# Part 1

# Basic concepts

Cambridge University Press
978-0-521-85852-6 - Lectures on the Combinatorics of Free Probability
Alexandru Nica and Roland Speicher
Excerpt
More information

LECTURE 1

# Non-commutative probability spaces and distributions

Since we are interested in the combinatorial aspects of free probability, we will focus on a framework which is stripped of its analytical structure (i.e. where we ignore the metric or topological structure of the spaces involved). The reason for the existence of this monograph is that even so (without analytical structure), the phenomenon of free independence is rich enough to be worth studying. The interesting combinatorial features of this phenomenon come from the fact that *we will allow the algebras of random variables to be non-commutative.* This certainly means that we have to consider a generalized concept of "random variable" (since in the usual meaning of the concept, where a random variable is a function on a probability space, the algebras of random variables would have to be commutative).

## Non-commutative probability spaces

DEFINITION 1.1. (1) A **non-commutative probability space** $(\mathcal{A}, \varphi)$ consists of a unital algebra $\mathcal{A}$ over $\mathbb{C}$ and a unital linear functional

$$\varphi : \mathcal{A} \to \mathbb{C}; \qquad \varphi(1_{\mathcal{A}}) = 1.$$

The elements $a \in \mathcal{A}$ are called **non-commutative random variables** in $(\mathcal{A}, \varphi)$. Usually, we will skip the adjective "non-commutative" and just talk about "random variables $a \in \mathcal{A}$."

An additional property which we will sometimes impose on the linear functional $\varphi$ is that it is a **trace**, i.e. it has the property that

$$\varphi(ab) = \varphi(ba), \quad \forall\, a, b \in \mathcal{A}.$$

When this happens, we say that the non-commutative probability space $(\mathcal{A}, \varphi)$ is **tracial.**

(2) In the framework of part (1) of the definition, suppose that $\mathcal{A}$ is a $*$-algebra, i.e. that $\mathcal{A}$ is also endowed with an antilinear $*$-operation $\mathcal{A} \ni a \mapsto a^* \in \mathcal{A}$, such that $(a^*)^* = a$ and $(ab)^* = b^* a^*$ for all $a, b \in \mathcal{A}$. If we have that

$$\varphi(a^* a) \geq 0, \quad \forall\, a \in \mathcal{A},$$

then we say that the functional $\varphi$ is **positive** and we will call $(\mathcal{A}, \varphi)$ a **$*$-probability space**.

(3) In the framework of a $*$-probability space we can talk about:

• **selfadjoint** random variables, i.e. elements $a \in \mathcal{A}$ with the property that $a = a^*$;

• **unitary** random variables, i.e. elements $u \in \mathcal{A}$ with the property that $u^*u = uu^* = 1$;

• **normal** random variables, i.e. elements $a \in \mathcal{A}$ with the property that $a^*a = aa^*$.

In these lectures we will be mostly interested in $*$-probability spaces, since this is the framework which provides us with a multitude of exciting examples. However, plain non-commutative probability spaces are also useful, because sometimes we encounter arguments relying solely on the linear and multiplicative structure of the algebra involved – these arguments are more easily understood when the $*$-operation is ignored (even if it happened that the algebra had a $*$-operation on it).

REMARKS 1.2. Let $(\mathcal{A}, \varphi)$ be a $*$-probability space.

(1) The functional $\varphi$ is selfadjoint, i.e. it has the property that

$$\varphi(a^*) = \overline{\varphi(a)}, \quad \forall a \in \mathcal{A}.$$

Indeed, since every $a \in \mathcal{A}$ can be written uniquely in the form $a = x+iy$ where $x, y \in \mathcal{A}$ are selfadjoint, the latter equation is immediately seen to be equivalent to the fact that $\varphi(x) \in \mathbb{R}$ for every selfadjoint element $x \in \mathcal{A}$. This in turn is implied by the positivity of $\varphi$ and the fact that every selfadjoint element $x \in \mathcal{A}$ can be written in the form $x = a^*a - b^*b$ for some $a, b \in \mathcal{A}$ (take e.g. $a = (x + 1_{\mathcal{A}})/2$, $b = (x - 1_{\mathcal{A}})/2$).

(2) Another consequence of the positivity of $\varphi$ is that we have:

$$|\varphi(b^*a)|^2 \leq \varphi(a^*a)\varphi(b^*b), \quad \forall\ a, b \in \mathcal{A}. \tag{1.1}$$

The inequality (1.1) is commonly called the **Cauchy–Schwarz inequality** for the functional $\varphi$. It is proved in exactly the same way as the usual Cauchy–Schwarz inequality (see Exercise 1.21 at the end of the lecture).

(3) If an element $a \in \mathcal{A}$ is such that $\varphi(a^*a) = 0$, then the Cauchy–Schwarz inequality (1.1) implies that $\varphi(ba) = 0$ for all $b \in \mathcal{A}$ (hence $a$ is in a certain sense a degenerate element for the functional $\varphi$). We will use the term "faithful" for the situation when no such degenerate elements exist, except for $a = 0$. That is, we make the following definition.

DEFINITION 1.3. Let $(\mathcal{A}, \varphi)$ be a $*$-probability space. If we have the implication:

$$a \in \mathcal{A}, \ \varphi(a^*a) = 0 \ \Rightarrow \ a = 0,$$

then we say that the functional $\varphi$ is **faithful**.

EXAMPLES 1.4. (1) Let $(\Omega, \mathcal{Q}, P)$ be a probability space in the classical sense, i.e. $\Omega$ is a set, $\mathcal{Q}$ is a $\sigma$-field of measurable subsets of $\Omega$ and $P : \mathcal{Q} \to [0,1]$ is a probability measure. Let $\mathcal{A} = L^\infty(\Omega, P)$, and let $\varphi$ be defined by

$$\varphi(a) = \int_\Omega a(\omega)\, dP(\omega), \quad a \in \mathcal{A}.$$

Then $(\mathcal{A}, \varphi)$ is a $*$-probability space (the $*$-operation on $\mathcal{A}$ is the operation of complex-conjugating a complex-valued function). The random variables appearing in this example are thus genuine random variables in the sense of "usual" probability theory.

The reader could object at this point that the example presented in the preceding paragraph only deals with genuine random variables that are bounded, and thus misses for instance the most important random variables from usual probability – those having a Gaussian distribution. We can overcome this problem by replacing the algebra $L^\infty(\Omega, P)$ with:

$$L^{\infty-}(\Omega, P) := \bigcap_{1 \le p < \infty} L^p(\Omega, P).$$

That is, we can make $\mathcal{A}$ become the algebra of genuine random variables which have finite moments of all orders. (The fact that $L^{\infty-}(\Omega, P)$ is indeed closed under multiplication follows by an immediate application of the Cauchy–Schwarz inequality in $L^2(\Omega, P)$ – cf. Exercise 1.22 at the end of the lecture.) In this enlarged version, our algebra of random variables will then contain the Gaussian ones.

Of course, one could also point out that in classical probability there are important cases of random variables which do not have moments of all orders. These ones, unfortunately, are beyond the scope of the present set of lectures – we cannot catch them in the framework of Definition 1.1.

(2) Let $d$ be a positive integer, let $M_d(\mathbb{C})$ be the algebra of $d \times d$ complex matrices with usual matrix multiplication, and let $\mathrm{tr} : M_d(\mathbb{C}) \to \mathbb{C}$ be the normalized trace,

$$\mathrm{tr}(a) = \frac{1}{d} \cdot \sum_{i=1}^{d} \alpha_{ii} \qquad \text{for} \qquad a = (\alpha_{ij})_{i,j=1}^{d} \in M_d(\mathbb{C}). \tag{1.2}$$

Then $(M_d(\mathbb{C}), \mathrm{tr})$ is a $*$-probability space (where the $*$-operation is given by taking both the transpose of the matrix and the complex conjugate of the entries).

(3) The above examples (1) and (2) can be "put together" into one example where the algebra consists of all the $d \times d$ matrices over $L^{\infty-}(\Omega, P)$ :

$$\mathcal{A} = M_d(\, L^{\infty-}(\Omega, P)\,),$$

and the functional $\varphi$ on it is

$$\varphi(a) := \int \mathrm{tr}(a(\omega))\, dP(\omega), \quad a \in \mathcal{A}.$$

The non-commutative random variables obtained here are thus **random matrices** over $(\Omega, \mathcal{Q}, P)$. (Observe that this example is obtained by starting with the space in Example 1.4.1 and by performing the $d \times d$ matrix construction described in Exercise 1.23.) We will elaborate more on random matrix examples later (cf. Lectures 22 and 23).

(4) Let $G$ be a group, and let $\mathbb{C}G$ denote its **group algebra**. That is, $\mathbb{C}G$ is a complex vector space having a basis indexed by the elements of $G$, and where the operations of multiplication and $*$-operation are defined in the natural way:

$$\mathbb{C}G := \Big\{ \sum_{g \in G} \alpha_g g \mid \alpha_g \in \mathbb{C}, \text{ only finitely many } \alpha_g \neq 0 \Big\},$$

with

$$\Big(\sum \alpha_g g\Big) \cdot \Big(\sum \beta_h h\Big) := \sum_{g,h} \alpha_g \beta_h (gh) = \sum_{k \in G} \Big( \sum_{g,h:\, gh=k} \alpha_g \beta_h \Big) k,$$

and

$$\Big(\sum \alpha_g g\Big)^* := \sum \bar{\alpha}_g g^{-1}.$$

Let $e$ be the unit element of $G$. The functional $\tau_G : \mathbb{C}G \to \mathbb{C}$ defined by the formula

$$\tau_G\Big(\sum \alpha_g g\Big) := \alpha_e$$

is called the canonical trace on $\mathbb{C}G$. Then $(\mathbb{C}G, \tau_G)$ is a $*$-probability space. It is easily verified that $\tau_G$ is indeed a trace (in the sense of Definition 1.1.1) and is faithful (in the sense of Definition 1.3).

(5) Let $\mathcal{H}$ be a Hilbert space and let $B(\mathcal{H})$ be the algebra of all bounded linear operators on $\mathcal{H}$. This is a $*$-algebra, where the adjoint $a^*$ of an operator $a \in B(\mathcal{H})$ is uniquely determined by the fact that

$$\langle a\xi, \eta \rangle = \langle \xi, a^* \eta \rangle, \quad \forall\, \xi, \eta \in \mathcal{H}.$$

Suppose that $\mathcal{A}$ is a unital $*$-subalgebra of $B(\mathcal{H})$ and that $\xi_o \in \mathcal{H}$ is a vector of norm one ($||\xi_o|| := \langle \xi_o, \xi_o \rangle^{1/2} = 1$). Then we get an example of $*$-probability space $(\mathcal{A}, \varphi)$, where $\varphi : \mathcal{A} \to \mathbb{C}$ is defined by:

$$\varphi(a) := \langle a\xi_o, \xi_o \rangle, \quad a \in \mathcal{A}. \tag{1.3}$$

A linear functional as defined in (1.3) is usually called a **vector-state** (on the algebra of operators $\mathcal{A}$).

Exercise 1.5. (1) Verify that in each of the examples described in 1.4, the functional considered as part of the definition of the $*$-probability space is indeed positive.

(2) Show that in Examples 1.4.1–1.4.4, the functional considered as part of the definition of the $*$-probability space is a faithful trace.

Definition 1.6. (1) A **morphism** between two $*$-probability spaces $(\mathcal{A}, \varphi)$ and $(\mathcal{B}, \psi)$ is a unital $*$-algebra homomorphism $\Phi : \mathcal{A} \to \mathcal{B}$ with the property that $\psi \circ \Phi = \varphi$.

(2) In the case when $(\mathcal{B}, \psi)$ is a $*$-probability space of the special kind discussed in Example 1.4.5, we will refer to a morphism $\Phi$ from $(\mathcal{A}, \varphi)$ to $(\mathcal{B}, \psi)$ as a **representation** of $(\mathcal{A}, \varphi)$. So, to be precise, giving a representation of $(\mathcal{A}, \varphi)$ amounts to giving a triple $(\mathcal{H}, \Phi, \xi_o)$ where $\mathcal{H}$ is a Hilbert space, $\Phi : \mathcal{A} \to B(\mathcal{H})$ is a unital $*$-homomorphism, and $\xi_o \in \mathcal{H}$ is a vector of norm one, such that $\varphi(a) = \langle \Phi(a)\xi_o, \xi_o \rangle$ for all $a \in \mathcal{A}$.

Remark 1.7. The $*$-probability spaces appearing in Examples 1.4.1, 1.4.2 and 1.4.4 have natural representations, on Hilbert spaces related to how the algebras of random variables were constructed – see Exercise 1.25 at the end of the lecture.

## $*$-distributions (case of normal elements)

A fundamental concept in the statistical study of random variables is that of distribution of a random variable. In the framework of a $*$-probability space $(\mathcal{A}, \varphi)$, the appropriate concept to consider is the $*$-*distribution* of an element $a \in \mathcal{A}$. Roughly speaking, the $*$-distribution of $a$ has to be some "standardized" way of reading the values of $\varphi$ on the unital $*$-subalgebra generated by $a$.

We start the discussion of $*$-distributions with the simpler case when $a \in \mathcal{A}$ is normal (i.e. is such that $a^*a = aa^*$). In this case the unital $*$-algebra generated by $a$ is

$$\mathcal{A} := \text{span}\{a^k (a^*)^l \mid k, l \geq 0\}; \tag{1.4}$$

the job of the $*$-distribution of $a$ must thus be to keep track of the values $\varphi(\, a^k (a^*)^l \,)$, where $k$ and $l$ run in $\mathbb{N} \cup \{0\}$. The kind of object which does this job and which we prefer to have whenever possible is a compactly supported probability measure on $\mathbb{C}$.

Definition 1.8. Let $(\mathcal{A}, \varphi)$ be a $*$-probability space and let $a$ be a normal element of $\mathcal{A}$. If there exists a compactly supported probability measure $\mu$ on $\mathbb{C}$ such that

$$\int z^k \, \bar{z}^l \, d\mu(z) = \varphi(a^k (a^*)^l), \qquad \text{for every } k, l \in \mathbb{N}, \tag{1.5}$$

then this $\mu$ is uniquely determined and we will call the probability measure $\mu$ the **$*$-distribution** of $a$.

Remarks 1.9. (1) The fact that a compactly supported probability measure $\mu$ on $\mathbb{C}$ is uniquely determined by how it integrates functions of the form $z \mapsto z^k \bar{z}^l$ with $k, l \in \mathbb{N}$ is an immediate consequence of the Stone–Weierstrass theorem. Or more precisely: due to Stone–Weierstrass, $\mu$ is determined as a linear functional on the space $C(K)$ of complex-valued continuous functions on $K$, where $K$ is the support of $\mu$; it is then well known that this in turn determines $\mu$ uniquely.

(2) It is not said that every normal element in a $*$-probability space has to have a $*$-distribution in the sense defined above. But this turns out to be true in a good number of important examples. Actually, this is *always* true when we look at $*$-probability spaces which have a representation on a Hilbert space, in the sense of the above Definition 1.6 (see Corollary 3.14 in Lecture 3); and civilized examples do have representations on Hilbert spaces – see Lecture 7.

Remark 1.10. *(The case of a selfadjoint element)*
Let $(\mathcal{A}, \varphi)$ be a $*$-probability space, and let $a$ be a selfadjoint element of $\mathcal{A}$ (that is, we have $a = a^*$, which implies in particular that $a$ is normal). Suppose that $a$ has $*$-distribution $\mu$, in the sense of Definition 1.8. Then $\mu$ is supported in $\mathbb{R}$. Indeed, we have

$$\begin{aligned} \int_{\mathbb{C}} |\, z - \bar{z} \,|^2 \, d\mu(z) &= \int_{\mathbb{C}} (z - \bar{z}\,)(\bar{z} - z) \, d\mu(z) \\ &= \int_{\mathbb{C}} 2z\bar{z} - z^2 - \bar{z}^2 \, d\mu(z) \\ &= 2\varphi(aa^*) - \varphi(a^2) - \varphi(\,(a^*)^2\,) = 0. \end{aligned}$$

Since $z \mapsto |\, z - \bar{z} \,|^2$ is a continuous non-negative function, we must have that $z - \bar{z}$ vanishes on the support $\mathrm{supp}(\mu)$ of our measure, and hence:

$$\mathrm{supp}(\mu) \subset \{z \in \mathbb{C} \mid z = \bar{z}\} = \mathbb{R}.$$

So in this case $\mu$ is really a measure on $\mathbb{R}$, and Equation (1.5) is better written in this case as

$$\int t^p \, d\mu(t) = \varphi(a^p), \qquad \forall \, p \in \mathbb{N}. \tag{1.6}$$

Conversely, suppose that we have a compactly supported measure $\mu$ on $\mathbb{R}$ such that (1.6) holds. Then clearly $\mu$ is the $*$-distribution of $a$ in the sense of Definition 1.8 (because $\int z^k \bar{z}^l \, d\mu(z)$ becomes $\int t^{k+l} \, d\mu(t)$, while $\varphi( a^k (a^*)^l )$ becomes $\varphi(a^{k+l})$).

The conclusion of this discussion is that for a selfadjoint element $a \in \mathcal{A}$ it would be more appropriate to talk about the **distribution** of $a$ (rather than talking about its $*$-distribution); this is defined as a compactly supported measure on $\mathbb{R}$ such that (1.6) holds. But there is actually no harm in treating $a$ as a general normal element, and in looking for its $*$-distribution, since in the end we arrive at the same result.

EXAMPLES 1.11. (1) Consider the framework of Example 1.4.1, where the algebra of random variables is $L^\infty(\Omega, P)$. Let $a$ be an element in $\mathcal{A}$; in other words, $a$ is a bounded measurable function, $a : \Omega \to \mathbb{C}$. Let us consider the probability measure $\nu$ on $\mathbb{C}$ which is called "the distribution of $a$" in usual probability; this is defined by

$$\nu(E) \ = \ P( \{\omega \in \Omega : a(\omega) \in E\} ), \quad E \subset \mathbb{C} \text{ Borel set.} \tag{1.7}$$

Note that $\nu$ is compactly supported. More precisely, if we choose a positive $r$ such that $|a(\omega)| \leq r$, $\forall \ \omega \in \Omega$, then it is clear that $\nu$ is supported in the closed disc centered at 0 and of radius $r$.

Now, $a$ is a normal element of $\mathcal{A}$ (all the elements of $\mathcal{A}$ are normal, since $\mathcal{A}$ is commutative). So it makes sense to place $a$ in the framework of Definition 1.8. We will show that the above measure $\nu$ is exactly the $*$-distribution of $a$ in this framework.

Indeed, Equation (1.7) can be read as

$$\int_{\mathbb{C}} f(z) \, d\nu(z) \ = \ \int_{\Omega} f( a(\omega) ) \, dP(\omega), \tag{1.8}$$

where $f$ is the characteristic function of the set $E$. By going through the usual process of taking linear combinations of characteristic functions, and then doing approximations of a bounded measurable function by step functions, we see that Equation (1.8) actually holds for every bounded measurable function $f : \mathbb{C} \to \mathbb{C}$. (The details of this are left to the reader.) Finally, let $k, l$ be arbitrary non-negative integers, and let $r > 0$ be such that $|a(\omega)| \leq r$ for every $\omega \in \Omega$. Consider a bounded measurable function $f : \mathbb{C} \to \mathbb{C}$ such that $f(z) = z^k \bar{z}^l$ for every $z \in \mathbb{C}$ having $|z| \leq r$. Since $\nu$ is supported in the closed disc of radius $r$ centered at 0, it follows that

$$\int_{\mathbb{C}} f(z)\, d\nu(z) \;=\; \int_{\mathbb{C}} z^k \bar{z}^l \, d\nu(z),$$

and, consequently, that

$$\int_{\Omega} f(\,a(\omega)\,)\, dP(\omega) \;=\; \int_{\Omega} a(\omega)^k\, \overline{a(\omega)}^{\,l}\, dP(\omega) \;=\; \varphi(\,a^k (a^*)^l\,).$$

Thus for this particular choice of $f$, Equation (1.8) gives us that

$$\int_{\mathbb{C}} z^k \bar{z}^l \, d\nu(z) \;=\; \varphi(\,a^k (a^*)^l\,),$$

and this is precisely (1.5), implying that $\nu$ is the $*$-distribution of $a$ in the sense of Definition 1.8.

(2) Consider the framework of Example 1.4.2, and let $a \in M_d(\mathbb{C})$ be a normal matrix. Let $\lambda_1, \ldots, \lambda_d$ be the eigenvalues of $a$, counted with multiplicities. By diagonalizing $a$ we find that

$$\mathrm{tr}(a^k (a^*)^l) = \frac{1}{d} \sum_{i=1}^{d} \lambda_i^k \bar{\lambda}_i^l, \qquad k, l \in \mathbb{N}.$$

The latter quantity can obviously be written as $\int z^k \bar{z}^l \, d\mu(z)$, where

$$\mu := \frac{1}{d} \sum_{i=1}^{d} \delta_{\lambda_i} \tag{1.9}$$

($\delta_\lambda$ stands here for the Dirac mass at $\lambda \in \mathbb{C}$). Thus it follows that $a$ has a $*$-distribution $\mu$, which is described by Equation (1.9). Usually this $\mu$ is called the **eigenvalue distribution** of the matrix $a$.

One can consider the question of how to generalize the above fact to the framework of random matrices (as in Example 1.4.3). It can be shown that the formula which appears in place of (1.9) in this case is

$$\mu := \frac{1}{d} \sum_{i=1}^{d} \int_{\Omega} \delta_{\lambda_i(\omega)} \, dP(\omega), \tag{1.10}$$

where $a = a^* \in M_d(L^{\infty-}(\Omega, P))$, and where $\lambda_1(\omega) \leq \cdots \leq \lambda_d(\omega)$ are the eigenvalues of $a(\omega)$, $\omega \in \Omega$. (Strictly speaking, Equation (1.10) requires an extension of the framework used in Definition 1.8, since the resulting averaged eigenvalue distribution $\mu$ will generally not have compact support. See Lecture 22 for more details about this.)

Our next example will be in connection to a special kind of element in a $*$-probability space, called a Haar unitary.