also invert a fiunction at a neighbourhood of a critical point. In this case the inverse will be multivalued.

Lemma 66 (Lagrange's inversion formula) *Suppose f is a function of a complex variable, which is holomorphic in a neighborhood of z_0 and has the Taylor expansion*

$$f(z) = w_0 + a_k (z - z_0)^k + \sum_{n=k+1}^{\infty} a_n (z - z_0)^n ,$$

with $a_k \neq 0$ and converging for all sufficiently small z. Then the functional inverse of $f(z)$ is a multi-valued analytic function in a neighborhood of w_0 that has a branch point at w_0 and it can be represented by the following series:

$$f^{-1}(w) = z_0 + \sum_{n=1}^{\infty} \left[\frac{1}{n} \operatorname*{res}_{z=z_0} \psi(z)^n \right] (w - w_0)^{n/k} ,$$

where $\mathrm{res}_{z=z_0}$ denotes the Cauchy residual at point z_0, and $\psi(z)$ denote any single-valued branch of the function $1/\left[f(z) - w_0\right]^{1/k}$.

For the proof see Theorem II.3.6 in Markushevich (1977).

8 Free Infinitely Divisible Distributions

8.1 Additive infinitely divisible distributions

8.1.1 Characterization

The measure μ is *infinitely-divisible* if for any n it can be represented as

$$\mu = \mu_{1/n} \boxplus \ldots \boxplus \mu_{1/n},$$

where $\boxplus$ denotes free additive convolution and there are n terms in the sum.

Analytically, this property means that the Voiculescu transform of the measure μ is n-divisible for every n. (See page 41 for the definition of the Voiculescu transform). That is, if $\varphi_\mu(z)$ denotes the Voiculescu transform of measure μ, then $\varphi_\mu(z)/n$ must be the Voiculescu transform of a probability measure for every n. Here are two examples:

Example 67 *Semicircle distribution*

The Voiculescu transform of the semicircle distribution is

$$\varphi_{SC}(z) = K_{SC}\left(\frac{1}{z}\right) - z = \frac{1}{z}.$$

This function is clearly n-divisible because $1/(zn)$ can be obtained simply by rescaling of the original measure.

Example 68 *Marchenko-Pastur distribution*

The Voiculescu transform of the Marchenko-Pastur distribution is:

$$\begin{aligned}\varphi_{MP}(z) &= K_{MP}\left(\frac{1}{z}\right) - z \\ &= \frac{\lambda z}{z-1}.\end{aligned}$$

Again, it is evident that this function is n-divisible: $\varphi_{MP}(z)/n$ is simply the Voiculescu transform of another Marchenko-Pastur distribution with parameter $\lambda/n.$

Example 69 *Compound Marchenko-Pastur distribution*

It is easy to check that $\varphi_{aX}(z) = a\varphi_X(z/a)$. Therefore, the Marchenko-Pastur distribution with parameter λ and scaled by $a,$ has the following Voiculescu transform

$$\varphi(z) = \frac{a\lambda z}{z-a}.$$

This distribution is also evidently infinitely-divisible, as well as the sum of a finite number of such distributions, which has the Voiculescu transform:

$$\varphi(z) = \sum_{i=1}^{n} \frac{b_i z}{z - a_i}. \tag{31}$$

If the following integral is well defined and corresponds to a Voiculescu transform of a probability distribution, then this distribution is also infinitely-divisible:

$$\varphi(z) = \int_{-\infty}^{\infty} \frac{sz}{z-s} d\sigma(s), \tag{32}$$

where $\sigma(s)$ is an non-decreasing function of $s.$ We can call this probability distribution a *compound Marchenko-Pastur distribution*. Intuitively, we can think about this distribution as the superposition of the Marchenko-Pastur distributions that have size $a = s$ and intensity $\lambda = s$. The amplitude (density) assigned to each of these distributions is given by $d\sigma(s)$. For example, if we write (31) in this formula then $\sigma(s)$ has jumps at a_i and the size of the jump is b_i/a_i.

This example motivates the following theorem:

Theorem 70 (Bercovici-Voiculescu (1993)) *(1) A probability measure is infinitely divisible if and only if it has an analytic coninuation defined everywhere on C^+ with values in $C^- \cup R$.*
(2) An analytic function $\varphi : C^+ \to C^-$ is a continuation of φ_μ for infinitely divisible μ if and only if

$$\lim_{|z|\to\infty, z\in\Gamma_\alpha} \varphi(z)/z = 0$$

for some $\alpha > 0$.
(3) The following representation holds for φ_μ when μ is an infinitely divisible probability measure:

$$\varphi_\mu = \alpha + \int_{-\infty}^{\infty} \frac{1+sz}{z-s} d\sigma(s), \tag{33}$$

where α is real and $d\sigma$ is a finite measure

Remark: Note that

$$\frac{1+zs}{z-s} = \left(\frac{sz}{z-s} - \frac{s}{1+s^2}\right)\frac{1+s^2}{s^2},$$

so what we did in passing from the compound Marchenko-Pastur distribution (32) to representation (33) is introducing certain normalization that handle the convergence in the case of very small jumps s.

A similar formula in the classical case is the famous Levy-Khintchine-Kolmogorov formula (see Section 18 on page 75 in Gnedenko and Kolmogorov (1959)). If $f(t)$ is the characteristic funtion of a (classically) infinitely-divisible representation, then

$$\log f(t) = iat + \int_{-\infty}^{\infty} \left(\exp(itu) - 1 - \frac{itu}{1+u^2}\right)\frac{1+u^2}{u^2} d\sigma(u),$$

where $d\sigma$ is a finite measure.

8.1.2 Stable laws

Consider sums of identically distributed free random variables that have the following form:

$$S^{(n)} = X_1^{(n)} + ... + X_{k_n}^{(n)}.$$

In other words we consider sums of rows in an object which is called a triangular array. A triangular array is a table of random variables that consists of infinite number of rows that have a finite but variable length. We assume that the length of the rows, k_n, increases as we go down the table, that is, as n grows. The question is when the distributions of sums converge a particular probability distribution. We will assume that the variables are self-adjoint and ask when the spectral distributions of sums $S^{(n)}$ weakly converge to a particular distribution.

We can write a problem in a different form. Let variables $X_i^{(n)}$ all have the spectral distribution μ_n. We write $k_n \circ \mu_n$ for the free convolution of k_n distributions μ_n, i.e.

$$k_n \circ \mu_n =: \underbrace{\mu_n \boxplus \ldots \boxplus \mu_n}_{k_n\text{-times}},$$

where we have k_n summands in the sum and $\boxplus$ sign denotes free convolution.

The question is when $k_n \circ \mu_n \to \nu$ for a fixed distribution ν. In this case we say that a sequence (k_n, μ_n) belongs to the *domain attraction* of the infinitely divisble ν. Slightly adjusting the arguments in the section about infinitely divisible distributions we can show that ν must be an infinitely divisible distribution. Essentially we need only to use division by k_n instead of division by n and use the assumption that $k_n \to \infty$.

Therefore we can write the φ-function of ν as

$$\varphi_\nu (z) = \alpha + \int_{-\infty}^{\infty} \frac{1 + zt}{z - t} d\sigma (t) .$$

It turns out that the conditions for a probability distribution to be in the domain of attaction of an infinitely divisible law has exactly the same form in the free probability case as in the classical case. In particular, that means that if a sequence of k_n classical convolutions of measure μ_n with itself converges to the normal law as $n \to \infty$, then the sequence of k_n free additive convlutions of μ_n converges to the semicircle law. Similarly, if the sequence of k_n classical convolutions of measure μ_n converges to the point measure, then the sequence of k_n free additive convolutions of μ_n also converges to the point measure. This means that at least for identically distributed summands the conditions for the Cental Limit Theorem and the Weak Law of Large Numbers are essentially the same in free probability case as in the classical case.

This correspondence of the free and the classical cases is puzzling and we will give an explanation in Section 18.

Theorem 71 (Bercovici-Pata (1999)) *$k_n \circ \mu_n \to \nu$ if and only if the following two conditions are satisfied:*
(1)

$$k_n \frac{t^2}{1+t^2} d\mu_n(t) \to d\sigma(t),$$

where the convergence is in the sense of weak convergence, and
(2)

$$k_n \int_{-\infty}^{\infty} \frac{t}{1+t^2} d\mu_n(t) \to \alpha.$$

Note that the conditions in this theorem are exactly the same as the conditions in the classical case.

Example 72 *Convergence to Semicircle (Wigner) Law*

Let ω be the semicircle distribution. When does $k_n \circ \mu_n \to \omega$? For the semicircle distribution we have $\sigma = c\delta_0$. Therefore, the conditions are that

$$k_n \frac{t^2}{t^2+1} \mu_n \to c\delta_0,$$

and that

$$\int_{-\infty}^{\infty} k_n \frac{t}{t^2+1} d\mu_n(t) \to 0.$$

Example 73 *Convergence to free Poisson (Marchenko-Pastur) Law*

Consider a sequence of free random variables with the *Bernoulli distribution*, i.e. with the distribution that puts weight p on 1 and $q = 1 - p$ on 0. Let in our triangular array the distribution of $X_i^{(n)}$ changes from raw to raw. Namely, let we have n summands in the n-th row and let the distribution for $X_i^{(n)}$ is Bernoulli with parameter $p_n = \lambda/n$. Do the sums of these random variables converge to a limiting distribution?

Let us write the relevant functions for the Bernoulli distribution. First, the Cauchy transform is

$$\begin{aligned} G_B(z) &= \frac{1}{z} + \frac{p}{z^2} + \frac{p}{z^3} + \ldots \\ &= \frac{1}{z}\frac{z-q}{z-1}. \end{aligned}$$

From this, we can calculate the K-transform:

$$K_B(u) = \frac{1 + u + \sqrt{(1-u)^2 + 4up}}{2u}.$$

If we take convolution of n Bernoulli variables with parameter p_n we get the following K-function:

$$K_n(u) = \frac{2 - n(1-u) + n\sqrt{(1-u)^2 + 4up_n}}{2u}.$$

If $p_n = \lambda/n$, then

$$K_n(u) = \frac{1}{u} + \frac{\lambda}{1-u} + o(1),$$

where $o(1)$ is with respect to $n \to \infty$. Therefore, the sums converge to the distribution with the K-function

$$K_{MP}(u) = \frac{1}{u} + \frac{\lambda}{1-u}.$$

We can recognize this distribution as the *Marchenko-Pastur distribution*, which we defined at page 39. This distibution is a free analogue of the *Poisson distribution* because in the classical case the sums of a similar sequence of independent Bernoulli random variables converge to the Poisson random variable.

8.2 Multiplicative infinitely divisible distributions

8.2.1 Measures on the unit circle

Recall that the main tool in the analysis of free multiplicative convolution is the S-transform as it was defined in (20), p. 48. It is also convenient to define a related function (Σ-*function*):

$$\Sigma(u) = S\left(\frac{u}{1-u}\right).$$

Let us use M_T to denote the set of probability distributions on the unit circle, and M_T^* to denote those of them that have non-zero expectation: $\mu \in M_T^*$ if and only if $\mu \in M_T$ and $\int z \, d\mu(z) \neq 0$.

An element μ of M_T is called *infinitely divisble* if for any n it can be represented as a free convolution of n identical measures μ_n:

$$\mu = \underbrace{\mu_n \boxtimes ... \boxtimes \mu_n}_{n\text{-times}},$$

where $\boxtimes$ denotes free multiplicative convolution and there are n terms in the product. We want to characterise the infinitely-divisible measures in terms of their S-transforms.

One other useful way to characterize measures on the unit circle is through their Poisson transforms. Recall that the *Poisson transform* of a measure μ supported on the unit circle is defined as

$$U_\mu(z) =: \int_{-\pi}^{\pi} P(r, \omega - \theta) \, d\mu(\theta),$$

where $z = re^{i\omega}$ and $P(r, \theta)$ is the *Poisson kernel*:

$$P(r, \theta) = \frac{1}{2\pi} \frac{1 - r^2}{1 - 2r\cos\theta + r^2}.$$

(Here we have identified measures on the unit circle and on the interval $[-\pi, \pi)$: $\mu(d\theta) = \mu\{\xi : |\xi| = 1 \text{ and } \arg\xi \in d\theta\}$).

Theorem 74 *The Poisson transforms $U_{\mu_j}(z)$ of measures μ_i converge to the Poisson transform $U_\mu(z)$ of measure μ uniformly on compact subsets of the unit disc if and only if μ_j weakly converges to μ.*

Proof: Indeed, if $\mu_j \to \mu$, then by the second theorem in Section I.D.1 of Koosis (1998), $U_{\mu_j}(z) \to U_\mu(z)$, where convergence is uniform on compact subsets of the unit disc. To make the reverse implication, note that the family of probability measures on the circle is tight, i.e., compact. Therefore, the only way in which $\{\mu_j\}$ can fail to converge is when there are two subsequences of $\{\mu_j\}$ that converge to different limits. Suppose that μ' and μ'' are those two different limits. Then by the first part of the proof and by assumption about $U_{\mu_j}(z)$, they must have the same Poisson integral: $U_{\mu'}(z) = U_{\mu''}(z)$. In other words, there exists a signed measure of finite variation $(\mu' - \mu'')$, which does not vanish identically and whose Poisson transform equals 0. This is impossible. (For instance, the impossibility follows from the last theorem in Section I.D.2 of Koosis (1998).) QED.

Note in particular that a particular consequence of this Theorem is that the measures on the unit disc are in one-to-one correspondence with their Poisson transforms.

It turns out that ψ-functions are closely related to the Poisson transform. Indeed, for the measures on the unit circlel, we can re-write the definition of the ψ-function as follows:

$$\psi_\mu(z) = \int_{|\xi|=1} \frac{d\mu(\xi)}{1 - z\xi} - 1.$$

Since

$$\operatorname{Re}\frac{1}{1-\xi z} = \frac{1 - r\cos(\omega - \theta)}{1 - 2r\cos(\omega - \theta) + r^2},$$

where $\xi = e^{-i\theta}$ and $z = re^{i\omega}$, therefore,

$$\operatorname{Re}\frac{1}{1-\xi z} = \frac{1}{2} + \pi P(r, \omega - \theta)$$

Hence,

$$U_\mu(z) = \frac{1}{\pi}\operatorname{Re}\psi_\mu(z) + \frac{1}{2\pi}. \tag{34}$$

This implies in particular, (1) that ψ-functions are in one to one correspondence with the measures on the unit circle, and (2) that the uniform convergence of ψ-functions on subsets of the unit disc is equivalent to the convergence of the corresponding measures. In turn, this implies also that the analogous properties hold for the S-transforms.

Theorem 75 *If $S_{\mu_i}(u)$ converges to a function $S(u)$ uniformly inside the unit disc then μ_i weakly converges to a measure ν and $S(u)$ is the S-transform of this measure.*

Now let us turn to the question of infinitely divisible measures in M_T.

Theorem 76 *If the expectation of infinitely divisible measure is zero than it must be the uniform measure on the unit circle.*

Proof: In this case, $\mu = \nu \boxtimes \nu$ and ν must have zero expectation. But in this case it is easy to check that the definition of freeness implies that all moments of μ equals zero, and this implies that the measure is uniform. QED.

It is useful to give examples of infinitely divisible measures from M_T^*.

Example 77 *For every $\lambda > 0$, the function*

$$S(z) = \exp\left(\lambda\left(z + \frac{1}{2}\right)\right)$$

is the S-transform of an infinitely divisible measure from M_T^*.

Indeed, we only need to check that for every $\lambda > 0$, the function $\exp(\lambda(z + 1/2))$ is an S-transform of a measure from M_T^*. Let $\varepsilon = \lambda/n$ and define $\mu_\varepsilon = (\delta_\zeta + \delta_{\overline{\zeta}})/2$, where $\zeta = \sqrt{1-\varepsilon} + i\sqrt{\varepsilon}$. The we can compute that the S-transform of μ_ε is as follows:

$$S_\varepsilon(z) = 1 + \varepsilon\lambda\left(z + \frac{1}{2}\right) + O\left(\varepsilon^2\right).$$

Then it is easy to see that μ_ε^n denote the measure μ_ε convolved n-times with itself, then the S-transform of μ_ε^n converges to $S(z) = \exp\left(\lambda\left(z + \frac{1}{2}\right)\right)$. By Proposition 75 we can conclude that the limit is S-transform of a probability distribution from M_T and it is easy to see that its expectation is not zero.

Example 78 *For every $\lambda > 0$ and $t \in \mathbb{R}$, the function*

$$S(z) = \exp\left(\frac{\lambda}{z + \frac{1}{2} + it}\right)$$

is the S-transform of an infinitely divisible measure from M_T^*.

We proceed as in the previous example. Let $\varepsilon = \lambda/n$ and define $\mu_\varepsilon = (1-\varepsilon)\delta_1 + \varepsilon\delta_\zeta$, where

$$\zeta = -\frac{\frac{1}{2} - it}{\frac{1}{2} + it}.$$

(Note that $|\zeta|^2 = 1$.) The S-transform of this measure is

$$S_\varepsilon(z) = 1 + \varepsilon\frac{\lambda}{z + \frac{1}{2} + it} + O\left(\varepsilon^2\right).$$

Then the S-transform of the n-time convolution of μ_ε with itself converges to

$$\exp\left(\lambda\left(z + 1/2 + it\right)^{-1}\right).$$

Therefore this convolution converges to a probability measure from M_T^* with the desired S-transform.

It can be seen that in terms of the Σ-function, this examples can be represented in a uniform way. Namely, in both cases the Σ-function can be represented as

$$\Sigma(z) = \exp\left(\lambda \frac{1+\zeta z}{1-\zeta z}\right) \tag{35}$$

for some $\lambda > 0$ and $\zeta \in T = \{z : |z| = 1\}$.

Theorem 79 *A measure $\mu \in M_T^*$ is infinitely divisible if and only if $\Sigma(z) = \exp(u(z))$ where $u(z)$ is a function, which is analytic in the unit disc and such that $\operatorname{Re} u(z) \geq 0$ if $z \in \{z : |z| < 1\}$.*

For proof, see Bercovici and Voiculescu (1992).

8.2.2 Measures on $\mathbb{R}^+$

The infinitely-divisible measures on $\mathbb{R}^+$ are defined similarly to infinitely-divisible measures on the unit circle. A measure on $\mathbb{R}^+$ is infinitely-divisible if and only if for any positive integer n we can find a measure μ_n on $\mathbb{R}^+$, such that

$$\mu = \underbrace{\mu_n \boxtimes \ldots \boxtimes \mu_n}_{n\text{-times}},$$

where $\boxtimes$ denotes free multiplicative convolution and there are n terms in the product.

We restrict our attention here to compactly-supported measures. As in the previous section define

$$\Sigma_\mu(z) = S_\mu\left(\frac{z}{1-z}\right).$$

The basic example of infinitely divisible distribution on the unit circle is given in the following Lemma:

Lemma 80 *For every $\lambda > 0$ and every $t \in \mathbb{R}^+$, the function*

$$\Sigma(z) = \exp\left(\frac{-\lambda z}{1-tz}\right)$$

is the Σ-transform of some probability measure μ compactly supported on $\mathbb{R}^+$.

The main theorem about multiplicatively infinitely-divisible measures is as follows.

Theorem 81 *A measure μ is infinitely-divisible if and only if $\Sigma_{\mu}(z) = \exp(u(z))$ where $u(z)$ has the following representation in the area $\mathbb{R}\backslash\mathbb{R}^{+}$*

$$u(z) = \alpha - \int_{-\infty}^{\infty} \frac{z}{1 - tz} d\sigma(t),$$

where α is real and σ is a finite measure.

For proof, see Bercovici and Voiculescu (1992).

9 Notes

For basic facts of operator algebra theory the reader can consult Bratteli and Robinson (1987). An introduction to quantum probability, which includes discussion of non-commutative probability spaces can be found in Parthasarathy (1992). The free probability came into the existence in Voiculescu (1983). Its first systematic description can be found in Voiculescu et al. (1992). An updated treatment of free probability theory that emphasizes its relation to random matrices is in Hiai and Petz (2000). Another textbook treatment of free probability that emphasize combinatorical aspect is in Nica and Speicher (2006)

The results about additive and multiplicative infinitely-divisible distributions are from Bercovici and Voiculescu (1993) and Bercovici and Voiculescu (1992), respectively. The domains of attractions and stable laws are studied in Bercovici et al. (1999).

Part II

Limit Theorems for Sums of Free Operators

10 CLT for Bounded Variables

In classical probability theory, one the most important theorem is the Central Limit Theorem (CLT). It has an analogue in non-commutative probability theory. First we formulate it for bounded random variables.

Theorem 82 *Let self-adjoint r.v.* X_i, $i = 1, 2, ...,$ *be free. Assume that* $E(X_i) = 0$, $\|X_i\| \leq L$ *and* $\lim_{n\to\infty} \left[E\left(X_1^2\right) + ... + E\left(X_n^2\right)\right]/n = 1$. *Then measures associated with r.v.* $n^{-1/2}\sum_{i=1}^{n} X_i$ *converge in distribution to an absolutely continuous measure with the following density:*

$$\phi(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\chi_{[-2,2]}(x).$$

This result was proven in Voiculescu (1983) and later generalized in Maassen (1992) to unbounded identically distributed variables that have a finite second moment. Another generalization can be found in Voiculescu (1998) and Pata (1996).

Proof of Theorem 82: We know that

$$K_{S_n}(z) = \sum_{i=1}^{n} K_{X_i}(z) - (n-1)z^{-1}$$

Consequently,

$$K_{S_n/\sqrt{n}}(z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} K_{X_i}\left(\frac{z}{\sqrt{n}}\right) - (n-1)z^{-1}.$$

Note that by Lemma 64, p.62, the power series for $K_{X_i}\left(\frac{z}{\sqrt{n}}\right)$ converges in $|z| \leq \sqrt{n}/(4L)$. Moreover, using Corollary 65 and the condition that $E(X_i) = 0$ we can write:

$$K_{S_n/\sqrt{n}}(z) = \frac{1}{z} + \left(\frac{1}{n}\sum_{i=1}^{n} E\left(X_i^2\right)\right)z + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} v_i\left(\frac{z}{\sqrt{n}}\right),$$

where for $|z| \leq \sqrt{n}/(2L)$ we can estimate $|v_i(z/\sqrt{n})| \leq 8Lz^2/n$. Therefore

$$\left| K_{S_n/\sqrt{n}}(z) - \frac{1}{z} - \left(\frac{1}{n} \sum_{i=1}^{n} E\left(X_i^2\right) \right) z \right| \leq \frac{1}{\sqrt{n}} 8Lz^2$$

for all z, such that $|z| \leq \sqrt{n}/(2L)$.

Therefore,

$$K_{S_n/\sqrt{n}}(z) - \frac{1}{z} \to z, \text{ as } n \to \infty,$$

and the convergence is uniform on the compact subsets of $\mathbb{C}$. This implies that

$$G_{S_n/\sqrt{n}}(z) \to G_\Phi(z), \text{ as } n \to \infty,$$

with the uniform convergence on the compact subsets. and this in term implies that the spectral distribution of $S_n/\sqrt{n}$ converges to the semicircle distribution. QED.

11 CLT, Proof by Lindeberg's Method

Freeness is a very strong condition imposed on operators and it is of interest to find out whether the Central Limit Theorem continues to hold if this condition is somewhat relaxed. This problems calls for a different proof of the non-commutative CLT because the existing proofs are based either on addivity of R-transform or on vanishing of mixed free cumulants, and both of these techniques are inextricably connected with the concept of freeness.

In this paper we give a proof of free CLT that avoids using either R-transforms or free cumulants. This allows us to give a generalization of the CLT to random variables that are not necessarily free but satisfy a weaker assumption. An example shows that this assumption is strictly weaker than assumption of freeness.

The proof that we use is a modification of the Lindeberg proof of the classical CLT (Lindeberg (1922)). The main difference is that we use polynomials instead of arbitrary functions from $C_c^3(\mathbb{R})$, and that more ingenuity is required to estimate the residual terms in the Taylor expansion formulas.

We will say that a sequence of zero-mean random variables $X_1, ..., X_n, ...$ satisfies *Condition A* if:

1. For every k, $E(X_k X_{i_1} ... X_{i_r}) = 0$ provided that $i_s \neq k$ for $s = 1, ..., r$.
2. For every $k \geq 2$, $E(X_k^2 X_{i_1} ... X_{i_r}) = E(X_k^2) E(X_{i_1} ... X_{i_r})$ provided that $i_s < k$ for $s = 1, ..., r$.

3. For every $k \geq 2$,

$$E\left(X_k X_{i_1} ... X_{i_p} X_k X_{i_{p+1}} ... X_{i_r}\right) = E\left(X_k^2\right) E\left(X_{i_1} ... X_{i_p}\right) E\left(X_{i_{p+1}} ... X_{i_r}\right)$$

provided that $i_s < k$ for $s = 1, ..., r$.

Intuitively, if we know how to calculate every moment of the sequence $X_1, ..., X_{k-1}$, then using Condition A we can also calculate the expectation of any product of random variables X_1, ..., X_k that involves no more than two occurences of variable X_k. Part 1 of Condition A is stronger than is needed for this calculation, since it involves variables with indices higher than k. However, we will need this additional strength in the proof of Lemma 93 below, which is essential for the proof of the main result.

Proposition 83 *Every sequence of free random variables X_1, ..., X_n, ... satisfies Condition A.*

This proposition can be checked by direct calculation that uses Proposition 8.

We will also need the following fact.

Proposition 84 *Let $X_1 ... X_l$ be zero-mean variables that satisfy Condition A(1), and let $Y_{l+1}, ..., Y_n$ be zero-mean variables which are free from each other and from the algebra generated by variables $X_1, ..., X_l$. Then $X_1, ..., X_l, Y_{l+1}, ... Y_n$ satisfies Condition A(1).*

Proof: Consider the moment $E\left(X_k A_{i_1} ... A_{i_s}\right)$, where A_{i_t} is either one of Y_j or one of X_i but with the exception that it can never be equal to X_k. Then we can use the fact that Y_j are free and write

$$E\left(X_k A_{i_1} ... A_{i_s}\right) = \sum_{\alpha} c_\alpha E\left(X_k X_{i_1(a)} ... X_{i_r(\alpha)}\right),$$

where none of $X_{i_t(\alpha)}$ equals X_k. Then, using the assumption that X_i satisfy Condition A(1), we conclude that $E\left(X_k A_{i_1} ... A_{i_s}\right) = 0$. Also $E\left(Y_k A_{i_1} ... A_{i_s}\right) = E\left(Y_k\right) E\left(A_{i_1} ... A_{i_s}\right) = 0$, provided that none of A_{i_t} equals Y_k. In sum, the sequence $X_1, ..., X_l, Y_{l+1}, ... Y_n$ satisfies Condition A(1). QED.

While freeness of random variables X_i is the same concept as freeness of the algebras that they generate, Condition A deals only with variables X_i only, and not with algebras that they generate. For example, it is conceivable that a sequence $\{X_i\}$

satisfy condition A but $\{X_i^2 - E(X_i^2)\}$ do not. In particular, this implies that Condition A requires checking a much smaller set of moment conditions than freeness. Below we will present an example of random variables which are not free but satisfy Condition A.

Recall that the standard semicircle law μ_{SC} is the probability distribution on $\mathbb{R}$ with the density $\pi^{-1}\sqrt{4-x^2}$ for $x \in [-2;2]$, and 0 otherwise. We are going to prove the following Theorem.

Theorem 85 *Suppose $\{\xi_i\}$ is a sequence of self-adjoint random variables that satisfies Condition A, and such that every ξ_i has the moments of all orders, $E\xi_i = 0$, $E\xi_i^2 = \sigma_i^2$, and that for every $k \geq 0$, the k^{th} absolute moments of ξ_i are uniformly bounded, i.e. $E|\xi_i|^k \leq \mu_k$ for all i. Suppose also that $s_N = (\sigma_1^2 + ... + \sigma_N^2)^{1/2}$ are such that $s_N/\sqrt{N} \to s > 0$ as $N \to \infty$. Then the spectral measure of $S_N = (\xi_1 + ... + \xi_N)/s_N$ converges in distribution to the semicircle law μ_{SC}.*

The contribution of this theorem is twofold. First, it shows that the semicircle central limit holds for a certain class of non-free variables. Second, it gives a proof of the free CLT which is different from the usual proof through R-transforms. It does not give improvement in conditions over a version of the free CLT which is formulated in Section 2.5 in Voiculescu (1998). (Note that the condition $\sum \varphi(a_i^2) \to 1$ in the statement of the theorem appears to be a typo and the condition is actually $n^{-1}\sum \varphi(a_i^2) \to 1$.)

11.1 Example

Let us present an example that suggest that Condition A is strictly weaker than freeness condition.

Let F be the free group with a countable number of generators f_k, $k = 1, ...$ Consider the set of relations $R = \{f_k f_{k-1} f_k f_{k-1} f_k f_{k-1} = e\}$, where $k \geq 2$, and define $G = F/\mathcal{R}$, that is, G is the group with generators f_k and relations generated by relations from R.

Here is a couple of useful consequences of these relationships:

1) $f_{k-1} f_k f_{k-1} f_k f_{k-1} f_k = e$.

(Indeed, $e = f_k^{-1}(f_k f_{k-1} f_k f_{k-1} f_k f_{k-1}) f_k = f_{k-1} f_k f_{k-1} f_k f_{k-1} f_k$.)

2) $f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} = e$ and $f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} = e$.

We are interested in the structure of the group G. For this purposes we will study the structure of $\mathcal{R}$, which is a subgroup of F generated by elements of R and their conjugates. We will represent elements of F by *words*, that is by sequences of generators. We will say that a word is *reduced* if does not have a subsequence of the form $f_k f_k^{-1}$ or $f_k^{-1} f_k$. It is *cyclically reduced* if it does not have the form of $f_k ... f_k^{-1}$ or $f_k^{-1} ... f_k$. We will call a number of elements in a reduced word w its length and denote it $|w|$. A set of relations R is *symmetrized* if for every word $r \in R$, the set R also contains its inverse r^{-1} and all cyclically reduced conjugates of both r and r^{-1}.

For our particular example, a symmetrized set of relations is given by the following list:

$$R = \left\{ \begin{array}{cc} f_k f_{k-1} f_k f_{k-1} f_k f_{k-1}, & f_{k-1} f_k f_{k-1} f_k f_{k-1} f_k, \\ f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1}, & f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} \end{array} \right\},$$

where k are all integers ≥ 2.

A word b is called a *piece* (relative to a symmetrized set R) if there exist two elements of R, r_1 and r_2, such that $r_1 = bc_1$ and $r_2 = bc_2$. In our case, each f_k and f_k^{-1} with index $k \geq 2$ is a piece because f_k is the initial part of relations $f_k f_{k-1} f_k f_{k-1} f_k f_{k-1}$ and $f_k f_{k+1} f_k f_{k+1} f_k f_{k+1}$, and f_k^{-1} is the initial part of the relations $f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1}$ and $f_k^{-1} f_{k+1}^{-1} f_k^{-1} f_{k+1}^{-1} f_k^{-1} f_{k+1}^{-1}$. There is no other pieces.

Now we introduce a condition of *small cancellation* for a symmetrized set R:

Condition 86 ($C'(\lambda)$) *If $r \in R$ and $r = bc$ where b is a piece, then $|b| < \lambda |r|$.*

Essentially, the condition says that if two relations are multiplied together, then a possible cancellation must be relatively small. Note that if R satisfies $C'(\lambda)$ then it satisfies $C'(\mu)$ for all $\mu \geq \lambda$.

In our example R satisfies $C'(1/5)$.

Another important condition is the triangle condition.

Condition 87 (T) *Let r_1, r_2, r_3 be three arbitrary elements of R such that $r_2 \neq r_1^{-1}$ and $r_3 \neq r_2^{-1}$ Then at least one of the products $r_1 r_2$, $r_2 r_3$, or $r_3 r_1$ is reduced without cancellation.*

In our example, Condition (T) is satisfied.

If s is a word in F, then $s > \lambda R$ means that there exists a word $r \in R$ such that $r = st$ and $|s| > \lambda |r|$. An important result from small cancellation theory that we will use later is the following theorem:

Theorem 88 (Greendlinger's Lemma) *Let R satisfy $C'(1/4)$ and T. Let w be a non-trivial, cyclically reduced word with $w \in \mathcal{R}$. Then either*
(1) $w \in R$,
or some cyclycally reduced conjugate w^* of w contains one of the following:
(2) two disjoint subwords, each $> \frac{3}{4}R$, or
(4) four disjoint subwords, each $> \frac{1}{2}R$.

This theorem is Theorem 4.6 on p. 251 in Lyndon and Schupp (1977).

Since in our example R satisfies both $C'(1/4)$ and T, we can infer that the conclusion of the theorem must hold in our case. For example, (2) means that we can find two disjoint subwords of w, s_1 and s_2, and two elements of R, r_1 and r_2, such that $r_i = s_i t_i$ and $|s_i| > (3/4)|r_i| = 9/2$. In particular, we can conclude that in this case $|w| \geq 10$. Similarly, in case (4), $|w| \geq 16$. One immediate application is that G does not collapse to the trivial group. Indeed, all f_i are not zero.

Let $L^2(G)$ be the functions of G that are square-summable with respect to the counting measure. G acts on $L^2(G)$ by left translations:

$$(L_g x)(h) = x(gh).$$

Let $\mathcal{A}$ be the group algebra of G. The action of G on $L^2(G)$ can be extended to the action of $\mathcal{A}$ on $L^2(G)$. Define the expectation on this group algebra by the following rule:

$$E(h) = \langle \delta_e, L_h \delta_e \rangle,$$

where $\langle , \rangle$ denotes the scalar product in $L^2(G)$. Alternatively, the expectation can be written as follows:

$$E(h) = a_e,$$

where $h = \sum_{g\in G} a_g g$ is a representation of a group algebra element h as a linear combination of elements $g \in G$. The expectation is clearly positive and finite by definition. It is also tracial because $g_1 g_2 = e$ if and only if $g_2 g_1 = e$.

If $L_h = \sum_{g\in G} a_g L_g$ is a linear operator corresponding to the element of group algebra $h = \sum_{g\in G} a_g g$, then its adjoint is $(L_h)^* = \sum_{g\in G} \overline{a_g} L_{g^{-1}}$, which correspond to the element $h^* = \sum_{g\in G} \overline{a}_g g^{-1}$.

Consider elements $X_i = f_i + f_i^{-1}$. They are self-adjoint and $E(X_i) = 0$. Also we can compute $E(X_i^2) = 2$. Indeed it is enough to note that $f_i^2 \neq e$, and this holds because insertion or deletion of an element from R changes the degree of f_i by a multiple of 3. Therefore, every word equal to zero must have the degree of every f_i equal to 0 modulo 3.

Proposition 89 *The sequence of variables $\{X_i\}$ is not free but it satisfies Condition A.*

Proof: The variables X_k are not free. Consider $X_2X_1X_2X_1X_2X_1$. Its expectation is 2, because $f_2f_1f_2f_1f_2f_1 = e$ and $f_2^{-1}f_1^{-1}f_2^{-1}f_1^{-1}f_2^{-1}f_1^{-1} = e$, and all other terms in the expansion of $X_2X_1X_2X_1X_2X_1$ are different from e. Indeed, the only terms that are not of the form above but still have the degree of all f_i equal to zero modulo 3 are $f_2f_1^{-1}f_2f_1^{-1}f_2f_1^{-1}$ and $f_2^{-1}f_1f_2^{-1}f_1f_2^{-1}f_1$, but they do not equal zero by application of Greendlinger's lemma. Therefore, $E\left(X_2X_1X_2X_1X_2X_1\right) = 2$. This contradicts the definition of freeness of variables X_2 and X_1.

Let us check Condition A. For A1, consider $E\left(f_kf_{i_1}...f_{i_n}\right)$, where $k \neq i_s$, and $i_s \neq i_{s+1}$ for every s. It is enough to check that this expectation equals 0. Indeed $f_kf_{i_1}...f_{i_n} \neq e$. This can be seen from the fact that an insertion or deletion of a relation can change the degree of f_k only by 3. Therefore $E\left(f_kf_{i_1}...f_{i_n}\right) = 0$. A similar argument works for $E\left(f_k^{-1}f_{i_1}...f_{i_n}\right) = 0$ and more generally for the expectation of every element of the form $f_k^{\varepsilon}f_{i_1}^{n_1}...f_{i_n}^{n_2}$, where $\varepsilon = \pm 1$ and n_s are integer.

Similarly, we can prove that $E\left(f_k^{\pm 2}f_{i_1}^{n_1}...f_{i_n}^{n_2}\right) = 0$ and this suffice to prove A2.

For A3 we have to consider elements of the form $f_k^{\varepsilon_1}f_{i_1}...f_{i_p}f_k^{\varepsilon_2}f_{i_{p+1}}...f_{i_q}$ Assume that neither $f_{i_1}...f_{i_p}$ nor $f_{i_{p+1}}...f_{i_q}$ can be reduced to e. Otherwise we can use property A2. The claim is that $E\left(f_k^{\varepsilon_1}f_{i_1}...f_{i_p}f_k^{\varepsilon_2}f_{i_{p+1}}...f_{i_q}\right) = 0$. This is clear when ε_1 and ε_2 have the same sign because of the fact that relation change the degree of the element f_k by a multiple of 3. A more difficult case is when $\varepsilon_1 = 1$ and $\varepsilon_2 = -1$. (The case with opposite signs is similar.) However, in this case we can conclude that $f_kf_{i_1}...f_{i_p}f_k^{-1}f_{i_{p+1}}...f_{i_q} \neq e$ by application of Greendlinger's lemma. Indeed, the only subwords that this word can contain, which would also be subwords of an element of R, are subwords of length 1 and 2. But these subwords fail to satisfy the requirement of either (2) or (4) in Greendlinger's lemma. Therefore, we can conclude that $f_kf_{i_1}...f_{i_p}f_k^{-1}f_{i_{p+1}}...f_{i_q} \neq e$ and A3 is also satisfied. Thus Condition A is satisfied by random variables $X_1, ..., X_k, ...$ in algebra $\mathcal{A}$, although these variables are not free. QED.

11.2 Proof of the main result

Outline of Proof: Our proof of the free CLT proceeds along the familiar lines of the Lindeberg method. We take a family of functions and evaluate an arbitrary function from this family on the sum $S_N = X_1 + ... + X_N$. The goal is to compare

$Ef\left(S_N\right)$ with $Ef\left(\widetilde{S}_N\right)$, where f are functions from a sufficiently large family, $\widetilde{S}_N = Y_1 + ... + Y_N$ and Y_i are independent semicircle variables chosen in such a way that $\operatorname{Var}\left(S_N\right) = \operatorname{Var}\left(\widetilde{S}_N\right)$. To estimate $\left|Ef\left(S_N\right) - Ef\left(\widetilde{S}_N\right)\right|$, we substitute the elements in S_N with the semicircle free random variables, one by one, and estimate the corresponding change in the expected value of $f\left(S_N\right)$. After that, we show that the total change, as all elements in the sum are substituted with semicircle random variables, is asymptotically small as $N \to \infty$. Finally, tightness of the selected family of functions allows us to conclude that the distribution of S_N must converge to the semicircle law as $N \to \infty$.

The usual choice of functions f are functions from $C_c^3\left(\mathbb{R}\right)$, that is, functions with continuous third derivative and compact support. In non-commutative setting this family of functions is not appropriate because the usual Taylor series formula is difficult to apply. Intuitively, it is difficult to develop $f\left(X + h\right)$ in power series of h if variables X and h do not commute. Since the Taylor formula is crucial for estimating the change in $Ef\left(S_N\right)$, we will still use it but restrict the family of functions to polynomials.

To show that the family of polynomials is sufficiently rich for our purposes, we use the following Proposition:

Proposition 90 *Suppose there is a unique d.f. F with the moments $\left\{m^{(r)}, r \geq 1\right\}$. Suppose that $\{F_N\}$ is a sequence of d.f., each of which has all its moments finite:*

$$m_N^{(r)} = \int_{-\infty}^{\infty} x^r dF_N.$$

Finally, suppose that for every $r \geq 1$:

$$\lim_{n\to\infty} m_N^{(r)} = m^{(r)}.$$

Then $F_N \to F$ vaguely.

See Theorem 4.5.5. in Chung (2001) for a proof.

Since the semicircle distribution is bounded and therefore is determined by its moments (see Corollary to Theorem II.12.7 Shiryaev (1995)), we only need to show that the moments of S_n converge to moments of the semicircle distribution.

Proof of Theorem 85: Define η_i as a sequence of random variables that are freely independent among themselves and also from all ξ_i. Suppose also that η_i have semicircle distributions with $E\eta_i = 0$ and $E\eta_i^2 = \sigma_i^2$. We are going to use as known the fact that the sum of free semicircle random variables is semicircle, and therefore, the spectral distribution of $(\eta_1 + ... + \eta_N) / \left(s\sqrt{N}\right)$ converges in distribution to the semicircle law μ_{SC} with zero expectation and unit variance. Let us define $X_i = \xi_i / s_N$ and $Y_i = \eta_i / s_N$. We will proceed by proving that moments of $X_1 + ... + X_N$ converge to moments of $Y_1 + ... + Y_N$ and applying Proposition 90. Let

$$\Delta f = Ef(X_1 + ... + X_N) - Ef(Y_1 + ... + Y_N),$$

where $f(x) = x^m$. We want to show that this difference approaches zero as N grows.

By assumption, $EY_i = EX_i = 0$ and $EY_i^2 = EX_i^2 = \sigma_i^2 / s_N^2$.

The first step is to write the difference Δf as follows:

$$\begin{aligned} \Delta f &= [Ef(X_1 + ... + X_{N-1} + X_N) - Ef(X_1 + ... + X_{N-1} + Y_N)] \\ &\quad + [Ef(X_1 + ... + X_{N-1} + Y_N) - Ef(X_1 + ... + Y_{N-1} + Y_N)] \\ &\quad + [Ef(X_1 + Y_2 ... + Y_{N-1} + Y_N) - Ef(Y_1 + Y_2 ... + Y_{N-1} + Y_N)]. \end{aligned}$$

We intend to estimate every difference in this sum. Let

$$Z_k = X_1 + ... + X_{k-1} + Y_{k+1} + ... + Y_N. \tag{36}$$

We are interested in

$$Ef(Z_k + X_k) - Ef(Z_k + Y_k).$$

We are goint to apply the Taylor expansion formula but first we define directional derivatives. Let $f'_{X_k}(Z_k)$ be the *derivative of f at Z_k in the direction X_k*, defined as follows:

$$f'_{X_k}(Z_k) = \lim_{t \downarrow 0} \frac{f(Z_k + tX_k) - f(Z_k)}{t}.$$

The higher order directional derivatives can be defined recursively. For example,

$$f''_{X_k}(Z_k) =: \left(f'_{X_k}\right)'_{X_k}(Z_k) = \lim_{t \downarrow 0} \frac{f'_{X_k}(Z_k + tX_k) - f'_{X_k}(Z_k)}{t}.$$

For polynomials this definition is equivalent to the following definition:

$$f''_{X_k}(Z_k) = 2 \lim_{t \downarrow 0} \frac{f(Z_k + tX_k) - f(Z_k) - tf'_{X_k}(Z_k)}{t^2}.$$

Example 91 *Operator directional derivatives of $f(x) = x^4$*

Let us compute $f'_X(Z)$ and $f''_X(Z)$ for $f(x) = x^4$. Using definitions we get

$$f'_X(Z) = Z^3X + Z^2XZ + ZXZ^2 + XZ^3$$

and

$$f''_X(Z) = 2\left(Z^2X^2 + ZXZX + XZ^2X + ZX^2Z + XZXZ + X^2Z^2\right). \quad (37)$$

The derivatives of f at $Z_k + \tau X_k$ in the direction X_k are defined similarly, for example:

$$\begin{aligned} & f'''_{X_k}(Z_k + \tau X_k) \\ = \ & 6 \lim_{t \downarrow 0} \frac{f(Z_k + (\tau + t) X_k) - f(Z_k + \tau X_k) - t f'_{X_k}(Z_k + \tau X_k) - \frac{1}{2} t^2 f''_{X_k}(Z_k + \tau X_k)}{t^3}. \end{aligned}$$

Next, let us write the Taylor formula for $f(Z_k + X_k)$:

$$f(Z_k + X_k) = f(Z_k) + f'_{X_k}(Z_k) + \frac{1}{2} f''_{X_k}(Z_k) + \frac{1}{2}\int_0^1 (1-\tau)^2 f'''_{X_k}(Z_k + \tau X_k)\, d\tau. \quad (38)$$

Formula (38) can be obtained by integration by parts from the expression

$$f(Z_k + X_k) - f(Z_k) = \int_0^1 f'_{X_k}(Z_k + \tau X_k)\, d\tau.$$

For polynomials it is easy to write the explicit expressions for $f^{(r)}_{X_k}(Z_k)$ or $f^{(r)}_{X_k}(Z_k + \tau X_k)$ although they can be quite cumbersome for polynomials of high degree. Very schematically, for a function $f(x) = x^m$, we can write

$$f'_{X_k}(Z_k) = X_k Z_k^{m-1} + Z_k X_k Z_k^{m-2} + \ldots + Z_k^{m-1} X_k, \quad (39)$$

and

$$f''_{X_k}(Z_k) = 2\left(X_k^2 Z_k^{m-2} + X_k Z_k X_k Z_k^{m-3} + \ldots + Z_k^{m-2} X_k^2\right), \quad (40)$$

Similar formulas hold for $f'_{Y_k}(Z_k)$ and $f''_{Y_k}(Z_k)$ with the change that Y_k should be put instead of X_k.

Using the assumptions that sequence $\{X_k\}$ satisfies Condition A and that variables Y_k are free, we can conclude that $Ef'_{Y_k}(Z_k) = Ef'_{X_k}(Z_k) = 0$ and that $Ef''_{Y_k}(Z_k) = Ef''_{X_k}(Z_k)$. Indeed, consider, for example, (40). We can use expression (36) for Z_k and the free independence of Y_i to expand (40) as

$$Ef''_{X_k}(Z_k) = \sum_\alpha c_\alpha P_\alpha \left(E\left(X_k \overline{X_1} X_k \overline{X_2}\right), E\left(X_k \overline{X_3} X_k \overline{X_4}\right), \ldots\right), \quad (41)$$

where $\overline{X_i}$ denotes certain monomials in variables $X_1, ..., X_{k-1}$ (i.e., $\overline{X_i} = X_{i_1}...X_{i_p}$ with $i_k \in \{1, ..., k-1\}$), and where α indexes certain polynomials P_α. In other words, using the free independence of Y_i and X_i we expand the expectations of polynomial $f''_{X_k}(Z_k)$ as a sum over polynomials in joint moments of variables X_j and Y_i where $j = 1, ..., k$ and $i = k+1, ..., N$. By freeness, we can achieve that the moments in this expression are either joint moments of variables X_j or joint moments of variables Y_i but never involve both X_j and Y_i. Moreover, we can explictly calculate the moments of Y_i (i.e., expectations of the products of Y_i) because their are mutually free. The resulting expansion is (41).

Let us try to make this process clearer by an example. Suppose that $f(x) = x^4$, $N = 4$, $k = 2$ and $Z_k = Z_2 = X_1 + Y_3 + Y_4$. We aim to compute $Ef''_{X_2}(Z_2)$. Using formula (37), we write:

$$\begin{aligned} Ef''_{X_2}(Z_2) &= 2E\left(Z_2^2 X_2^2 + ...\right) \\ &= 2E\left((X_1 + Y_3 + Y_4)^2 X_2^2 + ...\right) \\ &= 2\{E\left(X_1^2 X_2^2\right) + E\left(X_1 Y_3 X_2^2\right) + E\left(X_1 Y_4 X_2^2\right) \\ &\quad + E\left(Y_3 X_1 X_2^2\right) + E\left(Y_3^2 X_2^2\right) + E\left(Y_3 Y_4 X_2^2\right) \\ &\quad + E\left(Y_4 X_1 X_2^2\right) + E\left(Y_4 Y_3 X_2^2\right) + E\left(Y_4^2 X_2^2\right) + ...\}. \end{aligned}$$

Then, using freeness of Y_3 and Y_4 and the facts that $E(Y_i) = 0$ and $E(Y_i^2) = \sigma_i^2$, we continue as follows:

$$Ef''_{X_2}(Z_2) = 2\{E\left(X_1^2 X_2^2\right) + \sigma_3^2 E\left(X_2^2\right) + \sigma_4^2 E\left(X_2^2\right) + ...\},$$

which is the expression we wanted to obtain.

It is important to note that the coefficients c_α do not depend on variables X_j but only on Y_j, $j > k$, and on the locations, which Y_j take in the expansion of $f''_{X_k}(Z_k)$. Therefore, we can substitute Y_k for X_k and develop a similar formula for $Ef''_{Y_k}(Z_k)$:

$$Ef''_{Y_k}(Z_k) = \sum_\alpha c_\alpha P_\alpha\left(E\left(Y_k \overline{X_1} Y_k \overline{X_2}\right), E\left(Y_k \overline{X_3} Y_k \overline{X_4}\right), ...\right). \qquad (42)$$

In the example above, we will have

$$Ef''_{Y_2}(Z_2) = 2\{E\left(X_1^2 Y_2^2\right) + \sigma_3^2 E\left(Y_2^2\right) + \sigma_4^2 E\left(Y_2^2\right) + ...\}.$$

Formula (42) is exactly the same as formula (41) except that all X_k is substituted with Y_k. Finally, using Condition A we obtain that for every i:

$$\begin{aligned} E\left(Y_k \overline{X_i} Y_k \overline{X_{i+1}}\right) &= E\left(Y_k^2\right) E\left(\overline{X_i}\right) E\left(\overline{X_{i+1}}\right) \\ &= E\left(X_k^2\right) E\left(\overline{X_i}\right) E\left(\overline{X_{i+1}}\right) \\ &= E\left(X_k \overline{X_i} X_k \overline{X_{i+1}}\right), \end{aligned}$$

and therefore $Ef''_{Y_k}(Z_k) = Ef''_{X_k}(Z_k)$.

Consequently,

$$\begin{aligned} Ef(Z_k + X_k) - Ef(Z_k + Y_k) &= \frac{1}{2}\int_0^1 (1-\tau)^2 Ef'''_{X_k}(Z_k + \tau X_k)\, d\tau \\ &\quad -\frac{1}{2}\int_0^1 (1-\tau)^2 Ef'''_{Y_k}(Z_k + \tau Y_k)\, d\tau. \end{aligned}$$

Next, note that if f is a polynomial, then $f'''_{X_k}(Z_k + \tau X_k)$ is the sum of a finite number of terms, which are products of $Z_k + \tau X_k$ and X_k. The number of terms in this expansion is bounded by C_1, which depends only on the degree m of the polynomial f.

A typical term in the expansion looks like

$$E\left(Z_k + \tau X_k\right)^{m-7} X_k^3 \left(Z_k + \tau X_k\right)^3 X_k.$$

In addition, if we expand the powers of $Z_k + \tau X_k$, we will get another expansion that has the number of terms bounded by C_2, where C_2 depends only on m. A typical element of this new expansion is

$$E\left(Z_k^{m-7} X_k^3 Z_k^2 X_k^2\right).$$

Every term in this expansion has a total degree of X_k not less than 3, and, correspondingly, a total degree of Z_k not more than $m-3$. Our task is to show that as $n \to \infty$, these terms approach 0.

We will use the following lemma to estimate each of the summands in the expansion of $f'''_{X_k}(Z_k + \tau X_k)$.

Lemma 92 *Let X and Y be self-adjoint. Then*

$$\begin{aligned} & \left|E\left(X^{m_1} Y^{n_1} \ldots X^{m_r} Y^{n_r}\right)\right| \\ \leq\ & \left[E\left(X^{2^r m_1}\right)\right]^{2^{-r}} \left[E\left(Y^{2^r n_1}\right)\right]^{2^{-r}} \ldots \left[E\left(X^{2^r m_r}\right)\right]^{2^{-r}} \left[E\left(Y^{2^r n_r}\right)\right]^{2^{-r}}. \end{aligned}$$

Proof: For $r = 1$, this is the usual Cauchy-Schwartz inequality for traces:

$$|E(X^{m_1}Y^{n_1})|^2 \leq E\left(X^{2m_1}\right)E\left(Y^{2n_1}\right).$$

See, for example, Proposition I.9.5 on p. 37 in Takesaki (1979).

Next, we proceed by induction. We have two slightly different cases to consider. Assume first that r is even, $r = 2s$. Then, by the Cauchy-Schwartz inequality, we have:

$$\begin{aligned}
& |E(X^{m_1}Y^{n_1}...X^{m_r}Y^{n_r})|^2 \\
\leq\ & E(X^{m_1}Y^{n_1}...X^{m_s}Y^{n_s}Y^{n_s}X^{m_s}...Y^{n_1}X^{m_1}) \\
& \times E(Y^{n_r}X^{m_r}...Y^{n_{s+1}}X^{m_{s+1}}X^{m_{s+1}}Y^{n_{s+1}}...X^{m_r}Y^{n_r}) \\
=\ & E\left(X^{2m_1}Y^{n_1}...X^{m_s}Y^{2n_s}X^{m_s}...Y^{n_1}\right) \\
& \times E\left(Y^{2n_r}X^{m_r}...Y^{n_{s+1}}X^{2m_{s+1}}Y^{n_{s+1}}...X^{m_r}\right).
\end{aligned}$$

Applying the inductive hypothesis, we obtain:

$$\begin{aligned}
& |E(X^{m_1}Y^{n_1}...X^{m_r}Y^{n_r})|^2 \\
\leq\ & \left[E\left(X^{2^r m_1}\right)\right]^{2^{-r+1}}\left[E\left(Y^{2^r n_s}\right)\right]^{2^{-r+1}} \\
& \times\left[E\left(Y^{2^{r-1}n_1}\right)\right]^{2^{-r+2}}...\left[E\left(X^{2^{r-1}m_s}\right)\right]^{2^{-r+2}} \\
& \times\left[E\left(X^{2^r m_{s+1}}\right)\right]^{2^{-r+1}}\left[E\left(Y^{2^r n_r}\right)\right]^{2^{-r+1}} \\
& \times\left[E\left(Y^{2^{r-1}n_{s+1}}\right)\right]^{2^{-r+2}}...\left[E\left(X^{2^{r-1}m_r}\right)\right]^{2^{-r+2}}.
\end{aligned}$$

We recall that by the Markov inequality, $\left[E\left(Y^{2^{r-1}n_1}\right)\right]^{2^{-r+2}} \leq \left[E\left(Y^{2^r n_1}\right)\right]^{2^{-r+1}}$ and we get the desired inequality:

$$\begin{aligned}
& |E(X^{m_1}Y^{n_1}...X^{m_r}Y^{n_r})| \\
\leq\ & \left[E\left(X^{2^r m_1}\right)\right]^{2^{-r}}\left[E\left(Y^{2^r n_1}\right)\right]^{2^{-r}}...\left[E\left(X^{2^r m_r}\right)\right]^{2^{-r}}\left[E\left(Y^{2^r n_r}\right)\right]^{2^{-r}}.
\end{aligned}$$

Now let r be even, $r = 2s + 1$. Then

$$\begin{aligned}
& |E(X^{m_1}Y^{n_1}...X^{m_r}Y^{n_r})|^2 \\
\leq\ & E(X^{m_1}Y^{n_1}...Y^{n_s}X^{m_{s+1}}X^{m_{s+1}}Y^{n_s}...Y^{n_1}X^{m_1}) \\
& \times E(Y^{n_r}X^{m_r}...X^{m_{s+2}}Y^{n_{s+1}}Y^{n_{s+1}}X^{m_{s+2}}...X^{m_r}Y^{n_r}) \\
=\ & E\left(X^{2m_1}Y^{n_1}...Y^{n_s}X^{2m_{s+1}}Y^{n_s}...Y^{n_1}\right)E\left(Y^{2n_r}X^{m_r}...X^{m_{s+2}}Y^{2n_{s+1}}X^{m_{s+1}}...X^{m_r}\right).
\end{aligned}$$

After that we can use the inductive hypothesis and the Markov inequality and obtain that

$$\begin{aligned}
& \left|E\left(X^{m_1}Y^{n_1}...X^{m_r}Y^{n_r}\right)\right| \\
\leq\ & \left[E\left(X^{2^r m_1}\right)\right]^{2^{-r}} \left[E\left(Y^{2^r n_1}\right)\right]^{2^{-r}} ... \left[E\left(X^{2^r m_r}\right)\right]^{2^{-r}} \left[E\left(Y^{2^r n_r}\right)\right]^{2^{-r}}.
\end{aligned}$$

QED.

We apply Lemma 92 to estimate each of the summands in the expansion of $f'''_{X_k}(Z_k + \tau X_k)$. Consider a summand $E\left(Z_k^{m_1}X_k^{n_1}...Z_k^{m_r}X_k^{n_r}\right)$. Then by Lemma 92, we have

$$\begin{aligned}
& \left|E\left(Z_k^{m_1}X_k^{n_1}...Z_k^{m_r}X_k^{n_r}\right)\right| \\
\leq\ & \left[E\left(Z_k^{2^r m_1}\right)\right]^{2^{-r}} \left[E\left(X_k^{2^r n_1}\right)\right]^{2^{-r}} ... \left[E\left(Z_k^{2^r m_r}\right)\right]^{2^{-r}} \left[E\left(X_k^{2^r n_r}\right)\right]^{2^{-r}}.
\end{aligned} \tag{43}$$

Lemma 93 *Let $Z = (v_1 + ... + v_N)/N^{1/2}$, where v_i are self-adjoint and satisfy condition A and for each k, the k^{th} absolute moments of v_i are uniformly bounded, i.e., $E|v_i|^k \leq \mu_k$ for every i. Then, for every integer $r \geq 0$*

$$E(|Z|^r) = O(1) \text{ as } N \to \infty.$$

Proof: We will first treat the case of even r. In this case, $E(|Z|^r) = E(Z^r)$. Consider the expansion of $(v_1 + ... + v_N)^r$. Let us refer to the indices $1, ..., N$ as colors of the corresponding v. If a term in the expansion includes more than $r/2$ distinct colors, then one of the colors must be used by this term only once. Therefore, by the first part of condition A the expectation of such a term is 0.

Let us estimate a number of terms in the expansion that include no more than $r/2$ distinct colors. Consider a fixed combination of $\leq r/2$ colors. The number of terms that use colors only from this combination is $\leq (r/2)^r$. Indeed, consider the product

$$(v_1 + ... + v_N)(v_1 + ... + v_N)...(v_1 + ... + v_N)$$

with r product terms. We can choose an element from the first product term in $r/2$ possible ways, an element from the second product term in $r/2$ possible ways, etc. Therefore, the number of all possible choices is $(r/2)^r$. On the other hand, the number of possible different combinations of $k \leq r/2$ colors is

$$\frac{N!}{(N-k)!k!} \leq N^{r/2}.$$

Therefore, the total number of terms that use no more than $r/2$ colors is bounded from above by

$$(r/2)^r N^{r/2}.$$

Now let us estimate the expectation of an individual term in the expansion. In other words we want to estimate $E\left(v_{i_1}^{k_1} ... v_{i_s}^{k_s}\right)$, where $k_t \geq 1$, $k_1 + ... + k_s = r$, and $i_t \neq i_{t+1}$. First, note that

$$\left|E\left(v_{i_1}^{k_1} ... v_{i_s}^{k_s}\right)\right| \leq E\left(\left|v_{i_1}^{k_1} ... v_{i_s}^{k_s}\right|\right).$$

Indeed, using Cauchy-Schwartz inequality, for any operator X we can write

$$\begin{aligned} |E(X)|^2 &= E\left(U|X|^{1/2}|X|^{1/2}\right) \leq E\left(|X|^{1/2} U^* U |X|^{1/2}\right) E\left(|X|^{1/2}|X|^{1/2}\right) \\ &= E(|X| P) E(|X|), \end{aligned}$$

where U is a partial isometry and $P = U^*U$ is a projection. Note that from positivity of the expectation it follows that $E(|X| P) \leq E(|X|)$. Therefore, we can conclude that $|E(X)| \leq E(|X|)$.

Next, we use the Hölder inequality for traces of non-commutative operators (see Fack (1982), especially Corollary 4.4(iii) on page 324, for the case of the trace in a von Neumann algebra and Section III.7.2 in Gohberg and Krein (1969) for the case of compact operators and the usual operator trace). Note that

$$\underbrace{\frac{1}{s} + ... + \frac{1}{s}}_{s\text{-times}} = 1,$$

therefore, the Hölder inequality gives

$$E\left(\left|v_{i_1}^{k_1} ... v_{i_s}^{k_s}\right|\right) \leq \left[E\left(\left|v_{i_1}\right|^{k_1 s}\right) ... E\left(\left|v_{i_s}\right|^{k_s s}\right)\right]^{1/s}.$$

Using this result and uniform boundedness of the moments (from assumption of the lemma), we get:

$$\log\left|E\left(v_{i_1}^{k_1} ... v_{i_s}^{k_s}\right)\right| \leq \frac{1}{s} \sum_{i=1}^{s} \log \mu_{k_i s}.$$

Without loss of generality we can assume that bounds μ_k are increasing in k. Using the fact that $s \leq r$ and $k_i \leq r$, we obtain the bound:

$$\log\left|E\left(v_{i_1}^{k_1} ... v_{i_s}^{k_s}\right)\right| \leq \log \mu_{r^2},$$

or

$$\left|E\left(v_{i_1}^{k_1}...v_{i_s}^{k_s}\right)\right| \leq \mu_{r^2}.$$

Therefore,

$$E\left(v_1 + ... + v_N\right)^r \leq (r/2)^r \mu_{r^2} N^{r/2},$$

and

$$E\left(Z^r\right) \leq (r/2)^r \mu_{r^2}. \tag{44}$$

Now consider the case of odd r. In this case, we use the Lyapunov inequality to write:

$$\begin{aligned} E\left|Z\right|^r &\leq \left(E\left|Z\right|^{r+1}\right)^{\frac{r}{r+1}} \\ &\leq \left(\left(\frac{r+1}{2}\right)^{r+1} \mu_{(r+1)^2}\right)^{\frac{r}{r+1}} \\ &= \left(\frac{r+1}{2}\right)^r \left(\mu_{(r+1)^2}\right)^{\frac{r}{r+1}}. \end{aligned} \tag{45}$$

The important point is that bounds in (44) and (45) do not depend on n. QED.

By definition $Z_k = \left(\xi_1 + ... + \xi_{k-1} + \eta_{k+1} + ... + \eta_N\right)/s_N$ and by assumption ξ_i and η_i are uniformly bounded, and $s_N \sim \sqrt{N}$. Moreover, $\xi_1, ..., \xi_{k-1}$ satisfy Condition A by assumption, and $\eta_{k+1}, ..., \eta_N$ are free from each other and from $\xi_1, ..., \xi_{k-1}$. Therefore, $\xi_1, ..., \xi_{k-1}, \eta_{k+1}, ..., \eta_N$ satisfy condition A. Consequently, we can apply Lemma 93 to Z_k and conclude that $E\left|Z_k\right|^r$ is bounded by a constant that depends only on r but does not depend on N.

Using this fact, we can continue the estimate in (43) and write:

$$\begin{aligned} &\left|E\left(Z_k^{m_1} X_k^{n_1} ... Z_k^{m_r} X_k^{n_r}\right)\right| \\ \leq\ & C_4 \left[E\left(X_k^{2^r n_1}\right)\right]^{2^{-r}} ... \left[E\left(X_k^{2^r n_r}\right)\right]^{2^{-r}}, \end{aligned} \tag{46}$$

where the constant C_4 depends only on m.

Next we note that

$$\left[E\left(X_k^{2^r n_1}\right)\right]^{2^{-r}} \leq C\left(\frac{\mu_{2^r n_1}}{N^{2^{r-1} n_1}}\right)^{2^{-r}} = C\frac{\left(\mu_{2^r n_1}\right)^{2^{-r}}}{N^{n_1/2}}.$$

Next note that $n_1 + ... + n_r \geq 3$, therefore we can write

$$\left[E\left(X_k^{2^r n_1}\right)\right]^{2^{-r}} ... \left[E\left(X_k^{2^r n_r}\right)\right]^{2^{-r}} \leq C' N^{-3/2}.$$

In sum, we obtain the following Lemma:

Lemma 94

$$\left|Ef'''_{X_k}(Z_k+\tau X_k)\right| \leq C_5 N^{-3/2},$$

where C_5 depends only on the degree of polynomial f and the sequence of constants μ_k.

A similar result holds for $\left|Ef'''_{X_k}(Z_k+\tau Y_k)\right|$ and we can conclude that

$$|Ef(Z_k+X_k)-Ef(Z_k+Y_k)| \leq C_6 N^{-3/2}.$$

After we sum these inequalities over all $k=1,...,N$ we get

$$|Ef(X_1+...+X_N)-Ef(Y_1+...+Y_N)| \leq C_7 N^{-1/2}.$$

Clearly this approaches 0 as N grows. Applying Proposition 90, we conclude that the measure of $X_1+...+X_N$ converges to the measure of $Y_1+...+Y_N$ in distribution.

This finishes proof of the main theorem.

The key points of this proof are as follows: 1) We can substitute each random variable X_i in the sum S_N with a free random variable Y_i so that the first and the second derivatives of any polynomial with S_N in the argument remain unchanged. This depends on Condition A being satisfied by X_i. 2)We can estimate a change in the third derivative as we substitute Y_i for X_i by using the first part of Condition A and several matrix inequalities, valid for any collection of operators. Here Condition A is used only in the proof that the k-th moment of $(\xi_1+...+\xi_N)/N^{1/2}$ is bounded as $N\to\infty$.

It is interesting whether the ideas of this proof can be generalized to the case of the multivariate CLT.

12 CLT, Rate of Convergence

In this section we investigate the speed of convergence in free CLT and establish an inequality similar to the classical Berry-Esseen inequality.

Recall that without reference to operator theory, the free convolution can be defined as follows. Suppose μ_1 and μ_2 are two probability measures compactly supported on the real line. Define the *Cauchy transform* of μ_i as

$$G_i(z)=\int_{-\infty}^{\infty}\frac{d\mu_i(t)}{z-t}.$$

Each of $G_i(z)$ is well-defined and univalent for large enough z and we can define its functional inverse, which is well-defined in a neighborhood of 0. Let us call this inverse the K*-function* of μ_i and denote it as $K_i(z)$:

$$K_i(G_i(z)) = G_i(K_i(z)) = z.$$

Then, we define $K_3(z)$ by the following formula:

$$K_3(z) = K_1(z) + K_2(z) - \frac{1}{z}. \tag{47}$$

It turns out that $K_3(z)$ is the K-function of a probability measure, μ_3, which is the *free convolution* of μ_1 and μ_2.

Let us turn to issues of convergence. Let $d(\mu_1, \mu_2)$ denote the Kolmogorov distance between the probability measures μ_1 and μ_2. That is, if $\mathcal{F}_1$ and $\mathcal{F}_2$ are the distribution functions corresponding to measures μ_1 and μ_2, respectively, then

$$d(\mu_1, \mu_2) =: \sup_{x\in\mathbb{R}} |\mathcal{F}_1(x) - \mathcal{F}_2(x)|.$$

Let μ be a probability measure with the zero mean and unit variance and let m_3 be its third absolute moment. Then the classical Berry-Esseen inequality says that

$$d\left(\mu^{(n)}, \nu\right) \leq C m_3 \frac{1}{\sqrt{n}},$$

where ν is the standard Gaussian measure and $\mu^{(n)}$ is the normalized n-time convolution of measure μ with itself:

$$\mu^{(n)}(du) = \mu * ... * \mu\left(\sqrt{n}du\right).$$

This inequality was proved by Berry (1941) and Esseen (1945) for a more general situation of independent but not necessarily identical measures. A simple example with Bernoulli measures shows that in this inequality the order of $n^{-1/2}$ cannot be improved without further restrictions.

We aim to derive a similar inequality when the usual convolution of measures is replaced by free convolution. Namely, let

$$\mu^{(n)}(du) = \mu \boxplus ... \boxplus \mu\left(\sqrt{n}du\right)$$

and let ν denote the standard semicircle distribution. It is known that $\mu^{(n)}$ converges weakly to ν (Voiculescu (1983), Maassen (1992), Pata (1996), and Voiculescu

(1998)). We are interested in the speed of this convergence and we prove that if μ is supported on $[-L, L]$, then

$$d\left(\mu^{(n)}, \nu\right) \leq CL^3 \frac{1}{\sqrt{n}}. \tag{48}$$

An example shows that the rate of $n^{-1/2}$ cannot be improved without further restrictions, similar to the classical case.

The main tool in our proof of inequality (48) is Bai's theorem (1993) that relates the supremum distance between two probability measures to a distance between their Cauchy transforms. To estimate the distance between Cauchy transforms, we use the fact that as n grows, the K-function of $\mu^{(n)}$ approaches the K-function of the semicircle law. Therefore, the main problem in our case is to investigate whether the small distance between K-functions implies a small distance between the Cauchy transforms themselves. We approach this problem using the Lagrange formula for functional inverses.

The rest of the paper is organized as follows: Section 2 contains the formulation and the proof of the main result. It consists of several subsections. In Subsection 2.1 we formulate the result and outline its proof. Subsection 2.2 evaluates how fast the K-function of $\mu^{(n)}$ approaches the K-function of the semicircle law. Subsection 2.3 provides useful estimates on behavior of the Cauchy transform of the semicircle law and related functions. Subsection 2.4 introduces a functional equation for the Cauchy transforms and concludes the proof by estimating how fast the Cauchy transform of $\mu^{(n)}$ converges to the Cauchy transform of the semicircle law. An example in Section 3 shows that the rate of $n^{-1/2}$ cannot be improved.

12.1 Formulation and proof of the main result

12.1.1 Formulation of the result and outline of the proof

Let the *semicircle law* be the probability measure on the real line that has the following cumulative distribution function:

$$\Phi(x) = \frac{1}{\pi} \int_{-\infty}^{x} \sqrt{4 - t^2} \chi_{[-2,2]}(t)\, dt,$$

where $\chi_{[-2,2]}(t)$ is the characteristic function of the interval $[-2, 2]$.

Theorem 95 *Suppose that μ is a probability measure that has zero mean and unit variance, and is supported on interval $[-L, L]$. Let $\mu^{(n)}$ be the normalized n-time free convolution of measure μ with itself: $\mu^{(n)}(du) = \mu \boxplus \ldots \boxplus \mu\left(\sqrt{n}du\right)$. Let $\mathcal{F}_n(x)$ denote the cumulative distribution function of $\mu^{(n)}$. Then for large enough n the following bound holds:*

$$\sup_x |\mathcal{F}_n(x) - \Phi(x)| \leq CL^3 n^{-1/2},$$

where C is an absolute constant.

Remark: $C = 2^{16}$ will do, although this constant is far from the best possible.

Proof: First, we repeat here for convenience of the reader one of Bai's results (same as Theorem 59 on p. 58):

Theorem 96 (Bai (1993)) *Let measures with distribution functions $\mathcal{F}$ and $\mathcal{G}$ be supported on a finite interval $[-B, B]$ and let their Cauchy transforms be $G_{\mathcal{F}}(z)$ and $G_{\mathcal{G}}(z)$, respectively. Let $z = u + iv$. Then*

$$\begin{aligned} \sup_x |\mathcal{F}(x) - \mathcal{G}(x)| \quad &\leq \quad \frac{1}{\pi(1-\kappa)(2\gamma - 1)} \left[\int_{-A}^{A} |G_{\mathcal{F}}(z) - G_{\mathcal{G}}(z)| \, du \right. \\ &\quad \left. + \frac{1}{v} \sup_x \int_{|y| \leq 2vc} |\mathcal{G}(x+y) - \mathcal{G}(x)| \, dy \right], \end{aligned} \tag{49}$$

where $A > B$, $\kappa = \frac{4B}{\pi(A-B)(2\gamma-1)} < 1$, $\gamma > 1/2$, and c and γ are related by the following equality

$$\gamma = \frac{1}{\pi} \int_{|u|<c} \frac{1}{1+u^2} du.$$

Note that if $\mathcal{G}(x)$ is the semicircle distribution then $|\mathcal{G}'(x)| \leq \pi^{-1}$. Therefore $|\mathcal{G}(x+y) - \mathcal{G}(x)| \leq |y|/\pi$. Integrating this inequality, we obtain:

$$\frac{1}{v} \sup_x \int_{|y| \leq 2vc} |\mathcal{G}(x+y) - \mathcal{G}(x)| \, dy \leq \frac{4c^2}{\pi} v. \tag{50}$$

Hence, the main question is how fast v can be made to approach zero if the first integral in (49) is also required to approach zero.

Let G_Φ and G_n be the Cauchy transforms of the semicircle law and $\mu^{(n)}$, respectively. Assume for the moment that the following lemma holds:

Lemma 97 *Suppose $v = 2^{10}L^3/\sqrt{n}$. Then for all sufficiently large n, we have the following estimate:*

$$\int_{-8}^{8} |G_n(u+iv) - G_\Phi(u+iv)|\, du \leq \frac{2^{10}L^3}{\sqrt{n}}.$$

Let us apply Bai's theorem using Lemma 97 and inequality (50). Let $\Phi(x)$ and $\mathcal{F}_n(x)$ denote the cumulative distribution functions of the semicircle law and $\mu^{(n)}$, respectively. The semicircle law is supported on $[-2,2]$, and by taking n sufficiently large we can ensure that $\mu^{(n)}$ is supported on any fixed inteval that includes $[-2,2]$ (see Bercovici and Voiculescu (1995)). Suppose that n is so large that $\mu^{(n)}$ is supported on $\left[-2^{5/4}, 2^{5/4}\right]$. Then we can take $A = 8$, $B = 2^{5/4}$, $c = 6$, and calculate $\gamma = 0.895$ and $\kappa = 0.682$. Then Bai's theorem gives the following estimate:

$$\begin{aligned}\sup_x |\mathcal{F}_n(x) - \Phi(x)| &\leq 1.268\left(2^{10}L^3 n^{-1/2} + 46937 L^3 n^{-1/2}\right)\\ &\leq 2^{16}L^3 n^{-1/2}.\end{aligned}$$

QED.

Thus, the main task is to prove Lemma 97. Here is the plan of the proof. First, we estimate how close the K-functions of $\mu^{(n)}$ and Φ are to each other. Then we note that the Cauchy transforms of $\mu^{(n)}$ and Φ can be found from their functional equations:

$$K_n(G_n(z)) = z,$$

and

$$K_\Phi(G_\Phi(z)) = z,$$

where K_n and K_Φ denote the K-functions of $\mu^{(n)}$ and Φ. From the previous step we know that $K_n(z)$ is close to $K_\Phi(z)$. Our goal is to show that this implies that $G_n(z)$ is close to $G_\Phi(z)$.

If we introduce an extra parameter, t, then we can include these functional equations in a parametric family:

$$K_t(G_t(z)) = z. \tag{51}$$

Parameter $t = 0$ corresponds to Φ and $t = 1$ to $\mu^{(n)}$. Next, we fix z and consider G_t as a function of t. We develop this function in a power series in t:

$$G_t = G_\Phi + \sum_{k=1}^{\infty} c_k t^k,$$

where c_k are functions of z. Then we estimate I_k for each $k \geq 1$, where

$$I_k = \int_{-8}^{8} |c_k(u+iv)|\, du.$$

Then,

$$\int_{-8}^{8} |G_n(u+iv) - G_\Phi(u+iv)|\, du \leq \sum_{k=1}^{\infty} I_k,$$

and our estimates of I_k allow us to prove the claim of Lemma 97.

12.1.2 Speed of convergence of K-functions

In this Subsection, we derive an estimate for the speed of convergence of the K-functions of $\mu^{(n)}$ and the semicircle law. Let $K_n(z)$ denote the K-function of $\mu^{(n)}$. For the semicircle law the K-function is $K_\Phi(z) = z^{-1} + z$. Define $\varphi_n(z) = K_n(z) - z - z^{-1}$.

Lemma 98 *Suppose μ has zero mean and unit variance and is supported on $[-L, L]$. Then the function $\varphi_n(z)$ is holomorphic in $|z| \leq \sqrt{n}/(8L)$ and*

$$|\varphi_n(z)| \leq 32L^3 \frac{|z|^2}{\sqrt{n}}.$$

Proof: The measure $\mu^{(n)}$ is the n-time convolution of the measure $\widetilde{\mu}(dx) =: \mu(\sqrt{n}dx)$ with itself. Therefore, $K_n(z) = nK_{\widetilde{\mu}}(z) - (n-1)z^{-1}$. Since $\widetilde{\mu}$ is supported on $[-L/\sqrt{n}, L/\sqrt{n}]$, we can estimate $K_{\widetilde{\mu}}(z) - \frac{1}{z} - \frac{1}{n}z$ inside the circle $|z| = \sqrt{n}/(8L)$ by using the estimates for coefficients of $K_{\widetilde{\mu}}(z)$ from Lemma 64 on p. 62:

$$\begin{aligned}
\left| K_{\widetilde{\mu}}(z) - \frac{1}{z} - \frac{1}{n}z \right| &= \sum_{k=2}^{\infty} b_k z^k \leq \frac{2L}{\sqrt{n}} \sum_{k=2}^{\infty} \frac{1}{k} \left(\frac{4L}{\sqrt{n}} \right)^k |z|^k \\
&\leq 32 \left(\frac{L}{\sqrt{n}} \right)^3 |z|^2 \sum_{k=2}^{\infty} \frac{1}{k2^{k-1}} \\
&\leq 32 \left(\frac{L}{\sqrt{n}} \right)^3 |z|^2 .
\end{aligned}$$

Note that we used the assumption about the mean and variance of the measure μ in the first line by setting $b_1 = 0$ and $b_1 = 1/n$.

Using the summation formula (47) for K-functions, we further obtain:

$$\left| K_n(z) - \frac{1}{z} - z \right| \leq 32 \frac{L^3}{\sqrt{n}} |z|^2 .$$

QED.

Lemma 98 shows that as n grows, the radius of the convergence area of $\varphi_n(z)$, and therefore of $K_n(z)$, grows proportionally to $\sqrt{n}$. In particular, the radius of convergence will eventually cover every bounded domain. Lemma 98 also establishes the rate of convergence of $K_n(z)$ to its limit $K_\Phi(z) = z^{-1} + z$.

12.1.3 Useful estimates

Suppose $G_\Phi(z)$ is the Cauchy transform of the semicircle distribution.

Lemma 99 *1)* $|G_\Phi(z)| \leq 1$ *if* $\operatorname{Im} z > 0$;
2) $|z - 2G_\Phi(z)| \geq 2\sqrt{\operatorname{Im} z}$ *if* $\operatorname{Im} z \in (0, 2)$.

Proof: $G_\Phi(z) = \left(z - \sqrt{z^2 - 4}\right)/2$. If $z = u + iv$ and v is fixed, then the maximum of $|G_\Phi(z)|$ is reached for $u = 0$. Then $|G_\Phi(iv)| = \left(\sqrt{v^2 + 4} - v\right)/2$ and $\sup |G_\Phi(iv)| = 1$.

Next, $|z - 2G_\Phi(z)| = \left|\sqrt{u^2 - v^2 - 4 + i2uv}\right|$. If v is in $(0, 2)$ and fixed, the minimum of this expression is reached for $u = \pm\sqrt{4 - v^2}$ and equals $2\sqrt{v}$. QED.

Lemma 100 *If* $n \geq 64L^2$ *and* $\operatorname{Im} z > 0$, *then we have:*

$$|\varphi_n(G_\Phi(z))| \leq \frac{32L^3}{\sqrt{n}}.$$

Proof: This Lemma follows directly from Lemmas 98 and 99. QED.

12.1.4 Functional equation for the Cauchy transform

Let $G_n(z)$ denote the Cauchy transform of $\mu^{(n)}$. Let us write the following functional equation:

$$G(t, z) + \frac{1}{G(t, z)} + t\varphi_n(G(t, z)) = z, \tag{52}$$

where t is a complex parameter. For $t = 0$ the solution is $G_\Phi(z)$, and for $t = 1$ the solution is $G_n(z)$. Assume that $\varphi_n(z)$ is not identically zero. (If it is, then $\mu^{(n)}$ is semicircle and $d\left(\mu^{(n)}, \nu\right) = 0$.) Let us write equation (52) as

$$t = \frac{zG - G^2 - 1}{G\varphi_n(G)}. \tag{53}$$

We can think about z as a fixed complex parameter and about t as a function of the complex variable G, i.e., $t = f(G)$. Suppose $\varphi_n(G_\Phi(z))$ does not equal zero for a given value of z. (This holds for all but a countable number of values of parameter z.) Then, as a function of G, f is holomorphic in a neighborhood of $G_\Phi(z)$. What we would like to do is to invert this function f and write $G = f^{-1}(t)$. In particular we would like to develop $f^{-1}(t)$ in a series of t around $t = 0$. Then we would be able to estimate $\left|f^{-1}(1) - f^{-1}(0)\right|$, which is equal to $\left|G_n(z) - G_\Phi(z)\right|$. To perform this inversion, we use the Lagrange formula in Lemma 60 on p. 58.

Assume that z is fixed, and let us write G instead of $G(z)$ and G_Φ instead of $G_\Phi(z)$. By Lemma 60, we can write the solution of (53) as

$$G = G_\Phi + \sum_{k=1}^{\infty} c_k t^k, \tag{54}$$

where

$$c_k = \frac{1}{k} \operatorname*{res}_{G=G_\Phi} \left(\frac{G\varphi_n(G)}{zG - G^2 - 1} \right)^k. \tag{55}$$

We aim to estimate $I_k =: \int_{-8}^{8} \left|c_k(u + iv)\right| du$. In particular, we will show that for any $v \in (0,1)$, $I_1 = O\left(n^{-1/2}\right)$. In addition, we will show that if $v = b/\sqrt{n}$ for a suitably chosen b, then $\sum_{k=2}^{\infty} I_k = o\left(n^{-1/2}\right)$. This information is sufficient for a good estimate of

$$\int_{-8}^{8} \left|G_n(u + iv) - G_\Phi(u + iv)\right| du.$$

Let us consider first the case of $k = 1$. Then

$$\begin{aligned} c_1 &= \frac{G_\Phi \varphi_n(G_\Phi)}{G_2 - G_\Phi} \\ &= \frac{G_\Phi \varphi_n(G_\Phi)}{\sqrt{z^2 - 4}}, \end{aligned}$$

where G_2 denotes another root of the equation $G^2 - zG + 1$. Therefore, if $z = u + iv$, then we can calculate:

$$\begin{aligned} |c_1| &= \frac{|G_\Phi| \, |\varphi_n (G_\Phi)|}{\left[(u^2-4)^2 + 2 (u^2+4) v^2 + v^4 \right]^{1/4}} \\ &\leq \frac{32L^3}{\sqrt{n}} \frac{1}{\left[(u^2-4)^2 + 2 (u^2+4) v^2 + v^4 \right]^{1/4}}, \end{aligned}$$

where the last inequality holds by Lemma 100 for all $n \geq 64L^2$.

Lemma 101 *For every* $v \in (0,1)$,

$$\int_{-8}^{8} \left[\left(u^2 - 4 \right)^2 + 2 \left(u^2 + 4 \right) v^2 + v^4 \right]^{-1/4} du < 24.$$

Proof: Let us make substitution $x = u^2 - 4$. Then we get:

$$\begin{aligned} J &= \int_{-4}^{60} \left[x^2 + 2 (x+8) v^2 + v^4 \right]^{-1/4} \frac{dx}{\sqrt{x+4}} \\ &\leq \int_{-4}^{60} \frac{1}{(x^2+v^2)^{1/4}} \frac{dx}{\sqrt{x+4}}. \end{aligned}$$

Now we divide the interval of integration in two parts and write:

$$\begin{aligned} J &\leq \int_{-4}^{-2} ... + \int_{-2}^{60} ... \\ &\leq \frac{1}{\sqrt{2}} \int_{-4}^{-2} \frac{dx}{\sqrt{x+4}} + \frac{2}{\sqrt{2}} \int_{0}^{60} \frac{dx}{x^{1/2}} \\ &= 2 + \frac{4}{\sqrt{2}} \sqrt{60} < 24. \end{aligned}$$

QED.

Corollary 102 *For every* $v \in (0,1)$ *and all* $n \geq 64L^2$, *it is true that*

$$I_1 =: \int_{-8}^{8} |c_1 (u+iv)| \, du \leq \frac{768L^3}{\sqrt{n}}$$

Now we estimate c_k in (54) for $k \geq 2$. Define function $f_k(G)$ by the formula

$$f_k(G) =: \left(\frac{G\varphi_n(G)}{G_2 - G}\right)^k,$$

where G_2 denotes the root of the equation $G^2 - zG + 1$, which is different from G_Φ. Then formula (55) implies that kc_k equal to the coefficient before $(G - G_\Phi)^{k-1}$ in the expansion of $f_k(G)$ in power series of $(G - G_\Phi)$. To estimate this coefficient, we will use the Cauchy inequality:

$$|kc_k| \leq \frac{M_k(r)}{r^{k-1}},$$

where $M_k(r)$ is the maximum of $|f_k(G)|$ on the circle $|G - G_\Phi| = r$.

We will use $r = \sqrt{v}$ and our first goal is to estimate $M_k(\sqrt{v})$.

Lemma 103 *Let $z = u + iv$ and suppose that $v \in (0,1)$. If $n \geq 256L^2$, then*

$$M_k(v) \leq \left[\frac{512L^3}{\sqrt{n}} \frac{1}{\left((u^2-4)^2 + 2u^2v^2\right)^{1/4}}\right]^k.$$

Proof: Note that $|G| \leq |G_\Phi| + \sqrt{v}$ and therefore $|G| \leq 2$ provided that $v \in (0,1)$. Then Lemma 98 implies that if $n \geq 256L^2$, then $\varphi_n(G)$ is well defined and $|\varphi_n(G)| \leq 128L^3/\sqrt{n}$. It remains to estimate $|G_2 - G|$ from below. If we write $G = G_\Phi + e^{i\theta}\sqrt{v}$, then we have

$$\begin{aligned}
|G_2 - G| &= \left|\sqrt{z^2-4} - e^{i\theta}\sqrt{v}\right| \\
&\geq \left|\sqrt{z^2-4}\right| - \sqrt{v} \\
&= \left(\left(u^2-4\right)^2 + 2\left(u^2+4\right)v^2 + v^4\right)^{1/4} - \sqrt{v} \\
&> \left(\left(u^2-4\right)^2 + 2\left(u^2+4\right)v^2\right)^{1/4} - \sqrt{v} > 0.
\end{aligned}$$

From the concavity of function $t^{1/4}$ it follows that for positive A and B the following inequality holds:

$$[8(A+B)]^{1/4} - A^{1/4} \geq B^{1/4}.$$

Using v^2 as A, and $\left[\left(u^2-4\right)^2+2u^2v^2\right]/8$ as B, we can write this inequality as follows:

$$\left(\left(u^2-4\right)^2+\left(2u^2+4\right)v^2\right)^{1/4}-\sqrt{v}\geq\frac{1}{8^{1/4}}\left(\left(u^2-4\right)^2+2u^2v^2\right)^{1/4}>0.$$

Therefore

$$M_k\left(v\right)\leq\left[\frac{512L^3}{\sqrt{n}}\frac{1}{\left(\left(u^2-4\right)^2+2u^2v^2\right)^{1/4}}\right]^k.$$

QED.

Corollary 104 *For every* $v\in(0,1)$, $k\geq 2$, *and all* $n\geq 256L^2$, *it is true that*

$$\left|kc_k\left(u+iv\right)\right|\leq v^{-\frac{k-1}{2}}\left[\frac{512L^3}{\sqrt{n}}\frac{1}{\left(\left(u^2-4\right)^2+2u^2v^2\right)^{1/4}}\right]^k$$

Now we want to estimate integrals of $\left|c_k\left(u+iv\right)\right|$ when u changes from -8 to 8. The cases of $k=2$ and $k>2$ are slightly different and we treat them separately.

Let

$$I_k=:\int_{-8}^{8}\left|c_k\left(u+iv\right)\right|du$$

Lemma 105 *If* $v\in(0,1)$ *and* $n\geq 256L^2$*, then i)*

$$I_2\leq\frac{\log\left(60/v\right)}{\sqrt{v}}\frac{2^{19}L^6}{n},$$

and ii) if $k>2$, *then*

$$I_k\leq\frac{12}{k}v^{3/2}\left(\frac{512L^3}{v\sqrt{n}}\right)^k.$$

Proof: Using Corollary 104, we write:

$$I_k=:\int_{8}^{8}\left|c_k\left(u+iv\right)\right|du\leq\frac{1}{k}\frac{1}{v^{(k-1)/2}}\left(\frac{512L^3}{\sqrt{n}}\right)^k\int_{-8}^{8}\frac{du}{\left(\left(u^2-4\right)^2+2u^2v^2\right)^{k/4}}.$$

After substitution $x=u^2-4$, the integral in the right-hand side of the inequality can be re-written as

$$J_k=:\int_{-4}^{60}\frac{1}{\left(x^2+2\left(x+4\right)v^2\right)^{k/4}}\frac{dx}{\sqrt{x+4}}.$$

We divide the interval of integration into two portions and write the following inequality:

$$\begin{aligned} J_k &\leq \int_{-4}^{-1} \frac{dx}{\sqrt{x+4}} + \int_{-1}^{60} \frac{dx}{\left(x^2 + 2\left(x+4\right)v^2\right)^{k/4}} \\ &\leq 2\sqrt{3} + \int_{-1}^{60} \frac{dx}{\left(x^2+v^2\right)^{k/4}}. \end{aligned}$$

If we use substitution $s = x/v$, then we can write:

$$\begin{aligned} \int_{-1}^{60} \frac{dx}{\left(x^2+v^2\right)^{k/4}} &= \int_{-1/v}^{60/v} \frac{ds}{v^{\frac{k}{2}-1}\left(1+s^2\right)^{k/4}} \\ &\leq \frac{2}{v^{\frac{k}{2}-1}} \int_{0}^{60/v} \frac{ds}{\left(1+s^2\right)^{k/4}}. \end{aligned}$$

We again separate the interval of integration in two parts and write:

$$\frac{2}{v^{\frac{k}{2}-1}} \int_{0}^{60/v} \frac{ds}{\left(1+s^2\right)^{k/4}} \leq \frac{2}{v^{\frac{k}{2}-1}} \left[\int_0^1 ds + \int_1^{60/v} \frac{ds}{s^{k/2}} \right].$$

Here we have two different cases. If $k = 2$, then we evaluate the integrals as $1 + \log\left(60/v\right)$. Therefore,

$$\begin{aligned} J_2 &\leq 2\sqrt{3} + 2 + 2\log\left(60/v\right) \\ &\leq 4\log\left(60/v\right). \end{aligned}$$

Hence,

$$I_2 \leq \frac{\log\left(60/v\right)}{\sqrt{v}} \frac{2^{19} L^6}{n}.$$

If $k > 2$, then we have:

$$\begin{aligned} \frac{2}{v^{\frac{k}{2}-1}} \left[\int_0^1 ds + \int_1^{60/v} \frac{ds}{s^{k/2}} \right] &= \frac{2}{v^{\frac{k}{2}-1}} \left[1 + \frac{1}{-\frac{k}{2}+1} \left(\left(\frac{60}{v}\right)^{-\frac{k}{2}+1} - 1 \right) \right] \\ &\leq \frac{2}{v^{\frac{k}{2}-1}} \left(1 + \frac{1}{\frac{k}{2}-1} \right). \\ &\leq \frac{6}{v^{\frac{k}{2}-1}}. \end{aligned}$$

Therefore,

$$J_k \leq 2\sqrt{3} + \frac{6}{v^{\frac{k}{2}-1}} \leq \frac{12}{v^{\frac{k}{2}-1}},$$

and

$$\begin{aligned} I_k &\leq \frac{1}{k}\frac{1}{v^{(k-1)/2}}\left(\frac{512L^3}{\sqrt{n}}\right)^k \frac{12}{v^{\frac{k}{2}-1}} \\ &\leq \frac{12}{k}v^{3/2}\left(\frac{512L^3}{v\sqrt{n}}\right)^k. \end{aligned}$$

QED.

Corollary 106 *If* $v = 1024L^3/\sqrt{n}$, *and* $n \geq 256L^2$ *then*

$$I_2 \leq 2^{14}L^{9/2}\frac{\log\left(\frac{15}{256L^3}\sqrt{n}\right)}{n^{3/4}}.$$

In particular, $I_2 = o\left(n^{-1/2}\right)$ *as* $n \to \infty$.

Now we address the case when $k > 2$.

Corollary 107 *Suppose* $v = 1024L^3/\sqrt{n}$, *and* $n \geq 256L^2$. *Then*

$$\sum_{k=3}^{\infty} I_k \leq \frac{3}{2}v^{3/2} = 1536L^{9/2}\frac{1}{n^{3/4}}.$$

In particular, $\sum_{k=2}^{\infty} I_k = o\left(n^{-1/2}\right)$ *as* $n \to \infty$.

Joining results of Corollaries 102, 106, and 107, we get the following result.

Lemma 108 *Suppose* $v = 1024L^3/\sqrt{n}$. *Then for all sufficiently large* n, *we have the following estimate:*

$$\int_{-8}^{8} \left|G_n\left(u+iv\right) - G_\Phi\left(u+iv\right)\right| du \leq \frac{2^{10}L^3}{\sqrt{n}}.$$

Proof: From formula (54) we have

$$\left|G_n - G_\Phi\right| \leq \sum_{k=1}^{\infty} \left|c_k\right|.$$

Since the series has only positive terms, we can integrate it term by term and write:

$$\begin{aligned}\int_{-8}^{8} \left|G_n(u+iv) - G_\Phi(u+iv)\right| du &\leq \sum_{k=1}^{\infty} \int_{-8}^{8} \left|c_k(u+iv)\right| du \\ &\leq \frac{768L^3}{\sqrt{n}} + o\left(n^{-1/2}\right) \\ &\leq \frac{2^{10}L^3}{\sqrt{n}}\end{aligned}$$

for all sufficiently large n. QED.

Lemma 108 is identical to Lemma 97 and its proof completes the proof of Theorem 95.

12.2 Example

Consider a Bernoulli measure: $\mu\{-1/p\} = p$ and $\mu\{1/q\} = q \equiv 1-p$. This is a zero-mean measure with the variance equal to $(pq)^{-1}$. Let $\mu^{(n)}(dx) = \mu \boxplus ... \boxplus \mu\left(\sqrt{\frac{n}{pq}}dx\right)$ and let $\mathcal{F}_n(x)$ be the distribution function corresponding to $\mu^{(n)}$.

Proposition 109 *If $p \neq q$, then there exist such positive constants C_1 and C_2 that*

$$C_1 n^{-1/2} \leq \sup_x \left|\mathcal{F}_n(x) - \Phi(x)\right| \leq C_2 n^{-1/2}$$

for every n.

Proof: From the Voiculescu addition formula and the Stieljes inversion formula, it is easy to compute the density of the distribution of $\mu^{(n)}$:

$$f_n(x) = \frac{1}{2\pi} \frac{\sqrt{4 - x^2 + 2\frac{p-q}{\sqrt{pqn}}x - \frac{1}{pqn}}}{\left(1 + \frac{x}{\sqrt{nq/p}}\right)\left(1 - \frac{x}{\sqrt{np/q}}\right)},$$

if the square root is real, and if not, $f_n(x) = 0$. We compare this distribution with the semicircle distribution, which has the following density:

$$\phi(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\chi_{[-2,2]}(x).$$

More precisely, we seek to estimate

$$\sup_x \left| \int_{-\infty}^{x} \left(f_n\left(t\right) - \phi\left(t\right)\right) dt \right|.$$

The support of f_n is $\left[-2\sqrt{1-n^{-1}} + cn^{-1/2},\ 2\sqrt{1-n^{-1}} + cn^{-1/2}\right]$, where $c = \left(p-q\right)/\sqrt{pq}$. Suppose in the following that $p > q$ and introduce the new variable $u = x + 2\sqrt{1-n^{-1}} - c/\sqrt{n}$. Then,

$$\begin{aligned} 2\pi f_n\left(u\right) &= \sqrt{4u\sqrt{1-n^{-1}} - u^2} \\ &\sim \sqrt{4u - u^2}, \end{aligned}$$

where the asymptotic equivalence is for u fixed and $n \to \infty$ and we omit all terms that are $o\left(n^{-1/2}\right)$. Similarly,

$$\begin{aligned} 2\pi\phi\left(x\right) &= \sqrt{4 - \left(u - 2\left(1-n^{-1}\right) + cn^{-1/2}\right)^2} \\ &\sim \sqrt{4u - u^2 + 4cn^{-1/2} - 2cun^{-1/2}} \\ &= \sqrt{4u - u^2}\sqrt{1 + 2c\frac{1}{\sqrt{n}u}\frac{2-u}{4-u}}. \end{aligned}$$

Consequently,

$$\begin{aligned} \phi\left(u\right) - f_n\left(u\right) &\sim \sqrt{4u-u^2}\left[\sqrt{1 + 2c\frac{1}{\sqrt{n}u}\frac{2-u}{4-u}} - 1\right] \\ &\sim \sqrt{4u-u^2}\,c\frac{1}{\sqrt{n}u}\frac{2-u}{4-u} \\ &= c\frac{1}{\sqrt{nu}}\frac{2-u}{\sqrt{4-u}}. \end{aligned}$$

After integrating we get:

$$\left| \int_0^x \left(f_n\left(u\right) - \phi\left(u\right)\right) du \right| \sim c\frac{1}{\sqrt{n}} f\left(x\right),$$

where $f\left(x\right)$ is a continuous positive bounded function. From this expression it is clear that $\sup_x \left| \int_{-\infty}^{x} \left(f_n\left(t\right) - \phi\left(t\right)\right) dt \right|$ has the order of $n^{-1/2}$ provided that $p \neq q$. QED.

This example shows that the rate of $n^{-1/2}$ in Theorem 95 cannot be improved without further restrictions on measures. It would be interesting to extend Theorem 95 to measures with unbounded support or relate the constant in the inequality to moments of the convolved measures, similar to the classical case.

13 Superconvergence of Sums of Free Random Variables

In many respects, free probability theory parallels classical probability theory. There exist analogues of the central limit theorem (Voiculescu (1986)), the law of large numbers [Bercovici and Pata (1996)], and the classification of infinitely divisible and stable laws [Bercovici and Voiculescu (1992) and Bercovici et al. (1999)]. On the other hand, certain features of free and classical probability theories differ strikingly. Let $S_n = n^{-1/2} \sum_{i=1}^{n} X_i$, where X_i are identically distributed and free random variables. Then the law of S_n approaches the limit law in a completely different manner than in the classical case. To illustrate this, suppose that the support of X_i is $[-1, 1]$. Take a positive number $\alpha < 1$. Then, in the classical case the probability of $\{|S_n| > \alpha n\}$ is exponentially small but not zero. In contrast, in the non-commutative case the probability becomes identically zero for all sufficiently large n. This mode of convergence has been called *superconvergence* by Bercovici and Voiculescu (1995).

In this paper we extend the superconvergence result to a more general setting of non-identically distributed variables and estimate the rate of the superconvergence quantitatively. It turns out, in particular, that the support of S_n can deviate from the supporting interval of the limiting law by not more than $c/\sqrt{n}$, and we explicitly estimate the constant c. An example shows that the rate $n^{-1/2}$ in this estimate cannot be improved.

Related results have been obtained in random matrix literature. For example, Johnstone (2001) considers the distribution of the largest eigenvalue of an empirical covariance matrix for a sample of Gaussian vectors. This problem can be seen as a problem about the edge of the spectrum of a sum of n random rank-one operators in the N-dimensional vector space. More precisely, the question is about sums of the form $S_n = \sum_{i=1}^{n} x_i x_i'$, where x_i is a random N-vector with the entries distributed according to the Gaussian law with the normalized variance $1/N$. Then S_n is a matrix-valued random variable with the Wishart distribution.

Johnstone is interested in the asymptotic behavior of the distribution of the largest eigenvalue of S_n. The asymptotics is derived under the assumption that both n and N approach ∞, and that $\lim n/N = \gamma > 0$, $\gamma \neq \infty$. Johnstone finds that the largest eigenvalue has the variance of the order $n^{-2/3}$ and that after an appropriate normalization the distribution of the largest eigenvalue approaches the Tracy-Widom law. This law has a right-tail asymptotically equivalent to $\exp\left[-\left(2/3\right) s^{3/2}\right]$, and, in particular is unbounded from above. Johnstone's results have generalized the original breakthrough results by Tracy and Widom (1996) (see also Tracy and Widom (2000)) for selfadjoint random matrices without the covariance structure. In Soshnikov (1999) and (2002), it is shown that the results regarding the asymptotic distribution of the largest eigenvalue remain valid even if the matrix entries are not necessarily Gaussian.

In an earlier contribution, Bai and Silverstein (1998) also considered empirical covariance matrices of large random vectors that are non-neccesarily Gaussian and studied their largest eigenvalues. Again both n and N approach infinity and $\lim n/N = \gamma > 0, \gamma \neq \infty$. In contrast to Johnstone, Bai and Silverstein were interested in the behavior of the largest eigenvalue along a sequence of increasing random covariance matrices. Suppose the support of the limiting eigenvalue distribution is contained in the interior of a closed interval, I. Bai and Silverstein showed that the probability that the largest eigenvalue lies outside of I is zero for all sufficiently large n.

These results are not directly comparable with ours for several reasons. First, in our case the edge of the spectrum is not random in the classical sense and so it does not make sense to talk about its variance. Second, informally speaking, we are looking at the limit situation when $N = \infty$, $n \to \infty$. Because of this, we use much easier techniques than all these papers as we do not need to handle the interaction of the randomness and the passage to the asymptotic limit. Despite these differences, comparison of our results with the results of the random matrix literature is stimulating. In particular, the superconvergence in free probability theory can be thought as an analogue of the Bai-Silverstein result.

13.1 Results and examples

In the classical case the behavior of large deviations from the CLT is described by the Cramer theorem, the Bernstein inequality, and their generalizations. It turns out that in the non-commutative case, the behavior of large deviations is considerably different. The theorem below gives some quantitative bounds on how the distribution of a sum of free random variables differ from the limiting distribution.

Let $X_{n,i}$, $i = 1, ..., k_n$ be a double-indexed array of bounded self-adjoint random variables. The elements of each row, $X_{n,1}, ..., X_{n,k_n}$ are assumed to be free but are not necessarily identically distributed. Their associated probability measures are denoted $\mu_{n,i}$, their Cauchy transforms are $G_{n,i}(z)$, their k-th moments are $a_{n,i}^{(k)}$, etc. We define $S_n = X_{n,1} + ... + X_{n,k_n}$ and we will the study the behavior of the probability measure

μ_n associated with S_n.

We will assume that the first moments of the random variables $X_{n,i}$ are zero and that $\|X_{n,i}\| \leq L_{n,i}$. Let $v_n = a_{n,1}^{(2)} + ... + a_{n,k_n}^{(2)}$, $L_n = \max_i \{L_{n,i}\}$, and $T_n = L_{n,1}^3 + ... + L_{n,k_n}^3$.

Theorem 110 *Suppose that* $\limsup_{n\to\infty} T_n/v_n^{3/2} > 2^{-12}$. *Then for all sufficiently large* n *the support of* μ_n *belongs to*

$$I = \left(-2\sqrt{v_n} - cT_n/v_n, 2\sqrt{v_n} + cT_n/v_n\right),$$

where $c > 0$ *is an absolute constant.*

Remark 1: $c = 256$ will do although it is not best possible.

Remark 2: Informally, the assumption that $\limsup_{n\to\infty} T_n/v_n^{3/2} > 2^{-12}$ means that there are no large outliers. An example when the assumption is violated is provided by random variables with variance $a_{n,i}^{(2)} = n^{-1}$ and $L_{n,i} = 1$. Then $T_n = n$ and $v_n^{3/2} = 1$, so that $T_n/v_n^{3/2}$ increases when n grows.

Remark 3: The assumption in Theorem 110 are weaker than the assumptions of Theorem 7 in Bercovici and Voiculescu (1995). In particular, Theorem 110 allows making conclusions about random variables with non-uniformly bounded support. Consider, for example, random variables X_k, $k = 1, ..., n$, that are supported on intervals $\left[-k^{1/3}, k^{1/3}\right]$ and have variances of order $k^{2/3}$. Then T_n has the order of n^2 and v_n has the order of $n^{5/3}$. Therefore, $T_n/v_n^{3/2}$ has the order of $n^{-1/2}$ and Theorem 110 is applicable. It allows us to conclude that the support of $S_n = X_1 + ... + X_n$ is contained in the interval $\left(-2\sqrt{v_n} - cn^{1/3}, -2\sqrt{v_n} + cn^{1/3}\right)$.

Example 111 *Identically Distributed Variables.*

A particular case of the above scheme is the normalized sums of identically distributed, bounded, free r.v.: $S_n = (\xi_1 + ... + \xi_n)/\sqrt{n}$. If $\|\xi_i\| \leq L$ then $\|\xi_i/\sqrt{n}\| \leq L_{n,i} = L_n = L/\sqrt{n}$. Therefore $T_n = L^3/\sqrt{n}$. If the second moment of ξ_i is σ^2 then the second moment of the sum S_n is $v_n = \sigma^2$. Applying the theorem we obtain the result that starting with certain n, the support of the distribution of S_n belongs to $\left(-2\sigma - c\left(L^3/\sigma^2\right)n^{-1/2}, 2\sigma + c\left(L^3/\sigma^2\right)n^{-1/2}\right)$.

Example 112 *Free Poisson*

Let the n-th row of our scheme has $k_n = n$ identically distributed random variables $X_{n,i}$ with the Bernoulli distribution that places probability $p_{n,i}$ on 1 and $q_{n,i} = 1 - p_{n,i}$ on 0. (It is easy to normalize this distribution to have the zero mean by subtracting $p_{n,i}$). Suppose $\max_i p_{n,i} \to 0$ as $n \to \infty$ and

$$\sum_{i=1}^{n} p_{n,i} \to \lambda > 0$$

as $n \to \infty$. Then $L_{n,i} \sim 1$ and $a_{n,i}^2 = p_{n,i} (1 - p_{n,i})$ so that $T_n \sim n$ and $v_n \to \lambda$ as $n \to \infty$. Therefore, Theorem 110 does not apply. An easy calculation for the case $p_{n,i} = \lambda/n$ shows that superconvergence still holds. This example shows that the conditions of the theorem are not necessary for superconvergence to hold.

Example 113 *Identically Distributed Bernoulli Variables*

Let X_i be identically distributed with a distribution that puts positive weights p and q on $-\sqrt{q/p}$ and $\sqrt{p/q}$, respectively. Then $EX_i = 0$ and $EX_i^2 = 1$. It is not difficult to show that the support of $S_n = n^{-1/2} \sum_{i=1}^{n} X_i$ is the interval $I = [x_1, x_2]$, where

$$x_{1,2} = \pm 2\sqrt{1 - \frac{1}{n}} + \frac{q - p}{\sqrt{pq}} \frac{1}{\sqrt{n}}.$$

This example shows that rate of $n^{-1/2}$ in Theorem 110 cannot be improved without further restrictions. Note also that for $p > q$, L_n is $\sqrt{p/q}$, and therefore the coefficient before $n^{-1/2}$ is of order L_n. In the general bound the coefficient is L_n^3/σ^2. It is not clear whether it is possible to replace the coefficient in the general bound by a term of order L_n.

Recall the scaling properties of the Cauchy transform and its inverse:

Lemma 114 *i)* $G_{\alpha A}(z) = \alpha^{-1} G_A(z/a)$ *and ii)* $K_{\alpha A}(u) = \alpha K_A(\alpha u)$.

The claim of the lemma follows directly from definitions.

13.2 Proof

The key ideas of the proof are as follows:

1) We know that the Cauchy transform of the sum S_n is the Cauchy transform of a bounded r.v. (since by assumption each $X_{n,i}$ is bounded). Consequently the Cauchy transform of S_n is holomorphic in a certain circle around infinity (i.e., in the area

$|z| > R$ for some R). We want to estimate R and apply Lemma 56 to conclude that S_n is supported on $[-R, R]$.

2) Since the K-function of S_n, call it $K_n(z)$, is the sum of K-functions of $X_{n,i}$ and the latter are inverses of Cauchy transforms of $X_{n,i}$, it is an exercise in complex analysis to estimate the radius of convergence of power series of $K_n(z)$ at $z = 0$ and locate critical points of $K_n(z)$ (i.e., zeros of its derivative). Using this information, we can prove that the K-function of S_n takes real values and is a one-to-one function on a sufficiently large real interval around zero. Therefore, it has a differentiable inverse defined on a sufficiently large real interval around infinity (i.e. on the set $I = (-\infty, -A] \cup [A, \infty)$ for some A, which we can explicitly estimate). Moreover, with a little bit more effort we can show that this inverse function is well-defined and holomorphic in an open complex neighborhood of I. This shows that Lemma 56 is applicable, and the estimate for A provides the desired estimate for the support of S_n.

We will start with finding the radius of convergence of the Taylor series of $K_n(z)$. First we need to prove some preliminary facts about Cauchy transforms of $X_{n,i}$.

Define $g_{n,i}(z) = G_{n,i}(z^{-1})$. Since the series $G_{n,i}(z)$ are convergent everywhere in $|z| > L_{n,i}$, then the Taylor series for $g_{n,i}(z)$ converges everywhere in $|z| < L_{n,i}^{-1}$.

Assume that $R_{n,i}$ and $m_{n,i}$ are such that

1. $R_{n,i} \geq L_{n,i}$;
2. $|G_{n,i}(z)| \geq m_{n,i} > 0$ everywhere on $|z| = R_{n,i}$;
3. $g_{n,i}(z)$ is one-to-one in $|z| < R_{n,i}^{-1}$.

For example, we can take $R_{n,i} = 2L_{n,i}$ and $m_{n,i} = (4L_{n,i})^{-1}$. Indeed, for any z with $|z| = r > L_{n,i}$ we can estimate $G_{n,i}(z)$:

$$\begin{aligned} |G_{n,i}(z)| &\geq \frac{1}{r} - \left(\frac{a_{n,i}^2}{r^3} + \frac{|a_{n,i}^3|}{r^4} + \dots \right) \\ &\geq \frac{1}{r} - \left(\frac{L_{n,i}^2}{r^3} + \frac{L_{n,i}^3}{r^4} + \dots \right) \\ &= \frac{1}{r} - \frac{L_{n,i}^2}{r^2} \frac{1}{r - L_{n,i}}. \end{aligned}$$

In particular, taking $r = 2L_{n,i}$ we get the estimate:

$$|G_{n,i}(z)| \geq \frac{1}{4L_{n,i}},$$

valid for every i and everywhere on $|z| = 2L_{n,i}$.

It remains to show that $g_{n,i}(z)$ is one-to-one in $|z| < (2L_{n,i})^{-1}$. This is indeed so because

$$g_{n,i}(z) = z\left(1 + a_{n,i}^{(2)}z^2 + a_{n,i}^{(3)}z^3 + ...\right),$$

and we can estimate

$$\left|a_{n,i}^{(2)}z^2 + a_{n,i}^{(3)}z^3 + ...\right| \leq L_{n,i}^2\left(\frac{1}{2L_{n,i}}\right)^2 + L_{n,i}^3\left(\frac{1}{2L_{n,i}}\right)^3 + ... = \frac{1}{2}.$$

Definition 115 *Let* $R_n = \max_i\{R_{n,i}\}$, $m_n = \min_i\{m_{n,i}\}$, *and* $D_n = \sum_{i=1}^{k_n} R_{n,i}(m_{n,i})^{-2}$.

We are now able to investigate the region of convergence for the series $K_{n,i}(z)$.

Lemma 116 *The radius of convergence of* K*-series for measure* μ_n *is at least* m_n.

The lemma says essentially that if r.v. $X_{n,1}$, ..., X_{n,k_n} are all bounded by L_n, then K-series for $\sum_i X_{n,i}$ converge in the circle $|z| \leq 1/(4L_n)$.

Proof: Let us apply Lemma 61 to $G_{n,i}(z)$ with γ having radius $(R_{n,i})^{-1}$. By Lemma 61 the coefficients in the series for the inverse of $G_{n,i}(z)$ are

$$b_{n,i}^{(k)} = \frac{1}{2\pi ik}\oint_{\partial\gamma}\frac{dz}{z^2 g_{n,i}(z)^k},$$

and we can estimate them as

$$\left|b_{n,i}^{(k)}\right| \leq \frac{R_{n,i}}{k}(m_{n,i})^{-k}.$$

This implies that the radius of convergence of K-series for measures $\mu_{n,i}$ is $m_{n,i}$. Consequently, the radius of convergence of K-series for measure μ_n is at least m_n. QED.

Now we can investigate the behavior of $K_n(z)$ and its derivative inside its convergence circle.

Lemma 117 *For every* z *in* $|z| < m_n$, *the following inequalities are valid:*

$$\left|K_n(z) - \frac{1}{z} - v_n z\right| \leq D_n|z|^2, \qquad (56)$$

$$\left|K_n'(z) + \frac{1}{z^2} - v_n\right| \leq 2D_n|z|. \qquad (57)$$

Note that D_n is approximately $k_n L_n^3$, so the meaning of the lemma is that the growth of $K_n - z^{-1} - v_n z$ around $z = 0$ is bounded by a constant that depends on the norm of the variables $X_{n,1}$, ..., X_{n,k_n}.

Proof: Consider the circle with radius $m_{n,i}/2$. We can estimate $K_{n,i}$ inside this circle

$$\begin{aligned}\left|K_{n,i} - \frac{1}{z} - a_{n,i}^{(2)} z\right| &\leq \frac{R_{n,i}}{2} (m_{n,i})^{-2} |z|^2 + \frac{R_{n,i}}{3} (m_{n,i})^{-3} |z|^2 \frac{m_{n,i}}{2} \\ &\quad + \frac{R_{n,i}}{4} (m_{n,i})^{-3} |z|^2 \frac{m_{n,i}^2}{2^2} + ... \\ &= R_{n,i} (m_{n,i})^{-2} |z|^2 \left(\frac{1}{2} + \frac{1}{3}\frac{1}{2} + \frac{1}{4}\frac{1}{2^2} + ...\right) \leq R_{n,i} (m_{n,i})^{-2} |z|^2 .\end{aligned}$$

Consequently, using Voiculescu's addition formula we can estimate

$$\left|K_n (z) - \frac{1}{z} - v_n z\right| \leq D_n |z|^2 , \tag{58}$$

Similar argument leads to the estimate:

$$\left|K_n' (z) + \frac{1}{z^2} - v_n\right| \leq 2D_n |z| . \tag{59}$$

QED.

Lemma 118 *Suppose* $m_n > 4/\sqrt{v_n}$, $r_n < 1/\left(2\sqrt{v_n}\right)$, *and* $r_n \geq 4D_n/v_n^2$. *Then there are no zeros of* $K_n' (z)$ *inside* $|z| \leq 1/\sqrt{v_n} - r_n$.

Proof: On $|z| = v_n^{-1/2} - r_n$, we have $|z|^{-2} > v_n$. Also $\left|z - v_n^{-1/2}\right| \left|z + v_n^{-1/2}\right| > r_n v_n^{-1/2}$. This is easy to see by considering two cases $\operatorname{Re} z \geq 0$ and $\operatorname{Re} z \leq 0$. In the first case $\left|z - v_n^{-1/2}\right| \geq r_n$ and $\left|z + v_n^{-1/2}\right| > v_n^{-1/2}$. In the second case $\left|z - v_n^{-1/2}\right| > v_n^{-1/2}$ and $\left|z + v_n^{-1/2}\right| \geq r_n$. Therefore,

$$\begin{aligned}\left|-z^{-2} + v_n\right| &= \left|z^{-2}\right| v_n \left|z - v_n^{-1/2}\right| \left|z + v_n^{-1/2}\right| \\ &> r_n v_n^{3/2}.\end{aligned}$$

The circle γ lies entirely in the area where formula (57) applies to $K_n' (z)$. (Since by assumption $r_n < v_n^{-1/2}/2$, then $r_n + v_n^{-1/2} < 2v_n^{-1/2}$ and therfore $r_n + v_n^{-1/2} <$

$m_n/2$, provided that $m_n v_n^{1/2} > 4$, which holds by assumption.) Consequently, using (57) we can estimate

$$\left|K_n'(z) - \left(-z^{-2} + v_n\right)\right| \leq 4D_n v_n^{-1/2},$$

where we used $|z| \leq 2v_n^{-1/2}$. By assumption $r_n \geq 4D_n v_n^{-2}$, therefore $v_n^{3/2} r_n \geq 4D_n v_n^{-1/2}$, and Rouche's theorem is applicable. Both $K_n'(z)$ and $-z^{-2}+v_n$ have only one pole, which is of order two, in $|z| \leq v^{-1/2} - r_n$, and the function $-z^{-2} + v_n$ has no zeros inside $|z| \leq v^{-1/2} - r_n$. Therefore, Rouche's theorem implies that there is no zeros of $K_n'(z)$ inside $|z| \leq v^{-1/2} - r_n$. (Rouche's theorem is often formulated only for holomorphic functions but as a consequence of the argument principle (see e.g. Theorems II.2.3 and II.2.4 in Markushevich (1977)) it can be easily re-formulated for meromorphic functions. In this form it claims that a meromorphic function, $f(z)$, has the same difference between the number of zeros and number of poles inside a curve γ as another meromorphic function, $g(z)$, provided that $|f(z)| > |g(z) - f(z)|$. For this formulation see, e.g., Hille (1962), Theorem 9.2.3.) QED.

Condition 119 *Assume in the following* $r_n = 4D_n/v_n^2$ *and* $r_n < 1/\left(2\sqrt{v_n}\right)$.

Now we use our knowledge about the location of critical points of $K_n(z)$ to investigate how it behaves on the real interval around zero.

Lemma 120 *Suppose* $m_n > 4/\sqrt{v_n}$ *and* $D_n/v_n^{3/2} \leq 1/8$. *Then* $K_n(z)$ *maps the interval* $\left[-1/\sqrt{v_n} + r_n; 1/\sqrt{v_n} - r_n\right]$ *in a one-to-one fashion on the set that contains the union of two intervals* $\left(-\infty, -2\sqrt{v_n} - cD_n/v_n\right) \cup \left(2\sqrt{v_n} + cD_n/v_n, \infty\right)$, *where* c *is a constant that does not depend on* n.

Remark: For example, $c = 8$ will work.

Proof: The assumption that $m_n > 4/\sqrt{v_n}$ ensures that the series $\widetilde{K}_n(z)$ converges in $|z| \leq 4/\sqrt{v_n}$, $z \neq 0$. Note that $K_n(z)$ is real-valued on the set $I = \left[-1/\sqrt{v_n} + r_n, 0\right) \cup \left(0, 1/\sqrt{v_n} - r_n\right]$, because this set belongs to the area where the series $\widetilde{K}_n(z)$ converges and the coefficients of this series are real. Moreover, by Lemma 118, there is no critical points of $K_n(z)$ on I (i.e., for every $z \in I$, $K_n'(z) \neq 0$), therefore $K_n(z)$ must be strictly monotonic on subintervals $\left[-1/\sqrt{v_n} + r_n, 0\right)$ and $\left(0, 1/\sqrt{v_n} - r_n\right]$. Consequently, $K_n(I) = \left(-\infty, K_n\left(-1/\sqrt{v_n} + r_n\right)\right] \cup \left[K_n\left(1/\sqrt{v_n} - r_n\right), \infty\right)$. We claim that $K_n\left(1/\sqrt{v_n} - r_n\right) \leq 2\sqrt{v_n} + 8D_n/v_n$ and $K_n\left(-1/\sqrt{v_n} + r_n\right) \geq -2\sqrt{v_n} - 8D_n/v_n$.

Indeed, write

$$K_n(z) = \frac{1}{z} + v_n z + h(z),$$

then

$$K_n\left(\frac{1}{\sqrt{v_n}} - r_n\right) = \sqrt{v_n}\frac{1}{1 - r_n\sqrt{v_n}} + \sqrt{v_n}\left(1 - r_n\sqrt{v_n}\right) + h\left(\frac{1}{\sqrt{v_n}} - r_n\right).$$

According to our assumption $r_n\sqrt{v_n} = 4D_n/v_n^{3/2} < 1/2$. Therefore we can estimate

$$\frac{1}{1 - r_n\sqrt{v_n}} \leq 1 + 2r_n\sqrt{v_n},$$

and

$$K_n\left(\frac{1}{\sqrt{v_n}} - r_n\right) \leq 2\sqrt{v_n} + r_n v_n + \left|h\left(\frac{1}{\sqrt{v_n}} - r_n\right)\right|.$$

We can estimate the last term using Lemma 117 as

$$h\left(\frac{1}{\sqrt{v_n}} - r_n\right) \leq D_n\left|\frac{1}{\sqrt{v_n}}\right|^2 = D_n/v_n.$$

Altogether, after substituting $r_n = 4D_n/v_n^2$ we get

$$K_n\left(\frac{1}{\sqrt{v_n}} - r_n\right) \leq 2\sqrt{v_n} + 8D_n/v_n.$$

Similarly we can derive that

$$K_n\left(-\frac{1}{\sqrt{v_n}} + r_n\right) \geq -2\sqrt{v_n} - 8D_n/v_n.$$

QED.

From the previous Lemma we can conclude that $K_n(z)$ has a differentiable inverse defined on $\left(-\infty, -2\sqrt{v_n} - cD_n/v_n\right) \cup \left(2\sqrt{v_n} + cD_n/v_n, \infty\right)$. We can extend this conclusion to an open complex neighborhood of this interval. This is achieved in the next two lemmas.

Lemma 121 *As in previous Lemma suppose that* $m_n > 4/\sqrt{v_n}$ *and* $D_n/v_n^{3/2} \leq 1/8$. *Let* z *be an arbitrary point of the interval* $\left[-1/\sqrt{v_n} + r_n, 1/\sqrt{v_n} - r_n\right]$. *Then we can find a neighborhood* U_z *of* z *and a neighborhood* W_w *of* $w = K_n(z)$ *such that* K_n *is a one-to-one map of* U_z *on* W_w *and the inverse map* K_n^{-1} *is holomorphic everywhere in* W_w.

Proof: Since the series $\widetilde{K}_n(z) - z^{-1}$ converges in $|z| \leq 4/\sqrt{v_n}$, function $K_n(z)$ is holomorphic in $|z| \leq 4/\sqrt{v_n}$, $z \neq 0$. In addition, by Lemma 118, if

$$z_0 \in [-1/\sqrt{v_n} + r_n, 1/\sqrt{v_n} - r_n]$$

then z_0 is not a critical point of $K_n(z)$. Therefore for $z \neq 0$, the conclusion of the lemma follows from Theorems II.3.1 and II.3.2 in Markushevich (1977). For $z = 0$ the argument is parallel to the argument in Markushevich except for a different choice of local coordinates: Indeed, $f(z) = 1/K_n(z)$ is holomorphic at $z = 0$, it maps $z = 0$ to $w = 0$, and $f'(z) = 1 \neq 0$ at $z = 0$. Therefore, Theorems II.3.1 and II.3.2 in Markushevich (1977) are applicable to $f(z)$ and it has a well defined holomorphic inverse in a neighborhood of $w = 0$. This implies that $K_n(z)$ has a well-defined holomorphic inverse in a neighborhood of ∞, given by the formula $K_n^{-1}(z) = f^{-1}(1/z)$. QED.

Lemma 122 *Local inverses $K_n^{-1}(z)$ defined in the previous lemma are restriction of a function $G_n(z)$ which is defined and holomorphic everywhere in a neighborhood of $I = \{\infty\} \cup \left(-\infty, -2v_n^{1/2} - cD_n/v_n\right] \cup \left[2v_n^{1/2} + cD_n/v_n, \infty\right)$. The function $G_n(z)$ is the inverse of $K_n(z)$ in this neighborhood.*

Proof: By Lemma 120, for every point $w \in I$ we can find a unique

$$z \in [-1/\sqrt{v_n} + r_n, 1/\sqrt{v_n} - r_n]$$

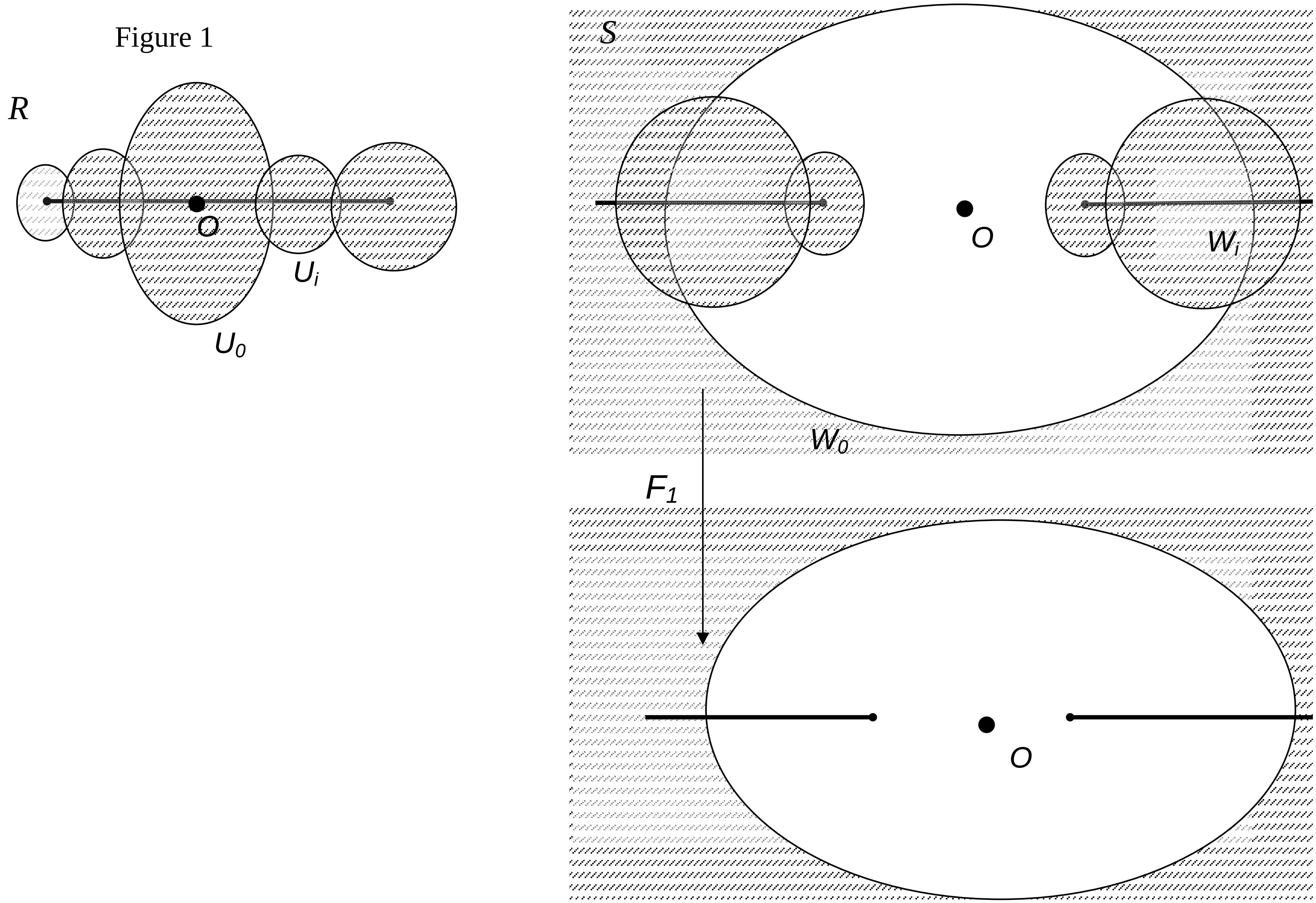
such that $K_n(z) = w$. Let U_z and W_w be the neighborhoods defined in the previous lemma. Also let us write (W_w, K_n^{-1}) to denote the local inverses defined in the previous lemma together with their areas of definition. Our task is to prove that these local inverses can be joined to form a function defined everywhere in a neighborhood of I. We will do it in several steps.

First, an examination of the proof of the previous lemma and Theorem II.3.1 in Markushevich (1977) shows that we can take each U_z in the form of a disc. Then, let $\widetilde{U}_z = U_z/3$, that is, define $\widetilde{U}_z$ as a disc that has the same center but 3 times smaller radius than U_z. Define $\widetilde{W}_w$ as $K_n\left(\widetilde{U}_z\right)$. These new sets are more convenient because of the following property: If $\widetilde{U}_{z_1} \cap \widetilde{U}_{z_2} \neq \emptyset$, then either $\widetilde{U}_{z_1} \cup \widetilde{U}_{z_2} \subset U_{z_1}$ or $\widetilde{U}_{z_1} \cup \widetilde{U}_{z_2} \subset U_{z_2}$. In particular, this means that if $\widetilde{U}_{z_1} \cap \widetilde{U}_{z_2} \neq \emptyset$ then $K_n(z)$ is a one-to-one map of $\widetilde{U}_{z_1} \cup \widetilde{U}_{z_2}$ on $\widetilde{W}_{w_1} \cup \widetilde{W}_{w_2}$. This is convenient because K_n is one-to-one not only on a particular neighborhood $\widetilde{U}_{z_1}$ but also on the union of every two intersecting neigborhoods $\widetilde{U}_{z_1}$ and $\widetilde{U}_{z_2}$. Let us call this *extended invertibility* property.

Next define even smaller $\tilde{\tilde{U}}_z$ with the following properties: 1) $\tilde{\tilde{U}}_z \subset \widetilde{U}_z$; 2) $\widetilde{\widetilde{W}}_w =: K_n\left(\tilde{\tilde{U}}_z\right)$ is either an open disk for $z \neq 0$ or the set $|w| > R$ for $z = 0$; and 3) $0 \notin \widetilde{\widetilde{W}}_w$. This is easy to achieve by taking an appropriate open subset of $\widetilde{W}_w$ as $\widetilde{\widetilde{W}}_w$ and applying K_n^{-1}. Note that the property of the previous paragraph remains valid for the new sets $\tilde{\tilde{U}}_z$.

Discs $\widetilde{\widetilde{W}}_w$ form an open cover of I and the corresponding sets $\tilde{\tilde{U}}_z$ form an open cover for $K_n^{-1}(I)$, which is a closed interval contained in $\left[-1/\sqrt{v_n} + r_n, 1/\sqrt{v_n} - r_n\right]$. Let U_i, $i = 0, ..., N$, be a finite cover of $K_n^{-1}(I)$, selected from $\left\{\tilde{\tilde{U}}_z\right\}$. (We can do it because of compactness of $K_n^{-1}(I)$.) And let $W_i =: K_n(U_i)$ be the corresponding cover of I, selected from $\left\{\widetilde{\widetilde{W}}_z\right\}$. For convenience, let W_0 denote the set $\widetilde{\widetilde{W}}_w$ for $w = \infty$. Finally let $R = \cup_{i=0}^{N} U_i$ and $S = \cup_{i=0}^{N} W_i$. Sets R and S are illustrated in Figure 1.

Figure 1



Clearly S is open. We aim to prove that S is simply connected in the extended complex plane $\mathbb{C} \cup \{\infty\}$. For this purpose let us define the deformation retraction F_1 of the set S by the formula: 1) if $z \in W_0$ then $z \to z$; 2) if $z \notin W_0$ then $z \to \operatorname{Re} z + (1-t) \operatorname{Im} z$. Here parameter t changes from 0 to 1. (For the definition and properties of deformation retractions see, e.g., Hatcher (2002), the definition is on page 2 and the main property is in Proposition 1.17.) This retraction reduces S to a homotopically equivalent set S' that consists of W_0 and two intervals of the real axis that do not include 0. After that we can use another deformation retraction F_2 that sends z to $(1-t)^{-1}z$. This retraction reduces S' to $S'' = \{\infty\}$ which is evidently simply-connected.

We know that there is a holomorphic inverse $K_n^{-1}(z)$ defined on each of W_i. Starting from one of these domains, say W_0, we can analytically continue $K_n^{-1}(z)$ to every other W_i. Indeed, take a point $z_0 \in U_0$ and $z_i \in U_i$ and connect them by a path that lies entirely in $R = \cup_{i=0}^{N} U_i$. This path corresponds to a chain $\{U_{k_s}\}$, $s = 1, .., n$ that connects U_0 and U_i. That is, $U_{k_1} = U_0$, $U_{k_n} = U_i$, and $U_{k_j} \cap U_{k_{j+1}} \neq \varnothing$. The corresponding $\{W_{k_s}\}$ form a chain that connects W_0 and W_j, that is, $W_{k_1} = W_0$, $W_{k_n} = W_i$, and $W_{k_j} \cap W_{k_{j+1}} \neq \varnothing$. By its definition, this chain of sets W_{k_s} has also a specific property that $K_n^{-1}\left(W_{k_j}\right) \cap K_n^{-1}\left(W_{k_{j+1}}\right) = U_{k_j} \cap U_{k_{j+1}} \neq \varnothing$.

Consider two adjacent sets, W_{k_j} and $W_{k_{j+1}}$, in this chain. Then the corresponding local inverses $\left(W_{k_j}, K_n^{-1}\right)$ and $\left(W_{k_{j+1}}, K_n^{-1}\right)$, which were defined in the previous lemma, coincide on an open non-empty set. Indeed, $K_n\left(U_{k_j} \cap U_{k_{j+1}}\right)$ is an open and non-empty set. Since $K_n\left(U_{k_j} \cap U_{k_{j+1}}\right) \subset K_n\left(U_{k_j}\right) \cap K_n\left(U_{k_{j+1}}\right) = W_{k_j} \cap W_{k_{j+1}}$, functions $\left(W_{k_j}, K_n^{-1}\right)$ and $\left(W_{k_{j+1}}, K_n^{-1}\right)$ are well defined on $K_n\left(U_{k_j} \cap U_{k_{j+1}}\right)$. Moreover, they must coincide on $K_n\left(U_{k_j} \cap U_{k_{j+1}}\right)$.

Indeed, by construction $U_{k_j} \cap U_{k_{j+1}} \neq \emptyset$ and, therefore, by the extended invertibility property, K_n is one-to-one on $U_{k_j} \cup U_{k_{j+1}}$. Hence there cannot exist two different z and $z' \in U_{k_j} \cup U_{k_{j+1}}$ that would map to one point in $K_n\left(U_{k_j} \cap U_{k_{j+1}}\right)$. Hence $\left(W_{k_j}, K_n^{-1}\right)$ and $\left(W_{k_{j+1}}, K_n^{-1}\right)$ must coincide on $K_n\left(U_{k_j} \cap U_{k_{j+1}}\right)$.

Using the property that if tow analytical functions coincide on an open set then each of them is an analytic continuation of the other, we conclude that the local inverse $\left(W_{k_j}, K_n^{-1}\right)$ can be analytically continued to $W_{k_{j+1}}$ where it coincides with the local inverse $\left(W_{k_{j+1}}, K_n^{-1}\right)$. Therefore, at least one analytic continuation of $\left(W_0, K_n^{-1}\right)$ is well-defined everywhere on S and has the property that when restricted to each of W_j it coincides with a local inverse of $K_n(z)$ defined in the previous lemma. Since S is simply connected, the analytic continuation is unique, that is, it does not depend on the choice of the chain of the neighborhoods that connect W_0 and W_j.

Let us denote the function resulting from this analytic continuation as $G_n(z)$. By construction, it is unambiguously defined for every W_j and the restrictions of $G_n(z)$ to W_j coincide with K_n^{-1}. Therefore, $G_n(z)$ satisfies the relations $K_n(G_n(z)) = z$ and $G_n(K_n(z)) = z$ everywhere on $R = \cup_{i=0}^N U_i$ and on $S = \cup_{i=0}^N W_i$. Every claim of the lemma is proved because S is an open neighborhood of I.

QED.

Lemma 123 *The function $G_n(z)$ constructed in the previous lemma is the Cauchy transform of S_n.*

By construction, $G_n^{-1}(z)$ is the inverse of $K_n(z)$ in a neighborhood of $\{\infty\} \cup \left(-\infty, -2v_n^{1/2} - cD_n/v_n\right) \cup \left(2v_n^{1/2} + cD_n/v_n, \infty\right)$. In particular, it is inverse of $K_n(z)$ in a neighborhood of infinity. Therefore, in this neighborhood it has the same power expansion as the Cauchy transform of S_n. Therefore, it coincides with the Cauchy transform of S_n in this neighborhood. Next we apply the principle that if two analytical functions coincide in an open domain then they coincide at every point where they can be continued analytically. QED.

Now it remains to apply Lemma 56 and we obtain the following Theorem.

Theorem 124 *Suppose that $i)$ $\liminf m_n\sqrt{v_n} > 4$, and $ii)$ $\limsup_{n\to\infty} D_n/v_n^{3/2} \leq 1/8$. Then for all sufficiently large n the support of μ_n belongs to*

$$I = \left(-2\sqrt{v_n} - cD_n/v_n, 2\sqrt{v_n} + cD_n/v_n\right),$$

where $c > 0$ is an absolute constant (e.g. $c = 8$).

Proof of Theorem 124. Let us collect the facts that we know about $G_n(z)$ that was defined in Lemma 122. First, by Lemma 123 it is the Cauchy transform of a bounded random variable S_n. Second, by Lemma 122 it is holomorphic at $z \in \mathbb{R}$, $|z| > 2v_n^{1/2} + cD_n/v_n$. Using Lemma 56 we conclude that the distribution of S_n is supported on the interval $\left[-2v_n^{1/2} - cD_n/v_n, 2v_n^{1/2} + cD_n\right]$. QED.

If we take $R_{n,i} = 2L_{n,i}$ and $m_{n,i} = (4L_{n,i})^{-1}$, then assumption i) is equivalent to

$$\liminf_{n\to\infty} \min_i \frac{\sqrt{v_n}}{4L_{n,i}} > 4,$$

which is equivalent to

$$\limsup_{n\to\infty} \frac{L_n}{\sqrt{v_n}} < 16.$$

From ii) we get

$$1/8 \geq \limsup_{n\to\infty} \frac{\sum_{i=1}^{k_n} R_{n,i} \left(m_{n,i}\right)^{-2}}{v_n^{3/2}} = \limsup_{n\to\infty} \frac{32 \sum_{i=1}^{k_n} L_{n,i}^3}{v_n^{3/2}},$$

which is equivalent to

$$\limsup_{n\to\infty} \frac{T_n}{v_n^{3/2}} \leq 1/256.$$

Finally note that the condition $\limsup_{n\to\infty} T_n/v_n^{3/2} \leq 2^{-12}$ implies that

$$\limsup_{n\to\infty} L_n/\sqrt{v_n} < 16.$$

Therefore, Theorem 110 is a consequence of Theorem 124.

Part III

Limit Theorems for Products of Free Operators

14 Norms of Products of Free Random Variables

14.1 Introduction

Suppose $X_1, X_2, \ldots, X_n$ are identically-distributed free random variables. These variables are infinite-dimensional linear operators but the reader may find it convenient to think of them as very large random matrices. The first question we will address in this paper is how the norm of $\Pi_n = X_1 X_2 \ldots X_n$ behaves. If X_i are all positive, then it is natural to look also at the symmetric product operation $\circ$ defined as follows: $X_1 \circ X_2 = X_1^{1/2} X_2 X_1^{1/2}$. The benefit is that unlike the usual operator product, this operation maps the set of positive variables to itself. For this operation we can ask how the norm of symmetric products $Y_n = X_1 \circ X_2 \circ \ldots \circ X_n$ behaves.[3]

Products of random matrices and their asymptotic behavior were originally studied by Bellman (1954). One of the decisive steps was made by Furstenberg and Kesten (1960), who investigated a matrix-valued stationary stochastic process $X_1, \ldots, X_n, \ldots$, and proved that the limit of $n^{-1} E(\log \|X_1 \ldots X_n\|)$ exists (but might equal $\pm\infty$) and that under certain assumptions $n^{-1} \log \|X_1 \ldots X_n\|$ converges to this limit almost surely. Essentially, the only facts that are used in the proof of this result are the ergodic theorem, the norm inequality $\|X_1 X_2\| \leq \|X_1\| \|X_2\|$ and the fact that the unit sphere is compact in finite-dimensional spaces. It is the lack of compactness of the unit sphere in the infinite-dimensional space that makes generalizations to infinite-dimensional operators non-trivial (see Ruelle (1982) for a generalization in the case of compact operators). More work on non-commutative products was done

[3] The operation $\circ$ is neither commutative, nor associative. By convention we multiply starting on the right, so, for example, $X_1 \circ X_2 \circ X_3 \circ X_4 = X_1 \circ (X_2 \circ (X_3 \circ X_4))$. However, this convention is not important for the question that we ask. First, it is easy to check that $X_1 \circ X_2$ has the same spectral distribution and therefore the same norm as $X_2 \circ X_1$. Second, if X_1, X_2, and X_3 are free, then the spectral distribution of $(X_1 \circ X_2) \circ X_3$ is the same as the spectral distribution of $X_1 \circ (X_2 \circ X_3)$, and therefore these two products have the same norm. In brief, if X_i are free, then the norm of $X_1 \circ X_2 \circ \ldots \circ X_n$ does not depend on the order in which X_i are multiplied by the operation $\circ$.

by Furstenberg (1963), Oseledec (1968), Kingman (1973), and others. The results are often called multiplicative ergodic theorems and they find many applications in mathematical physics. For example, see Ruelle (1984).

In this paper, we study products of free random variables. These variables are (non-compact) infinite-dimensional operators which can be thought of as a limiting case of large independent random matrices.

Suppose that X_i are free, identically-distributed, self-adjoint, and positive. Suppose also $E(X_i) = 1$. Then we show that the norm of $Y_n = X_1 \circ X_2 \circ ... \circ X_n$ grows no faster than a linear function of n. Precisely, we find that

$$\limsup_{n\to\infty} n^{-1} \|Y_n\| \leq c_1 \|X_i\| .$$

We are also able to show that if X_i is not concentrated at 1, then

$$\liminf_{n\to\infty} n^{-1/2} \|Y_n\| \geq c_2 > 0.$$

For the usual products $\Pi_n = X_1 X_2 ... X_n$ we can relax the assumption of self-adjointness. So, suppose that X_i are free and identically-distributed but not necessarily self-adjoint. Also, we do not require that $E(X_i) = 1$. Then we show that

$$\lim_{n\to\infty} n^{-1} \log \|\Pi_n\| = \log \sqrt{E(X_i^* X_i)}. \tag{60}$$

Another way to describe the behavior of Π_n is to look at how the norm of a fixed vector ξ changes when we consecutively apply free operators X_1, ..., X_n to it. More precisely, suppose that the action of the algebra of variables X_i on a Hilbert space H is described by a cyclic representation π and that the vector ξ is cyclic with respect to the expectation E. By definition, this means that $E(X) = \langle \xi, \pi(X)\xi \rangle$ for every operator X from a given algebra. Then we show that

$$n^{-1} \log \|\pi(\Pi_n)\xi\| = \log \sqrt{E(X_i^* X_i)}. \tag{61}$$

Note that we do not need to take the limit, since the equality holds for all n.

The reader may think of cyclic vectors as typical vectors. For example, if the representation π is cyclic and irreducible then cyclic vectors are dense in H. In colloquial terms, (60) says that for large n the product Π_n cannot increase the norm of any given vector ξ by more than $[E(X^* X)]^{n/2}$. And (61) says that for every cyclic vector ξ this growth rate is achieved.

One more way to capture the intuition of this result is to write

$$\lim_{n\to\infty} n^{-1} \log \|\Pi_n\| = \lim_{n\to\infty} n^{-1} \log \sup_{\|x\|=1} \|\pi(\Pi_n) x\|$$

We have shown that this limit is equal to

$$n^{-1} \log \|\pi(\Pi_n) \xi\|$$

where ξ is a cyclic vector. Thus, for large n the product Π_n acts uniformly in all directions. Its maximal dilation as measured by $\sup_{\|x\|=1} \|\pi(\Pi_n) x\|$ has the same exponential order of magnitude as the dilation in the direction of a typical vector ξ.

It is helpful to compare these results with the case of commutative random variables. Suppose for the moment that X_i are independent commutative random variables with positive values. Then,

$$\lim_{n\to\infty} n^{-1} \log \|X_1 ... X_n\| = \log \|X_i\|,$$

where the norm of a random variable is the essential supremum norm (i.e., $\|X\| = \operatorname{ess\,sup}_{\omega\in\Omega} |X(\omega)|$). Indeed, for every $\varepsilon > 0$ the measure of the set

$$\{\omega : |X_1(\omega) ... X_n(\omega)| \geq \|X_1\| ... \|X_n\| - \varepsilon\}$$

is positive. Therefore $\|X_1 ... X_n\| = \|X_1\|^n$. Note that $\log \sqrt{E(X_i^* X_i)} \leq \log \|X_i\|$ and therefore the norm of free products grows more slowly than we would expect from the classical case.

Another interesting comparison is that with results about products of random matrices. Let X_i be i.i.d. random $k \times k$ matrices. Then under suitable conditions,

$$\lim_{n\to\infty} n^{-1} \log \|X_n ... X_1\|$$

exists almost surely. Let us denote this limit as λ. Furstenberg (1963) developed a general formula for λ, and Cohen and Newman (1984) derived explicit results in the case when entries of X_i have a joint Gaussian distribution. In particular, if all entries of X_i are independent and have the distribution $\mathcal{N}(0, s_k^2)$ then

$$\lambda = (1/2) \left\{\log\left(s_k^2\right) + \log 2 + \psi(k/2)\right\}$$

where ψ is the digamma function ($\psi(x) = d \log \Gamma(x)/dx$). If the size of the matrices grows ($k \to \infty$) then $\lambda \sim (1/2) \log\left(k s_k^2\right)$. To compare this with our results, note that if $ks_k^2 \to s^2$, then the sequence of random matrices approximates a free random variable $\widetilde{X}_i$ with the spectral distribution that is uniform inside the circle of radius s. For this free variable, $E\left(\widetilde{X}_i^* \widetilde{X}_i\right) = s^2$, and our theorem shows that $\lim_{n\to\infty} n^{-1} \left\| \widetilde{X}_1 ... \widetilde{X}_n \right\| = \log s$. This limit agrees with the result for random matrices. Thus, our result can be seen as a limiting form of results for random matrices.

The results regarding $\|Y_n\|$ are also interesting. We can associate with X_i and Y_n probability measures μ_X and μ_{Y_n}, which are called the spectral probability measures of X_i and Y_n, respectively. Then the measure μ_{Y_n} is determined only by n and the measure μ_X and is called the n-time *free multiplicative convolution* of μ_X with itself:

$$\mu_{Y_n} = \underbrace{\mu_X \boxtimes ... \boxtimes \mu_X}_{n \text{ times}}.$$

The norm $\|Y_n\|$ is easy to interpret in terms of the distribution μ_{Y_n}. Indeed, it is the smallest number t such that the support of μ_{Y_n} is inside the interval $[0, t]$. Therefore, the growth in $\|Y_n\|$ measures the growth in the support of the spectral probability measure if the measure is convolved with itself using the operation of the free multiplicative convolution.

In the case of classical multiplicative convolutions of probability measures, the support grows exponentially, so that if μ_X is supported on $[0, L_X]$, then the measure $\mu_{X_1 ... X_n}$ is supported on $[0, (L_X)^n]$. What we have found in the case of free multiplicative convolutions is that if we fix $EX_i = 1$, then the support of the μ_{Y_n} grows no faster than a linear function of n, i.e., the support of μ_{Y_n} is inside the interval $[0, cnL_x]$ with an absolute constant c.

As was pointed out in the literature, a similar phenomenon occurs for sums of free random variables. The support of measures obtained by free additive convolutions grows much more slowly than in the case of classical additive convolutions. This effect was called superconvergence by Bercovici and Voiculescu (1995). Our finding about $\|Y_n\|$ can be considered as a superconvergence for free multiplicative convolutions.

The rest of the paper is organized as follows. Section 2 formulates the results. Section 3 contains the necessary technical background from free probability theory. Sections 4, 5, and 6 prove the results. And Section 7 concludes.

14.2 Results

Let X_1, X_2, ... , X_n be free identically-distributed positive random variables. Consider $\Pi_n = X_1 X_2 ... X_n$ and $Y_n = X_1 \circ X_2 \circ ... \circ X_n$ (by convention we multiply on the left, so that, for example, $X_1 \circ X_2 \circ X_3 \circ X_4 = X_1 \circ (X_2 \circ (X_3 \circ X_4))$). We will see later that these variables have the same moments: $E(\Pi_n)^k = E(Y_n)^k$. As a first step let us record some simple results about the expectation and variance of Y_n and Π_n. We define *variance* of a random variable A as

$$\sigma^2(A) =: E\left(A^2\right) - [E(A)]^2 .$$

Proposition 125 *Suppose that X_i are self-adjoint and $E(X_i) = 1$. Then $E(\Pi_n) = E(Y_n) = 1$ and $\sigma^2(\Pi_n) = \sigma^2(Y_n) = n\sigma^2(X_i)$.*

Note that the linear growth in the variance of $\Pi_n = X_1 ... X_n$ is in contrast with the classical case, where only the variance of $\log(X_1 ... X_n)$ grows linearly. We will prove this Proposition later when we have more technical tools available. Before that we are going to formulate the main results.

Let $\|A\|$ denote the usual operator norm of operator A.

Theorem 126 *Suppose that X_1, ..., X_n are identically-distributed positive self-adjoint free variables. Suppose also that $E(X_i) = 1$. Then*
(1) there exists such a constant, c, that $\|Y_n\| \leq c \|X_i\| n$;
and
(2) $\|Y_n\| \geq \sigma(X_i) \sqrt{n}$.

For the next theorem define

$$\gamma = \sigma\left(\frac{X_i^* X_i}{E(X_i^* X_i)}\right) \geq 0$$

Theorem 127 *Suppose that X_1, ..., X_n are free identically-distributed variables (not necessarily self-adjoint). Then*
(1) there exists such a constant, c, that $\|\Pi_n\| \leq c \|X_i\| \sqrt{n} \left[E(X_i^* X_i)\right]^{(n-1)/2}$;
and
(2) $\|\Pi_n\| \geq \gamma^{1/2} n^{1/4} \left[E(X_i^* X_i)\right]^{n/2}$.

Corollary 128 *Suppose that X_1, ..., X_n are free identically-distributed variables (not necessarily self-adjoint). Then*

$$\lim_{n\to\infty} n^{-1} \log \|\Pi_n\| = \log \sqrt{E(X_i^* X_i)}$$

Next, suppose that the algebra $\mathcal{A}$ acts on an (infinitely-dimensional) Hilbert space H. In other words, let π be a representation of $\mathcal{A}$. We call representation π *cyclic* if there exists such a vector $\xi \in H$ that $E(X) = \langle \xi, \pi(X)\xi \rangle$ for all operators $X \in \mathcal{A}$. The vectors with this property are also called *cyclic*.

Theorem 129 *Suppose π is a cyclic representation of $\mathcal{A}$, ξ is its cyclic vector, and X_1, ..., X_n are free identically-distributed variables from $\mathcal{A}$. Then*

$$n^{-1} \log \|\pi(\Pi_n)\xi\| = \log \sqrt{E(X_i^* X_i)}$$

Corollary 130 *If π and ξ are cyclic then*

$$\log \|\Pi_n\| \sim \log \|\pi(\Pi_n)\xi\| \sim n \log \|\pi(X_1)\xi\|$$

as $n \to \infty$.

14.3 Preliminaries

Let us write out several first terms in the power expansions for $\psi(z)$, $\psi^{-1}(z)$, and $S(z)$. Suppose for simplicity that $E(A) = 1$ and let $E\left(A^k\right) = m_k$. Then,

$$\begin{aligned}
\psi(z) &= z + m_2 z^2 + m_3 z^3 + ..., \\
\psi^{-1}(z) &= z - m_2 z^2 - \left(m_3 - 2m_2^2\right) z^3 + ..., \\
S(z) &= 1 + (1 - m_2) z + \left(2m_2^2 - m_2 - m_3\right) z^2 + ...
\end{aligned}$$

The Voiculescu theorem about multiplication of random variables implies that $S_{\Pi_n} = S_{Y_n} = (S_X)^n$, where S_X denotes the S-transform of any of X_i. Now it is easy to prove Proposition 125. Indeed, let us denote S_{Π_n} as S_n. Then, using the power expansions we can write:

$$\begin{aligned}
S_n(z) &= 1 + \left(1 - m_2^{(n)}\right) z + ... \\
&= (S_X)^n = 1 + n(1 - m_2) z + ...,
\end{aligned}$$

where $m_2^{(n)} =: E(\Pi_n)^2$ and $m_2 =: E(X_i)^2$. Then, using power expansion in (21), we conclude that $E(\Pi_n) = 1$. Next, by definition $\sigma^2(X_i) = m_2 - 1$ and $\sigma^2(\Pi_n) = m_2^{(n)} - 1$. Therefore, we can conclude that $\sigma^2(\Pi_n) = n\sigma^2(X)$. QED.

14.4 Proofs

Proof of Theorem 126

Throughout this section we assume that X_i are self-adjoint, $E(X_i) = 1$, and the support of the spectral distribution of X_i belongs to $[0, L]$.

Let us first go in a simpler direction and derive a lower bound on $\|Y_n\|$. That is, we are going to prove claim (2) of the theorem. From Proposition 125, we know that $E(Y_n) = 1$ and $\sigma^2(Y_n) = n\sigma^2(X_i)$. It is clear that for every positive random variable A, it is true that $E(A^2) \leq \|A\|^2$ and therefore $\|A\| \geq \sqrt{\sigma^2(A) + [E(A)]^2}$. Applying this to Y_n, we get $\|Y_n\| \geq \sqrt{n\sigma^2 + 1}$. In particular, $\|Y_n\| > \sigma\sqrt{n}$, so (2) is proved.

Now let us prove claim (1). By Theorem 49, $S_n(z) = (S_X(z))^n$. The idea of the proof is to investigate how $|S_X(z)|^n$ behaves for small z. It turns out that if z is of the order of n^{-1}, then $|S_X(z)|^n > c$ where c is a constant that does not depend on n. We will show that this fact implies that $\psi_n(z)$ (i.e., the ψ-function for Y_n) has the convergent power series in the area $|z| < (cn)^{-1}$ and that therefore the Cauchy transform of Y_n has the convergent power series in $|z| > cn$. This fact and the Perron-Stieltjes inversion formula imply that the support of the distribution of Y_n is inside $[-cn, cn]$.

Lemma 131 $E(X^k) \leq L^{k-1}$.

Proof:

$$E(X^k) = \int_0^L \lambda^k d\mu_X(\lambda) \leq L^{k-1} \int_0^L \lambda d\mu_X(\lambda) = L^{k-1},$$

where $d\mu_X$ denotes the spectral distribution of the variable X. QED.

Lemma 132 *The function $\psi_X(z)$ is one-to-one in $|z| \leq (4L)^{-1}$ and if $|z| = (4L)^{-1}$, then $|\psi_X(z)| \geq (6L)^{-1}$.*

Proof: If $|z| \leq (4L)^{-1}$ then

$$\begin{aligned} |\psi_X(z) - z| &\leq |z| \sum_{k=2}^{\infty} E(X^k) |z|^{k-1} \\ &\leq |z| \sum_{k=1}^{\infty} \frac{1}{4^k} = \frac{|z|}{3}. \end{aligned}$$

Therefore, $\psi_X(z)$ is one-to-one in this area.

If $|z| = (4L)^{-1}$, then

$$\begin{aligned}
|\psi_X(z)| &\geq |z| - \sum_{k=2}^{\infty} E\left(X^k\right)|z|^k \\
&\geq |z| \left(1 - \sum_{k=1}^{\infty} \frac{1}{4^k}\right) \\
&= \frac{1}{4L}\left(1 - \frac{1}{3}\right) = \frac{1}{6L}.
\end{aligned}$$

QED.

By Lemma 60, we can expand the functional inverse of $\psi_X(z)$ as follows:

$$\psi_X^{-1}(u) = u + \sum_{k=2}^{\infty} c_k u^k,$$

where

$$c_k = \frac{1}{2\pi i k}\int_\gamma \frac{dz}{[\psi_X(z)]^k}$$

Lemma 133 *If* $|u| \leq (72Ln)^{-1}$, *then*

$$\left|\frac{\psi_X^{-1}(u)}{u} - 1\right| \leq \frac{1}{7n}.$$

Proof: Using the previous lemma we can estimate c_k:

$$c_k \leq \frac{1}{k}\frac{1}{4L}(6L)^k \leq \frac{3}{2}(6L)^{k-1}.$$

Then

$$\begin{aligned}
\left|\frac{\psi_X^{-1}(u)}{u} - 1\right| &= \left|\sum_{k=2}^{\infty} c_k u^{k-1}\right| \\
&\leq \frac{3}{2}\sum_{k=1}^{\infty}\left(\frac{1}{12n}\right)^k = \frac{3}{2}\frac{1}{12n-1} \\
&= \frac{3}{2}\frac{12n}{12n-1}\frac{1}{12n} \leq \frac{1}{7n},
\end{aligned}$$

provided that $|u| \leq (72Ln)^{-1}$. QED.

Lemma 134 *If* $|u| \leq (72Ln)^{-1}$ *, then*

$$|1 - S_X(u)| \leq \frac{1}{6n}.$$

Proof: Recall that $S_X(u) = (1+u)\psi_X^{-1}(u)/u$. Then we can write:

$$\begin{aligned} |1 - S_X(u)| &= \left| u + (1+u)\left(\frac{\psi_X^{-1}(u)}{u} - 1\right)\right| \\ &\leq |u| + |1+u|\left|\frac{\psi_X^{-1}(u)}{u} - 1\right|. \end{aligned}$$

Then the previous lemma implies that for $|u| \leq (72Ln)^{-1}$ and $n \geq 2$, we have the estimate:

$$|1 - S_X(u)| \leq \frac{1}{72Ln} + \left|1 + \frac{1}{72Ln}\right|\frac{1}{7n}.$$

Note that $L \geq 1$ because $EX = 1$. Therefore,

$$|1 - S_X(u)| \leq \frac{1}{72n} + \frac{73}{72}\frac{1}{7n} \leq \frac{1}{6n}.$$

QED.

Lemma 135 *For all positive integer* n *if* $|u| \leq (72Ln)^{-1}$ *, then*

$$e^{1/6} \geq |S_X(u)|^n \geq e^{-1/3}.$$

Proof: Let us first prove the upper bound on $|S_X(u)|^n$. The previous lemma implies that

$$|S_X(u)|^n \leq \left(1 + \frac{1}{6n}\right)^n \leq e^{1/6}.$$

Now let us prove the lower bound. The previous lemma implies that

$$|S_X(u)|^n \geq \left(1 - \frac{1}{6n}\right)^n.$$

In an equivalent form,

$$n \log|S_X(u)| \geq n \log\left(1 - \frac{1}{6n}\right). \tag{62}$$

Recall the following elementary inequality: If $x \in [0, 1 - e^{-1}]$, then

$$\log(1 - x) \geq -2x.$$

Let $x = 1/(6n)$. Then

$$\log\left(1 - \frac{1}{6n}\right) \geq -\frac{1}{3n}.$$

Substituting this in (62), we get

$$n \log |S_X(u)| \geq -\frac{1}{3},$$

or

$$|S_X(u)|^n \geq e^{-1/3}.$$

QED.

By Theorem 49, $S_n(u) =: [S_X(u)]^n$ is the S-transform of the variable Y_n. The corresponding inverse ψ-function is $\psi_n^{-1}(u) = uS_n(u)/(1+u)$.

First, we estimate $S_n(u) - 1$.

Lemma 136 *If* $|u| \leq (72Ln)^{-1}$, *then*

$$|S_n(u) - 1| \leq \frac{1}{5}.$$

Proof: Write

$$\begin{aligned} |S_X(u)^n - 1| &\leq |S_X(u) - 1| \left(|S_X(u)|^{n-1} + |S_X(u)|^{n-2} + \dots + 1\right) \\ &\leq \frac{1}{6n} e^{1/6} n \leq \frac{1}{5}. \end{aligned}$$

QED.

Lemma 137 *The function* $\psi_n^{-1}(u)$ *is one-to-one in* $|u| = (72Ln)^{-1}$ *and if* $|u| = (72Ln)^{-1}$, *then*

$$\left|\psi_n^{-1}(u)\right| \geq \frac{1}{102Ln}.$$

Proof: Recall that by definition in (21), $\psi_n^{-1}(u) = uS_n(u)/(1+u)$. Therefore,

$$\left|\psi_n^{-1}(u) - u\right| = |u| \left|\frac{S_n(u) - (1+u)}{1+u}\right|$$

and by Lemma 136 we have the following estimate:

$$\begin{aligned} \left|\frac{S_n(u) - (1+u)}{1+u}\right| &\leq \frac{1}{1-|u|} |S_n(u) - 1| + \frac{|u|}{1-|u|} \\ &\leq \frac{72}{71}\frac{1}{5} + \frac{1}{71} \leq \frac{1}{4}. \end{aligned}$$

Therefore, $\psi_n^{-1}(u)$ is invertible in $|u| \leq (72Ln)^{-1}$.

Next, note that $\psi_n^{-1}(u) = uS_n(u)/(1+u)$ and if $|u| = (72Ln)^{-1}$, then

$$\left|\frac{u}{1+u}\right| \geq \frac{1}{72Ln} / \left(1 + \frac{1}{72Ln}\right) \geq \frac{1}{73Ln}.$$

Using Lemma 135, we get:

$$\left|\psi_n^{-1}(u)\right| \geq \frac{1}{73Ln} e^{-1/3} \geq \frac{1}{102Ln}.$$

QED.

Now we again apply Lemma 60 and obtain the following formula:

$$\psi_n(z) = z + \sum_{k=2}^{\infty} \left[\frac{1}{2\pi i k} \oint_\gamma \frac{du}{\left[\psi_n^{-1}(u)\right]^k}\right] z^k, \tag{63}$$

where we can take the circle $|u| = (72Ln)^{-1}$ as γ.

Lemma 138 *The radius of convergence of series (63) is at least* $(102Ln)^{-1}$.

Proof: By the previous lemma, the coefficient before z^k can be estimated as follows:

$$|c_k| \leq \frac{1}{k} \frac{1}{72Ln} (102Ln)^k.$$

This implies that series (63) converges at least for $|z| \leq (102Ln)^{-1}$. QED.

Lemma 139 *The support of the spectral distribution of* $Y_n = X_1 \circ X_2 \circ ... \circ X_n$ *belongs to the interval* $[-102Ln, 102Ln]$.

Proof: The variable Y_n is self-adjoint and has a well-defined spectral measure, $\mu_n(dx)$, supported on the real axis. We can infer the Cauchy transform of this measure from $\psi_n(z)$:

$$G_n(z) = z^{-1}\left[\psi_n\left(z^{-1}\right) + 1\right].$$

Using Lemma 138, we can conclude that the power series for $G_n(z)$ around $z = \infty$ converges in the area $|z| > 102Ln$. The coefficients of this series are real. Therefore, using the Perron-Stieltjes formula we conclude that $\mu_n(dx)$ is zero outside of the interval $[-102Ln, 102Ln]$. QED.

Lemma 139 implies the statement of Theorem 126.

Proof of Theorem 127

The norm of the operator Π_n coincides with the square root of the norm of the operator $\Pi_n^*\Pi_n$. Therefore, all we need to do is to estimate the norm of the self-adjoint operator $\Pi_n^*\Pi_n$.

Lemma 140 *For every bounded operator $X \in \mathcal{A}$, products X^*X and XX^* have the same spectral distributions.*

Proof: Since E is tracial, $E(X^*X)^k = E(XX^*)^k$. Therefore, X^*X and XX^* have the same sequences of moments and, therefore, the same distributions. QED.

If two variables A and B have the same sequences of moments, we say that they are *equivalent* and write $A \sim B$. In particular, if two self-adjoint variables have the same spectral distribution, then they are equivalent. Conversely, if two bounded self-adjoint variables are equivalent, then they have the same spectral distribution.

Lemma 141 *If $A \sim B$, A is free from C, and B is free from C, then $A+C \sim B+C$, $AC \sim BC$, and $CA \sim CB$.*

Proof: Since A and C are free, the moments of $A + C$ can be computed from the moments of A and C. The computation is exactly the same as for $B + C$, since B and C are also free. In addition we know that A and B have the same moments. Consequently, $A + C$ has the same moments as $B + C$, i.e., $A + C \sim B + C$. The other equivalences are obtained similarly. QED.

Lemma 142 *If $A \sim B$, then $S_A(z) = S_B(z)$. In words, if two variables are equivalent, then they have the same S-transform.*

Proof: From the definition of the ψ-function, it is clear that if $A \sim B$, then $\psi_A(z) = \psi_B(z)$. This implies that $\psi_A^{-1}(z) = \psi_B^{-1}(z)$ and therefore $S_A(z) = S_B(z)$. QED.

For example, since X_i are all identically distributed, all $S_{X_i}(z)$ are the same and we will denote this function as $S_X(z)$. Similarly, $S_{X_i^*X_i}(z)$ does not depend on i and we will denote it as $S_{X^*X}(z)$.

Lemma 143 *If X_1, ..., X_n are free then*

$$\Pi_n^*\Pi_n \sim X_n^*X_n...X_1^*X_1$$

and if X_1, ..., X_n are in addition identically distributed then

$$S_{\Pi_n^*\Pi_n} = S_{\Pi_n\Pi_n^*} = (S_{X^*X})^n$$

Proof: We will use induction. For $n = 1$, we have $\Pi_1^*\Pi_1 = X_1^*X_1$. Therefore $S_{\Pi_1^*\Pi_1} = S_{X^*X}$. Suppose that the statement is proved for $n-1$. Then

$$\begin{aligned} \Pi_n^*\Pi_n &= X_n^*...X_1^*X_1...X_n \\ &\sim X_nX_n^*X_{n-1}^*...X_1^*X_1...X_{n-1}, \end{aligned}$$

where the equivalence holds because E is tracial and it is easy to check that the products have the same moments. Therefore,

$$\begin{aligned} \Pi_n^*\Pi_n &\sim (X_nX_n^*)\,\Pi_{n-1}^*\Pi_{n-1} \\ &\sim (X_n^*X_n)\,\Pi_{n-1}^*\Pi_{n-1} \end{aligned}$$

by Lemmas 140 and 141. Then the inductive hypothesis implies that

$$\Pi_n^*\Pi_n \sim X_n^*X_n...X_1^*X_1.$$

Using Lemma 142 and Theorem 49, we write:

$$S_{\Pi_n^*\Pi_n} = (S_{X^*X})^n\,.$$

Since $\Pi_n^*\Pi_n \sim \Pi_n\Pi_n^*$, therefore $S_{\Pi_n^*\Pi_n} = S_{\Pi_n\Pi_n^*} = (S_{X^*X})^n$. QED.

We have managed to represent $S_{\Pi_n^*\Pi_n}$ as $(S_{X^*X})^n$ and therefore all the arguments of the previous section are applicable, except that we are interested in $(S_{X^*X})^n$ rather than in $(S_X)^n$. In particular, we can conclude that the following lemma holds:

Lemma 144 *Define*

$$\gamma = \sigma\left(\frac{X_i^*X_i}{E\left(X_i^*X_i\right)}\right).$$

Then
(1) $\|\Pi_n^*\Pi_n\| \leq 102\,\|X_i\|^2\,nE\left(X_i^*X_i\right)^{n-1}$, *and*
(2) $\|\Pi_n^*\Pi_n\| \geq \gamma\sqrt{n}E\left(X_i^*X_i\right)^n$.

Proof: Let us introduce variables $R_i = s^{-1}X_i$ where $s^2 = E\left(X^*X\right)$. Then $\|R_i^*R_i\| = \left(\|X_i\|/s\right)^2$ and $E\left(R_i^*R_i\right) = 1$. Let $\widetilde{\Pi}_n = R_1...R_n$. Then $\Pi_n^*\Pi_n = s^{2n}\widetilde{\Pi}_n^*\widetilde{\Pi}_n$ and the S-transform of $\widetilde{\Pi}_n^*\widetilde{\Pi}_n$ is $(S_{R^*R})^n$.

Note that $\widetilde{\Pi}_n^*\widetilde{\Pi}_n$ has the same S-transform and therefore the same distribution as $(R_1^*R_1)\circ...\circ(R_n^*R_n)$. Using Theorem 126, we conclude that $\left\|\widetilde{\Pi}_n^*\widetilde{\Pi}_n\right\| \leq 102\left(\|X_i\|/s\right)^2 n$. It follows that $\|\Pi_n^*\Pi_n\| \leq 102\,\|X_i\|^2\,s^{2n-2}n$. In addition, Theorem 126 implies that

$$\begin{aligned} \left\|\widetilde{\Pi}_n^*\widetilde{\Pi}_n\right\| &> \sqrt{n}\sigma\left(R_i^*R_i\right) \\ &= \gamma\sqrt{n} \end{aligned}$$

Consequently,

$$\|\Pi_n^*\Pi_n\| \geq \gamma\sqrt{n}s^{2n}.$$

QED.

From Lemma 144 we conclude that

$$\|\Pi_n\| \leq 11\,\|X_i\|\,\sqrt{n}E\,(X_i^*X_i)^{(n-1)/2}\,,$$

and

$$\|\Pi_n\| \geq \gamma^{1/2}n^{1/4}E\,(X_i^*X_i)^{n/2}$$

This completes the proof of Theorem 127.

Proof of Theorem 129

By definition of the cyclic vector, we have:

$$\begin{aligned}\|\pi\,(\Pi_n)\,\xi\|^2 &= \langle\pi\,(\Pi_n)\,\xi, \pi\,(\Pi_n)\,\xi\rangle \\ &= \langle\xi, \pi\,(\Pi_n^*\Pi_n)\,\xi\rangle \\ &= E\,(\Pi_n^*\Pi_n)\,.\end{aligned}$$

Using Lemma 143, we continue this as follows:

$$\begin{aligned}E\,(\Pi_n^*\Pi_n) &= E\,(X_n^*X_n...X_1^*X_1) \\ &= [E\,(X^*X)]^n\,.\end{aligned}$$

Consequently,

$$n^{-1}\log\|\Pi_n\xi\| = \frac{1}{2}\log E\,(X^*X)\,.$$

QED.

14.5 Conclusion

We have investigated how the norms of $\Pi_n = X_1...X_n$ and $Y_n = X_1 \circ ... \circ X_n$ grow as $n \to \infty$. For $\|\Pi_n\|$, we have shown that $\lim_{n\to\infty} n^{-1}\log\|\Pi_n\|$ exists and equals $\log\sqrt{E\,(X_i^*X_i)}$. For $\|Y_n\|$, we have proved that the growth rate of $\|Y_n\|$ is somewhere between $\sqrt{n}$ and n. There remains the question of whether $\lim_{n\to\infty} n^{-s}\|Y_n\|$ exists for some s.

Another interesting question, which is not resolved in this paper, is how the spectral radius of Π_n grows. Indeed, for Y_n, the norm coincides with the spectral radius. But for Π_n, the norm and the spectral radius are different because Π_n is not self-adjoint.

15 On Asymptotic Growth of the Support of Free Multiplicative Convolutions

15.1 Preliminaries and the main result

First, let us recall the definition of the free multiplicative convolution. Let a_k denote the moments of a compactly-supported probability measure μ, $a_k = \int t^k d\mu$, and let the ψ-transform of μ be $\psi(z) = \sum_{k=1}^{\infty} a_k z^k$. The inverse ψ-transform is defined as the functional inverse of $\psi(z)$ and denoted as $\psi^{(-1)}(z)$. It is a well-defined analytic function in a neighborhood of $z = 0$, provided that $a_1 \neq 0$.

Suppose that μ and ν are two probability measures supported on $\mathbb{R}^+ = \{x|x \geq 0\}$ and let $\psi_\mu^{(-1)}(z)$ and $\psi_\nu^{(-1)}(z)$ be their inverse ψ-transforms. Then, it is possible to show that

$$f(z) = \left(1 + z^{-1}\right) \psi_\mu^{(-1)}(z)\, \psi_\nu^{(-1)}(z)$$

is the inverse ψ-transform of a probability measure, which is called the free multiplicative convolution of measures μ and ν, and which is denoted as $\mu \boxtimes \nu$. The significance of this operation can be seen from the fact that if μ and ν are the distributions of singular values of two free operators X and Y, then $\mu \boxtimes \nu$ is the distribution of singular values of the product operator XY (assuming that the expectation is tracial). For more details about free convolutions and free probability theory, the reader can consult (Voiculescu, Dykema, and Nica 1992), (Hiai and Petz 2000), or (Nica and Speicher 2006).

Let the support of a measure μ be inside the interval $[0, L]$, let μ have the expectation 1 and variance v. We assume that the measure consists only of absolutely continuous and atomic parts and that the number of atoms is finite. We are interested in the support of the n-time free multiplicative convolution of the measure μ with itself, which we denote as μ_n.

Theorem 145 *Let L_n denote the upper boundary of the support of μ_n. Then*

$$\begin{aligned} \inf \frac{L_n}{n} &\geq v, \\ \limsup_{n\to\infty} \frac{L_n}{n} &\leq (1 + 2v) \exp\left(\frac{1+v}{1+2v}\right). \end{aligned}$$

Remarks: 1) Let X_i be free identically distributed variables and let $\Pi_n = X_1 ... X_n$. If μ is the spectral probability measure of $X_i^* X_i$, then μ_n is the spectral probability measure of $\Pi_n^*\Pi_n$. Assume that $E(X_i^* X_i) = 1$ and $E\left((X_i^* X_i)^2\right) = 1+v$, and define $\|\Pi_n\|_2 =: [E(\Pi_n^*\Pi_n)]^{1/2}$. Then our theorem implies that

$$\sqrt{n}\sqrt{v}\,\|\Pi_n\|_2 \leq \|\Pi_n\| \leq \sqrt{n}\sqrt{2(1+v)}\,\|\Pi_n\|_2$$

for all sufficiently large n. This result also holds if we relax the assumption $E(X_i^* X_i) = 1$ and define

$$v = \frac{E\left((X_i^* X_i)^2\right)}{[E(X_i^* X_i)]^2} - 1.$$

2) This theorem improves the result in Section 14 ((Kargin 2007)), where under the same conditions on the measure μ it was shown that $L_n/n \leq cL$ where c is a certain absolute constant. Thus the asymptotic growth in the support of free multiplicative convolutions μ_n is controlled by the variance of μ and not by the length of its support.

15.2 Proof of the main result

In our analysis we need a couple of estimates on the ψ-transform. Let the support of a measure μ be inside the interval $[0, L]$, let this measure have the unit expectation, and let it have the variance v. Note that

$$v = \int t^2 \mu(dt) - 1 \leq L - 1,$$

because we assumed that the expectation of measure μ is 1.

We want to estimate higher moments of measure μ in terms of v and L.

Lemma 146 *For every $k \geq 3$,*

$$E\left(X^k\right) \leq \begin{cases} \frac{L^k}{(L-1)^2} v + 1, & \textit{if } L \geq 2, \\ 2^k v + 1, & \textit{if } L < 2. \end{cases}$$

Proof: By binomial expansion,

$$\int x^k \mu(dx) = \sum_{j=0}^{k} \binom{k}{j} \left[\int (x-1)^{k-j} \mu(dx)\right].$$

If $L \geq 2$, then for each term of the sum with $j \leq k-2$ we can estimate this term by $(L-1)^{k-j-2} v$, and so we have:

$$\begin{aligned} E\left(X^k\right) &\leq \left[\binom{k}{0}(L-1)^{k-2} + \binom{k}{1}(L-1)^{k-3} + \ldots + \binom{k}{k-2} 1\right] v + 1 \\ &= \frac{1}{(L-1)^2}\left[\sum_{j=0}^{k}\binom{k}{j}(L-1)^{k-j}\right] v - k\frac{v}{L-1} - \frac{v}{(L-1)^2} + 1 \\ &\leq \frac{L^k}{(L-1)^2} v + 1. \end{aligned}$$

If $L < 2$, then we can estimate:

$$\begin{aligned} E\left(X^k\right) &\leq \left[\binom{k}{0} + \binom{k}{1} + \ldots + \binom{k}{k-2}\right] v + 1 \\ &\leq 2^k v + 1. \end{aligned}$$

QED.

Now we can estimate ψ and its derivative. For example, if $L \geq 2$, then we have:

$$\psi(z) \leq z + (1+v) z^2 + \frac{vL^3}{(L-1)^2}\frac{z^3}{1-Lz} + \frac{z^3}{1-z}$$

Similarly,

$$\psi'(z) \leq 1 + 2(1+v) z + \frac{vL^3}{(L-1)^2}\frac{3-2Lz}{(1-Lz)^2} z^2 + \frac{3-2z}{(1-z)^2} z^2. \tag{64}$$

If $L < 2$, then

$$\psi(z) \leq z + (1+v) z^2 + 8v\frac{z^3}{1-2z} + \frac{z^3}{1-z},$$

and

$$\psi'(z) \leq 1 + 2(1+v) z + 8v\frac{3-4z}{(1-2z)^2} z^2 + \frac{3-2z}{(1-z)^2} z^2.$$

These estimates are valid for all $z \in [0, L)$.

Proof of Theorem 145: The first claim of Theorem 145 is easy. Note the estimate:

$$nv = v_n = \int_0^{L_n} t^2 d\mu_n(t) - 1 \leq L_n - 1,$$

which is valid because the variances are additive relative to free multiplicative convolution, the expectation of the measure μ_n is 1, and the measure is supported on $[0,\infty)$. This implies that

$$\inf \frac{L_n}{n} \geq v.$$

The proof of the second claim is more involved. Let the ψ-transform of the n-time convolution be denoted as $\psi_n(z)$. The idea is that the Taylor series for $\psi_n(z)$ diverges for those real z which are greater in absolute value than $1/L_n$. Therefore the function $\psi_n(z)$ must have a singularity in the closed disc $|z| \leq 1/L_n$; i.e., there should exist a point z_0 in this disc, such that the Taylor series for $\psi_n(z)$ diverges at z_0. Since all the coefficients in this series are real and positive, we can take z_0 to be real and positive. Therefore, in order to bound $1/L_n$ from below, we are looking for the singularity of $\psi_n(z)$ which is located in $\mathbb{R}^+$ and which is closest to 0.

From the results in (Belinschi and Bercovici 2005) we know that for all large n, the measure μ_n does not have atoms in $\mathbb{R}^+ \setminus \{0\}$. Since atoms of μ_n correspond to poles of $\psi_n(z)$, therefore, we can assume that the function $\psi_n(z)$ does not have poles in $\mathbb{R}^+$. Hence, the problem is reduced to finding the branching point of $\psi_n(z)$, which would be closest to 0. This branching point of $\psi_n(z)$ equals the critical value of $\psi_n^{(-1)}(z)$. By Voiculescu's theorem,

$$\psi_n^{(-1)}(z) = \left(\frac{1+u}{u}\right)^{n-1} \left[\psi^{(-1)}(u)\right]^n,$$

where $\psi^{(-1)}(u)$ is the inverse ψ-function for measure μ. Therefore we can find critical points of $\psi_n^{(-1)}(z)$ from the equation:

$$\frac{d}{du}\left[n \log \psi^{(-1)}(u) + (n-1)\log\left(\frac{1+u}{u}\right)\right] = 0,$$

We can write this equation as

$$\frac{n}{n-1}\frac{d}{du}\log \psi^{(-1)}(u) - \frac{1}{u(1+u)} = 0.$$

Thus, our task is to estimate that root of this equation which is real, positive and closest to 0. In particular, if we succeed in proving that for all $u \in [0,b]$ the following inequality is valid:

$$\frac{d}{du}\log \psi^{(-1)}(u) > \frac{n-1}{n}\frac{1}{u(1+u)}, \tag{65}$$

then we can be sure that all the critical points of $\psi_n^{(-1)}(z)$ are greater than b. Since the inverse ψ-transform is increasing on the interval $[0,b]$, therefore we will be able to infer that the critical value of $\psi_n^{(-1)}(z)$ is greater than

$$\left[\psi^{(-1)}(b)\right]^n \left(\frac{1+b}{b}\right)^{n-1}.$$

Hence, we will be able to conclude that the upper boundary of the support of μ_n is less than

$$\frac{1}{\left[\psi^{(-1)}(b)\right]^n} \left(\frac{b}{1+b}\right)^{n-1}.$$

In order to proceed with this plan, we re-write inequality (65) as

$$\frac{1}{z\psi'(z)} > \frac{n-1}{n} \frac{1}{\psi(z)(1+\psi(z))}, \tag{66}$$

where $z = \psi^{(-1)}(u)$.

Consider first the case $L \geq 2$. Using estimate (64), we infer that inequality (66) is implied by the following inequality:

$$\frac{1}{z} \frac{1}{1+2(1+v)z + \frac{vL^3}{(L-1)^2}\frac{3-2Lz}{(1-Lz)^2}z^2 + \frac{3-2z}{(1-z)^2}z^2} > \frac{n-1}{n}\frac{1}{\psi(z)(1+\psi(z))},$$

Next we note that $\psi(z) \geq z$ because we assumed that the first moment is 1 and because all other moments are positive. Therefore, it is enough to show that

$$\frac{1}{1+2(1+v)z + \frac{vL^3}{(L-1)^2}\frac{3-2Lz}{(1-Lz)^2}z^2 + \frac{3-2z}{(1-z)^2}z^2} > \frac{n-1}{n}\frac{1}{1+z}.$$

Let us write this as

$$\frac{1}{n-1} + \frac{1}{n-1}z > (1+2v)z + \frac{3-2z}{(1-z)^2}z^2 + \frac{vL^3}{(L-1)^2}\frac{3-2Lz}{(1-Lz)^2}z^2.$$

If we fix an arbitrary $\varepsilon > 0$, then clearly for all $z \leq (n(1+2v+\varepsilon))^{-1}$ this inequality holds if n is sufficiently large.

A similar conclusion can be achieved in the case when $L < 2$. This implies that we can find such $c(\varepsilon, L, n)$ that inequality (66) holds for all real positive $z \leq z_0(n) = (n(1+2v+\varepsilon))^{-1} c(\varepsilon, L, n)$ and that $c(\varepsilon, L, n) \to 1$ as $n \to \infty$.

Next step is to estimate $[z_0/\psi(z_0)]^{n-1}$. Note that $\psi(z) = z + (1+v)z^2 + O(z^3)$, where the coefficient in the term $O(z^3)$ depends on L, but can be chosen uniformly for all $z \leq z_0(n)$ if n is sufficiently large. Therefore,

$$[z_0/\psi(z_0)]^{n-1} \to \exp\left(-\frac{1+v}{1+2v+\varepsilon}\right),$$

as $n \to \infty$. It follows that

$$z_0[z_0/\psi(z_0)]^{n-1} = (n(1+2v+\varepsilon))^{-1} \exp\left(-\frac{1+v}{1+2v+\varepsilon}\right) c'(\varepsilon, L, n),$$

where $c'(\varepsilon, L, n) \to 1$ as $n \to \infty$. It follows that the upper boundary of the support of the n-time convolution can be estimated as follows:

$$L_n \leq n(1+2v+\varepsilon) \exp\left(\frac{1+v}{1+2v+\varepsilon}\right) c''(\varepsilon, L, n),$$

where $c''(\varepsilon, L, n) \to 1$ as $n \to \infty$. In particular, this means that

$$\limsup_{n\to\infty} \frac{L_n}{n} \leq (1+2v) \exp\left(\frac{1+v}{1+2v}\right).$$

QED.

15.3 Conclusion

It would be interesting to find out whether the limit of $(nv)^{-1} L_n$ exists, and if it does, then whether it depends on the measure μ.

16 Lyapunov Exponents for Free Operators

16.1 Introduction

Suppose that at each moment of time, t_i, a system is described by a state function $\varphi(t_i)$ and evolves according to the law $\varphi(t_{i+1}) = X_i \varphi(t_i)$, where X_i is a sequence of linear operators. One can ask how small changes in the initial position of the system are reflected in its long-term behavior. If operators X_i do not depend on time, $X_i = X$, then the long-term behavior depends to a large extent on the spectrum of the operator X. If operators X_i do depend on time but can be modelled as a stationary stochastic process, then the long-term behavior of the system depends to a large extent on so-called *Lyapunov exponents* of the process X_i.

The largest Lyapunov exponent of a sequence of random matrices was investigated in a pioneering paper by Furstenberg and Kesten (1960). This study was followed by Oseledec (1968), which researched other Lyapunov exponents and finer aspects of the asymptotic behavior of matrix products. These investigations were greatly expanded and clarified by many other researchers. In particular, Ruelle (1982) developed a theory of Lyapunov exponents for random compact linear operators acting on a Hilbert space. Cohen and Newman (1984), Newman (1986b), Newman (1986a), and Isopi and Newman (1992) studied Lyapunov exponents for random $N \times N$ matrices when $N \to \infty$.

The goal of this paper is to investigate how the concept of Lyapunov exponents can be extended to the case of free linear operators. It was noted recently (Voiculescu (1991)) that the theory of free operators can be a natural asymptotic approximation for the theory of large random matrices. Moreover, it was noted that certain difficult calculations from the theory of large random matrices become significantly simpler if similar calculations are performed using free operators. For this reason it is interesting to study whether the concept of Lyapunov exponents is extendable to free operators, and what methods for calculation of Lyapunov exponents are available in this setting.

Free operators are not random in the traditional sense so the usual definition of Lyapunov exponents cannot be applied directly. Our definition of Lyapunov exponents is based on the observation that in the case of random matrices, the sum of logarithms of the k largest Lyapunov exponents equals the rate at which a random k-dimensional volume element grows asymptotically when we consecutively apply operators X_i.

In the case of free operators we employ the same idea. However, in this case we have to clarify how to measure the change in the "t-dimensional volume element" after we apply operators X_i. It turns out that we can measure this change by a suitable extension of the Fuglede-Kadison determinant. Given this extension, the definition proceeds as follows: Take a subspace of the Hilbert space, such that the corresponding projection is free from all X_i and have the dimension t relative to the given trace. Next, act on this subspace by the sequence of operators X_i. Apply the determinant to measure how the “volume element” in this subspace changes under these linear transformations. Use the asymptotic growth in the determinant to define the Lyapunov exponent corresponding to the dimension t.

It turns out that the growth of the t-dimensional volume element is exponential with a rate which is a function of the dimension t. We call this rate the *integrated Lyapunov exponent*. It is an analogue of the sum of the k largest Lyapunov exponents in the finite-dimensional case. The derivative of this function is called the *marginal Lyapunov exponent*. Its value at a point t is an analogue of the k-th largest Lyapunov exponent.

Next, we relate the marginal Lyapunov exponent $f_X(t)$ to the Voiculescu S-transform of the random variable $X_i^* X_i$. The relationship is very simple:

$$f_X(t) = -(1/2)\log\left[S_{X^*X}(-t)\right]. \tag{67}$$

Using this formula, we prove that the marginal Lyapunov exponent is decreasing in t, and derive an expression for the largest Lyapunov exponent. Formula (67) also allows us to prove the additivity of the marginal Lyapunov exponent with respect to operator product: $f_{XY}(t) = f_X(t) + f_Y(t)$ .The author is unaware if an analogous result holds for finite-dimensional random matrices.

As an example, we calculate Lyapunov exponents for variables X_i that have the Marchenko-Pastur distribution with parameter λ as the spectral probability distribution of $X_i^* X_i$. The case $\lambda = 1$ corresponds to the case considered in Newman (1986b), and the results of this paper are in agreement with Newman's "triangle" law. In addition, our results regarding the largest Lyapunov exponent agree with the results regarding the norm of products of large random matrices in Cohen and Newman (1984). Finally, our formula for computation of Lyapunov exponents seems to be easier to apply than the non-linear integral transformation developed in Newman (1986b).

An interesting by-product of our results is a relation between the extended Fuglede-Kadison determinant and the Voiculescu S-transform, which allows expressing each of them in terms of the other. Since under certain conditions the extended determinant retains the multiplicativity property of the original Fuglede-Kadison determinant, this relation sheds some additional light on the multiplicativity property of the S-transform.

The rest of the paper is organized as follows: Section 16.2 describes the extension of the Fuglede-Kadison determinant that we use in this paper. Section 16.3 defines the Lyapunov exponents of free operators, proves an existence theorem, and derives a formula for the calculation of Lyapunov exponents. Section 16.4 computes the Lyapunov exponents for a particular example. Section 16.5 connects the marginal

Lyapunov exponents and the S-transform, proves additivity and monotonicity of the marginal Lyapunov exponent, and derives a formula for the largest Lyapunov exponent. In addition, it derives a relation between the determinant and the S-transform. And Section 16.6 concludes.

16.2 A modification of the Fuglede-Kadison determinant

Let $\mathcal{A}$ be a finite von Neumann algebra and E be a trace in this algebra. Recall that if X is an element of $\mathcal{A}$ that has a bounded inverse, then the Fuglede-Kadison determinant ((Fuglede and Kadison 1952)) is defined by the following formula:

$$\det(X) = \exp \frac{1}{2} E \log(X^* X). \tag{68}$$

The most important property of the Fuglede-Kadison determinant is its multiplicativity:

$$\det(XY) = \det(X)\det(Y). \tag{69}$$

This determinant cannot be extended (non-trivially) to non-invertible X if we require that property (69) holds for all X and Y.

However, if we relax this property, then we can define an *extended determinant* as follows: Let $\log^{+\lambda}(t) =: \log t$ if $t > \lambda$ and $=: 0$ if $t \leq \lambda$. Note that on the interval $(0,1)$, $E\log^{+\lambda}(X^*X)$ is a (weakly) decreasing function of λ, and therefore it converges to a limit (possibly infinite) as $\lambda \to 0$.

Definition 147

$$\det(X) = \exp \frac{1}{2} \lim_{\lambda \downarrow 0} E \log^{+\lambda}(X^* X).$$

This extension of the Fuglede-Kadison determinant coincides with the extension introduced in Section 3.2 of Luck (2002) .

Example 148 *Zero Operator*

From Definition 147, if $X = 0$, then $\det X = 1$.

Example 149 *Finite dimensional algebra*

Consider the algebra of n-by-n matrices $M_n(C)$ with the trace given as the normalization of the usual matrix trace: $E(X) = n^{-1}\operatorname{Tr}(X)$. Then the original Fuglede-Kadison determinant is defined for all full-rank matrices and equals the product of the singular values of the operator in the power of $1/n$. It is easy to see that this equals the absolute value of the usual matrix determinant in the power of $1/n$. The extended Fuglede-Kadison determinant is defined for all matrices, including the matrices of rank $k < n$. It is equal to the product of non-zero singular values in the power of $1/n$.

We can write the definition of the determinant in a slightly different form. Recall that for a self-adjoint operator $X \in \mathcal{A}$ we can define its *spectral probability measure* as follows: First, we write the spectral decomposition as

$$X = \int_{-\infty}^{\infty} \lambda P_X(d\lambda),$$

where $\{P_X(\cdot)\}$ is a family of commuting projections. Then, the spectral probability measure of X is defined by the following formula:

$$\mu_X(S) = E(P_X(S)),$$

where S is an arbitrary Borel-measurable set. We can calculate the trace of any summable function of a self-adjoint variable A by using its spectral measure:

$$Ef(X) = \int_{-\infty}^{\infty} f(\lambda)\, d\mu_X(\lambda).$$

In particular the determinant of operator X can be written as

$$\det(X) = \exp \frac{1}{2} \lim_{\lambda \downarrow 0} \int_{\mathbb{R}^+} \log^{+\lambda}(t)\, d\mu_{X^*X}(t).$$

For arbitrary probability measure μ with support in $\mathbb{R}^+ = \{x | x \geq 0\}$, we write:

$$\det(\mu) = \exp \frac{1}{2} \lim_{\lambda \downarrow 0} \int_{\mathbb{R}^+} \log^{+\lambda}(t)\, d\mu(t).$$

For all invertible X the extended determinant defines the same object as the usual Fuglede-Kadison determinant. For non-invertible X, the multiplicativity property sometimes fails. However, it holds if a certain condition on images and domains of the multiplicands is fulfilled:

Proposition 150 *Let V be the closure of the range of the operator A. If B is an injective mapping on V and is the zero operator on $V^{\perp}$, then* $\det(BA) = \det(B)\det(A)$.

The claim of this proposition is a direct consequence of Theorem 3.14 and Lemma 3.15(7) in Luck (2002).

Now let us connect the determinant and the concepts of free probability theory. Following the conventions of free probability theory, we will call the pair $(\mathcal{A}, E)$ a *non-commutative probability space* if $\mathcal{A}$ is a finite von Neumann algebra, E is a trace in this algebra, and $E(I) = 1$. The trace E will be called the *expectation* by analogy with classical probability theory. Let $\mathcal{A}_{1,...,}\mathcal{A}_n$ be sub-algebras of algebra $\mathcal{A}$, and let a_i be elements of these sub-algebras such that $a_i \in \mathcal{A}_{k(i)}$.

Definition 151 *The sub-algebras $\mathcal{A}_{1,...,}\mathcal{A}_n$ (and their elements) are called* free *or* freely independent *if $E(a_1...a_m) = 0$ whenever the following two conditions hold:*
(a) $E(a_i) = 0$ for every i, and
(b) $k(i) \neq k(i+1)$ for every $i < m$, and $k(m) \neq k(1)$.

The random variables are called free or freely independent if the algebras that they generate are free. (See Voiculescu et al. (1992) or Hiai and Petz (2000) for more details on foundations of free probability theory.)

If μ is the spectral probability measure for X^*X and ν is the spectral probability measure for Y^*Y, then the spectral probability measure of Y^*X^*XY depends only on μ and ν. It is called the *free multiplicative convolution* of measures μ and ν, and denoted as $\mu \boxtimes \nu$.

Proposition 152 *Let μ and ν be two probability measures supported on $\mathbb{R}^+$. Suppose that they have no atoms at 0, i.e., $\mu(\{0\}) = \nu(\{0\}) = 0$. Then* $\det(\mu \boxtimes \nu) = \det(\mu)\det(\nu)$, *where $\mu \boxtimes \nu$ denotes the free multiplicative convolution of measures μ and ν.*

Proof: Let us take free, self-adjoint, and positive X and Y, such that μ and ν are the spectral probability measures for X^2 and Y^2, respectively. Then, by definition, $\mu \boxtimes \nu$ is the spectral probability measure for YX^2Y and we can write $\det(\mu \boxtimes \nu) = \det(XY) = \det(X)\det(Y) = \det(\mu)\det(\nu)$. The second equality holds true because the closure of the image of Y is the whole space (guaranteed by $\nu(\{0\}) = 0$), and X is injective on this image (guaranteed by $\mu(\{0\}) = 0$). QED.

16.3 Definition of Lyapunov exponents for free operators

Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of free identically-distributed operators. Let $\Pi_n = X_n \dots X_1$, and let P_t be a projection which is free of all X_i and has the dimension t, i.e., $E(P_t) = t$.

Definition 153 *The* integrated Lyapunov exponent *corresponding to the sequence X_i is a real-valued function of $t \in [0,1]$ which is defined as follows:*

$$F(t) = \lim_{n \to \infty} \frac{1}{n} \log \det (\Pi_n P_t),$$

provided that the limit exists.

Remark: In the case of random matrices, Π_n is the product of independent identically-distributed random matrices. In this case, it turns out that the function defined analogously to $F(t)$ equals the sum of the tN largest Lyapunov exponents divided by N, where N is the dimension of the matrices and t belongs to the set $\{0/N, 1/N, \dots, N/N\}$.

Our first task is to prove the existence of the limit in the previous definition.

Theorem 154 *Suppose that X_i are free identically-distributed operators. Let $u =: \dim \ker (X_i)$. Then*

$$F(t) = \begin{cases} \frac{1}{2} \lim_{\lambda \downarrow 0} E \log^{+\lambda} (P_t X_1^* X_1 P_t), & \text{if } t \leq 1-u, \\ \frac{1}{2} \lim_{\lambda \downarrow 0} E \log^{+\lambda} (P_u X_1^* X_1 P_u), & \text{if } t > 1-u. \end{cases}$$

Before proving this theorem, let us make some remarks. First, this theorem shows that the integrated Lyapunov exponent of the sequence $\{X_i\}$ exists and depends only on the spectral distribution of $X_i^* X_i$.

Next, suppose that we know that $F(t)$ is differentiable almost everywhere. Then we can define the *marginal Lyapunov exponent* as $f(t) = F'(t)$. We can also define the *distribution function of Lyapunov exponents* by the formula: $\mathcal{F}(x) = \mu \{t \in [0,1] : f(t) \leq x\}$, where μ is the usual Borel-Lebesgue measure. Intuitively, this function gives a measure of the set of the Lyapunov exponents which are less than a given threshold, x. In the finite-dimensional case it is simply the empirical distribution function of the Lyapunov exponents, i.e., the fraction of Lyapunov exponents that fall below the threshold x.

Proof of Theorem 154: The proof is through a sequence of lemmas. We will consider first the case of injective operators X_i and then will show how to generalize the argument to the case of arbitrary X_i.

Let P_A denote the projection on the closure of the range of operator A.

Lemma 155 *Suppose that A is an operator in a von Neumann probability space $\mathcal{A}$, that A is injective, and that P_t is a projection of dimension t. Then projection P_{AP_t} is equivalent to P_t. In particular, $E\left(P_{AP_t}\right) = t$.*

Proof: Recall that polar decomposition is possible in $\mathcal{A}$. (See Proposition II.3.14 on p. 77 in Takesaki (1979) for details.) Therefore, we can write $AP_t = WB$, where W is isometric and B is self-adjoint and where both W and B belong to $\mathcal{A}$. By definition, the range of W is $[\mathrm{Range}\,(AP_t)]$, and the domain of W is $[x : Bx = 0]^{\perp} = [x : AP_t x = 0]^{\perp} = [x : P_t x = 0]^{\perp} = [\mathrm{Range}\,(P_t)]$. Therefore, P_{AP_t} is equivalent to P_t, with the equivalence given by the isometric tranformation W. In particular, $\dim\left(P_{AP_t}\right) = \dim\left(P_t\right)$, i.e., $E\left(AP_t\right) = t$. QED.

Lemma 156 *If A, A^*, and P_t are free from an operator algebra $\mathcal{B}$, then P_{AP_t} is free from $\mathcal{B}$.*

Proof: P_{AP_t} belongs to the algebra generated by I, A, A^*, and P_t. By assumption, this algebra is free from $\mathcal{B}$. Hence, P_{AP_t} is also free from $\mathcal{B}$. QED.

Let us use the notation $Q_k = P_{X_k \ldots X_1 P_t}$ for $k \geq 1$ and $Q_0 = P_t$. Then by Lemma 156, Q_k is free from X_{k+1}. Besides, if all X_i are injective, then their product is injective and, therefore, by Lemma 155, Q_k is equivalent to P_t.

Lemma 157 *If all X_i are injective, then*

$$\det\left(\Pi_n P_t\right) = \prod\nolimits_{i=1}^{n} \det\left(X_i Q_{i-1}\right).$$

Proof: Note that $\Pi_n P_t = X_n Q_{n-1} X_{n-1} \ldots Q_1 X_1 Q_0$. We will proceed by induction. We need only to prove that

$$\det\left(X_{k+1} Q_k X_k \ldots Q_1 X_1 Q_0\right) = \det\left(X_{k+1} Q_k\right) \det\left(X_k \ldots Q_1 X_1 Q_0\right). \qquad (70)$$

Let V_k be the closure of the range of $X_k \ldots Q_1 X_1 Q_0$. Since X_{k+1} is injective and Q_k is the projector on V_k, therefore $X_{k+1} Q_k$ is injective on V_k and equal to zero on $V_k^{\perp}$. Consequently, we can apply Proposition 150 and obtain (70). QED.

Now we are ready to prove Theorem 154 for the case of injective X_i. Using Lemma 157, we write

$$n^{-1} \log\det\left(\Pi_n P_t\right) = \frac{1}{n} \sum\nolimits_{i=1}^{n} \log\det\left(X_i Q_{i-1}\right).$$

Note that X_i are identically distributed by assumption, Q_i have the same dimension by Lemma 155, and X_i and Q_{i-1} are free by Lemma 156. This implies that

$$\lim_{\lambda\downarrow 0} E\log^{+\lambda}\left(Q_{i-1}X_i^*X_iQ_{i-1}\right)$$

does not depend on i, and hence, $\det\left(X_iQ_{i-1}\right)$ does not depend on i. Hence, using $i=1$ we can write:

$$n^{-1}\log\det\left(\Pi_n P_t\right)=\log\det\left(X_1P_t\right).$$

This finishes the proof for the case of injective X_i. For the case of non-injective X_i, i.e., for the case when $\dim\ker\left(X_i\right)>0$, we need the following lemma.

Lemma 158 *Suppose that P_t is a projection operator free of A and such that $E\left(P_t\right)=t$. Then* $\dim\ker\left(AP_t\right)=\max\left\{1-t,\dim\ker\left(A\right)\right\}.$

Proof: Let $V=\left(\operatorname{Ker}A\right)^{\perp}$ and let P_V be the projection on V. Then $E\left(P_V\right)=1-\dim\operatorname{Ker}A$. Note that $Ax=0\iff P_Vx=0$. Consequently, $AP_tx=0\iff P_VP_tx=0$. Therefore, we have:

$$\begin{aligned}\dim\left\{x:AP_tx=0\right\}&=\dim\left\{x:P_VP_tx=0\right\}\\&=\dim\left\{x:P_tP_VP_tx=0\right\}.\end{aligned}$$

Since P_t and P_V are free, an explicit calculation of the distribution of $P_tP_VP_t$ shows that

$$\dim\left\{x:P_tP_VP_tx=0\right\}=\max\left\{1-t,1-\dim V\right\}.$$

QED.

Consider first the case when $0<\dim\ker X_i\leq 1-t$. This case is very similar to the case of injective X_i. Using Lemma 158 we conclude that $\dim\operatorname{Ker}\left(X_1P_t\right)=1-t$, and therefore that $E\left(P_{X_1P_t}\right)=t$. If, as before, we denote $P_{X_1P_t}$ as Q_1, then the projection Q_1 is free from X_2, and $E\left(Q_1\right)=t$.

Similarly, we obtain that $E\left(P_{X_2Q_1}\right)=t$. Proceeding inductively, we define $Q_k=P_{X_kQ_{k-1}}$ and conclude that Q_k is free from X_{k+1} and that $E\left(Q_k\right)=t$.

Next, we write $X_k...X_1P_t=X_kQ_{k-1}X_{k-1}Q_{k-2}...X_1Q_0$, where Q_0 denotes P_t, and note that X_kQ_{k-1} is injective on the range of Q_{k-1}. Indeed, if it were not injective, then we would have $\dim\left(\operatorname{Ker}\left(X_k\right)\cap\operatorname{Range}\left(Q_{k-1}\right)\right)>0$. But this would imply that $\dim(\mathrm{Ker}\, X_k Q_{k-1}) > \dim(\mathrm{Ker}\, Q_{k-1}) = 1 - t$, which contradicts the fact that $\dim(\mathrm{Ker}\, X_k Q_{k-1}) = 1 - t$. Therefore, Proposition 150 is applicable and

$$\begin{aligned} \det(X_k ... X_1 P_t) &= \det(X_k Q_{k-1}) ... \det(X_1 Q_0) \\ &= [\det(X_1 P_t)]^k . \end{aligned}$$

Now let us turn to the case when $\dim \ker X_i = u > 1-t$. Then $\dim \mathrm{Ker}\,(X_1 P_t) = 1-u$ and therefore $E(P_{X_1 P_t}) = u$. Proceeding as before, we conclude that $E(Q_k) = u$ for all $k \geq 1$, and we can write $X_k ... X_1 P_t = X_k Q_{k-1} X_{k-1} Q_{k-2} ... X_1 Q_0$, where we have denoted P_t as Q_0. Then we get the following formula:

$$\begin{aligned} \det(X_k ... X_1 P_t) &= \det(X_k Q_{k-1}) ... \det(X_2 Q_1) \det(X_1 Q_0) \\ &= [\det(X_1 P_u)]^{k-1} \det(X_1 P_t) . \end{aligned}$$

Therefore,

$$\lim_{n\to\infty} n^{-1} \log \det(\Pi_n P_t) = \log \det(X_1 P_u) .$$

QED.

16.4 Example

Let us compute the Lyapunov exponents for a random variable X that has the product $X^* X$ distributed according to the Marchenko-Pastur distribution. Recall that the continuous part of the Marchenko-Pastur probability distribution with parameter $\lambda > 0$ is supported on the interval $\left[\left(1-\sqrt{\lambda}\right)^2, \left(1+\sqrt{\lambda}\right)^2\right]$, and has the following density there:

$$p_\lambda(x) = \frac{\sqrt{4\lambda - (x-1-\lambda)^2}}{2\pi x}.$$

For $\lambda \in (0,1)$, this distribution also has an atom at 0 with the probability mass $(1-\lambda)$ assigned to it. The Marchenko-Pastur distribution is sometimes called the free Poisson distribution since it arises as a limit of free additive convolutions of the Bernoulli distribution, and a similar limit in the classical case equals the Poisson distribution. It can also be thought of as a scaled limit of the eigenvalue distribution of Wishart-distributed random matrices (see (Hiai and Petz 2000) for a discussion).

Proposition 159 *Suppose that X is a non-commutative random variable such that X^*X is distributed according to the Marchenko-Pastur distribution with parameter λ. If $\lambda \geq 1,$ then the distribution of Lyapunov exponents of X is*

$$\mathcal{F}(x) = \begin{cases} 0, & \text{if } x < (1/2)\log(\lambda - 1) \\ e^{2x} + 1 - \lambda, & \text{if } x \in \left[\frac{1}{2}\log(\lambda-1), \frac{1}{2}\log\lambda\right) \\ 1 & \text{if } x \geq \frac{1}{2}\log\lambda. \end{cases}.$$

If $\lambda < 1,$ then the distribution of Lyapunov exponents of X is

$$\mathcal{F}(x) = \begin{cases} e^{2x}, & \text{if } x < (1/2)\log(\lambda) \\ \lambda, & \text{if } x \in \left[\frac{1}{2}\log(\lambda), 0\right) \\ 1 & \text{if } x \geq 0. \end{cases}$$

Remark: If $\lambda = 1$, then the distribution is the exponential law discovered by C. M. Newman as a scaling limit of Lyapunov exponents of large random matrices. (See Newman (1986b), Newman (1986a), and Isopi and Newman (1992). This law is often called the "triangle" law since it implies that the exponentials of Lyapunov exponents converge to the law whose density function is in the form of a triangle.)

Proof of Proposition 159: It is easy to calculate that the continuous part of the distribution of P_tXP_t is supported on the interval $\left[\left(\sqrt{t}-\sqrt{\lambda}\right)^2, \left(\sqrt{t}+\sqrt{\lambda}\right)^2\right]$, and has the density function

$$p_{t,\lambda}(x) = \frac{\sqrt{4\lambda t - [x-(t+\lambda)]^2}}{2\pi x}.$$

This distribution also has an atom at $x = 0$ with the probability mass $\max\{1-\lambda, 1-t\}$. See for example, results in (Nica and Speicher 1996).

Next, we write the expression for the integrated Lyapunov exponent. If $\lambda \geq 1$, or $\lambda < 1$ but $\lambda \geq t$, then

$$\begin{aligned} F_\lambda(t) &= \frac{1}{2}\lim_{\varepsilon \downarrow 0} E\log^{+\varepsilon}(P_tX^*XP_t) \\ &= \frac{1}{2}\int_{\left(\sqrt{t}-\sqrt{\lambda}\right)^2}^{\left(\sqrt{t}+\sqrt{\lambda}\right)^2} \log x \frac{\sqrt{4\lambda t - [x-(t+\lambda)]^2}}{2\pi x} dx. \end{aligned} \tag{71}$$

If $\lambda < 1$ and $\lambda < t$, then

$$F_\lambda(t) = \frac{1}{2}\lim_{\varepsilon\downarrow 0} E\log^{+\varepsilon}\left(P_{1-\lambda}X^*XP_{1-\lambda}\right)$$

$$F_\lambda(t) = \frac{1}{2}\int_{\left(\sqrt{1-\lambda}-\sqrt{\lambda}\right)^2}^{\left(\sqrt{1-\lambda}+\sqrt{\lambda}\right)^2} \log x \frac{\sqrt{4\lambda(1-\lambda)-[x-1]^2}}{2\pi x}dx. \tag{72}$$

Differentiating (71) with respect to t, we obtain an expression for the marginal Lyapunov exponent:

$$f_\lambda(t) = \frac{1}{4\pi}\int_{\left(\sqrt{t}-\sqrt{\lambda}\right)^2}^{\left(\sqrt{t}+\sqrt{\lambda}\right)^2} \frac{\log x}{x}\frac{x-t+\lambda}{\sqrt{4\lambda x-[x-t+\lambda]^2}}dx. \tag{73}$$

Using substitutions $u = \left[x - \left(\sqrt{t}-\sqrt{\lambda}\right)^2\right] / \left(2\sqrt{\lambda t}\right) - 1$ and then $\theta = \arccos u$, this integral can be computed as

$$f_\lambda(t) = \frac{1}{2}\log(\lambda - t).$$

From this expression, we calculate the distribution of Lyapunov exponents for the case when $\lambda \geq 1$:

$$\mathcal{F}(x) = \begin{cases} 0, & \text{if } x < (1/2)\log(\lambda-1) \\ e^{2x}+1-\lambda, & \text{if } x \in \left[\frac{1}{2}\log(\lambda-1), \frac{1}{2}\log\lambda\right) \\ 1 & \text{if } x \geq \frac{1}{2}\log\lambda. \end{cases}$$

A similar analysis shows that for $\lambda < 1$, the distribution is as follows:

$$\mathcal{F}(x) = \begin{cases} e^{2x}, & \text{if } x < (1/2)\log(\lambda) \\ \lambda, & \text{if } x \in \left[\frac{1}{2}\log(\lambda), 0\right) \\ 1 & \text{if } x \geq 0. \end{cases}$$

QED.

16.5 A relation with the S-transform

In this section we derive a formula that makes the calculation of Lyapunov exponents easier and relates them to the S-transform of the operator X_i. Recall that the ψ-function of a bounded non-negative operator A is defined as $\psi_A(z) = \sum_{k=1}^{\infty} E\left(A^k\right)z^k$.

Then the S-transform is $S_A(z) = (1 + z^{-1}) \psi_A^{(-1)}(z)$, where $\psi_A^{(-1)}(z)$ is the functional inverse of $\psi_A(z)$ in a neighborhood of 0.

Theorem 160 *Let X_i be identically distributed free bounded operators. Let $Y = X_1^* X_1$ and suppose that $\mu_Y(\{0\}) = 1 - u \in [0, 1)$, where μ_Y denotes the spectral probability measure of Y. Then the marginal Lyapunov exponent of the sequence $\{X_i\}$ is given by the following formula:*

$$f_X(t) = \begin{cases} -\frac{1}{2}\log[S_Y(-t)] & \text{if } t < u, \\ 0 & \text{if } t > u, \end{cases}$$

where S_Y is the S-transform of the variable Y.

Remark: Note that if $X_1^* X_1$ has no atom at zero then the formula is simply $f_X(t) = -\frac{1}{2}\log[S_Y(-t)]$.

Proof: If $t > u$, then $f_X(t) = 0$ by Theorem 154. Assume in the following that $t < u$. Then $P_t X^* X P_t$ has an atom of mass $1 - t$ at 0. Let μ_t denote the spectral probability measure of $P_t X^* X P_t$, with the atom at 0 removed. (So the total mass of μ_t is t.) We start with the formula:

$$\log x = \log(c + x) - \int_0^c \frac{ds}{x + s},$$

and write:

$$\int_0^\infty \log x \, \mu_t(dx) = \lim_{c \to \infty} \left\{ t \log(c) + \int_0^c G_t(-s) \, ds \right\},$$

where G_t is the Cauchy transform of the measure μ_t.

Next, note that $G_t(-s) = -s^{-1}\psi_t(-s^{-1}) - ts^{-1}$ and substitute this into the previous equation:

$$\begin{aligned} \int_0^\infty \log x \, \mu_t(dx) &= \lim_{c \to \infty, \varepsilon \to 0} \left\{ t \log c - t \log c + t \log(\varepsilon) + \int_\varepsilon^c \frac{\psi_t(s^{-1})}{s} ds \right\} \\ &= \lim_{\varepsilon \to 0} \left\{ t \log(\varepsilon) + \int_\varepsilon^\infty \frac{\psi_t(s^{-1})}{s} ds \right\}. \end{aligned}$$

Using substitutions $v = -\log s$, and $A = -\log \varepsilon$, we can re-write this equation as follows:

$$\int_0^\infty \log x \, \mu_t(dx) = \lim_{A \to \infty} \left\{ -tA - \int_{-\infty}^A \psi_t(-e^v) \, dv \right\}.$$

The function $\psi_t(-e^v)$ monotonically decreases when v changes from $-\infty$ to ∞, and its value changes from 0 to $-t$. Let $s^* =: \psi_t(-e^0) = \psi_t(-1)$ and let $\xi_t(x)$ denote the functional inverse of $\psi_t(-e^v)$. The function $\xi_t(x)$ is defined on the interval $(-t, 0)$. In this interval it is monotonically decreasing from ∞ to $-\infty$. The only zero of $\xi_t(x)$ is at $x = s^*$.

It is easy to see that

$$\lim_{A\to\infty}\left\{-tA - \int_0^A \psi_t(-e^v)\, dv\right\} = -\int_{-t}^{s^*} \xi_t(x)\, dx,$$

and that

$$-\int_{-\infty}^{0} \psi_t(-e^v)\, dv = -\int_{s^*}^{0} \xi_t(x)\, dx.$$

Therefore,

$$\int_0^\infty \log x\, \mu_t(dx) = -\int_{-t}^{0} \xi_t(x)\, dx.$$

It remains to note that $\xi_t(x) = \log\left[-\psi_t^{(-1)}(x)\right]$, in order to conclude that

$$\int_0^\infty \log x\, \mu_t(dx) = -\int_{-t}^{0} \log\left[-\psi_t^{(-1)}(x)\right] dx.$$

The next step is to use Voiculescu's multiplication theorem and write: $\psi_t^{(-1)}(x) = \psi_Y^{(-1)}(x)(1+x)/(t+x)$. Then we have the formula:

$$\begin{aligned}\int_0^\infty \log x\, \mu_t(dx) &= -\int_{-t}^{0} \log\left[-\psi_Y^{(-1)}(x)\right] dx - \int_{-t}^{0} \log\left[\frac{1+x}{t+x}\right] dx \\ &= -\int_{-t}^{0} \log\left[-\psi_Y^{(-1)}(x)\right] dx + (1-t)\log(1-t) + t\log t.\end{aligned}$$

The integrated Lyapunov exponent is one half of this expression, and we can obtain the marginal Lyapunov exponent by differentiating over t:

$$\begin{aligned}f(t) &= \frac{1}{2}\left(-\log\left[-\psi_Y^{(-1)}(-t)\right] + \log t - \log(1-t)\right) \\ &= -\frac{1}{2}\log\left[\left(1-\frac{1}{t}\right)\psi_Y^{(-1)}(-t)\right] \\ &= -\frac{1}{2}\log\left[S_Y(-t)\right].\end{aligned}$$

QED.

Example

Let us consider again the case of identically distributed free X_i such that $X_i^* X_i$ has the Marchenko-Pastur distribution with the parameter $\lambda \geq 1$. In this case $S_Y(z) = (\lambda + z)^{-1}$. Hence, applying Theorem 160, we immediately obtain a formula for the marginal Lyapunov exponent:

$$f(t) = \frac{1}{2} \log(\lambda - t).$$

Inverting this formula, we obtain the formula for the distribution of Lyapunov exponents:

$$\mathcal{F}(x) = \begin{cases} 0, & \text{if } x < (1/2)\log(\lambda - 1), \\ e^{2x} + 1 - \lambda, & \text{if } x \in \left[\frac{1}{2}\log(\lambda - 1), \frac{1}{2}\log\lambda\right), \\ 1 & \text{if } x \geq \frac{1}{2}\log\lambda, \end{cases}$$

which is exactly the formula that we obtained earlier by a direct calculation from definitions. It is easy to check that a similar agreement holds also for $\lambda < 1$.

Corollary 161 *Let X and Y be such that X^*X and Y^*Y are bounded and have no atom at zero. Let f_X, f_Y, and f_{XY} denote the marginal Lyapunov exponents corresponding to variables X, Y and XY, respectively. Then*

$$f_{XY}(t) = f_X(t) + f_Y(t).$$

Proof: By Theorem 160,

$$\begin{aligned} f_{XY}(t) &= -\frac{1}{2}\log\left[S_{Y^*X^*XY}(-t)\right] \\ &= -\frac{1}{2}\log\left[S_{Y^*Y}(-t)\, S_{X^*X}(-t)\right] \\ &= f_X(t) + f_Y(t). \end{aligned}$$

QED.

Corollary 162 *If X is bounded and X^*X has no atom at zero, then the marginal Lyapunov exponent is (weakly) decreasing in t, i.e. $f_X'(t) \leq 0$.*

Proof: Because of Theorem 160, we need only to check that $S(t)$ is (weakly) decreasing on the interval $[-1, 0]$, and this was proved by Bercovici and Voiculescu in Proposition 3.1 on page 225 of Bercovici and Voiculescu (1992). QED.

Corollary 163 *If X is bounded and X^*X has no atom at zero, then the largest Lyapunov exponent equals* $(1/2)\log E(X^*X)$.

Proof: This follows from the previous Corollary and the fact that $S_Y(0) = 1/E(Y)$. QED.

Remark: It is interesting to compare this result with the result in Cohen and Newman (1984), which shows that the norm of the product of $N \times N$ i.i.d. random matrices $X_1, \ldots, X_n$ grows exponentially when n increases, and that the asymptotic growth rate approaches $\frac{1}{2}\log E(\mathrm{tr}\,(X_1^*X_1))$ if $N \to \infty$ and matrices are scaled appropriately. The assumption in Cohen and Newman (1984) about the distribution of matrix entries is that the distribution of $X_1^*X_1$ is invariant relative to orthogonal rotations of the ambient space. Since the growth rate of the norm of the product $X_1 \ldots X_n$ is another way to define the largest Lyapunov exponent of the sequence X_i, therefore the result in Cohen and Newman (1984) is in agreement with Corollary 163.

The main result of Theorem 160 can also be reformulated as the following interesting identity:

Corollary 164 *If Y is bounded, self-adjoint, and positive, and if $\{P_t\}$ is a family of projections which are free of Y and such that $E(P_t) = t$, then*

$$\begin{aligned}\log S_Y(-t) &= -\frac{d}{dt}\left[\lim_{\lambda\to 0} E\log^{+\lambda}(P_tYP_t)\right] \\ &= -2\frac{d}{dt}\left[\log\det\left(\sqrt{Y}P_t\right)\right].\end{aligned}$$

Conversely, we can express the determinant in terms of the S-transform:

Corollary 165 *If X is bounded and invertible, then*

$$\det(X) = \exp\left\{-\frac{1}{2}\int_0^1 \log S_{X^*X}(-t)\,dt\right\}.$$

16.6 Conclusion

One interesting remaining question is how the obtained results are related to the infinite-dimensional analogue of Newman's non-linear transformation, which can be defined as follows: Let $K(dt)$ be the spectral probability measure for $\sqrt{X_i^*X_i}$. Then for a certain range of x, we define $H(x)$ as the solution of the following integral equation:

$$\int \frac{t^2}{H(x)\,x^2 + (1 - H(x))\,t^2} K(dt) = 1.$$

A claim suggested by Newman's results about random matrices is that if $X_i^* X_i$ is invertible, then $H(x)$ is the distribution function for e^{μ}, where μ is the marginal Lyapunov exponent of X_i.

17 CLT for multiplicative free convolutions on the unit circle

17.1 Introduction

Suppose X is a unitary n-by-n matrix. Then X has n eigenvalues, which are all located on the unit circle. If we give each eigenvalue a weight of n^{-1}, then we can think about the distribution of these eigenvalues as a probability distribution supported on n points of the unit circle. More generally, if X is a unitary operator in a finite von Neumann algebra, then we can define a spectral probability distribution of X, which is supported on the unit circle (see, e.g., Section 1.1 in Hiai and Petz (2000)).

If we have several unitary operators X_1, ..., X_n, then it is natural to ask about the spectral distribution of their product. In general, we cannot determine this distribution without more information about relations among operators X_1, ..., X_n. However, if X_1, ..., X_n are infinite-dimensional and, in a certain sense, in a general position relative to each other, then the spectral distribution of their product is computable. The idea of a general position was formalized by Voiculescu in his concept of freeness of operators (see Voiculescu (1983), Voiculescu (1986), and a textbook by Hiai and Petz (2000)). If operators X_1, ..., X_n are free and unitary and their distributions are μ_1, ..., μ_n, respectively, then the distribution of their product is determined uniquely. This distribution is called the free multiplicative convolution of measures $\mu_1, ..., \mu_n$ and denoted as $\mu_1 \boxtimes ... \boxtimes \mu_n$.

What can we say about the asymptotic behavior of $\mu^{(n)} =: \mu_1 \boxtimes ... \boxtimes \mu_n$, as n increases to infinity? In particular, what are necessary and sufficient conditions on μ_i that ensure that $\mu^{(n)}$ converges to the uniform distribution on the unit circle?

To answer this question, let us define the expectation with respect to the measure μ_i. This is a functional that maps functions analytic in a neighborhood of the unit circle to complex numbers:

$$E_{\mu_i} f =: \int_{|\xi|=1} f(\xi)\, d\mu_i(\xi).$$

If unitary operator X_i has the spectral probability distribution μ_i, then we will also write:

$$Ef(X_i) =: E_{\mu_i} f.$$

In particular, EX_i denotes $\int_{|\xi|=1} \xi d\mu_i(\xi)$. Then the answer is given by the following theorem:

Theorem 166 *Suppose $\{X_i\}$ are free unitary operators with spectral measures μ_i. The measure $\mu^{(n)}$ of their product $\Pi_n = X_1 ... X_n$ converges to the uniform measure on the unit circle if and only if at least one of the following situations holds:*
(i) There exist two indices $i \neq j$ such that $EX_i = EX_j = 0$;
(ii) There exists exactly one index i such that $EX_i = 0$, and $\prod_{k=i+1}^{n} EX_k \to 0$ as $n \to \infty$;
(iii) There exists exactly one index i such that X_i has the uniform distribution;
(iv) $EX_k \neq 0$ for all k, and $\prod_{k=1}^{n} EX_k \to 0$ as $n \to \infty$.

In other words, convergence of $\mu^{(n)}$ to the uniform law implies that $\prod_{k=1}^{n} EX_k \to 0$, and the only case when the reverse implication fails is when $EX_i = 0$ for exactly one X_i, the measure μ_i is not uniform, and $\prod_{k=i+1}^{n} EX_k \nrightarrow 0$ as $n \to \infty$. Note that cases (ii) and (iii) above are not exclusive. It may happen that both μ_i is uniform and $\prod_{k=i+1}^{n} EX_k \to 0$ as $n \to \infty$. In this case, both (ii) and (iii) hold, and $\mu^{(n)}$ converges to the uniform law.

This theorem can be thought of as a limit theorem about free multiplicative convolutions of measures on the unit circle. There is some literature about traditional multiplicative convolutions of measures on the unit circle, or more generally, about convolutions of measures on compact groups. For the unit circle, this investigation was started by Levy (1939). Then it was continued by Kawada and Itô (1940), who studied compact groups, and Dvoretzky and Wolfowitz (1951) and Vorobev (1954), who both considered the case of commutative finite groups. These researchers found an important necessary condition for convergence of convolutions to the uniform law. This condition requires that there should be no normal subgroup such that the support of the convolved measures is supported entirely in an equivalence class relative to this subgroup. This condition is sufficient if summands are identically distributed. If they are not, then there are some sufficient and necessary conditions, which is are especially useful if the group is cyclic. A textbook presentation with further references can be found in Grenander (1963).

Recent investigations of convolutions on groups are mostly concerned with the speed of convergence of convolved measures to the uniform law. For a description of progress in this direction, the reader can consult surveys in Diaconis (1988) and Saloff-Coste (2004).

It turns out that free convolutions converge to the uniform law under much weaker conditions than usual convolutions. As an example, consider the distributions that are concentrated on -1 and $+1$. Let measure μ_k put the weight p_k on $+1$. Then usual convolutions remain concentrated on -1 and $+1$, and therefore they have no chance to converge to the uniform distribution on the unit circle. In contrast, we will show that free convolutions do converge to the uniform law, provided that either $\prod_{k=k_0}^{n}(2p_k - 1) \to 0$ for arbitrarily large k_0, or there exist two indices i and j such that $p_i = p_j = 1/2$.

The rest of this section is organized as follows. In Section 17.2 we outline the proof. Section 17.3 derives some auxiliar results that will be used in the proof. Section 17.4 proves the main result (Theorem 166). Section 17.5 derives the key estimate used in the proof. And Section 17.6 concludes.

17.2 Outline of the proof

Let $(\mathcal{A}, E)$ be a non-commutative probability space and $\{X_i\}_{i=1}^{\infty}$ be a sequence of free unitary operators (random variables) from this space. Let Π_n denote the partial products: $\Pi_n = X_1 ... X_n$. We denote $E(X_i)$ as a_i, and $E(\Pi_n)$ as $a_{(n)}$. First, note that it is enough to consider the case when all a_i are real and non-negative. Indeed, $e^{i\theta_n}\Pi_n$ converges in distribution to the uniform law if and only if Π_n converges in distribution to the uniform law. Therefore if a_i is not real and positive, then we can replace X_i with $e^{-i \arg a_i} X_i$ without affecting the convergence of Π_n.

We divide the analysis into the following cases:

Case I $a_{(n)} \nrightarrow 0$.

Case II $a_{(n)} \to 0$, and there are at least two indices, i and j, such that $a_i = a_j = 0$.

Case III $a_{(n)} \to 0$, and for all i, $a_i > 0$.

Subcase III.1 $\liminf a_i = 0$.

Subcase III.2 $\liminf a_i = \underline{a} > 0$.

Case IV $a_{(n)} \to 0$, and there exists exactly one index i, such that $a_i = 0$.

We will show that without loss of generality we can assume in this case that $a_1 = 0$, and $a_k > 0$ for all $k > 1$.

Subcase IV.1 X_1 has the uniform distribution.

Subcase IV.2 X_1 does not have the uniform distribution, and $\prod_{k=2}^n a_n \to 0$ as $n \to \infty$.

Subcase IV.3 X_1 does not have the uniform distribution, and $\prod_{k=2}^n a_n \nrightarrow 0$ as $n \to \infty$.

We will show that Π_n does not converge to the uniform law if and only if either Case I or Case IV.3 holds.

17.3 Auxiliary lemmas

We will use the Lemmas below:

Lemma 167 *Suppose A and B are free unitary operators, $|E(A)| \leq a$ and $|E(B)| \leq b$. Then*

$$\left|E\left[(AB)^k\right]\right| \leq M_k \max(a, b)$$

for some absolute constants M_k.

Proof: If we expand $E\left[(AB)^k\right]$ using Theorem 8, then we can observe that each term in the expansion contains either $E(A)$ or $E(B)$ as a separate multiple. The remaining multiples in this term are ≤ 1 in absolute value; therefore, we can bound each term by $\max(a, b)$. The number of terms in this expansion is bounded by a constant, M_k. Therefore, $\left|E\left[(AB)^k\right]\right|$ is bounded by $M_k \max(a, b)$. QED.

In the following lemmas we use the fact that the sequence of measures μ_i, supported on the unit circle, converges to the uniform law if and only if all moments converge to 0. (That is, for each k, $\int_{|\xi|=1} \xi^k d\mu_i(\xi) \to 0$ as $i \to \infty$.) For completeness we give a proof of this result below

Recall that the *ψ-function* of a bounded random variable X is defined as

$$\psi_X(z) =: \sum_{k=1}^{\infty} E\left(X^k\right) z^k. \tag{74}$$

If X is unitary operator with the spectral measure μ, then we can write:

$$\psi_\mu(z) = \int_{|\xi|=1} \frac{1}{1 - \xi z} d\mu(\xi) - 1.$$

Let us define $c_k^{(i)} =: E\left(X_i^k\right) = \int_{|\xi|=1} \xi^k d\mu_i(\xi)$. Note that for a fixed i, $c_k^{(i)}$ are coefficients in the Taylor series of $\psi_i(z)$, i.e., of the ψ-function of the measure μ_i.

Lemma 168 *Let μ_i be a sequence of measures supported on the unit circle. If for each k the coefficients $c_k^{(i)} \to 0$ as $i \to \infty$, then $\psi_i(z) \to 0$ uniformly on compact subsets of the unit disc.*

Proof: Let Ω be a compact subset of the unit disc, and let $\Omega \subset D_r$, where D_r denotes a closed disc with the radius $r < 1$. Fix an $\varepsilon \in (0,1)$. Then we can find such a k_0 that

$$\left|\sum_{k=k_0}^{\infty} c_k^{(j)} z^k\right| < \varepsilon/2$$

for all $z \in D_r$ and all j. Indeed, $\left|c_k^{(j)}\right| \leq 1$, and therefore,

$$\left|\sum_{k=k_0}^{\infty} c_k^{(j)} z^k\right| \leq \frac{r^{k_0}}{1-r},$$

so we can take $k_0 = \log(\varepsilon(1-r)/2)/\log r$.

Given k_0, we choose a j_0 so large that for all $j > j_0$ and all $k < k_0$, we have $\left|c_k^{(j)}\right| < \varepsilon/(2k_0)$. This is possible because by assumption for each k, coefficients $c_k^{(j)}$ converge to zero as $j \to \infty$, and we consider only a fixed finite number of possible k.

Consequently,

$$\left|\sum_{k=1}^{k_0-1} c_k^{(j)} z^k\right| \leq \sum_{k=1}^{k_0-1} \left|c_k^{(j)}\right| < \varepsilon/2$$

for every $j > j_0$ and all $z \in D_r$. Therefore,

$$\left|\sum_{k=1}^{\infty} c_k^{(j)} z^k\right| < \varepsilon$$

for every $j > j_0$ and all $z \in D_r$. Therefore, $\psi_j(z) \to 0$ uniformly on D_r, and therefore on Ω. Since Ω was arbitrary, we have proved that $\psi_j(z) \to 0$ uniformly on compact subsets of the unit disc. QED.

Lemma 168, formula (34), p. 70. and Theorem 74, p. 69, imply that if all moments of μ_j converge to 0, then $\mu_j \to \nu$, where ν is the uniform distribution on the unit circle. QED.

Lemma 169 *Suppose $\{A_n\}_{n=1}^{\infty}$ is a sequence of unitary operators that converges in distribution to the uniform law. Let $\{B_n\}_{n=1}^{\infty}$ be another sequence of unitary operators, and let the operator B_n be free of the operator A_n for every n. Then the sequence of products B_nA_n converges in distribution to the uniform law. Also, the sequence A_nB_n converges to the uniform law.*

Proof: Let $a_k^{(n)} = E\left((A_n)^k\right)$. By assumption, for each fixed k, the moment $a_k^{(n)} \to 0$ as $n \to \infty$. If we represent $E\left((B_nA_n)^k\right)$ as a polynomial in individual moments of B_n and A_n, then all terms of this polynomial contain at least one of the moments $a_i^{(n)}$, $i \leq k$, which are perhaps multiplied by some other moments. All of these other moments are less than 1 in absolute value. Therefore, we can write the following estimate:

$$\left| E\left((B_nA_n)^k\right) \right| \leq M_k \max_{i \leq k} \left\{ a_i^{(n)} \right\},$$

where M_k is the number of terms in the polynomial. If k is fixed and n is growing, then the assumption that A_n converges in distribution to the uniform law implies that $\max_{i \leq k} \left\{ a_i^{(n)} \right\}$ converges to zero. Therefore, all moments of B_nA_n converge to zero as $n \to \infty$, and therefore, by Lemma 168 and Theorem 74, p. 69, the sequence B_nA_n converges in distribution to the uniform law. A similar argument proves that A_nB_n converges in distribution to the uniform law. QED.

Lemma 170 *Suppose that B is a unitary operator, $\{A_n\}$ is a sequence of unitary operators, B is free from each of A_n, $E(B) \neq 0$, and the sequence A_n does not converge to uniform law. Then the sequence of products BA_n does not converge to the uniform law.*

Proof: The condition that the sequence A_n does not converge to the uniform law means that for some fixed k the sequence of k-th moments of A_n does not converge to zero as $n \to \infty$. Let k be the smallest of these indices. By selecting a subsequence we can assume that $\left| E\left(A_n^k\right) \right| > \alpha > 0$ for all n. Consider $E\left((BA_n)^k\right)$:

$$E\left((BA_n)^k\right) = [E(B)]^k \, E\left(A_n^k\right) + ...,$$

The number of the terms captured by ... is finite and depends only on k. Each of these terms includes at least one of $E\left(A_n^i\right)$ where $i < k$, and other multipliers in this term are less than 1 in absolute value. Therefore, each of these terms converges to zero. Hence, for any $\varepsilon > 0$, there exist such N that for all $n > N$, the sum of the terms captured by ... is less than ε in absolute value. Take $\varepsilon = |E(B)|^k \alpha/2$. Then for $n > N$, we have:

$$\left|E\left((BA_n)^k\right)\right| \geq |E(B)|^k \alpha/2.$$

Therefore, the sequence of products BA_n does not converge to the uniform law. QED.

Lemma 171 *Suppose that B is a unitary random variable, $\{A_n\}$ is a sequence of unitary random variables, B is free from each of A_n, B is not uniform, and the sequence of expectations $E(A_n)$ does not converge to zero. Then the sequence of products BA_n does not converge to the uniform law.*

Proof: By selecting a subsequence we can assume that $|E(A_n)| > \alpha > 0$ for all n. The assumption that B is not uniform means that for some k, $E\left(B^k\right) \neq 0$. Let k be the smallest of such k. Consider $E\left((BA_n)^k\right)$:

$$E\left((BA_n)^k\right) = [E(A_n)]^k E\left(B^k\right) + ...,$$

Each of the terms in ... includes one of $E(B^i)$ where $i < k$. Therefore, all terms in ... are zero. Hence,

$$\left|E\left((BA_n)^k\right)\right| = \left|[E(A_n)]^k E\left(B^k\right)\right| > \alpha^k \left|E\left(B^k\right)\right|.$$

Therefore, the sequence of products BA_n does not converge to the uniform law. QED.

17.4 Analysis

We use the following notation: ψ_i and S_i denote ψ- and S-functions for variables X_i (and measures μ_i), and $\psi_{(n)}$ and $S_{(n)}$ denote these functions for variables Π_n (and measures $\mu^{(n)}$).

Case I: $a_{(n)} \nrightarrow 0$**.**

Since $E(\Pi_n) = a_{(n)}$, therefore, if $a_{(n)} \nrightarrow 0$, then $E(\Pi_n) \nrightarrow 0$. Hence, Π_n cannot converge to the uniform measure on the unit circle.

Case II $a_{(n)} \rightarrow 0$, **and there are at least two indices** i **and** j **such that** $a_i = a_j = 0$.

Assume without loss of generality that $j > i$. Consider Π_n with $n \geq j$ and define $X =: X_1...X_i$ and $Y =: X_{i+1}..X_n$. Then $\Pi_n = XY$, and $E(Y) = E(X) = 0$. Using Lemma 167, we obtain that $\left|E\left[(\Pi_n)^k\right]\right| = 0$ for every $k > 0$. Therefore, the ψ-function of Π_n is zero, and Π_n has the uniform distribution on the unit circle.

Case III $a_{(n)} \to 0$, **and for all** i, $a_i > 0$.

Subcase III.1 $\liminf a_i = 0$.

In this case we can find a subsequence a_{n_i} that monotonically converges to zero.

Now, consider Π_j, where $j \in [n_i, n_{i+1})$. Then we can write $\Pi_j = XY$, where $X = X_1...X_{n_i-1}$, and $Y = X_{n_i}...X_j$. Then $EX \leq a_{n_{i-1}}$ and $EY \leq a_{n_i} \leq a_{n_{i-1}}$.

Applying Lemma 167 we get

$$\left|E\left(\Pi_j^k\right)\right| \leq M_k a_{n_{i-1}}.$$

This implies that for a fixed k, $\left|E\left(\Pi_j^k\right)\right|$ approaches zero as $j \to \infty$. By Lemma 168 and Proposition 74, this establishes that Π_j converges to the uniform law.

Case III $a_{(n)} \to 0$, **and for all** i, $a_i > 0$

Subcase III.2 $\liminf a_i = \underline{a} > 0$.

Let us choose such an a that $0 < a < \underline{a}$. Starting from some j_0, $a_j \in (a, 1)$. Let $\widetilde{\Pi}_n = X_{j_0}...X_{n+j_0-1}$. Then, by Lemmas 169 and 170, $\widetilde{\Pi}_n$ converges to the uniform law if and only if Π_n converges to the uniform law. Hence, without loss of generality we can restrict our attention to the case when $a_k \in (a, 1)$ for all k.

In this case we have to use the analytic apparatus developed by Voiculescu for free multiplicative convolutions. Let $\psi_X^{-1}(u)$ denote the functional inverse of $\psi_X(z)$ in a neighborhood of $z = 0$, where $\psi_X(z)$ is as defined in (74). (This inversion is possible provided that $E(X) \neq 0$.) Define also

$$S_X(u) = \frac{u+1}{u}\psi_X^{-1}(u).$$

By Voiculescu multiplication theorem, if X and Y are bounded free random variables and both $E(X)$ and $E(Y)$ are not zero, then $S_{XY}(z) = S_X(z) S_Y(z)$.

To prove convergence to the uniform law, we have to establish that for every $k > 0$ the coefficient $c_k^{(n)}$ in the Taylor expansion of function $\psi_{(n)}(z)$ approaches zero as $n \to \infty$. We know from Lemma 60, p. 58, that

$$kc_k^{(n)} = \operatorname*{res}_{z=0} \frac{1}{\left[\psi_{(n)}^{-1}(z)\right]^k};$$

therefore, our main task is to estimate this residual. This is the same as estimating the coefficient before the term z^{k-1} in the Taylor expansion of

$$f(z) = \left[\frac{z}{\psi_{(n)}^{-1}(z)}\right]^k.$$

We will approach this problem by using the Cauchy inequality (see Section 5.23 in Whittaker and Watson (1927)). Recall that this inequality says that

$$\left|kc_k^{(n)}\right| \leq \frac{M(r)}{r^{k-1}}, \tag{75}$$

where $r > 0$ is such that $f(z)$ is analytic inside $|z| = r$ and

$$M(r) =: \max_{|z|=r} |f(z)|.$$

It is easy to note that the constant in the Taylor expansion of $z/\psi_{(n)}^{-1}(z)$ is $a_{(n)}$ which approaches zero as $n \to \infty$. So $M(0) = a_{(n)}$. The main question is how large we can take r, so that $M(r)$ remains relatively small. In other words, we want to minimize the right-hand side of (75) by a suitable choice of r.

Proposition 172 *Suppose that $EX_i = a_i > a$ for each i and that $a_{(n)} =: \prod_{i=1}^n a_i \to 0$. Let $\alpha_i = 1 - a_i$. Then for all sufficiently large n, the following inequality holds:*

$$\left|c_k^{(n)}\right| \leq \left(\frac{C}{a^2}\right)^k \left[\left(\sum_{i=1}^n \alpha_i\right) \exp\left(-\sum_{i=1}^n \alpha_i\right)\right]^k,$$

where $C = 2^{17}$.

Proof: The main tool in the proof is the following proposition:

Proposition 173 *Suppose that $\alpha_i =: 1 - a_i < 1 - a$ for each i, and that z and n are such that*

$$|z| \leq \frac{a^2}{6684} \min\left\{1, \left(\sum_{i=1}^n \alpha_i\right)^{-1}\right\}.$$

Then,

$$\left|\frac{z}{\psi_{(n)}^{-1}(z)}\right|^k \leq \left(2e^2\right)^k \left(\prod_{i=1}^n a_i\right)^k.$$

We will prove this proposition in the next section and assume for now that it holds.

Lemma 174 *Suppose* $1 \geq a_k > 0$ *for all* k, *and let* $\alpha_i =: 1 - a_i$. *Then* $\prod_{i=1}^{n} a_i \to 0$ *if and only if* $\sum_{i=1}^{n} \alpha_i \to \infty$.

This is a standard result. For a proof see Section 2.7 in Whittaker and Watson (1927). Since $\log(1 - \alpha_i) \leq -\alpha_i$, we also have the following estimate.

$$\prod_{i=1}^{n} a_i \leq \exp\left(-\sum_{i=1}^{n} \alpha_i\right). \tag{76}$$

Let n_0 be so large that $\sum_{i=1}^{n_0} \alpha_i > 1$. (We can find such n_0 because by Lemma 174, $\sum_{i=1}^{n} \alpha_i \to \infty$ as $n \to \infty$.) In particular, this implies that $\sum_{i=1}^{n} \alpha_i > 1$ for every $n \geq n_0$. Define $r_n =: a^2 \left(\sum_{i=1}^{n} \alpha_i\right)^{-1}/6684$. Then, using Proposition 173 and formulas (75) and (76), we get:

$$\begin{aligned} \left|kc_k^{(n)}\right| &\leq (2e^2)^k \left(\prod\nolimits_{i=1}^{n} a_i\right)^k \left(\frac{6684}{a^2}\sum\nolimits_{i=1}^{n} \alpha_i\right)^{k-1} \\ &\leq \left[\frac{2^{17}}{a^2}\left(\sum\nolimits_{i=1}^{n} \alpha_i\right)\exp\left(-\sum\nolimits_{i=1}^{n} \alpha_i\right)\right]^k, \end{aligned}$$

provided that $n \geq n_0$. QED.

Using Lemma 174, we get the following Corollary:

Corollary 175 *If the assumptions of Proposition 172 hold, then for each* k, *the coefficient* $c_k^{(n)} \to 0$ *as* $n \to \infty$.

This Corollary shows that in Case III.2 the product Π_n converges to the uniform law.

Case IV $a_{(n)} \to 0$, **and there exists exactly one index** i, **such that** $a_i = 0$.

First, we want to show that without loss of generality we can assume in this case that $a_1 = 0$, and $a_k > 0$ for all $k > 1$. Indeed, suppose $a_i = 0$ for $i > 1$. Let $X = X_1 \ldots X_{i-1}$ and let $\widetilde{\Pi}_n = X_i \ldots X_{i+n-1}$. Then $E(X) \neq 0$, and using Lemmas 169 and 170, we conclude that Π_n converges to the uniform law if and only if $\widetilde{\Pi}_n$ converges to the uniform law.

Subcase IV.1 X_1 **has the uniform distribution.**

In this case all moments of X_1 are zero, i.e., $E\left(X_1^k\right) = 0$ for all $k > 0$, and Theorem 8 implies that all moments of Π_n are zero. Therefore, Π_n is uniform for all n.

Subcase IV.2 X_1 **does not have the uniform distribution, and** $\prod_{k=2}^{n} a_n \to 0$ **as** $n \to \infty$.

By Case III, the product $X_2 ... X_n$ converges to the uniform law, and using Lemma 169, we conclude that Π_n also converges to the uniform law.

Subcase IV.3 X_1 **does not have the uniform distribution and** $\prod_{k=2}^{n} a_n \nrightarrow 0$ **as** $n \to \infty$.

Applying Lemma 171 to $B = X_1$ and $A = X_2 ... X_n$, we conclude that Π_n does not converge to the uniform law.

17.5 Proof of Proposition 173

Let

$$f(z) =: \left(\frac{z}{\psi_{(n)}^{-1}(z)}\right)^k .$$

Using Theorem 49, 49, we can write this function as follows:

$$f(z) = \left(\frac{z^n}{(1+z)^{n-1}} \prod_{i=1}^{n} \frac{1}{\psi_i^{-1}(z)}\right)^k . \tag{77}$$

We want to estimate $|f(z)|$ for all sufficiently small z. The plan of the attack is as follows. First, we will show that if a_i is close to 1, then $\psi_i^{-1}(z)$ behaves approximately as $z/(a_i + z)$. The quantitative version of this statement is in Lemma 181. This behavior implies that $f(z)$ is approximately $\left[a_{(n)} \prod_{i=1}^{n} \left[(1 + z/a_i)/(1+z)\right]\right]^k$ and it is easy to show that the product is convergent if $z \le C \sum (1 - a_i)$. This implies the statement of Proposition 173.

To implement this plan, we start with some auxiliary estimates, which will later allow us to estimate $\psi_i(z)$, and then $\psi_i^{-1}(z)$ for small z.

Lemma 176 *Suppose μ is a probability measure on $[-\pi, \pi)$ such that*

$$\left|\int_{-\pi}^{\pi} \left(e^{i\theta} - 1\right) d\mu(\theta)\right| \le \alpha. \tag{78}$$

Then, i)

$$\int_{-\pi}^{\pi} \theta^2 d\mu(\theta) \le \frac{\pi^2}{2}\alpha < 5\alpha;$$

ii)

$$\left|\int_{-\pi}^{\pi} \theta d\mu(\theta)\right| \leq \left(1+\frac{\pi^3}{12}\right)\alpha < 3\alpha, \textit{ and}$$

iii) if $k > 2$, *then*

$$\int_{-\pi}^{\pi} |\theta|^k \, d\mu(\theta) \leq \frac{\pi^k}{2}\alpha.$$

Proof: Condition (78) implies that

$$\int_{-\pi}^{\pi} (1-\cos(\theta))\, d\mu(\theta) \leq \alpha$$

and that

$$\left|\int_{-\pi}^{\pi} \sin(\theta)\, d\mu(\theta)\right| \leq \alpha.$$

Since $1-\cos\theta \geq (2/\pi^2)\,\theta^2$, from the first of these inequalities we infer that:

$$\int_{-\pi}^{\pi} \theta^2 d\mu(\theta) \leq \left(\pi^2/2\right)\alpha,$$

which proves claim i) of the lemma.

Next, note that $|\sin\theta-\theta| \leq |\theta|^3/6$, and that

$$\frac{1}{6}\int_{-\pi}^{\pi} |\theta|^3 \, d\mu(\theta) \leq \frac{\pi}{6}\int_{-\pi}^{\pi} \theta^2 d\mu(\theta) \leq \frac{\pi^3}{12}\alpha.$$

Therefore,

$$\begin{aligned}\left|\int_{-\pi}^{\pi} \theta d\mu(\theta)\right| &\leq \left|\int_{-\pi}^{\pi} \sin(\theta)\, d\mu(\theta)\right| + \left|\int_{-\pi}^{\pi} (\theta-\sin(\theta))\, d\mu(\theta)\right| \\ &\leq \alpha + \frac{\pi^3}{12}\alpha.\end{aligned}$$

This proves claim ii) of the lemma.

For claim iii), note that

$$\int_{-\pi}^{\pi} |\theta|^k \, d\mu(\theta) \leq \pi^{k-2}\int_{-\pi}^{\pi} |\theta|^2 \, d\mu(\theta) \leq \frac{\pi^k}{2}\alpha.$$

QED.

Lemma 177 *Suppose Condition (78) holds, and k is a positive integer. Then*

$$\left|\int_{-\pi}^{\pi}\left(e^{ik\theta}-1\right)d\mu\left(\theta\right)\right|\leq 7k^{3}\alpha.$$

Proof: First, remark that $\left|1-\cos\left(k\theta\right)\right|\leq\left(k\theta\right)^{2}/2$ and therefore

$$\begin{aligned}\left|\int_{-\pi}^{\pi}\left(\cos k\theta-1\right)d\mu\left(\theta\right)\right| &\leq \frac{k^{2}}{2}\int_{-\pi}^{\pi}\theta^{2}d\mu\left(\theta\right)\\ &\leq \frac{\pi^{2}k^{2}}{4}\alpha.\end{aligned}$$

Next, we will use $\left|\sin\left(k\theta\right)-k\theta\right|\leq\left(k\left|\theta\right|\right)^{3}/6$ and write

$$\begin{aligned}\left|\int_{-\pi}^{\pi}\sin\left(k\theta\right)d\mu\left(\theta\right)\right| &\leq \left|\int_{-\pi}^{\pi}k\theta d\mu\left(\theta\right)\right|+\left|\frac{1}{6}\int_{-\pi}^{\pi}\left(k\left|\theta\right|\right)^{3}d\mu\left(\theta\right)\right|\\ &\leq k\left(1+\frac{\pi^{3}}{12}\right)\alpha+\frac{1}{6}k^{3}\frac{\pi^{3}}{2}\alpha\\ &\leq k^{3}\left(1+\frac{\pi^{3}}{6}\right)\alpha.\end{aligned}$$

Consequently,

$$\begin{aligned}\left|\int_{-\pi}^{\pi}\left(e^{ik\theta}-1\right)d\mu\left(\theta\right)\right| &\leq \alpha\sqrt{\frac{\pi^{4}k^{4}}{16}+k^{6}\left(1+\frac{\pi^{3}}{6}\right)^{2}}\\ &\leq 7k^{3}\alpha.\end{aligned}$$

QED.

Lemma 178 *Let X be unitary and $EX=a>0$. If $|z|\leq 1/2$ and $1-a\leq\alpha$, then*

$$\left|\psi_{X}\left(z\right)-\frac{az}{1-z}\right|\leq 716\alpha\left|z\right|^{2}.$$

Proof: We can write:

$$\psi_{X}\left(z\right)-\frac{az}{1-z}=\sum_{k=2}^{\infty}\left(E\left(X^{k}\right)-a\right)z^{k}.$$

Therefore, using Lemma 177, we estimate:

$$\begin{aligned}\left|\psi_X(z)-\frac{az}{1-z}\right| &\leq \sum_{k=2}^{\infty}\left(\left|E\left(X^k\right)-1\right|+|1-a|\right)z^k \\ &\leq 7\alpha|z|^2\sum_{k=0}^{\infty}\left[(k+2)^3+1/7\right]|z|^k \\ &\leq 716\alpha|z|^2\end{aligned}$$

QED.

To derive a similar estimate for $\psi_X^{-1}(z)$, we need a couple of preliminary lemmas.

Lemma 179 *Suppose X is unitary and $EX=a>0$. Then the function $\psi_X(z)$ has only one zero ($z=0$) in the area $|z|<a/3$. If $|z|=a/3$, then $|\psi_X(z)|\geq a^2/6$.*

Proof: Write the following estimate:

$$\begin{aligned}|\psi_X(z)-az| &= \left|\sum_{k=2}^{\infty}E\left(X^k\right)z^k\right| \\ &\leq \frac{|z|}{1-|z|}|z|<\frac{a}{2}|z|,\end{aligned}$$

if $|z|<a/3$. By Rouche's theorem, $\psi_X(z)$ has only one zero in $|z|<a/3$. The second claim also follows immediately from this estimate. QED.

Lemma 180 *Suppose X is unitary and $EX=a>0$. Then the function $\psi_X^{-1}(z)$ is analytical for $|z|<a^2/6$. In this area it can be represented as*

$$\psi_X^{-1}(z)=\frac{z}{a}(1+zv(z)),$$

where $v(z)$ is an analytical function. If $|z|\leq a^2/12$, then

$$|v(z)|\leq\frac{12}{a^2}.$$

Proof: Using Lagrange's formula, we can write

$$\psi_X^{-1}(z)=\frac{z}{a}+\sum_{k=2}^{\infty}c_kz^k,$$

where

$$c_k = \frac{1}{2\pi i} \frac{1}{k} \oint_{\gamma} \frac{du}{\left[\psi_X(u)\right]^k}.$$

By the previous Lemma, we can use the circle with the center at 0 and radius $a/3$ as γ, and then we can estimate c_k as follows:

$$|c_k| \leq \frac{a/3}{k\left(a^2/6\right)^k} = \frac{2}{ka}\left(\frac{6}{a^2}\right)^{k-1}.$$

It follows that the power series for $\psi_X^{-1}(z)$ converges in $|z| < a^2/6$. In particular it follows that $\psi_X^{-1}(z)$ can be represented as

$$\psi_X^{-1}(z) = \frac{z}{a}\left(1 + zv(z)\right),$$

where $v(z)$ is an analytical function. For $v(z)$ we have the following estimate:

$$|v(z)| \leq \frac{6}{a^2} \frac{1}{1 - \frac{6}{a^2}|z|}.$$

So if $|z| \leq a^2/12$, then

$$|v(z)| \leq \frac{12}{a^2}.$$

QED.

Lemma 181 *Let X be unitary and $EX = a > 0$. If $|z| \leq a^2/12$, and $\alpha \geq 1 - a$, then*

$$\left|\frac{\psi_X^{-1}(z)}{z/(a+z)} - 1\right| \leq \frac{3342\alpha}{a^2}|z|.$$

Proof: First of all, note that Lemma 180 implies that

$$\left|\psi_X^{-1}(z)\right| \leq \frac{2}{a}|z|$$

for $|z| \leq a^2/12$.

Now we use the functional equation for $\psi_X^{-1}(z)$:

$$\psi_X\left(\psi_X^{-1}(z)\right) = z.$$

If $|z| \leq a^2/12$, then $\left|\psi_X^{-1}(z)\right| \leq 2|z|/a \leq a/6 < 1/2$ and we can apply Lemma 178 to get:

$$\begin{aligned}\left|z - \frac{a\psi_X^{-1}(z)}{1-\psi_X^{-1}(z)}\right| &\leq 716\alpha\left|\psi_X^{-1}(z)\right|^2 \\ &\leq \frac{2864\alpha}{a^2}|z|^2.\end{aligned}$$

Next, we write this as

$$\begin{aligned}\left|z - (a+z)\psi_X^{-1}(z)\right| &\leq \left|1-\psi_X^{-1}(z)\right|\frac{2864\alpha}{a^2}|z|^2 \\ &\leq \frac{3342\alpha}{a^2}|z|^2.\end{aligned}$$

It follows that

$$\left|\frac{\psi_X^{-1}(z)}{z/(a+z)} - 1\right| \leq \frac{3342\alpha}{a^2}|z|.$$

QED.

Lemma 182 *Let $EX_i = a_i$ and assume that for each i, it is true that $a_i \geq a$. Assume also that $|z| \leq a^2/3342$ and let $\alpha_i =: 1 - a_i$. Then*

$$\left|\prod_{i=1}^n \frac{1}{\psi_i^{-1}(z)}\right| \leq \frac{\prod_{i=1}^n a_i}{|z|^n}\prod_{i=1}^n \frac{1}{1-c_i|z|}\left|\prod_{i=1}^n\left(1+\frac{z}{a_i}\right)\right|,$$

where $c_i = 3342\alpha_i/a_i^2$.

Proof: From Lemma 181 we infer that

$$\left|\psi_i^{-1}(z)\right| \geq \frac{z}{a_i}\frac{1}{1+z/a_i}\left(1-\frac{3342\alpha_i}{a_i^2}|z|\right).$$

Multiplying these inequalities together and inverting both sides, we get the desired result. QED.

Lemma 183 *Under the assumptions of the previous lemma, the following inequality holds:*

$$|f(z)| \leq \left(\frac{\prod_{i=1}^n a_i}{|1+z|}\prod_{i=1}^n\frac{1}{1-c_i|z|}\left|\prod_{i=1}^n\frac{1+z/a_i}{1+z}\right|\right)^k, \tag{79}$$

where $c_i = 3342\alpha_i/a_i^2$.

Proof: The claim of this lemma is a direct consequence of Lemma 182 and equality (77). QED.

We will estimate terms in the product on the right-hand side of (79) one by one.

Lemma 184 *Suppose that* $\alpha_i =: 1 - a_i < 1 - a$ *for each* i, *and that*

$$|z| \leq \frac{a^2}{6684} \min \left\{ 1, \left(\sum_{i=1}^{n} \alpha_i \right)^{-1} \right\}.$$

Then

$$\left| \prod_{i=1}^{n} \frac{1 + z/a_i}{1 + z} \right| \leq e.$$

Proof: We write:

$$\left| \prod_{i=1}^{n} \frac{1 + z/a_i}{1 + z} \right| = \exp \left(\operatorname{Re} \sum_{i=1}^{n} \log \left(1 + \frac{\alpha_i}{a_i} \frac{z}{1+z} \right) \right).$$

Recall that $\operatorname{Re} \log (1 + u) \leq |u|$ if $|u| < 1$. Under our assumption about $|z|$, it is true that

$$\left| \frac{\alpha_i}{a_i} \frac{z}{1+z} \right| < 1.$$

Therefore we can write:

$$\begin{aligned} \left| \prod_{i=1}^{n} \frac{1 + z/a_i}{1 + z} \right| &\leq \exp \left(\left| \frac{z}{1+z} \right| \sum_{i=1}^{n} \frac{\alpha_i}{a_i} \right) \\ &\leq \exp \left(\frac{2}{a} |z| \sum \alpha_i \right) \\ &\leq e. \end{aligned}$$

QED.

Lemma 185 *Suppose that* $\alpha_i =: 1 - a_i < 1 - a$ *for each* i, *and that*

$$|z| \leq \frac{a^2}{6684} \min \left\{ 1, \left(\sum_{i=1}^{n} \alpha_i \right)^{-1} \right\}.$$

Then,

$$\prod_{i=1}^{n} \frac{1}{1 - c_i |z|} \leq e,$$

where $c_i = 3342 \alpha_i / a_i^2$.

Proof: We use the inequality $\log(1-u) \geq -2u$, which is valid for $u \in (0, 1/2)$, and write:

$$\begin{aligned}
\prod_{i=1}^{n} \frac{1}{1 - c_i |z|} &= \exp\left(-\sum_{i=1}^{n} \log(1 - c_i |z|)\right) \\
&\leq \exp\left(2|z| \sum_{i=1}^{n} c_i\right) \\
&\leq \exp\left[\frac{6684}{a^2}\left(\sum_{i=1}^{n} \alpha_i\right)|z|\right] \\
&\leq e.
\end{aligned}$$

QED.

Finally, note that if $|z| \leq a^2/6684$, then $|z| \leq 1/2$ and, therefore, $\left|(1+z)^{-1}\right| \leq 2$. Collecting all the pieces, we obtain that if

$$|z| \leq \frac{a^2}{6684} \min\left\{1, \left(\sum_{i=1}^{n} \alpha_i\right)^{-1}\right\}$$

then:

$$|f(z)| \leq \left(2e^2\right)^k \left(\prod_{i=1}^{n} a_i\right)^k.$$

This completes the proof of Proposition 173.

17.6 Conclusion

We have derived sufficient and necessary conditions for the product of free unitary operators to converge in distribution to the uniform law. If *essential convergence* denotes the situation when the partial products continue to converge even after an arbitrary finite number of terms are removed, then the necessary and sufficient condition for essential convergence is that the products $\prod_{k_0}^{n} EX_i$ converges to zero for all k_0, that is, that the products of expectations essentially converge to zero. Essential convergence implies convergence. In addition, non-essential convergence can occur when there is either a term that has the uniform distribution, or there are two terms that have zero expectation. In the latter case convergence occurs because the product of these two terms has the uniform distribution.

Part IV

Free Point Processes and Free Extremes

In classical probability theory, one of the most important places is taken by the theory of extremal events. Recently, a similar theory has started to be developed in the context of free probability theory, in the role of independent random variables is played by freely independent non-commutative operators in a Hilbert space. In particular, (Ben Arous and Voiculescu 2006) have introduced and studied free extremal processes. The main object of this paper is to show that free extremes are naturally related to an object that we call a free point process.

The basic element in both classical and free theory of extremes is a measure μ. In the classical case, we take a sequence of i.i.d. random variables X_i, distributed according to μ, and then define a sequence of scaled maxima:

$$X^{(n)} = \max_{1 \le i \le n} \left\{ \frac{X_i - a_n}{b_n} \right\},$$

where a_n and b_n are certain constants.

If $F^{(n)}$ denotes the distribution function of $X^{(n)}$, then it was shown in the classical works by (Frechet 1927), (Fisher and Tippett 1928), and (Gnedenko 1943) that there are only 3 possible limit laws, to which a sequence of $F^{(n)}$ can converge, and that for a given measure μ, the distributions $F^{(n)}$ can converge to only one of these laws. In this case, it is said that the measure μ belongs to the domain of attraction of this limit law.

In the free case, a sequence of free self-adjoint operators X_i is taken, such that each of X_i has the spectral probability distribution μ. Using some ideas from (Ando 1989), (Ben Arous and Voiculescu 2006) have defined a maximum operation that maps any n-tuple of self-adjoint operators to another self-adjoint operator, which is called their maximum. By analogy with the classical case, the sequence of scaled maxima is defined as $X^{(n)} = \max_{1 \le i \le n} \left\{ \left(X_i - a_n I \right) / b_n \right\}$, and $F_f^{(n)}(x)$ is defined as the spectral distribution function of the self-adjoint operator $X^{(n)}$. Note that in general $F_f^{(n)}(x) \neq F^{(n)}(x)$.

The same question presents itself: When does the sequence of $F_f^{(n)}(x)$ converges?

Surprisingly, the answer to this question is very similar to the answer in the classical case: There are only 3 possible limit laws, and for a given μ, the distributions $F_f^{(n)}$ can converge to only one of them. As in the classical case, this allows defining domains of attraction of the free limit laws. A puzzling fact is that the limit laws are different in the classical and free cases, but the domains of attraction are the same!

In order to investigate this situation further, let us return to the classical case and consider the random point process N_n, which is defined by the following formula:

$$N_n = \sum_{i=1}^{n} \delta_{(X_i - a_n)/b_n}$$

When does this point process converges? It turns out that this question is intimately related to the convergence of the distributions $F^{(n)}$. If $F^{(n)}(x)$ converges to one of the classical limit laws $G(x)$, then the corresponding point process weakly converges to a Poisson random measure on every interval $[a, \infty)$, provided that $G(a) > 0$. Conversely, if N_n weakly converges on an interval $[a, \infty)$ to a Poisson random measure with the intensity measure $\lambda(dx)$, then $F^{(n)}(x)$ converges on the interval $[a, \infty)$ to a limit law $G(x)$. The limit law $G(x)$ and the intensity $\lambda(dx)$ are related by the equation $G(x) = \exp[-\lambda((x, \infty))]$.

What is the free analogue of the point process N_n? To motivate our definition, note that we can think about N_n as a linear functional on the space of measurable bounded functions: $\langle N_n, f \rangle =: \sum_{i=1}^{n} f((X_i - a_n)/b_n)$. This functional takes values in the space of bounded random variables. We define free point process analogously but prefer to work in a slightly greater generality and associate a free random process to any triangular array of free random variables

Let $\overline{\mathcal{A}}$ be the set of densely-defined closed operators affiliated with a von Neumann algebra $\mathcal{A}$, and let $\mathcal{B}_\infty(\mathbb{R})$ denote the set of all bounded, Borel measurable functions $f : \mathbb{R} \to \mathbb{R}$.

Definition 186 *Let $X_{i,n} \in \overline{\mathcal{A}}$, ($i = 1, \ldots, n$; $n = 1, \ldots$) be a triangular array of freely independent self-adjoint variables. Then the free point process M_n associated with the array $X_{i,n}$ is a sequence of $\mathcal{A}$-valued functionals on $\mathcal{B}_\infty(\mathbb{R})$, which are defined by the following formula:*

$$\langle M_n, f \rangle =: \sum_{i=1}^{n} f(X_{i,n}).$$

A triangular array of free variables that we use in applications to free extremes is, of course, $X_{i,n} = (X_i - a_n)/b_n$, where X_i is a sequence of free self-adjoint variables.

We can also define the concept of weak convergence of a free point process as a weak-$*$ convergence of the corresponding functionals. In the classical case, after a suitable scaling, the point process N_n converges to a Poisson random measure. It turns out that in the non-commutative case the free point process converges to a *free Poisson random measure*, which was recently defined in (Voiculescu 1998) and (Barndorff-Nielsen and Thorbjornsen 2005). Moreover, we have the following result.

Let $x^i = \inf\{x : G^i(x) > 0\}$, where $G^i(x)$, $i \in \{I, II, III\}$ is one of the classical extremal laws. On the interval $[x^i, \infty)$, let us define a measure $\lambda^i(dx)$ by the equality $\lambda^i((x,\infty)) = -\log G^i(x)$.

Theorem 187 *The following statements are equivalent:*
(i) μ belongs to the domain of attraction of the classical extremal limit law $G^i(x)$;
(ii) μ belongs to the domain of attraction of the free extremal limit law $G^i_f(x)$;
(iii) For some a_n and b_n, the point process N_n weakly converges on (x^i,∞) to the Poisson random measure with the intensity $\lambda^i(dx)$;
(iv) For some a_n and b_n, the free point process M_n weakly converges on (x^i,∞) to the free Poisson random measure with the intensity $\lambda^i(dx)$.

The equivalences of (i) and (iii) follows from the results in (Resnick 1987), and the equivalence of (i) and (ii) was proved in (Ben Arous and Voiculescu 2006). Thus, we only need to prove the equivalence of (i) and (iv).

Let $\mu_n(A) = \mu(b_n A + a_n)$. Note that (i) is equivalent to the statement that $n\mu_n(A) \to \lambda^i(A)$ for all Borel sets $A \subset (x^i,\infty)$. Indeed, suppose that μ is in the domain of attaction of $G^i(x)$. If $F(x)$ denote the distribution function of the measure μ, then

$$F^n(b_n x + a_n) \to G^i(x),$$

For every $x \in (x^i,\infty)$, $G^i(x)$ is positive. After taking logarithms on both sides, we get

$$n \log F(b_n x + a_n) \to \log G^i(x),$$

which is equivalent to

$$n(1 - F(b_n x + a_n)) \to -\log G^i(x).$$

Consequently,

$$n\mu_n\left((x,\infty)\right) \to \lambda^i\left((x,\infty)\right),$$

from which we conclude that $n\mu_n(A) \to \lambda^i(A)$ for all Borel sets $A \subset (x^i, \infty)$. The reverse implication is also clear.

Therefore, the equivalence of (i) and (iv) will be established if we prove the following general result about free point processes.

Recall that a measure is called Radon if $\mu(K) < \infty$ for every compact K.

Theorem 188 *Let $X_{i,n}$ be a triangular array of free, self-adjoint random variables and let the spectral probability measure of $X_{i,n}$ be μ_n. Let λ be a Radon measure on $D \subseteq \mathbb{R}$. The free point process M_n associated with the array $X_{i,n}$ converges weakly on D to a free Poisson measure M with the intensity measure λ if and only if*

$$n\mu_n(A) \to \lambda(A) \tag{80}$$

for every Borel set $A \subseteq D$.

We will prove this result in Section 18.

Theorem 187 tells us that the convergence of a free point process is related to the convergence of the distribution of free maxima. In addition, free processes can help us to define higher order free extreme processes. Indeed, in the classical case one way to calculate the probability distribution of the k-th order statistic is to calculate the probability that exactly $k-1$ variables X_i exceed a threshold t. This probability can be defined in terms of the corresponding point process.

In order to see that free point processes are necessary in the free case, note that the straightforward approach to the definition of higher order statistics does not work. While we can still define a set of space directions, which are dilated by exactly $k-1$ operators by the amount that exceeds t, this set is not a linear subspace except when $k=1$, and we cannot apply to this set the non-commutative analogue of probability – the dimension function. To break through this difficulty, we have to use free point processes.

Precisely, let $X_1, ..., X_n$ be freely independent self-adjoint variables and let X_i have the distribution F_i. Define projections $P_i(t) = 1_{(t,\infty)}(X_i)$ and consider the variable

$$Y(t) = \sum_{i=1}^{n} P_i(t).$$

Definition 189 *For every real* $k \geq 0$, *we say that* $F^{(n)}(t|k) =: E\left[1_{[0,k]}(Y(t))\right]$ *is the* distribution function of k-th order statistic *of the sequence* $X_1 ... X_n$, *and that it is the* k-th order free extremal convolution *of distributions* F_i.

To understand better the meaning of this definition, note that $Y(t) = \sum_{i=1}^{n} 1_{(t,\infty)}(X_i) = \left\langle M_n, 1_{(t,\infty)}\right\rangle$, where M_n is the free point process associated with the sequence X_i. Therefore the distribution $F^{(n)}(t|k)$ equals the expectation of $1_{[0,k]}\left(\left\langle M_n, 1_{(t,\infty)}\right\rangle\right)$. In the classical case we have an analogous expression $1_{[0,k]}\left(\left\langle N_n, 1_{(t,\infty)}\right\rangle\right)$, where $N_n = \sum_{i=1}^{n} \delta_{X_i}$ is the (classical) point process associated with the sequence of (classical) random variables X_i. In this case, the indicator function $1_{[0,k]}\left(\left\langle N_n, 1_{(t,\infty)}\right\rangle\right)$ corresponds to the event that the number of elements of $X_1, ..., X_n$ located in the interval (t, ∞) does not exceed k.

If $X^{(0)}$ denote the largest element of (classical) random variables, $X_1, ..., X_n$, $X^{(1)}$ denote the second largest one, and so on, then a realization of $X_1, ..., X_n$ will be counted by $1_{[0,k]}\left(\left\langle N_n, 1_{(t,\infty)}\right\rangle\right)$ if and only if $X^{(k)} \leq t$. It follows that $E1_{[0,k]}\left(\left\langle N_n, 1_{(t,\infty)}\right\rangle\right) = \Pr\left\{X^{(k)} \leq t\right\}$, i.e., this expectation gives the distribution of the $(k+1)$-st largest element of the sequence $X_1, ..., X_n$.

Thus, the expression $E1_{[0,k]}\left(\left\langle M_n, 1_{(t,\infty)}\right\rangle\right)$ can be interpreted as the non-commutative analogue of the distribution of the $(k+1)$-st order statistic. Note that the definition is valid not only for all integer k but also for all non-negative real k.

One question that immediately arises is whether we can define an operator, for which the distribution $F^{(n)}(t|k)$ would be a spectral distribution function? The answer to this question is positive. The condition $t' \geq t$ implies that $1_{[0,k]}(Y(t')) \geq 1_{[0,k]}(Y(t))$. Therefore, as t grows, the operators $1_{[0,k]}(Y(t))$ form an increasing family of projections and we can use this family to construct the required operator by the spectral resolution theorem.

Definition 190 *For every real* $k \geq 0$, *let*

$$Z^{(k)} = \int t \, d1_{[0,k]}(Y(t)).$$

We call $Z^{(k)}$ *the* k order statistic *of the family* X_i.

From the construction it is clear that $F^{(n)}(t|k)$ is the spectral distribution function of the operator $Z^{(k)}$.

In complete analogy with the classical case the limits of these free extremal convolutions can be computed using the limits of free point measures. If $G(x)$ is one of the classical limit laws, then we use $G^{(-1)}(x)$ to denote the functional inverse of $G(x)$. Let

$$\begin{aligned}
t_{-}(k) &= G^{(-1)}\left(\exp\left[-\left(1+\sqrt{k}\right)^{2}\right]\right), \\
t_{0}(k) &= G^{(-1)}\left(\frac{1}{e}\right), \\
t_{+}(k) &= G^{(-1)}\left(\exp\left[-\left(1-\sqrt{k}\right)^{2}\right]\right).
\end{aligned}$$

Let $\lambda(t) = -\log G(t)$ and $p_t(\xi) = (2\pi\xi)^{-1}\sqrt{4\xi - (1-\lambda(t)+\xi)^2}$.

Theorem 191 *Suppose that measure μ belongs to the domain of attraction of a (classical) limit law $G(x)$ and a_n, b_n are the corresponding norming constants. Assume that X_i are free self-adjoint variables with the spectral probability measure μ and let $F^{(n)}(t|k)$ denote the distribution of the k order statistic of the family $(X_i - a_n)/b_n$, where $i = 1, \ldots, n$. Then, as $n \to \infty$, the distribution $F^{(n)}(t|k)$ converges to a limit, $\overline{F}(t|k)$, which is given by the following formula:*

$$\overline{F}(t|k) = \begin{cases} 0, & \text{if } t < t_{-}, \\ \int_{\left(1-\sqrt{\lambda(t)}\right)^2}^{k} p_t(\xi)\, d\xi, & \text{if } t \in [t_{-}, t_0], \\ 1 - \lambda_t + \int_{\left(1-\sqrt{\lambda(t)}\right)^2}^{k} p_t(\xi)\, d\xi, & \text{if } (t_0, t_{+}], \\ 1 - \lambda(t)\, 1_{[0,1)}(k), & \text{if } t > t_{+}. \end{cases}$$

It turns out that in the particular case of the 0-order free extremal convolutions, their limits coincide with the limits discovered in (Ben Arous and Voiculescu 2006) (see Definition 6.8 and Theorems 6.9 and 6.11):

$$\begin{aligned}
\overline{F}^{(I)}(t|0) &= \left(1-e^{-x}\right) 1_{(0,\infty)}(x); \\
\overline{F}^{(II)}(t|0) &= \left(1-\frac{1}{x^{\alpha}}\right) 1_{(1,\infty)}(x); \text{ and} \\
\overline{F}^{(III)}(t|0) &= \left(1-|x|^{\alpha}\right) 1_{(-1,0)}(x) + 1_{[0,\infty)}(x),
\end{aligned}$$

where α is a positive parameter.

18 Convergence of Free Point Processes

18.1 Free Poisson random measure

Recall some facts about the *free Poisson distribution* with parameter ("intensity") λ. The continuous part of this distribution is supported on the interval $\left[\left(1-\sqrt{\lambda}\right)^2, \left(1+\sqrt{\lambda}\right)^2\right]$ and the density is

$$p_\lambda(x) = \frac{\sqrt{4x-(1-\lambda+x)^2}}{2\pi x}.$$

In addition, if $\lambda < 1$, then there is also an atom at zero with the probability weight $1-\lambda$. We call such an operator a (non-commutative) *Poisson random variable* with intensity λ and size 1.

The sum of two freely independent Poisson random variables of intensities λ_1 and λ_2 is again a Poisson random variable of intensity $\lambda_1 + \lambda_2$.

If we scale a non-commutative Poisson random variable by a, then we get a variable, which we call a *scaled (non-commutative) Poisson random variable* of intensity λ and size a.

Non-commutative Poisson random variables arise when we convolve a large number, N, of Bernoulli distributions that put probability λ/N on 1 and probability $1-\lambda/N$ on 0. The following result is well-known, see (Voiculescu 1998), or (Hiai and Petz 2000).

Proposition 192 *Suppose μ_n, $(n = 1, 2, ...)$ is a sequence of Bernoulli distributions, such that $\mu_n(\{1\}) \sim \lambda/n$ and $\mu_n(\{0\}) = 1 - \mu_n(\{1\})$. Define ν_n as follows:*

$$\nu_n = \underbrace{\mu_n \boxplus ... \boxplus \mu_n}_{n \text{ times}}.$$

Then ν_n weakly converges to the free Poisson distribution with intensity λ and size 1.

The following definition is basic for our investigation.

Definition 193 *Let $(\Theta, \mathcal{B}, \nu)$ be a measure space, and put*

$$\mathcal{B}_0 = \{B \in \mathcal{B} : \nu(B) < \infty\}.$$

Let further $(\mathcal{A}, E)$ be a W^*-probability space, and let $\mathcal{A}_+$ denote the cone of positive operators in $\mathcal{A}$. Then a free Poisson random measure *(*fPrm*) on $(\Theta, \mathcal{B}, \nu)$ with values*

in $(\mathcal{A}, E)$ is a mapping $M : \mathcal{B}_0 \to \mathcal{A}_+$, with the following properties:
(i) For any set B in $\mathcal{B}_0$, $M(B)$ is a free Poisson variable with parameter $\nu(B)$.
(ii) If $r \in \mathbb{N}$, and $B_1, ..., B_r \in \mathcal{B}_0$ are disjoint, then $M(B_1), ..., M(B_r)$ are free.
(iii) If $r \in \mathbb{N}$, and $B_1, ..., B_r \in \mathcal{B}_0$ are disjoint, then $M\left(\cup_{j=1}^r B_j\right) = \sum_{j=1}^r M(B_j)$.

The existence of a free Poisson measure for arbitrary spaces $(\Theta, \mathcal{B}, \nu)$ and $(\mathcal{A}, E)$ was shown in (Voiculescu 1998) and a different proof was given in (Barndorff-Nielsen and Thorbjornsen 2005).

If f is a real-valued simple function in $L^1(\Theta, \mathcal{B}, \nu)$, i.e,. if it can be written as

$$f = \sum_{i=1}^{r} a_i 1_{B_i},$$

for a system of disjoint $B_i \in \mathcal{B}_0$, then we define the integral with respect to a Poisson random measure M as follows:

$$\int_\Theta f\, dM = \sum_{i=1}^{r} a_i M(B_j).$$

It is possible to check that this definition is consistent. Moreover, as it is shown in (Barndorff-Nielsen and Thorbjornsen 2005), this concept can be extended to a larger class of functions:

Proposition 194 *Let f be a real-valued function in $L^1(\Theta, \mathcal{B}, \nu)$ and suppose that s_n is a sequence of real valued simple $\mathcal{B}$-measurable functions, satisfying the condition that there exists a positive ν-integrable function $h(\theta)$, such that $|s_n(\theta)| \leq h(\theta)$ for all n and θ. Suppose also that $\lim_{n\to\infty} s_n(\theta) = f(\theta)$ for all θ Then integrals $\int_\Theta s_n\, dM$ are well-defined and converge in probability to a self-adjoint (possibly unbounded) operator $I(f)$ affiliated with $\mathcal{A}$. Furthermore, the limit $I(f)$ is independent of the choice of approximating sequence s_n of simple functions.*

The resulting functional $I(f)$ is defined for all real valued functions f in $L^1(\Theta, \mathcal{B}, \nu)$ and is called the *integral with respect to the free Poisson random measure M*. It possesses all the usual properties of the integral: additivity, linear scaling, continuity, etc.

18.2 Free point process and weak convergence

Let $\overline{\mathcal{A}}$ be the set of densely-defined closed operators affiliated with a von Neumann algebra $\mathcal{A}$, and let $\mathcal{B}_\infty(\mathbb{R})$ denote the set of all bounded, Borel measurable functions $f : \mathbb{R} \to \mathbb{R}$.

Definition 195 *Let $X_{i,n} \in \overline{\mathcal{A}}$, ($i = 1, \ldots, n$; $n = 1, \ldots$) be a triangular array of freely independent self-adjoint variables Then the free point process M_n associated with the array $X_{i,n}$ is a sequence of A-valued functionals on $B_\infty(\mathbb{R})$, which are defined by the following formula:*

$$\langle M_n, f \rangle =: \sum_{i=1}^{n} f(X_{i,n}).$$

Note that we use terminology "point process" to emphasize the analogy with the classical case. In the classical case, an analogous functional N_n can be realized as a random real-valued measure $N_n(dx)$ which is concentrated on a (random) finite set of points in $\mathbb{R}$. In particular, $\langle N_n, f \rangle$ is a random variable. In the free case M_n is not random in the classical case but is completely determined by the array $X_{i,n}$. However, for each f, the bracket $\langle M_n, f \rangle$ is an operator in $\mathcal{A}$ and thus is a free random variable.

Next, we define the mode of convergence of free point measures that in the classical case corresponds to the weak convergence of point processes.

Let D be a Borel subset of $\mathbb{R}$ and let $\mathcal{F}_K^\infty(D)$ denote the space of bounded, Borel measurable functions that have compact support on D.

Definition 196 *We say that a free point process M_n* converges weakly on D *to a free Poisson random measure M, which is defined on $(D, \mathcal{B}, \lambda)$ and takes values in $\mathcal{A}$, if for every function $f \in \mathcal{F}_K^\infty(D)$ the following convergence holds:*

$$\langle M_n, f \rangle \xrightarrow{d} \int_{\mathbb{R}} f \, dM.$$

Sometimes we need to speak about convergence with respect to a class of functions, which is different from $\mathcal{F}_K^\infty(D)$.

Definition 197 *We say that a free point process M_n* converges weakly with respect to a class of functions $\mathcal{F}$ *to a free Poisson random measure M, if for every function $f \in \mathcal{F}$ the following convergence holds:*

$$\langle M_n, f \rangle \xrightarrow{d} \int_{\mathbb{R}} f \, dM.$$

Theorem 198 *Let $X_{i,n}$ be a triangular array of free, self-adjoint random variables and let the spectral probability measure of $X_{i,n}$ be μ_n. Let λ be a Radon measure on $D \subseteq \mathbb{R}$. The free point process M_n associated with the array $X_{i,n}$ converges weakly on D to a free Poisson measure M with the intensity measure λ if and only if*

$$n\mu_n(A) \to \lambda(A) \tag{81}$$

for every Borel set $A \subseteq D$.

We will prove Theorem 188 by considering initially the convergence of free point processes M_n with respect to the class of simple functions (i.e., finite sums of indicator functions), and then approximating functions from a more general class by simple functions.

18.3 Convergence with respect to simple functions

Let $\mathcal{S}(D)$ be the class of simple functions on $D \subset \mathbb{R}$, i.e., the class of finite sums of indicator functions of Borel sets belonging to D.

Proposition 199 *Let $X_{i,n}$ be a triangular array of free, self-adjoint random variables and let the spectral probability measure of $X_{i,n}$ be μ_n. Let λ be a Radon measure on $D \subseteq \mathbb{R}$. If*

$$n\mu_n(A) \to \lambda(A)$$

for each Borel set $A \subset D$, then the free point process M_n associated with the array $X_{i,n}$ converges weakly with respect to $\mathcal{S}(D)$ to a free Poisson random measure M with the intensity measure λ.

Before proving this theorem, we derive some auxiliary results.

Lemma 200 *Suppose $X_{i,n}$ is an array of free and identically distributed random variables with the spectral measure μ_n. Let $n\mu_n(A) \to \lambda(A) < \infty$ as $n \to \infty$. Let $Z_{i,n} = 1_A(X_{i,n})$. Then as $n \to \infty$, the sum $S_n = \sum_{i=1}^n Z_{i,n}$ converges in distribution to a free Poisson random variable with intensity $\lambda(A)$.*

Proof: Note that $Z_{i,n}$ are projections with expectation $\mu_n(A)$ and they are free. Therefore, $\sum_{i=1}^n Z_{i,n}$ is the sum of free projections and we can use Proposition 192 to infer the claim of the lemma. QED.

As the next step to the proof of Proposition 199 we need to check that if Borel sets A_k are disjoint, then the sums $S_k = \sum_{i=1}^{n} 1_{A_k}(X_{i,n})$ are asymptotically free with respect to growing n.

Recall the definition of the asymptotic freeness: Let $(\mathcal{A}_i, E_i)$ be a sequence of non-commutative probability spaces and let X_i and Y_i be two random variables in $\mathcal{A}_i$. Let also x and y be two free operators in a non-commutative probability space $(\mathcal{A}, E)$.

Definition 201 *The sequences X_i and Y_i are called* asymptotically free *if the sequence of pairs (X_i, Y_i) converges in distribution to the pair (x, y). That is, for every $\varepsilon > 0$ and every sequence of k-tuples $(n_1, ..., n_k)$ with non-negative integer n_j, there exists such i_0 that for $i \geq i_0$, the following inequality holds:*

$$\left|E_i\left(X_i^{n_1} Y_i^{n_2} ... X_i^{n_{k-1}} Y_i^{n_k}\right) - E\left(x^{n_1} y^{n_2} ... x^{n_{k-1}} y^{n_k}\right)\right| \leq \varepsilon.$$

At the cost of more complicated notation, this definiton can be generalized to the case of more than two variables.

Now, we can formulate a generalization of Proposition 192 that says that certain sums of projections not only converge to Poisson random variables, but are also asymptotically free.

Proposition 202 *Let $P_{i,n}$ be free projections of dimension λ/n and $Q_{i,n}$ be free projections of dimension μ/n Assume $P_{i,n}$ and $Q_{j,n}$ are free if $i \neq j$, and orthogonal to each other if $i = j$. Let $S_n = \sum_{i=1}^{n} P_{i,n}$ and $\Sigma_n = \sum_{i=1}^{n} Q_{i,n}$. Then the sequences S_n and Σ_n converge in distribution to free variables S and Σ, that have free Poisson distributions with parameters λ and μ, respectively. In particular, the sequences S_n and Σ_n are asymptotically free.*

Proof: The fact that each of the sequences S_n and Σ_n converge to a free Poisson distribution is clear from Proposition 192. The essential part is to prove that asymptotic freeness holds. Let us introduce variables $\widetilde{Q}_{i,n}$ which are (i) free among themselves, (ii) distributed according to the same distribution as $Q_{i,n}$, and (iii) are free from all of $P_{i,n}$. Define $\widetilde{\Sigma}_n = \sum_{i=1}^{n} \widetilde{Q}_{i,n}$. Clearly, S_n and $\widetilde{\Sigma}_n$ are freely independent and $\widetilde{\Sigma}_n$ has the same distribution as Σ_n. Consequently, the sequence of $\left(S_n, \widetilde{\Sigma}_n\right)$ converges in distribution to (S, Σ), where S and Σ are two freely independent random variables with free Poisson distributions.

What remains to prove is that if an integer $r > 0$ is fixed and $P_r(S_n, \Sigma_n)$ is an arbitrary non-commutative polynomial of degree r, then

$$E\left[P_r(S_n, \Sigma_n) - P_r\left(S_n, \widetilde{\Sigma}_n\right)\right] \to 0$$

as $n \to \infty$.

From this moment on we will omit the subscript n from variables $P_{i,n}$, $Q_{i,n}$, and $\widetilde{Q}_{i,n}$ to make the notation more transparent.

Let us expand $P_r(S_n, \Sigma_n)$ as a sum of products of variables P_i and Q_i. A similar expansion in products of P_i and $\widetilde{Q}_i$ holds for $P_r\left(S_n, \widetilde{\Sigma}_n\right)$. The expectations of the corresponding product terms in these expansions differ if and only if for some i, the product includes both P_i and Q_i (and P_i and $\widetilde{Q}_i$ in the corresponding term of the other expansion). Otherwise, these two products are the same from the distributional point of view and therefore must have the same expectations.

Example: Consider $E(S_n\Sigma_n) - E\left(S_n\widetilde{\Sigma}_n\right)$. Then we note that $E(P_iQ_j) = E\left(P_i\widetilde{Q}_j\right)$ if $i \neq j$, and therefore:

$$\begin{aligned} E(S_n\Sigma_n) - E\left(S_n\widetilde{\Sigma}_n\right) &= \sum_{i=1}^{n}\left[E(P_iQ_i) - E\left(P_i\widetilde{Q}_i\right)\right] \\ &= -\frac{\lambda\mu}{n}. \end{aligned}$$

End of Example.

Now consider one of the products that do have different expectations. Consider first the product in the expansion of $P_r(S_n, \Sigma_n)$. Let I is a set of all indices that are used in this product (without regard to whether it is the index of a P or a Q). For example, if the product is $P_1\widetilde{Q}_3P_7\widetilde{Q}_3P_3$, then the set I is $\{1, 3, 7\}$.

Let $I = \{i_1, ..., i_s\}$. Using the freeness of elements P_i and $\widetilde{Q}_i$ and the fact that they are projections and therefore $P_i^m = P_i$ and $\widetilde{Q}_i^m = \widetilde{Q}_i$ for every $m \geq 1$, we can compute the expectation of such a product as

$$\sum c(\varepsilon_1, ..., \varepsilon_s, \nu_1, ..., \nu_s)\, E(P_{i_1})^{\varepsilon_1} ... E(P_{i_s})^{\varepsilon_s}\, E\left(\widetilde{Q}_{i_1}\right)^{\nu_1} ... E\left(\widetilde{Q}_{i_s}\right)^{\nu_s}, \qquad (82)$$

where the sum is over $\varepsilon_i \in \{0, 1, ..., r\}$ and $\nu_i \in \{0, 1, ..., r\}$. Note that the following three conditions must hold for the terms in this sum: (i) For each i, either $\varepsilon_i \geq 1$, or $\nu_i \geq 1$, or both ≥ 1. (I.e., either $E(P_i)$ or $E\left(\widetilde{Q}_i\right)$ is present in the product, they cannot be both absent by the definition of the set I.) (ii) For at least one i, both ε_i and ν_i are ≥ 1. (This condition must hold because of our assumption that we consider only products with different expectations, so that for at least one i, both P_i and Q_i must be present in the product.) (iii) The coefficient $c(\varepsilon_1, ..., \varepsilon_s, \nu_1, ..., \nu_s)$ can depend on the number of elements P_i's and $\widetilde{Q}_i$'s and on their arrangement in the product, but it does not depend on n. In particular it can be bounded by a function of r.

(For example, $E\left(P_1\widetilde{Q}_3P_7\widetilde{Q}_3P_3\right) = E(P_1)\,E(P_3)\,E(P_7)\,E\left(\widetilde{Q}_3\right)$.)

Since $E(P_i) = \lambda/n$ and $E(Q_i) = \mu/n$, therefore we can estimate the expectation of this product as follows:

$$|E(\text{product})| \leq c'(r)\frac{1}{n^{\varepsilon_1+...+\varepsilon_s+\nu_1+...+\nu_s}} \leq c'(r)\frac{1}{n^{s+1}}.$$

If the degree of polynomial is r and the number elements in I is s, then the number of possible product terms with an index set I that consists of s elements is bounded by n^s (the upper bound on the number of possible choices of the set I that consists of s elements) multiplied by a certain function $f(r)$ which counts the possible arrangements of P_i and $\widetilde{Q}_i$, if the set I with s elements is fixed. Therefore we estimate:

$$|E(\text{sum of products with set } I \text{ that consists of } s \text{ elements})| \leq c(r)\frac{1}{n}.$$

Finally note that $s \leq r$, and therefore we can estimate the expectation of the sum of those products of P_i and $\widetilde{Q}_i$ that have expectations different from the corresponding products of P_i and Q_i as $rc(r)\,n^{-1}$.

Essentially the same argument can be used to estimate the expectation of the corresponding products of P_i and Q_i. Here, to derive formula (82) (with a possibly different coefficient c) we can use the freeness of pairs $\{P_i, Q_i\}$, and the fact that P_i and Q_i are projections, orthogonal to each other.

For example, to calculate $E(P_1Q_3Q_1P_3)$ we can first use freeness to write:

$$\begin{aligned} E(P_1Q_3Q_1P_3) &= E(P_1)\,E(Q_1)\,E(P_3Q_3) + E(P_3)\,E(Q_3)\,E(P_1Q_1) \\ &\quad -E(P_1)\,E(Q_1)\,E(P_3)\,E(Q_3). \end{aligned}$$

And then we can use orthogonality to finish this calculation as follows:

$$E(P_1Q_3Q_1P_3) = -E(P_1)\,E(Q_1)\,E(P_3)\,E(Q_3)$$

Given formula (82), the argument goes exactly as in the case with P_i and $\widetilde{Q}_i$, and allows us to conclude that the expectation of the sum of the relevant products of P_i and Q_i is bounded by $rc''(r)\,n^{-1}$.

Therefore, if an integer r is fixed, then, as n grows, $E\left[P_r\left(S_n,\Sigma_n\right)-P_r\left(S_n,\widetilde{\Sigma}_n\right)\right]\rightarrow 0$. Therefore $\left(S_n,\Sigma_n\right)$ converges in distribution to $\left(S,\Sigma\right)$, where S and Σ are free. This shows that S_n and Σ_n are asymptotically free. QED.

It is possible to extend this result to more than two families of projectors. This generalized result is as follows:

Proposition 203 *Let $P_{i,n}^{(k)}$, (where $n=1,2,...;$ $i=1,...,n,$ and $k=1,...,r$) be projections of dimension $\lambda^{(k)}/n$. Assume that for each n, algebras $\mathcal{A}_i$ generated by sets $\left\{P_{i,n}^{(k)}\right\}_{k=1}^{r}$ are free. Also assume that for each n and i, the projections $P_{i,n}^{(k)}$ are orthogonal to each other, i.e., $P_{i,n}^{(k)}P_{i,n}^{(k')}=0$ for every pair $k\neq k'$. Let $S_n^{(k)}=\sum_{i=1}^{n}P_{i,n}^{(k)}$. Then the sequences $S_n^{(k)}$ converge in distribution to freely independent variables $S^{(k)}$ that have free Poisson distributions with parameters $\lambda^{(k)}$, respectively. In particular, the sequences $S_n^{(k)}$ are asymptotically free.*

Now we can proceed to the proof of Proposition 199.

Proof: Let $f=\sum_{k=1}^{r}c_k 1_{A_k}\left(x\right)$, where A_k are disjoint Borel sets. Using the assumption that $n\mu_n\left(A_k\right)\rightarrow\lambda\left(A_k\right)$ and Lemma 200, we can find a free Poisson random measure M such that

$$\sum_{i=1}^{n}1_{A_k}\left(X_{i,n}\right)\xrightarrow{d}M\left(A_k\right)=\int_{\mathbb{R}}1_{A_k}\left(x\right)M\left(dx\right)$$

as $n\rightarrow\infty$. Indeed, it is enough to take a Poisson random measure M with the intensity measure λ.

In addition, by Proposition 203, sums $S_k=\sum_{i=1}^{n}1_{A_k}\left(X_{i,n}\right)$ become asymptotically free for different k as n grows. Since $M\left(A_k\right)$ are free by the definition of the free Poisson measure, this implies that

$$\sum_{k=1}^{r}c_k\sum_{i=1}^{n}1_{A_k}\left(X_{i,n}\right)\xrightarrow{d}\sum_{k=1}^{r}c_kM\left(A_k\right)=\sum_{k=1}^{r}c_k\int_{\mathbb{R}}1_{A_k}\left(x\right)M\left(dx\right).$$

as $n\rightarrow\infty$. Therefore,

$$\sum_{i=1}^{n}f\left(X_{i,n}\right)\xrightarrow{d}\int_{\mathbb{R}}f\left(x\right)M\left(dx\right),$$

where we used the additivity property of the integral with respect to a free Poisson random measure (see (Barndorff-Nielsen and Thorbjornsen 2005), Remark 4.2(b)). QED.

18.4 Convergence with respect to bounded, Borel measurable functions with compact support

The goal of this section is to prove our main Theorem 188.

Consider a bounded, Borel measurable, compactly supported function $f : D \to \mathbb{R}$, such that $0 < f \leq 1$. (A more general case of a function f, which satisfies $C_1 < f \leq C_2$, can be treated similarly.) For positive integers $N = 1, 2, \ldots$, and $k = 1, \ldots, N$, define the set

$$A_k^{(N)} = \left\{ x \in D : \frac{k-1}{N} < f(x) \leq \frac{k}{N} \right\}.$$

The sets $A_k^{(N)}$ are disjoint, measurable, and have finite λ-measure. Their union is D.

We define lower and upper approximations to the function f as follows:

$$l^N(x) = \sum_{k=1}^{N} \frac{k-1}{N} 1_{A_k^{(N)}}(x),$$

and

$$u^N(x) = \sum_{k=1}^{N} \frac{k}{N} 1_{A_k^{(N)}}(x),$$

We note that:

(i) $l^N(x) \leq u^N(x)$;

(ii) $l^N(x)$ is an increasing sequence of functions;

(iii) $u^N(x)$ is a decreasing sequence of functions, and

iv) $\lim_{N \to \infty} l^N(x) - u^N(x) = 0$ uniformly in x.

The functions $l^N(x)$ and $u^N(x)$ are simple: $l^N(x) = \sum_{i=1}^{N} c_k^{(N)} 1_{A_k^{(N)}}(x)$ and $u^N(x) = \sum_{i=1}^{N} d_k^{(N)} 1_{A_k^{(N)}}(x)$. Note also that $\sup_k \left(d_k^{(N)} - c_k^{(N)} \right) = 1/N$ converges to zero as $N \to \infty$.

Let us drop for convenience the superscript N when we consider it as fixed, and simply write $l(x) = \sum_{i=1}^{N} c_k 1_{A_k}(x)$ and $u(x) = \sum_{i=1}^{N} d_k 1_{A_k}(x)$, where A_k are disjoint Borel-measurable sets. By Proposition 199, as $n \to \infty$,

$$\sum_{i=1}^{n} l(X_{i,n}) \xrightarrow{d} \sum_{k=1}^{N} c_k M_k,$$

where M_k are freely independent Poisson random variables with intensities $\lambda_k = \lambda(A_k)$. Let $F_l(x)$ denote the distribution function of $\sum_{k=1}^{N} c_k M_k$.

Similarly,

$$\sum_{i=1}^{n} u\left(X_{i,n}\right) \xrightarrow{d} \sum_{k=1}^{N} d_k M_k,$$

and we denote the distribution function of $\sum_{k=1}^{N} d_k M_k$ as $F_u(x)$.

Let $F_{f,n}$ denote the distribution function of $\sum_{i=1}^{n} f\left(X_{i,n}\right)$ and let F_f be one of the limit points of this sequence of distribution functions.

Proposition 204 *F_f is a distribution function and $F_u(x) \leq F_f(x) \leq F_l(x)$ for every x.*

Proof: We will infer this from Lemma 205 below and its Corollary. This lemma is a particular case of Weyl's eigenvalue inequalities for operators in a von Neumann algebra of type II_1. If $F_A(x)$ is the spectral distribution function of a self-adjoint operator A, then we define the eigenvalue function $\lambda_A(t) = \inf\{x : F_A(x) \geq 1 - t\}$. The function $\lambda_A(t)$ is non-increasing and right-continuous.

Let us use notation $\lambda_A(t-0)$ to denote $\lim_{\varepsilon \downarrow 0} \lambda_A(t-\varepsilon)$. Then the following Lemma holds:

Lemma 205 *If A and B are two bounded self-adjoint operators from a W^*-probability space $\mathcal{A}$ and if B is non-negative definite, then*

$$\begin{aligned} \lambda_A(t) &\leq \lambda_{A+B}(t) \leq \lambda_A(t) + \|B\|, \text{ and} \\ \lambda_A(t-0) &\leq \lambda_{A+B}(t-0) \leq \lambda_A(t-0) + \|B\|. \end{aligned}$$

Corollary 206 *If $B \geq 0$, then $\mu_{A+B} \gg \mu_A$, that is, $F_{A+B}(x) \leq F_A(x)$ for each x.*

Proof of Lemma 205: This results easily follows from an inequality in (Bercovici and Li 2001) which states that if $(a-\varepsilon, a) \subset [0,1]$, $(b-\varepsilon, b) \subset [0,1]$, and $a+b \leq 1$, then

$$\int_{a+b-\varepsilon}^{a+by} \lambda_{A+B}(t)\, dt \leq \int_{a-\varepsilon}^{a} \lambda_A(t)\, dt + \int_{b-\varepsilon}^{b} \lambda_B(t)\, dt. \tag{83}$$

QED.

By Corollary 206, for each n the distribution $F_{f,n}$ is between the distribution functions of $\sum_{i=1}^{n} u\left(X_{i,n}\right)$ and $\sum_{i=1}^{n} l\left(X_{i,n}\right)$. As n grows, these two sequences of distribution functions approach $F_u(x)$ and $F_l(x)$, respectively. Therefore, every limit point of $F_{f,n}$ is between F_u and F_l. The claim that F_f is a distribution function follows from the fact that both F_u and F_l are distribution functions. QED.

Now we want to show that $F_u^{(N)}(x)$ approaches $F_l^{(N)}(x)$ as N grows.

Recall that the *Levy distance* between two distribution functions is defined as follows:

$$d_L(F_A, F_B) = \sup_x \inf \{ s \geq 0 : F_B(x - s) - s \leq F_A(x) \leq F_B(x + s) + s \} .$$

We can interpret this distance geometrically. Let Γ_A be the graph of function F_A, and at the points of discontinuity let us connect the left and right limits by a (vertical) straight line interval. Call the resulting curve $\widetilde{\Gamma}_A$. Similarly define $\widetilde{\Gamma}_B$. Let d be the maximum distance between $\widetilde{\Gamma}_A$ and $\widetilde{\Gamma}_B$ in the direction from the south-east to the north-west, i.e., in the direction which is obtained by rotating the vertical direction by $\pi/4$ counter-clockwise. Then $d_L(F_A, F_B) = d/\sqrt{2}$.

Proposition 207 *Let K be the sum of intensities of freely independent Poisson random variables M_k and let $F_l(x)$ and $F_u(x)$ be distribution functions of $\sum_{k=1}^N c_k M_k$ and $\sum_{k=1}^N d_k M_k$ Then*

$$d_L(F_l, F_u) \leq \left(2K + 3\sqrt{K} + 1\right) \sup_{1 \leq k \leq N} (d_k - c_k) .$$

Remark: In our case, the finiteness of K will be ensured by the assumptions that λ is Radon and that f has a compact support.

For the proof of this proposition we need two lemmas. Lemma 208 provides a bound on the norm of the sum of scaled Poisson random variables in terms of the sizes of these variables, and Lemma 209 relates the Levy distance between two random variables to the norm of their difference.

Lemma 208 *Let M_i, ($i = 1, ..., r$) be freely independent Poisson random variables, which have intensities λ_i, and let b_i be non-negative real numbers. Assume that $\sum_{i=1}^r \lambda_i \leq K$ and let $b = \sup_{1 \leq i \leq r} b_i$. Then*

$$\left\| \sum_{i=1}^r b_i M_i \right\| \leq b \left(2K + 3\sqrt{K} + 1\right) .$$

Proof: Let X_i be free self-adjoint random variables that have zero mean. Then by an inequality from (Voiculescu 1986):

$$\left\| \sum_{i=1}^r X_i \right\| \leq \max_{1 \leq i \leq r} \|X_i\| + \sqrt{\sum_{i=1}^r \mathrm{Var}(X_i)}.$$

If Y_i are free self-adjoint random variables with non-zero mean, and $X_i = Y_i - E(Y_i)$, then the previous inequality implies that

$$\begin{aligned}\left\|\sum_{i=1}^{r} Y_i\right\| &\leq \left|\sum_{i=1}^{r} E(Y_i)\right| + \left\|\sum_{i=1}^{r} X_i\right\| + \sqrt{\sum_{i=1}^{r} \mathrm{Var}(X_i)} \\ &\leq \left|\sum_{i=1}^{r} E(Y_i)\right| + \max_{1\leq i\leq r} d(Y_i) + \sqrt{\sum_{i=1}^{r} \mathrm{Var}(Y_i)}, \qquad (84)\end{aligned}$$

where $d(Y_i)$ is the diameter of the support of Y_i.

We will apply this inequality to $Y_i = b_i M_i$ and estimate each of the three terms on the right-hand side of (84) in turn:

1) Since $E(M_i) = \lambda_i$, and $\sum \lambda_i \leq K$, therefore $\sum_{i=1}^{r} b_i E(M_i) \leq bK$.

2) The diameter of the support of $b_i M_i$ is less or equal to $b_i\left(1+\sqrt{\lambda_i}\right)^2 \leq b\left(1+2\sqrt{K}+K\right)$.

3) Since $\mathrm{Var}(M_i) = \lambda_i$, therefore $\sqrt{\sum_{i=1}^{r} \mathrm{Var}(b_i M_i)} \leq b\sqrt{K}$.

In sum, $\left\|\sum_{i=1}^{r} b_i M_i\right\| \leq b\left(2K + 3\sqrt{K}+1\right)$. QED.

Lemma 209 *Let A and B be two bounded self-adjoint operators from a W^*-probability space $\mathcal{A}$ and assume that $B - A \geq 0$. Then*

$$d_L(F_A, F_B) \leq \|B - A\|.$$

Proof: Let F_A and F_B be distribution functions, and λ_A and λ_B be the corresponding λ-functions. Then we claim that

$$d_L(F_A, F_B) \leq \sup_{0\leq t\leq 1} |\lambda_A(t) - \lambda_B(t)|. \qquad (85)$$

Indeed, let the graphs of functions λ_A and λ_B be denoted as Λ_A and Λ_B, respectively. Connecting the left and right limits at the points of discontinuity gives us the curves $\widetilde{\Lambda}_A$ and $\widetilde{\Lambda}_B$. It is easy to see that these curves can be obtained from curves $\widetilde{\Gamma}_A$ and $\widetilde{\Gamma}_B$ (i.e., the graphs of $F_A(x)$ and $F_B(x)$ with connected limits at the points of discontinuity) by rotating them around the point $(0,1)$ counter-clockwise by the angle $\pi/2$ and then shifting the result of the rotation by vector $(0,-1)$. It follows that the distance d, which was used in the definition of the Levy distance can also be defined as the maximum distance between $\widetilde{\Lambda}_A$ and $\widetilde{\Lambda}_B$ in the direction from the south-west to the north-east, i.e., in the direction which is obtained by rotating the vertical direction by $\pi/4$ clockwise.

Since $\lambda_A(t)$ and $\lambda_B(t)$ are non-increasing functions, therefore

$$d \leq \sqrt{2} \sup_{0\leq t\leq 1} |\lambda_A(t) - \lambda_B(t)|.$$

This implies $d_L(F_A, F_B) \leq \sup_{0\leq t\leq 1} |\lambda_A(t) - \lambda_B(t)|$.

Inequality (85) and Lemma 205 imply the statement of the lemma.

QED.

Now we can prove Proposition 207:

Proof of Proposition 207: Let $X = \sum_{k=1}^{N} (d_k - c_k) M_k$. By Lemma 208,

$$\|X\| \leq b\left(2K + 3\sqrt{K} + 1\right),$$

where $b = \sup_{1\leq k\leq N} (d_k - c_k)$ and K is the sum of the intensities of M_k. By Lemma 209, this implies that

$$d_L(F_l, F_u) \leq b\left(2K + 3\sqrt{K} + 1\right).$$

QED.

Using Proposition 207**,** we can proceed to the proof of Theorem 188. By Proposition 199, we know that if N is fixed and $n \to \infty$, then

$$\sum_{i=1}^{n} l^N(X_{i,n}) \xrightarrow{d} \sum_{k=1}^{N} c_i^{(N)} M\left(A_k^{(N)}\right),$$

and

$$\sum_{i=1}^{n} u^N(X_{i,n}) \xrightarrow{d} \sum_{k=1}^{N} d_i^{(N)} M\left(A_k^{(N)}\right),$$

where M is a free Poisson random measure with intensity $\lambda(dx)$. Let the distributions of the right-hand sides be denoted as F_{l^N} and F_{u^N}.

By Corollary 206, F_{l^N} is a decreasing sequence and F_{u^N} is an increasing sequence of distribution functions. In addition, $F_{l^N}(x) \geq F_{u^N}(x)$ for every N and x. Since the sum of intensities of variables $M\left(A_k^{(N)}\right)$ is less than $\lambda(D) < \infty$ by assumption, therefore Proposition 207 is applicable and we can conclude that the Levy distance between F_{l^N} and F_{u^N} converges to zero as $N \to \infty$. Consequently, these two distributions (weakly) converge to a limit distribution function as $N \to \infty$.

Moreover, by the definition of the integral with respect to a free Poisson random measure, this limit equals the distribution function of $\int f(x) M(dx)$.

In addition, by Proposition 204 every limit point of the sequence of $F_{f,n}$ is between F_{l^N} and F_{u^N} for every N, and therefore the sequence of $F_{f,n}$ also converges to the distribution function of $\int f(x) M(dx)$. QED.

This completes the proof of Theorem 188.

19 Free Extremes

19.1 Definition of order statistics

Let us start with a heuristic motiviation of the definitions to follow. Let $X_1, ..., X_n$ be Hermitian positive-definite matrices and t be a real positive number. For each X_i we can define a linear subspace $V_i(t)$ spanned by eigenvectors of X_i with eigenvalues which are less or equal to t. Alternatively, we can think about $V_i(t)$ as a space of directions, which operators X_i^n dilate by less or equal to t^n for every integer $n \geq 1$. The intersection of these subpaces, $V^0(t)$, is the subspace of the directions, such that whatever direction $v \in V^0(t)$ and whatever $n \geq 1$ are given, none of operators X_i^n dilates the direction v by more than t^n times. The subspaces $V^0(t)$ are increasing with t and using the spectral resolution theorem we can associate an operator $\int t \, dP_{V^0(t)}(t)$ with this family of subspaces. This operator is natural to call the maximum of $X_1, \ldots, X_n$. The spectral distribution function of this operator evaluated at t equals the dimension of the subspace $V^0(t)$. This distribution function is the extremal (not necessarily free) convolution of distribution functions of $X_1, ..., X_n$ in the sense of (Ben Arous and Voiculescu 2006).

When we try to apply this reasoning and define the 2-nd order statistic instead of the maximum, we run into a difficulty. Indeed, in this case it is natural to look at vectors v for which there is exactly one of X_i that has the property that for some n, the operator X_i^n dilates v by more than t^n times. If $V^1(t)$ denotes this set of directions, then we can write this set algebraically as

$$V^1(t) = \bigcup_{i=1}^{n} \left[\left(\bigcap_{j \neq i} V_j(t) \right) \cap (H \backslash V_i(t)) \right],$$

where H is the space where X_i acts. A slightly different possibility would be to define this set as

$$V^1(t) = \bigcup_{i=1}^{n} \left[\left(\bigcap_{j \neq i} V_j(t) \right) \cap \left(V_i(t)^{\perp} \right) \right]$$

Unfortunately, whichever of these two definitions we use, $V^1(t)$ is not a linear subspace, so we can neither use the spectral resolution theorem, nor calculate its dimension.

One natural way out of this difficulty is to consider the linear span of the set $V^1(t)$. We prefer a similar but an analytically easier alternative. First, we define the sum of projections on subspaces $W_i(t) = V_i(t)^{\perp}$. Let us call this sum $Y(t)$. Note that if this sum of projections evaluated at a vector v equals zero, then x belongs to $\cap V_i(t) = V^0(t)$. Moreover, it is intuitively clear that if this sum is evaluated at v and the result of this evaluation, $\langle v, Y(t) v \rangle$, is small, then either v belongs to $V_i(t)$ for the majority of i, or v deviates outside of many $V_i(t)$ but only by a very small amount. In the first case, only a small number of operators X_i are such that for some n, the operator X_i^n dilates v by more than t^n. In the second case there can be many such X_i but then X_i^n dilate v by *not much* more than t^n.

This suggest introducing the set of directions on which the sum of projections on subspaces $W_i(t)$ is small. The great advantage of this new set is that it is a linear space and we can both measure its dimension and apply the spectral resolution theorem.

Now, after this intuitive introduction, we turn to a rigorous definition.

Let $X_1, ..., X_n$ be freely independent self-adjoint random variables and let X_i have the distribution F_i. Define projections $P_i(t) = 1_{(t,\infty)}(X_i)$ and consider the variable

$$Y_n(t) = \sum_{i=1}^{n} P_i(t).$$

Definition 2. *For every real $k \geq 0$, we say that $F^{(n)}(t|k) =: E\left[1_{[0,k]}(Y_n(t))\right]$ is the distribution function of k-th order statistic of the sequence $X_1 ... X_n$, and that it is the k-th order free extremal convolution of distributions F_i.*

We need to check that this is a consistent definition, and that $F^{(n)}(t|k)$ is indeed a distribution function for each $k \geq 0$.

For convenience we will omit index n in the following argument.

It is easy to see that $F(t|k)$ is non-decreasing in t. Indeed, let $t' \geq t$. Then for each i, $P_i(t') \leq P_i(t)$, and therefore, $Y(t') \leq Y(t)$. It follows that $1_{[0,k]}(Y(t')) \geq 1_{[0,k]}(Y(t))$, and therefore $F(t'|k) \geq F(t|k)$.

This function is also right-continuous in t. First, note that if $t_m \downarrow t$, then $P_i(t_m) \xrightarrow{d} P_i(t)$. Since operators P_i are freely independent for diffferent i, this implies that $Y(t_m) \xrightarrow{d} Y(t)$ as $t_m \downarrow t$. Indeed, the operators $Y(t_m)$ and $Y(t)$ are uniformly bounded ($\|Y(t_m)\| \leq n$ and $\|Y(t)\| \leq n$), and the moments of the distribution of $Y(t_m)$ converge to the corresponding moments of the distribution of $Y(t)$.

Let the distribution functions of $Y(t_m)$ and $Y(t)$ be denoted as $G_m(x)$ and $G(x)$, respectively. Then $E\left[1_{[0,k]}(Y(t_m))\right] = G_m(k)$ and $E\left[1_{[0,k]}(Y(t))\right] = G(k)$. The convergence $Y(t_m) \xrightarrow{d} Y(t)$ implies that $G_m(k) \to G(k)$ as $m \to \infty$, for all k at which $G(k)$ is continuous. We will prove that, moreover, even if $G(x)$ has a jump at $x = k$, then the sequence $G_m(k)$ still converges to $G(k)$. At this point of the argument, it is essential that t_m converges to t from above and therefore $G_m(k) \geq G(k)$.

Indeed, by seeking a contradiction, suppose that $G_m(k)$ does not converge to $G(k)$. Take ε such that $G_m(k) - G(k) > \varepsilon$ for all m, and take $k' > k$ such that (1) k' is a point of continuity of $G(x)$, and (2) $G(k') - G(k) < \varepsilon/2$. Such k' exists because $G(x)$ is right-continuous. Since $G_m(k)$ is increasing, we conclude that $G_m(k') - G(k') > \varepsilon/2$ for all m. But this means that $G_m(x)$ does not converge to $G(x)$ at a point of continuity of $G(x)$, namely, at k'. This is a contradiction, and we conclude that $G_m(k)$ converges to $G(k)$ for all k.

Finally, as $t \to \infty$, $P_i(t) \xrightarrow{d} 0$. Therefore $Y(t) \xrightarrow{d} 0$, and $1_{[0,k]}(Y(t)) \xrightarrow{d} I$. Hence $F(t|k) \to 1$ and we conclude that $F(t|k)$ is a valid distribution function.

Consider now the special case when $k = 0$. In this case $F^{(n)}(t|0)$ is the dimension of the nill-space of $Y(t)$, which equals to the dimension of the intersection of the nill-spaces of $P_i(t) =: 1_{(t,\infty)}(X_i)$. It is easy to see that this coincides with the definition of the free extremal convolution of the distributions $F_i(x)$, which was introduced in (Ben Arous and Voiculescu 2006).

19.2 Limits of the distributions of free extremes

Now let us investigate the question of the limiting behavior of the distributions $F^{(n)}(t|k)$ when $n \to \infty$. The limits are described in Theorem 191.

Proof of Theorem 191: For each n we re-define:

$$Y_n(t) = \sum_{i=1}^{n} 1_{(t,\infty)}\left(\frac{X_i - a_n}{b_n}\right) = \left\langle M_n, 1_{(t,\infty)}\right\rangle,$$

where M_n is the free point process associated with the triangular array $(X_i - a_n)/b_n$.

The bracket $\left\langle M_n, 1_{(t,\infty)}\right\rangle$ converges in distribution to a random variable C_t, which is a free Poisson random variable with the intensity $\lambda(t) = -\log G(t)$. Then, in order to calculate the limit of $F_n^{(k)}(t)$ for $n \to \infty$, we only need to calculate $E1_{[0,k]}(C_t)$, that is, the distribution function of C_t at k. Let us denote the distribution function of C_t as $G_t(x)$,

For $k < 0$, we have $G_t(k) = 0$. For $k = 0$,

$$G_t(0) = \begin{cases} 1-\lambda(t), & \text{if } \lambda(t) \leq 1, \\ 0, & \text{if } \lambda(t) > 1. \end{cases}$$

For $k > 0$,

$$G_t(k) = \begin{cases} G_t(0), & \text{if } k < \left(1-\sqrt{\lambda(t)}\right)^2, \\ G(0) + \int_{\left(1-\sqrt{\lambda(t)}\right)^2}^{k} p_t(\xi)\, d\xi, & \text{if } k \in \left[\left(1-\sqrt{\lambda(t)}\right)^2, \left(1+\sqrt{\lambda(t)}\right)^2\right], \\ 1 & \text{if } k > \left(1+\sqrt{\lambda(t)}\right)^2. \end{cases}$$

where

$$p_t(\xi) = \frac{\sqrt{4\xi - (1-\lambda(t)+\xi)^2}}{2\pi\xi}.$$

Then, we compute $G_t(k) \equiv \overline{F}(t|k)$ as a function of t for a fixed k. Let $\lambda^{-1}(x)$ denote the solution of the equation $\lambda_t = x$.

For $k = 0$:

$$\overline{F}(t|k) = \begin{cases} 0, & \text{if } t \leq \lambda^{-1}(1), \\ 1-\lambda(t), & \text{if } t > \lambda^{-1}(1). \end{cases}$$

For $k \in (0,1)$:

$$\overline{F^{(k)}}(t) = \begin{cases} 0, & \text{if } t < \lambda^{-1}\left(\left(1+\sqrt{k}\right)^2\right), \\ \int_{\left(1-\sqrt{\lambda(t)}\right)^2}^{k} p_t(\xi)\, d\xi, & \text{if } t \in \left[\lambda^{-1}\left(\left(1+\sqrt{k}\right)^2\right), \lambda^{-1}(1)\right], \\ 1-\lambda(t) + \int_{\left(1-\sqrt{\lambda(t)}\right)^2}^{k} p_t(\xi)\, d\xi, & \text{if } \left(t \in \lambda^{-1}(1), \lambda^{-1}\left(\left(1-\sqrt{k}\right)^2\right)\right], \\ 1-\lambda(t), & \text{if } t > \lambda^{-1}\left(\left(1-\sqrt{k}\right)^2\right). \end{cases}$$

For $k \geq 1$, we have:

$$\overline{F}(t|k) = \begin{cases} 0, & \text{if } t < \lambda^{-1}\left(\left(1+\sqrt{k}\right)^2\right), \\ \int_{\left(1-\sqrt{\lambda(t)}\right)^2}^{k} p_t(\xi)\,d\xi, & \text{if } t \in \left[\lambda^{-1}\left(\left(1+\sqrt{k}\right)^2\right), \lambda^{-1}(1)\right], \\ 1-\lambda(t)+\int_{\left(1-\sqrt{\lambda(t)}\right)^2}^{k} p_t(\xi)\,d\xi, & \text{if } \left(t \in \lambda^{-1}(1), \lambda^{-1}\left(\left(1-\sqrt{k}\right)^2\right)\right], \\ 1, & \text{if } t > \lambda^{-1}\left(\left(1-\sqrt{k}\right)^2\right). \end{cases}$$

Combinining these cases, we obtain the following equation:

$$\overline{F}(t|k) = \begin{cases} 0, & \text{if } t < \lambda^{-1}\left(\left(1+\sqrt{k}\right)^2\right), \\ \int_{\left(1-\sqrt{\lambda(t)}\right)^2}^{k} p_t(\xi)\,d\xi, & \text{if } t \in \left[\lambda^{-1}\left(\left(1+\sqrt{k}\right)^2\right), \lambda^{-1}(1)\right], \\ 1-\lambda(t)+\int_{\left(1-\sqrt{\lambda(t)}\right)^2}^{k} p_t(\xi)\,d\xi, & \text{if } \left(t \in \lambda^{-1}(1), \lambda^{-1}\left(\left(1-\sqrt{k}\right)^2\right)\right], \\ 1-\lambda(t)\,1_{[0,1)}(k), & \text{if } t > \lambda^{-1}\left(\left(1-\sqrt{k}\right)^2\right). \end{cases}$$

QED.

Example 210 *Distributions from the domain of attraction of Φ^{ν} law*

Consider the case of convergence to the law Φ^{ν}, when the constants a_n and b_n are chosen in such a way, that the limit law is $G(x) = \exp\left(-x^{-\lambda}\right)$ for $x > 0$.Then we can conclude that the limit distribution of the k order statistic is given as follows:

$$\overline{F}(t|k) = \begin{cases} 0, & \text{if } t < \left(1+\sqrt{k}\right)^{-2/\nu}, \\ \int_{\left(1-t^{-\nu/2}\right)^2}^{k} p_t(\xi)\,d\xi, & \text{if } t \in \left[\left(1+\sqrt{k}\right)^{-2/\nu}, 1\right], \\ 1-t^{-\nu}+\int_{\left(1-t^{-\nu/2}\right)^2}^{k} p_t(\xi)\,d\xi, & \text{if } t \in \left(1, \left(\left(1-\sqrt{k}\right)^2\right)^{-1/\nu}\right], \\ 1-t^{-\nu}1_{[0,1)}(k), & \text{if } t > \left(\left(1-\sqrt{k}\right)^2\right)^{-1/\nu}, \end{cases}$$

where

$$p_t(\xi) = \frac{\sqrt{4\xi - \left(1-t^{-\nu}+\xi\right)^2}}{2\pi\xi}.$$

We illustrate this result for some particular values of ν and k.

Consider $k = 0$. Then

$$\overline{F}\,(t|0) = \begin{cases} 0, & \text{if } t < 1, \\ 1 - t^{-\nu}, & \text{if } t \geq 1. \end{cases}$$

This is the Type 2 (“Pareto”) limit distribution in Definition 6.8 of (Ben Arous and Voiculescu 2006).

It is interesting to note that if $k > 1$, then for all sufficiently large t, $\overline{F}\,(t|k) = 1$. This can be interpreted as saying that the scaled k order statistic is guaranteed to be less than t_0 for a suffiiciently large k. In another interpretation, this result means that for our choice of scaling parameters a_n and b_n and for every $k > 1$,

$$\left\| \sum_{i=1}^{n} 1_{(b_n t + a_n)}\,(X_i) \right\| < k$$

if t and n are sufficiently large.

A similar situation occurs in the classical case if the initial distribution (i.e. the distribution of X_i) is bounded from above. In this case the limit distribution is also bounded from above. In contrast, in the free probability case this situation occurs even if the initial distribution is unbounded from above. Our previous example shows that this situation occurs even if the initial distribution has heavy tails.

A Appendix: Distributions and Their Transforms

	Semicircle
Density & atoms	$\frac{1}{2\pi}\sqrt{4-x^2}\chi_{[-2,2]}(x)$
Cauchy transform	$\frac{1}{2}\left(z-\sqrt{z^2-4}\right)$
K-function	$\frac{1}{z}+z$
S-transform	not defined
Moments	$m_{2k}=\frac{1}{k+1}\binom{2k}{k}$, $m_{2k+1}=0$.
Cumulants	$c_1=1$, $c_i=0$ for $i\neq 1$

	Marchenko-Pastur
Density & atoms	$\frac{\sqrt{4x-(1-\lambda+x)^2}}{2\pi x}\chi_{\left[\left(1-\sqrt{\lambda}\right)^2,\left(1+\sqrt{\lambda}\right)^2\right]}(x)$ and an atom at 0 with mass $(1-\lambda)$ if $\lambda<1$
Cauchy transform	$\frac{1-\lambda+z-\sqrt{(1-\lambda+z)^2-4z}}{2z}$
K-function	$\frac{1}{z}+\frac{\lambda}{1-z}$
S-transform	$\frac{1}{\lambda+z}$
Moments	
Cumulants	$c_i=\lambda$ for all i.

	Bernoulli
Density & atoms	Atom at 0 with probability q, and atom at 1 with probability p.
Cauchy transform	$\frac{z-q}{(z-1)z}$
K-function	$\frac{1+z-\sqrt{(1-z)^2+4pz}}{2z}$
S-transform	$\frac{1+z}{q+z}$
Moments	$m_i=q$ for all i
Cumulants	

	Arcsine
Density & atoms	$\frac{\chi_{[-2,2]}(t)}{\pi\sqrt{4-t^2}}$
Cauchy transform	$\frac{1}{\sqrt{z^2-4}}$
K-function	$\frac{1}{z}\sqrt{1+4z^2}$
S-transform	not defined
Moments	$m_n=\binom{2k}{k}$ if $n=2k$; $=0$ if $n=2k+1$
Cumulants	

Index

References

Akhieser, N. I. (1961). *The Classical Problem of Moments*. State Publishing House of Physical and Mathematical Literature, Moscow. In Russian. An English translation is available.

Ando, T. (1989). Majorization, doubly stochastic matrices, and comparison of eigenvalues. *Linear Algebra and Its Applications 118*, 163–248.

Avitzour, D. (1982). Free products of C*-algebras. *Transactions of the American Mathematical Society 271*, 423–435.

Bai, Z. D. (1993). Convergence rate of expected spectral distributions of large random matrices. part I. Wigner matrices. *Annals of Probability 21*, 625–648.

Bai, Z. D. and J. W. Silverstein (1998). No eigenvalues outside of the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Annals of Probability 26*, 316–345.

Barndorff-Nielsen, O. E. and S. Thorbjornsen (2005). The Levy-Ito decomposition in free probability. *Probability Theory and Related Fields 131*, 197–228.

Belinschi, S. T. and H. Bercovici (2004). Atoms and regularity for measures in a partially defined free convolution semigroup. *Mathematische Zeitschrift 248*, 665–674.

Belinschi, S. T. and H. Bercovici (2005). Partially defined semigroups relative to multiplicative free convolution. *International Mathematics Research Notices 2005*, 65–101.

Bellman, R. (1954). Limit theorems for non-commutative operations I. *Duke Mathematical Journal 21*, 491–500.

Ben Arous, G. and A. Guionnet (1997). Large deviations for Wigner′s law and Voiculescu′s non-commutative entropy. *Probability Theory and Related Fields 108*, 517–542.

Ben Arous, G. and D. Voiculescu (2006). Free extreme values. *Annals of Probability 34*, 2037–2059.

Bercovici, H. and W. S. Li (2001). Inequalities for eigenvalues of sums in a von Neumann algebra. In L. Kerchy, C. Foias, I. Gohberg, and H. Langer (Eds.), *Recent Advances in Operator Theory and Related Topics*, Volume 127 of *Op-*

erator Theory Advances and Applications, pp. 113–126. Birkhauser: Basel, Boston, Berlin.

Bercovici, H. and V. Pata (1996). The law of large numbers for free identically distributed random variables. *Annals of Probability 24*, 453–465.

Bercovici, H., V. Pata, and P. Biane (1999). Stable laws and domains of attraction in free probability theory. *Annals of Mathematics 149*, 1023–1060. with an appendix by Philippe Biane.

Bercovici, H. and D. Voiculescu (1992). Levy-Hincin type theorems for multiplicative and additive free convolution. *Pacific Journal of Mathematics 153*, 217–248.

Bercovici, H. and D. Voiculescu (1993). Free convolution of measures with unbounded support. *Indiana University Mathematics Journal 42*, 733–773.

Bercovici, H. and D. Voiculescu (1995). Superconvergence to the central limit and failure of the Cramer theorem for free random variables. *Probability Theory and Related Fields 102*, 215–222.

Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variables. *Transactions of American Mathematical Society 49*, 122–136.

Biane, P. (1998). Representations of symmetric groups and free probability. *Advances in Mathematics 138*, 126–181.

Bratteli, O. and D. W. Robinson (1987). *Operator Algebras and Quantum Statistical Mechanics 1* (2 ed.), Volume 1. Springer-Verlag.

Chistyakov, G. and F. Gotze (2006). Limit theorems in free probability theory. I. http://www.arxiv.org/math/OA/0602219.

Chung, K. L. (2001). *A Course in Probability Theory* (Third ed.). Academic Press.

Cohen, J. E. and C. M. Newman (1984). The stability of large random matrices and their products. *Annals of Probability 12*, 283–310.

Diaconis, P. (1988). *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics. Volume 11 in Lecture Notes - Monograph Series of IMS.

Dvoretzky, A. and J. Wolfowitz (1951). Sums of random integers reduced modulo m. *Duke Mathematical Journal 18*, 501–507.

Edelman, A. and N. R. Rao (2005). Random matrix theory. *Acta Numerica*, 233–297.

Esseen, C. G. (1945). Fourier analysis of distribution functions. *Acta Mathematica 77*, 1–125.

Fack, T. (1982). Sur la notion de valeur caracteristique. *Journal of Operator Theory 7*, 307–333.

Figa-Talamanca, A. and T. Steger (1994). Harmonic analysis for anisotropic random walks on homogeneous trees. Volume 110 of *Memoirs of the American Mathematical Society*. American Mathematical Society.

Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society 24*, 180–190.

Frechet, M. (1927). Sur la loi de probabilite de l'ecart maximum. *Annales de la societe polonaise de Mathematiques 6*, 93.

Fuglede, B. and R. V. Kadison (1952). Determinant theory in finite factors. *Annals of Mathematics 55*, 520–530.

Furstenberg, H. (1963). Noncommuting random products. *Transactions of the American Mathematical Society 108*, 377–428.

Furstenberg, H. and H. Kesten (1960). Products of random matrices. *Annals of Mathematical Statistics 31*, 457–469.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics 44*, 423–453.

Gnedenko, B. W. and A. N. Kolmogorov (1959). *Grenzverteilungen von Summen unabhangiger Zufallsgrossen*, Volume 9 of *Mathematische Lehrbucher und Monographien*. Akademie-Verlag, Berlin.

Gohberg, I. C. and M. G. Krein (1969). *Introduction to the Theory of Linear Nonselfadjoint Operators in Hilbert Space*, Volume 18 of *Translations of Mathematical Monographs*. American Mathematical Society.

Grenander, U. (1963). *Probabilities on Algebraic Structures*. John Wiley and Sons Inc.

Haagerup, U. (1997). On Voiculescus R- and S-transforms for free noncommuting random variables. In D.-V. Voiculescu (Ed.), *Free Probability The-*

ory, Volume 12 of *Fields Institute Communications*, pp. 127–148. American Mathematical Society.

Haagerup, U. and S. Thorbjornsen (2005). A new application of random matrices. *Annals of Mathematics 162*, 711–775.

Hatcher, A. (2002). *Algebraic Topology*. Cambridge University Press.

Hiai, F. and D. Petz (2000). *The Semicircle Law, Free Random Variables And Entropy* (1 ed.), Volume 77 of *Mathematical Surveys and Monographs*. American Mathematical Society.

Hille, E. (1962). *Analytic Function Theory*, Volume 1. Ginn and Company.

Isopi, M. and C. M. Newman (1992). The triangle law for Lyapunov exponents of large random matrices. *Communications in Mathematical Physics 143*, 591–598.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal component analysis. *Annals of Statistics 29*, 295–327.

Kargin, V. (2007). The norm of products of free random variables. *Probability Theory and Related Fields 139*, 397–413.

Kawada, Y. and K. Itô (1940). On the probability distribution on a compact group. *Proceedings of Physical and Mathematical Society of Japan 22*, 977–998.

Kingman, J. F. C. (1973). Subadditive ergodic theory. *Annals of Probability 1*, 883–899.

Koosis, P. (1998). *Introduction to H_p Spaces* (2 ed.). Cambridge University Press.

Lagrange (1869). *Oeuvres*, Volume 3. Gauthier-Villars, Paris.

Lehner, F. (1999). Computing norms of free operators with matrix coefficients. *American Journal of Mathematics 121*, 453–486.

Levy, P. (1939). L'addition des variables aleatoires definies sur une circonference. In D. Dugue (Ed.), *Oeuvres de Paul Levy*, Volume 3, pp. 452–492. Gauthier-Villars, (1976).

Lindeberg, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift 15*, 211–225.

Luck, W. (2002). *L^2-Invariants: Theory and Applications to Geometry and K-Theory*. Springer-Verlag, New York.

Lyndon, R. C. and P. E. Schupp (1977). *Combinatorial Group Theory*, Volume 89 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag, New York.

Maassen, H. (1992). Addition of freely independent random variables. *Journal of Functional Analysis 106*, 409–438.

Marcenko, V. and L. Pastur (1967). Distribution of eigenvalues of some sets of random matrices. *Mathematics in U.S.S.R 1*, 507–536.

Markushevich, A. I. (1977). *Theory of Functions of a Complex Variable* (2 ed.). Chelsea Publishing Company, New York.

McLaughlin, J. C. (1986). *Random Walks and Convolution Operators on Free Products*. Ph. D. thesis, NYU.

Newman, C. M. (1986a). The distribution of Lyapunov exponents: Exact results for random matrices. *Communications in Mathematical Physics 103*, 121–126.

Newman, C. M. (1986b). Lyapunov exponents for some products of random matrices: Exact expressions and asymptotic distributions. In J. E. Cohen, H. Kesten, and C. M. Newman (Eds.), *Random Matrices and Their Applications*, Volume 50 of *Contemporary Mathematics*, pp. 183–195. American Mathematical Society.

Nica, A. and R. Speicher (1996). On the multiplication of free n-tuples of noncommutative random variables. *American Journal of Mathematics 118*, 799–837.

Nica, A. and R. Speicher (2006). Lectures on the combinatorics of free probability. Volume 335 of *London Mathematical Society Lecture Note Series*. Cambridge University Press.

Oseledec, V. I. (1968). A multiplicative ergodic theorem. Ljapunov characteristic numbers for dynamical systems. *Transactions of the Moscow Mathematical Society 19*, 197–231.

Parthasarathy, K. R. (1992). *An introduction to Quantum Stochastic Calculus*. Birkhauser, Basel.

Pata, V. (1996). The central limit theorem for free additive convolution. *Journal of Functional Analysis 140*, 359–380.

Rao, N. R., J. Mingo, R. Speicher, and A. Edelman (2008). Statistical eigen-inference from large Wishart matrices. *Annals of Statistics*. forthcoming.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*, Volume 4 of *Applied Probability*. Springer-Verlag, New York.

Ruelle, D. (1982). Characterisic exponents and invariant manifolds in Hilbert space. *The Annals of Mathematics 115*, 243–290.

Ruelle, D. (1984). Characterisic exponents for a viscous fluid subjected to time dependent forces. *Communications in Mathematical Physics 93*, 285–300.

Saloff-Coste, L. (2004). Random walks on finite groups. In H. Kesten (Ed.), *Probability on Discrete Structures*, pp. 263–346. Volume 110 in Encyclopaedia of Mathematical Sciences.

Segal, I. E. (1947). Irreducible representations of operator algebras. *Bulletin of American Mathematical Society 53*, 73–88.

Shiryaev, A. N. (1995). *Probability* (Second ed.), Volume 95 of *Graduate Text in Mathematics*. Springer.

Soshnikov, A. (1999). Universality at the edge of the spectrum in Wigner random matrices. *Communications in Mathematical Physics 207*, 697–733.

Soshnikov, A. (2002). A note on the universality of the distribution of largest eigenvalues in certain sample covariance matrices. *Journal of Statistical Physics 108*, 1033–1056.

Speicher, R. (1990). A new example of independence and white noise. *Probability Theory and Related Fields 84*, 141–159.

Takesaki, M. (1979). *Theory of Operator Algebras*, Volume 1. Springer-Verlag, New York.

Tracy, C. A. and H. Widom (1996). On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics 177*, 727–754.

Tracy, C. A. and H. Widom (2000). The distribution of the largest eigenvalue in the gaussian ensembles. In J. van Diejen and L. Vinet (Eds.), *Calogero-Moser-Sutherland Models*, pp. 461–472. Springer, New York.

Voiculescu, D. (1983). Symmetries of some reduced free product C^*-algebras. In *Lecture Notes in Mathematics*, Volume 1132, pp. 556–588. Springer-Verlag, New York.

Voiculescu, D. (1986). Addition of certain non-commuting random variables. *Journal of Functional Analysis 66*, 323–346.

Voiculescu, D. (1987). Multiplication of certain non-commuting random variables. *Journal of Operator Theory 18*, 223–235.

Voiculescu, D. (1990). Circular and semicircular systems and free product factors. In A. Connes, M. Duflo, A. Joseph, and R. Rentschler (Eds.), *Operator Algebras, Unitary Representations, Enveloping Algebras, and Invariant Theory*, Volume 92 of *Progress in Mathematics*, pp. 45–60. Birkhauser, Boston.

Voiculescu, D. (1991). Limit laws for random matrices and free products. *Inventiones mathematicae 104*, 201–220.

Voiculescu, D. (1993). The analogues of entropy and of Fisher′s information measure in free probability theory I. *Communications in Mathematical Physics 155*, 71–92.

Voiculescu, D. (1994). The analogues of entropy and of Fisher′s information measure in free probability theory II. *Inventiones Mathematicae 118*, 411–440.

Voiculescu, D. (1996a). The analogues of entropy and of Fisher′s information measure in free probability theory III: The absence of Cartan subalgebras. *Geometric and Functional Analysis 6*, 172–199.

Voiculescu, D. (1996b). Appendix: Alternative proofs for the type II free Poisson variables and for the free compression results. *American Journal of Mathematics 118*, 832–837.

Voiculescu, D. (1998). Lectures on free probability theory. In P. Bernard (Ed.), *Lectures on Probability Theory and Statistics*, Volume 1738 of *Lecture Notes in Mathematics*, pp. 279–349. Springer.

Voiculescu, D., K. Dykema, and A. Nica (1992). *Free Random Variables*. A.M.S. Providence, RI. CRM Monograph series, No.1.

Vorobev, N. N. (1954). The addition of independent random variables on finite groups. *Matematicheskii Sbornik 34*, 89–126. In Russian.

Whittaker, E. T. and G. N. Watson (1927). *A Course of Modern Analysis* (4 ed.). Cambridge University Press.