# 11
# Free Probabilities

In the previous chapter we saw how to compute the spectrum of the sum of two large random matrices, first when one of them is a Wigner and later when one is "rotationally invariant" with respect to the other. In this chapter, we would like to formalize the notion of relative rotational invariance, which leads to the abstract concept of freeness.

The idea is as follows. In standard probability theory, one can work abstractly by defining expectation values (moments) of random variables. The concept of independence is then equivalent to the factorization of moments (e.g. $\mathbb{E}[A^3 B^2] = \mathbb{E}[A^3]\mathbb{E}[B^2]$ when $A$ and $B$ are independent).

However, random matrices do not commute in general and the concept of factorization of moments is not powerful enough to deal with non-commuting random objects. Following von Neumann, Voiculescu extended the concept of independence to non-commuting objects and called this property *freeness*. He then showed how to characterize the sum and the product of *free* variables. It was later realized that large rotationally invariant matrices provide an explicit example of (asymptotically) free random variables. In other words, free probabilities gave us very powerful tools to compute sums and products of large random matrices. We have already encountered the free addition; the free product will allow us to study sample covariance matrices in the presence of non-trivial true correlations.

This chapter may seem too dry and abstract for someone looking for applications. Bear with us, it is in fact not that complicated and we will keep the jargon to a minimum. The reward will be one of the most powerful and beautiful recent developments in random matrix theory, which we will expand upon in Chapter 12.

## 11.1 Algebraic Probabilities: Some Definitions

The ingredients we will need in this chapter are as follows:[1]

- A **ring** $\mathcal{R}$ of random variables, which can be non-commutative with respect to the multiplication.[2]

[1] In mathematical language, the first three items give a *-algebra, while $\tau$ gives a tracial state on this algebra.
[2] Recall that a ring is a set equipped with two binary operations that generalize the arithmetic operations of addition and multiplication.



https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

- A **field** of scalars, which is usually taken to be $\mathbb{C}$. The scalars commute with everything.
- An operation $*$, called involution. For instance, $*$ denotes the conjugate for complex numbers, the transpose for real matrices, and the conjugate transpose for complex matrices.
- A positive linear functional $\tau(.)$ ($\mathcal{R} \to \mathbb{C}$) that satisfies $\tau(AB) = \tau(BA)$ for $A, B \in \mathcal{R}$. By positive we mean $\tau(AA^*)$ is real non-negative. We also require that $\tau$ be *faithful*, in the sense that $\tau(AA^*) = 0$ implies $A = 0$. For instance, $\tau$ can be the expectation operator $\mathbb{E}[.]$ for standard probability theory, or the normalized trace operator $\frac{1}{N}\,\mathrm{Tr}(.)$ for a ring of matrices, or the combined operation $\frac{1}{N}\mathbb{E}[\mathrm{Tr}(.)]$.

We will call the elements in $\mathcal{R}$ the random variables and denote them by capital letters. For any $A \in \mathcal{R}$ and $k \in \mathbb{N}$, we call $\tau(A^k)$ the $k$th moment of $A$ and we assume in what follows that $\tau(A^k)$ is finite for all $k$. In particular, we call $\tau(A)$ the mean of $A$ and $\tau(A^2) - \tau(A)^2$ the variance of $A$. We will say that two elements $A$ and $B$ have the same distribution if they have the same moments of all orders.[3]

The ring of variables must have an element called $\mathbf{1}$ such that $A\mathbf{1} = \mathbf{1}A = A$ for every $A$. It satisfies $\tau(\mathbf{1}) = 1$. We will call $\mathbf{1}$ and its multiples $\alpha\mathbf{1}$ *constants*. Adding a constant simply shifts the mean as

$$\tau(A + \alpha\mathbf{1}) = \tau(A) + \alpha. \tag{11.1}$$

## 11.2 Addition of Commuting Variables

In this section, we recall some well-known properties of commuting random variables, i.e. such that

$$AB = BA, \quad \forall A, B \in \mathcal{R}. \tag{11.2}$$

Note that $A$ is not necessarily a real (or complex) number but can be an element of a more abstract ring. We will say that $A$ and $B$ are independent, if $\tau(p(A)q(B)) = \tau(p(A))\tau(q(B))$ for any polynomial $p, q$. This condition is equivalent to the factorization of moments.

From a scalar $\alpha$ we can build the constant $\alpha\mathbf{1}$ and write $A+\alpha$ to mean $A+\alpha\mathbf{1}$. Constants of the ring are independent of all other random variables, so if $A$ and $B$ are independent, $A + \alpha$ and $B$ are also independent. This setting recovers the classical probability theory of commutative random variables (with finite moments to every order).

### *11.2.1 Moments*

Now let us study the moments of the sum of independent random variables $A + B$. First we trivially have by linearity

$$\tau(A + B) = \tau(A) + \tau(B). \tag{11.3}$$

[3] This is of course not correct for standard commuting random variables: some distributions are not uniquely determined by their moments.

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

From now on we will assume $\tau(A) = \tau(B) = 0$, i.e. $A, B$ have mean zero, unless stated otherwise. For a non-zero mean variable $\tilde{A}$, we write $A = \tilde{A} - \tau(\tilde{A})$ such that $\tau(A) = 0$. One can recover the formulas for moments and cumulants of $\tilde{A}$ simply by substituting $A \to \tilde{A} - \tau(\tilde{A})$ in all formulas written for zero mean $A$. The procedure is straightforward but leads to rather cumbersome results.

For the second moment,

$$\begin{aligned}\tau\left((A+B)^2\right) &= \tau(A^2) + \tau(B^2) + 2\tau(AB) \\ &= \tau(A^2) + \tau(B^2) + 2\tau(A)\tau(B) = \tau(A^2) + \tau(B^2),\end{aligned} \tag{11.4}$$

i.e. the variance is also additive. For the third moment, we have

$$\tau\left((A+B)^3\right) = \tau(A^3) + \tau(B^3) + 3\tau(A)\tau(B^2) + 3\tau(B)\tau(A^2) = \tau(A^3) + \tau(B^3), \tag{11.5}$$

which is also additive. However, the fourth and higher moments are not additive anymore. For example we get, expanding $(A+B)^4$,

$$\tau\left((A+B)^4\right) = \tau(A^4) + \tau(B^4) + 6\tau(A^2)\tau(B^2). \tag{11.6}$$

### *11.2.2 Cumulants*

For zero mean variables the first three moments are additive but not the higher ones. Nevertheless, certain combinations of higher moments are additive; we call them cumulants and denote them as $\kappa_n$ for the $n$th cumulant. Note that for a variable with non-zero mean $\tilde{A}$, the second and third cumulants are the second and third moments of $A := \tilde{A} - \tau(\tilde{A})$:

$$\begin{aligned}\kappa_1(\tilde{A}) &= \tau(\tilde{A}), \\ \kappa_2(\tilde{A}) &= \tau(A^2) = \tau(\tilde{A}^2) - \tau(\tilde{A})^2, \\ \kappa_3(\tilde{A}) &= \tau(A^3) = \tau(\tilde{A}^3) - 3\tau(\tilde{A}^2)\tau(\tilde{A}) + 2\tau(\tilde{A})^3.\end{aligned} \tag{11.7}$$

For the fourth cumulant, let us consider for simplicity zero mean variables $A$ and $B$, and define $\kappa_4$ as

$$\kappa_4(A) := \tau(A^4) - 3\tau(A^2)^2. \tag{11.8}$$

Then we can verify that

$$\begin{aligned}\kappa_4(A+B) &= \tau\left((A+B)^4\right) - 3\tau\left((A+B)^2\right)^2 \\ &= \tau(A^4) + \tau(B^4) + 6\tau(A^2)\tau(B^2) - 3\left(\tau(A^2) + \tau(B^2)\right)^2 \\ &= \tau(A^4) - 3\tau(A^2)^2 + \tau(B^4) - 3\tau(B^2)^2 = \kappa_4(A) + \kappa_4(B),\end{aligned} \tag{11.9}$$

which is additive again.

In general, $\tau((A+B)^n)$ will be of the form $\tau(A^n) + \tau(B^n)$ plus some homogeneous mix of lower order terms. We can then define the $n$th cumulant $\kappa_n$ iteratively such that

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

$$\kappa_n(A+B) = \kappa_n(A) + \kappa_n(B), \tag{11.10}$$

where

$$\kappa_n(A) = \tau(A^n) + \text{lower order terms moments.} \tag{11.11}$$

In order to have a compact definition of cumulants, recall that we are looking for quantities that are additive for independent variables. But we already know that the log-characteristic function introduced in Eq. (8.4) is additive. In the present context, we define the characteristic function as[4]

$$\varphi_A(t) = \tau\left(\mathrm{e}^{\mathrm{i}tA}\right), \tag{11.12}$$

where the exponential function is formally defined through its power series:

$$\tau(\mathrm{e}^{\mathrm{i}tA}) = \sum_{\ell=0}^{\infty} \frac{(\mathrm{i}t)^\ell}{\ell!}\tau(A^\ell), \tag{11.13}$$

hence the characteristic function is also the moment generating function. Now, from the formal definition of the exponential and the factorization of moments one can easily show that for independent, commuting $A, B$,

$$\varphi_{A+B}(t) = \varphi_A(t)\varphi_B(t). \tag{11.14}$$

Here is an *algebraic* proof. For each $k$,

$$\tau((A+B)^k) = \sum_{i=0}^{k} \binom{k}{i} \tau(A^i)\tau(B^{k-i}), \tag{11.15}$$

with which we get

$$\begin{aligned}\varphi_{A+B}(t) &= \sum_{k=0}^{\infty}\sum_{i\le k} \frac{(\mathrm{i}t)^k}{k!}\binom{k}{i}\tau(A^i)\tau(B^{k-i}) = \sum_{i\le k} \frac{(\mathrm{i}t)^{k-i}\tau(B^{k-i})}{(k-i)!}\,\frac{(\mathrm{i}t)^i\tau(A^i)}{i!} \\ &= \left(\sum_i \frac{(\mathrm{i}t)^i\tau(A^i)}{i!}\right)\left(\sum_j \frac{(\mathrm{i}t)^j\tau(B^j)}{j!}\right) = \varphi_A(t)\varphi_B(t). \end{aligned} \tag{11.16}$$

We now define $H_A(t) := \log \varphi_A(t)$. Then, for independent, commuting $A, B$, we have

$$H_{A+B}(t) = H_A(t) + H_B(t). \tag{11.17}$$

We can expand $H_A(t)$ as a power series of $t$ and call the corresponding coefficients the cumulants, i.e.

$$H_A(t) = \log\tau\left(\mathrm{e}^{\mathrm{i}tA}\right) := \sum_{n=1}^{\infty} \frac{\kappa_n(A)}{n!}(\mathrm{i}t)^n. \tag{11.18}$$

[4] The factor i in the definition is not necessary in this setting as the formal power series of the exponential and the logarithm do not need to converge. We nevertheless include it by analogy with the Fourier transform.

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

From the additive property of $H$, the cumulants defined in the above way are automatically additive. In fact, using the power series for $\log(1+x)$, we have

$$H_A(t) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \left( \sum_{k=1}^{\infty} \frac{(\mathrm{i}t)^k}{k!} \tau(A^k) \right)^n \equiv \sum_{n=1}^{\infty} \frac{\kappa_n(A)}{n!} (\mathrm{i}t)^n. \tag{11.19}$$

Matching powers of $(\mathrm{i}t)$ we obtain an expression for $\kappa_n$ for any $n$. We can work out by hand the first few cumulants. For example, for $n = 1$, Eq. (11.19) readily yields $\kappa_1(A) = \tau(A)$. We now assume $A$ has mean zero, i.e. $\tau(A) = 0$. Then

$$\tau\left(\mathrm{e}^{\mathrm{i}tA}\right) = 1 + \frac{(\mathrm{i}t)^2}{2}\tau(A^2) + \frac{(\mathrm{i}t)^3}{6}\tau(A^3) + \frac{(\mathrm{i}t)^4}{24}\tau(A^4) + \cdots, \tag{11.20}$$

whereas the first few terms in the expansion of (11.19) are

$$H_A(t) = \frac{(\mathrm{i}t)^2}{2}\tau(A^2) + \frac{(\mathrm{i}t)^3}{6}\tau(A^3) + (\mathrm{i}t)^4 \left( \frac{\tau(A^4)}{24} - \frac{\tau(A^2)^2}{8} \right) + \cdots, \tag{11.21}$$

from which we recover the first four cumulants defined above:

$$\kappa_1(A) = 0, \quad \kappa_2(A) = \tau(A^2), \quad \kappa_3(A) = \tau(A^3), \quad \kappa_4(A) = \tau(A^4) - 3\tau(A^2)^2. \tag{11.22}$$

The expression for $\kappa_n$ soon becomes very cumbersome for larger $n$. Nevertheless, by exponentiating Eq. (11.18) and matching with Eq. (11.13), one can extract the following moment–cumulant relation for commuting variables:

$$\begin{aligned} \tau(A^n) &= \sum_{\substack{r_1, r_2, \ldots, r_n \geq 0 \\ r_1 + 2r_2 + \cdots + nr_n = n}} \frac{n!\, \kappa_1^{r_1} \kappa_2^{r_2} \cdots \kappa_n^{r_n}}{(1!)^{r_1} (2!)^{r_2} \cdots (n!)^{r_n} r_1!\, r_2! \cdots r_n!} \\ &= \kappa_n + \text{products of lower order terms} + \kappa_1^n. \end{aligned} \tag{11.23}$$

In particular, the scaling properties for the moments and cumulants (see (11.28) below) are consistent due to the relation $r_1 + 2r_2 + \cdots + nr_n = n$.

**Exercise 11.2.1 Cumulants of a constant**

Show that a constant $\alpha\mathbf{1}$ has $\kappa_1 = \alpha$ and $\kappa_n = 0$ for $n \geq 2$. (Hint: compute $H_{\alpha\mathbf{1}}(k) = \log\left(\tau\left(\mathrm{e}^{\mathrm{i}k\alpha\mathbf{1}}\right)\right)$.)

### *11.2.3 Scaling of Moments and Cumulants*

Moments and cumulants obey simple transformation rules under scalar addition and multiplication. For example, when adding a constant to a variable, $\tilde{A} := A + \alpha$, where $\tau(A) = 0$, we only change the first cumulant:

$$\kappa_1(\tilde{A}) = \alpha \quad \text{and} \quad \kappa_n(\tilde{A}) = \kappa_n(A) \quad \text{for} \quad n \geq 2. \tag{11.24}$$

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

For the case of multiplication by an arbitrary scalar $\alpha$, by commutativity of scalars and linearity of $\tau$ we have

$$\tau\left((\alpha A)^k\right) = \alpha^k \tau\left(A^k\right). \tag{11.25}$$

For the cumulant, we first look at the scaling of the log-characteristic function:

$$H_{\alpha A}(t) = \log\left(\tau\left(\mathrm{e}^{\mathrm{i}t\alpha A}\right)\right) = H_A(\alpha t). \tag{11.26}$$

And by (11.18), we have

$$H_{\alpha A}(t) = H_A(\alpha t) = \sum_{n=1}^{\infty} \frac{\alpha^n \kappa_n(A)}{n!} (\mathrm{i}t)^n. \tag{11.27}$$

Thus we have the simple scaling property

$$\kappa_n(\alpha A) = \alpha^n \kappa_n(A). \tag{11.28}$$

### *11.2.4 Law of Large Numbers and Central Limit Theorem*

Continuing our study of algebraic probabilities, we would like to recover two very important theorems in probability theory, namely the law of large numbers (LLN) and the central limit theorem (CLT). The first states that the sample average converges to the mean (a constant) as the number of observations $N \to \infty$, and the second that a large sum of properly centered and rescaled random variables converge to a Gaussian.

First we need to define in our context what we mean by a constant and a Gaussian. For simplicity, we can think of the variables in this section as standard random variables. We will later introduce non-commutating cumulants. The arguments of this section apply in the non-commutative case with independence replaced by freeness.

We have defined the constant variable $A = \alpha\mathbf{1}$, which satisfies

$$\kappa_1(A) = \alpha, \quad \kappa_\ell(A) = 0, \ \forall \ell > 1. \tag{11.29}$$

Then we define the "Gaussian" random variable as an element $A$ that satisfies

$$\kappa_2(A) \neq 0, \quad \kappa_\ell(A) = 0, \ \forall \ell > 2. \tag{11.30}$$

Note that this definition (in the scalar case) is equivalent to the standard Gaussian random variable with density

$$P_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{11.31}$$

with $\kappa_1 = \mu$ and $\kappa_2 = \sigma^2$.

By extension, we call $\kappa_1(A)$ the mean, and $\kappa_2(A)$ the variance. Now we can give a simple proof for the LLN and CLT within our algebraic setting. Let

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

$$S_K := \frac{1}{K}\sum_{i=1}^{K} A_i, \tag{11.32}$$

where $A_i$ are $K$ IID variables.[5] Then by (11.28) and the additive property of cumulants, we get that

$$\kappa_\ell(S_K) = \frac{K}{K^\ell}\kappa_\ell(A) \stackrel{K\to\infty}{\to} \begin{cases} \kappa_1(A), & \text{if } \ell = 1, \\ 0, & \text{if } \ell > 1. \end{cases} \tag{11.33}$$

In other words, $S_K$ converges to a constant in the sense of cumulants.

Assume now that $\kappa_1(A) = 0$ and consider

$$\widehat{S}_K := \frac{1}{\sqrt{K}}\sum_{i=1}^{K} A_i. \tag{11.34}$$

Then it is easy to see that

$$\kappa_\ell(\widehat{S}_K) = \frac{K}{K^{\ell/2}}\kappa_\ell(A) \to \begin{cases} 0, & \text{if } \ell = 1, \\ \kappa_2(A), & \text{if } \ell = 2, \\ 0, & \text{if } \ell > 2. \end{cases} \tag{11.35}$$

In other words, $\widehat{S}_K$ converges to a "Gaussian" random variable with variance $\kappa_2(A) = \tau(A^2)$, in the sense of cumulants.

In our algebraic probability setting we have made the assumption that the variables we consider have finite moments of all orders. This is a very strong assumption. In particular it excludes any variable whose probability decays as a power law. If we relaxed this assumption we would find that some sums of power-law distributed variables converge not to a Gaussian but to a Lévy-stable distribution. A similar concept exists in the non-commutative case, but it is beyond the scope of this book.

## 11.3 Non-Commuting Variables

We now return to our original goal of developing an extension of standard probabilities for non-commuting objects. One of the goals is to generalize the law of addition of independent variables. We consider a variable equal to $A + B$ where $A$ and $B$ are now non-commutative objects such as large random matrices. If we compute the first three moments of $A + B$, no particular problems arise thanks to the tracial property of $\tau$, and they behave as in the commutative case. For example, consider the third moment

[5] IID copies are variables $A_i$ that have exactly the same cumulants and are all independent. We did not define independence for more than two variables but the factorization of moments can be easily extended to more variables. Note that pairwise independence is not enough to assure independence as a group. For example if $x_1$, $x_2$ and $x_3$ are IID Gaussians, the Gaussian variables $x_1$, $x_2$ and $y_3 = \text{sign}(x_1 x_2)|x_3|$ are pairwise independent but not all independent as $\mathbb{E}[x_1 x_2 y_3] > 0$ whereas $\mathbb{E}[x_1]\mathbb{E}[x_2]\mathbb{E}[y_3] = 0$.

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

$$\tau\big((A+B)^3\big) = \tau(A^3) + \tau(A^2B) + \tau(ABA) + \tau(BA^2) \\ + \tau(AB^2) + \tau(BAB) + \tau(B^2A) + \tau(B^3). \tag{11.36}$$

But since $\tau(A^2B) = \tau(ABA) = \tau(BA^2)$ (and similarly when $A$ appears once and $B$ twice), the classic independence property

$$\tau(A^2B) = \tau(A^2)\tau(B) = 0 \tag{11.37}$$

appears to be sufficient. Things become more interesting for the fourth moment. Indeed,

$$\begin{aligned}\tau\big((A+B)^4\big) &= \tau(A^4) + \tau(A^3B) + \tau(A^2BA) + \tau(ABA^2) + \tau(BA^3) \\ &\quad + \tau(A^2B^2) + \tau(ABAB) + \tau(BA^2B) + \tau(AB^2A) + \tau(BABA) \\ &\quad + \tau(B^2A^2) + \tau(B^3A) + \tau(B^2AB) + \tau(BAB^2) + \tau(AB^3) + \tau(B^4) \\ &= \tau(A^4) + 4\tau(A^3B) + 4\tau(A^2B^2) + 2\tau(ABAB) + 4\tau(AB^3) + \tau(B^4),\end{aligned} \tag{11.38}$$

where in the second step we again used the tracial property of $\tau$. For commutative random variables, independence of $A, B$ means $\tau(A^2B^2) = \tau(A^2)\tau(B^2)$, and this is enough to treat all the terms above. In the non-commutative case, we also need to handle the term $\tau(ABAB)$. In general $ABAB$ is not equal to $A^2B^2$. "Independence" is therefore not enough to deal with this term, so we need a new concept. A radical solution would be to postulate that $\tau(ABAB)$ is zero whenever $\tau(A) = \tau(B) = 0$. As we compute higher moments of $A + B$ we will encounter more and more complicated similar mixed moments. The concept of *freeness* deals with all such terms at once.

### *11.3.1 Freeness*

Given two random variables $A, B$, we say they are **free** if for any polynomials $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$ such that

$$\tau(p_k(A)) = 0, \quad \tau(q_k(B)) = 0, \quad \forall k, \tag{11.39}$$

we have

$$\tau\left(p_1(A)q_1(B)p_2(A)q_2(B)\cdots p_n(A)q_n(B)\right) = 0. \tag{11.40}$$

We will call a polynomial (or a variable) traceless if $\tau(p(A)) = 0$. Note that $\alpha\mathbf{1}$ is free with respect to any $A \in \mathcal{R}$ because $\tau(p(\alpha\mathbf{1})) \equiv p(\alpha\mathbf{1})$, from the definition of $\mathbf{1}$. Hence,

$$\tau(p(\alpha\mathbf{1})) = 0 \Leftrightarrow p(\alpha\mathbf{1}) = 0. \tag{11.41}$$

Moreover, it is easy to see that if $A, B$ are free, then $p(A), q(B)$ are free for any polynomials $p, q$. By extension, $F(A)$ and $G(B)$ are also free for any function $F$ and $G$ defined by their power series.

The freeness is non-trivial only in the non-commutative case. For the commutative case, it is easy to check that $A, B$ are free if and only if either $A$ or $B$ is a constant. Free random

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

variables are "maximally" non-commuting, in a sense made more precise for the example of free random matrices in the next chapter. For example, for free and mean zero variables $A$ and $B$, we have $\tau(ABAB) = 0$ whereas $\tau(A^2B^2) = \tau((A^2 - \tau(A^2))B^2) + \tau(A^2)\tau(B^2) = \tau(A^2)\tau(B^2)$.

Assuming $A, B$ are free with $\tau(A) = \tau(B) = 0$, we can compute the moments of the free addition $A + B$. The second moment is easy:

$$\tau\big((A+B)^2\big) = \tau(A^2) + \tau(B^2) + 2\tau(AB) = \tau(A^2) + \tau(B^2), \tag{11.42}$$

because both $\tau(A), \tau(B)$ are zero. For the third and higher moments the trick is, as just above, to add and subtract quantities such that, in each term, at least one object of the form $(C - \tau(C))$ is present:

$$\begin{aligned}\tau\big((A+B)^3\big) &= \tau(A^3) + \tau(B^3) + 3\tau(A^2B) + 3\tau(AB^2)\\ &= \tau(A^3) + \tau(B^3) + 3\tau((A^2 - \tau(A^2))B) + 3\tau(A^2)\tau(B)\\ &\quad + 3\tau(A(B^2 - \tau(B^2))) + 3\tau(A)\tau(B^2)\\ &= \tau(A^3) + \tau(B^3),\end{aligned} \tag{11.43}$$

and for the fourth moment:

$$\begin{aligned}\tau\big((A+B)^4\big) &= \tau(A^4) + 4\tau(A^3B) + 4\tau(A^2B^2) + 2\tau(ABAB) + 4\tau(AB^3) + \tau(B^4)\\ &= \tau(A^4) + 4\tau((A^3 - \tau(A^3))B) + 4\tau(A^3)\tau(B)\\ &\quad + 4\tau\big(\big(A^2 - \tau(A^2)\big)\big(B^2 - \tau(B^2)\big)\big) + 4\tau(A^2)\tau(B^2)\\ &\quad + 4\tau\big(A\big(B^3 - \tau(B^3)\big)\big) + 4\tau(A)\tau(B^3) + \tau(B^4)\\ &= \tau(A^4) + \tau(B^4) + 4\tau(A^2)\tau(B^2).\end{aligned} \tag{11.44}$$

In particular, we find

$$\tau\big((A+B)^4\big) - 2\tau\big((A+B)^2\big)^2 = \tau(A^4) + \tau(B^4) - 2\tau(A^2)^2 - 2\tau(B^2)^2. \tag{11.45}$$

### 11.3.2 Free Cumulants

Let us define the cumulants of $A$ as

$$\kappa_1(A) = \tau(A), \quad \kappa_2(A) = \tau(A_0^2), \quad \kappa_3(A) = \tau(A_0^3), \quad \kappa_4(A) = \tau(A_0^4) - 2\tau(A_0^2)^2, \tag{11.46}$$

where $A_0$ is a short-hand for $A - \tau(A)\mathbf{1}$. Then these objects are additive for free random variables. The first three are the same as the commutative ones. But for the fourth cumulant, the coefficient in front of $\tau(A_0^2)^2$ is now 2 instead of 3. Higher cumulants all differ from their commutative counterparts.

As in the commutative case, we can define the $k$th free cumulant iteratively as

$$\kappa_k(A) = \tau(A^k) + \text{homogeneous products of lower order moments}, \tag{11.47}$$

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

such that

$$\kappa_k(A+B) = \kappa_k(A) + \kappa_k(B), \quad \forall k, \tag{11.48}$$

whenever $A, B$ are free.

An important example of non-commutative free random variables is two independent large random matrices where one of them is rotational invariant – see next chapter.[6]

### 11.3.3 Additivity of the R-Transform

In the previous chapter, we saw that the R-transform is additive for large rotationally invariant matrices. We will show here that we can define the R-transform in our abstract algebraic probability setting and that this R-transform is also additive for free variables. In the next chapter, we will dwell on why large rotationally invariant matrices are free.

First we define the Stieltjes transform as a moment generating function as in (2.22); we can define $\mathfrak{g}_A(z)$ for large $z$ as

$$\mathfrak{g}_A(z) = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \tau(A^k). \tag{11.49}$$

Then we can also define the R-transform as before:

$$R_A(g) := \mathfrak{z}_A(g) - \frac{1}{g} \tag{11.50}$$

for small enough $g$. Here the inverse function $\mathfrak{z}_A(g)$ is defined as the formal power series that satisfies $\mathfrak{g}_A(\mathfrak{z}_A(g)) = g$ to all orders.

We now claim that the R-transform is additive for free random variables, i.e.

$$R_{A+B}(g) = R_A(g) + R_B(g), \tag{11.51}$$

whenever $A, B$ are free.

We let $\mathfrak{z}_A(g)$ be the inverse function of

$$\mathfrak{g}_A(z) = \tau\left[(z-A)^{-1}\right], \tag{11.52}$$

whose power series is actually given by (11.49). Consider a fixed scalar $g$. By construction

$$\tau(g\mathbf{1}) = g = \mathfrak{g}_A(\mathfrak{z}_A(g)) = \tau\left[(\mathfrak{z}_A(g) - A)^{-1}\right]. \tag{11.53}$$

The arguments of $\tau(.)$ on the left and on the right of the above equation have the same mean but they are in general different, so let us define their difference as $gX_A$ via

$$gX_A := (z_A - A)^{-1} - g\mathbf{1}, \tag{11.54}$$

where $z_A := \mathfrak{z}_A(g)$. From its very definition, we have $\tau(X_A) = 0$.

[6] Freeness is only exact in the large $N$ limit of random matrices.

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

We can invert Eq. (11.54) and find

$$A - z_A = -\frac{1}{g}(1 + X_A)^{-1}. \tag{11.55}$$

Consider another variable $B$, free from $A$. For the same fixed $g$ we can find the scalar $z_B := \mathfrak{z}_B(g)$ and define $X_B$ with $\tau(X_B) = 0$ as for $A$, to find

$$B - z_B = -\frac{1}{g}(1 + X_B)^{-1}. \tag{11.56}$$

Since $X_A$ and $X_B$ are functions of $A$ and $B$, $X_A$ and $X_B$ are also free. Now,

$$\begin{aligned} A + B - z_A - z_B &= -\frac{1}{g}(1 + X_A)^{-1} - \frac{1}{g}(1 + X_B)^{-1} \\ &= -\frac{1}{g}(1 + X_A)^{-1}(2 + X_A + X_B)(1 + X_B)^{-1}. \end{aligned} \tag{11.57}$$

Hence, noting that $(1 + X_A)(1 + X_B) + 1 - X_A X_B = 2 + X_A + X_B$,

$$\begin{aligned} A + B - z_A - z_B + \frac{1}{g} &= -\frac{1}{g}(1 + X_A)^{-1}(1 - X_A X_B)(1 + X_B)^{-1}, \\ \left[A + B - \left(z_A + z_B - \frac{1}{g}\right)\right]^{-1} &= -g(1 + X_B)(1 - X_A X_B)^{-1}(1 + X_A). \end{aligned} \tag{11.58}$$

Using the identity

$$(1 - X_A X_B)^{-1} = \sum_{n=0}^{\infty} (X_A X_B)^n, \tag{11.59}$$

we can expand the expression

$$\tau\left[(1 + X_B)(1 - X_A X_B)^{-1}(1 + X_A)\right], \tag{11.60}$$

which will contain 1 plus terms of the form $\tau(X_A X_B X_A X_B \ldots X_B)$ where the initial and final factor might be either $X_A$ or $X_B$ but the important point is that $X_A$ and $X_B$ always alternate. By the freeness and zero mean of $X_A$ and $X_B$, all these terms are thus zero. Hence we get

$$\tau\left\{\left[A + B - \left(z_A + z_B - \frac{1}{g}\right)\right]^{-1}\right\} = -g \Rightarrow \mathfrak{g}_{A+B}(z_A + z_B - g^{-1}) = g, \tag{11.61}$$

finally leading to the announced result:[7]

$$\mathfrak{z}_{A+B} = z_A + z_B - g^{-1} \Rightarrow R_{A+B} = R_A + R_B. \tag{11.62}$$

[7] The above compact proof is taken from Tao [2012].

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

### 11.3.4 R-Transform and Cumulants

The R-transform is defined as a power series in $g$. We claim that the coefficients of this power series are in fact exactly the non-commutative cumulants defined earlier. In other words, $R_A(g)$ is the cumulant generating function:

$$R_A(g) := \sum_{k=1}^{\infty} \kappa_k(A) g^{k-1}. \tag{11.63}$$

To show that these coefficients are indeed the cumulants we first realize that the general equality $R(g) = \mathfrak{z}(g) - 1/g$ is equivalent to

$$z\mathfrak{g}_A(z) - 1 = \mathfrak{g}_A(z) R_A(\mathfrak{g}_A(z)). \tag{11.64}$$

We can compute the power series of the two sides of this equality:

$$z\mathfrak{g}_A(z) - 1 = \sum_{k=1}^{\infty} \frac{m_k}{z^k}, \tag{11.65}$$

where $m_k := \tau(A^k)$ denotes the $k$th moment, and

$$\mathfrak{g}_A(z) R_A(\mathfrak{g}_A(z)) = \sum_{k=1}^{\infty} \kappa_k \left( \frac{1}{z} + \sum_{\ell=1}^{\infty} \frac{m_l}{z^{\ell+1}} \right)^k. \tag{11.66}$$

Equating the right hand sides of Eqs. (11.65) and (11.66) and matching powers of $1/z$ we get recursive relations between moments ($m_k$) and cumulants ($\kappa_k$):

$$\begin{aligned}
m_1 &= \kappa_1 && \Rightarrow \quad m_1 = \kappa_1, \\
m_2 &= \kappa_2 + \kappa_1 m_1 && \Rightarrow \quad m_2 = \kappa_2 + \kappa_1^2, \\
m_3 &= \kappa_3 + 2\kappa_2 m_1 + \kappa_1 m_2 && \Rightarrow \quad m_3 = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3, \\
m_4 &= \kappa_4 + 4\kappa_3 m_1 + \kappa_2[2m_2 + m_1^2] + \kappa_1 m_3 && \Rightarrow \quad m_4 = \kappa_4 + 6\kappa_2\kappa_1^2 + 2\kappa_2^2 + 4\kappa_3\kappa_1 + \kappa_1^4.
\end{aligned} \tag{11.67}$$

By looking at the $z^{-k}$ term coming from the $[1/z + \cdots]^k$ term in Eq. (11.66) we realize that $m_k = \kappa_k + \cdots$ where "$\cdots$" are homogeneous combinations of lower order $\kappa_k$ and $m_k$. Since the coefficients of the power series Eq. (11.63) are additive under addition of free variables and obey the property

$$\kappa_k(A) = \tau(A^k) + \text{homogeneous products of lower order moments}, \tag{11.68}$$

they are therefore the cumulants defined in Section 11.3.2.

### 11.3.5 Cumulants and Non-Crossing Partitions

We saw that Eq. (11.64) can be used to compute cumulants iteratively. Actually that equation can be translated into a systematic relation between moments and cumulants:

$$m_n = \sum_{\pi \in \mathrm{NC}(n)} \kappa_{\pi_1} \cdots \kappa_{\pi_{\ell_\pi}}, \tag{11.69}$$

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

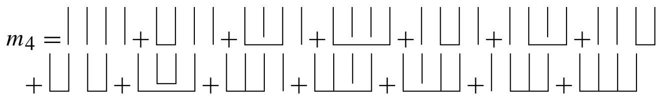

Figure 11.1 List of all non-crossing partitions of four elements. In Eq. (11.69) for $m_4$, the first partition contributes $\kappa_1\kappa_1\kappa_1\kappa_1 = \kappa_1^4$. The next six all contribute $\kappa_2\kappa_1^2$ and so forth. We read $m_4 = \kappa_1^4 + 6\kappa_2\kappa_1^2 + 2\kappa_2^2 + 4\kappa_3\kappa_1 + \kappa_4$.

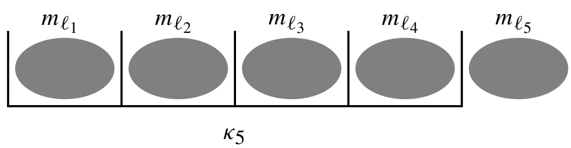

Figure 11.2 A typical non-crossing diagram contributing to a large moment $m_n$. In this example the first element is connected to four others (giving a factor of $\kappa_5$) which breaks the diagram into five disjoint non-crossing diagrams contributing a factor $m_{\ell_1}m_{\ell_2}m_{\ell_3}m_{\ell_4}m_{\ell_5}$. Note that we must have $\ell_1 + \ell_2 + \ell_3 + \ell_4 + \ell_5 = n$.

where $\pi \in \mathrm{NC}(n)$ indicates that the sum is over all possible non-crossing partitions of $n$ elements. For any such partition $\pi$ the integers $\{\pi_1, \pi_2, \ldots, \pi_{\ell_\pi}\}$ $(1 \le \ell_\pi \le n)$ equal the number elements in each group (see Fig. 11.3). They satisfy

$$n = \sum_{k=1}^{\ell_\pi} \pi_k. \tag{11.70}$$

We will show that, provided we define cumulants by Eq. (11.69), we recover Eq. (11.64). But before we do so, let us first show this relation on a simple example. Figure 11.1 shows the computation of the fourth moment in terms of the cumulants.

The argument is very similar to the recursion relation obtained for Catalan numbers where we considered non-crossing pair partitions (see Section 3.2.3). Here the argument is slightly more complicated as we have partitions of all possible sizes. We consider the moment $m_n$ for $n \ge 1$. We break down the sum over all non-crossing partitions of $n$ elements by looking at $\ell$, the size of the set containing the first element (for example in Fig. 11.3, the first element belongs to a set of size $\ell = 5$). The size of this first set can be $1 \le \ell \le n$. This initial set breaks the partition into $\ell$ (possibly empty) disjoint smaller partitions. They must be disjoint, otherwise there would be a crossing. In Figure 11.2 we show how an initial 5-set breaks the full partition into 5 blocks. In each of these blocks, every non-crossing partition is possible, the only constraint is that the total size of the partition must be $n$. The sum over all possible non-crossing partitions of size $k$ of the relevant $\kappa$'s is the moment $m_k$. Note that the empty partition contributes a multiplicative factor 1, so we define $m_0 \equiv 1$. Putting everything together we obtain the following recursion relation for $m_n$:

$$m_n = \sum_{\ell=1}^{n} \kappa_\ell \prod_{\substack{k_1,k_2,\ldots,k_\ell \ge 0 \\ k_1+k_2+\cdots+k_\ell = n-\ell}} m_{k_1} m_{k_2} \ldots m_{k_\ell}. \tag{11.71}$$

Let us multiply both sides of this equation by $z^{-n}$ and sum over $n$ from 1 to $\infty$. The left hand side gives $zg(z) - 1$, by definition of $g(z)$. The right hand side reads

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

Figure 11.3 Generic non-crossing partition of 23 elements with two singletons, five doublets, two triplets, and one quintent, such that $23 = 5 + 2 \cdot 3 + 5 \cdot 2 + 2 \cdot 1$. In Eq. (11.69), this particular partition appears for $m_{23}$ and contributes $\kappa_5 \kappa_3^2 \kappa_2^5 \kappa_1^2$.

$\tau(A_1 A_2 A_3) =$ 

Figure 11.4 List of all non-crossing partitions of three elements. From this we get Eq. (11.74) for three elements: $\tau(A_1A_2A_3) = \kappa_1(A_1)\kappa_1(A_2)\kappa_1(A_3) + \kappa_2(A_1, A_2)\kappa_1(A_3) + \kappa_2(A_1, A_3)\kappa_1(A_2) + \kappa_2(A_2, A_3)\kappa_1(A_3) + \kappa_3(A_1, A_2, A_3)$.

$$\sum_{n=1}^{\infty} \sum_{\ell=1}^{n} \kappa_\ell \prod_{k_1, k_2, \ldots, k_\ell \geq 0} \delta_{k_1+k_2+\cdots+k_\ell = n-\ell} \frac{m_{k_1} m_{k_2} \ldots m_{k_\ell}}{z^{1+k_1} z^{1+k_2} \ldots z^{1+k_\ell}}, \tag{11.72}$$

which can be transformed into

$$\sum_{\ell=1}^{\infty} \kappa_\ell \left[ \sum_{k_1=0}^{\infty} \frac{m_{k_1}}{z^{1+k_1}} \right]^\ell = g(z) R(g(z)), \tag{11.73}$$

where we have used Eq. (11.63). We thus recover exactly Eq. (11.64), showing that the relation (11.69) is equivalent to our previous definition of the free cumulants.

It is interesting to contrast the moment–cumulant relation in the standard (commutative) case (Eq. (11.23)) and the free (non-commutative) case (Eq. (11.69)). Both can be written as a sum over all partitions on $n$ elements; in the standard case all partitions are allowed, while in the free case the sum is only over non-crossing partitions.

### *11.3.6 Freeness as the Vanishing of Mixed Cumulants*

We have defined freeness in Section 11.3.1 as the property of two variables $A$ and $B$ such that the trace of any mixed combination of traceless polynomials in $A$ and in $B$ vanishes. There exists another equivalent definition of freeness, namely that every mixed cumulant of $A$ and $B$ vanish. To make sense of this definition we first need to introduce cumulants of several variables. They are defined recursively by

$$\tau(A_1 A_2 \ldots A_n) = \sum_{\pi \in \mathrm{NC}(n)} \kappa_\pi (A_1 A_2 \ldots A_n), \tag{11.74}$$

where the $A_i$'s are not necessarily distinct and $\mathrm{NC}(n)$ is the set of all non-crossing partitions of $n$ elements. Here

$$\kappa_\pi (A_1 A_2 \ldots A_n) = \kappa_{\pi_1}(\ldots) \ldots \kappa_{\pi_\ell}(\ldots) \tag{11.75}$$

are the products of cumulants of variables belonging to the same group of the corresponding partition – see Figure 11.4 for an illustration. We also call these generalized cumulants the free cumulants.

When all the variables in Eq. (11.74) are the same ($A_i = A$) we recover the previous definition of cumulants with a slightly different notation (e.g. $\kappa_3(A, A, A) \equiv \kappa_3(A)$).

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

Cumulants with more than one variable are called mixed cumulants (e.g. $\kappa_4(A, A, B, A)$). By applying Eq. (11.74) we find for the low generalized cumulants of two variables

$$\begin{aligned} m_1(A) &= \kappa_1(A), \\ m_2(A, B) &= \kappa_1(B)\kappa_1(B) + \kappa_2(A, B), \\ m_3(A, A, B) &= \kappa_1(A)^2\kappa_1(B) + \kappa_2(A, A)\kappa_1(B) + 2\kappa_2(A, B)\kappa_1(A) + \kappa_3(A, A, B). \end{aligned} \tag{11.76}$$

We can now state more precisely the alternative definition of freeness: a set of variables is free if and only if all their mixed cumulants vanish. For example, in the low cumulants listed above, freeness of $A$ and $B$ implies that $\kappa_2(A, B) = \kappa_3(A, A, B) = 0$.

This definition of freeness is easy to generalize to a collection of variables, i.e. a collection of variables is free if all its mixed cumulants are zero. As noted at the end of Section 11.3.7 below, pairwise freeness is not enough to ensure that a collection is free.

We remark that vanishing of mixed cumulants implies that free cumulants are additive. In Speicher's notation, $\kappa_k(A, B, C, \ldots)$ is a multi-linear function in each of its arguments, where $k$ gives the number of variables. Thus we have

$$\begin{aligned} \kappa_k(A + B, A + B, \ldots) &= \kappa_k(A, A, \ldots) + \kappa_k(B, B, \ldots) + \text{mixed cumulants} \\ &= \kappa_k(A, A, \ldots) + \kappa_k(B, B, \ldots), \end{aligned} \tag{11.77}$$

i.e. $\kappa_k$ is additive.

We will give a concrete application of the formalism of free mixed cumulants in Section 12.2.

### *11.3.7 The Central Limit Theorem for Free Variables*

We can now go back and re-read Section 11.2.4. We can replace every occurrence of the word *independent* with *free*, and *cumulant* with *free cumulant*. The LLN now states that the sum of $K$ free identically distributed (FID) variables normalized by $K^{-1}$ converges to a constant (also called a scalar) with the same mean.

Let us define a free Wigner variable as a variable with second free cumulant $\kappa_2 > 0$ and all other free cumulants equal to zero. In other words, a free Wigner variable is such that $R_W(x) = \kappa_2 x$. The CLT then states that the sum of $K$ zero-mean free identical variables normalized by $K^{-1/2}$ converges to a free Wigner variable with the same second cumulant.

In the case where our free random variables are large symmetric random matrices, the Wigner defined here by its cumulant coincides with the Wigner matrices defined in Chapter 2. We indeed saw that the R-transform of a Wigner matrix is given by $R(x) = \sigma^2 x$, i.e. the cumulant generating function has a single term corresponding to $\kappa_2 = \sigma^2$.

Alternatively, we note that the moments of a Wigner are given by the sum over non-crossing *pair* partitions (Eq. (3.26)). Comparing with Eq. (11.69), we realize that partitions containing anything other than pairs must contribute zero, hence only the second cumulant of the Wigner is non-zero.

The LLN and the CLT require variables to be collectively free, in the sense that all mixed cumulants are zero. As is the case with *independence*, pairwise freeness is not enough to ensure freeness as a collection (see footnote on page 161). Indeed, in Section 12.5 we will encounter variables that are pairwise free but not free as a collection. One can have $A$ and $B$ mutually free and both free with respect to $C$ but $A + B$ is not free with respect

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

to $C$. This does not happen for rotationally invariant large matrices but can arise in other constructions. The definition of a *free collection* is just an extension of definition (11.40) including traceless polynomials in all variables in the collection. With this definition, sums of variables in the collection are free from those not included in the sum (e.g. $A + B$ is free from $C$).

### 11.3.8 Subordination Relation for Addition of Free Variables

We now introduce the subordination relation for free addition, which is just a rewriting of the addition of R-transforms. For free $A$ and $B$, we have

$$R_A(g) + R_B(g) = R_{A+B}(g) \Rightarrow \mathfrak{z}_A(g) + R_B(g) = \mathfrak{z}_{A+B}(g), \tag{11.78}$$

where

$$\mathfrak{g}_{A+B}(\mathfrak{z}_{A+B}) = g = \mathfrak{g}_A(z_A) = \mathfrak{g}_A\left(\mathfrak{z}_{A+B} - R_B(g)\right). \tag{11.79}$$

We call $z := \mathfrak{z}_{A+B}(g)$, then the above relations give

$$\mathfrak{g}_{A+B}(z) = \mathfrak{g}_A\left(z - R_B(\mathfrak{g}_{A+B}(z))\right), \tag{11.80}$$

which is called a subordination relation (compare with Eq. (10.2)).

## 11.4 Free Product

In the previous section, we have studied the property of the sum of free random variables. In the case of commuting variables, the question of studying the product of (positive) random variables is trivial, since taking the logarithm of this product we are back to the problem of sums again. In the case of non-commuting variables, things are more interesting. We will see below that one needs to introduce the so-called S-transform, which is the counterpart of the R-transform for products of free variables.

We start by noticing that the free product of traceless variables is trivial. If $A, B$ are free and $\tau(A) = \tau(B) = 0$, we have

$$\tau((AB)^k) = \tau(ABAB \ldots AB) = 0. \tag{11.81}$$

### 11.4.1 Low Moments of Free Products

We will now compute the first few moments of the free products of two variables with a non-zero trace: $C := AB$ where $A, B$ are free and $\tau(A) \neq 0$, $\tau(B) \neq 0$. Without loss of generality, we can assume that $\tau(A) = \tau(B) = 1$ by rescaling $A$ and $B$. Then

$$\tau(C) = \tau\left((A - \tau(A))(B - \tau(B))\right) + \tau(A)\tau(B) = \tau(A)\tau(B) = 1. \tag{11.82}$$

We can also use (11.74) to get

$$\tau(C) = \kappa_2(A, B) + \kappa_1(A)\kappa_1(B) = \kappa_1(A)\kappa_1(B) = 1, \tag{11.83}$$

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

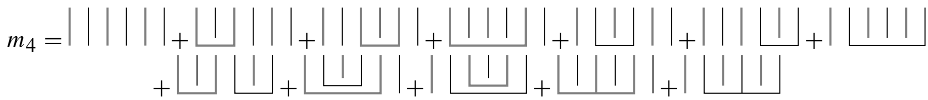


Figure 11.5 List of all non-crossing partitions of six elements contributing to $\tau(ABABAB)$ excluding mixed $AB$ terms. Terms involving $A$ are in thick gray and $B$ in black. Equation (11.86) can be read off from these diagrams. Note that $\kappa_1(A) = \kappa_2(B) = 1$.

since mixed cumulants are zero for mutually free variables. Similarly, using Eq. (11.74) we can get that (see Fig. 11.1 for the non-crossing partitions of four elements)

$$\begin{aligned} \tau(C^2) = \tau(ABAB) &= \kappa_1(A)^2\kappa_1(B)^2 + \kappa_2(A)\kappa_1(B)^2 + \kappa_1(A)^2\kappa_2(B) \\ &= 1 + \kappa_2(A) + \kappa_2(B), \end{aligned} \tag{11.84}$$

which gives

$$\kappa_2(C) := \tau(C^2) - \tau(C)^2 = \kappa_2(A) + \kappa_2(B). \tag{11.85}$$

For the third moment of $C = AB$, we can follow Figure 11.5 and get

$$\begin{aligned} \tau(C^3) &= \tau(ABABAB) \\ &= 1 + 3\kappa_2(A) + 3\kappa_2(B) + 3\kappa_2(A)\kappa_2(B) + \kappa_3(A) + \kappa_3(B), \end{aligned} \tag{11.86}$$

leading to

$$\begin{aligned} \kappa_3(C) :&= \tau(C^3) - 3\tau(C^2)\tau(C) + 2\tau(C)^3 \\ &= \kappa_3(A) + \kappa_3(B) + 3\kappa_2(A)\kappa_2(B). \end{aligned} \tag{11.87}$$

Under free products of unit-trace variables, the mean remains equal to one and the variance is additive. The third cumulant is not additive; it is strictly greater than the sum of the third cumulants unless one of the two variables is the identity (unit scalar).

### *11.4.2 Definition of the S-Transform*

We will now show that the above relations can be encoded into the S-transform $S(t)$ which is multiplicative for products of free variables:

$$S_{AB}(t) = S_A(t)S_B(t) \tag{11.88}$$

for $A$ and $B$ free. To define the S-transform, we first introduce the T-transform as

$$\begin{aligned} \mathfrak{t}_A(\zeta) &= \tau\left[(1-\zeta^{-1}A)^{-1}\right] - 1 \\ &= \zeta\mathfrak{g}_A(\zeta) - 1 \\ &= \sum_{k=1}^{\infty} \frac{m_k}{\zeta^k}. \end{aligned} \tag{11.89}$$

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

The behavior at infinity of the T-transform depends explicitly on $m_1$, the first moment of $A$ ($\mathfrak{t}(\zeta) \sim m_1/\zeta$), unlike the Stieltjes transform which always behaves as $1/z$.

The T-transform has the same singularities as the Stieltjes transform except maybe at zero. When $A$ is a matrix, one can recover the continuous part of its eigenvalue density $\rho(x)$ using the following T-version of the Sokhotski–Plemelj formula:

$$\lim_{\eta \to 0+} \operatorname{Im} \mathfrak{t}(x - i\eta) = \pi x \rho(x). \tag{11.90}$$

Poles in the T-transform indicate Dirac masses. If $\mathfrak{t}(\zeta) \sim A/(\zeta - \lambda_0)$ near $\lambda_0$ then the density is a Dirac mass of amplitude $A/\lambda_0$ at $\lambda_0$. The behavior at zero of the T-transform is a bit different from that of the Stieltjes transform. A regular density at zero gives a regular Stieltjes and hence $\mathfrak{t}(0) = -1$. Deviations from this value indicate a Dirac mass at zero, hence when $\mathfrak{t}(0) \neq -1$, the density has a Dirac at zero of amplitude $\mathfrak{t}(0) + 1$.

The T-transform can also be written as

$$\mathfrak{t}_A(\zeta) = \tau\left[A\,(\zeta - A)^{-1}\right]. \tag{11.91}$$

We define $\zeta_A(t)$ to be the inverse function of $\mathfrak{t}_A(\zeta)$. When $m_1 \neq 0$, $\mathfrak{t}_A$ is invertible for large $\zeta$, and hence $\zeta_A$ exists for small enough $t$. We then define the S-transform as[8]

$$S_A(t) := \frac{t+1}{t\zeta_A(t)}, \tag{11.92}$$

for variables $A$ such that $\tau(A) \neq 0$.

Let us compute the S-transform of the identity $S_{\mathbf{1}}(t)$:

$$\mathfrak{t}_{\mathbf{1}}(\zeta) = \frac{1}{\zeta - 1} \Rightarrow \zeta_{\mathbf{1}}(t) = \frac{t+1}{t} \Rightarrow S_{\mathbf{1}}(t) = 1, \tag{11.93}$$

as expected as the identity is free with respect to any variable. The S-transform scales in a simple way with the variable $A$. To find its scaling we first note that

$$\mathfrak{t}_{\alpha A}(\zeta) = \tau\left[(1 - (\alpha^{-1}\zeta)^{-1}A)^{-1}\right] - 1 = \mathfrak{t}_A(\zeta/\alpha), \tag{11.94}$$

which gives

$$\zeta_{\alpha A}(t) = \alpha\zeta_A(t). \tag{11.95}$$

Then, using (11.92), we get that

$$S_{\alpha A}(t) = \alpha^{-1} S_A(t). \tag{11.96}$$

The above scaling is slightly counterintuitive but it is consistent with the fact that $S_A(0) = 1/\tau(A)$. We will be focusing on unit trace objects such that $S(0) = 1$.

[8] Most authors prefer to define the S-transform in terms of the moment generating function $\psi(z) := \mathfrak{t}(1/z)$. The definition $S(t) = \psi^{-1}(t)(t+1)/t$ is equivalent to ours ($\psi^{-1}(t)$ is the functional inverse of $\psi(z)$). We prefer to work with the T-transform as the function $t(\zeta)$ has an analytic structure very similar to that of $g(z)$. The function $\psi(z)$ is analytic near zero and singular for large values of $z$ corresponding to the reciprocal of the eigenvalues.

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

The construction of the S-transform relies on the properties of mixed moments of free variables. In that respect it is closely related to the R-transform. Using the relation $t_A(\zeta) = \zeta \mathfrak{g}_A(\zeta) - 1$, one can get the following relationships between $R_A$ and $S_A$:

$$S_A(t) = \frac{1}{R_A(tS_A(t))}, \quad R_A(g) = \frac{1}{S_A(gR_A(g))}. \tag{11.97}$$

### 11.4.3 Multiplicativity of the S-Transform

We can now show the multiplicative property (11.88). The proof is similar to the one given for the additive case and is adapted from it.

We fix $t$ and let $\zeta_A$ and $\zeta_B$ be the inverse T-transforms of $\mathfrak{t}_A$ and $\mathfrak{t}_B$. Then we define $E_A$ through

$$1 + t + E_A = (1 - A/\zeta_A)^{-1}, \tag{11.98}$$

and similarly for $E_B$. We have $\tau(E_A) = 0$, $\tau(E_B) = 0$, and, since $A$ and $B$ are free, $E_A, E_B$ are also free. Then we have

$$\frac{A}{\zeta_A} = 1 - (1 + t + E_A)^{-1}, \tag{11.99}$$

which gives

$$\begin{aligned}\frac{AB}{\zeta_A \zeta_B} &= \left[1 - (1 + t + E_A)^{-1}\right]\left[1 - (1 + t + E_B)^{-1}\right] \\ &= (1 + t + E_A)^{-1}\left[(t + E_A)(t + E_B)\right](1 + t + E_B)^{-1}.\end{aligned} \tag{11.100}$$

Using the identity

$$t(E_A + E_B) = \frac{t}{1+t}\left[(1 + t + E_A)(1 + t + E_B) - (1+t)^2 - E_A E_B\right], \tag{11.101}$$

we can rewrite the above expression as

$$\begin{aligned}&\frac{AB}{\zeta_A \zeta_B} = \frac{t}{1+t} + (1 + t + E_A)^{-1}\left[-t + \frac{E_A E_B}{1+t}\right](1 + t + E_B)^{-1} \\ &\Rightarrow 1 - \frac{1+t}{t}\frac{AB}{\zeta_A \zeta_B} = (1+t)(1 + t + E_A)^{-1}\left[1 - \frac{E_A E_B}{t(1+t)}\right](1 + t + E_B)^{-1} \\ &\Rightarrow \left[1 - \frac{1+t}{t}\frac{AB}{\zeta_A \zeta_B}\right]^{-1} = \frac{1}{1+t}(1 + t + E_B)\left[1 - \frac{E_A E_B}{t(1+t)}\right]^{-1}(1 + t + E_A).\end{aligned} \tag{11.102}$$

Using the expansion

$$\left[1 - \frac{E_A E_B}{t(1+t)}\right]^{-1} = \sum_{n=0}\left(\frac{E_A E_B}{t(1+t)}\right)^n, \tag{11.103}$$

one can check that

$$\tau\left[(1 + t + E_B)\left[1 - \frac{E_A E_B}{t(1+t)}\right]^{-1}(1 + t + E_A)\right] = (1+t)^2, \tag{11.104}$$

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

where we used the freeness condition for $E_A$ and $E_B$. Thus we get that

$$\tau\left\{\left[1-\frac{1+t}{t}\frac{AB}{\zeta_A\zeta_B}\right]^{-1}\right\}=1+t \quad\Rightarrow\quad \mathfrak{t}_{AB}\left(\frac{t\zeta_A\zeta_B}{1+t}\right)=t, \tag{11.105}$$

which gives that

$$S_{AB}(t)=S_A(t)S_B(t) \tag{11.106}$$

thanks to the definition (11.92).

### *11.4.4 Subordination Relation for the Free Product*

We next derive a subordination relation for the free product using (11.88) and (11.92):

$$S_{AB}(t)=S_A(t)S_B(t) \quad\Rightarrow\quad \zeta_{AB}(t)=\frac{\zeta_A(t)}{S_B(t)}, \tag{11.107}$$

where

$$\mathfrak{t}_{AB}(\zeta_{AB}(t))=t=\mathfrak{t}_A(\zeta_A(t))=\mathfrak{t}_A(\zeta_{AB}(t)S_B(t)). \tag{11.108}$$

We call $\zeta := \zeta_{AB}(t)$, then the above relations give

$$\mathfrak{t}_{AB}(\zeta)=\mathfrak{t}_A\left(\zeta S_B(\mathfrak{t}_{AB}(\zeta))\right), \tag{11.109}$$

which is the subordination relation for the free product. In fact, the above is true even when $S_A$ does not exist, e.g. when $\tau(A)=0$.

When applied to free random matrices, the form $AB$ is not very useful since it is not necessarily symmetric even if $A$ and $B$ are. But if $A \succeq 0$ (i.e. $A$ is positive semi-definite symmetric) and $B$ is symmetric, then $A^{\frac{1}{2}}BA^{\frac{1}{2}}$ has the same moments as $AB$ and is also symmetric. In our applications below we will always encounter the case $A \succeq 0$ and call $A^{\frac{1}{2}}BA^{\frac{1}{2}}$ the free product of $A$ and $B$.

**Exercise 11.4.1 Properties of the S-transform**

(a) Using Eq. (11.92), show that

$$R(x)=\frac{1}{S(xR(x))}. \tag{11.110}$$

Hint: define $t=xR(x)=zg-1$ and identify $x$ as $g$.

(b) For a variable such that $\tau(M)=\kappa_1=1$, write $S(t)$ as a power series in $t$, compute the first few terms of the powers series, up to (and including) the $t^2$ term, using Eq. (11.110) and Eq. (11.63). You should find

$$S(t)=1-\kappa_2 t+(2\kappa_2^2-\kappa_3)t^2+O(t^3). \tag{11.111}$$

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

(c) We have shown that, when **A** and **B** are mutually free with unit trace,

$$\tau(\mathbf{AB}) = 1, \tag{11.112}$$

$$\tau(\mathbf{ABAB}) - 1 = \kappa_2(\mathbf{A}) + \kappa_2(\mathbf{B}), \tag{11.113}$$

$$\tau(\mathbf{ABABAB}) = \kappa_3(\mathbf{A}) + \kappa_3(\mathbf{B}) + 3\kappa_2(\mathbf{A})\kappa_2(\mathbf{B}) + 3(\kappa_2(\mathbf{A}) + \kappa_2(\mathbf{B})) + 1. \tag{11.114}$$

Show that these relations are compatible with $S_{\mathbf{AB}}(t) = S_{\mathbf{A}}(t)S_{\mathbf{B}}(t)$ and the first few terms of your power series in (b).

(d) Consider $\mathbf{M}_1 = \mathbf{1} + \sigma_1\mathbf{W}_1$ and $\mathbf{M}_2 = \mathbf{1} + \sigma_2\mathbf{W}_2$ where $\mathbf{W}_1$ and $\mathbf{W}_2$ are two different (free) unit Wigner matrices and both $\sigma$'s are less than $1/2$. $\mathbf{M}_1$ and $\mathbf{M}_2$ have $\kappa_3 = 0$ and are positive definite in the large $N$ limit. What is $\kappa_3(\mathbf{M}_1\mathbf{M}_2)$?

**Exercise 11.4.2 S-transform of the matrix inverse**

(a) Consider $\mathbf{M}$ an invertible symmetric random matrix and $\mathbf{M}^{-1}$ its inverse. Using Eq. (11.89), show that

$$\mathfrak{t}_{\mathbf{M}}(\zeta) + \mathfrak{t}_{\mathbf{M}^{-1}}\left(\frac{1}{\zeta}\right) + 1 = 0. \tag{11.115}$$

(b) Using Eq. (11.115), show that

$$S_{\mathbf{M}^{-1}}(x) = \frac{1}{S_{\mathbf{M}}(-x-1)}. \tag{11.116}$$

Hint: write $u(x) = 1/\zeta(t)$ where $u(x)$ is such that $x = \mathfrak{t}_{\mathbf{M}^{-1}}(u(x))$. Equation (11.115) is then equivalent to $x = -1 - t$.

## *Bibliographical Notes*

- The concept of freeness was introduced in
  - D. Voiculescu. Symmetries of some reduced free product C*-algebras. In H. Araki et al. (eds.), *Operator Algebras and Their Connections with Topology and Ergodic Theory*. Lecture Notes in Mathematics, volume 1132. Springer, Berlin, Heidelberg, 1985,
  - D. Voiculescu. Limit laws for random matrices and free products. *Inventiones mathematicae*, 104(1):201–220, 1991,

  with the second reference making the link to RMT.
- For general introductions to freeness, see
  - J. A. Mingo and R. Speicher. *Free Probability and Random Matrices*. Springer, New York, 2017,

https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press

  - J. Novak. Three lectures on free probability. In *Random Matrix Theory, Interacting Particle Systems, and Integrable Systems*. Cambridge University Press, Cambridge, 2014.


- For a more operational approach to freeness, see

  - A. M. Tulino and S. Verdú. *Random Matrix Theory and Wireless Communications*. Now publishers, Hanover, Mass., 2004,.

- The proof of the additivity of the R-transform presented here is from

  - T. Tao. *Topics in Random Matrix Theory*. American Mathematical Society, Providence, Rhode Island, 2012,


  see also

  - P. Zinn-Justin. Adding and multiplying random matrices: A generalization of Voiculescu's formulas. *Physical Review E*, 59:4884–4888, 1999,


  for an alternative point of view.
- Central limit theorems and Lévy-stable free variables:

  - H. Bercovici, V. Pata, and P. Biane. Stable laws and domains of attraction in free probability theory. *Annals of Mathematics*, 149(3):1023–1060, 1999,
  - Z. Burda, J. Jurkiewicz, M. A. Nowak, G. Papp, and I. Zahed. Free random Lévy and Wigner-Lévy matrices. *Physical Review E*, 75:051126, 2007.


https://doi.org/10.1017/9781108768900.014 Published online by Cambridge University Press