Variables and moments

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. Variable, Moment

Abstract. This is an introduction to the study of random variables using moments, and related techniques from combinatorics and classical analysis. We first discuss the classical moment problem, with illustrations for the various classical discrete and continuous laws. Then we get into a more advanced study of the variables and moments, and their various functional transforms, using partitions and cumulants. We then investigate a number of more advanced analytic aspects, notably with a study of the orthogonal polynomials. Finally, we have a look at the complex random variables, with a discussion of the moment problem here, and of related combinatorial and analytic aspects.

Preface

This is an introduction to the study of random variables using moments, and related techniques from combinatorics and classical analysis, ranging from basic to advanced, with all needed preliminaries included. The book is organized in 4 parts, as follows:

I - We first discuss the classical moment problem, with illustrations for the various classical discrete and continuous laws.

II - Then we get into a more advanced study of the variables and moments, and their various functional transforms, using partitions and cumulants.

III - We then investigate a number of more specialized analytic aspects, notably with a study of the associated orthogonal polynomials.

IV - Finally, we have a look at the complex random variables, with a discussion of the moment problem here, and of related combinatorial and analytic aspects.

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

Cergy, January 2026
Teo Banica

Contents

Part I

Variables, moments

CHAPTER 1

Discrete laws

1a. Variables, moments

You surely know a bit about random variables $f : X \to \mathbb{R}$, their means, also called expectations, $E(f) \in \mathbb{R}$, and about their variances $V(f) = E(f^2) - E(f)^2 \geq 0$ too. However, this is not the end of the story, because in order to have a fine understanding of your variable $f : X \to \mathbb{R}$, you might need the higher moments $M_k = E(f^k)$ too. In fact, the sequence of moments, besides encoding the mean and the variance, via $E(f) = M_1$ and $V(f) = M_2 - M_1^2$, determines the law, and can be useful for many purposes.

We will be talking about this, in this book. In this present Part I we would first like to compute the moments of all standard probability distributions, discrete or continuous, as to get familiar with this notion. Then, we would like to understand how the sequence of moments determines the law, in general, and with illustrations too. And then, in what concerns more advanced aspects, we have Parts II, III and IV for this.

Getting started now, there are many possible entry points to probability, with a quite standard one, focusing on the discrete case, which is the simplest, being 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:

$$X=\Big\{1,\ldots,6\Big\}\quad,\quad P(i)=p_i\quad,\quad p_i\geq0\ ,\ \sum_i p_i=1$$

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:

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

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 ahead now, let us go back to the context of Definition 1.1, which is the most convenient one, technically speaking. As usual in probability, we are mainly interested in winning. But, winning what? In case we are dealing with a usual die, what we win is what the die says, and on average, what we win is the following quantity:

$$E=\frac{1+2+3+4+5+6}{6}=3.5$$

In case we are dealing with the biased die in Example 1.3, again what we win is what the die says, and on average, what we win is the following quantity:

$$E=\sum_i i\times p_i$$

With this understood, what about coins? Here, before doing any computation, we have to assign some numbers to our events, and a standard choice here is as follows:

$$f:\{\text{heads, tails}\}\to\mathbb{R}\quad,\quad f(\text{heads})=1\quad,\quad f(\text{tails})=0$$

With this choice made, what we can expect to win is the following quantity:

$$\begin{aligned}E(f) &= f(\text{heads})\times P(\text{heads})+f(\text{tails})\times P(\text{tails})\\ &= 1\times\frac{1}{2}+0\times\frac{1}{2}\\ &= \frac{1}{2}\end{aligned}$$

Of course, in the case where the coin is biased, this computation will lead to a different outcome. And also, with a different convention for f, we will get a different outcome too. Moreover, we can combine if we want these two degrees of flexibility.

In short, you get the point. In order to do some math, in the context of Definition 1.1, we need a random variable $f:X\to\mathbb{R}$, and the math will consist in computing the expectation of this variable, $E(f)\in\mathbb{R}$. Alternatively, in order to do some business in the context of Definition 1.1, we need some form of "money", and our random variable $f:X\to\mathbb{R}$ will stand for that money, and then $E(f)\in\mathbb{R}$, for the average gain.

Let us axiomatize this situation as follows:

DEFINITION 1.5. *A random variable on a probability space X is a function*

$$f:X\to\mathbb{R}$$

and the expectation of such a random variable is the quantity

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

which is best thought as being the average gain, when the game is played.

Here the word "game" refers to the probability space interpretation from Conclusion 1.4. Indeed, in that context, with our discrete set of events X being thought of as corresponding to a generalized die, and by thinking of f as representing some sort of money, the above quantity $E(f)$ is what we win, on average, when playing the game.

We have already seen some good illustrations for Definition 1.5, so time now to get into more delicate aspects. Imagine that you want to set up some sort of business, with your variable $f:X\to\mathbb{R}$. You are of course mostly interested in the expectation $E(f)\in\mathbb{R}$, but passed that, the way this expectation comes in matters too. For instance:

(1) When your variable is constant, $f=c$, you certainly have $E(f)=c$, and your business will run smoothly, with not so many surprises on the way.

(2) On the opposite, for a complicated variable satisfying $E(f) = c$, your business will be more bumpy, with wins or loses on the way, depending on your skills.

In short, and extrapolating now from business to mathematics, physics, chemistry and everything else, we must complement Definition 1.5 with something finer, regarding the "quality" of the expectation $E(f) \in \mathbb{R}$ appearing there. And the first thought here, which is the correct one, goes to the following number, called variance of our variable:

$$\begin{aligned} V(f) &= E\left((f - E(f))^2\right) \\ &= E(f^2) - E(f)^2 \end{aligned}$$

However, let us not stop here. For a total control of your business, be that of financial, mathematical, physical or chemical type, you will certainly want to know more about your variable $f : X \to \mathbb{R}$. Which leads us into general moments, constructed as follows:

Definition 1.6. *The moments of a variable $f : X \to \mathbb{R}$ are the numbers*

$$M_k = E(f^k)$$

which satisfy $M_0 = 1$, then $M_1 = E(f)$, and then $V(f) = M_2 - M_1^2$.

And, good news, with this we have all the needed tools in our bag for doing some good business. To put things in a very compacted way, M_0 is about foundations, M_1 is about running some business, M_2 is about running that business well, and M_3 and higher are advanced level, about ruining all the competing businesses.

As a further piece of basic probability, coming this time as a theorem, we have:

Theorem 1.7. *Given a random variable $f : X \to \mathbb{R}$, if we define its law as being*

$$\mu = \sum_{x \in X} P(x)\delta_{f(x)}$$

regarded as probability measure on $\mathbb{R}$, then the moments are given by the formula

$$E(f^k) = \int_{\mathbb{R}} y^k d\mu(y)$$

with the usual convention that each Dirac mass integrates up to 1.

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

(1) To start with, given a random variable $f : X \to \mathbb{R}$, we can certainly talk about its law μ, as being the formal linear combination of Dirac masses in the statement. Our claim is that this is a probability measure on $\mathbb{R}$, in the sense of Definition 1.1. Indeed, the weight of each point $y \in \mathbb{R}$ is the following quantity, which is positive, as it should:

$$d\mu(y) = \sum_{f(x)=y} P(x)$$

Moreover, the total mass of this measure is 1, as it should, due to:

$$\begin{aligned}\sum_{y\in\mathbb{R}} d\mu(y) &= \sum_{y\in\mathbb{R}}\sum_{f(x)=y} P(x)\\ &= \sum_{x\in X} P(x)\\ &= 1\end{aligned}$$

Thus, we have indeed a probability measure on $\mathbb{R}$, in the sense of Definition 1.1.

(2) Still talking basics, let us record as well the following alternative formula for the law, which is clear from definitions, and that we will often use, in what follows:

$$\mu = \sum_{y\in\mathbb{R}} P(f=y)\delta_y$$

(3) Now let us compute the moments of f. With the usual convention that each Dirac mass integrates up to 1, as mentioned in the statement, we have:

$$\begin{aligned}E(f^k) &= \sum_{x\in X} P(x)f(x)^k\\ &= \sum_{y\in\mathbb{R}} y^k \sum_{f(x)=y} P(x)\\ &= \int_{\mathbb{R}} y^k d\mu(y)\end{aligned}$$

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

The above theorem is quite interesting, because we can see here a relation with integration, as we know it from calculus. In view of this, it is tempting to further go this way, by formulating the following definition, which is something purely mathematical:

Definition 1.8. *Given a set X, which can be finite, countable, or even uncountable, a discrete probability measure on it is a linear combination as follows,*

$$\mu = \sum_{x\in X} \lambda_x \delta_x$$

with the coefficients $\lambda_i \in \mathbb{R}$ satisfying $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. For $f : X \to \mathbb{R}$ we set

$$\int_X f(x)d\mu(x) = \sum_{x\in X} \lambda_x f(x)$$

with the convention that each Dirac mass integrates up to 1.

Observe that, with this, we are now into pure mathematics. However, and we insist on this, it is basic probability, as developed before, which is behind all this. Now by staying abstract for a bit more, with Definition 1.8 in hand, we can recover our previous basic probability notions, from Definition 1.1 and from Theorem 1.7, as follows:

Theorem 1.9. *With the above notion of discrete probability measure in hand:*

(1) *A discrete probability space is simply a space X, with a discrete probability measure on it ν. In this picture, the probability function is $P(x) = d\nu(x)$.*
(2) *Each random variable $f : X \to \mathbb{R}$ has a law, which is a discrete probability measure on $\mathbb{R}$. This law is given by $\mu = f_*\nu$, push-forward of ν by f.*

Proof. This might look a bit scary, but is in fact a collection of trivialities, coming straight from definitions, the details being as follows:

(1) Nothing much to say here, with our assertion being plainly clear, just by comparing Definition 1.1 and Definition 1.8. As a interesting comment, however, in the general context of Definition 1.8, a probability measure $\mu = \sum_{x\in X} \lambda_x\delta_x$ as there depends only on the following function, called density of our probability measure:

$$\varphi : X \to \mathbb{R} \quad , \quad \varphi(x) = \lambda_x$$

And, with this notion in hand, our equation $P(x) = d\nu(x)$ simply says that the probability function P is the density of ν. Which is something which is good to know.

(2) Pretty much the same story here, with our first assertion being clear, just by comparing Theorem 1.7 and Definition 1.8. As for the second assertion, consider more generally a probability space (X, ν), and a function $f : X \to Y$. We can then construct a probability measure $\mu = f_*\nu$ on Y, called push-forward of ν by f, as follows:

$$\nu = \sum_{x\in X} \lambda_x\delta_x \implies \mu = \sum_{y\in Y} \left(\sum_{f(x)=y} \lambda_x\right)\delta_y$$

Alternatively, at the level of the corresponding measures of the parts $Z \subset Y$, we have the following abstract formula, which looks more conceptual:

$$\mu(Z) = \nu(f^{-1}(Z))$$

In any case, one way or another we can talk about push-forward measures $\mu = f_*\nu$, and in the case of a random variable $f : X \to \mathbb{R}$, we obtain in this way the law of f. $\square$

Very nice all this, and needless to say, welcome to measure theory. In what follows we will rather go back to probability theory developed in the old way, as in the beginning of the present chapter, and keep developing that material, because we still have many interesting things to be learned. But, let us keep Definition 1.8 and Theorem 1.9, which are quite interesting, somewhere in our head. We will be back to these later.

1b. Binomial laws

Let us talk now about the key notion in probability, which is independence. Motivated by what happens when flipping a biased coin several times in a row, we have:

Definition 1.10. *Given $p\in[0,1]$, the Bernoulli law of parameter p is given by:*

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

More generally, the k-th binomial law of parameter p, with $k\in\mathbb N$, is given by

$$P(s)=p^s(1-p)^{k-s}\binom{k}{s}$$

with the Bernoulli law appearing at $k=1$, with $s=1,0$ here standing for win and lose.

Let us try now to understand the relation between the Bernoulli and binomial laws. Indeed, we know that the Bernoulli laws produce the binomial laws, simply by iterating the game, from 1 throw to $k\in\mathbb N$ throws. Obviously, what matters in all this is the "independence" of our coin throws, so let us record this finding, as follows:

Theorem 1.11. *The following happen, in the context of the biased coin game:*

(1) *The Bernoulli laws μ_{ber} produce the binomial laws μ_{bin}, by iterating the game $k\in\mathbb N$ times, via the independence of the throws.*
(2) *We have in fact $\mu_{bin}=\mu_{ber}^{*k}$, with $*$ being the convolution operation for real probability measures, given by $\delta_x*\delta_y=\delta_{x+y}$, and linearity.*

Proof. Obviously, this is something a bit informal, but let us prove this as stated, and we will come back later to it, with precise definitions, general theorems and everything. In what regards the first assertion, nothing to be said there, this is what life teaches us. As for the second assertion, the formula $\mu_{bin}=\mu_{ber}^{*k}$ there certainly looks like mathematics, so job for us to figure out what this exactly means. And, this can be done as follows:

(1) The first idea is to encapsulate the data from Definition 1.10 into the probability measures associated to the Bernoulli and binomial laws. For the Bernoulli law, the corresponding measure is as follows, with the δ symbols standing for Dirac masses:

$$\mu_{ber}=(1-p)\delta_0+p\delta_1$$

As for the binomial law, here the measure is as follows, constructed in a similar way, you get the point I hope, again with the δ symbols standing for Dirac masses:

$$\mu_{bin}=\sum_{s=0}^k p^s(1-p)^{k-s}\binom{k}{s}\delta_s$$

(2) Getting now to independence, the point is that, as we will soon discover abstractly, the mathematics there is that of the following formula, with $*$ standing for the convolution

operation for the real measures, which is given by $\delta_x * \delta_y = \delta_{x+y}$ and linearity:

$$\mu_{bin} = \underbrace{\mu_{ber} * \ldots * \mu_{ber}}_{k\ terms}$$

(3) To be more precise, this latter formula does hold indeed, as a straightforward application of the binomial formula, the formal proof being as follows:

$$\begin{aligned}
\mu_{ber}^{*k} &= \big((1-p)\delta_0 + p\delta_1\big)^{*k} \\
&= \sum_{s=0}^{k} p^s (1-p)^{k-s} \binom{k}{s} \delta_0^{*(k-s)} * \delta_1^{*s} \\
&= \sum_{s=0}^{k} p^s (1-p)^{k-s} \binom{k}{s} \delta_s \\
&= \mu_{bin}
\end{aligned}$$

(4) Summarizing, save for some uncertainties regarding what independence exactly means, mathematically speaking, and more on this in a moment, theorem proved. □

Getting to formal mathematical work now, let us start with the following straightforward definition, inspired by what happens for coins, dice and cards:

Definition 1.12. *We say that two variables $f, g : X \to \mathbb{R}$ are independent when*

$$P(f = x, g = y) = P(f = x)P(g = y)$$

happens, for any $x, y \in \mathbb{R}$.

As already mentioned, this is something very intuitive, inspired by what happens for coins, dice and cards. As a first result now regarding independence, we have:

Theorem 1.13. *Assuming that $f, g : X \to \mathbb{R}$ are independent, we have:*

$$E(fg) = E(f)E(g)$$

More generally, we have the following formula, for the mixed moments,

$$E(f^k g^l) = E(f^k)E(g^l)$$

and the converse holds, in the sense that this formula implies the independence of f, g.

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

$$\begin{aligned}
E(f^kg^l) &= \sum_{xy}x^ky^lP(f=x,g=y)\\
&= \sum_{xy}x^ky^lP(f=x)P(g=y)\\
&= \sum_x x^kP(f=x)\sum_y y^lP(g=y)\\
&= E(f^k)E(g^l)
\end{aligned}$$

As for the last assertion, this is clear too, because having the above computation work, for any $k,l\in\mathbb N$, amounts in saying that the independence formula for f,g holds. □

Regarding now the convolution operation, motivated by what we found before, in Theorem 1.11, let us start with the following abstract definition:

Definition 1.14. *Given a space X with a sum operation $+$, we can define the convolution of any two discrete probability measures on it,*

$$\mu=\sum_i a_i\delta_{x_i}\quad,\quad \nu=\sum_j b_j\delta_{y_j}$$

as being the discrete probability measure given by the following formula:

$$\mu*\nu=\sum_{ij}a_ib_j\delta_{x_i+y_j}$$

That is, the convolution operation $*$ is defined by $\delta_x*\delta_y=\delta_{x+y}$, and linearity.

As a first observation, our operation is well-defined, with $\mu*\nu$ being indeed a discrete probability measure, because the weights are positive, $a_ib_j\geq 0$, and their sum is:

$$\sum_{ij}a_ib_j=\sum_i a_i\sum_j b_j=1\times 1=1$$

Also, the above definition agrees with what we did before with coins, and Bernoulli and binomial laws. We have in fact the following general result:

Theorem 1.15. *Assuming that $f,g:X\to\mathbb R$ are independent, we have*

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

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

PROOF. We have indeed the following straightforward verification, based on the independence formula from Definition 1.12, and on Definition 1.14:

$$\begin{aligned}\mu_{f+g} &= \sum_{x\in\mathbb{R}} P(f+g=x)\delta_x \\ &= \sum_{y,z\in\mathbb{R}} P(f=y,g=z)\delta_{y+z} \\ &= \sum_{y,z\in\mathbb{R}} P(f=y)P(g=z)\delta_y * \delta_z \\ &= \left(\sum_{y\in\mathbb{R}} P(f=y)\delta_y\right) * \left(\sum_{z\in\mathbb{R}} P(g=z)\delta_z\right) \\ &= \mu_f * \mu_g\end{aligned}$$

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

Before going further, let us attempt as well to find a proof of Theorem 1.15, based on the moment characterization of independence, from Theorem 1.13. For this purpose, we will need the following standard fact, which is of certain theoretical interest:

THEOREM 1.16. *The sequence of moments*

$$M_k = \int_{\mathbb{R}} x^k d\mu(x)$$

uniquely determines the law.

PROOF. Indeed, assume that the law of our variable is as follows:

$$\mu = \sum_i \lambda_i \delta_{x_i}$$

The sequence of moments is then given by the following formula:

$$M_k = \sum_i \lambda_i x_i^k$$

But it is then standard calculus to recover the numbers $\lambda_i, x_i \in \mathbb{R}$, and so the measure μ, out of the sequence of numbers M_k. Indeed, assuming that the numbers x_i are $0 < x_1 < \ldots < x_n$ for simplifying, in the $k \to \infty$ limit we have the following formula:

$$M_k \sim \lambda_n x_n^k$$

Thus, we got the parameters $\lambda_n, x_n \in \mathbb{R}$ of our measure μ, and then by substracting them and doing an obvious recurrence, we get the other parameters $\lambda_i, x_i \in \mathbb{R}$ as well. We will leave the details here as an instructive exercise, and come back to this problem later in this book, with more advanced and clever methods for dealing with it. □

Getting back now to our philosophical question above, namely recovering Theorem 1.15 via moment technology, we can now do this, the result being as follows:

THEOREM 1.17. *Assuming that $f,g:X\to\mathbb R$ are independent, the measures*

$$\mu_{f+g}\quad,\quad\mu_f*\mu_g$$

have the same moments, and so, they coincide.

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, we have as well the following computation:

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

Thus, job done, and theorem proved, or rather Theorem 1.15 reproved. □

Getting back now to the basic theory of independence, here is now a second result, coming as a continuation of Theorem 1.15, which is something more advanced:

THEOREM 1.18. *Assuming that $f,g:X\to\mathbb R$ are independent, we have*

$$F_{f+g}=F_fF_g$$

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

PROOF. We have the following computation, using Theorem 1.15:

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

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.

Let us do now some computations. We recall from the above that the k-th binomial law of parameter $p \in (0,1)$, with $k \in \mathbb{N}$, is given by the following formula:

$$P(s) = p^s (1-p)^{k-s} \binom{k}{s}$$

As a first concrete result about these laws, we have:

THEOREM 1.19. *The mean of the k-th binomial law of parameter $p \in (0,1)$ is:*

$$E = kp$$

As for the variance and higher moments, these are given by similar formulae.

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

(1) In what regards the mean, this can be computed as follows:

$$\begin{aligned}
E &= \sum_{s=0}^{k} P(s)s \\
&= \sum_{s=0}^{k} p^s(1-p)^{k-s}\binom{k}{s}s \\
&= \sum_{s=1}^{k} p^s(1-p)^{k-s}\binom{k}{s}s \\
&= \sum_{s=1}^{k} p^s(1-p)^{k-s}\frac{k!}{(s-1)!(k-s)!} \\
&= k\sum_{s=1}^{k} p^s(1-p)^{k-s}\frac{(k-1)!}{(s-1)!(k-s)!} \\
&= k\sum_{r=0}^{k-1} p^{r+1}(1-p)^{k-r-1}\frac{(k-1)!}{r!(k-r-1)!} \\
&= kp\sum_{r=0}^{k-1} p^{r}(1-p)^{k-r-1}\frac{(k-1)!}{r!(k-r-1)!} \\
&= kp(p+(1-p))^{k-1} \\
&= kp
\end{aligned}$$

(2) As for the variance, and the higher moments too, these can be computed in a similar way, essentially by using the binomial formula. □

1c. Poisson limits

At a more advanced level, we have the Poisson Limit Theorem (PLT), that we would like to explain now. Let us start with the following definition:

Definition 1.20. *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 1.21. *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 1.22. *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 1.23. *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 1.21, 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 1.24 (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 1.25. *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 1.26. *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.

1d. Further distributions

We would like to discuss now some technical generalizations of the main laws that we saw so far, namely the binomial ones and the Poisson ones.

To start with, we can talk about negative binomial laws, also called Pascal laws. Many things can be said about these, in analogy with what we know about binomial laws.

In relation with Poisson laws, we have work to do too. Indeed, we have the following notion, extending the Poisson limit theory developed in the previous section:

DEFINITION 1.27. *Associated to any compactly supported positive measure ν on $\mathbb{C}$, not necessarily of mass 1, is the probability measure*

$$p_\nu=\lim_{n\to\infty}\left(\left(1-\frac{t}{n}\right)\delta_0+\frac{1}{n}\nu\right)^{*n}$$

where $t=mass(\nu)$, called compound Poisson law.

In what follows we will be mainly interested in the case where the measure ν is discrete, as is for instance the case for $\nu=t\delta_1$ with $t>0$, which produces the Poisson laws. The following standard result allows one to detect compound Poisson laws:

PROPOSITION 1.28. *For $\nu=\sum_{i=1}^s t_i\delta_{z_i}$ with $t_i>0$ and $z_i\in\mathbb{C}$, we have*

$$F_{p_\nu}(y)=\exp\left(\sum_{i=1}^s t_i(e^{iyz_i}-1)\right)$$

where F denotes the Fourier transform.

PROOF. Let η_n be the measure in Definition 1.27, under the convolution sign:

$$\eta_n=\left(1-\frac{t}{n}\right)\delta_0+\frac{1}{n}\nu$$

We have then the following computation:

$$\begin{aligned}F_{\eta_n}(y)=\left(1-\frac{t}{n}\right)+\frac{1}{n}\sum_{i=1}^s t_ie^{iyz_i}\quad&\implies\quad F_{\eta_n^{*n}}(y)=\left(\left(1-\frac{t}{n}\right)+\frac{1}{n}\sum_{i=1}^s t_ie^{iyz_i}\right)^n\\&\implies\quad F_{p_\nu}(y)=\exp\left(\sum_{i=1}^s t_i(e^{iyz_i}-1)\right)\end{aligned}$$

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

We have as well the following result, providing an alternative to Definition 1.27, and which will be our formulation here of the Compound Poisson Limit Theorem:

THEOREM 1.29 (CPLT). *For $\nu = \sum_{i=1}^s t_i\delta_{z_i}$ with $t_i > 0$ and $z_i \in \mathbb{C}$, we have*

$$p_\nu = \text{law}\left(\sum_{i=1}^s z_i\alpha_i\right)$$

where the variables α_i are Poisson (t_i), independent.

PROOF. Let α be the sum of Poisson variables in the statement, namely:

$$\alpha = \sum_{i=1}^s z_i\alpha_i$$

By using some standard Fourier transform formulae, we have:

$$\begin{aligned} F_{\alpha_i}(y) = \exp(t_i(e^{iy} - 1)) \quad &\implies \quad F_{z_i\alpha_i}(y) = \exp(t_i(e^{iyz_i} - 1)) \\ &\implies \quad F_\alpha(y) = \exp\left(\sum_{i=1}^s t_i(e^{iyz_i} - 1)\right) \end{aligned}$$

Thus we have indeed the same formula as in Proposition 1.28, as desired. □

At the level of main examples of compound Poisson laws, we have:

DEFINITION 1.30. *The Bessel law of level $s \in \mathbb{N} \cup \{\infty\}$ and parameter $t > 0$ is*

$$b_t^s = p_{t\varepsilon_s}$$

with ε_s being the uniform measure on the s-th roots of unity. The measures

$$b_t = b_t^2 \quad , \quad B_t = b_t^\infty$$

are called real Bessel law, and complex Bessel law.

In practice now, we first have to study the measures b_t^s in our standard way, meaning density, moments, Fourier, semigroup property, limiting theorems, and other aspects. In what regards limiting theorems, the measures b_t^s appear by definition via the CPLT, so done with that. As a consequence of this, however, let us record:

PROPOSITION 1.31. *The Bessel laws are given by*

$$b_t^s = \text{law}\left(\sum_{k=1}^s w^k a_k\right)$$

where $a_1, \ldots, a_s$ are Poisson (t) independent, and $w = e^{2\pi i/s}$.

PROOF. This follows indeed from Theorem 1.29. □

Many other things can be said about the Bessel laws. We will be back to this.

Finally, we can talk about hypergeometric laws. Again, we will be back to this.

1e. Exercises

This was a standard introduction to discrete probability, and as exercises, we have:

Exercise 1.32. *Compute all probabilities at poker.*

Exercise 1.33. *Compute the moments of binomial laws.*

Exercise 1.34. *Compute the moments of negative binomial laws.*

Exercise 1.35. *Compute the moments of real Bessel laws.*

Exercise 1.36. *Compute the moments of complex Bessel laws.*

Exercise 1.37. *Compute the moments of general Bessel laws.*

Exercise 1.38. *What can you say about the moments of compound Poisson laws?*

Exercise 1.39. *Compute the moments of hypergeometric laws.*

As bonus exercise, look up and learn some more discrete laws, not discussed here.

CHAPTER 2

Continuous laws

2a. Continuous laws

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

DEFINITION 2.1. *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 2.2. *Two variables $f, g \in L^\infty(X)$ are called independent when*

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

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 2.3. *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:

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

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 2.4. *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 2.3 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. □

This was for the foundations of probability theory, quickly explained. For further reading, a classical book is Feller [31]. A nice, more modern book is Durrett [28].

2b. Central limits

The main result in classical probability is the Central Limit Theorem (CLT), that we will explain now. Let us first discuss the normal distributions, that we will see later to appear as limiting laws in the CLT. We will need the following standard result:

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 integral in the statement. By using polar coordinates, namely $x = r\cos t$, $y = r\sin t$, with the corresponding Jacobian being r, we have:

$$\begin{aligned} I^2 &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-x^2-y^2} dx dy \\ &= \int_0^{2\pi} \int_0^\infty e^{-r^2} r\, dr dt \\ &= 2\pi \int_0^\infty \left(-\frac{e^{-r^2}}{2} \right)' dr \\ &= \pi \end{aligned}$$

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

We can now introduce the normal distributions, as follows:

DEFINITION 2.6. *The normal law of parameter* 1 *is the following measure:*

$$g_1 = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

More generally, the normal law of parameter $t > 0$ *is the following measure:*

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

These are also called Gaussian distributions, with "g" standing for Gauss.

The above laws are usually denoted $\mathcal{N}(0,1)$ and $\mathcal{N}(0,t)$, but since we will be doing in this book all kinds of probability, we will use simplified notations for all our measures. Let us mention as well that the normal laws traditionally have 2 parameters, the mean and the variance, but here we will not need the mean, all our theory using centered laws. Finally, observe that the above laws have indeed mass 1, as they should, due to:

$$\int_{\mathbb{R}} e^{-x^2/2t} dx = \int_{\mathbb{R}} e^{-y^2} \sqrt{2t}\, dy = \sqrt{2\pi t}$$

Generally speaking, the normal laws appear as bit everywhere, in real life. The reasons for this come from the Central Limit Theorem (CLT), that we will explain in a moment, after developing some more general theory. As a first result, we have:

PROPOSITION 2.7. *We have the variance formula*

$$V(g_t) = t$$

valid for any $t > 0$.

PROOF. The first moment is 0, because our normal law g_t is centered. As for the second moment, this can be computed as follows:

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

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

Here is another result, which is widely useful in practice:

THEOREM 2.8. *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. The Fourier transform formula can be established as follows:

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

As for $g_s * g_t = g_{s+t}$, this follows via Theorem 2.4, $\log F_{g_t}$ being linear in t. ☐

We are now ready to state and prove the CLT, as follows:

THEOREM 2.9 (CLT). *Given real variables $f_1, f_2, f_3, \ldots \in L^\infty(X)$ 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 $n \to \infty$, in moments.

PROOF. In terms of moments, the Fourier transform is given by:

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

Thus, the Fourier transform of the variable in the statement is:

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

But this function being the Fourier transform of g_t, we obtain the result. ☐

Let us discuss now some further properties of the normal law. We first have:

PROPOSITION 2.10. *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. We have the following computation, valid for any integer $k \in \mathbb{N}$:

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

Now recall from the proof of Proposition 2.7 that we have $M_0 = 1$, $M_1 = 0$. Thus by recurrence, we are led to the formula in the statement. □

We have the following alternative formulation of the above result:

PROPOSITION 2.11. *The moments of the normal law are the numbers*

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

where $P_2(k)$ *is the set of pairings of* $\{1, \ldots, k\}$.

PROOF. Let us count the pairings of $\{1, \ldots, k\}$. 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:

$$|P_2(k)| = (k-1)|P_2(k-2)|$$

As for the initial data, this is $P_1 = 0$, $P_2 = 1$. Thus, we are led to the result. □

We are not done yet, and here is one more improvement of the above:

THEOREM 2.12. *The moments of the normal law are the numbers*

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

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

PROOF. This follows indeed from Proposition 2.11, because the number of blocks of a pairing of $\{1, \ldots, k\}$ is trivially $k/2$, independently of the pairing. □

We will see later in this book that many other interesting probability distributions are subject to similar formulae regarding their moments, involving partitions.

2c. Rayleigh variables

Regarding now the Rayleigh laws, we have the following result:

THEOREM 2.13. *The moments of the Rayleigh law, given by*

$$R_t = law\left(\frac{a^2+b^2}{2}\right)$$

with $a, b \sim g_t$, independent, are given by the following formula, at $t = 1$,

$$M_p = p!$$

and are given by the formula $M_p = t^p p!$, in general.

PROOF. We have indeed the following computation, to start with:

$$\begin{aligned}
M_p &= \frac{1}{2^p}\int (a^2+b^2)^p \\
&= \frac{1}{2^p}\sum_r \binom{p}{r}\int a^{2r}\int b^{2p-2r} \\
&= \frac{1}{2^p}\sum_r \binom{p}{r}(2r)!!(2p-2r)!! \\
&= \frac{1}{2^p}\sum_r \frac{p!}{r!(p-r)!}\cdot\frac{(2r)!}{2^r r!}\cdot\frac{(2p-2r)!}{2^{p-r}(p-r)!} \\
&= \frac{p!}{4^p}\sum_r \binom{2r}{r}\binom{2p-2r}{p-r}
\end{aligned}$$

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 our moment computation, as follows:

$$M_p = \frac{p!}{4^p}\times 4^p = p!$$

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

2d. Further distributions

Further distributions.

2e. Exercises

Exercises:

Exercise 2.14.

Exercise 2.15.

Exercise 2.16.

Exercise 2.17.

Exercise 2.18.

Exercise 2.19.

Exercise 2.20.

Exercise 2.21.

Bonus exercise.

CHAPTER 3

Stieltjes inversion

3a. Complex analysis

Complex analysis.

3b. Cauchy transform

Cauchy transform.

3c. Stieltjes inversion

We have learned some basic probability in chapters 1-2, and in view of the material there, an interesting question is how to recover a probability measure out of its moments. And the answer here, which is something non-trivial, is as follows:

Theorem 3.1. *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. □

There is a discussion here regarding the atoms, too.

3d. Basic illustrations

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

FACT 3.2. *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, we will leave them as an instructive exercise.

Needless to say, counting loops on graphs, as in Fact 3.2, 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 3.3. *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 3.1. 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 3.4. *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 3.1. We obtain, a bit as before in Theorem 3.3, 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 3.5. *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 standard 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 3.1. □

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

Theorem 3.6. *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:

$$E_{2k} = \binom{2k}{k} = \frac{(2k)!}{k!k!} = D_k$$

$$E_{2k-1} = \binom{2k-1}{k} = \frac{(2k-1)!}{k!(k-1)!} = \frac{D_k}{2}$$

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

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

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. □

There are of course many other illustrations, for the Stieltjes inversion formula.

3e. Exercises

Exercises:

Exercise 3.7.

Exercise 3.8.

Exercise 3.9.

Exercise 3.10.

Exercise 3.11.

Exercise 3.12.

Exercise 3.13.

Exercise 3.14.

Bonus exercise.

CHAPTER 4

Hankel determinants

4a. Functional analysis

Functional analysis.

4b. Hankel determinants

The Stieltjes inversion formula 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, $E \geq 0$ shows that must have:

$$M_2 \geq M_1^2$$

In answer now, we have the following result:

THEOREM 4.1. *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:

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

(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:

$$\int_{\mathbb{R}} \left| \sum_{i=1}^n c_i x^i \right|^2 d\mu(x) = \int_{\mathbb{R}} \sum_{i,j=1}^n c_i \bar{c}_j x^{i+j} d\mu(x) = \sum_{i,j=1}^n c_i \bar{c}_j M_{i+j}$$

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

4c. Basic computations

Basic computations.

4d. The moment problem

The moment problem.

4e. Exercises

Exercises:

Exercise 4.2.

Exercise 4.3.

Exercise 4.4.

Exercise 4.5.

Exercise 4.6.

Exercise 4.7.

Exercise 4.8.

Exercise 4.9.

Bonus exercise.

Part II

Partitions, cumulants

CHAPTER 5

Cumulants

5a. Cumulants

Getting back to the basics now, probability and combinatorics at large, of quite general type, we have here the following key definition, due to Rota:

DEFINITION 5.1. *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 5.2. *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. Here all the theoretical assertions regarding moments and cumulants are clear from definitions, and the numerics are clear from definitions too. To be more precise, we know from Definition 5.1 that the cumulants are defined by the following formula:

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

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. □

Obviously, there should be some explicit formulae for the correspondences in Proposition 5.2. This is indeed the case, but things here are quite tricky, and we will discuss this later, once we will have enough motivations for the study of the cumulants.

5b. Basic examples

In order to get familiar with the cumulants, let us work out some examples. In what regards the basic probability measures, that we know from Part I, the cumulants are always given by simple formulae, as shown by the following result:

THEOREM 5.3. *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 $t, t, t, \ldots$*
(4) *For $\mu = b_t$ the cumulants are $0, t, 0, t, \ldots$*

PROOF. We have 4 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.

(4) For $\mu = b_t$ we have the following computation:

$$\begin{aligned} K_\mu(\xi) &= \log F_\mu(-i\xi) \\ &= \log\exp\left[\left(\frac{e^\xi + e^{-\xi}}{2} - 1\right)t\right] \\ &= \left(\frac{e^\xi + e^{-\xi}}{2} - 1\right)t \end{aligned}$$

But the plain coefficients of this series are the numbers $(1+(-1)^n)t/n!$, so the Taylor coefficients of this series are the numbers $0, t, 0, t, \ldots$, as claimed. □

5c. Linearization

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 5.4. *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 the following computation, valid for any $c, d \in \mathbb{R}$, which gives the results:

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

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. □

5d. Further examples

We have the following result, generalizing Theorem 5.3 (3) and (4), and which is something very useful, when dealing with the compound Poisson laws:

THEOREM 5.5. *For a compound Poisson law p_ν we have*

$$k_n(p_\nu) = M_n(\nu)$$

valid for any integer $n \geq 1$.

PROOF. We recall from the end of chapter 1 that associated to any compactly supported positive measure ν on $\mathbb{R}$, which is not necessarily of mass 1, is the following probability measure, where $t = mass(\nu)$, called compound Poisson law:

$$p_\nu = \lim_{n\to\infty} \left(\left(1 - \frac{t}{n}\right) \delta_0 + \frac{1}{n}\nu \right)^{*n}$$

In relation with our problem, we can assume, by using a continuity argument, that our measure ν is discrete, as follows, with $t_i > 0$ and $z_i \in \mathbb{R}$, and the sum being finite:

$$\nu = \sum_i t_i \delta_{z_i}$$

In this case, according to the compound PLT from chapter 1, we have the following formula, with the variables α_i being Poisson (t_i), and independent:

$$p_\nu = \mathrm{law}\left(\sum_{i=1}^s z_i \alpha_i \right)$$

Finally, let us recall as well from chapter 1 the formula of the Fourier transform of p_ν, which is as follows, and which can be used for proving the compound CLT:

$$F_{p_\nu}(y) = \exp\left(\sum_{i=1}^s t_i (e^{iyz_i} - 1) \right)$$

By using now this Fourier transform formula for p_ν, we obtain:

$$\begin{aligned}
K_{p_\nu}(\xi) &= \log F_{p_\nu}(-i\xi)\\
&= \log\exp\left[\sum_i t_i(e^{\xi z_i}-1)\right]\\
&= \sum_i t_i\sum_{n\geq 1}\frac{(\xi z_i)^n}{n!}\\
&= \sum_{n\geq 1}\frac{\xi^n}{n!}\sum_i t_i z_i^n\\
&= \sum_{n\geq 1}\frac{\xi^n}{n!}M_n(\nu)
\end{aligned}$$

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

5e. Exercises

Exercises:

Exercise 5.6.

Exercise 5.7.

Exercise 5.8.

Exercise 5.9.

Exercise 5.10.

Exercise 5.11.

Exercise 5.12.

Exercise 5.13.

Bonus exercise.

CHAPTER 6

Inversion formula

6a. Sets, partitions

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. We recall from chapter 5 that we have the following formulae, for the cumulants in terms of moments:

$$k_1 = M_1$$

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

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

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

$$\vdots$$

Also, we have the following formulae, for the moments in terms of cumulants:

$$M_1 = k_1$$

$$M_2 = k_1^2 + k_2$$

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

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

$$\vdots$$

In order to understand what exactly is going on, with moments and cumulants, which reminds a bit the Möbius inversion formula, we need to do some combinatorics, in relation with the set-theoretic partitions. We first have the following definition:

Definition 6.1. *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.

Thus, what we have is in fact a Möbius matrix, $M_{\pi\nu} = \mu(\pi, \nu)$. As a first example, the Möbius matrix $M_{\pi\nu} = \mu(\pi, \nu)$ of the lattice $P(2) = \{||, \sqcap\}$ is as follows:

$$M = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

At $k = 3$ now, we have the following formula for the Möbius matrix $M_{\pi\nu} = \mu(\pi, \nu)$, once again written with the indices picked increasing in $P(3) = \{|||, \sqcap|, \sqcap\!\!\!|\,, |\sqcap, \sqcap\!\sqcap\}$:

$$M = \begin{pmatrix} 1 & -1 & -1 & -1 & 2 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

In general now, the Möbius matrix of $P(k)$ looks a bit like the above matrices at $k = 2, 3$, being upper triangular, with 1 on the diagonal, and so on.

6b. Möbius function

Back to the general case now, the main interest in the Möbius function comes from the Möbius inversion formula, which can be formulated as follows:

THEOREM 6.2. *We have the following implication,*

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

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

PROOF. The above formula is in fact a linear algebra result, so let us start with some linear algebra. 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 6.1, so to the Möbius matrix M. Thus we have:

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

Now by applying this equality of matrices to vectors, regarded as complex functions on $P(n)$, we are led to the inversion formula in the statement. $\square$

As a first illustration, for $P(2)$ the formula $M=A^{-1}$ appears as follows:

$$\begin{pmatrix}1&-1\\0&1\end{pmatrix}=\begin{pmatrix}1&1\\0&1\end{pmatrix}^{-1}$$

At $k=3$ now, the formula $M=A^{-1}$ for $P(3)$ reads:

$$\begin{pmatrix}1&-1&-1&-1&2\\0&1&0&0&-1\\0&0&1&0&-1\\0&0&0&1&-1\\0&0&0&0&1\end{pmatrix}=\begin{pmatrix}1&1&1&1&1\\0&1&0&0&1\\0&0&1&0&1\\0&0&0&1&1\\0&0&0&0&1\end{pmatrix}^{-1}$$

In general, the inversion formula $M=A^{-1}$ looks quite similar.

6c. Inversion formula

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

Definition 6.3. *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 6.4. *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) First, it is clear from our conventions, from Definition 6.3, that the first set of formulae is equivalent to the second set of formulae, by multiplicativity over blocks.

(2) The other observation is that, due to the Möbius inversion formula, from Theorem 6.2, in the second set of formulae, the two formulae there are in fact equivalent.

(3) Summarizing, the 4 formulae in the statement are all equivalent. In what follows we will focus on the first 2 formulae, which are the most useful, in practice.

(4) 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 numerics:

$$M_1 = k_| = k_1$$

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

$$M_3 = k_{|||} + k_{\sqcap|} + k_{\sqcap\!\!\!|} + k_{|\sqcap} + k_{\sqcap\!\sqcap} = k_1^3 + 3k_1k_2 + k_3$$

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 the numerics:

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

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

$$k_1 = M_| = M_1$$

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

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

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 the numerics:

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

(5) After all these preliminaries, time now to get to work, and prove the result. As mentioned above, our formulae are all equivalent, and it is enough to prove just one of them. We will prove in what follows the first formula, namely:

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

(6) 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:

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

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

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

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

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

(9) 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:

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

(10) We are now in position to conclude. According to (6,7,9), 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. □

6d. Basic examples

Basic examples.

6e. Exercises

Exercises:

Exercise 6.5.

Exercise 6.6.

Exercise 6.7.

Exercise 6.8.

Exercise 6.9.

Exercise 6.10.

Exercise 6.11.

Exercise 6.12.

Bonus exercise.

CHAPTER 7

Further cumulants

7a. Noncrossing partitions

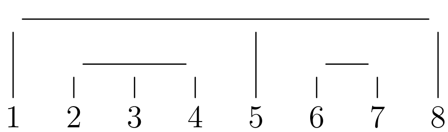
Many interesting things can be said about noncrossing partitions, notably with:

THEOREM 7.1. *We have a bijection as follows,*

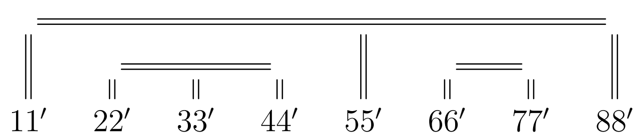
$$NC(k) \simeq NC_2(2k)$$

obtained by fattening the partitions, and by shrinking the pairings.

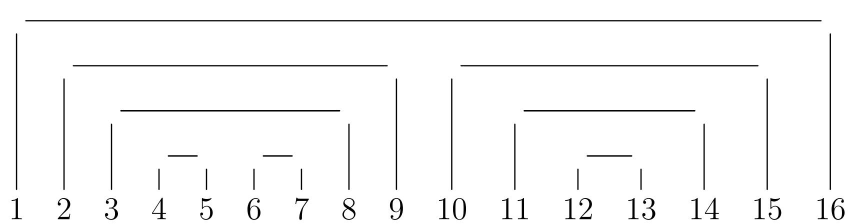
PROOF. This is something self-explanatory, and in order to see how this works, let us discuss an example. Consider a noncrossing partition, say the following one:



Now let us "fatten" this partition, by doubling everything, as follows:



We can see emerging here a noncrossing pairing, and by relabeling the points $1,\ldots,16$, and properly redrawing the picture, what we have is indeed a noncrossing pairing:



As for the reverse operation, that is obviously obtained by "shrinking" our pairing, by collapsing pairs of consecutive neighbors, that is, by identifying $1 = 2$, then $3 = 4$, then $5 = 6$, and so on. Thus, we are led to the conclusion in the statement. □

7b. Free cumulants

In what follows we discuss the free analogues of the classical cumulants, following Speicher [82], and subsequent work. We first have the following definition:

DEFINITION 7.2. *The free cumulants $\kappa_n(a)$ of a variable $a \in A$ are defined by*

$$R_a(\xi) = \sum_{n=1}^{\infty} \kappa_n(a)\xi^{n-1}$$

with the R-transform being defined as usual by the formula

$$G_a\left(R_a(\xi) + \frac{1}{\xi}\right) = \xi$$

where $G_a(\xi) = \int_{\mathbb{R}} \frac{d\mu(t)}{\xi - t}$ with $\mu = \mu_a$ is the corresponding Cauchy transform.

As before with classical cumulants, we have a number of basic examples and illustrations, and a number of basic general results. Let us start with some numerics:

PROPOSITION 7.3. *The free cumulants $\kappa_1, \kappa_2, \kappa_3, \ldots$ appear as a modification of the moments $M_1, M_2, M_3, \ldots$, and uniquely determine μ. We have*

$$\kappa_1 = M_1$$

$$\kappa_2 = -M_1^2 + M_2$$

$$\kappa_3 = 2M_1^3 - 3M_1M_2 + M_3$$

$$\kappa_4 = -5M_1^4 + 10M_1^2M_2 - 2M_2^2 - 4M_1M_3 + M_4$$

$$\vdots$$

in one sense, and in the other sense we have

$$M_1 = \kappa_1$$

$$M_2 = \kappa_1^2 + \kappa_2$$

$$M_3 = \kappa_1^3 + 3\kappa_1\kappa_2 + \kappa_3$$

$$M_4 = \kappa_1^4 + 6\kappa_1^2\kappa_2 + 2\kappa_2^2 + 4\kappa_1\kappa_3 + \kappa_4$$

$$\vdots$$

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

Proof. Here all theoretical assertions regarding moments and cumulants are clear from definitions, and the numerics are clear from definitions too, after some computations based on Definition 7.2. Let us actually present these computations, which are quite instructive, more complicated than the classical ones, and that we will need, later on:

(1) We know that the Cauchy transform is the following function:

$$G(\xi)=\sum_{n=0}^\infty\frac{M_n}{\xi^{n+1}}$$

Consider the inverse of this Cauchy transform G, with respect to composition:

$$G(K(\xi))=K(G(\xi))=\xi$$

According to Definition 7.2, the free cumulants κ_n appear then as follows:

$$K(\xi)=\frac{1}{\xi}+\sum_{n=1}^\infty\kappa_n\xi^{n-1}$$

Thus, we can compute moments in terms of free cumulants, and vice versa, by using either of the inversion formulae $G(K(\xi))=\xi$ and $K(G(\xi))=\xi$.

(2) This was for the theory. In practice now, playing with the original inversion formula from Definition 7.2, namely $G(K(\xi))=\xi$, proves to be something quite complicated, so we will choose to use instead the other inversion formula, namely:

$$K(G(\xi))=\xi$$

Thus, the equation that we want to use is as follows, with $G=G(\xi)$:

$$\frac{1}{G}+\sum_{n=1}^\infty\kappa_nG^{n-1}=\xi$$

(3) With $\xi=z^{-1}$ our equation takes the following form, with $G=G(z^{-1})$:

$$\frac{1}{G}+\sum_{n=1}^\infty\kappa_nG^{n-1}=z^{-1}$$

Now by multiplying by z, our equation takes the following form:

$$\frac{z}{G}+z\sum_{n=1}^\infty\kappa_nG^{n-1}=1$$

Equivalently, our equation is as follows, with $G=G(z^{-1})$ as before:

$$\frac{z}{G}+\sum_{n=1}^\infty\kappa_nz^n\left(\frac{G}{z}\right)^{n-1}=1$$

(4) Observe now that we have the following formula:

$$\frac{G}{z}=\frac{G(z^{-1})}{z}=\frac{\sum_{n=0}^\infty M_nz^{n+1}}{z}=\sum_{n=0}^\infty M_nz^n$$

This suggests introducing the following quantity:

$$F=\sum_{n=1}^\infty M_nz^n$$

Indeed, we have then $G/z=1+F$, and our equation becomes:

$$\frac{1}{1+F}+\sum_{n=1}^\infty\kappa_nz_n(1+F)^{n-1}=1$$

(5) By expanding the fraction on the left, our equation becomes:

$$\sum_{n=0}^\infty(-F)^n+\sum_{n=1}^\infty\kappa_nz_n(1+F)^{n-1}=1$$

Moreover, we can cancel the 1 term on both sides, and our equation becomes:

$$\sum_{n=1}^\infty(-F)^n+\sum_{n=1}^\infty\kappa_nz_n(1+F)^{n-1}=0$$

Alternatively, we can write our equation as follows:

$$\sum_{n=1}^\infty\kappa_nz_n(1+F)^{n-1}=-\sum_{n=1}^\infty(-F)^n$$

(6) Good news, this latter equation is something that we are eventually happy with. By remembering that we have $F=\sum_{n=1}^\infty M_nz^n$, our equation looks as follows:

$$\begin{aligned}&\kappa_1z+\kappa_2z^2(1+M_1z+M_2z^2+\ldots)+\kappa_3z^3(1+M_1z+M_2z^2+\ldots)^2+\ldots\\=\ &(M_1z+M_2z^2+\ldots)-(M_1z+M_2z^2+\ldots)^2+(M_1z+M_2z^2+\ldots)^3-\ldots\end{aligned}$$

(7) This was for the hard part, carefully fine-tuning our equation, as to have it as simple as possible, before getting to numeric work. The rest is routine. Indeed, by looking at the terms of order $1,2,3,4$ we obtain, instantly or almost, the formulae of $\kappa_1,\kappa_2,\kappa_3,\kappa_4$ in the statement. As for the formulae for M_1,M_2,M_3,M_4, these follow from these.

(8) To be more precise, the equations that we get at order $1,2,3,4$ are as follows:

$$\kappa_1=M_1$$

$$\kappa_2=M_2-M_1^2$$

$$\kappa_2M_1+\kappa_3=M_3-2M_1M_2+M_1^3$$

$$\kappa_4+2\kappa_3M_1+\kappa_2M_2=M_4-2M_1M_3-M_2^2+3M_1^2M_2-M_1^4$$

Thus, we are led to the formulae of $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ in the statement, and then to the formulae of M_1, M_2, M_3, M_4 in the statement, as desired. □

Observe the similarity with the formulae for classical cumulants. In fact, a careful comparison here is worth the effort, leading to the following conclusion:

Conclusion 7.4. *The first three classical and free cumulants coincide,*

$$k_1 = \kappa_1 \quad , \quad k_2 = \kappa_2 \quad , \quad k_3 = \kappa_3$$

but the formulae for the fourth classical and free cumulants are different,

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

$$\kappa_4 = -5M_1^4 + 10M_1^2M_2 - 2M_2^2 - 4M_1M_3 + M_4$$

and the same happens at higher order as well.

This is something quite interesting, and we will back later with a conceptual explanation for this, via partitions, the idea being that all this comes from:

$$P(n) = NC(n) \iff n \leq 3$$

But more on this later. At the level of basic general results, we first have:

Theorem 7.5. *The free cumulants have the following properties:*

(1) $\kappa_n(\lambda a) = \lambda^n \kappa_n(a)$.
(2) $\kappa_n(a+b) = \kappa_n(a) + \kappa_n(b)$, *if a, b are free.*

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

(1) We have the following Cauchy transform computation:

$$\begin{aligned} G_{\lambda a}(\xi) &= \int_{\mathbb{R}} \frac{d\mu_{\lambda a}(t)}{\xi - t} \\ &= \int_{\mathbb{R}} \frac{d\mu_a(s)}{\xi - \lambda s} \\ &= \frac{1}{\lambda} \int_{\mathbb{R}} \frac{d\mu_a(s)}{\xi/\lambda - s} \\ &= \frac{1}{\lambda} G_a\left(\frac{\xi}{\lambda}\right) \end{aligned}$$

But this gives the following formula, by using the definition of the R-transform:

$$\begin{aligned} G_{\lambda a}\left(\lambda R_a(\lambda\xi) + \frac{1}{\xi}\right) &= \frac{1}{\lambda} G_a\left(R_a(\lambda\xi) + \frac{1}{\lambda\xi}\right) \\ &= \frac{1}{\lambda} \cdot \lambda\xi \\ &= \xi \end{aligned}$$

Thus we have the formula $R_{\lambda a}(\xi) = \lambda R_a(\lambda\xi)$, which gives (1).

(2) This follows from the standard fact, that we know well from chapter 9, that the R-transform linearizes the free convolution operation. □

Again in analogy with the classical case, at the level of examples, we have:

THEOREM 7.6. *The sequence of free cumulants $\kappa_1, \kappa_2, \kappa_3, \ldots$ is as follows:*
1. *For $\mu = \delta_c$ the free cumulants are $c, 0, 0, \ldots$*
2. *For $\mu = \gamma_t$ the free cumulants are $0, t, 0, 0, \ldots$*
3. *For $\mu = \pi_t$ the free cumulants are $t, t, t, \ldots$*
4. *For $\mu = \beta_t$ the free cumulants are $0, t, 0, t, \ldots$*

Also, for compound free Poisson laws the free cumulants are $k_n(\pi_\nu) = M_n(\nu)$.

PROOF. The proofs are analogous to those from the classical case, as follows:

(1) For $\mu = \delta_c$ we have $G_\mu(\xi) = 1/(\xi - c)$, and so $R_\mu(\xi) = c$, as desired.

(2) For $\mu = \gamma_t$ we have, as computed before, $R_\mu(\xi) = t\xi$, as desired.

(3) For $\mu = \pi_t$ we have, also from before, $R_\mu(\xi) = t/(1-\xi)$, as desired.

(4) For $\mu = \beta_t$ this follows from the formulae from before, but the best is to prove directly the last assertion, which generalizes (3,4). With $\nu = \sum_i c_i \delta_{z_i}$ we have:

$$\begin{aligned}
R_{\pi_\nu}(\xi) &= \sum_i \frac{c_i z_i}{1 - \xi z_i} \\
&= \sum_i c_i z_i \sum_{n \geq 0} (\xi z_i)^n \\
&= \sum_{n \geq 0} \xi^n \sum_i c_i z_i^{n+1} \\
&= \sum_{n \geq 1} \xi^{n-1} \sum_i c_i z_i^n \\
&= \sum_{n \geq 1} \xi^{n-1} M_n(\nu)
\end{aligned}$$

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

7c. Inversion formula

As before in the classical case, we can define now generalized free cumulants, $\kappa_\pi(a)$ with $\pi \in P(k)$, by starting with the numeric free cumulants $\kappa_n(a)$, as follows:

DEFINITION 7.7. *We define free cumulants $\kappa_\pi(a)$, depending on partitions*

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

by starting with $\kappa_n(a)$, 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 free cumulant is the usual one, namely:

$$\kappa_{1_n}(a) = \kappa_n(a)$$

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:

$$\kappa_\pi(a) = \kappa_{b_1}(a) \ldots \kappa_{b_s}(a)$$

With this convention, we have the following result, due to Speicher [82]:

Theorem 7.8. *We have the moment-cumulant formulae*

$$M_n(a) = \sum_{\nu \in NC(n)} \kappa_\nu(a) \quad , \quad \kappa_n(a) = \sum_{\nu \in NC(n)} \mu(\nu, 1_n) M_\nu(a)$$

or, equivalently, we have the moment-cumulant formulae

$$M_\pi(a) = \sum_{\nu \leq \pi} \kappa_\nu(a) \quad , \quad \kappa_\pi(a) = \sum_{\nu \leq \pi} \mu(\nu, \pi) M_\nu(a)$$

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

Proof. As before in the classical case, the 4 formulae in the statement are equivalent, via Möbius inversion. Thus, it is enough to prove one of them, and we will prove the first formula, which in practice is the most useful one. Thus, we must prove that:

$$M_n(a) = \sum_{\nu \in NC(n)} \kappa_\nu(a)$$

(1) In order to prove this formula, let us get back to the construction of the free cumulants, from Definition 7.2. The Cauchy transform of a is the following function:

$$G_a(\xi) = \sum_{n=0}^\infty \frac{M_n(a)}{\xi^{n+1}}$$

Consider the inverse of this Cauchy transform G_a, with respect to composition:

$$G_a(K_a(\xi)) = K_a(G_a(\xi)) = \xi$$

According to Definition 7.2, the free cumulants $\kappa_n(a)$ appear then as follows:

$$K_a(\xi) = \frac{1}{\xi} + \sum_{n=1}^\infty \kappa_n(a)\xi^{n-1}$$

Thus, we can compute moments in terms of free cumulants by using either of the inversion formulae $G_a(K_a(\xi)) = \xi$ and $K_a(G_a(\xi)) = \xi$.

(2) In practice, as explained in the proof of Proposition 7.3, the best is to use the second inversion formula, $K_a(G_a(\xi)) = \xi$, which after some manipulations reads:

$$\begin{aligned} & \kappa_1 z + \kappa_2 z^2(1 + M_1 z + M_2 z^2 + \ldots) + \kappa_3 z^3 (1 + M_1 z + M_2 z^2 + \ldots)^2 + \ldots \\ = \ & (M_1 z + M_2 z^2 + \ldots) - (M_1 z + M_2 z^2 + \ldots)^2 + (M_1 z + M_2 z^2 + \ldots)^3 - \ldots \end{aligned}$$

We have already seen, in the proof of Proposition 7.3, how to exploit this formula at order $n = 1, 2, 3, 4$. The same method works in general, and after some computations, this leads to the formula that we want to establish, namely:

$$M_n(a) = \sum_{\nu \in NC(n)} \kappa_\nu(a)$$

(3) We are therefore led to the conclusions in the statement. All this was of course quite brief, and for details here, we refer for instance to Nica-Speicher [68]. $\square$

Observe that the above result leads among others to a more conceptual explanation for Conclusion 7.4, with the equalities and non-equalities there simply coming from:

$$P(n) = NC(n) \iff n \leq 3$$

Many other things can be said about free cumulants, along these lines.

7d. Further cumulants

Further cumulants.

7e. Exercises

Exercises:

Exercise 7.9.

Exercise 7.10.

Exercise 7.11.

Exercise 7.12.

Exercise 7.13.

Exercise 7.14.

Exercise 7.15.

Exercise 7.16.

Bonus exercise.

CHAPTER 8

Infinite divisibility

8a. Infinite divisibility

Infinite divisibility.

8b. Basic examples

Basic examples.

8c. Further convolutions

We have the following notion of independence, generalizing the usual one:

DEFINITION 8.1. *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 8.2. *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 8.1.

(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 8.1, 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 [87]:

DEFINITION 8.3. *Two subalgebras $A,B\subset C$ are called free when the following condition is satisfied, for any $a_i\in A$ and $b_i\in B$:*

$$tr(a_i)=tr(b_i)=0\implies tr(a_1b_1a_2b_2\ldots)=0$$

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 8.1 and Definition 8.3, 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 8.4. *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 8.1, 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 [87], which clarifies the analogy between independence and freeness:

THEOREM 8.5. *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 8.4.*

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 8.1, 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 8.2 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 8.2, 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 8.4, 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. □

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 [87]:

Theorem 8.6. *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 8.5, 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 8.4, 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 $a + b$. □

The idea now is that with a bit of luck, the basic theory from the classical case, namely the Fourier transform, and the CLT, should have free extensions. Let us start with:

THEOREM 8.7. *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 8.8. *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. $\square$

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 8.7 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 8.9. *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. $\square$

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 8.10. *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 8.9, 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 8.9, which were both clear as well. □

The advantage with this latter model comes from the following result, from [87], which has a very simple formulation, without linear combinations or anything:

Proposition 8.11. *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 8.7 how to model the individual distributions $\mu \in \mathcal{P}(\mathbb{R})$, and by combining this with Proposition 8.11, 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 Voiculescu [87], we have:

THEOREM 8.12. *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 8.11, 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:

$$X=S^*+a_0+a_1S+a_2S^2+\ldots$$

$$Y=T^*+b_0+b_1T+a_2T^2+\ldots$$

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:

$$r=\frac{s+t}{\sqrt2}\quad,\quad u=\frac{s-t}{\sqrt2}$$

The new shifts, associated to these vectors $r,u\in H$, are then given by:

$$R=\frac{S+T}{\sqrt2}\quad,\quad U=\frac{S-T}{\sqrt2}$$

By using now these two new shifts, which are free according to Proposition 8.11, we obtain the following equality of distributions:

$$\begin{aligned}X+Y&=S^*+T^*+\sum_ka_kS^k+b_kT^k\\&=\sqrt2R^*+\sum_ka_k\left(\frac{R+U}{\sqrt2}\right)^k+b_k\left(\frac{R-U}{\sqrt2}\right)^k\\&\sim\sqrt2R^*+\sum_ka_k\left(\frac{R}{\sqrt2}\right)^k+b_k\left(\frac{R}{\sqrt2}\right)^k\\&\sim R^*+\sum_ka_kR^k+b_kR^k\end{aligned}$$

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 8.7 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 [88], is as follows:

THEOREM 8.13. *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 8.12, the operation $\mu \to f$ from Theorem 8.7 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 8.7:

$$X = S^* + f(S)$$

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:

$$Sq_z = z^{-1}(q_z - \delta_0) \quad , \quad S^* q_z = z q_z$$

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:

$$(z^{-1} + f(z) - X^*)^{-1}\delta_0 = z q_z$$

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 [87].

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 [87]:

THEOREM 8.14. *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 8.13, 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 8.13 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 8.8, 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 [88], we have:

THEOREM 8.15 (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 8.13, 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:

$$\frac{1}{2\pi t} \int_{-2\sqrt{t}}^{2\sqrt{t}} \sqrt{4t - x^2} x^{2k} dx = t^k C_k$$

On the other hand, we know from before that we have the following formula:

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

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

$$\begin{aligned}G_{\gamma_t}(y) &= y^{-1}\sum_{k=0}^{\infty} t^k C_k y^{-2k}\\ &= y^{-1}\cdot\frac{1-\sqrt{1-4ty^{-2}}}{2ty^{-2}}\\ &= \frac{y}{2t}\left(1-\sqrt{1-4ty^{-2}}\right)\\ &= \frac{y}{2t}-\frac{1}{2t}\sqrt{y^2-4t}\end{aligned}$$

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 8.13 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. □

8d. Transforms, bijections

With the cumulant theory in hand, developed before, we can now formulate the following simple definition, making the connection between classical and free:

Definition 8.16. *We say that a real probability measure*

$$m \in \mathcal{P}(\mathbb{R})$$

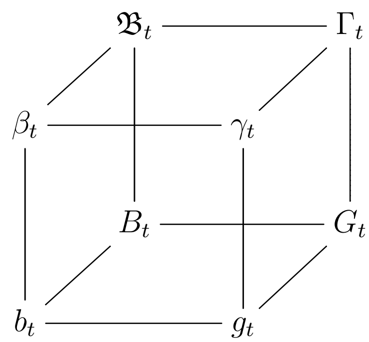
is the classical version of another measure, called its free version, or liberation

$$\mu \in \mathcal{P}(\mathbb{R})$$

when the classical cumulants of m coincide with the free cumulants of μ.

As a first observation, this definition fits with all the classical and free probability theory developed in the above, in this whole book so far, and notably with the measures from the standard cube, and to start with, we have the following result:

THEOREM 8.17. *In the standard cube of basic probability measures,*



the upper measures appear as the free versions of the lower measures.

PROOF. This follows indeed from our various cumulant formulae found before. □

In order to reach now to a more advanced theory, depending this time on a parameter $t > 0$, which is something essential, and whose importance will become clear later on, let us formulate, following Bercovici-Pata [15], and the subsequent work in [68]:

DEFINITION 8.18. *A convolution semigroup of measures*

$$\{m_t\}_{t>0} \quad : \quad m_s * m_t = m_{s+t}$$

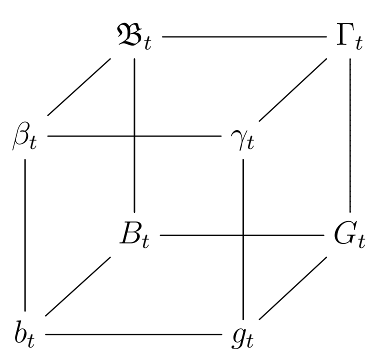
is in Bercovici-Pata bijection with a free convolution semigroup of measures

$$\{\mu_t\}_{t>0} \quad : \quad \mu_s \boxplus \mu_t = \mu_{s+t}$$

when the classical cumulants of m_t coincide with the free cumulants of μ_t.

As before, this fits with all the theory developed so far in this book, and notably with the measures from the standard cube, and we have the following result:

THEOREM 8.19. *In the standard cube of basic semigroups of measures,*



the upper semigroups are in Bercovici-Pata bijection with the lower semigroups.

PROOF. This is a technical improvement of Theorem 8.17, based on the fact that the upper measures in the above diagram form indeed free convolution semigroups, and that the lower measures form indeed classical convolution semigroups, which itself is something that we know well, from the various semigroup results established in above. □

Back to the examples now, there are many other, and we will be back to this. But, before anything, let us formulate the following surprising result, from [**14**]:

THEOREM 8.20. *The normal law g_1 is freely infinitely divisible.*

PROOF. This is something tricky, involving all sorts of not very intuitive computations, and for full details here, we refer here to the original paper [**14**]. □

The above result shows that the normal law g_1 should have a "classical analogue" in the sense of the Bercovici-Pata bijection. And isn't that puzzling. The problem, however, is that this latter law is difficult to compute, and interpret. See [**14**].

Still in relation with the Bercovici-Pata bijection, let us also mention that there are many interesting analytic aspects, coming from the combinatorics of the infinitely divisible laws, classical or free. For this, and other analytic aspects, we refer to [**15**].

8e. Exercises

Exercises:

EXERCISE 8.21.

EXERCISE 8.22.

EXERCISE 8.23.

EXERCISE 8.24.

EXERCISE 8.25.

EXERCISE 8.26.

EXERCISE 8.27.

EXERCISE 8.28.

Bonus exercise.

Part III

Orthogonal polynomials

CHAPTER 9

Hilbert spaces

9a. Scalar products

Quantum physics tells us to look as well at the infinite dimensional complex spaces, such as the space of wave functions $\psi:\mathbb{R}^3\to\mathbb{C}$ of the electron. In order to do some mathematics on these spaces, we will need scalar products. So, let us start with:

DEFINITION 9.1. *A scalar product on a complex vector space H is a binary operation $H\times H\to\mathbb{C}$, denoted $(x,y)\to<x,y>$, satisfying the following conditions:*

1. *$<x,y>$ is linear in x, and antilinear in y.*
2. *$\overline{<x,y>}=<y,x>$, for any x,y.*
3. *$<x,x>>0$, for any $x\neq 0$.*

As before in the previous chapters, we use here mathematicians' convention for scalar products, that is, $<,>$ linear at left, as opposed to physicists' convention, $<,>$ linear at right. The reasons for this are quite subtle, coming from the fact that, while basic quantum mechanics looks better with $<,>$ linear at right, advanced quantum mechanics looks better with $<,>$ linear at left. Or at least that's what my cats say.

As a basic example for Definition 9.1, we have the finite dimensional vector space $H=\mathbb{C}^N$, with its usual scalar product, namely:

$$<x,y>=\sum_i x_i\bar{y}_i$$

We will see later in this chapter that in finite dimensions, this is in fact the only example, the point being that algebrically we must have $H\simeq\mathbb{C}^N$, for some $N\in\mathbb{N}$, and then we can always change the basis, as to make it orthogonal with respect to $<,>$, which in practice makes $<,>$ to be given by the above formula. More on this in a moment.

In infinite dimensions now, there are many interesting examples of spaces naturally coming with scalar products, and notably various spaces of L^2 functions, which appear for instance in various problems coming from physics. We will discuss them later.

Summarizing, what we have in Definition 9.1 is a potentially useful generalization of the usual scalar product $<,>$ on the simplest complex vector space, $\mathbb{C}^N$. In order to study now the scalar products, let us formulate the following definition:

Definition 9.2. *The norm of a vector $x \in H$ is the following quantity:*

$$||x|| = \sqrt{<x, x>}$$

We also call this number length of x, or distance from x to the origin.

The terminology comes from what happens in $\mathbb{C}^N$, where the length of the vector, as defined above, coincides with the usual length, given by:

$$||x|| = \sqrt{\sum_i |x_i|^2}$$

In analogy with what happens in finite dimensions, we have two important results regarding the norms. First we have the Cauchy-Schwarz inequality, as follows:

Theorem 9.3. *We have the Cauchy-Schwarz inequality,*

$$|<x, y>| \leq ||x|| \cdot ||y||$$

and the equality case holds precisely when x, y are proportional.

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

(1) Consider, and we will understand why in a moment, the following quantity, depending on a real variable $t \in \mathbb{R}$, and on a variable on the unit circle, $w \in \mathbb{T}$:

$$f(t) = ||twx + y||^2$$

By developing f, we see that this is a degree 2 polynomial in t:

$$\begin{aligned} f(t) &= <twx + y, twx + y> \\ &= t^2 <x, x> + tw <x, y> + t\bar{w} <y, x> + <y, y> \\ &= t^2||x||^2 + 2tRe(w <x, y>) + ||y||^2 \end{aligned}$$

(2) Since f is obviously positive, its discriminant must be negative:

$$4Re(w <x, y>)^2 - 4||x||^2 \cdot ||y||^2 \leq 0$$

But this is equivalent to the following condition:

$$|Re(w <x, y>)| \leq ||x|| \cdot ||y||$$

Now the point is that we can arrange for the number $w \in \mathbb{T}$ to be such that the quantity $w <x, y>$ is real. Thus, we obtain the Cauchy-Schwarz inequality:

$$|<x, y>| \leq ||x|| \cdot ||y||$$

(3) Finally, the study of the equality case is straightforward, by using the fact that the discriminant of f vanishes precisely when we have a root. Indeed, this shows that having equality in Cauchy-Schwarz is the same as asking for the following to happen:

$$f(t) = 0$$

But this latter condition is very easy to process, as follows:

$$\begin{aligned} f(t)=0 &\iff ||twx+y||^2=0 \\ &\iff ||twx+y||=0 \\ &\iff twx+y=0 \\ &\iff x\sim y \end{aligned}$$

Thus we are led to the conclusion in the statement, namely that in order to have equality in the Cauchy-Schwarz inequality, the vectors x,y must be proportional. □

As a second main result now, we have the Minkowski inequality:

THEOREM 9.4. *We have the Minkowski inequality*

$$||x+y||\leq||x||+||y||$$

and the equality case holds precisely when x,y are proportional.

PROOF. This follows indeed from the Cauchy-Schwarz inequality, as follows:

$$\begin{aligned} & ||x+y||\leq||x||+||y|| \\ \iff & ||x+y||^2\leq(||x||+||y||)^2 \\ \iff & ||x||^2+||y||^2+2Re<x,y>\leq||x||^2+||y||^2+2||x||\cdot||y|| \\ \iff & Re<x,y>\leq||x||\cdot||y|| \end{aligned}$$

As for the equality case, this is clear from Cauchy-Schwarz as well. □

In abstract terms, the Minkowski inequality tells us that the following happens:

PROPOSITION 9.5. *The following function is a norm on H,*

$$||x||=\sqrt{<x,x>}$$

in the usual sense, that of the abstract normed spaces.

PROOF. Recall indeed that a normed space is an abstract vector space X with a function $||.||:X\to[0,\infty)$, called norm, subject to the following conditions:

– $||x||>0$ for $x\neq0$.

– $||\lambda x||=|\lambda|\cdot||x||$.

– $||x+y||\leq||x||+||y||$.

In our case, the first two axioms are trivially satisfied, and the third axiom, called triangle inequality, is the Minkowski inequality. Thus, the result holds indeed. □

Alternatively, and perhaps more illustrating, we have the following result:

Theorem 9.6. *The following function is a distance on H,*

$$d(x,y) = ||x-y||$$

in the usual sense, that of the abstract metric spaces.

Proof. This follows indeed from the Minkowski inequality, which corresponds to the triangle inequality, the other two axioms being trivially satisfied. To be more precise:

(1) Let us first recall that a metric space is an abstract space X with a function $d : X \times X \to [0,\infty)$, called distance, which is subject to the following conditions:

– $d(x,y) > 0$ for $x \neq y$, and $d(x,x) = 0$.

– $d(x,y) = (y,x)$.

– $d(x,y) \leq d(x,z) + d(y,z)$.

(2) Now let us try to check these axioms for $d(x,y) = ||x-y||$. The first axiom is clear, and so is the second axiom, so we are led with checking the third axiom, the triangle inequality one, which in practice means to establish the following inequality:

$$||x-y|| \leq ||x-z|| + ||y-z||$$

(3) But this is clear, because with $x' = x - z$ and $y' = z - y$, our estimate reads:

$$||x'+y'|| \leq ||x'|| + ||y'||$$

And this being the Minkowski inequality, done with the axiom check, as desired. □

The above result is quite important, because it shows that we can normally do geometry and analysis in our present setting, a bit as in the finite dimensional case. In order to do such abstract geometry, we will often need the following key result:

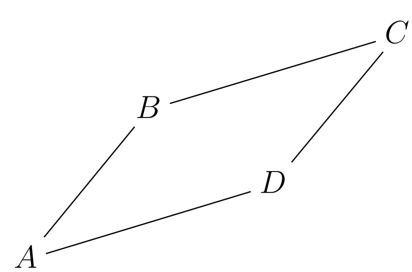
Theorem 9.7. *The distances on H are subject to the identity*

$$||x+y||^2 + ||x-y||^2 = 2(||x||^2 + ||y||^2)$$

called parallelogram identity.

Proof. This is something quite fundamental, the idea being as follows:

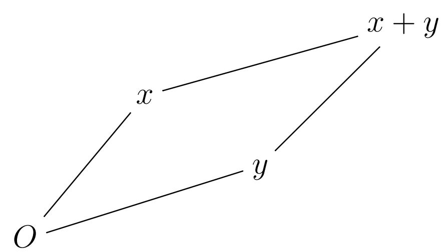
(1) To start with, there is a relation here with a basic result from plane geometry, that you might know or not. Consider indeed a parallelogram in the plane:



The above-mentioned formula from plane geometry is then as follows:

$$AC^2 + BD^2 = AB^2 + BC^2 + CD^2 + DA^2$$

But this is more or less the formula in the statement. Indeed, if we choose the origin to be A, and relabel x, y the points B, D, our parallelogram becomes:



Now with this done, observe we have the following two formulae:

$$AC^2 + BD^2 = ||x + y||^2 + ||x - y||^2$$

$$AB^2 + BC^2 + CD^2 + DA^2 = 2(||x||^2 + ||y||^2)$$

Thus, the plane geometry formula is the same as the formula in the statement.

(2) In practice now, all this remains a mere remark, because our spaces are complex instead of real, have arbitrary dimension instead of 2, and also because we have not said in the above how the proof of the elementary geometry formula goes. So, better forget about all this, and try to prove the formula in the statement, from scratch.

(3) But here, things are in fact quite straightforward, because we have:

$$\begin{aligned} & ||x + y||^2 + ||x - y||^2 \\ = \ & < x + y, x + y > + < x - y, x - y > \\ = \ & ||x||^2 + ||y||^2 + < x, y > + < y, x > + ||x||^2 + ||y||^2 - < x, y > - < y, x > \\ = \ & 2(||x||^2 + ||y||^2) \end{aligned}$$

Thus, we have proved our formula, and as a bonus, we have understood as well how the above-mentioned plane geometry formula works. Indeed, our computation above obviously works as well for the real scalar products, and this gives the result. □

As a second result now, which is something fundamental too, everything can be formally recovered in terms of distances, as follows:

THEOREM 9.8. *The scalar products can be recovered from distances, via the formula*

$$4 < x, y > = ||x + y||^2 - ||x - y||^2 + i||x + iy||^2 - i||x - iy||^2$$

called complex polarization identity.

Proof. This is something that we have already met in finite dimensions. In arbitrary dimensions the proof is similar, as follows:

$$\begin{aligned} & \|x+y\|^2-\|x-y\|^2+i\|x+iy\|^2-i\|x-iy\|^2 \\ = & \|x\|^2+\|y\|^2-\|x\|^2-\|y\|^2+i\|x\|^2+i\|y\|^2-i\|x\|^2-i\|y\|^2 \\ & +2Re(<x,y>)+2Re(<x,y>)+2iIm(<x,y>)+2iIm(<x,y>) \\ = & 4<x,y> \end{aligned}$$

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

Summarizing, all the basic formulae involving scalar products and norms, that we know well from linear algebra, do hold in our abstract vector space setting. As a word of warning here, however, not ever to be forgotten, we have:

Warning 9.9. *Unlike other things, the basic formula for real scalar products,*

$$<x,y>=||x||\cdot||y||\cdot\cos\alpha$$

does not hold, in our complex vector space setting.

To be more precise here, in what regards the above formula, you certainly know from plane geometry that the formula holds indeed for $\mathbb{R}^2$, and you might know too, from space geometry, that the formula holds as well for $\mathbb{R}^3$. The same goes for any $\mathbb{R}^N$, with similar proof, and going a bit abstract, for any real vector space coming with a scalar product, and this because by Cauchy-Schwarz we have $|<x,y>|\leq||x||\cdot||y||$, so the above formula can stand as a definition for the angle $\alpha\in[0,\pi)$ between our vectors x,y.

In the complex space setting, however, this does not work. Indeed, we still have the Cauchy-Schwarz inequality, telling us that $|<x,y>|\leq||x||\cdot||y||$, but the scalar product being now a complex number, $<x,y>\in\mathbb{C}$, so is its quotient by $||x||\cdot||y||\in\mathbb{R}$, so we cannot come with an angle $\alpha\in[0,\pi)$ whose cosine equals this quotient.

9b. Hilbert spaces

In order to do analysis on our spaces, we need the Cauchy sequences that we construct to converge. This is something which is automatic in finite dimensions, but in arbitrary dimensions, this can fail. It is convenient here to formulate a detailed new definition, as follows, which will be the starting point for our various considerations to follow:

Definition 9.10. *A Hilbert space is a complex vector space H given with a scalar product $<x,y>$, satisfying the following conditions:*

1. *$<x,y>$ is linear in x, and antilinear in y.*
2. *$\overline{<x,y>}=<y,x>$, for any x,y.*
3. *$<x,x>>0$, for any $x\neq0$.*
4. *H is complete with respect to the norm $||x||=\sqrt{<x,x>}$.*

In other words, what we did here is to take Definition 9.1, and add the condition that H must be complete with respect to the norm $||x|| = \sqrt{< x, x >}$, that we know indeed to be a norm, according to the Minkowski inequality proved above. As a basic example, as before, we have the space $H = \mathbb{C}^N$, with its usual scalar product:

Proposition 9.11. *The space $H = \mathbb{C}^N$, with its usual scalar product, namely*

$$< x, y >= \sum_i x_i \bar{y}_i$$

is a Hilbert space, which is finite dimensional.

Proof. Here the fact that $< x, y >= \sum_i x_i \bar{y}_i$ is indeed a scalar product on $\mathbb{C}^N$ is something that we know well, and the completness condition is automatic. □

We will see later in this chapter, when talking about orthogonal bases for our spaces, that any finite dimensional Hilbert space H appears as above, $H \simeq \mathbb{C}^N$. Thus, at least we know one thing, done with finite dimensions, no bad surprises here.

More generally now, we have the following construction of Hilbert spaces:

Proposition 9.12. *The sequences of numbers (x_i) which are square-summable,*

$$\sum_i |x_i|^2 < \infty$$

form a Hilbert space $l^2(\mathbb{N})$, with the following scalar product:

$$< x, y >= \sum_i x_i \bar{y}_i$$

In fact, given any index set I, we can construct a Hilbert space $l^2(I)$, in this way.

Proof. There are several things to be proved, as follows:

(1) Our first claim is that $l^2(\mathbb{N})$ is a vector space. For this purpose, we must prove that $x, y \in l^2(\mathbb{N})$ implies $x + y \in l^2(\mathbb{N})$. But this leads us into proving $||x + y|| \leq ||x|| + ||y||$, where $||x|| = \sqrt{< x, x >}$. Now since we know this inequality to hold on each subspace $\mathbb{C}^N \subset l^2(\mathbb{N})$ obtained by truncating, this inequality holds everywhere, as desired.

(2) Our second claim is that $<, >$ is well-defined on $l^2(\mathbb{N})$. But this follows from the Cauchy-Schwarz inequality, $| < x, y > | \leq ||x|| \cdot ||y||$, which can be established by truncating, a bit like we established the Minkowski inequality in (1) above.

(3) It is also clear that $<, >$ is a scalar product on $l^2(\mathbb{N})$, so it remains to prove that $l^2(\mathbb{N})$ is complete with respect to $||x|| = \sqrt{< x, x >}$. But this is clear, because if we pick a Cauchy sequence $\{x^n\}_{n \in \mathbb{N}} \subset l^2(\mathbb{N})$, then each numeric sequence $\{x_i^n\}_{i \in \mathbb{N}} \subset \mathbb{C}$ is Cauchy, and by setting $x_i = \lim_{n \to \infty} x_i^n$, we have $x^n \to x$ inside $l^2(\mathbb{N})$, as desired.

(4) Finally, the same arguments extend to the case of an arbitrary index set I, leading to a Hilbert space $l^2(I)$, and with the remark here that there is absolutely no problem of taking about quantities of type $||x||^2 = \sum_{i\in I} |x_i|^2 \in [0,\infty]$, even if the index set I is uncountable, because we are summing positive numbers. □

Even more generally, we have the following construction of Hilbert spaces:

THEOREM 9.13. *Given a measured space X, the functions $f : X \to \mathbb{C}$, taken up to equality almost everywhere, which are square-summable,*

$$\int_X |f(x)|^2 dx < \infty$$

form a Hilbert space $L^2(X)$, with the following scalar product:

$$< f, g >= \int_X f(x)\overline{g(x)}dx$$

In the case $X = I$, with the counting measure, we obtain in this way the space $l^2(I)$.

PROOF. This is a straightforward generalization of Proposition 9.12, with the arguments from the proof of Proposition 9.12 carrying over in our case, as follows:

(1) The first part, regarding Cauchy-Schwarz and Minkowski, extends without problems, by using this time approximation by step functions.

(2) Regarding the fact that $<,>$ is indeed a scalar product on $L^2(X)$, there is a subtlety here, because if we want $< f, f >> 0$ for $f \neq 0$, we must declare that $f = 0$ when $f = 0$ almost everywhere, and so that $f = g$ when $f = g$ almost everywhere.

(3) Regarding the fact that $L^2(X)$ is complete with respect to $||f|| = \sqrt{< f, f >}$, this is again basic measure theory, by picking a Cauchy sequence $\{f_n\}_{n\in\mathbb{N}} \subset L^2(X)$, and then constructing a pointwise, and hence L^2 limit, $f_n \to f$, almost everywhere.

(4) Finally, the last assertion is clear, because the integration with respect to the counting measure is by definition a sum, and so $L^2(I) = l^2(I)$ in this case. □

As a conclusion to what we did so far, the Hilbert spaces are now axiomatized, and the main examples discussed. In order to do now some geometry on our spaces, in analogy with what we know from finite dimensions, let us start with the following definition:

DEFINITION 9.14. *Let H be a Hilbert space.*

1. *We call two vectors orthogonal, $x \perp y$, when $< x, y >= 0$.*
2. *Given a subset $S \subset H$, we set $S^\perp = \left\{x \in H \middle| x \perp y, \forall y \in S\right\}$.*

Here the first notion is something very familiar and intuitive, with the comment however that in the present complex space setting, orthogonality does not exactly mean that "we have a right angle between our vectors x, y", as explained in Warning 9.9.

In what regards (2), this is something very familiar too, and as an observation here, the subset $S^\perp \subset H$ constructed there is a closed linear space. More on this later.

Getting now to what can be done with orthogonality, we have here:

THEOREM 9.15. *Let H be a Hilbert space, and $E \subset H$ be a closed subspace.*

(1) *Given $x \in H$, we can find a unique $y \in E$, minimizing $||x - y||$.*
(2) *With x, y as above, we have $x = y + z$, for a certain $z \in E^\perp$.*
(3) *Thus, we have a direct sum decomposition $H = E \oplus E^\perp$.*
(4) *In terms of $H = E \oplus E^\perp$, the projection $x \to y$ is given by $P(x, y) = x$.*

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

(1) Given $x \in H$ and two vectors $v, w \in E$, we have the following estimate:

$$\begin{aligned} ||x-v||^2 + ||x-w||^2 &= 2\left(\left|\left|x - \frac{v+w}{2}\right|\right|^2 + \left|\left|\frac{v-w}{2}\right|\right|^2\right) \\ &\geq 2d(x,E)^2 + \frac{||v-w||^2}{2} \end{aligned}$$

But this shows that any sequence in E realizing the inf in the definition of $d(x, E)$ is Cauchy, so it converges to a vector y. Since E is closed we have $y \in E$, so y realizes the inf. Moreover, again from the above inequality, such a y realizing the inf is unique.

(2) In order to prove $x - y \in E^\perp$, let $v \in E$ and choose $w \in \mathbb{T}$ such that $w < x - y, v >$ is a real number. For any $t \in \mathbb{R}$ we have the following equality:

$$||x - y + twv|| = ||x - y||^2 + 2tw < x - y, v > + t^2||v||^2$$

By construction of the vector y we know that this function has a minimum at $t = 0$. But this function is a degree 2 polynomial, so the middle term must vanish:

$$2w < x - y, v >= 0$$

Now since this must hold for any $v \in E$, we must have $x - y \in E^\perp$, as desired.

(3) This is consequence of what we found in (1,2).

(4) This is also a consequence of what we found in (1,2). □

Many things can be said, as a continuation of the above, as for instance with:

PROPOSITION 9.16. *For a closed subspace $E \subset H$, we have:*

$$E^{\perp\perp} = E$$

More generally, for an arbitrary linear subspace $E \subset H$, we have

$$E^{\perp\perp} = \bar{E}$$

and with the closing operation being needed, in infinite dimensions.

Proof. All this comes indeed as an elementary application of our orthogonal projection technology from Theorem 9.15, and we will leave the details here as an exercise. □

Moving forward now, let us discuss some abstract aspects of the Hilbert spaces. You might know a bit, or not, about the Banach spaces, which are something more general than the Hilbert spaces. In view of this, our goal now will be to see what the general Banach space theory has to say, in the particular case of the Hilbert spaces.

And here, things are very simple, because we have, as a main result:

Theorem 9.17. *Given a Hilbert space H and a closed subspace $E\subset H$, any linear form $f:E\to\mathbb C$ can be extended into a linear form*

$$\tilde{f}:H\to\mathbb C$$

having the same norm, and this by using $H=E\oplus E^\perp$, and setting $\tilde{f}=0$ on $E^\perp$.

Proof. This is indeed something self-explanatory. Observe that what we have here is the Hahn-Banach theorem, for the Hilbert spaces, coming with a trivial proof. □

Still talking abstract functional analysis, the few other basic Banach space results trivialize in the case of Hilbert spaces, as shown by the following result:

Theorem 9.18. *Let H be a Hilbert space.*

(1) *Any linear form $f:H\to\mathbb C$ must be of type $f(y)=<z,y>$, with $z\in H$.*
(2) *Thus, we have a Banach space isomorphism $H^*\simeq\bar{H}$.*
(3) *In particular, H is reflexive as Banach space, $H^{**}=H$.*

Proof. This is something that you might already know from Banach space theory, but we have an elementary proof for this, as follows:

(1) Consider a linear form $f:H\to\mathbb C$. Choose $v\in H$ such that $f(v)\neq 0$. By linearity we may assume $f(v)=1$. Then each $z\in H$ decomposes in the following way:

$$z=(z-f(z)v)+f(z)v$$

This shows that we have a direct sum decomposition of H, as follows:

$$H=\ker(f)\oplus\mathbb Cv$$

Now pick $z\in\ker(f)^\perp$ and consider the kernel of the linear form $f_z(y)=<z,y>$:

$$Ker(f_z)=\{y\in H|<z,y>=0\}\supset Ker(f)$$

The linear forms f_z and f are then given by the following formulae:

$$f_z(a+\lambda v)=\lambda f_z(v)\quad,\quad f(a+\lambda v)=\lambda$$

It follows that we have $f=\mu f_z$, with $\mu=f_z(v)^{-1}$, and so that we have, as desired:

$$f=f_{\overline{\mu}z}$$

(2) This is just an abstract reformulation of what we found in (1).

(3) This follows from (2), because we have $H^{**} = \bar{H}^* = \bar{\bar{H}} = H$. □

As a conclusion to all this, which is really pleasant, the various general Banach space results are all clear in the Hilbert space setting. However, do not worry, the Hilbert spaces have their own amount of mystery, that we will explore in what follows.

9c. Gram-Schmidt

At a more advanced level now, we can talk about orthonormal bases, and the related notion of dimension of a Hilbert space. However, this is something quite tricky, in the present infinite dimensional setting, that will take us some time to understand.

Let us start with the following result, that you surely know from linear algebras:

THEOREM 9.19. *Any system of linearly independent vectors $\{f_1,\ldots,f_n\}$ can be turned into an orthogonal system $\{e_1,\ldots,e_n\}$ by using the Gram-Schmidt procedure,*

$$e_1 = f_1$$
$$e_2 = f_2 + \alpha_1 f_1$$
$$e_3 = f_3 + \beta_1 f_1 + \beta_2 f_2$$
$$e_4 = f_4 + \gamma_1 f_1 + \gamma_2 f_2 + \gamma_3 f_3$$
$$\vdots$$

with the needed scalars $\alpha_i,\beta_i,\gamma_i,\ldots$ being uniquely determined.

PROOF. Many things can be said here, depending on how sharp you want to be, with the essentials of what is to be known being as follows:

(1) Let us first study the case $n = 2$. With $e_1 = f_1$ and $e_2 = f_2 + \alpha_1 f_1$ as in the statement, the needed orthogonality condition can be processed as follows:

$$\begin{aligned}
e_1 \perp e_2 &\iff\ < f_1, f_2 + \alpha_1 f_1 >= 0\\
&\iff \alpha_1 < f_1, f_1 >= - < f_1, f_2 >\\
&\iff \alpha_1 = -\frac{< f_1, f_2 >}{< f_1, f_1 >}
\end{aligned}$$

Thus, we get our result, and with the remark that, alternatively, we can set:

$$e_2 = f_2 - Proj_{e_1}(f_2)$$

Indeed, with the above formula of α_1 in hand, the vector $e_2 = f_2 + \alpha_1 f_1$ that we get is precisely this one. Or, we can simply argue that this latter vector e_2 does the job, and with some basic linear algebra telling us that this vector e_2 is indeed unique.

(2) At $n=3$ now, with e_1,e_2 already constructed, and with $e_3=f_3+\beta_1f_1+\beta_2f_2$ as in the statement, the first orthogonality condition can be processed as follows:

$$\begin{aligned} e_1\perp e_3 &\iff <f_1,f_3+\beta_1f_1+\beta_2f_2>=0\\ &\iff \beta_1<f_1,f_1>+\beta_2<f_1,f_2>=-<f_1,f_3> \end{aligned}$$

As for the second orthogonality condition, this can be now processed as follows:

$$\begin{aligned} e_2\perp e_3 &\iff <f_2,f_3+\beta_1f_1+\beta_2f_2>=0\\ &\iff \beta_1<f_2,f_1>+\beta_2<f_2,f_2>=-<f_2,f_3> \end{aligned}$$

Thus, we are led to the following system, for the parameters β_1,β_2:

$$\beta_1<f_1,f_1>+\beta_2<f_1,f_2>=-<f_1,f_3>$$

$$\beta_1<f_2,f_1>+\beta_2<f_2,f_2>=-<f_2,f_3>$$

Now let us compute the determinant of this system. This is given by:

$$\begin{aligned} D &= \begin{vmatrix} <f_1,f_1> & <f_1,f_2>\\ <f_2,f_1> & <f_2,f_2> \end{vmatrix}\\ &= <f_1,f_1><f_2,f_2>-<f_1,f_2><f_2,f_1>\\ &= \|f_1\|^2\|f_2\|^2-|<f_1,f_2>|^2 \end{aligned}$$

But this is exactly the quantity from the Cauchy-Schwarz inequality, so we have $D\geq 0$, with equality when f_1,f_2 are proportional. Now since f_1,f_2 were assumed to be linearly independent, we conclude that we have $D>0$, so our system has indeed solutions.

(3) Alternatively, we can say at $n=3$ that with the vectors e_1,e_2 being already constructed, we can construct the vector e_3 as follows, obviously doing the orthogonality job, and with its uniqueness coming from some standard linear algebra:

$$e_3=f_3-Proj_{e_1}(f_3)-Proj_{e_2}(f_3)$$

(4) Summarizing, we have two possible proofs for our result. Getting now to the general case, as a first proof, which is perhaps the most straightforward, we can set:

$$e_1=f_1$$

$$e_2=f_2-Proj_{e_1}(f_2)$$

$$e_3=f_3-Proj_{e_1}(f_3)-Proj_{e_2}(f_3)$$

$$e_4=f_4-Proj_{e_1}(f_4)-Proj_{e_2}(f_4)-Proj_{e_3}(f_4)$$

$$\vdots$$

Indeed, these vectors do indeed the needed orthogonality job, and their uniqueness is clear too, via some basic linear algebra, that we will leave here as an exercise.

(5) Alternatively, by doing some explicit computations, as in (1) and (2), we must prove that a certain determinant is nonzero. To be more precise, at step $k+1$ of the orthogonalization algorithm, the system to be solved is as follows:

$$x_1 < f_1, f_1 > + x_2 < f_1, f_2 > + \ldots + x_k < f_1, f_k > = - < f_1, f_{k+1} >$$

$$x_1 < f_2, f_1 > + x_2 < f_2, f_2 > + \ldots + x_k < f_2, f_k > = - < f_2, f_{k+1} >$$

$$\vdots$$

$$x_1 < f_k, f_1 > + x_2 < f_k, f_2 > + \ldots + x_k < f_k, f_k > = - < f_k, f_{k+1} >$$

Thus, the determinant to be studied, in order to prove that our system has indeed solutions, is the Gram determinant of $f_1, \ldots, f_k$, given by the following formula:

$$D_k = \begin{vmatrix} < f_1, f_1 > & < f_1, f_2 > & \ldots & < f_1, f_k > \\ < f_2, f_1 > & < f_2, f_2 > & \ldots & < f_2, f_k > \\ & \vdots & \vdots & \\ < f_k, f_1 > & < f_k, f_2 > & \ldots & < f_k, f_k > \end{vmatrix}$$

(6) Now in relation with this latter question, we have already seen in (2) that we have $D_2 > 0$, but with this being something quite complicated, coming from Cauchy-Schwarz. So, not very good news, but fortunately, linear algebra comes to the rescue. Consider the square matrix formed by our vectors $f_1, \ldots, f_k$, arranged horizontally, as follows:

$$F = \begin{pmatrix} (f_1)_1 & \ldots & (f_1)_k \\ & \vdots & \\ (f_k)_1 & \ldots & (f_k)_k \end{pmatrix}$$

We have then the following computation, for any two indices i, j:

$$\begin{aligned} (FF^*)_{ij} &= \sum_l F_{il}(F^*)_{lj} \\ &= \sum_l F_{il}\bar{F}_{jl} \\ &= \sum_l (f_i)_l \overline{(f_j)_l} \\ &= < f_i, f_j > \end{aligned}$$

We conclude that at the matrix level, we have the following formula:

$$FF^* = \begin{pmatrix} < f_1, f_1 > & < f_1, f_2 > & \ldots & < f_1, f_k > \\ < f_2, f_1 > & < f_2, f_2 > & \ldots & < f_2, f_k > \\ & \vdots & \vdots & \\ < f_k, f_1 > & < f_k, f_2 > & \ldots & < f_k, f_k > \end{pmatrix}$$

Thus, at the level of the corresponding determinants we obtain, as desired:

$$D_k = \det(FF^*) = |\det F|^2 > 0$$

(7) Finally, and getting back now to the system, we can work out some explicit formulae for e_i, alternative to those in (4), based on this. To be more precise, we have:

$$e_k = \frac{1}{D_{k-1}} \begin{vmatrix} < f_1, f_1 > & < f_1, f_2 > & \ldots & < f_1, f_k > \\ < f_2, f_1 > & < f_2, f_2 > & \ldots & < f_2, f_k > \\ & \vdots & \vdots & \\ < f_{k-1}, f_1 > & < f_{k-1}, f_2 > & \ldots & < f_{k-1}, f_k > \\ f_1 & f_2 & \ldots & f_k \end{vmatrix}$$

And we will leave some illustrations here as an instructive exercise, and please do better than my students, who usually stop after 2-3 steps. □

9d. Bases, separability

Getting back now to our Hilbert space questions, we have the following result:

THEOREM 9.20. *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. □

So long for abstract Hilbert space questions, and orthonormal bases, and many other things can be said here. In practice now, and getting to the essentials, according to Theorem 9.20, there is only one separable Hilbert space, up to isomorphism.

In order to further comment on this, let us recall the following result:

THEOREM 9.21 (Weierstrass). *Any continuous function on a closed interval*

$$f:[a,b]\to\mathbb{R}$$

can be uniformly approximated by polynomials.

PROOF. This is indeed something very classical, with a well-known, constructive proof, being by using an approximation by suitable Bernstein polynomials, namely:

$$f_n(x)=\sum_{k=0}^n f\left(\frac{k}{n}\right)b_{kn}(x)$$

To be more precise, we assume here that $[a,b]=[0,1]$, and we set:

$$b_{kn}(x)=\binom{n}{k}x^k(1-x)^{n-k}$$

As for the proof of this, this is something well-known, which goes as follows:

(1) Consider indeed the basic Bernstein polynomials b_{kn}, as constructed above. These remind the binomial laws, so it is with some probability that we will start. We have the following formulae, which are all elementary to establish, and which in probabilistic terms are dealing with the moments of order $0,1,2$ of the binomial laws:

$$\sum_k\binom{n}{k}x^k(1-x)^{n-k}=1$$

$$\sum_k\frac{k}{n}\binom{n}{k}x^k(1-x)^{n-k}=x$$

$$\sum_k\left(x-\frac{k}{n}\right)^2\binom{n}{k}x^k(1-x)^{n-k}=\frac{x(1-x)}{n}$$

(2) In terms of the basic Bernstein polynomials b_{kn}, the above formulae read:

$$\sum_k b_{kn}(x)=1$$

$$\sum_k\frac{k}{n}\cdot b_{kn}(x)=x$$

$$\sum_k\left(x-\frac{k}{n}\right)^2 b_{kn}(x)=\frac{x(1-x)}{n}$$

(3) Now consider our arbitrary continuous function $f:[0,1]\to\mathbb{R}$, and construct for any $n\in\mathbb{N}$ the approximation indicated above, namely:

$$f_n(x)=\sum_{k=0}^n f\left(\frac{k}{n}\right)b_{kn}(x)$$

In order to estimate the error $|f_n-f|$, we use the uniform continuity property of f. So, pick $\varepsilon>0$, and then $\delta>0$ such that the following happens:

$$|x-y|<\delta\implies|f(x)-f(y)|<\varepsilon$$

Now with this done, we have the following estimate, using the first formula in (2) at the first step, the uniform continuity at the last step, and with $M=\sup|f|$:

$$\begin{aligned}
&|f_n(x)-f(x)|\\
=&\left|\sum_k\left(f\left(\frac{k}{n}\right)-f(x)\right)b_{kn}(x)\right|\\
\leq&\sum_k\left|f\left(\frac{k}{n}\right)-f(x)\right|b_{kn}(x)\\
=&\sum_{\left|x-\frac{k}{n}<\delta\right|}\left|f\left(\frac{k}{n}\right)-f(x)\right|b_{kn}(x)+\sum_{\left|x-\frac{k}{n}\geq\delta\right|}\left|f\left(\frac{k}{n}\right)-f(x)\right|b_{kn}(x)\\
\leq&\ \varepsilon+M\sum_{\left|x-\frac{k}{n}\geq\delta\right|}b_{kn}(x)
\end{aligned}$$

(4) The point now is that the last sum on the right can be estimated by using the Chebycheff inequality, based on the third formula from (2), and we obtain:

$$\begin{aligned}
\sum_{\left|x-\frac{k}{n}\geq\delta\right|}b_{kn}(x)&\leq\sum_k\delta^{-2}\left(x-\frac{k}{n}\right)^2b_{kn}(x)\\
&=\delta^{-2}\,\frac{x(1-x)}{n}\\
&\leq\frac{\delta^{-2}}{4n}
\end{aligned}$$

(5) Now by putting everything together, we obtain the following estimate:

$$|f_n(x)-f(x)|\leq\varepsilon+\frac{\delta^{-2}M}{4n}$$

Thus we have indeed $|f_n-f|\to0$, uniform convergence, as desired.

(6) Summarizing, present theorem proved, modulo some learning in relation with the Chebycheff inequality that we used above, that we will leave as an exercise. □

There are many other versions of the Weierstrass theorem, and interesting illustrations too. Exercise for you to learn more about all this, and the more, the better.

Back now to the Hilbert spaces, we recall that, according to Theorem 9.20, there is only one separable such space, up to isomorphism. As a first result regarding this unique separable Hilbert space that we are interested in, we have:

THEOREM 9.22. *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 9.20.

(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 working out the details here as an instructive exercise. □

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

CONCLUSION 9.23. *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 the following fomula, X being a separable measured space,

$$H = L^2(X)$$

or simply say that H is an abstract separable Hilbert space.

It is also possible to make some physics comments here, with this unique separable Hilbert space being, and no surprise here, the space that we live in.

9e. Exercises

Exercises:

Exercise 9.24.

Exercise 9.25.

Exercise 9.26.

Exercise 9.27.

Exercise 9.28.

Exercise 9.29.

Exercise 9.30.

Exercise 9.31.

Bonus exercise.

CHAPTER 10

Orthogonal polynomials

10a. Orthogonal polynomials

The separability results from the previous chapter are something quite subtle. In practice, all this leads us into orthogonal polynomials, which are defined as follows:

Definition 10.1. *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.

Observe that the orthogonal polynomials exist indeed for any real measure $d\mu(x) = f(x)dx$, as explained above. It is possible to be a bit more explicit here, as follows:

Theorem 10.2. *The orthogonal polynomials with respect to μ are given by*

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

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:

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

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 10.3. *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 following differential equation,

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

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

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

called Bonnet recurrence formula, as well as 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. As a first observation, we are not lost somewhere in abstract math, because of the occurrence of the Legendre equation, which is something quite fundamental in physics. As for the proof of the result, this is quite standard, going as follows:

(1) The first assertion is clear, because the Gram-Schmidt procedure applied to the Weierstrass basis $\{x^k\}$ can only lead to a certain family of polynomials $\{P_k\}$, with each P_k being of degree k, and also unique, if we assume that it has positive leading coefficient, with this $\pm$ choice being needed, as usual, at each step of Gram-Schmidt.

(2) In order to have now an idea about these beasts, here are the first few of them, which can be obtained say via a straightforward application of Gram-Schmidt:

$$\begin{aligned}
P_0 &= 1\\
P_1 &= x\\
P_2 &= (3x^2-1)/2\\
P_3 &= (5x^3-3x)/2\\
P_4 &= (35x^4-30x^2+3)/8\\
P_5 &= (63x^5-70x^3+15x)/8
\end{aligned}$$

(3) Now thinking about what Gram-Schmidt does, this is certainly something by recurrence. And examining the recurrence leads to the Legendre equation in the statement, which is the well-known Legendre equation from physics, namely:

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

(4) As for the Bonnet recurrence formula, which is as follows, the story here is similar:

$$(k+1)P_{k+1}(x)=(2k+1)xP_k(x)-kP_{k-1}(x)$$

(5) Finally, let us examine the Rodrigues formula in the statement, namely:

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

In order to prove this, by uniqueness there is no really need to try to understand where this formula comes from, and we have two choices here, either by verifying that $\{P_k\}$ is orthonormal, or by verifying the Legendre equation. And both methods work. $\square$

Many other things can be said about the Legendre polynomials, as a continuation of the above, involving both mathematics and physics. We will be back to this.

10b.

10c.

10d.

10e. Exercises

Exercises:

Exercise 10.4.

Exercise 10.5.

Exercise 10.6.

Exercise 10.7.

Exercise 10.8.

Exercise 10.9.

Exercise 10.10.

Exercise 10.11.

Bonus exercise.

CHAPTER 11

Classical polynomials

11a. Classical polynomials

The results from chapter 10 regarding the Legendre polynomials are just the tip of the iceberg, and as a continuation of that material, we have:

THEOREM 11.1. *The orthogonal polynomials for $L^2[-1,1]$, with measure*

$$d\mu(x) = (1-x)^a(1+x)^b dx$$

called Jacobi polynomials, satisfy as well a degree 2 equation, namely

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

as well as an order 2 recurrence relation, and are given by the following formula:

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

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

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

(1) To start with, in what regards the precise statement, the order 2 recurrence relation mentioned there is something quite complicated, as follows:

$$\begin{aligned}
& 2k(k+a+b)(2k+a+b-2)P_k(x) \\
= \ & (2k+a+b-1)\left[(2k+a+b)(2k+a+b-2)x+a^2-b^2\right]P_{k-1}(x) \\
- \ & 2(k+a-1)(k+b-1)(2k+a+b)P_{k-2}(x)
\end{aligned}$$

(2) Regarding now the proof, the statement itself appears as a generalization of the result for Legendre polynomials, which corresponds to the particular case $a=b=0$, and the proof is quite similar. We will leave learning more about all this as an exercise.

(3) For completness, let us record as well a few numerics, as follows:

$$\begin{aligned} P_0 &= 1 \\ P_1 &= (a+1)+(a+b+2)\,\frac{x-1}{2} \\ P_2 &= \frac{(a+1)(a+2)}{2}+(a+2)(a+b+3)\,\frac{x-1}{2} \\ &+ \frac{(a+b+3)(a+b+4)}{2}\left(\frac{x-1}{2}\right)^2 \end{aligned}$$

(4) Regarding now the main particular cases of the Jacobi polynomials, these are the Gegenbauer polynomials, appearing at $a=b$. However, there is not that much of a simplification when passing from general parameters a,b to equal parameters, $a=b$, so in practice, the main particular cases are those indicated in the statement, namely:

– The Legendre polynomials, that we know well from chapter 10, appearing at the simplest values of the parameters, namely $a=b=0$.

– The Chebycheff polynomials of the first kind T_k, which are given by the formula $T_k(\cos t)=\cos(kt)$ from trigonometry, appearing at $a=b=-\frac{1}{2}$.

– The Chebycheff polynomials of the second kind U_k, which are given by the formula $U_k(\cos t)\sin t=\sin((k+1)t)$, appearing at $a=b=\frac{1}{2}$.

(5) So, this was for the story of the Jacobi polynomials, and their main particular cases, and in practice, we will leave some further learning here as an exercise, coming as a continuation of the further learning about the Legendre polynomials. □

Getting now to other spaces $X\subset\mathbb{R}$, we have the following result, which complements well the theory of Legendre polynomials, for the needs of basic quantum mechanics:

Theorem 11.2. *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 $\{P_k\}$, satisfying the following differential equation,

$$xP_k''(x)+(1-x)P_k'(x)+kP_k(x)=0$$

as well as the following order 2 recurrence relation,

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

and which are 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 and Jacobi polynomials, and many further things can be said here, with exercise for you to learn a bit about all this. Let us record as well a few numeric values, for the Laguerre polynomials:

$$\begin{aligned}
P_0 &= 1\\
P_1 &= 1-x\\
P_2 &= (x^2-4x+2)/2\\
P_3 &= (-x^3+9x^2-18x+6)/6\\
P_4 &= (x^4-16x^3+72x^2-96x+24)/24
\end{aligned}$$

Finally, for the story to be complete, no discussion about the Laguerre polynomials would be complete without a word about their use, in quantum mechanics. □

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

THEOREM 11.3. *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 $\{P_k\}$, satisfying the following differential equation,

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

as well as the following order 2 recurrence relation,

$$P_{k+1}(x)=2xP_k(x)-2kP_{k-1}(x)$$

and which are 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. Let us record as well a few numeric values, for the Hermite polynomials:

$$\begin{aligned}
P_0 &= 1\\
P_1 &= 2x\\
P_2 &= 4x^2-2\\
P_3 &= 8x^3-12x\\
P_4 &= 16x^4-48x^2+12\\
P_5 &= 32x^5-160x^3+120x\\
P_6 &= 64x^6-480x^4+720x^2-120
\end{aligned}$$

With of course, exercise for you to deduce all these formulae. □

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 11.4. *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.*

Let us end this chapter with more about the Chebycheff polynomials, which are perhaps the most important, in the above list. As mentioned in the proof of Theorem 11.1, these naturally appear in the context of trigonometry, via the following formulae:

$$T_k(\cos t) = \cos(kt) \quad , \quad U_k(\cos t)\sin t = \sin((k+1)t)$$

However, much more can be said. Let us start with the following standard result:

PROPOSITION 11.5. *The eigenvalues of the segment graph are as follows:*

(1) *At* $N = 2$ *we have* $-1, 1$.
(2) *At* $N = 3$ *we have* $0, \pm\sqrt{2}$.
(3) *At* $N = 4$ *we have* $\pm\frac{\sqrt{5}\pm 1}{2}$.
(4) *At* $N = 5$ *we have* $0, \pm 1, \pm\sqrt{3}$.
(5) *At* $N = 6$ *we have the solutions of* $x^6 - 5x^4 + 6x^2 - 1 = 0$.

PROOF. These are some straightforward linear algebra computations, with some tricks being needed only at $N = 4$, the details being as follows:

(1) At $N = 2$ the adjacency matrix and its eigenvalues are as follows:

$$d = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \begin{vmatrix} x & -1 \\ -1 & x \end{vmatrix} = x^2 - 1 = (x-1)(x+1)$$

(2) At $N = 3$ the adjacency matrix and its eigenvalues are as follows:

$$d = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad , \quad \begin{vmatrix} x & -1 & 0 \\ -1 & x & -1 \\ 0 & -1 & x \end{vmatrix} = x^3 - 2x = x(x^2 - 2)$$

(3) At $N = 4$ the adjacency matrix and its characteristic polynomial are as follows:

$$d = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad , \quad \begin{vmatrix} x & -1 & 0 & 0 \\ -1 & x & -1 & 0 \\ 0 & -1 & x & -1 \\ 0 & 0 & -1 & x \end{vmatrix} = x^4 - 3x^2 + 1$$

Now by solving the degree 2 equation, and fine-tuning the answer, we obtain:

$$x = \pm\sqrt{\frac{3 \pm \sqrt{5}}{2}} = \pm\frac{\sqrt{5} \pm 1}{2}$$

(4) At $N=5$ the adjacency matrix and its eigenvalues are as follows:

$$d=\begin{pmatrix}0&1&0&0&0\\1&0&1&0&0\\0&1&0&1&0\\0&0&1&0&1\\0&0&0&1&0\end{pmatrix}\quad,\quad\begin{vmatrix}x&-1&0&0&0\\-1&x&-1&0&0\\0&-1&x&-1&0\\0&0&-1&x&-1\\0&0&0&-1&x\end{vmatrix}=x(x^2-1)(x^2-3)$$

(5) At $N=6$ the adjacency matrix and its eigenvalues are as follows:

$$d=\begin{pmatrix}0&1&0&0&0&0\\1&0&1&0&0&0\\0&1&0&1&0&0\\0&0&1&0&1&0\\0&0&0&1&0&1\\0&0&0&0&1&0\end{pmatrix}\quad,\quad\begin{vmatrix}x&-1&0&0&0&0\\-1&x&-1&0&0&0\\0&-1&x&-1&0&0\\0&0&-1&x&-1&0\\0&0&0&-1&x&-1\\0&0&0&0&-1&x\end{vmatrix}=x^6-5x^4+6x^2-1$$

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

All the above does not look very good. However, as a matter of having the dirty job fully done, with mathematical pride, let us look as well into eigenvectors. At $N=2$ things are quickly settled, with the diagonalization of the adjacency matrix being as follows:

$$\begin{pmatrix}0&1\\1&0\end{pmatrix}=\frac{1}{2}\begin{pmatrix}1&1\\1&-1\end{pmatrix}\begin{pmatrix}1&0\\0&-1\end{pmatrix}\begin{pmatrix}1&1\\1&-1\end{pmatrix}$$

At $N=3$ the diagonalization formula becomes more complicated, as follows:

$$\begin{pmatrix}0&1&0\\1&0&1\\0&1&0\end{pmatrix}=\frac{1}{4}\begin{pmatrix}1&1&1\\0&\sqrt{2}&-\sqrt{2}\\-1&1&1\end{pmatrix}\begin{pmatrix}0&0&0\\0&\sqrt{2}&0\\0&0&-\sqrt{2}\end{pmatrix}\begin{pmatrix}2&0&-2\\1&\sqrt{2}&1\\1&-\sqrt{2}&1\end{pmatrix}$$

At $N=4$ now, in view of the eigenvalue formula that we found, $x^4-3x^2+1=0$, we must proceed with care. The equation for the eigenvectors $dv=xv$ is as follows:

$$\begin{pmatrix}0&1&0&0\\1&0&1&0\\0&1&0&1\\0&0&1&0\end{pmatrix}\begin{pmatrix}a\\b\\c\\d\end{pmatrix}=x\begin{pmatrix}a\\b\\c\\d\end{pmatrix}$$

In other words, the equations to be satisfied are as follows:

$$b=xa$$

$$a+c=xb$$

$$b+d=xc$$

$$c=xd$$

With the choice $a = 1$, the solutions of these equations are as follows:

$$a = 1$$
$$b = x$$
$$c = x^2 - 1$$
$$d = (x^2 - 1)/x$$

In order to compute now the c components of the eigenvectors, we can use the formula $x = (\pm 1 \pm \sqrt{5})/2$ from Proposition 11.5. Indeed, this formula gives:

$$\begin{aligned}\left(\frac{\pm 1 \pm \sqrt{5}}{2}\right)^2 - 1 &= \frac{6 \pm 2\sqrt{5}}{4} - 1\\ &= \frac{3 \pm \sqrt{5}}{2} - 1\\ &= \frac{1 \pm \sqrt{5}}{2}\end{aligned}$$

Thus, almost done, and we deduce that the passage matrix is as follows:

$$P = \begin{pmatrix} 1 & 1 & 1 & 1\\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} & \frac{-1+\sqrt{5}}{2} & \frac{-1-\sqrt{5}}{2}\\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} & \frac{1+\sqrt{5}}{2}\\ 1 & 1 & -1 & -1 \end{pmatrix}$$

To be more precise, according to our equations above, the first row must consist of $a = 1$ entries. Then the second row must consist of $b = x$ entries, with $x = (\pm 1 \pm \sqrt{5})/2$. Then the third row must consist of $c = x^2 - 1$ entries, but these are easily computed, as explained above. Finally, the fourth row must consist of $d = (x^2 - 1)/x$ entries, which means that the fourth row appears by dividing the third row by the second row, which is easily done too. In case you wonder why $d = \pm 1$, here is another proof of this:

$$\begin{aligned}d = \pm 1 &\iff x^2 - 1 = \pm x\\ &\iff (x^2 - 1)^2 = x^2\\ &\iff x^4 - 2x^2 + 1 = x^2\\ &\iff x^4 - 3x^2 + 1 = 0\end{aligned}$$

Very nice all this, and leaving the computation of P^{-1} for you, here are a few more observations, in relation with what we found in Proposition 11.5:

(1) In the case $N = 5$ the eigenvectors can be computed too, and the diagonalization finished, via the standard method, namely system of equations, and with the numerics involving powers of the eigenvalues that we found. Exercise for you.

(2) The same stays true at $N = 6$, again with the eigenvector numerics involving powers of the eigenvalues, and with these eigenvalues being explicitly computable, via the Cardano formula for degree 3 equations. Have fun with this too, of course.

(3) However, all this does not look very good, and at $N = 7$ and higher we will certainly run into difficult questions, and save for some interesting remarks, normally depending on the parity of N, we will not be able to fully solve the diagonalization problem.

So, what to do? Work some more, of course, the hard way. Proposition 11.5 and its proof look quite trivial, but if you get into the full details of the computations, and let me assign here this, as a key exercise, you will certainly notice that, when computing that determinants, you can sort of use a recurrence method. And, this leads to:

THEOREM 11.6. *The characteristic polynomials P_N of the segment graphs satisfy*

$$P_0 = 1 \quad , \quad P_1 = x \quad , \quad P_{N+1} = xP_N - P_{N-1}$$

and are the well-known Chebycheff polynomials from trigonometry.

PROOF. Obviously, many things going on here, ranging from precise to definitional, or even informal, the idea with all this being as follows:

(1) To start with, by computing determinants, we are led to the recurrence formula in the statement. Here is the proof at $N = 4$, the general case being similar:

$$\begin{aligned} P_5 &= \begin{vmatrix} x & -1 & 0 & 0 & 0 \\ -1 & x & -1 & 0 & 0 \\ 0 & -1 & x & -1 & 0 \\ 0 & 0 & -1 & x & -1 \\ 0 & 0 & 0 & -1 & x \end{vmatrix} \\ &= x \begin{vmatrix} x & -1 & 0 & 0 \\ -1 & x & -1 & 0 \\ 0 & -1 & x & -1 \\ 0 & 0 & -1 & x \end{vmatrix} + \begin{vmatrix} -1 & -1 & 0 & 0 \\ 0 & x & -1 & 0 \\ 0 & -1 & x & -1 \\ 0 & 0 & -1 & x \end{vmatrix} \\ &= x \begin{vmatrix} x & -1 & 0 & 0 \\ -1 & x & -1 & 0 \\ 0 & -1 & x & -1 \\ 0 & 0 & -1 & x \end{vmatrix} - \begin{vmatrix} x & -1 & 0 \\ -1 & x & -1 \\ 0 & -1 & x \end{vmatrix} \\ &= xP_4 - P_3 \end{aligned}$$

(2) Regarding now the intial values for our recurrence, according to Proposition 11.5 these should be normally the following two polynomials:

$$P_2 = x^2 - 1 \quad , \quad P_3 = x^3 - 2x$$

(3) However, we can formally add the values $P_0 = 1$ and $P_1 = x$, as for the final statement to look better, with this being justified by the following formulae:

$$P_0 = 1$$
$$P_1 = x$$
$$P_2 = xP_1 - P_0 = x^2 - 1$$
$$P_3 = xP_2 - P_1 = x^3 - 2x$$

(4) Thus, we have the recurrence formula in the statement, and with the initial values there, and our polynomials follow to be the Chebycheff polynomials, as stated. □

11b.

11c.

11d.

11e. Exercises

Exercises:

Exercise 11.7.

Exercise 11.8.

Exercise 11.9.

Exercise 11.10.

Exercise 11.11.

Exercise 11.12.

Exercise 11.13.

Exercise 11.14.

Bonus exercise.

CHAPTER 12

Further polynomials

12a. Further polynomials

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

Complex variables

CHAPTER 13

Complex measures

13a. Complex measures

Let us discuss now the complex analogues of all the above, first with a notion of complex normal, or Gaussian law. To start with, we have the following definition:

Definition 13.1. *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 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 13.2. *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 follows indeed from the real result, namely $g_s * g_t = g_{s+t}$, established in chapter 2, simply by taking the real and imaginary parts. □

We have as well the following complex analogue of the CLT:

Theorem 13.3 (CCLT). *Given complex variables $f_1, f_2, f_3, \ldots \in L^\infty(X)$ 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 $n \to \infty$, in moments.

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. □

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 13.4. *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.

Observe that, since $f, \bar{f}$ commute, we can permute terms, and restrict the attention to exponents of type $k = \ldots\circ\circ\circ\bullet\bullet\bullet\ldots$, if we want to. However, our results about the complex Gaussian laws, and other complex laws too, later on, will actually look better without doing is, so we will use Definition 13.4 as stated. We first have:

Theorem 13.5. *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 $k = \circ\bullet\bullet\circ\ldots$, of the variable from Definition 13.1, 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. □

As before with the real Gaussian laws, a better-looking statement is in terms of partitions. Given a colored integer $k = \circ\bullet\bullet\circ\ldots$, we say that a pairing $\pi \in P_2(k)$ is matching when it pairs $\circ - \bullet$ symbols. With this convention, we have the following result:

THEOREM 13.6. *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 13.5. Indeed, we can assume that we are in the case $t=1$, and here we know from Theorem 13.5 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 13.7 (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*

$$E\left(f_{i_1}^{k_1}\ldots f_{i_s}^{k_s}\right)=t^{s/2}\#\left\{\pi\in\mathcal{P}_2(k)\Big|\pi\leq\ker i\right\}$$

where $k=k_1\ldots k_s$ and $i=i_1\ldots i_s$, for the joint moments of these variables, 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 13.6, 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 13.8 (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 13.7. With some changes in the indices and notations, the formula there reads:

$$E\left(f_{I_1}^{K_1}\dots f_{I_s}^{K_s}\right)=t^{s/2}\#\left\{\sigma\in\mathcal{P}_2(K)\Big|\sigma\leq\ker I\right\}$$

Now observe that we have $\mathcal{P}_2(K)=\emptyset$, unless the colored integer $K=K_1\dots K_s$ 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 $K=K_1\dots K_s$ is of the following special form:

$$K=\underbrace{\circ\circ\dots\circ}_k\ \underbrace{\bullet\bullet\dots\bullet}_k$$

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 $I=I_1\dots I_s$ in the following way:

$$I=i_1\dots i_k\ j_1\dots j_k$$

With these changes made, the above usual Wick formula reads:

$$E\left(f_{i_1}\dots f_{i_k}f_{j_1}^*\dots f_{j_k}^*\right)=t^k\#\left\{\sigma\in\mathcal{P}_2(K)\Big|\sigma\leq\ker(ij)\right\}$$

The point now is that the matching pairings $\sigma\in\mathcal{P}_2(K)$, with $K=\circ\dots\circ\bullet\dots\bullet$, 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:

$$E\left(f_{i_1}\dots f_{i_k}f_{j_1}^*\dots f_{j_k}^*\right)=t^k\#\left\{\pi\in S_k\Big|i_{\pi(r)}=j_r,\forall r\right\}$$

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 13.9 (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*

$$E\left(f_{i_1}f_{j_1}^*\dots f_{i_k}f_{j_k}^*\right)=t^k\#\left\{\pi\in S_k\Big|i_{\pi(r)}=j_r,\forall r\right\}$$

for the non-vanishing joint moments of these variables.

PROOF. This follows from our second Wick formula, from Theorem 13.8, simply by permuting the terms, as to have an alternating sequence of plain and conjugate variables. Alternatively, we can start with Theorem 13.7, and then perform the same manipulations as in the proof of Theorem 13.8, but with the exponent being this time as follows:

$$K=\underbrace{\circ\bullet\circ\bullet\dots\dots\circ\bullet}_{2k}$$

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

13b.

13c.

13d.

13e. Exercises

Exercises:

Exercise 13.10.

Exercise 13.11.

Exercise 13.12.

Exercise 13.13.

Exercise 13.14.

Exercise 13.15.

Exercise 13.16.

Exercise 13.17.

Bonus exercise.

CHAPTER 14

Moment problem

14a. Moment problem

14b.

14c.

14d.

14e. Exercises

Exercises:

EXERCISE 14.1.

EXERCISE 14.2.

EXERCISE 14.3.

EXERCISE 14.4.

EXERCISE 14.5.

EXERCISE 14.6.

EXERCISE 14.7.

EXERCISE 14.8.

Bonus exercise.

CHAPTER 15

Cumulant theory

15a. Cumulant theory

15b.

15c.

15d.

15e. Exercises

Exercises:

Exercise 15.1.

Exercise 15.2.

Exercise 15.3.

Exercise 15.4.

Exercise 15.5.

Exercise 15.6.

Exercise 15.7.

Exercise 15.8.

Bonus exercise.

CHAPTER 16

Special functions

16a. Special functions

16b.

16c.

16d.

16e. Exercises

Congratulations for having read this book, and no exercises for this final chapter.

Bibliography

[1] G.W. Anderson and B. Farrell, Asymptotically liberating sequences of random unitary matrices, *Adv. Math.* **255** (2014), 381–413.

[2] G.W. Anderson, A. Guionnet and O. Zeitouni, An introduction to random matrices, Cambridge Univ. Press (2010).

[3] M. Anshelevich, Free Meixner states, *Comm. Math. Phys.* **276** (2007), 863–899.

[4] O. Arizmendi, I. Nechita and C. Vargas, On the asymptotic distribution of block-modified random matrices, *J. Math. Phys.* **57** (2016), 1–27.

[5] G. Aubrun, Partial transposition of random states and non-centered semicircular distributions, *Random Matrices Theory Appl.* **1** (2012), 125–145.

[6] V.I. Arnold, Mathematical methods of classical mechanics, Springer (1974).

[7] M.F. Atiyah, The geometry and physics of knots, Cambridge Univ. Press (1990).

[8] T. Banica, Principles of operator algebras (2024).

[9] T. Banica, S.T. Belinschi, M. Capitaine and B. Collins, Free Bessel laws, *Canad. J. Math.* **63** (2011), 3–37.

[10] T. Banica and S. Curran, Decomposition results for Gram matrix determinants, *J. Math. Phys.* **51** (2010), 1–14.

[11] T. Banica and R. Speicher, Liberation of orthogonal Lie groups, *Adv. Math.* **222** (2009), 1461–1501.

[12] R.J. Baxter, Exactly solved models in statistical mechanics, Academic Press (1982).

[13] S.T. Belinschi and H. Bercovici, Partially defined semigroups relative to multiplicative free convolution, *Int. Math. Res. Not.* **2** (2005), 65–101.

[14] S.T. Belinschi, M. Bożejko, F. Lehner and R. Speicher, The normal distribution is $\boxplus$-infinitely divisible, *Adv. Math.* **226** (2011), 3677–3698.

[15] H. Bercovici and V. Pata, Stable laws and domains of attraction in free probability theory, *Ann. of Math.* **149** (1999), 1023–1060.

[16] H. Bercovici and D.V. Voiculescu, Free convolutions of measures with unbounded support, *Indiana Univ. Math. J.* **42** (1993), 733–773.

[17] I. Bengtsson and K. Życzkowski, Geometry of quantum states, Cambridge Univ. Press (2006).

[18] N. Berline, E. Getzler and M. Vergne, Heat kernels and Dirac operators, Springer (2004).

[19] A. Bose, Random matrices and non-commutative probability, CRC Press (2021).

[20] B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability, *Int. Math. Res. Not.* **17** (2003), 953–982.

[21] B. Collins and S. Matsumoto, On some properties of orthogonal Weingarten functions, *J. Math. Phys.* **50** (2009), 1–18.

[22] B. Collins and P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic groups, *Comm. Math. Phys.* **264** (2006), 773–795.

[23] A. Connes, Noncommutative geometry, Academic Press (1994).

[24] S. Curran, Quantum rotatability, *Trans. Amer. Math. Soc.* **362** (2010), 4831–4851.

[25] S. Curran and R. Speicher, Quantum invariant families of matrices in free probability, *J. Funct. Anal.* **261** (2011), 897–933.

[26] P. Di Francesco, Meander determinants, *Comm. Math. Phys.* **191** (1998), 543–583.

[27] P.A.M. Dirac, Principles of quantum mechanics, Oxford Univ. Press (1930).

[28] R. Durrett, Probability: theory and examples, Cambridge Univ. Press (1990).

[29] A. Einstein, Relativity: the special and the general theory, Dover (1916).

[30] L.C. Evans, Partial differential equations, AMS (1998).

[31] W. Feller, An introduction to probability theory and its applications, Wiley (1950).

[32] E. Fermi, Thermodynamics, Dover (1937).

[33] M. Février and A. Nica, Infinitesimal non-crossing cumulants and free probability of type B, *J. Funct. Anal.* **258** (2010), 2983–3023.

[34] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics, Caltech (1963).

[35] P. Flajolet and R. Sedgewick, Analytic combinatorics, Cambridge Univ. Press (2009).

[36] M. Fukuda and P. Śniady, Partial transpose of random quantum states: exact formulas and meanders, *J. Math. Phys.* **54** (2013), 1–31.

[37] C. Godsil and G. Royle, Algebraic graph theory, Springer (2001).

[38] D.J. Griffiths, Introduction to electrodynamics, Cambridge Univ. Press (2017).

[39] D.J. Griffiths and D.F. Schroeter, Introduction to quantum mechanics, Cambridge Univ. Press (2018).

[40] D.J. Griffiths, Introduction to elementary particles, Wiley (2020).

[41] A. Guionnet, M. Krishnapur and O. Zeitouni, The single ring theorem, *Ann. of Math.* **174** (2011), 1189–1217.

[42] J. Harris, Algebraic geometry, Springer (1992).

[43] R. Hartshorne, Algebraic geometry, Springer (1977).

[44] F. Hiai and D. Petz, The semicircle law, free random variables and entropy, AMS (2000).

[45] H. Hofer and E. Zehnder, Symplectic invariants and Hamiltonian dynamics, Birkhäuser (1994).

[46] L. Hörmander, The analysis of linear partial differential operators, Springer (1983).

[47] R.A. Horn and C.R. Johnson, Matrix analysis, Cambridge Univ. Press (1985).

[48] K. Huang, Introduction to statistical physics, CRC Press (2001).

[49] J.E. Humphreys, Introduction to Lie algebras and representation theory, Springer (1972).

[50] V.F.R. Jones, Subfactors and knots, AMS (1991).

[51] L.P. Kadanoff, Statistical physics: statics, dynamics and renormalization, World Scientific (2000).

[52] T. Kibble and F.H. Berkshire, Classical mechanics, Imperial College Press (1966).

[53] C. Köstler, R. Speicher, A noncommutative de Finetti theorem: invariance under quantum permutations is equivalent to freeness with amalgamation, *Comm. Math. Phys.* **291** (2009), 473–490.

[54] M. Kumar, Quantum: Einstein, Bohr, and the great debate about the nature of reality, Norton (2009).

[55] T. Lancaster and K.M. Blundell, Quantum field theory for the gifted amateur, Oxford Univ. Press (2014).

[56] L.D. Landau and E.M. Lifshitz, Mechanics, Pergamon Press (1960).

[57] L.D. Landau and E.M. Lifshitz, The classical theory of fields, Addison-Wesley (1951).

[58] L.D. Landau and E.M. Lifshitz, Quantum mechanics: non-relativistic theory, Pergamon Press (1959).

[59] S. Lang, Algebra, Addison-Wesley (1993).

[60] P. Lax, Linear algebra and its applications, Wiley (2007).

[61] P. Lax, Functional analysis, Wiley (2002).

[62] W. Liu, General de Finetti type theorems in noncommutative probability, *Comm. Math. Phys.* **369** (2019), 837–866.

[63] V.A. Marchenko and L.A. Pastur, Distribution of eigenvalues in certain sets of random matrices, *Mat. Sb.* **72** (1967), 507–536.

[64] M.L. Mehta, Random matrices, Elsevier (2004).

[65] J.A. Mingo and A. Nica, Annular noncrossing permutations and partitions, and second-order asymptotics for random matrices, *Int. Math. Res. Not.* **28** (2004), 1413–1460.

[66] J.A. Mingo and M. Popa, Freeness and the transposes of unitarily invariant random matrices, *J. Funct. Anal.* **271** (2016), 883–921.

[67] J.A. Mingo and R. Speicher, Free probability and random matrices, Springer (2017).

[68] A. Nica and R. Speicher, Lectures on the combinatorics of free probability, Cambridge Univ. Press (2006).

[69] M.A. Nielsen and I.L. Chuang, Quantum computation and quantum information, Cambridge Univ. Press (2000).

[70] R.K. Pathria and and P.D. Beale, Statistical mechanics, Elsevier (1972).

[71] A. Peres, Quantum theory: concepts and methods, Kluwer (1993).

[72] P. Petersen, Linear algebra, Springer (2012).

[73] W. Rudin, Principles of mathematical analysis, McGraw-Hill (1964).

[74] W. Rudin, Real and complex analysis, McGraw-Hill (1966).

[75] W. Rudin, Functional analysis, McGraw-Hill (1973).

[76] W. Rudin, Fourier analysis on groups, Dover (1974).

[77] B. Ryden, Introduction to cosmology, Cambridge Univ. Press (2002).

[78] D.V. Schroeder, An introduction to thermal physics, Oxford Univ. Press (1999).

[79] J. Schwinger and B.H. Englert, Quantum mechanics: symbolism of atomic measurements, Springer (2001).

[80] I.R. Shafarevich, Basic algebraic geometry, Springer (1974).

[81] R. Shankar, Fundamentals of physics, Yale Univ. Press (2014).

[82] R. Speicher, Multiplicative functions on the lattice of noncrossing partitions and free convolution, *Math. Ann.* **298** (1994), 611–628.

[83] R. Speicher, Combinatorial theory of the free product with amalgamation and operator-valued free probability theory, *Mem. Amer. Math. Soc.* **132** (1998).

[84] S. Sternberg, Dynamical systems, Dover (2010).

[85] J.R. Taylor, Classical mechanics, Univ. Science Books (2003).

[86] N.H. Temperley and E.H. Lieb, Relations between the "percolation" and "colouring" problem and other graph-theoretical problems associated with regular planar lattices: some exact results for the "percolation" problem, *Proc. Roy. Soc. London* **322** (1971), 251–280.

[87] D.V. Voiculescu, Addition of certain noncommuting random variables, *J. Funct. Anal.* **66** (1986), 323–346.

[88] D.V. Voiculescu, Multiplication of certain noncommuting random variables, *J. Operator Theory* **18** (1987), 223–235.

[89] D.V. Voiculescu, Limit laws for random matrices and free products, *Invent. Math.* **104** (1991), 201–220.

[90] D.V. Voiculescu, K.J. Dykema and A. Nica, Free random variables, AMS (1992).

[91] J. von Neumann, Mathematical foundations of quantum mechanics, Princeton Univ. Press (1955).

[92] J. von Neumann and O. Morgenstern, Theory of games and economic behavior, Princeton Univ. Press (1944).

[93] J. Watrous, The theory of quantum information, Cambridge Univ. Press (2018).

[94] S. Weinberg, Foundations of modern physics, Cambridge Univ. Press (2011).

[95] S. Weinberg, Lectures on quantum mechanics, Cambridge Univ. Press (2012).

[96] H. Weyl, The theory of groups and quantum mechanics, Princeton Univ. Press (1931).

[97] H. Weyl, The classical groups: their invariants and representations, Princeton Univ. Press (1939).

[98] H. Weyl, Space, time, matter, Princeton Univ. Press (1918).

[99] E. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, *Ann. of Math.* **62** (1955), 548–564.

[100] B. Zwiebach, A first course in string theory, Cambridge Univ. Press (2004).

Index