CHAPTER 4

Elements of free probability theory

1. Cumulants and moments in classical probability

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. For random variables X_i on this probability space, define $m_n[X_1, \ldots, X_n] = \mathbf{E}[\prod_{j=1}^n X_j]$ whenever the expectation exists. We will also write $m_0 = 1$. The function $m_\cdot[\cdot]$ is called the *moment* function.

Let $\mathcal{P}_n$ denote the set of all set-partitions of $[n]$. For example, $\mathcal{P}_3$ consists of the five partitions $\{\{1,2,3\}\}$, $\{\{1,2\},\{3\}\}$, $\{\{1,3\},\{2\}\}$, $\{\{2,3\},\{1\}\}$ and $\{\{1\},\{2\},\{3\}\}$. The sets that make up a partition are referred to as *blocks*. Note that the order of the blocks, or of the elements in individual blocks are irrelevant (in other words, the partition $\{\{3\},\{2,1\}\}$ is the same as $\{\{1,2\},\{3\}\}$). For a partition Π we denote the number of blocks by ℓ_Π and the individual blocks by Π_j, $1 \le j \le \ell_\Pi$. If we ever need to be more definite, we shall let Π_1 be the block containing 1, Π_2 to be the block containing the least element not in Π_1 etc.

Definition 77. Define the *cumulant function* $\kappa_n[X_1, \ldots, X_n]$ by the equations

$$m_n[X_1, \ldots, X_n] = \sum_{\Pi \in \mathcal{P}_n} \prod_{j=1}^{\ell_\Pi} \kappa_{|\Pi_j|}[X[\Pi_j]]. \tag{41}$$

Here if $\Pi_j = \{k_1, \ldots, k_r\}$ with $k_1 < k_2 < \ldots < k_r$, then $|\Pi_j| := r$ and $[X[\Pi_j]]$ is the short form for $[X_{k_1}, \ldots, X_{k_r}]$.

Rewrite the first three equations as

$$\kappa_1[X] = m_1[X], \qquad \kappa_2[X,Y] = m_2[X,Y] - \kappa_1[X]\kappa_1[Y]$$
$$\kappa_3[X,Y,Z] = m_3[X,Y,Z] - \kappa_2[X,Y]\kappa_1[Z] - \kappa_2[X,Z]\kappa_1[Y] - \kappa_2[Y,Z]\kappa_1[X] + \kappa_1[X]\kappa_1[Y]\kappa_1[Z]$$

It is clear that we can define κ_1 from the first equation, κ_2 from the second and so on, inductively.

For any $\Pi \in \mathcal{P}_n$, introduce the notation

$$m_\Pi[X_1, \ldots, X_n] = \prod_{j=1}^{\ell_\Pi} m_{|\Pi_j|}[X[\Pi_j]], \qquad \kappa_\Pi[X_1, \ldots, X_n] = \prod_{j=1}^{\ell_\Pi} \kappa_{|\Pi_j|}[X[\Pi_j]].$$

In this notation, the equations defining cumulants may be written as $m_n[X] = \sum_{\Pi \in \mathcal{P}_n} \kappa_\Pi[X]$ where $X = (X_1, \ldots, X_n)$.

Exercise 78. Show that $\kappa_n[X] = \sum_{\Pi \in \mathcal{P}_n} (-1)^{\ell_\Pi - 1} (\ell_\Pi - 1)! m_\Pi[X]$.

The following lemma collects some basic properties of cumulants.

Lemma 79. *(1) Cumulant function is multilinear:* $\kappa_n[cX_1 + dX_1', X_2, \ldots, X_n] = c\kappa_n[X_1, X_2, \ldots, X_n] + d\kappa_n[X_1', X_2, \ldots, X_n]$ *and similarly in each of the other co-ordinates. Further,* κ_n *is symmetric in its arguments. For* $\Pi \in \mathcal{P}_n$, κ_Π *is multilinear but not necessarily symmetric.*

(2) *Assume that* $X = (X_1, \ldots, X_d)$ *is such that* $\mathbf{E}[e^{\langle \mathbf{t}, X\rangle}] < \infty$ *for* $\mathbf{t}$ *in a neighbourhood of* 0 *in* $\mathbb{R}^d$. *Let* $\varphi_X(\mathbf{t}) = \mathbf{E}\left[e^{\langle \mathbf{t}, X\rangle}\right]$ *and* $\psi_X(\mathbf{t}) = \log \mathbf{E}\left[e^{\langle \mathbf{t}, X\rangle}\right]$. *Then,*

$$\varphi_X(\mathbf{t}) = \sum_{n=0}^{\infty} \sum_{i_1,\ldots,i_n=1}^{d} \frac{t_{i_1} \ldots t_{i_n}}{n!} m_n[X_{i_1}, \ldots, X_{i_n}],$$

$$\psi_X(\mathbf{t}) = \sum_{n=1}^{\infty} \sum_{i_1,\ldots,i_n=1}^{d} \frac{t_{i_1} \ldots t_{i_n}}{n!} \kappa_n[X_{i_1}, \ldots, X_{i_n}].$$

(3) *Let* $U = (X_1, \ldots, X_k)$ *and* $V = (X_{k+1}, \ldots, X_d)$. *Then, the following are equivalent.*
   (i) U *and* V *are independent.*
   (ii) $\kappa_n[X_{i_1}, \ldots, X_{i_n}] = 0$ *for any* $n \geq 1$ *and any* $i_1, \ldots, i_n \in [d]$ *whenever there is least one* p *such that* $i_p \leq k$ *and at least one* q *such that* $i_q > k$.

PROOF. (1) Obvious.

(2) Expand $e^{\langle \mathbf{t}, X\rangle} = \sum_n \langle \mathbf{t}, X\rangle^n / n!$ and $\langle \mathbf{t}, X\rangle^n = \sum_{i_1,\ldots,i_n=1}^{d} t_{i_1} \ldots t_{i_n} X_{i_1} \ldots X_{i_n}$. Taking expectations gives the expansion for $\varphi_X(t)$. To get the expansion for $\Psi_X(\mathbf{t})$, let $\psi(\mathbf{t}) = \sum_{n=1}^{\infty} \sum_{i_1,\ldots,i_n=1}^{d} \frac{t_{i_1} \ldots t_{i_n}}{n!} \kappa_n[X_{i_1}, \ldots, X_{i_n}]$ and consider

$$e^{\psi(\mathbf{t})} \quad = \quad \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{k_1,\ldots,k_n=1}^{d} \kappa_{k_1}$$

(3) $U = (X_1, \ldots, X_m)$ is independent of $V = (X_{m+1}, \ldots, X_n)$ if and only if $\psi_{(U,V)}(t, s) = \psi_U(t) + \psi_V(s)$ for all $t \in \mathbb{R}^m$, $s \in \mathbb{R}^{n-m}$. By part (b), ψ_U (respectively, ψ_V) has an expansion involving $\kappa_k[X_{i_1}, \ldots, X_{i_k}]$ where $i_1, \ldots, i_k \leq m$ (respectively, $i_1, \ldots, i_k > m$). However, $\psi_{(U,V)}$ has coefficients $\kappa_k[X_{i_1}, \ldots, X_{i_k}]$ where i_r range over all of $[n]$. Thus, U and V are independent if and only if $\kappa_k[X_{i_1}, \ldots, X_{i_k}] = 0$ whenever there are p, q such that $i_p \leq m$ and $i_q > m$. This proves the equivalence of the two statements. ■

Part (c) of the lemma is the reason why cumulants are useful in studying independent random variables. We shall illustrate this by a quick proof of the central limit theorem (for a restricted class of random variables). However, first we make a few remarks on cumulants of one random variable which the reader may be familiar with.

Let X be a real-valued random variable such that $\mathbf{E}[e^{tX}] < \infty$ for t in a neighbourhood of zero. Then $\varphi_X(t) = \sum_{n=0}^{\infty} m_n(X) t^n / n!$ and $\psi_X(t) = \sum_{n=1}^{\infty} \kappa_n(X) t^n / n!$ where $m_n = m_n[X, \ldots, X]$ and $\kappa_n[X, \ldots, X]$. The relationship between moments and cumulants becomes

$$m_n(X) = \sum_{\Pi \in \mathcal{P}_n} \prod_{j=1}^{\ell_\Pi} \kappa_{|\Pi_j|}(X), \qquad \kappa_n(X) = \sum_{\Pi \in \mathcal{P}_n} (-1)^{\ell_\Pi - 1} (\ell_\Pi - 1)! \prod_{j=1}^{\ell_\Pi} \kappa_{|\Pi_j|}(X).$$

The cumulant sequence (or the moment sequence) determines the moment generating function and hence the distribution of X. Thus knowing the cumulant sequence is sufficient to answer every question about X (in principle). Of course, a quantity like $\mathbf{P}(1 < X < 2)$ is not easy to express in terms of cumulants, so the "in principle" phrase must be taken seriously. There is an additional issue of loss of generality in considering only random variables with moments. For these reasons usually one does not base probability theory on moments or cumulants exclusively. However, there are features that can be captured well in terms of cumulants. Independence is one of them, as part (c) of Lemma 79 shows.

Summing of independent random variables is also captured nicely in terms of cumulants. Indded, if X and Y are independent random variables, by part (a) of Lemma 79

we can write $\kappa_n(X+Y) = \kappa_n[X+Y,\ldots,X+Y]$ as a sum of 2^n terms. By part (c) of the same lemma, using independence, all but two of these vanish and we get $\kappa_n(X+Y) = \kappa_n(X)+\kappa_n(Y)$. A particular case is when $Y=c$, a constant, in which case $\kappa_n(X+c) = \kappa_n(X)+c\delta_{n,1}$. Observe that in contrast, $m_n(X+c)$ has a relatively more complicated expression in terms of moments of X.

Exercise 80.
(1) If $X \sim N(\mu,\sigma^2)$, then $\kappa_1[X]=\mu$, $\kappa_2[X]=\sigma^2$ and $\kappa_n[X]=0$ for $n\geq 3$.
(2) Conversely, if $\kappa_n[X]=0$ for all $n\geq 3$, then $X \sim N(\kappa_1,\kappa_2)$.
(3) If X,Y are i.i.d random variables and $X+Y \stackrel{d}{=} \sqrt{2}X$, show that $X \sim N(0,\sigma^2)$ for some σ^2.

Example 81. Let $X\sim\exp(1)$. Then $\varphi_X(t) = (1-t)^{-1} = \sum_{n\geq 0} t^n$ for $t<1$. Hence $m_n = n!$. $\psi_X(t) = -\log(1-t) = \sum_{n\geq 1} n^{-1}t^n$ which shows that $\kappa_n = (n-1)!$. If $Y\sim \text{Gamma}(\nu,1)$ then for integer values of ν it is a sum of i.i.d exponentials, hence $\kappa_n(Y) = \nu(n-1)!$. It may be verified directly that this is also true for any $\nu>0$.

Example 82. Let $X\sim\text{Pois}(1)$. Then $\mathbf{E}[e^{tX}] = e^{-1+e^t}$. Expanding this, one can check that $m_n = e^{-1}\sum_{k=0}^{\infty}\frac{k^n}{k!}$. It is even easier to see that $\psi_X(t) = -1+e^t$ and hence $\kappa_n = 1$ for all $n\geq 1$ and hence also $\kappa_\Pi = 1$. But then, the defining equation for cumulants in terms of moments shows that $m_n = \sum_{\Pi\in\mathcal{P}_n}\kappa_\Pi = |\mathcal{P}_n|$. Thus as a corollary, we have the non-trivial relation $|\mathcal{P}_n| = e^{-1}\sum_{k=0}^{\infty}\frac{k^n}{k!}$, known as Dobinsky's formula.

Remark 83. The relationship between m_n and κ_n just comes from the connection that $\log\varphi = \psi$ where $m_n/n!$ are the coefficient of φ and $\kappa_n/n!$ are coefficients of ψ. The same is true for coefficients of any two power series related this way. A closer look at the expressions for m_n in terms of κ_n or the reverse one shows that if m_n counts some combinatorial objects, then κ_n counts the connected pieces of the same combinatorial object.

For example, in Example 81, $m_n = n!$ counts the number of permutations on n letters while $\kappa_n = (n-1)!$ counts the number of cyclic permutations. As any permutation may be written as a product of disjoint cycles, it makes sense to say that cycles are the only connected permutations.

In Example 82, $m_n = |\mathcal{P}_n|$ while $\kappa_n = 1$. Indeed, the only "connected partition" is the one having only one block $\{1,2,\ldots,n\}$.

In case of $N(0,1)$, we know that m_n counts the number of matching of $[n]$. What are connected matchings? If $n>2$, there are no connected matchings! Hence, $\kappa_n = 0$ for $n\geq 3$.

Now we turn to the promised proof of CLT. By part (c) of Exercise 80, if $S_n/\sqrt{n}$ were to converge to a limit, then it is easy to see that the limit random would have to satisfy the recursive distributional equation $U+V \stackrel{d}{=} \sqrt{2}U$ where U,V are i.i.d copies of the limit variable and hence $U\sim N(0,\sigma^2)$. Using cumulants we can actually show that this is the case.

PROOF OF CENTRAL LIMIT THEOREM ASSUMING MGF EXISTS. Suppose $X_1,X_2,\ldots$ are i.i.d with zero mean and unit variance and such that the mgf of X_1 exists in a neighbourhood of zero, then for any fixed $p\geq 1$,

$$\kappa_p[S_n/\sqrt{n}] \;=\; n^{-\frac{p}{2}}\kappa[S_n,\ldots,S_n] \;=\; n^{-\frac{p}{2}}\sum_{\mathbf{i}\in[n]^p}\kappa[X_{i_1},\ldots,X_{i_p}]$$

by multilinearity of cumulants. If $X_{i_r} \neq X_{i_s}$, the corresponding summand will vanish by the independence of X_js. Therefore,

$$\kappa_p[S_n/\sqrt{n}] = n^{-\frac{p}{2}} \sum_{j=1}^{n} \kappa[X_j, X_j, \ldots, X_j] = n^{-\frac{p}{2}+1} \kappa_p[X_1]$$

which goes to zero for $p \geq 3$. As the first two cumulants are 0 and 1 respectively, we see that the cumulants of $S_n/\sqrt{n}$ converge to cumulants of $N(0,1)$ and hence the moments converge also. Thus, $S_n/\sqrt{n}$ converges in distribution to $N(0,1)$. ■

2. Non-commutative probability spaces

We define[1] three notions of non-commutative probability space, of which the first one is sufficient for our purposes. In the next section we shall introduce the notion of independence in such spaces.

Definition 84. A *non-commutative probability space* is a pair $(\mathcal{A}, \varphi)$ where $\mathcal{A}$ is a *unital algebra* over complex numbers and φ is a linear functional on $\mathcal{A}$ such that $\varphi(\mathbf{1}) = \mathbf{1}$.

A unital algebra $\mathcal{A}$ is a vector space over $\mathbb{C}$ endowed with a multiplication operation $(a,b) \to ab$ which is assumed to be associative and also distributive over addition and scalar multiplication. In addition we assume that there is a *unit*, denoted 1, such that $a\mathbf{1} = a = \mathbf{1}a$ for all $a \in \mathcal{A}$.

Example 85. Let $\mathcal{A}$ be the space of all polynomials in one variable with complex coefficients. This is a unital algebra with the obvious operations. Fix a complex Borel measure μ on $\mathbb{R}$ such that $\mu(\mathbb{R}) = 1$. Define $\varphi(P) = \int P(x)\mu(dx)$ for any $P \in \mathcal{A}$. Then, $(\mathcal{A}, \varphi)$ is a (commutative!) ncps. This leads us to a smaller class of ncps. If we considered polynomials in three variables and μ a measure on $\mathbb{R}^3$, we would again get a ncps. The difference is that in one dimension, at least if μ is compactly supported, then $(\mathcal{A}, \varphi)$ has all the information in the classical measure space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mu)$.

In the example above, of particular interest are probability measures. We have assumed that $\mu(\mathbb{R}) = 1$, but positivity is an extra condition which can be framed by saying that $\varphi(P) \geq 0$ if $P(x) \geq 0$ for all $x \in \mathbb{R}$. Observe that there is no clear way to introduce the notion of positivity in a general unital algebra. This leads us to a smaller sub-class of ncps.

Definition 86. Let $\mathcal{A}$ be a C^*-algebra[2] with a unit Let $\varphi : \mathcal{A} \to \mathbb{C}$ be a linear functional such that $\varphi(aa^*) \geq 0$ for all $a \in \mathcal{A}$ (we say that φ is a positive linear functional). Assume also that $\varphi(\mathbf{1}) = 1$. Then, we say that φ is a state. $(\mathcal{A}, \varphi)$ is called a C^*-probability space.

Observe that φ is necessarily bounded. In fact, for any self-adjoint a, $a - \|a\|1$ and $\|a\|1 - a$ are non-negative elements (can be written as b^*b for some b). Hence $|\varphi(a)| \leq \|a\|$ as $\varphi(1) = 1$. If a is any element of the algebra, it can be written in a unique way as $x + iy$ where x, y are self-adjoint and hence $|\varphi(a)| \leq 2$.

---

[1] Much of our presentation of free probability is taken from three sources. The St. Flour lecture notes of Voiculescu **?**, various lecture notes of Roland Speicher available on his home page, and the book of Anderson Guionnet and Zeitouni.

[2] By definition, this means that $\mathcal{A}$ has three structures. (a) That of a complex Banach space, (b) that of an algebra and finally, (c) an *involution* $* : \mathcal{A} \to \mathcal{A}$. These operations respect each other as follows. The algebra operations are continuous and respect the norm in the sense that $\|ab\| \leq \|a\|\|b\|$. The involution is idempotent ($(a^*)^* = a$) and satisfies $(ab)^* = b^*a^*$. In addition it is norm-preserving, and conjugate linear (and hence also continuous). Lastly, we have the identity $\|aa^*\| = \|a\|^2$ for all $a \in \mathcal{A}$. We say that a is Hermitian if $a^* = a$ and that a is positive is $a = bb^*$ for some $b \in \mathcal{A}$.

Example 87. Let $\mathcal{A} := \mathcal{B}(H)$ be the algebra of bounded linear operators on a Hilbert space H. This is a C^*-algebra where the identity I is the unit and taking adjoints is the involution. Let $u \in H$ be a unit vector and define $\varphi(T) = \langle Tu, u\rangle$. Then, φ is a linear functional and $\varphi(I) = 1$. Further, $\varphi(T^*T) = \|Tu\|^2 \geq 0$. Thus, $(\mathcal{A}, \varphi)$ is a C^*-probability space. Here multiplication is truly non-commutative.

If $\psi(T) = \langle Tv, v\rangle$ for a different unit vector v, then for $0 < s < 1$, the pair $(\mathcal{A}, s\varphi + (1-s)\psi)$ is also a C^*-probability space. φ is called a pure state while $s\varphi + (1-s)\psi$ is called a mixed state. Any closed subalgebra of $\mathcal{B}(H)$ that is closed under adjoints is also a C^*-algebra. We only consider those that contain the unit element.

Example 88. Let K be a compact metric space and let $\mathcal{A} = C(K)$ (continuous complex-valued functions). The operations are obvious (involution means taking the conjugate of a function). Let μ be any Borel probability measure on K and define $\varphi(f) = \int_K f d\mu$. Then $(\mathcal{A}, \varphi)$ is a C^*-probability space.

Example 89. The same applies to $C_b(\mathbb{R})$ and $\varphi(f) = \int f d\mu$ for some Borel probability measure μ. It is a commutative C^*-algebra. In fact this is not different from the previous example, as $C_b(\mathbb{R}) = C(K)$ where K is the Stone-Cech compactification of $\mathbb{R}$.

As these examples show, a C^*-probability space generalizes the idea of presenting a probability measure on $\mathbb{R}$ by giving the integrals of all bounded continuous functions which is more than giving the integral of polynomials only. However, for later purposes, it is useful to remark that C^*-probability space is like the algebra of *complex-valued* random variables, not real valued ones. A third level is to specify a probability measure μ by giving the integrals of bounded measurable functions.

Definition 90. Let H be a Hilbert space and let $\mathcal{A} \subseteq \mathcal{B}(H)$ be a W^*-algebra[3] We assume that it contains the identity. Let u b a unit vector in H and define $\varphi(T) = \langle Tu, u\rangle$ for $T \in \mathcal{A}$ (a pure state). Then we say that $(\mathcal{A}, \varphi)$ is a W^*-probability space.

Example 91.
(1) Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and let $\mathcal{A} = L^\infty(\mathbf{P})$. We can think of $\mathcal{A}$ as a subalgebra of $\mathcal{B}(L^2(\mu))$ by the map $M : \mathcal{A} \to \mathcal{B}(L^2(\mu))$ by $f \to M_f$ where $M_f(g) = f \cdot g$. Then we leave it as an exercise to check that $\mathcal{A}$ is closed under weak operator topology.

Let $\mathbf{1}$ be the constant random variable 1. Then $\mathcal{A}$ is a unital algebra. Let $\varphi(X) := \mathbf{E}[X] = \langle M_X \mathbf{1}, \mathbf{1}\rangle$ for $X \in \mathcal{A}$. This satisfies the definition of a n.c.p.s. Of course $(\mathcal{A}, \varphi)$ is commutative and not of the main interest to us here, but this example explains the phrase "probability space" in the n.c.p.s. In this case there is a notion of positivity, and $\varphi(X) \geq 0$ for $X \geq 0$.

(2) The example 87 is a W^*-probability space too. Subalgebras of $\mathcal{B}(H)$ that are closed in weak operator topology are also W^*-probability spaces.

Example 92 (The prime example - 1)**.** Let $\mathcal{M}_n$ be the space of $n \times n$ complex matrices. This is a W^*-algebra (it is $\mathcal{B}(H)$ where $H = \mathbb{C}^n$). If $\mathbf{e}_k$ is the k^{th} standard co-ordinate vector, then $\varphi_k(T) = \langle T\mathbf{e}_k, \mathbf{e}_k\rangle$ defines a pure state on $\mathcal{M}_n$. Average over k to get a new positive linear functional $\hat{\text{tr}}_n(T) := n^{-1}\text{tr}(T)$. In other words, $\hat{\text{tr}}$ is the mean of the ESD of T.

Example 93 (The prime example - 2)**.** Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and let $\mathcal{A} = L^\infty(\mathbf{P}) \otimes \mathcal{M}_n$ be the space of all random matrices $X = (X_{i,j})_{i,j \leq n}$ where $X_{i,j}$ are bounded,

[3]This means that $\mathcal{A}$ is a C^*-subalgebra of $\mathcal{B}(H)$ and in addition is closed under weak operator topology. That is, if $T \in \mathcal{B}(H)$ and T_α is a net in $\mathcal{A}$ such that $\langle T_\alpha u, v\rangle \to \langle Tu, v\rangle$ for all $u, v \in H$, then $T \in \mathcal{A}$.

complex-valued random variables on Ω. Then, define $\varphi_n(X) = \mathbf{E}[\hat{\mathrm{tr}}(X)]$, the mean of the expected ESD. Then $(\mathcal{A}, \varphi)$ is a ncps, in fact a C^* probability space.

Boundedness of entries is too restrictive as it does not even allow GUE matrices. Instead, we may consider the space $\mathcal{A}$ of random matrices $X = (X_{i,j})_{i,j\le n}$ where $X_{i,j} \in \bigcap_{p<\infty} L^p(\Omega, \mathcal{F}, \mathbf{P})$. Define $\varphi_n(X) = \mathbf{E}[\hat{\mathrm{tr}}(X)]$ as before. This is a non-commutative probability space, although not a C^* probability space.

3. Distribution of non-commutative random variables and Free independence

Let $(\mathcal{A}, \varphi)$ be a ncps. Any element $a \in \mathcal{A}$ are referred to as a *non-commutative random variable* and $\varphi(a)$ as its *non-commutative expectation*.

Define the non-commutative moment function as $m_n[a_1, \dots, a_n] = \varphi(a_1 a_2 \dots a_n)$. As in the classical case, $m_n[\cdot]$ is multilinear, but not symmetric because of non-commutativity. If $a_1, \dots, a_k$ are ncrvs on the same ncps, then the collection of all moments $\{m_n[a_{i_1}, \dots, a_{i_n}] : 1 \le i_1, \dots, i_n \le k\}$ is called the *joint distribution* of $a_1, \dots, a_n$. For one variable, this is just the collection of moments $\varphi(a^n)$, $n \ge 1$.

In classical probability, the distribution of a bounded real-valued random variable X can be recovered from its moments $\mathbf{E}[X^n]$, $n \ge 1$. However, for a complex-valued random variable (even if bounded), one needs joint moments of the real and imaginary parts of X, or equivalently, that of X and $\overline{X}$, to recover the distribution of X. This motivates the following definition.

In a C^* or W^* probability space, the joint distribution of a and a^* is called the $*$-distribution of a. Observe that this involves specifying $\varphi(P(a, a^*))$ for any non-commutative polynomial P (with complex coefficients) in two variables. Similarly one defines the $*$-distribution for more than one variable. As we remarked earlier, an element of a C^*-probability space is analogous to a complex valued random variable. For a probability measure on the complex plane, the moments , $\{\int z^n \mu(dz) : n \ge 1\}$ does not determine the measure. For example, any radially symmetric μ has $\int z^n \mu(dz) = 0$ for $n \ge 1$. Instead, one should specify the joint moments of the real and imaginary parts, or equivalently, $\int z^m \bar{z}^n \mu(dz)$. Thus, the $*$-distribution is what corresponds to the distribution of a complex-valued random variable.

In the special, but important case when a is *Hermitian* (to be considered analogous to real-valued random variables), the the $*$-distribution is the same as the distribution of a. Further, the following fact is important.

Proposition 94. *If a is a self-adjoint element of a C^*-probability space, then there exists a unique Borel probability measure μ_a on $\mathbb{R}$ such that $m_n(a) = \int x^n \mu_a(dx)$.*

Assuming the fact, by abuse of terminology we may refer to μ_a as the distribution of a. Thus, for self-adjoint elements of a C^*-probability space, the distribution refers to a p=classical probability measure on $\mathbb{R}$. Observe that this does not hold for non self-adjoint elements, or for joint distribution of several ncrvs.

PROOF OF 94. $m_n[a] = \varphi(a^n)$. Let $P(a) = \sum_{k=0}^n c_k a^k$. By the positivity of φ, we see that

$$0 \le \varphi\left(P(a)P(a)^*\right) = \sum_{k,\ell=0}^{n} c_k \bar{c}_\ell \varphi(a^{k+\ell})$$

which means that the infinite matrix $\left(\varphi(a^{i+j})\right)_{i,j\ge 0}$ is a positive definite matrix. Therefore, there exists at least one probability measure μ with moments $\varphi(a^n)$. However, by the

boundedness of φ (we showed earlier that $\|\varphi\| \leq 2$) and the properties of norm in a C^*-algebra, we see that $\varphi(a^n) \leq 2\|a^n\| \leq 2\|a\|^n$. Thus, the moments of μ satisfy $\int x^n \mu(dx) \leq 2\|a\|^n$. This implies that μ must be compactly supported in $[-\|a\|, \|a\|]$. Since the moments of a compactly supported measure determines the measure, we also see that μ is unique. ■

Remark 95. Alternately, restrict to the example of a C^*-probability space given in 87. Then a is a self-adjoint operator on H and by the spectral theorem, there is a spectral measure of a at the vector u satisfying $\int x^n \mu(dx) = \langle a^n u, u\rangle = \varphi(a^n)$. This is the μ we require. Since we know that the spectral measure is supported on the spectrum, and the spectrum is contained in $B(0, \|a\|)$ and the spectrum of a self-adjoint element is real, it follows that μ is supported on $[-\|a\|, \|a\|]$.

We now illustrate with an example.

Example 96. Let $H = \ell^2(\mathbb{N})$ and $\mathbf{e}_0 := (1,0,0,\ldots)$. Let $\mathcal{A} = \mathcal{B}(H)$ and $\varphi(T) = \langle T\mathbf{e}_0, \mathbf{e}_0\rangle$. Now let $L(x_0, x_1, \ldots) = (x_1, x_2, \ldots)$ define the left-shift operator. Its adjoint is the right shift operator $L^*(x_0, x_1, \ldots) = (0, x_0, x_1, x_2, \ldots)$. It is easy to see that $\varphi(L^n) = \varphi(L^{*n}) = 1$ for $n = 0$ and equal to 0 for $n \geq 1$. Let $S = L + L^*$, a self-adjoint variable. Then $\varphi(S^n) = \langle (L + L^*)^n \mathbf{e}_0, \mathbf{e}_0\rangle$. It is easy to check that the latter is zero for n odd and is equal to the Catalan number $C_k = \frac{1}{k+1}\binom{2k}{k}$ for $n = 2k$. These are the (classical) moments of the semicircle law supported on $[-2, 2]$. Hence the non-commutative distribution of S is $\mu_{s.c}$.

If we define $\psi(T) = \langle T\mathbf{e}_1, \mathbf{e}_1\rangle$ where $\mathbf{e}_1 = (0,1,0,\ldots)$, can you find the distribution of S in the new ncps $(\mathcal{A}, \psi)$?

Example 97. Let $H = \ell^2(\mathbb{Z})$ and let $\mathbf{e}_0$ be the vector $\mathbf{e}_0(k) = \delta_{k,0}$. Then define the left shift operator L and its adjoint L^* (the right shift operator) in the obvious way. Again, $m_n(L) = m_n(L^*) = \delta_{n,0}$. Let $S = L + L^*$. Now, it is easy to check that $m_n(S)$ is $\binom{2k}{k}$ if $n = 2k$ and equal to zero if n is odd. These are the moments of the *arcsine* distribution with density $\frac{1}{\pi\sqrt{4-x^2}}$ on $[-2, 2]$. Hence S has arc-sine distribution on $[-2, 2]$.

4. Free independence and free cumulants

Independence is a central concept in probability theory. What is the analogue in the non-commutative setting? There is more than one possible notion of independence in non-commutative probability spaces, but there is a particular one that relates to random matrix theory.

Definition 98. Let $(\mathcal{A}, phi)$ be a ncps and let $\mathcal{A}_i$ be a collection of unital subalgebras of $\mathcal{A}$. We say that $\mathcal{A}_i$ are *freely independent* if $\varphi(a_1 a_2 \ldots a_n) = 0$ for any $n \geq 1$ and any $a_i \in \mathcal{A}_{k_i}$ where $k_1 \neq k_2 \neq k_3 \ldots \neq k_n$ (consecutive elements come from different subalgebras). Elements $b_1, b_2, \ldots$ are said to be freely independent if the unital subalgebras generated by b_1, by b_2 etc., are freely independent.

Example 99. So far, classical probability spaces were special cases of non-commutative probability spaces. However, classically independent random variables are almost never freely independent. For example, if X, Y are random variables on $(\Omega, \mathcal{F}, \mathbf{P})$, for them to be freely independent we must have $\mathbf{E}[XYXY] = 0$ by this happens if and only if at least one of X and Y is degenerate at zero.

Example 100. We construct two non-trivial variables that are freely independent. Let $H = \mathbb{C}^2$ with orthonormal basis $\mathbf{e}_1, \mathbf{e}_2$. Then for $n \geq 2$ we define $H^{\otimes n}$ as a 2^n-dimensional space whose basis elements we denote by $\mathbf{e}_{i_1} \otimes \mathbf{e}_{i_2} \otimes \ldots \otimes \mathbf{e}_{i_n}$ where $i_1, \ldots, i_n \in \{1, 2\}$. Let

$H^{\otimes 0} = \mathbb{C}$ with orthonormal basis $\mathbf{e}_0 = 1$ (thus $\mathbf{e}_0 = \pm 1$).Then set $\mathcal{H} := \bigoplus_{n\geq 0} H^{\otimes n}$. $\mathcal{H}$. This is called the *full Fock space* corresponding to H and clearly $\{\mathbf{e}_{i_1} \otimes \mathbf{e}_{i_2} \otimes \ldots \otimes \mathbf{e}_{i_n} : n \geq 1,\ i_k = 1,2\} \cup \{\mathbf{e}_0\}$. It is evident how to generalize this definition for any Hilbert space H, not just $\mathbb{C}^2$.

Define the state $\varphi(T) = \langle T\mathbf{e}_0, \mathbf{e}_0\rangle$ for $T \in \mathcal{B}(\mathcal{H})$. This is a C^*-probability space.

We define $L_1, L_2 \in \mathcal{B}(\mathcal{H})$ as follows. Let $L_1(\mathbf{e}_{i_1} \otimes \mathbf{e}_{i_2} \otimes \ldots \otimes \mathbf{e}_{i_n}) = \mathbf{e}_1 \otimes \mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_n}$ and extend linearly to $\mathcal{H}$. Likewise define L_2 using $\mathbf{e}_2$. The adjoints are given by

$$L_1^*(\mathbf{e}_{i_1} \otimes \mathbf{e}_{i_2} \otimes \ldots \otimes \mathbf{e}_{i_n}) = \begin{cases} \mathbf{e}_{i_2} \otimes \ldots \otimes \mathbf{e}_{i_n} & \text{if } i_1 = 1. \\ 0 & \text{otherwise} \end{cases}$$

and likewise for L_2^*. By the same logic as in example 97 it is easy to see that the non-commutative distribution of $T := L_1 + L_1^*$ and $S := L_2 + L_2^*$ are both semicircle distribution on $[-2,2]$. We now claim that they are freely independent. In fact the algebras $\mathcal{A}_1 = \langle L_1, L_1^*\rangle$ and $\mathcal{A}_2 = \langle L_2, L_2^*\rangle$ are freely independent.

We shall only consider the simplest non-trivial example and leave the full proof to the reader. Since $\varphi(T) = \varphi(S) = 0$, we must show that $\varphi(TSTS) = 0$. For this, consider $\langle (L_1 + L_1^*)(L_2 + L_2^*)(L_1 + L_1^*)(L_2 + L_2^*)\mathbf{e}_0, \mathbf{e}_0\rangle$, expand the product and observe that each term vanishes.

I have not written the next few sections fully or properly. Please refer to the books of Anderson, Guionnet and Zeitouni or the various lecture notes of Roland Speicher available on his homepage. If I find time, I shall write this stuff and post it here. For now, just a summary of what we covered in class.

Topics covered next:

(i) Free cumulants defined through free moments by a similar formula to the classical case, but summing only over non-crossing partitions.

(ii) Free independence is equivalent to vanishing of mixed cumulants.

(iii) Free central limit theorem - once the previous section is in place, this follows by copying word by word the proof of classical CLT using cumulants.

(iv) Relationship to random matrix theory - Random matrices $X = (X_{i,j})_{i,j\leq n}$ where $X_{i,j}$ are random variables on $(\Omega, \mathcal{F}, \mathcal{P})$ can be considered also as elements of the non-commutative probability space as described in Example 93.

(v) The crucial connecting fact is that in many cases, large random matrices that are independent in the classical sense, are asymptotically (as the matrix size grows) freely independent. In particular this holds for the following pairs of random matrices.

   (a) Let D be a real diagonal whose ESD converges to a compactly supported measure on $\mathbb{R}$. Let $X^{(i)}$ be (scaled by $1/\sqrt{n}$) independent Wigner matrices with entries that have all moments. Then $D, X^{(1)}, X^{(2)}, \ldots$ are freely independent.