

Chapter 2

The Free Central Limit Theorem and Free Cumulants

Recall from Chapter 1 that if (\mathcal{A}, φ) is a non-commutative probability space and $\mathcal{A}_1, \dots, \mathcal{A}_s$ are subalgebras of \mathcal{A} which are free with respect to φ , then freeness gives us in principle a rule by which we can evaluate $\varphi(a_1 a_2 \cdots a_k)$ for any alternating word in random variables a_1, a_2, \dots, a_k . Thus we can in principle calculate all mixed moments for a system of free random variables. However, we do not yet have any concrete idea of the structure of this factorization rule. This situation will be greatly clarified by the introduction of *free cumulants*. Classical cumulants appeared in Chapter 1, where we saw that they are intimately connected with the combinatorial notion of set partitions. Our free cumulants will be linked in a similar way to the lattice of non-crossing set partitions; the latter were introduced in combinatorics by Kreweras [113]. We will motivate the appearance of free cumulants and non-crossing partition lattices in free probability theory by examining in detail a proof of the central limit theorem by the method of moments.

The combinatorial approach to free probability was initiated by Speicher in [159, 161], in order to get alternative proofs for the free central limit theorem and the main properties of the R -transform, which had been treated before by Voiculescu in [176, 177] by more analytic tools. Nica showed a bit later in [135] how this combinatorial approach connects in general to Voiculescu's operator-theoretic approach in terms of creation and annihilation operators on the full Fock space. The combinatorial path was pursued much further by Nica and Speicher; for more details on this, we refer to the standard reference [137].

2.1 The classical and free central limit theorems

Our setting is that of a non-commutative probability space (\mathcal{A}, φ) and a sequence $(a_i)_{i \in \mathbb{N}} \subset \mathcal{A}$ of centred and identically distributed random variables. This means that $\varphi(a_i) = 0$ for all $i \geq 1$ and that $\varphi(a_i^n) = \varphi(a_j^n)$ for any $i, j, n \geq 1$. We assume that our random variables $a_i, i \geq 1$ are either classically independent or freely independent as defined in Chapter 1. Either form of independence gives us a *factorization rule* for calculating mixed moments in the random variables.

For $k \geq 1$, set

$$S_k := \frac{1}{\sqrt{k}}(a_1 + \cdots + a_k). \quad (2.1)$$

The Central Limit Theorem is a statement about the limit distribution of the random variable S_k in the large k limit. Let us begin by reviewing the kind of convergence we shall be considering.

Recall that given a real-valued random variable X on a probability space, we have a probability measure μ_X on \mathbb{R} , called the *distribution* of X . The distribution of X is defined by the equation

$$E(f(X)) = \int f(t) d\mu_X(t) \text{ for all } f \in C_b(\mathbb{R}) \quad (2.2)$$

where $C_b(\mathbb{R})$ is the C^* -algebra of all bounded continuous functions on \mathbb{R} . We say that a probability measure μ on \mathbb{R} is *determined by its moments* if μ has moments $\{\alpha_k\}_k$ of all orders and μ is the only probability measure on \mathbb{R} with moments $\{\alpha_k\}_k$. If the moment generating function of μ has a positive radius of convergence, then μ is determined by its moments (see Billingsley [41, Theorem 30.1]).

Exercise 1. Show that a compactly supported measure is determined by its moments.

A more general criterion is the Carleman condition (see Akhiezer [3, p. 85]) which says that a measure μ is determined by its moments $\{\alpha_k\}_k$ if we have $\sum_{k \geq 1} (\alpha_{2k})^{-1/(2k)} = \infty$.

Exercise 2. Using the Carleman condition, show that the Gaussian measure is determined by its moments.

A sequence of probability measures $\{\mu_n\}_n$ on \mathbb{R} is said to converge *weakly* to μ if $\{\int f d\mu_n\}_n$ converges to $\int f d\mu$ for all $f \in C_b(\mathbb{R})$. Given a sequence $\{X_n\}_n$ of real-valued random variables, we say that $\{X_n\}_n$ *converges in distribution* (or *converges in law*) if the probability measures $\{\mu_{X_n}\}_n$ converge weakly.

If we are working in a non-commutative probability space (\mathcal{A}, φ) , we call an element a of \mathcal{A} a *non-commutative random variable*. Given such an a , we may define μ_a by $\int p d\mu_a = \varphi(p(a))$ for all polynomials $p \in \mathbb{C}[x]$. At this level of generality, we may not be able to define $\int f d\mu_a$ for all functions $f \in C_b(\mathbb{R})$, so we call the linear functional $\mu_a : \mathbb{C}[x] \rightarrow \mathbb{C}$ the *algebraic distribution* of a , even if it is not a probability measure. However when it is clear from the context we shall just call μ_a the distribution of a . Note that if a is a self-adjoint element of a C^* -algebra and φ is positive and has norm 1, then μ_a extends from $\mathbb{C}[x]$ to $C_b(\mathbb{R})$ and thus μ_a becomes a probability measure on \mathbb{R} .

Definition 1. Let $(\mathcal{A}_k, \varphi_k)$, for $k \in \mathbb{N}$, and (\mathcal{A}, φ) be non-commutative probability spaces.

- 1) Let $(b_k)_{k \in \mathbb{N}}$ be a sequence of non-commutative random variables with $b_k \in \mathcal{A}_k$, and let $b \in \mathcal{A}$. We say that b_k converges in distribution to b , denoted by $b_k \xrightarrow{\text{distr}} b$, if

$$\lim_{k \rightarrow \infty} \varphi_k(b_k^n) = \varphi(b^n) \quad (2.3)$$

for any fixed $n \in \mathbb{N}$.

- 2) More generally, let I be an index set. For each $i \in I$, let $b_k^{(i)} \in \mathcal{A}_k$ for $k \in \mathbb{N}$ and $b^{(i)} \in \mathcal{A}$. We say that $(b_k^{(i)})_{i \in I}$ converges in distribution to $(b^{(i)})_{i \in I}$, denoted by $(b_k^{(i)})_{i \in I} \xrightarrow{\text{distr}} (b^{(i)})_{i \in I}$, if

$$\lim_{k \rightarrow \infty} \varphi_k(b_k^{(i_1)} \cdots b_k^{(i_n)}) = \varphi(b^{(i_1)} \cdots b^{(i_n)}) \quad (2.4)$$

for all $n \in \mathbb{N}$ and all $i_1, \dots, i_n \in I$.

Note that this definition is neither weaker nor stronger than weak convergence of the corresponding distributions. For real-valued random variables, the convergence in (2.3) is sometimes called convergence in moments. However there is an important case where the two conditions coincide. If we have a sequence of probability measures $\{\mu_k\}_k$ on \mathbb{R} , each having moments of all orders and a probability measure μ determined by its moments, such that for every n we have $\int t^n d\mu_k(t) \rightarrow \int t^n d\mu$ as $k \rightarrow \infty$, then $\{\mu_k\}_k$ converges weakly to μ (see Billingsley [41, Theorem 30.2]). To see that weak convergence does not imply convergence in moments, consider the sequence $\{\mu_k\}_k$ where $\mu_k = (1 - 1/k)\delta_0 + (1/k)\delta_k$ and δ_k is the probability measure with an atom at k of mass 1.

Exercise 3. Show that $\{\mu_k\}_k$ converges weakly to δ_0 but that we do not have convergence in moments.

We want to make a statement about convergence in distribution of the random variables $(S_k)_{k \in \mathbb{N}}$ from (2.1) (which all come from the same underlying non-commutative probability space). Thus we need to do a moment calculation. Let $[k] = \{1, \dots, k\}$ and $[n] = \{1, \dots, n\}$. We have

$$\varphi(S_k^n) = \frac{1}{k^{n/2}} \sum_{r: [n] \rightarrow [k]} \varphi(a_{r_1} \cdots a_{r_n}).$$

It turns out that the fact that the random variables a_1, \dots, a_k are independent and identically distributed makes the task of calculating this sum less complex than it initially appears. The key observation is that because of (classical or free) independence of the a_i 's and the fact that they are identically distributed, the value of $\varphi(a_{r_1} \cdots a_{r_n})$ depends not on all details of the multi-index r , but just on the

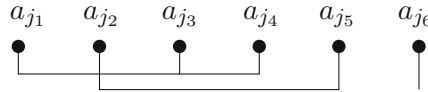


Fig. 2.1 Suppose $j_1 = j_3 = j_4$ and $j_2 = j_5$ but $\{j_1, j_2, j_6\}$ are distinct. Then $\ker(j) = \{(1, 3, 4), (2, 5), (6)\}$

information where the indices are the same and where they are different. Let us recall some notation from the proof of Theorem 1.1.

Notation 2. Let $i = (i_1, \dots, i_n)$ be a multi-index. Then its kernel, denoted by $\ker i$, is that partition in $\mathcal{P}(n)$ whose blocks correspond exactly to the different values of the indices (Fig. 2.1),

$$k \text{ and } l \text{ are in the same block of } \ker i \iff i_k = i_l.$$

Lemma 3. With this notation we have that $\ker i = \ker j$ implies $\varphi(a_{i_1} \cdots a_{i_n}) = \varphi(a_{j_1} \cdots a_{j_n})$.

Proof: To see this note first that $\ker i = \ker j$ implies that the i -indices can be obtained from the j -indices by the application of some permutation σ , i.e. $(j_1, \dots, j_n) = (\sigma(i_1), \dots, \sigma(i_n))$. We know that the random variables a_1, \dots, a_k are (classically or freely) independent. This means that we have a factorization rule for calculating mixed moments in a_1, \dots, a_k in terms of the moments of individual a_i 's. In particular this means that $\varphi(a_{i_1} \cdots a_{i_n})$ can be written as some expression in moments $\varphi(a_i^r)$, while $\varphi(a_{j_1} \cdots a_{j_n})$ can be written as that same expression except with $\varphi(a_i^r)$ replaced by $\varphi(a_{\sigma(i)}^r)$. However, since our random variables all have the same distribution, then $\varphi(a_i^r) = \varphi(a_{\sigma(i)}^r)$ for any i, j , and thus $\varphi(a_{i_1} \cdots a_{i_n}) = \varphi(a_{j_1} \cdots a_{j_n})$. \square

Let us denote the common value of $\varphi(a_{i_1} \cdots a_{i_n})$ for all i with $\ker i = \pi$, for some $\pi \in \mathcal{P}(n)$, by $\varphi(\pi)$. Consequently, we have

$$\varphi(S_k^n) = \frac{1}{k^{n/2}} \sum_{\pi \in \mathcal{P}(n)} \varphi(\pi) \cdot |\{i : [n] \rightarrow [k] \mid \ker i = \pi\}|.$$

It is not difficult to see that

$$\#\{i : [n] \rightarrow [k] \mid \ker i = \pi\} = k(k - 1) \cdots (k - \#\pi + 1)$$

because we have k choices for the first block of π , $k - 1$ choices for the second block of π , and so on until the last block where we have $k - \#\pi + 1$.

Then what we have proved is that

$$\varphi(S_k^n) = \frac{1}{k^{n/2}} \sum_{\pi \in \mathcal{P}(n)} \varphi(\pi) \cdot k(k-1) \cdots (k - \#\pi + 1).$$

The great advantage of this expression over what we started with is that the number of terms does not depend on k . Thus we are in a position to take the limit as $k \rightarrow \infty$, provided we can effectively estimate each term of the sum.

Our first observation is the most obvious one, namely, we have

$$k(k-1) \cdots (k - \#\pi + 1) \sim k^{\#\pi} \quad \text{as } k \rightarrow \infty.$$

Next observe that if π has a block of size 1, then we will have $\varphi(\pi) = 0$. Indeed suppose that $\pi = \{V_1, \dots, V_m, \dots, V_s\} \in \mathcal{P}(n)$ with $V_m = \{l\}$ for some $l \in [n]$. Then we will have

$$\varphi(\pi) = \varphi(a_{j_1} \cdots a_{j_{l-1}} a_{j_l} a_{j_{l+1}} \cdots a_{j_n})$$

where $\ker(j) = \pi$ and thus $j_l \notin \{j_1, \dots, j_{l-1}, j_{l+1}, \dots, j_n\}$. Hence we can write $\varphi(\pi) = \varphi(ba_{j_l}c)$, where $b = a_{j_1} \cdots a_{j_{l-1}}$ and $c = a_{j_{l+1}} \cdots a_{j_n}$ and thus

$$\varphi(\pi) = \varphi(ba_{j_l}c) = \varphi(a_{j_l})\varphi(bc) = 0,$$

since a_{j_l} is (classically or freely) independent of $\{b, c\}$. (For the free case, this factorization was considered in Equation (1.13) in the last chapter. In the classical case, it is obvious, too.) Of course, for this part of the argument, it is crucial that we assume our variables a_i to be centred.

Thus the only partitions which contribute to the sum are those with blocks of size at least 2. Note that such a partition can have at most $n/2$ blocks. Now,

$$\lim_{k \rightarrow \infty} \frac{k^{\#\pi}}{k^{n/2}} = \begin{cases} 1, & \text{if } \#\pi = n/2 \\ 0, & \text{if } \#\pi < n/2 \end{cases}.$$

Hence the only partitions which contribute to the sum in the $k \rightarrow \infty$ limit are those with *exactly* $n/2$ blocks, i.e. partitions each of whose blocks has size 2. Such partitions are called *pairings*, and the set of pairings is denoted $\mathcal{P}_2(n)$.

Thus we have shown that

$$\lim_{k \rightarrow \infty} \varphi(S_k^n) = \sum_{\pi \in \mathcal{P}_2(n)} \varphi(\pi).$$

Note that in particular if n is odd, then $\mathcal{P}_2(n) = \emptyset$, so that the odd limiting moments vanish. In order to determine the even limiting moments, we must distinguish between the setting of classical independence and free independence.

2.1.1 Classical central limit theorem

In the case of classical independence, our random variables commute and factorize completely with respect to φ . Thus if we denote by $\varphi(a_i^2) = \sigma^2$ the common variance of our random variables, then for any pairing $\pi \in \mathcal{P}_2(n)$ we have $\varphi(\pi) = \sigma^n$. Thus we have

$$\lim_{k \rightarrow \infty} \varphi(S_k^n) = \sum_{\pi \in \mathcal{P}_2(n)} \sigma^n = \begin{cases} \sigma^n (n-1)(n-3) \dots 5 \cdot 3 \cdot 1, & \text{if } n \text{ even} \\ 0, & \text{if } n \text{ odd} \end{cases}.$$

From Section 1.1, we recognize these as exactly the moments of a Gaussian random variable of mean 0 and variance σ^2 . Since by Exercise 2 the normal distribution is determined by its moments, and hence our convergence in moments is the same as the classical convergence in distribution, we get the following form of the classical central limit theorem: if $(a_i)_{i \in \mathbb{N}}$ are classically independent random variables which are identically distributed with $\varphi(a_i) = 0$ and $\varphi(a_i^2) = \sigma^2$, and having all moments, then S_k converges in distribution to a Gaussian random variable with mean 0 and variance σ^2 . Note that one can see the derivation above also as a proof of the Wick formula for Gaussian random variables if one takes the central limit theorem for granted.

2.1.2 Free central limit theorem

Now we want to deal with the case where the random variables are freely independent. In this case, $\varphi(\pi)$ will not be the same for all pair partitions $\pi \in \mathcal{P}_2(2n)$ (we focus on the even moments now because we already know that the odd ones are zero). Let's take a look at some examples:

$$\begin{aligned} \varphi(\{(1, 2), (3, 4)\}) &= \varphi(a_1 a_1 a_2 a_2) = \varphi(a_1^2) \varphi(a_2^2) = \sigma^4 \\ \varphi(\{(1, 4), (2, 3)\}) &= \varphi(a_1 a_2 a_2 a_1) = \varphi(a_1^2) \varphi(a_2^2) = \sigma^4 \\ \varphi(\{(1, 3), (2, 4)\}) &= \varphi(a_1 a_2 a_1 a_2) = 0. \end{aligned}$$

The last equality is just from the definition of freeness, because $a_1 a_2 a_1 a_2$ is an alternating product of centred free variables.

In general, we will get $\varphi(\pi) = \sigma^{2n}$ if we can successively remove neighbouring pairs of identical random variables in the word corresponding to π so that we end with a single pair (see Fig. 2.2); if we cannot we will have $\varphi(\pi) = 0$ as in the example $\varphi(a_1 a_2 a_1 a_2) = 0$ above. Thus the only partitions that give a non-zero contribution are the *non-crossing* ones (see [137, p. 122] for details). Non-crossing pairings were encountered already in Chapter 1, where we denoted the set of non-crossing pairings by $NC_2(2n)$. Then we have as our free central limit theorem that

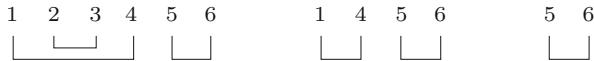


Fig. 2.2 We start with the pairing $\{(1, 4), (2, 3), (5, 6)\}$ and remove the pair $(2, 3)$ of adjacent elements (middle figure). Next we remove the pair $(1, 4)$ of adjacent elements. We are then left with a single pair; so the pairing must have been non-crossing to start with

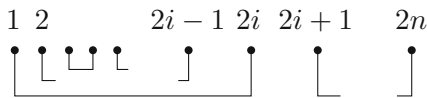


Fig. 2.3 We have C_{i-1} possible pairings on $[2, 2i - 1]$ and C_{n-i} possible pairings on $[2i + 1, 2n]$

$$\lim_{k \rightarrow \infty} \varphi(S_k^{2n}) = \sigma^{2n} \cdot |NC_2(2n)|.$$

In Chapter 1 we already mentioned that the cardinality $C_n := |NC_2(2n)|$ is given by the *Catalan numbers*. We want now to elaborate on the proof of this claim.

A very simple method is to show that the pairings are in a bijective correspondence with the Dyck paths; by using André’s reflection principle, one finds that there are $\binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1} \binom{2n}{n}$ such paths (see [137, Prop. 2.11] for details).

Our second method for counting non-crossing pairings is to find a simple recurrence which they satisfy. The idea is to look at the block of a pairing which contains the number 1. In order for the pairing to be non-crossing, 1 must be paired with some even number in the set $[2n]$, else we would necessarily have a crossing. Thus 1 must be paired with $2i$ for some $i \in [n]$. Now let i run through all possible values in $[n]$, and count for each the number of non-crossing pairings that contain this pair, as in the diagram (Fig. 2.3).

In this way we see that the cardinality C_n of $NC_2(2n)$ must satisfy the recurrence relation

$$C_n = \sum_{i=1}^n C_{i-1} C_{n-i}, \tag{2.5}$$

with initial condition $C_0 = 1$. One can then check using a generating function that the Catalan numbers satisfy this recurrence; hence $C_n = \frac{1}{n+1} \binom{2n}{n}$.

Exercise 4. Let $f(z) = \sum_{n=0}^{\infty} C_n z^n$ be the generating function for $\{C_n\}_n$, where $C_0 = 1$ and C_n satisfies the recursion (2.5).

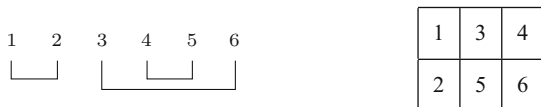


Fig. 2.4 In the bijection between $NC_2(6)$ and 2×3 standard Young tableaux, the pairing $\{(1, 2), (3, 6), (4, 5)\}$ gets mapped to the tableaux on the right

- (i) Show that $1 + zf(z)^2 = f(z)$.
- (ii) Show that f is also the power series for $\frac{1-\sqrt{1-4z}}{2z}$.
- (iii) Show that $C_n = \frac{1}{n+1} \binom{2n}{n}$.

We can also prove directly that $C_n = \frac{1}{n+1} \binom{2n}{n}$ by finding a bijection between $NC_2(2n)$ and some standard set of objects which we can see directly is enumerated by the Catalan numbers. A reasonable choice for this “canonical” set is the collection of $2 \times n$ standard Young tableaux. A standard Young tableaux of shape $2 \times n$ is a filling of the squares of a $2 \times n$ grid with the numbers $1, \dots, 2n$ which is strictly increasing in each of the two rows and each of the n columns. The number of these standard Young tableaux is very easy to calculate, using a famous and fundamental result known as the *hook-length formula* [167, Vol. 2, Corollary 7.21.6]. The hook-length formula tells us that the number of standard Young tableaux on the $2 \times n$ rectangle is

$$\frac{(2n)!}{(n+1)!n!} = \frac{1}{n+1} \binom{2n}{n}. \tag{2.6}$$

Thus we will have proved that $|NC_2(2n)| = \frac{1}{n+1} \binom{2n}{n}$ if we can bijectively associate to each pair partition $\pi \in NC_2(2n)$ a standard Young tableaux on the $2 \times n$ rectangular grid. This is very easy to do. Simply take the “left-halves” of each pair in π and write them in increasing order in the cells of the first row. Then take the “right-halves” of each pair of π and write them in increasing order in the cells of the second row. Figure 2.4 shows the bijection between $NC_2(6)$ and standard Young tableaux on the 2×3 rectangle.

Definition 4. A self-adjoint random variable s with odd moments $\varphi(s^{2n+1}) = 0$ and even moments $\varphi(s^{2n}) = \sigma^{2n} C_n$, where C_n is the n -th Catalan number and $\sigma > 0$ is a constant, is called a *semi-circular element* of variance σ^2 . In the case $\sigma = 1$, we call it the *standard semi-circular element*.

The argument we have just provided gives us the *free central limit theorem*.

Theorem 5. If $(a_i)_{i \in \mathbb{N}}$ are self-adjoint, freely independent, and identically distributed with $\varphi(a_i) = 0$ and $\varphi(a_i^2) = \sigma^2$, then S_k converges in distribution to a semi-circular element of variance σ^2 as $k \rightarrow \infty$.

This free central limit theorem was proved as one of the first results in free probability theory by Voiculescu already in [176]. His proof was much more operator theoretic; the proof presented here is due to Speicher [159] and was the first hint at a relation between free probability theory and the combinatorics of non-crossing partitions. (An early concrete version of the free central limit theorem, before the notion of freeness was isolated, appeared also in the work of Bożejko [43] in the context of convolution operators on free groups.)

Recall that in Chapter 1 it was shown that for a random matrix X_N chosen from $N \times N$ GUE we have that

$$\lim_{N \rightarrow \infty} E[\text{tr}(X_N^n)] = \begin{cases} 0, & \text{if } n \text{ odd} \\ C_{n/2}, & \text{if } n \text{ even} \end{cases} \quad (2.7)$$

so that a GUE random matrix is a semi-circular element in the limit of large matrix size, $X_N \xrightarrow{\text{distr}} s$.

We can also define a family of semi-circular random variables.

Definition 6. Suppose (\mathcal{A}, φ) is a $*$ -probability space. A self-adjoint family $(s_i)_{i \in I} \subset \mathcal{A}$ is called a *semi-circular family* of covariance $C = (c_{ij})_{i,j \in I}$ if $C \geq 0$ and for any $n \geq 1$ and any n -tuple $i_1, \dots, i_n \in I$ we have

$$\varphi(s_{i_1} \cdots s_{i_n}) = \sum_{\pi \in NC_2(n)} \varphi_\pi[s_{i_1}, \dots, s_{i_n}],$$

where

$$\varphi_\pi[s_{i_1}, \dots, s_{i_n}] = \prod_{(p,q) \in \pi} c_{i_p i_q}.$$

If C is diagonal, then $(s_i)_{i \in I}$ is a free semi-circular family.

This is the free analogue of Wick's formula. In fact, using this language and our definition of convergence in distribution from Definition 1, it follows directly from Lemma 1.9 that if X_1, \dots, X_r are matrices chosen independently from GUE, then, in the large N limit, they converge in distribution to a semi-circular family s_1, \dots, s_r of covariance $c_{ij} = \delta_{ij}$.

Exercise 5. Show that if $\{x_1, \dots, x_n\}$ is a semi-circular family and $A = (a_{ij})$ is an invertible matrix with real entries, then $\{y_1, \dots, y_n\}$ is a semi-circular family where $y_i = \sum_j a_{ij} x_j$.

Exercise 6. Let $\{x_1, \dots, x_n\}$ be a semi-circular family such that for all i and j we have $\varphi(x_i x_j) = \varphi(x_j x_i)$. Show that by diagonalizing the covariance matrix we can find an orthogonal matrix $O = (o_{ij})$ such that $\{y_1, \dots, y_n\}$ is a free semi-circular family where $y_i = \sum_j o_{ij} x_j$.

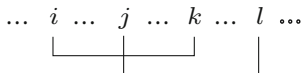


Fig. 2.5 A crossing in a partition

Exercise 7. Formulate and prove a multidimensional version of the free central limit theorem.

2.2 Non-crossing partitions and free cumulants

We begin by recalling some relevant definitions concerning non-crossing partitions from Section 1.8.

Definition 7. A partition $\pi \in \mathcal{P}(n)$ is called *non-crossing* if there do not exist numbers $i, j, k, l \in [n]$ with $i < j < k < l$ such that i and k are in the same block of π and j and l are in the same block of π , but i and j are not in the same block of π . The collection of all non-crossing partitions of $[n]$ was denoted $NC(n)$.

Figure 2.5 should make it clear what a *crossing* in a partition is; a non-crossing partition is a partition with no crossings.

Note that $\mathcal{P}(n)$ is partially ordered by

$$\pi_1 \leq \pi_2 \iff \text{each block of } \pi_1 \text{ is contained in a block of } \pi_2. \tag{2.8}$$

We also say that π_1 is a *refinement* of π_2 . $NC(n)$ is a subset of $\mathcal{P}(n)$ and inherits this partial order, so $NC(n)$ is an induced sub-poset of $\mathcal{P}(n)$. In fact both are lattices; they have well-defined *join* \vee and *meet* \wedge operations (though the join of two non-crossing partitions in $NC(n)$ does not necessarily agree with their join when viewed as elements of $\mathcal{P}(n)$). Recall that the join $\pi_1 \vee \pi_2$ in a lattice is the smallest σ with the property that $\sigma \geq \pi_1$ and $\sigma \geq \pi_2$ and that the meet $\pi_1 \wedge \pi_2$ is the largest σ with the property that $\sigma \leq \pi_1$ and $\sigma \leq \pi_2$.

We now define the important *free cumulants* of a non-commutative probability space (\mathcal{A}, φ) . They were introduced by Speicher in [161]. For other notions of cumulants and the relation between them, see [11, 74, 117, 153].

Definition 8. Let (\mathcal{A}, φ) be a non-commutative probability space. The corresponding *free cumulants* $\kappa_n : \mathcal{A}^n \rightarrow \mathbb{C}$ ($n \geq 1$) are defined inductively in terms of moments by the *moment-cumulant formula*

$$\varphi(a_1 \cdots a_n) = \sum_{\pi \in NC(n)} \kappa_\pi(a_1, \dots, a_n), \tag{2.9}$$

where, by definition, if $\pi = \{V_1, \dots, V_r\}$, then

$$\kappa_\pi(a_1, \dots, a_n) = \prod_{\substack{V \in \pi \\ V = (i_1, \dots, i_l)}} \kappa_l(a_{i_1}, \dots, a_{i_l}). \quad (2.10)$$

Remark 9. In Equation (2.10) and below, we always mean that the elements i_1, \dots, i_l of V are in increasing order. Note that Equation (2.9) has a formulation using Möbius inversion which we might call the *cumulant-moment formula*. To present this we need the moment version of Equation (2.10). For a partition $\pi \in \mathcal{P}(n)$ with $\pi = \{V_1, \dots, V_r\}$, we set

$$\varphi_\pi(a_1, \dots, a_n) = \prod_{\substack{V \in \pi \\ V = (i_1, \dots, i_l)}} \varphi(a_{i_1} \cdots a_{i_l}). \quad (2.11)$$

We also need the Möbius function μ for $NC(n)$ (see [137, Lecture 10]). Then our cumulant-moment relation can be written

$$\kappa_n(a_1, \dots, a_n) = \sum_{\pi \in NC(n)} \mu(\pi, 1_n) \varphi_\pi(a_1, \dots, a_n). \quad (2.12)$$

One could use Equation (2.12) as the definition of free cumulants; however for practical calculations Equation (2.9) is usually easier to work with.

Example 10. (1) For $n = 1$, we have $\varphi(a_1) = \kappa_1(a_1)$, and thus

$$\kappa_1(a_1) = \varphi(a_1). \quad (2.13)$$

(2) For $n = 2$, we have

$$\varphi(a_1 a_2) = \kappa_{\{(1,2)\}}(a_1, a_2) + \kappa_{\{(1),(2)\}}(a_1, a_2) = \kappa_2(a_1, a_2) + \kappa_1(a_1)\kappa_1(a_2).$$

Since we know from the $n = 1$ calculation that $\kappa_1(a_1) = \varphi(a_1)$, this yields

$$\kappa_2(a_1, a_2) = \varphi(a_1 a_2) - \varphi(a_1)\varphi(a_2). \quad (2.14)$$

(3) For $n = 3$, we have

$$\begin{aligned} \varphi(a_1 a_2 a_3) &= \kappa_{\{(1,2,3)\}}(a_1, a_2, a_3) + \kappa_{\{(1,2),(3)\}}(a_1, a_2, a_3) + \kappa_{\{(1),(2,3)\}}(a_1, a_2, a_3) \\ &\quad + \kappa_{\{(1,3),(2)\}}(a_1, a_2, a_3) + \kappa_{\{(1),(2),(3)\}}(a_1, a_2, a_3) \\ &= \kappa_3(a_1, a_2, a_3) + \kappa_2(a_1, a_2)\kappa_1(a_3) + \kappa_2(a_2, a_3)\kappa_1(a_1) \\ &\quad + \kappa_2(a_1, a_3)\kappa_1(a_2) + \kappa_1(a_1)\kappa_1(a_2)\kappa_1(a_3). \end{aligned}$$

Thus we find that

$$\begin{aligned} \kappa_3(a_1, a_2, a_3) &= \varphi(a_1 a_2 a_3) - \varphi(a_1) \varphi(a_2 a_3) \\ &\quad - \varphi(a_2) \varphi(a_1 a_3) - \varphi(a_3) \varphi(a_1 a_2) + 2\varphi(a_1) \varphi(a_2) \varphi(a_3). \end{aligned} \quad (2.15)$$

These three examples outline the general procedure of recursively defining κ_n in terms of the mixed moments. It is easy to see that κ_n is an n -linear function.

Exercise 8. (i) Show the following: if φ is a trace, then the cumulant κ_n is, for each $n \in \mathbb{N}$, invariant under cyclic permutations, i.e. for all $a_1, \dots, a_n \in \mathcal{A}$, we have

$$\kappa_n(a_1, a_2, \dots, a_n) = \kappa_n(a_2, \dots, a_n, a_1).$$

(ii) Let us assume that all moments with respect to φ are invariant under all permutations of the entries, i.e. that we have for all $n \in \mathbb{N}$ and all $a_1, \dots, a_n \in \mathcal{A}$ and all $\sigma \in S_n$ that $\varphi(a_{\sigma(1)} \cdots a_{\sigma(n)}) = \varphi(a_1 \cdots a_n)$. Is it then true that also the free cumulants κ_n ($n \in \mathbb{N}$) are invariant under all permutations?

Let us also point out how the definition appears when $a_1 = \cdots = a_n = a$, i.e. when all the random variables are the same. Then we have

$$\varphi(a^n) = \sum_{\pi \in NC(n)} \kappa_\pi(a, \dots, a).$$

Thus if we write $\alpha_n^a := \varphi(a^n)$ and $\kappa_\pi^a := \kappa_\pi(a, \dots, a)$, this reads

$$\alpha_n^a = \sum_{\pi \in NC(n)} \kappa_\pi^a. \quad (2.16)$$

Note the similarity to Equation (1.3) for classical cumulants.

Since the Catalan number is the number of non-crossing pairings of $[2n]$ as well as the number of non-crossing partitions of $[n]$, we can use Equation (2.16) to show that the cumulants of the standard semi-circle law are all 0 except $\kappa_2 = 1$.

Exercise 9. Use Equation (2.16) to show that for the standard semi-circle law all cumulants are 0, except κ_2 which equals 1.

As another demonstration of the simplifying power of the moment-cumulant formula (2.16), let us use the formula to find a simple expression for the moments and free cumulants of the *Marchenko-Pastur law*. This is a probability measure on $\mathbb{R}^+ \cup \{0\}$ that is as fundamental as the semi-circle law (see Section 4.5). Let $0 < c < \infty$ be a positive real number. For each c we shall construct a probability measure ν_c . Set $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$. For $c \geq 1$, ν_c has as support the interval $[a, b]$ and the density $\sqrt{(b-x)(x-a)}/(2\pi x)$; that is

$$d\nu_c(x) = \frac{\sqrt{(b-x)(x-a)}}{2\pi x} dx.$$

For $0 < c < 1$, ν_c has the same density on $[a, b]$ and in addition has an atom at 0 of mass $1 - c$; thus

$$d\nu_c(x) = (1 - c)\delta_0 + \frac{\sqrt{(b-x)(x-a)}}{2\pi x} dx.$$

Note that when $c = 1$, $a = 0$ and the density has a “pole” of order 1/2 at 0 and thus is still integrable.

Exercise 10. In this exercise we shall show that ν_c is a probability measure for all c . Let $R = -x^2 + (a + b)x - ab$, and then write

$$\frac{\sqrt{R}}{x} = \frac{R}{x\sqrt{R}} = \frac{1}{2} \frac{-2x + (a + b)}{\sqrt{R}} + \frac{1}{2} \frac{a + b}{\sqrt{R}} - \frac{ab}{x\sqrt{R}}.$$

- (i) Show that the integral of the first term on $[a, b]$ is 0.
- (ii) Using the substitution $t = (x - (1 + c))/\sqrt{c}$, show that the integral of the second term over $[a, b]$ is $\pi(a + b)/2$.
- (iii) Let $u = (b - a)/(2ab)$, $v = (b + a)/(2ab)$ and $t = u^{-1}(v - x^{-1})$. With this substitution show that the integral of the third term over $[a, b]$ is $-\pi\sqrt{ab}$.
- (iv) Using the first three parts, show that ν_c is a probability measure.

Definition 11. The *Marchenko-Pastur distribution* is the law with distribution ν_c with $0 < c < \infty$. We shall see in Exercise 11 that all free cumulants of ν_c are equal to c . By analogy with the classical cumulants of the Poisson distribution, ν_c is also called the *free Poisson law* (of rate c). We should also note that we have chosen a different normalization than that used by other authors in order to make the cumulants simple; see Remark 12 and Exercise 12 below.

Exercise 11. In this exercise we shall find the moments and free cumulants of the Marchenko-Pastur law.

- (i) Let α_n be the n^{th} moment. Use the substitution $t = (x - (1 + c))/\sqrt{c}$ to show that

$$\alpha_n = \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \frac{1}{k+1} \binom{n-1}{2k} \binom{2k}{k} (1+c)^{n-2k-1} c^{1+k}.$$

- (ii) Expand the expression $(1 + c)^{n-2k-1}$ to obtain that

$$\alpha_n = \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \sum_{l=k}^{n-k-1} \frac{(n-1)!}{k!(k+1)!(l-k)!(n-k-l-1)!} c^{l+1}.$$

(iii) Interchange the order of summation and use Vandermonde convolution ([79, (5.23)]) to show that

$$\alpha_n = \sum_{l=1}^n \frac{c^l}{n} \binom{n}{l-1} \binom{n}{l}.$$

(iv) Finally use the fact ([137, Cor. 9.13]) that $\frac{1}{n} \binom{n}{l-1} \binom{n}{l}$ is the number of non-crossing partitions of $[n]$ with l blocks to show that

$$\alpha_n = \sum_{\pi \in NC(n)} c^{\#(\pi)}.$$

Use this formula to show that $\kappa_n = c$ for all $n \geq 1$.

Remark 12. Given $y > 0$, let $a' = (1 - \sqrt{y})^2$ and $b' = (1 + \sqrt{y})^2$. Let ρ_y be the probability measure on \mathbb{R} given by $\sqrt{(b' - t)(t - a')}/(2\pi y t) dt$ on $[a', b']$ when $y \leq 1$ and $(1 - y^{-1})\delta_0 + \sqrt{(b' - t)(t - a')}/(2\pi y t) dt$ on $\{0\} \cup [a', b']$ when $y > 1$. As above δ_0 is the Dirac mass at 0. This might be called the standard form of the Marchenko-Pastur law. In the exercise below, we shall see that ρ_y is related to ν_c in a simple way and the cumulants of ρ_y are not as simple as those of ν_c .

Exercise 12. Show that by setting $c = 1/y$ and making the substitution $t = x/c$ we have

$$\int x^k d\nu_c(x) = c^k \int t^k d\rho_y(t).$$

Show that the free cumulants of ρ_y are given by $\kappa_n = c^{1-n}$.

There is a combinatorial formula by Krawczyk and Speicher [111] for expanding cumulants whose arguments are products of random variables. For example, consider the expansion of $\kappa_2(a_1 a_2, a_3)$. This can be written as

$$\kappa_2(a_1 a_2, a_3) = \kappa_3(a_1, a_2, a_3) + \kappa_1(a_1)\kappa_2(a_2, a_3) + \kappa_2(a_1, a_3)\kappa_1(a_2). \quad (2.17)$$

A more complicated example is given by:

$$\begin{aligned} & \kappa_2(a_1 a_2, a_3 a_4) \\ &= \kappa_4(a_1, a_2, a_3, a_4) + \kappa_1(a_1)\kappa_3(a_2, a_3, a_4) + \kappa_1(a_2)\kappa_3(a_1, a_3, a_4) \\ & \quad + \kappa_1(a_3)\kappa_3(a_1, a_2, a_4) + \kappa_1(a_4)\kappa_3(a_1, a_2, a_3) + \kappa_2(a_1, a_4)\kappa_2(a_2, a_3) \\ & \quad + \kappa_2(a_1, a_3)\kappa_1(a_2)\kappa_1(a_4) + \kappa_2(a_1, a_4)\kappa_1(a_2)\kappa_1(a_3) \\ & \quad + \kappa_1(a_1)\kappa_2(a_2, a_3)\kappa_1(a_4) + \kappa_1(a_1)\kappa_2(a_2, a_4)\kappa_1(a_3). \end{aligned} \quad (2.18)$$

In general, the evaluation of a free cumulant with products of entries involves summing over all π which have the property that they connect all different product strings. Here is the precise formulation, for the proof we refer to [137, Theorem 11.12]. Note that this is the free counter part of the formula (1.16) for classical cumulants.

Theorem 13. *Suppose n_1, \dots, n_r are positive integers and $n = n_1 + \dots + n_r$. Consider a non-commutative probability space (\mathcal{A}, φ) and $a_1, a_2, \dots, a_n \in \mathcal{A}$. Let*

$$A_1 = a_1 \cdots a_{n_1}, \quad A_2 = a_{n_1+1} \cdots a_{n_1+n_2}, \quad \dots, \quad A_r = a_{n_1+\dots+n_{r-1}+1} \cdots a_n.$$

Then

$$\kappa_r(A_1, \dots, A_r) = \sum_{\substack{\pi \in NC(n) \\ \pi \vee \tau = 1_n}} \kappa_\pi(a_1, \dots, a_n) \quad (2.19)$$

where the summation is over those $\pi \in NC(n)$ which connect the blocks corresponding to A_1, \dots, A_r . More precisely, this means that $\pi \vee \tau = 1_n$ where

$$\tau = \{(1, \dots, n_1), (n_1 + 1, \dots, n_1 + n_2), \dots, (n_1 + \dots + n_{r-1} + 1, \dots, n)\}$$

and $1_n = \{(1, 2, \dots, n)\}$ is the partition with only one block.

Exercise 13. (i) Let $\tau = \{(1, 2), (3)\}$. List all $\pi \in NC(3)$ such that $\pi \vee \tau = 1_3$. Check that these are exactly the terms appearing on the right-hand side of Equation (2.17).

(ii) Let $\tau = \{(1, 2), (3, 4)\}$. List all $\pi \in NC(4)$ such that $\pi \vee \tau = 1_4$. Check that these are exactly the terms on the right-hand side of Equation (2.18).

The most important property of free cumulants is that we may characterize free independence by the vanishing of “mixed” cumulants. Let (\mathcal{A}, φ) be a non-commutative probability space and $\mathcal{A}_1, \dots, \mathcal{A}_s \subset \mathcal{A}$ unital subalgebras. A cumulant $\kappa_n(a_1, a_2, \dots, a_n)$ is *mixed* if each a_i is in one of the subalgebras, but a_1, a_2, \dots, a_n do not all come from the same subalgebra.

Theorem 14. *The subalgebras $\mathcal{A}_1, \dots, \mathcal{A}_s$ are free if and only if all mixed cumulants vanish.*

The proof of this theorem relies on formula (2.19) and on the following proposition which is a special case of Theorem 14. For the details of the proof of Theorem 14, we refer again to [137, Theorem 11.15].

Proposition 15. *Let (\mathcal{A}, φ) be a non-commutative probability space and let κ_n , $n \geq 1$ be the corresponding free cumulants. For $n \geq 2$, $\kappa_n(a_1, \dots, a_n) = 0$ if $1 \in \{a_1, \dots, a_n\}$.*

Proof: We consider the case where the last argument a_n is equal to 1 and proceed by induction on n .

For $n = 2$,

$$\kappa_2(a, 1) = \varphi(a1) - \varphi(a)\varphi(1) = 0.$$

So the base step is done.

Now assume for the induction hypothesis that the result is true for all $1 \leq k < n$.

We have that

$$\begin{aligned} \varphi(a_1 \cdots a_{n-1} 1) &= \sum_{\pi \in NC(n)} \kappa_{\pi}(a_1, \dots, a_{n-1}, 1) \\ &= \kappa_n(a_1, \dots, a_{n-1}, 1) + \sum_{\substack{\pi \in NC(n) \\ \pi \neq 1_n}} \kappa_{\pi}(a_1, \dots, a_{n-1}, 1). \end{aligned}$$

According to our induction hypothesis, a partition $\pi \neq 1_n$ can have $\kappa_{\pi}(a_1, \dots, a_{n-1}, 1)$ different from zero only if (n) is a one-element block of π , i.e. $\pi = \sigma \cup \{(n)\}$ for some $\sigma \in NC(n-1)$. For such a partition, we have

$$\kappa_{\pi}(a_1, \dots, a_{n-1}, 1) = \kappa_{\sigma}(a_1, \dots, a_{n-1})\kappa_1(1) = \kappa_{\sigma}(a_1, \dots, a_{n-1}),$$

hence

$$\begin{aligned} \varphi(a_1 \cdots a_{n-1} 1) &= \kappa_n(a_1, \dots, a_{n-1}, 1) + \sum_{\sigma \in NC(n-1)} \kappa_{\sigma}(a_1, \dots, a_{n-1}) \\ &= \kappa_n(a_1, \dots, a_{n-1}, 1) + \varphi(a_1 \cdots a_{n-1}). \end{aligned}$$

Since $\varphi(a_1 \cdots a_{n-1} 1) = \varphi(a_1 \cdots a_{n-1})$, we have proved that $\kappa_n(a_1, \dots, a_{n-1}, 1) = 0$. \square

Whereas Theorem 14 gives a useful characterization for the freeness of subalgebras, its direct application to the case of random variables would not yield a satisfying characterization in terms of the vanishing of mixed cumulants in the subalgebras generated by the variables. By invoking again the product formula for free cumulants, Theorem 13, it is quite straightforward to get the following much more useful characterization in terms of mixed cumulants of the variables.

Theorem 16. *Let (\mathcal{A}, φ) be a non-commutative probability space. The random variables a_1, \dots, a_s in \mathcal{A} are free if and only if all mixed cumulants of the a_1, \dots, a_s vanish. That is, a_1, \dots, a_s are free if and only if whenever we choose $i_1, \dots, i_n \in \{1, \dots, s\}$ in such a way that $i_k \neq i_l$ for some $k, l \in [n]$, then $\kappa_n(a_{i_1}, \dots, a_{i_n}) = 0$.*

2.3 Products of free random variables

We want to understand better the calculation rule for mixed moments of free variables. Thus we will now derive the basic combinatorial description for such mixed moments.

Let $\{a_1, \dots, a_r\}$ and $\{b_1, \dots, b_r\}$ be free random variables, and consider

$$\varphi(a_1 b_1 a_2 b_2 \cdots a_r b_r) = \sum_{\pi \in NC(2r)} \kappa_{\pi}(a_1, b_1, a_2, b_2, \dots, a_r, b_r).$$

Since the a 's are free from the b 's, we only need to sum over those partitions π which do not connect the a 's with the b 's. Each such partition may be written as $\pi = \pi_a \cup \pi_b$, where π_a denotes the blocks consisting of a 's and π_b the blocks consisting of b 's. Hence by the definition of free cumulants

$$\begin{aligned} \varphi(a_1 b_1 a_2 b_2 \cdots a_r b_r) &= \sum_{\pi_a \cup \pi_b \in NC(2r)} \kappa_{\pi_a}(a_1, \dots, a_r) \cdot \kappa_{\pi_b}(b_1, \dots, b_r) \\ &= \sum_{\pi_a \in NC(r)} \kappa_{\pi_a}(a_1, \dots, a_r) \cdot \left(\sum_{\substack{\pi_b \in NC(r) \\ \pi_a \cup \pi_b \in NC(2r)}} \kappa_{\pi_b}(b_1, \dots, b_r) \right). \end{aligned}$$

It is now easy to see that, for a given $\pi_a \in NC(r)$, there exists a biggest $\sigma \in NC(r)$ with the property that $\pi_a \cup \sigma \in NC(2r)$. This σ is called the *Kreweras complement* of π_a and is denoted by $K(\pi_a)$; see [137, Def. 9.21]. This $K(\pi_a)$ is given by connecting as many b 's as possible in a non-crossing way without getting crossings with the blocks of π_a . The mapping K is an order-reversing bijection on the lattice $NC(r)$.

But then the summation condition on the internal sum above is equivalent to the condition $\pi_b \leq K(\pi_a)$. Summing κ_{π} over all $\pi \in NC(r)$ gives the corresponding r -th moment, which extends easily to

$$\sum_{\substack{\pi \in NC(r) \\ \pi \leq \sigma}} \kappa_{\pi}(b_1, \dots, b_r) = \varphi_{\sigma}(b_1, \dots, b_r),$$

where φ_{σ} denotes, in the same way as in κ_{π} , the product of moments along the blocks of σ ; see Equation (2.11).

Thus we get as the final conclusion of our calculations that

$$\varphi(a_1 b_1 a_2 b_2 \cdots a_r b_r) = \sum_{\pi \in NC(r)} \kappa_{\pi}(a_1, \dots, a_r) \cdot \varphi_{K(\pi)}(b_1, \dots, b_r). \quad (2.20)$$

Let us consider some simple examples for this formula. For $r = 1$, there is only one $\pi \in NC(1)$, which is its own complement, and we get

$$\varphi(a_1 b_1) = \kappa_1(a_1) \varphi(b_1).$$

As $\kappa_1 = \varphi$, this gives the usual factorization formula

$$\varphi(a_1 b_1) = \varphi(a_1) \varphi(b_1).$$

For $r = 2$, there are two elements in $NC(2)$, \sqcup and \sqcap , and we have

$$K(\sqcup) = \sqcap \quad \text{and} \quad K(\sqcap) = \sqcup$$

and the formula above gives

$$\varphi(a_1 b_1 a_2 b_2) = \kappa_2(a_1, a_2) \varphi(b_1) \varphi(b_2) + \kappa_1(a_1) \kappa_1(a_2) \varphi(b_1 b_2).$$

With $\kappa_1(a) = \varphi(a)$ and $\kappa_2(a_1, a_2) = \varphi(a_1 a_2) - \varphi(a_1) \varphi(a_2)$, this reproduces formula (1.14).

The formula above is not symmetric between the a 's and the b 's (the former appear with cumulants, the latter with moments). Of course, one can also exchange the roles of a and b , in which case one ends up with

$$\varphi(a_1 b_1 a_2 b_2 \cdots a_r b_r) = \sum_{\pi \in NC(r)} \varphi_{K^{-1}(\pi)}(a_1, \dots, a_r) \cdot \kappa_{\pi}(b_1, \dots, b_r). \quad (2.21)$$

Note that K^2 is not the identity, but a cyclic rotation of π .

Formulas (2.20) and (2.21) are particularly useful when one of the sets of variables has simple cumulants, as is the case for semi-circular random variables $b_i = s$. Then only the second cumulants $\kappa_2(s, s) = 1$ are non-vanishing, i.e. in effect the sum is only over non-crossing pairings. Thus, if s is semi-circular and free from $\{a_1, \dots, a_r\}$, then we have

$$\varphi(a_1 s a_2 s \cdots a_r s) = \sum_{\pi \in NC_2(r)} \varphi_{K^{-1}(\pi)}(a_1, \dots, a_r). \quad (2.22)$$

Let us also note in passing that one can rewrite the Equations (2.20) and (2.21) above in the symmetric form (see [137, (14.4)])

$$\kappa_r(a_1 b_1, a_2 b_2, \dots, a_r b_r) = \sum_{\pi \in NC(r)} \kappa_{\pi}(a_1, \dots, a_r) \cdot \kappa_{K(\pi)}(b_1, \dots, b_r). \quad (2.23)$$

2.4 Functional relation between moment series and cumulant series

Notice how much more efficient the result on the description of freeness in terms of cumulants is in checking freeness of random variables than the original definition of free independence. In the cumulant framework, we can forget about centredness and weaken “alternating” to “mixed”. Also, the problem of adding two freely independent random variables becomes easy on the level of free cumulants. If $a, b \in (\mathcal{A}, \varphi)$ are free with respect to φ , then

$$\begin{aligned}\kappa_n^{a+b} &= \kappa_n(a + b, \dots, a + b) \\ &= \kappa_n(a, \dots, a) + \kappa_n(b, \dots, b) + (\text{mixed cumulants in } a, b) \\ &= \kappa_n^a + \kappa_n^b.\end{aligned}$$

Thus the problem of calculating moments is shifted to the relation between cumulants and moments. We already know that the moments are polynomials in the cumulants, according to the moment-cumulant formula (2.16), but we want to put this relationship into a framework more amenable to performing calculations.

For any $a \in \mathcal{A}$, let us consider formal power series in an indeterminate z defined by

$$\begin{aligned}M(z) &= 1 + \sum_{n=1}^{\infty} \alpha_n^a z^n, && \text{moment series of } a \\ C(z) &= 1 + \sum_{n=1}^{\infty} \kappa_n^a z^n, && \text{cumulant series of } a.\end{aligned}$$

We want to translate the moment-cumulant formula (2.16) into a statement about the relationship between the moment and cumulant series.

Proposition 17. *The relation between the moment series $M(z)$ and the cumulant series $C(z)$ of a random variable is given by*

$$M(z) = C(zM(z)). \tag{2.24}$$

Proof: The idea is to sum first over the possibilities for the block of π containing 1, as in the derivation of the recurrence for C_n . Suppose that the first block of π looks like $V = \{1, v_2, \dots, v_s\}$, where $1 < v_1 < \dots < v_s \leq n$. Then we build up the rest of the partition π out of smaller “nested” non-crossing partitions π_1, \dots, π_s with $\pi_1 \in NC(\{2, \dots, v_2 - 1\})$, $\pi_2 \in NC(\{v_2 + 1, \dots, v_3 - 1\})$, etc. Hence if we denote $i_1 = |\{2, \dots, v_2 - 1\}|$, $i_2 = |\{v_2 + 1, \dots, v_3 - 1\}|$, etc., then we have

$$\begin{aligned}
\alpha_n &= \sum_{s=1}^n \sum_{\substack{i_1, \dots, i_s \geq 0 \\ s+i_1+\dots+i_s=n}} \sum_{\pi = V \cup \pi_1 \cup \dots \cup \pi_s} \kappa_s \kappa_{\pi_1} \cdots \kappa_{\pi_s} \\
&= \sum_{s=1}^n \sum_{\substack{i_1, \dots, i_s \geq 0 \\ s+i_1+\dots+i_s=n}} \kappa_s \left(\sum_{\pi_1 \in NC(i_1)} \kappa_{\pi_1} \right) \cdots \left(\sum_{\pi_s \in NC(i_s)} \kappa_{\pi_s} \right) \\
&= \sum_{s=1}^n \sum_{\substack{i_1, \dots, i_s \geq 0 \\ s+i_1+\dots+i_s=n}} \kappa_s \alpha_{i_1} \cdots \alpha_{i_s}.
\end{aligned}$$

Thus we have

$$\begin{aligned}
1 + \sum_{n=1}^{\infty} \alpha_n z^n &= 1 + \sum_{n=1}^{\infty} \sum_{s=1}^n \sum_{\substack{i_1, \dots, i_s \geq 0 \\ s+i_1+\dots+i_s=n}} \kappa_s z^s \alpha_{i_1} z^{i_1} \cdots \alpha_{i_s} z^{i_s} \\
&= 1 + \sum_{s=1}^{\infty} \kappa_s z^s \left(\sum_{i=0}^{\infty} \alpha_i z^i \right)^s.
\end{aligned}$$

□

Now consider the *Cauchy transform* of a :

$$G(z) := \varphi\left(\frac{1}{z-a}\right) = \sum_{n=0}^{\infty} \frac{\varphi(a^n)}{z^{n+1}} = \frac{1}{z} M(1/z) \quad (2.25)$$

and the *R-transform* of a defined by

$$R(z) := \frac{C(z) - 1}{z} = \sum_{n=0}^{\infty} \kappa_{n+1}^a z^n. \quad (2.26)$$

Also put $K(z) = R(z) + \frac{1}{z} = \frac{C(z)}{z}$. Then we have the relations

$$K(G(z)) = \frac{1}{G(z)} C(G(z)) = \frac{1}{G(z)} C\left(\frac{1}{z} M\left(\frac{1}{z}\right)\right) = \frac{1}{G(z)} z G(z) = z.$$

Note that M and C are in $\mathbb{C}[[z]]$, the ring of formal power series in z , $G \in \mathbb{C}[[\frac{1}{z}]]$, and $K \in \mathbb{C}((z))$, the ring of formal Laurent series in z , i.e. $zK(z) \in \mathbb{C}[[z]]$. Thus $K \circ G \in \mathbb{C}((\frac{1}{z}))$ and $G \circ K \in \mathbb{C}[[z]]$. We then also have $G(K(z)) = z$.

Thus we recover the following theorem of Voiculescu, which is the main result on the *R-transform*. Voiculescu's original proof in [177] was much more operator theoretic. One should also note that this computational machinery for the *R-transform* was also found independently and about the same time by Woess [204, 205], Cartwright and Soardi [49], and McLaughlin [125], in a more restricted setting of random walks on free product of groups. Our presentation here is based on the approach of Speicher in [161].

Theorem 18. For a random variable a , let $G_a(z)$ be its Cauchy transform, and define its R -transform $R_a(z)$ by

$$G_a[R_a(z) + 1/z] = z. \quad (2.27)$$

Then, for a and b freely independent, we have

$$R_{a+b}(z) = R_a(z) + R_b(z). \quad (2.28)$$

Let us write, for a and b free, the above as:

$$z = G_{a+b}[R_{a+b}(z) + 1/z] = G_{a+b}[R_a(z) + R_b(z) + 1/z]. \quad (2.29)$$

If we now put $w := R_{a+b}(z) + 1/z$, then we have $z = G_{a+b}(w)$ and we can continue Equation (2.29) as:

$$G_{a+b}(w) = z = G_a[R_a(z) + 1/z] = G_a[w - R_b(z)] = G_a[w - R_b[G_{a+b}(w)]].$$

Thus we get the *subordination* functions ω_a and ω_b given by

$$\omega_a(z) = z - R_b[G_{a+b}(z)] \quad \text{and} \quad \omega_b(z) = z - R_a[G_{a+b}(z)]. \quad (2.30)$$

We have $\omega_a, \omega_b \in \mathbb{C}(\frac{1}{z})$, so $G_a \circ \omega_a \in \mathbb{C}[\frac{1}{z}]$. These satisfy the subordination relations

$$G_{a+b}(z) = G_a[\omega_a(z)] = G_b[\omega_b(z)]. \quad (2.31)$$

We say that G_{a+b} is subordinate to both G_a and G_b . The name comes from the theory of univalent functions; see [65, Ch. 6] for a general discussion.

Exercise 14. Show that $\omega_a(z) + \omega_b(z) - 1/G_a(\omega_a(z)) = z$.

Exercise 15. Suppose we have formal Laurent series $\omega_a(z)$ and $\omega_b(z)$ in $\frac{1}{z}$ such that

$$G_a(\omega_a(z)) = G_b(\omega_b(z)) \quad \text{and} \quad \omega_a(z) + \omega_b(z) - 1/G_a(\omega_a(z)) = z. \quad (2.32)$$

Let G be the formal power series $G(z) = G_a(\omega_a(z))$ and $R(z) = G^{(-1)}(z) - z^{-1}$. ($G^{(-1)}$ denotes here the inverse under composition of G .) By replacing z by $G^{(-1)}(z)$ in the second equation of (2.32), show that $R(z) = R_a(z) + R_b(z)$. These equations can thus be used to define the distribution of the sum of two free random variables.

At the moment these are identities on the level of formal power series. In the next chapter, we will elaborate on their interpretation as identities of analytic functions; see Theorem 3.43.

2.5 Subordination and the non-commutative derivative

One might wonder about the relevance of the subordination formulation in (2.31). Since it has become more and more evident that the subordination formulation of free convolution is in many cases preferable to the (equivalent) description in terms of the R -transform, we want to give here some idea why subordination is a very natural concept in the context of free probability. When subordination appeared in this context first in papers of Voiculescu [181] and Biane [34], it was more an ad hoc construction – its real nature was only revealed later in the paper [190] of Voiculescu, where he related it to the non-commutative version of the derivative operation.

We will now introduce the basics of this non-commutative derivative; as before in this chapter, we will ignore all analytic questions and just deal with formal power series. In Chapter 8 we will have more to say about the analytic properties of the non-commutative derivatives.

Let $\mathbb{C}\langle x \rangle$ be the algebra of polynomials in the variable x . Then we define the *non-commutative derivative* ∂_x as a linear mapping $\partial_x : \mathbb{C}\langle x \rangle \rightarrow \mathbb{C}\langle x \rangle \otimes \mathbb{C}\langle x \rangle$ by the requirements that it satisfies the Leibniz rule

$$\partial_x(qp) = \partial_x(q) \cdot 1 \otimes p + q \otimes 1 \cdot \partial_x(p)$$

and by

$$\partial_x 1 = 0, \quad \partial_x x = 1 \otimes 1.$$

This means that it is given more explicitly as the linear extension of

$$\partial_x x^n = \sum_{k=0}^{n-1} x^k \otimes x^{n-1-k}. \quad (2.33)$$

We can also (and will) extend this definition from polynomials to infinite formal power series.

Exercise 16. (i) Let, for some $z \in \mathbb{C}$ with $z \neq 0$, f be the formal power series

$$f(x) = \frac{1}{z-x} = \sum_{n=0}^{\infty} \frac{x^n}{z^{n+1}}.$$

Show that we have then $\partial_x f = f \otimes f$.

(ii) Let f be a formal power series in x with the property that $\partial_x f = f \otimes f$. Show that f must then be either zero or of the form $f(x) = 1/(z-x)$ for some $z \in \mathbb{C}$, with $z \neq 0$.

We will now consider polynomials and formal power series in two non-commuting variables x and y . In this context, we still have the notion of ∂_x

(and also of ∂_y), and now their character as “partial” derivatives becomes apparent. Namely, we define $\partial_x : \mathbb{C}\langle x, y \rangle \rightarrow \mathbb{C}\langle x, y \rangle \otimes \mathbb{C}\langle x, y \rangle$ by the requirements that it should be a derivation, i.e. satisfy the Leibniz rule, and by the prescriptions:

$$\partial_x x = 1 \otimes 1, \quad \partial_x y = 0, \quad \partial_x 1 = 0.$$

For a monomial $x_{i_1} \cdots x_{i_n}$ in x and y (where we put $x_1 := x$ and $x_2 := y$), this means explicitly

$$\partial_x x_{i_1} \cdots x_{i_n} = \sum_{k=1}^n \delta_{1i_k} x_{i_1} \cdots x_{i_{k-1}} \otimes x_{i_{k+1}} \cdots x_{i_n}. \quad (2.34)$$

Again it is clear that we can extend this definition also to formal power series in non-commuting variables.

Let us note that we may define the derivation ∂_{x+y} on $\mathbb{C}\langle x + y \rangle$ exactly as we did ∂_x . Namely, $\partial_{x+y}(1) = 0$ and $\partial_{x+y}(x + y) = 1 \otimes 1$. Note that ∂_{x+y} can be extended to all of $\mathbb{C}\langle x, y \rangle$ but not in a unique way unless we specify another basis element. Since $\mathbb{C}\langle x + y \rangle \subset \mathbb{C}\langle x, y \rangle$, we may apply ∂_x to $\mathbb{C}\langle x + y \rangle$ and observe that $\partial_x(x + y) = 1 \otimes 1 = \partial_{x+y}(x + y)$. Thus

$$\partial_x(x + y)^n = \sum_{k=1}^n (x + y)^{k-1} \otimes (x + y)^{n-k} = \partial_{x+y}(x + y)^n.$$

Hence

$$\partial_x|_{\mathbb{C}\langle x+y \rangle} = \partial_{x+y}. \quad (2.35)$$

If we are given a polynomial $p(x, y) \in \mathbb{C}\langle x, y \rangle$, then we will also consider $E_x[p(x, y)]$, the conditional expectation of $p(x, y)$ onto a function of just the variable x , which should be the best approximation to p among such functions. There is no algebraic way of specifying what best approximation means; we need a state φ on the $*$ -algebra generated by self-adjoint elements x and y for this. Given such a state, we will require that the difference between $p(x, y)$ and $E_x[p(x, y)]$ cannot be detected by functions of x alone; more precisely, we ask that

$$\varphi(q(x) \cdot E_x[p(x, y)]) = \varphi(q(x) \cdot p(x, y)) \quad (2.36)$$

for all $q \in \mathbb{C}\langle x \rangle$. If we are going from the polynomials $\mathbb{C}\langle x, y \rangle$ over to the Hilbert space completion $L^2(x, y, \varphi)$ with respect to the inner product given by $\langle f, g \rangle := \varphi(g^* f)$, then this amounts just to an orthogonal projection from the space $L^2(x, y, \varphi)$ onto the subspace $L^2(x, \varphi)$ generated by polynomials in the variable x . (Let us assume that φ is positive and faithful so that we get an inner product.) Thus, on the Hilbert space level, the existence and uniqueness of $E_x[p(x, y)]$ are clear. In general, though, it might not be the case that the projection of a polynomial in x and

y is a polynomial in x – it will just be an L^2 -function. If we assume, however, that x and y are free, then we claim that this projection maps polynomials to polynomials. In fact for this construction to work at the algebraic level we only need assume that $\varphi|_{\mathbb{C}\langle x \rangle}$ is non-degenerate as this shows that E_x is well defined by (2.36). It is clear from Equation (2.36) that $\varphi(E_x(a)) = \varphi(a)$ for all $a \in \mathbb{C}\langle x, y \rangle$.

Let us consider some examples. Assume that x and y are free. Then it is clear that we have

$$E_x[x^n y^m] = x^n \varphi(y^m)$$

and more generally

$$E_x[x^{n_1} y^m x^{n_2}] = x^{n_1+n_2} \varphi(y^m).$$

It is not so clear what $E_x[yx yx]$ might be. Before giving the general rule, let us make some simple observations.

Exercise 17. Let $\mathcal{A}_1 = \mathbb{C}\langle x \rangle$ and $\mathcal{A}_2 = \mathbb{C}\langle y \rangle$ with x and y free and $\varphi|_{\mathcal{A}_1}$ non-degenerate.

- (i) Show that $E_x[\mathring{\mathcal{A}}_2] = 0$.
- (ii) For $\alpha_1, \dots, \alpha_n \in \{1, 2\}$ with $\alpha_1 \neq \dots \neq \alpha_n$ and $n \geq 2$, show that $E_x[\mathring{\mathcal{A}}_{\alpha_1} \cdots \mathring{\mathcal{A}}_{\alpha_n}] = 0$.

Exercise 18. Let \mathcal{A}_1 and \mathcal{A}_2 be as in Exercise 17. Since \mathcal{A}_1 and \mathcal{A}_2 are free, we can use Equation (1.12) from Exercise 1.9 to write

$$\mathcal{A}_1 \vee \mathcal{A}_2 = \mathcal{A}_1 \oplus \mathring{\mathcal{A}}_2 \oplus \sum_{n \geq 2}^{\oplus} \sum_{\alpha_1 \neq \dots \neq \alpha_n}^{\oplus} \mathring{\mathcal{A}}_{\alpha_1} \mathring{\mathcal{A}}_{\alpha_2} \cdots \mathring{\mathcal{A}}_{\alpha_n}.$$

We have just shown that if E_x is a linear map satisfying Equation (2.36), then E_x is the identity on the first summand and 0 on all remaining summands. Show that by defining E_x this way we get the existence of a linear mapping from $\mathcal{A}_1 \vee \mathcal{A}_2$ to \mathcal{A}_1 satisfying Equation (2.36). An easy consequence of this is that for $q_1(x), q_2(x) \in \mathbb{C}\langle x \rangle$ and $p(x, y) \in \mathbb{C}\langle x, y \rangle$ we have $E_x[q_1(x)p(x, y)q_2(x)] = q_1(x)E_x[p(x, y)]q_2(x)$.

Let $a_1 = y^{n_1}, a_2 = y^{n_2}$ and $b = x^{m_1}$. To compute $E_x(y^{n_1} x^{m_1} y^{n_2})$ we follow the same centring procedure used to compute $\varphi(a_1 b a_2)$ in Section 1.12. From Exercise 17 we see that

$$\begin{aligned} E_x[a_1 b a_2] &= E_x[\mathring{a}_1 b a_2] + \varphi(a_1) b \varphi(a_2) \\ &= E_x[\mathring{a}_1 \mathring{b} a_2] + \varphi(\mathring{a}_1 a_2) \varphi(b) + \varphi(a_1) b \varphi(a_2) \\ &= \varphi(\mathring{a}_1 a_2) \varphi(b) + \varphi(a_1) b \varphi(a_2) \\ &= \varphi(a_1 a_2) \varphi(b) - \varphi(a_1) \varphi(b) \varphi(a_2) + \varphi(a_1) b \varphi(a_2). \end{aligned}$$

Thus

$$\begin{aligned} E_x[y^{n_1}x^{m_1}y^{n_2}x^{m_2}] &= \varphi(y^{n_1+n_2})\varphi(x^{m_1})x^{m_2} + \varphi(y^{n_1})x^{m_1}\varphi(y^{n_2})x^{m_2} \\ &\quad - \varphi(y^{n_1})\varphi(x^{m_1})\varphi(y^{n_2})x^{m_2}. \end{aligned}$$

The following theorem (essentially in the work [34] of Biane) gives the general recipe for calculating such expectations. As usual the formulas are simplified by using cumulants. To give the rule, we need the following bit of notation. Given $\sigma \in \mathcal{P}(n)$ and $a_1, \dots, a_n \in \mathcal{A}$, we define $\tilde{\varphi}_\sigma(a_1, \dots, a_n)$ in the same way as φ_σ in Equation (2.11) except we do not apply φ to the last block, i.e. the block containing n . For example, if $\sigma = \{(1, 3, 4), (2, 6), (5)\}$, then $\tilde{\varphi}_\sigma(a_1, a_2, a_3, a_4, a_5, a_6) = \varphi(a_1 a_3 a_4) \varphi(a_5) a_2 a_6$. More explicitly, for $\sigma = \{V_1, \dots, V_s\} \in NC(r)$ with $r \in V_s$, we put

$$\tilde{\varphi}_\sigma(a_1, \dots, a_r) = \varphi\left(\prod_{i_1 \in V_1} a_{i_1}\right) \cdots \varphi\left(\prod_{i_{s-1} \in V_{s-1}} a_{i_{s-1}}\right) \cdot \prod_{i_s \in V_s} a_{i_s}.$$

Theorem 19. *Let x and y be free. Then for $r \geq 1$ and $n_1, m_1, \dots, n_r, m_r \geq 0$, we have*

$$E_x[y^{n_1}x^{m_1} \cdots y^{n_r}x^{m_r}] = \sum_{\pi \in NC(r)} \kappa_\pi(y^{n_1}, \dots, y^{n_r}) \cdot \tilde{\varphi}_{K(\pi)}(x^{m_1}, \dots, x^{m_r}). \quad (2.37)$$

Let us check that this agrees with our previous calculation of $E_x[y^{n_1}x^{m_1}y^{n_2}x^{m_2}]$.

$$\begin{aligned} E_x[y^{n_1}x^{m_1}y^{n_2}x^{m_2}] &= \kappa_{\{(1,2)\}}(y^{n_1}, y^{n_2}) \cdot \tilde{\varphi}_{\{(1),(2)\}}(x^{m_1}, x^{m_2}) + \kappa_{\{(1),(2)\}}(y^{n_1}, y^{n_2}) \cdot \tilde{\varphi}_{\{(1,2)\}}(x^{m_1}, x^{m_2}) \\ &= \kappa_2(y^{n_1}, y^{n_2})\varphi(x^{m_1})x^{m_2} + \kappa_1(y^{n_1})\kappa_1(y^{n_2})x^{m_1+m_2} \\ &= (\varphi(y^{n_1+n_2}) - \varphi(y^{n_1})\varphi(y^{n_2}))\varphi(x^{m_1}) \cdot x^{m_2} + \varphi(y^{n_1})\varphi(y^{n_2}) \cdot x^{m_1+m_2}. \end{aligned}$$

The proof of the theorem is outlined in the exercise below.

Exercise 19. (i) Given $\pi \in NC(n)$, let π' be the non-crossing partition of $[n'] = \{0, 1, 2, 3, \dots, n\}$ obtained by joining 0 to the block of π containing n . For $a_0, a_1, \dots, a_n \in \mathcal{A}$, show that $\varphi_{\pi'}(a_0, a_1, a_2, \dots, a_n) = \varphi(a_0 \tilde{\varphi}_\pi(a_1, \dots, a_n))$.

(ii) Suppose that $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ are unital subalgebras of \mathcal{A} which are free with respect to the state φ . Let $x_0, x_1, \dots, x_n \in \mathcal{A}_1$ and $y_1, y_2, \dots, y_n \in \mathcal{A}_2$. Show that

$$\varphi(x_0 y_1 x_1 y_2 x_2 \cdots y_n x_n) = \sum_{\pi \in NC(n)} \kappa_\pi(y_1, \dots, y_n) \varphi_{K(\pi)'}(x_0, x_1, \dots, x_n).$$

Prove Theorem 19 by showing that with the expression given in (2.37) one has for all $m \geq 0$

$$\varphi(x^m \cdot E_x[y^{n_1} x^{m_1} \cdots y^{n_r} x^{m_r}]) = \varphi(x^m \cdot y^{n_1} x^{m_1} \cdots y^{n_r} x^{m_r}).$$

Exercise 20. Use the method of Exercise 19 to work out $E_x[x^{m_1} y^{n_1} \cdots x^{m_r} y^{n_r}]$.

By linear extension of Equation (2.37), one can thus get the projection onto one variable x of any non-commutative polynomial or formal power series in two free variables x and y . We now want to identify the projection of resolvents in $x + y$. To achieve this we need a crucial intertwining relation between the partial derivative and the conditional expectation.

Lemma 20. *Suppose φ is a state on $\mathbb{C}\langle x, y \rangle$ such that x and y are free and $\varphi|_{\mathbb{C}\langle x \rangle}$ is non-degenerate. Then*

$$E_x \otimes E_x \circ \partial_{x+y}|_{\mathbb{C}\langle x+y \rangle} = \partial_x \circ E_x|_{\mathbb{C}\langle x+y \rangle}. \quad (2.38)$$

Proof: We let $\mathcal{A}_1 = \mathbb{C}\langle x \rangle$ and $\mathcal{A}_2 = \mathbb{C}\langle y \rangle$. We use the decomposition from Exercise 1.9

$$\mathcal{A}_1 \vee \mathcal{A}_2 \ominus \mathcal{A}_1 = \mathring{\mathcal{A}}_2 \oplus \sum_{n \geq 2}^{\oplus} \sum_{\alpha_1 \neq \cdots \neq \alpha_n}^{\oplus} \mathring{\mathcal{A}}_{\alpha_1} \cdots \mathring{\mathcal{A}}_{\alpha_n}$$

and examine the behaviour of $E_x \otimes E_x \circ \partial_x$ on each summand. We know that ∂_x is 0 on $\mathring{\mathcal{A}}_2$ by definition. For $n \geq 2$

$$\begin{aligned} & E_x \otimes E_x \circ \partial_x(\mathring{\mathcal{A}}_{\alpha_1} \cdots \mathring{\mathcal{A}}_{\alpha_n}) \\ & \subseteq \sum_{k=1}^n \delta_{1, \alpha_k} E_x(\mathring{\mathcal{A}}_{\alpha_1} \cdots \mathring{\mathcal{A}}_{\alpha_{k-1}}(\mathbb{C}1 \oplus \mathring{\mathcal{A}}_{\alpha_k})) \otimes E_x((\mathbb{C}1 \oplus \mathring{\mathcal{A}}_{\alpha_k})\mathring{\mathcal{A}}_{\alpha_{k+1}} \cdots \mathring{\mathcal{A}}_{\alpha_n}). \end{aligned}$$

By Exercise 17, in each term, one or both of the factors is 0. Thus $E_x \otimes E_x \circ \partial_x|_{\mathcal{A}_1 \vee \mathcal{A}_2 \ominus \mathcal{A}_1} = 0$. Hence

$$E_x \otimes E_x \circ \partial_x|_{\mathcal{A}_1 \vee \mathcal{A}_2} = E_x \otimes E_x \circ \partial_x \circ E_x|_{\mathcal{A}_1 \vee \mathcal{A}_2} = \partial_x \circ E_x|_{\mathcal{A}_1 \vee \mathcal{A}_2},$$

and then by Equation (2.35) we have

$$E_x \otimes E_x \circ \partial_{x+y}|_{\mathbb{C}\langle x+y \rangle} = E_x \otimes E_x \circ \partial_x|_{\mathbb{C}\langle x+y \rangle} = \partial_x \circ E_x|_{\mathbb{C}\langle x+y \rangle}.$$

□

Theorem 21. *Let x and y be free. For every $z \in \mathbb{C}$ with $z \neq 0$, there exists a $w \in \mathbb{C}$ such that*

$$\mathbb{E}_x \left[\frac{1}{z - (x + y)} \right] = \frac{1}{w - x}. \quad (2.39)$$

In other words, the best approximation for a resolvent in $x + y$ by a function of x is again a resolvent.

By applying the state φ to both sides of (2.39), one obtains the subordination for the Cauchy transforms, and thus it is clear that the w from above must agree with the subordination function from (2.31), $w = \omega(z)$.

Proof: We put

$$f(x, y) := \frac{1}{z - (x + y)}.$$

By Exercise 16, part (i), we know that $\partial_{x+y} f = f \otimes f$. By Lemma 20 we have that for functions g of $x + y$

$$\partial_x \mathbb{E}_x [g(x + y)] = \mathbb{E}_x \otimes \mathbb{E}_x [\partial_{x+y} g(x + y)]. \quad (2.40)$$

By applying (2.40) to f , we obtain

$$\partial_x \mathbb{E}_x [f] = \mathbb{E}_x \otimes \mathbb{E}_x [\partial_{x+y} f] = \mathbb{E}_x \otimes \mathbb{E}_x [f \otimes f] = \mathbb{E}_x [f] \otimes \mathbb{E}_x [f].$$

Thus, by the second part of Exercise 16, we know that $\mathbb{E}_x [f]$ is a resolvent in x and we are done. \square