Normal random variables

Teo Banica

Department of Mathematics, University of Cergy-Pontoise, F-95000 Cergy-Pontoise, France. teo.banica@gmail.com

2010 *Mathematics Subject Classification.* 60C05

Key words and phrases. Random variable, Normal variable

Abstract. This is an introduction to the various types of normal random variables, with all needed preliminaries included. We first discuss the probability basics, standard central limits, and the theory of the usual, real normal variables, with mathematics, examples, illustrations, and all needed formulae. Then we go on a similar discussion regarding the complex normal variables, and the Rayleigh variables too, again with formulae and illustrations, and with a look into invariance questions too. We then move to arbitrary dimensions, with a discussion regarding the Gaussian vectors, featuring some functional analysis, and some geometry and physics too. Finally, we provide an introduction to the various quantum versions of the central limits and normal variables, and notably to those coming from free probability and random matrices.

Preface

What is a normal variable? Good question, depending on the type of measurements that you make. Indeed, we have here real normal variables, complex normal variables, general vector normal variables, and even some quantum versions of these.

This is an introduction to the various types of normal random variables, with all needed preliminaries included. The book is organized in 4 parts, as follows:

I - We first discuss the probability basics, central limits, and the theory of the usual, real normal variables, with mathematics, examples, illustrations, and formulae.

II - Then we go on a similar discussion regarding the complex normal variables, and the Rayleigh variables too, and with a look into invariance questions too.

III - We then move to arbitrary dimensions, with a discussion regarding the Gaussian vectors, featuring some functional analysis, and some geometry and physics too.

IV - Finally, we provide an introduction to the various quantum versions of the normal variables, and notably to those coming from free probability and random matrices.

Many thanks to my cats, for precious help with some of the asymptotics.

Cergy, January 2026
Teo Banica

Contents

Part I

Normal variables

CHAPTER 1

Random variables

1a. Random variables

Generally speaking, probability theory is best learned by flipping coins, rolling dice, or playing cards, and with such activities falling under the "discrete probability theory" banner. In order to discuss this, let us start with an abstract definition, as follows:

DEFINITION 1.1. *A discrete probability space is a set X, usually finite or countable, whose elements $x \in X$ are called events, together with a function*

$$P : X \to [0, \infty)$$

called probability function, which is subject to the condition

$$\sum_{x \in X} P(x) = 1$$

telling us that the overall probability for something to happen is 1.

As a first comment, our condition $\sum_{x \in X} P(x) = 1$ perfectly makes sense, and this even if X is uncountable, because the sum of positive numbers is always defined, as a number in $[0, \infty]$, and this no matter how many positive numbers we have.

As a second comment, we have chosen in the above not to assume that X is finite or countable, and this for instance because we want to be able to regard any probability function on $\mathbb{N}$ as a probability function on $\mathbb{R}$, by setting $P(x) = 0$ for $x \notin \mathbb{N}$.

As a third comment, once we have a probability function $P : X \to [0, \infty)$ as above, with $P(x) \in [0, 1]$ telling us what the probability for an event $x \in X$ to happen is, we can compute what the probability for a set of events $Y \subset X$ to happen is, by setting:

$$P(Y) = \sum_{y \in Y} P(y)$$

But more on this, mathematical aspects of discrete probability theory, later, when further building on Definition 1.1. For the moment, what we have above will do.

With this discussed, let us explore now the basic examples, coming from the real life. And here, there are many things to be learned. As a first example, we have:

EXAMPLE 1.2. *Flipping coins.*

Here things are simple and clear, because when you flip a coin the corresponding discrete probability space, together with its probability measure, is as follows:

$$X=\Big\{\text{heads, tails}\Big\}\quad,\quad P(\text{heads})=P(\text{tails})=\frac{1}{2}$$

In the case where the coin is biased, as to land on heads with probability $2/3$, and on tails with probability $1/3$, the corresponding probability space is as follows:

$$X=\Big\{\text{heads, tails}\Big\}\quad,\quad P(\text{heads})=\frac{2}{3}\quad,\quad P(\text{tails})=\frac{1}{3}$$

More generally, given any number $p\in[0,1]$, we have an abstract probability space as follows, where we have replaced heads and tails by win and lose:

$$X=\Big\{\text{win, lose}\Big\}\quad,\quad P(\text{win})=p\quad,\quad P(\text{lose})=1-p$$

Finally, things become more interesting when flipping a coin, biased or not, several times in a row. We will be back to this in a moment, with details.

EXAMPLE 1.3. *Rolling dice.*

Again, things here are simple and clear, because when you throw a die the corresponding probability space, together with its probability measure, is as follows:

$$X=\Big\{1,\ldots,6\Big\}\quad,\quad P(i)=\frac{1}{6}\ ,\ \forall i$$

As before with coins, we can further complicate this by assuming that the die is biased, say landing on face i with probability $p_i\in[0,1]$. In this case the corresponding probability space, together with its probability measure, is as follows:

$$X=\Big\{1,\ldots,6\Big\}\quad,\quad P(i)=p_i\quad,\quad p_i\geq 0\ ,\ \sum_i p_i=1$$

Also as before with coins, things become more interesting when throwing a die several times in a row, or equivalently, when throwing several identical dice at the same time. In this latter case, with n identically biased dice, the probability space is as follows:

$$X=\Big\{1,\ldots,6\Big\}^n\quad,\quad P(i_1\ldots i_n)=p_{i_1}\ldots p_{i_n}\quad,\quad p_i\geq 0\ ,\ \sum_i p_i=1$$

Observe that the sum 1 condition in Definition 1.1 is indeed satisfied, and with this proving that our dice modeling is bug-free, due to the following computation:

$$\begin{aligned}\sum_{i\in X}P(i) &= \sum_{i_1,\ldots,i_n}P(i_1\ldots i_n)\\ &= \sum_{i_1,\ldots,i_n}p_{i_1}\ldots p_{i_n}\\ &= \sum_{i_1}p_{i_1}\ldots\sum_{i_n}p_{i_n}\\ &= 1\times\ldots\times 1\\ &= 1\end{aligned}$$

Getting back now to theory, in the general context of Definition 1.1, we can see that what we have there is very close to the biased die, from Example 1.3. Indeed, in the general context of Definition 1.1, we can say that what happens is that we have a die with $|X|$ faces, which is biased such that it lands on face i with probability $P(i)$.

Which is something quite interesting, allowing us to have some intuition on what is going on, in discrete probability. So, let us record this finding, as follows:

CONCLUSION 1.4. *Discrete probability can be understood as being about throwing a general die, having an arbitrary number of faces, and which is arbitrarily biased too.*

Moving on, with some further probability computations, at a more advanced level, which is playing cards, we have the following result, which is very useful in practice:

THEOREM 1.5. *The probabilities at poker are as follows:*

(1) *One pair:* 0.533.
(2) *Two pairs:* 0.120.
(3) *Three of a kind:* 0.053.
(4) *Full house:* 0.006.
(5) *Straight:* 0.005.
(6) *Four of a kind:* 0.001.
(7) *Flush:* 0.000.
(8) *Straight flush:* 0.000.

PROOF. Let us consider indeed our deck of 32 cards, $7,8,9,10,J,Q,K,A$. The total number of possibilities for a poker hand is:

$$\binom{32}{5}=\frac{32\cdot 31\cdot 30\cdot 29\cdot 28}{2\cdot 3\cdot 4\cdot 5}=32\cdot 31\cdot 29\cdot 7$$

(1) For having a pair, the number of possibilities is:

$$N = \binom{8}{1}\binom{4}{2} \times \binom{7}{3}\binom{4}{1}^3 = 8 \cdot 6 \cdot 35 \cdot 64$$

Thus, the probability of having a pair is:

$$P = \frac{8 \cdot 6 \cdot 35 \cdot 64}{32 \cdot 31 \cdot 29 \cdot 7} = \frac{6 \cdot 5 \cdot 16}{31 \cdot 29} = \frac{480}{899} = 0.533$$

(2) For having two pairs, the number of possibilities is:

$$N = \binom{8}{2}\binom{4}{2}^2 \times \binom{24}{1} = 28 \cdot 36 \cdot 24$$

Thus, the probability of having two pairs is:

$$P = \frac{28 \cdot 36 \cdot 24}{32 \cdot 31 \cdot 29 \cdot 7} = \frac{36 \cdot 3}{31 \cdot 29} = \frac{108}{899} = 0.120$$

(3) For having three of a kind, the number of possibilities is:

$$N = \binom{8}{1}\binom{4}{3} \times \binom{7}{2}\binom{4}{1}^2 = 8 \cdot 4 \cdot 21 \cdot 16$$

Thus, the probability of having three of a kind is:

$$P = \frac{8 \cdot 4 \cdot 21 \cdot 16}{32 \cdot 31 \cdot 29 \cdot 7} = \frac{3 \cdot 16}{31 \cdot 29} = \frac{48}{899} = 0.053$$

(4) For having full house, the number of possibilities is:

$$N = \binom{8}{1}\binom{4}{3} \times \binom{7}{1}\binom{4}{2} = 8 \cdot 4 \cdot 7 \cdot 6$$

Thus, the probability of having full house is:

$$P = \frac{8 \cdot 4 \cdot 7 \cdot 6}{32 \cdot 31 \cdot 29 \cdot 7} = \frac{6}{31 \cdot 29} = \frac{6}{899} = 0.006$$

(5) For having a straight, the number of possibilities is:

$$N = 4\left[\binom{4}{1}^4 - 4\right] = 16 \cdot 63$$

Thus, the probability of having a straight is:

$$P = \frac{16 \cdot 63}{32 \cdot 31 \cdot 29 \cdot 7} = \frac{9}{2 \cdot 31 \cdot 29} = \frac{9}{1798} = 0.005$$

(6) For having four of a kind, the number of possibilities is:

$$N = \binom{8}{1}\binom{4}{4} \times \binom{7}{1}\binom{4}{1} = 8 \cdot 7 \cdot 4$$

Thus, the probability of having four of a kind is:

$$P = \frac{8 \cdot 7 \cdot 4}{32 \cdot 31 \cdot 29 \cdot 7} = \frac{1}{31 \cdot 29} = \frac{1}{899} = 0.001$$

(7) For having a flush, the number of possibilities is:

$$N = 4\left[\binom{8}{4} - 4\right] = 4 \cdot 66$$

Thus, the probability of having a flush is:

$$P = \frac{4 \cdot 66}{32 \cdot 31 \cdot 29 \cdot 7} = \frac{33}{4 \cdot 31 \cdot 29 \cdot 7} = \frac{9}{25172} = 0.000$$

(8) For having a straight flush, the number of possibilities is:

$$N = 4 \cdot 4$$

Thus, the probability of having a straight flush is:

$$P = \frac{4 \cdot 4}{32 \cdot 31 \cdot 29 \cdot 7} = \frac{1}{2 \cdot 31 \cdot 29 \cdot 7} = \frac{1}{12586} = 0.000$$

Thus, we have obtained the numbers in the statement. □

Summarizing, probability is basically about binomials and factorials, and ultimately about numbers. We will see later that, in connection with more advanced questions, of continuous nature, some standard calculus comes into play as well.

Let us discuss now the general theory. The fundamental result in probability is the Central Limit Theorem (CLT), and our next task will be that of explaining this. With the idea in mind of doing things a bit abstractly, our starting point will be:

Definition 1.6. *Let X be a probability space, that is, a space with a probability measure, and with the corresponding integration denoted E, and called expectation.*

1. *The random variables are the real functions $f \in L^\infty(X)$.*
2. *The moments of such a variable are the numbers $M_k(f) = E(f^k)$.*
3. *The law of such a variable is the measure given by $M_k(f) = \int_{\mathbb{R}} x^k d\mu_f(x)$.*

Here the fact that μ_f exists indeed is well-known. By linearity, we would like to have a real probability measure making hold the following formula, for any $P \in \mathbb{R}[X]$:

$$E(P(f)) = \int_{\mathbb{R}} P(x) d\mu_f(x)$$

By using a standard continuity argument, it is enough to have this formula for the characteristic functions χ_I of the measurable sets of real numbers $I \subset \mathbb{R}$:

$$E(\chi_I(f)) = \int_{\mathbb{R}} \chi_I(x) d\mu_f(x)$$

But this latter formula, which reads $P(f\in I)=\mu_f(I)$, can serve as a definition for μ_f, and we are done. Alternatively, assuming some familiarity with measure theory, μ_f is the push-forward of the probability measure on X, via the function $f:X\to\mathbb R$.

Next in line, we need to talk about independence. We can do this as follows:

DEFINITION 1.7. *Two variables $f,g\in L^\infty(X)$ are called independent when*

$$E(f^kg^l)=E(f^k)\,E(g^l)$$

happens, for any $k,l\in\mathbb N$.

Again, this definition hides some non-trivial things. Indeed, by linearity, we would like to have a formula as follows, valid for any polynomials $P,Q\in\mathbb R[X]$:

$$E[P(f)Q(g)]=E[P(f)]\,E[Q(g)]$$

By using a continuity argument, it is enough to have this formula for characteristic functions χ_I,χ_J of the measurable sets of real numbers $I,J\subset\mathbb R$:

$$E[\chi_I(f)\chi_J(g)]=E[\chi_I(f)]\,E[\chi_J(g)]$$

Thus, we are led to the usual definition of independence, namely:

$$P(f\in I,g\in J)=P(f\in I)\,P(g\in J)$$

All this might seem a bit abstract, but in practice, the idea is of course that f,g must be independent, in an intuitive, real-life sense. As a first result now, we have:

PROPOSITION 1.8. *Assuming that $f,g\in L^\infty(X)$ are independent, we have*

$$\mu_{f+g}=\mu_f*\mu_g$$

where $*$ is the convolution of real probability measures.

PROOF. We have the following computation, using the independence of f,g:

$$\begin{aligned}M_k(f+g)&=E((f+g)^k)\\&=\sum_r\binom{k}{r}E(f^rg^{k-r})\\&=\sum_r\binom{k}{r}M_r(f)M_{k-r}(g)\end{aligned}$$

On the other hand, by using the Fubini theorem, we have as well:

$$\begin{aligned}\int_{\mathbb{R}} x^k d(\mu_f * \mu_g)(x) &= \int_{\mathbb{R}\times\mathbb{R}} (x+y)^k d\mu_f(x)d\mu_g(y)\\ &= \sum_r \binom{k}{r} \int_{\mathbb{R}} x^r d\mu_f(x) \int_{\mathbb{R}} y^{k-r} d\mu_g(y)\\ &= \sum_r \binom{k}{r} M_r(f) M_{k-r}(g)\end{aligned}$$

Thus μ_{f+g} and $\mu_f * \mu_g$ have the same moments, so they coincide, as desired. □

Here is now a second result on independence, which is something more advanced:

THEOREM 1.9. *Assuming that $f, g \in L^\infty(X)$ are independent, we have*

$$F_{f+g} = F_f F_g$$

where $F_f(x) = E(e^{ixf})$ is the Fourier transform.

PROOF. We have the following computation, using Proposition 1.8 and Fubini:

$$\begin{aligned}F_{f+g}(x) &= \int_{\mathbb{R}} e^{ixz} d\mu_{f+g}(z)\\ &= \int_{\mathbb{R}} e^{ixz} d(\mu_f * \mu_g)(z)\\ &= \int_{\mathbb{R}\times\mathbb{R}} e^{ix(z+t)} d\mu_f(z) d\mu_g(t)\\ &= \int_{\mathbb{R}} e^{ixz} d\mu_f(z) \int_{\mathbb{R}} e^{ixt} d\mu_g(t)\\ &= F_f(x) F_g(x)\end{aligned}$$

Thus, we are led to the conclusion in the statement. □

As a comment here, you might wonder what that $i \in \mathbb{C}$ number in the definition of the Fourier transform is good for. Good question, which will be answered, in due time.

This was for the foundations of probability theory, quickly explained. For further reading, a classical book is Feller [28]. A nice, more modern book is Durrett [25].

1b. Poisson limits

Let us look more in detail at discrete probability theory. The mathematics here will involve the Poisson laws p_t, which appear via the Poisson Limit Theorem (PLT), that we would like to explain now. Let us start with the following definition:

Definition 1.10. *The Poisson law of parameter* 1 *is the following measure,*

$$p_1=\frac{1}{e}\sum_{k\in\mathbb{N}}\frac{\delta_k}{k!}$$

and the Poisson law of parameter $t>0$ *is the following measure,*

$$p_t=e^{-t}\sum_{k\in\mathbb{N}}\frac{t^k}{k!}\,\delta_k$$

with the letter "p" standing for Poisson.

As a first observation, the above laws have indeed mass 1, as they should, due to the following key formula, which is actually the key formula of all mathematics:

$$e^t=\sum_{k\in\mathbb{N}}\frac{t^k}{k!}$$

We will see in the moment why these measures appear a bit everywhere, in discrete contexts, the reasons for this coming from the Poisson Limit Theorem (PLT). Let us first develop some general theory. We first have the following result:

Theorem 1.11. *The mean and variance of* p_t *are given by:*

$$E=t\quad,\quad V=t$$

In particular for the Poisson law p_1 *we have* $E=1,V=1$*.*

Proof. We have two computations to be performed, as follows:

(1) Regarding the mean, this can be computed as follows:

$$\begin{aligned}E&=e^{-t}\sum_{k\geq0}\frac{t^k}{k!}\cdot k\\&=e^{-t}\sum_{k\geq1}\frac{t^k}{(k-1)!}\\&=e^{-t}\sum_{l\geq0}\frac{t^{l+1}}{l!}\\&=te^{-t}\sum_{l\geq0}\frac{t^l}{l!}\\&=t\end{aligned}$$

(2) For the variance, we first compute the second moment, as follows:

$$\begin{aligned} M_2 &= e^{-t}\sum_{k\geq 0}\frac{t^k}{k!}\cdot k^2 \\ &= e^{-t}\sum_{k\geq 1}\frac{t^k k}{(k-1)!} \\ &= e^{-t}\sum_{l\geq 0}\frac{t^{l+1}(l+1)}{l!} \\ &= te^{-t}\sum_{l\geq 0}\frac{t^l l}{l!}+te^{-t}\sum_{l\geq 0}\frac{t^l}{l!} \\ &= te^{-t}\sum_{l\geq 1}\frac{t^l}{(l-1)!}+t \\ &= t^2e^{-t}\sum_{m\geq 0}\frac{t^m}{m!}+t \\ &= t^2+t \end{aligned}$$

(3) Thus the variance is given by the following formula:

$$\begin{aligned} V &= M_2-E^2 \\ &= (t^2+t)-t^2 \\ &= t \end{aligned}$$

We are therefore led to the conclusions in the statement. □

At the theoretical level now, we have the following result:

THEOREM 1.12. *We have the following formula, for any $s,t>0$,*

$$p_s*p_t=p_{s+t}$$

so the Poisson laws form a convolution semigroup.

PROOF. By using $\delta_k * \delta_l = \delta_{k+l}$ and the binomial formula, we obtain:

$$\begin{aligned} p_s * p_t &= e^{-s}\sum_k \frac{s^k}{k!}\,\delta_k * e^{-t}\sum_l \frac{t^l}{l!}\,\delta_l \\ &= e^{-s-t}\sum_n \delta_n \sum_{k+l=n}\frac{s^kt^l}{k!l!} \\ &= e^{-s-t}\sum_n \frac{\delta_n}{n!}\sum_{k+l=n}\frac{n!}{k!l!}s^kt^l \\ &= e^{-s-t}\sum_n \frac{(s+t)^n}{n!}\,\delta_n \\ &= p_{s+t} \end{aligned}$$

Thus, we are led to the conclusion in the statement. □

Next in line, we have the following result, which is fundamental as well:

THEOREM 1.13. *The Poisson laws appear as formal exponentials*

$$p_t = \sum_k \frac{t^k(\delta_1 - \delta_0)^{*k}}{k!}$$

with respect to the convolution of measures $*$.

PROOF. By using the binomial formula, the measure on the right is:

$$\begin{aligned} \mu &= \sum_k \frac{t^k}{k!}\sum_{r+s=k}(-1)^s\frac{k!}{r!s!}\delta_r \\ &= \sum_k t^k \sum_{r+s=k}(-1)^s\frac{\delta_r}{r!s!} \\ &= \sum_r \frac{t^r\delta_r}{r!}\sum_s \frac{(-1)^s}{s!} \\ &= \frac{1}{e}\sum_r \frac{t^r\delta_r}{r!} \\ &= p_t \end{aligned}$$

Thus, we are led to the conclusion in the statement. □

Regarding now the Fourier transform computation, this is as follows:

THEOREM 1.14. *The Fourier transform of* p_t *is given by*

$$F_{p_t}(y) = \exp\left((e^{iy} - 1)t\right)$$

for any $t > 0$.

PROOF. We have indeed the following computation:

$$\begin{aligned}F_{p_t}(y) &= e^{-t}\sum_k \frac{t^k}{k!}F_{\delta_k}(y)\\ &= e^{-t}\sum_k \frac{t^k}{k!}\,e^{iky}\\ &= e^{-t}\sum_k \frac{(e^{iy}t)^k}{k!}\\ &= \exp(-t)\exp(e^{iy}t)\\ &= \exp\left((e^{iy}-1)t\right)\end{aligned}$$

Thus, we obtain the formula in the statement. □

Observe that the above formula gives an alternative proof for Theorem 1.12, by using the fact that the logarithm of the Fourier transform linearizes the convolution.

As another application of the above Fourier transform formula, which is of key importance, we can now establish the Poisson Limit Theorem, as follows:

THEOREM 1.15 (PLT). *We have the following convergence, in moments,*

$$\left(\left(1-\frac{t}{n}\right)\delta_0+\frac{t}{n}\,\delta_1\right)^{*n}\to p_t$$

for any $t>0$.

PROOF. Let us denote by ν_n the measure under the convolution sign, namely:

$$\nu_n=\left(1-\frac{t}{n}\right)\delta_0+\frac{t}{n}\,\delta_1$$

We have the following computation, for the Fourier transform of the limit:

$$\begin{aligned}F_{\delta_r}(y)=e^{iry} \quad&\implies\quad F_{\nu_n}(y)=\left(1-\frac{t}{n}\right)+\frac{t}{n}\,e^{iy}\\ &\implies\quad F_{\nu_n^{*n}}(y)=\left(\left(1-\frac{t}{n}\right)+\frac{t}{n}\,e^{iy}\right)^n\\ &\implies\quad F_{\nu_n^{*n}}(y)=\left(1+\frac{(e^{iy}-1)t}{n}\right)^n\\ &\implies\quad F(y)=\exp\left((e^{iy}-1)t\right)\end{aligned}$$

Thus, we obtain indeed the Fourier transform of p_t, as desired. □

At the level of moments now, things are quite subtle for Poisson laws. We first have the following result, dealing with the simplest case, where the parameter is $t=1$:

THEOREM 1.16. *The moments of p_1 are the Bell numbers,*

$$M_k(p_1) = |P(k)|$$

where $P(k)$ is the set of partitions of $\{1, \ldots, k\}$.

PROOF. The moments of p_1 are given by the following formula:

$$M_k = \frac{1}{e} \sum_r \frac{r^k}{r!}$$

We therefore have the following recurrence formula for these moments:

$$\begin{aligned}
M_{k+1} &= \frac{1}{e} \sum_r \frac{(r+1)^{k+1}}{(r+1)!} \\
&= \frac{1}{e} \sum_r \frac{r^k}{r!} \left(1 + \frac{1}{r}\right)^k \\
&= \frac{1}{e} \sum_r \frac{r^k}{r!} \sum_s \binom{k}{s} r^{-s} \\
&= \sum_s \binom{k}{s} \cdot \frac{1}{e} \sum_r \frac{r^{k-s}}{r!} \\
&= \sum_s \binom{k}{s} M_{k-s}
\end{aligned}$$

With this done, let us try now to find a recurrence for the Bell numbers:

$$B_k = |P(k)|$$

A partition of $\{1, \ldots, k+1\}$ appears by choosing s neighbors for 1, among the k numbers available, and then partitioning the $k - s$ elements left. Thus, we have:

$$B_{k+1} = \sum_s \binom{k}{s} B_{k-s}$$

Thus, our moments M_k satisfy the same recurrence as the numbers B_k. Regarding now the initial values, in what concerns the first moment of p_1, we have:

$$M_1 = \frac{1}{e} \sum_r \frac{r}{r!} = 1$$

Also, by using the above recurrence for the numbers M_k, we obtain from this:

$$M_2 = \sum_s \binom{1}{s} M_{k-s} = 1 + 1 = 2$$

On the other hand, $B_1 = 1$ and $B_2 = 2$. Thus we obtain $M_k = B_k$, as claimed. □

More generally now, we have the following result, dealing with the case $t > 0$:

THEOREM 1.17. *The moments of p_t with $t > 0$ are given by*

$$M_k(p_t) = \sum_{\pi \in P(k)} t^{|\pi|}$$

where $|.|$ is the number of blocks.

PROOF. The moments of the Poisson law p_t with $t > 0$ are given by:

$$M_k = e^{-t} \sum_r \frac{t^r r^k}{r!}$$

We have the following recurrence formula for these moments:

$$\begin{aligned}
M_{k+1} &= e^{-t} \sum_r \frac{t^{r+1}(r+1)^{k+1}}{(r+1)!} \\
&= e^{-t} \sum_r \frac{t^{r+1} r^k}{r!} \left(1 + \frac{1}{r}\right)^k \\
&= e^{-t} \sum_r \frac{t^{r+1} r^k}{r!} \sum_s \binom{k}{s} r^{-s} \\
&= \sum_s \binom{k}{s} \cdot e^{-t} \sum_r \frac{t^{r+1} r^{k-s}}{r!} \\
&= t \sum_s \binom{k}{s} M_{k-s}
\end{aligned}$$

Regarding now the initial values, the first moment of p_t is given by:

$$M_1 = e^{-t} \sum_r \frac{t^r r}{r!} = e^{-t} \sum_r \frac{t^r}{(r-1)!} = t$$

Now by using the above recurrence we obtain from this:

$$M_2 = t \sum_s \binom{1}{s} M_{k-s} = t(1+t) = t + t^2$$

On the other hand, consider the numbers in the statement, namely:

$$S_k = \sum_{\pi \in P(k)} t^{|\pi|}$$

Since a partition of $\{1,\ldots,k+1\}$ appears by choosing s neighbors for 1, among the k numbers available, and then partitioning the $k-s$ elements left, we have:

$$S_{k+1}=t\sum_s\binom{k}{s}S_{k-s}$$

As for the initial values of these numbers, these are $S_1=t$, $S_2=t+t^2$. Thus the initial values coincide, and so these numbers are the moments of p_t, as stated. □

Summarizing, we have so far a quite good understanding of discrete probability theory. Of course, this is just the beginning of things, and we will be back to this, later.

1c. Central limits

Moving on, and in relation with what we want to do in this book, normal variables, you have certainly heard about bell-shaped curves, and perhaps even observed them in physics or chemistry class, because any routine measurement leads to such curves.

Mathematically, here is the question that we would like to solve:

QUESTION 1.18. *Given random variables $f_1,f_2,f_3,\ldots$, say taken discrete, which are i.i.d., centered, and with common variance $t>0$, do we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\sim g_t$$

in the $n\to\infty$ limit, for some bell-shaped density g_t? And, what is the formula of g_t?

Observe that this question perfectly makes sense, with the probability theory that we know, say by assuming that our random variables $f_1,f_2,f_3,\ldots$ are discrete, as said above. But of course, we would like to solve this question in general too.

As for the $1/\sqrt{n}$ factor, there is certainly need for a normalization factor there, as for things to have a chance to converge, and the good factor is $1/\sqrt{n}$, as shown by:

PROPOSITION 1.19. *In order for a sum of the following type to have a chance to converge, with $f_1,f_2,f_3,\ldots$ being i.i.d., centered, and with common variance $t>0$,*

$$S=\sum_{i=1}^n f_i$$

we must normalize this sum by a $1/\sqrt{n}$ factor, as in Question 1.18.

PROOF. The idea here is to look at the moments of S. Since all variables f_i are centered, $E(f_i)=0$, so is their sum, $E(S)=0$, and no contradiction here. However, when looking at the variance of S, which equals the second moment, due to $E(S) = 0$, things become interesting, due to the following computation:

$$\begin{aligned} V(S) &= E(S^2) \\ &= E\left(\sum_{ij} f_i f_j\right) \\ &= \sum_{ij} E(f_i f_j) \\ &= \sum_i E(f_i^2) + \sum_{i \neq j} E(f_i)E(f_j) \\ &= \sum_i E(f_i^2) \\ &= nt \end{aligned}$$

Thus, we are in need a normalization factor α, in order for our sum to have a chance to converge. But, repeating the computation with S replaced by αS gives:

$$V(\alpha S) = \alpha^2 nt$$

Thus, the good normalization factor is $\alpha = 1/\sqrt{n}$, as claimed. □

So far, so good, we have a nice problem above, and time now to make a plan, in order to solve it. With the tools that we have, from this book so far, here is such a plan:

PLAN 1.20. *In order to solve our central limiting question, we have to:*

(1) *Apply Fourier and let $n \to \infty$, as to compute the Fourier transform of g_t.*
(2) *Do some combinatorics and calculus, as to compute the moments of g_t.*
(3) *Recover g_t out of its moments, again via combinatorics and calculus.*

Getting to work now, let us start with (1). Things are quickly done here, by using the standard linearization results for convolution, which lead to:

THEOREM 1.21. *Given discrete variables $f_1, f_2, f_3, \ldots$, which are i.i.d., centered, and with common variance $t > 0$, we have*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_i \sim g_t$$

with $n \to \infty$, with g_t being the law having $F(x) = e^{-tx^2/2}$ as Fourier transform.

PROOF. There are several things going on here, the idea being as follows:

(1) Observe first that in terms of moments, the Fourier transform of an arbitrary random variable $f : X \to \mathbb{R}$ is given by the following formula:

$$\begin{aligned}
F_f(x) &= E(e^{ixf}) \\
&= E\left(\sum_{k=0}^{\infty} \frac{(ixf)^k}{k!}\right) \\
&= \sum_{k=0}^{\infty} \frac{(ix)^k E(f^k)}{k!} \\
&= \sum_{k=0}^{\infty} \frac{i^k M_k(f)}{k!}\, x^k
\end{aligned}$$

(2) In particular, in the case of a centered variable, $E(f) = 0$, as those that we are interested in, the Fourier transform formula that we get is as follows:

$$F_f(x) = 1 - \frac{M_2(f)}{2} \cdot x^2 - i\,\frac{M_3(f)}{6} \cdot x^3 + \ldots$$

Moreover, by further assuming that the Fourier variable is small, $x \simeq 0$, the Fourier transform formula that we get, that we will use in what follows, becomes:

$$F_f(x) = 1 - \frac{M_2(f)}{2} \cdot x^2 + O(x^2)$$

(3) In addition to this, we will also need to know what happens to the Fourier transform when rescaling. But the formula here is very easy to find, as follows:

$$\begin{aligned}
F_{\alpha f}(x) &= E(e^{ix\alpha f}) \\
&= E(e^{i\alpha x f}) \\
&= F_f(\alpha x)
\end{aligned}$$

(4) Good news, we can now do our computation. By using the above formulae in (2) and (3), the Fourier transform of the variable in the statement is given by:

$$\begin{aligned}
F(x) &= \left[F_f\left(\frac{x}{\sqrt{n}}\right)\right]^n \\
&= \left[1 - \frac{M_2(f)}{2} \cdot \frac{x^2}{n} + O(n^{-2})\right]^n \\
&= \left[1 - \frac{tx^2}{2n} + O(n^{-2})\right]^n \\
&\simeq \left[1 - \frac{tx^2}{2n}\right]^n \\
&\simeq e^{-tx^2/2}
\end{aligned}$$

(5) Thus, we are led to the conclusion in the statement, modulo the fact that we do not know yet that a density g_t having as Fourier transform $F(x) = e^{-tx^2/2}$ really exists.

(6) Summarizing, we can declare our theorem proved, modulo finding that law g_t, which still remains to be done. But no worries here, we will do this, very soon. ☐

Getting now to step (2) of our Plan 1.20, that is easy to work out too, via some elementary one-variable calculus, with the result here being as follows:

THEOREM 1.22. *The "normal" law g_t, having as Fourier transform*

$$F(x) = e^{-tx^2/2}$$

must have all odd moments zero, and its even moments must be the numbers

$$M_k(g_t) = t^{k/2} \times k!!$$

where $k!! = (k-1)(k-3)(k-5)\ldots$, for $k \in 2\mathbb{N}$.

PROOF. Again, several things going on here, the idea being as follows:

(1) To start with, at the level of formalism and notations, in view of Question 1.18 and of Theorem 1.21, we have adopted the term "normal" for the mysterious law g_t that we are looking for, the one having $F(x) = e^{-tx^2/2}$ as Fourier transform.

(2) Getting towards the computation of the moments, as a first useful observation, according to Theorem 1.21 this normal law g_t must be centered, as shown by:

$$\begin{aligned} f_i = \text{centered} \quad &\implies \quad \sum_{i=1}^n f_i = \text{centered} \\ &\implies \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n f_i = \text{centered} \\ &\implies \quad g_t = \text{centered} \end{aligned}$$

Moreover, the same argument works by replacing "centered" with "having an even function as density", and this shows, via some standard calculus, that we will leave here as an exercise, that the odd moments of our normal law must vanish:

$$M_{2l+1}(g_t) = 0$$

Thus, first assertion proved, and we only have to care about the even moments.

(3) As a comment here, as we will see in a moment, our study below of the moments computes in fact the odd moments too, as being all equal to 0, this time without making reference to Theorem 1.21. Thus, definitely no worries with the odd moments.

(4) Getting to work now, we must reformulate the equation $F(x)=e^{-tx^2/2}$, in terms of moments. We know from the proof of Theorem 1.21 that we have:

$$F(x)=\sum_{k=0}^\infty\frac{i^kM_k(g_t)}{k!}\,x^k$$

On the other hand, we have the following formula, for the exponential:

$$e^{-tx^2/2}=\sum_{r=0}^\infty(-1)^r\frac{t^rx^{2r}}{2^rr!}$$

Thus, our equation $F(x)=e^{-tx^2/2}$ takes the following form:

$$\sum_{k=0}^\infty\frac{i^kM_k(g_t)}{k!}\,x^k=\sum_{r=0}^\infty(-1)^r\frac{t^rx^{2r}}{2^rr!}$$

(5) As a first observation, the odd moments must vanish, as said in (2) above. As for the even moments, these can be computed as follows:

$$\begin{aligned}
M_k(g_t) &= k!\times\frac{t^{k/2}}{2^{k/2}(k/2)!}\\
&= t^{k/2}\times\frac{k!}{2^{k/2}(k/2)!}\\
&= t^{k/2}\times\frac{2\cdot3\cdot4\ldots(k-1)\cdot k}{2\cdot4\cdot6\ldots(k-2)\cdot k}\\
&= t^{k/2}\times 3\cdot5\ldots(k-3)(k-1)\\
&= t^{k/2}\times k!!
\end{aligned}$$

Thus, we are led to the formula in the statement. □

The moment formula that we found is quite interesting, and before going ahead with step (3) of our Plan 1.20, let us look a bit at this, and see what we can further say.

To be more precise, in analogy with what we know from before about the Poisson laws, making reference to interesting combinatorics and partitions, when it comes to computing moments, we have the following result, regarding the normal laws:

THEOREM 1.23. *The moments of the normal law g_t are given by*

$$M_k(g_t)=t^{k/2}|P_2(k)|$$

for any $k\in\mathbb N$, with $P_2(k)$ standing for the pairings of $\{1,\ldots,k\}$.

Proof. This is a reformulation of Theorem 1.22, the idea being as follows:

(1) We know from Theorem 1.22 that the moments of the normal law $M_k = M_k(g_t)$ that we are interested in are given by the following formula, with the convention $k!! = 0$ for k odd, and $k!! = (k-1)(k-3)(k-5)\ldots$ for k even, for the double factorials:

$$M_k(g_t) = t^{k/2} \times k!!$$

Now observe that, according to our above convention for the double factorials, these are subject to the following recurrence relation, with initial data $1!! = 0, 2!! = 1$:

$$k!! = (k-1)(k-2)!!$$

We conclude that the moments of the normal law $M_k = M_k(g_t)$ are subject to the following recurrence relation, with initial data $M_1 = 0, M_2 = t$:

$$M_k = t(k-1)M_{k-2}$$

(2) On the other hand, let us first count the pairings of the set $\{1,\ldots,k\}$. In order to have such a pairing, we must pair 1 with one of the numbers $2,\ldots,k$, and then use a pairing of the remaining $k-2$ numbers. Thus, we have the following recurrence formula for the number P_k of such pairings, with the initial data $P_1 = 0, P_2 = 1$:

$$P_k = (k-1)P_{k-2}$$

Now by multiplying by $t^{k/2}$, the resulting numbers $N_k = t^{k/2}P_k$ will be subject to the following recurrence relation, with initial data $N_1 = 0, N_2 = t$:

$$N_k = t(k-1)N_{k-2}$$

(3) Thus, the moments $M_k = M_k(g_t)$ and the numbers $N_k = t^{k/2}P_k$ are subject to the same recurrence relation, with the same initial data, so they are equal, as claimed. □

Still in analogy with what we know from before about the Poisson laws, we can further process what we found in Theorem 1.23, and we are led in this way to:

Theorem 1.24. *The moments of the normal law g_t are given by*

$$M_k(g_t) = \sum_{\pi \in P_2(k)} t^{|\pi|}$$

where $P_2(k)$ is the set of pairings of $\{1,\ldots,k\}$, and $|.|$ is the number of blocks.

Proof. This is a quick reformulation of Theorem 1.23, with the number of blocks of a pairing of $\{1,\ldots,k\}$ being trivially $k/2$, independently of the pairing. □

As a philosophical conclusion now to all this, let us formulate:

CONCLUSION 1.25. *The normal laws g_t have properties which are quite similar to those of the Poisson laws p_t, and combinatorially, the passage*

$$p_t \to g_t$$

appears by replacing the partitions with the pairings.

Which sounds quite conceptual, and promising, hope you agree with me. In the meantime, however, we still need to know what the density of g_t is.

1d. Density search

So, let us get now to step (3) of our Plan 1.20. This does not look obvious at all, but some partial integration know-how leads us to the following statement:

THEOREM 1.26. *The normal laws are given by*

$$g_t = \frac{1}{\sqrt{2t} \cdot I} e^{-x^2/2t} dx$$

with the constant on the bottom being $I = \int_{\mathbb{R}} e^{-x^2} dx$.

PROOF. This comes from partial integration, as follows:

(1) Let us first do a naive computation. Consider the following quantities:

$$M_k = \int_{\mathbb{R}} x^k e^{-x^2} dx$$

It is quite obvious that by partial integration we will get a recurrence formula for these numbers, similar to the one that we have for the moments of the normal laws. So, let us do this. By partial integration we obtain the following formula, for any $k \in \mathbb{N}$:

$$\begin{aligned} M_k &= -\frac{1}{2} \int_{\mathbb{R}} x^{k-1} \left(e^{-x^2}\right)' dx \\ &= \frac{1}{2} \int_{\mathbb{R}} (k-1) x^{k-2} e^{-x^2} dx \\ &= \frac{k-1}{2} \cdot M_{k-2} \end{aligned}$$

(2) Thus, we are on the good way, with the recurrence formula that we got being the same as that for the moments of $g_{1/2}$. Now let us fine-tune this, as to reach to the same recurrence as for the moments of g_t. Consider the following quantities:

$$N_k = \int_{\mathbb{R}} x^k e^{-x^2/2t} dx$$

By partial integration as before, we obtain the following formula:

$$\begin{aligned} N_k &= \int_{\mathbb{R}} (tx^{k-1}) \left(-e^{-x^2/2t}\right)' dx \\ &= \int_{\mathbb{R}} t(k-1)x^{k-2}e^{-x^2/2t}dx \\ &= t(k-1)\int_{\mathbb{R}} x^{k-2}e^{-x^2/2t}dx \\ &= t(k-1)N_{k-2} \end{aligned}$$

(3) Thus, almost done, and it remains to discuss normalization. We know from the above that we must have a formula as follows, with I_t being a certain constant:

$$g_t = \frac{1}{I_t} \cdot e^{-x^2/2t}\, dx$$

But the constant I_t must be the one making g_t of mass 1, and so:

$$\begin{aligned} I_t &= \int_{\mathbb{R}} e^{-x^2/2t}\, dx \\ &= \int_{\mathbb{R}} e^{-2ty^2/2t}\sqrt{2t}dy \\ &= \sqrt{2t}\int_{\mathbb{R}} e^{-y^2}dy \end{aligned}$$

Thus, we are led to the formula in the statement. □

What we did in the above is good work, and it remains to compute the constant I appearing in Theorem 1.26. So, almost done, modulo solving the following question:

$$\int_{\mathbb{R}} e^{-x^2}dx = ?$$

However, and here comes the bad news, this integral seems impossible to compute, with the usual tools of calculus. So, let us formulate the following question:

QUESTION 1.27. *What is the value of the following integral,*

$$I = \int_{\mathbb{R}} e^{-x^2}dx$$

that we need as input, for our normal variable theory?

And good question this is. We will see in the next chapter that this question can be solved indeed, but only with some advanced integration know-how.

1e. Exercises

This was a standard introduction to probability, mostly focusing on enumerating, counting, and other discrete aspects, and as exercises on this, we have:

EXERCISE 1.28. *What happens when flipping a coin several times in a row?*

EXERCISE 1.29. *What happens when rolling a die several times in a row?*

EXERCISE 1.30. *Learn about the binomial laws, and their various properties.*

EXERCISE 1.31. *Redo all the poker computations, using ordered hands.*

EXERCISE 1.32. *Clarify if needed all the basics, including learning Fubini.*

EXERCISE 1.33. *What can be $i \in \mathbb{C}$ in the definition of Fourier good for?*

EXERCISE 1.34. *Learn more about the Poisson laws, and their properties.*

EXERCISE 1.35. *Spend a few days in trying to compute $\int_{\mathbb{R}} e^{-x^2}dx$.*

As bonus exercise, read some measure theory, which is needed for a good understanding of probability theory, and with this being an excellent investment.

CHAPTER 2

Normal variables

2a. Calculus, Gauss

We recall from the previous chapter that we have unfinished business with the central limits, with the remaining problem being as follows:

$$\int_{\mathbb{R}} e^{-x^2} dx = ?$$

To be more precise, this integral is impossible to compute, with one-variable techniques. However, we can solve it by using two dimensions, as follows:

THEOREM 2.1. *We have the following formula,*

$$\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}$$

called Gauss integral formula.

PROOF. As already mentioned, this is something which is nearly impossible to prove, with bare hands. However, this can be proved by using two dimensions, as follows:

$$\begin{aligned}
\int_{\mathbb{R}}\int_{\mathbb{R}} e^{-x^2-y^2} dxdy &= 4\int_0^\infty\int_0^\infty e^{-x^2-y^2} dxdy \\
&= 4\int_0^\infty\int_0^\infty e^{-t^2y^2-y^2} ydtdy \\
&= 4\int_0^\infty\int_0^\infty ye^{-y^2(1+t^2)} dydt \\
&= 2\int_0^\infty\int_0^\infty \left(-\frac{e^{-y^2(1+t^2)}}{1+t^2}\right)' dydt \\
&= 2\int_0^\infty \frac{dt}{1+t^2} \\
&= 2\int_0^\infty (\arctan t)' dt \\
&= \pi
\end{aligned}$$

Thus, we are led to the conclusion in the statement. □

Before going further, we would like to present as well a second proof for the Gauss formula, which is quite standard too, using polar coordinates. Let us start with:

PROPOSITION 2.2. *We have the change of variable formula*

$$\int_a^b f(x)dx = \int_c^d f(\varphi(t))\varphi'(t)dt$$

where $c = \varphi^{-1}(a)$ *and* $d = \varphi^{-1}(b)$.

PROOF. This follows with $f = F'$, from the following differentiation rule:

$$(F\varphi)'(t) = F'(\varphi(t))\varphi'(t)$$

Indeed, by integrating between c and d, we obtain the result. □

In several variables now, things are quite similar, the result being as follows:

THEOREM 2.3. *Given a transformation* $\varphi = (\varphi_1, \ldots, \varphi_N)$, *we have*

$$\int_E f(x)dx = \int_{\varphi^{-1}(E)} f(\varphi(t))|J_\varphi(t)|dt$$

with the J_φ *quantity, called Jacobian, being given by*

$$J_\varphi(t) = \det\left[\left(\frac{d\varphi_i}{dx_j}(x)\right)_{ij}\right]$$

and with this generalizing the usual formula from one variable calculus.

PROOF. This is something quite tricky, the idea being as follows:

(1) Observe first that this generalizes indeed the change of variable formula in 1 dimension, from Proposition 2.2, the point here being that the absolute value on the derivative appears as to compensate for the lack of explicit bounds for the integral.

(2) As a second observation, we can assume if we want, by linearity, that we are dealing with the constant function $f = 1$. For this function, our formula reads:

$$vol(E) = \int_{\varphi^{-1}(E)} |J_\varphi(t)|dt$$

In terms of $D = \varphi^{-1}(E)$, this amounts in proving that we have:

$$vol(\varphi(D)) = \int_D |J_\varphi(t)|dt$$

Now since this latter formula is additive with respect to D, it is enough to prove it for small cubes D. And here, as a first remark, our formula is clear for the linear maps φ, by using the definition of the determinant of real matrices, as a signed volume.

(3) However, the extension of this to the case of non-linear maps φ is something which looks non-trivial, so we will not follow this path, in what follows. So, while the above $f = 1$ discussion is certainly something nice, our theorem is still in need of a proof.

(4) In order to prove the theorem, as stated, let us rather focus on the transformations used φ, instead of the functions to be integrated f. Our first claim is that the validity of the theorem is stable under taking compositions of such transformations φ.

(5) In order to prove this claim, consider a composition, as follows:

$$\varphi : E \to F \quad , \quad \psi : D \to E \quad , \quad \varphi \circ \psi : D \to F$$

Assuming that the theorem holds for φ, ψ, we have the following computation:

$$\begin{aligned}
\int_F f(x)dx &= \int_E f(\varphi(s))|J_\varphi(s)|ds \\
&= \int_D f(\varphi \circ \psi(t))|J_\varphi(\psi(t))| \cdot |J_\psi(t)|dt \\
&= \int_D f(\varphi \circ \psi(t))|J_{\varphi\circ\psi}(t)|dt
\end{aligned}$$

Thus, our theorem holds as well for $\varphi \circ \psi$, and we have proved our claim.

(6) Next, as a key ingredient, let us examine the case where we are in $N = 2$ dimensions, and our transformation φ has one of the following special forms:

$$\varphi(x, y) = (\psi(x, y), y) \quad , \quad \varphi(x, y) = (x, \psi(x, y))$$

By symmetry, it is enough to deal with the first case. Here the Jacobian is $d\psi/dx$, and by replacing if needed $\psi \to -\psi$, we can assume that this Jacobian is positive, $d\psi/dx > 0$. Now by assuming as before that $D = \varphi^{-1}(E)$ is a rectangle, $D = [a, b] \times [c, d]$, we can prove our formula by using the change of variables in 1 dimension, as follows:

$$\begin{aligned}
\int_E f(s)ds &= \int_{\varphi(D)} f(x, y)dxdy \\
&= \int_c^d \int_{\psi(a,y)}^{\psi(b,y)} f(x, y)dxdy \\
&= \int_c^d \int_a^b f(\psi(x, y), y)\frac{d\psi}{dx}\,dxdy \\
&= \int_D f(\varphi(t))J_\varphi(t)dt
\end{aligned}$$

(7) But with this, we can now prove the theorem, in $N = 2$ dimensions. Indeed, given a transformation $\varphi = (\varphi_1, \varphi_2)$, consider the following two transformations:

$$\phi(x, y) = (\varphi_1(x, y), y) \quad , \quad \psi(x, y) = (x, \varphi_2 \circ \phi^{-1}(x, y))$$

We have then $\varphi = \psi \circ \phi$, and by using (6) for ψ, ϕ, which are of the special form there, and then (3) for composing, we conclude that the theorem holds for φ, as desired.

(8) Thus, theorem proved in $N = 2$ dimensions, and the extension of the above proof to arbitrary N dimensions is straightforward, that we will leave this as an exercise. □

In order to discuss now the applications, in 2 dimensions, let us start with:

PROPOSITION 2.4. *We have polar coordinates in 2 dimensions,*

$$\begin{cases} x & = & r\cos t \\ y & = & r\sin t \end{cases}$$

the corresponding Jacobian being $J = r$.

PROOF. This is something elementary, with the Jacobian being as follows:

$$\begin{aligned} J &= \begin{vmatrix} \frac{d(r\cos t)}{dr} & \frac{d(r\cos t)}{dt} \\ \\ \frac{d(r\sin t)}{dr} & \frac{d(r\sin t)}{dt} \end{vmatrix} \\ &= \begin{vmatrix} \cos t & -r\sin t \\ \sin t & r\cos t \end{vmatrix} \\ &= r\cos^2 t + r\sin^2 t \\ &= r \end{aligned}$$

Thus, we have indeed the formula in the statement. □

We can now compute the Gauss integral, which is the best calculus formula ever:

THEOREM 2.5. *We have the following formula,*

$$\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}$$

called Gauss integral formula.

Proof. Let I be the above integral. By using polar coordinates, we obtain:

$$\begin{aligned}
I^2 &= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{-x^2-y^2}dxdy \\
&= \int_0^{2\pi}\int_0^\infty e^{-r^2}rdrdt \\
&= 2\pi\int_0^\infty\left(-\frac{e^{-r^2}}{2}\right)'dr \\
&= 2\pi\left[0-\left(-\frac{1}{2}\right)\right] \\
&= \pi
\end{aligned}$$

Thus, we are led to the formula in the statement. □

Finally, let us record as well the following result, that we will need at some point:

Theorem 2.6. *We have spherical coordinates in 3 dimensions,*

$$\begin{cases}
x = r\cos s \\
y = r\sin s\cos t \\
z = r\sin s\sin t
\end{cases}$$

the corresponding Jacobian being $J(r,s,t) = r^2\sin s$.

Proof. The fact that we have indeed spherical coordinates is clear. Regarding now the Jacobian, this is given by the following formula:

$$\begin{aligned}
&J(r,s,t) \\
&= \begin{vmatrix}\cos s & -r\sin s & 0\\ \sin s\cos t & r\cos s\cos t & -r\sin s\sin t\\ \sin s\sin t & r\cos s\sin t & r\sin s\cos t\end{vmatrix} \\
&= r^2\sin s\sin t\begin{vmatrix}\cos s & -r\sin s\\ \sin s\sin t & r\cos s\sin t\end{vmatrix} + r\sin s\cos t\begin{vmatrix}\cos s & -r\sin s\\ \sin s\cos t & r\cos s\cos t\end{vmatrix} \\
&= r\sin s\sin^2 t\begin{vmatrix}\cos s & -r\sin s\\ \sin s & r\cos s\end{vmatrix} + r\sin s\cos^2 t\begin{vmatrix}\cos s & -r\sin s\\ \sin s & r\cos s\end{vmatrix} \\
&= r\sin s(\sin^2 t+\cos^2 t)\begin{vmatrix}\cos s & -r\sin s\\ \sin s & r\cos s\end{vmatrix} \\
&= r\sin s\times 1\times r \\
&= r^2\sin s
\end{aligned}$$

Thus, we have indeed the formula in the statement. □

Very nice all this, and getting back now to probability theory, and to our central limiting question, raised some time ago, back in chapter 1, we can now fully answer this question, and formulate the Central Limit Theorem (CLT), as follows:

THEOREM 2.7 (CLT). *Given discrete random variables $f_1, f_2, f_3, \ldots$, which are i.i.d., centered, and with common variance $t > 0$, we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i \sim g_t$$

in the $n \to \infty$ limit, in moments, with the limiting mesure being

$$g_t = \frac{1}{\sqrt{2\pi t}}e^{-x^2/2t}dx$$

called normal, or Gaussian law of parameter $t > 0$.

PROOF. This follows indeed from our various results above, and more specifically from the results in chapter 1, complemented by Theorem 2.1, or Theorem 2.5. □

2b. Normal variables

Let us study now more in detail the laws that we found. Normally we already have everything that is needed, but it is instructive at this point to do some computations, based on the explicit formula of g_t found in Theorem 2.7. We first have:

PROPOSITION 2.8. *We have the variance formula*

$$V(g_t) = t$$

valid for any $t > 0$.

PROOF. We already know this, but we can establish this as well directly, starting from our formula of g_t from Theorem 2.7. Indeed, the first moment is 0, because our normal law g_t is centered. As for the second moment, this can be computed as follows:

$$\begin{aligned} M_2 &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} x^2 e^{-x^2/2t}dx \\ &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} (tx)\left(-e^{-x^2/2t}\right)' dx \\ &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} te^{-x^2/2t}dx \\ &= t \end{aligned}$$

We conclude from this that the variance is $V = M_2 = t$, as claimed. □

More generally, we can recover in this way the computation of all moments:

THEOREM 2.9. *The even moments of the normal law are the numbers*

$$M_k(g_t) = t^{k/2} \times k!!$$

where $k!! = (k-1)(k-3)(k-5)\ldots$, *and the odd moments vanish.*

PROOF. Again, we already know this, but we can establish this as well directly, starting from our formula above of g_t. Indeed, we have the following computation:

$$\begin{aligned} M_k &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} y^k e^{-y^2/2t}dy \\ &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} (ty^{k-1})\left(-e^{-y^2/2t}\right)' dy \\ &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} t(k-1)y^{k-2}e^{-y^2/2t}dy \\ &= t(k-1)\times\frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} y^{k-2}e^{-y^2/2t}dy \\ &= t(k-1)M_{k-2} \end{aligned}$$

Thus by recurrence, we are led to the formula in the statement. □

Here is another result, which is the key one for the study of the normal laws:

THEOREM 2.10. *We have the following formula, valid for any* $t > 0$*:*

$$F_{g_t}(x) = e^{-tx^2/2}$$

In particular, the normal laws satisfy $g_s * g_t = g_{s+t}$*, for any* $s, t > 0$.

PROOF. As before, we already know this, but we can establish now the Fourier transform formula as well directly, by using the explicit formula of g_t, as follows:

$$\begin{aligned} F_{g_t}(x) &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} e^{-y^2/2t+ixy}dy \\ &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} e^{-(y/\sqrt{2t}-\sqrt{t/2}ix)^2-tx^2/2}dy \\ &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} e^{-z^2-tx^2/2}\sqrt{2t}dz \\ &= \frac{1}{\sqrt{\pi}}e^{-tx^2/2}\int_{\mathbb{R}} e^{-z^2}dz \\ &= e^{-tx^2/2} \end{aligned}$$

As for the last assertion, this follows from the fact that $\log F_{g_t}$ is linear in t. □

Observe that, thinking retrospectively, the above computation formally solves the question raised in chapter 1, and so could have been used there, afterwards. However, all this remains based on the Gauss integral formula, and there is no escape from that.

2c. CLT, revised

CLT, revised.

2d. Basic illustrations

Basic illustrations.

2e. Exercises

This was a standard chapter on the normal laws, and as exercises, we have:

Exercise 2.11. *Clarify the details in the first proof of the Gauss formula.*

Exercise 2.12. *Clarify the details in the proof of the change of variables theorem.*

Exercise 2.13. *Does the proof of this theorem simplify, for the polar coordinates?*

Exercise 2.14. *What about spherical coordinates in* $3D$*, any simplifications there?*

Exercise 2.15. *By the way, clarify the range of the angles, in that* $3D$ *formula.*

Exercise 2.16. *And also, learn about the stereographic projection too.*

Exercise 2.17. *Learn some other proofs of the Gauss formula.*

Exercise 2.18. *Further work on the exact convergence in the CLT.*

As bonus exercise, do some experiments, reaching to normal law readings.

CHAPTER 3

Advanced formulae

3a. Cumulants

We have seen a lot of interesting combinatorics in the previous chapter, but this is not the end of the story. Following Rota, let us formulate indeed the following definition:

Definition 3.1. *Associated to any real probability measure $\mu = \mu_f$ is the following modification of the logarithm of the Fourier transform $F_\mu(\xi) = E(e^{i\xi f})$,*

$$K_\mu(\xi) = \log E(e^{\xi f})$$

called cumulant-generating function. The Taylor coefficients $k_n(\mu)$ of this series, given by

$$K_\mu(\xi) = \sum_{n=1}^\infty k_n(\mu)\,\frac{\xi^n}{n!}$$

are called cumulants of the measure μ. We also use the notations k_f, K_f for these cumulants and their generating series, where f is a variable following the law μ.

In other words, the cumulants are more or less the coefficients of the logarithm of the Fourier transform $\log F_\mu$, up to some normalizations. To be more precise, we have $K_\mu(\xi) = \log F_\mu(-i\xi)$, so the formula relating $\log F_\mu$ to the cumulants $k_n(\mu)$ is:

$$\log F_\mu(-i\xi) = \sum_{n=1}^\infty k_n(\mu)\,\frac{\xi^n}{n!}$$

Equivalently, the formula relating $\log F_\mu$ to the cumulants $k_n(\mu)$ is:

$$\log F_\mu(\xi) = \sum_{n=1}^\infty k_n(\mu)\,\frac{(i\xi)^n}{n!}$$

We will see in a moment the reasons for the above normalizations, namely change of variables $\xi \to -i\xi$, and Taylor coefficients instead of plain coefficients, the idea being that for simple laws like g_t, p_t, we will obtain in this way very simple quantities. Let us also mention that there is a reason for indexing the cumulants by $n = 1, 2, 3, \ldots$ instead of $n = 0, 1, 2, \ldots$, and more on this later, once we will have some theory and examples.

As a first observation, the sequence of cumulants $k_1, k_2, k_3, \ldots$ appears as a modification of the sequence of moments $M_1, M_2, M_3, \ldots$, the numerics being as follows:

Proposition 3.2. *The sequence of cumulants $k_1, k_2, k_3, \ldots$ appears as a modification of the sequence of moments $M_1, M_2, M_3, \ldots$, and uniquely determines μ. We have*

$$k_1 = M_1$$

$$k_2 = -M_1^2 + M_2$$

$$k_3 = 2M_1^3 - 3M_1M_2 + M_3$$

$$k_4 = -6M_1^4 + 12M_1^2M_2 - 3M_2^2 - 4M_1M_3 + M_4$$

$$\vdots$$

in one sense, and in the other sense we have

$$M_1 = k_1$$

$$M_2 = k_1^2 + k_2$$

$$M_3 = k_1^3 + 3k_1k_2 + k_3$$

$$M_4 = k_1^4 + 6k_1^2k_2 + 3k_2^2 + 4k_1k_3 + k_4$$

$$\vdots$$

with in both cases the correspondence being polynomial, with integer coefficients.

Proof. We know from Definition 3.1 that the cumulants are given by:

$$\log E(e^{\xi f}) = \sum_{s=1}^{\infty} k_s(f) \frac{\xi^s}{s!}$$

By exponentiating, we obtain from this the following formula:

$$E(e^{\xi f}) = \exp\left(\sum_{s=1}^{\infty} k_s(f) \frac{\xi^s}{s!}\right)$$

Now by looking at the terms of order $1, 2, 3, 4$, this gives the above formulae. □

The interest in cumulants comes from the fact that $\log F_\mu$, and so the cumulants $k_n(\mu)$ too, linearize the convolution. To be more precise, we have the following result:

Theorem 3.3. *The cumulants have the following properties:*

1. $k_n(cf) = c^n k_n(f)$.
2. $k_1(f + d) = k_1(f) + d$, *and* $k_n(f + d) = k_n(f)$ *for* $n > 1$.
3. $k_n(f + g) = k_n(f) + k_n(g)$, *if* f, g *are independent.*

PROOF. Here (1) and (2) are both clear from definitions, because we have:

$$\begin{aligned} K_{cf+d}(\xi) &= \log E(e^{\xi(cf+d)}) \\ &= \log[e^{\xi d} \cdot E(e^{\xi cf})] \\ &= \xi d + K_f(c\xi) \end{aligned}$$

As for (3), this follows from the fact that the Fourier transform $F_f(\xi) = E(e^{i\xi f})$ satisfies the following formula, whenever f, g are independent random variables:

$$F_{f+g}(\xi) = F_f(\xi)F_g(\xi)$$

Indeed, by applying the logarithm, we obtain the following formula:

$$\log F_{f+g}(\xi) = \log F_f(\xi) + \log F_g(\xi)$$

With the change of variables $\xi \to -i\xi$, we obtain the following formula:

$$K_{f+g}(\xi) = K_f(\xi) + K_g(\xi)$$

Thus, at the level of coefficients, we obtain $k_n(f+g) = k_n(f) + k_n(g)$, as claimed. □

At the level of examples now, we have the following result:

THEOREM 3.4. *The sequence of cumulants $k_1, k_2, k_3, \ldots$ is as follows:*

(1) *For $\mu = \delta_c$ the cumulants are $c, 0, 0, \ldots$*
(2) *For $\mu = g_t$ the cumulants are $0, t, 0, 0, \ldots$*
(3) *For $\mu = p_t$ the cumulants are $t, t, t, \ldots$*

PROOF. We have 3 computations to be done, the idea being as follows:

(1) For $\mu = \delta_c$ we have the following computation:

$$\begin{aligned} K_\mu(\xi) &= \log E(e^{c\xi}) \\ &= \log(e^{c\xi}) \\ &= c\xi \end{aligned}$$

But the plain coefficients of this series are the numbers $c, 0, 0, \ldots$, and so the Taylor coefficients of this series are these same numbers $c, 0, 0, \ldots$, as claimed.

(2) For $\mu = g_t$ we have the following computation:

$$\begin{aligned} K_\mu(\xi) &= \log F_\mu(-i\xi) \\ &= \log\exp\left[-t(-i\xi)^2/2\right] \\ &= t\xi^2/2 \end{aligned}$$

But the plain coefficients of this series are the numbers $0, t/2, 0, 0, \ldots$, and so the Taylor coefficients of this series are the numbers $0, t, 0, 0, \ldots$, as claimed.

(3) For $\mu = p_t$ we have the following computation:

$$\begin{aligned} K_\mu(\xi) &= \log F_\mu(-i\xi) \\ &= \log\exp\left[(e^{i(-i\xi)}-1)t\right] \\ &= (e^\xi - 1)t \end{aligned}$$

But the plain coefficients of this series are the numbers $t/n!$, and so the Taylor coefficients of this series are the numbers $t, t, t, \ldots$, as claimed. $\square$

There are many other interesting illustrations. We will be back to this.

3b. Inversion formula

Getting back to theory now, the sequence of cumulants $k_1, k_2, k_3, \ldots$ appears as a modification of the sequence of moments $M_1, M_2, M_3, \ldots$, and understanding the relation between moments and cumulants will be our next task. Let us start with:

Definition 3.5. *The Möbius function of any lattice, and so of P, is given by*

$$\mu(\pi,\nu) = \begin{cases} 1 & \text{if } \pi = \nu \\ -\sum_{\pi \leq \tau < \nu} \mu(\pi,\tau) & \text{if } \pi < \nu \\ 0 & \text{if } \pi \not\leq \nu \end{cases}$$

with the construction being performed by recurrence.

As an illustration here, for $P(2) = \{||, \sqcap\}$, we have by definition:

$$\mu(||,||) = \mu(\sqcap,\sqcap) = 1$$

Also, $|| < \sqcap$, with no intermediate partition in between, so we obtain:

$$\mu(||,\sqcap) = -\mu(||,||) = -1$$

Finally, we have $\sqcap \not\leq ||$, and so we have as well the following formula:

$$\mu(\sqcap,||) = 0$$

The main interest in the Möbius function comes from the Möbius inversion formula, which in linear algebra terms can be stated and proved as follows:

Theorem 3.6. *We have the following implication,*

$$f(\pi) = \sum_{\nu \leq \pi} g(\nu) \quad \Longrightarrow \quad g(\pi) = \sum_{\nu \leq \pi} \mu(\nu,\pi) f(\nu)$$

valid for any two functions $f, g : P(n) \to \mathbb{C}$.

PROOF. Consider the adjacency matrix of P, given by the following formula:

$$A_{\pi\nu} = \begin{cases} 1 & \text{if } \pi \leq \nu \\ 0 & \text{if } \pi \not\leq \nu \end{cases}$$

Our claim is that the inverse of this matrix is the Möbius matrix of P, given by:

$$M_{\pi\nu} = \mu(\pi, \nu)$$

Indeed, the above matrix A is upper triangular, and when trying to invert it, we are led to the recurrence in Definition 3.5, so to the Möbius matrix M. Thus we have:

$$M = A^{-1}$$

Thus, in practice, we are led to the inversion formula in the statement. □

With these ingredients in hand, let us go back to probability. We first have:

DEFINITION 3.7. *We define quantities $M_\pi(f), k_\pi(f)$, depending on partitions*

$$\pi \in P(k)$$

by starting with $M_n(f), k_n(f)$, and using multiplicativity over the blocks.

To be more precise, the convention here is that for the one-block partition $1_n \in P(n)$, the corresponding moment and cumulant are the usual ones, namely:

$$M_{1_n}(f) = M_n(f) \quad , \quad k_{1_n}(f) = k_n(f)$$

Then, for an arbitrary partition $\pi \in P(k)$, we decompose this partition into blocks, having sizes $b_1, \ldots, b_s$, and we set, by multiplicativity over blocks:

$$M_\pi(f) = M_{b_1}(f) \ldots M_{b_s}(f) \quad , \quad k_\pi(f) = k_{b_1}(f) \ldots k_{b_s}(f)$$

With this convention, following Rota and others, we can now formulate a key result, fully clarifying the relation between moments and cumulants, as follows:

THEOREM 3.8. *We have the moment-cumulant formulae*

$$M_n(f) = \sum_{\nu \in P(n)} k_\nu(f) \quad , \quad k_n(f) = \sum_{\nu \in P(n)} \mu(\nu, 1_n) M_\nu(f)$$

or, equivalently, we have the moment-cumulant formulae

$$M_\pi(f) = \sum_{\nu \leq \pi} k_\nu(f) \quad , \quad k_\pi(f) = \sum_{\nu \leq \pi} \mu(\nu, \pi) M_\nu(f)$$

where μ is the Möbius function of $P(n)$.

Proof. There are several things going on here, the idea being as follows:

(1) According to our conventions above, the first set of formulae is equivalent to the second set of formulae. Also, due to the Möbius inversion formula, in the second set of formulae, the two formulae there are in fact equivalent. Thus, the 4 formulae in the statement are all equivalent. In what follows we will focus on the first 2 formulae.

(2) Let us first work out some examples. At $n=1,2,3$ the moment formula gives the following equalities, which are in tune with the findings from Proposition 3.2:

$$M_1 = k_{|} = k_1$$

$$M_2 = k_{|\,|} + k_{\sqcap} = k_1^2 + k_2$$

$$M_3 = k_{|\,|\,|} + k_{\sqcap|} + k_{\sqcap\hskip-5pt|\,} + k_{|\sqcap} + k_{\sqcap\sqcap} = k_1^3 + 3k_1k_2 + k_3$$

At $n=4$ now, which is a case which is of particular interest for certain considerations to follow, the computation is as follows, again in tune with Proposition 3.2:

$$\begin{aligned} M_4 &= k_{|\,|\,|} + (\underbrace{k_{\sqcap|\,|} + \ldots}_{6\ terms}) + (\underbrace{k_{\sqcap\sqcap} + \ldots}_{3\ terms}) + (\underbrace{k_{\sqcap\!\sqcap|} + \ldots}_{4\ terms}) + k_{\sqcap\!\sqcap\!\sqcap} \\ &= k_1^4 + 6k_1^2k_2 + 3k_2^2 + 4k_1k_3 + k_4 \end{aligned}$$

As for the cumulant formula, at $n=1,2,3$ this gives the following formulae for the cumulants, again in tune with the findings from Proposition 3.2:

$$k_1 = M_{|} = M_1$$

$$k_2 = (-1)M_{|\,|} + M_{\sqcap} = -M_1^2 + M_2$$

$$k_3 = 2M_{|\,|\,|} + (-1)M_{\sqcap|} + (-1)M_{\sqcap\hskip-5pt|\,} + (-1)M_{|\sqcap} + M_{\sqcap\sqcap} = 2M_1^3 - 3M_1M_2 + M_3$$

Finally, at $n=4$, after computing the Möbius function of $P(4)$, we obtain the following formula for the fourth cumulant, again in tune with Proposition 3.2:

$$\begin{aligned} k_4 &= (-6)M_{|\,|\,|} + 2(\underbrace{M_{\sqcap|\,|} + \ldots}_{6\ terms}) + (-1)(\underbrace{M_{\sqcap\sqcap} + \ldots}_{3\ terms}) + (-1)(\underbrace{M_{\sqcap\!\sqcap|} + \ldots}_{4\ terms}) + M_{\sqcap\!\sqcap\!\sqcap} \\ &= -6M_1^4 + 12M_1^2M_2 - 3M_2^2 - 4M_1M_3 + M_4 \end{aligned}$$

(3) Time now to get to work, and prove the result. As mentioned above, the formulae in the statement are all equivalent, and it is enough to prove the first one, namely:

$$M_n(f) = \sum_{\nu \in P(n)} k_\nu(f)$$

In order to do this, we use the very definition of the cumulants, namely:

$$\log E(e^{\xi f}) = \sum_{s=1}^\infty k_s(f)\,\frac{\xi^s}{s!}$$

By exponentiating, we obtain from this the following formula:

$$E(e^{\xi f})=\exp\left(\sum_{s=1}^\infty k_s(f)\,\frac{\xi^s}{s!}\right)$$

(4) Let us first compute the function on the left. This is easily done, as follows:

$$E(e^{\xi f})=E\left(\sum_{n=0}^\infty\frac{(\xi f)^n}{n!}\right)=\sum_{n=0}^\infty M_n(f)\,\frac{\xi^n}{n!}$$

(5) Regarding now the function on the right, this is given by:

$$\begin{aligned}\exp\left(\sum_{s=1}^\infty k_s(f)\,\frac{\xi^s}{s!}\right) &= \sum_{p=0}^\infty\frac{\left(\sum_{s=1}^\infty k_s(f)\,\frac{\xi^s}{s!}\right)^p}{p!}\\ &= \sum_{p=0}^\infty\frac{1}{p!}\sum_{s_1=1}^\infty k_{s_1}(f)\,\frac{\xi^{s_1}}{s_1!}\ldots\ldots\sum_{s_p=1}^\infty k_{s_p}(f)\,\frac{\xi^{s_p}}{s_p!}\\ &= \sum_{p=0}^\infty\frac{1}{p!}\sum_{s_1=1}^\infty\ldots\sum_{s_p=1}^\infty k_{s_1}(f)\ldots k_{s_p}(f)\,\frac{\xi^{s_1+\ldots+s_p}}{s_1!\ldots s_p!}\end{aligned}$$

But the point now is that all this leads us into partitions. Indeed, we are summing over indices $s_1,\ldots,s_p\in\mathbb{N}$, which can be thought of as corresponding to a partition of $n=s_1+\ldots+s_p$. So, let us rewrite our sum, as a sum over partitions. For this purpose, recall that the number of partitions $\nu\in P(n)$ having blocks of sizes $s_1,\ldots,s_p$ is:

$$\binom{n}{s_1,\ldots,s_p}=\frac{n!}{p_1!\ldots p_s!}$$

Also, when resumming over partitions, there will be a $p!$ factor as well, coming from the permutations of $s_1,\ldots,s_p$. Thus, our sum can be rewritten as follows:

$$\begin{aligned}\exp\left(\sum_{s=1}^\infty k_s(f)\,\frac{\xi^s}{s!}\right) &= \sum_{n=0}^\infty\sum_{p=0}^\infty\frac{1}{p!}\sum_{s_1+\ldots+s_p=n}k_{s_1}(f)\ldots k_{s_p}(f)\,\frac{\xi^n}{s_1!\ldots s_p!}\\ &= \sum_{n=0}^\infty\frac{\xi^n}{n!}\sum_{p=0}^\infty\frac{1}{p!}\sum_{s_1+\ldots+s_p=n}\binom{n}{s_1,\ldots,s_p}k_{s_1}(f)\ldots k_{s_p}(f)\\ &= \sum_{n=0}^\infty\frac{\xi^n}{n!}\sum_{\nu\in P(n)}k_\nu(f)\end{aligned}$$

(6) We are now in position to conclude. According to (3,4,5), we have:

$$\sum_{n=0}^\infty M_n(f)\,\frac{\xi^n}{n!}=\sum_{n=0}^\infty\frac{\xi^n}{n!}\sum_{\nu\in P(n)}k_\nu(f)$$

Thus, we have the following formula, valid for any $n \in \mathbb{N}$:

$$M_n(f) = \sum_{\nu \in P(n)} k_\nu(f)$$

We are therefore led to the conclusions in the statement. □

3c. Stieltjes inversion

An interesting question, that we met since chapter 1, is how to recover a probability measure out of its moments. And the answer here, which is non-trivial, is as follows:

THEOREM 3.9. *The density of a real probability measure μ can be recaptured from the sequence of moments $\{M_k\}_{k\geq 0}$ via the Stieltjes inversion formula*

$$d\mu(x) = \lim_{t \searrow 0} -\frac{1}{\pi} Im\left(G(x+it)\right) \cdot dx$$

where the function on the right, given in terms of moments by

$$G(\xi) = \xi^{-1} + M_1\xi^{-2} + M_2\xi^{-3} + \ldots$$

is the Cauchy transform of the measure μ.

PROOF. The Cauchy transform of our measure μ is given by:

$$\begin{aligned} G(\xi) &= \xi^{-1}\sum_{k=0}^{\infty} M_k \xi^{-k} \\ &= \int_{\mathbb{R}} \frac{\xi^{-1}}{1-\xi^{-1}y}\, d\mu(y) \\ &= \int_{\mathbb{R}} \frac{1}{\xi - y}\, d\mu(y) \end{aligned}$$

Now with $\xi = x + it$, we obtain the following formula:

$$\begin{aligned} Im(G(x+it)) &= \int_{\mathbb{R}} Im\left(\frac{1}{x-y+it}\right) d\mu(y) \\ &= \int_{\mathbb{R}} \frac{1}{2i}\left(\frac{1}{x-y+it} - \frac{1}{x-y-it}\right) d\mu(y) \\ &= -\int_{\mathbb{R}} \frac{t}{(x-y)^2+t^2}\, d\mu(y) \end{aligned}$$

By integrating over $[a,b]$ we obtain, with the change of variables $x=y+tz$:

$$\begin{aligned}\int_a^b Im(G(x+it))dx &= -\int_{\mathbb{R}}\int_a^b \frac{t}{(x-y)^2+t^2}\,dx\,d\mu(y)\\ &= -\int_{\mathbb{R}}\int_{(a-y)/t}^{(b-y)/t}\frac{t}{(tz)^2+t^2}\,t\,dz\,d\mu(y)\\ &= -\int_{\mathbb{R}}\int_{(a-y)/t}^{(b-y)/t}\frac{1}{1+z^2}\,dz\,d\mu(y)\\ &= -\int_{\mathbb{R}}\left(\arctan\frac{b-y}{t}-\arctan\frac{a-y}{t}\right)d\mu(y)\end{aligned}$$

Now observe that with $t\searrow 0$ we have:

$$\lim_{t\searrow 0}\left(\arctan\frac{b-y}{t}-\arctan\frac{a-y}{t}\right)=\begin{cases}\frac{\pi}{2}-\frac{\pi}{2}=0 & (y<a)\\ \frac{\pi}{2}-0=\frac{\pi}{2} & (y=a)\\ \frac{\pi}{2}-(-\frac{\pi}{2})=\pi & (a<y<b)\\ 0-(-\frac{\pi}{2})=\frac{\pi}{2} & (y=b)\\ -\frac{\pi}{2}-(-\frac{\pi}{2})=0 & (y>b)\end{cases}$$

We therefore obtain the following formula:

$$\lim_{t\searrow 0}\int_a^b Im(G(x+it))dx=-\pi\left(\mu(a,b)+\frac{\mu(a)+\mu(b)}{2}\right)$$

Thus, we are led to the conclusion in the statement. □

Before getting further, let us mention that the above result does not fully solve the moment problem, because we still have the question of understanding when a sequence of numbers $M_1,M_2,M_3,\ldots$ can be the moments of a measure μ. For instance we certainly must have $M_0=1$, but we must have as well the following inequality:

$$M_2\geq M_1^2$$

In answer now, we have the following result, complementing Theorem 3.9:

THEOREM 3.10. *A sequence of numbers $M_0,M_1,M_2,M_3,\ldots\in\mathbb{R}$, with $M_0=1$, is the series of moments of a real probability measure μ precisely when:*

$$\begin{vmatrix}M_0\end{vmatrix}\geq 0\quad,\quad\begin{vmatrix}M_0&M_1\\M_1&M_2\end{vmatrix}\geq 0\quad,\quad\begin{vmatrix}M_0&M_1&M_2\\M_1&M_2&M_3\\M_2&M_3&M_4\end{vmatrix}\geq 0\quad,\quad\ldots$$

That is, the associated Hankel determinants must be all positive.

PROOF. This is something a bit more advanced, the idea being as follows:

(1) As a first observation, the positivity conditions in the statement tell us that the following associated linear forms must be positive:

$$\sum_{i,j=1}^n c_i\bar{c}_jM_{i+j}\geq 0$$

(2) But this is something very classical, in one sense the result being elementary, coming from the following computation, which shows that we have positivity indeed:

$$\begin{aligned}\int_\mathbb{R}\left|\sum_{i=1}^nc_ix^i\right|^2d\mu(x)&=\int_\mathbb{R}\sum_{i,j=1}^nc_i\bar{c}_jx^{i+j}d\mu(x)\\&=\sum_{i,j=1}^nc_i\bar{c}_jM_{i+j}\end{aligned}$$

(3) As for the other sense, here the result comes once again from the above formula, this time via some standard functional analysis. □

Getting back now to more concrete things, the point is that we have:

FACT 3.11. *Given a graph X, with distinguished vertex $*$, we can talk about the probability measure μ having as k-th moment the number of length k loops based at $*$:*

$$M_k=\left\{*-i_1-i_2-\ldots-i_k=*\right\}$$

As basic examples, for the graph $\mathbb{N}$ the moments must be the Catalan numbers C_k, and for the graph $\mathbb{Z}$, the moments must be the central binomial coefficients D_k.

To be more precise, the first assertion, regarding the existence and uniqueness of μ, follows from a basic linear algebra computation, by diagonalizing the adjacency matrix of X. As for the examples, involving the graphs $\mathbb{N}$ and $\mathbb{Z}$, these are both very standard.

Needless to say, counting loops on graphs, as in Fact 3.11, is something important in applied mathematics, and physics. So, back to our business now, motivated by all this, as a basic application of the Stieltjes formula, let us solve the moment problem for the Catalan numbers C_k, and for the central binomial coefficients D_k. We first have:

THEOREM 3.12. *The real measure having as even moments the Catalan numbers, $C_k=\frac{1}{k+1}\binom{2k}{k}$, and having all odd moments 0 is the measure*

$$\gamma_1=\frac{1}{2\pi}\sqrt{4-x^2}dx$$

called Wigner semicircle law on $[-2,2]$.

PROOF. In order to apply the inversion formula, our starting point will be the well-known formula for the generating series of the Catalan numbers, namely:

$$\sum_{k=0}^{\infty} C_k z^k = \frac{1-\sqrt{1-4z}}{2z}$$

By using this formula with $z = \xi^{-2}$, we obtain the following formula:

$$\begin{aligned} G(\xi) &= \xi^{-1}\sum_{k=0}^{\infty} C_k \xi^{-2k} \\ &= \xi^{-1}\cdot\frac{1-\sqrt{1-4\xi^{-2}}}{2\xi^{-2}} \\ &= \frac{\xi}{2}\left(1-\sqrt{1-4\xi^{-2}}\right) \\ &= \frac{\xi}{2}-\frac{1}{2}\sqrt{\xi^2-4} \end{aligned}$$

Now let us apply Theorem 3.9. The study here goes as follows:

(1) According to the general philosophy of the Stieltjes formula, the first term, namely $\xi/2$, which is "trivial", will not contribute to the density.

(2) As for the second term, which is something non-trivial, this will contribute to the density, the rule here being that the square root $\sqrt{\xi^2-4}$ will be replaced by the "dual" square root $\sqrt{4-x^2}\,dx$, and that we have to multiply everything by $-1/\pi$.

(3) As a conclusion, by Stieltjes inversion we obtain the following density:

$$d\mu(x) = -\frac{1}{\pi}\cdot -\frac{1}{2}\sqrt{4-x^2}\,dx = \frac{1}{2\pi}\sqrt{4-x^2}dx$$

Thus, we have obtained the mesure in the statement, and we are done. □

We have the following version of the above result:

THEOREM 3.13. *The real measure having as sequence of moments the Catalan numbers, $C_k = \frac{1}{k+1}\binom{2k}{k}$, is the measure*

$$\pi_1 = \frac{1}{2\pi}\sqrt{4x^{-1}-1}\,dx$$

called Marchenko-Pastur law on $[0,4]$.

PROOF. As before, we use the standard formula for the generating series of the Catalan numbers. With $z = \xi^{-1}$ in that formula, we obtain the following formula:

$$\begin{aligned} G(\xi) &= \xi^{-1}\sum_{k=0}^\infty C_k\xi^{-k} \\ &= \xi^{-1}\cdot\frac{1-\sqrt{1-4\xi^{-1}}}{2\xi^{-1}} \\ &= \frac{1}{2}\left(1-\sqrt{1-4\xi^{-1}}\right) \\ &= \frac{1}{2}-\frac{1}{2}\sqrt{1-4\xi^{-1}} \end{aligned}$$

With this in hand, let us apply now the Stieltjes inversion formula, from Theorem 3.9. We obtain, a bit as before in Theorem 3.12, the following density:

$$d\mu(x) = -\frac{1}{\pi}\cdot -\frac{1}{2}\sqrt{4x^{-1}-1}\,dx = \frac{1}{2\pi}\sqrt{4x^{-1}-1}\,dx$$

Thus, we are led to the conclusion in the statement. □

Regarding now the central binomial coefficients, we have here:

THEOREM 3.14. *The real probability measure having as moments the central binomial coefficients, $D_k = \binom{2k}{k}$, is the measure*

$$\alpha_1 = \frac{1}{\pi\sqrt{x(4-x)}}\,dx$$

called arcsine law on $[0,4]$.

PROOF. We have the following computation, using some well-known formulae:

$$\begin{aligned} G(\xi) &= \xi^{-1}\sum_{k=0}^\infty D_k\xi^{-k} \\ &= \frac{1}{\xi}\sum_{k=0}^\infty D_k\left(-\frac{t}{4}\right)^k \\ &= \frac{1}{\xi}\cdot\frac{1}{\sqrt{1-4/\xi}} \\ &= \frac{1}{\sqrt{\xi(\xi-4)}} \end{aligned}$$

But this gives the density in the statement, via Theorem 3.9. □

Finally, we have the following version of the above result:

THEOREM 3.15. *The real probability measure having as moments the middle binomial coefficients, $E_k = \binom{k}{[k/2]}$, is the following law on $[-2,2]$,*

$$\sigma_1 = \frac{1}{2\pi}\sqrt{\frac{2+x}{2-x}}\,dx$$

called modified the arcsine law on $[-2,2]$.

PROOF. In terms of the central binomial coefficients D_k, we have:

$$E_{2k} = D_k \quad , \quad E_{2k-1} = \frac{D_k}{2}$$

Standard calculus based on the Taylor formula for $(1+t)^{-1/2}$ gives:

$$\frac{1}{2x}\left(\sqrt{\frac{1+2x}{1-2x}} - 1\right) = \sum_{k=0}^{\infty} E_k x^k$$

With $x = \xi^{-1}$ we obtain the following formula for the Cauchy transform:

$$\begin{aligned} G(\xi) &= \xi^{-1}\sum_{k=0}^{\infty} E_k \xi^{-k} \\ &= \frac{1}{\xi}\left(\sqrt{\frac{1+2/\xi}{1-2/\xi}} - 1\right) \\ &= \frac{1}{\xi}\left(\sqrt{\frac{\xi+2}{\xi-2}} - 1\right) \end{aligned}$$

By Stieltjes inversion we obtain the density in the statement. □

Finally, the above technology applies of course to the normal laws too.

3d. Orthogonal polynomials

Let us start our discussion with the following standard result:

THEOREM 3.16. *Any Hilbert space H has an orthonormal basis $\{e_i\}_{i\in I}$, which is by definition a set of vectors whose span is dense in H, and which satisfy*

$$< e_i, e_j >= \delta_{ij}$$

with δ being a Kronecker symbol. The cardinality $|I|$ of the index set, which can be finite, countable, or uncountable, depends only on H, and is called dimension of H. We have

$$H \simeq l^2(I)$$

in the obvious way, mapping $\sum \lambda_i e_i \to (\lambda_i)$. The Hilbert spaces with $\dim H = |I|$ being countable, such as $l^2(\mathbb{N})$, are all isomorphic, and are called separable.

Proof. We have many assertions here, the idea being as follows:

(1) In finite dimensions an orthonormal basis $\{e_i\}_{i\in I}$ can be constructed by starting with any vector space basis $\{f_i\}_{i\in I}$, and using the Gram-Schmidt procedure. As for the other assertions, these are all clear, from basic linear algebra.

(2) In general, the same method works, namely Gram-Schmidt, with a subtlety coming from the fact that the basis $\{e_i\}_{i\in I}$ will not span in general the whole H, but just a dense subspace of it, as it is in fact obvious by looking at the standard basis of $l^2(\mathbb{N})$.

(3) And there is a second subtlety as well, coming from the fact that the recurrence procedure needed for Gram-Schmidt must be replaced by some sort of "transfinite recurrence", using standard tools from logic, and more specifically the Zorn lemma.

(4) Finally, everything at the end, regarding our notion of separability for the Hilbert spaces, is clear from definitions, and from our various results above. □

According to Theorem 3.16, there is only one separable Hilbert space, up to isomorphism. There are many interesting things that can be said, about this magic and unique Hilbert space. As a first result such result, which is something theoretical, we have:

Theorem 3.17. *The following happen, in relation with separability:*

1. *The Hilbert space $H = L^2[-1,1]$ is separable, with orthonormal basis coming by applying Gram-Schmidt to the basis $\{x^k\}_{k\in\mathbb{N}}$, coming from Weierstrass.*
2. *In fact, any $H = L^2(\mathbb{R},\mu)$, with $d\mu(x) = f(x)dx$, is separable, and the same happens in higher dimensions, for $H = L^2(\mathbb{R}^N,\mu)$, with $d\mu(x) = f(x)dx$.*
3. *More generally, given a separable abstract measured space X, the associated Hilbert space of square-summable functions $H = L^2(X)$ is separable.*

Proof. Many things can be said here, the idea being as follows:

(1) The fact that $H = L^2[-1,1]$ is separable is clear indeed from the Weierstrass density theorem, which provides us with the algebraic basis $g_k = x^k$, which can be orthogonalized by using the Gram-Schmidt procedure, as explained in Theorem 3.16.

(2) Regarding now more general spaces, of type $H = L^2(\mathbb{R},\mu)$, we can use here the same argument, after modifying if needed our measure μ, in order for the functions $g_k = x^k$ to be indeed square-summable. As for higher dimensions, the situation here is similar, because we can use here the multivariable polynomials $g_k(x) = x_1^{k_1}\ldots x_N^{k_N}$.

(3) Finally, the last assertion, regarding the general spaces of type $H = L^2(X)$, which generalizes all this, comes as a consequence of general measure theory, and we will leave some learning, and working out the details here, as an instructive exercise. □

As a conclusion to all this, which is a bit philosophical, we have:

CONCLUSION 3.18. *We are interested in one space, namely the unique separable Hilbert space H, but due to various technical reasons, it is often better to forget that we have $H = l^2(\mathbb{N})$, and say instead that we have $H = L^2(X)$, with X being a separable measured space, or simply say that H is an abstract separable Hilbert space.*

It is also possible to make some physics comments here, with the unique separable Hilbert space H from Conclusion 3.18, that we will be presumably obsessed with, in what follows, being, and no surprise here, the space that we live in.

Let us go back now to Theorem 3.17 and its proof, which is something quite subtle. That material leads us into orthogonal polynomials, which are defined as follows:

DEFINITION 3.19. *The orthogonal polynomials with respect to $d\mu(x) = f(x)dx$ are polynomials $P_k \in \mathbb{R}[x]$ of degree $k \in \mathbb{N}$, which are orthogonal inside $H = L^2(\mathbb{R}, \mu)$:*

$$\int_{\mathbb{R}} P_k(x)P_l(x)f(x)dx = 0 \quad , \quad \forall k \neq l$$

Equivalently, these orthogonal polynomials $\{P_k\}_{k\in\mathbb{N}}$, which are each unique modulo scalars, appear from the Weierstrass basis $\{x^k\}_{k\in\mathbb{N}}$, by doing Gram-Schmidt.

As a first observation, the orthogonal polynomials exist indeed for any real measure $d\mu(x) = f(x)dx$, because we can obtain them from the monomials x^k via Gram-Schmidt, as indicated above. It is possible to be a bit more explicit here, as follows:

THEOREM 3.20. *The orthogonal polynomials with respect to μ are given by*

$$P_k = c_k \begin{vmatrix} M_0 & M_1 & \dots & M_k \\ M_1 & M_2 & \dots & M_{k+1} \\ \vdots & \vdots & & \vdots \\ M_{k-1} & M_k & \dots & M_{2k-1} \\ 1 & x & \dots & x^k \end{vmatrix}$$

where $M_k = \int_{\mathbb{R}} x^k d\mu(x)$ are the moments of μ, and $c_k \in \mathbb{R}^*$ can be any numbers.

PROOF. Let us first see what happens at small values of $k \in \mathbb{N}$. At $k = 0$ our formula is as follows, stating that the first polynomial P_0 must be a constant, as it should:

$$P_0 = c_0|M_0| = c_0$$

At $k = 1$ now, again by using $M_0 = 1$, the formula is as follows:

$$P_1 = c_1 \begin{vmatrix} M_0 & M_1 \\ 1 & x \end{vmatrix} = c_1(x - M_1)$$

But this is again the good formula, because the degree is 1, and we have:

$$\begin{aligned} < 1, P_1 > &= c_1 < 1, x - M_1 > \\ &= c_1(< 1, x > - < 1, M_1 >) \\ &= c_1(M_1 - M_1) \\ &= 0 \end{aligned}$$

At $k = 2$ now, things get more complicated, with the formula being as follows:

$$P_2 = c_2 \begin{vmatrix} M_0 & M_1 & M_2 \\ M_1 & M_2 & M_3 \\ 1 & x & x^2 \end{vmatrix}$$

However, no need for big computations here, in order to check the orthogonality, because by using the fact that x^k integrates up to M_k, we obtain:

$$< 1, P_2 >= \int_{\mathbb{R}} P_2(x)d\mu(x) = c_2 \begin{vmatrix} M_0 & M_1 & M_2 \\ M_1 & M_2 & M_3 \\ M_0 & M_1 & M_2 \end{vmatrix} = 0$$

Similarly, again by using the fact that x^k integrates up to M_k, we have as well:

$$< x, P_2 >= \int_{\mathbb{R}} xP_2(x)d\mu(x) = c_2 \begin{vmatrix} M_0 & M_1 & M_2 \\ M_1 & M_2 & M_3 \\ M_1 & M_2 & M_3 \end{vmatrix} = 0$$

Thus, result proved at $k = 0, 1, 2$, and the proof in general is similar. □

In practice now, all this leads us to a lot of interesting combinatorics, and countless things can be said. For the simplest measured space $X \subset \mathbb{R}$, which is the interval $[-1, 1]$, with its uniform measure, the orthogonal basis problem can be solved as follows:

THEOREM 3.21. *The orthonormal polynomials for $L^2[-1, 1]$, subject to*

$$\int_{-1}^{1} P_k(x)P_l(x)\,dx = \delta_{kl}$$

and called Legendre polynomials, satisfy the equation

$$(1 - x^2)P_k''(x) - 2xP_k'(x) + k(k + 1)P_k(x) = 0$$

which is the Legendre equation from physics. Moreover, we have the formula

$$P_k(x) = \frac{1}{2^k k!} \cdot \frac{d^k}{dx^k} \left(1 - x^2\right)^k$$

called Rodrigues formula for the Legendre polynomials.

PROOF. The idea here is that thinking at what Gram-Schmidt does, this is certainly something by recurrence. And examining the recurrence leads to the Legendre equation. As for the Rodrigues formula, we have two choices here, either by verifying that $\{P_k\}$ is orthonormal, or by verifying the Legendre equation. And both methods work. □

The above result is just the tip of the iceberg, and as a continuation, we have:

THEOREM 3.22. *The orthogonal polynomials for $L^2[-1,1]$, with measure*

$$d\mu(x) = (1-x)^\alpha(1+x)^\beta dx$$

called Jacobi polynomials, satisfy as well a degree 2 equation, and are given by:

$$P_k(x) = \frac{(-1)^k}{2^k k!}(1-x)^{-\alpha}(1+x)^{-\beta}\frac{d^k}{dx^k}\left[(1-x)^\alpha(1+x)^\beta(1-x^2)^k\right]$$

At $\alpha = \beta = 0$ we recover the Legendre polynomials, and at $\alpha = \beta = \pm\frac{1}{2}$ we recover the Chebycheff polynomials of the first and second kind, from trigonometry.

PROOF. Obviously, many things going on here, and much more can be added, but the idea is quite simple, namely that this appears as a generalization of Theorem 10.31. We will leave learning more about all this as an interesting exercise. □

Getting now to other spaces $X \subset \mathbb{R}$, of particular interest here is the following result, which complements well Theorem 10.31, for the needs of basic quantum mechanics:

THEOREM 3.23. *The orthogonal polynomials for $L^2[0,\infty)$, with scalar product*

$$< f, g >= \int_0^\infty f(x)g(x)e^{-x}\, dx$$

are the Laguerre polynomials $\{P_k\}$, given by the following formula,

$$P_k(x) = \frac{e^x}{k!}\cdot\frac{d^k}{dx^k}\left(e^{-x}x^k\right)$$

called Rodrigues formula for the Laguerre polynomials.

PROOF. The story here is very similar to that of the Legendre polynomials, and many further things can be said here, with exercise for you to learn a bit about all this. □

Finally, regarding the space $X = \mathbb{R}$ itself, we have here the following result:

THEOREM 3.24. *The orthogonal polynomials for $L^2(\mathbb{R})$, with scalar product*

$$< f, g >= \int_0^\infty f(x)g(x)e^{-x^2}\, dx$$

are the Hermite polynomials $\{P_k\}$, given by the following formula,

$$P_k(x) = (-1)^k e^{x^2}\cdot\frac{d^k}{dx^k}\left(e^{-x^2}\right)$$

called Rodrigues formula for the Hermite polynomials.

PROOF. As before, the story here is quite similar to that of the Legendre and other orthogonal polynomials, and exercise for you to learn a bit about all this. □

And with this, good news, end of the story with the orthogonal polynomials, at least at the very introductory level, and this due to the following fact, which is something quite technical, and that we will not attempt to prove, or even explain in detail here:

FACT 3.25. *From an abstract point of view, coming from degree 2 equations, and Rodrigues formulae for the solutions, there are only three types of "classical" orthogonal polynomials, namely the Jacobi, Laguerre and Hermite ones, discussed above.*

Finally, as already mentioned, the above results are very useful in the context of basic quantum mechanics, and more specifically, for solving the hydrogen atom, following Heisenberg and Schrödinger. Again, exercise for you to learn a bit about this.

3e. Exercises

We had a lot of combinatorics in this chapter, and as exercises, we have:

EXERCISE 3.26. *Compute the cumulants of other known measures.*

EXERCISE 3.27. *Learn more about Möbius functions, and their applications.*

EXERCISE 3.28. *Clarify what happens for Dirac masses, in Stieltjes inversion.*

EXERCISE 3.29. *Clarify what we said above, regarding Hankel determinants.*

EXERCISE 3.30. *Do the Stieltjes inversion for the normal laws.*

EXERCISE 3.31. *Do the Stieltjes inversion for other known laws.*

EXERCISE 3.32. *Learn more about orthogonal polynomials, and their properties.*

EXERCISE 3.33. *In particular, get to know everything about Chebycheff polyomials.*

As bonus exercise, read more about Hilbert spaces, and about operators too.

CHAPTER 4

Geometric aspects

4a. Spheres, Wallis

Let us work out now the general spherical coordinate formula, in arbitrary N dimensions. The formula here, which generalizes those at $N=2,3$, is as follows:

THEOREM 4.1. *We have spherical coordinates in N dimensions,*

$$\begin{cases} x_1 &= r\cos t_1 \\ x_2 &= r\sin t_1\cos t_2 \\ \vdots \\ x_{N-1} &= r\sin t_1\sin t_2\ldots\sin t_{N-2}\cos t_{N-1} \\ x_N &= r\sin t_1\sin t_2\ldots\sin t_{N-2}\sin t_{N-1} \end{cases}$$

the corresponding Jacobian being given by the following formula,

$$J(r,t)=r^{N-1}\sin^{N-2}t_1\sin^{N-3}t_2\,\ldots\,\sin^2 t_{N-3}\sin t_{N-2}$$

and with this generalizing the known formulae at $N=2,3$.

PROOF. As before, the fact that we have spherical coordinates is clear. Regarding now the Jacobian, also as before, by developing over the last column, we have:

$$\begin{aligned} J_N &= r\sin t_1\ldots\sin t_{N-2}\sin t_{N-1}\times\sin t_{N-1}J_{N-1} \\ &+ r\sin t_1\ldots\sin t_{N-2}\cos t_{N-1}\times\cos t_{N-1}J_{N-1} \\ &= r\sin t_1\ldots\sin t_{N-2}(\sin^2 t_{N-1}+\cos^2 t_{N-1})J_{N-1} \\ &= r\sin t_1\ldots\sin t_{N-2}J_{N-1} \end{aligned}$$

Thus, we obtain the formula in the statement, by recurrence. □

As a comment here, the above convention for spherical coordinates is one among many, designed to best work in arbitrary N dimensions. Also, in what regards the precise range of the angles $t_1,\ldots,t_{N-1}$, we will leave this to you, as an instructive exercise.

As an application, let us compute the volumes of spheres. For this purpose, we must understand how the products of coordinates integrate over spheres. Let us start with the case $N=2$. Here the sphere is the unit circle $\mathbb{T}$, and with $z=e^{it}$ the coordinates are $\cos t,\sin t$. We can first integrate arbitrary powers of these coordinates, as follows:

THEOREM 4.2 (Wallis). *We have the following formulae,*

$$\int_0^{\pi/2}\cos^p t\,dt=\int_0^{\pi/2}\sin^p t\,dt=\left(\frac{\pi}{2}\right)^{\varepsilon(p)}\frac{p!!}{(p+1)!!}$$

where $\varepsilon(p)=1$ if p is even, and $\varepsilon(p)=0$ if p is odd, and where

$$m!!=(m-1)(m-3)(m-5)\ldots$$

with the product ending at 2 if m is odd, and ending at 1 if m is even.

PROOF. Let us first compute the integral on the left in the statement:

$$I_p=\int_0^{\pi/2}\cos^p t\,dt$$

We do this by partial integration. We have the following formula:

$$\begin{aligned}(\cos^p t\sin t)' &= p\cos^{p-1}t(-\sin t)\sin t+\cos^p t\cos t\\ &= p\cos^{p+1}t-p\cos^{p-1}t+\cos^{p+1}t\\ &= (p+1)\cos^{p+1}t-p\cos^{p-1}t\end{aligned}$$

By integrating between 0 and $\pi/2$, we obtain the following formula:

$$(p+1)I_{p+1}=pI_{p-1}$$

Thus we can compute I_p by recurrence, and we obtain:

$$\begin{aligned}I_p &= \frac{p-1}{p}\,I_{p-2}\\ &= \frac{p-1}{p}\cdot\frac{p-3}{p-2}\,I_{p-4}\\ &= \frac{p-1}{p}\cdot\frac{p-3}{p-2}\cdot\frac{p-5}{p-4}\,I_{p-6}\\ &\ \ \vdots\\ &= \frac{p!!}{(p+1)!!}\,I_{1-\varepsilon(p)}\end{aligned}$$

But $I_0=\frac{\pi}{2}$ and $I_1=1$, so we get the result. As for the second formula, this follows from the first one, with $t=\frac{\pi}{2}-s$. Thus, we have proved both formulae in the statement. □

We can now compute the volume of the sphere, as follows:

THEOREM 4.3. *The volume of the unit sphere in $\mathbb{R}^N$ is given by*

$$V=\left(\frac{\pi}{2}\right)^{[N/2]}\frac{2^N}{(N+1)!!}$$

with our usual convention $N!!=(N-1)(N-3)(N-5)\ldots$

PROOF. Let us denote by B^+ the positive part of the unit sphere, or rather unit ball B, obtained by cutting this unit ball in 2^N parts. At the level of volumes, we have:

$$V = 2^N V^+$$

We have the following computation, using spherical coordinates:

$$\begin{aligned}
V^+ &= \int_{B^+} 1 \\
&= \int_0^1 \int_0^{\pi/2} \ldots \int_0^{\pi/2} r^{N-1} \sin^{N-2} t_1 \ldots \sin t_{N-2}\, dr dt_1 \ldots dt_{N-1} \\
&= \int_0^1 r^{N-1}\, dr \int_0^{\pi/2} \sin^{N-2} t_1\, dt_1 \ldots \int_0^{\pi/2} \sin t_{N-2} dt_{N-2} \int_0^{\pi/2} 1 dt_{N-1} \\
&= \frac{1}{N} \times \left(\frac{\pi}{2}\right)^{[N/2]} \times \frac{(N-2)!!}{(N-1)!!} \cdot \frac{(N-3)!!}{(N-2)!!} \ldots \frac{2!!}{3!!} \cdot \frac{1!!}{2!!} \cdot 1 \\
&= \frac{1}{N} \times \left(\frac{\pi}{2}\right)^{[N/2]} \times \frac{1}{(N-1)!!} \\
&= \left(\frac{\pi}{2}\right)^{[N/2]} \frac{1}{(N+1)!!}
\end{aligned}$$

Here we have used the following formula, for computing the exponent of $\pi/2$:

$$\begin{aligned}
\varepsilon(0) + \varepsilon(1) + \varepsilon(2) + \ldots + \varepsilon(N-2) &= 1 + 0 + 1 + \ldots + \varepsilon(N-2) \\
&= \left[\frac{N-2}{2}\right] + 1 \\
&= \left[\frac{N}{2}\right]
\end{aligned}$$

Thus, we are led to the formula in the statement. □

As main particular cases of the above formula, we have:

THEOREM 4.4. *The volumes of the low-dimensional spheres are as follows:*

(1) *At $N = 1$, the length of the unit interval is $V = 2$.*
(2) *At $N = 2$, the area of the unit disk is $V = \pi$.*
(3) *At $N = 3$, the volume of the unit sphere is $V = \frac{4\pi}{3}$*
(4) *At $N = 4$, the volume of the corresponding unit sphere is $V = \frac{\pi^2}{2}$.*

PROOF. Some of these results are well-known, but we can obtain all of them as particular cases of the general formula in Theorem 4.3, as follows:

(1) At $N = 1$ we obtain $V = 1 \cdot \frac{2}{1} = 2$.

(2) At $N = 2$ we obtain $V = \frac{\pi}{2} \cdot \frac{4}{2} = \pi$.

(3) At $N=3$ we obtain $V=\frac{\pi}{2}\cdot\frac{8}{3}=\frac{4\pi}{3}$.

(4) At $N=4$ we obtain $V=\frac{\pi^2}{4}\cdot\frac{16}{8}=\frac{\pi^2}{2}$. □

We can compute in the same way the area of the sphere, the result being:

THEOREM 4.5. *The area of the unit sphere in $\mathbb{R}^N$ is given by*

$$A=\left(\frac{\pi}{2}\right)^{[N/2]}\frac{2^N}{(N-1)!!}$$

with the our usual convention for double factorials, namely:

$$N!!=(N-1)(N-3)(N-5)\ldots$$

In particular, at $N=2,3,4$ we obtain respectively $A=2\pi,4\pi,2\pi^2$.

PROOF. Regarding the first assertion, there is no need to compute again, because the formula in the statement can be deduced from Theorem 4.3, as follows:

(1) We can either use a standard "pizza" argument, as in 1 dimension, which shows that the area and volume of the sphere in $\mathbb{R}^N$ are related by the following formula:

$$A=N\cdot V$$

Together with the formula in Theorem 4.3 for V, this gives the result.

(2) Or, we can start the computation in the same way as we started the proof of Theorem 4.3, the beginning of this computation being as follows:

$$vol(S^+)=\int_0^{\pi/2}\ldots\int_0^{\pi/2}\sin^{N-2}t_1\ldots\sin t_{N-2}\,dt_1\ldots dt_{N-1}$$

Now by comparing with the beginning of the proof of Theorem 4.3, the only thing that changes is the following quantity, which now dissapears:

$$\int_0^1 r^{N-1}\,dr=\frac{1}{N}$$

Thus, we have $vol(S^+)=N\cdot vol(B^+)$, and so we obtain the following formula:

$$vol(S)=N\cdot vol(B)$$

But this means $A=N\cdot V$, and together with the formula in Theorem 4.3 for V, this gives the result. As for the last assertion, this can be either worked out directly, or deduced from the results for volumes that we have so far, by multiplying by N. □

Let us record as well the asymptotics, obtained via Stirling, as follows:

THEOREM 4.6. *The volume of the unit sphere in $\mathbb{R}^N$ is given by*

$$V \simeq \left(\frac{2\pi e}{N}\right)^{N/2} \frac{1}{\sqrt{\pi N}}$$

in the $N \to \infty$ limit. As for the area, this is $A = N \cdot V$.

PROOF. This is something very standard, the idea being as follows:

(1) We use the exact formula found in Theorem 4.3, namely:

$$V = \left(\frac{\pi}{2}\right)^{[N/2]} \frac{2^N}{(N+1)!!}$$

(2) But the double factorials can be estimated by using the Stirling formula. Indeed, in the case where $N = 2K$ is even, we have the following computation:

$$\begin{aligned}
(N+1)!! &= 2^K K! \\
&\simeq \left(\frac{2K}{e}\right)^K \sqrt{2\pi K} \\
&= \left(\frac{N}{e}\right)^{N/2} \sqrt{\pi N}
\end{aligned}$$

As for the case where $N = 2K - 1$ is odd, here the estimate goes as follows:

$$\begin{aligned}
(N+1)!! &= \frac{(2K)!}{2^K K!} \\
&\simeq \frac{1}{2^K}\left(\frac{2K}{e}\right)^{2K} \sqrt{4\pi K} \left(\frac{e}{K}\right)^K \frac{1}{\sqrt{2\pi K}} \\
&= \left(\frac{2K}{e}\right)^K \sqrt{2} \\
&= \left(\frac{N+1}{e}\right)^{(N+1)/2} \sqrt{2} \\
&= \left(\frac{N}{e}\right)^{N/2} \left(\frac{N+1}{N}\right)^{N/2} \sqrt{\frac{N+1}{e}} \cdot \sqrt{2} \\
&\simeq \left(\frac{N}{e}\right)^{N/2} \sqrt{e} \cdot \sqrt{\frac{N}{e}} \cdot \sqrt{2} \\
&= \left(\frac{N}{e}\right)^{N/2} \sqrt{2N}
\end{aligned}$$

(3) Now back to the spheres, when N is even, the estimate goes as follows:

$$\begin{aligned} V &= \left(\frac{\pi}{2}\right)^{N/2} \frac{2^N}{(N+1)!!} \\ &\simeq \left(\frac{\pi}{2}\right)^{N/2} 2^N \left(\frac{e}{N}\right)^{N/2} \frac{1}{\sqrt{\pi N}} \\ &= \left(\frac{2\pi e}{N}\right)^{N/2} \frac{1}{\sqrt{\pi N}} \end{aligned}$$

As for the case where N is odd, here the estimate goes as follows:

$$\begin{aligned} V &= \left(\frac{\pi}{2}\right)^{(N-1)/2} \frac{2^N}{(N+1)!!} \\ &\simeq \left(\frac{\pi}{2}\right)^{(N-1)/2} 2^N \left(\frac{e}{N}\right)^{N/2} \frac{1}{\sqrt{2N}} \\ &= \sqrt{\frac{2}{\pi}} \left(\frac{2\pi e}{N}\right)^{N/2} \frac{1}{\sqrt{2N}} \\ &= \left(\frac{2\pi e}{N}\right)^{N/2} \frac{1}{\sqrt{\pi N}} \end{aligned}$$

Thus, we are led to the uniform formula in the statement. □

4b. Spherical integrals

Let us discuss now the computation of arbitrary integrals over the sphere. We will need an extension of the previous Wallis formula, from Theorem 4.2, as follows:

THEOREM 4.7 (Wallis). *We have the following formula,*

$$\int_0^{\pi/2} \cos^p t \sin^q t \, dt = \left(\frac{\pi}{2}\right)^{\varepsilon(p)\varepsilon(q)} \frac{p!!q!!}{(p+q+1)!!}$$

where $\varepsilon(p) = 1$ if p is even, and $\varepsilon(p) = 0$ if p is odd, and where

$$m!! = (m-1)(m-3)(m-5)\ldots$$

with the product ending at 2 if m is odd, and ending at 1 if m is even.

PROOF. We use the same idea as in the proof of Theorem 4.2. Let I_{pq} be the integral in the statement. In order to do the partial integration, observe that we have:

$$\begin{aligned} (\cos^p t \sin^q t)' &= p \cos^{p-1} t(-\sin t) \sin^q t \\ &+ \cos^p t \cdot q \sin^{q-1} t \cos t \\ &= -p \cos^{p-1} t \sin^{q+1} t + q \cos^{p+1} t \sin^{q-1} t \end{aligned}$$

By integrating between 0 and $\pi/2$, we obtain, for $p, q > 0$:

$$pI_{p-1,q+1} = qI_{p+1,q-1}$$

Thus, we can compute I_{pq} by recurrence. When q is even we have:

$$\begin{aligned} I_{pq} &= \frac{q-1}{p+1} I_{p+2,q-2} \\ &= \frac{q-1}{p+1} \cdot \frac{q-3}{p+3} I_{p+4,q-4} \\ &= \frac{q-1}{p+1} \cdot \frac{q-3}{p+3} \cdot \frac{q-5}{p+5} I_{p+6,q-6} \\ &= \vdots \\ &= \frac{p!!q!!}{(p+q)!!} I_{p+q} \end{aligned}$$

But the last term comes from Theorem 4.2, and we obtain the result:

$$\begin{aligned} I_{pq} &= \frac{p!!q!!}{(p+q)!!} I_{p+q} \\ &= \frac{p!!q!!}{(p+q)!!} \left(\frac{\pi}{2}\right)^{\varepsilon(p+q)} \frac{(p+q)!!}{(p+q+1)!!} \\ &= \left(\frac{\pi}{2}\right)^{\varepsilon(p)\varepsilon(q)} \frac{p!!q!!}{(p+q+1)!!} \end{aligned}$$

Observe that this gives the result for p even as well, by symmetry. Indeed, we have $I_{pq} = I_{qp}$, by using the following change of variables:

$$t = \frac{\pi}{2} - s$$

In the remaining case now, where both p, q are odd, we can use once again the formula $pI_{p-1,q+1} = qI_{p+1,q-1}$ established above, and the recurrence goes as follows:

$$\begin{aligned} I_{pq} &= \frac{q-1}{p+1} I_{p+2,q-2} \\ &= \frac{q-1}{p+1} \cdot \frac{q-3}{p+3} I_{p+4,q-4} \\ &= \frac{q-1}{p+1} \cdot \frac{q-3}{p+3} \cdot \frac{q-5}{p+5} I_{p+6,q-6} \\ &= \vdots \\ &= \frac{p!!q!!}{(p+q-1)!!} I_{p+q-1,1} \end{aligned}$$

In order to compute the last term, observe that we have:

$$\begin{aligned}
I_{p1} &= \int_0^{\pi/2} \cos^p t \sin t \, dt \\
&= -\frac{1}{p+1}\int_0^{\pi/2} (\cos^{p+1} t)' \, dt \\
&= \frac{1}{p+1}
\end{aligned}$$

Thus, we can finish our computation in the case p,q odd, as follows:

$$\begin{aligned}
I_{pq} &= \frac{p!!q!!}{(p+q-1)!!} I_{p+q-1,1} \\
&= \frac{p!!q!!}{(p+q-1)!!}\cdot\frac{1}{p+q} \\
&= \frac{p!!q!!}{(p+q+1)!!}
\end{aligned}$$

Thus, we obtain the formula in the statement, the exponent of $\pi/2$ appearing there being $\varepsilon(p)\varepsilon(q) = 0\cdot 0 = 0$ in the present case, and this finishes the proof. □

We can now integrate over the spheres, as follows:

THEOREM 4.8. *The polynomial integrals over the unit sphere $S^{N-1}_{\mathbb{R}} \subset \mathbb{R}^N$, with respect to the normalized, mass 1 measure, are given by the following formula,*

$$\int_{S^{N-1}_{\mathbb{R}}} x_1^{k_1}\ldots x_N^{k_N} \, dx = \frac{(N-1)!!k_1!!\ldots k_N!!}{(N+\Sigma k_i - 1)!!}$$

valid when all exponents k_i are even. If an exponent k_i is odd, the integral vanishes.

PROOF. Assume first that one of the exponents k_i is odd. We can make then the following change of variables, which shows that the integral in the statement vanishes:

$$x_i \to -x_i$$

Assume now that all exponents k_i are even. As a first observation, the result holds indeed at $N=2$, due to the formula from Theorem 4.7, which reads:

$$\int_0^{\pi/2} \cos^p t \sin^q t \, dt = \left(\frac{\pi}{2}\right)^{\varepsilon(p)\varepsilon(q)} \frac{p!!q!!}{(p+q+1)!!} = \frac{p!!q!!}{(p+q+1)!!}$$

In the general case now, where the dimension $N \in \mathbb{N}$ is arbitrary, the integral in the statement can be written in spherical coordinates, as follows:

$$I = \frac{2^N}{A}\int_0^{\pi/2}\ldots\int_0^{\pi/2} x_1^{k_1}\ldots x_N^{k_N} J \, dt_1\ldots dt_{N-1}$$

Here A is the area of the sphere, J is the Jacobian, and the 2^N factor comes from the restriction to the $1/2^N$ part of the sphere where all the coordinates are positive. According to Theorem 4.5, the normalization constant in front of the integral is:

$$\frac{2^N}{A}=\left(\frac{2}{\pi}\right)^{[N/2]}(N-1)!!$$

As for the unnormalized integral, this is given by:

$$\begin{aligned}I'=\int_0^{\pi/2}\ldots\int_0^{\pi/2}\quad&(\cos t_1)^{k_1}(\sin t_1\cos t_2)^{k_2}\\&\vdots\\&(\sin t_1\sin t_2\ldots\sin t_{N-2}\cos t_{N-1})^{k_{N-1}}\\&(\sin t_1\sin t_2\ldots\sin t_{N-2}\sin t_{N-1})^{k_N}\\&\sin^{N-2}t_1\sin^{N-3}t_2\ldots\sin^2 t_{N-3}\sin t_{N-2}\\&dt_1\ldots dt_{N-1}\end{aligned}$$

By rearranging the terms, we obtain:

$$\begin{aligned}I'\;&=\;\int_0^{\pi/2}\cos^{k_1}t_1\sin^{k_2+\ldots+k_N+N-2}t_1\,dt_1\\&\quad\int_0^{\pi/2}\cos^{k_2}t_2\sin^{k_3+\ldots+k_N+N-3}t_2\,dt_2\\&\quad\vdots\\&\quad\int_0^{\pi/2}\cos^{k_{N-2}}t_{N-2}\sin^{k_{N-1}+k_N+1}t_{N-2}\,dt_{N-2}\\&\quad\int_0^{\pi/2}\cos^{k_{N-1}}t_{N-1}\sin^{k_N}t_{N-1}\,dt_{N-1}\end{aligned}$$

Now by using the above-mentioned formula at $N=2$, this gives:

$$\begin{aligned}I'\;&=\;\frac{k_1!!(k_2+\ldots+k_N+N-2)!!}{(k_1+\ldots+k_N+N-1)!!}\left(\frac{\pi}{2}\right)^{\varepsilon(N-2)}\\&\quad\frac{k_2!!(k_3+\ldots+k_N+N-3)!!}{(k_2+\ldots+k_N+N-2)!!}\left(\frac{\pi}{2}\right)^{\varepsilon(N-3)}\\&\quad\vdots\\&\quad\frac{k_{N-2}!!(k_{N-1}+k_N+1)!!}{(k_{N-2}+k_{N-1}+l_N+2)!!}\left(\frac{\pi}{2}\right)^{\varepsilon(1)}\\&\quad\frac{k_{N-1}!!k_N!!}{(k_{N-1}+k_N+1)!!}\left(\frac{\pi}{2}\right)^{\varepsilon(0)}\end{aligned}$$

Now let F be the part involving the double factorials, and P be the part involving the powers of $\pi/2$, so that $I' = F \cdot P$. Regarding F, by cancelling terms we have:

$$F = \frac{k_1!! \ldots k_N!!}{(\Sigma k_i + N - 1)!!}$$

As in what regards P, by summing the exponents, we obtain $P = \left(\frac{\pi}{2}\right)^{[N/2]}$. We can now put everything together, and we obtain:

$$\begin{aligned} I &= \frac{2^N}{A} \times F \times P \\ &= \left(\frac{2}{\pi}\right)^{[N/2]} (N-1)!! \times \frac{k_1!! \ldots k_N!!}{(\Sigma k_i + N - 1)!!} \times \left(\frac{\pi}{2}\right)^{[N/2]} \\ &= \frac{(N-1)!! k_1!! \ldots k_N!!}{(\Sigma k_i + N - 1)!!} \end{aligned}$$

Thus, we are led to the conclusion in the statement. □

Let us record as well the following useful version of the above formula:

THEOREM 4.9. *We have the following integration formula over $S^{N-1}_{\mathbb{R}} \subset \mathbb{R}^N$, with respect to the normalized, mass 1 measure, valid for any exponents $k_i \in \mathbb{N}$,*

$$\int_{S^{N-1}_{\mathbb{R}}} |x_1^{k_1} \ldots x_N^{k_N}| \, dx = \left(\frac{2}{\pi}\right)^{\Sigma(k_1,\ldots,k_N)} \frac{(N-1)!! k_1!! \ldots k_N!!}{(N + \Sigma k_i - 1)!!}$$

with $\Sigma = [odds/2]$ if N is odd and $\Sigma = [(odds+1)/2]$ if N is even, where "odds" denotes the number of odd numbers in the sequence $k_1, \ldots, k_N$.

PROOF. As before, the formula holds at $N = 2$, due to Theorem 4.7. In general, the integral in the statement can be written in spherical coordinates, as follows:

$$I = \frac{2^N}{A} \int_0^{\pi/2} \ldots \int_0^{\pi/2} x_1^{k_1} \ldots x_N^{k_N} \, J \, dt_1 \ldots dt_{N-1}$$

Here A is the area of the sphere, J is the Jacobian, and the 2^N factor comes from the restriction to the $1/2^N$ part of the sphere where all the coordinates are positive. The normalization constant in front of the integral is, as before:

$$\frac{2^N}{A} = \left(\frac{2}{\pi}\right)^{[N/2]} (N-1)!!$$

As for the unnormalized integral, this can be written as before, as follows:

$$\begin{aligned} I' &= \int_0^{\pi/2}\cos^{k_1}t_1\sin^{k_2+\ldots+k_N+N-2}t_1\,dt_1\\ &\quad\int_0^{\pi/2}\cos^{k_2}t_2\sin^{k_3+\ldots+k_N+N-3}t_2\,dt_2\\ &\quad\vdots\\ &\quad\int_0^{\pi/2}\cos^{k_{N-2}}t_{N-2}\sin^{k_{N-1}+k_N+1}t_{N-2}\,dt_{N-2}\\ &\quad\int_0^{\pi/2}\cos^{k_{N-1}}t_{N-1}\sin^{k_N}t_{N-1}\,dt_{N-1}\end{aligned}$$

Now by using the formula at $N=2$, we get:

$$\begin{aligned} I' &= \frac{\pi}{2}\cdot\frac{k_1!!(k_2+\ldots+k_N+N-2)!!}{(k_1+\ldots+k_N+N-1)!!}\left(\frac{2}{\pi}\right)^{\delta(k_1,k_2+\ldots+k_N+N-2)}\\ &\quad\frac{\pi}{2}\cdot\frac{k_2!!(k_3+\ldots+k_N+N-3)!!}{(k_2+\ldots+k_N+N-2)!!}\left(\frac{2}{\pi}\right)^{\delta(k_2,k_3+\ldots+k_N+N-3)}\\ &\quad\vdots\\ &\quad\frac{\pi}{2}\cdot\frac{k_{N-2}!!(k_{N-1}+k_N+1)!!}{(k_{N-2}+k_{N-1}+k_N+2)!!}\left(\frac{2}{\pi}\right)^{\delta(k_{N-2},k_{N-1}+k_N+1)}\\ &\quad\frac{\pi}{2}\cdot\frac{k_{N-1}!!k_N!!}{(k_{N-1}+k_N+1)!!}\left(\frac{2}{\pi}\right)^{\delta(k_{N-1},k_N)}\end{aligned}$$

In order to compute this quantity, let us denote by F the part involving the double factorials, and by P the part involving the powers of $\pi/2$, so that we have:

$$I' = F\cdot P$$

Regarding F, there are many cancellations there, and we end up with:

$$F = \frac{k_1!!\ldots k_N!!}{(\Sigma k_i+N-1)!!}$$

As in what regards P, the δ exponents on the right sum up to the following number:

$$\Delta(k_1,\ldots,k_N) = \sum_{i=1}^{N-1}\delta(k_i,k_{i+1}+\ldots+k_N+N-i-1)$$

In other words, with this notation, the above formula reads:

$$\begin{aligned} I' &= \left(\frac{\pi}{2}\right)^{N-1} \frac{k_1!!k_2!!\ldots k_N!!}{(k_1+\ldots+k_N+N-1)!!}\left(\frac{2}{\pi}\right)^{\Delta(k_1,\ldots,k_N)} \\ &= \left(\frac{2}{\pi}\right)^{\Delta(k_1,\ldots,k_N)-N+1} \frac{k_1!!k_2!!\ldots k_N!!}{(k_1+\ldots+k_N+N-1)!!} \\ &= \left(\frac{2}{\pi}\right)^{\Sigma(k_1,\ldots,k_N)-[N/2]} \frac{k_1!!k_2!!\ldots k_N!!}{(k_1+\ldots+k_N+N-1)!!} \end{aligned}$$

To be more precise, here the formula relating Δ to Σ follows from a number of simple observations, the first of which being the fact that, due to obvious parity reasons, the sequence of δ numbers appearing in the definition of Δ cannot contain two consecutive zeroes. Together with $I = (2^N/V)I'$, this gives the formula in the statement. □

4c. Hyperspherical laws

We can go back now to probability, and we have the following result:

THEOREM 4.10. *The moments of the hyperspherical variables are*

$$\int_{S^{N-1}_\mathbb{R}} x_i^p dx = \frac{(N-1)!!p!!}{(N+p-1)!!}$$

and the rescaled variables $y_i = \sqrt{N}x_i$ become normal and independent with $N \to \infty$.

PROOF. The moment formula in the statement follows from the general formula from Theorem 4.8. As a consequence, with $N \to \infty$ we have the following estimate:

$$\begin{aligned} \int_{S^{N-1}_\mathbb{R}} x_i^p dx &\simeq N^{-p/2} \times p!! \\ &= N^{-p/2} M_p(g_1) \end{aligned}$$

Thus, the rescaled variables $\sqrt{N}x_i$ become normal with $N \to \infty$, as claimed. As for the proof of the asymptotic independence, this is standard too, once again by using the formula in Theorem 4.8. Indeed, the joint moments of $x_1, \ldots, x_N$ are given by:

$$\begin{aligned} \int_{S^{N-1}_\mathbb{R}} x_1^{k_1} \ldots x_N^{k_N}\, dx &= \frac{(N-1)!!k_1!!\ldots k_N!!}{(N+\Sigma k_i-1)!!} \\ &\simeq N^{-\Sigma k_i} \times k_1!!\ldots k_N!! \end{aligned}$$

By rescaling, the joint moments of the variables $y_i = \sqrt{N}x_i$ are given by:

$$\int_{S^{N-1}_\mathbb{R}} y_1^{k_1} \ldots y_N^{k_N}\, dx \simeq k_1!!\ldots k_N!!$$

Thus, we have multiplicativity, and so independence with $N \to \infty$, as claimed. □

4d. Rotation groups

Importantly, we can recover the normal laws as well in connection with the rotation groups. Indeed, we have the following reformulation of Theorem 4.10:

THEOREM 4.11. *We have the integration formula*

$$\int_{O_N} U_{ij}^p\,dU = \frac{(N-1)!!p!!}{(N+p-1)!!}$$

and the rescaled variables $V_{ij} = \sqrt{N}U_{ij}$ become normal and independent with $N \to \infty$.

PROOF. We use the basic fact that the rotations $U \in O_N$ act on the points of the real sphere $z \in S^{N-1}_{\mathbb{R}}$, with the stabilizer of $z = (1, 0, \ldots, 0)$ being the subgroup $O_{N-1} \subset O_N$. In algebraic terms, this gives an identification as follows:

$$S^{N-1}_{\mathbb{R}} = O_N/O_{N-1}$$

In functional analytic terms, this result provides us with an embedding as follows, for any i, which makes correspond the respective integration functionals:

$$C(S^{N-1}_{\mathbb{R}}) \subset C(O_N) \quad , \quad x_i \to U_{1i}$$

With this identification made, the result follows from Theorem 4.10. □

In order to go beyond this, we will need an advanced result, as follows:

THEOREM 4.12. *The Haar integration over a closed subgroup $G \subset_v U_N$ is given on the dense subalgebra of smooth functions by the Weingarten type formula*

$$\int_G g_{i_1j_1}^{e_1}\ldots g_{i_kj_k}^{e_k}\,dg = \sum_{\pi,\nu\in D(k)} \delta_\pi(i)\delta_\sigma(j)W_k(\pi,\nu)$$

valid for any colored integer $k = e_1 \ldots e_k$ and any multi-indices i, j, where $D(k)$ is a linear basis of $Fix(v^{\otimes k})$, the associated generalized Kronecker symbols are given by

$$\delta_\pi(i) = <\pi, e_{i_1} \otimes \ldots \otimes e_{i_k}>$$

and $W_k = G_k^{-1}$ is the inverse of the Gram matrix, $G_k(\pi,\nu) = <\pi,\nu>$.

PROOF. This is something very standard, coming from the fact that the above integrals form altogether the orthogonal projection P^k onto the following space:

$$Fix(v^{\otimes k}) = span(D(k))$$

Consider now the following linear map, with $D(k) = \{\xi_k\}$ being as in the statement:

$$E(x) = \sum_{\pi\in D(k)} <x, \xi_\pi> \xi_\pi$$

By a standard linear algebra computation, it follows that we have $P = WE$, where W is the inverse of the restriction of E to the following space:

$$K = span\left(T_\pi\Big|\pi \in D(k)\right)$$

But this restriction is the linear map given by the matrix G_k, and so W is the linear map given by the inverse matrix $W_k = G_k^{-1}$, and this gives the result. □

In the easy case, we have the following more concrete result:

THEOREM 4.13. *For an easy group $G \subset U_N$, coming from a category of partitions $D = (D(k,l))$, we have the Weingarten formula*

$$\int_G g_{i_1 j_1}^{e_1} \ldots g_{i_k j_k}^{e_k}\, dg = \sum_{\pi,\nu \in D(k)} \delta_\pi(i)\delta_\nu(j) W_{kN}(\pi,\nu)$$

for any $k = e_1 \ldots e_k$ and any i,j, where $D(k) = D(\emptyset, k)$, δ are usual Kronecker type symbols, checking whether the indices match, and $W_{kN} = G_{kN}^{-1}$, with

$$G_{kN}(\pi,\nu) = N^{|\pi \vee \nu|}$$

where $|.|$ is the number of blocks.

PROOF. We use the abstract Weingarten formula, from Theorem 4.12. Indeed, the Kronecker type symbols there are then the usual ones, as shown by:

$$\begin{aligned}
\delta_{\xi_\pi}(i) &= < \xi_\pi, e_{i_1} \otimes \ldots \otimes e_{i_k} > \\
&= \left\langle \sum_j \delta_\pi(j_1,\ldots,j_k) e_{j_1} \otimes \ldots \otimes e_{j_k}, e_{i_1} \otimes \ldots \otimes e_{i_k} \right\rangle \\
&= \delta_\pi(i_1,\ldots,i_k)
\end{aligned}$$

The Gram matrix being as well the correct one, we obtain the result. □

Let us go back now to the general easy groups $G \subset U_N$, with the idea in mind of computing the laws of truncated characters. First, we have the following formula:

PROPOSITION 4.14. *The moments of truncated characters are given by the formula*

$$\int_G (g_{11} + \ldots + g_{ss})^k dg = Tr(W_{kN} G_{ks})$$

where G_{kN} and $W_{kN} = G_{kN}^{-1}$ are the associated Gram and Weingarten matrices.

PROOF. We have indeed the following computation:

$$\begin{aligned}\int_G (g_{11}+\ldots+g_{ss})^k dg &= \sum_{i_1=1}^s \ldots \sum_{i_k=1}^s \int_G g_{i_1i_1}\ldots g_{i_ki_k}\,dg\\ &= \sum_{\pi,\nu\in D(k)} W_{kN}(\pi,\nu)\sum_{i_1=1}^s\ldots\sum_{i_k=1}^s \delta_\pi(i)\delta_\nu(i)\\ &= \sum_{\pi,\nu\in D(k)} W_{kN}(\pi,\nu)G_{ks}(\nu,\pi)\\ &= Tr(W_{kN}G_{ks})\end{aligned}$$

Thus, we have reached to the formula in the statement. □

In order to process now the above formula, and reach to concrete results, we must impose on our group a uniformity condition. Let us start with:

PROPOSITION 4.15. *For an easy group $G=(G_N)$, coming from a category of partitions $D\subset P$, the following conditions are equivalent:*

1. *$G_{N-1}=G_N\cap U_{N-1}$, via the embedding $U_{N-1}\subset U_N$ given by $u\to diag(u,1)$.*
2. *$G_{N-1}=G_N\cap U_{N-1}$, via the N possible diagonal embeddings $U_{N-1}\subset U_N$.*
3. *D is stable under the operation which consists in removing blocks.*

If these conditions are satisfied, we say that $G=(G_N)$ is uniform.

PROOF. The equivalence $(1)\iff(2)$ comes from the inclusion $S_N\subset G_N$, which makes everything S_N-invariant. As for $(1)\iff(3)$, this is something standard too. □

Now back to the laws of truncated characters, we have the following result:

THEOREM 4.16. *For a uniform easy group $G=(G_N)$, we have the formula*

$$\lim_{N\to\infty}\int_{G_N}\chi_t^k=\sum_{\pi\in D(k)} t^{|\pi|}$$

with $D\subset P$ being the associated category of partitions.

PROOF. We use Proposition 4.14. With $s=[tN]$, the formula there becomes:

$$\int_{G_N}\chi_t^k=Tr(W_{kN}G_{k[tN]})$$

The point now is that in the uniform case the Gram matrix, and so the Weingarten matrix too, is asymptotically diagonal. Thus, we obtain the following estimate:

$$\begin{aligned}\int_{G_N}\chi_t^k &\simeq \sum_{\pi\in D(k)} W_{kN}(\pi,\pi)G_{k[tN]}(\pi,\pi)\\ &\simeq \sum_{\pi\in D(k)} N^{-|\pi|}(tN)^{|\pi|}\\ &= \sum_{\pi\in D(k)} t^{|\pi|}\end{aligned}$$

Thus, we are led to the formula in the statement. □

We can now enlarge our collection of truncated character results, and we have:

THEOREM 4.17. *With $N\to\infty$, the laws of truncated characters are as follows:*

(1) *For O_N we obtain the Gaussian law g_t.*
(2) *For S_N we obtain the Poisson law p_t.*
(3) *For H_N we obtain the Bessel law b_t.*

PROOF. We already know these results at $t=1$. In the general case, $t>0$, these follow via some standard combinatorics, from the formula in Theorem 4.16. □

4e. Exercises

This was a quite exciting geometric chapter, and as exercises, we have:

EXERCISE 4.18. *Clarify the range of angles in the spherical coordinate formula.*

EXERCISE 4.19. *Memorize what comes from the first Wallis formula, at small p.*

EXERCISE 4.20. *Compute the volume of the unit sphere, by some other means.*

EXERCISE 4.21. *Learn if needed the proof of the Stirling formula, with full details.*

EXERCISE 4.22. *Talking double factorials, learn about the gamma function too.*

EXERCISE 4.23. *Memorize what comes from the Wallis 2 formula, at small p,q.*

EXERCISE 4.24. *Compute the hyperspherical laws, at small values of N.*

EXERCISE 4.25. *Learn in detail the Weingarten formula, and its applications.*

As bonus exercise, read a bit about analysis on manifolds, and about Lie groups too.

Part II

Complex variables

CHAPTER 5

Complex variables

5a. Complex CLT

We have seen so far a number of interesting results regarding the normal laws, and their geometric interpretation. As a main topic for this present Part II, let us discuss now the complex analogues of all this. To start with, we have the following definition:

Definition 5.1. *A complex random variable is a variable $f:X\to\mathbb{C}$. In the discrete case, the law of such a variable is the complex probability measure*

$$\mu=\sum_i\alpha_i\delta_{z_i}\quad,\quad \alpha_i\geq 0\quad,\quad \sum_i\alpha_i=1\quad,\quad z_i\in\mathbb{C}$$

given by the following formula, with P being the probability over X,

$$\mu=\sum_{z\in\mathbb{C}}P(f=z)\delta_z$$

with the sum being finite or countable, as per our discretness assumption.

Observe the similarity with the analogous notions introduced in chapter 1, for the real variables $f:X\to\mathbb{R}$. In fact, what we are doing here is to extend the formalism from chapter 1, from real to complex, in a straightforward way. As a basic example for this, any real variable $f:X\to\mathbb{R}$ can be regarded as a complex variable $f:X\to\mathbb{C}$.

In order to understand the precise relation with the real theory, from chapter 1, we can decompose any complex variable $f:X\to\mathbb{C}$ as a sum, as follows:

$$f=g+ih\quad,\quad g=Re(f),\ h=Im(f)$$

With this done, we have the following computation, for the corresponding law:

$$\begin{aligned}\mu &= \sum_{z\in\mathbb{C}} P(f=z)\delta_z\\ &= \sum_{x,y\in\mathbb{R}} P(f=x+iy)\delta_{x+iy}\\ &= \sum_{x,y\in\mathbb{R}} P(g+ih=x+iy)\delta_{x+iy}\\ &= \sum_{x,y\in\mathbb{R}} P(g=x,h=y)\delta_{x+iy}\end{aligned}$$

In the case where the real and imaginary parts $g,h:X\to\mathbb{R}$ are independent, we can say more about this, with the above computation having the following continuation:

$$\begin{aligned}\mu &= \sum_{x,y\in\mathbb{R}} P(g=x,h=y)\delta_{x+iy}\\ &= \sum_{x,y\in\mathbb{R}} P(g=x)P(h=y)\delta_{x+iy}\\ &= \sum_{x,y\in\mathbb{R}} P(g=x)P(h=y)\delta_x*\delta_{iy}\\ &= \left(\sum_{x\in\mathbb{R}} P(g=x)\delta_x\right)*\left(\sum_{y\in\mathbb{R}} P(h=y)\delta_{iy}\right)\\ &= \mu_g*i\mu_h\end{aligned}$$

To be more precise, we have used here in the beginning the independence of the variables $h,g:X\to\mathbb{R}$, and at the end we have denoted the measure on the right, which is obtained from μ_h by putting this measure on the imaginary axis, by $i\mu_h$.

All this is quite interesting, going beyond what we know so far about basic probability, in the real case, so let us record this finding, along with a bit more, as follows:

THEOREM 5.2. *For a discrete complex random variable $f:X\to\mathbb{C}$, decomposed into real and imaginary parts as $f=g+ih$, and with g,h assumed independent, we have*

$$\mu_f=\mu_g*i\mu_h$$

with $*$ being the usual convolution operation, $\delta_z*\delta_t=\delta_{z+t}$, and with $\mu\to i\mu$ denoting the rotated version, $\mathbb{R}\to i\mathbb{R}$. If g,h are not independent, this formula does not hold.

PROOF. We already know that the first assertion holds, as explained in the above. As for the second assertion, this follows by carefully examining the above computation.

Indeed, we have used only at one point the independence of g,h, so for the formula $\mu_f=\mu_g*i\mu_h$ to hold, the equality used at that point, which is as follows, must hold:

$$\sum_{x,y\in\mathbb R}P(g=x,h=y)\delta_{x+iy}=\sum_{x,y\in\mathbb R}P(g=x)P(h=y)\delta_{x+iy}$$

But this is the same as saying that the following must hold, for any x,y:

$$P(g=x,h=y)=P(g=x)P(h=y)$$

We conclude that, in order for the decomposition formula $\mu_f=\mu_g*i\mu_h$ to hold, the real and imaginary parts $g,h:X\to\mathbb R$ must be independent, as stated. □

Many other things can be said, along the same lines, inspired by the basic theory of the complex numbers. Indeed, what we used in the above was the fact that any complex number decomposes as $z=x+iy$ with $x,y\in\mathbb R$, but at a more advanced level, we can equally use formulae of type $z=re^{it}$, or $|z|^2=z\bar z$ and so on, and we are led in this way to a whole collection of results, connecting real and complex probability theory.

Going now straight to the point, probabilistic limiting theorems, let us discuss the complex analogue of the CLT. We have the following statement, to start with:

Theorem 5.3. *Given discrete complex variables $f_1,f_2,f_3,\ldots$ whose real and imaginary parts are i.i.d., centered, and with common variance $t>0$, we have*

$$\frac{1}{\sqrt n}\sum_{i=1}^n f_i\sim C_t$$

with $n\to\infty$, in moments, where C_t is the law of a complex variable whose real and imaginary parts are independent, and each following the law g_t.

Proof. This follows indeed from the real CLT, established in chapter 2, simply by taking the real and imaginary parts of all the variables involved. □

It is tempting at this point to call Theorem 5.3 the complex CLT, or CCLT, but before doing that, let us study a bit more all this. We would like to have a better understanding of the limiting law C_t at the end, and for this purpose, let us look at a sum as follows, with a,b being real independent variables, both following the normal law g_t:

$$c=a+ib$$

To start with, this variable is centered, in a complex sense, because we have:

$$\begin{aligned}E(c)&=E(a+ib)\\&=E(a)+iE(b)\\&=0+i\cdot 0\\&=0\end{aligned}$$

Regarding now the variance, things are more complicated, because the usual variance formula from the real case, which is $V(c) = E(c^2)$ in the centered case, will not provide us with a positive number, in the case where our variable is not real. So, in order to have a variance which is real, and positive too, we must rather use a formula of type $V(c) = E(|c|^2)$, in the centered case. And, with this convention for the variance, we have then the following computation, for the variance of the above variable c:

$$\begin{aligned} V(c) &= E(|c|^2) \\ &= E(a^2 + b^2) \\ &= E(a^2) + E(b^2) \\ &= V(a^2) + V(b^2) \\ &= t + t \\ &= 2t \end{aligned}$$

But this suggests to divide everything by $\sqrt{2}$, as to have in the end a variable having complex variance t, in our sense, and we are led in this way into:

Definition 5.4. *The complex normal, or Gaussian law of parameter $t > 0$ is*

$$G_t = law\left(\frac{1}{\sqrt{2}}(a + ib)\right)$$

where a, b are real and independent, each following the law g_t.

In short, the complex normal laws appear as natural complexifications of the real normal laws. As in the real case, these measures form convolution semigroups:

Proposition 5.5. *The complex Gaussian laws have the property*

$$G_s * G_t = G_{s+t}$$

for any $s, t > 0$, and so they form a convolution semigroup.

Proof. This follows indeed from the real result, namely $g_s * g_t = g_{s+t}$, established in chapter 2, simply by taking real and imaginary parts. □

We have as well the following complex analogue of the CLT:

Theorem 5.6 (CCLT). *Given discrete complex variables $f_1, f_2, f_3, \ldots$ whose real and imaginary parts are i.i.d. and centered, and having variance $t > 0$, we have*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_i \sim G_t$$

with $n \to \infty$, in moments.

Proof. This follows indeed from our previous CCLT result, from Theorem 5.3, by dividing everything by $\sqrt{2}$, as explained in the above. □

5b. Wick formula

Regarding now the moments, the situation here is more complicated than in the real case, because in order to have good results, we have to deal with both the complex variables, and their conjugates. Let us formulate the following definition:

DEFINITION 5.7. *The moments a complex variable $f \in L^\infty(X)$ are the numbers*

$$M_k = E(f^k)$$

depending on colored integers $k = \circ\bullet\bullet\circ\ldots$, with the conventions

$$f^\emptyset = 1 \quad , \quad f^\circ = f \quad , \quad f^\bullet = \bar{f}$$

and multiplicativity, in order to define the colored powers f^k.

As an illustration for this notion, which is something very intuitive, here are the formulae of the four possible order 2 moments of a complex variable f:

$$M_{\circ\circ} = E(f^2) \quad , \quad M_{\circ\bullet} = E(f\bar{f})$$

$$M_{\bullet\circ} = E(\bar{f}f) \quad , \quad M_{\bullet\bullet} = E(\bar{f}^2)$$

Observe that, since $f, \bar{f}$ commute, we have the following identity, which shows that there is a bit of redundancy in our above definition, as formulated:

$$M_{\circ\bullet} = M_{\bullet\circ}$$

In fact, again since $f, \bar{f}$ commute, we can permute terms, in the general context of Definition 5.7, and restrict the attention to exponents of the following type:

$$k = \ldots\circ\circ\circ\bullet\bullet\bullet\bullet\ldots$$

However, our results about the complex Gaussian laws, and other complex laws, later on, not to talk about laws of matrices, random matrices and other noncommuting variables, that will appear later too, will look better without doing this. So, we will use Definition 5.7 as stated. Getting to work now, we first have the following result:

THEOREM 5.8. *The moments of the complex normal law are given by*

$$M_k(G_t) = \begin{cases} t^p p! & (k \text{ uniform, of length } 2p) \\ 0 & (k \text{ not uniform}) \end{cases}$$

where $k = \circ\bullet\bullet\circ\ldots$ is called uniform when it contains the same number of $\circ$ and $\bullet$.

PROOF. We must compute the moments, with respect to colored integer exponents $k = \circ\bullet\bullet\circ\ldots$ as above, of the variable from Definition 5.4, namely:

$$f = \frac{1}{\sqrt{2}}(a + ib)$$

We can assume that we are in the case $t = 1$, and the proof here goes as follows:

(1) As a first observation, in the case where our exponent $k = \circ\bullet\bullet\circ\ldots$ is not uniform, a standard rotation argument shows that the corresponding moment of f vanishes. To be more precise, the variable $f' = wf$ is complex Gaussian too, for any complex number $w \in \mathbb{T}$, and from $M_k(f) = M_k(f')$ we obtain $M_k(f) = 0$, in this case.

(2) In the uniform case now, where the exponent $k = \circ\bullet\bullet\circ\ldots$ consists of p copies of $\circ$ and p copies of $\bullet$, the corresponding moment can be computed as follows:

$$\begin{aligned}
M_k &= \int (f\bar{f})^p \\
&= \frac{1}{2^p}\int (a^2+b^2)^p \\
&= \frac{1}{2^p}\sum_r \binom{p}{r}\int a^{2r}\int b^{2p-2r} \\
&= \frac{1}{2^p}\sum_r \binom{p}{r}(2r)!!(2p-2r)!! \\
&= \frac{1}{2^p}\sum_r \frac{p!}{r!(p-r)!}\cdot\frac{(2r)!}{2^r r!}\cdot\frac{(2p-2r)!}{2^{p-r}(p-r)!} \\
&= \frac{p!}{4^p}\sum_r \binom{2r}{r}\binom{2p-2r}{p-r}
\end{aligned}$$

(3) In order to finish now the computation, let us recall that we have the following formula, coming from the generalized binomial formula, or from the Taylor formula:

$$\frac{1}{\sqrt{1+t}} = \sum_{q=0}^{\infty}\binom{2q}{q}\left(\frac{-t}{4}\right)^q$$

By taking the square of this series, we obtain the following formula:

$$\frac{1}{1+t} = \sum_p \left(\frac{-t}{4}\right)^p \sum_r \binom{2r}{r}\binom{2p-2r}{p-r}$$

Now by looking at the coefficient of t^p on both sides, we conclude that the sum on the right equals 4^p. Thus, we can finish the moment computation in (2), as follows:

$$M_k = \frac{p!}{4^p}\times 4^p = p!$$

We are therefore led to the conclusion in the statement. □

What we have in Theorem 5.8 is usually what is needed in practice, when dealing with moments. But, as before with the real normal laws, or even before with the Poisson laws, a better-looking statement regarding the moments is in terms of partitions.

Indeed, given a colored integer $k = \circ \bullet \bullet \circ \ldots$, let us say that $\pi \in P_2(k)$ is matching when it pairs $\circ - \bullet$ symbols. With this convention, we have the following result:

THEOREM 5.9. *The moments of the complex normal law are the numbers*

$$M_k(G_t) = \sum_{\pi \in \mathcal{P}_2(k)} t^{|\pi|}$$

where $\mathcal{P}_2(k)$ are the matching pairings of $\{1, \ldots, k\}$, and $|.|$ is the number of blocks.

PROOF. This is a reformulation of Theorem 5.8. Indeed, we can assume that we are in the case $t = 1$, and here we know from Theorem 5.8 that the moments are:

$$M_k = \begin{cases} (|k|/2)! & (k \text{ uniform}) \\ 0 & (k \text{ not uniform}) \end{cases}$$

On the other hand, the numbers $|\mathcal{P}_2(k)|$ are given by exactly the same formula. Indeed, in order to have a matching pairing of k, our exponent $k = \circ \bullet \bullet \circ \ldots$ must be uniform, consisting of p copies of $\circ$ and p copies of $\bullet$, with $p = |k|/2$. But then the matching pairings of k correspond to the permutations of the $\bullet$ symbols, as to be matched with $\circ$ symbols, and so we have $p!$ such pairings. Thus, we have the same formula as for the moments of f, and we are led to the conclusion in the statement. □

In practice, we also need to know how to compute joint moments. We have here:

THEOREM 5.10 (Wick formula). *Given independent variables f_i, each following the complex normal law G_t, with $t > 0$ being a fixed parameter, we have the formula*

$$E\left(f_{i_1}^{k_1} \ldots f_{i_s}^{k_s}\right) = t^{s/2} \# \left\{ \pi \in \mathcal{P}_2(k) \middle| \pi \leq \ker i \right\}$$

where $k = k_1 \ldots k_s$ and $i = i_1 \ldots i_s$, for the joint moments of these variables, where $\pi \leq \ker i$ means that the indices of i must fit into the blocks of π, in the obvious way.

PROOF. This is something well-known, which can be proved as follows:

(1) Let us first discuss the case where we have a single variable f, which amounts in taking $f_i = f$ for any i in the formula in the statement. What we have to compute here are the moments of f, with respect to colored integer exponents $k = \circ \bullet \bullet \circ \ldots$, and the formula in the statement tells us that these moments must be:

$$E(f^k) = t^{|k|/2} |\mathcal{P}_2(k)|$$

But this is the formula in Theorem 5.9, so we are done with this case.

(2) In general now, when expanding the product $f_{i_1}^{k_1} \ldots f_{i_s}^{k_s}$ and rearranging the terms, we are left with doing a number of computations as in (1), and then making the product of the expectations that we found. But this amounts in counting the partitions in the statement, with the condition $\pi \leq \ker i$ there standing for the fact that we are doing the various type (1) computations independently, and then making the product. □

The above statement is one of the possible formulations of the Wick formula, and there are many more formulations, which are all useful. For instance, we have:

THEOREM 5.11 (Wick formula 2). *Given independent variables f_i, each following the complex normal law G_t, with $t > 0$ being a fixed parameter, we have the formula*

$$E\left(f_{i_1}\ldots f_{i_k}f_{j_1}^*\ldots f_{j_k}^*\right)=t^k\#\left\{\pi\in S_k\Big|i_{\pi(r)}=j_r,\forall r\right\}$$

for the non-vanishing joint moments of these variables.

PROOF. This follows from the usual Wick formula, from Theorem 5.10. With some changes in the indices and notations, the formula there reads:

$$E\left(f_{I_1}^{K_1}\ldots f_{I_s}^{K_s}\right)=t^{s/2}\#\left\{\sigma\in\mathcal{P}_2(K)\Big|\sigma\leq\ker I\right\}$$

Now observe that we have $\mathcal{P}_2(K)=\emptyset$, unless the colored integer $K=K_1\ldots K_s$ is uniform, in the sense that it contains the same number of $\circ$ and $\bullet$ symbols. Up to permutations, the non-trivial case, where the moment is non-vanishing, is the case where the colored integer $K=K_1\ldots K_s$ is of the following special form:

$$K=\underbrace{\circ\circ\ldots\circ}_k\ \underbrace{\bullet\bullet\ldots\bullet}_k$$

So, let us focus on this case, which is the non-trivial one. Here we have $s=2k$, and we can write the multi-index $I=I_1\ldots I_s$ in the following way:

$$I=i_1\ldots i_k\ j_1\ldots j_k$$

With these changes made, the above usual Wick formula reads:

$$E\left(f_{i_1}\ldots f_{i_k}f_{j_1}^*\ldots f_{j_k}^*\right)=t^k\#\left\{\sigma\in\mathcal{P}_2(K)\Big|\sigma\leq\ker(ij)\right\}$$

The point now is that the matching pairings $\sigma\in\mathcal{P}_2(K)$, with $K=\circ\ldots\circ\bullet\ldots\bullet$, of length $2k$, as above, correspond to the permutations $\pi\in S_k$, in the obvious way. With this identification made, the above modified usual Wick formula becomes:

$$E\left(f_{i_1}\ldots f_{i_k}f_{j_1}^*\ldots f_{j_k}^*\right)=t^k\#\left\{\pi\in S_k\Big|i_{\pi(r)}=j_r,\forall r\right\}$$

Thus, we have reached to the formula in the statement, and we are done. □

Finally, here is one more formulation of the Wick formula, useful as well:

THEOREM 5.12 (Wick formula 3). *Given independent variables f_i, each following the complex normal law G_t, with $t > 0$ being a fixed parameter, we have the formula*

$$E\left(f_{i_1}f_{j_1}^*\ldots f_{i_k}f_{j_k}^*\right)=t^k\#\left\{\pi\in S_k\Big|i_{\pi(r)}=j_r,\forall r\right\}$$

for the non-vanishing joint moments of these variables.

PROOF. This follows from our second Wick formula, from Theorem 5.11, simply by permuting the terms, as to have an alternating sequence of plain and conjugate variables. Alternatively, we can start with Theorem 5.10, and then perform the same manipulations as in the proof of Theorem 5.11, but with the exponent being this time as follows:

$$K=\underbrace{\circ\bullet\circ\bullet\ldots\ldots\circ\bullet}_{2k}$$

Thus, we are led to the conclusion in the statement. □

5c. Complex spheres

Getting now to geometric aspects, we have the following variation of the formula for spherical integrals from chapter 4, dealing now with the complex sphere:

THEOREM 5.13. *We have the following integration formula over the complex sphere $S^{N-1}_\mathbb{C}\subset\mathbb{C}^N$, with respect to the normalized uniform measure,*

$$\int_{S^{N-1}_\mathbb{C}}|z_1|^{2k_1}\ldots|z_N|^{2k_N}\,dz=\frac{(N-1)!k_1!\ldots k_n!}{(N+\sum k_i-1)!}$$

valid for any exponents $k_i\in\mathbb{N}$. As for the other polynomial integrals in $z_1,\ldots,z_N$ and their conjugates $\bar{z}_1,\ldots,\bar{z}_N$, these all vanish.

PROOF. Consider an arbitrary polynomial integral over $S^{N-1}_\mathbb{C}$, containing the same number of plain and conjugated variables, as to not vanish trivially, written as follows:

$$I=\int_{S^{N-1}_\mathbb{C}}z_{i_1}\bar{z}_{i_2}\ldots z_{i_{2k-1}}\bar{z}_{i_{2k}}\,dz$$

By using transformations of type $p\to\lambda p$ with $|\lambda|=1$, we see that this integral I vanishes, unless each z_a appears as many times as $\bar{z}_a$ does, and this gives the last assertion. So, assume now that we are in the non-vanishing case. Then the k_a copies of z_a and the k_a copies of $\bar{z}_a$ produce by multiplication a factor $|z_a|^{2k_a}$, so we have:

$$I=\int_{S^{N-1}_\mathbb{C}}|z_1|^{2k_1}\ldots|z_N|^{2k_N}\,dz$$

Now by using the standard identification $S^{N-1}_\mathbb{C}\simeq S^{2N-1}_\mathbb{R}$, we obtain:

$$\begin{aligned}I&=\int_{S^{2N-1}_\mathbb{R}}(x_1^2+y_1^2)^{k_1}\ldots(x_N^2+y_N^2)^{k_N}\,d(x,y)\\&=\sum_{r_1\ldots r_N}\binom{k_1}{r_1}\ldots\binom{k_N}{r_N}\int_{S^{2N-1}_\mathbb{R}}x_1^{2k_1-2r_1}y_1^{2r_1}\ldots x_N^{2k_N-2r_N}y_N^{2r_N}\,d(x,y)\end{aligned}$$

By using the formula from chapter 4, for the real spheres, we obtain:

$$\begin{aligned}
I &= \sum_{r_1\ldots r_N}\binom{k_1}{r_1}\ldots\binom{k_N}{r_N}\frac{(2N-1)!!(2r_1)!!\ldots(2r_N)!!(2k_1-2r_1)!!\ldots(2k_N-2r_N)!!}{(2N+2\sum k_i-1)!!}\\
&= \sum_{r_1\ldots r_N}\binom{k_1}{r_1}\ldots\binom{k_N}{r_N}\frac{2^{N-1}(N-1)!\prod(2r_i)!/(2^{r_i}r_i!)\prod(2k_i-2r_i)!/(2^{k_i-r_i}(k_i-r_i)!)}{2^{N+\sum k_i-1}(N+\sum k_i-1)!}\\
&= \sum_{r_1\ldots r_N}\binom{k_1}{r_1}\ldots\binom{k_N}{r_N}\frac{(N-1)!(2r_1)!\ldots(2r_N)!(2k_1-2r_1)!\ldots(2k_N-2r_N)!}{4^{\sum k_i}(N+\sum k_i-1)!r_1!\ldots r_N!(k_1-r_1)!\ldots(k_N-r_N)!}
\end{aligned}$$

Now observe that can rewrite this quantity in the following way:

$$\begin{aligned}
I &= \sum_{r_1\ldots r_N}\frac{k_1!\ldots k_N!(N-1)!(2r_1)!\ldots(2r_N)!(2k_1-2r_1)!\ldots(2k_N-2r_N)!}{4^{\sum k_i}(N+\sum k_i-1)!(r_1!\ldots r_N!(k_1-r_1)!\ldots(k_N-r_N)!)^2}\\
&= \sum_{r_1}\binom{2r_1}{r_1}\binom{2k_1-2r_1}{k_1-r_1}\ldots\sum_{r_N}\binom{2r_N}{r_N}\binom{2k_N-2r_N}{k_N-r_N}\frac{(N-1)!k_1!\ldots k_N!}{4^{\sum k_i}(N+\sum k_i-1)!}\\
&= 4^{k_1}\times\ldots\times 4^{k_N}\times\frac{(N-1)!k_1!\ldots k_N!}{4^{\sum k_i}(N+\sum k_i-1)!}\\
&= \frac{(N-1)!k_1!\ldots k_N!}{(N+\sum k_i-1)!}
\end{aligned}$$

Thus, we are led to the formula in the statement. □

In what regards now the hyperspherical laws, the result here is as follows:

THEOREM 5.14. *The rescalings $\sqrt{N}z_i$ of the unit complex sphere coordinates*

$$z_i : S^{N-1}_{\mathbb{C}} \to \mathbb{C}$$

as well as the rescalings $\sqrt{N}U_{ij}$ of the unitary group coordinates

$$U_{ij} : U_N \to \mathbb{C}$$

become complex Gaussian and independent with $N \to \infty$.

PROOF. We have several assertions to be proved, the idea being as follows:

(1) According to the formula in Theorem 5.13, the polynomials integrals in $z_i, \bar{z}_i$ vanish, unless the number of $z_i, \bar{z}_i$ is the same. In this latter case these terms can be grouped together, by using $z_i\bar{z}_i = |z_i|^2$, and the relevant integration formula is:

$$\int_{S^{N-1}_{\mathbb{C}}} |z_i|^{2k}\,dz = \frac{(N-1)!k!}{(N+k-1)!}$$

Now with $N \to \infty$, we obtain from this the following estimate:

$$\int_{S^{N-1}_\mathbb{C}} |z_i|^{2k} dx \simeq N^{-k} \times k!$$

Thus, the rescaled variables $\sqrt{N} z_i$ become normal with $N \to \infty$, as claimed.

(2) As for the proof of the asymptotic independence, this is standard too, again by using the formula in Theorem 5.13. Indeed, the joint moments of $z_1, \ldots, z_N$ are given by:

$$\begin{aligned}\int_{S^{N-1}_\mathbb{R}} |z_1|^{2k_1} \ldots |z_N|^{2k_N} \, dx &= \frac{(N-1)! k_1! \ldots k_n!}{(N + \sum k_i - 1)!} \\ &\simeq N^{-\Sigma k_i} \times k_1! \ldots k_N!\end{aligned}$$

By rescaling, the joint moments of the variables $y_i = \sqrt{N} z_i$ are given by:

$$\int_{S^{N-1}_\mathbb{R}} |y_1|^{2k_1} \ldots |y_N|^{2k_N} \, dx \simeq k_1! \ldots k_N!$$

Thus, we have multiplicativity, and so independence with $N \to \infty$, as claimed.

(3) Regarding the last assertion, we can use here the basic fact that the rotations $U \in U_N$ act on the points of the sphere $z \in S^{N-1}_\mathbb{C}$, with the stabilizer of $z = (1, 0, \ldots, 0)$ being the subgroup $U_{N-1} \subset U_N$. In algebraic terms, this gives an equality as follows:

$$S^{N-1}_\mathbb{C} = U_N / U_{N-1}$$

In functional analytic terms, this result provides us with an embedding as follows, for any i, which makes correspond the respective integration functionals:

$$C(S^{N-1}_\mathbb{C}) \subset C(U_N) \quad , \quad x_i \to U_{1i}$$

With this identification made, the result follows from (1,2). □

5d. Unitary groups

We can enlarge as well our collection of character results, as follows:

THEOREM 5.15. *With $N \to \infty$, the laws of truncated characters are as follows:*

(1) *For O_N we obtain the Gaussian law g_t.*
(2) *For U_N we obtain the complex Gaussian law G_t.*
(3) *For S_N we obtain the Poisson law p_t.*
(4) *For H_N we obtain the Bessel law b_t.*
(5) *For H_N^s we obtain the generalized Bessel law b_t^s.*
(6) *For K_N we obtain the complex Bessel law B_t.*

PROOF. We already know these results in the real case, and in the complex case the proof is similar, based on the real version of the Weingarten formula from chapter 4. □

5e. Exercises

Exercises:

EXERCISE 5.16.

EXERCISE 5.17.

EXERCISE 5.18.

EXERCISE 5.19.

EXERCISE 5.20.

EXERCISE 5.21.

EXERCISE 5.22.

EXERCISE 5.23.

Bonus exercise.

CHAPTER 6

Rayleigh variables

6a. Rayleigh variables

As a consequence of the moment computations from chapter 5, we have:

THEOREM 6.1. *The moments of the Rayleigh law, given by*

$$R_t = law(|G_t|)$$

are given by the following formula, at the parameter value $t = 1$,

$$M_p = p!$$

and are given by the formula $M_p = t^p p!$, in general.

PROOF. This follows indeed from the moment computations from chapter 5. □

Many other things can be said about the Rayleigh laws, which are quite interesting mathematical objects, notably with some further formulae, regarding the Fourier transform, and the cumulants, which can be obtained as a consequence of Theorem 6.1.

Let us record as well the following statement, called Rayleigh Central Limiting Theorem, which is something of theoretical importance:

THEOREM 6.2 (RCLT). *Given discrete complex variables $f_1, f_2, f_3, \ldots$ whose real and imaginary parts are i.i.d. and centered, and having variance $t > 0$, we have*

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_i \right| \sim R_t$$

with $n \to \infty$, in moments.

PROOF. This follows indeed from our previous Central Limiting result, namely the CCLT from chapter 5, by taking absolute values on both sides. □

As a comment here, observe that the Rayleigh laws, while being certainly real, capture the essentials of the two-dimensional nature of the normal laws.

6b.

6c.

6d.

6e. Exercises

Exercises:

Exercise 6.3.

Exercise 6.4.

Exercise 6.5.

Exercise 6.6.

Exercise 6.7.

Exercise 6.8.

Exercise 6.9.

Exercise 6.10.

Bonus exercise.

CHAPTER 7

Formulae, analysis

7a. Formulae, analysis

7b.

7c.

7d.

7e. Exercises

Exercises:

EXERCISE 7.1.

EXERCISE 7.2.

EXERCISE 7.3.

EXERCISE 7.4.

EXERCISE 7.5.

EXERCISE 7.6.

EXERCISE 7.7.

EXERCISE 7.8.

Bonus exercise.

CHAPTER 8

Invariance questions

8a. Invariance questions

8b.

8c.

8d.

8e. Exercises

Exercises:

EXERCISE 8.1.

EXERCISE 8.2.

EXERCISE 8.3.

EXERCISE 8.4.

EXERCISE 8.5.

EXERCISE 8.6.

EXERCISE 8.7.

EXERCISE 8.8.

Bonus exercise.

Part III

Higher dimensions

CHAPTER 9

Gaussian vectors

9a. Gaussian vectors

9b.

9c.

9d.

9e. Exercises

Exercises:

Exercise 9.1.

Exercise 9.2.

Exercise 9.3.

Exercise 9.4.

Exercise 9.5.

Exercise 9.6.

Exercise 9.7.

Exercise 9.8.

Bonus exercise.

CHAPTER 10

Functional analysis

10a. Functional analysis

10b.

10c.

10d.

10e. Exercises

Exercises:

Exercise 10.1.

Exercise 10.2.

Exercise 10.3.

Exercise 10.4.

Exercise 10.5.

Exercise 10.6.

Exercise 10.7.

Exercise 10.8.

Bonus exercise.

CHAPTER 11

The heat kernel

11a. Heat diffusion

Time now for some applications of what we learned, to theoretical physics, and more specifically, to thermodynamics. And with this being something significant, because it is in relation with questions from thermodynamics that probability theory really shines.

As a main question, that we would like to investigate in this chapter, we have:

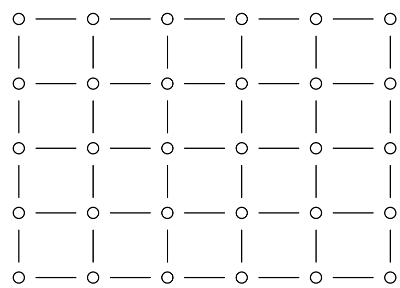
QUESTION 11.1. *How does heat propagate, in the context of two of several materials put in contact, or even in the case of a single material, not uniformly heated?*

As a first observation, this is a quite wide-ranging and tricky question, and even in the case of gases, assuming the Boltzmann theory developed, talking about such things will full mathematical rigor will be no easy task. In short, we are here deep into physics, or very physical type, and of material science flavor, so modesty, and time to discuss all this, with a mixture of general laws, experimental findings, and some math too.

Regarding several materials put in contact, we already had a flavor of that in the context of basic thermodynamics, when talking about the Clapeyron equation, with the "materials" in that case being the liquid and gaseous form of the same material.

However, all this is rather something quite advanced. The simplest diffusion problems appear in fact when putting several gases, at different temperatures, in contact, and with these several gases being allowed to be actually samples of the same gas, but at different temperatures. And there are countless things that can be said here, both at the level of basic thermodynamics, and at the level of more advanced theory.

In practice, the simplest heat diffusion question, studied and understood since long, concerns a container containing two gases, having initial different temperatures $T_1 < T_2$, separated by a membrane. Heat transfer goes on, in this setting, and obviously, we can model this by focusing on the membrane, with a basic grid model for it, as follows:



There is some sort of "game" played by the two gases, over this grid, and we can model this, and then recover the known results about heat diffusion, in this setting.

At a more advanced level, we can remove the membrane. Again, there is some sort of "game" here, played by the two gases, which can be 2D or 3D, depending on modelling. Also, in this setting, we can actually keep the membrane, but allow it to inflate.

11b. Some calculus

In order to further study the heat diffusion, we will need some standard multivariable calculus. Let us start with a straightforward definition, as follows:

Definition 11.2. *We say that a map $f:\mathbb{R}^N\to\mathbb{R}^M$ is differentiable at $x\in\mathbb{R}^N$ if*

$$f(x+t)\simeq f(x)+f'(x)t$$

for some linear map $f'(x):\mathbb{R}^N\to\mathbb{R}^M$, called derivative of f at the point $x\in\mathbb{R}^N$.

But is this the correct definition. I can hear you screaming that we are probably going the wrong way, because for functions $f:\mathbb{R}\to\mathbb{R}$ the derivative is something much simpler, as follows, and that we should try to imitate, in our higher dimensional setting:

$$f'(x)=\lim_{t\to0}\frac{f(x+t)-f(x)}{t}$$

However, this is not possible, for a number of reasons, that are worth discussing in detail. So, here is the discussion, answering all kinds of questions that you might have:

(1) First of all, the above formula does not make any sense for a function $f:\mathbb{R}^N\to\mathbb{R}^M$ with $N\neq M$, because we cannot divide oranges by apples. And it doesn't make sense either at $N=M\in\mathbb{N}$, because, well, here we have $\mathbb{R}^N$ oranges, I agree with you, but there is no way of dividing these oranges, unless we are in the special cases $N=1,2$.

(2) More philosophically know, we have seen that having $f'(x)$ defined as a number is difficult, but the question is, do we really want to have $f'(x)$ defined as a number? And my claim here is that, this would be a pity. Think at the case where $f:\mathbb{R}^N\to\mathbb{R}^M$ is linear. Such a map is just "perfect", and so should equal its own derivative, $f=f'$.

(3) Summarizing, our Definition 11.2 is perfection, and is waiting for some further study, and this is what we will do. And in case you're still secretly dreaming about having $f'(x)$ defined as some sort of number, wait for it. When $N = M$ at least, there is indeed a lucky number, namely $\det(f'(x))$, called Jacobian, but more on this later.

Getting back now to Definition 11.2 as formulated, and agreed upon, we have there a linear map $f'(x) : \mathbb{R}^N \to \mathbb{R}^M$, waiting to be further understood. So, time now to use our linear algebra knowledge. We know from there that such linear maps correspond to rectangular matrices $A \in M_{M\times N}(\mathbb{R})$, and we are led in this way to:

QUESTION 11.3. *Given a differentiable map $f : \mathbb{R}^N \to \mathbb{R}^M$, in the abstract sense of Definition 7.2, what exactly is its derivative*

$$f'(x) : \mathbb{R}^N \to \mathbb{R}^M$$

regarded as a rectangular matrix, $f'(x) \in M_{M\times N}(\mathbb{R})$?

Again, I might hear scream you here, arguing that you come after a long battle, just agreeing that the derivative is a linear map, and not a number, and now what, we are trying to replace this linear map by a matrix, and so by a bunch of numbers. Good point, and I have no good answer to this. What we are doing here, Definition 11.2, then Question 11.3, and finally Theorem 11.4 to follow, are things that took mankind several centuries to develop, and that we are now presenting in a compressed form.

In any case, hope that you're still with me, and here is the answer to Question 11.3:

THEOREM 11.4. *The derivative of a differentiable function $f : \mathbb{R}^N \to \mathbb{R}^M$, making the approximation formula*

$$f(x+t) \simeq f(x) + f'(x)t$$

work, is the matrix of partial derivatives at x, namely

$$f'(x) = \left(\frac{df_i}{dx_j}(x)\right)_{ij} \in M_{M\times N}(\mathbb{R})$$

acting on the vectors $t \in \mathbb{R}^N$ by usual multiplication.

PROOF. As a first observation, the formula in the statement makes sense indeed, as an equality, or rather approximation, of vectors in $\mathbb{R}^M$, as follows:

$$f\begin{pmatrix} x_1+t_1 \\ \vdots \\ x_N+t_N \end{pmatrix} \simeq f\begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} \frac{df_1}{dx_1}(x) & \ldots & \frac{df_1}{dx_N}(x) \\ \vdots & & \vdots \\ \frac{df_M}{dx_1}(x) & \ldots & \frac{df_M}{dx_N}(x) \end{pmatrix} \begin{pmatrix} t_1 \\ \vdots \\ t_N \end{pmatrix}$$

In order to prove now this formula, which does make sense, the idea is as follows:

(1) First of all, at $N=M=1$ what we have is a usual 1-variable function $f:\mathbb{R}\to\mathbb{R}$, and the formula in the statement is something that we know well, namely:

$$f(x+t)\simeq f(x)+f'(x)t$$

(2) Let us discuss now the case $N=2,M=1$. Here what we have is a function $f:\mathbb{R}^2\to\mathbb{R}$, and by using twice the basic approximation result from (1), we obtain:

$$\begin{aligned}
f\begin{pmatrix}x_1+t_1\\x_2+t_2\end{pmatrix} &\simeq f\begin{pmatrix}x_1+t_1\\x_2\end{pmatrix}+\frac{df}{dx_2}(x)t_2\\
&\simeq f\begin{pmatrix}x_1\\x_2\end{pmatrix}+\frac{df}{dx_1}(x)t_1+\frac{df}{dx_2}(x)t_2\\
&= f\begin{pmatrix}x_1\\x_2\end{pmatrix}+\begin{pmatrix}\frac{df}{dx_1}(x)&\frac{df}{dx_2}(x)\end{pmatrix}\begin{pmatrix}t_1\\t_2\end{pmatrix}
\end{aligned}$$

(3) More generally, we can deal in this way with the general case $M=1$, as follows:

$$\begin{aligned}
f\begin{pmatrix}x_1+t_1\\\vdots\\x_N+t_N\end{pmatrix} &\simeq f\begin{pmatrix}x_1\\\vdots\\x_N\end{pmatrix}+\frac{df}{dx_1}(x)t_1+\ldots+\frac{df}{dx_N}(x)t_N\\
&= f\begin{pmatrix}x_1\\\vdots\\x_N\end{pmatrix}+\begin{pmatrix}\frac{df}{dx_1}(x)&\ldots&\frac{df}{dx_N}(x)\end{pmatrix}\begin{pmatrix}t_1\\\vdots\\t_N\end{pmatrix}
\end{aligned}$$

(4) But this gives the result in the case where both $N,M\in\mathbb{N}$ are arbitrary too. Indeed, consider a function $f:\mathbb{R}^N\to\mathbb{R}^M$, and let us write it as follows:

$$f=\begin{pmatrix}f_1\\\vdots\\f_M\end{pmatrix}$$

We can apply (3) to each of the components $f_i:\mathbb{R}^N\to\mathbb{R}$, and we get:

$$f_i\begin{pmatrix}x_1+t_1\\\vdots\\x_N+t_N\end{pmatrix}\simeq f_i\begin{pmatrix}x_1\\\vdots\\x_N\end{pmatrix}+\begin{pmatrix}\frac{df_i}{dx_1}(x)&\ldots&\frac{df_i}{dx_N}(x)\end{pmatrix}\begin{pmatrix}t_1\\\vdots\\t_N\end{pmatrix}$$

But this collection of M formulae tells us precisely that the following happens, as an equality, or rather approximation, of vectors in $\mathbb{R}^M$:

$$f\begin{pmatrix}x_1+t_1\\\vdots\\x_N+t_N\end{pmatrix}\simeq f\begin{pmatrix}x_1\\\vdots\\x_N\end{pmatrix}+\begin{pmatrix}\frac{df_1}{dx_1}(x)&\ldots&\frac{df_1}{dx_N}(x)\\\vdots&&\vdots\\\frac{df_M}{dx_1}(x)&\ldots&\frac{df_M}{dx_N}(x)\end{pmatrix}\begin{pmatrix}t_1\\\vdots\\t_N\end{pmatrix}$$

Thus, we are led to the conclusion in the statement. □

Moving forward, let us formulate something nice, namely:

DEFINITION 11.5. *Given $f:\mathbb{R}^N\to\mathbb{R}^N$, its Jacobian at $x\in\mathbb{R}^N$ is the number*

$$\det(f'(x))\in\mathbb{R}$$

measuring the infinitesimal rate of the volume inflation by f, at the point x.

Here the first part is standard, because when $N=M$, as above, the derivative is a linear map $f'(x):\mathbb{R}^N\to\mathbb{R}^N$, which is the same as a square matrix $f'(x)\in M_N(\mathbb{R})$, and so we can consider the determinant of this matrix, $\det(f'(x))\in\mathbb{R}$. As for the second part, this comes from our knowledge of the determinant from linear algebra.

All this is very nice, and as a first observation, according to our formula of $f'(x)$ as being the matrix formed by the partial derivatives, we have:

$$\det(f'(x))=\begin{vmatrix}\frac{df_1}{dx_1}(x)&\ldots&\frac{df_1}{dx_N}(x)\\ \vdots&&\vdots\\ \frac{df_N}{dx_1}(x)&\ldots&\frac{df_N}{dx_N}(x)\end{vmatrix}$$

Now back to Theorem 11.4, generally speaking, that is what you need to know for upgrading from calculus to multivariable calculus. As a standard result here, we have:

THEOREM 11.6. *We have the chain derivative formula*

$$(f\circ g)'(x)=f'(g(x))\cdot g'(x)$$

as an equality of matrices.

PROOF. Consider indeed a composition of functions, as follows:

$$f:\mathbb{R}^N\to\mathbb{R}^M\quad,\quad g:\mathbb{R}^K\to\mathbb{R}^N\quad,\quad f\circ g:\mathbb{R}^K\to\mathbb{R}^M$$

According to Theorem 11.4, the derivatives of these functions are certain linear maps, corresponding to certain rectangular matrices, as follows:

$$f'(g(x))\in M_{M\times N}(\mathbb{R})\quad,\quad g'(x)\in M_{N\times K}(\mathbb{R})\qquad (f\circ g)'(x)\in M_{M\times K}(\mathbb{R})$$

Thus, our formula makes sense indeed. As for proof, this comes from:

$$\begin{aligned}(f\circ g)(x+t)&=f(g(x+t))\\&\simeq f(g(x)+g'(x)t)\\&\simeq f(g(x))+f'(g(x))g'(x)t\end{aligned}$$

Thus, we are led to the conclusion in the statement. □

Let us recall now from one variable calculus that we have the following result:

Proposition 11.7. *We have the change of variable formula*

$$\int_a^b f(x)dx = \int_c^d f(\varphi(t))\varphi'(t)dt$$

where $c = \varphi^{-1}(a)$ and $d = \varphi^{-1}(b)$.

Proof. This follows with $f = F'$, from the following differentiation rule, that we know well, and whose proof is something elementary:

$$(F\varphi)'(t) = F'(\varphi(t))\varphi'(t)$$

Indeed, by integrating between c and d, we obtain the result. □

In several variables now, we can only expect the above $\varphi'(t)$ factor to be replaced by something similar, a sort of "derivative of φ, arising as a real number". We are led to:

Theorem 11.8. *Given a transformation $\varphi = (\varphi_1, \ldots, \varphi_N)$, we have*

$$\int_E f(x)dx = \int_{\varphi^{-1}(E)} f(\varphi(t))|J_\varphi(t)|dt$$

with the J_φ quantity, called Jacobian, being given by

$$J_\varphi(t) = \det\left[\left(\frac{d\varphi_i}{dx_j}(x)\right)_{ij}\right]$$

and with this generalizing the formula from Proposition 11.7.

Proof. This is something quite tricky, the idea being as follows:

(1) Observe first that this generalizes indeed the change of variable formula in 1 dimension, from Proposition 11.7, the point here being that the absolute value on the derivative appears as to compensate for the lack of explicit bounds for the integral.

(2) In general now, we can first argue that, the formula in the statement being linear in f, we can assume $f = 1$. Thus we want to prove $vol(E) = \int_{\varphi^{-1}(E)} |J_\varphi(t)|dt$, and with $D = \varphi^{-1}(E)$, this amounts in proving $vol(\varphi(D)) = \int_D |J_\varphi(t)|dt$.

(3) Now since this latter formula is additive with respect to D, it is enough to prove that $vol(\varphi(D)) = \int_D J_\varphi(t)dt$, for small cubes D, and assuming $J_\varphi > 0$. But this basically follows by using the definition of the determinant of real matrices, as a volume. □

All the above was of course quite tricky, and you may wonder if there is a simpler proof for this. Good question, and in answer, yes this is a difficult problem, which is actually open, and every now and then mathematicians still publish papers about this. In the hope that one day we will see a smart new paper on this from you too, reader.

Moving now towards second derivatives, that we will need too, we have here:

THEOREM 11.9. *The second derivative of a function $\varphi : \mathbb{R}^N \to \mathbb{R}$, making the formula*

$$\varphi(x+h) \simeq \varphi(x) + \varphi'(x)h + \frac{< \varphi''(x)h, h >}{2}$$

work, is its Hessian matrix $\varphi''(x) \in M_N(\mathbb{R})$, given by the following formula:

$$\varphi''(x) = \left(\frac{d^2\varphi}{dx_i dx_j} \right)_{ij}$$

Moreover, this Hessian matrix is symmetric, $\varphi''(x)_{ij} = \varphi'(x)_{ji}$.

PROOF. There are several things going on here, the idea being as follows:

(1) As a first observation, at $N = 1$ the Hessian matrix constructed above is simply the 1×1 matrix having as entry the second derivative $\varphi''(x)$, and the formula in the statement is something that we know well from basic calculus, namely:

$$\varphi(x+h) \simeq \varphi(x) + \varphi'(x)h + \frac{\varphi''(x)h^2}{2}$$

(2) At $N = 2$ now, we obviously need to differentiate φ twice, and the point is that we come in this way upon the following formula, called Clairaut formula:

$$\frac{d^2\varphi}{dxdy} = \frac{d^2\varphi}{dydx}$$

But, is this formula correct or not? As an intuitive justification for it, let us consider a product of power functions, $\varphi(z) = x^p y^q$. We have then our formula, due to:

$$\frac{d^2\varphi}{dxdy} = \frac{d}{dx}\left(\frac{dx^p y^q}{dy}\right) = \frac{d}{dx}\left(qx^p y^{q-1}\right) = pqx^{p-1}y^{q-1}$$

$$\frac{d^2\varphi}{dydx} = \frac{d}{dy}\left(\frac{dx^p y^q}{dx}\right) = \frac{d}{dy}\left(px^{p-1} y^q\right) = pqx^{p-1}y^{q-1}$$

Next, let us consider a linear combination of power functions, $\varphi(z) = \sum_{pq} c_{pq} x^p y^q$, which can be finite or not. We have then, by using the above computation:

$$\frac{d^2\varphi}{dxdy} = \frac{d^2\varphi}{dydx} = \sum_{pq} c_{pq} pqx^{p-1}y^{q-1}$$

Thus, we can see that our commutation formula for derivatives holds indeed, due to the fact that the functions in x, y commute. Of course, all this does not fully prove our formula, in general. But exercise for you, to have this idea fully working, or to look up the standard proof of the Clairaut formula, using the mean value theorem.

(3) Moving now to $N = 3$ and higher, we can use here the Clairaut formula with respect to any pair of coordinates, which gives the Schwarz formula, namely:

$$\frac{d^2\varphi}{dx_i dx_j} = \frac{d^2\varphi}{dx_j dx_i}$$

Thus, the second derivative, or Hessian matrix, is symmetric, as claimed.

(4) Getting now to the main topic, namely approximation formula in the statement, in arbitrary N dimensions, this is in fact something which does not need a new proof, because it follows from the one-variable formula in (1), applied to the restriction of φ to the following segment in $\mathbb{R}^N$, which can be regarded as being a one-variable interval:

$$I = [x, x + h]$$

To be more precise, let $y \in \mathbb{R}^N$, and consider the following function, with $r \in \mathbb{R}$:

$$f(r) = \varphi(x + ry)$$

We know from (1) that the Taylor formula for f, at the point $r = 0$, reads:

$$f(r) \simeq f(0) + f'(0)r + \frac{f''(0)r^2}{2}$$

And our claim is that, with $h = ry$, this is precisely the formula in the statement.

(5) So, let us see if our claim is correct. By using the chain rule, we have the following formula, with on the right, as usual, a row vector multiplied by a column vector:

$$f'(r) = \varphi'(x + ry) \cdot y$$

By using again the chain rule, we can compute the second derivative as well:

$$\begin{aligned} f''(r) &= (\varphi'(x + ry) \cdot y)' \\ &= \left(\sum_i \frac{d\varphi}{dx_i}(x + ry) \cdot y_i\right)' \\ &= \sum_i \sum_j \frac{d^2\varphi}{dx_i dx_j}(x + ry) \cdot \frac{d(x + ry)_j}{dr} \cdot y_i \\ &= \sum_i \sum_j \frac{d^2\varphi}{dx_i dx_j}(x + ry) \cdot y_i y_j \\ &= < \varphi''(x + ry)y, y > \end{aligned}$$

(6) Time now to conclude. We know that we have $f(r) = \varphi(x + ry)$, and according to our various computations above, we have the following formulae:

$$f(0) = \varphi(x) \quad , \quad f'(0) = \varphi'(x) \quad , \quad f''(0) = < \varphi''(x)y, y >$$

Buit with this data in hand, the usual Taylor formula for our one variable function f, at order 2, at the point $r = 0$, takes the following form, with $h = ry$:

$$\begin{aligned}\varphi(x+ry) &\simeq \varphi(x) + \varphi'(x)ry + \frac{<\varphi''(x)y, y> r^2}{2}\\ &= \varphi(x) + \varphi'(x)t + \frac{<\varphi''(x)h, h>}{2}\end{aligned}$$

Thus, we have obtained the formula in the statement. □

As before in the one variable case, many more things can be said, as a continuation of the above. For instance the local minima and maxima of $\varphi : \mathbb{R}^N \to \mathbb{R}$ appear at the points $x \in \mathbb{R}^N$ where the derivative vanishes, $\varphi'(x) = 0$, and where the second derivative $\varphi''(x) \in M_N(\mathbb{R})$ is positive, respectively negative. But, you surely know all this.

Let us just record here the following key fact, that we will need later:

Proposition 11.10. *Intuitively, the following quantity, called Laplacian of φ,*

$$\Delta\varphi = \sum_{i=1}^N \frac{d^2\varphi}{dx_i^2}$$

measures how much different is $\varphi(x)$, compared to the average of $\varphi(y)$, with $y \simeq x$.

Proof. This is something a bit heuristic, but good to know. Let us write the formula in Theorem 7.9, as such, and with $h \to -h$ too:

$$\varphi(x+h) \simeq \varphi(x) + \varphi'(x)h + \frac{<\varphi''(x)h, h>}{2}$$

$$\varphi(x-h) \simeq \varphi(x) - \varphi'(x)h + \frac{<\varphi''(x)h, h>}{2}$$

By making the average, we obtain the following formula:

$$\frac{\varphi(x+h) + \varphi(x-h)}{2} \simeq \varphi(x) + \frac{<\varphi''(x)h, h>}{2}$$

Thus, thinking a bit, we are led to the conclusion in the statement, modulo some discussion about integrating all this, that we will not really need, in what follows. □

With this understood, the problem is now, what can we say about the mathematics of Δ? We have here the following straightforward question, inspired by linear algebra:

Question 11.11. *The Laplace operator being linear,*

$$\Delta(a\varphi + b\psi) = a\Delta\varphi + b\Delta\psi$$

what can we say about it, inspired by usual linear algebra?

In answer now, the space of functions $\varphi:\mathbb R^N\to\mathbb R$, on which Δ acts, being infinite dimensional, the usual tools from linear algebra do not apply as such, and we must be extremely careful. For instance, we cannot really expect to diagonalize Δ, via some sort of explicit procedure, as we usually do in linear algebra, for the usual matrices.

Thinking some more, there is actually a real bug too with our problem, because at $N=1$ this problem becomes "what can we say about the second derivatives $\varphi'':\mathbb R\to\mathbb R$ of the functions $\varphi:\mathbb R\to\mathbb R$, inspired by linear algebra", with answer "not much".

And by thinking even more, still at $N=1$, there is a second bug too, because if $\varphi:\mathbb R\to\mathbb R$ is twice differentiable, nothing will guarantee that its second derivative $\varphi'':\mathbb R\to\mathbb R$ is twice differentiable too. Thus, we have some issues with the domain and range of Δ, regarded as linear operator, and these problems will persist at higher N.

So, shall we trash Question 11.11? Not so quick, because, very remarkably, some magic comes at $N=2$ and higher in relation with complex analysis, according to:

Principle 11.12. *The functions $\varphi:\mathbb R^N\to\mathbb R$ which are 0-eigenvectors of Δ,*

$$\Delta\varphi=0$$

called harmonic functions, have the following properties:

(1) *At $N=1$, nothing spectacular, these are just the linear functions.*
(2) *At $N=2$, these are, locally, the real parts of holomorphic functions.*
(3) *At $N\geq 3$, these still share many properties with the holomorphic functions.*

In order to understand this, or at least get introduced to it, let us first look at the case $N=2$. Here, any $\varphi:\mathbb R^2\to\mathbb R$ can be regarded as function $\varphi:\mathbb C\to\mathbb R$, depending on $z=x+iy$. Thus, it is natural to enlarge the attention to the functions $\varphi:\mathbb C\to\mathbb C$, and ask which of these functions are harmonic, $\Delta\varphi=0$. And here, we have:

Theorem 11.13. *Any holomorphic function $\varphi:\mathbb C\to\mathbb C$, when regarded as function*

$$\varphi:\mathbb R^2\to\mathbb C$$

is harmonic. Moreover, the conjugates $\bar\varphi$ of holomorphic functions are harmonic too.

Proof. The first assertion comes from the following computation, with $z=x+iy$:

$$\begin{aligned}\Delta z^n &= \frac{d^2z^n}{dx^2}+\frac{d^2z^n}{dy^2}\\ &= \frac{d(nz^{n-1})}{dx}+\frac{d(inz^{n-1})}{dy}\\ &= n(n-1)z^{n-2}-n(n-1)z^{n-2}\\ &= 0\end{aligned}$$

As for the second assertion, this follows from $\Delta\bar\varphi=\overline{\Delta\varphi}$. □

Many more things can be said, along these lines, notably a proof of the assertion (2) in Principle 11.12, which is however a quite tough piece of mathematics, and then with a clarification of the assertion (3) too, from that same principle, which again requires some substantial mathematics. We will be back to both these topics, in due time.

11c. Heat equation

Back to physics, as a second question that we would like to investigate in this chapter, we have the problem of understanding how heat will get diffused over time $t>0$ inside a piece of a material, which is unevenly heated, initially. The result here is as follows:

THEOREM 11.14. *Heat diffusion is described by the heat equation*

$$\dot{\varphi}=\alpha\Delta\varphi$$

where $\alpha>0$ is a constant, called thermal diffusivity of the medium, and

$$\Delta=\sum_i\frac{d}{dx_i^2}$$

is the Laplace operator.

PROOF. The study here can be done by using lattice models, as follows:

(1) To start with, as an intuitive explanation for the equation, since the second derivative φ'' in one dimension, or the quantity $\Delta\varphi$ in general, computes the average value of a function φ around a point, minus the value of φ at that point, the heat equation as formulated above tells us that the rate of change $\dot{\varphi}$ of the temperature of the material at any given point must be proportional, with proportionality factor $\alpha>0$, to the average difference of temperature between that given point and the surrounding material.

(2) The heat equation as formulated above is of course something approximative, and several improvements can be made to it, first by incorporating a term accounting for heat radiation, and then doing several fine-tunings, depending on the material involved. But more on this later, for the moment let us focus on the heat equation above.

(3) In relation with our modelling questions, we can recover this equation by using a basic lattice model. Indeed, let us first assume that we are in the one-dimensional case, $N=1$. Here our model looks as follows, with distance $l>0$ between neighbors:

$$-\!\!-\!\!-\ \circ_{x-l}\overset{l}{-\!\!-\!\!-}\circ_x\overset{l}{-\!\!-\!\!-}\circ_{x+l}\ -\!\!-\!\!-$$

In order to model heat diffusion, we have to implement the intuitive mechanism explained above, namely "the rate of change of the temperature of the material at any given point must be proportional, with proportionality factor $\alpha>0$, to the average difference of temperature between that given point and the surrounding material".

(4) In practice, this leads to a condition as follows, expressing the change of the temperature φ, over a small period of time $\delta > 0$:

$$\varphi(x,t+\delta)=\varphi(x,t)+\frac{\alpha\delta}{l^2}\sum_{x\sim y}\left[\varphi(y,t)-\varphi(x,t)\right]$$

To be more precise, we have made several assumptions here, as follows:

– General heat diffusion assumption: the change of temperature at any given point x is proportional to the average over neighbors, $y\sim x$, of the differences $\varphi(y,t)-\varphi(x,t)$ between the temperatures at x, and at these neighbors y.

– Infinitesimal time and length conditions: in our model, the change of temperature at a given point x is proportional to small period of time involved, $\delta > 0$, and is inverse proportional to the square of the distance between neighbors, l^2.

(5) Regarding these latter assumptions, the one regarding the proportionality with the time elapsed $\delta > 0$ is something quite natural, physically speaking, and mathematically speaking too, because we can rewrite our equation as follows, making it clear that we have here an equation regarding the rate of change of temperature at x:

$$\frac{\varphi(x,t+\delta)-\varphi(x,t)}{\delta}=\frac{\alpha}{l^2}\sum_{x\sim y}\left[\varphi(y,t)-\varphi(x,t)\right]$$

As for the second assumption that we made above, namely inverse proportionality with l^2, this can be justified on physical grounds too, but again, perhaps the best is to do the math, which will show right away where this proportionality comes from.

(6) So, let us do the math. In the context of our 1D model the neighbors of x are the points $x\pm l$, and so the equation that we wrote above takes the following form:

$$\frac{\varphi(x,t+\delta)-\varphi(x,t)}{\delta}=\frac{\alpha}{l^2}\Big[(\varphi(x+l,t)-\varphi(x,t))+(\varphi(x-l,t)-\varphi(x,t))\Big]$$

Now observe that we can write this equation as follows:

$$\frac{\varphi(x,t+\delta)-\varphi(x,t)}{\delta}=\alpha\cdot\frac{\varphi(x+l,t)-2\varphi(x,t)+\varphi(x-l,t)}{l^2}$$

(7) We recognize on the right the usual approximation of the second derivative, coming from calculus, and more specifically from the Taylor formula, at order 2. Thus, when taking the continuous limit of our model, $l\to 0$, we obtain the following equation:

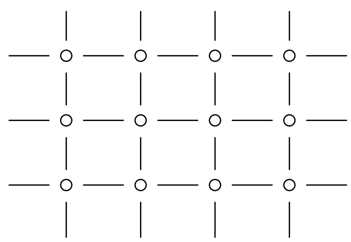
$$\frac{\varphi(x,t+\delta)-\varphi(x,t)}{\delta}=\alpha\cdot\varphi''(x,t)$$

Now with $t\to 0$, we are led in this way to the heat equation, namely:

$$\dot{\varphi}(x,t)=\alpha\cdot\varphi''(x,t)$$

(8) In practice now, there are of course still a few details to be discussed, in relation with all this, for instance at the end, in relation with the precise order of the limiting operations $l \to 0$ and $\delta \to 0$ to be performed, but these remain minor aspects, because our equation makes it clear, right from the beginning, that time and space are separated, and so that there is no serious issue with all this. And so, fully done with 1D.

(9) With this done, let us discuss now 2 dimensions. Here we can use a lattice model as follows, with all lengths being $l > 0$, for simplifying:



(10) We have to implement now the physical heat diffusion mechanism, namely "the rate of change of the temperature of the material at any given point must be proportional, with proportionality factor $\alpha > 0$, to the average difference of temperature between that given point and the surrounding material". In practice, this leads to a condition as follows, expressing the change of the temperature φ, over a small period of time $\delta > 0$:

$$\varphi(x,y,t+\delta) = \varphi(x,y,t) + \frac{\alpha\delta}{l^2} \sum_{(x,y)\sim(u,v)} [\varphi(u,v,t) - \varphi(x,y,t)]$$

In fact, we can rewrite our equation as follows, making it clear that we have here an equation regarding the rate of change of temperature at x:

$$\frac{\varphi(x,y,t+\delta) - \varphi(x,y,t)}{\delta} = \frac{\alpha}{l^2} \sum_{(x,y)\sim(u,v)} [\varphi(u,v,t) - \varphi(x,y,t)]$$

(11) So, let us do the math. In the context of our 2D model the neighbors of x are the points $(x \pm l, y \pm l)$, so the equation above takes the following form:

$$\begin{aligned}
& \frac{\varphi(x,y,t+\delta) - \varphi(x,y,t)}{\delta} \\
= \ & \frac{\alpha}{l^2}\Big[(\varphi(x+l,y,t) - \varphi(x,y,t)) + (\varphi(x-l,y,t) - \varphi(x,y,t))\Big] \\
+ \ & \frac{\alpha}{l^2}\Big[(\varphi(x,y+l,t) - \varphi(x,y,t)) + (\varphi(x,y-l,t) - \varphi(x,y,t))\Big]
\end{aligned}$$

Now observe that we can write this equation as follows:

$$\begin{aligned}\frac{\varphi(x,y,t+\delta)-\varphi(x,y,t)}{\delta} &= \alpha\cdot\frac{\varphi(x+l,y,t)-2\varphi(x,y,t)+\varphi(x-l,y,t)}{l^2}\\ &+ \alpha\cdot\frac{\varphi(x,y+l,t)-2\varphi(x,y,t)+\varphi(x,y-l,t)}{l^2}\end{aligned}$$

(12) As it was the case before in one dimension, we recognize on the right the usual approximation of the second derivative, coming from calculus. Thus, when taking the continuous limit of our model, $l\to 0$, we obtain the following equation:

$$\frac{\varphi(x,y,t+\delta)-\varphi(x,y,t)}{\delta}=\alpha\left(\frac{d^2\varphi}{dx^2}+\frac{d^2\varphi}{dy^2}\right)(x,y,t)$$

Now with $t\to 0$, we are led in this way to the heat equation, namely:

$$\dot{\varphi}(x,y,t)=\alpha\cdot\Delta\varphi(x,y,t)$$

Finally, in arbitrary N dimensions the same argument carries over, namely a straightforward lattice model, and gives the heat equation, as formulated in the statement. □

Observe that we can use if we want different lengths $l>0$ on the vertical and on the horizontal, because these will simplify anyway due to proportionality. Also, for some further mathematical fun, we can build our model on a cylinder, or a torus.

Also, as mentioned before, our heat equation above is something approximative, and several improvements can be made to it, first by incorporating a term accounting for heat radiation, and also by doing several fine-tunings, depending on the material involved. Some of these improvements can be implemented in the lattice model setting.

11d. Into the heat

Let us go back now to the heat equation, and try to solve it. To start with, as a result often used by mathematicians, as to assume $\alpha=1$ for their mathematics, we have:

PROPOSITION 11.15. *Up to a time rescaling, we can assume $\alpha=1$, as to deal with*

$$\dot{\varphi}=\Delta\varphi$$

called normalized heat equation.

PROOF. This is clear physically speaking, because according to our model, changing the parameter $\alpha>0$ will result in accelerating or slowing the heat diffusion, in time $t>0$. Mathematically, this follows via a change of variables, for the time variable t. □

Regarding now the resolution of the heat equation, we have here:

THEOREM 11.16. *The heat equation, normalized as $\dot{\varphi} = \Delta\varphi$, and with initial condition $\varphi(x,0) = f(x)$, has as solution the function*

$$\varphi(x,t) = (K_t * f)(x)$$

where the function $K_t : \mathbb{R}^N \to \mathbb{R}$, called heat kernel, is given by

$$K_t(x) = (4\pi t)^{-N/2} e^{-||x||^2/4t}$$

with $||x||$ being the usual norm of vectors $x \in \mathbb{R}^N$.

PROOF. According to the definition of the convolution operation $*$, we have to check that the following function satisfies $\dot{\varphi} = \Delta\varphi$, with initial condition $\varphi(x,0) = f(x)$:

$$\varphi(x,t) = (4\pi t)^{-N/2} \int_{\mathbb{R}^N} e^{-||x-y||^2/4t} f(y)dy$$

But both checks are elementary, coming from definitions. □

Getting back now to calculus, in order to get more advanced results about the heat equation, for instance regarding the case where we have a point heat source, we will need a tough piece of mathematics, namely the formula of Δ in spherical coordinates.

Ready for this? First, we have the following result, that you know well:

THEOREM 11.17. *We have spherical coordinates in 3 dimensions,*

$$\begin{cases} x &= r\cos s \\ y &= r\sin s\cos t \\ z &= r\sin s\sin t \end{cases}$$

the corresponding Jacobian being $J(r,s,t) = r^2\sin s$.

PROOF. The fact that we have indeed spherical coordinates is clear. Regarding now the Jacobian, this is given by the following formula:

$$\begin{aligned}
& J(r,s,t)\\
= & \begin{vmatrix}\cos s & -r\sin s & 0\\ \sin s\cos t & r\cos s\cos t & -r\sin s\sin t\\ \sin s\sin t & r\cos s\sin t & r\sin s\cos t\end{vmatrix}\\
= & r^2\sin s\sin t\begin{vmatrix}\cos s & -r\sin s\\ \sin s\sin t & r\cos s\sin t\end{vmatrix}+r\sin s\cos t\begin{vmatrix}\cos s & -r\sin s\\ \sin s\cos t & r\cos s\cos t\end{vmatrix}\\
= & r\sin s\sin^2 t\begin{vmatrix}\cos s & -r\sin s\\ \sin s & r\cos s\end{vmatrix}+r\sin s\cos^2 t\begin{vmatrix}\cos s & -r\sin s\\ \sin s & r\cos s\end{vmatrix}\\
= & r\sin s(\sin^2 t+\cos^2 t)\begin{vmatrix}\cos s & -r\sin s\\ \sin s & r\cos s\end{vmatrix}\\
= & r\sin s\times 1\times r\\
= & r^2\sin s
\end{aligned}$$

Thus, we have indeed the formula in the statement. □

We will need the formula of the Laplace operator Δ in spherical coordinates. The result here, and its proof, which are quite tricky, are as follows:

THEOREM 11.18. *The Laplace operator in spherical coordinates is*

$$\Delta=\frac{1}{r^2}\cdot\frac{d}{dr}\left(r^2\cdot\frac{d}{dr}\right)+\frac{1}{r^2\sin s}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d}{ds}\right)+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2}{dt^2}$$

with our conventions above for the spherical coordinates.

PROOF. There are several proofs here, a short, elementary one being as follows:

(1) Let us first see how Δ behaves under a change of coordinates $\{x_i\}\to\{y_i\}$, in arbitrary N dimensions. Our starting point is the chain rule for derivatives:

$$\frac{d}{dx_i}=\sum_j\frac{d}{dy_j}\cdot\frac{dy_j}{dx_i}$$

By using this rule, then Leibnitz for products, then again this rule, we obtain:

$$\begin{aligned}
\frac{d^2f}{dx_i^2} &= \sum_j \frac{d}{dx_i}\left(\frac{df}{dy_j}\cdot\frac{dy_j}{dx_i}\right)\\
&= \sum_j \frac{d}{dx_i}\left(\frac{df}{dy_j}\right)\cdot\frac{dy_j}{dx_i}+\frac{df}{dy_j}\cdot\frac{d}{dx_i}\left(\frac{dy_j}{dx_i}\right)\\
&= \sum_j\left(\sum_k \frac{d}{dy_k}\cdot\frac{dy_k}{dx_i}\right)\left(\frac{df}{dy_j}\right)\cdot\frac{dy_j}{dx_i}+\frac{df}{dy_j}\cdot\frac{d^2y_j}{dx_i^2}\\
&= \sum_{jk}\frac{d^2f}{dy_kdy_j}\cdot\frac{dy_k}{dx_i}\cdot\frac{dy_j}{dx_i}+\sum_j\frac{df}{dy_j}\cdot\frac{d^2y_j}{dx_i^2}
\end{aligned}$$

(2) Now by summing over i, we obtain the following formula, with A being the derivative of $x \to y$, that is to say, the matrix of partial derivatives dy_i/dx_j:

$$\begin{aligned}
\Delta f &= \sum_{ijk}\frac{d^2f}{dy_kdy_j}\cdot\frac{dy_k}{dx_i}\cdot\frac{dy_j}{dx_i}+\sum_{ij}\frac{df}{dy_j}\cdot\frac{d^2y_j}{dx_i^2}\\
&= \sum_{ijk}A_{ki}A_{ji}\frac{d^2f}{dy_kdy_j}+\sum_{ij}\frac{d^2y_j}{dx_i^2}\cdot\frac{df}{dy_j}\\
&= \sum_{jk}(AA^t)_{jk}\frac{d^2f}{dy_kdy_j}+\sum_j\Delta(y_j)\frac{df}{dy_j}
\end{aligned}$$

(3) So, this will be the formula that we will need. Observe that this formula can be further compacted as follows, with all the notations being self-explanatory:

$$\Delta f = Tr(AA^tH_y(f)) + <\Delta(y),\nabla_y(f)>$$

(4) Getting now to spherical coordinates, $(x,y,z)\to(r,s,t)$, the derivative of the inverse, obtained by differentiating x,y,z with respect to r,s,t, is given by:

$$A^{-1}=\begin{pmatrix}\cos s & -r\sin s & 0\\ \sin s\cos t & r\cos s\cos t & -r\sin s\sin t\\ \sin s\sin t & r\cos s\sin t & r\sin s\cos t\end{pmatrix}$$

The product $(A^{-1})^tA^{-1}$ of the transpose of this matrix with itself is then:

$$\begin{pmatrix}\cos s & \sin s\cos t & \sin s\sin t\\ -r\sin s & r\cos s\cos t & r\cos s\sin t\\ 0 & -r\sin s\sin t & r\sin s\cos t\end{pmatrix}\begin{pmatrix}\cos s & -r\sin s & 0\\ \sin s\cos t & r\cos s\cos t & -r\sin s\sin t\\ \sin s\sin t & r\cos s\sin t & r\sin s\cos t\end{pmatrix}$$

But everything simplifies here, and we have the following remarkable formula, which by the way is something very useful, worth to be memorized:

$$(A^{-1})^t A^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2 s \end{pmatrix}$$

Now by inverting, we obtain the following formula, in relation with the above:

$$AA^t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/r^2 & 0 \\ 0 & 0 & 1/(r^2 \sin^2 s) \end{pmatrix}$$

(5) Let us compute now the Laplacian of r, s, t. We first have the following formula, that we will use many times in what follows, and is worth to be memorized:

$$\begin{aligned} \frac{dr}{dx} &= \frac{d}{dx}\sqrt{x^2+y^2+z^2} \\ &= \frac{1}{2} \cdot \frac{2x}{\sqrt{x^2+y^2+z^2}} \\ &= \frac{x}{r} \end{aligned}$$

Of course the same computation works for y, z too, and we therefore have:

$$\frac{dr}{dx} = \frac{x}{r} \quad , \quad \frac{dr}{dy} = \frac{y}{r} \quad , \quad \frac{dr}{dz} = \frac{z}{r}$$

(6) By using the above formulae, twice, we can compute the Laplacian of r:

$$\begin{aligned} \Delta(r) &= \Delta\left(\sqrt{x^2+y^2+z^2}\right) \\ &= \frac{d}{dx}\left(\frac{x}{r}\right) + \frac{d}{dy}\left(\frac{y}{r}\right) + \frac{d}{dz}\left(\frac{z}{r}\right) \\ &= \frac{r^2-x^2}{r^3} + \frac{r^2-y^2}{r^3} + \frac{r^2-z^2}{r^3} \\ &= \frac{2}{r} \end{aligned}$$

(7) In what regards now s, the computation here goes as follows:

$$\begin{aligned}
\Delta(s) &= \Delta\left(\arccos\left(\frac{x}{r}\right)\right)\\
&= \frac{d}{dx}\left(-\frac{\sqrt{r^2-x^2}}{r^2}\right)+\frac{d}{dy}\left(\frac{xy}{r^2\sqrt{r^2-x^2}}\right)+\frac{d}{dz}\left(\frac{xz}{r^2\sqrt{r^2-x^2}}\right)\\
&= \frac{2x\sqrt{r^2-x^2}}{r^4}+\frac{r^2(z^2-2y^2)+2x^2y^2}{r^4\sqrt{r^2-x^2}}+\frac{r^2(y^2-2z^2)+2x^2z^2}{r^4\sqrt{r^2-x^2}}\\
&= \frac{2x\sqrt{r^2-x^2}}{r^4}+\frac{x(2x^2-r^2)}{r^4\sqrt{r^2-x^2}}\\
&= \frac{x}{r^2\sqrt{r^2-x^2}}\\
&= \frac{\cos s}{r^2\sin s}
\end{aligned}$$

(8) Finally, in what regards t, the computation here goes as follows:

$$\begin{aligned}
\Delta(t) &= \Delta\left(\arctan\left(\frac{z}{y}\right)\right)\\
&= \frac{d}{dx}(0)+\frac{d}{dy}\left(-\frac{z}{y^2+z^2}\right)+\frac{d}{dz}\left(\frac{y}{y^2+z^2}\right)\\
&= 0-\frac{2yz}{(y^2+z^2)^2}+\frac{2yz}{(y^2+z^2)^2}\\
&= 0
\end{aligned}$$

(9) We can now plug the data from (4) and (6,7,8) in the general formula that we found in (2) above, and we obtain in this way:

$$\begin{aligned}
\Delta f &= \frac{d^2f}{dr^2}+\frac{1}{r^2}\cdot\frac{d^2f}{ds^2}+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2f}{dt^2}+\frac{2}{r}\cdot\frac{df}{dr}+\frac{\cos s}{r^2\sin s}\cdot\frac{df}{ds}\\
&= \frac{2}{r}\cdot\frac{df}{dr}+\frac{d^2f}{dr^2}+\frac{\cos s}{r^2\sin s}\cdot\frac{df}{ds}+\frac{1}{r^2}\cdot\frac{d^2f}{ds^2}+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2f}{dt^2}\\
&= \frac{1}{r^2}\cdot\frac{d}{dr}\left(r^2\cdot\frac{df}{dr}\right)+\frac{1}{r^2\sin s}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{df}{ds}\right)+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2f}{dt^2}
\end{aligned}$$

Thus, we are led to the formula in the statement. □

Regarding now our discretization questions, things here are quite tricky. In relation with Theorem 11.16, and with the heat kernel, the first thought towards discretization goes to the Central Limit Theorem (CLT) from probability theory, which produces the normal laws, in dimension $N=1$, but also in general, in arbitrary $N\geq 1$ dimensions.

11e. Exercises

Exercises:

Exercise 11.19.

Exercise 11.20.

Exercise 11.21.

Exercise 11.22.

Exercise 11.23.

Exercise 11.24.

Exercise 11.25.

Exercise 11.26.

Bonus exercise.

CHAPTER 12

Into manifolds

12a. Into manifolds

12b.

12c.

12d.

12e. Exercises

Exercises:

Exercise 12.1.

Exercise 12.2.

Exercise 12.3.

Exercise 12.4.

Exercise 12.5.

Exercise 12.6.

Exercise 12.7.

Exercise 12.8.

Bonus exercise.

Part IV

Quantum versions

CHAPTER 13

Free probability

13a. Freeness

Welcome to free probability. We have met some already, and in this chapter and in the next three ones we discuss the foundations and main results of free probability, in analogy with the foundations and main results of classical probability.

The common framework for classical and free probability is "noncommutative probability". This is something very general. Let us start with the following definition:

DEFINITION 13.1. *A C^*-algebra is a complex algebra A, having a norm $||.||$ making it a Banach algebra, and an involution $*$, related to the norm by the formula*

$$||aa^*|| = ||a||^2$$

which must hold for any $a \in A$.

As a basic example, the algebra $B(H)$ of the bounded linear operators $T : H \to H$ on a complex Hilbert space H is a C^*-algebra, with the usual norm and involution:

$$||T|| = \sup_{||x||=1} ||Tx|| \quad , \quad < Tx, y >=< x, T^*y >$$

More generally, any closed $*$-subalgebra of $B(H)$ is a C^*-algebra. It is possible to prove that any C^*-algebra appears in this way, as follows:

$$A \subset B(H)$$

In finite dimensions we have $H = \mathbb{C}^N$, and so the operator algebra $B(H)$ is the usual matrix algebra $M_N(\mathbb{C})$, with the usual norm and involution, namely:

$$||M|| = \sup_{||x||=1} ||Mx|| \quad , \quad (M^*)_{ij} = \bar{M}_{ji}$$

As explained in chapter 4, in the context of Peter-Weyl theory, some algebra shows that the finite dimensional C^*-algebras are the direct sums of matrix algebras:

$$A = M_{n_1}(\mathbb{C}) \oplus \ldots \oplus M_{n_k}(\mathbb{C})$$

Summarizing, the C^*-algebra formalism is something in between the $*$-algebras, which are purely algebraic objects, and whose theory basically leads nowhere, and the fully advanced operator algebras, which are the von Neumann algebras. More on this later.

As yet another class of examples now, which are of particular importance for us, we have various algebras of functions $f : X \to \mathbb{C}$. The theory here is as follows:

THEOREM 13.2. *The commutative C^*-algebras are the algebras of type $C(X)$, with X being a compact space, the correspondence being as follows:*

(1) *Given a compact space X, the algebra $C(X)$ of continuous functions $f : X \to \mathbb{C}$ is a commutative C^*-algebra, with norm and involution as follows:*

$$||f|| = \sup_{x \in X} |f(x)| \quad , \quad f^*(x) = \overline{f(x)}$$

(2) *Conversely, any commutative C^*-algebra can be written as $A = C(X)$, with its "spectrum" appearing as the space of Banach algebra characters of A:*

$$X = \left\{\chi : A \to \mathbb{C}\right\}$$

In view of this, given an arbitrary C^*-algebra A, not necessarily commutative, we agree to write $A = C(X)$, and call the abstract space X a compact quantum space.

PROOF. This is something that we know from before, the idea being as follows:

(1) First of all, the fact that $C(X)$ is a Banach algebra is clear, because a uniform limit of continuous functions must be continuous. As for the formula $||ff^*|| = ||f||^2$, this is something trivial for functions, because on both sides we obtain $\sup_{x \in X} |f(x)|^2$.

(2) Given a commutative C^*-algebra A, the character space $X = \{\chi : A \to \mathbb{C}\}$ is indeed compact, and we have an evaluation morphism $ev : A \to C(X)$. The tricky point, which follows from basic spectral theory, is to prove that ev is indeed isometric. □

The above result is quite interesting for us, because it allows one to formally write any C^*-algebra as $A = C(X)$, with X being a noncommutative compact space. This is certainly something very nice, and in order to do now some probability theory over such spaces X, we would need probability measures μ. But, the problem is that these measures μ are impossible to define, because our spaces X have no points in general.

However, we can trick, and do probability theory just by using expectations functionals $E : A \to \mathbb{C}$, instead of the probability measures μ themselves. These expectations are called traces, are are denoted $tr : A \to \mathbb{C}$, and their axiomatization is as follows:

DEFINITION 13.3. *A trace, or expectation, or integration functional, on a C^*-algebra A is a linear form $tr : A \to \mathbb{C}$ having the following properties:*

(1) *tr is unital, and continuous.*
(2) *tr is positive, $a \geq 0 \implies \varphi(a) \geq 0$.*
(3) *tr has the trace property $tr(ab) = tr(ba)$.*

We call tr faithful when $a > 0 \implies \varphi(a) > 0$.

In the commutative case, $A = C(X)$, the Riesz theorem shows that the positive traces $tr : A \to \mathbb{C}$ appear as integration functionals with respect to positive measures μ:

$$tr(f) = \int_X f(x)\, d\mu(x)$$

Moreover, the unitality of tr corresponds to the fact that μ has mass one, and the faithfulness of tr corresponds to the faithfulness of μ. Thus, in general, when A is no longer commutative, in order to do probability theory on the underlying noncommutative compact space X, what we need is a faithful trace $tr : A \to \mathbb{C}$ as above.

So, this will be our philosophy in what follows, a noncommutative probability space (X, μ) being something abstract, corresponding in practice to a pair (A, tr). This is of course something a bit simplified, because associated to any space X, noncommutative or even classical, there are in fact many possible C^*-algebras of functions $f : X \to \mathbb{C}$, such as $C(X)$, $L^\infty(X)$ and so on, and for a better theory, we would have to make a choice between these various C^*-algebras associated to X. But let us not worry with this for the moment, what we have is good for starting some computations, so let us just do these computations, see what we get, and we will come back later to more about formalism.

Going ahead with definitions, everything in what follows will be based on:

Definition 13.4. *Let A be a C^*-algebra, given with a trace $tr : A \to \mathbb{C}$.*

1. *The elements $a \in A$ are called random variables.*
2. *The moments of such a variable are the numbers $M_k(a) = tr(a^k)$.*
3. *The law of such a variable is the functional $\mu : P \to tr(P(a))$.*

Here $k = \circ \bullet \bullet \circ \ldots$ is by definition a colored integer, and the corresponding powers a^k are defined by the following formulae, and multiplicativity:

$$a^\emptyset = 1 \quad , \quad a^\circ = a \quad , \quad a^\bullet = a^*$$

As for the polynomial P, this is a noncommuting $*$-polynomial in one variable:

$$P \in \mathbb{C} < X, X^* >$$

Observe that the law is uniquely determined by the moments, because we have:

$$P(X) = \sum_k \lambda_k X^k \implies \mu(P) = \sum_k \lambda_k M_k(a)$$

Generally speaking, the above definition is something quite abstract, but there is no other way of doing things, at least at this level of generality. However, in certain special cases, the formalism simplifies, and we recover more familiar objects, as follows:

THEOREM 13.5. *Assuming that $a\in A$ is normal, $aa^*=a^*a$, its law corresponds to a probability measure on its spectrum $\sigma(a)\subset\mathbb{C}$, according to the following formula:*

$$tr(P(a))=\int_{\sigma(a)}P(x)d\mu(x)$$

When the trace is faithful we have $supp(\mu)=\sigma(a)$. Also, in the particular case where the variable is self-adjoint, $a=a^*$, this law is a real probability measure.

PROOF. This is something very standard, coming from the continuous functional calculus for the C^*-algebras. In fact, we can deduce from this that more is true, in the sense that the following formula holds, for any $f\in C(\sigma(a))$:

$$tr(f(a))=\int_{\sigma(a)}f(x)d\mu(x)$$

In addition, assuming that we are in the case $A\subset B(H)$, the measurable functional calculus tells us that the above formula holds in fact for any $f\in L^\infty(\sigma(a))$. □

We have the following independence notion, generalizing the one from chapter 1:

DEFINITION 13.6. *Two subalgebras $A,B\subset C$ are called independent when the following condition is satisfied, for any $a\in A$ and $b\in B$:*

$$tr(ab)=tr(a)tr(b)$$

Equivalently, the following condition must be satisfied, for any $a\in A$ and $b\in B$:

$$tr(a)=tr(b)=0\implies tr(ab)=0$$

Also, two variables $a,b\in C$ are called independent when the algebras that they generate,

$$A=<a>\quad,\quad B=<b>$$

are independent inside C, in the above sense.

Observe that the above two independence conditions are indeed equivalent, with this following from the following computation, with the convention $a'=a-tr(a)$:

$$\begin{aligned}tr(ab)&=tr[(a'+tr(a))(b'+tr(b))]\\&=tr(a'b')+t(a')tr(b)+tr(a)tr(b')+tr(a)tr(b)\\&=tr(a'b')+tr(a)tr(b)\\&=tr(a)tr(b)\end{aligned}$$

The other remark is that the above notion generalizes indeed the usual notion of independence, from the classical case, the precise result here being as follows:

THEOREM 13.7. *Given two compact measured spaces X,Y, the algebras*

$$C(X)\subset C(X\times Y)\quad,\quad C(Y)\subset C(X\times Y)$$

are independent in the above sense, and a converse of this fact holds too.

PROOF. We have two assertions here, the idea being as follows:

(1) First of all, given two abstract compact spaces X, Y, we have embeddings of algebras as in the statement, defined by the following formulae:

$$f \to [(x,y) \to f(x)] \quad , \quad g \to [(x,y) \to g(y)]$$

In the measured space case now, the Fubini theorems tells us that we have:

$$\int_{X \times Y} f(x)g(y) = \int_X f(x) \int_Y g(y)$$

Thus, the algebras $C(X), C(Y)$ are independent in the sense of Definition 13.6.

(2) Conversely, assume that $A, B \subset C$ are independent, with C being commutative. Let us write our algebras as follows, with X, Y, Z being certain compact spaces:

$$A = C(X) \quad , \quad B = C(Y) \quad , \quad C = C(Z)$$

In this picture, the inclusions $A, B \subset C$ must come from quotient maps, as follows:

$$p : Z \to X \quad , \quad q : Z \to Y$$

Regarding now the independence condition from Definition 13.6, in the above picture, this tells us that the following equality must happen:

$$\int_Z f(p(z))g(q(z)) = \int_Z f(p(z)) \int_X g(q(z))$$

Thus we are in a Fubini type situation, and we obtain from this:

$$X \times Y \subset Z$$

Thus, the independence of the algebras $A, B \subset C$ appears as in (1) above. □

It is possible to develop some theory here, but this is ultimately not very interesting. As a much more interesting notion now, we have Voiculescu's freeness [89]:

DEFINITION 13.8. *Two subalgebras $A, B \subset C$ are called free when the following condition is satisfied, for any $a_i \in A$ and $b_i \in B$:*

$$tr(a_i) = tr(b_i) = 0 \implies tr(a_1 b_1 a_2 b_2 \ldots) = 0$$

Also, two variables $a, b \in C$ are called free when the algebras that they generate,

$$A = < a > \quad , \quad B = < b >$$

are free inside C, in the above sense.

In short, freeness appears by definition as a kind of "free analogue" of usual independence, taking into account the fact that the variables do not necessarily commute. As a first observation, of theoretical nature, there is actually a certain lack of symmetry between Definition 13.6 and Definition 13.8, because in contrast to the former, the latter does not include an explicit formula for the quantities of the following type:

$$tr(a_1b_1a_2b_2\ldots)$$

However, this is not an issue, and is simply due to the fact that the formula in the free case is something more complicated, the precise result being as follows:

PROPOSITION 13.9. *Assuming that $A, B\subset C$ are free, the restriction of tr to $< A, B >$ can be computed in terms of the restrictions of tr to A, B. To be more precise,*

$$tr(a_1b_1a_2b_2\ldots)=P\Big(\{tr(a_{i_1}a_{i_2}\ldots)\}_i,\{tr(b_{j_1}b_{j_2}\ldots)\}_j\Big)$$

where P is certain polynomial in several variables, depending on the length of the word $a_1b_1a_2b_2\ldots$, and having as variables the traces of products of type

$$a_{i_1}a_{i_2}\ldots\quad,\quad b_{j_1}b_{j_2}\ldots$$

with the indices being chosen increasing, $i_1<i_2<\ldots$ and $j_1<j_2<\ldots$

PROOF. This is something a bit theoretical, so let us begin with an example. Our claim is that if a, b are free then, exactly as in the case where we have independence:

$$tr(ab)=tr(a)tr(b)$$

Indeed, let us go back to the computation performed after Definition 13.6, which was as follows, with the convention $a'=a-tr(a)$:

$$\begin{aligned}tr(ab)&=tr[(a'+tr(a))(b'+tr(b))]\\&=tr(a'b')+t(a')tr(b)+tr(a)tr(b')+tr(a)tr(b)\\&=tr(a'b')+tr(a)tr(b)\\&=tr(a)tr(b)\end{aligned}$$

Our claim is that this computation perfectly works under the sole freeness assumption. Indeed, the only non-trivial equality is the last one, which follows from:

$$tr(a')=tr(b')=0\implies tr(a'b')=0$$

In general, the situation is of course more complicated than this, but the same trick applies. To be more precise, we can start our computation as follows:

$$\begin{aligned}tr(a_1b_1a_2b_2\ldots)&=tr\Big[(a_1'+tr(a_1))(b_1'+tr(b_1))(a_2'+tr(a_2))(b_2'+tr(b_2))\ldots\ldots\Big]\\&=tr(a_1'b_1'a_2'b_2'\ldots)+\text{other terms}\\&=\text{other terms}\end{aligned}$$

Observe that we have used here the freeness condition, in the following form:

$$tr(a_i')=tr(b_i')=0\implies tr(a_1'b_1'a_2'b_2'\ldots)=0$$

Now regarding the "other terms", those which are left, each of them will consist of a product of traces of type $tr(a_i)$ and $tr(b_i)$, and then a trace of a product still remaining to be computed, which is of the following form, for some elements $\alpha_i \in A$ and $\beta_i \in B$:

$$tr(\alpha_1\beta_1\alpha_2\beta_2\ldots)$$

To be more precise, the variables $\alpha_i \in A$ appear as ordered products of those $a_i \in A$ not getting into individual traces $tr(a_i)$, and the variables $\beta_i \in B$ appear as ordered products of those $b_i \in B$ not getting into individual traces $tr(b_i)$. Now since the length of each such alternating product $\alpha_1\beta_1\alpha_2\beta_2\ldots$ is smaller than the length of the original product $a_1b_1a_2b_2\ldots$, we are led into of recurrence, and this gives the result. □

Let us discuss now some models for independence and freeness. We have the following result, from [89], which clarifies the analogy between independence and freeness:

THEOREM 13.10. *Given two algebras (A, tr) and (B, tr), the following hold:*

1. *A, B are independent inside their tensor product $A\otimes B$, endowed with its canonical tensor product trace, given by $tr(a\otimes b) = tr(a)tr(b)$.*
2. *A, B are free inside their free product $A * B$, endowed with its canonical free product trace, given by the formulae in Proposition 13.9.*

PROOF. Both the above assertions are clear from definitions, as follows:

(1) This is clear with either of the definitions of the independence, from Definition 13.6, because we have by construction of the product trace:

$$\begin{aligned} tr(ab) &= tr[(a\otimes 1)(1\otimes b)] \\ &= tr(a\otimes b) \\ &= tr(a)tr(b) \end{aligned}$$

Observe that there is a relation here with Theorem 13.7 as well, due to the following formula for compact spaces, with $\otimes$ being a topological tensor product:

$$C(X\times Y) = C(X)\otimes C(Y)$$

To be more precise, the present statement generalizes the first assertion in Theorem 13.7, and the second assertion tells us that this generalization is more or less the same thing as the original statement. All this comes of course from basic measure theory.

(2) This is clear too from definitions, the only point being that of showing that the notion of freeness, or the recurrence formulae in Proposition 13.9, can be used in order to construct a canonical free product trace, on the free product of the algebras involved:

$$tr : A * B \to \mathbb{C}$$

But this can be checked for instance by using a GNS construction. Indeed, consider the GNS constructions for the algebras (A, tr) and (B, tr):

$$A \to B(l^2(A)) \quad , \quad B \to B(l^2(B))$$

By taking the free product of these representations, we obtain a representation as follows, with the $*$ on the right being a free product of pointed Hilbert spaces:

$$A * B \to B(l^2(A) * l^2(B))$$

Now by composing with the linear form $T \to < T\xi, \xi >$, where $\xi = 1_A = 1_B$ is the common distinguished vector of $l^2(A)$, $l^2(B)$, we obtain a linear form, as follows:

$$tr : A * B \to \mathbb{C}$$

It is routine then to check that tr is indeed a trace, and this is the "canonical free product trace" from the statement. Then, an elementary computation shows that A, B are free inside $A * B$, with respect to this trace, and this finishes the proof. See [89]. □

13b. Free convolution

All the above was quite theoretical, and as a concrete application of the above results, bringing us into probability, we have the following result, from [90]:

THEOREM 13.11. *We have a free convolution operation* $\boxplus$ *for the distributions*

$$\mu : \mathbb{C} < X, X^* > \to \mathbb{C}$$

which is well-defined by the following formula, with a, b *taken to be free:*

$$\mu_a \boxplus \mu_b = \mu_{a+b}$$

This restricts to an operation, still denoted $\boxplus$*, on the real probability measures.*

PROOF. We have several verifications to be performed here, as follows:

(1) We first have to check that given two variables a, b which live respectively in certain C^*-algebras A, B, we can recover inside some C^*-algebra C, with exactly the same distributions μ_a, μ_b, as to be able to sum them and talk about μ_{a+b}. But this comes from Theorem 13.10, because we can set $C = A * B$, as explained there.

(2) The other verification which is needed is that of the fact that if two variables a, b are free, then the distribution μ_{a+b} depends only on the distributions μ_a, μ_b. But for this purpose, we can use the general formula from Proposition 13.9, namely:

$$tr(a_1 b_1 a_2 b_2 \ldots) = P\Big(\{tr(a_{i_1} a_{i_2} \ldots)\}_i, \{tr(b_{j_1} b_{j_2} \ldots)\}_j\Big)$$

Now by plugging in arbitrary powers of a, b as variables a_i, b_j, we obtain a family of formulae of the following type, with Q being certain polyomials:

$$tr(a^{k_1} b^{l_1} a^{k_2} b^{l_2} \ldots) = Q\Big(\{tr(a^k)\}_k, \{tr(b^l)\}_l\Big)$$

Thus the moments of $a + b$ depend only on the moments of a, b, with of course colored exponents in all this, according to our moment conventions, and this gives the result.

(3) Finally, in what regards the last assertion, regarding the real measures, this is clear from the fact that if the variables a,b are self-adjoint, then so is their sum $a+b$. $\square$

Along the same lines, but with some technical subtleties this time, we can talk as well about multiplicative free convolution, following [91], as follows:

THEOREM 13.12. *We have a free convolution operation $\boxtimes$ for the distributions*

$$\mu:\mathbb{C}<X,X^*>\to\mathbb{C}$$

which is well-defined by the following formula, with a,b taken to be free:

$$\mu_a\boxtimes\mu_b=\mu_{ab}$$

In the case of the self-adjoint variables, we can equally set

$$\mu_a\boxtimes\mu_b=\mu_{\sqrt{a}b\sqrt{a}}$$

and so we have an operation, still denoted $\boxtimes$, on the real probability measures.

PROOF. We have two statements here, the idea being as follows:

(1) The verifications for the fact that $\boxtimes$ as above is indeed well-defined at the general distribution level are identical to those done before for $\boxplus$, with the result basically coming from the formula in Proposition 13.9, and with Theorem 13.10 invoked as well, in order to say that we have a model, and so we can indeed use this formula.

(2) Regarding now the last assertion, regarding the real measures, this was something trivial for $\boxplus$, but is something trickier now for $\boxtimes$, because if we take a,b to be self-adjoint, thier product ab will in general not be self-adjoint, and definitely it will be not if we want a,b to be free, and so the formula $\mu_a\boxtimes\mu_b=\mu_{ab}$ will apparently makes us exit the world of real probability measures. However, this is not exactly the case. Indeed, let us set:

$$c=\sqrt{a}b\sqrt{a}$$

This new variable is then self-adjoint, and its moments are given by:

$$\begin{aligned}
tr(c^k) &= tr[(\sqrt{a}b\sqrt{a})^k]\\
&= tr[\sqrt{a}ba\ldots ab\sqrt{a}]\\
&= tr[\sqrt{a}\cdot\sqrt{a}ba\ldots ab]\\
&= tr[(ab)^k]
\end{aligned}$$

Thus, we are led to the conclusion in the statement. $\square$

We would like now to have linearization results for $\boxplus$ and $\boxtimes$, in the spirit of the known results for $*$ and $\times$. We will do this slowly, in several steps. As a first objective, we would like to convert our one and only modeling result so far, namely Theorem 13.10, which is a rather abstract result, into something more concrete. Let us start with:

Theorem 13.13. *Let Γ be a discrete group, and consider the complex group algebra $\mathbb{C}[\Gamma]$, with involution given by the fact that all group elements are unitaries:*

$$g^*=g^{-1}\quad,\quad\forall g\in\Gamma$$

The maximal C^*-seminorm on $\mathbb{C}[\Gamma]$ is then a C^*-norm, and the closure of $\mathbb{C}[\Gamma]$ with respect to this norm is a C^*-algebra, denoted $C^*(\Gamma)$. Moreover,

$$tr(g)=\delta_{g1}$$

defines a positive unital trace $tr:C^*(\Gamma)\to\mathbb{C}$, which is faithful on $\mathbb{C}[\Gamma]$.

Proof. We have two assertions to be proved, the idea being as follows:

(1) In order to prove the first assertion, regarding the maximal seminorm which is a norm, we must find a $*$-algebra embedding as follows, with H being a Hilbert space:

$$\mathbb{C}[\Gamma]\subset B(H)$$

For this purpose, consider the Hilbert space $H=l^2(\Gamma)$, having the family $\{h\}_{h\in\Gamma}$ as orthonormal basis. Our claim is that we have an embedding, as follows:

$$\pi:\mathbb{C}[\Gamma]\subset B(H)\quad,\quad\pi(g)(h)=gh$$

Indeed, since $\pi(g)$ maps the basis $\{h\}_{h\in\Gamma}$ into itself, this operator is well-defined and bounded, and is an isometry. It is also clear from the formula $\pi(g)(h)=gh$ that $g\to\pi(g)$ is a morphism of algebras, and since this morphism maps the unitaries $g\in\Gamma$ into isometries, this is a morphism of $*$-algebras. Finally, the faithfulness of π is clear.

(2) Regarding the second assertion, we can use here once again the above construction. Indeed, we can define a linear form on the image of $C^*(\Gamma)$, as follows:

$$tr(T)=<T\delta_1,\delta_1>$$

This functional is then positive, and is easily seen to be a trace. Moreover, on the group elements $g\in\Gamma$, this functional is given by the following formula:

$$tr(g)=\delta_{g1}$$

Thus, it remains to show that tr is faithful on $\mathbb{C}[\Gamma]$. But this follows from the fact that tr is faithful on the image of $C^*(\Gamma)$, which contains $\mathbb{C}[\Gamma]$. □

As an illustration, we have the following more precise result, in the abelian case:

Proposition 13.14. *Given a discrete abelian group Γ, we have an isomorphism*

$$C^*(\Gamma)\simeq C(G)$$

where $G=\widehat{\Gamma}$ is its Pontrjagin dual, formed by the characters $\chi:\Gamma\to\mathbb{T}$. Moreover,

$$tr(g)=\delta_{g1}$$

corresponds in this way to the Haar integration over G.

PROOF. We have two assertions to be proved, the idea being as follows:

(1) Since Γ is abelian, $A = C^*(\Gamma)$ is commutative, so by the Gelfand theorem we have $A = C(X)$. The spectrum $X = Spec(A)$, consisting of the characters $\chi : C^*(\Gamma) \to \mathbb{C}$, can be then identified with the Pontrjagin dual $G = \widehat{\Gamma}$, and this gives the result.

(2) Regarding now the last assertion, we must prove here that we have:

$$tr(f) = \int_G f(x)dx$$

But this is clear via the above identifications, for instance because the linear form $tr(g) = \delta_{g1}$, when viewed as a functional on $C(G)$, is left and right invariant. □

Getting back now to our questions, we can now formulate a general modelling result for independence and freeness, providing us with large classes of examples, as follows:

THEOREM 13.15. *We have the following results, valid for group algebras:*

(1) $C^*(\Gamma), C^*(\Lambda)$ *are independent inside* $C^*(\Gamma \times \Lambda)$.
(2) $C^*(\Gamma), C^*(\Lambda)$ *are free inside* $C^*(\Gamma * \Lambda)$.

PROOF. In order to prove these results, we have two possible methods:

(1) We can either use the general results in Theorem 13.10, along with the following two isomorphisms, which are both standard:

$$C^*(\Gamma \times \Lambda) = C^*(\Lambda) \otimes C^*(\Gamma) \quad , \quad C^*(\Gamma * \Lambda) = C^*(\Lambda) * C^*(\Gamma)$$

(2) Or, we can prove this directly, by using the fact that each algebra is spanned by the corresponding group elements. Indeed, this shows that it is enough to check the independence and freeness formulae on group elements, which is in turn trivial. □

13c. Linearization

We have seen so far the foundations of free probability, in analogy with those of classical probability, taken with a functional analysis touch. The idea now is that with a bit of luck, the basic theory from the classical case, namely the Fourier transform, and then the CLT, should have free extensions. Let us being our discussion with the following definition, from [90], coming from the theory developed in the above:

DEFINITION 13.16. *The real probability measures are subject to operations $*$ and $\boxplus$, called classical and free convolution, given by the formulae*

$$\mu_a * \mu_b = \mu_{a+b} \quad , \quad \mu_\alpha \boxplus \mu_\beta = \mu_{\alpha+\beta}$$

with a, b being independent, and α, β being free, and all variables being self-adjoint.

The problem now is that of linearizing these operations $*$ and $\boxplus$. In what regards $*$, we know from chapter 1 that this operation is linearized by the logarithm $\log F$ of the Fourier transform, which in the present setting, where $E = tr$, is given by:

$$F_a(x) = tr(e^{ixa})$$

In order to find a similar result for $\boxplus$, we need some efficient models for the pairs of free random variables (a, b). This is a priori not a problem, because once we have $a \in A$ and $b \in B$, we can form the free product $A * B$, which contains a, b as free variables.

However, the initial choice, that of the variables $a \in A$, $b \in B$ modeling some given laws $\mu, \nu \in \mathcal{P}(\mathbb{R})$, matters a lot. Indeed, any kind of abstract choice here would lead us into an abstract algebra $A * B$, and so into the abstract combinatorics of the free convolution, that cannot be solved with bare hands, and that we want to avoid.

In short, we must be tricky, at least in what concerns the beginning of our computation. Following [**90**], the idea will be that of temporarily lifting the self-adjointness assumption on our variables a, b, and looking instead for random variables α, β, not necessarily self-adjoint, modelling in integer moments our given laws $\mu, \nu \in \mathcal{P}(\mathbb{R})$, as follows:

$$tr(\alpha^k) = M_k(\mu) \quad , \quad tr(\beta^k) = M_k(\nu)$$

To be more precise, assuming that α, β are indeed not self-adjoint, the above formulae are not the general formulae for α, β, simply because these latter formulae involve colored integers $k = \circ \bullet \bullet \circ \ldots$ as exponents. Thus, in the context of the above formulae, μ, ν are not the distributions of α, β, but just some "parts" of these distributions.

Now with this idea in mind, due to Voiculescu and quite tricky, the solution to the law modelling problem comes in a quite straightforward way, involving the good old Hilbert space $H = l^2(\mathbb{N})$ and the good old shift operator $S \in B(H)$, as follows:

THEOREM 13.17. *Consider the shift operator on the space $H = l^2(\mathbb{N})$, given by $S(e_i) = e_{i+1}$. The variables of the following type, with $f \in \mathbb{C}[X]$ being a polynomial,*

$$S^* + f(S)$$

model then in moments, up to finite order, all the distributions $\mu : \mathbb{C}[X] \to \mathbb{C}$.

PROOF. We have already met the shift S before, as the simplest example of an isometry which is not a unitary, $S^*S = 1, SS^* = 1$, with this coming from:

$$S^*(e_i) = \begin{cases} e_{i-1} & (i > 0) \\ 0 & (i = 0) \end{cases}$$

Consider now a variable as in the statement, namely:

$$T = S^* + a_0 + a_1 S + a_2 S^2 + \ldots + a_n S^n$$

The computation of the moments of T is then as follows:

– We first have $tr(T) = a_0$.

– Then the computation of $tr(T^2)$ will involve a_1.

– Then the computation of $tr(T^3)$ will involve a_2.

– And so on.

Thus, we are led to a certain recurrence, that we will not attempt to solve now, with bare hands, but which definitely gives the conclusion in the statement. □

Before getting further, with free products of such models, let us work out a very basic example, which is something fundamental, that we will need in what follows:

PROPOSITION 13.18. *In the context of the above correspondence, the variable*

$$T = S + S^*$$

follows the Wigner semicircle law, $\gamma_1 = \frac{1}{2\pi}\sqrt{4-x^2}dx$.

PROOF. In order to compute the law of variable T in the statement, we can use the moment method. The moments of this variable are as follows:

$$\begin{aligned} M_k &= tr(T^k) \\ &= tr((S+S^*)^k) \\ &= \#(1 \in (S+S^*)^k) \end{aligned}$$

Now since the S shifts to the right on $\mathbb{N}$, and S^* shifts to the left, while remaining positive, we are left with counting the length k paths on $\mathbb{N}$ starting and ending at 0. Since there are no such paths when $k = 2r+1$ is odd, the odd moments vanish:

$$M_{2r+1} = 0$$

In the case where $k = 2r$ is even, such paths on $\mathbb{N}$ are best represented as paths in the upper half-plane, starting at 0, and going at each step NE or SE, depending on whether the original path on $\mathbb{N}$ goes at right or left, and finally ending at $k \in \mathbb{N}$. With this picture we are led to the following formula for the number of such paths:

$$M_{2r+2} = \sum_s M_{2s} M_{2r-s}$$

But this is exactly the recurrence formula for the Catalan numbers, and so:

$$M_{2r} = \frac{1}{r+1}\binom{2r}{r}$$

Summarizing, the odd moments of T vanish, and the even moments are the Catalan numbers. But these numbers being the moments of the Wigner semicircle law γ_1, as explained before, we are led to the conclusion in the statement. □

Getting back now to our linearization program for $\boxplus$, the next step is that of taking a free product of the model found in Theorem 13.17 with itself. There are two approaches here, one being a bit abstract, and the other one being more concrete. We will explain in what follows both of them. The abstract approach, which is quite nice, making a link with our main modeling result so far, involving group algebras, is as follows:

Proposition 13.19. *We can talk about semigroup algebras $C^*(\Gamma)\subset B(l^2(\Gamma))$, exactly as we did for the group algebras, and at the level of examples:*

(1) *With $\Gamma=\mathbb N$ we recover the shift algebra $A=<S>$ on $H=l^2(\mathbb N)$.*
(2) *With $\Gamma=\mathbb N*\mathbb N$, we obtain the algebra $A=<S_1,S_2>$ on $H=l^2(\mathbb N*\mathbb N)$.*

Proof. We can talk indeed about semigroup algebras $C^*(\Gamma)\subset B(l^2(\Gamma))$, exactly as we did for the group algebras, the only difference coming from the fact that the semigroup elements $g\in\Gamma$ will now correspond to isometries, which are not necessarily unitaries. Now this construction in hand, both the assertions are clear, as follows:

(1) With $\Gamma=\mathbb N$ we recover indeed the shift algebra $A=<S>$ on the Hilbert space $H=l^2(\mathbb N)$, the shift S itself being the isometry associated to the element $1\in\mathbb N$.

(2) With $\Gamma=\mathbb N*\mathbb N$ we recover the double shift algebra $A=<S_1,S_2>$ on the Hilbert space $H=l^2(\mathbb N*\mathbb N)$, the two shifts S_1,S_2 themselves being the isometries associated to two copies of the element $1\in\mathbb N$, one for each of the two copies of $\mathbb N$ which are present. □

In what follows we will rather use an equivalent, second approach to our problem, which is exactly the same thing, but formulated in a less abstract way, as follows:

Proposition 13.20. *We can talk about the algebra of creation operators*

$$S_x:v\to x\otimes v$$

on the free Fock space associated to a real Hilbert space H, given by

$$F(H)=\mathbb C\Omega\oplus H\oplus H^{\otimes 2}\oplus\ldots$$

and at the level of examples, we have:

(1) *With $H=\mathbb C$ we recover the shift algebra $A=<S>$ on $H=l^2(\mathbb N)$.*
(2) *With $H=\mathbb C^2$, we obtain the algebra $A=<S_1,S_2>$ on $H=l^2(\mathbb N*\mathbb N)$.*

Proof. We can talk indeed about the algebra $A(H)$ of creation operators on the free Fock space $F(H)$ associated to a real Hilbert space H, with the remark that, in terms of the abstract semigroup notions from Proposition 13.19, we have:

$$A(\mathbb C^k)=C^*(\mathbb N^{*k})\quad,\quad F(\mathbb C^k)=l^2(\mathbb N^{*k})$$

As for the assertions (1,2) in the statement, these are both clear, either directly, or by passing via (1,2) from Proposition 13.19, which were both clear as well. □

The advantage with this latter model comes from the following result, from [90], which has a very simple formulation, without linear combinations or anything:

PROPOSITION 13.21. *Given a real Hilbert space H, and two orthogonal vectors $x \perp y$, the corresponding creation operators S_x and S_y are free with respect to*

$$tr(T) = < T\Omega, \Omega >$$

called trace associated to the vacuum vector.

PROOF. In standard tensor product notation for the elements of the free Fock space $F(H)$, the formula of a creation operator associated to a vector $x \in H$ is as follows:

$$S_x(y_1 \otimes \ldots \otimes y_n) = x \otimes y_1 \otimes \ldots \otimes y_n$$

As for the formula of the adjoint of this creation operator, called annihilation operator associated to the vector $x \in H$, this is as follows:

$$S_x^*(y_1 \otimes \ldots \otimes y_n) = < x, y_1 > \otimes y_2 \otimes \ldots \otimes y_n$$

We obtain from this the following formula, which holds for any two vectors $x, y \in H$:

$$S_x^* S_y = < x, y > id$$

With these formulae in hand, the result follows by doing some elementary computations, in the spirit of those done for the group algebras, in the above. □

With this technology in hand, let us go back to our linearization program for $\boxplus$. We know from Theorem 13.17 how to model the individual distributions $\mu \in \mathcal{P}(\mathbb{R})$, and by combining this with Proposition 13.10 and Proposition 13.21, we therefore know how to freely model pairs of distributions $\mu, \nu \in \mathcal{P}(\mathbb{R})$, as required by the convolution problem. We are therefore left with doing the sum in the model, and then computing its distribution. And the point here is that, still following [90], we have:

THEOREM 13.22. *Given two polynomials $f, g \in \mathbb{C}[X]$, consider the variables*

$$S^* + f(S) \quad , \quad T^* + g(T)$$

where S, T are two creation operators, or shifts, associated to a pair of orthogonal norm 1 vectors. These variables are then free, and their sum has the same law as

$$R^* + (f + g)(R)$$

with R being the usual shift on $l^2(\mathbb{N})$.

PROOF. We have two assertions here, the idea being as follows:

(1) The freeness assertion comes from the general freeness result from Proposition 13.21, via the various identifications coming from the previous results.

(2) Regarding the second assertion, the idea is that this comes from a 45° rotation trick. Let us write indeed the two variables in the statement as follows:

$$X = S^* + a_0 + a_1 S + a_2 S^2 + \ldots$$

$$Y = T^* + b_0 + b_1 T + a_2 T^2 + \ldots$$

Now let us perform the following 45° base change, on the real span of the vectors $s,t\in H$ producing our two shifts S,T, as follows:

$$r=\frac{s+t}{\sqrt{2}}\quad,\quad u=\frac{s-t}{\sqrt{2}}$$

The new shifts, associated to these vectors $r,u\in H$, are then given by:

$$R=\frac{S+T}{\sqrt{2}}\quad,\quad U=\frac{S-T}{\sqrt{2}}$$

By using now these two new shifts, which are free according to Proposition 13.21, we obtain the following equality of distributions:

$$\begin{aligned}X+Y&=S^*+T^*+\sum_k a_kS^k+b_kT^k\\&=\sqrt{2}R^*+\sum_k a_k\left(\frac{R+U}{\sqrt{2}}\right)^k+b_k\left(\frac{R-U}{\sqrt{2}}\right)^k\\&\sim\sqrt{2}R^*+\sum_k a_k\left(\frac{R}{\sqrt{2}}\right)^k+b_k\left(\frac{R}{\sqrt{2}}\right)^k\\&\sim R^*+\sum_k a_kR^k+b_kR^k\end{aligned}$$

To be more precise, here at the end we have used the freeness property of R,U in order to cut U from the computation, as it cannot bring anything, and then we did a basic rescaling at the very end. Thus, we are led to the conclusion in the statement. □

As a conclusion, the operation $\mu\to f$ from Theorem 13.17 linearizes $\boxplus$. In order to reach now to something concrete, we are left with a computation inside $C^*(\mathbb{N})$, which is elementary, and whose conclusion is that $R_\mu=f$ can be recaptured from μ via the Cauchy transform G_μ. The precise result here, due to Voiculescu [90], is as follows:

THEOREM 13.23. *Given a real probability measure μ, define its R-transform as follows:*

$$G_\mu(\xi)=\int_{\mathbb{R}}\frac{d\mu(t)}{\xi-t}\implies G_\mu\left(R_\mu(\xi)+\frac{1}{\xi}\right)=\xi$$

The free convolution operation is then linearized by this R-transform.

PROOF. This can be done by using the above results, in several steps, as follows:

(1) According to Theorem 13.22, the operation $\mu\to f$ from Theorem 13.17 linearizes the free convolution operation $\boxplus$. We are therefore left with a computation inside $C^*(\mathbb{N})$. To be more precise, consider a variable as in Theorem 13.17:

$$X=S^*+f(S)$$

In order to establish the result, we must prove that the R-transform of X, constructed according to the procedure in the statement, is the function f itself.

(2) In order to do so, we fix $|z|<1$ in the complex plane, and we set:

$$q_z=\delta_0+\sum_{k=1}^\infty z_k\delta_k$$

The shift and its adjoint act then on this vector as follows:

$$Sq_z=z^{-1}(q_z-\delta_0)\quad,\quad S^*q_z=zq_z$$

It follows that the adjoint of our operator X acts on this vector as follows:

$$\begin{aligned}X^*q_z&=(S+f(S^*))q_z\\&=z^{-1}(q_z-\delta_0)+f(z)q_z\\&=(z^{-1}+f(z))q_z-z^{-1}\delta_0\end{aligned}$$

Now observe that the above formula can be written as follows:

$$z^{-1}\delta_0=(z^{-1}+f(z)-X^*)q_z$$

The point now is that when $|z|$ is small, the operator appearing on the right is invertible. Thus, we can rewrite the above formula as follows:

$$(z^{-1}+f(z)-X^*)^{-1}\delta_0=zq_z$$

Now by applying the trace, we are led to the following formula:

$$\begin{aligned}tr\left[(z^{-1}+f(z)-X^*)^{-1}\right]&=\left\langle(z^{-1}+f(z)-X^*)^{-1}\delta_0,\delta_0\right\rangle\\&=<zq_z,\delta_0>\\&=z\end{aligned}$$

(3) Let us apply now the procedure in the statement to the real probability measure μ modelled by X. The Cauchy transform G_μ is then given by:

$$\begin{aligned}G_\mu(\xi)&=\overline{tr((\xi-X)^{-1})}\\&=tr\Big((\bar{\xi}-X^*)^{-1}\Big)\\&=tr((\xi-X^*)^{-1})\end{aligned}$$

Now observe that, with the choice $\xi=z^{-1}+f(z)$ for our complex variable, the trace formula found in (2) above tells us that we have:

$$G_\mu\big(z^{-1}+f(z)\big)=z$$

Thus, by definition of the R-transform, we have the following formula:

$$R_\mu(z)=f(z)$$

But this finishes the proof, as explained before in step (1) above. □

Summarizing, the situation in free probability is quite similar to the one in classical probability, the product spaces needed for the basic properties of the Fourier transform being replaced by something "noncommutative", namely the free Fock space models. This is of course something quite surprising, and the credit for this remarkable discovery, which has drastically changed operator algebras, goes to Voiculescu's paper [90].

13d. Central limits

With the above linearization technology in hand, we can do many things. First, we have the following free analogue of the CLT, at variance 1, due to Voiculescu [90]:

THEOREM 13.24. *Given self-adjoint variables $x_1, x_2, x_3, \ldots$ which are f.i.d., centered, with variance 1, we have, with $n \to \infty$, in moments,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i \sim \gamma_1$$

with the limiting measure being the Wigner semicircle law on $[-2,2]$:

$$\gamma_1 = \frac{1}{2\pi}\sqrt{4-x^2}\,dx$$

Due to this, we also call this Wigner law free Gaussian law.

PROOF. We follow the same idea as in the proof of the CLT, from chapter 2:

(1) The R-transform of the variable in the statement on the left can be computed by using the linearization property from Theorem 13.23, and is given by:

$$R(\xi) = nR_x\left(\frac{\xi}{\sqrt{n}}\right) \simeq \xi$$

(2) Regarding now the right term, our first claim here is that the Cauchy transform of the Wigner law γ_1 satisfies the following equation:

$$G_{\gamma_1}\left(\xi + \frac{1}{\xi}\right) = \xi$$

Indeed, we know from before that the even moments of γ_1 are given by:

$$\frac{1}{2\pi}\int_{-2}^2 \sqrt{4-x^2}x^{2k}dx = C_k$$

On the other hand, we also know from before that the generating series of the Catalan numbers is given by the following formula:

$$\sum_{k=0}^\infty C_k z^k = \frac{1-\sqrt{1-4z}}{2z}$$

By using this formula with $z = y^{-2}$, we obtain the following formula:

$$\begin{aligned} G_{\gamma_1}(y) &= y^{-1}\sum_{k=0}^\infty C_k y^{-2k} \\ &= y^{-1}\cdot\frac{1-\sqrt{1-4y^{-2}}}{2y^{-2}} \\ &= \frac{y}{2}\left(1-\sqrt{1-4y^{-2}}\right) \\ &= \frac{y}{2}-\frac{1}{2}\sqrt{y^2-4} \end{aligned}$$

Now with $y = \xi+\xi^{-1}$, this formula becomes, as claimed in the above:

$$\begin{aligned} G_{\gamma_1}\left(\xi+\frac{1}{\xi}\right) &= \frac{\xi+\xi^{-1}}{2}-\frac{1}{2}\sqrt{\xi^2+\xi^{-2}-2} \\ &= \frac{\xi+\xi^{-1}}{2}-\frac{\xi^{-1}-\xi}{2} \\ &= \xi \end{aligned}$$

(3) We conclude from the formula found in (2) and from Theorem 13.23 that the R-transform of the Wigner semicircle law γ_1 is given by the following formula:

$$R_{\gamma_1}(\xi) = \xi$$

Observe that this follows in fact as well from the following formula, coming from Proposition 13.18, and from the technical details of the R-transform:

$$S + S^* \sim \gamma_1$$

Thus, the laws in the statement have the same R-transforms, so they are equal. □

Summarizing, we have proved the free CLT at $t = 1$. The passage to the general case, where $t > 0$ is arbitrary, is routine, and still following Voiculescu [90], we have:

THEOREM 13.25 (Free CLT). *Given self-adjoint variables $x_1, x_2, x_3, \ldots$ which are f.i.d., centered, with variance $t > 0$, we have, with $n \to \infty$, in moments,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i \sim \gamma_t$$

with the limiting measure being the Wigner semicircle law on $[-2\sqrt{t}, 2\sqrt{t}]$:

$$\gamma_t = \frac{1}{2\pi t}\sqrt{4t-x^2}\,dx$$

Due to this, we also call this Wigner law free Gaussian law.

PROOF. We follow the above proof at $t=1$, by making changes where needed:

(1) The R-transform of the variable in the statement on the left can be computed by using the linearization property from Theorem 13.23, and is given by:

$$R(\xi)=nR_x\left(\frac{\xi}{\sqrt{n}}\right)\simeq t\xi$$

(2) Regarding now the right term, our claim here is that we have:

$$G_{\gamma_t}\left(t\xi+\frac{1}{\xi}\right)=\xi$$

Indeed, we know from before that the even moments of γ_t are given by:

$$\frac{1}{2\pi t}\int_{-2\sqrt{t}}^{2\sqrt{t}}\sqrt{4t-x^2}x^{2k}dx=t^kC_k$$

On the other hand, we know from before that we have the following formula:

$$\sum_{k=0}^\infty C_kz^k=\frac{1-\sqrt{1-4z}}{2z}$$

By using this formula with $z=ty^{-2}$, we obtain the following formula:

$$\begin{aligned}G_{\gamma_t}(y)&=y^{-1}\sum_{k=0}^\infty t^kC_ky^{-2k}\\&=y^{-1}\cdot\frac{1-\sqrt{1-4ty^{-2}}}{2ty^{-2}}\\&=\frac{y}{2t}\left(1-\sqrt{1-4ty^{-2}}\right)\\&=\frac{y}{2t}-\frac{1}{2t}\sqrt{y^2-4t}\end{aligned}$$

Now with $y=t\xi+\xi^{-1}$, this formula becomes, as claimed in the above:

$$\begin{aligned}G_{\gamma_t}\left(t\xi+\frac{1}{\xi}\right)&=\frac{t\xi+\xi^{-1}}{2t}-\frac{1}{2t}\sqrt{t^2\xi^2+\xi^{-2}-2t}\\&=\frac{t\xi+\xi^{-1}}{2t}-\frac{\xi^{-1}-t\xi}{2t}\\&=\xi\end{aligned}$$

(3) We conclude from the formula found in (2) and from Theorem 13.23 that the R-transform of the Wigner semicircle law γ_t is given by the following formula:

$$R_{\gamma_t}(\xi)=t\xi$$

Thus, the laws in the statement have the same R-transforms, so they are equal. □

Regarding the limiting measures γ_t, one problem that we were having was that of understanding how γ_t exactly appears, out of γ_1. We can now solve this question:

THEOREM 13.26. *The Wigner semicircle laws have the property*

$$\gamma_s \boxplus \gamma_t = \gamma_{s+t}$$

so they form a 1*-parameter semigroup with respect to free convolution.*

PROOF. This follows either from Theorem 13.25, or from Theorem 13.23, by using the fact that the R-transform of γ_t, which is given by $R_{\gamma_t}(\xi) = t\xi$, is linear in t. $\square$

As a conclusion to what we have so far, we have:

THEOREM 13.27. *The Gaussian laws g_t, given by*

$$g_t = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx$$

and the Wigner laws γ_t, given by

$$\gamma_t = \frac{1}{2\pi t} \sqrt{4t - x^2} dx$$

have the following properties:

1. *They appear via the CLT, and the free CLT.*
2. *They form semigroups with respect to $*$ and $\boxplus$.*
3. *Their transforms are $\log F_{g_t}(x) = -tx^2/2$, $R_{\gamma_t}(x) = tx$.*
4. *Their moments are $M_k = \sum_{\pi \in D(k)} t^{|\pi|}$, with $D = P_2, NC_2$.*

PROOF. These are all results that we already know, the idea being as follows:

(1,2) These assertions follow from (3,4), via the general theory.

(3,4) These assertions follow by doing some combinatorics and calculus. $\square$

To summarize, our initial purpose for this chapter was to vaguely explore the basics of free probability, but all of a sudden, due to the power of Voiculescu's R-transform [90], we are now into stating and proving results which are on par with what we have been doing in the first part of this book, namely reasonably advanced probability theory.

This is certainly quite encouraging, and we will keep developing free probability in what follows, in the remainder of this book, with free analogues of everything, or almost, of what we have been doing in chapters 1-12, in relation with classical probability and its applications, and also with some results about the random matrices.

13e. Exercises

Exercises:

EXERCISE 13.28.

EXERCISE 13.29.

EXERCISE 13.30.

EXERCISE 13.31.

EXERCISE 13.32.

EXERCISE 13.33.

EXERCISE 13.34.

EXERCISE 13.35.

Bonus exercise.

CHAPTER 14

Random matrices

14a. Spectral measures

Welcome to the random matrices, which are first class mathematics and physics. In order to talk about such matrices and their spectral measures, we need to do some more linear algebra in infinite dimensions, as a continuation of our operator theory discussion from the previous chapters. It is convenient to upgrade our formalism, as follows:

Definition 14.1. *An abstract operator algebra, or C^*-algebra, is a complex algebra A having a norm $||.||$ and an involution $*$, subject to the following conditions:*

(1) *A is closed with respect to the norm.*
(2) *We have $||aa^*|| = ||a||^2$, for any $a \in A$.*

In other words, what we did here is to axiomatize the abstract properties of the operator algebras $A \subset B(H)$, coming from the various general results about linear operators from chapter 13, without any reference to the ambient Hilbert space H.

As basic examples here, we have the usual matrix algebras $M_N(\mathbb{C})$, with the norm and the involution being the usual matrix norm and involution, given by:

$$||A|| = \sup_{||x||=1} ||Ax|| \quad , \quad (A^*)_{ij} = \overline{A}_{ji}$$

Some other basic examples are the algebras $L^\infty(X)$ of essentially bounded functions $f : X \to \mathbb{C}$ on a measured space X, with the usual norm and involution, namely:

$$||f|| = \sup_{x \in X} |f(x)| \quad , \quad f^*(x) = \overline{f(x)}$$

We can put these two basic classes of examples together, as follows:

Proposition 14.2. *The random matrix algebras $A = M_N(L^\infty(X))$ are C^*-algebras, with their usual norm and involution, given by:*

$$||Z|| = \sup_{x \in X} ||Z_x|| \quad , \quad (Z^*)_{ij} = \overline{Z}_{ij}$$

These algebras generalize both the algebras $M_N(\mathbb{C})$, and the algebras $L^\infty(X)$.

Proof. The fact that the C^*-algebra axioms are satisfied is clear from definitions. As for the last assertion, this follows by taking $X = \{.\}$ and $N = 1$, respectively. □

We can in fact say more about the above algebras, as follows:

THEOREM 14.3. *Any algebra of type $L^\infty(X)$ is an operator algebra, as follows:*

$$L^\infty(X)\subset B(L^2(X))\quad,\quad f\to(g\to fg)$$

More generally, any random matrix algebra is an operator algebra, as follows,

$$M_N(L^\infty(X))\subset B\left(\mathbb C^N\otimes L^2(X)\right)$$

with the embedding being the above one, tensored with the identity.

PROOF. We have two assertions to be proved, the idea being as follows:

(1) Given $f\in L^\infty(X)$, consider the following operator, acting on $H=L^2(X)$:

$$T_f(g)=fg$$

Observe that T_f is indeed well-defined, and bounded as well, because:

$$||fg||_2=\sqrt{\int_X|f(x)|^2|g(x)|^2d\mu(x)}\leq||f||_\infty||g||_2$$

The application $f\to T_f$ being linear, involutive, continuous, and injective as well, we obtain in this way a C^*-algebra embedding $L^\infty(X)\subset B(H)$, as desired.

(2) Regarding the second assertion, this is best viewed in the following way:

$$\begin{aligned}M_N(L^\infty(X))&=M_N(\mathbb C)\otimes L^\infty(X)\\&\subset M_N(\mathbb C)\otimes B(L^2(X))\\&=B\left(\mathbb C^N\otimes L^2(X)\right)\end{aligned}$$

Here we have used (1), and some standard tensor product identifications. □

Our purpose in what follows is to develop the spectral theory of the C^*-algebras, and in particular that of the random matrix algebras $A=M_N(L^\infty(X))$ that we are interested in, one of our objectives being that of talking about spectral measures, in the normal case, in analogy with what we know about the usual matrices. Let us start with:

THEOREM 14.4. *Given an element $a\in A$ of a C^*-algebra, define its spectrum as:*

$$\sigma(a)=\left\{\lambda\in\mathbb C\Big|a-\lambda\notin A^{-1}\right\}$$

The following spectral theory results hold, exactly as in the $A=B(H)$ case:

1. *We have $\sigma(ab)\cup\{0\}=\sigma(ba)\cup\{0\}$.*
2. *We have $\sigma(f(a))=f(\sigma(a))$, for any $f\in\mathbb C(X)$ having poles outside $\sigma(a)$.*
3. *The spectrum $\sigma(a)$ is compact, non-empty, and contained in $D_0(||a||)$.*
4. *The spectra of unitaries $(u^*=u^{-1})$ and self-adjoints $(a=a^*)$ are on $\mathbb T,\mathbb R$.*
5. *The spectral radius of normal elements $(aa^*=a^*a)$ is given by $\rho(a)=||a||$.*

In addition, assuming $a\in A\subset B$, the spectra of a with respect to A and to B coincide.

Proof. Here the assertions (1-5), which are of course formulated a bit informally, are well-known for the full operator algebra $A = B(H)$, and the proof in general is similar:

(1) Assuming that $1 - ab$ is invertible, with inverse c, we have $abc = cab = c - 1$, and it follows that $1 - ba$ is invertible too, with inverse $1 + bca$. Thus $\sigma(ab), \sigma(ba)$ agree on $1 \in \mathbb{C}$, and by linearity, it follows that $\sigma(ab), \sigma(ba)$ agree on any point $\lambda \in \mathbb{C}^*$.

(2) The formula $\sigma(f(a)) = f(\sigma(a))$ is clear for polynomials, $f \in \mathbb{C}[X]$, by factorizing $f - \lambda$, with $\lambda \in \mathbb{C}$. Then, the extension to the rational functions is straightforward, because $P(a)/Q(a) - \lambda$ is invertible precisely when $P(a) - \lambda Q(a)$ is.

(3) By using $1/(1-b) = 1 + b + b^2 + \ldots$ for $||b|| < 1$ we obtain that $a - \lambda$ is invertible for $|\lambda| > ||a||$, and so $\sigma(a) \subset D_0(||a||)$. It is also clear that $\sigma(a)$ is closed, so what we have is a compact set. Finally, assuming $\sigma(a) = \emptyset$ the function $f(\lambda) = \varphi((a-\lambda)^{-1})$ is well-defined, for any $\varphi \in A^*$, and by Liouville we get $f = 0$, contradiction.

(4) Assuming $u^* = u^{-1}$ we have $||u|| = 1$, and so $\sigma(u) \subset D_0(1)$. But with $f(z) = z^{-1}$ we obtain via (2) that we have as well $\sigma(u) \subset f(D_0(1))$, and this gives $\sigma(u) \subset \mathbb{T}$. As for the result regarding the self-adjoints, this can be obtained from the result for the unitaries, by using (2) with functions of type $f(z) = (z + it)/(z - it)$, with $t \in \mathbb{R}$.

(5) It is routine to check, by integrating quantities of type $z^n/(z - a)$ over circles centered at the origin, and estimating, that the spectral radius is given by $\rho(a) = \lim ||a^n||^{1/n}$. But in the self-adjoint case, $a = a^*$, this gives $\rho(a) = ||a||$, by using exponents of type $n = 2^k$, and then the extension to the general normal case is straightforward.

(6) Regarding now the last assertion, the inclusion $\sigma_B(a) \subset \sigma_A(a)$ is clear. For the converse, assume $a - \lambda \in B^{-1}$, and set $b = (a - \lambda)^*(a - \lambda)$. We have then:

$$\sigma_A(b) - \sigma_B(b) = \left\{\mu \in \mathbb{C} - \sigma_B(b) \Big| (b - \mu)^{-1} \in B - A\right\}$$

Thus this difference in an open subset of $\mathbb{C}$. On the other hand b being self-adjoint, its two spectra are both real, and so is their difference. Thus the two spectra of b are equal, and in particular b is invertible in A, and so $a - \lambda \in A^{-1}$, as desired. □

We can now a prove a key result, as follows:

Theorem 14.5 (Gelfand). *If X is a compact space, the algebra $C(X)$ of continuous functions on it $f : X \to \mathbb{C}$ is a C^*-algebra, with usual norm and involution, namely:*

$$||f|| = \sup_{x \in X} |f(x)| \quad , \quad f^*(x) = \overline{f(x)}$$

Conversely, any commutative C^*-algebra is of this form, $A = C(X)$, with

$$X = \left\{\chi : A \to \mathbb{C} \ , \ \text{normed algebra character}\right\}$$

with topology making continuous the evaluation maps $ev_a : \chi \to \chi(a)$.

PROOF. There are several things going on here, the idea being as follows:

(1) The first assertion is clear from definitions. Observe that we have indeed:

$$||ff^*|| = \sup_{x\in X} |f(x)|^2 = ||f||^2$$

Observe also that the algebra $C(X)$ is commutative, because $fg = gf$.

(2) Conversely, given a commutative C^*-algebra A, let us define X as in the statement. Then X is compact, and $a \to ev_a$ is a morphism of algebras, as follows:

$$ev : A \to C(X)$$

(3) We first prove that ev is involutive. We use the following formula, which is similar to the $z = Re(z) + iIm(z)$ decomposition formula for usual complex numbers:

$$a = \frac{a + a^*}{2} + i \cdot \frac{a - a^*}{2i}$$

Thus it is enough to prove $ev_{a^*} = ev_a^*$ for the self-adjoint elements a. But this is the same as proving that $a = a^*$ implies that ev_a is a real function, which is in turn true, by Theorem 14.4, because $ev_a(\chi) = \chi(a)$ is an element of $\sigma(a)$, contained in $\mathbb{R}$.

(4) Since A is commutative, each element is normal, so ev is isometric:

$$||ev_a|| = \rho(a) = ||a||$$

It remains to prove that ev is surjective. But this follows from the Stone-Weierstrass theorem, because $ev(A)$ is a closed subalgebra of $C(X)$, which separates the points. □

As a main consequence of the Gelfand theorem, we have:

THEOREM 14.6. *For any normal element $a \in A$ we have an identification as follows:*

$$< a >= C(\sigma(a))$$

In addition, given a function $f \in C(\sigma(a))$, we can apply it to a, and we have

$$\sigma(f(a)) = f(\sigma(a))$$

which generalizes the previous rational calculus formula, in the normal case.

PROOF. Since a is normal, the C^*-algebra $< a >$ that is generates is commutative, so if we denote by X the space of the characters $\chi :< a >\to \mathbb{C}$, we have:

$$< a >= C(X)$$

Now since the map $X \to \sigma(a)$ given by evaluation at a is bijective, we obtain:

$$< a >= C(\sigma(a))$$

Thus, we are dealing here with usual functions, and this gives all the assertions. □

In order to get now towards noncommutative probability, we first have to develop the theory of positive elements, and linear forms. First, we have the following result:

PROPOSITION 14.7. *For an element $a \in A$, the following are equivalent:*

1. *a is positive, in the sense that $\sigma(a) \subset [0, \infty)$.*
2. *$a = b^2$, for some $b \in A$ satisfying $b = b^*$.*
3. *$a = cc^*$, for some $c \in A$.*

PROOF. This is something very standard, as follows:

(1) $\implies$ (2) Observe first that $\sigma(a) \subset \mathbb{R}$ implies $a = a^*$. Thus the algebra $< a >$ is commutative, and by using Theorem 14.6, we can set $b = \sqrt{a}$.

(2) $\implies$ (3) This is trivial, because we can simply set $c = b$.

(2) $\implies$ (1) This is clear too, because we have:

$$\sigma(a) = \sigma(b^2) = \sigma(b)^2 \subset \mathbb{R}^2 = [0, \infty)$$

(3) $\implies$ (1) We proceed by contradiction. By multiplying c by a suitable element of $< cc^* >$, we are led to the existence of an element $d \neq 0$ satisfying:

$$-dd^* \geq 0$$

By writing now $d = x + iy$ with $x = x^*, y = y^*$ we have:

$$dd^* + d^*d = 2(x^2 + y^2) \geq 0$$

Thus $d^*d \geq 0$, which is easily seen to contradict the condition $-dd^* \geq 0$. □

We can talk as well about positive linear forms, as follows:

DEFINITION 14.8. *Consider a linear map $\varphi : A \to \mathbb{C}$.*

1. *φ is called positive when $a \geq 0 \implies \varphi(a) \geq 0$.*
2. *φ is called faithful and positive when $a \geq 0, a \neq 0 \implies \varphi(a) > 0$.*

In the commutative case, $A = C(X)$, the positive linear forms appear as follows, with μ being positive, and strictly positive if we want φ to be faithful and positive:

$$\varphi(f) = \int_X f(x) d\mu(x)$$

In general, the positive linear forms can be thought of as being integration functionals with respect to some underlying "positive measures". We have:

DEFINITION 14.9. *Let A be a C^*-algebra, given with a positive trace $tr : A \to \mathbb{C}$.*

1. *The elements $a \in A$ are called random variables.*
2. *The moments of such a variable are the numbers $M_k(a) = tr(a^k)$.*
3. *The law of such a variable is the functional $\mu_a : P \to tr(P(a))$.*

Here the exponent $k = \circ\bullet\bullet\circ\ldots$ is by definition a colored integer, and the powers a^k are defined by the following formulae, and multiplicativity:

$$a^\emptyset = 1 \quad , \quad a^\circ = a \quad , \quad a^\bullet = a^*$$

As for the polynomial P, this is a noncommuting $*$-polynomial in one variable:

$$P \in \mathbb{C} < X, X^* >$$

Observe that the law is uniquely determined by the moments, because we have:

$$P(X) = \sum_k \lambda_k X^k \implies \mu_a(P) = \sum_k \lambda_k M_k(a)$$

At the level of the general theory, we have the following key result, extending the various results that we have, regarding the self-adjoint and normal matrices:

THEOREM 14.10. *Let A be a C^*-algebra, with a trace tr, and consider an element $a \in A$ which is normal, in the sense that $aa^* = a^*a$.*

1. *μ_a is a complex probability measure, satisfying $supp(\mu_a) \subset \sigma(a)$.*
2. *In the self-adjoint case, $a = a^*$, this measure μ_a is real.*
3. *Assuming that tr is faithful, we have $supp(\mu_a) = \sigma(a)$.*

PROOF. This is something very standard, that we already know for the usual complex matrices, and whose proof in general is quite similar, as follows:

(1) In the normal case, $aa^* = a^*a$, the Gelfand theorem, or rather the subsequent continuous functional calculus theorem, tells us that we have:

$$< a >= C(\sigma(a))$$

Thus the functional $f(a) \to tr(f(a))$ can be regarded as an integration functional on the algebra $C(\sigma(a))$, and by the Riesz theorem, this latter functional must come from a probability measure μ on the spectrum $\sigma(a)$, in the sense that we must have:

$$tr(f(a)) = \int_{\sigma(a)} f(z) d\mu(z)$$

We are therefore led to the conclusions in the statement, with the uniqueness assertion coming from the fact that the elements a^k, taken as usual with respect to colored integer exponents, $k = \circ\bullet\bullet\circ\ldots$, generate the whole C^*-algebra $C(\sigma(a))$.

(2) This is something which is clear from definitions.

(3) Once again, this is something which is clear from definitions. □

As a first concrete application now, by getting back to the random matrices, and to the various questions raised in the beginning of this chapter, we have:

THEOREM 14.11. *Given a random matrix $Z \in M_N(L^\infty(X))$ which is normal,*

$$ZZ^* = Z^*Z$$

its law, which is by definition the following abstract functional,

$$\mu : \mathbb{C} < X, X^* > \to \mathbb{C} \quad , \quad P \to \frac{1}{N}\int_X tr(P(Z))$$

when restricted to the usual polynomials in two variables,

$$\mu : \mathbb{C}[X, X^*] \to \mathbb{C} \quad , \quad P \to \frac{1}{N}\int_X tr(P(Z))$$

must come from a probability measure on the spectrum $\sigma(Z) \subset \mathbb{C}$, as follows:

$$\mu(P) = \int_{\sigma(T)} P(x)d\mu(x)$$

We agree to use the symbol μ for all these notions.

PROOF. This follows indeed from what we know from Theorem 14.10, applied to the normal element $a = Z$, belonging to the C^*-algebra $A = M_N(L^\infty(X))$. □

14b. Gaussian matrices

We have now all the needed ingredients for launching some explicit random matrix computations. Our goal will be that of computing the asymptotic moments, and then the asymptotic laws, with $N \to \infty$, for the main classes of large random matrices.

Let us begin by specifying the precise classes of matrices that we are interested in. First we have the complex Gaussian matrices, which are constructed as follows:

DEFINITION 14.12. *A complex Gaussian matrix is a random matrix of type*

$$Z \in M_N(L^\infty(X))$$

which has i.i.d. centered complex normal entries.

To be more precise, the assumption in this definition is that all the matrix entries Z_{ij} are independent, and follow the same complex normal law G_t, for a fixed value of $t > 0$. We will see that the above matrices have an interesting, and "central" combinatorics, among all kinds of random matrices, with the study of the other random matrices being usually obtained as a modification of the study of the Gaussian matrices.

As a somewhat surprising remark, using real normal variables in Definition 14.12, instead of the complex ones appearing there, leads nowhere. The correct real versions of the Gaussian matrices are the Wigner random matrices, constructed as follows:

Definition 14.13. *A Wigner matrix is a random matrix of type*

$$Z\in M_N(L^\infty(X))$$

which has i.i.d. centered complex normal entries, up to the constraint $Z=Z^*$.

This definition is something a bit compacted, and to be more precise, a Wigner matrix is by definition a random matrix as follows, with the diagonal entries being real normal variables, $a_i\sim g_t$, for some $t>0$, the upper diagonal entries being complex normal variables, $b_{ij}\sim G_t$, the lower diagonal entries being the conjugates of the upper diagonal entries, as indicated, and with all the variables a_i,b_{ij} being independent:

$$Z=\begin{pmatrix}a_1&b_{12}&\ldots&\ldots&b_{1N}\\ \bar{b}_{12}&a_2&\ddots&&\vdots\\ \vdots&\ddots&\ddots&\ddots&\vdots\\ \vdots&&\ddots&a_{N-1}&b_{N-1,N}\\ \bar{b}_{1N}&\ldots&\ldots&\bar{b}_{N-1,N}&a_N\end{pmatrix}$$

As a comment here, for many concrete applications the Wigner matrices are in fact the central objects in random matrix theory, and in particular, they are often more important than the Gaussian matrices. In fact, these are the random matrices which were first considered and investigated, a long time ago, by Wigner himself.

However, as we will soon discover, the Gaussian matrices are somehow more fundamental than the Wigner matrices, at least from an abstract point of view, and this will be the point of view that we will follow here, with the Gaussian matrices coming first.

Finally, we will be interested as well in the complex Wishart matrices, which are the positive versions of the above random matrices, constructed as follows:

Definition 14.14. *A complex Wishart matrix is a random matrix of type*

$$Z=YY^*\in M_N(L^\infty(X))$$

with Y being a complex Gaussian matrix.

As before with the Gaussian and Wigner matrices, there are many possible comments that can be made here, of technical or historical nature. As a first key fact, using real Gaussian variables instead of complex ones leads to a less interesting combinatorics. Also, these matrices were introduced and studied by Marchenko and Pastur not long after Wigner, and so historically came second. Finally, in what regards their combinatorics and applications, these matrices quite often come first, before both the Gaussian and the Wigner ones, with all this being of course a matter of knowledge and taste.

Summarizing, we have three main types of random matrices, which can be thought of as being “complex”, “real” and “positive”, and that we will study in what follows, in this precise order, with this order being the one that fits us best here. Let us also mention that there are many other interesting classes of random matrices, which are more specialized, usually appearing as modifications of the above. More on these later.

In order to compute the asymptotic laws of the Gaussian, Wigner and Wishart matrices, we use the moment method. We first have the following result:

THEOREM 14.15. *Given a sequence of Gaussian random matrices*

$$Z_N \in M_N(L^\infty(X))$$

having independent G_t variables as entries, for some fixed $t > 0$, we have

$$M_k\left(\frac{Z_N}{\sqrt{N}}\right) \simeq t^{|k|/2}|\mathcal{NC}_2(k)|$$

for any colored integer $k = \circ\bullet\bullet\circ\ldots$, in the $N \to \infty$ limit.

PROOF. This is something standard, which can be done as follows:

(1) We fix $N \in \mathbb{N}$, and we let $Z = Z_N$. Let us first compute the trace of Z^k. With $k = k_1 \ldots k_s$, and with the convention $(ij)^\circ = ij, (ij)^\bullet = ji$, we have:

$$\begin{aligned} Tr(Z^k) &= Tr(Z^{k_1} \ldots Z^{k_s}) \\ &= \sum_{i_1=1}^N \ldots \sum_{i_s=1}^N (Z^{k_1})_{i_1 i_2} (Z^{k_2})_{i_2 i_3} \ldots (Z^{k_s})_{i_s i_1} \\ &= \sum_{i_1=1}^N \ldots \sum_{i_s=1}^N (Z_{(i_1 i_2)^{k_1}})^{k_1} (Z_{(i_2 i_3)^{k_2}})^{k_2} \ldots (Z_{(i_s i_1)^{k_s}})^{k_s} \end{aligned}$$

(2) Next, we rescale our variable Z by a $\sqrt{N}$ factor, as in the statement, and we also replace the usual trace by its normalized version, $tr = Tr/N$. Our formula becomes:

$$tr\left(\left(\frac{Z}{\sqrt{N}}\right)^k\right) = \frac{1}{N^{s/2+1}} \sum_{i_1=1}^N \ldots \sum_{i_s=1}^N (Z_{(i_1 i_2)^{k_1}})^{k_1} (Z_{(i_2 i_3)^{k_2}})^{k_2} \ldots (Z_{(i_s i_1)^{k_s}})^{k_s}$$

Thus, the moment that we are interested in is given by:

$$M_k\left(\frac{Z}{\sqrt{N}}\right) = \frac{1}{N^{s/2+1}} \sum_{i_1=1}^N \ldots \sum_{i_s=1}^N \int_X (Z_{(i_1 i_2)^{k_1}})^{k_1} (Z_{(i_2 i_3)^{k_2}})^{k_2} \ldots (Z_{(i_s i_1)^{k_s}})^{k_s}$$

(3) Let us apply now the Wick formula, that we know well from before. We conclude that the moment that we are interested in is given by the following formula:

$$\begin{aligned}
&M_k\left(\frac{Z}{\sqrt{N}}\right)\\
&= \frac{t^{s/2}}{N^{s/2+1}}\sum_{i_1=1}^N\ldots\sum_{i_s=1}^N\#\left\{\pi\in\mathcal{P}_2(k)\Big|\pi\leq\ker\left((i_1i_2)^{k_1},(i_2i_3)^{k_2},\ldots,(i_si_1)^{k_s}\right)\right\}\\
&= t^{s/2}\sum_{\pi\in\mathcal{P}_2(k)}\frac{1}{N^{s/2+1}}\#\left\{i\in\{1,\ldots,N\}^s\Big|\pi\leq\ker\left((i_1i_2)^{k_1},(i_2i_3)^{k_2},\ldots,(i_si_1)^{k_s}\right)\right\}
\end{aligned}$$

(4) Our claim now is that in the $N\to\infty$ limit the combinatorics of the above sum simplifies, with only the noncrossing partitions contributing to the sum, and with each of them contributing precisely with a 1 factor, so that we will have, as desired:

$$\begin{aligned}
M_k\left(\frac{Z}{\sqrt{N}}\right) &= t^{s/2}\sum_{\pi\in\mathcal{P}_2(k)}\left(\delta_{\pi\in NC_2(k)}+O(N^{-1})\right)\\
&\simeq t^{s/2}\sum_{\pi\in\mathcal{P}_2(k)}\delta_{\pi\in NC_2(k)}\\
&= t^{s/2}|\mathcal{NC}_2(k)|
\end{aligned}$$

(5) In order to prove this, the first observation is that when k is not uniform, in the sense that it contains a different number of $\circ,\bullet$ symbols, we have $\mathcal{P}_2(k)=\emptyset$, and so:

$$M_k\left(\frac{Z}{\sqrt{N}}\right)=t^{s/2}|\mathcal{NC}_2(k)|=0$$

(6) Thus, we are left with the case where k is uniform. Let us examine first the case where k consists of an alternating sequence of $\circ$ and $\bullet$ symbols, as follows:

$$k=\underbrace{\circ\bullet\circ\bullet\ldots\ldots\circ\bullet}_{2p}$$

In this case it is convenient to relabel our multi-index $i=(i_1,\ldots,i_s)$, with $s=2p$, in the form $(j_1,l_1,j_2,l_2,\ldots,j_p,l_p)$. With this done, our moment formula becomes:

$$M_k\left(\frac{Z}{\sqrt{N}}\right)=t^p\sum_{\pi\in\mathcal{P}_2(k)}\frac{1}{N^{p+1}}\#\left\{j,l\in\{1,\ldots,N\}^p\Big|\pi\leq\ker\left(j_1l_1,j_2l_1,j_2l_2,\ldots,j_1l_p\right)\right\}$$

Now observe that, with k being as above, we have an identification $\mathcal{P}_2(k)\simeq S_p$, obtained in the obvious way. With this done too, our moment formula becomes:

$$M_k\left(\frac{Z}{\sqrt{N}}\right)=t^p\sum_{\pi\in S_p}\frac{1}{N^{p+1}}\#\left\{j,l\in\{1,\ldots,N\}^p\Big|j_r=j_{\pi(r)+1},l_r=l_{\pi(r)},\forall r\right\}$$

(7) We are now ready to do our asymptotic study, and prove the claim in (4). Let indeed $\gamma \in S_p$ be the full cycle, which is by definition the following permutation:

$$\gamma = (1\,2\,\ldots\,p)$$

In terms of γ, the conditions $j_r = j_{\pi(r)+1}$ and $l_r = l_{\pi(r)}$ found above read:

$$\gamma\pi \leq \ker j \quad , \quad \pi \leq \ker l$$

Counting the number of free parameters in our moment formula, we obtain:

$$M_k\left(\frac{Z}{\sqrt{N}}\right) = \frac{t^p}{N^{p+1}}\sum_{\pi\in S_p} N^{|\pi|+|\gamma\pi|} = t^p\sum_{\pi\in S_p} N^{|\pi|+|\gamma\pi|-p-1}$$

(8) The point now is that the last exponent is well-known to be ≤ 0, with equality precisely when the permutation $\pi \in S_p$ is geodesic, which in practice means that π must come from a noncrossing partition. Thus we obtain, in the $N \to \infty$ limit, as desired:

$$M_k\left(\frac{Z}{\sqrt{N}}\right) \simeq t^p|\mathcal{NC}_2(k)|$$

This finishes the proof in the case of the exponents k which are alternating, and the case where k is an arbitrary uniform exponent is similar, by permuting everything. □

The above result is very nice, but the resulting asymptotic measure is still in need to be interpreted. For more on all this, we refer to free probability theory [**91**].

14c. Wigner and Wishart

Regarding now the Wigner matrices, we have here the following result, coming as a consequence of Theorem 14.15, via some simple algebraic manipulations:

THEOREM 14.16. *Given a sequence of Wigner random matrices*

$$Z_N \in M_N(L^\infty(X))$$

having independent G_t variables as entries, with $t > 0$, up to $Z_N = Z_N^*$, we have

$$M_k\left(\frac{Z_N}{\sqrt{N}}\right) \simeq t^{k/2}|NC_2(k)|$$

for any integer $k \in \mathbb{N}$, in the $N \to \infty$ limit.

PROOF. This can be deduced from a direct computation based on the Wick formula, similar to that from the proof of Theorem 14.15, but the best is to deduce this result from Theorem 14.15 itself. Indeed, we know from there that for Gaussian matrices $Y_N \in M_N(L^\infty(X))$ we have the following formula, valid for any colored integer $K = \circ\bullet\bullet\circ\ldots$, in the $N \to \infty$ limit, with $\mathcal{NC}_2$ standing for noncrossing matching pairings:

$$M_K\left(\frac{Y_N}{\sqrt{N}}\right) \simeq t^{|K|/2}|\mathcal{NC}_2(K)|$$

By doing some combinatorics, we deduce from this that we have the following formula for the moments of the matrices $Re(Y_N)$, with respect to usual exponents, $k \in \mathbb{N}$:

$$\begin{aligned} M_k\left(\frac{Re(Y_N)}{\sqrt{N}}\right) &= 2^{-k} \cdot M_k\left(\frac{Y_N}{\sqrt{N}} + \frac{Y_N^*}{\sqrt{N}}\right) \\ &= 2^{-k} \sum_{|K|=k} M_K\left(\frac{Y_N}{\sqrt{N}}\right) \\ &\simeq 2^{-k} \sum_{|K|=k} t^{k/2} |\mathcal{NC}_2(K)| \\ &= 2^{-k} \cdot t^{k/2} \cdot 2^{k/2} |\mathcal{NC}_2(k)| \\ &= 2^{-k/2} \cdot t^{k/2} |NC_2(k)| \end{aligned}$$

Now since the matrices $Z_N = \sqrt{2} Re(Y_N)$ are of Wigner type, this gives the result. □

Now by putting everything together, we obtain the Wigner theorem, as follows:

THEOREM 14.17. *Given a sequence of Wigner random matrices*

$$Z_N \in M_N(L^\infty(X))$$

which by definition have i.i.d. complex normal entries, up to $Z_N = Z_N^*$, we have

$$Z_N \sim \gamma_t$$

in the $N \to \infty$ limit, where $\gamma_t = \frac{1}{2\pi t}\sqrt{4t - x^2}dx$ is the Wigner semicircle law.

PROOF. This follows indeed from Theorem 14.16, via some combinatorics, that we know from before, in order to recover the Wigner law, out of the Catalan numbers. □

Let us discuss now the Wishart matrices, which are the positive analogues of the Wigner matrices. Quite surprisingly, the computation here leads to the Catalan numbers, but not in the same way as for the Wigner matrices, the result being as follows:

THEOREM 14.18. *Given a sequence of complex Wishart matrices*

$$W_N = Y_N Y_N^* \in M_N(L^\infty(X))$$

with Y_N being $N \times N$ complex Gaussian of parameter $t > 0$, we have

$$M_k\left(\frac{W_N}{N}\right) \simeq t^k C_k$$

for any exponent $k \in \mathbb{N}$, in the $N \to \infty$ limit.

PROOF. There are several possible proofs for this result, as follows:

(1) A first method is by using the formula that we have in Theorem 14.15, for the Gaussian matrices Y_N. Indeed, we know from there that we have the following formula, valid for any colored integer $K = \circ\bullet\bullet\circ\ldots$, in the $N\to\infty$ limit:

$$M_K\left(\frac{Y_N}{\sqrt{N}}\right)\simeq t^{|K|/2}|\mathcal{NC}_2(K)|$$

With $K = \circ\bullet\circ\bullet\ldots$, alternating word of lenght $2k$, with $k\in\mathbb{N}$, this gives:

$$M_k\left(\frac{Y_NY_N^*}{N}\right)\simeq t^k|\mathcal{NC}_2(K)|$$

Thus, in terms of the Wishart matrix $W_N = Y_NY_N^*$ we have, for any $k\in\mathbb{N}$:

$$M_k\left(\frac{W_N}{N}\right)\simeq t^k|\mathcal{NC}_2(K)|$$

The point now is that, by doing some combinatorics, we have:

$$|\mathcal{NC}_2(K)| = |NC_2(2k)| = C_k$$

Thus, we are led to the formula in the statement.

(2) A second method, that we will explain now as well, is by proving the result directly, starting from definitions. The matrix entries of our matrix $W = W_N$ are given by:

$$W_{ij}=\sum_{r=1}^N Y_{ir}\bar{Y}_{jr}$$

Thus, the normalized traces of powers of W are given by the following formula:

$$\begin{aligned}
tr(W^k) &= \frac{1}{N}\sum_{i_1=1}^N\ldots\sum_{i_k=1}^N W_{i_1i_2}W_{i_2i_3}\ldots W_{i_ki_1}\\
&= \frac{1}{N}\sum_{i_1=1}^N\ldots\sum_{i_k=1}^N\sum_{r_1=1}^N\ldots\sum_{r_k=1}^N Y_{i_1r_1}\bar{Y}_{i_2r_1}Y_{i_2r_2}\bar{Y}_{i_3r_2}\ldots Y_{i_kr_k}\bar{Y}_{i_1r_k}
\end{aligned}$$

By rescaling now W by a $1/N$ factor, as in the statement, we obtain:

$$tr\left(\left(\frac{W}{N}\right)^k\right)=\frac{1}{N^{k+1}}\sum_{i_1=1}^N\ldots\sum_{i_k=1}^N\sum_{r_1=1}^N\ldots\sum_{r_k=1}^N Y_{i_1r_1}\bar{Y}_{i_2r_1}Y_{i_2r_2}\bar{Y}_{i_3r_2}\ldots Y_{i_kr_k}\bar{Y}_{i_1r_k}$$

By using now the Wick rule, we obtain the following formula for the moments, with $K = \circ\bullet\circ\bullet\ldots$, alternating word of lenght $2k$, and with $I = (i_1r_1, i_2r_1, \ldots, i_kr_k, i_1r_k)$:

$$\begin{aligned} M_k\left(\frac{W}{N}\right) &= \frac{t^k}{N^{k+1}}\sum_{i_1=1}^N\ldots\sum_{i_k=1}^N\sum_{r_1=1}^N\ldots\sum_{r_k=1}^N\#\left\{\pi\in\mathcal{P}_2(K)\Big|\pi\leq\ker(I)\right\}\\ &= \frac{t^k}{N^{k+1}}\sum_{\pi\in\mathcal{P}_2(K)}\#\left\{i,r\in\{1,\ldots,N\}^k\Big|\pi\leq\ker(I)\right\}\end{aligned}$$

In order to compute this quantity, we use the standard bijection $\mathcal{P}_2(K)\simeq S_k$. By identifying the pairings $\pi\in\mathcal{P}_2(K)$ with their counterparts $\pi\in S_k$, we obtain:

$$M_k\left(\frac{W}{N}\right) = \frac{t^k}{N^{k+1}}\sum_{\pi\in S_k}\#\left\{i,r\in\{1,\ldots,N\}^k\Big|i_s=i_{\pi(s)+1},r_s=r_{\pi(s)},\forall s\right\}$$

Now let $\gamma\in S_k$ be the full cycle, which is by definition the following permutation:

$$\gamma = (1\,2\,\ldots\,k)$$

The general factor in the product computed above is then 1 precisely when following two conditions are simultaneously satisfied:

$$\gamma\pi\leq\ker i \quad,\quad \pi\leq\ker r$$

Counting the number of free parameters in our moment formula, we obtain:

$$M_k\left(\frac{W}{N}\right) = t^k\sum_{\pi\in S_k}N^{|\pi|+|\gamma\pi|-k-1}$$

The point now is that the last exponent is well-known to be ≤ 0, with equality precisely when the permutation $\pi\in S_k$ is geodesic, which in practice means that π must come from a noncrossing partition. Thus we obtain, in the $N\to\infty$ limit:

$$M_k\left(\frac{W}{N}\right)\simeq t^kC_k$$

Thus, we are led to the conclusion in the statement. □

We are led in this way to the following result:

THEOREM 14.19. *Given a sequence of complex Wishart matrices*

$$W_N = Y_NY_N^*\in M_N(L^\infty(X))$$

with Y_N being $N\times N$ complex Gaussian of parameter $t>0$, we have

$$\frac{W_N}{tN}\sim\frac{1}{2\pi}\sqrt{4x^{-1}-1}\,dx$$

with $N\to\infty$, with the limiting measure being the Marchenko-Pastur law π_1.

PROOF. This follows indeed from Theorem 14.18, via some standard combinatorics, in order to recover the Marchenko-Pastur law, out of the Catalan numbers. □

Let us discuss now a generalization of the above results, motivated by a whole array of concrete questions, and bringing into the picture a "true" parameter $t > 0$, which is different from the parameter $t > 0$ used above, which is something quite trivial.

For this purpose, let us go back to the definition of the Wishart matrices. There were as follows, with Y being a $N \times N$ matrix with i.i.d. entries, each following the law G_t:

$$W = YY^*$$

The point now is that, more generally, we can use in this construction a $N \times M$ matrix Y with i.i.d. entries, each following the law G_t, with $M \in \mathbb{N}$ being arbitrary. Thus, we have a new parameter, and by ditching the old parameter $t > 0$, we are led to the following definition, which is the "true" definition of the Wishart matrices:

DEFINITION 14.20. *A complex Wishart matrix is a $N \times N$ matrix of the form*

$$W = YY^*$$

where Y is a $N \times M$ matrix with i.i.d. entries, each following the law G_1.

In order to see now what is going on, combinatorially, let us compute moments. The result here is substantially more interesting than that for the previous Wishart matrices, with the new revelant numeric parameter being now the number $t = M/N$, as follows:

THEOREM 14.21. *Given a sequence of complex Wishart matrices*

$$W_N = Y_N Y_N^* \in M_N(L^\infty(X))$$

with Y_N being $N \times M$ complex Gaussian of parameter 1, we have

$$M_k\left(\frac{W_N}{N}\right) \simeq \sum_{\pi \in NC(k)} t^{|\pi|}$$

for any exponent $k \in \mathbb{N}$, in the $M = tN \to \infty$ limit.

PROOF. This is something which is very standard, as follows:

(1) Before starting, let us clarify the relation with our previous Wishart matrix results. In the case $M = N$ we have $t = 1$, and the formula in the statement reads:

$$M_k\left(\frac{W_N}{N}\right) \simeq |NC(k)|$$

Thus, what we have here is the previous Wishart matrix formula, in full generality, at the value $t = 1$ of our old parameter $t > 0$.

(2) Observe also that by rescaling, we can obtain if we want from this the previous Wishart matrix formula, in full generality, at any value $t > 0$ of our old parameter. Thus, things fine, we are indeed generalizing what we did before.

(3) In order to prove now the formula in the statement, we proceed as usual, by using the Wick formula. The matrix entries of our Wishart matrix $W = W_N$ are given by:

$$W_{ij} = \sum_{r=1}^{M} Y_{ir} \bar{Y}_{jr}$$

Thus, the normalized traces of powers of W are given by the following formula:

$$\begin{aligned} tr(W^k) &= \frac{1}{N} \sum_{i_1=1}^{N} \ldots \sum_{i_k=1}^{N} W_{i_1 i_2} W_{i_2 i_3} \ldots W_{i_k i_1} \\ &= \frac{1}{N} \sum_{i_1=1}^{N} \ldots \sum_{i_k=1}^{N} \sum_{r_1=1}^{M} \ldots \sum_{r_k=1}^{M} Y_{i_1 r_1} \bar{Y}_{i_2 r_1} Y_{i_2 r_2} \bar{Y}_{i_3 r_2} \ldots Y_{i_k r_k} \bar{Y}_{i_1 r_k} \end{aligned}$$

By rescaling now W by a $1/N$ factor, as in the statement, we obtain:

$$tr\left(\left(\frac{W}{N}\right)^k\right) = \frac{1}{N^{k+1}} \sum_{i_1=1}^{N} \ldots \sum_{i_k=1}^{N} \sum_{r_1=1}^{M} \ldots \sum_{r_k=1}^{M} Y_{i_1 r_1} \bar{Y}_{i_2 r_1} Y_{i_2 r_2} \bar{Y}_{i_3 r_2} \ldots Y_{i_k r_k} \bar{Y}_{i_1 r_k}$$

(4) By using now the Wick rule, we obtain the following formula for the moments, with $K = \circ \bullet \circ \bullet \ldots$, alternating word of lenght $2k$, and $I = (i_1 r_1, i_2 r_1, \ldots, i_k r_k, i_1 r_k)$:

$$\begin{aligned} M_k\left(\frac{W}{N}\right) &= \frac{1}{N^{k+1}} \sum_{i_1=1}^{N} \ldots \sum_{i_k=1}^{N} \sum_{r_1=1}^{M} \ldots \sum_{r_k=1}^{M} \#\left\{\pi \in \mathcal{P}_2(K) \Big| \pi \leq \ker I\right\} \\ &= \frac{1}{N^{k+1}} \sum_{\pi \in \mathcal{P}_2(K)} \#\left\{i \in \{1, \ldots, N\}^k, r \in \{1, \ldots, M\}^k \Big| \pi \leq \ker I\right\} \end{aligned}$$

(5) In order to compute this quantity, we use the standard bijection $\mathcal{P}_2(K) \simeq S_k$. By identifying the pairings $\pi \in \mathcal{P}_2(K)$ with their counterparts $\pi \in S_k$, we obtain:

$$M_k\left(\frac{W}{N}\right) = \frac{1}{N^{k+1}} \sum_{\pi \in S_k} \#\left\{i \in \{1, \ldots, N\}^k, r \in \{1, \ldots, M\}^k \Big| i_s = i_{\pi(s)+1}, r_s = r_{\pi(s)}\right\}$$

Now let $\gamma \in S_k$ be the full cycle, which is by definition the following permutation:

$$\gamma = (1\, 2\, \ldots\, k)$$

The general factor in the product computed above is then 1 precisely when following two conditions are simultaneously satisfied:

$$\gamma\pi \leq \ker i \quad , \quad \pi \leq \ker r$$

Counting the number of free parameters in our expectation formula, we obtain:

$$M_k\left(\frac{W}{N}\right)=\frac{1}{N^{k+1}}\sum_{\pi\in S_k}N^{|\gamma\pi|}M^{|\pi|}=\sum_{\pi\in S_k}N^{|\gamma\pi|-k-1}M^{|\pi|}$$

(6) Now by using the same arguments as in the case $M=N$, from the proof of Theorem 15.18, we conclude that in the $M=tN\to\infty$ limit the permutations $\pi\in S_k$ which matter are those coming from noncrossing partitions, and so that we have:

$$M_k\left(\frac{W}{N}\right)\simeq\sum_{\pi\in NC(k)}N^{-|\pi|}M^{|\pi|}=\sum_{\pi\in NC(k)}t^{|\pi|}$$

We are therefore led to the conclusion in the statement. □

In order to recapture now the density out of the moments, we can of course use the Stieltjes inversion formula, but the computations here are a bit opaque. So, inspired from what happens at $t=1$, let us cheat a bit, and formulate things as follows:

Definition 14.22. *The Marchenko-Pastur law π_t of parameter $t>0$ is given by:*

$$a\sim\gamma_t\implies a^2\sim\pi_t$$

That is, π_t the law of the square of a variable following the law γ_t.

This is certainly nice and simple, and we know that at $t=1$ we obtain indeed the Marchenko-Pastur law π_1, as constructed above. In general, we have:

Proposition 14.23. *The Marchenko-Pastur law of parameter $t>0$ is*

$$\pi_t=\max(1-t,0)\delta_0+\frac{\sqrt{4t-(x-1-t)^2}}{2\pi x}\,dx$$

the support being $[0,4t^2]$, and the moments of this measure are

$$M_k=\sum_{\pi\in NC(k)}t^{|\pi|}$$

exactly as for the asymptotic moments of the complex Wishart matrices.

Proof. This follows as usual, by doing some computations, either combinatorics, or calculus. To be more precise, we have three formulae for π_t to be connected, namely the one in Definition 14.22, and the two ones from the present statement, and the connections between them can be established exactly as we did before, at $t=1$. □

Now back to the complex Wishart matrices that we are interested in, we can now formulate a final result regarding them, as follows:

Theorem 14.24. *Given a sequence of complex Wishart matrices*

$$W_N = Y_N Y_N^* \in M_N(L^\infty(X))$$

with Y_N being $N \times M$ complex Gaussian of parameter 1, we have

$$\frac{W_N}{N} \sim \max(1-t,0)\delta_0 + \frac{\sqrt{4t-(x-1-t)^2}}{2\pi x}\,dx$$

with $M = tN \to \infty$, with the limiting measure being the Marchenko-Pastur law π_t.

Proof. This follows indeed from Theorem 14.21 and Proposition 14.23. □

Many other things can be said, along these lines, and for more on all this, we refer to free probability theory [**91**], which has answers to nearly all potential questions that can be asked, regarding the various classes of random matrices investigated above.

14d. Block modifications

Our goal now will be that of explaining a surprising result, stating that when suitably block-transposing the entries of a complex Wishart matrix, we obtain as asymptotic distribution a shifted version of Wigner's semicircle law. Let us start with:

Definition 14.25. *The partial transpose of a complex Wishart matrix W of parameters (dn, dm) is the matrix*

$$\tilde{W} = (id \otimes t)W$$

where id is the identity of $M_d(\mathbb{C})$, and t is the transposition of $M_n(\mathbb{C})$.

In more familiar terms of bases and indices, the standard decomposition $\mathbb{C}^{dn} = \mathbb{C}^d \otimes \mathbb{C}^n$ induces an algebra decomposition $M_{dn}(\mathbb{C}) = M_d(\mathbb{C}) \otimes M_n(\mathbb{C})$, and with this convention made, the partial transpose matrix $\tilde{W}$ constructed above has entries as follows:

$$\tilde{W}_{ia,jb} = W_{ib,ja}$$

Our goal in what follows will be that of computing the law of $\tilde{W}$, first when d, n, m are fixed, and then in the $d \to \infty$ regime. For this purpose, we will need a number of standard facts regarding the noncrossing partitions. Let us start with:

Proposition 14.26. *For a permutation $\sigma \in S_p$, we have the formula*

$$|\sigma| + \#\sigma = p$$

where $|\sigma|$ is the number of cycles of σ, and $\#\sigma$ is the minimal $k \in \mathbb{N}$ such that σ is a product of k transpositions. Also, the following formula defines a distance on S_p,

$$(\sigma, \pi) \to \#(\sigma^{-1}\pi)$$

and the set of permutations $\sigma \in S_p$ which saturate the triangular inequality

$$\#\sigma + \#(\sigma^{-1}\gamma) = \#\gamma = p - 1$$

where $\gamma \in S_p$ is a full cycle, is in bijection with the set $NC(p)$.

PROOF. All this is standard combinatorics, that we will leave here as an exercise. □

We will need as well the following well-known result:

PROPOSITION 14.27. *The number $||\pi||$ of blocks having even size is given by*

$$1+||\pi||=|\pi\gamma|$$

for every noncrossing partition $\pi\in NC(p)$.

PROOF. We use a recurrence over the number of blocks of π. If π has just one block, its associated geodesic permutation is γ and we have:

$$|\gamma^2|=\begin{cases}1&(p\text{ odd})\\2&(p\text{ even})\end{cases}$$

For partitions π with more than one block, we can assume without loss of generality that $\pi=\hat{1}_k\sqcup\pi'$, where $\hat{1}_k$ is a contiguous block of size k. Recall that the number of blocks of the permutation $\pi\gamma$ is given by the following formula, where $\rho_{14}\in P_2(2p)$ is the pair partition which pairs an element i with $i+(-1)^{i+1}3$:

$$|\pi\gamma|=|\widetilde{\pi}\vee\rho_{14}|$$

If k is an even number, $k=2r$, consider the following partition, which contains the block $(1\,4\,5\,8\,\ldots 4r-3\,4r)$, along with the blocks coming from elements of the form $4i+2,4i+3$ from $\{1,\ldots,4r\}$ and from π':

$$\sigma=\widetilde{\hat{1}_{2r}\sqcup\pi'}\vee\rho_{14}$$

We can count the blocks of the join of two partitions by drawing them one beneath the other and counting the number of connected components of the curve, without taking into account the possible crossings. We conclude that we have the following formula, where ρ_{14}' is ρ_{14} restricted to the set $\{2k+1,2k+2\ldots,2p\}$:

$$|\widetilde{\pi}\vee\rho_{14}|=1+|\widetilde{\pi'}\vee\rho_{14}'|$$

If k is odd, $k=2r+1$, there is no extra block appearing, so we have:

$$|\widetilde{\pi}\vee\rho_{14}|=|\widetilde{\pi'}\vee\rho_{14}'|$$

Thus, we are led to the conclusion in the statement. □

We can now investigate the block-transposed Wishart matrices, and we have:

THEOREM 14.28. *For any $p\geq 1$ we have the formula*

$$\lim_{d\to\infty}(E\circ tr)\big(m\tilde{W}\big)^p=\sum_{\pi\in NC(p)}m^{|\pi|}n^{||\pi||}$$

where $|.|$ and $||.||$ are the number of blocks, and the number of blocks of even size.

Proof. The matrix elements of the partial transpose matrix are given by:

$$\tilde{W}_{ia,jb} = W_{ib,ja} = (dm)^{-1}\sum_{k=1}^{d}\sum_{c=1}^{m} G_{ib,kc}\bar{G}_{ja,kc}$$

This gives the following formula:

$$\begin{aligned}
tr(\tilde{W}^p) &= (dn)^{-1}(dm)^{-p}\sum_{i_1,\ldots,i_p=1}^{d}\sum_{a_1,\ldots,a_p=1}^{n}\prod_{s=1}^{p} W^{\Gamma}_{i_s a_s, i_{s+1}a_{s+1}}\\
&= (dn)^{-1}(dm)^{-p}\sum_{i_1,\ldots,i_p=1}^{d}\sum_{a_1,\ldots,a_p=1}^{n}\prod_{s=1}^{p} W_{i_s a_{s+1}, i_{s+1}a_s}\\
&= (dn)^{-1}(dm)^{-p}\sum_{i_1,\ldots,i_p=1}^{d}\sum_{a_1,\ldots,a_p=1}^{n}\prod_{s=1}^{p}\sum_{j_1,\ldots,j_p=1}^{d}\sum_{b_1,\ldots,b_p=1}^{m} G_{i_s a_{s+1}, j_s b_s}\bar{G}_{i_{s+1}a_s, j_s b_s}
\end{aligned}$$

The average of the general term can be computed by the Wick rule, namely:

$$E\left(\prod_{s=1}^{p} G_{i_s a_{s+1}, j_s b_s}\bar{G}_{i_{s+1}a_s, j_s b_s}\right) = \sum_{\pi\in S_p}\prod_{s=1}^{p}\delta_{i_s, i_{\pi(s)+1}}\delta_{a_{s+1}, a_{\pi(s)}}\delta_{j_s, j_{\pi(s)}}\delta_{b_s, b_{\pi(s)}}$$

Let $\gamma\in S_p$ be the full cycle $\gamma = (1\,2\,\ldots\,p)^{-1}$. The general factor in the above product is 1 if and only if the following four conditions are simultaneously satisfied:

$$\gamma^{-1}\pi\leq\ker i \quad , \quad \pi\gamma\leq\ker a \quad , \quad \pi\leq\ker j \quad , \quad \pi\leq\ker b$$

Counting the number of free parameters in the above equation, we obtain:

$$\begin{aligned}
(E\circ tr)(\tilde{W}^p) &= (dn)^{-1}(dm)^{-p}\sum_{\pi\in S_p} d^{|\pi|+|\gamma^{-1}\pi|}m^{|\pi|}n^{|\pi\gamma|}\\
&= \sum_{\pi\in S_p} d^{|\pi|+|\gamma^{-1}\pi|-p-1}m^{|\pi|-p}n^{|\pi\gamma|-1}
\end{aligned}$$

The exponent of d in the last expression on the right is:

$$\begin{aligned}
N(\pi) &= |\pi|+|\gamma^{-1}\pi|-p-1\\
&= p-1-(\#\pi+\#(\gamma^{-1}\pi))\\
&= p-1-(\#\pi+\#(\pi^{-1}\gamma))
\end{aligned}$$

As explained in the beginning of this section, this quantity is known to be ≤ 0, with equality iff π is geodesic, hence associated to a noncrossing partition. Thus:

$$(E\circ tr)(\tilde{W}^p) = (1+O(d^{-1}))m^{-p}n^{-1}\sum_{\pi\in NC(p)} m^{|\pi|}n^{|\pi\gamma|}$$

Together with $|\pi\gamma| = ||\pi||+1$, this gives the result. □

We would like now to find an equation for the moment generating function of the asymptotic law of $m\tilde{W}$. This moment generating function is defined by:

$$F(z)=\lim_{d\to\infty}(E\circ tr)\left(\frac{1}{1-zm\tilde{W}}\right)$$

We have the following result, regarding this moment generating function:

THEOREM 14.29. *The moment generating function of $m\tilde{W}$ satisfies the equation*

$$(F-1)(1-z^2F^2)=mzF(1+nzF)$$

in the $d\to\infty$ limit.

PROOF. We use the formula in Theorem 14.28. If we denote by $N(p,b,e)$ the number of partitions in $NC(p)$ having b blocks and e even blocks, we have:

$$\begin{aligned}F&=1+\sum_{p=1}^\infty\sum_{\pi\in NC(p)}z^pm^{|\pi|}n^{||\pi||}\\&=1+\sum_{p=1}^\infty\sum_{b=0}^\infty\sum_{e=0}^\infty z^pm^bn^eN(p,b,e)\end{aligned}$$

Let us try to find a recurrence formula for the numbers $N(p,b,e)$. If we look at the block containing 1, this block must have $r\geq 0$ other legs, and we get:

$$\begin{aligned}N(p,b,e)&=\sum_{r\in2\mathbb{N}}\sum_{p=\Sigma p_i+r+1}\sum_{b=\Sigma b_i+1}\sum_{e=\Sigma e_i}N(p_1,b_1,e_1)\ldots N(p_{r+1},b_{r+1},e_{r+1})\\&+\sum_{r\in2\mathbb{N}+1}\sum_{p=\Sigma p_i+r+1}\sum_{b=\Sigma b_i+1}\sum_{e=\Sigma e_i+1}N(p_1,b_1,e_1)\ldots N(p_{r+1},b_{r+1},e_{r+1})\end{aligned}$$

Here $p_1,\ldots,p_{r+1}$ are the number of points between the legs of the block containing 1, so that we have $p=(p_1+\ldots+p_{r+1})+r+1$, and the whole sum is split over two cases, r even or odd, because the parity of r affects the number of even blocks of our partition. Now by multiplying everything by a $z^pm^bn^e$ factor, and by carefully distributing the various powers of z,m,b on the right, we obtain the following formula:

$$\begin{aligned}z^pm^bn^eN(p,b,e)&=m\sum_{r\in2\mathbb{N}}z^{r+1}\sum_{p=\Sigma p_i+r+1}\sum_{b=\Sigma b_i+1}\sum_{e=\Sigma e_i}\prod_{i=1}^{r+1}z^{p_i}m^{b_i}n^{e_i}N(p_i,b_i,e_i)\\&+mn\sum_{r\in2\mathbb{N}+1}z^{r+1}\sum_{p=\Sigma p_i+r+1}\sum_{b=\Sigma b_i+1}\sum_{e=\Sigma e_i+1}\prod_{i=1}^{r+1}z^{p_i}m^{b_i}n^{e_i}N(p_i,b_i,e_i)\end{aligned}$$

Let us sum now all these equalities, over all $p\geq 1$ and over all $b,e\geq 0$. According to the definition of F, at left we obtain $F-1$. As for the two sums appearing on the right, that is, at right of the two z^{r+1} factors, when summing them over all $p\geq 1$ and over all $b,e\geq 0$, we obtain in both cases F^{r+1}. So, we have the following formula:

$$\begin{aligned} F-1 &= m\sum_{r\in 2\mathbb{N}}(zF)^{r+1}+mn\sum_{r\in 2\mathbb{N}+1}(zF)^{r+1}\\ &= m\,\frac{zF}{1-z^2F^2}+mn\,\frac{z^2F^2}{1-z^2F^2}\\ &= mzF\,\frac{1+nzF}{1-z^2F^2}\end{aligned}$$

But this gives the formula in the statement, and we are done. □

We can reformulate Theorem 14.29 as follows:

THEOREM 14.30. *The Cauchy transform of $m\tilde{W}$ satisfies the equation*

$$(\xi G-1)(1-G^2)=mG(1+nG)$$

in the $d\to\infty$ limit. Moreover, this equation simply reads

$$R=\frac{m}{2}\left(\frac{n+1}{1-z}-\frac{n-1}{1+z}\right)$$

with the substitutions $G\to z$ and $\xi\to R+z^{-1}$.

PROOF. We have two assertions to be proved, the idea being as follows:

(1) Consider the equation of F, found in Theorem 14.29, namely:

$$(F-1)(1-z^2F^2)=mzF(1+nzF)$$

With $z\to\xi^{-1}$ and $F\to\xi G$, so that $zF\to G$, we obtain, as desired:

$$(\xi G-1)(1-G^2)=mG(1+nG)$$

(2) Thus, we have our equation for the Cauchy transform, and with this in hand, we can try to go ahead, and use somehow the Stieltjes inversion formula, in order to reach to a formula for the density. This is certainly possible, but our claim is that we can do better, by performing first some clever manipulations on the Cauchy transform.

(3) To be more precise, with $\xi\to K$ and $G\to z$, this equation becomes:

$$(zK-1)(1-z^2)=mz(1+nz)$$

The point now is that with $K\to R+z^{-1}$ this latter equation becomes:

$$zR(1-z^2)=mz(1+nz)$$

But the solution of this latter equation is trivial to compute, given by:

$$R=m\,\frac{1+nz}{1-z^2}=\frac{m}{2}\left(\frac{n+1}{1-z}-\frac{n-1}{1+z}\right)$$

Thus, we are led to the conclusion in the statement. □

All the above suggests the following definition:

Definition 14.31. *Given a real probability measure μ, define its R-transform by:*

$$G_\mu(\xi)=\int_\mathbb{R}\frac{d\mu(t)}{\xi-t}\implies G_\mu\left(R_\mu(\xi)+\frac{1}{\xi}\right)=\xi$$

That is, the R-transform is the inverse of the Cauchy transform, up to a ξ^{-1} factor.

Getting back now to our questions, we would like to find the probability measure having as R-transform the function in Theorem 14.30. But here, we can only expect to find some kind of modification of the Marchenko-Pastur law, so as a first piece of work, let us just compute the R-transform of the Marchenko-Pastur law. We have here:

Proposition 14.32. *The R-transform of the Marchenko-Pastur law π_t is*

$$R_{\pi_t}(\xi)=\frac{t}{1-\xi}$$

for any $t>0$.

Proof. This can be done in two steps, as follows:

(1) At $t=1$, we know that the moments of π_1 are the Catalan numbers, $M_k=C_k$, and we obtain that the Cauchy transform is given by the following formula:

$$G(\xi)=\frac{1}{2}-\frac{1}{2}\sqrt{1-4\xi^{-1}}$$

Now with $R(\xi)=\frac{1}{1-\xi}$ being the function in the statement, at $t=1$, we have:

$$\begin{aligned}G\left(R(\xi)+\frac{1}{\xi}\right)&=G\left(\frac{1}{1-\xi}+\frac{1}{\xi}\right)\\&=G\left(\frac{1}{\xi-\xi^2}\right)\\&=\frac{1}{2}-\frac{1}{2}\sqrt{1-4\xi+4\xi^2}\\&=\frac{1}{2}-\frac{1}{2}(1-2\xi)\\&=\xi\end{aligned}$$

Thus, the function $R(\xi)=\frac{1}{1-\xi}$ is indeed the R-transform of π_1, in the above sense.

(2) In the general case, $t>0$, the proof is similar, by using the moment formula for π_t, that we know from the above, and we will leave this as an exercise. □

All this is very nice, and we can now further build on Theorem 14.30, as follows:

THEOREM 14.33. *The R-transform of $m\tilde{W}$ is given by*

$$R = R_{\pi_s} - R_{\pi_t}$$

in the $d \to \infty$ limit, where $s = m(n+1)/2$ and $t = m(n-1)/2$.

PROOF. We know from Theorem 14.30 that the R-transform of $m\tilde{W}$ is given by:

$$R = \frac{m}{2}\left(\frac{n+1}{1-z} - \frac{n-1}{1+z}\right)$$

By using now the formula in Proposition 14.32, this gives the result. □

We can now formulate a final result, due to Aubrun, as follows:

THEOREM 14.34. *For a block-transposed Wishart matrix $\tilde{W} = (id \otimes t)W$ we have, in the $n = \beta m \to \infty$ limit, with $\beta > 0$ fixed, the formula*

$$\frac{\tilde{W}}{d} \sim \gamma_\beta^1$$

with γ_β^1 being the shifted version of the semicircle law γ_β, with support centered at 1.

PROOF. This follows from Theorem 14.33. Indeed, in the $n = \beta m \to \infty$ limit, with $\beta > 0$ fixed, we are led to the following formula for the Stieltjes transform:

$$f(x) = \frac{\sqrt{4\beta - (1-x)^2}}{2\beta\pi}$$

But this is the density of the shifted semicircle law having support as follows:

$$S = [1 - 2\sqrt{\beta}, 1 + 2\sqrt{\beta}]$$

Thus, we are led to the conclusion in the statement. □

14e. Exercises

This was a quite exciting chapter, and as exercises on this, we have:

EXERCISE 14.35. *Compute the spectral measures of various matrices, of your choice.*

EXERCISE 14.36. *Learn more about C^*-algebras, including the GNS theorem.*

EXERCISE 14.37. *Importantly, learn as well about the von Neumann algebras.*

EXERCISE 14.38. *Learn as well some other approaches to the spectral measures.*

EXERCISE 14.39. *Is the asymptotic law of Gaussian matrices a "circular law"?*

EXERCISE 14.40. *Is the Wigner semicircle law some sort of "free Gaussian law"?*

EXERCISE 14.41. *Is the Marchenko-Pastur law some sort of "free Poisson law"?*

EXERCISE 14.42. *Read more, from Aubrun and others, about block modifications.*

As bonus exercise, learn some free probability theory, from [91].

CHAPTER 15

Circular systems

15a. Circular variables

We have seen so far that free probability theory leads to a remarkable free analogue of the CLT, with the limiting measure being the Wigner semicircle law. This is certainly something very interesting, theoretically speaking, and by reminding the fact that the Wigner laws appear in connection with many fundamental questions in mathematics, in relation with random walks on graphs, with Lie groups, and with random matrices as well, there are certainly many things to be done, as a continuation of this.

However, no hurry, and we will do this slowly. As a first objective, which is something quite straightforward, now that we have a free CLT, we would like to have as well a free analogue of the complex central limiting theorem (CCLT), adding to the classical CCLT, and providing us with free analogues Γ_t of the complex Gaussian laws G_t.

This will be something quite technical, and in order to get started, let us begin by recalling the theory of the complex Gaussian laws G_t. We first have:

DEFINITION 15.1. *The complex Gaussian law of parameter $t > 0$ is*

$$G_t = law\left(\frac{1}{\sqrt{2}}(a + ib)\right)$$

where a, b are independent, each following the law g_t.

There are many things that can be said about these laws, simply by adapting the known results from the real case, regarding the usual normal laws g_t. As a first such result, the above measures form convolution semigroups:

PROPOSITION 15.2. *The complex Gaussian laws have the property*

$$G_s * G_t = G_{s+t}$$

for any $s, t > 0$, and so they form a convolution semigroup.

PROOF. This is something that we know from chapter 5, coming from $g_s * g_t = g_{s+t}$, by taking the real and imaginary parts of all variables involved. □

We have as well the following complex analogue of the CLT:

THEOREM 15.3 (CCLT). *Given complex variables $f_1, f_2, f_3, \ldots \in L^\infty(X)$ which are i.i.d., centered, and with variance $t > 0$, we have, with $n \to \infty$, in moments,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_i \sim G_t$$

where G_t is the complex Gaussian law of parameter t.

PROOF. This is something that we know too from chapter 5, which follows from the real CLT, by taking real and imaginary parts. Indeed, let us write:

$$f_i = \frac{1}{\sqrt{2}}(x_i + iy_i)$$

The variables x_i satisfy then the assumptions of the CLT, so their rescaled averages converge to a normal law g_t, and the same happens for the variables y_i. The limiting laws that we obtain being independent, their rescaled sum is complex Gaussian, as desired. □

Regarding now the moments, we have here the following result:

PROPOSITION 15.4. *The moments of the complex normal law are the numbers*

$$M_k(G_t) = t^{|k|/2} |\mathcal{P}_2(k)|$$

where $\mathcal{P}_2(k)$ is the set of matching pairings of $\{1, \ldots, k\}$.

PROOF. This is again something that we know well too, from chapter 5, the idea being as follows, with $c = \frac{1}{\sqrt{2}}(a + ib)$ being the variable in Definition 15.1:

(1) In the case where k contains a different number of $\circ$ and $\bullet$ symbols, a rotation argument shows that the corresponding moment of c vanishes. But in this case we also have $\mathcal{P}_2(k) = \emptyset$, so the formula in the statement holds indeed, as $0 = 0$.

(2) In the case left, where k consists of p copies of $\circ$ and p copies of $\bullet$, the corresponding moment is the p-th moment of $|c|^2$, which by some calculus is $t^p p!$. But in this case we have as well $|\mathcal{P}_2(k)| = p!$, so the formula in the statement holds indeed, as $t^p p! = t^p p!$. □

As a final basic result regarding the laws G_t, we have the Wick formula:

THEOREM 15.5. *Given independent variables X_i, each following the complex normal law G_t, with $t > 0$ being a fixed parameter, we have the Wick formula*

$$E\left(X_{i_1}^{k_1} \ldots X_{i_s}^{k_s}\right) = t^{s/2} \# \left\{ \pi \in \mathcal{P}_2(k) \Big| \pi \leq \ker i \right\}$$

where $k = k_1 \ldots k_s$ and $i = i_1 \ldots i_s$, for the joint moments of these variables.

PROOF. This is something from chapter 5 too, the idea being as follows:

(1) In the case where we have a single complex normal variable X, we have to compute the moments of X, with respect to colored integer exponents $k = \circ\bullet\bullet\circ\ldots$, and the formula in the statement coincides with the one in Theorem 15.4, namely:

$$E(X^k) = t^{|k|/2}|\mathcal{P}_2(k)|$$

(2) In general now, when expanding $X_{i_1}^{k_1}\ldots X_{i_s}^{k_s}$ and rearranging the terms, we are left with doing a number of computations as in (1), then making the product of the numbers that we found. But this amounts in counting the partitions in the statement. □

Let us discuss now the free analogues of the above results. As in the classical case, there is actually not so much work to be done here, in order to get started, because we can obtain the free convolution and central limiting results, simply by taking the real and imaginary parts of our variables. Following Voiculescu [89], [90], we first have:

Definition 15.6. *The Voiculescu circular law of parameter $t > 0$ is given by*

$$\Gamma_t = law\left(\frac{1}{\sqrt{2}}(a+ib)\right)$$

where a, b are free, each following the Wigner semicircle law γ_t.

In other words, the passage $\gamma_t \to \Gamma_t$ is by definition entirely similar to the passage $g_t \to G_t$ from the classical case, by taking real and imaginary parts. As before in other similar situations, the fact that Γ_t is indeed well-defined is clear from definitions.

Let us start with a number of straightforward results, obtained by complexifying the free probability theory that we have. As a first result, we have, as announced above:

Proposition 15.7. *The Voiculescu circular laws have the property*

$$\Gamma_s \boxplus \Gamma_t = \Gamma_{s+t}$$

so they form a 1-parameter semigroup with respect to free convolution.

Proof. This follows from our result from chapter 13 stating that the Wigner laws γ_t have the free semigroup convolution property, by taking real and imaginary parts. □

Next in line, also as announced above, and also from [90], we have the following natural free analogue of the complex central limiting theorem (CCLT):

Theorem 15.8 (Free CCLT). *Given random variables $x_1, x_2, x_3, \ldots$ which are f.i.d., centered, with variance $t > 0$, we have, with $n \to \infty$, in moments,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i \sim \Gamma_t$$

where Γ_t is the Voiculescu circular law of parameter t.

PROOF. This follows indeed from the free CLT, established in chapter 13, by taking real and imaginary parts. Indeed, let us write:

$$x_i = \frac{1}{\sqrt{2}}(y_i + iz_i)$$

The variables y_i satisfy then the assumptions of the free CLT, and so their rescaled averages converge to a semicircle law γ_t, and the same happens for the variables z_i:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n y_i \sim \gamma_t \quad , \quad \frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \sim \gamma_t$$

Now since the two limiting semicircle laws that we obtain in this way are free, their rescaled sum is circular, in the sense of Definition 15.6, and this gives the result. □

Summarizing, we have so far complex analogues of both the classical and free CLT, and the basic theory of the limiting measures, including their semigroup property. As a conclusion to all this, let us formulate the following statement:

THEOREM 15.9. *We have classical and free limiting theorems, as follows,*

$$\begin{array}{ccc} FCLT & — & FCCLT \\ | & & | \\ CLT & — & CCLT \end{array}$$

the limiting laws being the following measures,

$$\begin{array}{ccc} \gamma_t & — & \Gamma_t \\ | & & | \\ g_t & — & G_t \end{array}$$

which form classical and free convolution semigroups.

PROOF. This follows indeed from the various results established above. To be more precise, the results about the left edge of the square are from the previous chapter, and the results about the right edge are those discussed in the above. □

Going ahead with more study of the Voiculescu circular variables, less trivial now is the computation of their moments. We will do this in what follows, among others in order to expand Theorem 15.9 into something much sharper, involving as well moments.

For our computations, we will need explicit models for the circular variables. Following [90], and the material in chapter 13, let us start with the following key result:

Proposition 15.10. *Let H be the complex Hilbert space having as basis the colored integers $k = \circ\bullet\bullet\circ\ldots$, and consider the shift operators on this space:*

$$S : k \to \circ k \quad , \quad T : k \to \bullet k$$

We have then the following equalities of distributions,

$$S + S^* \sim \gamma_1 \quad , \quad S + T^* \sim \Gamma_1$$

with respect to the state $\varphi(T) =< Te, e >$, where e is the empty word.

Proof. This is standard free probability, the idea being as follows:

(1) The first formula, namely $S + S^* \sim \gamma_1$, is something that we already know, in a slightly different formulation, from chapter 13, when proving the CLT.

(2) As for the second formula, $S + T^* \sim \Gamma_1$, this follows from the first formula, by using the freeness results and the rotation tricks established in chapter 13. □

At the combinatorial level now, we have the following result, which is in analogy with the moment theory of the Wigner semicircle law, developed above:

Theorem 15.11. *A variable $a \in A$ follows the law Γ_1 precisely when its moments are*

$$tr(a^k) = |\mathcal{NC}_2(k)|$$

for any colored integer $k = \circ\bullet\bullet\circ\ldots$

Proof. By using Proposition 15.10, it is enough to do the computation in the model there. To be more precise, we can use the following explicit formulae for S, T:

$$S : k \to \circ k \quad , \quad T : k \to \bullet k$$

With these formulae in hand, our claim is that we have the following formula:

$$< (S + T^*)^k e, e >= |\mathcal{NC}_2(k)|$$

In order to prove this formula, we can proceed as for the semicircle laws, in chapter 9 above. Indeed, let us expand the quantity $(S + T^*)^k$, and then apply the state φ.

With respect to the previous computation, from chapter 13, what happens is that the contributions will come this time via the following formulae, which must succesively apply, as to collapse the whole product of S, S^*, T, T^* variables into a 1 quantity:

$$S^*S = 1 \quad , \quad T^*T = 1$$

As before, in the proof for the semicircle laws, from chapter 13, these applications of the rules $S^*S = 1$, $T^*T = 1$ must appear in a noncrossing manner, but what happens now, in contrast with the computation from the proof in chapter 9 where $S + S^*$ was self-adjoint, is that at each point where the exponent k has a $\circ$ entry we must use $T^*T = 1$, and at each point where the exponent k has a $\bullet$ entry we must use $S^*S = 1$. Thus the contributions, which are each worth 1, are parametrized by the partitions $\pi \in \mathcal{NC}_2(k)$. Thus, we obtain the above moment formula, as desired. □

More generally now, by rescaling, we have the following result:

THEOREM 15.12. *A variable $a \in A$ is circular, $a \sim \Gamma_t$, precisely when its moments are given by the formula*

$$tr(a^k) = t^{|k|/2} |\mathcal{NC}_2(k)|$$

for any colored integer $k = \circ \bullet \bullet \circ \ldots$

PROOF. This follows indeed from Theorem 15.11, by rescaling. Alternatively, we can get this as well directly, by suitably modifying Proposition 15.10 first. □

Even more generally now, we have the following free version of the Wick rule:

THEOREM 15.13. *Given free variables a_i, each following the Voiculescu circular law Γ_t, with $t > 0$ being a fixed parameter, we have the Wick type formula*

$$tr(a_{i_1}^{k_1} \ldots a_{i_s}^{k_s}) = t^{s/2} \# \left\{ \pi \in \mathcal{NC}_2(k) \Big| \pi \leq \ker i \right\}$$

where $k = k_1 \ldots k_s$ and $i = i_1 \ldots i_s$, for the joint moments of these variables, with the inequality $\pi \leq \ker i$ on the right being taken in a technical, appropriate sense.

PROOF. This follows a bit as in the classical case, the idea being as follows:

(1) In the case where we have a single complex normal variable a, we have to compute the moments of a, with respect to colored integer exponents $k = \circ \bullet \bullet \circ \ldots$, and the formula in the statement coincides with the one in Theorem 15.12, namely:

$$tr(a^k) = t^{|k|/2} |\mathcal{NC}_2(k)|$$

(2) In general now, when expanding the product $a_{i_1}^{k_1} \ldots a_{i_s}^{k_s}$ and rearranging the terms, we are left with doing a number of computations as in (1), and then making the product of the expectations that we found. But this amounts precisely in counting the partitions in the statement, with the condition $\pi \leq \ker i$ there standing precisely for the fact that we are doing the various type (1) computations independently. □

All the above was a bit brief, based on Voiculescu's original paper [**90**], and on his foundational free probability book with Dykema and Nica [**91**]. The combinatorics of the free families of circular variables, called "circular systems", is something quite subtle, and there has been a lot of work developed in this direction, since [**91**].

Getting back now to the case of the single variables, from Theorem 15.12, the formula there has the following more conceptual interpretation:

THEOREM 15.14. *The moments of the Voiculescu laws are the numbers*

$$M_k(\Gamma_t) = \sum_{\pi \in \mathcal{NC}_2(k)} t^{|\pi|}$$

with "$\mathcal{NC}_2$" standing for the noncrossing matching pairings.

Proof. This follows from the formula in Theorem 15.12. Indeed, we know from there that a variable $a \in A$ is circular, of parameter $t > 0$, precisely when we have the following formula, for any colored integer $k = \circ \bullet \bullet \circ \ldots$:

$$tr(a^k) = t^{|k|/2} |\mathcal{NC}_2(k)|$$

Now since the number of blocks of a pairing $\pi \in \mathcal{NC}_2(k)$ is given by $|\pi| = |k|/2$, this formula can be written in the following alternative way:

$$tr(a^k) = \sum_{\pi \in \mathcal{NC}_2(k)} t^{|\pi|}$$

Thus, we are led to the conclusion in the statement. □

All this is quite nice, when compared with the similar results from the classical case, regarding the complex Gaussian laws, that we established above, and with other results of the same type as well. As a conclusion to these considerations, we can now formulate a global result regarding the classical and free complex Gaussian laws, as follows:

Theorem 15.15. *The complex Gaussian laws G_t and the circular Voiculescu laws Γ_t, given by the formulae*

$$G_t = law\left(\frac{1}{\sqrt{2}}(a + ib)\right) \quad , \quad \Gamma_t = law\left(\frac{1}{\sqrt{2}}(\alpha + i\beta)\right)$$

where $a, b/\alpha, \beta$ are independent/free, following g_t/γ_t, have the following properties:

1. *They appear via the complex CLT, and the free complex CLT.*
2. *They form semigroups with respect to the operations $*$ and $\boxplus$.*
3. *Their moments are $M_k = \sum_{\pi \in D(k)} t^{|\pi|}$, with $D = \mathcal{P}_2, \mathcal{NC}_2$.*

Proof. This is a summary of results that we know, the idea being as follows:

(1) This is something quite straightforward, by using the linearization results provided by the logarithm of the Fourier transform, and by the R-transform.

(2) This is quite straightforward, too, once again by using the linearization results provided by the logarithm of the Fourier transform, and by the R-transform.

(3) This comes by doing some combinatorics and calculus in the classical case, and some combinatorics and operator theory in the free case, as explained above. □

More generally now, we can put everything together, with some previous results included as well, and we have the following result at the level of the moments of the asymptotic laws that we found so far, in classical and free probability:

THEOREM 15.16. *The moments of the various central limiting measures, namely*

$$\begin{array}{ccc} \gamma_t & - & \Gamma_t \\ | & & | \\ g_t & - & G_t \end{array}$$

are always given by the same formula, involving partitions, namely

$$M_k = \sum_{\pi \in D(k)} t^{|\pi|}$$

where the sets of partitions $D(k)$ in question are respectively

$$\begin{array}{ccc} NC_2 & \leftarrow & \mathcal{NC}_2 \\ \downarrow & & \downarrow \\ P_2 & \leftarrow & \mathcal{P}_2 \end{array}$$

and where $|.|$ is the number of blocks.

PROOF. This follows by putting together the various moment results that we have, from the previous chapter, and from Theorem 15.15. □

Summarizing, we are done with the combinatorial program outlined in the beginning of the present chapter. We will be back to this in the next chapter, by adding some new laws to the picture, coming from the classical and free PLT and CPLT, and then in the chapter afterwards, 16 below, with full conceptual explanations for all this.

15b. Multiplicative results

With the above basic combinatorial study done, let us discuss now a number of more advanced results regarding the Voiculescu circular laws Γ_t, which are of multiplicative nature, and quite often have no classical counterpart. Things here will be quite technical, and all that follows will be rather an introduction to the subject.

In general now, in order to deal with multiplicative questions for the free random variables, we are in need of results regarding the multiplicative free convolution operation $\boxtimes$. Let us recall from chapter 13 that we have the following result:

DEFINITION 15.17. *We have a free convolution operation $\boxtimes$, constructed as follows:*

(1) *For abstract distributions, via $\mu_a \boxtimes \mu_b = \mu_{ab}$, with a, b free.*
(2) *For real measures, via $\mu_a \boxtimes \mu_b = \mu_{\sqrt{a}b\sqrt{a}}$, with a, b self-adjoint and free.*

All this is quite tricky, explained in chapter 13, the idea being that, while (1) is straightforward, (2) is not, and comes by considering the variable $c=\sqrt{a}b\sqrt{a}$, which unlike ab is always self-adjoint, and whose moments are given by:

$$\begin{aligned} tr(c^k) &= tr[(\sqrt{a}b\sqrt{a})^k] \\ &= tr[\sqrt{a}ba\ldots ab\sqrt{a}] \\ &= tr[\sqrt{a}\cdot\sqrt{a}ba\ldots ab] \\ &= tr[(ab)^k] \end{aligned}$$

As a remark here, observe that we have used in the above, and actually for the first time since talking about freeness, the trace property of the trace, namely:

$$tr(ab)=tr(ba)$$

This is quite interesting, philosophically speaking, because in the operator algebra world there are many interesting examples of subalgebras $A\subset B(H)$ coming with natural linear forms $\varphi:A\to\mathbb{C}$ which are continuous and positive, but which are not traces. It is possible to do a bit of free probability on such algebras, but not much.

Quite remarkably, the free multiplicative convolution operation $\boxtimes$ can be linearized, in analogy with what happens for the usual multiplicative convolution $\times$, and the additive operations $*,\boxplus$ as well. We have here the following result, due to Voiculescu [90]:

Theorem 15.18. *The free multiplicative convolution operation $\boxtimes$ for the real probability measures $\mu\in\mathcal{P}(\mathbb{R})$ can be linearized as follows:*

(1) *Start with the sequence of moments M_k, then compute the moment generating function, or Stieltjes transform of the measure:*

$$f(z)=1+M_1z+M_2z^2+M_3z^3+\ldots$$

(2) *Perform the following operations to the Stieltjes transform:*

$$\psi(z)=f(z)-1$$

$$\psi(\chi(z))=z$$

$$S(z)=\left(1+\frac{1}{z}\right)\chi(z)$$

(3) *Then $\log S$ linearizes the free multiplicative convolution, $S_{\mu\boxtimes\nu}=S_\mu S_\nu$.*

Proof. There are several proofs here, with the original proof of Voiculescu being quite similar to the proof of the R-transform theorem, using free Fock space models, then with a proof by Haagerup, obtained by further improving on this, and finally with the proof from the book of Nica and Speicher, using pure combinatorics. The proof of Haagerup, which is the most in tune with the present book, is as follows:

(1) According to our conventions from Definition 15.17, we want to prove that, given noncommutative variables a, b which are free, we have the following formula:

$$S_{\mu_{ab}}(z) = S_{\mu_a}(z)S_{\mu_b}(z)$$

(2) For this purpose, consider the orthogonal shifts S, T on the free Fock space, as in chapter 9. By using the algebraic arguments from chapter 13, from the proof of the R-transform theorem, we can assume as there that our variables have a special form, that fits our present objectives, and to be more specifically, the following form:

$$a = (1+S)f(S^*) \quad , \quad b = (1+T)g(T^*)$$

Our claim, which will prove the theorem, is that we have the following formulae, for the S-transforms of the various variables involved:

$$S_{\mu_a}(z) = \frac{1}{f(z)} \quad , \quad S_{\mu_b}(z) = \frac{1}{g(z)} \quad , \quad S_{\mu_{ab}}(z) = \frac{1}{f(z)g(z)}$$

(3) Let us first compute S_{μ_a}. We know that we have $a = (1+S)f(S^*)$, with S being the shift on $l^2(\mathbb{N})$. Given $|z| < 1$, consider the following vector:

$$p = \sum_{k\geq 0} z^k e_k$$

The shift and its adjoint act on this vector in the following way:

$$Sp = \sum_{k\geq 0} z^k e_{k+1} = \frac{p - e_0}{z}$$

$$S^*p = \sum_{k\geq 1} z^k e_{k-1} = zp$$

Thus $f(S^*)p = f(z)p$, and we deduce from this that we have:

$$\begin{aligned} ap &= (1+S)f(z)p \\ &= f(z)(p + Sp) \\ &= f(z)\left(p + \frac{p-e_0}{z}\right) \\ &= \left(1 + \frac{1}{z}\right)f(z)p - \frac{f(z)}{z}e_0 \end{aligned}$$

By dividing everything by $(1 + 1/z)f(z)$, this formula becomes:

$$\frac{z}{1+z} \cdot \frac{1}{f(z)}\, ap = p - \frac{e_0}{1+z}$$

We can write this latter formula in the following way:

$$\left(1 - \frac{z}{1+z} \cdot \frac{1}{f(z)}\, a\right) p = \frac{e_0}{1+z}$$

Now by inverting, we obtain from this the following formula:

$$\left(1-\frac{z}{1+z}\cdot\frac{1}{f(z)}\,a\right)^{-1}e_0=(1+z)p$$

(4) But this gives us the formula of S_{μ_a}. Indeed, consider the following function:

$$\rho(z)=\frac{z}{1+z}\cdot\frac{1}{f(z)}$$

With this notation, the formula that we found in (3) becomes:

$$(1-\rho(z)a)^{-1}e_0=(1+z)p$$

By using this, in terms of $\varphi(T)=<Te_0,e_0>$, we obtain:

$$\begin{aligned}\varphi\left((1-\rho(z)a)^{-1}\right)&=&<(1-\rho(z)a)^{-1}e_0,e_0>\\&=&<(1+z)p,e_0>\\&=&1+z\end{aligned}$$

Thus the above function ρ is the inverse of the following function:

$$\psi(z)=\varphi\left(\frac{1}{1-za}\right)-1$$

But this latter function is the ψ function from the statement, and so ρ is the function χ from the statement, and we can finish our computation, as follows:

$$\begin{aligned}S_{\mu_a}(z)&=&\frac{1+z}{z}\cdot\rho(z)\\&=&\frac{1+z}{z}\cdot\frac{z}{1+z}\cdot\frac{1}{f(z)}\\&=&\frac{1}{f(z)}\end{aligned}$$

(5) A similar computation, or just a symmetry argument, gives $S_{\mu_b}(z)=1/g(z)$. In order to compute now $S_{\mu_{ab}}(z)$, we use a similar trick. Consider the following vector of $l^2(\mathbb{N}*\mathbb{N})$, with the primes and double primes referring to the two copies of $\mathbb{N}$:

$$q=e_0+\sum_{k\geq1}(e_1'+e_1''+e_1'\otimes e_1'')^{\otimes k}$$

The adjoints of the shifts S,T act as follows on this vector:

$$S^*q=z(1+T)q\quad,\quad T^*q=zq$$

By using these formulae, we have the following computation:

$$\begin{aligned} abq &= (1+S)f(S^*)(1+T)g(T^*)q \\ &= (1+S)f(S^*)(1+T)g(z)q \\ &= g(z)(1+S)f(S^*)(1+T)q \end{aligned}$$

In order to compute the last term, observe that we have:

$$\begin{aligned} S^*(1+T)q &= (S^*+S^*T)q \\ &= S^*q \\ &= z(1+T)q \end{aligned}$$

Thus $f(S^*)(1+T)q = f(z)(1+T)q$, and back to our computation, we have:

$$\begin{aligned} abq &= g(z)(1+S)f(z)(1+T)q \\ &= f(z)g(z)(1+S)(1+T)q \\ &= f(z)g(z)\left(\frac{1+z}{z}\cdot q - \frac{e_0}{z}\right) \end{aligned}$$

Now observe that we can write this formula as follows:

$$\left(1 - \frac{z}{1+z}\cdot\frac{1}{f(z)g(z)}\cdot ab\right)q = \frac{e_0}{1+z}$$

By inverting, we obtain from this the following formula:

$$\left(1 - \frac{z}{1+z}\cdot\frac{1}{f(z)g(z)}\cdot ab\right)^{-1} e_0 = (1+z)q$$

(6) But this formula that we obtained is similar to the formula that we obtained at the end of (3) above. Thus, we can use the same argument as in (4), and we obtain:

$$S_{\mu_{ab}}(z) = \frac{1}{f(z)g(z)}$$

We are therefore done with the computations, and this finishes the proof. □

Getting back now to the circular variables, let us look at the polar decomposition of such variables. In order to discuss this, let us start with a well-known result:

THEOREM 15.19. *We have the following results:*

1. *Any matrix $T \in M_N(\mathbb{C})$ has a polar decomposition, $T = U|T|$.*
2. *Assuming $T \in A \subset M_N(\mathbb{C})$, we have $U, |T| \in A$.*
3. *Any operator $T \in B(H)$ has a polar decomposition, $T = U|T|$.*
4. *Assuming $T \in A \subset B(H)$, we have $U, |T| \in \bar{A}$, weak closure.*

Proof. All this is standard, the idea being as follows:

(1) In each case under consideration, the first observation is that the matrix or general operator T^*T being positive, it has a square root:

$$|T| = \sqrt{T^*T}$$

(2) With this square root extracted, in the invertible case we can compare the action of T and $|T|$, and we conclude that we have $T = U|T|$, with U being a unitary. In the general, non-invertible case, a similar analysis leads to the conclusion that we have as well $T = U|T|$, but with U being this time a partial isometry.

(3) In what regards now algebraic and topological aspects, in finite dimensions the extraction of the square root, and so the polar decomposition itself, takes place over the matrix blocks of the ambient algebra $A \subset M_N(\mathbb{C})$, and so takes place inside A itself.

(4) In infinite dimensions however, we must take the weak closure, an illustrating example here being the functions $f \in A$ belonging to the algebra $A = C(X)$, represented on $H = L^2(X)$, whose polar decomposition leads into the bigger algebra $\bar{A} = L^\infty(X)$. □

Summarizing, we have a basic linear algebra result, regarding the polar decomposition of the usual matrices, and in infinite dimensions pretty much the same happens, with the only subtlety coming from the fact that the ambient operator algebra $A \subset B(H)$ must be taken weakly closed. We will be back to this, with more details, in chapter 16 below, when talking about such algebras $A \subset B(H)$, which are called von Neumann algebras.

In connection with our probabilistic questions, we first have the following result:

Proposition 15.20. *The polar decomposition of semicircular variables is $s = eq$, with the variables e, q being as follows:*

1. *e has moments $1, 0, 1, 0, 1, \ldots$*
2. *q is quarter-circular.*
3. *e, q are independent.*

Proof. It is enough to prove the result in a model of our choice, and the best choice here is the most straightforward model for the semicircular variables, namely:

$$s = x \in L^\infty\Big([-2, 2], \gamma_1\Big)$$

To be more precise, we endow the interval $[-2, 2]$ with the probability measure γ_1, and we consider here the variable $s = x = (x \to x)$, which is trivially semicircular. The polar decomposition of this variable is then $s = eq$, with e, q being as follows:

$$e = sgn(x) \quad , \quad q = |x|$$

Now since e has moments $1, 0, 1, 0, 1, \ldots$, and also q is quarter-circular, and finally e, q are independent, this gives the result in our model, and so in general. □

Less trivial now is the following result, due to Voiculescu [90]:

THEOREM 15.21. *The polar decomposition of circular variables is $c = uq$, with the variables u, q being as follows:*

(1) *u is a Haar unitary.*
(2) *q is quarter-circular.*
(3) *u, q are free.*

PROOF. This is something which looks quite similar to Proposition 15.20, but which is more difficult, and can be however proved, via various techniques:

(1) The original proof, by Voiculescu in [90], uses Gaussian random matrix models for the circular variables. We will discuss this proof at the end of the present chapter, after developing the needed Gaussian random matrix model technology.

(2) A second proof can be obtained by pure combinatorics, in the spirit of Theorem 15.13, regarding the free Wick formula, and of Theorem 15.18, regarding the S-transform, or rather in the spirit of the underlying combinatorics of these results.

(3) Finally, there is as well a third known proof, more in the spirit of the free Fock space proofs for the R and S transform results, from [90], using a suitable generalization of the free Fock spaces. We will discuss this proof right below. □

15c. Semigroup models

We discuss here the direct approach to Theorem 15.21, with purely algebraic techniques. We will use semigroup algebras, jointly generalizing the main models that we have, namely group algebras, and free Fock spaces. Let us start with:

DEFINITION 15.22. *We call "semigroup" a unital semigroup, embeddable into a group:*

$$M \subset G$$

For such a semigroup M, we use the notation

$$M^{-1} = \left\{ m^{-1} \middle| m \in M \right\}$$

regarded as a subset of some group G containing M, as above.

As a first observation, the above embeddability assumption $M \subset G$ tells us that the usual group cancellation rules hold in M, namely:

$$ab = ac \implies b = c$$

$$ba = ca \implies b = c$$

Regarding the precise relation between M and the various groups G containing it, it is possible to talk here about the Grothendieck group G associated to such a semigroup M. However, we will not need this in what follows, and use Definition 15.22 as such.

With the above definition in hand, we have the following construction, which unifies the main models that we have, namely the group algebras, and the free Fock spaces:

PROPOSITION 15.23. *Let M be a semigroup. By using the left simplifiability of M we can define, as for the discrete groups, an embedding of semigroups, as follows:*

$$(M,\cdot)\to(B(l^2(M)),\circ)$$

$$m\to\lambda_M(m)=[\delta_n\to\delta_{mn}]$$

Via this embedding, the C^*-algebra $C^*(M)\subset B(l^2(M))$ generated by $\lambda_M(M)$, together with the following canonical state, is a noncommutative random variable algebra:

$$\tau_M(T)=<T\delta_e,\delta_e>$$

Also, the operators in $\lambda_M(M)$ are isometries, but not necessarily unitaries.

PROOF. Everything here is standard, as for the usual group algebras, with the only subtlety appearing at the level of the isometry property of the operators $\lambda_M(m)$. To be more precise, for every $m\in M$, the adjoint operator $\lambda_M(m)^*$ is given by:

$$\lambda_M(m)^*(\delta_n)=\sum_{x\in M}<\lambda_M(m)^*\delta_n,\delta_x>\delta_x=\sum_{x\in M}\delta_{n,mx}\delta_x$$

Thus we have indeed the isometry property for these operators, namely:

$$\lambda_M(m)^*\lambda_M(m)=1$$

As for the unitarity propety of the such operators, this definitely holds in the usual discrete group case, $M=G$, but not in general. As a basic example here, for the semigroup $M=\mathbb{N}$, which satisfies of course the assumptions in Definition 15.22, the operator $\lambda_M(m)$ associated to the element $m=1\in\mathbb{N}$ is the usual shift:

$$\lambda_{\mathbb{N}}(1)=S\in B(l^2(\mathbb{N}))$$

But this shift S, that we know well from the above, is an isometry which is not a unitary. Thus, we are led to the conclusions in the statement. □

At the level of examples now, as announced above, we have:

PROPOSITION 15.24. *The construction $M\to C^*(M)$ is as follows:*

1. *For the discrete groups, $M=G$, we obtain in this way the usual discrete group algebras $C^*(G)$, as previously constructed in the above.*
2. *For a free semigroup, $M=\mathbb{N}^{*I}$, we obtain the algebra of creation operators over the full Fock space over $\mathbb{R}^I$, with the state associated to the vacuum vector.*

PROOF. All this is clear from definitions, with (1) being obvious, and (2) coming via our usual identifications for the free Fock spaces and related algebras. □

As a key observation now, enabling us to do some probability, we have:

Proposition 15.25. *If $M\subset N$ are semigroups satisfying the condition*

$$M(N-M)=N-M$$

then for every family $\{a_i\}_{i\in I}$ of elements in M, we have the formula

$$\{\lambda_N(a_i)\}_{i\in I}\sim\{\lambda_M(a_i)\}_{i\in I}$$

as an equality of joint distributions, with respect to the canonical states.

Proof. Assuming $M\subset N$ we have $l^2(M)\subset l^2(N)$, and for $m,m'\in M$ we have:

$$\lambda_M(m)\delta_{m'}=\lambda_N(m)\delta_{m'}$$

Thus if we suppose $M(N-M)=N-M$, as in the statement, then we have:

$$\begin{aligned}\lambda_M(m)^*\delta_{m'}&=\sum_{x\in M}\delta_{m',mx}\delta_x\\&=\sum_{x\in N}\delta_{m',mx}\delta_x\\&=\lambda_N(m)^*\delta_{m'}\end{aligned}$$

In particular, if $m_1,\ldots,m_k\in M$, and $\alpha_1,\ldots,\alpha_k$ are exponents in $\{1,*\}$, then:

$$\lambda_M(m_1)^{\alpha_1}\ldots\lambda_M(m_k)^{\alpha_k}\delta_e=\lambda_N(m_1)^{\alpha_1}\ldots\lambda_N(m_k)^{\alpha_k}\delta_e$$

Thus, we are led to the conclusion in the statement. □

Following [8], let us introduce the following technical notion:

Definition 15.26. *Let N be a semigroup. Consider the following order on it:*

$$a\preceq_N b\iff b\in aN$$

We say that N is in the class E if it satisfies one of the following equivalent conditions:

1. *For $\preceq_N$ every bounded subset is totally ordered.*
2. *$a\preceq c,b\preceq c\implies a\preceq b$ or $b\preceq a$.*
3. *$aN\cap bN\neq\emptyset\implies aN\subset bN$ or $bN\subset aN$.*
4. *$NN^{-1}\cap N^{-1}N=N\cup N^{-1}$.*

Also by following [8], let us introduce as well the following notion, which is something standard in the combinatorial theory of semigroups:

Definition 15.27. *Let $(a_i)_{i\in I}$ be a family of elements in a semigroup N.*

1. *We say that $(a_i)_{i\in I}$ is a code if the semigroup $M\subset N$ generated by the a_i is isomorphic to $\mathbb{N}^{*I}$, via $a_i\to e_i$, and satisfies $M(N-M)=N-M$.*
2. *We say that $(a_i)_{i\in I}$ is a prefix if $a_i\in a_jN\implies i=j$, which means that the elements a_i are not comparable via the order relation $\preceq_N$.*

In our probabilistic setting, the notion of code is of interest, due to:

Proposition 15.28. *Assuming that $(a_i,b_i)_{i\in I}$ is a code, the family*

$$\left(\frac{1}{2}(\lambda_N(a_i)+\lambda_N(b_i)^*)\right)_{i\in I}$$

is a circular family, in the sense of free probability theory.

Proof. Let $(a_i,b_i)_{i\in I}$ be a code, and consider the following family:

$$\Big(\lambda_N(a_i),\lambda_N(b_i)\Big)_{i\in I}\in B(l^2(N))$$

By using Proposition 15.25, this family has the same distribution as a family of creation operators associated to a family of $2I$ orthonormal vectors, on the free Fock space:

$$\Big(\lambda_{\mathbb{N}^{*I}}(e_i),\lambda_{\mathbb{N}^{*I}}(f_i)\Big)_{i\in I}\in B(l^2(N^{*I}))$$

Thus, we obtain the result, via the standard facts about the circular systems on free Fock spaces, that we know from chapter 13. □

In view of this, the following result provides us with a criterion for finding circular systems in the algebras of the semigroups in the class E, from Definition 15.26:

Proposition 15.29. *For a semigroup $N\in E$, a family*

$$(a_i)_{i\in I}\subset N$$

having at least two elements is a prefix if and only if it is a code.

Proof. We have two implications to be proved, as follows:

(1) Let first $(a_i)_{i\in I}$ be a code which is not a prefix, for instance because we have $a_i=a_jn$ with $i\neq j,n\in N$. Then n is in the semigroup M generated by the a_k and $a_i=a_jn$ with $i\neq j$, so M cannot be free, and this is a contradiction, as desired.

(2) Conversely, suppose now that $(a_i)_{i\in I}$ is a prefix and let, with $m\in N$:

$$A=a_{i_1}^{\alpha_1}\ldots a_{i_n}^{\alpha_n}m=a_{j_1}^{\beta_1}\ldots a_{j_s}^{\beta_s}$$

We have then $a_{i_1}\preceq A$, $a_{j_1}\preceq A$, and so $i_1=j_1$. We can therefore simplify A to the left by a_{i_1}. A reccurence on $\sum\alpha_i$ shows then that we have $n\leq s$ and:

$$a_{i_k}=a_{j_k}\quad,\quad\forall k\leq n$$
$$\alpha_k=\beta_k\quad,\quad\forall k<n$$
$$\alpha_n\leq\beta_n$$
$$m=a_{j_n}^{\beta_n-\alpha_n}a_{j_{n+1}}^{\beta_{n+1}}\ldots a_{j_s}^{\beta_s}$$

Finally, we know that m is in the semigroup generated by the a_i, so we have a code. Moreover, for $m=e$ we obtain that we have $n=s$, $a_{j_k}=a_{i_k}$ and $\alpha_k=\beta_k$ for any $k\leq n$. Thus the variables a_i freely generate the semigroup M, and so the family $(a_i)_{i\in I}$ is a code. Thus, we are led to the conclusion in the statement. □

Summarizing, we have some good freeness results, for our semigroups. Before getting into applications, let us discuss now the examples. We have here the following result:

PROPOSITION 15.30. *The class E has the following properties:*

1. *All the groups are in E.*
2. *The positive parts of totally ordered abelian groups are in E.*
3. *If G is a group and $M \in E$, then $M \times G \in E$.*
4. *If A_1, A_2 are in E, then the free product $A_1 * A_2$ is in E.*

PROOF. This is something elementary, whose proof goes as follows:

(1) This is obvious, coming from definitions.

(2) This is obvious as well, because M is here totally ordered by $\preceq_M$.

(3) Let G be a group and $M \in E$. We have then, as desired:

$$\begin{aligned}
& (M \times G)(M \times G)^{-1} \cap (M \times G)^{-1}(M \times G) \\
= & (M \times G)(M^{-1} \times G) \cap (M^{-1} \times G)(M \times G) \\
= & (MM^{-1} \times G) \cap (M^{-1}M \times G) \\
= & (MM^{-1} \cap M^{-1}M) \times G \\
= & (M \cup M^{-1}) \times G \\
= & (M \times G) \cup (M^{-1} \times G) \\
= & (M \times G) \cup (M \times G)^{-1}
\end{aligned}$$

(4) Let $a, b, c \in A_1 * A_2$ such that $ab = c$. We write, as reduced words:

$$a = x_1 \dots x_n \quad , \quad b = y_1 \dots y_m \quad , \quad c = z_1 \dots z_p$$

Now let s be such that the following equalities happen:

$$x_n y_1 = 1 \quad , \quad \dots \quad , \quad x_{n-s+1} y_s = 1 \quad , \quad x_{n-s} y_{s+1} \neq 1$$

Consider now the following element:

$$u = x_{n-s+1} \dots x_n = (y_1 \dots y_s)^{-1}$$

We have then the following computation:

$$c = ab = x_1 \dots x_{n-s} y_{s+1} \dots y_m$$

Now let $i \in \{1, 2\}$ be such that $z_{n-s} \in A_i$. There are two cases:

– If $x_{n-s} \in A_1$ and $y_{s+1} \in A_2$ or if $x_{n-s} \in A_2$ and $y_{s+1} \in A_1$, then $x_1 \dots x_{n-s} y_{s+1} \dots y_m$ is a reduced word. In particular, we have $x_1 = z_1$, $x_2 = z_2$, and so on up to $x_{n-s} = z_{n-s}$. Thus we have $a = z_1 \dots z_{n-s} u$, with u invertible.

– If $x_{n-s}, y_{s+1} \in A_i$ then $x_1 = z_1$ and so on, up to $x_{n-s-1} = z_{n-s-1}$ and $x_{n-s} y_{s+1} = z_{n-s}$. In this case we have $a = z_1 \dots z_{n-s-1} x_{n-s} u$, with u invertible.

Now observe that in both cases we obtained that a is of the form $z_1 \ldots z_f xu$ for some f, with u invertible and such that if $z_{f+1} \in A_i$, then there exists $y \in A_i$ such that:

$$xy = z_{f+1}$$

Indeed, we can take $f = n - s - 1$ and $x = z_{n-s}, y = 1$ in the first case, and $x = x_{n-s}, y = y_{s+1}$ in the second one. Suppose now that $A_1, A_2 \in E$ and let $a, b, a', b' \in A_1 * A_2$ such that $ab = a'b'$. Let $z_1 \ldots z_p$ be the decomposition of $ab = a'b'$ as a reduced word. Then we can decompose our words, as above, in the following way:

$$a = z_1 \ldots z_f xu \quad , \quad a' = z_1 \ldots z_{f'} x'u'$$

We have to show that $a = a'm$ or that $a' = am$ for some $m \in A_1 * A_2$. But this is clear in all three cases that can appear, namely $f < f'$, $f' < f$, $f = f'$. □

We can now formulate a main result about semigroup freeness, as follows:

Theorem 15.31. *The following happen:*

(1) *Given $M \subset N$, both in the class E, satisfying $M(N - M) = N - M$, any x in the $*$-algebra generated by $\lambda(M)$ can be written as follows, with $p_i, q_i \in M$:*

$$x = \sum_i a_i \lambda_N(p_i) \lambda_N(q_i)^*$$

(2) *Asssume $A, B \in E$, and let x be an element of the $*$-algebra generated by $\lambda_{A*B}(A)$ such that $\tau(x) = 0$. If W_A, W_B are respectively the sets of reduced words beginning by an element of A, B, then x acts as follows:*

$$l^2(W_B \cup \{e\}) \to l^2(W_A)$$

(3) *Let $A, B \in E$. Then $\lambda_{A*B}(A)$ and $\lambda_{A*B}(B)$ are free.*

Proof. This follows from our results so far, the idea being is as follows:

(1) It is enough to prove this for elements of the form $x = \lambda(m)^* \lambda(n)$ with $m, n \in M$, because the general case will follow easily from this. In order to do so, observe that $x = \lambda(m)^* \lambda(n)$ is different from 0 precisely when there exist $a, b \in N$ such that:

$$< \lambda(m)^* \lambda(n) \delta_a, \delta_b > \neq 0$$

That is, the following condition must be satisfied:

$$na = mb$$

We know that there exists $c \in N$ with $n = mc$ or with $m = nc$. Moreover, as $M(N - M) = N - M$, it follows that $c \in M$. Thus $x = \lambda(m)^* \lambda(n) \neq 0$ implies that $x = \lambda(c)$ or $x = \lambda(c)^*$ with $c \in M$, and this finishes the proof.

(2) We apply (1) with $M = A$ and $N = A * B$ for writing, with $p_i, q_i \in A$:

$$x = \sum_i a_i \lambda(p_i) \lambda(q_i)^*$$

Consider now the following element:

$$\tau(\lambda(p_i)\lambda(q_i)^*) = \sum_x \delta_{e,p_ix}\delta_{e,q_ix}$$

This element is nonzero precisely when $p_i = q_i$ is invertible, and in this case:

$$\lambda(p_i)\lambda(q_i)^* = 1$$

Now since we assumed $\tau(x) = 0$, it follows that we can write:

$$x = \sum a_i\lambda(p_i)\lambda(q_i)^* \quad , \quad \tau(\lambda(p_i)\lambda(q_i)^*) = 0$$

By linearity, it is enough to prove the result for $x = \lambda(p_i)\lambda(q_i)^*$. Let $m \in W_B \cup \{e\}$ and suppose that $x\delta_m \neq 0$. Then $\lambda(q_i)^*\delta_m \neq 0$ implies that $m = q_ic$ for some word $c \in A * B$. As $q_i \in A$ and $m \in W_B \cup \{e\}$, it follows that q_i is invertible. Now observe that:

$$p_iq_i^{-1} = 1 \implies \tau(x) = 1$$

It follows that we have, as desired:

$$x\delta_m = \delta_{p_iq_i^{-1}m} \in l^2(W_A)$$

(3) This follows from (2) above. Indeed, let $P = x_n \ldots x_1$ be a product of elements in $ker(\tau)$, such that x_{2k} is in the $*$-algebra generated by $\lambda(B)$ and x_{2k+1} is in the $*$-algebra generated by $\lambda(A)$. Then $x_1\delta_e \in l^2(W_A)$. Thus $x_2x_1\delta_e \in l^2(W_B)$, and so on. By a reccurence, $P\delta_e$ is in $l^2(W_A)$ or in $l^2(W_B)$. But this implies that $\tau(P) = 0$, as desired. □

As a main application of the above semigroup technology, we have:

THEOREM 15.32. *Consider a Haar unitary u, free from a semicircular s. Then*

$$c = us$$

is a circular variable.

PROOF. Denote by z the image of $1 \in \mathbb{Z}$ and by n the image of $1 \in \mathbb{N}$ by the canonical embeddings into the free product $\mathbb{Z} * \mathbb{N}$. Let $\lambda = \lambda_{\mathbb{Z}*\mathbb{N}}$. We know that $\mathbb{Z} * \mathbb{N} \in E$. Also (zn, nz^{-1}) is obviously a prefix, so it is a code. Thus, the following variable is circular:

$$c = \frac{1}{2}(\lambda(zn) + \lambda(nz^{-1})^*)$$

The point now is that we have the following formula:

$$\frac{1}{2}(\lambda(zn) + \lambda(nz^{-1})^*) = us$$

But this gives the result, in our model and so in general as well, because $u = \lambda(z)$ is a Haar-unitary, $s = 1/2(\lambda(n) + \lambda(n)^*)$ is semicircular, and u and s are free. □

We can now recover the Voiculescu polar decomposition result for the circular variables, obtained in [90], by using random matrix techniques, as follows:

THEOREM 15.33. *Consider the polar decomposition of a circular variable, in some von Neumann algebraic probability space with faithful normal state:*

$$x = vb$$

Then v is Haar unitary, b is quarter-circular, and (v, b) are free.

PROOF. This follows by suitably manipulating Theorem 15.32, as to replace the semicircular element there by a quarter-circular. Consider indeed the following group:

$$G = \mathbb{Z} * (\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z})$$

Let z, t, a be the images of the following elements, into this group G:

$$1 \in \mathbb{Z} \quad , \quad (1, \hat{0}) \in \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z}) \quad , \quad (0, \hat{1}) \in \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$$

Let $u = \lambda_G(z)$, $d = \lambda_G(a)$ and choose a quarter-circular $q \in C^*(\lambda_G(t))$. Then (q, d) are independent, so dq is semicircular, and so $c = udq$ is circular, and:

– The module of c is q, which is a quarter-circular.

– The polar part of c is ud, which is obviously a Haar unitary.

– Consider the automorphism of G which is the identity on $\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ and maps $z \to za$. This extends to a trace-preserving automorphism of $C^*(G)$ which maps:

$$u \to ud \quad , \quad q \to q$$

Since u, q are free, it follows that ud, q are free too, finishing the proof. □

15d. Gaussian matrices

As an application of the semicircular and circular variable theory developed so far, and of free probability in general, let us go back now to the random matrices. Following Voiculescu's paper [90], we will prove now a number of key freeness results for them, complementing the basic random matrix theory developed in chapter 14. As a first result, completing our asymptotic law study for the Gaussian matrices, we have:

THEOREM 15.34. *Given a sequence of complex Gaussian matrices*

$$Z_N \in M_N(L^\infty(X))$$

having independent G_t variables as entries, with $t > 0$, we have

$$\frac{Z_N}{\sqrt{N}} \sim \Gamma_t$$

in the $N \to \infty$ limit, with the limiting measure being Voiculescu's circular law.

PROOF. We know from chapter 14, with this having been actually our very first moment computation for random matrices, in this book, that the asymptotic moments of the complex Gaussian matrices are given by the following formula:

$$M_k\left(\frac{Z_N}{\sqrt{N}}\right)\simeq t^{|k|/2}|\mathcal{NC}_2(k)|$$

On the other hand, we also know from the above that an abstract noncommutative variable $a\in A$ is circular, following the law Γ_t, precisely when its moments are:

$$M_k(a)=t^{|k|/2}|\mathcal{NC}_2(k)|$$

Thus, we are led to the conclusion in the statement. □

The above result is of course something quite theoretical, and having it formulated as such is certainly something nice. However, and here comes our point, it is actually possible to use free probability theory in order to go well beyond this, with this time some truly "new" results on the random matrices. We will explain this now, following Voiculescu's paper [90]. Let us begin with the Wigner matrices. We have here:

THEOREM 15.35. *Given a family of sequences of Wigner matrices,*

$$Z_N^i\in M_N(L^\infty(X))\quad,\quad i\in I$$

with pairwise independent entries, each following the complex normal law G_t, with $t>0$, up to the constraint $Z_N^i=(Z_N^i)^*$, the rescaled sequences of matrices

$$\frac{Z_N^i}{\sqrt{N}}\in M_N(L^\infty(X))\quad,\quad i\in I$$

become with $N\to\infty$ semicircular, each following the Wigner law γ_t, and free.

PROOF. This is something quite subtle, the idea being as follows:

(1) First of all, we know from chapter 14 that for any $i\in I$ the corresponding sequence of rescaled Wigner matrices becomes semicircular in the $N\to\infty$ limit:

$$\frac{Z_N^i}{\sqrt{N}}\simeq\gamma_t$$

(2) Thus, what is new here, and that we have to prove, is the asymptotic freeness assertion. For this purpose we can assume that we are dealing with the case of 2 sequences of matrices, $|I|=2$. So, assume that we have Wigner matrices as follows:

$$Z_N,Z_N'\in M_N(L^\infty(X))$$

We have to prove that these matrices become asymptotically free, with $N\to\infty$.

(3) But this something that can be proved directly, via various routine computations with partitions, which simplify as usual in the $N\to\infty$ limit, and bring freeness.

(4) However, we can prove this as well by using a trick, based on the result in Theorem 15.34. Consider indeed the following random matrix:

$$Y_N = \frac{1}{\sqrt{2}}(Z_N + iZ'_N)$$

This is then a complex Gaussian matrix, and so by using Theorem 15.34, we obtain that in the limit $N \to \infty$, we have:

$$\frac{Y_N}{\sqrt{N}} \simeq \Gamma_t$$

Now recall that the circular law Γ_t was by definition the law of the following variable, with a, b being semicircular, each following the law γ_t, and free:

$$c = \frac{1}{\sqrt{2}}(a + ib)$$

We are therefore in the situation where the variable $(Z_N + iZ'_N)/\sqrt{N}$, which has asymptotically semicircular real and imaginary parts, converges to the distribution of $a + ib$, equally having semicircular real and imaginary parts, but with these real and imaginary parts being free. Thus Z_N, Z'_N become asymptotically free, as desired. □

Getting now to the complex case, we have a similar result here, as follows:

THEOREM 15.36. *Given a family of sequences of complex Gaussian matrices,*

$$Z_N^i \in M_N(L^\infty(X)) \quad , \quad i \in I$$

with pairwise independent entries, each following the complex normal law G_t, with $t > 0$, the rescaled sequences of matrices

$$\frac{Z_N^i}{\sqrt{N}} \in M_N(L^\infty(X)) \quad , \quad i \in I$$

become with $N \to \infty$ circular, each following the Voiculescu law Γ_t, and free.

PROOF. This follows from Theorem 15.35, which applies to the real and imaginary parts of our complex Gaussian matrices, and gives the result. □

The above results are interesting for both free probability and random matrices. As an illustration here, we have the folowing application to free probability:

THEOREM 15.37. *Consider the polar decomposition of a circular variable in some von Neumann algebraic probability space with faithful normal state:*

$$x = vb$$

Then v is Haar-unitary, b is quarter-circular and (v, b) are free.

PROOF. This is indeed easy to see in the Gaussian matrix model provided by Theorem 15.36, and for details here, we refer to Voiculescu's paper [90]. □

There are many other applications along these lines, and conversely, free probability can be used as well for the detailed study of the Wigner and Gaussian matrices.

15e. Exercises

Exercises:

Exercise 15.38.

Exercise 15.39.

Exercise 15.40.

Exercise 15.41.

Exercise 15.42.

Exercise 15.43.

Exercise 15.44.

Exercise 15.45.

Bonus exercise.

CHAPTER 16

Discrete versions

16a. Poisson limits

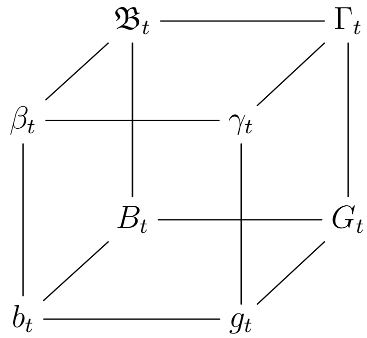
We have seen that free probability leads to two key limiting theorems, namely the free analogues of the CLT and CCLT. The limiting measures are the Wigner semicircle laws γ_t and the Voiculescu circular laws Γ_t. Together with the Gaussian laws g_t and G_t coming from the classical CLT and CCLT, these laws form a square diagram, as follows:

$$\begin{matrix}
\gamma_t & - & \Gamma_t \\
| & & | \\
g_t & - & G_t
\end{matrix}$$

Motivated by this, in this chapter we develop more free limiting theorems. First, we will find a free analogue of the PLT, with the corresponding limiting measures, appearing as the free analogues of the Poisson laws p_t, being the Marchenko-Pastur laws π_t. This will lead to an extension to the above square diagram, into a rectangle, as follows:

$$\begin{matrix}
\pi_t & - & \gamma_t & - & \Gamma_t \\
| & & | & & | \\
p_t & - & g_t & - & G_t
\end{matrix}$$

More generally, we will find a free analogue of the compound Poisson limit theorem (CPLT), that we know from before. At the level of the philosophy, and of the above diagram, there are no complex analogues of p_t, π_t, but by using certain measures found via the classical and free CPLT, namely the real and purely complex Bessel laws b_t, B_t discussed before, and their free analogues $\beta_t, \mathfrak{B}_t$ to be discussed here, we will be able to modify and then fold the diagram, as to complete it into a cube, as follows:



Which is of course quite nice, theoretically speaking, because this leads to a kind of 3D orientation inside classical and free probability, which is something very useful.

Getting started now, we would first like to have a free analogue of the Poisson Limit Theorem (PLT). Although elementary from what we have, this was something not done by Voiculescu himself, and not appearing in the foundational book [**91**], and only explained later, in the book of Hiai and Petz. The statement is as follows:

Theorem 16.1 (Free PLT). *The following limit converges, for any* $t > 0$,

$$\lim_{n\to\infty}\left(\left(1-\frac{t}{n}\right)\delta_0+\frac{t}{n}\delta_1\right)^{\boxplus n}$$

and we obtain the Marchenko-Pastur law of parameter t,

$$\pi_t = \max(1-t,0)\delta_0 + \frac{\sqrt{4t-(x-1-t)^2}}{2\pi x}\,dx$$

also called free Poisson law of parameter t.

Proof. Consider the measure in the statement, under the convolution sign:

$$\eta = \left(1-\frac{t}{n}\right)\delta_0+\frac{t}{n}\delta_1$$

The Cauchy transform of this measure is easy to compute, and is given by:

$$G_\eta(\xi) = \left(1-\frac{t}{n}\right)\frac{1}{\xi}+\frac{t}{n}\cdot\frac{1}{\xi-1}$$

In order to prove the result, we want to compute the following R-transform:

$$R = R_{\eta^{\boxplus n}}(y) = nR_\eta(y)$$

According to the formula of G_η, the equation for this function R is as follows:

$$\left(1-\frac{t}{n}\right)\frac{1}{1/y+R/n}+\frac{t}{n}\cdot\frac{1}{1/y+R/n-1} = y$$

By multiplying both sides by n/y, this equation can be written as:

$$\frac{t + yR}{1 + yR/n} = \frac{t}{1 + yR/n - y}$$

With $n \to \infty$ things simplify, and we obtain the following formula:

$$t + yR = \frac{t}{1 - y}$$

Thus we have the following formula, for the R-transform that we are interested in:

$$R = \frac{t}{1 - y}$$

But this gives the result, since R_{π_t} is elementary to compute from what we have, by "doubling" the results for the Wigner law γ_t, and is given by the same formula. □

As in the continuous case, most of the basic theory of π_t was already done before, with all this partly coming from the theory of SO_3, at $t = 1$. One thing which was missing there, however, was that of understanding how the law π_t, with parameter $t > 0$, exactly appears, out of π_1. We can now solve this question, as follows:

THEOREM 16.2. *The Marchenko-Pastur laws have the property*

$$\pi_s \boxplus \pi_t = \pi_{s+t}$$

so they form a 1*-parameter semigroup with respect to free convolution.*

PROOF. This follows either from Theorem 16.1, or from the fact that the R-transform of π_t, computed in the proof of Theorem 16.1, is linear in t. □

All this is very nice, conceptually speaking, and we can now summarize the various discrete probability results that we have, classical and free, as follows:

THEOREM 16.3. *The Poisson laws p_t and the Marchenko-Pastur laws π_t, given by*

$$p_t = e^{-t} \sum_k \frac{t^k}{k!}\,\delta_k$$

$$\pi_t = \max(1-t, 0)\delta_0 + \frac{\sqrt{4t - (x-1-t)^2}}{2\pi x}\,dx$$

have the following properties:

1. *They appear via the PLT, and the free PLT.*
2. *They form semigroups with respect to $*$ and $\boxplus$.*
3. *Their transforms are $\log F_{p_t}(x) = t(e^{ix} - 1)$, $R_{\pi_t}(x) = t/(1-x)$.*
4. *Their moments are $M_k = \sum_{\pi \in D(k)} t^{|\pi|}$, with $D = P, NC$.*

Proof. These are all results that we already know, from here and from the previous chapters. To be more precise:

(1) The PLT is from before, and the FPLT is from here.

(2) The semigroup properties are from before, and from here.

(3) The formula for F_{p_t} is from before, and the one for R_{π_t}, from here.

(4) The moment formulae follow from the formulae of functional transforms. □

We can in fact merge this with our previous continuous results, and we obtain:

Theorem 16.4. *The moments of the various central limiting measures, namely*

$$\begin{matrix} \pi_t & - & \gamma_t & - & \Gamma_t \\ | & & | & & | \\ p_t & - & g_t & - & G_t \end{matrix}$$

are always given by the same formula, involving partitions, namely

$$M_k = \sum_{\pi \in D(k)} t^{|\pi|}$$

where the sets of partitions $D(k)$ in question are respectively

$$\begin{matrix} \pi_t & - & \gamma_t & - & \Gamma_t \\ | & & | & & | \\ p_t & - & g_t & - & G_t \end{matrix}$$

and where $|.|$ is the number of blocks.

Proof. This follows indeed by putting together the various results that we have, from chapter 10 for the square on the right, and from here for the edge on the left. □

We will later some more conceptual explanations for all this, featuring classical and free cumulants, classical and free quantum groups, and many more.

Moving ahead now, let us try to find a free analogue of the CPLT. We will follow the CPLT material from before, by performing modifications where needed, as to replace everywhere classical probability with free probability. Let us start with the following straightforward definition, similar to the one from the classical case:

Definition 16.5. *Associated to any compactly supported positive measure ρ on $\mathbb{C}$ is the probability measure*

$$\pi_\rho = \lim_{n\to\infty}\left(\left(1-\frac{c}{n}\right)\delta_0+\frac{1}{n}\rho\right)^{\boxplus n}$$

where $c = mass(\rho)$, called compound free Poisson law.

In what follows we will be mostly interested in the case where ρ is discrete, as is for instance the case for the measure $\rho = t\delta_1$ with $t > 0$, which produces the free Poisson laws. The following result allows one to detect compound free Poisson laws:

Proposition 16.6. *For a discrete measure, written as*

$$\rho = \sum_{i=1}^s c_i\delta_{z_i}$$

with $c_i > 0$ and $z_i \in \mathbb{C}$, we have the following formula,

$$R_{\pi_\rho}(y) = \sum_{i=1}^s \frac{c_i z_i}{1-yz_i}$$

where R denotes as usual the Voiculescu R-transform.

Proof. In order to prove this result, let η_n be the measure appearing in Definition 16.5, under the free convolution sign, namely:

$$\eta_n = \left(1-\frac{c}{n}\right)\delta_0+\frac{1}{n}\rho$$

The Cauchy transform of η_n is then given by the following formula:

$$G_{\eta_n}(\xi) = \left(1-\frac{c}{n}\right)\frac{1}{\xi}+\frac{1}{n}\sum_{i=1}^s\frac{c_i}{\xi-z_i}$$

Consider now the R-transform of the measure $\eta_n^{\boxplus n}$, which is given by:

$$R_{\eta_n^{\boxplus n}}(y) = nR_{\eta_n}(y)$$

By using the general theory of the R-transform, from chapter 13, the above formula of G_{η_n} shows that the equation for $R = R_{\eta_n^{\boxplus n}}$ is as follows:

$$\begin{aligned}
&\left(1-\frac{c}{n}\right)\frac{1}{1/y+R/n}+\frac{1}{n}\sum_{i=1}^s\frac{c_i}{1/y+R/n-z_i} = y\\
\implies\quad &\left(1-\frac{c}{n}\right)\frac{1}{1+yR/n}+\frac{1}{n}\sum_{i=1}^s\frac{c_i}{1+yR/n-yz_i} = 1
\end{aligned}$$

Now multiplying by n, then rearranging the terms, and letting $n \to \infty$, we get:

$$\frac{c+yR}{1+yR/n}=\sum_{i=1}^s\frac{c_i}{1+yR/n-yz_i}\implies c+yR_{\pi_\rho}(y)=\sum_{i=1}^s\frac{c_i}{1-yz_i}$$
$$\implies R_{\pi_\rho}(y)=\sum_{i=1}^s\frac{c_iz_i}{1-yz_i}$$

Thus, we are led to the conclusion in the statement. □

We have as well the following result, providing an alternative to Definition 16.5, and which, together with Definition 16.5, can be thought of as being the free CPLT:

THEOREM 16.7. *For a discrete measure, written as*

$$\rho=\sum_{i=1}^s c_i\delta_{z_i}$$

with $c_i>0$ and $z_i\in\mathbb C$, we have the formula

$$\pi_\rho=\mathrm{law}\left(\sum_{i=1}^s z_i\alpha_i\right)$$

where the variables α_i are free Poisson(c_i), free.

PROOF. Let α be the sum of free Poisson variables in the statement:

$$\alpha=\sum_{i=1}^s z_i\alpha_i$$

In order to prove the result, we will show that the R-transform of α is given by the formula in Proposition 16.6. We have the following computation:

$$R_{\alpha_i}(y)=\frac{c_i}{1-y}\implies R_{z_i\alpha_i}(y)=\frac{c_iz_i}{1-yz_i}$$
$$\implies R_\alpha(y)=\sum_{i=1}^s\frac{c_iz_i}{1-yz_i}$$

Thus we have the same formula as in Proposition 16.6, and we are done. □

All the above is quite general, and in practice, in order to obtain concrete results, the simplest measures that we can use as “input” for the CPLT are the same measures as those that we used in the classical case, namely the measures of type $\rho=t\varepsilon_s$, with $t>0$, and with ε_s being the uniform measure on the s-th roots of unity. We will discuss this in what follows, by following the literature on the subject.

16b. Bessel laws

As mentioned above, for various reasons, including the construction of the "standard cube" discussed in the beginning of this chapter, we are interested in the applications of the free CPLT with the "simplest" input measures, with these simplest measures being those of type $\rho = t\varepsilon_s$, with $t > 0$, and with ε_s being the uniform measure on the s-th roots of unity. We are led in this way the following class of measures:

DEFINITION 16.8. *The Bessel and free Bessel laws, depending on parameters $s \in \mathbb{N} \cup \{\infty\}$ and $t > 0$, are the following compound Poisson and free Poisson laws,*

$$b_t^s = p_{t\varepsilon_s} \quad , \quad \beta_t^s = \pi_{t\varepsilon_s}$$

with ε_s being the uniform measure on the s-th roots of unity. In particular:

(1) *At $s = 1$ we recover the Poisson laws p_t, π_t.*
(2) *At $s = 2$ we have the real Bessel laws b_t, β_t.*
(3) *At $s = \infty$ we have the complex Bessel laws $B_t, \mathfrak{B}_t$.*

The terminology here comes from the fact, that we know from before, that the density of the measure b_t, appearing at $s = 2$, is a Bessel function of the first kind.

Our next task will be that upgrading our results about the free Poisson law π_t in this setting, using a parameter $s \in \mathbb{N} \cup \{\infty\}$. First, we have the following result:

THEOREM 16.9. *The free Bessel laws have the property*

$$\beta_t^s \boxplus \beta_{t'}^s = \beta_{t+t'}^s$$

so they form a 1-parameter semigroup with respect to free convolution.

PROOF. This follows indeed from the fact that the R-transform of β_t^s is linear in t, which is something that we already know, from the above. □

Let us discuss now some more advanced aspects. Given a real probability measure μ, one can ask whether the convolution powers $\mu^{\boxtimes s}$ and $\mu^{\boxplus t}$ exist, for various values of the parameters $s, t > 0$. For the free Poisson law, the answer to this is as follows:

PROPOSITION 16.10. *The free convolution powers of the free Poisson law*

$$\pi^{\boxtimes s} \quad , \quad \pi^{\boxplus t}$$

exist for any positive values of the paremeters, $s, t > 0$.

PROOF. We have two measures to be studied, the idea being as follows:

(1) The free Poisson law π is by definition the $t = 1$ particular case of the free Poisson law of parameter t, or Marchenko-Pastur law of parameter $t > 0$, given by:

$$\pi_t = \max(1 - t, 0)\delta_0 + \frac{\sqrt{4t - (x - 1 - t)^2}}{2\pi x}\,dx$$

The Cauchy transform of this measure is given by:

$$G(\xi)=\frac{(\xi+1-t)+\sqrt{(\xi+1-t)^2-4\xi}}{2\xi}$$

We can compute now the R transform, by proceeding as follows:

$$\begin{aligned}\xi G^2+1=(\xi+1-t)G \quad &\Longrightarrow \quad Kz^2+1=(K+1-t)z\\ &\Longrightarrow \quad Rz^2+z+1=(R+1-t)z+1\\ &\Longrightarrow \quad Rz=R-t\\ &\Longrightarrow \quad R=t/(1-z)\end{aligned}$$

The last expression being linear in t, the measures π_t form a semigroup with respect to free convolution. Thus we have $\pi_t=\pi^{\boxplus t}$, which proves the second assertion.

(2) Regarding now the measure $\pi^{\boxtimes s}$, there is no explicit formula for its density. However, we can prove that this measure exists, by using some abstract results. Indeed, we have the following computation for the S transform of π_t:

$$\begin{aligned}\xi G^2+1=(\xi+1-t)G \quad &\Longrightarrow \quad zf^2+1=(1+z-zt)f\\ &\Longrightarrow \quad z(\psi+1)^2+1=(1+z-zt)(\psi+1)\\ &\Longrightarrow \quad \chi(z+1)^2+1=(1+\chi-\chi t)(z+1)\\ &\Longrightarrow \quad \chi(z+1)(t+z)=z\\ &\Longrightarrow \quad S=1/(t+z)\end{aligned}$$

In particular at $t=1$ we have the following formula:

$$S(z)=\frac{1}{1+z}$$

Thus the Σ transform of π, which is by definition $\Sigma(z)=S(z/(1-z))$, is given by:

$$\Sigma(z)=1-z$$

On the other hand, it is well-known from the general theory of the S-transform that the Σ transforms of the probability measures which are $\boxtimes$-infinitely divisible are the functions of the form $\Sigma(z)=e^{v(z)}$, where $v:\mathbb{C}-[0,\infty)\to\mathbb{C}$ is analytic, satisfying:

$$v(\bar{z})=\bar{v}(z) \quad , \quad v(\mathbb{C}^+)\subset\mathbb{C}^-$$

Now in the case of the free Poisson law, the function $v(z)=\log(1-z)$ satisfies these properties, and we are led to the conclusion in the statement. □

Getting now towards the free Bessel laws, we have the following remarkable identity, in relation with the above convolution powers of π:

THEOREM 16.11. *We have the formula*

$$\pi^{\boxtimes s-1}\boxtimes\pi^{\boxplus t}=((1-t)\delta_0+t\delta_1)\boxtimes\pi^{\boxtimes s}$$

valid for any $s\geq 1$*, and any* $t\in(0,1]$*.*

PROOF. We know from the previous proof that the S transform of the free Poisson law π is given by the following formula:

$$S_1(z)=\frac{1}{1+z}$$

We also know from there that the S transform of $\pi^{\boxplus t}$ is given by:

$$S_t(z)=\frac{1}{t+z}$$

Thus the measure on the left in the statement has the following S transform:

$$S(z)=\frac{1}{(1+z)^{s-1}}\cdot\frac{1}{t+z}$$

The S transform of $\alpha_t=(1-t)\delta_0+t\delta_1$ can be computed as follows:

$$\begin{aligned}f=1+tz/(1-z)\quad&\implies\quad\psi=tz/(1-z)\\&\implies\quad z=t\chi/(1-\chi)\\&\implies\quad\chi=z/(t+z)\\&\implies\quad S=(1+z)/(t+z)\end{aligned}$$

Thus the measure on the right in the statement has the following S transform:

$$S(z)=\frac{1}{(1+z)^s}\cdot\frac{1+z}{t+z}$$

Thus the S transforms of our two measures are the same, and we are done. □

The relation with the free Bessel laws, as previously defined, comes from:

THEOREM 16.12. *The free Bessel law is the real probability measure* β_t^s*, with*

$$(s,t)\in(0,\infty)\times(0,\infty)-(0,1)\times(1,\infty)$$

defined concretely as follows:

(1) *For* $s\geq 1$ *we set* $\beta_t^s=\pi^{\boxtimes s-1}\boxtimes\pi^{\boxplus t}$.
(2) *For* $t\leq 1$ *we set* $\beta_t^s=((1-t)\delta_0+t\delta_1)\boxtimes\pi^{\boxtimes s}$.

PROOF. This follows indeed from the above results. To be more precise, these results show that the measures constructed in the statement exist indeed, and coincide with the free Bessel laws, as previously defined, as compound free Poisson laws. □

In view of the above, we can regard the free Bessel law β_t^s as being a natural two-parameter generalization of the free Poisson law π, in connection with Voiculescu's free convolution operations $\boxtimes$ and $\boxplus$. Observe that we have the following formulae:

$$\begin{cases}\beta_1^s = \pi^{\boxtimes s}\\ \beta_t^1 = \pi^{\boxplus t}\end{cases}$$

As a comment here, concerning the precise range of the parameters (s,t), the above results can be probably improved. The point is that the measure β_t^s still exists for certain points in the critical rectangle $(0,1)\times(1,\infty)$, but not for all of them.

Next in line, we have the following result:

Proposition 16.13. *The Stieltjes transform of β_t^s satisfies:*

$$f = 1 + zf^s(f+t-1)$$

In particular at $t=1$ we have the formula $f = 1 + zf^{s+1}$.

Proof. We have the following computation:

$$\begin{aligned} S = \frac{1}{(1+z)^{s-1}}\cdot\frac{1}{t+z} \quad &\implies \quad \chi = \frac{z}{(1+z)^s}\cdot\frac{1}{t+z}\\ &\implies \quad z = \frac{\psi}{(1+\psi)^s}\cdot\frac{1}{t+\psi}\\ &\implies \quad z = \frac{f-1}{f^s}\cdot\frac{1}{t+f-1}\end{aligned}$$

Thus, we obtain the equation in the statement. □

At $t=1$, we have in fact the following result, which is more explicit:

Theorem 16.14. *The Stieltjes transform of β_1^s with $s\in\mathbb{N}$ is given by*

$$f(z) = \sum_{p\in NC_s} z^{k(p)}$$

where NC_s is the set of noncrossing partitions all whose blocks have as size multiples of s, and where $k: NC_s\to\mathbb{N}$ is the normalized length.

Proof. With the notation $C_k = \#NC_s(k)$, where $NC_s(k)\subset NC_s$ consists of the partitions of $\{1,\ldots,sk\}$ belonging to NC_s, the sum on the right is:

$$f(z) = \sum_k C_k z^k$$

For a given partition $p \in NC_s(k+1)$ we can consider the last s legs of the first block, and make cuts at right of them. This gives a decomposition of p into $s+1$ partitions in NC_s, and we obtain in this way the following recurrence formula for the numbers C_k:

$$C_{k+1} = \sum_{\Sigma k_i = k} C_{k_0} \ldots C_{k_s}$$

By multiplying now by z^{k+1}, and then summing over k, we obtain that the generating series of these numbers C_k satisfies the following equation:

$$f - 1 = zf^{s+1}$$

But this is the equation found in Proposition 16.13, so we obtain the result. □

Next, we have the following result, dealing with the case $t > 0$:

THEOREM 16.15. *The Stieltjes transform of β_t^s with $s \in \mathbb{N}$ is given by:*

$$f(z) = \sum_{p \in NC_s} z^{k(p)} t^{b(p)}$$

where $k, b : NC_s \to \mathbb{N}$ are the normalized length, and the number of blocks.

PROOF. With notations from the previous proof, let F_{kb} be the number of partitions in $NC_s(k)$ having b blocks, and set $F_{kb} = 0$ for other integer values of k, b. All sums will be over integer indices ≥ 0. The sum on the right in the statement is then:

$$f(z) = \sum_{kb} F_{kb} z^k t^b$$

The recurrence formula for the numbers C_k in the previous proof becomes:

$$\sum_b F_{k+1,b} = \sum_{\Sigma k_i = k} \sum_{b_i} F_{k_0 b_0} \ldots F_{k_s b_s}$$

In this formula, each term contributes to $F_{k+1,b}$ with $b = \Sigma b_i$, except for those of the form $F_{00} F_{k_1 b_1} \ldots F_{k_s b_s}$, which contribute to $F_{k+1,b+1}$. We get:

$$\begin{aligned} F_{k+1,b} &= \sum_{\Sigma k_i = k} \sum_{\Sigma b_i = b} F_{k_0 b_0} \ldots F_{k_s b_s} \\ &+ \sum_{\Sigma k_i = k} \sum_{\Sigma b_i = b-1} F_{k_1 b_1} \ldots F_{k_s b_s} \\ &- \sum_{\Sigma k_i = k} \sum_{\Sigma b_i = b} F_{k_1 b_1} \ldots F_{k_s b_s} \end{aligned}$$

This gives the following formula for the polynomials $P_k = \sum_b F_{kb} t^b$:

$$P_{k+1} = \sum_{\Sigma k_i = k} P_{k_0} \ldots P_{k_s} + (t-1) \sum_{\Sigma k_i = k} P_{k_1} \ldots P_{k_s}$$

Consider now the following generating function:

$$f=\sum_k P_k z^k$$

In terms of this generating function, we get the following equation:

$$f-1=zf^{s+1}+(t-1)zf^s$$

But this is the same as the equation of the Stieltjes transform of β_t^s, namely:

$$f=1+zf^s(f+t-1)$$

Thus, we are led to the conclusion in the statement. □

Let us discuss now the computation of the moments of the free Bessel laws. The idea will be that of expressing these moments in terms of generalized binomial coefficients. We recall that the coefficient corresponding to $\alpha\in\mathbb{R}$, $k\in\mathbb{N}$ is:

$$\binom{\alpha}{k}=\frac{\alpha(\alpha-1)\ldots(\alpha-k+1)}{k!}$$

We denote by $m_1,m_2,m_3,\ldots$ the sequence of moments of a given probability measure. With this convention, we first have the following result:

THEOREM 16.16. *The moments of β_1^s with $s>0$ are*

$$m_k=\frac{1}{sk+1}\binom{sk+k}{k}$$

which are the Fuss-Catalan numbers.

PROOF. In the case $s\in\mathbb{N}$, we know that we have $m_k=\#NC_s(k)$. The formula in the statement follows then by counting such partitions. In the general case $s>0$, observe first that the Fuss-Catalan number in the statement is a polynomial in s:

$$\frac{1}{sk+1}\binom{sk+k}{k}=\frac{(sk+2)(sk+3)\ldots(sk+k)}{k!}$$

Thus, in order to pass from the case $s\in\mathbb{N}$ to the case $s>0$, it is enough to check that the k-th moment of π_{s1} is analytic in s. But this is clear from the equation $f=1+zf^{s+1}$ of the Stieltjes transform of π_{s1}, and this gives the result. □

We have as well the following result, which deals with the general case $t>0$:

THEOREM 16.17. *The moments of β_t^s with $s>0$ are*

$$m_k=\sum_{b=1}^k\frac{1}{b}\binom{k-1}{b-1}\binom{sk}{b-1}t^b$$

which are the Fuss-Narayana numbers.

PROOF. In the case $s \in \mathbb{N}$, we know from the above that we have the following formula, where F_{kb} is the number of partitions in $NC_s(k)$ having b blocks:

$$m_k = \sum_b F_{kb} t^b$$

With this observation in hand, the formula in the statement follows by counting such partitions, with this count being well-known. This result can be then extended to any parameter $s > 0$, by using a standard complex variable argument, as before. □

In the case $s \notin \mathbb{N}$, the moments of β_t^s can be further expressed in terms of gamma functions. In the case $s = 1/2$, the result is as follows:

THEOREM 16.18. *The moments of $\beta_1^{1/2}$ are given by the following formulae:*

$$m_{2p} = \frac{1}{p+1}\binom{3p}{p}$$

$$m_{2p-1} = \frac{2^{-4p+3}p}{(6p-1)(2p+1)} \cdot \frac{p!(6p)!}{(2p)!(2p)!(3p)!}$$

PROOF. According to our various results above, the even moments of the free Bessel law β_t^s with $s = n - 1/2$, $n \in \mathbb{N}$, are given by:

$$\begin{aligned} m_{2p} &= \frac{1}{(n-1/2)(2p)+1}\binom{(n+1/2)2p}{2p} \\ &= \frac{1}{(2n-1)p+1}\binom{(2n+1)p}{2p} \end{aligned}$$

With $n = 1$ we get the formula in the statement. Now for the odd moments, we can use here the following well-known identity:

$$\binom{m-1/2}{k} = \frac{4^{-k}}{k!} \cdot \frac{(2m)!}{m!} \cdot \frac{(m-k)!}{(2m-2k)!}$$

With $m = 2np + p - n$ and $k = 2p - 1$ we get:

$$\begin{aligned} m_{2p-1} &= \frac{1}{(n-1/2)(2p-1)+1}\binom{(n+1/2)(2p-1)}{2p-1} \\ &= \frac{2}{(2n-1)(2p-1)+2}\binom{(2np+p-n)-1/2}{2p-1} \\ &= \frac{2^{-4p+3}}{(2p-1)!} \cdot \frac{(4np+2p-2n)!}{(2np+p-n)!} \cdot \frac{(2np-p-n+1)!}{(4np-2p-2n+3)!} \end{aligned}$$

In particular with $n = 1$ we obtain:

$$\begin{aligned} m_{2p-1} &= \frac{2^{-4p+3}}{(2p-1)!} \cdot \frac{(6p-2)!}{(3p-1)!} \cdot \frac{p!}{(2p+1)!} \\ &= \frac{2^{-4p+3}(2p)}{(2p)!} \cdot \frac{(6p)!(3p)}{(3p)!(6p-1)6p} \cdot \frac{p!}{(2p)!(2p+1)} \end{aligned}$$

But this gives the formula in the statement. □

16c. The standard cube

Let us get back now to the fundamental question, mentioned in the beginning of this chapter, of arranging the main probability measures that we know, classical and free, into a cube, and this for having some kind of 3D orientation, inside probability. We have:

THEOREM 16.19. *The various classical and free central limiting measures,*

$$\begin{matrix} \beta_t^s & - & \gamma_t & - & \Gamma_t \\ | & & | & & | \\ b_t^s & - & g_t & - & G_t \end{matrix}$$

have moments always given by the same formula, involving partitions, namely

$$M_k = \sum_{\pi \in D(k)} t^{|\pi|}$$

where the sets of partitions $D(k)$ in question are respectively

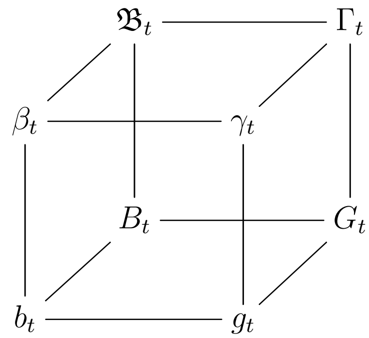
$$\begin{matrix} NC^s & \leftarrow & NC_2 & \leftarrow & \mathcal{NC}_2 \\ \downarrow & & \downarrow & & \downarrow \\ P^s & \leftarrow & P_2 & \leftarrow & \mathcal{P}_2 \end{matrix}$$

and where $|.|$ is the number of blocks.

PROOF. This follows by putting together the various moment results that we have. □

The above result is quite nice, and is complete as well, containing all the moment results that we have established so far, throughout this book. However, forgetting about being as general as possible, we can in fact do better. Nothing in life is better than having some 3D orientation, and as a main application of the above, we can modify a bit the above diagram, as to have a nice-looking cube, as follows:

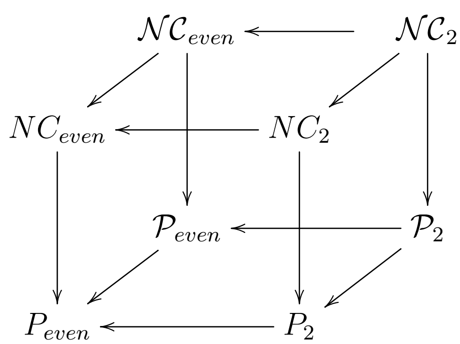
THEOREM 16.20. *The moments of the main central limiting measures,*



are always given by the same formula, involving partitions, namely

$$M_k = \sum_{\pi \in D(k)} t^{|\pi|}$$

where the sets of partitions $D(k)$ in question are respectively



and where $|.|$ is the number of blocks.

PROOF. This follows by putting together the various moment results that we have. To be more precise, the result follows from Theorem 16.19, by restricting the attention on the left to the cases $s = 2, \infty$, which can be thought of as being "fully real" and "purely complex", and then folding the 8-measure diagram into a cube, as above. □

16d. Matrix models

Let us discuss now the relation between the above free PLT theory and the random matrices. As a starting point, the free Poisson laws π_t that we found in the above, via the free PLT, coincide with the Marchenko-Pastur laws, shown in chapter 14 to appear as limiting laws for the complex Wishart matrices. This is certainly nice, conceptually speaking, but the point is that we can now truly improve the Marchenko-Pastur result from chapter 14, with an asymptotic freeness statement added, as follows:

THEOREM 16.21. *Given a family of sequences of complex Wishart matrices,*

$$Z_N^i = Y_N^i (Y_N^i)^* \in M_N(L^\infty(X)) \quad , \quad i \in I$$

with each Y_N^i being a $N \times M$ matrix, with entries following the normal law G_1, and with all these entries being pairwise independent, the rescaled sequences of matrices

$$\frac{Z_N^i}{N} \in M_N(L^\infty(X)) \quad , \quad i \in I$$

become with $M = tN \to \infty$ Marchenko-Pastur, each following the law π_t, and free.

PROOF. Here the first assertion is the Marchenko-Pastur theorem, and the second assertion follows from the freeness result for the Gaussian matrices, from chapter 15. □

At a more technical level now, many things to be done, which promise to be quite technical. Let us start with the multiplicative models. We will first restrict attention to the case $t = 1$, since we have $\beta_t^s = \pi^{\boxtimes s-1} \boxtimes \pi^{\boxplus t}$, and therefore matrix models for β_t^s will follow from matrix models for $\pi^{\boxtimes s}$. We first have the following result:

THEOREM 16.22. *Let $G_1, \ldots, G_s$ be a family of $N \times N$ independent matrices formed by independent centered Gaussian variables, of variance $1/N$. Then with*

$$M = G_1 \ldots G_s$$

the moments of the spectral distribution of MM^* converge, up to a normalization, to the corresponding moments of β_1^s, as $N \to \infty$.

PROOF. We prove this by recurrence. At $s = 1$ it is well-known that MM^* is a model for $\beta_1^1 = \pi$. So, assume that the result holds for $s - 1 \geq 1$. We have:

$$\begin{aligned} tr(MM^*)^k &= tr(G_1 \ldots G_s G_s^* \ldots G_1^*)^k \\ &= tr\big(G_1 (G_2 \ldots G_s G_s^* \ldots G_1^* G_1)^{k-1} G_2 \ldots G_s G_s^* \ldots G_1^*\big) \end{aligned}$$

We can pass the first G_1 matrix to the right, and we get:

$$\begin{aligned} tr(MM^*)^k &= tr\big((G_2 \ldots G_s G_s^* \ldots G_1^* G_1)^{k-1} G_2 \ldots G_s G_s^* \ldots G_1^* G_1\big) \\ &= tr(G_2 \ldots G_s G_s^* \ldots G_1^* G_1)^k \\ &= tr((G_2 \ldots G_s G_s^* \ldots G_2^*)(G_1^* G_1))^k \end{aligned}$$

We know that $G_1^* G_1$ is a Wishart matrix, hence is a model for π:

$$G_1^* G_1 \sim \pi$$

Also, we know by recurrence that $G_2 \ldots G_s G_s^* \ldots G_2^*$ gives a matrix model for β_1^{s-1}:

$$G_2 \ldots G_s G_s^* \ldots G_2^* \sim \beta_1^{s-1}$$

Now since the matrices $G_1^* G_1$ and $G_2 \ldots G_s G_s^* \ldots G_2^*$ are asymptotically free, their product gives a matrix model for $\pi_{s-1,1} \boxtimes \pi_{11} = \beta_1^s$, and we are done. □

We have as well the following result, which is of different nature:

THEOREM 16.23. *If W is a complex Wishart matrix of parameters (sN,N) and*

$$D=\begin{pmatrix}1_N&0&&0\\0&w1_N&&0\\&&\ddots&\\0&0&&w^{s-1}1_N\end{pmatrix}$$

with $w=e^{2\pi i/s}$ then the moments of the spectral distribution of $(DW)^s$ converge, up to a normalization, to the corresponding moments of β_1^s, as $N\to\infty$.

PROOF. We use the following complex Wishart matrix formula of Graczyk, Letac and Massam, whose proof is via standard combinatorics:

$$E(Tr(DW)^K)=\sum_{\sigma\in S_K}\frac{M^{\gamma(\sigma^{-1}\pi)}}{M^K}\,r_\sigma(D)$$

Here W is by definition a complex Wishart matrix of parameters (M,N), and D is a deterministic $M\times M$ matrix. As for the right term, this is as follows:

(1) π is the cycle $(1,\ldots,K)$.

(2) $\gamma(\sigma)$ is the number of disjoint cycles of σ.

(3) If we denote by $C(\sigma)$ the set of such cycles and for any cycle c, by $|c|$ its length, then the function on the right is given by:

$$r_\sigma(D)=\prod_{c\in C(\sigma)}Tr(D^{|c|})$$

In our situation we have $K=sk$ and $M=sN$, and we get:

$$E(Tr(DW)^{sk})=\sum_{\sigma\in S_{sk}}\frac{(sN)^{\gamma(\sigma^{-1}\pi)}}{(sN)^{sk}}\,r_\sigma(D)$$

Now since D is uniformly formed by s-roots of unity, we have:

$$Tr(D^p)=\begin{cases}sN & \text{if } s|p\\ 0 & \text{if } s\not|p\end{cases}$$

Thus if we denote by S_{sk}^s the set of permutations $\sigma\in S_{sk}$ having the property that all the cycles of σ have length multiple of s, the above formula reads:

$$E(Tr(DW)^{sk})=\sum_{\sigma\in S_{sk}^s}\frac{(sN)^{\gamma(\sigma^{-1}\pi)}}{(sN)^{sk}}\,(sN)^{\gamma(\sigma)}$$

In terms of the normalized trace tr, we obtain the following formula:

$$E(tr(DW)^{sk})=\sum_{\sigma\in S_{sk}^s}(sN)^{\gamma(\sigma^{-1}\pi)+\gamma(\sigma)-sk-1}$$

The exponent on the right, say L_σ, can be estimated by using the distance on the Cayley graph of S_{sk}, in the following way:

$$\begin{aligned} L_\sigma &= \gamma(\sigma^{-1}\pi)+\gamma(\sigma)-sk-1 \\ &= (sk-d(\sigma,\pi))+(sk-d(e,\sigma))-sk-1 \\ &= sk-1-(d(e,\sigma)+d(\sigma,\pi)) \\ &\leq sk-1-d(e,\pi) \\ &= 0 \end{aligned}$$

Now when taking the limit $N\to\infty$ in the above formula of $E(tr(DW)^{sk})$, the only terms that count are those coming from permutations $\sigma\in S_{sk}^s$ having the property $L_\sigma=0$, which each contribute with a 1 value. We therefore obtain:

$$\begin{aligned} \lim_{N\to\infty} E(tr(DW)^{sk}) &= \#\{\sigma\in S_{sk}^s \mid L_\sigma=0\} \\ &= \#\{\sigma\in S_{sk}^s \mid d(e,\sigma)+d(\sigma,\pi)=d(e,\pi)\} \\ &= \#\{\sigma\in S_{sk}^s \mid \sigma\in[e,\pi]\} \end{aligned}$$

But this number that we obtained is well-known to be the same as the number of noncrossing partitions of $\{1,\ldots,sk\}$ having all blocks of size multiple of s. Thus we have reached to the sets $NC_s(k)$ from the above, and we are done. □

As a consequence of the above random matrix formula, we have the following alternative approach to the free CPLT, in the case of the free Bessel laws:

THEOREM 16.24. *The moments of the free Bessel law π_{s1} with $s\in\mathbb{N}$ coincide with those of the variable*

$$\left(\sum_{k=1}^s w^k\alpha_k\right)^s$$

where $\alpha_1,\ldots,\alpha_s$ are free random variables, each of them following the free Poisson law of parameter $1/s$, and $w=e^{2\pi i/s}$.

PROOF. This is something that we already know, coming from the combinatorics of the free CPLT, but we can prove this now by using random matrices as well. For this purpose, let $G_1,\ldots,G_s$ be a family of independent $sN\times N$ matrices formed by independent, centered complex Gaussian variables, of variance $1/(sN)$. The following matrices $H_1,\ldots,H_s$ are then complex Gaussian and independent as well:

$$H_k=\frac{1}{\sqrt{s}}\sum_{p=1}^s w^{kp}G_p$$

Thus the following matrix provides a model for the variable $\Sigma w^k \alpha_k$:

$$\begin{aligned}
M &= \sum_{k=1}^{s} w^k H_k H_k^* \\
&= \frac{1}{s}\sum_{k=1}^{s}\sum_{p=1}^{s}\sum_{q=1}^{s} w^{k+kp-kq} G_p G_q^* \\
&= \sum_{p=1}^{s}\sum_{q=1}^{s}\left(\frac{1}{s}\sum_{k=1}^{s}\left(w^{1+p-q}\right)^k\right) G_p G_q^* \\
&= G_1G_2^* + G_2G_3^* + \ldots + G_{s-1}G_s^* + G_sG_1^*
\end{aligned}$$

Now observe that this matrix can be written as follows:

$$\begin{aligned}
M &= \begin{pmatrix} G_1 & G_2 & \ldots & G_{s-1} & G_s \end{pmatrix} \begin{pmatrix} G_2^* \\ G_3^* \\ \vdots \\ G_s^* \\ G_1^* \end{pmatrix} \\
&= \begin{pmatrix} G_1 & G_2 & \ldots & G_{s-1} & G_s \end{pmatrix} \begin{pmatrix} 0 & 1_N & 0 & \ldots & 0 \\ 0 & 0 & 1_N & \ldots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \ldots & 1_N \\ 1_N & 0 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} G_1^* \\ G_2^* \\ \vdots \\ G_{s-1}^* \\ G_s^* \end{pmatrix} \\
&= GOG^*
\end{aligned}$$

In this formula $G = (G_1 \ \ldots \ G_s)$ is the $sN \times sN$ Gaussian matrix obtained by concatenating $G_1, \ldots, G_s$, and O is the matrix in the middle. But this latter matrix is of the form $O = UDU^*$ with U unitary, so and we have:

$$M = GUDU^*G^*$$

Now since GU is a Gaussian matrix, M has the same law as $M' = GDG^*$. By using this, we obtain the following moment formula:

$$\begin{aligned}
E\left(\left(\sum_{l=1}^{s} w^l \alpha_l\right)^{sk}\right) &= \lim_{N\to\infty} E(tr(M^{sk})) \\
&= \lim_{N\to\infty} E(tr(GDG^*)^{sk}) \\
&= \lim_{N\to\infty} E(tr(D(G^*G))^{sk})
\end{aligned}$$

Thus with $W = G^*G$ we get the result. $\square$

As a last topic regarding the free CPLT, which is perhaps the most important, let us review now the results regarding the block-modified Wishart matrices from chapter 14, with free probability tools. We will see in particular that the laws obtained there are free combinations of free Poisson laws, or compound free Poisson laws.

Consider a complex Wishart matrix of parameters (dn, dm). In other words, we start with a $dn \times dm$ matrix Y having independent complex G_1 entries, and we set:

$$W = YY^*$$

This matrix has size $dn \times dn$, and is best thought of as being a $d \times d$ array of $n \times n$ matrices. We will be interested here in the study of the block-modified versions of W, obtained by applying to the $n \times n$ blocks a given linear map, as follows:

$$\varphi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$$

We recall from chapter 8 that we have the following asymptotic moment formula, extending the usual moment computation for the Wishart matrices:

Theorem 16.25. *The asymptotic moments of a block-modified Wishart matrix*

$$\widetilde{W} = (id \otimes \varphi)W$$

with parameters $d, m, n \in \mathbb{N}$, as above, are given by the formula

$$\lim_{d\to\infty} M_e\left(\frac{\widetilde{W}}{d}\right) = \sum_{\sigma\in NC_p} (mn)^{|\sigma|}(M_e^\sigma \otimes M_e^\gamma)(\Lambda)$$

where $\Lambda \in M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$ is the square matrix associated to $\varphi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$.

Proof. This is something that we know well from chapter 14, coming from the Wick formula, and with the correspondence between linear maps $\varphi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ and square matrices $\Lambda \in M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$ being as well explained there. □

As explained in chapter 14, it is possible to further build on the above result, with some concrete applications, by doing some combinatorics and calculus. With the free probability theory that we learned so far, we can now clarify all this. We first have:

Proposition 16.26. *Given a square matrix $\Lambda \in M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$, having distribution*

$$\rho = law(\Lambda)$$

the moments of the compound free Poisson law $\pi_{mn\rho}$ are given by

$$M_e(\pi_{mn\rho}) = \sum_{\sigma\in NC_p} (mn)^{|\sigma|}(M_e^\sigma \otimes M_e^\sigma)(\Lambda)$$

for any choice of the extra parameter $m \in \mathbb{N}$.

PROOF. This can be proved in several ways, as follows:

(1) A first method is by a straightforward computation, based on the general formula of the R-transform of the compound free Poisson laws, given in the above, and we will leave the computations here, which are all elementary, as an instructive exercise.

(2) Another method is by using the well-known fact that the free cumulants of $\pi_{mn\rho}$ coincide with the moments of $mn\rho$. Thus, these free cumulants are given by:

$$\begin{aligned}\kappa_e(\pi_{mn\rho}) &= M_e(mn\rho)\\ &= mn\cdot M_e(\Lambda)\\ &= mn\cdot(M_e^\gamma\otimes M_e^\gamma)(\Lambda)\end{aligned}$$

By using now Speicher's free moment-cumulant formula, this gives the result. □

We can see now an obvious similarity with the formula in Theorem 16.25. In order to exploit this similarity, let us introduce:

DEFINITION 16.27. *We call a square matrix $\Lambda\in M_n(\mathbb{C})\otimes M_n(\mathbb{C})$ multiplicative when*

$$(M_e^\sigma\otimes M_e^\gamma)(\Lambda)=(M_e^\sigma\otimes M_e^\sigma)(\Lambda)$$

holds for any $p\in\mathbb{N}$, any exponents $e_1,\ldots,e_p\in\{1,*\}$, and any $\sigma\in NC_p$.

This notion is something quite technical, but we will see many examples in what follows. For instance, the square matrices Λ coming from the basic linear maps φ appearing in chapter 14 are all multiplicative. Now with the above notion in hand, we can formulate an asymptotic result regarding the block-modified Wishart matrices, as follows:

THEOREM 16.28. *Consider a block-modified Wishart matrix*

$$\widetilde{W}=(id\otimes\varphi)W$$

and assume that the matrix $\Lambda\in M_n(\mathbb{C})\otimes M_n(\mathbb{C})$ associated to φ is multiplicative. Then

$$\frac{\widetilde{W}}{d}\sim\pi_{mn\rho}$$

holds, in moments, in the $d\to\infty$ limit, where $\rho=law(\Lambda)$.

PROOF. By comparing the moment formulae in Theorem 16.25 and in Proposition 16.26, we conclude that the asymptotic formula $\frac{\widetilde{W}}{d}\sim\pi_{mn\rho}$ is equivalent to the following equality, which should hold for any $p\in\mathbb{N}$, and any exponents $e_1,\ldots,e_p\in\{1,*\}$:

$$\sum_{\sigma\in NC_p}(mn)^{|\sigma|}(M_e^\sigma\otimes M_e^\gamma)(\Lambda)=\sum_{\sigma\in NC_p}(mn)^{|\sigma|}(M_e^\sigma\otimes M_e^\sigma)(\Lambda)$$

Now by assuming that Λ is multiplicative, in the sense of Definition 16.27, these two sums are trivially equal, and this gives the result. □

Many other things can be said, as a continuation of the above.

16e. Exercises

Congratulations for having read this book, and no exercises for this final chapter.

Bibliography

[1] V.I. Arnold, Ordinary differential equations, Springer (1973).

[2] V.I. Arnold, Mathematical methods of classical mechanics, Springer (1974).

[3] V.I. Arnold, Catastrophe theory, Springer (1974).

[4] V.I. Arnold, Lectures on partial differential equations, Springer (1997).

[5] V.I. Arnold and B.A. Khesin, Topological methods in hydrodynamics, Springer (1998).

[6] M.F. Atiyah, K-theory, CRC Press (1964).

[7] M.F. Atiyah, The geometry and physics of knots, Cambridge Univ. Press (1990).

[8] T. Banica, Calculus and applications (2026).

[9] T. Banica, Measure and integration (2025).

[10] T. Banica, Methods of free probability (2024).

[11] R.J. Baxter, Exactly solved models in statistical mechanics, Academic Press (1982).

[12] I. Bengtsson and K. Życzkowski, Geometry of quantum states, Cambridge Univ. Press (2006).

[13] N. Berline, E. Getzler and M. Vergne, Heat kernels and Dirac operators, Springer (2004).

[14] S.J. Blundell and K.M. Blundell, Concepts in thermal physics, Oxford Univ. Press (2006).

[15] S.M. Carroll, Spacetime and geometry, Cambridge Univ. Press (2004).

[16] A.R. Choudhuri, Astrophysics for physicists, Cambridge Univ. Press (2012).

[17] A. Connes, Noncommutative geometry, Academic Press (1994).

[18] J.B. Conway, A course in functional analysis, Springer (1985).

[19] W.N. Cottingham and D.A. Greenwood, An introduction to the standard model of particle physics, Cambridge Univ. Press (2012).

[20] P.A. Davidson, Introduction to magnetohydrodynamics, Cambridge Univ. Press (2001).

[21] P.A.M. Dirac, Principles of quantum mechanics, Oxford Univ. Press (1930).

[22] M.P. do Carmo, Differential geometry of curves and surfaces, Dover (1976).

[23] M.P. do Carmo, Riemannian geometry, Birkhäuser (1992).

[24] S. Dodelson, Modern cosmology, Academic Press (2003).

[25] R. Durrett, Probability: theory and examples, Cambridge Univ. Press (1990).

[26] A. Einstein, Relativity: the special and the general theory, Dover (1916).

[27] L.C. Evans, Partial differential equations, AMS (1998).

[28] W. Feller, An introduction to probability theory and its applications, Wiley (1950).

[29] E. Fermi, Thermodynamics, Dover (1937).

[30] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics I: mainly mechanics, radiation and heat, Caltech (1963).

[31] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics II: mainly electromagnetism and matter, Caltech (1964).

[32] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics III: quantum mechanics, Caltech (1966).

[33] R.P. Feynman and A.R. Hibbs, Quantum mechanics and path integrals, Dover (1965).

[34] P. Flajolet and R. Sedgewick, Analytic combinatorics, Cambridge Univ. Press (2009).

[35] A.P. French, Special relativity, Taylor and Francis (1968).

[36] W. Fulton, Algebraic topology, Springer (1995).

[37] W. Fulton and J. Harris, Representation theory, Springer (1991).

[38] C. Godsil and G. Royle, Algebraic graph theory, Springer (2001).

[39] H. Goldstein, C. Safko and J. Poole, Classical mechanics, Addison-Wesley (1980).

[40] M.B. Green, J.H. Schwarz and E. Witten, Superstring theory, Cambridge Univ. Press (2012).

[41] D.J. Griffiths, Introduction to electrodynamics, Cambridge Univ. Press (2017).

[42] D.J. Griffiths and D.F. Schroeter, Introduction to quantum mechanics, Cambridge Univ. Press (2018).

[43] D.J. Griffiths, Introduction to elementary particles, Wiley (2020).

[44] G.H. Hardy and E.M. Wright, An introduction to the theory of numbers, Oxford Univ. Press (1938).

[45] J. Harris, Algebraic geometry, Springer (1992).

[46] R. Hartshorne, Algebraic geometry, Springer (1977).

[47] A. Hatcher, Algebraic topology, Cambridge Univ. Press (2002).

[48] H. Hofer and E. Zehnder, Symplectic invariants and Hamiltonian dynamics, Birkhäuser (1994).

[49] L. Hörmander, The analysis of linear partial differential operators, Springer (1983).

[50] R.A. Horn and C.R. Johnson, Matrix analysis, Cambridge Univ. Press (1985).

[51] K. Huang, Introduction to statistical physics, CRC Press (2001).

[52] K. Huang, Quantum field theory, Wiley (1998).

[53] K. Huang, Quarks, leptons and gauge fields, World Scientific (1982).

[54] J.E. Humphreys, Introduction to Lie algebras and representation theory, Springer (1972).

[55] K. Ireland and M. Rosen, A classical introduction to modern number theory, Springer (1982).

[56] V.F.R. Jones, Subfactors and knots, AMS (1991).

[57] L.P. Kadanoff, Statistical physics: statics, dynamics and renormalization, World Scientific (2000).

[58] T. Kibble and F.H. Berkshire, Classical mechanics, Imperial College Press (1966).

[59] M. Kumar, Quantum: Einstein, Bohr, and the great debate about the nature of reality, Norton (2009).

[60] T. Lancaster and K.M. Blundell, Quantum field theory for the gifted amateur, Oxford Univ. Press (2014).

[61] L.D. Landau and E.M. Lifshitz, Mechanics, Pergamon Press (1960).

[62] L.D. Landau and E.M. Lifshitz, The classical theory of fields, Addison-Wesley (1951).

[63] L.D. Landau and E.M. Lifshitz, Quantum mechanics: non-relativistic theory, Pergamon Press (1959).

[64] S. Lang, Algebra, Addison-Wesley (1993).

[65] P. Lax, Linear algebra and its applications, Wiley (2007).

[66] P. Lax, Functional analysis, Wiley (2002).

[67] P. Lax and M.S. Terrell, Calculus with applications, Springer (2013).

[68] P. Lax and M.S. Terrell, Multivariable calculus with applications, Springer (2018).

[69] D. McDuff and D. Salamon, Introduction to symplectic topology, Oxford Univ. Press (2017).

[70] M.L. Mehta, Random matrices, Elsevier (2004).

[71] M.A. Nielsen and I.L. Chuang, Quantum computation and quantum information, Cambridge Univ. Press (2000).

[72] R.K. Pathria and and P.D. Beale, Statistical mechanics, Elsevier (1972).

[73] A. Peres, Quantum theory: concepts and methods, Kluwer (1993).

[74] P. Petersen, Linear algebra, Springer (2012).

[75] W. Rudin, Principles of mathematical analysis, McGraw-Hill (1964).

[76] W. Rudin, Real and complex analysis, McGraw-Hill (1966).

[77] W. Rudin, Functional analysis, McGraw-Hill (1973).

[78] W. Rudin, Fourier analysis on groups, Dover (1974).

[79] B. Ryden, Introduction to cosmology, Cambridge Univ. Press (2002).

[80] D.V. Schroeder, An introduction to thermal physics, Oxford Univ. Press (1999).

[81] J. Schwinger and B.H. Englert, Quantum mechanics: symbolism of atomic measurements, Springer (2001).

[82] J.P. Serre, A course in arithmetic, Springer (1973).

[83] J.P. Serre, Linear representations of finite groups, Springer (1977).

[84] I.R. Shafarevich, Basic algebraic geometry, Springer (1974).

[85] R. Shankar, Fundamentals of physics, Yale Univ. Press (2014).

[86] A.M. Steane, Thermodynamics, Oxford Univ. Press (2016).

[87] S. Sternberg, Dynamical systems, Dover (2010).

[88] J.R. Taylor, Classical mechanics, Univ. Science Books (2003).

[89] D.V. Voiculescu, Addition of certain noncommuting random variables, *J. Funct. Anal.* **66** (1986), 323–346.

[90] D.V. Voiculescu, Limit laws for random matrices and free products, *Invent. Math.* **104** (1991), 201–220.

[91] D.V. Voiculescu, K.J. Dykema and A. Nica, Free random variables, AMS (1992).

[92] J. von Neumann, Mathematical foundations of quantum mechanics, Princeton Univ. Press (1955).

[93] J. von Neumann and O. Morgenstern, Theory of games and economic behavior, Princeton Univ. Press (1944).

[94] J. Watrous, The theory of quantum information, Cambridge Univ. Press (2018).

[95] S. Weinberg, Foundations of modern physics, Cambridge Univ. Press (2011).

[96] S. Weinberg, Lectures on quantum mechanics, Cambridge Univ. Press (2012).

[97] H. Weyl, The theory of groups and quantum mechanics, Princeton Univ. Press (1931).

[98] H. Weyl, The classical groups: their invariants and representations, Princeton Univ. Press (1939).

[99] H. Weyl, Space, time, matter, Princeton Univ. Press (1918).

[100] B. Zwiebach, A first course in string theory, Cambridge Univ. Press (2004).