The Combinatorial Motivation of the R-transform and Voiculescu's Characterization, a Reference

Nagoya University Mathematics Department
Alexander Jerschow

Abstract

This document serves as a personal reference on the combinatorial approach to free probability theory, created because I often forget the flow from the definition of non-commutative probability spaces through cumulants to the R-transform and so on. Much of this material was recycled for my first biannual research report for my master's program at Nagoya University (中間研究報告書). The text covers basic definitions of non-commutative probability spaces, combinatorial aspects including non-crossing partitions and multiplicative functionals, and the analytization process leading to the R-transform. We conclude with Voiculescu's characterization of the R-transform using Nevanlinna-Pick theory.

Contents

References and Abbreviations

In this document, I use the following references with the indicated abbreviations:

- **Voiculescu** = Voiculescu, D. V. (1986). *Addition of certain non-commuting random variables.* Journal of Functional Analysis, 66(3), 323-346.

- **Mingo Speicher** = Mingo, J. A., & Speicher, R. (2017). *Free Probability and Random Matrices.* Fields Institute Monographs, Vol. 35, Springer.

- **Zeitouni Anderson Guionnet** = Anderson, G. W., Guionnet, A., & Zeitouni, O. (2010). *An Introduction to Random Matrices.* Cambridge University Press.

- **Speicher** = Speicher, R. (2019). *Lecture Notes on Free Probability Theory.* arXiv:1908.08125.

1 Basic Definitions

The following discusses various distinctions, for instance, non-commutative probability spaces, $*$-probability spaces, C^*-probability spaces, etc. For the sake of brevity, we often give results in relation to one type of setting (usually non-commutative probability spaces or $*$-probability spaces) without giving a full account of possible analogs in related settings.

A non-commutative probability space is defined as a pair (A, φ) where A is a unital algebra over $\mathbb{C}$ and $\varphi : A \to \mathbb{C}$ is a unital[1] linear functional. The elements of A are called non-commutative random variables.

For unital subalgebras $(A_i)_{i\in I}$ of A, we say they are free if for any $k \in \mathbb{N}$, indices $i(1) \neq i(2) \neq \cdots \neq i(k)$, and elements $a_j \in A_{i(j)}$ with $\varphi(a_j) = 0$, we have (note $[k] := (i)_{i=1}^k$):

$$\varphi\left(\overset{\rightarrow}{\prod_{i\in[k]}} a_i\right) = \varphi(a_1 a_2 \cdots a_k) = 0.$$

Here we make use of directed product notation because we are in the non-commutative setting. Note that $i(1) \neq i(2) \neq \cdots \neq i(k)$ signifies that adjacent elements in the product come from different subalgebras. Note also that we do not require $i(1) \neq i(k)$. When A is a $*$-algebra and φ is positive[2], we call (A, φ) a $*$-probability space. If for all $a \in A \setminus \{0\}$, we have $\varphi(a^*a) > 0$, we call φ faithful. In the context of a $*$-algebra, we recall the standard definitions:

- a is self-adjoint $\overset{def}{\iff}$ $a = a^*$
- a is unitary $\overset{def}{\iff}$ $a^*a = aa^* = 1$

[1] unitality $\overset{def}{\iff}$ $\varphi(1) = 1$

[2] positivity $\overset{def}{\iff}$ $\varphi(a^*a) \geq 0 \ \forall a \in A$

- a is normal $\overset{def}{\iff}$ $a^*a = aa^*$

Further distinctions arise based on the structure of A. For instance, when A is a C^*-algebra or W^*-algebra, we obtain C^*-probability spaces or W^*-probability spaces, respectively.

For a general non-commutative probability space (A, φ) (not necessarily a $*$-probability space), we say φ is tracial if $\varphi(ab) = \varphi(ba)$ for all $a, b \in A$. In this case, (A, φ) is called a tracial probability space.

Important examples of non-commutative/$*$-probability spaces include:

- The algebra $M_n(\mathbb{C})$ of $n \times n$ complex matrices equipped with the trace $\mathrm{Tr}(A) = \sum_{k=1}^n A_{kk}$ forms a $*$-probability space. The normalized trace $\frac{1}{n}\mathrm{Tr}$ is particularly important in random matrix theory, where large random matrix distributions exhibit so-called "asymptotic freeness", revealing deep connections between random matrix theory and free probability.

- For a classical probability space $(\Omega, \mathcal{F}, P)$, the space $L^\infty(\Omega, P)$ of essentially bounded random variables with $\varphi(a) = \int a dP$ forms a $*$-probability space, where the involution is given by complex conjugation.

While freeness was defined for subalgebras, we extend it to elements: elements $(a_i)_{i \in I}$ in A are free if the unital algebras they generate are free. Similarly, for subsets $B_i \subset A$, we say they are free if the generated unital subalgebras are free. Additionally, in the case of a $*$-probability space, we have the notion of $*$-freeness where we replace "unital subalgebras", or in the case of elements/sets, "generated unital subalgebras" with "unital $*$-subalgebras" or "generated unital $*$-subalgebras", respectively.

For a non-commutative probability space (A, φ), the moments of $a \in A$ are the values $\varphi(a^n)$ for $n \in \mathbb{N}$. The joint moments of elements $a_1, \ldots, a_s \in A$ are the values $\varphi(a_{i_1} \cdots a_{i_k})$ for $k \in \mathbb{N}$ and $i_j \in \{1, \ldots, s\}$. In a $*$-probability space, the $*$-moments of a are the joint moments of (a, a^*), and the joint $*$-moments of $(a_i)_{i=1}^s$ are the joint moments of $(a_i, a_i^*)_{i=1}^s$.

The collection of all moments of a is called its distribution. Similarly, the collection of all moments of $(a_i)_{i=1}^s$ is called their joint distribution. In the $*$-case, we similarly speak of $*$-moments/joint $*$-moments to define $*$-distributions and joint $*$-distributions, respectively.

Freeness serves as the non-commutative analog of classical independence. A fundamental result shows that, similarly to the joint distribution classically independent random variables, the joint distribution of free random variables is determined by their individual distributions:

[**Mingo Speicher 1.2 Proposition 13**] Let (A, φ) be a non-commutative probability space and $(A_i)_{i \in I}$ free unital subalgebras. If B is the algebra generated by the A_i, then $\varphi|_B$ can be uniquely recovered from $(\varphi|_{A_i})_{i \in I}$.

This result extends to $*$-probability spaces since $*$-distributions are determined by (non-$*$) distributions of generated self-adjoint collections.

In classical probability theory, distributions are defined via measures on σ-algebras. In non-commutative probability, we lack such measures in general. Instead, we define distributions through the functional φ by considering collections of non-commutative monomials. This approach seems natural given by the fact that in classical probability, many distributions are characterized by their moments (e.g., those satisfying Carleman's condition). Thus, we could view φ as a "generalized non-commutative expectation functional".

When considering C^*-,W^*-probability spaces, our definition of distributions also makes sense from a spectral theoretic point of view. By definition of the $*$-distribution of a normal element a, we have the values of φ on all monomials of a, a^*. But, by linearity of φ, this yields the values of φ on all polynomials, which, by the continuous functional calculus, is dense in the generated C^*-algebra of a with respect to the given C^* norm. Because φ is, in particular, a state, it is bounded with norm 1 by C^* theory (i.e. continuous). This means, in the C^* framework, our definition of $*$-distribution does not merely yield values of φ on the $*$-algebra generated by a, but on its closure, the generated C^*-algebra (polynomial approximation and continuity of φ).

Further, in the case our C^*-algebra is in fact a W^*-algebra and our φ is normal[3], we can use the measurable functional calculus and the density of the continuous functions in the bounded measurable ones (with respect to the topology of pointwise convergence) to see that the operators corresponding to the continuous functions are dense in their generated W^*-algebra, which is nothing but their weak, σ-weak, strong, or σ-strong closure (these operators form a convex set, upon which the 4 closures coincide by W^* theory). By φ's normality, we obtain φ on this generated W^*-algebra. Therefore, at least for normal elements, if we have a $*$-distribution, we may speak of φ on respective generated C^*-,W^*-algebras.

A fundamental difference between commutative and non-commutative prob-

[3]normality $\overset{def}{\Longleftrightarrow}$ continuity with respect to the weak operator topology on the domain

ability becomes clear through the following result:

[**Speicher Proposition 1.10**] For a $*$-probability space (A, φ) with faithful φ and self-adjoint elements x, y, if x and y commute and are free, then one of them is constant.

This shows that in the context of $*$-probability spaces with faithful expectations (a substantial class of non-commutative probability spaces), the non-commutative framework becomes trivial when applied to commutative situations. Given that the non-commutative setting underlies a very rich theory, we see that there is a big distinction between the classical and non-commutative cases.

2 Combinatorial Background

2.1 Non-Crossing Partitions

To understand the role of combinatorics in calculating moments of non-commutative random variables, consider the following motivating examples:

(i) For the simplest non-trivial case (let a and b be free):

$$0 = \varphi((a - \varphi(a))(b - \varphi(b)) = \varphi(ab) - \varphi(a)\varphi(b)$$

implies as in the commutative setting with expectation: $\varphi(ab) = \varphi(a)\varphi(b)$. Note that to use our freeness assumption, we must have $\varphi(a_{i(j)}) = 0$ for all j, so we consider the "centered" variables $a_{i(j)} - \varphi(a_{i(j)})$.

(ii) Next, one concludes using (i) (let $a1, a2, b$ be free):

$$0 = \varphi((a_1 - \varphi(a_1))(b - \varphi(b))(a_2 - \varphi(a_2))) = \varphi(a_1 b a_2) - \varphi(a_1 a_2)\varphi(b)$$

i.e. $\varphi(a_1 a_2)\varphi(b) = \varphi(a_1 b a_2)$.

(iii) For a more complex case, one can use the above to calculate:

$$\varphi((a_1 - \varphi(a_1))(a_2 - \varphi(a_2))(a_1 - \varphi(a_1))(a_2 - \varphi(a_2))) =$$

$$\varphi(a_1 a_2 a_1 a_2) - \varphi(a_1^2)\varphi(a_2)^2 - \varphi(a_1)^2\varphi(a_2^2) + \varphi(a_1)^2\varphi(a_2)^2$$

i.e. $\varphi(a_1 a_2 a_1 a_2) = \varphi(a_1)^2\varphi(a_2^2) + \varphi(a_1^2)\varphi(a_2)^2 - \varphi(a_1)^2\varphi(a_2)^2$.

(iv) For $\{a_i\}, \{b_i\}, c, d$ free we may calculate the following:

$$\varphi(a_1 b_1 c b_2 a_2 d a_3) = \varphi((a_1)(b_1 c b_2)(a_2 d a_3)) = \varphi((a_1)(a_2 d a_3))\varphi(b_1 c b_2) =$$

$$\varphi(a_1a_2a_3)\varphi(d)\varphi(b_1b_2)\varphi(c).$$

Note that in the second equality, we combined elements to obtain three new ones, all free from each other, and applied (ii). We similarly applied (i) in the last equality.

If we "group" product elements from the same subalgebra (by connecting them with lines lying underneath the product), we get a "partition of the product" (for illustration, please refer to Speicher's lecture notes, chapter 2). Through such calculations, one observes that products resulting in "partitions that do not cross" (so-called non-crossing partitions) work out nicely, whereas those that do cross do not. This suggests a fundamental relationship between these "nice products" and non-crossing partitions.

[**Zeitouni Anderson Guionnet Definition 2.1.4**] Let $P(n)$ denote the set of all partitions of $[n] = \{1, 2, \dots, n\}$.

(i) A partition $\pi \in P(n)$ is called *crossing* if there exist $p_1 < q_1 < p_2 < q_2$ such that $p_1 \sim_\pi p_2$ and $q_1 \sim_\pi q_2$. A partition is called *non-crossing* if it is not crossing.

(ii) Let $NC(n)$ denote the set of all non-crossing partitions of $[n]$.

(iii) For $\pi, \sigma \in NC(n)$, we say $\pi \leq \sigma$ if every block of π is completely contained in some block of σ. This relation equips $NC(n)$ with a partial order. We may also write $\pi < \sigma$ if $\pi \leq \sigma$ and $\pi \neq \sigma$.

(iv) The unique maximal element of $NC(n)$ with respect to this order is $1_n := \{\{1, 2, \dots, n\}\}$.

(v) The unique minimal element of $NC(n)$ is $0_n := \{\{1\}, \{2\}, \dots, \{n\}\}$.

[**Example**]

(i) $\{\{1\}, \{2\}\} \leq \{\{1, 2\}\}$

(ii) $\{\{1, 4\}, \{2, 3\}, \{5, 6\}, \{7\}\} \leq \{\{1, 4, 5, 6\}, \{2, 3\}, \{7\}\}$ and $\{\{1, 4\}, \{2, 3\}, \{5, 6, 7\}\} \leq \{\{1, 2, 3, 4, 7\}, \{5, 6\}\}$

The set $NC(n)$ possesses rich structure—it is not just a partially ordered set but actually forms a lattice, meaning any two elements have a unique "sup" as well as an "inf", i.e. a smallest element larger than the two and a largest element smaller than the two:

[**Definition**] A partially ordered set P is a *lattice* if for all $a, b \in P$ there exists a *meet* $a \wedge b := \max\{c \in P \mid c \leq a, c \leq b\}$ and a *join* $a \vee b := \min\{c \in$

$P \mid c \geq a, c \geq b\}$.

[**Example**] The meet of $\{\{1,4\},\{2,3\},\{5\}\}$ and $\{\{1,5\},\{2,3,4\}\}$ is $\{\{1\},\{2,3\},\{4\},\{5\}\}$ and their join is the maximal element 1_5.

$P(n)$ is also a lattice and its meet restricted to the sub-lattice $NC(n)$ corresponds to $NC(n)$'s meet. However, the join operation in $NC(n)$ differs from that in $P(n)$. The join in $NC(n)$ can be obtained by taking the join in $P(n)$ and then merging crossing partition blocks. This distinction is illustrated by: $\{\{1,3\},\{2\},\{4\}\} \vee \{\{1\},\{2,4\},\{3\}\} = \{\{1,3\},\{2,4\}\}$ in $P(4)$ whereas $\{\{1,3\},\{2\},\{4\}\} \vee \{\{1\},\{2,4\},\{3\}\} = 1_4$ in $NC(4)$. One sees that the join in $P(n)$ need not always lie in $NC(n)$. Finally, the following example shows that, even in $P(n)$, a naive relation such as $p \sim_{\pi\vee\sigma} q \iff p \sim_\pi q$ or $p \sim_\sigma q$ does not hold:

[**Example**]

(i) For $\pi = \{\{1\},\{2,3\}\}$ and $\sigma = \{\{1,2\},\{3\}\}$ in $P(3)$, their join is 1_3, even though neither $1 \sim_\pi 3$ nor $1 \sim_\sigma 3$ holds.

(ii) For $\pi = \{\{1\},\{2,3\},\{4\},\{5,6\}\}$ and $\sigma = \{\{1,2\},\{3\},\{4,5\},\{6\}\}$ in $P(6)$, their join is 1_6, again without direct connections in the original partitions. Thus, for the equivalence to $p \sim_{\pi\vee\sigma} q$, a more complex (nevertheless non-inductive) relation such as $\exists p = p_1, p_2, ..., p_n = q | \forall i \in [n-1] p_i \sim_\pi p_{i+1}$ or $p_i \sim_\sigma p_{i+1}$ also does not work for arbitrary $NC(m)$. The key is one must take n arbitrary (i.e. there exists an n and a chain $p_1, ..., p_n$ such that $p_i \sim_\pi p_{i+1}$ or $p_i \sim_\sigma p_{i+1}$)

2.2 Convolutions of Non Crossing Partitions

Although our primary interest lies in $NC(n)$, the following concepts apply to arbitrary partially ordered sets. For this reason, we denote our poset generically by P.

[**Speicher Definition 3.3**] Let $P^{(2)} := \{(\pi,\sigma) \in P^2 \mid \pi \leq \sigma\}$ and let $F, G : P^{(2)} \to \mathbb{C}$ be complex-valued functions. Let $f : P \to \mathbb{C}$ be another complex-valued function.

(i) The *convolution* of F and G is defined as:

$$(F * G)(\pi,\sigma) := \sum_{\tau \in P,\, \pi \leq \tau \leq \sigma} F(\pi,\tau) G(\tau,\sigma)$$

(ii) The *convolution* of f and G is defined as:

$$(f * G)(\sigma) := \sum_{\tau \in P,\, \tau \leq \sigma} f(\tau) G(\tau,\sigma)$$

(iii) The *delta function* $\delta(\pi, \sigma) := \delta_{\pi,\sigma}$ serves as the unit in this convolution algebra $(\mathbb{C}^{P^{(2)}}, *)$, called the incidence algebra.

One may think of (ii) as a special case of (i) by considering $F : P^{(2)} \to \mathbb{C}$ such that $\forall \sigma \in P, F(0, \sigma) = f(\sigma)$. Additionally, one may think of the convolution as summing over the "interval" $[\pi, \sigma] := \{\tau \in P | \pi \leq \tau \leq \sigma\}$. Finally, it is easy to check that the incidence algebra is associative but, in general, not commutative.

A particularly important function in the incidence algebra is the zeta function $\zeta(\pi, \sigma) := 1$ for all $\pi \leq \sigma$. Its inverse is of fundamental importance:

[**Speicher Proposition 3.5**] For a finite poset P, there exists a unique inverse ζ^{-1} of ζ, called the Möbius function and denoted by μ, such that $\zeta * \mu = \mu * \zeta = \delta$.

The proof proceeds by recursively solving for μ using the definition of convolution, and then by viewing these functions as matrix transformations. A crucial corollary is:

[**Speicher Corollary 3.6**] Let P be a finite poset with Möbius function μ. For any functions $f, g : P \to \mathbb{C}$, the following are equivalent:

(i) $f = g * \zeta$, i.e., $f(\sigma) = \sum_{\pi \leq \sigma} g(\pi)$ for all $\sigma \in P$

(ii) $g = f * \mu$, i.e., $g(\sigma) = \sum_{\pi \leq \sigma} f(\pi)\mu(\pi, \sigma)$ for all $\sigma \in P$

This Möbius inversion formula will play a crucial role in connecting combinatorial structures with moment calculations in free probability. In the next section we make this connection and calculate some values of the Möbius function.

3 Multiplicative Functionals and Moments

Let us denote $\bigcup_{n \in \mathbb{N}} NC(n)$ by NC.

[**Speicher Definition 3.8**]

(i) Let $(p_n)_{n \in \mathbb{N}}$ be a sequence of n-linear functionals on A^n. We extend this family to $(p_\pi)_{\pi \in NC}$ via:

$$p_\pi(a_1, \ldots, a_n) := \prod_{V \in \pi} p_{\#V}(a_1, \ldots, a_n|_V)$$

where if $V = \{i_1, \ldots, i_s\}$ with $i_1 < \cdots < i_s$, then $p_s(a_1, \ldots, a_n|_V) := p_s(a_{i_1}, \ldots, a_{i_s})$. We call $(p_\pi)_{\pi \in NC}$ the family of *multiplicative functionals* on NC determined by $(p_n)_{n \in \mathbb{N}}$.

(ii) Let $\phi_n(a_1, \ldots, a_n) := \phi(a_1 \cdots a_n)$. For all $\pi \in NC$, let ϕ_π be the extended functional given by (i).

(iii) For fixed (a_i), consider $\phi_{(\cdot)}(a_1, \ldots, a_n) : NC(n) \to \mathbb{C}$ as a function on $NC(n)$. We define $\kappa_{(\cdot)}(a_1, \ldots, a_n) := \phi_{(\cdot)}(a_1, \ldots, a_n) * \mu$, where μ is the Möbius function. The resulting κ_π defined on NC are called *free cumulants.*

From the Möbius inversion formula (inverted), we immediately obtain:

$$\phi(a_1 \cdots a_n) = \phi_n(a_1, \ldots, a_n) = \sum_{\pi \in NC(n)} \kappa_\pi(a_1, \ldots, a_n) =$$

$$\sum_{\pi \in NC(n)} \prod_{V \in \pi} \kappa_{\#V}(a_1, \ldots, a_n|_V)$$

This shows that ϕ_n and κ_n determine each other uniquely. Note:

[**Speicher Proposition 3.11**] For a non-commutative probability space (A, φ) with free cumulants $(\kappa_\pi)_{\pi \in NC}$:

(i) For all $n \in \mathbb{N}$ and $\pi \in NC(n)$, $\kappa_\pi : A^n \to \mathbb{C}$ is n-linear.

(ii) The family $(\kappa_\pi)_{\pi \in NC}$ is a family of multiplicative functionals on NC determined by $(\kappa_n)_{n \in \mathbb{N}}$.

The following fundamental theorem justifies our combinatorial digression and highlights the power of free cumulants:

[**Mingo Speicher 2.2 Theorem 14**] In a non-commutative probability space (A, φ), elements $(a_i)_{i \in I}$ are free if and only if all mixed cumulants vanish. That is, for all $n \in \mathbb{N}$ and indices $i(1), \ldots, i(n) \in I$, if there exist $k \neq l$ with $i(k) \neq i(l)$, then:

$$\kappa_n(a_{i(1)}, \ldots, a_{i(n)}) = 0$$

Of course this theorem can be extended to $*$-probability spaces once one notes that a, b are $*$-free $\iff$ a^*, b are. This theorem reveals one of the great advantages of working with cumulants: their additivity under "free convolution" (additivity under sums of free variables). If a and b are free, then:

$$\kappa_n(a + b, \ldots, a + b) = \kappa_n(a, \ldots, a) + \kappa_n(b, \ldots, b)$$

since all mixed terms vanish by the theorem.

However, free cumulants also have drawbacks. While moments are associative:

$$\phi_{n-1}(a_1, \ldots, a_i a_{i+1}, \ldots, a_n) = \phi(a_1 \cdots a_i a_{i+1} \cdots a_n) = \phi_{n-1}(a_1, \ldots, a_{i-1} a_i, \ldots, a_n)$$

cumulants do not share this property, as the discussion after the following example demonstrates.

[**Speicher Example 3.14**] Let (A, φ) be a non-commutative probability space. Using the the corollary on Möbius inversion of the last section, we calculate some κ_n and obtain some values for μ along the way. Note below that we use $\varphi_n, \varphi_{1_n}, \varphi_{\{[n]\}}$ interchangeably (similarly for $\kappa, 0_n$ etc.).

(i) $\varphi_1(a_1) = \sum_{\pi \leq 1_1} \kappa_\pi(a_1)\zeta(\pi, 1_1) = \kappa_1(a_1)$ yields $\kappa_1(a_1) = \varphi_1(a_1) = \sum_{\pi \leq 1_1} \kappa_\pi(a_1)\mu(\pi, 1_1) = \kappa_1(a_1)mu(1_1, 1_1)$, which yields $\mu(1_1, 1_1) = 1$. Actually, by μ's (and ζ's) definition, we get that for any $\pi \in NC$ $\mu(\pi, \pi) = 1$ $(1 = \delta(\pi, \pi) = (\mu * \zeta)(\pi, \pi) = \sum_{\pi \leq \sigma \leq \pi} \mu(\pi, \sigma)\zeta(\sigma, \pi) = \mu(\pi, \pi)\sigma(\pi, \pi) = \mu(\pi, \pi))$.

(ii) $\varphi_2(a_1, a_2) = \kappa_{0_2}(a_1, a_2) + \kappa_{1_2}(a_1, a_2)$ and by (i)

$$\kappa_{0_2}(a_1, a_2) = \kappa_1(a_1)\kappa_1(a_2) = \varphi_1(a_1)\varphi_1(a_2).$$

Thus, by rearranging and using $\kappa = \varphi * \mu$ we obtain $\varphi_2(a_1, a_2) - \varphi_1(a_1)\varphi_1(a_2) = \kappa_2(a_1, a_2) = \varphi_2(a_1, a_2)\mu(1_2, 1_2) + \varphi_{0_2}(a_1, a_2)\mu(0_2, 1_2)$ wherefrom we deduce $\mu(0_2, 1_2) = -1$.

(iii) $\varphi_3(a_1, a_2, a_3) = \kappa_3(a_1, a_2, a_3) + \kappa_{\{\{1,2\},3\}}(a_1, a_2, a_3) + \kappa_{\{\{1,3\},\{2\}\}}(a_1, a_2, a_3) + \kappa_{\{\{1\},\{2,3\}\}}(a_1, a_2, a_3) + \kappa_{0_n}(a_1, a_2, a_3)$, whereupon, similarly to above, rearranging and using (i), (ii), we obtain a formula for κ_3 as well as values of μ: $\mu(\{\{1,2\},\{3\}\}, 1_3) = \mu(\{\{1,3\},\{2\}\}, 1_3) = \mu(\{\{1\},\{2,3\}\}, 1_3) = 1$, $\mu(0_3, 1_3) = 2$.

Note that for values of $\mu(\sigma, \pi)$, where $\pi, \sigma \in NC(n), \pi \neq 1_n$ for some n, we can embed $\{\tau \in NC(n) | \tau \leq \pi\}$ via an order preserving lattice isomorphism into some $NC(m), m < n$, whereby π is mapped to 1_m. This observation allows us to conclude that, to determine μ, as a function on NC, it is sufficient to solve for $\mu(\cdot, 1_n)$ $\forall n \in \mathbb{N}$.

From the example above we can give an example of "non-associativity" of κ: $\kappa(a_1a_2, a_3) = \varphi_2(a_1a_2, a_3) - \varphi_1(a_1a_2)\varphi_1(a_3) = \varphi(a_1a_2a_3) - \varphi(a_1a_2)\varphi(a_3)$, which is not equal to $\varphi(a_1a_2a_3) - \varphi(a_1)\varphi(a_2a_3)$ (take, for instance, $\varphi(a_1) \neq 0 = \varphi(a_3), a_2 = a_3 \neq 0$ so that $\varphi(a_2a_3) = \varphi(a_2^2) \neq 0$).

Free cumulants satisfy the following, more complicated, combinatorial rule instead of associativity:

[**Mingo Speicher 2.2 Theorem 13**] For a non-commutative probability space (A, φ) with free cumulants $(\kappa_\pi)_{\pi \in NC}$, let $i : [m] \to [n]$ satisfy $1 \leq$

$i(1) < \cdots < i(m) = n$ and define:

$$\begin{aligned} A_1 &:= a_1 \cdots a_{i(1)} \\ A_2 &:= a_{i(1)+1} \cdots a_{i(2)} \\ &\vdots \\ A_m &:= a_{i(m-1)+1} \cdots a_{i(m)} \end{aligned}$$

Then for $\tau \in NC(m)$:

$$\kappa_\tau(A_1, \dots, A_m) = \sum_{\substack{\pi \in NC(n) \\ \pi \vee \hat{0}_m = \hat{\tau}}} \kappa_\pi(a_1, \dots, a_n)$$

Here, $\hat{\sigma}$ denotes the embedded partition of $\sigma \in NC(m)$ in $NC(n)$. Specifically, $\hat{0}_m$ is the partition:

$$\hat{0}_m = \{\{a_1, \dots, a_{i(1)}\}, \{a_{i(1)+1}, \dots, a_{i(2)}\}, \dots, \{a_{i(m-1)+1}, \dots, a_{i(m)}\}\}$$

In general, one obtains the embedding of $\sigma \in NC(m)$ by treating the blocks of $\hat{0}_m$ as elements of a set of length m and grouping them according to σ. For instance, for $m = 4$, $n = 7$, with $A_1 = a_1a_2$, $A_2 = a_3$, $A_3 = a_4a_5a_6$, $A_4 = a_7$, and $\sigma = \{\{1\}, \{2, 4\}, \{3\}\}$, we have:

$$\hat{\sigma} = \{\{1, 2\}, \{3, 7\}, \{4, 5, 6\}\}$$

4 Analytization

Thus far, we have developed a combinatorial framework for understanding the relationship between moments and cumulants. However, this approach can be unwieldy for practical calculations, motivating us to "analytize" the setting by introducing analytic transforms. We need the following facts, which can be proven without too much trouble:

[Speicher Facts 4.2]

(i) Let (A, φ) be a $*$-probability space and x self-adjoint. If the moments of x are exponentially bounded, i.e., there exists $M > 0$ such that for all $n \in \mathbb{N}$, $|\varphi(x^n)| \leq M^n$, then there exists a unique $\mu \in \mathbb{R}$ (a probability measure on $\mathbb{R}$) such that $\varphi(x^n) = \int t^n d\mu$ and μ is supported on $B_M(0)$. In this case, we may refer to μ as the distribution of x.

(ii) For any compactly supported measure $\mu \in \mathcal{P}(\mathbb{R})$, there exists a $*$-probability space and a self-adjoint element x within it such that $\varphi(x^n) = \int t^n d\mu$.

From now on, as in the context of the facts above, we assume appearing non commutative probability spaces are $*$-probability spaces and measures are in $\mathcal{P}(\mathbb{R})$ and have compact support. For a given $a \in A$ in a probability space (A, φ), let $m_n := \varphi_n(a, \ldots, a) = \varphi(a^n)$ (the moments of a) and $\kappa_n := \kappa_n(a, \ldots, a)$ (the cumulants of a). Note that compactly supported measures are uniquely determined via their moments, which are, by definition, uniquely determined via corresponding cumulants. Thus, a compactly supported measure is uniquely determined via its free cumulants.

[**Definition**] The following are definitions as formal series (for the time being).

(i) The *moment series* of a is defined as:

$$M(z) := 1 + \sum_{n \in \mathbb{N}} m_n z^n$$

(ii) The *cumulant series* of a is defined as:

$$C(z) := 1 + \sum_{n \in \mathbb{N}} \kappa_n z^n$$

In order to "analytize" our current framework of formal series, we consider the Cauchy transform, which we review below (those familiar may skip the italicized part below).

*[**Mingo Speicher 3.1 Definition 1**] Let $\mu \in \mathcal{P}(\mathbb{R})$ (many results hold for Borel measures as well).*

(i) The Cauchy transform $G_\mu : \mathbb{C} \setminus \mathbb{R} \to \mathbb{C}$ is defined as follows:

$$G_\mu(z) := \int_{\mathbb{R}} \frac{1}{z - t} \mu(dt).$$

(ii) we denote the upper, resp. lower, half planes by $\mathbb{C}^+, \mathbb{C}^-$, respectively.

(iii) we denote the set of probability measures over a set X with $\mathcal{P}(X)$.

(iv) One often encounters the Stieltjes transform S_μ instead of the Cauchy transform. It is just $-G_\mu$.

We collect some properties of the Cauchy transform:

*[**Mingo Speicher 3.1 Lemma 2**] G_μ is analytic on $\mathbb{C}^+$ with range in $\mathbb{C}^-$.*

We recall the Stieltjes Inversion Formula below, which "inverts" the Cauchy transform to recover the corresponding measure:

*[**Mingo Speicher 3.1 Theorem 6**] Let $\nu \in P(\mathbb{R})$. For arbitrary $a < b$, we have:*

$$\nu((a, b)) + \frac{\nu(\{a, b\})}{2} = -\lim_{y \to 0^+} \frac{1}{\pi} \int_a^b Im(G(x + iy))dx.$$

We note, finally, that for compactly supported μ, the Cauchy transform admits a series expansion about infinity:

[Speicher Proposition 4.10] *For $\mu \in \mathcal{P}(\mathbb{R})$ supported on $[-r, r]$, and for all $|z| > r$, $z \in \mathbb{C}^+$, we have:*

$$G_\mu(z) = \sum_{n \in \mathbb{N}_0} \frac{m_n}{z^{n+1}}$$

where m_n denotes the n-th moment of μ.

The relation $M(1/z)/z = G_\mu(z)$ together with the analyticity of G_μ establishes convergence on some open domain for $M(1/z)/z$, giving us a convergent series (i.e. we understand $M(1/z)/z$ as more than just a formal series). Moreover, we obtain the following fundamental connection:

[Speicher Theorem 3.26] For a $*$-probability space (A, φ) and $a \in A$ such that $a = a^*$, the following are equivalent:

(i) $\kappa * \zeta = m$, i.e., for all $n \in \mathbb{N}$:

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

(ii) $C(zM(z)) = M(z)$

(iii) $M(z/C(z)) = C(z)$

where the last two points represent equalities of formal power series.

Using these formal equalities and the analytic identity $M(1/z)/z = G_\mu(z)$, we obtain analytic equalities on appropriate domains (abbreviating G_μ as G):

$$C(G(z)) = C(M(1/z)/z) = M(z) = zG(z)$$

Defining $K(z) := C(z)/z$, we obtain $K(G(z)) = z$ and thus $G(K(z)) = z$ (analytically on some domain). Furthermore, K has one simple pole at 0, allowing us to write:

$$K(z) = \frac{1}{z} + R(z)$$

where $R(z) := \sum_{n \in \mathbb{N}} \kappa_n z^{n-1}$ is called the *R-transform*. Please see Mingo Speicher 3.3 Theorem 17 for a proper analytic justification of the above-stated.

We want to demonstrate a nice consequence of this analytization, namely, calculation of certain "free convolutions". That is, whereas the classical convolution describes the distribution of the sum of two classically independent random variables, the free convolution describes that of the sum of two free random variables. We obtain existence in the case of compactly supported measures by the following theorem:

[**Speicher Theorem 4.5**] For $\mu, \nu \in \mathcal{P}(\mathbb{R})$ compactly supported, let (A, φ) contain $x, y \in A$ such that $x \sim \mu$, $y \sim \nu$, and x and y are free. Then the distribution of $x + y$ corresponds to a unique compactly supported measure $\mu \boxplus \nu$, called the *free convolution* of μ and ν. The measure $\mu \boxplus \nu$ is independent of the particular realizations x and y.

Note that the analytic series relations obtained above, together with the Stieltjes Inversion Formula yield a recipe for the free convolution of two (compactly supported) measures μ, ν. This can be viewed as a nice application of the R-transform:

1. Calculate the Cauchy transforms G_μ, G_ν
2. Solve for K_μ, K_ν (using $G_{(\cdot)}(K_{(\cdot)}(z)) = z$)
3. Solve for R_μ, R_ν (using $K_{(\cdot)}(z) = 1/z + R_{(\cdot)}(z)$)
4. Additivity of cumulants yields $R_{\mu\boxplus\nu} = R_\mu + R_\nu$
5. Reverse the process to obtain $G_{\mu\boxplus\nu}$
6. Apply the Stieltjes Inversion Formula

[**Zeitouni Anderson Guionnet Exercise 5.3.27**] Using the recipe above, we will calculate $\mu \boxplus \nu$ for $\mu = \nu = \frac{\delta_{-1}+\delta_1}{2}$. First, the Cauchy transform can be evaluated as $G(z) = \frac{z}{z^2-1}$. Using step 2 we obtain $K(z) = \frac{1\pm\sqrt{1+4z^2}}{2z}$. Subtracting $1/z$ yields $R(z) = \frac{\sqrt{1+4z^2}-1}{2z}$ (note $R(0) = \kappa_1 = \mu_1 = \int x d\mu = 1/2 - 1/2 = 0$ allows us to reject $R(z) = \frac{-\sqrt{1+4z^2}-1}{2z}$). We multiply by two to obtain $R_{\mu\boxplus\nu}$, add $1/z$ to obtain $K_{\mu\boxplus\nu}(z)$, and use $z = K_{\mu\boxplus\nu}(G_{\mu\boxplus\nu}(z))$ to conclude $G_{\mu\boxplus\nu}(z) = (z^2-4)^{-\frac{1}{2}}$. We can apply the Stieltjes Inversion Formula to see that $\mu \boxplus \nu$ is a "semicircular distribution" (actually the standard semicircular distribution, characterized by having variance 1) with density:

$$\begin{cases} \frac{1}{\pi\sqrt{4-t^2}} & |t| \leq 2 \\ 0 & |t| > 2 \end{cases}$$

The semicircular distribution is the free analog of the normal distribution. Indeed, one can prove a "free central limit theorem", in which one shows that the distribution of the sum of n free, identically distributed, centered variables divided by the familiar normalizing constant $\sqrt{n}$ converges to a semicircular distribution of variance equal to that of the individual variables. See Mingo Speicher 2.1.2 Theorem 5 (as well as Definition 4 for the specification of semicircular distributions of given variance).

In particular, from the example above, we observe that the free convolution of two discrete distributions yields a continuous one. If we "translate" this statement into the classical setting—there exist independent discrete random variables (finite range), whose sum yields a continuous random variable distributed according to the standard Gaussian (uncountable range)—we immediately see the impossibility in the classical case. This can be viewed as another instance of the stark difference between commutative and non-commutative settings.

We finish with Voiculescu's characterization of the R-transform using Nevanlinna-Pick theory:

*[**Pick Interpolation Theorem**] Let $\xi_1, \dots, \xi_n \in \mathbb{C}^+$ be distinct and $\xi'_1, \dots, \xi'_n \in \mathbb{C}^+$. The following are equivalent:*

(i) The matrix B with entries $B_{ij} = \frac{\xi'_i - \bar{\xi}'_j}{\xi_i - \bar{\xi}_j}$

(ii) There is a Pick function (holomorphic function on $\mathbb{C}^+$ with image in closure$(\mathbb{C}^+)$) mapping ξ_i to ξ'_i.

[**Voiculescu Lemma 3.3**] Let $f(z) = \sum_{n \in \mathbb{N}_0} a_{n+1} z^n$. The existence of a compactly supported measure $\mu \in \mathcal{P}(\mathbb{R})$ with corresponding R-transform equal to f is equivalent to all of the following three conditions holding true:

(i) There exists $C > 0$ such that f converges on $B_C(0)$.

(ii) $f(\bar{\cdot}) = \overline{f(\cdot)}$.

(iii) There exists $C_1 \geq 1/C$ such that for all $m \in \mathbb{N}, z_1, \dots, z_m \in \{z \mid |z| \geq C_1, \mathrm{Im}(z) > 0\}$ and $\xi_1, \dots, \xi_m \in \mathbb{C}$:

$$\sum_{j,k \in [m]} \frac{z_j - \bar{z}_k}{z_j - \bar{z}_k + f(z_j^{-1}) - f(\bar{z}_k^{-1})} \xi_j \bar{\xi}_k \geq 0$$

Note that in Voiculescu's paper, $K(z)$ is defined as $z + R(1/z)$ instead of $1/z + R(z)$ and $H(z)$ as (Voiculescu's) $K(z)$'s inverse (which here is the inverse of $K(1/z)$, i.e. $1/G(z)$).

Proof: $(\Rightarrow)$: Since, by Speicher Proposition 4.10 above, $G_\mu =: G$ has a convergent power series expansion about infinity with the moments of μ as coefficients (which are real), we have $G(\bar{\cdot}) = \bar{G}(\cdot)$. Furthermore, using the convolution sum relation between cumulants and moments, it could be an exercise to prove that the moments of a measure are exponentially bounded if and only if the corresponding cumulants are, i.e. $\exists C > 0 : |m_n| \leq C^n \iff \exists C > 0 : |\kappa_n| \leq C^n$ (see also Speicher Proposition 4.4). Using compact support of μ, we have exponential boundedness of μ's moments (take i.e. $C := \sup_{x \in supp(\mu)} |x|$ to get $\int_{supp(\mu)} |t|^n d\mu \leq \mu(supp(\mu)) C^n = C^n$), implying that of κ_n, allowing us to conclude convergence in a region about 0 for f (the coefficients of $f = R_\mu$ are exactly the κ_n). Moreover, by the moment-cumulant convolution relation, we easily see that the κ_n are real and thus, $f(\bar{\cdot}) = \bar{f}(\cdot)$.

Since G maps the upper half plane to the lower half plane, $1/G$ maps the upper half plane to itself, from which we may conclude the following by the Pick interpolation theorem: $\forall n \in \mathbb{N}, \omega_1, \ldots, \omega_n \in \mathbb{C}^+$, the matrix B with entries:

$$B_{ij} = \frac{1/G(\omega_i) - 1/\bar{G}(\omega_j)}{\omega_i - \bar{\omega}_j}$$

is positive semi-definite. Taking $z_i := 1/G(\omega_i)$ and noting that the inverse of $1/G(z)$ is $z + R_\mu(1/z) = z + f(1/z)$, we obtain:

$$B_{ij} = \frac{z_i - \bar{z}_j}{z_i - \bar{z}_j + f(1/z_i) - \bar{f}(1/z_j)},$$

yielding (ii).
($\Leftarrow$): We use the opposite direction of the Pick interpolation theorem to obtain a Pick function $H(z)$ agreeing with $1/G(z)$ in a neighborhood. We conclude that H is $1/G$'s analytic continuation to $\mathbb{C}^+$. We thus obtain holomorphicity (and thus an analytic expansion about infinity) of G in a neighborhood of infinity, $\bar{G} = G(\bar{\cdot})$ and $Ran(G) \subset \mathbb{C}^-$, meaning G is the Cauchy transform of a measure μ with moments being the coefficients of G's expansion. ■

It should be noted that at the time of Voiculescu's paper, Roland Speicher had not yet developed free cumulants. In fact, Voiculescu takes an operator theoretic approach to define the R-transform, using Toeplitz creation and annihilation operators instead of the more combinatorial approach used here.

References

[1] Voiculescu, D. V. (1986). *Addition of certain non-commuting random variables*. Journal of Functional Analysis, 66(3), 323-346. DOI: https://doi.org/10.1016/0022-1236(86)90062-5

[2] Mingo, J. A., & Speicher, R. (2017). *Free Probability and Random Matrices*. Fields Institute Monographs, Vol. 35, Springer. URL: https://mast.queensu.ca/ mingo/mingo_speicher_2017.pdf

[3] Anderson, G. W., Guionnet, A., & Zeitouni, O. (2010). *An Introduction to Random Matrices*. Cambridge University Press. URL: https://www.wisdom.weizmann.ac.il/ zeitouni/cupbook.pdf

[4] Speicher, R. (2019). *Lecture Notes on Free Probability Theory*. arXiv:1908.08125.