Introduction to non-commutative probability

Isak Wulff Mottelson

February 24, 2012

Supervisor: Magdalena Musat

Abstract

In this project several fundamental concepts of free probability are presented. We define the $\mathfrak{R}$ transform and prove that it linearizes free additive convolution. This is used to prove the free central limit theorem, which in turn is used to prove a variant of the celebrated Wigner's Law of random matrices. The project is in english.

Contents

Introduction

Free probability was introduced in the 80's by Dan Voiculescu to attack the elusive 'free group factors ismomorphism problem' of operator algebras. Although not sufficient to solve the problem completely, the theory enjoyed a great deal of success and, as a result, a great deal of interest. Later, important connections to many other areas of mathematics has been discovered, warranting indenpendent interest in free probability. This is the point of view of the present presentation.

In section 1 we introduce the fundamental concepts of free probability and present a number of examples. The proofs are mostly elementary and may be skipped on a first reading. In section 2 The $\mathfrak{R}$ transform is introduced and we investigate how it interacts with free convolution. The proofs of this section are mainly due to Uffe Haagerup [4]. In section 3 the free central limit theorem is proved and it is used in the final section to prove a variant of Wigner's Law due to Dan Voiculescu [1]. This result, though originally intended for other purposes, illustrates how methods from free probability can be used to study problems of independent interest.
In the appendix a short introduction to formal power series is given, including a proof of the Lagrange Inversion Formula, used in the discussion of the $\mathfrak{R}$ transform.

I wish to thank my advisor Magdalena Musat for her help and advice. Although our communication have not always been optimal, I sincerely appreciate the respect and support she has shown in the proccess.

Isak Mottelson
Department of mathematical sciences
University of Copenhagen
January, 2012.

1 Free probability

1.1 Non-commutative probability spaces

The prefix 'non-commutative' can be added to many classical areas of mathematics. The programme typically runs roughly as follows:

> Classically one studies a space of some kind. Often, this naturally suggests some algebra of functions on the space, and one then strives to express the interesting properties of the space in terms of this algebra. To earn the prefix 'non-commutative' one then forgets the space, and replaces the algebra with a non-commutative algebra with corresponding structure.

Examples include non-commutative topological spaces (C^*-agebras), non-commutative measure spaces (von Neumann algebras), non-commutative manifolds and many more. Many constructions carry over (loosely) to the non-commutative setting as well, such as non-commutative one-point compactification (unitalization), non-commutative Stone-Čech compactification (multiplier algebras) and non-commutative L^p-spaces. In the following we introduce the concepts of a non-commutative probability space, as well as some of the corresponding contructions, and as we shall see, very closely follow the outline given above.

The fundamental notion in probability theory is that of a probability space, i.e., the triple $(\Omega, \mathcal{S}, P)$ with Ω a set, $\mathcal{S}$ a σ-algebra on Ω and P a probability measure on $(\Omega, \mathcal{S})$. We may then study the random variables, i.e., the measurable maps $X : \Omega \to \mathbb{C}$ (or $X : \Omega \to \mathbb{R}$). This algebra is too big to study even in the commutative case, so it is a question to which smaller algebra we should try to apply the above procedure. An example could be the algebra

$$L = \bigcap_{p=1}^{\infty} L^p(\Omega, \mathcal{S}, P)$$

of random variables with moments of all orders, or one may study the subalgebra $L^\infty(\Omega, \mathcal{S}, P) \subseteq L$ of essentially bounded random variables[1]. On these algebras we have a natural functional $\mathbb{E}$ given by

$$\mathbb{E}(f) = \int_\Omega f dP, \qquad f \in L$$

and note that $\mathbb{E}(1) = 1$ with 1 being the constant function 1 on Ω. This suggests the following definition:

Definition 1.1. *A non-commutative probability space is a pair (A, ϕ) where A is an algebra with unit, and ϕ is a linear functional on A taking the value 1 at the indentity element of A*

It is customary to refer to the elements of A as non-commutative stochastic variables (or random variables) and to the functional as the expectation. This definition is very generous in that it allows any algebra and any functional as

[1] We gloss over the fact that the above algebras actually consist of *equivalence classes* of functions.

input data, which sometimes can make arguments simpler. On the other hand, because of its generality it permits no analytic arguments of any kind. Further, the notion of a positive random variable cannot be expressed in this level of generality. To remedy this, we have a more restrictive definition:

Definition 1.2. *A C^*-probability space is a pair (A, ϕ) where A is a C^*-algebra with unit and ϕ is a state.*

To illustrate the above definitions, we give a few examples:

Example 1.3. *Let I be any set and denote by $\mathbb{C}\langle X_i \mid i \in I\rangle$ the non-commutative polynomials in the symbols $(X_i)_{i\in I}$, i.e., linear combinations of $X_{i_1}^{m_1}\cdots X_{i_n}^{m_n}$, $n \in \mathbb{N}$, $m_1, \ldots m_n \in \mathbb{N}$ and $i_1, \ldots i_n \in I$ such that $i_1 \neq i_2 \neq \cdots \neq i_n$. Note that the possibility that $i_1 = i_3$ is not precluded. Together with any functional μ taking the value 1 at the unit of the algebra $(\mathbb{C}\langle X_i \mid i \in I\rangle, \mu)$, this is a non-commutative probability space.*

The next example will serve as a model space for phenomena occuring in any non-commutative probability space. It has both very algebraic and very analytic properties at the same time:

Example 1.4 (Full Fock spaces)**.** *Let $\mathfrak{H}_n$ denote an n-dimensional Hilbert space with orthonormal basis $e_1, \ldots, e_n$. Consider* the full Fock space over $\mathfrak{H}_n$ *defined by*

$$\mathfrak{F}(\mathfrak{H}_n) = \mathbb{C}\Omega \oplus \bigoplus_{k=1}^{\infty} \mathfrak{H}_{\mathfrak{n}}{}^{\otimes k}.$$

Here Ω denotes some distinguished unit vector. Consider the algebra $B(\mathfrak{F}(\mathfrak{H}_n))$ of bounded linear operators on the full Fock space over $\mathfrak{H}_n$, together with the linear functional ω given by

$$\omega(T) = \langle T\Omega, \Omega\rangle, \qquad T \in B(\mathfrak{F}(\mathfrak{H}_n)).$$

Since $B(\mathfrak{F}(\mathfrak{H}_n))$ is a unital C^*-algebra, and ω is a state (a vector state, even) the pair $(B(\mathfrak{F}(\mathfrak{H}_n)), \omega)$ is a C^*-probability space for any Hilbert space $\mathfrak{H}_n$.

1.2 Distributions and freeness

A classical radom variable $X : \Omega \to \mathbb{C}$ gives rise to a measure $X(P)$ on $\mathbb{C}$ defined by

$$X(P)(E) = P(X^{-1}(E)), \qquad E \in \mathbb{B}(\mathbb{C}).$$

We refer to $X(P)$ as the distribution of X (with respect to P). Note that in a certain sense, a measure is exactly a functional on the functions on the measure space given by integration with respect to the measure. And if X has moments of all orders we have the change of variable formula

$$\int_{\mathbb{R}} p dX(P) = \int_{\Omega} p \circ X dP = \mathbb{E}(p \circ X),$$

for each polynomial $p \in \mathbb{C}[X]$. For large classes of random variables, it is even true that the measure is completely determined by the values of the above funtional on polynomials. This suggests the following definition:

Definition 1.5. *If a is a random variable in a non-commutative probability space (A, ϕ) we denote by μ_a the functional $\mu_a : \mathbb{C}[X] \to \mathbb{C}$ given by*

$$\mu_a\left(\sum_{n=1}^{k} c_n X^n\right) = \sum_{n=1}^{k} c_n \phi(a^n).$$

That is, $\mu_a(p) = \phi(p(a))$ for all $p \in \mathbb{C}[X]$. We refer to μ_a as the distribution of a.

More generally, we define

Definition 1.6. *If I is an index set and $(a_i)_{i\in I}$ a family of random variables in a non-commutative probability space (A, ϕ), define the joint distribution $\mu_{(a_i)_{i\in I}}$ as the functional on $\mathbb{C}\langle X_i \mid i \in I\rangle$ given by*

$$\mu_{(a_i)_{i\in I}}\left(X_{i_1}^{m_1} \cdots X_{i_n}^{m_n}\right) = \phi\left(a_{i_1}^{m_1} \cdots a_{i_n}^{m_n}\right)$$

Note that this identifies the distribution of an element/family of elements with the collection of (non-commutative) moments of the element/family.
The following example illustrates the approach:

Example 1.7. *Let (A, ϕ) be a C^*-probability space, and let a be a normal element in A. By definition the distribution of a is the map*

$$\mu_a : \mathbb{C}[X] \to \mathbb{C} \quad \text{given by} \quad \mu_a(P) = \phi(P(a)).$$

However, in this case we can use the Spectral Theorem to extend this functional to the C^*-algebra $C(\sigma(a))$ simply by defining $\mu_a(f) = \phi(f(a))$. By the properties of the continuous functional calculus, to wit it is continuous and preserves positivity, this gives us a state on $C(\sigma(a))$. Now the Riesz Representation Theorem yields a unique regular probability measure μ such that

$$\mu_a(f) = \int_{\sigma(a)} f d\mu \qquad \forall\ f \in C(\sigma(a)).$$

Since the above extension of μ_a to $C(\sigma(a))$ is the only continuous one (the polynomials being dense in $C(\sigma(a))$) μ is the only measure with this property. Hence, if the C^*-algebra is already an algebra of stochastic variables on some probability space, the above construction gives back the classically defined distribution.

A very important concept in classical probability is the notion of independent random variables. Classically, random variables $X_1, \ldots, X_n$ are said to be independent if the joint distribution $(X_1, \ldots, X_n)(P)$ on $\mathbb{C}^n$ is the product measure $X_1(P) \otimes \cdots \otimes X_n(P)$. In particular, if $p_1, \ldots p_n$ are polynomials in one variable, we get (by Fubini's Theorem) that

$$\mathbb{E}\left(p_1(X_1) \cdots p_n(X_n)\right) = \prod_{i=1}^{n} \mathbb{E}\left(p_i(X_i)\right).$$

We can easily translate this notion directly to the setting of non-commutative probability. We might say that non-commutative random variables $x_1, \ldots x_n$ in (A, ϕ) are independent if

$$\phi(p_1(x_1)p_2(x_2) \cdots p_n(x_n)) = \phi(p(x_1))\phi(p_2(x_2)) \cdots \phi(p_n(x_n)),$$

but that seems ill advised for the following reason. In the commutative case, the desirable aspect of independent variables is that their joint distribution is completely determined from the marginal distributions. If we simply require the above condition for the non-commutative situation this will no longer hold, since it gives no information on the expectations of non-commutative polynomials in the given variables, e.g., $\phi(x_1x_2x_1)$. One solution is to additionally require that the variables under consideration commute, but that seems to defeat the purpose of introducing non-commutative probability theory. Instead, we introduce the notion of freeness, which is a purely non-commutative phenomenon:

Definition 1.8. *Let (A, ϕ) be a non-commutative probability space and let $(\mathcal{A}_i)_{i\in I}$ be a family of unital subalgebras. We say that the family is free if*

$$\phi(a_1 \cdots a_n) = 0$$

whenever $a_i \in \mathcal{A}_{m_i}$, $m_1 \neq m_2 \neq \cdots \neq m_n$ with $\phi(A_i) = 0$ for $i = 1, \ldots n$. A family of elements $(a_i)_{i\in I}$ is said to be free if $(\mathrm{alg}(a_i))_{i\in I}$ is[2] free.

Similarly to classically independent variables we have that the distribution of $a + b$ only depends on the marginal distributions:

Proposition 1.9. *Let a, b be free random variables in a non-commutative probability space $(\mathcal{A}, \phi)$. Then μ_{a+b} depends only on μ_a and μ_b. More specifically, there exists polynomials P_n independent of a and b such that*

$$\mu_{a+b}(X^n) = P_n\left(\mu_a(X), \ldots, \mu_a(X^n), \mu_b(X), \ldots, \mu_b(X^n)\right).$$

We then denote $\mu_{a+b} = \mu_a \boxplus \mu_b$ - the additive free convolution.

Proof. The author remarks that although it may seem complicated, the proof is really simple, and may perhaps even be skipped by the confident reader. Start by writing

$$\mu_{a+b}(X^n) = \phi\left((a+b)^n\right)$$

and expand into monomials. Let $A_1 = \mathrm{alg}(a, 1)$, $A_2 = \mathrm{alg}(b, 1)$ and write the monomials as

$$c = c_1c_2 \cdots c_m, \quad m \in \mathbb{N} \quad c_j \in A_{i_j} \quad i_1 \neq i_2 \neq \cdots \neq i_m$$

where c_j is a power n_j of either a or b for all j and $\sum n_j \leq n$. Now, for any $x \in \mathcal{A}$ put $\overset{o}{x} = x - \phi(x)1$. Then calculate:

$$\begin{aligned}\phi(c_1 \cdots c_m) &= \phi\left((\overset{o}{c_1} + \phi(c_1)1) \cdots (\overset{o}{c_m} + \phi(c_m))\right) \\ &= \sum_{S\subseteq\{1,\ldots,m\}} \left(\prod_{j\in\{1,\ldots,m\}\setminus S} \phi(c_j)\right) \phi\left(\prod_{j\in S} \overset{o}{c_j}\right)\end{aligned}$$

The crucial observation is then the fact that if $S = \{1, \ldots, m\}$, then

$$\phi\left(\prod_{j\in S} \overset{o}{c_j}\right) = \phi(\overset{o}{c_1} \cdots \overset{o}{c_m}) = 0$$

[2]Or $(C^*(a_i))_{i\in I}$ in the case of a C^*-probability space.

by assumption. Hence we may assume $S \subsetneq \{1, \ldots, m\}$. Now employ the same manoeuvre:

$$\phi\left(\prod_{j\in S}{}^{o}_{} c_j\right) = \sum_{R\subseteq S}\left(\prod_{j\in S\setminus R}(-\phi(c_j))\right)\phi\left(\prod_{j\in R} c_j\right)$$

Now note that $\prod_{j\in R} c_j$ can be written as $\tilde{c_1}\tilde{c_2}\cdots\tilde{c_q}$ with[3] $\tilde{c_j} \in A_{k_j}$, $k_1 \neq \cdots \neq k_q$ for some $q \leq |R| < m$ Hence, a lot of moments of a and b aside we are back where we started, only with strictly fewer factors. An induction argument then finishes the proof. □

Note that the proof shows that we can calculate the expectation of any non-commutative polynomial in free variables only knowing the moments of the individual variables.
An important exmple to illustrate the above concept is in order:

1.3 The model example

Let $\mathfrak{H}_n$ denote an n-dimensional Hilbert space with orthonormal basis $e_1, \ldots, e_n$ and recall the non-commutative probability space $(B(\mathfrak{F}(\mathfrak{H}_n)), \omega)$ of example 1.4. Define the *creation operators* $\ell_k : \mathfrak{F}(\mathfrak{H}_n) \to \mathfrak{F}(\mathfrak{H}_n)$, $k = 1, \ldots, n$ on $\mathfrak{F}(\mathfrak{H}_n)$ by

$$\ell_k(e_{j_1} \otimes \cdots e_{j_m}) = e_k \otimes e_{j_1} \otimes \cdots e_{j_m}, \quad \ell_k\Omega = e_k.$$

These maps are clearly linear isometries and hence in $B(\mathfrak{F}(\mathfrak{H}_n))$. Further, we immediately see that

$$\ell_j^*(e_{i_1} \otimes \cdots \otimes e_{i_m}) = \langle e_j, e_{i_1}\rangle\, e_{i_2} \otimes \cdots \otimes e_{i_m}, \quad \ell_j^* e_j = \Omega.$$

We now have

Lemma 1.10. *The families $\left\{\ell_j, \ell_j^*\right\}$ are free in $(B(\mathfrak{F}(\mathfrak{H}_n)), \omega)$, viewed as either a non-commutative probability space or a C^*-probability space. We say that the ℓ_js are $*$-free.*

Proof. Since the ℓ_js are isometries we have $\ell_j^*\ell_j = 1$. Therefore

$$A_j = \mathrm{alg}(\ell_j, \ell_j^*) = \mathrm{span}\left\{\ell_j^m(\ell_j^*)^n \mid m, n \in \mathbb{N}_0\right\}.$$

Furthermore, we see that

$$\omega\left(\ell_j^m(\ell_j^*)^n\right) = \left\langle(\ell_j^*)^m\Omega, (\ell_j^*)^n\Omega\right\rangle = \begin{cases} 1 & \text{for } n = m = 0 \\ 0 & \text{else} \end{cases}$$

Note that this actually proves that $\omega(x) = 0$ for all non-scalar $x \in A_j$.
Now let $j_1 \neq j_2 \neq \cdots \neq j_k$ and assume

$$\omega\left(\ell_{j_i}^{m_i}(\ell_{j_i}^*)^{n_i}\right) = 0, \quad i = 1, \ldots, k, \ \omega(\ell_{j_1}^{m_1}(\ell_{j_1}^*)^{n_1}\cdots\ell_{j_k}^{m_k}(\ell_{j_k}^*)^{n_k}) \neq 0.$$

Now calculate:

$$\begin{aligned}\omega\left(\ell_{j_1}^{m_1}(\ell_{j_1}^*)^{n_1}\cdots\ell_{j_k}^{m_k}(\ell_{j_k}^*)^{n_k}\right) &= \left\langle\ell_{j_1}^{m_1}(\ell_{j_1}^*)^{n_1}\cdots\ell_{j_k}^{m_k}(\ell_{j_k}^*)^{n_k}\Omega, \Omega\right\rangle \\ &= \left\langle(\ell_{j_1}^*)^{n_1}\cdots\ell_{j_k}^{m_k}(\ell_{j_k}^*)^{n_k}\Omega, (\ell_{j_1}^*)^{m_1}\Omega\right\rangle.\end{aligned}$$

[3] We even still have that each factor is a power of a or b.

If $\omega(\ell_{j_1}^{m_1}(\ell_{j_1}^*)^{n_1}) = 0$, then either m_1 or n_1 is different from 0. If $m_1 \neq 0$, then $(\ell_{j_1}^*)^{m_1}\Omega = 0$ and the above expression is zero. If $m_1 = 0$, then (as $n_1 \neq 0$) we must have $m_2 = 0$, since $\ell_j^*\ell_k = 0$ for $j \neq k$. Inductively, we get $m_1 = m_2 = \cdots = m_k = 0$. But then $n_k \neq 0$, and the expectation is zero anyway (since $\ell_{j_k}^*\Omega = 0$).

By continuity, the above proof also shows that the ℓ_j's are $*$-free in the C^*-algebra sense. □

For independent classical random variables X, Y it is well known that $(X + Y)(P) = X(P) * Y(P)$ (convolution). To analyze such quantities one has the Fourier transform, which turns convolution into multiplication. One may then even take the logarithm to completely linearize the before so complicated phenomenon. In the next section we introduce the $\mathfrak{R}$-transform of Voiculescu and prove that it has a similar property with respect to free convolution.

2 The $\mathfrak{R}$ transform

2.1 Definition and properties

The present work approaches the $\mathfrak{R}$ transform from the perspective of formal power series, developed in the appendix. For an element a in a non-commutative probability space (A, ϕ), define the 'formal Cauchy transform' of a as a formal power series $G_a(\lambda)$ by

$$G_a(\lambda) = \sum_{n=0}^{\infty} \lambda^{-n-1}\phi(a^n).$$

Strictly speaking, this is may not be a bona fide Laurent series, since arbitrarily many coefficients correspoding to negative powers of λ may occur. Hence, for simplicity, we consider $\psi(t) \in \mathbb{C}[[t]]$ defined by

$$\psi(t) = \sum_{n=0}^{\infty} t^{n+1}\phi(a^n).$$

The philosophy of this object is to encode all the moments $m_n = \phi(a^n)$ of a in a single power series. Since there is no constant term in ψ and the first order term has coefficient 1, the series can be inverted in $\mathbb{C}[[t]]$ with respect to composition (see the appendix) to get a series $\psi^{(-1)}(z) \in \mathbb{C}[[z]]$. We then makte the following definition.

Definition 2.1. *If a is a random variable in a non-commutative probability space (A, ϕ) we define the $\mathfrak{R}$ transform of a as the formal power series*

$$\mathfrak{R}_a(z) = \frac{1}{\psi^{(-1)}(z)} - \frac{1}{z}.$$

Here $\psi(t)$ is the formal power series

$$\psi(t) = t\left(1 + \sum_{n=1}^{\infty} m_n t^n\right).$$

We now claim that $\mathfrak{R}_a(z) \in \mathbb{C}[[z]]$ and denote it

$$\mathfrak{R}_a(z) = \sum_{p=0}^{\infty} r_{p+1} z^p.$$

The coefficients of the series $(r_p)_{p=1}^{\infty}$ are called the free cumulants of a. That this actually defines a formal power series follows directly from the composition and inversion formulas in the appendix. The reason for introducing this strange object is primarily the following two facts:

(I) The moments $(m_k)_{k=1}^{\infty}$ are completely determined by the $\mathfrak{R}$ transform. Furthermore, there exist fixed polynomials (independent of the random variable in question) expressing the first k cumulants in terms of the first k moments and vice versa.

(II) If a and b are free random variables in (A, ϕ), then

$$\mathfrak{R}_{a+b}(z) = \mathfrak{R}_a(z) + \mathfrak{R}_b(z)$$

The first point follows from the Lagrange inversion formula, proved in the appendix. More precisely one has:

Theorem 2.2 (The moment-cumulant formulas)**.** *The cumulants may be expressed in terms of the moments as*

$$r_p = m_p + \sum_{k=2}^{p} \frac{(-1)^{k-1}}{k} \binom{p+k-2}{k-1} \sum_{Q_k} m_{q_1} \cdots m_{q_k}. \tag{1}$$

The moments may be expressed in terms of cumulants as

$$m_p = r_p + \sum_{k=2}^{p} \frac{1}{k} \binom{p}{k-1} \sum_{Q_k} r_{q_1} \cdots r_{q_k}. \tag{2}$$

Here Q_k denotes the set

$$Q_k = \left\{ (q_1, \ldots, q_k) \in \mathbb{N}^k \mid \sum_{i=1}^{k} q_i = p \right\}$$

The first few cumulants are expressed below:

$$\begin{aligned} r_1 &= m_1 \\ r_2 &= m_2 - m_1^2 \\ r_3 &= m_3 - 3m_1 m_2 + 2m_1^3. \end{aligned}$$

We now turn our attention towards the property II.

2.2 The $\mathfrak{R}$ transform and free convolution

The strategy is first to prove the result in a very easy model example in $(B(\mathfrak{F}(\mathfrak{H}_2)), \omega)$, and then to extend this to an arbitrary case. To accomplish the first step we prove the following theorem:

Theorem 2.3. *a) Let f be a polynomial, and let*

$$a = \ell_1 + f(\ell_1^*).$$

Then

$$\mathfrak{R}_a(z) = f(z) \quad z \in \mathbb{C}.$$

b) Let f and g be two polynomials, and let

$$a = \ell_1 + f(\ell_1^*)$$
$$b = \ell_2 + g(\ell_2^*).$$

Then a and b are free and

$$\mathfrak{R}_{a+b}(z) = \mathfrak{R}_a(z) + \mathfrak{R}_b(z), \quad z \in \mathbb{C}.$$

Before we begin the proof, note that the above claims are made in a C^*-probability space. Hence the 'formal Cauchy transform' $G(\lambda)$ actually converges, at least for $|\lambda| > \|a\|$. Hence, in this case the $\mathfrak{R}$ transform is actually an analytic function in some domain and

$$\mathfrak{R}(z) = G^{(-1)}(z) - \frac{1}{z}.$$

We use this to employ methods from complex analysis in the following proof.

Proof. a) The strategy is to calculate the formal Cauchy transform by relating it to the resolvent by the Neumann series.
For $z \in \mathbb{C},\ |z| < 1$ define the vector

$$\omega_z = (1 - z\ell_1)^{-1}\Omega = \Omega + \sum_{n=1}^{\infty} z^n e_1^{\otimes n}.$$

Then

$$\ell_1\omega_z = \sum_{n=0}^{\infty} z^n e_1^{\otimes(n+1)} = \frac{1}{z}(\omega_z - \Omega), \quad 0 < |z| < 1.$$

Since we have $\ell_1^*\Omega = 0$ and $\ell_1^*\left(e_1^{\otimes n}\right)$, (here we set $e_1^{\otimes 0} = \Omega$) we get

$$\ell_1^*\omega_z = \sum_{n=1}^{\infty} z^n e_1^{\otimes(n-1)} = z\omega_z, \quad |z| < 1.$$

Therefore we have that

$$(\ell_1^*)^n \omega_z = z^n\omega_z, \quad n \in \mathbb{N}, \quad |z| < 1.$$

Hence for our $a = \ell_1 + f(\ell_1^*)$ we have proved

$$\begin{aligned} a\omega_z &= \frac{1}{z}(\omega_z - \Omega) + f(z)\omega_z \\ &= \left(\frac{1}{z} + f(z)\right)\omega_z - \frac{1}{z}\Omega, \quad 0 < |z| < 1. \end{aligned}$$

This means that

$$\left(\left(\tfrac{1}{z}+f(z)\right)-a\right)\omega_z=\frac{1}{z}\Omega.$$

Since

$$\lim_{z\to 0}\left|\tfrac{1}{z}+f(z)\right|=\infty$$

we can choose $\delta\in\mathbb{R}$, $0<\delta\leq 1$ such that

$$\left|\tfrac{1}{z}+f(z)\right|>\|a\|,\quad \text{when } 0<|z|<\delta.$$

Using the Neumann series we see that $\left(\tfrac{1}{z}+f(z)\right)-a$ is invertible for $0<|z|<\delta$, and

$$\left\{\left(\tfrac{1}{z}+f(z)\right)-a\right\}^{-1}\Omega=z\omega_z.$$

By definition of ω this gives

$$\omega\left(\left\{\left(\tfrac{1}{z}+f(z)\right)-a\right\}^{-1}\right)=z\left\langle\omega_z,\Omega\right\rangle=z.$$

Now we are ready to finish the proof:
We need to set

$$G_a(\lambda)=\sum_{n=0}^{\infty}\lambda^{-n-1}\omega(a^n)=\omega\left((\lambda-a)^{-1}\right),\quad |\lambda|>\|a\|,$$

again using the Neumann series. The above discussion then gives

$$G_a\left(\tfrac{1}{z}+f(z)\right)=z,\quad 0<|z|<\delta.$$

That is, G_a is invertible near infinity, and

$$G_a^{(-1)}(z)=\frac{1}{z}+f(z),\quad 0<|z|<\delta.$$

By the discussion preceding the proof, this gives

$$\mathfrak{R}_a(z)=f(z)$$

as desired.

b) First note that ℓ_1 and ℓ_2 are free, as proved in Lemma 1.10. Now the proof proceeds very much like in the proof of a). That is, we introduce

$$\rho_z=\left(1-z(\ell_1+\ell_2)\right)^{-1}\Omega,\quad |z|<1/2.$$

Then, arguing as above and using that $\ell_1^*\ell_2=\ell_2^*\ell_1=0$ we get

$$\ell_1^*\rho_z=z\rho_z,\quad \ell_2^*\rho_z=z\rho_z.$$

In the same way as in the previous calculations this gives

$$(a+b)\rho_z=\left\{\left(\frac{1}{z}+f(z)+g(z)\right)-(a+b)\right\}\rho_z=\frac{1}{z}\Omega.$$

Inverting this for small z as above we get

$$\omega\left(\left\{\left(\frac{1}{z}+f(z)+g(z)\right)-(a+b)\right\}^{-1}\right)=(z\rho_z,\Omega)=z.$$

As in the first part we conclude

$$\mathfrak{R}_{a+b}(z)=f(z)+g(z)=\mathfrak{R}_a(z)+\mathfrak{R}_b(z)$$

which is what we wanted to prove.

□

This gives (II) remarkably easily in view of (I):

Corollary 2.4. *Let (A,ϕ) be a non-commutative probability space, and let $a,b\in A$ be free. Then*

$$\mathfrak{R}_{a+b}(z)=\mathfrak{R}_a(z)+\mathfrak{R}_b(z)$$

as formal power series.

Proof. Put

$$\mathfrak{R}_a(z)=\sum_{p=0}^{\infty}r_{p+1}(a)z^p$$
$$\mathfrak{R}_b(z)=\sum_{p=0}^{\infty}r_{p+1}(b)z^p$$

Let $n\geq 1$ be given. We prove that $r_p(a+b)=r_p(a)+r_p(b)$ for all $p\leq n$. Define

$$a'=\ell_1+\sum_{p=0}^{n}r_{p+1}(a)(\ell_1^*)^p$$
$$b'=\ell_2+\sum_{p=0}^{n}r_{p+1}(b)(\ell_2^*)^p$$

in $B(\mathfrak{F}(\mathfrak{H}_2))$. Then $\mathfrak{R}_a(z)$ and $\mathfrak{R}_{a'}(z)$ agrees on the first n coefficients. By (I) this implies that the first n moments of a and a' agree (and the same for b and b'). By Proposition 1.9 and Lemma 1.10 this gives

$$\phi\left((a+b)^p\right)=\omega\left((a'+b')^p\right)$$

for $p\leq n$. Hence by (I) again $\mathfrak{R}_{a+b}(z)$ coincides with $\mathfrak{R}_{a'+b'}(z)$ on the first n coefficients. But by Theorem 2.3 we have $r_p(a'+b')=r_p(a')+r_p(b')=r_p(a)+r_p(b)$ as we wanted to prove. □

3 The central limit theorem

Recall the central limit theorem of classical probability (see [5], Theorem 2.7):

Theorem 3.1 (The central limit theorem)**.** *Let $(X_n)_{n=1}^{\infty}$ be a sequence of independent identically distributed random variables with mean μ and variance σ^2. Then*

$$X_N = \sqrt{N}\left(\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) - \mu\right)$$

converges in distribution towards a random variable with distribution $\mathcal{N}(0, \sigma^2)$.

If one imposes further conditions on the variables one may relax the assumption of identical distributions.
In this section we prove the free analogue of this result. The notion of convergence in distribution is a natural one in non-commutative probability:

Definition 3.2. *If $\mu_n : \mathbb{C}[X] \to \mathbb{C}$ is a sequence of distributions, we say that μ_n converges to a distribution $\mu : \mathbb{C}[X] \to \mathbb{C}$ if*

$$\mu_n(P) \to \mu(P)$$

For all $P \in \mathbb{C}[X]$. Now let (A, ϕ) be a non-commutative probability space, and $(a_n)_{n\in\mathbb{N}}$ be a sequence of random variables with distributions μ_n. Then we say that a_n converges to μ in distribution if μ_n converges to μ.

Before we state the free central limit theorem, let us introduce the distribution playing the role of the normal distribution of classical probability:

Definition 3.3. *The* semi-circle *distribution with center a and radius $r > 0$ is the distribution $\gamma_{a,r} : \mathbb{C}[X] \to \mathbb{C}$ such that*

$$\gamma_{a,r}(P) = \frac{2}{r^2\pi}\int_{a-r}^{a+r} P(t)\sqrt{r^2 - (t-a)^2}dt,$$

for all $P \in \mathbb{C}[X]$.

Equivalently, we could have just listed all the moments of the semi-circle distribution, and since this is the way we approach distributions in non-commutative probability, we calculate them below:

Lemma 3.4. *The semi-circle distribution $\gamma_{0,r}$ has moments*

$$m_{2n} = \left(\frac{r}{2}\right)^{2n}\frac{(2n)!}{n!(n+1)!}.$$

The odd moments are 0 since the density is even.

Proof. This is just integration.

$$\begin{aligned} m_{2n} &= \frac{2}{\pi r^2}\int_{-r}^{r} x^{2n}\sqrt{r^2 - x^2}dx \\ &= \frac{2}{\pi}r^{2n}\int_{-1}^{1} t^{2n}\sqrt{1-t^2}dt \\ &= \frac{2r^{2n}}{\pi}\int_{-\pi/2}^{\pi/2} \sin^{2n} t\cos^2 t dt \\ &= \frac{r^{2n}}{\pi(2n+1)}\int_{0}^{2\pi} \sin^{2n+2} t dt. \end{aligned} \tag{3}$$

Here the second and third lines follow from elementary substitions and the last line follow from integration by parts and using that sin is odd. We now employ standard methods from complex analysis to finish the calculation:

$$\begin{aligned}\int_0^{2\pi} \sin^{2n} t dt &= \int_{\mathbb{T}} -iz^{-1}\left(\frac{z-z^{-1}}{2i}\right)^{2n} dz \\ &= \frac{-i(-1)^n}{4^n}\int_{\mathbb{T}} z^{-(2n+1)}\sum_{k=0}^{2n}\binom{2n}{k} z^{2k}(-1)^{2n-k} dz \\ &= 2\pi i \cdot \frac{-i(-1)^n}{4^n}\binom{2n}{n}(-1)^n \\ &= \frac{2\pi}{4^n}\binom{2n}{n}, \end{aligned} \tag{4}$$

using the Cauchy Residue Theorem in the penultimate line. Plugging this into equation (3) we get

$$\begin{aligned} m_{2n} &= \frac{2\pi}{4^{n+1}}\binom{2n+2}{n+1}\frac{r^{2n}}{\pi(2n+1)} \\ &= \left(\frac{r}{2}\right)^{2n}\frac{(2n)!}{n!(n+1)!}. \end{aligned} \tag{5}$$

This finishes the lemma. For $r = 2$ we get the moments

$$m_{2n} = \frac{(2n)!}{n!(n+1)!} = \binom{2n}{n} - \binom{2n}{n-1}$$

which are known in combinatorics as the Catalan numbers. □

Like many classical proofs of the Central Limit Theorem the strategy is to exploit the 'Fourier transformation' - in this case the $\mathfrak{R}$ transform. To that end we calculate the $\mathfrak{R}$ transform of the semi-circle distribution.

Sadly it is not readily computed directly, so we employ a trick. We consider the random variable $\ell_1 + \ell_1^*$, which clearly has $\mathfrak{R}$ transform z (in view of Lemma 2.3). We then realize, using some combinatorics, that this variable has the same distribution as $\gamma_{0,2}$ and use the moment-cumulant formula.

Lemma 3.5. *The semi-circle distribution $\gamma_{0,2}$ has $\mathfrak{R}$ transform $\mathfrak{R}_{\gamma_{0,2}}(z) = z$.*

Proof. Let T denote the operator $\ell_1 + \ell_1^*$ in $B(\mathfrak{F}(\mathfrak{H}))$. Let us show that the moments of T coincide with those of $\gamma_{0,2}$ calculated in Lemma 3.4. That is, we need to calculate

$$\omega\left(T^n\right) = \left\langle \left(\ell_1 + \ell_1^*\right)^n \Omega, \Omega \right\rangle.$$

Expanding the power of T we get

$$T^n = \sum_{w \in \mathfrak{W}_{2,2n}} w\left(\ell_1, \ell_1^*\right).$$

Here $\mathfrak{W}_{2,2n}$ denotes the set of words in 2 letters of length $2n$. As in the proof of Lemma 1.10 we see that the expectation of a term in the above sum can only

be different from 0 if ℓ_1 and ℓ_1^* appear the same number of times. This proves that the odd moments are 0. Consider a word w in 2 letters of length $2n$:

$$w(\ell_1, \ell_1^*) = \ell_1^{k_1}(\ell_1^*)^{k_2} \cdots \ell_1^{k_{m-1}}(\ell_1^*)^{k_m}, \quad \sum_{i=1}^{m} k_i = 2n.$$

Recall that (1) $\ell_1^*\Omega = 0$ and that (2) $\ell_1^*\ell_1 = 1$. When we then use (2) to make cancellations in the right end of the word we see from (1) that if the rightmost operator is ℓ_1^* at any point, the expectation is 0. On the other hand, if this never happens and the number of ℓ_1s and ℓ_1^*s is the same, the entire term is 1, and therefore has expectation 1. That is, we only need to count the number of such terms. In combinatorics terms, we need to count the number of non-negative walks in $\mathbb{Z}$ of length $2n$ that begin and end in 0. Such a walk is called a Dyck path in the literature. Let B_n denote the total number of walks of length $2n$ beginning and ending in 0. Clearly,

$$B_n = \binom{2n}{n}.$$

It suffices to count the number $\overline{B_n}$ of the above paths that hit the negative axis at least once. The first time such a path hits -1, we get a path ending in -2 by reversing every subsequent step in the sequence. We see that one can get every path of length $2n$ starting at 0 and ending in -2 in this way, so we have proved

$$\omega\left(T^{2n}\right) = B_n - \overline{B_n} = \binom{2n}{n} - \binom{2n}{n-1}$$

as desired (see Lemma 3.4).

□

By a simple scaling argument $\frac{r}{2}T$ has distribution $\gamma_{0,r}$.

3.1 The free Central Limit Theorem

The statement of the theorem is the following:

Theorem 3.6 (Central limit theorem)**.** *Let (A, ϕ) be a non-commutative probability space, and let $(a_n)_{n\in\mathbb{N}}$ be a sequence of random variables in A Assume that:*

1. *For all $n \in \mathbb{N}$, $\phi(a_n) = 0$.*
2. *There exist $r > 0$ such that*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \phi(a_n^2) = \frac{r^2}{4}.$$

3. *For all $k \geq 2$ we have*

$$\sup_{n\in\mathbb{N}} |\phi(a_n^k)| < \infty.$$

Put

$$s_N = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} a_n.$$

Then s_N converges to $\gamma_{0,r}$ in distribution.

We rely on a few relatively simple lemmas in addition to Corollary 2.4:

Lemma 3.7. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of distributions with $\mathfrak{R}$ transforms*

$$\mathfrak{R}_{\mu_n}(z) = \sum_{p=0}^{\infty} r_{n,p+1} z^p.$$

Then if μ is a distribution with $\mathfrak{R}$ transform $\mathfrak{R}_\mu(z) = \sum_{p=0}^{\infty} r_{p+1} z^p$, we have that μ_n converges to μ if and only if

$$r_{n,p+1} \to r_{p+1} \quad \text{for} \quad n \to \infty$$

for all $p \in \mathbb{N}$.

Proof. This is an immediate consequence of Proposition 1.9. □

We need to investigate how the $\mathfrak{R}$-transform interacts with scaling:

Lemma 3.8. *Let a be a non-commutative random variable. Then for any $r > 0$ we have*

$$\mathfrak{R}_{ra}(z) = r\mathfrak{R}_a(rz).$$

Proof. Let

$$\mathfrak{R}_a(z) = \sum_{k=0}^{\infty} r_{k+1} z^k$$

Now, $\ell_1 + \sum_{k=0}^{N} r_{k+1}(\ell_1^*)^k$ shares the first N free cumulants with a, and hence shares the first N moments, as well (by the moment-cumulant formula). Therefore, $r\ell_1 + \sum_{k=0}^{N} r \cdot r_{k+1}(\ell_1^*)^k$ has the same first N moments as ra. Now we claim that for any sequence (α_k) of complex numbers, the distribution of

$$r\ell_1 + \sum_{k=0}^{N} \beta_{k+1}(\ell_1^*)^k$$

equals that of

$$\ell_1 + \sum_{k=0}^{N} \beta_{k+1}(r\ell_1^*)^k.$$

This follows from the fact that a monomial in ℓ_1 and ℓ_1^* can only have non-zero expectation if ℓ_1 and ℓ_1^* appear an equal number of times just as in Lemma 3.5.[4] We use this to see that

$$r\ell_1 + \sum_{k=0}^{N} r \cdot r_{k+1}(\ell_1^*)^k$$

[4]Recall that in non-commutative probability the distribution is just the collection of moments, which for the above random variables may be expanded into monomials in ℓ_1 and ℓ_1^*.

has the same distribution as

$$\ell_1 + \sum_{k=0}^{N} r \cdot r_{k+1}(r\ell_1^*)^k.$$

This finally implies that $\mathfrak{R}_{ra}(z)$ and $r\mathfrak{R}_a(rz)$ agree in the first N terms (again by the moment-cumulant formula). Since N was arbitrary this proves the lemma. □

3.2 Proof of Theorem 3.6

We want to prove that the distribution

$$\mu_{s_N} = \mu_{N^{-1/2}(a_1+\cdots+a_N)}$$

converges to $\gamma_{0,r}$. By Lemma 3.8 and Corollary 2.4 we get

$$\mathfrak{R}_{s_N}(z) = \frac{1}{\sqrt{N}}\mathfrak{R}_{a_1+\cdots+a_N}\left(\frac{1}{\sqrt{N}}z\right) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\mathfrak{R}_{a_n}\left(\frac{1}{\sqrt{N}}z\right)$$

Let

$$\mathfrak{R}_{a_n}(z) = \sum_{k=0}^{\infty} r_{k+1}^{(n)} z^k$$

denote the $\mathfrak{R}$ transform of the individual random variables. Then we have

$$\begin{aligned}\mathfrak{R}_{s_N}(z) &= \sum_{k=0}^{\infty}\left(N^{-\frac{k+1}{2}}\sum_{n=1}^{N} r_{k+1}^{(n)}\right) z^k \\ &= \sum_{k=0}^{\infty}\beta_{k+1}^{(N)} z^k\end{aligned}$$

where

$$\beta_k^{(N)} = N^{-\frac{k}{2}}\sum_{n=0}^{N} r_k^{(n)}.$$

By Lemma 3.7 and Lemma 3.5 we have to prove that

$$\beta_2^{(N)} \to \frac{r^2}{4} \qquad \text{and} \qquad \beta_k^{(N)} \to 0 \text{ for } k \neq 2$$

as $N \to \infty$. Since $m_1 = r_1$ for all random variables, we immediately have $\beta_1^{(N)} = 0$ for all N, by assumption. Similarly, since $r_2^{(n)} = \phi(a_n^2) - \phi(a_n)^2 = \phi(a_n^2)$ we get

$$\beta_2^{(N)} = \frac{1}{N}\sum_{n=1}^{N}\phi(a_n^2) \to \frac{r^2}{4}$$

by assumption. Expressing the kth cumulant as a polynomial $Q_k(m_1, \ldots, m_k)$ (this can be done as we saw in the previous chapter) we get from condition *3.* that the kth cumulant of a_n is uniformly bounded in n. This gives

$$|\beta_k^{(N)}| \leq N^{-\frac{k}{2}} NC \to 0 \text{ for } N \to \infty$$

if $k > 2$. Now Lemma 3.7 and Lemma 3.5 together prove the theorem.

4 Random matrices

4.1 Introduction and preliminaries

Interest in random matrices (that is, matrices with random variables as entries), and in particular their eigenvalues originally stems from various areas of applied mathematics. In the beginning of the 20th century Wishart used them to study statistical data analysis and they were later used by Wigner to describe heavy-nuclei atoms. Since then is has become clear that random matrices play an important role in pure mathematics as well. The main result of this section (this project even) is Theorem 4.4, which may be viewed as a variant of a famous result by Wigner. The present theorem is due to Dan Voiculescu who used it as a tool to study the group von Neumann algebras of free groups. Here on the other hand we view the theorem as an application of free probability to an apparently unrelated area of classical mathematics. This shows (in the authors oppinion) that free probability warrants our attention whether or not we are interested in operator algebra.

The type of results we are looking for conerns limit behaviours of certain classes of random matrices. To enable the connection to free probability, we make the following definition.

Definition 4.1. *Let $I = \cup_{j\in J} I_j$ be a partition of a set I. A sequence of families $(\{T_i(n)|i \in I_j\})_{j\in J}$ of sets of random variables in a non-commutative probability space is said to be* asymptotically free *as $n \to \infty$, if the joint distributions has a limit μ and if $(\{X_i \mid i \in I_j\})_{j\in J}$ is a free family of sets of random variables in $(\mathbb{C}\langle X_i \mid i \in I\rangle, \mu)$.*

We start with a technical result, which may not be of independent interest. In essence, we prove that in order to get asymptotic freeness of central limit type sums, one does not need the families themselves to be free. It suffices to assume freeness 'up to second order'. More precisely, one has

Theorem 4.2. *Let $(T_j)_{j\in\mathbb{N}}$ be a sequence of random variables in a non-commutative probability space (A, ϕ), with ϕ being a trace. Assume further that*

1. *For all $m \in \mathbb{N}$,*
$$\sup_{j_1,j_2,\ldots,j_m} |\phi(T_{j_1}T_{j_2}\cdots T_{j_m})| < \infty \ .$$
2. *(a) If $j_1 \neq j_q$ for all $q = 2, \ldots m$ then*
$$\phi(T_{j_1}T_{j_2}\cdots T_{j_m}) = 0.$$
(b) If $j_1 = j_2$ and $j_p = q$ for at most 2 values of p for each q, then
$$\phi(T_{j_1}T_{j_2}\cdots T_{j_m}) = \phi(T_{j_3}\cdots T_{j_m}).$$
(c) If $j_1 \neq j_2 \neq \ldots \neq j_m \neq j_1$ then
$$\phi(T_{j_1}T_{j_2}\cdots T_{j_m}) = 0.$$

Furthermore if $\beta : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ *is an injection, define*

$$A_{m,n} = n^{-1/2} \sum_{q=1}^{n} T_{\beta(m,q)}.$$

Then

(A) $(A_{m,n})_{m\in\mathbb{N}}$ *is asymptotically free as* $n \to \infty$.

(B) The distribution of $A_{m,n}$ *converges to a semi-circle distribution as* $n \to \infty$.

Note that under the assumtions (1) and (2), we can calculate expectations of the form

$$\phi(T_{j_1} \cdots T_{j_m})$$

as long as each T_{j_i} occurs at most twice, using that ϕ is a trace.

Proof. The strategy of the proof is to first prove that the conditions (1) and (2) are satisfied if $(T_j)_{j\in\mathbb{N}}$ is a free family with $\phi(T_j) = 0$, $\phi(T_j^2) = 1$ and $\sup_{j\in\mathbb{N}} |\phi(T_j^k)| < \infty$ for all $j, k \in \mathbb{N}$. Next, one proves that the limit distribution exists for a family satisfying (1) and (2) and that the value of the limit distribution can be calculated only using these conditions. Therefore, it suffices to prove the result under the assumtion that the family is free, which is then the final step of the proof.
(I): Prove that (1) and (2) hold for a free family, subject to $\phi(T_j) = 0$, $\phi(T_j^2) = 1$ and $\sup_{j\in\mathbb{N}} |\phi(T_j^k)| < \infty$ for all $j, k \in \mathbb{N}$.
(1): Since the family is free we have that $\phi(T_{j_1} \cdots T_{j_m})$ is one of finitely many polynomials in the $\phi(T_{j_i}^k)$ by proposition 1.9. The boundedness then follows from the j-independent bound on $\phi(T_j^k)$.
(2a): If j_1 appears only once, then $\{T_{j_1}\}$ and $\{T_{j_2}, \ldots, T_{j_m}\}$ are free families. Hence

$$\phi(T_{j_1} \cdots T_{j_m}) = \phi(T_{j_1})\phi(T_{j_2} \cdots T_{j_m}) = 0.$$

(2b): As above, if T_{j_1} appears only in the first two positions, then the families $\{T_{j_1}\}$ and $\{T_{j_3}, \ldots, T_{j_m}\}$ are free and

$$\phi(T_{j_1}^2 T_{j_3} \cdots T_{j_m}) = \phi(T_{j_1}^2)\phi(T_{j_3} \cdots T_{j_m}) = \phi(T_{j_3} \cdots T_{j_m}).$$

(2c): This is clear from the freeness condition. This leads us to:
(II): Prove that the limit distribution of $(A_{m,n})_{m\in\mathbb{N}}$ exists as $n \to \infty$ and that it can be calculated using only (1) and (2).
For this, we have to calculate

$$\lim_{n\to\infty} \phi(A_{m_1,n} A_{m_2,n} \cdots A_{m_r,n})$$

Seeking to keep track of the n-dependence, we write out the A's and sum in a more convenient order to give us (after applying some combinatorics)

$$A_{m_1,n} \cdots A_{m_r,n} = n^{-r/2} \sum_{I\in P^r} c_I \pi_I \qquad (*)$$

where $P = \{m_1, \ldots, m_r\} \times \{1, \ldots, n\}$, c_I is a number equal to either[5] 0 or 1 and

$$\pi_I = T_{\beta(a_1,b_1)} T_{\beta(a_2,b_2)} \cdots T_{\beta(a_r,b_r)} \qquad \text{for } I = ((a_1,b_1), \ldots, (a_r,b_r)) \in P^r.$$

We now claim that

$$\lim_{n\to\infty} \phi(A_{m_1,n} \cdots A_{m_r,n}) = \lim_{n\to\infty} n^{-r/2} \sum_{I \in Q_r} c_I \pi_I \qquad (**)$$

where

$$Q_r = \left\{ I = ((a_1,b_1), \ldots (a_r,b_r)) \in P^r \;\middle|\; \begin{array}{l} \text{each pair } (a_i,b_i) \text{ appears twice} \\ \text{or not at all} \end{array} \right\}.$$

Recalling that any terms $I = (p_1, \ldots, p_r)$ in $(*)$ where some p_i appears only once has $\phi(\pi_I) = 0$ by condition (2a), we need to prove that the limit in $(**)$ exists and that

$$\lim_{n\to\infty} n^{-r/2} \phi\left(\sum_{I \in R_r} c_I \pi_I \right) = 0$$

where

$$R_r = \left\{ I = (p_1, \ldots, p_r) \in P^r \;\middle|\; \begin{array}{l} \text{for all } i,\ p_i \text{ appears at least twice, and one} \\ p_i \text{ appears at least thrice} \end{array} \right\}.$$

Because of (1), the $\phi(\pi_I)$ are uniformly bounded, so it suffices to prove that $|R_r| = o(n^{r/2})$ as $n \to \infty$. This is a counting argument: If $I = (p_1, \ldots, p_r) \in R_r$ then $|\{p_1, \ldots, p_r\}| = s < r/2$. Hence, in order to get an element in R_r, one first has to choose $s < r/2$ of the $r \times n$ pairs in P^r, and then fill in the r positions with those s pairs, with the constraint that each pair appear at least twice and one appears at least three times. Omitting the last condition, we obtain the estimate

$$|R_r| \leq \sum_{1 \leq s < r/2} \binom{nr}{s} s^r < \sum_{1 \leq s < r/2} (nr)^s s^r,$$

and using that $s \leq r/2 - 1/2$ we get

$$|R_r| \leq C_r n^{r/2 - 1/2}$$

for some constant not depending on n. This proves the second claim. In order to prove the first claim, we are allowed to assume the stronger conditions from (I) (see the remark just before the proof). But then the claim is almost trivial, since a free family is certainly asymptotically free, and the limit distribution of each $A_{m,n}$ is the semi-circle distribution by the Central Limit Theorem. □

We now indtroduce the class of random matrices we will be working with (many other classes are studied in the literature)

Definition 4.3. *Let $(\Omega, \mathcal{S}, P)$ be a classical probability space large enough to support a sequence of independent Gaussian random variables defined on it, and let* $\mathrm{SGRM}(n, \sigma^2)$ *denote the set of random matrices $[a_{ij}]_{1 \leq i,j \leq n}$ with a_{ij} such that*

[5] In fact, $c_I \neq 0$ only for those $I = ((a_1,b_1), \ldots, (a_r,b_r)) \in P^r$ such that $a_1 = m_1, \ldots, a_r = m_r$.

(i) *The entries a_{ij}, $1 \leq i \leq j \leq n$ form a set of $\frac{1}{2}n(n+1)$ independent random variables.*

(ii) $a_{ii} \sim N(0, \sigma^2)$.

(iii) $\mathrm{Re}(a_{ij}), \mathrm{Im}(a_{ij}) \sim N\left(0, \frac{\sigma^2}{2}\right)$ *and* $\mathrm{Re}(a_{ij})$ *and* $\mathrm{Im}(a_{ij})$ *are independent.*

(iv) $a_{ij} = \overline{a_{ji}}$.

Further, let L denote the unital algebra of complex measurable functions on $(\Omega, \mathcal{S})$ with moments of any order with respect to P. Finally, define the non-commutative probability space $(L \otimes M_n, \phi_n)$ where $\phi_n = \mathbb{E} \otimes \mathrm{tr}_n$ with tr_n being the normalized trace on M_n. Note that ϕ_n is a trace, and that the algebra generated by $\mathrm{SGRM}(n, \sigma^2)$ is contained in $L \otimes M_n$. We say that elements A, B of $L \otimes M_n$ are independent *if each element of A is independent of each element of B.*

The above space is often called the GUE (for Gaussian unitary ensemble) in the applications.

4.2 Wigner's Law

We now state and prove the main result of the section (cf. [1] Theorem 4.1.2).

Theorem 4.4. *Let $(Y_s(n))_{s \in \mathbb{N}} = ([a_{ij}]_{1 \leq i,j \leq n})$ be independent elements of $\mathrm{SGRM}\left(n, \frac{1}{n}\right)$. Then $(Y_s(n))_{s \in \mathbb{N}}$ is asymptotically free as $n \to \infty$ and each $Y_s(n)$ converges in distribution to $\gamma_{0,2}$.*

The strategy of the proof is as follows:
(I) Prove that

$$\sup_{n, s_1, \ldots, s_m} |\phi_n(Y_{s_1}(n) \cdots Y_{s_m}(n))| < \infty$$

This is the main part of the proof in terms of difficulty.

(II) Let $\omega \in \beta\mathbb{N} \backslash \mathbb{N}$ be a free ultrafilter, and let $\mu_\omega(T) = \lim_{n \to \omega} \mu_n(T)$ for each $T \in \mathbb{C}\langle X_i \mid i \in \mathbb{N} \rangle$ with μ_n the distribution of $(Y_s(n))_{s \in \mathbb{N}}$ (note that this limit exists in view of (I)). Now prove that $(X_i)_{i \in \mathbb{N}} \subseteq \mathbb{C}\langle X_i \mid i \in \mathbb{N} \rangle$ satisfies the conditions of Theorem 4.2 with respect to μ_ω.

(III) Let $\beta : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be an injection and define the random variables

$$B_{m,n,N} = N^{-1/2} \sum_{q=1}^{N} Y_{\beta(q,m)}(n).$$

Let $\nu_{n,N}$ denote the distribution of $(B_{m,n,N})_{m \in \mathbb{N}}$ and $\nu_{\omega,N} = \lim_{n \to \omega} \nu_{n,N}$ (pointwise). Then, by standard properties of Gaussian random variables, we get $\nu_{n,N} = \mu_n$ for each $n, N \in \mathbb{N}$. By (I) and (II) we know that $(X_i)_{i \in \mathbb{N}}$ satisfies the conditions of Theorem 4.2 with respect to μ_ω and note that the distribution of the corresponding $A_{m,N}$ from Theorem 4.2 is exactly $\nu_{\omega,N}$. Hence $(A_{m,N})_{m \in \mathbb{N}}$ is asymptotically free and semi-circularly distributed as N tends to ∞. That is, the respective distributions satisfy

$$*_{i \in \mathbb{N}} \gamma_{0,2}(T) = \lim_{N \to \infty} \nu_{\omega,N}(T) = \lim_{N \to \infty} \mu_\omega(T) = \mu_\omega(T)$$

where $*_{i\in\mathbb{N}}\gamma_{0,2}$ is the distribution of a family $(X_i)_{i\in\mathbb{N}}$ of free variables each with distribution $\gamma_{0,2}$. But this is independent of the free ultrafilter, which at last yields the desired result.

Before we start, we recall elementary facts concerning the Gauss distribution (see [5], Example 21.1).

Lemma 4.5. *The following holds:*

1. *If $X \sim N\left(0, \frac{1}{n}\right)$. Then*
   (a) $\mathbb{E}(X^{2p-1}) = 0$ *for* $p \in \mathbb{N}$.
   (b) $\mathbb{E}(X^{2p}) = (2p-1)!!n^{-p}$ *for* $p \in \mathbb{N}$
2. *If $X = Y + iZ$ with Y, Z independent and distributed as $N\left(0, \frac{1}{2n}\right)$ then*
   (a) $\mathbb{E}(X^p\overline{X^q}) = 0$ *if* $p \neq q$.
   (b) $\mathbb{E}(|X|^{2p}) = p!n^{-p}$ *for* $p \in \mathbb{N}$.

We are now ready to prove 'Wigner's law':

4.3 Proof of Theorem 4.4

(I) We want to find a constant $C(m)$ such that

$$|\phi_n(Y_{s_1}(n)\dots Y_{s_m}(n))| \leq C(m) < \infty$$

From the definition of ϕ_n we get (with arithmetics in the index set modulo m)

$$\begin{aligned}\phi_n(Y_{s_1}(n)\dots Y_{s_m}(n)) &= \frac{1}{n}\sum_{\substack{1\leq i_1,\dots,i_m\leq n\\ j_q=i_{q+1}}} \mathbb{E}\left(a_{i_1j_1}(s_1,n)\dots a_{i_mj_m}(s_m,n)\right)\\ &= \frac{1}{n}\sum_{1\leq i_1,\dots i_m\leq n} \mathbb{E}\left(a_{i_1i_2}(s_1,n)\dots a_{i_mi_1}(s_m,n)\right) \qquad (\Delta)\end{aligned}$$

Now observe that if a term in (Δ) is non-zero, then by Lemma 4.5 we have

(i) Each $a_{ii}(s,n)$ occurs an even number of times.

(ii) If $a_{ij}(s,n)$ occurs q times, then $a_{ji}(s,n)$ occurs exactly q times, as well.

Particularly, m must be even. Now (i) and (ii) imply the existence of a permutation $\gamma : \{1,\dots,m\} \to \{1,\dots,m\}$ such that for all $q \in \{1,\dots,m\}$:

$$\gamma^2 = \gamma\circ\gamma = \mathrm{id}, \gamma(q) \neq q,\ (i_q, j_q) = (j_{\gamma(q)}, i_{\gamma(q)}),\ s_q = s_{\gamma(q)}$$

just by ordering the groupings from (i) and (ii) into pairs and letting γ be the permutation that switches the indices of the pairs. Grouping the terms in (Δ) and using Hölder's inequality we get

$$\begin{aligned}\mathbb{E}\left(a_{i_1i_2}(s_1,n)\dots a_{i_mi_1}(s_m,n)\right) &= \mathbb{E}\prod_{q<\gamma(q)} |a_{i_qi_{q+1}}(s_q,n)|^2\\ &\leq \left(\prod_{q<\gamma(q)} \mathbb{E}|a_{i_qi_{q+1}}(s_q,n)|^m\right)^{2/m}.\end{aligned}$$

From Lemma 4.5 we get that

$$\mathbb{E}(|a_{i_q i_{q+1}}(s_q,n)|^m) = \left(\frac{m}{2}\right)! n^{-m/2}, \qquad i_q = i_{q+1}$$
$$\mathbb{E}(|a_{i_q i_{q+1}}(s_q,n)|^m) = (m-1)!! n^{-m/2}, \qquad i_q \neq i_{q+1}$$

In any case, we get that

$$\mathbb{E}(a_{i_1 i_2}(s_1,n)\dots a_{i_m i_1}(s_m,n)) \leq \left(\prod_{q<\gamma(q)} n^{-m/2}(m-1)!!\right)^{2/m} = (m-1)!! n^{-m/2}.$$

Inserting in (Δ) we get

$$|\phi_n(Y_{s_1}(n)\dots Y_{s_m}(n))| \leq n^{-(m/2+1)}(m-1)!! \sum_{\gamma\in\Gamma_m} \theta_n(\gamma) \qquad (\Delta\Delta)$$

with[6]

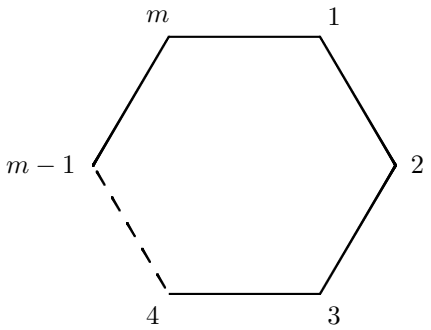
$$\Gamma_m = \{\gamma \in S_m \mid \gamma^2 = \mathrm{id},\ \gamma(q) \neq q,\ s_{\gamma(q)} = s_q\}$$
$$\theta_n(\gamma) = |\{(i_1,\dots,i_m) \in \{1,\dots n\}^m \mid i_{\gamma(q)} = i_{q+1},\ i_q = i_{\gamma(q)+1}\}|$$

Now let $d(\gamma)$ denote the number of equivalence classes of $\{1,\dots,m\}$ under the relation generated by

$$\gamma(q) \sim q+1, \qquad q \sim \gamma(q)+1$$

(using that $\gamma^2 = \mathrm{id}$ one sees that the two relations are actually the same). Since in the definition of θ_n one only has to choose one i_q for each equivalence class, we see that $\theta_n(\gamma) = n^{d(\gamma)}$. Hence we need to prove that $d(\gamma) \leq \frac{m}{2}+1$. This is done using graph theory.

Consider the polygonal graph $\mathcal{G}$ with m vertices and m edges (see figure)



Now obtain another graph $\mathcal{G}'$ from $\mathcal{G}$ as follows: The vertices $e_1,\dots,e_{d(\gamma)}$ are the equivalence classes of $\{1,\dots,m\}$ under $\sim$ and the edges are such that e_i and e_j are connected by an edge iff there is $u \in e_i,\ v \in e_j$ such that u and v are connected by an edge in $\mathcal{G}$ - i.e., $\mathcal{G}'$ is the quotient graph with the appropriate identifications. Note that $\mathcal{G}'$ is still connected, since one may map any path in

[6] Note that although Γ_m depends on $s_1,\dots,s_m$, the cardinality of Γ_m is bounded (e.g., by $m!$), independently of $s_1,\dots s_m$.

$\mathcal{G}$ to the quotient graph. We want to control the number of edges in $\mathcal{G}'$: Hence, assume there is an edge between e_i and e_j in $\mathcal{G}'$, i.e., there exist $u \in e_i,\ v \in e_j$ such that $u = v+1$ or $v = u+1 \pmod m$. Assume $u = v+1 \pmod m$.
Now note that

$$\gamma(v)+1 \sim v \in e_j \quad \text{and} \quad \gamma(v) \sim v+1 = u \in e_i.$$

Hence both of the edges $[v+1, v]$ and $[\gamma(v), \gamma(v)+1]$ in $\mathcal{G}$ represent $[e_i, e_j]$. But if $m > 2$ (the case $m = 2$ is easy) the two edges in $\mathcal{G}$ are distinct since otherwise

$$v+1 = \gamma(v) \text{ and } v = \gamma(v)+1, \quad \text{i.e.} \quad v+1 = v-1 \pmod m$$

which is a contradiction. Therefore each edge in $\mathcal{G}'$ is represented by at least two distinct edges in $\mathcal{G}$, and hence the number of edges in $\mathcal{G}'$ is at most $m/2$. Hence, since the Euler characteristic of any connected graph is at most 1 we get

$$d(\gamma) - \tfrac{m}{2} \le \chi(\mathcal{G}') = |\{\text{vertices in } \mathcal{G}'\}| - |\{\text{edges in } \mathcal{G}'\}| \le 1$$

which gives the desired inequality $d(\gamma) \le m/2 + 1$.
Inserting this in $(\Delta\Delta)$ we get

$$|\phi_n(Y_{s_1}(n) \dots Y_{s_m}(n))| \le (m-1)!!|\Gamma_m| \le (m-1)!!m!$$

which finally proves (I).

(II) Since (I) proves 1 of Theorem 4.2 we need to prove 2 (a),(b),(c) for $(X_i)_{i\in\mathbb{N}}$ with respect to μ_ω.
2 (a): If $s_1 \neq s_q$ for all $q = 2, \dots m$ then using (Δ) we get

$$\begin{aligned}
\mu_\omega(X_{s_1} \dots X_{s_m}) &= \lim_{n\to\omega} \mu_n(X_{s_1} \dots X_{s_m}) \\
&= \lim_{n\to\omega} \phi_n(Y_{s_1}(n) \dots Y_{s_m}(n)) \\
&= \lim_{n\to\omega} \frac{1}{n} \sum_{1\le 1_1,\dots,i_m\le n} \mathbb{E}(a_{i_1 i_2}(s_1,n) \dots a_{i_m i_1}(s_m,n)) \\
&= \lim_{n\to\omega} \frac{1}{n} \sum_{1\le i_1,\dots,i_m\le n} \mathbb{E}(a_{i_1 i_2}(s_1,n))\mathbb{E}(a_{i_2 i_3}(s_2,n) \dots a_{i_m i_1}(s_m,n)) \\
&= 0
\end{aligned}$$

as desired.

2 (b): Assume $s_1 = s_2$ and $j_1 \neq j_p$ for $p > 2$ and let $Z = Y_{s_3}(n) \dots Y_{s_m}(n) = [b_{jk}]_{1\le j,k\le n}$. Then as above we have

$$\begin{aligned}
\mu_\omega(X_{s_1} \cdots X_{s_m}) &= \lim_{n\to\omega} \phi_n\left(\left[\sum_{q=1}^n \sum_{i=1}^n a_{ji}(s_1,n) a_{iq}(s_1,n) b_{qk} \right]_{1\le j,k\le n} \right) \\
&= \lim_{n\to\omega} \frac{1}{n} \sum_{j=1}^n \mathbb{E}\left(\sum_{q=1}^n \sum_{i=1}^n a_{ji}(s_1,n) a_{iq}(s_1,n) b_{qj} \right) \\
&= \lim_{n\to\omega} \frac{1}{n} \sum_{j=1}^n \sum_{q=1}^n \sum_{i=1}^n \mathbb{E}(a_{ji}(s_1,n) a_{iq}(s_1,n))\mathbb{E}(b_{qj}) \\
&= \lim_{n\to\omega} \frac{1}{n} \sum_{j=1}^n \mathbb{E}(b_{jj}) = \lim_{n\to\omega} \phi_n(Z) = \mu_\omega(X_{s_3} \cdots X_{s_m})
\end{aligned}$$

using that $\mathbb{E}(a_{ji}(s_n)a_{iq}(s,n)) = \frac{1}{n}\delta_{j,q}$.

2 (c): If $s_1 \neq s_2 \neq \ldots \neq s_m \neq s_1$, then by $(\Delta\Delta)$ we have

$$|\phi_n(Y_{s_1} \ldots Y_{s_m})| \leq n^{-(m/2+1)}(m-1)!! \sum_{\gamma \in \Gamma_m} n^{d(\gamma)}$$

We will prove $d(\gamma) \leq m/2$ under the above assumptions. To that end assume that for some $\gamma \in \Gamma_m$ some equivalence class $[q]_\sim$ contains only a single element. In other words $q = \gamma(q) + 1 \pmod m$, i.e. $s_{q-1} = s_{\gamma(q)} = s_q$ (indeices still calculated mod m) by the definition of Γ_m. But this contradicts the asssumtions. Hence each equivalence class contains at least two elements, whence $d(\gamma) \leq m/2$ which by $(\Delta\Delta)$ implies that $\phi_n(Y_{s_1}(n) \ldots Y_{s_m}(n)) \to 0$ as $n \to \infty$.

(III) This argument was given briefly in the strategy section, but is recounted here in detail for convenience:
Let $\beta : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be an injection and define the following random variables:

$$B_{m,n,N} = N^{-1/2} \sum_{q=1}^{N} Y_{\beta(q,m)}.$$

By the convolution properties of Gaussian distributions we have that $B_{m,n,N} \in \mathrm{SGRM}\left(n, \frac{1}{n}\right)$ and clearly $(B_{m,n,N})_{m\in\mathbb{N}}$ are independent. Let $\nu_{n,N}$ denote the distribution of $(B_{m,n,N})_{m\in\mathbb{N}}$ and put $\nu_\omega(T) = \lim_{n\to\omega} \mu_n(T)$ for each $T \in \mathbb{C}\langle X_i \mid i \in \mathbb{N}\rangle$. The above observation then implies that $\mu_n = \nu_{n,N}$ for each $n, N \in \mathbb{N}$ and therefore $\mu_\omega = \nu_{\omega,N}$ as well. As in Theorem 4.2 define the random variables in $(\mathbb{C}(X_i | i \in \mathbb{N}, \mu_\omega))$

$$A_{m,N} = N^{-1/2} \sum_{q=1}^{N} X_{\beta(q,m)}.$$

Note that the distribution of $A_{m,N}$ is exactly $\nu_{\omega,N}$. The conclusion of Theorem 4.2 is then that $\lim_{N\to\infty} \nu_{\omega,N}(T) = (*_{i\in\mathbb{N}}\gamma_{0,2})(T)$, where

$$(*_{i\in\mathbb{N}}\gamma_{0,2}) : \mathbb{C}\langle X_i \mid i \in \mathbb{N}\rangle \to \mathbb{C}$$

is the distribution of a free family $(X_i)_{i\in\mathbb{N}}$ each with distribution $\gamma_{0,2}$. Hence we have, for each $T \in \mathbb{C}\langle X_i \mid i \in \mathbb{N}\rangle$,

$$(*_{i\in\mathbb{N}}\gamma_{0,2})(T) = \lim_{N\to\infty} \nu_{\omega,N}(T) = \lim_{N\to\infty} \mu_\omega(T) = \mu_\omega(T).$$

This yields the following conclusions:

- The limit distribution μ_ω is independent of $\omega \in \beta\mathbb{N}\backslash\mathbb{N}$, and therefore $(Y_s(n))_{s\in\mathbb{N}}$ actually has an asymptotic distribution μ.
- The equation above actually proves that $\mu = (*_{i\in\mathbb{N}}\gamma_{0,2})$. That is, $(Y_s(n))_{n\in\mathbb{N}}$ is asymptoticaly free as $n \to \infty$, and the limit distribution of each $Y_s(n)$ is a semi-circle distribution.

This concludes the proof.

A Formal power series

In this section we develop the theory of formal power series and prove the Lagrange inversion formula.
As a set, the ring of formal power series $\mathbb{C}[[X]]$ is $\mathbb{C}^{\mathbb{N}_0}$ - the set of complex sequences. The ring structure is

$$(a_n)_{n\in\mathbb{N}_0} + (b_n)_{n\in\mathbb{N}_0} = (a_n + b_n)_{n\in\mathbb{N}_0}$$
$$(a_n)_{n\in\mathbb{N}_0} \times (b_n)_{n\in\mathbb{N}_0} = \left(\sum_{k=0}^{n} a_k b_{n-k}\right)_{n\in\mathbb{N}_0}.$$

This makes $\mathbb{C}[[X]]$ a ring with unit $1 = (1, 0, \ldots)$ and $0 = (0, 0, \ldots)$. We use the formal expression

$$(a_n)_{n\in\mathbb{N}_0} = \sum_{n=0}^{\infty} a_n X^n.$$

We now note a number of elementary facts about algebra in $\mathbb{C}[[X]]$.

- If n is a natural number and $f(X) = \sum_{k=0}^{\infty} a_k X^k$, then

  $$f(X)^n = \sum_{k=0}^{\infty} c_k X^k.$$

  Here

  $$c_0 = a_0^n, \qquad c_m = \frac{1}{m a_0} \sum_{k=1}^{m} (kn - m + k) a_k c_{m-k}$$

  if $a_0 \neq 0$[7]. Note that the first m coefficients of $f(X)^n$ are determined from the first m coefficients of $f(X)$.

- If $a_0 \neq 0$ then $f(X) = \sum_{k=0}^{\infty} a_k X^k$ is invertible in $\mathbb{C}[[X]]$ with

  $$f(X)^{-1} = \sum_{k=0}^{\infty} b_k X^k.$$

  Here

  $$b_0 = \frac{1}{a_0}, \qquad b_n = -\frac{1}{a_0} \sum_{k=1}^{n} a_k b_{n-k} \quad n \geq 1.$$

  As above, the first m coefficients of $f(X)^{-1}$ are determined by the first m coefficients of $f(X)$. A combination of the two previous points prove that if $a_0 = 0$ we get that

  $$(1 + f(X))^p = \sum_{k=0}^{\infty} \binom{p}{k} f(X)^k \tag{6}$$

  for any k (positive or negative)[8].

[7] If this is not the case write $f(X) = X^m \sum_{k=0}^{\infty} a'_k X^k$ for some sequence a'_k with $a'_0 \neq 0$.
[8] If n is positive we have $\binom{-n}{k} = (-1)^k \binom{n+k-1}{k}$.

- Let two formal power series be given:

$$f(X) = \sum_{k=0}^{\infty} a_k X^k, \qquad g(X) = \sum_{k=0}^{\infty} b_k X^k.$$

If $a_0 = 0$ we define the *composition* $g \circ f$ of g and f by[9]

$$g \circ f(X) = \sum_{k=0}^{\infty} b_k f(X)^k = \sum_{n=0}^{\infty} c_n X^n$$

Here the coefficients c_n can be found by expanding the powers of f:

$$c_n = \sum_{\pi=\{B_1,\dots B_k} a_{|B_1|} \cdots a_{|B_k|} b_k.$$

Here π runs through the set of all partitions of $\{1, \dots, n\}$ and $B_1, \dots, B_n$ denotes the blocks of the partitions. Lastly, $|B_k|$ denotes the number of elements in the kth block. This formula is sometimes called the Faá di Bruno formula. If

$$f(X) = \sum_{k=1}^{\infty} a_k X^k \quad g(X) = \sum_{k=0}^{\infty} b_k X^k,$$

we have

$$g \circ f(X) = b_0 + b_1 a_1 X + (b_1 a_2 + b_2 a_1^2) X^2 + (b_1 a_3 + 2 b_2 a_1 a_2 + b_3 a_1^3) X^3 + \cdots$$

As in the above examples, we have that the first m coefficients of $g \circ f$ are determined by the first m coefficients of f and g.

- If $f(X) = \sum_{k=0}^{\infty} a_k X^k \in \mathbb{C}[[X]]$ with $a_0 = 0$ and $a_1 \neq 0$ then there exist a compositional inverse $f^{(-1)}(X) \in \mathbb{C}[[X]]$[10]. The coefficients can be found recursively from the above formula. One can also use the celebrated Lagrange inversion formula proved below.

We introduce a few extra concepts in order to get to the Lagrange inversion formula.
As a generalization of $\mathbb{C}[[X]]$, we define the ring of formal Laurent series $\mathbb{C}((X))$ as the set of formal sums

$$f(X) = \sum_{k=-\infty}^{\infty} a_k X^k$$

with $a_k \neq 0$ for only finitely many negative $k \in \mathbb{Z}$. The product is now

$$\left(\sum_{k \in \mathbb{Z}} a_k X^k \right) \times \left(\sum_{k \in \mathbb{Z}} b_k X^k \right) = \sum_{k \in \mathbb{N}} \left(\sum_{n \in \mathbb{Z}} a_n b_{k-n} \right) X^k.$$

[9] It may seem strange that we require $a_0 = 0$. We assume this to completely avoid analytic arguments. If we do not assume this we need to add infinitely many terms to get the constant term of the composition, which can not be done for a general formal power series.
[10] Of course the compositional neutral element is $f(X) = X$.

Here the sum over n is finite since a_i and b_i are only non-zero for finitely many negative i. The sum in $\mathbb{C}((X))$ is the obvious extension of the sum in $\mathbb{C}[[X]]$, and the composition is the same, as well. Lastly, define the order of $f(X) \in \mathbb{C}((X))$ to be $\mathrm{ord}(f) = \min\{k \in \mathbb{Z} \mid a_k \neq 0\}$

If $f(X) = \sum_{k\in\mathbb{Z}} a_k X^k \in \mathbb{C}((X)))$ we define the extraction functionals $[X^m]$ for each $m \in \mathbb{Z}$ by

$$[X^m]f(X) = a_m$$

In particular we denote for $m = -1$

$$[X^{-1}]f(X) = \mathrm{Res}\,(f(X))$$

Define the formal differentiation map $D : \mathbb{C}((X))) \to \mathbb{C}((X)))$ by

$$D\left(\sum_{k\in\mathbb{Z}} a_k X^k\right) = \sum_{k\in\mathbb{Z}} k a_k X^{k-1}.$$

We also denote $Df = f'$. Note that is f is a non-constant Laurent series then $\mathrm{ord}(f') = \mathrm{ord}(f) - 1$. It is easy to see that D is a derivation (in particular linear) and that it satisfies

$$\mathrm{ker}D = \mathbb{C}, \quad \mathrm{im}D = \left\{f \in \mathbb{C}((X)) \,|[X^{-1}]f = 0\right\} = \mathrm{ker}\mathrm{Res}.$$

By the above, we have an exact sequence

$$0 \to \mathbb{C} \to \mathbb{C}((X)) \overset{D}{\to} \mathbb{C}((X)) \overset{\mathrm{Res}}{\to} \mathbb{C} \to 0.$$

We prove a lemma describing some relations between the maps D and Res.

Lemma A.1. *For $f(X) \in \mathbb{C}((X))$ we have*

1. $\mathrm{Res}(f') = 0$.
2. $\mathrm{Res}(fg') = -\mathrm{Res}(f'g)$.
3. $\mathrm{Res}\,(f'/f) = \mathrm{ord}(f)$ *as longs as* $f \neq 0$.
4. $\mathrm{Res}\,((f \circ g)g') = \mathrm{ord}(f)\mathrm{Res}(g)$, *if* $\mathrm{ord}(g) > 0$.
5. $[X^n]f(X) = \mathrm{Res}(X^{-n-1}f(X))$

Proof. Property *1.* is a part of the exactness of the above sequence. Part *2.* follows from *1.* and the fact that D is a derivation. To see *3.* note that any $f \in \mathbb{C}((X))$ can be written

$$f(X) = X^m g(X)$$

where $m = \mathrm{ord}(f)$ and $\mathrm{ord}(g) = 0$. Hence using that D is a derivation we get

$$f'/f = mX^{-1} + g'/g.$$

As $\mathrm{ord}(g) = 0$ we have that both g, g' and g^{-1} are elements of $\mathbb{C}[[X]]$. Since $m = \mathrm{ord}(f)$ *3.* follows.
For the proof of *4.* note that since $\mathrm{im}D = \ker \mathrm{Res}$ we can write

$$f(X) = f_{-1}X^{-1} + F'(X)$$

for some $F(X) \in \mathbb{C}((X))$. Consequently we have

$$(f \circ g) = f_{-1}g^{-1}g' + (F' \circ g)g' = f_{-1}g^{-1}g' + (F' \circ g)'.$$

From *1.* and *3.* we then get *4.*
Property *5.* follows directly from the definition. □

We are now ready to formulate and prove the Lagrange inversion formula:

Theorem A.2 (The Lagrange inversion formula)**.** *For $f(X) \in \mathbb{C}[[X]]$ with $[X^0]f(X) = 0$ and $[X^1]f(X) \neq 0$, the compostional inverse $f^{(-1)}(X) \in \mathbb{C}[[X]]$ satisfies for all $k, n \in \mathbb{Z}$*

$$k[X^k]g^n = n[X^{-n}]f^{-k} \qquad (\textit{multiplicative powers})$$

In particular, for $n = 1$ and $k \geq 0$

$$[X^k]g = \frac{1}{k}\mathrm{Res}\left(f^{-k}\right)$$

Proof. Using Lemma A.1, we simply compute:

$$\begin{aligned} k[X^k]g^n &= k\mathrm{Res}\left(X^{-k-1}g^n\right) = k\mathrm{Res}\left(X^n f^{-k-1}f'\right) = -\mathrm{Res}\left(X^n\left(f^{-k}\right)'\right) \\ &= \mathrm{Res}\left(\left(X^n\right)' f^{-k}\right) = n\mathrm{Res}\left(X^{n-1}f^{-k}\right) = n[X^{-n}]f^{-k}, \end{aligned}$$

as desired. □

We now use this theroem to prove the moment-cumulant formulas. That is, we let

$$\psi(t) = t + m_1t^2 + m_2t^3 + \ldots = t\left(1 + \sum_{p=1}^{\infty} m_p t^p\right).$$

Then we want to calculate the $p-1$ coefficient of the series[11]

$$\mathfrak{R}(z) = \psi^{(-1)}(z)^{-1} - \frac{1}{z}.$$

We calculate:

$$\begin{aligned} [z^{p-1}]\mathfrak{R}(z) = [z^{p-1}]\psi^{(-1)}(z)^{-1} &= -\frac{1}{p-1}[z^1]\psi(z)^{-(p-1)} \\ &= -\frac{1}{p-1}[z^p]\left(1 + \sum_{n=1}^{\infty} m_n z^n\right)^{-(p-1)} \\ &= -\frac{1}{p-1}[z^p]\left(\sum_{k=0}^{\infty}(-1)^k\binom{p+k-2}{k}\left(\sum_{n=1}^{\infty} m_n z^n\right)^k\right) \\ &= \sum_{k=1}^{p}\frac{(-1)^{k+1}}{k}\binom{p+k-2}{k-1}\sum_{\substack{q_1,\ldots,q_k \geq 1 \\ q_1+\cdots+q_k=p}} m_{q_1}\cdots m_{q_k} \end{aligned}$$

[11] Recall that the indexation of the $\mathfrak{R}$ transform is $\mathfrak{R}(z) = \sum_{p=0}^{\infty} r_{p+1}z^p$.

This is the desired formula. For the other formula we perform similar computations: From the definition of the $\mathfrak{R}$ transform we have

$$\psi^{(-1)}(z) = \frac{z}{z\mathfrak{R}(z)+1} = z\left(1+\sum_{n=1}^{\infty} r_n z^n\right)^{-1}$$

We now calculate:

$$\begin{aligned} m_p = [t^{p+1}]\psi(t) &= \frac{1}{p+1}[t^{-1}]\left(\psi^{(-1)}(t)\right)^{-(p+1)} \\ &= \frac{1}{p+1}[t^p]\left(1+\sum_{n=1}^{\infty} r_n t^n\right)^{p+1} \\ &= \frac{1}{p+1}[t^p]\left(\sum_{k=0}^{p+1}\binom{p+1}{k}\left(\sum_{n=1}^{\infty} r_n t^n\right)^k\right) \\ &= \sum_{k=1}^{p}\frac{1}{k}\binom{p}{k-1}\sum_{\substack{q_1,\ldots,q_k\geq 1 \\ q_1+\cdots+q_k=p}} r_{q_1}\cdots r_{q_k} \end{aligned}$$

as desired.[12]

References

[1] D.V. Voiculescu, K.J. Dykema, A. Nica, *Free Random Variables.* CRM Monograph Series, vol. 1, American Mathematical Society, Providence, RI, 1992.

[2] S. Thorbjørnsen, *Mixed moments of Voiculescu's Gaussian random matrices.* J. Funct. Anal. 176 (2000), 213-246.

[3] U. Haagerup and S. Thorbjørnsen, *Random matrices and K-theory for exact C^*-algebras*, Documenta Math. 4 (1999), 341-450.

[4] U. Haagerup, *On Voiculescu's R- and S-Transforms for free Non-Commuting Random Variables.* Fields Institute Communications, Volume 12, 1997

[5] P. Billingsley - *Probability and measure*, John Wiley & Sons Inc., New York third edition, 1995. A Wiley-Interscience Publication

[12] For both formulas we have restricted the index set in the last line so that terms that are always zero does not appear.