FREE PROBABILITY THEORY AND RANDOM MATRICES

ROLAND SPEICHER

ABSTRACT. Free probability theory originated in the context of operator algebras, however, one of the main features of that theory is its connection with random matrices. Indeed, free probability can be considered as *the* theory providing concepts and notations, without relying on random matrices, for dealing with the limit $N \to \infty$ of $N \times N$-random matrices.

One of the basic approaches to free probability, on which I will concentrate in this lecture, is of a combinatorial nature and centers around so-called free cumulants. In the spirit of the above these arise as the combinatorics (in leading order) of $N \times N$-random matrices in the limit $N = \infty$. These free cumulants are multilinear functionals which are defined in combinatorial terms by a formula involving non-crossing partitions.

I will present the basic definitions and properties of non-crossing partitions and free cumulants and outline its relations with freeness and random matrices. As examples, I will consider the problems of calculating the eigenvalue distribution of the sum of randomly rotated matrices and of the compression (upper left corner) of a randomly rotated matrix.

1. Random matrices and freeness

Free probability theory, due to Voiculescu, originated in the context of operator algebras, however, one of the main features of that theory is its connection with random matrices. Indeed, free probability can be considered as *the* theory providing concepts and notations, without relying on random matrices, for dealing with the limit $N \to \infty$ of $N \times N$-random matrices.

Let us consider a sequence $(A_N)_{N\in\mathbb{N}}$ of selfadjoint $N \times N$-random matrices A_N. In which sense can we talk about the limit of these matrices? Of course, such a limit does not exist as a $\infty \times \infty$-matrix and

Research supported by a grant of NSERC, Canada.

Lectures at the European Summer School 'Asymptotic Combinatorics with Applications to Mathematical Physics', St. Petersburg (Russia), July 2001.

there is no convergence in the usual topologies connected to operators. What converges and survives in the limit are the moments of the random matrices.

To talk about moments we need in addition to the random matrices also a state. This is given in a canonical way by the averaged trace: Let tr_N be the normalized trace on $N \times N$-matrices, i.e. for $A = (a_{ij})_{i,j=1}^N$ we have

$$\mathrm{tr}_N(A) := \frac{1}{N} \sum_{i=1}^N a_{ii}.$$

In the same way, we get the averaged trace $\mathrm{tr}_N \otimes \mathbb{E}$ for $N \times N$-random matrices, i.e. for $A = (a_{ij}(\omega))_{i,j=1}^N$ (where the entries a_{ij} are random variables on some probability space Ω equipped with a probability measure P) we have

$$\mathrm{tr}_N \otimes \mathbb{E}(A) := \frac{1}{N} \sum_{i=1}^N \int_\Omega a_{ii}(\omega) dP(\omega).$$

Given these states $\mathrm{tr}_N \otimes E$, we can now talk about the k-th moment $\mathrm{tr}_N \otimes \mathbb{E}(A_N^k)$ of our random matrix A_N, and it is well known that for nice random matrix ensembles these moments converge for $N \to \infty$. So let us denote by α_k the limit of the k-th moment,

$$\lim_{N\to\infty} \mathrm{tr}_N \otimes E(A_N^k) =: \alpha_k.$$

Thus we can say that the limit $N = \infty$ consists exactly of the collection of all these moments α_k. But instead of talking about a collection of numbers α_k we prefer to identify these numbers as moments of some variable A. Abstractly it is no problem to find such an A, we just take a free algebra $\mathcal{A}$ with generator A and define a state φ on $\mathcal{A}$ by setting

$$\varphi(A^k) := \alpha_k.$$

Of course, nothing deep has happened, this is just a shift in language, but it provides us with a more conceptual way of looking at the limit of our random matrices. Now we can say that our random matrices A_N converge to the variable A in distribution (which just means that the moments of A_N converge to the moments of A). We will denote this by $A_N \to A$. Note that the nature of the limit $N = \infty$ is quite different from the case of finite N. In the latter case the A_N live in classical probability spaces of $N \times N$-random matrices, whereas the $N = \infty$ limit object A is not of a classical nature any more, but lives in a 'non-classical probability space' given by some algebra $\mathcal{A}$ and a state φ.

1.1. **Remark.** One should note that for a selfadjoint operator $A = A^*$, the collection of moments (or, equivalently, the state φ corresponding to these moments) corresponds also to a probability measure μ_A on the real line, determined by

$$\varphi(A^k) = \int_{\mathbb{R}} t^k d\mu_A(t).$$

(We can ignore the problem of non-uniqueness of this moment problem, because usually our operators A are bounded, which ensures uniqueness.) In particular, for a selfadjoint $N \times N$-matrix $A = A^*$ this measure is given by the eigenvalue distribution of A, i.e. it puts mass $1/N$ on each of the eigenvalues of A (counted with multiplicity):

$$\mu_A = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i},$$

where $\lambda_1, \dots, \lambda_N$ are the eigenvalues of A. In the same way, for a random matrix A, μ_A is given by the averaged eigenvalue distribution of A. Thus, moments of random matrices with respect to the averaged trace $\mathrm{tr}_N \otimes \mathbb{E}$ contain exactly that type of information in which one is usually interested when dealing with random matrices.

1.2. **Example.** Let us consider the basic example of random matrix theory, expressed in our new language. Let G_N be the usual selfadjoint Gaussian $N \times N$-random matrices (i.e., entries above the diagonal are independently and normally distributed). Then the famous theorem of Wigner can be stated in our language in the form that

$$G_N \to s, \qquad \text{where } s \text{ is a semi-circular variable,}$$

where semi-circular just means that the measure μ_s is given by the semi-circular distribution (or, equivalently, the even moments of the even variable s are given by the Catalan numbers).

Up to now, nothing crucial has happened, we have just shifted a bit the usual way of looking on things. A new and crucial concept, however, appears if we go over from the case of one variable to the case of more variables. Of course, again joint moments are the surviving quantities in multi-matrix models (even if it is now not so clear any more how to prove this convergence in concrete models) and we can adapt our way of looking on things to this situation by making the following definition.

1.3. **Definition.** Consider $N \times N$-random matrices $A_N^{(1)}, \dots, A_N^{(m)}$ and variables $A_1, \dots, A_m$ (living in some abstract algebra $\mathcal{A}$ equipped with a state φ). We say that

$$(A_N^{(1)}, \dots, A_N^{(m)}) \to (A_1, \dots, A_m) \qquad \text{in distribution,}$$

if

$$\lim_{N\to\infty} \mathrm{tr}_N \otimes \mathbb{E}[A_N^{(i_1)} \cdots A_N^{(i_k)}] = \varphi(A_{i_1} \cdots A_{i_k})$$

for all choices of k, $1 \le i_1, \dots, i_k \le m$.

1.4. **Remark.** The $A_1, \dots, A_m$ arising in the limit of random matrices are a priori abstract elements in some algebra $\mathcal{A}$, but it is good to know that in many cases they can also be concretely realized by some kind of creation and annihilation operators on a full Fock space. Indeed, free probability theory was introduced by Voiculescu for investigating the structure of special operator algebras generated by these type of operators. In the beginning, free probability had nothing to do with random matrices.

1.5. **Example.** Let us now consider the example of two independent Gaussian random matrices $G_N^{(1)}, G_N^{(2)}$ (i.e., each of them is a selfadjoint Gaussian random matrix and all entries of $G_N^{(1)}$ are independent from all entries of $G_N^{(2)}$). Then one knows that all joint moments converge, and we can say that $(G_N^{(1)}, G_N^{(2)}) \to (s_1, s_2)$, where Wigner tells us that both s_1 and s_2 are semi-circular. The question is: What is the relation between s_1 and s_2? Does the independence between $G_N^{(1)}$ and $G_N^{(2)}$ survive in some form also in the limit? The answer is yes and is provided by a basic theorem of Voiculescu which says that s_1 and s_2 are **free** in the following sense.

1.6. **Definition.** Let $\mathcal{A}$ be a unital algebra and $\varphi : \mathcal{A} \to \mathbb{C}$ a linear functional on $\mathcal{A}$, which is unital, i.e., $\varphi(1) = 1$. Then $a_1, \dots, a_m \in \mathcal{A}$ are called *free* (with respect to φ) if

$$\varphi[p_1(a_{i(1)}) \cdots p_k(a_{i(k)})] = 0$$

whenever

- $p_1, \dots, p_k$ are polynomials in one variable
- $i(1) \ne i(2) \ne i(3) \ne \cdots \ne i(k)$ (only neigbouring elements are required to be distinct)
- $\varphi[p_j(a_{i(j)})] = 0$ for all $j = 1, \dots, k$

1.7. **Remark.** Note that the definition of freeness can be considered as a way of organizing the information about all joint moments of free variables in a systematic and conceptual way. Indeed, the above definition allows to calculate mixed moments of free variables in terms of moments of the single variables. For example, if a, b are free, then the definiton of freeness requires that

$$\varphi[(a - \varphi(a) \cdot 1)(b - \varphi(b) \cdot 1)] = 0,$$

which implies that

$$\varphi(ab) = \varphi(a) \cdot \varphi(b) \qquad \text{if } a, b \text{ are free.}$$

In the same way,

$$\varphi[(a - \varphi(a) \cdot 1)(b - \varphi(b) \cdot 1)(a - \varphi(a) \cdot 1)(b - \varphi(b) \cdot 1)] = 0$$

leads finally to

$$\varphi(abab) = \varphi(aa) \cdot \varphi(b) \cdot \varphi(b) + \varphi(a) \cdot \varphi(a) \cdot \varphi(bb) - \varphi(a) \cdot \varphi(b) \cdot \varphi(a) \cdot \varphi(b).$$

Analogously, all mixed moments can (at least in principle) be calculated by reducing them to alternating products of centered variables as in the definition of freeness.

Thus the statement 's_1, s_2 are free and each of them is semicircular' determines all joint moments in s_1 and s_2.

Formulating our knowledge about the joint moments of s_1 and s_2 in this peculiar way might look not very illuminating on first sight, but it will turn out that recognizing this notion of freeness as the organizing principle for the collection of moments adds a new perspective on the limit of random matrices.

In particular, we are now in the context of non-commutative probability theory which consists mainly of the doctrine that one should use notations and ideas from classical probability theory in order to understand problems about non-commutative algebras.

Free probability theory can be described as that part of non-commutative probability theory where the notion of 'freeness' plays an essential role. Furthermore, according to the basic philosophy of Voiculescu this notion of freeness should be considered (and investigated) in analogy with the classical notion of 'independence' - both freeness and independence prescribe special relations between joint moments of some variables. (Of course, both cases correspond to very special, but also very fundamental situations.)

One of the most interesting features of freeness is that this concept appears in at least two totally different mathematical situations. Originally it was introduced by Voiculescu in the context of operator algebras, later it turned out that there is also some relation, as described above, with random matrices. This gives a non-trivial connection between these two different fields. For example, modelling operator algebras by random matrices has led to some of the most impressive results about operator algebras in the last years.

Furthermore, apart from the concrete manifestations of freeness via random matrices or operator algebras, there exist also an abstract probabilistic theory of freeness, which shows the depth of this concept and which I want to address in the following.

2. Combinatorial approach to free probability: non-crossing partitions and free cumulants

'Freeness' of random variables is defined in terms of mixed moments; namely the defining property is that very special moments (alternating and centered ones) have to vanish. This requirement is not easy to handle in concrete calculations. Thus we will present here another approach to freeness, more combinatorial in nature, which puts the main emphasis on so called 'free cumulants'. These are some polynomials in the moments which behave much better with respect to freeness than the moments. The nomenclature comes from classical probability theory where corresponding objects are also well known and are usually called 'cumulants' or 'semi-invariants'. There exists a combinatorial description of these classical cumulants, which depends on partitions of sets. In the same way, free cumulants can also be described combinatorially, the only difference to the classical case is that one has to replace all partitions by so called 'non-crossing partitions'.

This combinatorial description of freeness is due to me [8, 9] (see also [3]); in a series of joint papers with A. Nica [4, 5, 6] it was pursued very far and yielded a lot of new results in free probability theory. For more information on other aspects of freeness, in particular the original analytical approach of Voiculescu, one should consult the papers [10, 11, 13], the collection of various articles [13], or the monographs [14, 1]

2.1. **Definitions.** A *partition* of the set $S := \{1, \dots, n\}$ is a decomposition

$$\pi = \{V_1, \dots, V_r\}$$

of S into disjoint and non-empty sets V_i, i.e.

$$V_i \cap V_j = \emptyset \quad (i, j = 1, \dots, r; i \neq j) \qquad \text{and} \qquad S = \bigcup_{i=1}^{r} V_i.$$

We call the V_i the *blocks* of π.
For $1 \leq p, q \leq n$ we write

$$p \sim_\pi q \qquad \text{if } p \text{ and } q \text{ belong to the same block of } \pi.$$

A partition π is called *non-crossing* if the following does not occur: There exist $1 \leq p_1 < q_1 < p_2 < q_2 \leq n$ with

$$p_1 \sim_\pi p_2 \not\sim_\pi q_1 \sim_\pi q_2.$$

The set of all non-crossing partitions of $\{1, \dots, n\}$ is denoted by *NC(n)*.

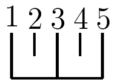
Non-crossing partitions were introduced by Kreweras [2] in a purely combinatorial context without any reference to probability theory.

2.2. **Examples.** We will also use a graphical notation for our partitions; the term 'non-crossing' will become evident in such a notation. Let

$$S = \{1, 2, 3, 4, 5\}.$$

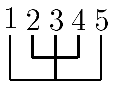
Then the partition

$$\pi = \{(1, 3, 5), (2), (4)\} \qquad \hat{=}$$



is non-crossing, whereas

$$\pi = \{(1, 3, 5), (2, 4)\} \qquad \hat{=}$$



is crossing.

2.3. **Remarks.** 1) In an analogous way, non-crossing partitions $NC(S)$ can be defined for any linearly ordered set S; of course, we have

$$NC(S_1) \cong NC(S_2) \qquad \text{if} \qquad \#S_1 = \#S_2.$$

2) In most cases the following recursive description of non-crossing partitions is of great use: a partition π ist non-crossing if and only if at least one block $V \in \pi$ is an interval and $\pi \backslash V$ is non-crossing; i.e. $V \in \pi$ has the form

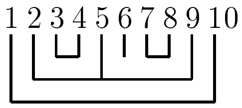
$$V = (k, k + 1, \dots, k + p) \qquad \text{for some } 1 \leq k \leq n \text{ and } p \geq 0,\ k + p \leq n$$

and we have

$$\pi \backslash V \in NC(1, \dots, k - 1, k + p + 1, \dots, n) \cong NC(n - (p + 1)).$$

Example: The partition

$$\{(1,10),(2,5,9),(3,4),(6),(7,8)\} \quad \hat{=}$$



can, by successive removal of intervals, be reduced to

$$\{(1,10),(2,5,9)\}\hat{=}\{(1,5),(2,3,4)\}$$

and finally to

$$\{(1,5)\}\hat{=}\{(1,2)\}.$$

3) By writing a partition π in the form $\pi = \{V_1, \dots, V_r\}$ we will always assume that the elements within each block V_i are ordered in increasing order.

2.4. **Definition.** Let $(\mathcal{A}, \varphi)$ be a probability space, i.e. $\mathcal{A}$ is a unital algebra and $\varphi : \mathcal{A} \to \mathbb{C}$ is a unital linear functional. We define the *(free or non-crossing) cumulants*

$$k_n : \mathcal{A}^n \to \mathbb{C} \qquad (n \in \mathbb{N})$$

(indirectly) by the following system of equations:

$$\varphi(a_1 \dots a_n) = \sum_{\pi \in NC(n)} k_\pi[a_1, \dots, a_n] \qquad (a_1, \dots, a_n \in \mathcal{A}),$$

where k_π denotes a product of cumulants according to the block structure of π:

$$k_\pi[a_1, \dots, a_n] := k_{V_1}[a_1, \dots, a_n] \dots k_{V_r}[a_1, \dots, a_n] \qquad \text{for } \pi = \{V_1, \dots, V_r\} \in NC(n)$$

and

$$k_V[a_1, \dots, a_n] := k_{\#V}(a_{v_1}, \dots, a_{v_l}) \qquad \text{for} \qquad V = (v_1, \dots, v_l).$$

2.5. **Remarks and Examples.** 1) Note: the above equations have the form

$$\varphi(a_1 \dots a_n) = k_n(a_1, \dots, a_n) + \text{smaller order terms}$$

and thus they can be resolved for the $k_n(a_1, \dots, a_n)$ in a unique way.
2) Examples:

- $n = 1$

$$\varphi(a_1) = k_{|}[a_1] = k_1(a_1),$$

thus

$$k_1(a_1) = \varphi(a_1).$$

- $n = 2$

$$\begin{aligned}\varphi(a_1a_2) &= k_{\sqcup}[a_1, a_2] + k_{|\,|}[a_1, a_2]\\ &= k_2(a_1, a_2) + k_1(a_1)k_1(a_2),\end{aligned}$$

thus

$$k_2(a_1, a_2) = \varphi(a_1a_2) - \varphi(a_1)\varphi(a_2).$$

- $n = 3$

$$\begin{aligned}\varphi(a_1a_2a_3) &= k_{\sqcup\!\sqcup}[a_1, a_2, a_3] + k_{|\,\sqcup}[a_1, a_2, a_3] + k_{\sqcup\,|}[a_1, a_2, a_3]\\ &\quad + k_{\underline{|\,{\scriptscriptstyle|}\,|}}[a_1, a_2, a_3] + k_{|\,|\,|}[a_1, a_2, a_3]\\ &= k_3(a_1, a_2, a_3) + k_1(a_1)k_2(a_2, a_3) + k_2(a_1, a_2)k_1(a_3)\\ &\quad + k_2(a_1, a_3)k_1(a_2) + k_1(a_1)k_1(a_2)k_1(a_3),\end{aligned}$$

and thus

$$\begin{aligned}k_3(a_1, a_2, a_3) &= \varphi(a_1a_2a_3) - \varphi(a_1)\varphi(a_2a_3) - \varphi(a_1a_3)\varphi(a_2)\\ &\quad - \varphi(a_1a_2)\varphi(a_3) + 2\varphi(a_1)\varphi(a_2)\varphi(a_3).\end{aligned}$$

3) For $n = 4$ we consider the special case where all $\varphi(a_i) = 0$. Then we have

$$k_4(a_1, a_2, a_3, a_4) = \varphi(a_1a_2a_3a_4) - \varphi(a_1a_2)\varphi(a_3a_4) - \varphi(a_1a_4)\varphi(a_2a_3).$$

4) The k_n are multi-linear functionals in their n arguments.

The meaning of the concept 'cumulants' for freeness is shown by the following theorem.

2.6. **Theorem.** Consider $a_1, \dots, a_m \in \mathcal{A}$. Then the following two statements are equivalent:

i) $a_1, \dots, a_m$ are free.
ii) mixed cumulants vanish, i.e.: We have for all $n \geq 2$ and for all $1 \leq i(1), \dots, i(n) \leq m$:

$$k_n(a_{i(1)}, \dots, a_{i(n)}) = 0,$$

whenever there exist $1 \leq p, q \leq n$ with $i(p) \neq i(q)$.

2.7. **Remarks.** 1) An example of the vanishing of mixed cumulants is that for a, b free we have $k_3(a, a, b) = 0$, which, by the definition of k_3 just means that

$$\varphi(aab) - \varphi(a)\varphi(ab) - \varphi(aa)\varphi(b) - \varphi(ab)\varphi(a) + 2\varphi(a)\varphi(a)\varphi(b) = 0.$$

This vanishing of mixed cumulants in free variables is of course just a reorganization of the information about joint moments of free variables – but in a form which is much more useful for many applications.
2) The above characterization of freeness in terms of cumulants is the

translation of the definition of freeness in terms of moments – by using the relation between moments and cumulants from Definition 2.4. One should note that in contrast to the characterization in terms of moments we do not require that $i(1) \neq i(2) \neq \cdots \neq i(n)$ or $\varphi(a_i) = 0$. (That's exactly the main part of the proof of that theorem: to show that on the level of cumulants the assumption 'centered' is not needed and that 'alternating' can be weakened to 'mixed'.) Hence the characterization of freeness in terms of cumulants is much easier to use in concrete calculations.

3. **Addition of free variables**

3.1. **Notation.** For a random variable $a \in \mathcal{A}$ we put

$$k_n^a := k_n(a, \dots, a)$$

and call $(k_n^a)_{n\geq 1}$ the *(free) cumulants of a.*

Our main theorem on the vanishing of mixed cumulants in free variables specialises in this one-dimensional case to the linearity of the cumulants.

3.2. **Proposition.** Let a and b be free. Then we have

$$k_n^{a+b} = k_n^a + k_n^b \qquad \text{for all } n \geq 1.$$

Proof. We have

$$\begin{aligned} k_n^{a+b} &= k_n(a+b, \dots, a+b) \\ &= k_n(a, \dots, a) + k_n(b, \dots, b) \\ &= k_n^a + k_n^b, \end{aligned}$$

because cumulants which have both a and b as arguments vanish by Theorem 2.6. □

Thus, the addition of free random variables is easy to describe on the level of cumulants; the cumulants are additive under this operation. It remains to make the connection between moments and cumulants as explicit as possible. On a combinatorial level, our definition specializes in the one-dimensional case to the following relation.

3.3. **Proposition.** Let $(m_n)_{n\geq 1}$ and $(k_n)_{n\geq 1}$ be the moments and free cumulants, respectively, of some random variable. The connection between these two sequences of numbers is given by

$$m_n = \sum_{\pi \in NC(n)} k_\pi,$$

where

$$k_\pi := k_{\#V_1} \cdots k_{\#V_r} \qquad \text{for} \qquad \pi = \{V_1, \dots, V_r\}.$$

Example: For $n = 3$ we have

$$\begin{aligned} m_3 &= k_{\sqcup\!\sqcup} + k_{\shortmid\,\sqcup} + k_{\sqcup\,\shortmid} + k_{\llcorner\!\shortmid\!\lrcorner} + k_{\shortmid\,\shortmid\,\shortmid} \\ &= k_3 + 3k_1k_2 + k_1^3. \end{aligned}$$

For concrete calculations, however, one would prefer to have a more analytical description of the relation between moments and cumulants. This can be achieved by translating the above relation to corresponding formal power series.

3.4. **Theorem.** Let $(m_n)_{n\geq 1}$ and $(k_n)_{n\geq 1}$ be two sequences of complex numbers and consider the corresponding formal power series

$$\begin{aligned} M(z) &:= 1 + \sum_{n=1}^{\infty} m_n z^n, \\ C(z) &:= 1 + \sum_{n=1}^{\infty} k_n z^n. \end{aligned}$$

Then the following three statements are equivalent:

(i) We have for all $n \in \mathbb{N}$

$$m_n = \sum_{\pi \in NC(n)} k_\pi = \sum_{\pi = \{V_1,\dots,V_r\} \in NC(n)} k_{\#V_1} \dots k_{\#V_r}.$$

(ii) We have for all $n \in \mathbb{N}$ (where we put $m_0 := 1$)

$$m_n = \sum_{s=1}^{n} \sum_{\substack{i_1,\dots,i_s \in \{0,1,\dots,n-s\} \\ i_1+\cdots+i_s = n-s}} k_s m_{i_1} \dots m_{i_s}.$$

(iii) We have

$$C[zM(z)] = M(z).$$

Proof. We rewrite the sum

$$m_n = \sum_{\pi \in NC(n)} k_\pi$$

in the way that we fix the first block V_1 of π (i.e. that block which contains the element 1) and sum over all possibilities for the other blocks; in the end we sum over V_1:

$$m_n = \sum_{s=1}^{n} \sum_{V_1 \text{ with } \#V_1 = s} \sum_{\substack{\pi \in NC(n) \\ \text{where } \pi = \{V_1, \dots\}}} k_\pi.$$

If

$$V_1 = (v_1 = 1, v_2, \dots, v_s),$$

then $\pi = \{V_1, \dots\} \in NC(n)$ can only connect elements lying between some v_k and v_{k+1}, i.e. $\pi = \{V_1, V_2, \dots, V_r\}$ such that we have for all $j = 2, \dots, r$: there exists a k with $v_k < V_j < v_{k+1}$. There we put

$$v_{s+1} := n + 1.$$

Hence such a π decomposes as

$$\pi = V_1 \cup \tilde{\pi}_1 \cup \dots \cup \tilde{\pi}_s,$$

where

$$\tilde{\pi}_j \text{ is a non-crossing partition of } \{v_j + 1, v_j + 2, \dots, v_{j+1} - 1\}.$$

For such π we have

$$k_\pi = k_{\#V_1} k_{\tilde{\pi}_1} \dots k_{\tilde{\pi}_s} = k_s k_{\tilde{\pi}_1} \dots k_{\tilde{\pi}_s},$$

and thus we obtain

$$\begin{aligned}
m_n &= \sum_{s=1}^{n} \sum_{1=v_1<v_2<\dots<v_s\le n} \sum_{\substack{\pi = V_1 \cup \tilde{\pi}_1 \cup \dots \cup \tilde{\pi}_s \\ \tilde{\pi}_j \in NC(v_j+1, \dots, v_{j+1}-1)}} k_s k_{\tilde{\pi}_1} \dots k_{\tilde{\pi}_s} \\
&= \sum_{s=1}^{n} k_s \sum_{1=v_1<v_2<\dots<v_s\le n} \Big(\sum_{\tilde{\pi}_1 \in NC(v_1+1, \dots, v_2-1)} k_{\tilde{\pi}_1} \Big) \dots \Big(\sum_{\tilde{\pi}_s \in NC(v_s+1, \dots, n)} k_{\tilde{\pi}_s} \Big) \\
&= \sum_{s=1}^{n} k_s \sum_{1=v_1<v_2<\dots<v_s\le n} m_{v_2-v_1-1} \dots m_{n-v_s} \\
&= \sum_{s=1}^{n} \sum_{\substack{i_1, \dots, i_s \in \{0,1,\dots,n-s\} \\ i_1+\dots+i_s+s=n}} k_s m_{i_1} \dots m_{i_s} \qquad (i_k := v_{k+1} - v_k - 1).
\end{aligned}$$

This yields the implication (i) $\Longrightarrow$ (ii).
We can now rewrite (ii) in terms of the corresponding formal power

series in the following way (where we put $m_0 := k_0 := 1$):

$$\begin{aligned}
M(z) &= 1 + \sum_{n=1}^{\infty} z^n m_n \\
&= 1 + \sum_{n=1}^{\infty} \sum_{s=1}^{n} \sum_{\substack{i_1,\dots,i_s \in \{0,1,\dots,n-s\} \\ i_1+\dots+i_s = n-s}} k_s z^s m_{i_1} z^{i_1} \dots m_{i_s} z^{i_s} \\
&= 1 + \sum_{s=1}^{\infty} k_s z^s \big(\sum_{i=0}^{\infty} m_i z^i\big)^s \\
&= C[zM(z)].
\end{aligned}$$

This yields (iii).
Since (iii) describes uniquely a fixed relation between the numbers $(k_n)_{n\geq 1}$ and the numbers $(m_n)_{n\geq 1}$, this has to be the relation (i). □

If we rewrite the above relation between the formal power series in terms of the Cauchy transform

$$G(z) := \sum_{n=0}^{\infty} \frac{m_n}{z^{n+1}}$$

and the R-transform

$$R(z) := \sum_{n=0}^{\infty} k_{n+1} z^n,$$

then we obtain Voiculescu's formula.

3.5. **Corollary.** The relation between the Cauchy transform $G(z)$ and the R-transform $R(z)$ of a random variable is given by

$$G[R(z) + \frac{1}{z}] = z.$$

Proof. We just have to note that the formal power series $M(z)$ and $C(z)$ from Theorem 3.4 and $G(z)$, $R(z)$, and $K(z) = R(z) + \frac{1}{z}$ are related by:

$$G(z) = \frac{1}{z} M(\frac{1}{z})$$

and

$$C(z) = 1 + zR(z) = zK(z), \qquad \text{thus} \qquad K(z) = \frac{C(z)}{z}.$$

This gives

$$K[G(z)] = \frac{1}{G(z)} C[G(z)] = \frac{1}{G(z)} C[\frac{1}{z} M(\frac{1}{z})] = \frac{1}{G(z)} M(\frac{1}{z}) = z,$$

thus $K[G(z)] = z$ and hence also

$$G[R(z) + \frac{1}{z}] = G[K(z)] = z.$$

□

3.6. **Free convolution.** The above results give us a quite effective tool for calculating the distribution of the sum $a + b$ of free variables from the distribution of a and the distribution of b. In analogy with the usual convolution (which corresponds to the sum of independent random variables) we introduce the notion $\boxplus$ of *free convolution* as operation on probability measures by

$$\mu_{a+b} = \mu_a \boxplus \mu_b \qquad \text{if } a, b \text{ are free.}$$

Then we know that free cumulants and the R-transform linearize this free convolution.

In particular, we also have the free convolution powers

$$\mu^{\boxplus r} := \mu \boxplus \cdots \boxplus \mu \qquad (r\text{-times})$$

of μ, which are in terms of cumulants characterized by

$$k_n(\mu^{\boxplus r}) = r \cdot k_n(\mu).$$

If we are given free variables a and b and we want to calculate the distribution of $a + b$, then we calculate the R-transforms R_a and R_b and get thus by the linearization property the R-transform of $a + b$,

$$R_{a+b} = R_a + R_b.$$

It remains to extract the distribution out of this. From the R-transform we can get the Cauchy transform G_{a+b} by Corollary 3.5, and then we use the classical Stietljes inversion formula for extracting the distribution from this. In general, the relation between R-transform and Cauchy transform might lead to equations which have no analytic solutions, however, in many concrete case these equations can be solved. For example, if we put $\mu = \frac{1}{2}(\delta_0 + \delta_1)$, then the above machinery shows that the distribution $\mu \boxplus \mu$ is given by the arcsine law.

3.7. **Application: random sum of matrices** [7]**.** Fix two (compactly supported) probability measures μ and ν on the real line and consider deterministic (e.g., diagonal) $N \times N$-matrices A_N and C_N, whose eigenvalue distribution converges, for $N \to \infty$, towards the given measures. To put it in the language introduced in Section 1, we assume that

$$A_N \to a \qquad \text{with} \qquad \mu_a = \mu$$

and

$$C_N \to c \qquad \text{with} \qquad \mu_c = \nu.$$

Now we rotate C_N against A_N randomly by replacing C_N by

$$B_N := U_N C_N U_N^*,$$

where U_N is a random Haar unitary matrix from the ensemble of unitary $N \times N$-matrices equipped with the Haar measure. Of course, the eigenvalue distribution of B_N is the same as the one of C_N, however, any definite relation between the eigenspaces of A_N and the eigenspaces of C_N has now been destroyed. A_N and B_N are in the limit $N \to \infty$ generic realizations of the given eigenvalue distributions μ and ν. The question which we want to address is: What is the eigenvalue distribution of the sum $A_N + B_N$ in the limit $N \to \infty$, i.e. what can we say about

$$A_N + B_N \to ?$$

A version of the theorem of Voiculescu about the connection between random matrices and freeness tells us that A_N and B_N become free in the limit $N \to \infty$, i.e. it yields that

$$(A_N, B_N) \to (a, b) \qquad \text{with } \mu_a = \mu,\ \mu_b = \nu, \text{ and } a, b \text{ free.}$$

Thus we know that the eigenvalue distribution of $A_N + B_N$ converges towards the distribution of $a+b$ where a and b are free. But the distribution of $a+b$ can be calculated with our tools from free probability in a very effective and systematic way by using the R-transform machinery. For example, if we take the generic sum of two projections of trace $1/2$, (i.e., $\mu = \nu = \frac{1}{2}(\delta_0 + \delta_1)$, then our example from above shows us that the distribution of

$$\begin{pmatrix} 1 &&&&& \\ & 0 &&&& \\ && 1 &&& \\ &&& 0 && \\ &&&& 1 & \\ &&&&& \ddots \end{pmatrix} + U_N \begin{pmatrix} 1 &&&&& \\ & 0 &&&& \\ && 1 &&& \\ &&& 0 && \\ &&&& 1 & \\ &&&&& \ddots \end{pmatrix} U_N^*,$$

is, in the limit $N \to \infty$, given by the arcsine law.

4. Multiplication of free variables

Finally, to show that our description of freeness in terms of cumulants has also a significance apart from dealing with additive free convolution, we will apply it to the problem of the product of free random variables: Consider $a_1, \dots, a_n, b_1, \dots, b_n$ such that $\{a_1, \dots, a_n\}$ and $\{b_1, \dots, b_n\}$

are free. We want to express the distribution of the random variables $a_1 b_1, \dots, a_n b_n$ in terms of the distribution of the a's and of the b's.

4.1. **Notation.** 1) Analogously to k_π we define for

$$\pi = \{V_1, \dots, V_r\} \in NC(n)$$

the expression

$$\varphi_\pi[a_1 \dots, a_n] := \varphi_{V_1}[a_1, \dots, a_n] \dots \varphi_{V_r}[a_1, \dots, a_n],$$

where

$$\varphi_V[a_1, \dots, a_n] := \varphi(a_{v_1} \dots a_{v_l}) \qquad \text{for} \qquad V = (v_1, \dots, v_l).$$

Examples:

$$\begin{aligned}
\varphi_{\sqcup\!\sqcup}[a_1, a_2, a_3] &= \varphi(a_1 a_2 a_3) \\
\varphi_{|\,\sqcup}[a_1, a_2, a_3] &= \varphi(a_1)\varphi(a_2 a_3) \\
\varphi_{\sqcup\,|}[a_1, a_2, a_3] &= \varphi(a_1 a_2)\varphi(a_3) \\
\varphi_{\sqcup\!\!\!\!\!\;^{\scriptscriptstyle |}\;}[a_1, a_2, a_3] &= \varphi(a_1 a_3)\varphi(a_2) \\
\varphi_{|\,|\,|}[a_1, a_2, a_3] &= \varphi(a_1)\varphi(a_2)\varphi(a_3)
\end{aligned}$$

2) Let $\sigma, \pi \in NC(n)$. Then we write

$$\sigma \leq \pi$$

if each block of σ is contained as a whole in some block of π, i.e. σ can be obtained out of π by refinement of the block structure.
Example:

$$\{(1), (2,4), (3), (5,6)\} \leq \{(1,5,6), (2,3,4)\}$$

With these notations we can generalize the relation

$$\varphi(a_1 \dots a_n) = \sum_{\pi \in NC(n)} k_\pi[a_1, \dots, a_n]$$

in the following way.

$$\varphi_\sigma[a_1, \dots, a_n] = \sum_{\substack{\pi \in NC(n) \\ \pi \leq \sigma}} k_\pi[a_1, \dots, a_n].$$

Consider now

$$\{a_1, \dots, a_n\}, \{b_1, \dots, b_n\} \qquad \text{free.}$$

We want to express alternating moments in a and b in terms of moments of a and moments of b. We have

$$\varphi(a_1 b_1 a_2 b_2 \dots a_n b_n) = \sum_{\pi \in NC(2n)} k_\pi[a_1, b_1, a_2, b_2, \dots, a_n, b_n].$$

Since the a's are free from the b's, Theorem 2.6 tells us that only such π contribute to the sum whose blocks do not connect a's with b's. But this means that such a π has to decompose as

$$\pi = \pi_1 \cup \pi_2 \qquad \text{where} \quad \begin{aligned} &\pi_1 \in NC(1, 3, 5, \dots, 2n-1) \\ &\pi_2 \in NC(2, 4, 6, \dots, 2n). \end{aligned}$$

Thus we have

$$\begin{aligned} &\varphi(a_1 b_1 a_2 b_2 \dots a_n b_n) \\ &\quad = \sum_{\substack{\pi_1 \in NC(odd), \pi_2 \in NC(even) \\ \pi_1 \cup \pi_2 \in NC(2n)}} k_{\pi_1}[a_1, a_2, \dots, a_n] \cdot k_{\pi_2}[b_1, b_2, \dots, b_n] \\ &\quad = \sum_{\pi_1 \in NC(odd)} \Big(k_{\pi_1}[a_1, a_2, \dots, a_n] \cdot \sum_{\substack{\pi_2 \in NC(even) \\ \pi_1 \cup \pi_2 \in NC(2n)}} k_{\pi_2}[b_1, b_2, \dots, b_n] \Big). \end{aligned}$$

Note now that for a fixed π_1 there exists a maximal element σ with the property $\pi_1 \cup \sigma \in NC(2n)$ and that the second sum is running over all $\pi_2 \leq \sigma$.

4.2. **Definition.** Let $\pi \in NC(n)$ be a non-crossing partition of the numbers $1, \dots, n$. Introduce additional numbers $\bar{1}, \dots, \bar{n}$, with alternating order between the old and the new ones, i.e. we order them in the way

$$1\bar{1}2\bar{2}\dots n\bar{n}.$$

We define the *complement* $K(\pi)$ of π as the maximal $\sigma \in NC(\bar{1}, \dots, \bar{n})$ with the property

$$\pi \cup \sigma \in NC(1, \bar{1}, \dots, n, \bar{n}).$$

If we present the partition π graphically by connecting the blocks in $1, \dots, n$, then σ is given by connecting as much as possible the numbers $\bar{1}, \dots, \bar{n}$ without getting crossings among themselves and with π.

(This natural notation of the complement of a non-crossing partition is also due to Kreweras [2]. Note that there is no analogue of this for the case of all partitions.)

With this definition we can continue our above calculation as follows:

$$\begin{aligned}
&\varphi(a_1b_1a_2b_2\dots a_nb_n)\\
&\qquad = \sum_{\pi_1\in NC(n)} \Big(k_{\pi_1}[a_1,a_2,\dots,a_n]\cdot \sum_{\substack{\pi_2\in NC(n)\\ \pi_2\leq K(\pi_1)}} k_{\pi_2}[b_1,b_2,\dots,b_n]\Big)\\
&\qquad = \sum_{\pi_1\in NC(n)} k_{\pi_1}[a_1,a_2,\dots,a_n]\cdot \varphi_{K(\pi_1)}[b_1,b_2,\dots,b_n].
\end{aligned}$$

Thus we have proved the following result.

4.3. **Theorem.** Consider

$$\{a_1,\dots,a_n\},\{b_1,\dots,b_n\} \qquad \text{free.}$$

Then we have

$$\varphi(a_1b_1a_2b_2\dots a_nb_n) = \sum_{\pi\in NC(n)} k_{\pi}[a_1,a_2,\dots,a_n]\cdot \varphi_{K(\pi)}[b_1,b_2,\dots,b_n].$$

Similar to the additive case one can translate this combinatorial description of the product of free variables in an analytic way in terms of the so-called S-transfrom. However, this is more complicated as in the case of the R-transform and we will not address this problem here. Instead, we want to show that the above combinatorial description of the product of free variables can lead to quite explicit (and unexpected) results without running through an analytic reformulation. Such a result is given in our final application to the problem of the compression of a random matrix.

4.4. **Application: Compression of random matrix.** Consider, as in Section 3.7, a sequence of deterministic $N\times N$-matrices C_N with prescribed eigenvalue distribution μ in the limit $N\to\infty$ and consider the randomly rotated version $A_N := U_NC_NU_N^*$ of this matrix. The question we want to address is the following: Can we calculate the eigenvalue distribution of upper left corners of the matrices A_N. Formally, we get these corners by compressing A_N with projections of the form

$$P_N := \begin{pmatrix} 1 &&&&& \\ & \ddots &&&& \\ && 1 &&& \\ &&& 0 && \\ &&&& \ddots & \\ &&&&& 0 \end{pmatrix},$$

where $\mathrm{tr}_N(P_N) \to \alpha$ for some fixed α with $0 < \alpha \leq 1$. Thus we ask for the eigenvalue distribution of $P_N A_N P_N$ in the limit $N \to \infty$. (However, we have to calculate this in the compressed space, throwing away the bunch of trivial zero eigenvalues outside of the non-trivial corner of $P_N A_N P_N$, i.e., we consider $P_N A_N P_N$ not as $N \times N$-matrix, but as $\alpha N \times \alpha N$-matrix.)

Now, again by Voiculescu's theorem about asymptotic freeness of random matrices, we know that

$$(A_N, P_N) \to (a, p),$$

where a has the prescribed distribution μ, p is a projection of trace α and a and p are free. Thus, the answer for our question on the distribution of corners in randomly rotated matrices is provided by calculating the distribution of pap in the compressed space, i.e. by calculating the renormalized moments

$$\frac{1}{\alpha}\varphi[(pap)^n],$$

which is, by the trace property of φ and the projection property $p^2 = p$, the same as

$$\frac{1}{\alpha}\varphi[(ap)^n].$$

This fits now exactly in the above frame of calculating the moments of products of free variables, in the special case where the second variable is a projection of trace α. Using $p^k = p$ for all $k \geq 1$ and $\varphi(p) = \alpha$ gives

$$\varphi_{K(\pi)}[p, p, \dots, p] = \varphi(p \dots p)\varphi(p \dots p) \cdots = \alpha^{|K(\pi)|},$$

where $|K(\pi)|$ denotes the number of blocks of $K(\pi)$. We can express this number of blocks also in terms of π, since we always have the relation

$$|\pi| + |K(\pi)| = n + 1.$$

Thus we can continue our calculation of Theorem 4.3 in this case as

$$\begin{aligned}\frac{1}{\alpha}\varphi[(ap)^n] &= \frac{1}{\alpha} \sum_{\pi \in NC(n)} k_\pi[a, \dots, a]\alpha^{n+1-|\pi|} \\ &= \sum_{\pi \in NC(n)} \frac{1}{\alpha^{|\pi|}} k_\pi[\alpha a, \dots, \alpha a],\end{aligned}$$

which shows that

$$k_n(pap, \dots, pap) = \frac{1}{\alpha} k_n(\alpha a, \dots, \alpha a)$$

for all n. By our remarks on the additive free convolution, this gives the surprising result that the renormalized distribution of pap is given by

$$\mu_{pap} = \mu_{\alpha a}^{\boxplus 1/\alpha}.$$

In particular, for $\alpha = 1/2$, we have

$$\mu_{pap} = \mu_{1/2a}^{\boxplus 2} = \mu_{1/2a} \boxplus \mu_{1/2a}.$$

This means that the distribution of the upper left corner of size 1/2 of a randomly rotated matrix is, apart from rescaling with the factor 1/2, the same as the distribution of the sum of the considered matrix and another randomly rotated copy of itself. E.g., if we take the example $\mu = \frac{1}{2}(\delta_0 + \delta_1)$, then the corner of size 1/2 of such a randomly rotated projection has as eigenvalue distribution the arcsine law.

References

1. F. Hiai and D. Petz, *The semicircle law, free random variables and entropy* (Mathematical Surveys and Monographs, Vol. 77), AMS, 2000.
2. G. Kreweras, *Sur les partitions non-croisees d'un cycle*, Discrete Math. **1** (1972), 333–350.
3. A. Nica, *R-transforms of free joint distributions, and non-crossing partitions*, J. Funct. Anal. **135** (1996), 271–296.
4. A. Nica and R. Speicher, *On the multiplication of free n-tuples of non-commutative random variables* (with an appendix by D. Voiculescu), Amer. J. Math. **118** (1996), 799–837.
5. A. Nica and R. Speicher, *R-diagonal pairs—a common approach to Haar unitaries and circular elements*, Free Probability Theory (D.-V. Voiculescu, ed.), AMS, 1997, pp. 149–188.
6. A. Nica and R. Speicher, *Commutators of free random variables*, Duke Math. J. **92** (1998), 553–592.
7. R. Speicher, *Free convolution and the random sum of matrices*, RIMS **29** (1993), 731–744.
8. R. Speicher, *Multiplicative functions on the lattice of non-crossing partitions and free convolution*, Math. Ann. **298** (1994), 611–628.
9. R. Speicher, *Combinatorial theory of the free product with amalgamation and operator-valued free probability theory*, Memoirs of the AMS **627** (1998).
10. D. Voiculescu, *Addition of certain non-commuting random variables*, J. Funct. Anal. **66** (1986), 323–346.
11. D. Voiculescu, *Limit laws for random matrices and free products*, Invent. math. **104** (1991), 201–220.
12. D. Voiculescu, *Free probability theory: random matrices and von Neumann algebras*, Proceedings of the ICM 1994, Birkhäuser, 1995, pp. 227–241.
13. D. Voiculescu (ed.), *Free Probability Theory* (Fields Institute Communications, vol. 12), AMS, 1997.
14. D.V. Voiculescu, K.J. Dykema, and A. Nica, *Free Random Variables* (CRM Monograph Series, vol. 1), AMS, 1993.

Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada

E-mail address: speicher@mast.queensu.ca