Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea Magistrale in Fisica

A Random Matrix Theory approach to Complex Networks

Relatore:
Prof. Armando Bazzani

Presentata da:
Pierfrancesco Dionigi

Anno Accademico 2018/2019

"*Je n'ai pas le temps...*"
◊ *Évariste Galois* ◊

Abstract

It is presented a formal mathematical approach to complex networks through Random Matrix Theory (RMT). Wigner semicircular law is presented as a generalization of Central Limit Theorem for certain classes of matrix ensembles. The different ways for dealing with the spectral distribution of random matrices are shown and the connections and the differences between them are highlighted. Then it is studied how RMT can be linked to Free Probability theory. It is exposed how two different kinds of random graphs, apparently similar, posses different spectral properties analyzing their adjacency matrices. From this analysis some geometrical and topological properties of the graphs are deduced and correlations between vertices can be analyzed. Then a random walk is constructed on the graph through a Markov chain, defining the transition matrix of the process with the weighted adjacency matrix of the network. Finally it is showed how the dynamical behavior of the random walk is deeply connected with the eigenvalues of the transition matrix, and the main relation are shown.

***

Si presenta un approccio matematico formale ai complex networks tramite l'uso della Random Matrix Theory (RMT). La legge del semicerchio di Wigner viene presentata come una generalizzazione del Teorema del Limite Centrale per determinati ensemble di matrici random. Sono presentati inoltre i principali metodi per calcolare la distribuzione spettrale delle matrici random e se ne sottolineano le differenze. Si è poi studiato come la RMT sia collegata alla Free Probability. Si è studiato come due tipi di grafi random apparentemente uguali, posseggono proprietà spettrali differenti analizzando le loro matrici di adiacenza. Da questa analisi si deducono alcune proprietà geometriche e topologiche dei grafi e si può analizzare la correlazione statistica tra i vertici. Si è poi costruito sul grafo un passeggiata aleatoria tramite catene di Markov, definendo la matrice di transizione del processo tramite la matrice di adiacenza del network opportunamente normalizzata. Infine si è mostrato come il comportamento dinamico della passeggiata aleatoria sia profondamente connesso con gli autovalori della matrice di transizione, e le principali relazioni sono mostrate.

Contents

Chapter 1

Introduction

While the world is becoming more complex and interconnected, we need more powerful tools to deal with new emerging issues or simply for making previsions. Today's technological means give us good possibilities in gathering crucial informations from data and thus having faithful data driven models. But despite the usefulness of these models sometimes our knowledge of the underlying mechanisms is not as good as the technics we have to analyze the data. In this view also a theoretical and system modeling work is nowadays still necessary. On the one hand we need it for its intrinsically importance as fundamental research work in understanding what data are saying us about the underlying process under study; on the other because it could provide us new insights and clues in situations where there is a lack of appropriate data and, of course, for deliver new data-inquiring methods or improving the existing ones. Some of the most important areas of application of complex system have become economics (to the point that nowadays we label the various physical approaches to economics by the name of *econophysics*), biology, sociology and, of course, statistical physics, in particular the areas of out-of-equilibrium thermodynamics and dynamical systems. In this thesis we will focus on the branch of complex systems which deals with complex networks. Network theory has become nowadays a central tool for analyzing large systems where many discrete entities interact by the means of different mutual relations. But the dealing with networks requires deep analytical tools to understand networks' behavior. The main point in theoretical modeling of networks is that, outside toy models, we face with real networks which dimension and complexity requires often a statistical approach. Here we encounter *Random matrix theory* (RMT in the sequel). The statistical

study of networks through random matrix theory can give us a lot of information about the network topology, and a sort of measure on the statistical dependence of the nodes in the network. As an invitation we will give a brief revision of the applications of RMT and complex networks in modern science.

1.1 Random Matrix Theory

Random matrix theory (RMT) has its beginning during the 1950s, when the spectra of large nuclei was investigated by Wigner and Dyson. The high complexity of the problem required a brand new approach. As is well know by any physicist (an hard truth to live with, though), analytical results are an exception rather than a rule. Often a statistical approach is the key recipe to tackle the problem. Transition probabilities in large random nuclei are well approximated by large hermitian matrices, where the entries are affected by such a noise that they can be regarded as random. Thus a study on the statistical behavior of the spectra of this ensemble of matrices can give precious insights on the overall behavior of the phenomenon under inquiry. In modern times RMT had a wide range of applications which covers both fundamental and applied physics but also pure mathematics problems. A good recap of different physical application of RMT can be found in [32]. In the realm of Quantum theory, RMT enters in the definition of Quantum Chaos. Indeed thanks to Bohigas-Giannoni-Schmit conjecture [11], chaotical quantum systems should have a spectra consistent with the one of Gaussian Orthogonal Ensemble. Random matrix theory has been found linked to plane tiling problems [37], and it didn't took too much that an application to 2-dimensional quantum gravity was found [19]. Application and connections also with Quantum Field Theory and Conformal field theory can be found in [32] chapter 7. Nevertheless we can affirm that, mostly, RMT theory has his widespread in statistical physics and complex system theory. We find applications of RMT in Anderson Localization [1], many-body systems [14] and other problems in solid state physics which deal with a mean field approach. From a pure mathematical point of view, RMT find application in number theory thank to Hilbert-Pólya conjecture [4], probability, plane tiling and Stochastic optimal control. The importance of RMT in the context of probability theory is self evident. A theory of non commuting random variables and their asymptotic behavior, which is called Free probability and

we will expose in section 4, is the natural evolution of classical probability theory, and has since the first 70s attracted many probabilists interested in stochastic processes, given the deep connection found in these two areas. RMT is also well used in interdisciplinary fields, like econophysics, finance, computer science and in general fields where a the system under study is big enough to deserve a statistical treatment. For example in [59] is exploited the universality result of RMT, Wigner semicircle law: examining the correlation matrices of stocks' price fluctuations one could separate the random noise components from the deterministic component using the methods presented in [13] to see the trend of the considered stocks.

Like often happen in subjects studied by academics coming from different backgrounds, RMT, and in particular the computation of the Empirical Spectral Distribution (ESD) can be approached by a lot of different methods. Traditionally, mathematicians developed rigorous methods, based on similarities with standard probability theory, while physicist tackled the problem with many powerful methods coming from different areas of physics; we can mention replica method [24], cavity method [56] and conformal field theory methods [40] for citing the main ones. The powerfulness of these methods, and the simplicity with which they get to the searched results is often faulted by lack of mathematical rigor, especially in situation where some kind of analytical continuation is performed.

1.2 Network Theory and Thesis Outline

The usages of complex networks are so numerous and destined to grow that is difficult to make an exhaustive review. One interesting reference text could be [52]. As already said networks can be used for modeling systems where discrete agents interact. The possible diversity of the type of interactions, of the agents and the topological relationship a network can be constructed with, give scientists an adaptable, flexible and feasible tool for analyzing real systems. Often we are interested to define a sort of dynamics over the networks, trying to simulate real processes going on real networks. In biological networks [**?**], for example, the interaction between different species are studied. Different parameters of the network correspond to real properties of the system that can be observed. For example a biological network constructed with a bipartite graph, should describe a situation where two different agents, i.e. species, interact only with agents of the other category. The connectiv-

ity of the graph represents how many interaction an individual has with the remaining environment. Often we are interested in global properties of the network, such as its resilience, i.e. its resistance and stability to defaults of some of its parts. This is a crucial aspect for example in the study of the *web* networks , *power grids*.

The principal aim of this thesis is to introduce, without pretending of being exhaustive, the setting in which a rigorous mathematical analysis of RMT can be developed, with the goal to analyze with rigorous mathematical tools random walks on complex networks. This kind of stochastic processes are particular studied nowadays, where a large number of entities (from humans to different kind of various system, like biological systems or IT frameworks) get in touch and interact with a large set of different objects and this interaction must be somewhat analyzed with some kind of model. In the recent years, complex networks, shed a light also in neural network theory, giving great contribution to the development of a consistent AI theory.

In chapter 2 we will set the general mathematical framework: the main results of probability theory we are going to use are presented, with a brief focus on stochastic processes. In chapter random matrix theory and the fundamental concept of concentration of measure are introduced, giving the main result of Wigner semicircular law. In chapter the mathematical setting of complex networks, random graph theory, is presented, we analyze two types of random graph, and apply the consequences of the spectral analysis of those to random walk defined on them, viewed as a special kind of Markov chain (section 2.3) on a graph.

Chapter 2

Stochastic processes: Markov chains and Brownian motion

Probability theory and its application to stochastic processes gives us powerful tools for make models concerning complex networks. Our aim in this chapter is to formalize the idea of Markov chain. Heuristically we can think a Markov chain like the formalization of a walk in which each step is stochastic but restricted under certain rules.

2.1 A brief revision of probability concepts

The axiomatization of probability theory throughout the use of measure theory is one of the most powerful and intriguing results of the early 20^{th} century. The main idea is to give a mathematical shape to the naive concept of an event related to the process we are studying (think about and experiment and its possible outcomes) and its probability. The setting is simple. We have a set Ω containing all the possible **outcomes** of our experiment. An **event** is a collection of outcomes, or, stated in math terms, a subset of Ω. The set of al possible events, $\mathcal{F}$, is a subset of the power set of Ω ($\mathcal{F} \subseteq 2^{\Omega}$) called $\sigma-$**algebra** or $\sigma-$**field**, is constrained to some necessary properties for permitting the basic operations with events like union or intersection avoiding problematical situations.

Definition 2.1. $\sigma-$***algebra***
A $\sigma-$***algebra*** $\mathcal{F}$ *of set* Ω *is a subset of the power set* 2^{Ω} *with the following properties:*

1. $\Omega \in \mathcal{F}$

2. *If* $\mathcal{B} \in \mathcal{F}$*, then the complement event* $\bar{\mathcal{B}} := \Omega / \mathcal{B}$ *is in* $\mathcal{F}$

3. *If* $\mathcal{I}$ *is a countable index set and for each* $\forall \iota \in \mathcal{I}$*,* $B_\iota \in \mathcal{F}$*, we have* $\bigcup_{\iota \in \mathcal{I}} \mathcal{B}_\iota \in \mathcal{F}$

Given the formal definition of the *space of events* is now possible to define a probability on it. This problem is equivalent to define probability measure on Ω compatible with the $\sigma-$algebra $\mathcal{F}$.

Definition 2.2. ***Probability Measure***
A ***probability measure*** $\mathbb{P}$ *is a function* $\mathbb{P} : \mathcal{F} \to [0, 1]$ *with the following properties:*

1. $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$

2. *If* $\mathcal{I}$ *is a countable index set and for each* $\forall \jmath, \iota \in \mathcal{I}$ *we have that* $\mathcal{B}_\iota, \mathcal{B}_\jmath \in \mathcal{F}$*,* $\mathcal{B}_\iota \bigcap \mathcal{B}_\jmath = \emptyset$*, then*

$$\mathbb{P}(\bigcup_{\iota \in \mathcal{I}} \mathcal{B}_\iota) = \sum_{\iota \in \mathcal{I}} \mathbb{P}(\mathcal{B}_\iota) \tag{2.1}$$

This definition of probabilty measure automatically satisfy **Kolmogorov Axioms**. For *non disjointed* sets we can restate (2) with the **union bound**, which will be used in the sequel

$$\mathbb{P}(\bigcup_{\iota \in \mathcal{I}} \mathcal{B}_\iota) \leq \sum_{\iota \in \mathcal{I}} \mathbb{P}(\mathcal{B}_\iota) \tag{2.2}$$

The triple of mathematical objects we have mantioned $(\Omega, \mathcal{F}, \mathbb{P})$ is itself an unique entity, the **probability space**, and it is the cornerstone upon which we will construct our theory. Another (often unmentioned) assumption we make is that the probability space, the sample space in particular, must be **extendible**. By this we mean that having two different probability spaces $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$, $(\Omega^2, \mathcal{F}^2, \mathbb{P}^2)$ of different experiments we always assume that exists a third probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that there exists a projection map $\pi_{1/2} : \Omega \to \Omega^{1/2}$ which is *surjective*, *measurable*[1] and *probability preserving*,

[1] Given two measurable spaces, i.e. $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{F}')$, a *measurable function* is a function $f : \Omega \to \Omega'$ such that $f^{-1}(E \in \mathcal{F}') \in \mathcal{F}$, where with f^{-1} we denote the *preimage*.

i.e. $\mathbb{P}(\pi_{1/2}^{-1}(\mathcal{B} \in \mathcal{F}^{1/2})) = \mathbb{P}_{1/2}(\mathcal{B})$. See further details in [65]. We can now define another main object of probability theory, the **random variable**:

Definition 2.3. ***Random Variable***
Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a ***random variable*** *(r.v.) taking values in a measurable space $(R, \mathcal{R})$ is a measurable function $X : \Omega \to R$.*

In view of application to stochastic processes, we will often call $(R, \mathcal{R})$ **state space**. Given a r.v. taking values in $(R, \mathcal{R})$ and a$\mathcal{B} \in \mathcal{F}$, we will denote with a little abuse of notation the probability of the event $\{\omega \in \Omega : X(\omega) \in \mathcal{B}\}$ with $\mathbb{P}(X \in \mathcal{B})$. This probability takes the role of the probability measure naturally induced on $(R, \mathcal{R})$ by X and it is called the **distribution** μ_X of X (thus we have $\mu_X(\mathcal{B} \in \mathcal{R}) = \mathbb{P}(X \in \mathcal{B})$). This probability, for each measurable $\mathcal{B}$, can be expressed using the *Radon-Nikodym theorem* can be expressed by

$$\mu_X(\mathcal{B}) = \int_{\mathcal{B}} f dm$$

where dm is a reference $\sigma-$*finite*[2] measure (for example Lebesgue measure if $R \equiv \text{Re}$) defined on R and $f \in L^1(R, dm)$ such that $\int_R f dm = 1$. We have then that $d\mu_X = f dm$. sometimes f is called **density function** if dm is a Lebesgue measure (see [22] for further details). We will work in the next pages with several random variables. It is convenient sometimes to work with r.v. with the same distribution (*identically distributed* or in short *i.d.*). If X and Y share a common distribution μ, we will write $X \equiv Y$. A set of r.v. which are independent and identical distributed will be denoted by the acronym *i.i.d.*. Is it also possible to define random variables through algebraic operation on other r.v.[3]. The case we are interested in is the one of the sum of two r.v. $Z = X + Y$ with distribution μ and ν respectively. The distribution ζ of Z is the convolution of the two distributions, $\zeta = \mu * \nu$. Thanks to Radon-Nikodym theorem, we can express this rule with the density functions $g(x)$ of μ and $f(x)$ of ν

$$h(x) = (f * g)(x) = \int f(x - y)g(y)dy \tag{2.3}$$

It is now possible for us to define what is the *most likely* outcome for a random variable X. This notion is called **expected value** $\mathbb{E}X$ (often called

[2] A measurable space $(\Omega, \mathcal{F})$ with measure μ is said $\sigma-$finite if there exist a sequence of $\mathcal{B}_\iota \in \mathcal{F}$ with $\iota \in \mathcal{I}$ countable, such that $\forall \iota, \mu(\mathcal{B}_\iota) < \infty$ and the $\mathcal{B}_i$s cover Ω, $\bigcup_{\iota \in \mathcal{I}} \mathcal{B}_\iota = \Omega$

[3] See [61] as a reference.

mean), and is performed with an integration. The general formula is

$$\mathbb{E}X := \int_\Omega X(\omega) d\mathbb{P}(\omega) \tag{2.4}$$

If X takes values in the reals $\mathbb{R}$, endowed with a $\sigma-$algebra $\mathcal{R}$, $(\mathbb{R}, \mathcal{R})$, we can rewrite this integral; if $X(\omega) > 0, (\textit{unsigned r.v.})\ \forall \omega \in \Omega$ we have

$$\mathbb{E}X = \int_0^\infty x d\mu_X(x)$$

Otherwise (*signed r.v.*), if $X \in \text{Re}$ we can take positive and negative part of X and then summing the expected values $\mathbb{E}X = \mathbb{E}X^+ + \mathbb{E}X^-$, if the single expectation values are finite. This is given for free by the definition of the expected value itself, inheriting the properties of the relative integral, like linearity. In general it will be sufficient in the rest of the text when dealing with expected values in the presence of a signed r.v. to substitute X with $|X|$. A concept that will be fundamental in the sequel is the $k^{\textbf{th}}$ **moment** of a random variable X, defined as $m_k(X) := \mathbb{E}X^k$, if the absolute integrability condition $\mathbb{E}|X|^k < \infty$ holds. For $k = 1$ we have the above expected value, called also **mean**. An important quantity related to the 2^{nd} moment is the **variance**

$$\mathbf{Var}(X) := \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

For a centered r.v. (i.e. with mean 0) the variance is identical to the 2^{nd} moment $\mathbf{Var}(X) = \mathbb{E}X^2$. We also mention for later convenience

- the **exponential moments** $\mathbb{E}e^{t \cdot X}$, where $t \cdot X$ is an inner product defined on the state space of X taking values on the reals.
- the **characteristic function** $\mathbb{E}e^{it \cdot X}$, where $t \cdot X$ is as above.
- the **resolvent** $\mathbb{E}\frac{1}{X - z}$ with complex z.

Moments are useful for bounding probabilities. Indeed we have using first moment **Markov's inequality**

$$\mathbb{P}(X \geq \lambda) \leq \frac{1}{\lambda}\mathbb{E}X \tag{2.5}$$

remembering to sobstitute X with $|X|$ in case of a signed r.v. Using the variance (thus the second moment) we then have **Chebyshev's inequality**

$$\mathbb{P}(|X - \mathbb{E}X| \geq \lambda) \leq \frac{\mathbf{Var}(X)}{\lambda^2} \tag{2.6}$$

Another important concept is that of **independece**. We say that two events are independent if

$$\mathbb{P}(\mathcal{B} \bigcap \mathcal{E}) = \mathbb{P}(\mathcal{B})\mathbb{P}(\mathcal{E})$$

Extending this to r.v. thanks to the relations between $\mathbb{P}$ and μ_X, we say that two r.v. X and Y are independent if

$$\mathbb{P}(X \in \mathcal{B}, Y \in \mathcal{E}) = \mathbb{P}(X \in \mathcal{B})\mathbb{P}(Y \in \mathcal{E}) \iff \mu_{(X,Y)} = \mu_X \times \mu_Y$$

this is equivalent to say that the **joint distribution**[4] is the **product measure** of the single variables. More generally, if we have a family of r.v. $\{X_\alpha\}_{\alpha \in \mathcal{J}}$, that could be even uncountable, and taking values on R_α, is said, jointly independent if the joint distribution coincides with the product measure. We can have in the family $\{X_\alpha\}_{\alpha \in \mathcal{J}}$ a weaker condition of independence. If there is a max k for which for every subsequence in $\mathcal{J}$ of k elements we have that $\{X_{\alpha_1} \dots X_{\alpha_k}\}$ are independent, we say that the family is **k-wise independent**. It is a result of probability theory that given a probability space and a distribution one can create a set of independent random variables with the same distribution (i.i.d.). An important result in this aim is **Kolmogorov extension theorem** that will be discussed in the next section because of its centrality in random processes' existence. Another important issue is the asymptotic behavior of events' probability, in particular when we are dealing with events connected to random variables. It is often the case that a r.v. X depends on an index n, and thus the behavior of $X = X_n$, and the probability of the occurrence of events related to it (e.g. the event $\mathcal{F} \ni \mathcal{B} = X \in \mathcal{E}$ with $\mathcal{E} \in \mathcal{R}$), are linked to the values of the parameter. We are often interested in the asymptotic behavior of these events $\mathcal{B}_n$, i.e. their probability to hold given $n \to \infty$[5]

- An event $\mathcal{B}_n$ holds **surely** is its complement is *equal* to the null event $\emptyset$.
- An event $\mathcal{B}_n$ holds **almost surely** if $\mathbb{P}(\mathcal{B}_t) = 1$, i.e. the complement event $\overline{\mathcal{B}_n}$ is set with null measure.
- An event $\mathcal{B}_n$ holds with **overwhelming probability** if $\forall \alpha$ fixed, it holds with probability $1 - O_\alpha(n^{-1\alpha})$.

[4] i.e. the distribution of the single random variable formed by the couple (X,Y)

[5] We will use the familiar notations $f(x) = O(g(x))$ and $f(x) = o(g(x))$ for respectively $|f(x)| \leq Cg(x)$ and $|f(x)| \leq c(x)g(x)$ with $x \to \infty$ and constant $C \in \mathbb{R}$ and $c(x) \to 0$. If the constant C depends on some parameter α we will write O_α. See **??** for reference.

- An event $\mathcal{B}_n$ holds with **high probability** if it holds with probability $1 - O(n^{-c})$, for some c independent from t.
- An event $\mathcal{B}_n$ holds **asymptotically almost surely** if it holds with probability $1 - o(1)$.

These notations are listed in decreasing order of certainty, i.e. the one below is implied by the one above. Another fundamental question that will be central in 3 is about the convergence of r.v. towards a certain distribution. We have as before a set of r.v. depending on a parameter n that goes to infinity, and we ask whether the set of variables X_n and their distribution converge to a limit r.v. X with a specific distribution. Let work with a set of variables X_n and X taking values on a state space $(R, \mathcal{R})$ that is taken to be $\sigma-$compact[6] metric space with distance d and $\mathcal{R}$ Borel $\sigma-$algebra . These are different types of convergence, ordered in decreasing strength:

- X_n converges **almost surely** to X if

$$\mathbb{P}(\limsup_{n\to\infty} d(X_n, X) \leq \epsilon) = 1 \quad \forall \epsilon > 0 \tag{2.7}$$

- X_n converges **in probability** to X if

$$\liminf_{n\to\infty} \mathbb{P}(d(X_n, X) \leq \epsilon) = 1 \quad \forall \epsilon > 0 \tag{2.8}$$

- X_n converges **in distribution** to X if

$$\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)) \tag{2.9}$$

  where $f : R \to \mathbb{R}$ is a bounded function.

The next step will be conditioning our knowledge on the events. Often we dispose of data in our experiments that allow us to restrict the sample space. Let say that we know that on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we know that and event $\mathcal{E} \in \mathcal{F}$ holds. Then the probability of a generic event $\mathcal{B}$ must be restricted to $\mathcal{B} \bigcap \mathcal{E}$, and the resulting probability becomes

$$\mathbb{P}(\mathcal{B} \mid \mathcal{E}) = \frac{\mathbb{P}(\mathcal{B} \bigcap \mathcal{E})}{\mathbb{P}(\mathcal{E})}$$

[6] A $\sigma-$compact topological space is the countable union of compact subspaces.

The expected value of a random variable is straightforward[7]: we have to consider the new random variable $(X \mid \mathcal{E})$ and the related distribution

$$\mu_{(X|\mathcal{E})}(\mathcal{B}) = \frac{\mathbb{P}(\{X \in \mathcal{B}\} \bigcap \mathcal{E})}{\mathbb{P}(\mathcal{E})}$$

Often, the notation used for this kind of conditioning is

$$\mathbb{E}(X; \mathcal{E}) = \mathbb{E}(X\mathbf{1}_{\mathcal{E}}) = \int_{\mathcal{E}} X d\mathbb{P}$$

where $\mathbf{1}_{\mathcal{B}}$ is the indicator function of $\mathcal{B}$. Generalizing, if we have a collection of events, or more formally, a sub $\sigma-$algebra $\widetilde{\mathcal{F}} \subseteq \mathcal{F}$ we can conditioning in the same way, defining in a similar fashion $\mathbb{P}(\mathcal{B} \mid \widetilde{\mathcal{F}})$ and $\mathbb{E}(X \mid \widetilde{\mathcal{F}})$ working with the events $\mathcal{I} \in \widetilde{\mathcal{F}}$. For settling the thing we can say:

Definition 2.4. ***Conditional Expectation***
Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a r.v. X taking values in the state space $(R, \mathcal{R})$ we define the conditional expectation of X conditioned by the sub $\sigma-$algebra $\widetilde{\mathcal{F}} \subseteq \mathcal{F}$ the r.v. $Y \equiv \mathbb{E}(X \mid \widetilde{\mathcal{F}})$ with the following properties:

1. *Y is $\widetilde{\mathcal{F}}-$measurable.*
2. *For all $\mathcal{B} \in \widetilde{\mathcal{F}}$ we have*

$$\int_{\mathcal{B}} X d\mathbb{P} = \int_{\mathcal{B}} Y d\mathbb{P}$$

It follows directly by the definition that the quantity $\mathbb{E}(\mathbb{E}(X \mid \widetilde{\mathcal{F}}))$ is nothing but $\mathbb{E}(X)$ itself. In fact by the properties of conditional expectation, given two sub $\sigma-$algebra $\mathcal{F}^1 \subseteq \mathcal{F}^2 \subseteq \mathcal{F}$ we have[8]

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{F}^2) \mid \mathcal{F}^1) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}^1) \mid \mathcal{F}^2) = \mathbb{E}(X \mid \mathcal{F}^1)$$

taking $\mathcal{F}^1 = \{\emptyset, \Omega\}$ we get the searched equality. We have seen that conditioning is strongly built over $\sigma-$algebras and sub $\sigma-$algebras. It is thus important thus to identify the main sub $\sigma-$algebras which we will have to deal with. Given a family of subsets $\mathcal{A} = \{\mathcal{B}_\iota\}_{\iota \in \mathcal{I}}$ of Ω we will denote by

[7]Here it is necessary a little remark: while for discrete r.v. this concept is trivial, for continuous one we need a construction called **disintegration**. See [38] for further details.

[8]See [22], theorem 5.1.6, page 228

$\sigma(\mathcal{A})$ the smallest $\sigma-$algebra containing $\mathcal{A}$. We will indicate with $\sigma(X)$ the $\sigma-$algebra generated by a r.v. X on the state space $(R, \mathcal{R})$ given by the collection of preimages $\{X^{-1}(\mathcal{B}) : \mathcal{B} \in \mathcal{R})\} \subseteq \mathcal{F}$. If we have a sequence of random variables $\{X_1,\ X_2,\ X_3 \ldots\}$ we denote by $\mathcal{F}_n = \sigma(X_1,\ X_2,\ X_3 \ldots X_n)$ the $\sigma-$algebra generated by the first n r.v. This notion will be important in the study of stochastic processes, where it will represent the information known untill the n^{th} step of the process. Turning back to expectations, given two r.v. X and Y, we define $\mathbb{E}(X \mid Y)$ to be the r.v. by the $\sigma-$algebra $\sigma(Y)$, $\mathbb{E}(X \mid Y) := \mathbb{E}(X \mid \sigma(Y))$. It is important to remark that the $\sigma-$algebra on which we make the conditioning is the information available at the moment. Thus it is intuitive that under the constraint of existence $\mathbb{E}|Y|, \mathbb{E}|YX| < \infty$ with $X \in \widetilde{\mathcal{F}}$ we have that $\mathbb{E}(XY \mid \widetilde{\mathcal{F}}) = X\mathbb{E}(Y \mid \widetilde{\mathcal{F}})$.

Another important topic, that will be the principal goal in chapter 3 is the estimate on the tail of the distribution of a random variable X. This kind of estimate is obtained through the exploitation of inequalities such as (2.5) and (2.6). This is the first example in which we highlight the importance of the knowledge of various moments $\mathbb{E}X^k$ of our distribution in order to estimating some specific behavior. This kind of reasoning will be central in the sequel. These types of tail characterization are listed in decreasing order of strenght.

- X is said to be **surely bounded** (**almost surely bounded**) if there exists a costant $C > 0$ such that $|X| \leq C$ surely (almost surely).
- X is said to be **sub-Gaussian** if $\exists C, c > 0$ constants such that

$$\mathbb{P}(|X| \geq \lambda) \leq Ce^{-c\lambda^2} \quad \forall > 0 \tag{2.10}$$

- X is said to have **sub-exponential tail** if $\exists C, c, a > 0$ such that

$$\mathbb{P}(|X| \geq \lambda^a) \leq Ce^{-c\lambda^2} \quad \forall > 0 \tag{2.11}$$

- X is said to have **finite k^{th} moment** if $\exists C > 0$ such that

$$\mathbb{E}(|X|^k) \leq C \tag{2.12}$$

  for some $k \geq 1$.

- X is said to have **absolutely integrable** if $\mathbb{E}|X| < \infty$.

2.2 Stochastic processes

The principal goal of a stochastic process is to formalize the description of a random process which depends on an evolution parameter. For obvious reasons this evolution parameter is mostly taken as time. Thus we have random variables of the form $X(t,\omega) := X_t(\omega)$

Definition 2.5. ***Stochastic Process***
Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a state space endowed with a sigma algebra (S, Σ) and an ***index set***[9] *T, a* ***stochastic process*** *is a measurable function*

$$X : \Omega \times T \to S$$

The set of all the values at any time of the stochastic process will be noted by S^T.

The first issue with stochastic processes is their existence. We have already mentioned that, given a known finite set of probability distributions μ_α, is a basic result of probability theory to create a finite number of independent r.v. But how does it work whet it comes to work with a possible *uncountable* set of r.v., how we can construct their probability space? The next theorem will be of fundamental importance:

Theorem 2.2.1. ***Kolmogorov Extension Theorem***
Given an arbitrary index set T and a collection of measurable Hausdorff topological spaces $\{\Omega_t, \mathcal{F}_t\}_{t\in T}$ endowed with topology τ_t. For each finite subset $\mathcal{I} \subset T$ let's define $\mu_\mathcal{I}$ an inner regular probability measure on the product sigma field $\mathcal{F}_\mathcal{I} = \prod_{t\in\mathcal{I}} \mathcal{F}_t$ with respect to the product topology $\tau_\mathcal{I} = \prod_{t\in\mathcal{I}} \tau_t$. We further require that for each finite subsets $\mathcal{K} \subset \mathcal{I} \subset T$ $\mu_\mathcal{K}$ and $\mu_\mathcal{I}$ should be ***compatible****: given the projection map $\pi : \mathcal{I} \to \mathcal{K}$ we must have that the pushforward measure*[10] *$\pi_*(\mu_\mathcal{I}) = \mu_\mathcal{K}$. Then there exists a unique probability measure μ_T on $\mathcal{F}_T$ such that, for all finite subgroups $\mathcal{I} \subset T$, we have*

$$(\pi_\mathcal{I})_* \mu_T = \mu_\mathcal{I}$$

[9] The index set T will be mostly a subset of the real line, such as an interval or the set of natural numbers.

[10] Given two measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ with measure μ on $(\Omega_1, \mathcal{F}_1)$ and a map $\pi : \Omega_1 \to \Omega_2$ we define the pushforward measure $f_*(\mu)$ the measure such that

$$f_*(\mu)(\mathcal{B} \in \mathcal{F}_2) = \mu(f^{-1}(B))$$

Kolmogorov extension theorem is a powerful result of measure theory, an its general form requires some abstraction to deal with generic state spaces. We will use a simplified version keeping in mind that, usually, we work with state spaces of the type $\mathbb{R}^n$ and with a discrete index set. In this situation the types of measure we are dealing with are probability measures on $(\mathbb{R}^n, \mathcal{R}^n)$ that thanks to Radon-Nicodym theorem can always be described in terms of Lebesgue measures. The compatibility criterion now states, for a collection $\mathcal{B}_1 \cdots \mathcal{B}_n$ of elements of $\mathcal{R}^n$

$$\mu_{n+1}(\mathcal{B}_1 \times \cdots \times \mathcal{B}_n \times \mathbb{R}) = \mu_n(\mathcal{B}_1 \times \cdots \times \mathcal{B}_n)$$

Further material on Kolmogorov extension theorem can be finded in [22, 38, 64]. In a topological space, a **Borel set** is any set that can be formed through the of countable union or intersection and complement (relative or absolute) of the open sets defined by the topology. The particular sigma algebra formed by Borel sets is called Borel algebra, which ultimately is a $\sigma-$algebra compatible with the topology. The particular class of product sigma fields associated with a specific product topology forms a particular kind of mathematical object called **cylinder $\sigma-$algebra** formed by the underlying topological cylinder sets. As this concept is useful to define the $\sigma-$algebra generated by a stochastic process and gives a measure-theorethical definition of stochastic process, we describe it here. We start by identifying the state space of our whole stochastic process X_t with $\mathbb{R}^T$, the space of real valued function from T given by $\{X_t(\omega)\}_{t\in T}$, sometimes called product state space. Considering the relative Borel $\sigma-$algebra $\mathcal{B}(\mathbb{R}^n)$ we can suppose that our stochastic process takes values within $\mathbb{R}^T \supset \mathcal{E} = \{X(\cdot, \omega) : X(t, \omega) \in \mathbb{R}, t \in T, \omega \in \Omega\}$ the set of values that our process can take in the product state space $\mathbb{R}^T$. Given a finite index set $\mathcal{I}$, a cylinder subset of $\mathcal{E}$ is defined as

$$C_{t_1,\cdots,t_n}(\mathcal{B}_1, \cdots, \mathcal{B}_n) = \{X(\cdot, \omega) \in \mathcal{E} : X(t_i) \in \mathcal{B}_i, i \in \mathcal{I}, \mathcal{B}_i \in \mathcal{B}(\mathbb{R}), \omega \in \Omega\}$$

The collection of these sets keeping fixed the index set and varying the $\mathcal{B}_\iota$s in $\mathcal{B}(\mathbb{R})$ forms a $\sigma-$algebra $\Sigma_\mathcal{I}$. Varying the index set in T and taking the product we get the cylinder $\sigma-$algebra $\mathcal{F}_\mathcal{E} = \sigma(\prod_{\mathcal{I}\subseteq T} \Sigma_\mathcal{I})$. We can now see a stochastic process like a measurable function between the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the product state space, or one of its restrictions, $(\mathcal{E} \subset \mathbb{R}^T, \mathcal{F}_\mathcal{E})$ endowed with the cylinder $\sigma-$algebra $X : \Omega \to \mathcal{E}$. The sigma algebra generated by a stochastic process is now

$$\sigma(X) = \{X^{-1}(\mathcal{B}) : \mathcal{B} \in \mathcal{F}_\mathcal{E}\}$$

Usually the knowledge of the entire process is precluded, because of its evolution which follows a random behavior (e.g. $\forall t$, $X_t(\omega)$ could follow a particular distribution). Indeed one of the goals of the study of stochastic processes is to infer the evolution of the process itself using its history till the time t. Thus what is required is, it is possible, a restriction on the values of $X_{t+\delta}(\omega)$ ($X_{n+1}(\omega)$ for discrete and ordered index set) given $X_t(\omega)$ (respectively $X_n(\omega)$). We immediately highlight that if the set of $X_t(\omega)$ in T has some property of independence, e.g. joint independence, the values of $X_{t+\delta}(\omega)$ (resp. $X_{n+1}(\omega)$) will not depend at all, or weakly for weaker independence conditions, from the previous values $X_t(\omega)$ (resp. $X_n(\omega)$). In these cases what is studied is the overall behavior, studying properties such as weather or not the process will pass near a point, if a value is likely to recur more than once or neither once. Considering a processes evolving in t, the information available at a specific time $\tilde{t}$ is given by the $\sigma-$algebra induced by all the X_σ for $\sigma \leq \tilde{t}$. In the case of a discrete time index set $T = \{1, 2, \cdots\}$ this coincides with the $\sigma-$algebra $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ that we have described at the end of the previous section. In this optic will be important also the concept of **filtration**:

Definition 2.6. ***Filtration***
*Given a $(\Omega, \mathcal{F}, \mathbb{P})$ and an arbitrary index set $\mathcal{I}$, let be, $\forall \iota \in \mathcal{I}$, $\mathcal{F}_\iota$ a sub $\sigma-$algebra of $\mathcal{F}$ with the property that $\forall \iota, \jmath \in \mathcal{I} : \iota < \jmath \Rightarrow \mathcal{F}_\iota \subseteq \mathcal{F}_\jmath$. The family $(\mathcal{F}_\iota)_{\iota \in \mathcal{I}}$ is called **filtration**. A probability space endowed with a filtration is said **filtered probability space** $(\Omega, \mathcal{F}, (\mathcal{F}_\iota)_{\iota \in \mathcal{I}}, \mathbb{P})$. A process X_t with values in $(R, \mathcal{R})$ is said **adapted** to the filtration $(\mathcal{F}_\theta)_{\theta \in T}$, with $\mathcal{I} \equiv T$, if $\forall t \in T$ we have that $X_t : \Omega \to R$ is measurable with respect to the couple $(\mathcal{F}_t, \mathcal{R})$.*

It is straightforward that a stochastic process X_t comes with a *natural* filtration given by $(\mathcal{F}^X_\theta)_{\theta \in T}$ whose elements $\mathcal{F}^X_\theta, \theta \in T$, are defined by $\mathcal{F}^X_\theta = \{\sigma(X_\tau) : \tau \leq \theta\}$ which is called indeed **natural** or **induced filtration**. Given a deterministic time $t \in T$, it is natural to define the set of events w.r.t. the filtration that are the future ($\mathcal{F}^+_t = \bigcap_{\theta > t} \mathcal{F}_\theta$) and past events ($\mathcal{F}^-_t = \bigcap_{\theta \leq t} \mathcal{F}_\theta$) at t.

Sometimes the quantity of information we have on the stochastic process depends on some quantities which are linked with the flow of the time parameter t in X_t and act like a trigger on the measuration of the properties of our system. This idea is that of **stopping time** or **stopping rule**:

Definition 2.7. ***Stopping time***

Given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_\theta)_{\theta\in T}, \mathbb{P})$ *and a index set* T*, a* ***stopping time*** τ *is a random variable with values in* $\overline{T} = T \cup \sup(T)$ *such that*

$$\{\tau \leq t\} \in \mathcal{F}_t, \ \forall t \in T$$

We also define the $\sigma-$*algebra of the past and future events associated with the stopping time* τ

$$\mathcal{F}_\tau^- = \{\mathcal{B} \in \mathcal{F} : \mathcal{B} \cap \{\tau = \theta\} \in \mathcal{F}_\theta^-, \forall \theta \in T\}$$
$$\mathcal{F}_\tau^+ = \{\mathcal{B} \in \mathcal{F} : \mathcal{B} \cap \{\tau = \theta\} \in \mathcal{F}_\theta^+, \forall \theta \in T\}$$

Euristically, the stopping time is a r.v. which entails some specific information at a time $\tilde{t}$ on the process that does not need the knowledge of events future with respect to $\tilde{t}$. If my stochastic process is the toss of a coin, we could define the stopping time τ as the time when we have a certain number n of heads, for example. This time does not entail any information about the future, but only on the past of the stochastic process. For a discrete time index set $T = \mathbb{N} \cup \{\infty\}$ the stopping time could only have an integer value in $1-1$ correspondence with the indexing of the stochastic process $\{X_1, X_2, \cdots, X_n\}$, and thus the event $\{\tau = n\}$ is conteined in the $\sigma-$algebra generated by the first n r.v. of the stochastic process $\{\tau = n\} \in \sigma(X_1, X_2, \cdots, X_n)$. An example of stopping time could be the *hitting time of a set* $\mathcal{B}$, the smallest time $\tilde{t}$ at which the stochastic process hits the set $\mathcal{B} \subset R$ in the state space

$$\{\tau = \sigma\} = \{X_{t<\sigma} \in \mathcal{B}^c, X_\sigma \in \mathcal{B}\} \in \mathcal{F}_\sigma$$

we are now ready to introduce some stochastic processes.

2.3 Markov processes and Markov chains

Markov processes, and in particular Markov chains, will be fundamental in their application to random graphs, networks and related random walks. A **Markov process** is a stochastic process for which the following property holds:

Definition 2.8. ***Markov property***
Given an adapted stochastic process X_t *on the filtrated probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_\theta)_{\theta\in T}, \mathbb{P})$ *taking values on the measurable state space* $(R, \mathcal{R})$ *is said to be* ***markovian*** *or to have the* ***Markov property*** *if* $\forall \mathcal{B} \in \mathcal{R}$, $\forall \theta, \sigma \in T$ *with* $\theta > \sigma$ *we have:*

$$\mathbb{P}(X_\theta \in \mathcal{B} \mid \mathcal{F}_\sigma) = \mathbb{P}(X_\theta \in \mathcal{B} \mid X_\sigma) \tag{2.13}$$

We note that we could have used the natural filtration $(\mathcal{F}_\theta^X)_\theta \in T$, respect to which the stochastic process X_t is *naturally* adapted. In substance a Markov process is a stochastic process in which the future depends only on the present (or most recent) value of the r.v. X_t. Another way to state Markov property is through expectation, thanks to (2.4): given a bounded and measurable

$$\mathbb{E}(X_\theta \in \mathcal{B} \mid \mathcal{F}_\sigma) = \mathbb{P}(X_\theta \in \mathcal{B} \mid X_\sigma) \tag{2.14}$$

A fundamental role in Markov processes is held by **transition function** or **transition probability**

$$p(\cdot, \cdot, \cdot) : T \times R \times \mathcal{R} \to [0, 1]$$

which, gives the probability for passing from a value $X_t \in R$ to a set of values $\mathcal{B} \in \mathcal{R}$ after a time δ, otherwise stated the probabilty of $\mathbb{P}(X_{t+\delta} \in \mathcal{B} \mid X_t)$. Transition probability are indicated by

$$p(\delta, X_t, \mathcal{B}) = \mathbb{P}(X_{t+\delta} \in \mathcal{B} \mid \mathcal{F}_t^X) = \mathbb{P}(X_{t+\delta} \in \mathcal{B} \mid X_t) \tag{2.15}$$

in which the last equality follows from the Markov property. From now on for the sake of clarity and also for important implication on the applications we assume that our Markov process has an origin and no end, $T \equiv [0, \infty)$. These functions are subject to some properties:

1. $p(t, x, \cdot)$ for $t \in T$, $x \in R$, is a *measure probability* on $(R, \mathcal{R})$
2. $p(0, x, \cdot) = \delta_x$ is a *point measure*, $x \in R$
3. $p(\cdot, \cdot, \mathcal{B})$ is $(T \times R)$-measurable
4. (*Chapman-Kolmogorov equation*)

$$p(t + s, x, \mathcal{B}) = \int p(s, y, \mathcal{B}) p(t, x, dy) \tag{2.16}$$

The equation (2.16) has an intuitive interpretation: the probability for getting to a state $X_{t+s} \in \mathcal{B}$ from a state $X_t = x \in R$ is obtained integrating over all intermediate states $y \in R$ which are sent to $\mathcal{B}$ by another transition function[11]. We can see through the properties of the integral in (2.16) that transition functions are *time homogeneous*. Thanks to (2.15) the whole

[11] Equation (2.16) entails the **semigroup** property of transition functions.

Markov process is *time homogeneous*[12] There is a straightforward extension of Markov property. We can indeed extend this independence of Markov process when the sigma algebra which entails the information is the one generated by a stopping time.

Definition 2.9. ***Strong Markov property***
Let be X_t a stochastic process like before, and let τ be a stopping time. For each value of the stopping time, conditioning on the event $\{\tau < \infty\}$, we say that the process has the ***strong Markov property*** *if $\forall t$:*

$$\mathbb{P}(X_{\tau+t} \in \mathcal{B} \mid \mathcal{F}_\tau^-) = \mathbb{P}(X_{\tau+t} \in \mathcal{B} \mid X_\tau) \tag{2.17}$$

with $\mathcal{F}_\tau^-$ is the past sigma algebra w.r.t. stopping time τ. Stated with transition functions we have

$$\mathbb{P}(X_{\tau+t} \in \mathcal{B} \mid \mathcal{F}_\tau^-) = p(t, X_\tau, \mathcal{B}) \tag{2.18}$$

The formulation with expectations is straightforward.

If the stopping time τ is a discrete r.v. the strong Markov propertiy follows directly from the standard one[13].

Once one is given with the **initial distribution** μ in $(R, \mathcal{R})$, i.e. the probability distribution $\mu(\mathcal{B}) = \mathbb{P}(X(0) \in \mathcal{B})$, with some regularity condition and a realization of the transition function $p(t, x, \mathcal{B})$, we have the whole distribution for a finite sample of X_t:

Theorem 2.3.1. ***Finite dimensional distribution for Markov process***[14]
Let be μ an initial distribution on $(R, \mathcal{R})$, and $p(t, x, \mathcal{B})$ a transition function like above. Then if $\int p(t, x, \cdot) d\mu(x)$ is ***tight***[15] *there a Markov process*

[12] For the sake of clarity (2.15) is self-given when we deal with Markov processes in which there is no difference the two transitions $(X_t = x \to X_{t+\delta} = y)$ and $(X_\theta = x \to X_{\theta+\delta} = y)$ which are called indeed *time homogeneous* Markov processes and transitions depends only on initial and final state $(x \to y)$ and on the interval δ in which the transition takes act. Because we deal only with Markov processes of this kind we take (2.15) as a definition.

[13] See [27], *proposition 4.1.3*

[14] See [27], *theorem 4.1.1* for the demonstration.

[15] A measure μ on a space Ω is said to be **tight** if $\forall \epsilon, \exists \mathcal{B} \subset \Omega : \mu(\mathcal{B}) = 1 - \epsilon$, with $\mathcal{B}$ compact in Ω.

on $(R, \mathcal{R})$ *with transition probability* $p(t, x, \mathcal{B})$ *and initial distribution* μ *finite* $n-$*dimensional distribution given by:*

$$\mathbb{P}(X_0 \in \mathcal{B}_0, X_{t_1} \in \mathcal{B}_1, \cdots, X_{t_n} \in \mathcal{B}_n) = \int_{\mathcal{B}_0} \cdots \int_{\mathcal{B}_{n-1}} p(t_n - t_{n-1}, y_{n-1}, \mathcal{B}_n) p(t_{n-1} - t_{n-2}, y_{n-1}, dy_{n-1}) \cdots p(t_1, y_0, dy_1) d\mu(y_0) \tag{2.19}$$

We will denote the pro probability distribution (2.19) with $\mathbb{P}_\mu$, in particular for $\mu \equiv \delta_x$ we will write $\mathbb{P}_x$ and with $\mathbb{E}_\mu$, $\mathbb{E}_x$ the respective expectations. Using this probability and the time homogeneity for Markov processes we have that

$$\mathbb{P}_\mu(X_{t+\delta} = x \mid X_t = y) = \mathbb{P}_y(X_\delta = x) = p(\delta, x, y) \tag{2.20}$$

for $x, y \in R$ and a time shift δ. We have handled until now the most general of Markov processes, those with continuous index set T and possibly uncountable state space $(R, \mathcal{R})$. For our applications will be enough restricts our scope to Markov process with a discrete but possible infinite index set $T = \mathbb{Z}^+$ and a discrete, countable or even finite, state space $R = \{i, j, k \cdots\}$. These particular Markov processes are called **Markov chains**. A Markov chain X consists of a series of r.v. at discrete timese $X = \{X_0, X_1, X_2, \cdots, X_n, \cdots\}$. The transition functions now become thanks to time homogeneity and the equal spacing of time increments and using (2.20):

$$p_{ij} = \mathbb{P}_i(X_1 = j) = \mathbb{P}_\mu(X_{n+1} = j \mid X_n = i) \tag{2.21}$$

while the probability measure μ becomes a discrete probability measure. The (2.16) becomes

$$p_{ij} = \sum_{k \in R} p_{ik} p_{kj}$$

and the equation (2.19) becomes

$$\mathbb{P}_\mu(X_0 = i, X_1 = j, X_2 = k \cdots, X_n = m) = \mu(i) p_{ij} p_{jk} \cdots p_{lm} \tag{2.22}$$

We can thus define the **transition matrix** $\boldsymbol{P} = (p_{ij})_{i,j \in R}$, such that $\boldsymbol{p}_i = (p_{ij})_{j \in R}$ is a **probability vector**, i.e. $\sum_i \boldsymbol{p}_i = 1$: this is the simple observation that given a state j the probability of getting to j from all the other

states must add up to 1. In chapter 5 and in general in statistical physics contexts, $\boldsymbol{P}$ is called **stochastic matrix**; moreover if the rows of $\boldsymbol{P}$ sum up to 1, $\sum_j p_{ij}$, then it is called **doubly stochastic**. Given a Markov chain $X = \{X_0, X_1, X_2, \cdots, X_n, \cdots\}$ we can extract a Markov subchain Y of fined fixed step m, such that $Y = \{Y_0 = X_0, Y_1 = X_m, Y_2 = X_{2m} \cdots, Y_n = X_{nm}, \cdots\}$; it's easy to see that the transition matrix for Y is $\boldsymbol{P}^m$, namely the m^{th} power of the one of X.

With each state i is possible to define a hitting time variable $\tau(i)$ defined such that

$$\tau(i) = \inf\{n \geq 1 : X_n = i\}$$

It's costumary to set $\tau(i) = \infty$ if such n doesn't exists. Let be $\mathcal{N}_i = \sum_{n=1}^{\infty} 1_{X_n=i}$ the number of visits to the site i. We have that the two events $\{\tau(i) < \infty\}$ and $\mathcal{N}_i > 0$ coincide. With $\tau^m(i)$ we indicate the time m^{th} visit at t. Clearly $\tau^m(i) > \tau^{m-1}(i)$, and we can define it by recurrence or time homogeneity setting $X_{\tau^{m-1}(i)}$ as the new starting point. We will call the state i **recurrent** if $\mathbb{P}_i(\tau(i) < \infty) = 1$. A recurrent state is a state starting from which it is certain for som $n > 0$ that $X_n = i$. A state that is not recurrent is named **transient**, a state starting from which is the probability of returning is less than 1. There are other ways for identifying recurrence:

Proposition 2.3.2. *Let be $i \in R$ a fixed state. Than the following conditions are equivalent*

1. *i is recurrent*
2. *The event $\{\mathcal{N}_i = \infty\}$ is almost sure[16] w.r.t. $\mathbb{P}_i$.*
3. *$\mathbb{E}_i \mathcal{N}_i = \infty$*

An analogous proposition old for transient state, modifying the events in an obvious way. We define two states i and j to **communicate** if for some m, $p_{ij}^m > 0$ and $p_{ji}^m > 0$. If in R all states do communicate is then called **irreducible**. This is extendible also to subsets of R. We remark that is not true in general that $p_{ij} = p_{ji}$, thus it is not a trivial fact. We indicate this relation with $\leftrightarrow$. If i is recurrent and that j is reachable from i, then i and j do communicate, $i \leftrightarrow j$, and even j is recurrent. We have in fact that

[16] An event $\mathcal{B}$ is almost sure if the complementary event $\overline{\mathcal{B}}$ has null probability w.r.t. the defined probability measure.

$\mathbb{P}_j(\tau(i) < \infty) = 1$ otherwise there would be an evolution of X starting from i and passing from j such that $\mathbb{P}_i(\tau(i) = \infty) > 0$; moreover we have

$$\mathbb{E}_j \mathcal{N}_j = \sum_{m=1}^{\infty} p_{jj}^m \geq \sum_{m=1}^{\infty} p_{ji}^{m_1} p_{ii} p_{ij}^{m_2} = \infty$$

Thus reccurence with $\leftrightarrow$ is an equivalence relation, and our state space R is partitioned in disjoint **recurrent classes** R_is such that

$$R = T \cup R_1 \cup R_2 \cdots \tag{2.23}$$

where the T is the set of transient states. The different recurrent classes have the absorbing property and are called **closed**: namely when we start in R_k we stay in R_k

$$\mathbb{P}_i(X_n \in R_k, \forall n) = 1, \quad \forall i \in R_k$$

The aim of this decomposition is that we can always restrict ourselves to think that we are in an irreducible state space, because if it is not the case we can break down R like above. If we start with an irreducible R it could else be made of all transient states or is made up by a unique recurrent class. Another distinction depends on whether $\mathbb{E}_i \tau(i)$ is finite or not: we will call i **positive recurrent** in the first case and **null recurrent** in the second. If there is a common divisor d_i in the set of recurrence times $\{\tau^m(i)\}_{m \in \mathbb{N}}$ it is called **period**; if $d_i = 1$ the chain is called **aperiodic**. We have the following[17]:

Proposition 2.3.3.
Given a recurrent class R, the states in it are all either positive or null recurrent and have the same period.

2.3.1 Stationarity and asymptotic behavior of Markov chains

Given a $\boldsymbol{\mu}(n) = (\mu_i(n))_{i \in R}$ probability measure on R at the time $t = n$, and considering it as a row vector we can define by matrix multiplication a new probability distribution at time $t = n + 1$,

$$\boldsymbol{\mu}(n+1) = \boldsymbol{\mu}(n) \boldsymbol{P} \tag{2.24}$$

[17] See [5, 22] for demonstrations.

where $\boldsymbol{P}$ is the transition matrix of the Markov chain. We will indicate the dependence by time with a subscript index, e.g. $\boldsymbol{\mu}_0$ will be the provability distribution over R at time $t = 0$, i.e. the initial distribution. We say that the distribution $\boldsymbol{\mu}$ is **stationary** if it is a left eigenvector of $\boldsymbol{P}$ with eigenvalue 1, $\boldsymbol{\mu} = \boldsymbol{\mu}\boldsymbol{P}$. The name stationary is refers to the fact that if we start with an $\boldsymbol{\mu}_0$ we will have that $\widetilde{\boldsymbol{\mu}}(n+1) = \boldsymbol{\mu}_0$ for each time step, i.e. the probability of being in a state j is the same as the initial. A brief analysis shows us that $\boldsymbol{\mu}_0\boldsymbol{P}$ is nothing than the $\mathbb{P}_\mu$ distribution of X_1, in fact:

$$(\boldsymbol{\mu}_0\boldsymbol{P})_j = \sum_{i \in R} \mu(0)_i p_{ij} = \mathbb{P}_\mu(X_1 = j) = \mu_j(1) \tag{2.25}$$

Generalizing, we have that the $\mathbb{P}_\mu$ distribution of X_n is given by

$$\boldsymbol{\mu}_0\boldsymbol{P}^n = \boldsymbol{\mu}_n \tag{2.26}$$

A fundamental result is the following:

Theorem 2.3.4. ***Existence of stationary distributions***
Given an irreducible and closed chain a stationary and positive (i.e. $\forall j, \mu_j > 0$) measure $\boldsymbol{\mu}$ always exists and it is unique up to a constant. If the chain is positive recurrent $\boldsymbol{\mu}$ is given by

$$(\boldsymbol{\mu})_j = \frac{1}{\mathbb{E}_j \tau(j)}$$

Given that an irreducible Markov chain with finite state space is always positive recurrent, the above result is quite useful. It is also possible to define a stationary measure in more general chains, see [5, 22]. An important case of stationary measure for physical models are **reversible measures**. A reversible measure has the property that

$$\mu_i p_{ij} = \mu_j p_{ji} \tag{2.27}$$

this property is often called in physical context **detailed balance**. Detailed balance is a stronger property than stationarity and implies it as one can see summing in i both sides in (2.27). A necessary and sufficient condition for the existence of a reversible measure is the following one:

Theorem 2.3.5.
A reversible measure exists if and only if

1. *$p_{ij} > 0$ implies $p_{ji} > 0$*

2. *for any loop* $X_0, X_1, \cdots, X_n \equiv X_0$ *the quantity*

$$\prod_{i=1}^{n} \frac{p_{i-1,i}}{p_{i,i-1}}$$

exists and is equal to 1.

Given a partition in classes like the one (2.23) we can define stationary distribution in each R_i and then take $\boldsymbol{\mu} = \sum_i \boldsymbol{\mu}_i$ as a stationary measure for the whole chain. Another important question is the asymptotic behaviour of the quantities p_{ij}^n in the limit $n \to \infty$. We state the two principal results.

Proposition 2.3.6.
Given a state j transient or null recurrent, we have $p_{ij}^n \to 0, \forall i \in R$ as $n \to \infty$.

Thus, a non trivial behavior in the limit of $n \to \infty$ is expected for positive recurrent chains.

Theorem 2.3.7. ***Ergodic theorem for markov chains***
For an irreducible, positive recurrent and aperiodic chain with stationary distribution $\boldsymbol{\mu}$ we have that

$$p_{ij}^n \to \mu_j, \forall j$$

or otherwise stated

$$\mathbf{P}^n \to \boldsymbol{e}^T \boldsymbol{\mu}$$

where $\boldsymbol{e}^T$ is the row vector of all ones.

this result can be interpreted that saying that no matter what the initial state the limiting distribution of X_n will be $\boldsymbol{\mu}$. The result can be generalized to periodic chains with some trickery, see [5]. Markov chains that satisfy theorem **??** are thus said **ergodic**.

2.4 Wiener process and Brownian motion

One special type of Markov processes are Wiener processes. In non-mathematical context **Wiener process** is usually called **Brownian motion**; actually Wiener process is the name of the mathematical model describing the physical problem of Brownian motion, thus we will use this last name for identifying both, as is usage nowadays. For our discussion is important because it

will be used to derive in **??** the distribution function of some special matrix ensembles. From a theoretical point of view it is utterly important for its wide range of uses and its rich structure provides multiple connection with physics modeling and mathematical physics. It is also the first place in which we introduce the **normal distribution** that will be fundamental in 3. We will indicate the normal distribution by $N(\mu, \sigma)$ a distribution whose density function $f(x)$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.28}$$

where μ is the mean of the distribution[18], σ^2 is the variance and σ is called **standard deviation**. $N(0,1)$ is often referred to as **standard normal distribution**. We are now ready to define the one dimensional Brownian motion.

Definition 2.10. ***Brownian motion***
A one dimensional ***Brownian motion*** *is a stochastic process B_t with time index set $T = \mathbb{R}_{\geq 0}$ taking values in $(\mathbb{R}, \mathcal{R})$ with the following properties:*

1. *$B_0 = x$ with $x \in \mathbb{R}$ deterministic.*
2. *The map $t \mapsto B_t$ is almost surely continuous..*
3. *For each set of ordered variables $t_0 < t_1 < \cdots < t_n$ the increments $B_{t_n} - B_{t_{n-1}}, \cdots, B_{t_2} - B_{t_1}$ are independent.*
4. *$\forall t_+, t_- \in T$, $t_+ > t_-$, $B_{t_+} - B_{t_-}$ has $N(0, t_+ - t_-)$ as distribution. In particular $B_t = B_t - B_0$ has as distribution $N(x,t)$.*

If the time set T is discrete we will talk about **discrete Brownian motion**. We can always traslate the motion in such a way that B_0 is always set equal to 0. If x in the first property is not a deterministic point but is a random variable X_0, we will have that the overall process is given by $B_t^* = X_0 + B_t$, with B_t standard Brownian motion. The distribution of B_t^* is thus the convolution of the distributions of X_0 and B_t as B_t^* is the sum of two random variables and (**??**) holds. Third property is referred to as *independent increments property*. The fourth one entails two different properties:

[18] Unfortunately the standard symbol in literature for distributions and mean do coincide, but it will not create any ambiguity and the context will distinguish between the two usages.

the property that $B_{t_+} - B_{t_-}$ has the same of $B_{t_+ - t_-}$ is referred to *stationary increments*. A process with these properties is called **Levy process**. Thus Brownian motion is a Levy process with **gaussian** stationary increments. The power of Brownian motion is inside this gaussian property. In fact, as we will see in 3 normal distribution plays a special role in probability theory. We can generalize Brownian motion to a $n-$dimensional Brownian motion requiring that each component $B_t^1, \cdots B_t^n$ is a one dimensional Brownian motion. It turns out that using Kolmogorov extension theorem is possible to demonstrate the existence of Brownian motion and its uniqueness[19]. Moreover it possess normal and strong Markov property.

For what concern Brownian motion trajectory it is a result that it is *Hölder continuous* with exponent $\theta < \frac{1}{2}$ but they are not *Lipschitz continuous*. Hence Brownian motion as a map $t \mapsto B_t$ is nowherwe differentiable. As in Markov process we can talk about transience and recurrence. The dimension of the Brownian motion is central in this topic. In fact it turns out that Brownian motion in dimensions $d \geq 3$ is *transient*, while it is *recurrent* for $d = 1$. For $d = 2$ it is *neighborhood recurrent*, which means that given a point in the plane x, the Brownian motion will visit almost surely any neighborhood of x arbitrary many times, but will never visit x at any time.

2.4.1 Heat equation

Brownian motion is deeply linked with **heat equation**. In fact working with a smooth $F : \mathbb{R} \to \mathbb{R}$ with bounded derivatives, we have that expanding with Taylor $F(x)$ where the increment corresponds to $dx := x(t + dt) - x(t)$, we have

$$F(x + dx) = F(x) + \frac{d}{dx}F(x)dx + \frac{1}{2}\frac{d^2}{dx^2}F(x)|dx|^2 + O(|dx|^3)$$

Considering x a function of a parameter t, $t \mapsto x(t)$ we have that $dx = x(t + dt) - x(t)$ and we can identify $x(t)$ with our Brownian process B_t:

$$F(B_{t+dt}) = F(B_t) + \frac{d}{dx}F(B_t)dB_t + \frac{1}{2}\frac{d^2}{dx^2}F(B_t)|dB_t|^2 + O(|dB_t|^3)$$

Taking expectations and remembering that $dB = B_{t+dt} - Bt$ is independent of $B(t)$ and has mean 0 for the properties of Brownian motion we have taking

[19] See [22, 27, 51, 65] as references on Brownian motion.

expectation that

$$\mathbb{E}F(B_{t+dt}) = \mathbb{E}F(B_t) + \mathbb{E}\frac{1}{2}\frac{d^2}{dx^2}F(B_t)|dB_t|^2 + o(dt)$$

Performing then a limit on dt we obtain:

$$\frac{d}{dt}\mathbb{E}F(B_t) = \frac{1}{2}\mathbb{E}\frac{d^2}{dx^2}F(B_t) \tag{2.29}$$

The process B_t has a probability density function depending on time t. In fact we know that for $t = 0$ we have a Dirac mass at $x_0 = B_0$. Denoting this density function with $\rho(x, t)$ we have

$$\mathbb{E}F(B_t) = \int_{\mathbb{R}} F(x)\rho(x, t)dx$$

Applying equation (2.29) we have that in a tempered distributional sense

$$\frac{\partial}{\partial t}\rho(x, t) = \frac{\partial^2}{\partial x^2}\rho(x, t) \tag{2.30}$$

which is the heat equation. It is known from PDE theory that the **heat kernel** is given by

$$\rho(x, t) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{|x-\mu|^2}{2t}}$$

For an $n-$dimensional Brownian motion we will have the heat kernel given by $N(\mu, \sigma^2 I)_{\mathbb{R}^n}$, the $n-$dimensional multivariate normal distribution with covariance matrix $\sigma^2 I$. Is thanks to the integrability of $N(\mu, \sigma^2 I)_{\mathbb{R}^n}$ for $n \geq 3$ that we have transient behavior of Brownian motion; viceversa the different types of divergence for $n = 1, 2$ give the different recurrence properties. We also observe that $n-$dimensional is rotation invariant thanks to its underlying distribution $N(\mu, \sigma^2 I)_{\mathbb{R}^n}$. This will play a central role in section **??**.

Chapter 3

Random Matrix Theory

Random matrix theory[1] (RMT for short) sees his birth at the beginning of the 20^{th} century, and it was introduced for the study of the spectra of certain heavy atoms. Since then RMT became an important object of study. We begin our discussion with the description of concentration of measure phenomenon and continue with the main results of RMT.

3.1 Concentration of measure

When studying an ensemble of random variables $\{X_1, X_2 \cdots, X_n\}$ taking values in the state space $(R, \mathcal{R})$ (we will deal with variables taking values in $\mathbb{R}$ or $\mathbb{C}$) we may be interested in the behavior of some "nice" function $F : R^n \to \mathbb{R}$. The classic example is the estimate of partial sums of the type $F(X_1, \cdots, X_n) \equiv S_n = X_1 + X_2 + \cdots + X_n$. The main point is that if the set of random variables has independence property ($k-$wise or even jointly independent), their fluctuation will cancel mutually out, concentrating the value sharply around the expected value. Obviously this concentration will be more evident in the great n limit, where the random noise due to fluctuations starts to flatten the value to a limit. The degree of knowledge on the moments and on their boundedness and the degree of independence of the variables we can exploit will affect the sharpness of the concentration. We start with some inequalities, describing how moments and other tools can

[1] The textbooks that we use as reference in this section are [3, 8, 65] in general, but mostly [65] which gives an overall view on the subject from different mathematical point of views. One quite old but still masterpiece reference is the classic text by Mehta [46].

be used for tail estimate in paragraph 3.1.1. Then we will use these powerful tools for deriving a masterpiece result in probability theory, the central limit theorem in paragraph 3.1.2.

3.1.1 Concentration inequalities and the Moment Method

Let's study the behavior of the sum $S_n = X_1 + X_2 + \cdots + X_n$. For simplicity we will suppose that the X_i will take value in $\mathbb{R}$. The case for complex variables will be easily deduced exploiting $\mathbb{C} \sim \mathbb{R}^2$. We will work with moments, and using expectation properties it is easy to note that $\mathbb{E}(S_n) = \sum_{i=1}^n \mathbb{E}(X_i)$. Thus normalizing $S_n - \mathbb{E}S_n = \sum_i = 1^n (X_i - \mathbb{E}X_i)$ we can always assume that the r.v. X_n have mean zero taking $S_n \to S_n - \mathbb{E}S_n$. Directly from equations (2.1), (2.5) and (2.6) we can have some preliminary results. In particular equations (2.1) and (2.5) adapted to S_n results in

$$\mathbb{P}(S_n \neq 0) \leq \sum_{i=1}^n \mathbb{P}(X_i \neq 0) \tag{3.1}$$

$$\mathbb{E}|S_n| \leq \sum_{i=1}^n \mathbb{E}|X_i| \tag{3.2}$$

In particular eq. () is trivial, but will be useful when we will deal with non bounded r.v. On the other side eq. () gives a better bound, but a weak one. The point is that we are not taking advantage of the independence of $\{X_1, X_2 \cdots, X_n\}$. Using the second moment we make our first assumption of independence. We have:

$$\mathbb{E}(|S_n|^2) = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}X_i X_j$$

Now, the quantity $\mathbb{E}(X_i X_j)$ is equal to zero, if we have pairwise independent r.v. unless $i = j$ and it is equal to $\mathbf{Var}(X_i)$ thanks to the zero mean assumption. Using then (2.6) we have

$$\mathbb{P}(|S_n| \geq \lambda) \leq \frac{1}{\lambda^2} \sum_{i=1}^n \mathbf{Var}(X_i) \tag{3.3}$$

Inserting in this equation $\lambda = \sqrt{\sum_{i=1}^n \mathbf{Var}(X_i)}$ we get that equation (3.3) asserts that $S_n = O(\sqrt{\sum_{i=1}^n \mathbf{Var}(X_i)})$. Euristically the values of S_n are concentrated inside an interval dominated by square root of the variances. We

remark that this result is not trivial. If the X_ns are Bernoulli variables with values $\{0,1\}$, they have unitary variance, and thus $\sqrt{\sum_{i=1}^{n} \mathbf{Var}(X_i)} = \sqrt{(n)}$ which in the limit $n \to \infty$ is much more powerfull than the trivial bound given by n. We can now turn to higher moments, working with variables with unit variance and let us further assume that they are bounded, $|X_i| \leq K$; we will see in the sequel how to deal with this simplification. Considering

$$\mathbb{E}(|S_n|^k) = \sum_{1 \leq i_1, \cdots, i_k \leq n} \mathbb{E} X_{i_1} X_{i_2} \cdots X_{i_k}$$

we note that to pursue a result we must require $k-$wise independence. This expectation is null otherwise we have that each X_i appears at least twice. But, for example, if $k = 9$, we cannot have that each X_i appears twice, but we must have at least an odd moment, let's say $\mathbb{E}X_j^3$. It turns out that working with odd moments does not provide control on the distribution of S_n because of we should account of $\mathbb{E}|S_n|^3 \neq \mathbb{E}S_n^3$. Thus we can limit our discussion to the case of even k, and we have at most $k/2$ terms. We can have in general that we can have a number inferior than $k/2$ if some X_i appears more than twice, say $k/2 - l$. Given the boundedness of the variables $|X_i| \leq K$ we can make an upper bound that will be linked to the number l such that $k/2 - l$ is the number of indices that appears. In fact the other variables that appears twice will contribute with 1, given the unit variance assumption. The bound is given thus by K^{2l} for each l multiplied by the number of configuration N_l, i.e. the number of ways that the set of integers $\{i_1, \cdots, i_k\}$ can be chosen in $\{1, \cdots, n\}$ such that we have l variables in "excess" from the couple condition. We have

$$\mathbb{E}|S_n|^k \leq \sum_{l=0}^{k/2} K^{2l} N_l$$

A combinatory argument, with some bounds lead us to the final result

$$\mathbb{P}(|S_n| \geq \lambda\sqrt{n}) \leq 2 \left(\frac{\sqrt{ek/2}}{\lambda} \right)^k \tag{3.4}$$

This inequality can be optimized when our variables are not just $k-$wise independent for each fixed k, but are jointly independent. We end up with:

$$\mathbb{P}(|S_n| \geq \lambda\sqrt{n}) \leq C \exp\Big(-c\lambda^2\Big) \tag{3.5}$$

The control of all the moments give thus a strong bound of sub-Gaussian type. The control of each moment is somewhat included in the control of

the exponential moment $\mathbb{E}\exp(tS_n)$ which can be used to demonstrate the following theorem:

Theorem 3.1.1. ***Chernoff inequality***
Given a set of independent and almost surely bounded r.v. $\{X_1, X_2 \cdots, X_n\}$ *with* $|X_i| \leq K$ *with mean* μ_i *and variance* σ_i^2. *We have* $\forall \lambda > 0$

$$\mathbb{P}(|S_n - \mu| \geq \lambda\sigma) \leq C \max(\exp\Big(-c\lambda^2\Big), \exp(-c\lambda\sigma/K)) \tag{3.6}$$

with $C, c > 0$ *constants and* $\mu = \sum_{i=1}^n \mu_i$, $\sigma^2 = \sum_{i=1}^n \sigma_i^2$.

We note that the exponential moment relies heavily on the commutativity of the random variables, thus cannot be used in more general contexts. The condition of jointly independent r.v. can be somewhat relaxed to a **martingale property**

$$\mathbb{E}(X_i \mid X_1, \cdots, X_{i-1}) = 0$$

which give raise to a large deviation bound of sub-Gaussian type and it is called **Azuma's inequality**.

Until now we have assumed that our variables are bounded, but what we can do if they are not? **Truncation method** provides a trick for doing this. We divide our variables X_i in the sum of two different random variables $X_{i,\leq N}$ and $X_{i,>N}$, where $X_{i,\leq N} = X_i \mathbf{1}(|X_i| \leq N)$ and $X_{i,>N}$ is analogous, i.e. we split in two the variable with respect to some truncation parameter N. The two new sums

$$S_{n,\leq N} = X_{1,\leq N} + \cdots + X_{n,\leq N}$$

$$S_{n,>N} = X_{1,>N} + \cdots + X_{n,>N}$$

can be analyzed by different means: $S_{n,\leq N}$ is made up with bounded variables, so we can use the concentration inequalities that we have seen up to now. $S_{n,>N}$ on the other end is not made by bounded variables, but the fact that we usually work with non *heavy-tailed* distribution give us the possibility of bounding $S_{n,>N}$ in a such way that its contribution becomes negligible in the limit $n \to \infty$. In fact usually we can use for these tail r.v. equations (3.1.1) and (3.1.1) which usually give a nice asymptotic behavior.

We can now analyze the case when F is not just the sum or a linear combination of the variables $\{X_1, X_2 \cdots, X_n\}$, but a function taking

real or complex values. It is necessary to make some regularity condition on F. Let us assume that F has some kind of **Lipschitz**[2] behavior. We can start with a Lipschitz behavior on the single variables, i.e. $|F(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n) - F(x_1, \cdots, x_{i-1}, \overline{x_i}, x_{i+1}, \cdots, x_n)| \leq k_i$. We can further assume that the variables take values in different state spaces R_i. We have the following:

Theorem 3.1.2. ***McDiarmid's inequality***
Given a set of independent r.v. $\{X_1, X_2 \cdots, X_n\}$ *taking values in* $R_1, \cdots, R_n$ *and a function* $F : R_1, \cdots, R_n \to \mathbb{R}$ *such that we have for each component* i*:*

$$|F(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n) - F(x_1, \cdots, x_{i-1}, \overline{x_i}, x_{i+1}, \cdots, x_n)| \leq k_i \tag{3.7}$$

where $x_k \in Im(X_k)$ *and* $\overline{x_i} \in Im(X_i)$*. Then we have for* $\lambda > 0$*:*

$$\mathbb{P}\left(|F(X) - \mathbb{E}F(X)| \geq \lambda\sigma^2\right) \leq C\exp\left(-c\lambda^2\right) \tag{3.8}$$

with $C, c > 0$ *constants and* $\sigma^2 = \sum_{i=1}^n k_i^2$*.*

If we require joint Lipschitz regularity for all variables varied at once, then we come at powerful results. The idea back this kind inequalities that will be fundamental even in the next paragraph is that the universality behavior of a sufficiently regular and large set of random variables is well described by gaussian variables. Any set of regular variables will be in the great n limit described faithfully except for errors that go to zero for $n \to \infty$. Gaussian r.v. are concentrated under global $1-$Lipschitz functions, the matter reduce to demonstrate that this is an universal behavior. The fundamental theorem from Talagrand ([62]) is a cornerstone in concentration of measure theory:

Theorem 3.1.3. ***Talagrand concentration inequality***
Given a set of independent r.v. $\{X_1, X_2 \cdots, X_n\}$ *bounded by a constant* $K > 0$ *such that* $|X_i| \leq K \ \forall i$*, let be* F *a* 1*-Lipschitz convex function, then we have for any* λ*:*

$$\mathbb{P}\left(|F(X) - \mathbb{E}F(X)| \geq \lambda K\right) \leq C\exp\left(-c\lambda^2\right) \tag{3.9}$$

with $C, c > 0$ *constants.*

[2] A real valued function $f(x)$ is said to be **Lipschitz** if $|f(x) - f(y)| \leq K|x - y|$ for some constant K.

A corollary for Talagrand concentration will be useful in chapter **??**.

Corollary 3.1.4. *Given a set of independent complex r.v.* $\{X_1, X_2 \cdots, X_n\}$, *with mean zero and unit variance, bounded almost certainly by* K. *Let* V *be a* $d-$*dimentional subspace of* $\mathbb{C}^n$. *Then we have, using the Talagrand inequality with* $F \equiv d(X, V)$ *the distance of a vector* X *from a subspace* V, *that*

$$\mathbb{P}\left(\left|d(X,V) - \sqrt{n-d}\right| \geq \lambda K\right) \leq C\exp\left(-c\lambda^2\right) \tag{3.10}$$

3.1.2 Central Limit Theorem

We now try to give formal rigor to the assertion of universality we made in the previous paragraph. Let's take S_n. The point is that the statistical fluctuations of the variables X_i of a certain ensemble of r.v. under study tend to rule out each other under certain regularity assumption in a large n limit. This results not only in a large deviation inequalities like equation (3.6), but the overall behavior of the variables converges to an *universal* distribution. This result goes by the name of Central Limit Theorem.

Theorem 3.1.5. ***Central Limit Theorem***
Consider the normalized sum

$$Z_n := \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

where the variables X_i *are iid real r.c. with finite mean and variance* σ^2. *For* $n \to \infty$, Z_n *converges in distribution to the standard normal distribution* $N(0,1)_{\mathbb{R}}$.

The factor in Z_n, $\frac{1}{\sqrt{n}}$, is tight, in the sense that we cannot under the general assumption of CLT find any weaker regularization. There are many ways to demonstrate CLT. They all try to demonstrate the convergence of Z_n to $N(0,1)$ through the convergence of other quantities that are easier to work with. One of these is an analytical method and goes through the analysis of the distributions. It takes the characteristic function $\mathbb{E}e^{itX}$, that is the measure equivalent of the standard Fourier transform and uses a tool named **Lévy continuity theorem** to establish the result. We will use the moment method, that will be central even in the analysis of matrices; in fact

Fourier method relies heavily on the commutativity of the random variables. We quickly note that given the Taylor expansion

$$\mathbb{E}X^{itX} = \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\mathbb{E}X^k$$

the two methods must be linked. The following continuity theorem will be useful even in the future

Theorem 3.1.6. ***Carleman continuity theorem***
Consider a set of sub-Gaussian variables X_i, and a sub-Gaussian X, the the following statements are equivalent

- *For every k $\mathbb{E}X_n^k$ converges pointwise to $\mathbb{E}X^k$*
- *X_n converges in distribution to X.*

So, to verify that a distribution X is the limit of a series X_n, it will suffice to show the convergence of the moments as $n \to \infty$. The moment of a gaussian variable $G \equiv N(0,1)_{\mathbb{R}}$ are easy to obtain:

$$\mathbb{E}G^k = \int x^k \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} = \begin{cases} 0, & \text{if } k \text{ is odd} \\ \frac{k!}{2^{k/2}(k/2)!}, & \text{if } k \text{ is even} \end{cases} \tag{3.11}$$

It remains to evaluate the moments of Z_n. As before we can assume that the iid random variables are distributed with unit variance, zero mean and we can always apply truncation method and considering bounded r.v. In fact if $Z_{n,\leq}$ converges in distribution to $N(0,1)$ so will do $Z_{n,>}$ to zero from dominated convergence theorem. The fact that we can consider X bounded give us the possibility of applying Chernoff inequality (3.6) and use Carleman's theorem.Thus we can formally study the convergence of Z_n through analysis of the moments of the distribution. This is done as follows. Considering the k^{th} moment of the distribution we have

$$\mathbb{E}X_n^k = \frac{1}{n^{k/2}} \sum_{1\leq i_1,\cdots,i_k\leq n} \mathbb{E}X_{i_1}\cdots X_{i_k}$$

As before using mean zero property and unit variance we arrive to considering sums of the type

$$\frac{1}{n^{k/2}} \sum_{1\leq i_1,\cdots,i_m\leq n} c(k,a_1,\cdots a_m)\mathbb{E}X_{i_1}^{a_1}\cdots X_{i_m}^{a_m} \tag{3.12}$$

where now $j_1 < \cdots < j_m$ are a subset of $\{1, \cdots, n\}$. It's easy to see that the factor $c(k, a_1, \cdots a_m)$ is the multinomial coefficent

$$c(k, a_1, \cdots a_m) = \frac{k!}{a_1! \cdots a_m!}$$

Now, most of the terms $X_{j_i}^{a_i}$ vanish ora are infinitesimal in the limit $n \to \infty$ with respect to a partition where each term have an even power a_i of at least 2. Thus as in the previous paragraph, we have at most $k/2$ elements. The number of summands in (3.12) is $\binom{n}{m}$. When we have effectively $k/2 - l$ different elements, the boundedness give elements that are $O(n^l)$ for $n \to \infty$. Thus we must look for only the terms where $m = k/2$. For odd k we have always an exceeding term that cancel $\mathbb{E}X_n^k$. For even k, given the unit variance assumption, each $X_i^2 = 1$. We remain with

$$\frac{1}{n^{k/2}} \binom{n}{m} c(k, 2, \cdots, 2) = \frac{k!}{2^{k/2}(k/2)!} + o(1) \tag{3.13}$$

which coincides with (3.11) for $n \to \infty$.

The generalization of the idea behind CLT leads to the famous **Lindeberg swapping trick**. In order to perform calculations (even when the variables appear inside a *nice* function F) on variables that are compatible with theorem 3.1.5 we can work with directly with gaussians G_i in place of the r.v. X_i. This trick is useful when we are interested in rate of convergence of CLT, such as in **Berry-Esséen theorem** which we present for later convenience[3]:

Theorem 3.1.7. ***Berry-Esséen theorem***
Consider a set of centered iid random variables X_i with bounded third moment $\mathbb{E}|X_i|^3 = \rho < \infty$. If $\mu(x)$ is the distribution of Z_n we have

$$|\mu(x) - \mathcal{N}(x)| \leq 3\frac{\rho}{\sqrt{n}}\sigma^3$$

where $\mathcal{N}(x)$ is the cumulative distribution of $N(0,1)_{\mathbb{R}}$

[3] See [22, 44, 65].

3.2 Operator norm Bound and Bai-Yin theorem

We are now ready to deal with matrices and random matrix theory, as it will be the principal argument of this chapter. We will deal principally with square matrices $n \times n$, but some of the results of this chapter are extendible to rectangular matrices $m \times n$, see as reference [8]. We can always interpret a matrix as an array of random variables $M = (\xi_{ij})_{1 \leq i,j \leq n}$. Depending on the characterization of the ξ_{ij} we will have

- **Iid matrix ensembles** where the $\xi_{ij} \equiv \xi$ are iid variables.
- **Symmetric Wigner ensembles** where the ξ_{ij} are independent real random variables for the upper triangular sector $j > i$, but we impose symmetric behavior $\xi_{ij}) = \xi_{ji}$. Usually the elements ξ_{ij}, $j > i$, are taken to be identically distributed, but it is not a binding request. For our studies diagonal entries will play a marginal role, because we will see that in a large n limit their influence goes to zero, but for definition we require that ξ_{ii} to be iid with a different distribution. For example, the **Gaussian Orthogonal Ensemble** (GOE) has upper triangular elements distributed under $N(0,1)_{\mathbb{R}}$ law, while diagonal elements are r.v. with distribution $N(0,2)_{\mathbb{R}}$ for convenience.
- **Hermitian Wigner ensembles** are complex matrices, with hermitian property $\xi_{ij} = \overline{\xi_{ji}}$. The same details we have said for symmetric matrices hold here. An example is **Gaussian Unitary Ensemble** (GUE), where off diagonal elements have as distribution $N(0,1)_{\mathbb{C}}$ and the diagonal ones have $N(0,1)_{\mathbb{R}}$.

One of the first quantities we are going to analyze is statistical behavior of the operator norm of a $n \times n$ matrix M in a given ensemble, defined by

$$\|M\|_{op} := \sup_{x \in \mathbb{C}^n : \|x\|=1} \|Mx\| \tag{3.14}$$

where $\|\cdot\|$ is the usual norm in $\mathbb{C}^n$, in the large n limit. This value coincides with the largest **singular value** $\sigma_1(M)$[4], i.e. the square root biggest eigenvalue of the matrix $M^\dagger M$, where † represents adjoint. In this section we deal

[4]We will suppose the eigenvalues and singular values ordered in decreasing order, i.e. $\lambda_1 \geq, \lambda_2 \geq \cdots \lambda_n$.

with results that don't require the matrix M to be symmetric or hermitian. Even if these assumptions will be made for solving the problem in a more direct way, the generalization to a generic matrix is possible. We are going to use concentration of measure arguments like in the previous section, and in particular the moment method for obtaining bounds of the type:

$$\mathbb{P}(\|M\|_{op} \geq \lambda) \leq f(\lambda, n)$$

which will in general depend on the dimension n of the matrix and the threshold λ. We note here that te matrix J_n, i.e. the matrix $n \times n$ containing all 1s has $\|J_n\|_{op} = n$, fact that will be needed later. Like in the analysis made for $|S_n|$, a rough bound is given through J_n: multiplying his entries by $O(1)$ we expect an operator norm of $O(n)$. This results, as before, does not take in account the independence of the entries and the mutual cancellations. One key result derives directly from theorem 3.1.3:

Proposition 3.2.1. ***Concentration of measure for*** $\|M\|_{op}$
Given a matrix M with independent entries M_{ij} of mean 0 bounded by one, $|M_{ij}| \leq 1$. Then for any $\lambda > 0$ we have

$$\mathbb{P}\Big(\Big|\|M\|_{op} - \mathbb{E}\left(\|M\|_{op}\right)\Big| \geq \lambda\Big) \leq C \exp\Big(-c\lambda^2\Big) \tag{3.15}$$

for some constants $C, c > 0$.

Basically we are going to take advantage of the convexity and the $1-$Lipschitz behavior of $\|M\|_{op}$, seen as a function on $\mathbb{C}^{n^2}$ of the entries of the matrix M. $1-$Lipschitz behavior is easily obtained by bounding $\|M\|_{op}$ with Frobenius norm

$$\|M\|_F = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} |M_{ij}|^2\right)^{\frac{1}{2}}$$

. We can then apply directly theorem 3.1.3. It is left to calculate $\mathbb{E}\left(\|M\|_{op}\right)$. We will use moment method[5] applied to real symmetric matrices because it is of easier implementation, in particular in combinatorial result. The generalization to hermitian matrix requires some little effort in the analysis we are gonna give in what follows. The problem for a generic can be tackled considering an augmented matrix of the tipe

$$\tilde{M} = \begin{bmatrix} 0 & M \\ M^\dagger & 0 \end{bmatrix}$$

[5]We will follow the presentation given in [65].

reducing the problem to the study of an hermitian matrix. In the symmetric or hermitian case $\|M\|_{op}$ is directly linked to eigenvalues of M. In fact we have that

$$\sigma_1(M) = \max_{1\leq i\leq n} |\lambda_i(M)|$$

that is also a first direct informations on the spectrum of our matrix. The moments of a matrix M are defined w.r.t. the **Empirical Spectral Distribution** (ESD)

$$\mu_M = \frac{1}{n}\sum_{j=1}^{n} \delta_{\lambda_j(M)} \tag{3.16}$$

The ESD is simply the distribution that gives a weight to λ_j according to its multiplicity. Otherwise stated, $\mu_M(\mathcal{A})$ counts the number of eigenvalues of M in a given interval $\mathcal{A}$:

$$\mu_M(\mathcal{A}) = \frac{1}{n}\sharp\{\lambda_j(M) \mid \lambda_j(M) \in \mathcal{A}\}$$

where $\sharp$ indicates cardinality of the set. Considering an ensemble of random matrices, it should be noted that μ_M is not a simple distribution, but rather a *random* distribution. Indeed it gives us the eigenvalues of an element drawn from the ensemble. It is itself a random variable in the space of distribution over the real line. Now we can see that moments are defined to be

$$\begin{aligned} m_1(M) &= \int_{\mathbb{R}} t d\mu_M(t) = \frac{1}{n}\operatorname{tr} M \\ &\vdots \\ m_k(M) &= \int_{\mathbb{R}} t^k d\mu_M(t) = \frac{1}{n}\operatorname{tr} M^k \end{aligned} \tag{3.17}$$

$\mathbb{E}\mu_M$ can be defined through *Riesz representation theorem* as the distribution which solves the following identity:

$$\int_{\mathbb{R}} \phi(x) d\mathbb{E}\mu_M(x) = \mathbb{E}\int_{\mathbb{R}} \phi(x) d\mu_M(x) \tag{3.18}$$

Given the connection of moments and $\|M\|_{op}$, with eigenvalues its easy to see that the knowledge of the formers poses bonds on the latter. In fact given that $\|M\|_{op} = \max_{1\leq i\leq n} |\lambda_i(M)|$, we have that

$$\|M\|_{op}^k \leq \operatorname{tr} M^k \leq n\|M\|_{op}^k \tag{3.19}$$

For example, given that the second moment $m_2(M) = \frac{1}{n}\operatorname{tr} M^2 = \frac{1}{n}\|M\|_F^2$, in a matrix with off-diagonal entries of mean 0 and variance 1 and diagonal entries of mean zero and bounded variance, taking M^2 and by the weak law of large numbers and applying equation 3.19, we have

$$(1+o(1))\sqrt{n} \leq \|M\|_{op} \leq (1+o(1))n$$

asymptotically almost surely. Like before, we now work with even moments with $k > 2$, assuming that we M_{ij} has mean zero, unit variance and are bounded by constant K. The general term $\operatorname{tr} M^k$ is given by

$$\operatorname{tr}\Big\{M^k\Big\} = \operatorname{tr}(M\cdots M) = \sum_{1\leq i_1,i_2,\cdots,i_k\leq n} M_{i_1i_2}M_{i_2i_3}M_{i_3i_4}\cdots M_{i_{k-1}i_k}M_{i_ki_1} \tag{3.20}$$

We can view this sum graphically. Each i_k is a node of a graph (see chapter 5), and couples of indices that appear in (3.2) are the edges that appear in the graph, i.e. the sequence $\{i_1,i_2\},\{i_2,i_3\},\cdots\{i_{k-1},i_k\},\{i_k,i_1\}$ is a closed path on the graph of possible indices. We study the expected value $\mathbb{E}(\operatorname{tr} M^k)$, because we can take advantage of mean zero and independence of the entries. The only non vanishing terms, like in the dissertation of the moment method for Z_n, are those in which there are at most $k/2$ terms, and thus at most $k/2+1$ nodes. We note that we can divide the cycles in classes of equivalence, and work with classes. For example, with $k=6$, all cycles composed by three distinct edge, give the same contribution to the sum (). Considering a class of j edges, with multiplicity $a_1,\cdots,a_j$ that sum up to k, choosing the first node from the n possible ones, we can identify the cycle once we are given with the j edges. Each edge is a choice between n elements (we allow to remain in the same node). Thus we can have at most n^{j+1} cycles of this type. Using the conditions on M_{ij}, we have that if $\{i,j\}$ occurs a times in the cycle, it will contribute with a factor K^{a-2}, where a is the multiplicity $a_j \geq 2$ corresponding to the edge $\{i,j\}$. Thus the single summand in (3.2) gives a contribution of K^{k-2j}, and having n^{j+1} elements in the same class it contributes with a factor $n^{j+1}K^{k-2j}$. Summing over the classes, and using the bound

$$n^{j+1}K^{k-2j} \leq n^{k/2+1}\max\left(1,\left(\frac{K}{\sqrt{n}}\right)^{k-1}\right)$$

we obtain that

$$\mathbb{E}(\operatorname{tr} M^k) \leq \left(\frac{k}{2}\right)^k n^{k/2+1} \max\left(1, \left(\frac{K}{\sqrt{n}}\right)^{k-1}\right)$$

$$\Big\Downarrow \text{ by eq. (3.19)} \tag{3.21}$$

$$\mathbb{E}(\|M\|_{op}^k) \leq \left(\frac{k}{2}\right)^k n^{k/2+1} \max\left(1, \left(\frac{K}{\sqrt{n}}\right)^{k-1}\right)$$

And we get by Markov inequality (2.5)

$$\mathbb{P}\left(\|M\|_{op} \geq \lambda\right) \leq \lambda^{-k} \left(\frac{k}{2}\right)^k n^{k/2+1} \max\left(1, \left(\frac{K}{\sqrt{n}}\right)^{k-1}\right)$$

Optimizing in k, taking it comparable with $\log n$, we have that with high, probability

$$\|M\|_{op} = O\left(\sqrt{n} \log(n) \max\left(1, \frac{K}{\sqrt{n}}\right)\right) \tag{3.22}$$

It is easy to see that if $K = o(\sqrt{n})$, each class contribution to the sum $n^{j+1}K^{k-2j}$ will be $o_k(n^{k/2+1})$, where the subscript k means that the asymptotic behavior depends on it. On the other hand, if we have $j = \frac{k}{2}$, each edge appears twice in this type of cycle, and by the unit variance assumption, the contribution of the whole cycle is 1. The determination of the expectation of (3.2) is thus reduced to the count of this type of cycles, in particular those which travers exactly $k/2+1$ nodes. Indeed if they were less than $k/2+1$ we would have an infinitesimal contribution, because the total number of cycle of $k/2$ distinct edges that cross less than $k/2+1$ is an $o_k(n^{k/2+1})$ and can be neglected. This combinatorial problem is equivalent to find the so called **Dick words** of length k. This number is given by $C_{\frac{k}{2}} n(n-1) \cdot (n-k/2)$, where

$$C_{\frac{k}{2}} = \frac{k!}{\left(\frac{k}{2}+1\right)!(\frac{k}{2})!} \tag{3.23}$$

is the $\frac{k}{2}^{\text{th}}$ **Catalan number**[6]. This is exactly analogous to the factor $\frac{k!}{2^{k/2}(k/2)!}$ we have found in (3.13). We have thus demonstrated that, for a symmetric

[6] See [65] Lemma 2.3.15, page 122.

matrix M with jointly independent elements with mean zero and variance one, bounded in magnitude by $o(\sqrt{n})$, the even expected moments are given by

$$\mathbb{E}(\operatorname{tr} M^k) = \left(C_{\frac{k}{2}} + o_k(1)\right) n^{\frac{k}{2}+1} \tag{3.24}$$

Using **Stirling formula**

$$n! = (1 + o(1))\sqrt{2\pi n} n^n e^{?n}.$$

we see that $\left(C_{\frac{k}{2}}\right)^{1/k} \to 2$ as $k \to \infty$. We have obtained

Theorem 3.2.2. ***Lower Bai-Yin theorem***
Given a real symmetric random matrix M with upper triangular elements independent and with mean zero and variance one and bounded in magnitude by $O(1)$. We have that

$$\|M\|_{op} \geq (2 - o(1))\sqrt{n}$$

For a sharp upper bound we need to improve the bound in (3.22); in particular the logarithm term can be improved. Omitting the combinatorial tedious results, we state the main results of this chapter.

Theorem 3.2.3. ***Upper Bai-Yin theorem***
Given a real symmetric random matrix M with upper triangular elements iid and with mean zero, variance one and 4^{th} moment bounded in magnitude by $O(1)$. Then we have asymptotically almost surely that

$$\|M\|_{op} \leq (2 + o(1))\sqrt{n}$$

We note that the 4^{th} moment boundedness is the best possible.

3.3 Wigner Semicircular Law

We are now going to study the asymptotical behavior of the ESD (3.16) of symmetric matrices. We will pursue this, result through two methods, moment method, using the results of the previous chapter, and the Stiltjes transform. The main result will be **Wigner semicircular distribution**, which is the asymptotic limit of the ESD of this type of matrices.

Theorem 3.3.1. ***Wigner semicircular distribution***
Let be A_n a symmetric wigner matrix, and $M_n = \frac{1}{\sqrt{n}} A_n$ the normalized version of A_n. Then the ESDs μ_{M_n} will converge almost surely to ***Wigner***

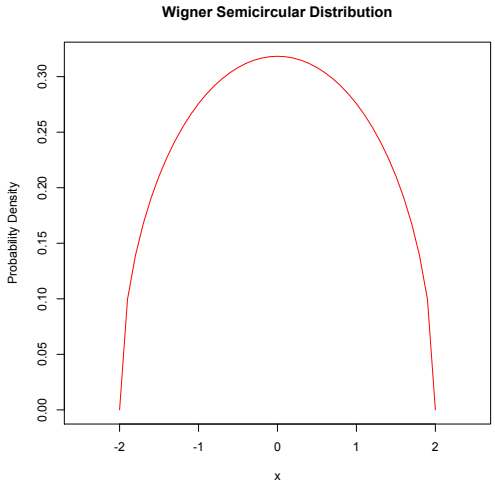


Figure 3.1: Wigner semicircular distribution

semicircular distribution

$$\mu_{sc} = \frac{1}{2\pi}\left(4 - |x|^2\right)_+^{\frac{1}{2}} dx \tag{3.25}$$

The demonstration will require that the matrix is centered. The following result state that even non centered have the same distribution.

Theorem 3.3.2. ***Rank inequality***
Let be A and B two Hermitan matrices and $\mu_A^c(x) = \mu_A(]-\infty, x])$, $\mu_B^c(x) = \mu_B(]-\infty, x])$ their cumulative ESD, we have:

$$\|\mu_A^c(x) - \mu_B^c(x)\|_\infty \leq \frac{1}{n}\operatorname{rank}|A - B| \tag{3.26}$$

where $\|f(x)\|_\infty = \sup_x |f(x)|$

In fact a non centered random matrix M_n can be written as $M_n = W_n + mJ_n$, where J_n is the matrix of all 1s and m is the mean of the distribution of the M_{ij}. We will deal with sparse matrices in chapter 5. Another result, similar to theorem 3.3.2, says that if we deform $\frac{1}{\sqrt{n}}M_n$ whose distribution converges a.s. to a limit μ by a sequence N_n with $\frac{1}{n^2}\|N_n\|_F^2 \to 0$ a.s. then

$\mu_{\frac{1}{\sqrt{n}}(M_n+N_n)} \to \mu$ a.s. Noting from the CLT that the diagonal entries of Wigner matrix have Frobenius norm of order $O(\sqrt{n})$, the result we have just stated says that they can be ruled out in our calculation without loss of generality. In the optics

Defining the sparsity parameter α as the number of the expected non zero values in a row, Wigner law is believed to hold for $\alpha \geq \log(n)$[7]. This is the threshold under which at which, reasoning with the arguments of moment method we used in the last section, a cycle starting from a point cannot reach some points, invalidating the reasoning we made. We will investigate more on this point in chapter 5. Typically, theorem , is used in the case where the entries are iid, or at most that the diagonal entries and the upper triangular elements have two different distributions. In [67] is proved the following:

Theorem 3.3.3. ***Lindeberg condition for convergence***
Let be A_n a symmetric matrix with diagonal and upper triangular independent entries, with mean zero and bounded variances $\sigma_{ij} < C$. Then considering the normalized matrix $M_n = \frac{1}{n}A_n$ the following two results imply each other

- *$\mu_{M_n} \to \mu_{sc}$ in probability, $\frac{\sigma_{ij}}{n} \to 0$ for $n \to \infty$*

$$\Updownarrow$$

- *$\lim_{n\to\infty} \sum_{j=1}^{n} \sigma_{ij}^2 = 1$, $i = 1, \cdots, n$ and the following Lindeberg condition:*

$$\forall \lambda > 0, \quad i = 1, \cdots, n : \lim_{n\to\infty} \sum_{j=1}^{n} \mathbb{E}\left((M_n)_{ij}^2 \mathbf{1}_{|(M_n)_{ij}>\lambda|}\right) = 0 \tag{3.27}$$

where $\mathbf{1}_{\mathcal{B}}$ is the indicator function of $\mathcal{B}$.

The name of 3.27 is another prove that Wigner semicircular law is the non abelian counterpart of central limit theorem. In fact equation (3.27) is nothin more than an adaptation of the more famous Lindeberg CLT to the matrix case. This equation is the base for many generalization of semicircular law, for example where some kind of dependence are introduced in M_n, see [31].

[7] See [34, 41].

Wigner semicircular law has also stability properties w.r.t. perturbations. We have in fact that from **Weilandt-Hoffmann inequality**

$$\sum_{j=1}^{n} |\lambda_j(A+B) - \lambda_j(A)|^2 \leq \|B\|_F^2$$

We can transpose this inequality in terms of ESD. In fact $\forall \lambda, \epsilon,\ \epsilon > 0$, we have given the normalized matrices A and B that:

$$\mu_{A+B}(]-\infty, \lambda]) \leq \mu_A(]-\infty, \lambda+\epsilon]) + \frac{1}{\epsilon^2 n^2}\|B\|_F^2$$
$$\mu_{A+B}(]-\infty, \lambda]) \leq \mu_A(]-\infty, \lambda-\epsilon]) + \frac{1}{\epsilon^2 n^2}\|B\|_F^2$$

We are ready to prove convergence to semicircular law.

3.3.1 Moment method

We are gonna use a straightforward adapted version of theorem 3.1.6 to the case of ESD. The fact that the moments are sub-gaussian is a direct consequence of theorem 3.2.1. What remains to show is that the moments $\mathbb{E}\,\mathrm{tr}\left(\frac{1}{n}M_n^k\right)$ tend to those of semicircular distribution

$$\int_{\mathbb{R}} x^k d\mu_{sc}(x) = C_{\frac{k}{2}} \tag{3.28}$$

where $C_{\frac{k}{2}}$ is the $\frac{k}{2}^{\text{th}}$ Catalan number defined in (3.23). Indeed we have for even moment $k = 2m$ that:

$$\begin{aligned}
\int_{\mathbb{R}} x^{2m} d\mu_{sc}(x) = & \\
&= \frac{1}{2\pi}\int_{-2}^{2} x^{2m}\sqrt{4-x^2}dx \\
&= \frac{1}{\pi}\int_{0}^{2} x^{2m}\sqrt{4-x^2}dx \\
&= \frac{2^{2k+1}}{\pi}\int_{0}^{1} y^{k-\frac{1}{2}}\sqrt{1-y}dy \\
&= \frac{2^{2k+1}}{\pi}\frac{\Gamma(k+1/2)\Gamma(3/2)}{\Gamma(k+2)} = \frac{1}{k+1}\binom{2k}{k}
\end{aligned} \tag{3.29}$$

which coincides with 3.23 for $2k \to k$. Remembering by equation (3.24) that $\mathbb{E}(\mathrm{tr}\, M^k) = (C_{\frac{k}{2}} + o_k(1))n^{\frac{k}{2}+1}$, we cam apply Carleman continuity theorem 3.1.6 and assert that $\mu_{\frac{1}{\sqrt{n}}M_n} \to \mu_{sc}$ a.s.

3.3.2 Stieltjes transform method

As in classical central limit theorem, moment method works with algebraic computations, but in a roundabout way: we already know which is the limit distribution. In fact Carleman's theorem 3.1.6 relies heavily on the knowledge of the target distribution moments. Looking at the classical setting, the straightforward method for the retrieval of the limit distribution without knowing it is *Fourier method* which relies on the Fourier transform of the distribution, which is also called *characteristic function* (see section 2.1). We note that is the knowledge of the characteristic function of a distribution is in a certain sense a control on it more powerful than a control via moment method. In fact, we can do a Taylor expansion of $F_X(t) := \mathbb{E}e^{it\cdot X}$

$$\mathbb{E}e^{it\cdot X} = \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\mathbb{E}X^k$$

we retrieve the information on the moments of the considered distribution. Thus the point is how the fourier transform could be generalized to our case. We indeed highlight that standard fourier method relies on the commutative properties of the random variables whose distribution is under study. One possible generalization could be *Harish-Chandra-Itzykson-Zuber integral formula* (see [33, 35, 65]) and we will say something more on it in section **??**. In fact this method can be used in an effective way only for highly symmetric ensembles, as GOE and GUE (see the beginning of section 3.2). Another way for recollecting our wanted result relies on *Stieltjes transform method* ([8, 25, 65]). We define the Stieltjes transform[8] as

$$s_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z} d\mu(x), \quad z \in \mathbb{C} \tag{3.30}$$

where $\mu(x)$ is the distribution we are interested in. In our case we are interested in the ESD of the elements of matrix ensemble we are considering, i.e. (including the normalization factor) $\mu_{\frac{1}{\sqrt{n}}M_n}$; this could be represented by the ***resolvent*** of our matrix

$$\int_{\mathbb{R}} \frac{1}{x-z} d\mu_{\frac{1}{\sqrt{n}}M_n}(x) = \frac{1}{n}\operatorname{tr}\left(\frac{1}{\sqrt{n}}M_n - zI\right)^{-1} \tag{3.31}$$

[8] Certain authors with a more *functional* or *complex analysis* background call (3.30) *Cauchy transform*, sometimes multiplied by a factor -1 (i.e. $1/(x-z)$).

As for Fourier method, Stieltjes transform has a relation with the knowledge of the moments of our distribution. Indeed expanding $s_{\mu_{\frac{1}{\sqrt{n}}M_n}}(z)$ we obtain:

$$s_{\mu_{\frac{1}{\sqrt{n}}M_n}}(z) = -\frac{1}{z} - \frac{1}{z^2 n^{\frac{3}{2}}} \operatorname{tr} M_n - \frac{1}{z^3 n^2} \operatorname{tr} M_n^2 - \dots \tag{3.32}$$

which entails somewhat a knowledge on the moments of the inquired distribution. The power of Stieltjes transform relies on the complex analytic tools that we can exploit using it (in the same way as classical CLT uses the properties of Fourier transforms in deriving the limit distribution). First we note that $s_\mu(z)$ is well defined for $\operatorname{Im} z > 0$ or $\operatorname{Im} z < 0$. In particular, writing $z = a + ib$ we see that

$$\operatorname{Im} \frac{1}{x - z} = \frac{b}{(x-a)^2 + b^2} > 0$$

Thus if we take $\operatorname{Im} z > 0$, having $\operatorname{Im}(s_\mu(z)) > 0$, we remain in the upper plane. This property of being a complex analytic map which maps the upper plane in itself defines a special class of bounded analytic function with nice properties (see [29] theorem 3.9 part *c*); in particular, when an asymptotic condition of the form

$$s_\mu(z) = \frac{-1 + o_\mu(1)}{z} \tag{3.33}$$

for z going to infinity with $\frac{|\mathrm{Re}(z)|}{|Im(z)|} = O(1)$ ($o_\mu(1)$ goes to zero for any fixed μ) is respected (as it is for Stieltjes transform), we have a type of *Herglotz function* (see ibidem). In this view the imaginary part of the Stieltjes transform can be seen as the convolution of the measure μ with the poisson kernel $P_b(x) := \frac{1}{\pi} \frac{b}{x^2 + b^2}$, i.e.

$$\operatorname{Im}(s_\mu(a + ib)) = \pi \mu * P_b(a)$$

Thus we can recover the distribution μ from the knowledge of its transform $s_\mu(z)$

$$\lim_{b \to 0^+} \frac{1}{\pi} \operatorname{Im} s_\mu(a + ib) \to \mu, \quad \forall a \in \mathbb{R} \tag{3.34}$$

where the above limit is to be taken with respect to the vague topology[9]. We can thus affirm a theorem similar to 3.1.6

[9]The *vague topology* is the weak$-^*$topology defined on the space of complex Radon measure over a locally compact Hausdorff space X. The vague convergence is performed by $\int_X f d\mu_n \to \int_X f d\mu$, $\forall f \in C_0(X)$ space of the test functions.

Theorem 3.3.4. ***Stieltjes continuity theorem***
Given a sequence of random probability measures μ_n on $\mathbb{R}$, and let be μ a fixed probability measure. Then μ_n converges almost surely to μ with respect to the vague topology if and only if $s_{\mu_n}(z)$ converges in probability to $s_\mu(z)$, $\forall z : \operatorname{Im} z > 0$.

An analogous statement holds for convergence in probability and convergence in expectation. Equation (3.34) is sometimes referred to as "*Stieljes Inversion Theorem´´*. Stieltjes transform can also be used for the retrieval of a distribution μ from the knowledge of its moments using equation (3.34). Indeed, given that moments $m_k(\mu) = \int_{\mathbb{R}} x^k d\mu(x)$ of our distribution exist, we can define a ***moment-generating function***[10]

$$M^\mu(z) = \sum_{k=1}^{\infty} m_k z^k, \quad z \in \mathbb{C} \tag{3.35}$$

which is an analytical function for sufficiently small z inside the convergence radius. Then it is easy to see that, by Taylor expanding the integrand in (3.30), we get the following equation which relates the Stieltjes transform to the moments of the distribution under study (and it is in a certain sense the formal result behind the naive result (3.32)):

$$s_\mu(z) = \frac{1}{z} M^\mu\left(\frac{1}{z}\right) \tag{3.36}$$

for sufficiently large $|z|$. Then using (3.34) we can recover μ. We note that the inversion formula (3.34) works for a fixed μ; that means that in for a series of measures μ_n that equation could give us the analytic form of μ_n for each *fixed* n, but an analytic continuation performed by taking also the limit $n \to \infty$ could bring some convergence issues. The key point is to study the asymptotic behavior of the difference

$$s_{\mu_n} - s_{\mu_{n-1}} \tag{3.37}$$

In particular we are interested in the case where the distribution is the ESD, $\mu_n \equiv \mu_{\frac{1}{\sqrt{n}} M_n}$, and we will focus on this special situation (we will use the shortcut μ_n). For this purpose we slightly modify the way we look at the sequence μ_n. Up to know we considered the matrices M_n as drawn from the

[10] For an introduction to the usage of generating functions see [71].

ensemble of Wigner random matrices. We can look at theorem 3.3 as if the matrix M_{n+1} is M_n where we have added one row and one column compatibile with the ensemble we are studying. An important bound concerning the eigenvalues of M_{n+1} to the ones of M_n is the following:

Lemma 3.3.5. ***Cauchy Interlacing Law***
Given an $n \times n$ Hermitian matrix M_n, we designate by M_{n-1} its top left minor of dimension $n-1 \times n-1$. Then, for all $1 \leq i < n$ we have

$$\lambda_{i+1}(M_n) \leq \lambda_i(M_{n-1}) \leq \lambda_i M_n \tag{3.38}$$

We proceed thus to analyze the difference (3.37). Using the expression (3.16) in the Stieltjes transform, with $z = a+ib$ and $b > 0$, and using Lemma 3.3.5 we get:

$$\begin{gathered}
\sum_{j=1}^{n} \frac{b}{\left(\frac{\lambda_j(M_n)}{\sqrt{n}} - a\right)^2 + b^2} - \sum_{j=1}^{n-1} \frac{b}{\left(\frac{\lambda_j(M_{n-1})}{\sqrt{n}} - a\right)^2 + b^2} \\
\Downarrow \quad \text{expressing through } s_\mu(z) \\
n s_{\mu_{\frac{1}{\sqrt{n}}M_n}}(a+ib) - \sqrt{n(n-1)}\, s_{\mu_{\frac{1}{\sqrt{n-1}}M_{n-1}}}\left(\frac{\sqrt{n}}{\sqrt{n-1}}(a+ib)\right) = O(1) \\
\Downarrow \\
s_{\mu_{\frac{1}{\sqrt{n}}M_n}}(a+ib) = s_{\mu_{\frac{1}{\sqrt{n-1}}M_{n-1}}}(a+ib) + O\left(\frac{1}{n}\right)
\end{gathered} \tag{3.39}$$

where the last passage comes from by expanding $s_{\mu_{\frac{1}{\sqrt{n-1}}M_{n-1}}}\left(\frac{\sqrt{n}}{\sqrt{n-1}}(a+ib)\right)$ and using the complex derivation formula for an integral representation $\frac{d^k}{dz^k}f(z) = \int_\gamma \frac{\partial^k D(z,z')}{\partial z^k} g(z')dz'$, where in our case the integral kernel is the Cauchy integral kernel[11] $\frac{1}{z-z'}$, the contour of integration γ is the real line $\mathbb{R}$ and $g(z')dz' \equiv d\mu(x)$. We get using the trivial condition $\left|\frac{1}{x-z}\right| \leq \frac{1}{|\mathrm{Im}(z)|}$ the general (with respect to the considered μ) the bound

$$\left|\frac{d^k}{dz^k} s_\mu(z)\right| = O_k\left(\frac{1}{|\mathrm{Im}(z)|^{j+1}}\right) \tag{3.40}$$

[11] See footnote 8.

Applying this bound to the case under study we get the last equation of 3.39. From this last result we can obtain two conclusion. First the removal of a row and a column perturbate the Stieltjes transform in a weak way. Indeed after a permutation of the rows (for example) we end up with the same result, where in M_{n-1} the removed row and column are different. This tell us that each row and column can influence the sequence of Stieltjes transforms $s_{\mu_{\frac{1}{\sqrt{n}}M_n}}$ by at most $O\left(\frac{1}{n}\right)$, a result that can be linked to Theorem 3.3.2. Then the independence of the elements in the upper triangular portion and Theorem 3.1.2

$$\mathbb{P}\left(|s_{\mu_n}(z) - \mathbb{E}s_{\mu_n}(z)| \geq \frac{\lambda}{\sqrt{n}}\right) \leq Ce^{-c\lambda^2}$$

for all $\lambda > 0$ and costants $C, c > 0$. Applying the Borel-Cantelli lemma we find that $s_{\mu_n}(z) - \mathbb{E}s_{\mu_n}(z)$ converges almost surely, and thus if $s_{\mu_n}(z)$ converges in expectation it also converges almost surely. So what remains is to determinate what $\mathbb{E}s_{\mu_n}(z)$ is. Taking the expectation of equation (3.31), using linearity and the properties of the trace we have

$$\mathbb{E}s_{\mu_n}(z) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left(\frac{1}{\sqrt{n}}M_n - zI_n\right)^{-1}_{jj} = \mathbb{E}\left(\frac{1}{\sqrt{n}}M_n - zI_n\right)^{-1}_{nn} \tag{3.41}$$

where the last passage follows from the fact that each element of the diagonal of the resolvent is iid. Using *Schur complement*[12], we can express the last term in the diagonal of an inverse matrix as a function of the inverse of the minors. Indeed we have for a matrix A_n

$$(A_n^{-1})_{nn} = \frac{1}{a_{nn} - Y^* A_{n-1}^{-1} X}$$

where a_{nn} is the last element in the diagonal of the initial matrix, while Y^* and X are respectively the last $n-1$ dimensional row and column (respectively) of A_n where a_{nn} has been removed and are used as row and column vectors. Specializing this relation to our situation we get

$$\mathbb{E}s_{\mu_n}(z) = -\mathbb{E}\frac{1}{z + \frac{1}{n}X^*\left(\frac{1}{\sqrt{n}}M_{n-1} - zI_{n-1}\right)^{-1}X - \frac{1}{\sqrt{n}}\xi_{nn}} \tag{3.42}$$

[12]See for reference [47] or [8] chapter $A.1.3$.

where we posed $Y^* = X^*$ because our matrix is Hermitian and ξ_{nn} is the last diagonal entry of M_n. The main point is to note that the quantity $\frac{1}{n}X^*\left(\frac{1}{\sqrt{n}}M_{n-1} - zI_{n-1}\right)^{-1}X$ is the trace of the resolvent of M_{n-1}. Indeed, using the alias for clarity $R := \left(\frac{1}{\sqrt{n}}M_{n-1} - zI_{n-1}\right)^{-1}$, we have that $(X^*RX)^{1/2} = \left\|R^{1/2}X\right\|$ is Lipschitz bounded function in X. Using theorem 3.1.3, we conclude that X^*RX is concentrated near its mean:

$$\mathbb{P}\left(|X^*RX - \mathbb{E}X^*RX| \geq \lambda\sqrt{n}\right) \leq Ce^{-c\lambda^2} \tag{3.43}$$

It remains to study what is the mean. Expressing in components:

$$\mathbb{E}X^*RX = \sum_{i=1}^{n-1}\sum_{j=i}^{n-1}\mathbb{E}\overline{\xi_{in}}r_{ij}\xi_{jn} \tag{3.44}$$

Conditioning on a single realization of R, and thus considering it as fixed, we have that equation (3.44) is exactly the trace of the particular realization R. Using the *disintegration theorem* we can return to the case of random R, getting that

$$\begin{aligned}\mathbb{E}X^*RX = \mathrm{tr}\{R\} =& n\frac{\sqrt{n}}{\sqrt{n-1}}s_{\mu_{n-1}}\left(\frac{\sqrt{n}}{\sqrt{n-1}}z\right)\\ &\Big\Downarrow \quad \text{using (3.39)}\\ \mathrm{tr}\{R\} = n\left(s_{\mu_n}(z) + o(1)\right) &\Rightarrow X^*RX = n\left(\mathbb{E}s_{\mu_n}(z) + o(1)\right)\end{aligned} \tag{3.45}$$

Using the results we have just obtained in equation (3.42), taking the $n \to \infty$ limit and using *Arzelá-Ascoli theorem* we have the final self consistent equation

$$\mathbb{E}s_{\mu_n}(z) = -\frac{1}{z + \mathbb{E}s_{\mu_n}(z)} + o(1) \Rightarrow s_\mu(z) = -\frac{1}{z + s_\mu(z)} \tag{3.46}$$

Solving it and taking the branch corresponding to condition (3.33) we get

$$s_\mu(z) = \frac{-z + \sqrt{z^2 - 4}}{z} \Rightarrow \mu = \frac{1}{2\pi}\left(4 - x^2\right)_+^{\frac{1}{2}}dx \tag{3.47}$$

where the last passage come from Stieltjes inversion formula (3.34).

3.3.3 Rate of convergence

The use of Stieltjes transform machinery described in the last section can be used to answer to an important question: at what rate do the distributions $\mu_{\frac{1}{\sqrt{n}}M_n}$ converge to the limit spectral distribution μ_{sc} of equation (3.25). We remark that the usage of moment method is not suitable for this purpose and this question has been left without answer until Zhidong Bai's papers in 1993 [6,7]. We will give only the basic results of this issue, because it requires some mathematical effort. The details can be found in the two mentioned papers by Bai and in [8]. As before we will suppose that our matrix M_n has centered entries ξ_{ij} with unit variance in the triangular elements but we let the diagonal elements having variance σ^2. We further require that for any n, the 3$^{\text{rd}}$ and 6$^{\text{th}}$ moments of the entries are bounded by a constant M, $\mathbb{E}\left|\xi_{ij}^3\right|, \mathbb{E}\left|\xi_{ij}^6\right| \leq M$. The point is to evaluate the $o(1)$ term in the first part of equation (3.46) which, of course, depends on n. We will call it for the moment δ_n. Analyzing equation (3.42), we see that the error in 3.46) can be expressed by

$$\delta_n = -\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\frac{\frac{1}{\sqrt{n}}\xi_{nn} - \frac{1}{n}X^*\left(\frac{1}{\sqrt{n}}M_{n-1} - zI_{n-1}\right)^{-1}X + \mathbb{E}s_{\mu_n}(z)}{\left(z + \mathbb{E}s_{\mu_n}(z)\right)\left(z - \frac{1}{\sqrt{n}}\xi_{nn} + \frac{1}{n}X^*\left(\frac{1}{\sqrt{n}}M_{n-1} - zI_{n-1}\right)^{-1}X\right)}$$

Working on this quantity as described in the references, we can conclude the following:

Theorem 3.3.6. ***Rate of convergence***
Under the above assumptions we have the following results

$$\|\mathbb{E}\mu_n - \mu_{sc}\| = O\left(\frac{1}{\sqrt{n}}\right) \tag{3.48}$$

$$\|\mathbb{E}\mu_n - \mu_{sc}\| = O_p\left(\frac{1}{n^{\frac{2}{5}}}\right) \tag{3.49}$$

$$\|\mathbb{E}\mu_n - \mu_{sc}\| = O_{a.s.}\left(\frac{1}{n^{\frac{2}{5}+\varepsilon}}\right), \quad \forall\varepsilon > 0 \tag{3.50}$$

$$\tag{3.51}$$

where $\|\cdot\|$ denotes the Kolmogorov distance as used in theorem 3.3.2.

Above O_p and $O_{a.s.}$ denotes the convergence rate in *probability* and *almost surely* (e.g. for $X_n = O_p(n^{-\alpha})$ means that $\forall\delta > 0$, there exists two constants

$M, N > 0$ such that$\mathbb{P}\left(|X_n n^\alpha| > M\right) < \delta$, $\forall n > N$; an analogous definition holds for the almost surely condition).

3.3.4 Physical methods

The analysis of random matrix in statistical physics is made through two different methods. The first one, from a conceptual and historical point of view is ***replica method***, followed by its evolution (in a certain sense) the ***cavity method***.

Replica method. Initially developed for spin glasses models by Edwards and Anderson in their seminal paper [23] in 1975, had more and more attention on it after its use made by Sherrington and Kirkpatrick in [39] for developing their famous model. Always Edwards with Jones in 1976 extended this method to the analysis of large random matrices [24]. A modern review can be found in [41] adapted. Substantially, the replica method tries to evaluate the mean free energy f of a given system that is proportional to the logarithm of the partition function $\mathcal{Z}$. Being an observable depending on some random variables which represent the inner disorder of the system under study, through $\mathcal{Z}$, the average on the disorder of the system must be done over the free energy itself to obtain its mean value. This create an issue, because even for the most simple kind of disorder, the gaussian one, entailed in the distribution of some variables inside the partition function, the average process is prohibitive through the logarithm

$$\langle f \rangle \sim \int \prod_{ij \cdot m} d(f(x_{ij\cdots m})) \log \mathcal{Z}(x_{ij\cdots n})$$

where the $x_{ij\cdots m}$ is a set of variables on which we are performing the average and $d(f(x_{ij\cdots m}))$ is their distribution. This inconvenience is avoided recurring to a trick. In fact $\log x$ can be always expressed by the limit

$$\log x = \lim_{n\to 0} \frac{x^n - 1}{n} \tag{3.52}$$

Thus we have, taking also the thermodynamic limit ($N := \sum x_{ij\cdots m} \to \infty$) on our mean free energy

$$\lim_{N\to\infty} \langle f \rangle \sim \lim_{N\to\infty} \int \prod_{ij\cdots m} d(f(x_{ij\cdots m})) \lim_{n\to 0} \left(\frac{\mathcal{Z}^n - 1}{n} \right)$$

Now the average can be done more easily. An assumption in the replica trick is that in $\mathcal{Z}^n = \prod_\alpha^n \mathcal{Z}_\alpha$ each $\mathcal{Z}_\alpha$ is independent from the other. We are thus *replicating* the system under study in several independent copies, and the searched result will be given taking the zero limit of this ensemble. Since its publication the replica trick suffered of criticism for the entailed mathematical problems. For example in [68] doubts are posed on the analytic continuation required passing from integer n to real n performing the limit, and on the tricky exchange of limits $\lim_{N\to\infty} \leftrightarrow \lim_{n\to 0}$. As said pointed out by Bai in [9] replica approach seems to start at the same point of Stieltjes transform in paragraph 3.3.2, but while the Stieltjes method proceeds in a rigorous way, the introduction in the replica approach of the Hubbard-Stratonovich transform for decoupling the variables could be the *non rigorous* and most tricky point. The problematical connection of the replica approach to one of the known and well behaved mathematical methods is far from being solved, but as pointed out in [63] it is probable that more conditions should be required fo perform safely the replica trick.

The replica method for ESD of large matrices starts from the observation that equation (3.16) can be expressed by *Sokhotski-Plemelj theorem*. Indeed taken a real symmetric $N \times N$[13] matrix M_N, with eigenvalues $\{\lambda_i(M_N)\}_{1\le i\le n}$, adding a small negative imaginary part $-i\varepsilon$ we get

$$\lim_{\varepsilon\to 0^+} \frac{1}{\lambda - \lambda_i(M_N) - i\varepsilon} = \mathcal{P}\left(\frac{1}{\lambda - \lambda_i(M_N)}\right) + i\pi\delta\left(\lambda - \lambda_i(M_N)\right) \tag{3.53}$$

Thus we can express μ_{M_n} of equation (3.16) as

$$\begin{aligned} \mu_{M_N} = \frac{1}{N}\sum_{j=1}^{N} \delta_{\lambda_j(M_N)} &= \lim_{\varepsilon\to 0^+} \frac{1}{\pi N} \sum_{i=1}^{N} \operatorname{Im}\left(\frac{1}{\lambda - \lambda_i(M_N) - i\varepsilon}\right) \\ &= \lim_{\varepsilon\to 0^+} \operatorname{tr}\left(\frac{1}{(\lambda - i\varepsilon)I - M_N}\right) \end{aligned} \tag{3.54}$$

At this point we are going to introduce the logarithm to take advantage of

[13] We have switched from n to N for the dimension of the matrices because we will use n for the replica index as is done in literature.

its properties.

$$\ln\left(\det((\lambda - i\varepsilon)I - M_n)\right) = \ln\left(\prod_{j=1}^{N}((\lambda - i\varepsilon) - \lambda_j(M_n))\right)$$
$$= \sum_{j=1}^{N} \ln\left((\lambda - i\varepsilon) - \lambda_j(M_n)\right) \tag{3.55}$$
$$\Downarrow$$
$$\frac{\partial}{\partial\lambda}\left(\ln\det\left((\lambda - i\varepsilon)I - M_n\right)\right) = \mathrm{tr}\left((\lambda - i\varepsilon)I - M_N\right)^{-1} \tag{3.56}$$
$$\Downarrow$$
$$\mu_{M_N} = \lim_{\varepsilon\to 0^+} \frac{1}{\pi N}\,\mathrm{Im}\frac{\partial}{\partial\lambda}\left[\ln\det\Big((\lambda - i\varepsilon)I - M_n\Big)\right] \tag{3.57}$$

Now what we are left to inspect is the quantity $\det\left((\lambda - i\varepsilon)I - M_n\right)$. As already done for the Stieltjes transform, we will work with the complex variable $z := \lambda - i\varepsilon$ and the issues will be present near the real axis. From the standard use of gaussian integral in physics (for example in statistical physics or in quantum field theory) it is well known that if we treat $(zI - M_N)$ as as the matrix building up the quadratic form corresponding to the hamiltonian of a physical system, the quantity $\det^{-\frac{1}{2}}(zI - M_N)$ can be calculated through a *Fresnel integral*:

$$\det{}^{-\frac{1}{2}}(zI - M_N) =$$
$$= \left(\frac{e^{i\frac{\pi}{4}}}{\sqrt{\pi}}\right)^N \int_{-\infty}^{+\infty} \prod_{k=1}^{N} dx_k \exp\left(-i\sum_{i,j=1}^{N} x_i (zI - M_N)_{ij} x_j\right) \tag{3.58}$$

Before continuing two little remarks are needed. The first point that must be highlighted is that it is easy to see looking at equation (3.56) that this is the same starting point of Stieltjes transform method. While we used formal mathematical manipulation in the dedicated section, here we perform a calculation based with a statistical physics analogy. The replica trick (3.52), which will used in the following lines, relies on an analytical continuation which could bring formal problems and also conceptual ones, and on some ansatz that may not be satisfied. The second point is that while the logarithm appears in a natural way in the evaluation of the *mean free*

energy in spin glass models, here it is not actually needed conceptually, but as tool to perform computation. We are thus mapping our pure mathematical problem to one interacting particle problem that we can solve, where the statistical mechanics is introduced via the partition function

$$\overline{\mathcal{Z}_{M_N}}(z) = \left(\frac{e^{i\frac{\pi}{4}}}{\sqrt{\pi}}\right)^N \int_{-\infty}^{+\infty} \prod_{k=1}^{N} dx_k \exp\left(-i \sum_{i,j=1}^{N} x_i (zI - M_N)_{ij} x_j\right) \tag{3.59}$$

$$\Downarrow \quad \text{substituting in (3.57)}$$

$$\mu_{M_N} = \lim_{\varepsilon \to 0^+} \frac{-2}{\pi N} \operatorname{Im}\left(\frac{\partial}{\partial z} \ln \mathcal{Z}_{M_N}(z)\right)_{z=\lambda - i\varepsilon} \tag{3.60}$$

It must be noted that this is a fictional correspondence. Indeed what we would call the *hamiltonian* of our system is a complex function $\mathcal{H}_{\mathcal{M}_\mathcal{N}} = i \sum_{i,j=1}^{N} x_i \left(zI - M_N\right)_{ij} x_j$, and we cannot define a proper measure over the ensemble of the states of the system. We introduced the notation $\overline{\mathcal{Z}_{M_N}}(z)$ for distinguish this fictional complex partition function to the real one of the system that we will exploit for *cavity method*. Nevertheless, continuing with the statistical physics analogy, we have the free energy of the system (which ultimately is the ESD μ_{M_N}) given by $f = \ln \mathcal{Z}$. The random nature of M_N suggest us to perform and average over the possible elements of the ensembles. For simplifying the calculations, we will take the $(M_N)_{ij}$ distributed by the normal law $N(0, \sigma^2)$. The limit empirical spectral distribution μ_{ens} for the ensemble is thus given by

$$\mu_{ens} = \int \mu_{M_N} \prod_{i,j=1}^{N} \left[\frac{\exp\left(\frac{-(M_N)_{ij}^2 N^2}{2J^2}\right)}{\left(2\pi \frac{J^2}{N}\right)^{\frac{1}{2}}} d(M_N)_{ij}\right] \tag{3.61}$$

where we have substituted $\sigma^2 = \frac{J^2}{N}$. It is straightforward that the factor N appearing there is nothing else than the $\frac{1}{\sqrt{N}}$ normalization that we have done in the previous paragraphs. Now we will apply the trick (**??**) to the above equation. As already said, we will have $\mathcal{Z}^n = \prod_\alpha^n \mathcal{Z}_\alpha$ each $\mathcal{Z}_\alpha$ is independent from the other, and α is the replica index. We are thus doing conceptually as if we have replicated our system in an integer number of copies, taking then the $n \to 0$ limit passing through not only integer numbers. Thus the

expression of μ_{M_N} to be inserted in (3.61) is

$$\mu_{M_N} = \frac{-2}{N\pi}\,\mathrm{Im}\,\frac{\partial}{\partial z}\lim_{n\to 0}\frac{1}{n}\left[\left(\frac{e^{i\frac{\pi}{4}}}{\sqrt{\pi}}\right)^{nN}\int_{-\infty}^{+\infty}\left(\prod_{\beta=0}^{n}\prod_{k=1}^{N}dx_k^\beta\right)\times\right.$$
$$\left.\times\exp\left(-i\sum_{\alpha=0}^{n}\sum_{i,j=1}^{N}x_i^\alpha(zI-M_N)_{ij}x_j^\alpha\right)-1\right] \tag{3.62}$$

The replica trick permits us to perform the average on the ensemble. Thus, substituting the above in equation (3.61), carrying out the gaussian integration and retaining only the terms linear in n and leading-order in N in view of the two limits we are going to perform (the $n\to 0$ and the thermodynamical one $N\to\infty$) we are left with:

$$\mu_{ens} = \frac{-2}{N\pi}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\lim_{n\to 0}\frac{1}{n}\left[\left\{\left(\frac{e^{i\frac{\pi}{4}}}{\sqrt{\pi}}\right)^{N}\times\right.\right.$$
$$\left.\left.\times\int_{-\infty}^{+\infty}\left(\prod_{k=1}^{N}dx_k\right)\exp\left[-i\lambda\sum_{i=1}^{N}x_i^2-\frac{J^2}{N}\left(\sum_{i=1}^{N}x_i^2\right)^2\right]\right\}^n-1\right] \tag{3.63}$$

Where the replica index α has been factored out and absorbed in the n exponent. The integral in (3.63) may be simplified by an *Hubbard-Stratonovich* transform

$$\exp\left[\frac{-J^2}{N}\left(\sum_i(x_i)^2\right)^2\right] = \left(\frac{N}{2\pi}\right)^{\frac{1}{2}}\frac{\lambda}{\sqrt{2J^2}}\int_{-\infty}^{+\infty}ds\exp\left(\frac{-\lambda^2}{4J^2}Ns^2-i\lambda s\sum_i(x_i)^2\right)$$

Substituting back in (3.63) and doing the Fresnel integral we are left with an integral inside that equation of the form

$$I = \left[\left(\frac{n}{2\pi}\right)^{\frac{1}{2}}\left(\frac{\pi^N}{2J^2}\right)^{\frac{1}{2}}\lambda\exp\left(\frac{-N}{2}\ln\lambda\right)\int_{-\infty}^{+\infty}ds\exp\left(-Ng(s)\right)\right] \tag{3.64}$$

$$g(s) = \frac{\lambda^2 s^2}{4J^2}+\frac{1}{2}\ln[i(1+s)]$$

This last integral must be performed through *saddle-point* integration. We end up, after the $n\to 0$ limit with

$$\mu_{ens} = \frac{-2}{N\pi}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\left(-Ng(s_0)-\frac{N}{2}\ln\lambda\right)$$

Where $s_0 = \frac{1}{2}\left[1 - i\left(\frac{4J^2}{\lambda^2} - 1\right)^{\frac{1}{2}}\right]$ is the correct saddle point of our integration[14]. Carrying out the differentiation we end up with

$$\mu_{ens} = \frac{1}{2\pi J^2}\left(4J^2 - \lambda^2\right)^{\frac{1}{2}} \equiv \mu_{sc}$$

So we have retrieved Wigner semicircular law where the distribution of the entries of the *normalized* matrix has variance J^2. Despite the apparent clarity, the replica method even for the ESD determination, present several mathematical issues. One of the first paper criticizing the systematic usage of this method for the ESD retrieval was [69].

Cavity method The other approach used is cavity method, developed by Parisi with Mézard and Virasoro [48]. Cavity method is the natural evolution of TAP equation used for a mean field approach to spin glasses. It focus on a precise site j, and studies the marginal probabilities of the branches attached to j in the approximation (reasonable for big systems locally treelike) that these are independent once we excluded the site j from the system. On each subtree a recursive approach is applied, giving a **belief propagation** algorithm that applied recursively provides a quick convergence, provided that the system is not too much correlated. A modern review applied to sparse matrices can be found in [56]. Incidentally we highlight that the philosophy at the base of cavity method reduces to the study of certain paths on the system and is substantially the same of moment method, although more light should be shed on this point, especially for the non trivial requirement in cavity method of decoupled branches starting from a point j. For the graph theoretical treatment used, this is a reasonable requirement, because for sufficiently diluted random graphs, the creation of loops has a decreasing probability with the asymptotic of $O(N^{-1})$ (see [52] for example). The starting point of the cavity method, as for the replica approach is the analogy with a mechanical system of N interacting particles whose positions are indexed by x_i. Thus we arrive in the same ways described above to equation (3.60) which is the analogous of the free energy of the system. The point here is that we want to define (unlike for the replica approach in which this correspondence was more conceptual and non-essential) a *Gibbs-Boltzman* measure for the states of our system, thus, we have to deal with the fact

[14] See [24] for details on the integration process.

that the hamiltonian of our system is a complex function. First we can get rid of the complex unit making a variable substitution, and changing rotating the integration path in the complex plan such that its new extrema are $]-\theta\cdot\infty, +\theta\cdot\infty[$ with $\theta^2 = i$. The negative imaginary part of z will ensure convergence of the integral. At this point it is possible to introduce a proper partition function and the Gibbs measure for our states:

$$\left.\begin{array}{l} \mathcal{Z}_{M_N}(z) = \int_{-\theta\cdot\infty}^{+\theta\cdot\infty} \frac{d\boldsymbol{x}}{(2\pi)^{\frac{N}{2}}} \exp\left(-\mathcal{H}_{M_N}(\boldsymbol{x}, z)\right) \\ \\ \mathcal{H}_{M_N}(\boldsymbol{x}, z) = \frac{1}{2}\sum_{i,j\in\mathcal{G}_{\mathcal{M}_N}} x_i (zI - M_N)_{ij} x_j \end{array}\right\} \Longrightarrow \mathbb{P}_{M_N}(\boldsymbol{x}) = \frac{1}{\mathcal{Z}_{M_N}(z)} e^{-\mathcal{H}_{M_N}(\boldsymbol{x},z)} \tag{3.65}$$

where $\boldsymbol{x}$ is the vector of the position of the interacting particles and restricted the summation to a graph $\mathcal{G}_{M_N} = (V, E)$[15], where the vertices V of the graph represent our particles, and an edge will be present in the set of the graph's edges E if the two related particles do interact. This is conceptually the same as looking M_N as a special weighted adjacency matrix of an undirected graph. Before continuing we point out that the formal realization of a statistical mechanics through the passages we have made (including the change of variables and the path of integration in the complex plane) is sometimes avoided in literature as in [56] without particular worries on imaginary factors and convergence issues. The same author in [57] has presented a treatise in which he avoids the complete statistical mechanics analogy working with pseudo physical complex quantities, e.g. the partition function in (3.59). We preferred to expose instead the various passages which brought us to equation (3.3.4) and will omit the integration extrema for clearness in the following formulas. In this formal setting we can thus rewrite equation (3.60) in the following way

$$\mu_{M_N} = \lim_{\varepsilon\to 0^+} \frac{1}{\pi N} \sum_{i=1}^{N} \operatorname{Im}\left[\langle x_i^2\rangle_z\right]_{z=\lambda-i\varepsilon} \tag{3.66}$$

where $\langle\cdots\rangle_z$ indicates the average with respect to Gibbs distribution (3.3.4). If we suppose that the graph obtained from the particle analogy is *locally treelike*, choosing a site x_i with connectivity k_i, the k_i branches originating from it are reciprocally poorly interacting except through x_i itself. Speaking with the probability distribution $\mathbb{P}_{M_N}(\boldsymbol{x})$, which is nothing else that the *Joint Probabilty Density Function* (JPDF) for the variables $x_1, \ldots, x_N$, can

[15] See section 5 for the graph theory definitions.

be expressed as

$$\mathbb{P}_{M_N}(\boldsymbol{x}) = \frac{1}{\mathcal{Z}_{M_N}(z)} e^{-\mathcal{H}_{M_N}(\boldsymbol{x},z)} = \frac{1}{\mathcal{Z}_{M_N}(z)} \prod_{(i,j)\in E} \psi_{ij}(x_i, x_j) \prod_{i\in V} \phi_i(x_i) \quad (3.67)$$

where ψ represent the interaction part corresponding to M_N in $\mathcal{H}_{M_N}$ and ϕ the self-interacting part competing to zI. If the graph is a tree as we supposed it to be, removing a the selected site x_i, equation (3.67) should factor out as

$$\mathbb{P}^{(i)}_{M_N}(\boldsymbol{x}^{(i)}) = \frac{1}{\mathcal{Z}^{(i)}_{M_N}(z)} \prod_{\ell=1}^{k_i} \left(\prod_{(h,j)\in E^{(i)}_\ell} \psi_{hj}(x_h, x_j) \prod_{j\in V^{(i)}_\ell} \phi_j(x_j) \right) \quad (3.68)$$

where in general the superscript (i) denotes that we are considering the graph where we have removed the i^{th} vertex and its related features (and hence $E^{(i)}_\ell$ and $V^{(i)}_\ell$ represent respectively the edges and the vertices of the ℓ^{th} branch departing from x_i). This type of graph is known as *cavity graph*, and the treelike approximation[16] we are doing is sometimes referred to as *Bethe approximation*. Each of the single k_i factors appearing in formula (3.68) is the joint probability distribution of the considered branch. The cavity method gives us a straightforward way to calculate the marginal probability distribution[17] of each site x_i, $\mathbb{P}_i(x_i)$. Indeed, indicating with ∂i the set of neighbors sites of i, we have:

$$\begin{aligned}
\langle x_i^2 \rangle_z &= \frac{1}{\mathcal{Z}_{M_N}(z)} \int \left(\prod_{k=1}^{N} \frac{dx_k}{\sqrt{2\pi}} \right) x_i^2 e^{\left(-\mathcal{H}_{M_N}(\boldsymbol{x},z)\right)} = \\
&= \frac{1}{\mathcal{Z}_{M_N}(z)} \int \left(\prod_{k=1}^{N} \frac{dx_k}{\sqrt{2\pi}} \right) x_i^2 e^{\left(-\frac{1}{2}\sum_{i,j\in\mathcal{G}_{\mathcal{M}_N}} x_i (zI - M_N)_{ij} x_j\right)} = \\
&= \frac{1}{\mathcal{Z}_{M_N}(z)} \int \frac{dx_i}{\sqrt{2\pi}} x_i^2 \exp\left(\frac{-zx_i^2}{2}\right) \int \prod_{k\in V\setminus i}^{N} \left(\frac{dx_k}{\sqrt{2\pi}}\right) \times \\
&\times \exp\left(x_i \sum_{\ell\in\partial i} \{M_N\}_{i\ell} x_\ell \right) \exp\left(\sum_{h\in V\setminus\{i\}} \frac{-zx_h^2}{2} + \sum_{h,j\in V\setminus\{i\}} x_h \{M_N\}_{hj} x_j \right)
\end{aligned} \quad (3.69)$$

[16] We call it approximation because in real lattice systems the treelike structure is reached only locally and neglecting eventual cycles present in the graph.

[17] Given a set of random variables X_i whose joint probability distribution is known, the marginal probability of X_k is obtained by averaging the joint distribution with respect to all the other variables

Thus we see that the marginal distribution at the site i is

$$\begin{aligned}\mathbb{P}_i(x_i) = \exp\left(\frac{-zx_i^2}{2}\right)\int\prod_{k\in V\setminus i}^{N}\left(\frac{dx_k}{\sqrt{2\pi}}\right)\times \\ \exp\left(x_i\sum_{\ell\in\partial i}\{M_N\}_{i\ell}x_\ell\right)\exp\left(\sum_{h\in V\setminus\{i\}}\frac{-zx_h^2}{2}+\sum_{h,j\in V\setminus\{i\}}x_h\{M_N\}_{hj}x_j\right)\end{aligned} \tag{3.70}$$

which is nothing than

$$\int\left(\prod_{k\in V\setminus i}^{N}\frac{dx_k}{\sqrt{2\pi}}\right)\mathbb{P}_{M_N}(\boldsymbol{x})$$

consistently with the marginal probability definition. Now it is not difficult to see that the second exponential which we are integrating in (3.70) can be written in terms of the probability distribution over the disconnected branches of the cavity graph $\mathcal{G}_{M_N}^{(i)}$, $\mathbb{P}_j^{(i)}(x_j)$, which are called the *cavity distributions*. These can be expressed recursively (i.e. applying again the cavity argument on each branch starting from x_i) by

$$\mathbb{P}_j^{(i)}(x_j) = \frac{e^{-\frac{1}{2}zx_j^2}}{\mathcal{Z}_j^{(i)}}\int d\boldsymbol{x}_{\partial j\setminus i}\exp\left(x_j\sum_{\ell\in\partial j\setminus i}\{M_N\}_{j\ell}x_\ell\right)\prod_{k\in\partial j\setminus i}\mathbb{P}_k^{(j)}(x_k) \tag{3.71}$$

where with $\boldsymbol{x}_{\partial i}$ we denote the set of variables adjacent to x_i. The marginal distribution of x_i can be thus written as

$$\mathbb{P}_i(x_i) = \exp\left(\frac{-zx_i^2}{2}\right)\int d\boldsymbol{x}_{\partial i}\exp\left(x_i\sum_{\ell\in\partial i}\{M_N\}_{i\ell}x_\ell\right)\prod_{\ell\in\partial i}\mathbb{P}_\ell^{(i)}(x_\ell) \tag{3.72}$$

The equation of cavity distribution is automatically fulfilled for the $\mathbb{P}_j^{(i)}(x_j)$ in the form

$$\mathbb{P}_j^{(i)}(x_j) = \frac{1}{\sqrt{2\pi\Delta_\ell^{(i)}}}e^{-\frac{1}{2\Delta_\ell^{(i)}}x^2}$$

With this substitution in equation (3.3.4) we obtain the equation for the variances $\Delta_j^{(i)}(z)$

$$\Delta_j^{(i)} = \frac{1}{z-\sum_{\ell\in\partial j\setminus i}\{M_N^2\}_{j\ell}\Delta_\ell^{(j)}(z)} \tag{3.73}$$

With this solution even the marginals are Gaussian with variances Δ_i, and equation (3.72) gives the following equation for the variances:

$$\Delta_i(z) = \frac{1}{z - \sum_{\ell \in \partial i} \{M_N^2\}_{i\ell} \Delta_\ell^{(i)}(z)} \tag{3.74}$$

It must be noted that because z is a complex number, cavity variances are in general complex. Setting $M_N = \frac{A}{\sqrt{c}}$, where $c = \frac{1}{N}\sum_{i=1}^{N} k_i$ is the *mean connectivity* and A is a symmetric gaussian matrix with entries with zero mean and variance J^2, we are interested in the large c limit. Defining

$$\Delta = \lim_{c\to\infty} \frac{1}{c} \sum_{\ell \in \partial i} \Delta_\ell$$

we get

$$\lim_{c\to\infty} \sum_{\ell \in \partial i} \{M_N^2\}_{i\ell} \Delta_\ell^{(i)} = \lim_{c\to\infty} \frac{1}{c} \sum_{\ell \in \partial i} J_{i\ell}^2 \Delta_\ell^{(i)} = J^2 \Delta$$

. Performing the limit and using equation (3.74) and we retrieve Wigner semicircle law:

$$\Delta = \frac{1}{z - J^2\Delta} \Longrightarrow \mu_{M_N} = \frac{1}{2\pi J^2} \sqrt{4J^2 - \lambda^2} \tag{3.75}$$

Chapter 4

Free Probability

In chapter 2 we described the formal setting of probability theory. In this framework, random variables were defined (see definition 2.3) as measurable functions from the considered set of possible events Ω to a target space R. The explicit reference to the underlying probability space in the definition of random variables gives some constraints we want to get rid of. We already mentioned the necessity of defining a formal way to extend our probability space when we make particular operations such as taking limits. Indeed the underlying space of events Ω could depend upon the index on which we are taking the limit. This point is almost always understood, but, anyway, remains a not negligible feature of our theory, in situation such as random matrix theory where a lot of these operations are taken. For what is more, if we could abstract our theory we should be able of making interesting generalizations. The idea we use is inherited from *non-commutative geometry* [18]. As we can look at the algebra of functions on a manifold to extract informations about the manifold, here we look at the *algebra* of random variables as an independent object without caring too much on the set of possible events it lives upon. Classical random variables belong to *abelian algebra*. The main point we want to generalize is the possibility that our random variables could not be commutative objects, as for example the random matrices of last chapter. A key feature of classic probability theory is *expectation*. We have difined it through and integral[1], which is of course a linear functional. Based on the above assumption, we can make the following definition:

[1] A sum for discrete probability distributions, but we could always use the integral with the Dirac measure, as we have done for ESD.

Definition 4.1. ***Non-commutative probability space***
A ***non-commutative probability space*** *is a* **unital** $*-$**algebra** $\mathcal{A}$ *with a* $*-$*linear functional* $\tau{:}\mathcal{A} \to \mathbb{C}$*, preserving the identity* $\tau(\mathbf{1}_{\mathcal{A}}) = 1$ *and* ***positive****. i.e. with* $\tau(XX^*) \geq 0$*,* $\forall X \in \mathcal{A}$*; if we have that* $\tau(XX^*) = 0 \iff X = 0$*, we say that* τ *is* ***faithful****. Given* $X, Y \in \mathcal{A}$ *if we further have that* $\tau(XY) = \tau(YX)$*,* τ *is termed* ***trace****.*

Important examples of the algebra $\mathcal{A}$ are C^*-algebras, Von Neumann algebras (which are sometimes called W^*-algebras) and bounded operator algebras on an Hilbert space $\mathcal{B}(\mathcal{H})$ (though this necessitates the *Gelfand-Naimark-Segal construction*, see [65]). A powerful result in characterization of Von Neumann algebras is the following:

Theorem 4.0.1. ***Von Neumann double commutant theorem***
Let $\mathcal{M} \subseteq \mathcal{B}(\mathcal{H})$ *a complex algebra of bounded over an Hilbert space* $\mathcal{H}$*, which is closed under taking adjoints and is unital. Then the closure of* $\mathcal{M}$ *under the strong operator topology is the same as the closure under weak operator topology. The closure is given by the double commutant of* $\mathcal{M}$*),* $\mathcal{M}''$*, and is a Von Neumann algebra, where*

$$\begin{aligned} \mathcal{M}'' &= (\mathcal{M}')' \\ \mathcal{M}' &= \{Y \in \mathcal{B}(\mathcal{H}) \mid YX = XY, \quad \forall X \in \mathcal{M}\} \end{aligned} \tag{4.1}$$

For our purposes, and what we have seen until this chapter it will be useful to introduce the concept of *random matrix* within this formal setting. The naive definition of a random matrix is that of an array of a given dimension filled with random variables. Random variables are, of course, classical commuting objects belonging to the algebra of summable functions; in particular we want functions where all moments are well defined, i.e. $X \in \overline{L}(\Omega) := \cup_{k=1}^{\infty} L^k(\Omega)$. Thus we have that $\mathcal{A} \equiv \overline{L}$. The functional τ is represented by the expectation $\mathbb{E}$, while the *involution* $* : \mathcal{A} \to \mathcal{A}$ is performed taking the complex conjugate for complex random variables, i.e. $*(X) \mapsto \overline{X}$. The trace property of the expectation is straightforward, while the positivity comes from the properties of complex numbers. A deterministic $n \times n$ matrix with complex entries belongs to the space $\mathcal{M}_{n\times n}(\mathbb{C})$. It is easy to see that in this case $\mathcal{A} \equiv \mathcal{M}_{n\times n}(\mathbb{C})$ possesses all the properties required and in this case the space is non-commutative. The involution $*$ is given by the adjoint of the matrix M, $*(M) = M^\dagger$. As we have already seen in section 3.2, a natural candidate for the linear functional τ is the reduced trace of the matrix, $\tau(M) := \frac{1}{n}\operatorname{tr}(M)$. The positivity and the tracial

properties of τ come naturally from standard trace properties. Now the idea of representing our random matrix M_r (where the subscript stresses out the *random* character of the matrix) through a randomly filled array can be formalized; we can indeed take the algebra $\mathcal{A} := \overline{L}(\Omega) \otimes \mathcal{M}_{n\times n}(\mathbb{C})$, which is in a naive sense the composition of the already analyzed cases. For what concerns the involution, this remains as for $\mathcal{M}_{n\times n}(\mathbb{C})$: indeed the adjoint operation already includes the standard complex involution we have described above. Taking the reduced trace of an element M_r of our algebra, we end up with a sum of random variables. Indeed we have $\frac{1}{n}\operatorname{tr}(M_r) = \frac{1}{n}\sum_i^n X_{ii}$, where the X_i are the random variables appearing in the diagonal of M_r, and which actually can itself be seen as random variable in $\overline{L}(\Omega)$. The most natural and correct way to define a linear functional compatible with our algebra is then taking the expectation of the obtained trace, i.e. $\tau(M_r) = \mathbb{E}\frac{1}{n}\operatorname{tr}(M_r)$. As a self-consistence check, we can look at the measure introduced by this linear functional. Suppressing the subscript r, we must have:

$$\tau(M^k) = \int_{\mathbb{C}} z^k d\mu_M(z) = \mathbb{E}\frac{1}{n}\operatorname{tr}(M) \tag{4.2}$$

We see clearly that we must retrieve the ESD $\mu_M = \frac{1}{n}\sum_{j=1}^n \delta_{\lambda_j(M)}$ if M is deterministic, or the expected[2] ESD $\mu_M = \mathbb{E}\frac{1}{n}\sum_{j=1}^n \delta_{\lambda_j(M)}$ if M is a random variable. Dealing with non commuting random variables we can ask if the quantity

$$\tau\left(M^k (M^\dagger)^\ell\right) = \int_{\mathbb{C}} z^k \overline{z}^\ell d\mu_M(z) \tag{4.3}$$

From spectral theorem, we know that if M is normal, i.e. $MM^\dagger = M^\dagger M$, the above expression is well defined, but in general it is not, and this is a first big difference with classical probability theory. Generalizing to general algebras with possibly non-adjoint elements X, X^*, we ask ourselves if given a function $f(X, X^\dagger)$ taking values in $\mathcal{A}$, is always possible to define a possibly complex measure such that the quantity

$$\tau\left(f(X, X^*)\right) = \int_{\mathbb{C}} f(z, \overline{z}) d\mu(z)$$

is always well defined. We can simplify the problem using *Stone-Weierstrass* theorem, and reducing to calculate the above quantity for f being a polynomial in X and X^*, $P(X, X^*)$. If we work with self-adjoint elements, which

[2]Rember the discussion on the expected empirical distribution of section 3.2.

is the case, we will need only to evaluate $P(X)$. Using *Riesz representation* theorem with some expedients, we get the following one

Theorem 4.0.2. ***Operator Spectral Theorem***
Given a probability space $(\mathcal{A}, \tau)$ with $X \in \mathcal{A}$ a bounded self-adjoint element, there always exists an unique Borel measure μ_X on $[-\rho(X), \rho(X)]$, with $\rho(X)$ spectral radius of X, such that

$$\tau(P(x)) = \int_{\mathbb{C}} P(z) d\mu_X(z)$$

for any polynomial $P : \mathbb{C} \to \mathbb{C}$.

Given the measure μ_X is it possible to define the Stieltjes transform of it $s_X(z) = \tau\left((X - z)^{-1}\right)$, and theorem 3.1.6 holds.

4.1 Independence and Freeness

The next step is to deal with the independence of our random variables. In chapter 2 we defined two random variables X, Y independent if $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$, where now the expectation is defined through the functional $\mathbb{E}X = \phi(X)$. For example, due to the commutativity of the independent variables, it is easy to see that $\phi(XYX) = \phi(X^2)\phi(Y)$. Now that our variables are not supposed to be moved without issues the definition of the quantity $\phi(XYX)$ must somewhat be settled. Nevertheless it is possible to extend the classical definition of independence for sub-algebras $\mathcal{A}_\backslash$ which commute one with the other.

Definition 4.2. ***Independence for Algebras***
Given a non-commutative probability space $(\mathcal{A}, \tau)$, any family of sub-algebras $\{\mathcal{A}_\alpha \subseteq \mathcal{A}\}_{\alpha \in \mathcal{I}}$ will be called ***independent*** *when*

- $[\mathcal{A}_{\alpha_i}, \mathcal{A}_{\alpha_j}] = 0$ *when* $\alpha_i \neq \alpha_j$,
- $\tau(X_1 X_2 \cdots X_n) = \tau(X_1)\tau(X_2)\cdots\tau(X_n)$ *for* $X_k \in \mathcal{A}_{\alpha_k}$, $\lambda_i \neq \lambda_j$ *when* $i \neq j$

where with $[\cdot, \cdot]$ we indicated the commutator of two sub-algebras.

Of course this definition extends immediately to single random variables X_j considering the algebras generated by them. But how to deal with non commuting sub-algebras? The right way to tackle this situation comes from group theory, in particular the theory of *free groups*. Indeed, given a group $\mathcal{G}$ generated by two of its subgroups $\mathcal{G}_1$ and $\mathcal{G}_2$, we can ask ourselves how can be a generic element $g \in \mathcal{G}$ can be represented with respect to $\mathcal{G}_1$ and $\mathcal{G}_2$. $\mathcal{G}$ is then said to be the *free product* of $\mathcal{G}_1$ and $\mathcal{G}_2$, $\mathcal{G} = \mathcal{G}_1 * \mathcal{G}_2$, whenever a element g can be represented as a *formal word* of consecutive alternating elements $h_i \in \mathcal{G}_1$ and $w_j \in \mathcal{G}_2$, i.e.

$$g = h_1 w_1 h_2 w_2 \cdots h_k w_k$$

for a certain index k, and where the product rule of the group is intended. In this way we can say that the two subgroups are independent in the sense that no relation within their elements could be settled. This could be easily extended to sub-algebras, defining a generic element $a \in \mathcal{A}$ as the product of subsequent different elements a_i of the different sub algebras. It is easy to gather that at this point the value of $\tau(a)$ depends in a certain sense on the different values the linear functional takes on the different elements a_i of the sub-algebras. We have thus the following definition:

Definition 4.3. ***Freeness of algebras***
Given a non-commutative probability space $(\mathcal{A}, \tau)$ and a family of unital sub-algebras $\mathcal{A}_1, \cdots, \mathcal{A}_n \subseteq \mathcal{A}$, we say that a family is ***free*** *if $\tau(a_1, a_2 \cdots a_n) = 0$ when:*

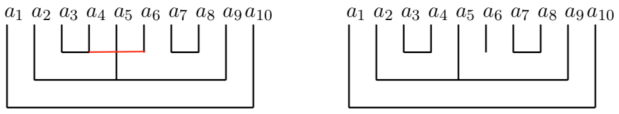
- *$a_i \in \mathcal{A}_{i(j)}$ where subsequent indices are different, $i(j) \neq i(j+1)$,*
- *$\tau(a_i) = 0$, $\forall i$.*

The family of sub-algebras is required to be unital because we remember we should have $\tau(e) = 1$, where e is the identity of the algebra. In the same way we can talk about the *freeness* of random variables looking at the algebras generated by them. We see that remarkably that free random variables have some properties very similar to usual independence of classical random variables. For example it follows that for free r.v. X_1, X_2 we have that $\tau(X_1, X_2) = \tau(X_1)\tau(X_2)$ or that $\tau(X_1, X, X_2) = \tau(X)\tau(X_1, X_2)$ where $X \in \mathcal{A}_i$ and $X_1, X, 2 \in \mathcal{A}_j$, with $i \neq j$. Despite of being very similar from certain algebrical aspects, freeness and independence are quite different. Let's take for example two free r.v. X, Y. If $\tau(X) = \tau(Y) = 0$ we will

have from the definition of freeness that $\tau(XYXY) = 0$; this is not the case of independent X and Y, where we would have that $\tau(XYXY) = \tau(X^2)\tau(Y^2)$ which actually is (except exceptional cases) never zero. We can extend freeness to non centered r.v. by just looking at the quantity $X - \tau(X)$[3]We understate the identity element.) and also to polynomials of r.v. $P(X)$, getting the most general form

$$\tau\left(\prod_{k=1}^{n}(P_k(X_{i_k}) - \tau(P_k(X_{i_k})))\right) = 0 \tag{4.4}$$

whenever the variables X_{i_k} are free and no two adjacent indices are equal. The first thing we notice, is that while the concept of independence is of more practical use, the definition of freeness for non commuting random variables appears rather cumbersome. So we ask if there is a more practical way of discerning if two random variables are actually free. The answer to the last question comes from combinatorics. The relationships between combinatorics and free probability have been developed as an alternative way of tackling problems in the theory, which is actually born as branch of functional analysis. The pioneer in this area was R. Speicher with its texts [49, 53]. Given a set S, a partition π of it will be a collection of blocks $\{V_1, \cdots V_s\}$ that are disjoint subset of S and together cover the whole set. When the set is ordered we can define a partition to be *crossing* ($CP(S)$) if given the elements $x_1 \leq y_1 \leq x_2 \leq y_2$ we have that x_1 and x_2 belong to the same block while y_1 and y_2 belong to different blocks. Analogously we can define *non-crossing* partitions ($NCP(S)$), and we are particularly interested to it.



(a) $CP(10)$. **(b)** $NCP(10)$.

Figure 4.1: Two partition of the set $\{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}\}$, one *crossing* and one *non-crossing*.

In particular partitions where the blocks are made of two elements are called *pairings* ($NCP_2(S)/CP_2(S)$). It is a standard result in combinatorics

[3](

that the number of elements in $NCP_2(n)$, where with the argument n we simply intend the set $\{1, \cdots n\}$, is exactly 0 for odd n or the $\frac{n^{th}}{2}$ Catalan number $C_{\frac{n}{2}}$ for n even. Here we see the first connection with random matrix theory, and semicircular law. We can now define the *free cumulants*.

Definition 4.4. ***Free cumulants***
Given a non-commutative probability space $(\mathcal{A}, \tau)$, we define free cumulants $k_n : \mathcal{A}^n \to \mathbb{C}$ to be the multilinear functionals defined by the recursive ***moment-cumulant formula****:*

$$k_1(X) = \tau(X) \qquad \tau(X_1 \dots X_n) = \sum_{\pi \in NCP(n)} k_\pi(X_1, \dots, X_n) \tag{4.5}$$

where given $\pi = \{V_1, V_2, \dots V_s\}$, we define $k_\pi = \prod_i^s k_{V_i}(X_1, \dots, X_n)$, and $k_{V_i}(X_1, \dots, X_n) = \{X_{v(1)}, \dots X_{v(r)}\}$ where $V_i = \{v(1), \dots, v(r)\}$.

For giving an example of how k_π decomposes, let's take $S = \{1, 2, 3, 4, 5\}$, and $\pi(S) = \{(1, 3, 5), (2), (4)\}$ we have $k_\pi(X_1 \dots X_5) = k_3(X_1, X_3, X_5)\tau(X_2)\tau(X_4)$. It is worth to notice that free cumulants are well defined. Indeed at the n order only one free cumulant acting on n elements appears. Then it could be defined recursively inverting the *moment-cumulant formula*, and calculating in order the formulas for $n-1$ order up to arriving to $k_1 = \tau(X)$. Once we settled free cumulants we can now state the result which connects them with freeness:

Theorem 4.1.1. ***Free cumulant theorem***
Let $(\mathcal{A}, \tau)$ be a non-commutative probability space, and let $X_1, \dots, X_n \in \mathcal{A}$ be random variables. Then the family $\{X_1, ..., X_n\}$ is free if and only if the free cumulants $k_s(X_{i(1)}, \dots, X_{i(s)})$ is equal to zero whenever $2 \leq s \leq n$, and there exist k and j such that $i(k) \neq i(j)$.

Thus, working with cumulants gives us a better way to see if a set of random variables is free or not.

4.2 Free CLT & Free convolution

In the last section we have seen the first connection between random matrix theory and free probability. Indeed, the moments of Wigner semicircle law are given by Catalan numbers, which also give us the number of elements

in $NCP_2(n)$. One thing we have already noticed is the closeness of the quantities (3.23) and (3.13). This is not a case, but a deep insight of what is going on.

Theorem 4.2.1. ***Free Central Limit Theorem***
Given a non-commutative probability space $(\mathcal{A}, \tau)$ a family of X_i selfadjoint free identical distributed r.v., with mean zero, $\tau(X_i) = 0$, and variance one, $\tau(X_i^2) = 1$ the quantity

$$S_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

converges in distribution to a random variable whose probability distribution μ_X is Wigner semicircle law (3.25).

From the above theorem it is now clear that the *universality* role that normal distribution plays in classical probability theory is played by Wigner's law in free probability. The difference between the two situations, as we already noticed, is that the non commutativity poses constraint in the computation of moments which became in the free case *non-crossing* partition, while are *crossing* in the classical one. One of the tools used in classical probability theory for dealing with sums of random variables is convolution (2.3). We would like to generalize this notion. First by multilinearity of free cumulants it is easy to se that if we have two free random variables X and Y and we want to compute their sum $Z = X + Y$, we have that

$$k_n^Z = k_n^{X+Y} = k_n^X + k_n^Y$$

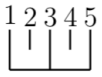
Thus inverting formula (4.5) we can retrieve the moments of μ_Z and its analytical form. We will indicate the probability arising from this procedure by $\mu_Z = \mu_X \boxplus \mu_Y$. Free convolution has strong connection with Stieltjes transform. Defining the $\mathcal{R}-$transform of a distribution μ as the following generating function

$$\mathcal{R}^\mu(z) = \sum_{n=0}^{\infty} k_{n+1} z^n \tag{4.6}$$

we see immediately that $R^{X+Y}(z) = \mathcal{R}^X(z) + \mathcal{R}^Y(z)$. Thus we guess that there should be link with the moment generating function $M(z)$ (3.35) and Stieltjes transform $s_\mu(z)$. Indeed we have

$$\mathcal{R}^\mu(s_\mu(z)) + \frac{1}{s_\mu(z)} = z \qquad s_\mu(\mathcal{R}(z) + \frac{1}{z}) = z \tag{4.7}$$

The above equation, with formula (3.34) can permit us to retrieve the unknown distribution of the variable Z.



Chapter 5

Random Graphs and Complex Networks

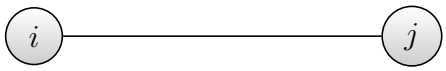
Complex networks are a powerful tool in mathematical physics. They can model the interaction of discrete system under certain assumption, and have far application in physics, biology, economics and also sociology models. We will explore the mathematical setting of Complex Networks (CN), and then try to apply the mathematical machinery developed until now to give a rigorous mathematical description of some CN's properties, in particular random walks on complex networks.

5.1 Random graphs

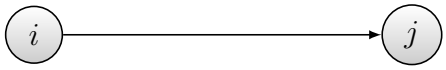
A graph $G = (V, E)$ is a discrete collection of points V (also called vertices or nodes) and edges (E). The edges are link between two points, and are in general specified by indicating two such points (i, j) (we have supposed that the nodes are ordered in such way to indicate the as $\{1, \cdots, n\}$); in the general setting a link can connect a point with itself. The cardinality of the set of vertices $|V|$ is called **order** of a graph. The cardinality of the set of edges $|E|$ is called **size**. The number of edges linked to a site i is named **degree** of the site, or in physical context **connectivity** d. A **path**, is a sequence of edges. A path in which a starting node is also the last one is callet **cyclic path** or **closed path** or simply **cycle**. In the case of a graph in which there are no more than one edge connecting two nodes a path is identified by the sequence of nodes touched $\{i \to j \to m \to k\}$

We have the following properties:

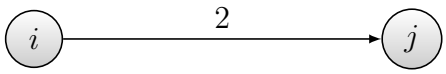
- **Undirected graph**: a graph is undirected if the edge between (i, j) is symmetric, i.e. $(i, j) = (j, i)$.



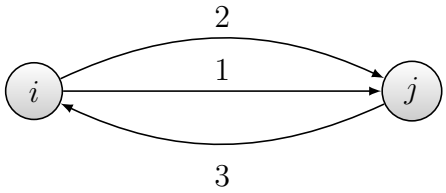
- **Directed graph**: a graph is directed if the edge between (i, j) is an ordered pair, i.e. $(i, j) \neq (j, i)$.



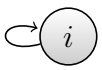
- **Weighted graph**: a graph is said to be weighted if each edge is associated to a number, called weight. The resulting graph is indicated with $G(V, E, W)$ where W_{ij} is the matrix carrying the weights.



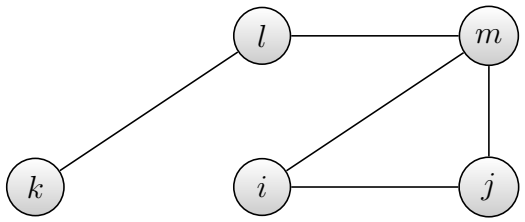
- **Multiple edges**: a graph is said to have multiple edges if two nodes i, j are linked by more than one edge. Usually different nodes have different weight.



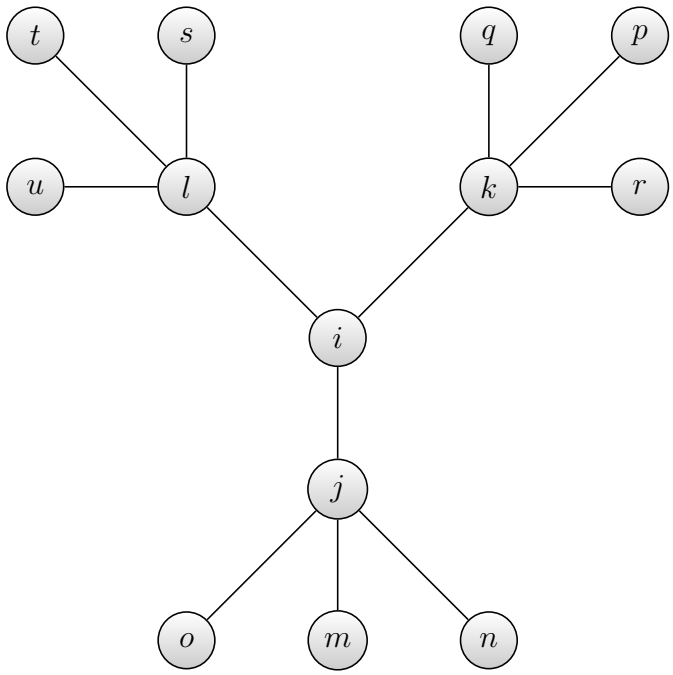
- **Site loops**: a graph is said to have a loop at the site i if there is an edge (i, i). A remark: in physical contexts, or more generally in context outside pure graph theory, a closed path, i.e. a path where the starting node is also the last one of the path, goes by the word loop. We will work mainly with simple graphs (see below) so no ambiguities will raise.



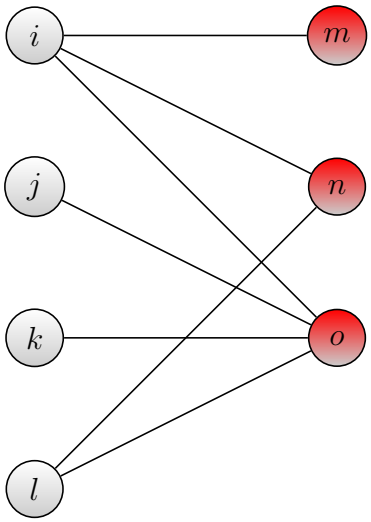
- **Simple graph**: a graph which has not multiple edges nor loops. In a simple graph with n nodes we have at most $n-1$ edges.



- **Regular graph**: a graph in which each node has the same degree.
- **Complete graph**: a graph in which each node has full degree (i.e. $n-1$).
- **Connected graph**: an undirected graph is called connected if, for every couple of point there is a path that join them. A disconnected graph may have many connected components, i.e. maximal connected subgraphs of G.
- **Tree**: a connected and acyclic graph. Usually they appear in context with where we have locally a regular graph of constant degree k, and thus they are usually represented as regular.



- **Bipartite**: a graph is said to be bipartite if the nodes can be divided in two sets such that no edges link two nodes of the same set.



Connected with a graph we can define many matrix quantities.

- **Adjacency matrix**: after ordering the nodes $1, \cdots, n$, the adjacency matrix A for simple graph, is a matrix $|V| \times |V|$ with entries 1 in position

A_{ij} if nodes (i, j) are connected by an edge, and 0 otherwise. Being simple, we note that the diagonalFor non simple graph, the adjacency matrix A_{ij} contains at in the position (i, j) the number of directed edges from $i \to j$. For loops we can take either a 1 or a 2, depending on the convention.

- **Degree matrix**: a diagonal matrix D containing in the position (i, i) the number of edges connected to the node i.
- **Laplacian matrix**: defined as $L = D - A$. Sometimes is useful to define a symmetric normalized laplacian $L^{sym} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ and will have an important role in the study of random walks.

A **random graph** is a graph generated by a random process. For example given a graph $G(V, E)$ of a fixed order and size, we could pick up randomly couple of nodes until we saturate the number of edges. In view of our applications to complex networks we will introduce two classes of random graphs:

- **Erdős-Rényi graph**: there are two equivalent ways of realizing an Erdős-Rényi graph.

  Case 1: $G(n, m)$
  In this case, the random graph is picked up from with uniform probability between the set of graphs with n nodes and m edges. Thus the fixed parameter is the number of edges, leaving the randomness of the generation process in the stochastic choice between the various instance of $G(n, m)$ uniformly distributed.

  Case 2: $G(n, p)$
  This kind of random graph is generated by a Bernoulli process. $G(n, p)$ is a set of n nodes, and the links beetween the nodes are created with a probability p. Usually we let be $p = p(n)$. Mean connectivity is given by $\langle d \rangle = np$. The probability of getting a graph with m edges is given by

  $$p^m (1 - p)^{\binom{n}{2} - m}$$

  The two types of Erdős-Rényi graph will be equivalent in the following way. For the $G(n, p)$ the expected value of edges is $\binom{n}{2} p$. We can than

evaluate that there will be correspondency for $m = \binom{n}{2}p(1-p)$ with standard deviation given by $\sigma = \sqrt{\binom{n}{2}p(1-p)}$.Thus the two models $G(n,p)$ and $G(n, m = \binom{n}{2}p)$ are asymptotical coincident for the law of large numbers, provided that $p(1-p)\binom{n}{2} \to \infty$ for $n \to \infty$[1].

- $k-$**regular random graph** $G_{n,d}$ where each node has degree d. It is important to remark that here the nodes comes by default with d free edges attached that must be linked through a random process. We will see that this fixed value introduces significant differences with $G(n,p)$ graphs' spectra.

The presence of the constraint of fixed degree, introduce correlation inside the graph and its adjacency matrix in a such way that Lindeberg condition (3.27) is no more satisfied. We remember that we are interested in large n behavior, and we can take $d(n)$ and $p(n)$. The study of spectra of random graph, i.e. of the ESD of the adjacency matrix of the graph, gives interesting results.

5.1.1 $G(n,p)$ spectra

The spectra of Erdős-Rényi graph depends heavily on the asymptotic behavior of $p(n)$. It results that

- $p = \omega(\frac{1}{n})$[2] in this case the sparsity of the matrix does not undermine the wigner semicircular law. We have the following[3]

  Theorem 5.1.1. ***Wigner distribution for Erdős-Rényi random graphs***
  Let A_n be the adjacency matrix of a Erdős-Rényi random graph $G(n,p)$, with $p(n) = \omega(\frac{1}{n})$. Then, the ESD of $\frac{1}{\sigma\sqrt{n}}A_n$, with $\sigma^2 = p(1-p)$ converges in distribution to Wigner's semicircular law

$$\rho_{sc}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2} & \text{if } |x| \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

[1] See [12, 36], *theorem* 2.2 and *proposition* 1.12rispectively.
[2] We use the asymptotical notation $f(x) = \omega(g(x))$ if $\frac{f(x)}{g(x)} = \infty$ for $x \to \infty$.
[3] See [15, 66]

- $p = O(\frac{1}{n})$: in this cases we have a probability of the form $p = \frac{\alpha}{n}$, and we will have a sparse graph. It is known that for $p = \frac{\log n}{n}$ the graph has a sharp threshold over which it is connected almost surely, i.e. $\forall \epsilon > 0$ we have that the $G(n, p)$ with $p = \frac{\log n}{n}(1 - \epsilon)$ is disconnected a.s. while for $p = \frac{\log n}{n}(1 + \epsilon)$ the graph will be connected a.s. From [26] we have the following results:

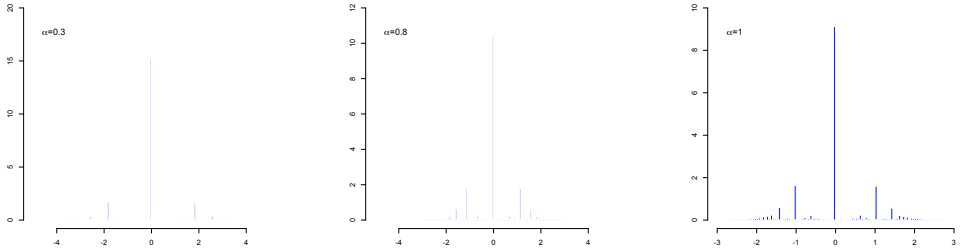
  Case 1: $\alpha < 1$
  In this case, G will be disconnected, the size of its component being $O(\log n)$. The extreme sparsity is encoded in the presence of trees mostly that will contribute with discrete spectra.

  Case 2: $\alpha = 1$
  G we will have isolated components of size $O(n^{\frac{2}{3}})$. This is the percolation transition threshold.

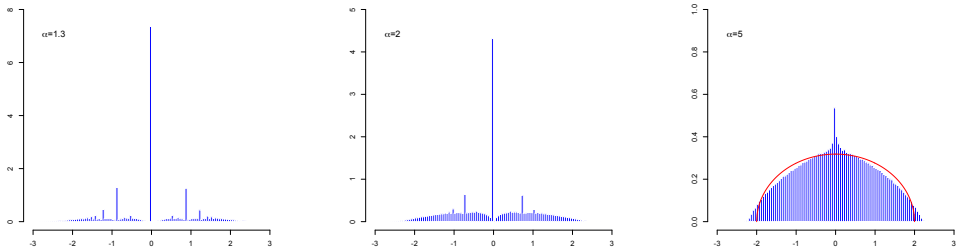
  Case 3: $\alpha > 1$
  In this case, G will have a giant component of size $O(n)$ which in the limit $\alpha \to \infty$ covers all the graph. The other components have size $O(\log n)$. The giant component will contribute to the continuum part of the spectrum while the small components, being mostly tree, will give discrete contribution.

  The presence of delta peaks in the case of sparse graph is linked to the formation of treelike components without cycles. For an increasing $\alpha > 1$ these components are more and more unlikely to have an evident role in the spectra which becomes dominated by the continuous component[4].

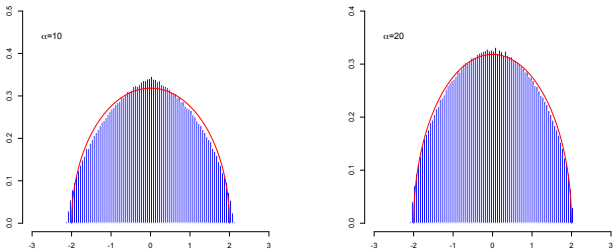
[4] See [10] chapter 5.2 for a dissertation. Other results may be found in [60]



(a) $\alpha = 0.3$. **(b)** $\alpha = 0.8$. **(c)** $\alpha = 1$.

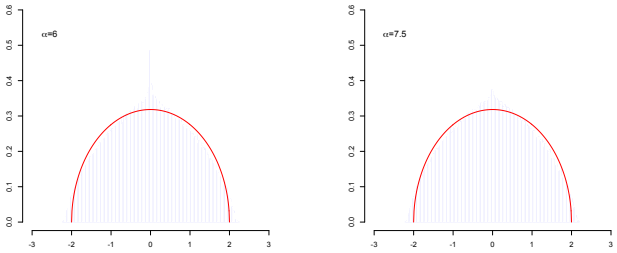


(d) $\alpha = 1.3$. **(e)** $\alpha = 2$. **(f)** $\alpha = 5$.



(g) $\alpha = 10$. **(h)** $\alpha = 20$.

Figure 5.1: Spectral distribution of the normalized and centered adjacency matrix $M_n = \frac{1}{\sigma\sqrt{n}}(A_n - pJ_n)$ (J_n is the matrix of all 1s), with $\sigma^2 = p(1-p)$ of $G(n,p)$, for various values of α. We took $n = 1000$ and a population of 100 random graphs. In red the Wigner semicircular law (3.25)



(a) $\alpha = 6$. **(b)** $\alpha = 7.5$.

Figure 5.2: Spectral distribution for two particular values of α with $n = 1000$ and a population of 100 samples. In fact these two values are beside the threshold $\log 1000 \sim 6.91$ above which the graph becomes connected a.s.

5.1.2 $G_{n,d}$ spectra

$d-$regular graphs in the low connectivity limit have an adjacency matrix with highly dependent entries. In fact we can think each node is given with d legs to be connected with other legs. While for the first node the choice is random, with an uniform distribution over the other nodes, the more we continue the more the choice is not free. McKay studied in [45] the spectral behavior of these graphs.

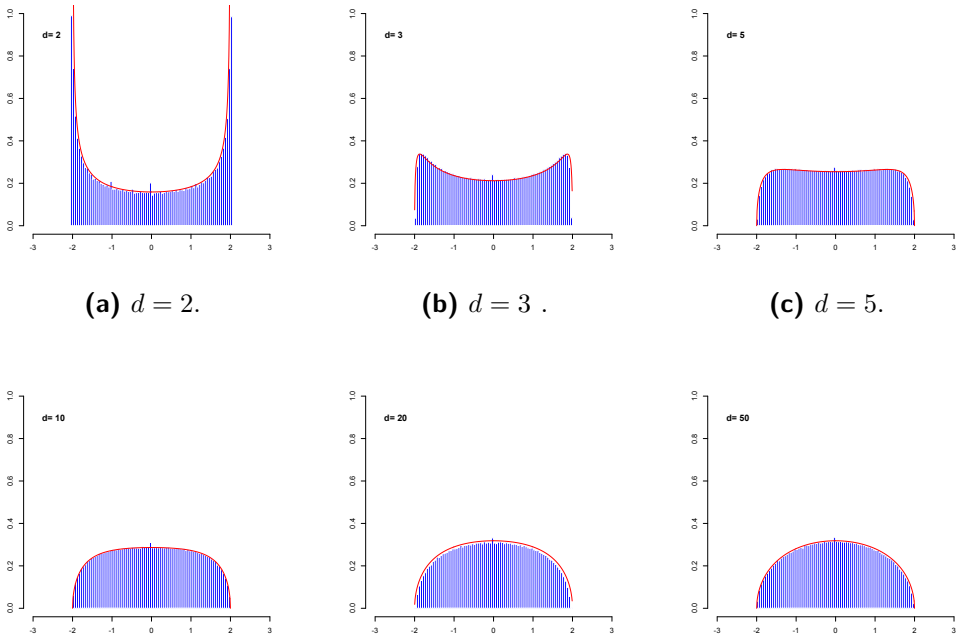
Theorem 5.1.2. ***McKay distribution***
Let be $G_{n,d}$ a random $d-$regular graph with n nodes, and let A_n be the adjacency matrix of the graph. Then, for $n \to \infty$ we have that ESD of A_n converges to

$$\rho_{mk}(x) = \begin{cases} \frac{d\sqrt{4(d-1)-x^2}}{2\pi(d^2-x^2)} & \text{if } |x| \leq 2\sqrt{d-1} \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

Continuing with our heuristic reasoning, if d approaches to ∞ with n, the constraints at late nodes becomes weaker, until, in the limit of a complete graph, they tend to disappear. This heuristic argument is not wrong: in fact in [66] the main result is the following:

Theorem 5.1.3. ***Convergence of McKay distribution to Wigner semicircular law***

Let A_n be the adjacency matrix of a $d-$regular random graph $G_{n,d}$ as before. Letting $M_n = \frac{1}{\sqrt{\frac{d}{n}(1-\frac{d}{n})}} A_n$, we have that the distribution of the M_n converges in distribution to $\rho_{sc}(x)$.



(a) $d = 2$. **(b)** $d = 3$. **(c)** $d = 5$.

(d) $d = 10$. **(e)** $\alpha = 20$. **(f)** $d = 50$.

Figure 5.3: Spectral distribution of the normalized centered adjacency matrix $M'_n = \frac{1}{\sqrt{\frac{d}{n}(1-\frac{d}{n})}}(A_n - \frac{d}{n}J_n)$ (J_n is the matrix of all 1s) of $G_{n,d}$, for various values of d. We took $n = 1000$ and a population of 100 random graphs. In red McKay distribution (5.1).

5.1.3 $G(n,p)$ vs $G_{n,d}$

The generating random process for the two graphs, even if the two sound simile at a first glance, are deeply different and bring to very different results. In fact even if we expect for a graph $G(n,p)$ with expected connectivity value $\langle d \rangle \equiv np = \overline{d}$ a behavior similar to $G_{n,\overline{d}}$ we have two very different kind of graph. The main reason for that, is that the pairing legs process for the $G_{n,d}$

brings a high correlation between the nodes in the graph, destroying the independence condition which is fundamental for Wigner-type results. For understanding why this is so different let's take the complete graph of order n, $G_{full}(n)$. The two type of graph are obtainable from this one removing edges with a certain law. Let's start with $G(n,p)$. For each node i, the process consist in the removal of the edge (i,j) with probability $(1-p)$. This process repeats itself in the same way, *for each node*, disregarding what the process has already done at the node j in precedence, because the probability to shut down an edge will be the same. With $G_{n,d}$ the situation is quite different. We must remove $(n-1)-d$ edges for the first node. This process is totally free, we will chose $(n-1)-d$ randomly with uniform probability. But it is easy to gather that going on with the nodes the process will be not be free. In fact at each new node we must take in account what has happened to the previous steps. Having performed the process on the first node, it is left with d edges. Let's take as second node one of the d nodes linked to first one. For each of these the removable edges are not $(n-1)-d$ but $(n-1)-d-1$ because the first node is now untouchable. This difference results in an overall greater connectedness of the graph, disadvantaging the formation of isolated treelike structure. We also note that the graph for great n and $\frac{d}{n} \sim 0$ results mostly without cycles, thus treelike at least a.s. locally and with higher probability globally depending on the behavior of $\frac{d}{n}$. It could be asked if this different behavior is given for differences in the number of edges, given $\langle d \rangle = np = d$[5]. For having a quantitative measure of this we can make a comparison between the number of expected edges and we take in account the fluctuations. We have already mentioned that the average number of edges for a $G(n,p)$ graph is $\binom{n}{2}p$. For calculating the number of edges of $G_{n,d}$ the following lemma is useful:

Lemma 5.1.4. ***Handshake lemma***
Let $G(V,E)$ a graph with $|V| = n$, $V = \{v_1, \cdots, v_q\}$ and $|E| = m$.

$$\sum_{i=1}^{n} deg(v_i) = 2m$$

where $deg(v_i)$ is the degree of v_i.

[5]Another interesting related question is the expected number of isolated sites, the results are large deviation inequalities which bound the fluctuation of this number of sites. See [30] for more.

Thus having in a $d-$regular graph with n nodes $deg(v_i) = d, \forall i$, we have

$$\sum_{i=1}^{n} deg(v_i) = nd = 2m \rightarrow m = \frac{nd}{2}$$

Setting in $G(n,p)$, $\langle d \rangle = np = d$ where d is the fixed connectivity of $G_{n,d}$ we have

$$\begin{cases} m^{G(n,p)} = \binom{n}{2} p = \frac{n!}{2!(n-2)!} p = \frac{n(n-1)}{2} p = \frac{d(n-1)}{2} \\ m^{G_{n,d}} = \frac{nd}{2} \end{cases}$$

At a first glance we have similar values. What is different is that $m^{G_{n,d}}$ is deterministic while $m^{G(n,p)}$ is an expected value that has a fluctuation. Indeed we have considering the formation of the edges like a sum of $\binom{n}{2}$ independent Bernoulli variables that the number of edge is given by $S_n = \sum_{i=1}^{\binom{n}{2}} X_i$, where X_i is the random Bernoulli variable that takes value 1 with probability p, and 0 with probability $(1-p)$. We have of course that

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{i=1}^{\binom{n}{2}} X_i\right) = \binom{n}{2} p$$

for the independence of the events. We now search some deviation inequality and we can apply the results of section 3.1. Let's search the probability

$$\mathbb{P}(|S_n - \mathbb{E}(S_n)| \geq \epsilon \mathbb{E}(S_n))$$

This can be analyzed by dividing the event $|S_n - \mathbb{E}(S_n)|$ in two cases

$$\begin{cases} S_n \geq (1+\epsilon)\mathbb{E}S_n \\ S_n \leq (1-\epsilon)\mathbb{E}S_n \end{cases}$$

i.e. the cases where S_n exceeds (be less than, respectively) the expected value of more (less) than a quantity $(1+\epsilon)$ (resp. $1-\epsilon$). We will deal with the case $S_n \geq (1+\epsilon)\mathbb{E}S_n$, the other result will be straightforward. The probability of this event is the same of that of the event $\{e^{x(S_n - (1+\epsilon)\mathbb{E}S_n)} \geq 1\}$ and applying Markov's inequality (2.5) with $\lambda = 1$ we have

$$\mathbb{P}(e^{x(S_n-(1+\epsilon)\mathbb{E}S_n)} \geq 1) \leq \mathbb{E}\left(e^{x(S_n-(1+\epsilon)\mathbb{E}S_n)}\right) = \mathbb{E}\left(e^{xS_n}\right) e^{-x(1+\epsilon)\mathbb{E}S_n}$$

where x is a parameter to be optimized later. It is easy to see that $\mathbb{E}(e^{xX_i}) = pe^x + (1-p)$, thus applying the definition of S_n we have that

$$\mathbb{E}\left(e^{xS_n}\right) = (pe^x + (1-p))^{\binom{n}{2}}$$

Working in the same way with the event $S_n \leq (1-\epsilon)\mathbb{E}S_n$ we have the two bound:

$$\begin{cases} \mathbb{P}(S_n \geq (1+\epsilon)\mathbb{E}S_n) \leq (pe^x + (1-p))^{\binom{n}{2}} e^{-x(1+\epsilon)p\binom{n}{2}} \\ \mathbb{P}(S_n \leq (1-\epsilon)\mathbb{E}S_n) \leq (pe^{-x} + (1-p))^{\binom{n}{2}} e^{x(1-\epsilon)p\binom{n}{2}} \end{cases} \tag{5.2}$$

It remains to optimize the bound in x.

$$\begin{cases} \frac{d}{dx}\left[(pe^x + (1-p))^{\binom{n}{2}} e^{-x(1+\epsilon)p\binom{n}{2}}\right] = 0 \\ \frac{d}{dx}\left[(pe^{-x} + (1-p))^{\binom{n}{2}} e^{x(1-\epsilon)p\binom{n}{2}}\right] = 0 \end{cases} \tag{5.3}$$

$$\Updownarrow$$

$$\begin{cases} e^{-x(1+\epsilon)p\binom{n}{2}}\binom{n}{2}\left[(pe^x + (1-p))^{\binom{n}{2}-1} pe^x - (pe^x + (1-p))^{\binom{n}{2}}(1+\epsilon)p\right] = 0 \\ e^{x(1-\epsilon)p\binom{n}{2}}\binom{n}{2}\left[- (pe^x + (1-p))^{\binom{n}{2}-1} pe^x + (pe^x + (1-p))^{\binom{n}{2}}(1-\epsilon)p\right] = 0 \end{cases}$$

We give a concrete example: for $n = 1000$ nodes and $G_{n,d}$ with $d = 2$ and $G(n,p)$ with and $\langle d \rangle = np = 2 \iff p = 0.002$. Substituting this values of p, n in equation (5.2) for a fluctuation of 50 edges in $G(n,p)$ (that is with an average $m^{G(n,p)} = \frac{2(1000-1)}{2} = 999$ the 5%) we have a probability of almost 0.3 in both directions. Thus the differences between the two graphs are mostly accountable to the different method of generation; in particular the behavior of $G_{n,d}$ in the low d limit is characterized by strongly correlation.

5.2 Random Walks

We now turn back to the discussion of Markov chains in section 2.3. With the machinery of that section we can define over the graph[6] a **random walk**. The random walk will be simply a Markov chain in which the random variable at the time $t = n+1$, X_{n+1}, will be node at which we are, given the position at time $t = n$ X_n, i.e. we will identify the state space R with the vertex set V. Markov matrix $\boldsymbol{P}$ will be thus constructed starting from the adjacency matrix in some way. Indeed the interaction encoded by $\boldsymbol{P}$ in graph jargon is represented by weighted edges of the graph normalized in a such way to have a stochastic matrix. A model that we can keep in mind as an

[6] In this section we will indicate with n the time step, while the order of the graph will be denoted by N.

example is the one of a particle moving between the edges at each discrete and normalized time step $\Delta t = 1$. In this physical view the process is called **random walk on complex network**. The network is called *complex* because its topological properties are relevant for what concerns the behavior of the walk, like the asymptotic behavior. This type of process has manifestly the Markov property: we don't need to know where the particle was at the time $t = n - 1$ for determine where will be at time $t = n + 1$, all we have to know is $\boldsymbol{P}$ and where we are at time $t = n$. Taking advantage of time homogeneity this problem will be equal formally to giving a Dirac mass distribution δ_j for localizing the particle at the site j at the starting point, and then describing its evolution applying equation (2.25) with $(\boldsymbol{\mu}_0)_i = \delta_{ij}$. Generalizing we can assume that we are given with a probability distribution over the nodes that gives us the probability of finding at time $t = 0$ the particle in one of the nodes. As we have seen, this evolution is given by (2.24) with a distribution over the nodes μ, seen as a row vector. Alternatively we can follow the evolution of the column probability vector $\overline{\boldsymbol{\mu}}_n := \boldsymbol{\mu}_n^T$ given by

$$\boldsymbol{\mu}_{n+1} = \boldsymbol{\mu}_n \boldsymbol{P} \iff \boldsymbol{\mu}_{n+1}^T = (\boldsymbol{\mu}_n \boldsymbol{P})^T = \boldsymbol{P}^T \overline{\boldsymbol{\mu}}_n$$

where ^{T} is the transpose operator. But given a graph G, how we can construct our Markov matrix? Considering a simple graph we have that the probability of staying at a node i after a time step is null, because in simple graph loops[7] are prohibited. Thus given the adjacency matrix A_ij we can suppose that from a site i we can jump to any site j connected with it with the same probability. Thus we have

$$(\boldsymbol{P})_{ij} = \frac{A_{ij}}{deg(i)} \tag{5.4}$$

where $deg(i) = \sum_j A_{ij}$ is the degree of the node i. This relation in matrix notation is $\boldsymbol{P} = D^{-1}A$, where D is the degree matrix. This kind of random walk is called **unbiased**. If we let the the graph being weighted, then the probability of the jump is given by

$$(\boldsymbol{P})_{ij} = \frac{A_{ij}W_{ij}}{deg_W(i)} \tag{5.5}$$

[7]Keep attention, here we mean loops in the graph theoretical sense, even if we have already mentioned that in physical context the word loop is often used for referring to cycles.

where $deg_W(i) = \sum_j A_{ij}W_{ij}$. In matrix notation we can write $\boldsymbol{P} = D_W^{-1}(A \circ W)$, where $\circ$ denotes the Hadamard product and D_W^{-1} is the inverse degree weighted matrix with $D_{ij} = \sum_j A_{ij}$. Often is assumed that the information of the adjacency matrix is included in the weight matrix, which will have zeros in correspondence of an absence of edges; the relation can thus be writable $\boldsymbol{P} = D_W^{-1}W$. We note that even if W is not symmetric, it can be symmetrized through a similarity transformation, thus it has real spectrum. Indeed we have for example for an unbiased random walk

$$D^{-1}W = D^{-\frac{1}{2}}\left(D^{-\frac{1}{2}}WD^{-\frac{1}{2}}\right)D^{\frac{1}{2}} = D^{-\frac{1}{2}}(I - L)D^{\frac{1}{2}}$$

where L is the laplacian of the graph defined by $L = D - A$ and I (here and in the sequel) is the identity matrix. In simple and not bipartite graph, the random walk so created is an irreducible and aperiodic Markov chain, and by construction reversible. We note that by construction the elements of $\boldsymbol{P}$ are **non-negative**, and so is the matrix. For further studying the spectral distribution of Markov matrices we need the following results, preceded by a definition.

Definition 5.1. ***Irreducible Matrix***
We say that a matrix M is irreducible if there doesn't exist a permutation matrix P such that M can be trasformed in block upper triangular form by P:

$$P^{-1}MP \neq \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} \tag{5.6}$$

where A, B and C are non trivial matrices of the right dimension.

The name of this condition is named to the homonymous property of Markov chains and graphs, see [47]. In fact an irreducible Markov chain has an irreducible transition matrix. We have the following theorem for identifying irreducible matrices.

Theorem 5.2.1. ***Criterion for Irreducible Matrices***[8]
Given a non-negative square matrix M $n \times n$, it is irreducible if and only if

$$(I + M)^{n-1}$$

is a strictly positive matrix.

[8] See [47], page 672.

We can now state the following powerful result

Theorem 5.2.2. ***Perron-Frobenius theorem***
Given a non-negative irreducible matrix M the following results holds.

- *The spectral radius $\rho(M)$ of M is positive, $\rho(M) > 0$, is in the spectrum of M, $\rho(M) \in \sigma(M)$ and is called* ***Perron root****.*
- *$\rho(M)$ is a simple eigenvalue.*
- *Exists and is unique a vector $\boldsymbol{v}$ called* ***Perron vector*** *identified by the following properties:*

$$\begin{aligned} M\boldsymbol{v} &= \rho(M)\boldsymbol{v} \\ \|\boldsymbol{v}\| &= 1 \\ \boldsymbol{v} &> 0 \end{aligned}$$

Morover if $\rho(M)$ is the only eigenvalue on the spectral cirle of M, M is said ***primitive****.*

The primitivity condition will have a central role in limit theory of transition matrices.

Proposition 5.2.3. ***Primitivity criterions.***
A non-negative irreducible matrix M is primitive if and only if one of the following occurs:

- $\lim_{k\to\infty}(\frac{M}{\rho(M)})^k$ *exists.*
- $M^k > 0$ *for some* $k > 0$
- *The Markov chain associated to M is aperiodic.*

We now look to limit theory of transition matrices. As we have seen this is linked to study equation (2.26) this problem resolves in the study of the asymptotic behavior of $\boldsymbol{P}^k$ or $(\boldsymbol{P}^T)^k$. We suppose that the transition $\boldsymbol{P}$ matrix describes an irreducible Markov process, and we also require primitive. All these properties extend obviously to the transpose $(\boldsymbol{P}^T)^k$. The assumption we make on $\boldsymbol{P}$ are exactly the requirements of theorem **??**: we are restating those result analyzing only transition matrix. Given such $\boldsymbol{P}$ we can apply Perron's theorem. $\boldsymbol{P}$ is a stochastic matrix thus we can see that

$\boldsymbol{e}^T = \{1, \cdots, 1\}$ is a left eigenvalue with eigenvalue 1, i.e. our Perron vector, or, with the jargon of section 2.3, our stationary distribution:

$$\boldsymbol{e}^T = \boldsymbol{e}^T \boldsymbol{P}$$

We now work for simplicity with $\boldsymbol{P}^T := \overline{\boldsymbol{P}}$. We call the Perron vector of $\overline{\boldsymbol{P}}$, $\boldsymbol{v}_0$. This can be normalized to give a probability distribution over the states. Given an intial probability $\boldsymbol{p}(0)$ we can expand it in terms of the eigenvectors of $\overline{\boldsymbol{P}}$, $\{\boldsymbol{v}_i\}_{0 \leq i \leq n}$. We have that

$$\boldsymbol{p}(0) = \sum_i \alpha_i(0) \boldsymbol{v}_i \tag{5.7}$$

where $\alpha_i(0) = \langle \boldsymbol{p}(0), \boldsymbol{v}_i \rangle$ is the projection of $\boldsymbol{p}(0)$ onto the i^{th} eigenvector $\boldsymbol{v}_i$. Applying $\overline{\boldsymbol{P}}^k$, (thus the transpose of equation (2.26)) we have

$$\boldsymbol{p}(k) = \overline{\boldsymbol{P}}^k \left(\boldsymbol{p}(0)\right) = \overline{\boldsymbol{P}}^k \left(\sum_i \alpha_i(0) \boldsymbol{v}_i \right) = \alpha_0(0) \boldsymbol{v}_0 + \sum_{i>0} \lambda_i^k \alpha_i(0) \boldsymbol{v}_i \tag{5.8}$$

where the λ_i are the eigenvalues of $\boldsymbol{v}_i$, and from Perron's theorem we know that $|\lambda_i| < 1$, where 1 is $\rho(\overline{\boldsymbol{P}})$ is the Perron eigenvalue. For the property $|\lambda_i| < 1$ we recover the result of theorem **??**

$$\lim_{k \to \infty} \overline{\boldsymbol{P}}^k = \boldsymbol{v}_0 \cdot \boldsymbol{e}^T \tag{5.9}$$

Thus we can make estimate that the reaching of the asymptotical state $\boldsymbol{v}^{eq} = \alpha_0^k(0) \boldsymbol{v}_0$ in time steps k is bounded by

$$\|\boldsymbol{p}(k) - \boldsymbol{v}^{eq}\|^2 = \left\| \sum_{i>0} \lambda_i \alpha_i(0) \boldsymbol{v}_i \right\|^2 \leq \sum_{i>0} \|\lambda_i \alpha_i(0) \boldsymbol{v}_i\|^2 \leq \lambda_*^{2k} \sum_{i>0} \alpha_i^2(0) \leq \lambda_*^{2k}. \tag{5.10}$$

where $\lambda_* = \max\{\lambda_i : \lambda_i \neq 1\} < 1$ is the second largest eigenvalue after Perron eigenvalue. Taking the Poissonian limit for the times steps in equation (5.8) we have

$$\sum_i \lambda_i^k \alpha_i(0) \boldsymbol{v}_i \to \sum_i \alpha_i(0) \boldsymbol{v}_i \sum_m \lambda_i^m \frac{e^{-k} k^m}{m!} = \sum_i \alpha_i(0) \boldsymbol{v}_i e^{-(1-\lambda_i)k}$$

This expression is again dominated by $\gamma_* = (1 - \lambda_*)$, which rules the asymptotic behavior, and the quantity $\tau_* = \frac{1}{(1-\lambda_*)}$ is the relaxation time.

For non primitive transition matrix the limit

$$\lim_{k\to\infty}(\frac{\overline{\boldsymbol{P}}}{\rho(\overline{\boldsymbol{P}})})^k$$

does not exist. In the language of Markov chains we have a periodic process. One way to deal with situation is to study the **Cesàro limit**

$$\lim_{k\to\infty}\left[\frac{\boldsymbol{p}(0)+\boldsymbol{p}(1)+\cdots+\boldsymbol{p}(k-1)}{k}\right] \tag{5.11}$$

It results that this limit converges to the stationary distribution, like for equation (5.9):

$$\lim_{k\to\infty}\left[\frac{I+\boldsymbol{P}+\boldsymbol{P}^2+\cdots+\boldsymbol{P}^{k-1}}{k}\right] \tag{5.12}$$

The heuristic motivation for this is that the j^{th} component of stationary distribution represent the fraction of time spent on the node j. This is given by the mean over the total time-steps of the j^{th} component of probability distributions $\{(k)\}_{k\geq 0}$, i.e. exactly the j^{th} component of the Cesàro limit (5.11)[9].

5.2.1 Laplacian matrix

We recall that the normalized laplacian matrix L^{sym} of the graph is defined by

$$L^{symm} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$$

where $L = D - A$ is the standard laplacian. The spectral distribution of L and L^{sym} Have important meanings. It results that $\sigma(L) > 0$. The dimension of the eigenspace V_0 of L related to the null eigenvalue is the number of the connected components of a graph. Thus, the second smaller eigenvalue gives us information on the connectedness of the graph and it is called **Fiedler eigenvalue** or **algebraic connectivity**. The same eigenvalue for the normalized laplacian L^{sym} is exactly $\gamma_* = (1 - \lambda_*)$ appearing in the limit theory of transition matrix. Thus the study of Laplacian spectrum gives us important topological and dynamical information on the graph and the random walk defined on it. The spectral gap between the algebraic connectivity and

[9]Formal threatment can be found in [47], chapter 8

the null eigenvalue is under study, and could give important information on the robustness[10]. Further material can be found in [16, 28, 50, 58].

5.2.2 Spectra of transition matrices and comparisons with simulations

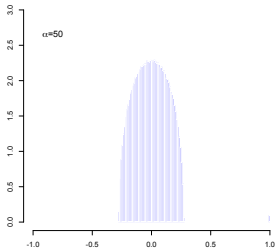
The calculation of the spectra of transition matrices is straightforward. The definitions (5.4) and (5.5) of $\boldsymbol{P}$ in the unbiased and weighted cases tells us that thanks to Perron-Frobenius theorem the greatest eigenvalue of $\boldsymbol{P}$ is 1. Working with irreducible chains, isolated sites, i.e. null rows in the matrix corresponding to zero values in D, are forbidden. Keeping in mind that $\boldsymbol{P}$ is not centered and considering the results on centered matrices of theorem 3.2.3, tells us that the greatest of normalized eigenvalues of a $G(n,p)$ concentrates at the value 2, where the normalization if provided by the factor $\frac{1}{\sqrt{np(1-p)}}$. Thus eigenvalues of W (and also the others) are shrinked by a factor $\frac{\sqrt{np(1-p)}}{np}$. In particular the second greatest eigenvalue will fluctuate, with variance given by theorem 3.2.3 opportunely scaled, around $2\sqrt{\frac{1-p}{np}}$. For what concerns $G_{n,d}$ the discussion is the same, but this time the scaling factor is given by $\sqrt{\frac{1-\frac{d}{n}}{d}}$, which is essentially the same factor adapted to $G_{n,d}$.

The behavior in the pictures ?? can be explained in the following way. We have said that $\boldsymbol{P}$ can be symmetrized by a similarity transform $D^{-\frac{1}{2}}$. We can always find a base on which the matrix acts constructed in the following way: picking the Perron eigenvector we choose the other $n-1$ eigenvector in a such way that they form a subspace orthogonal to that. This results in a complessive rotation S. In a such way W decompose in

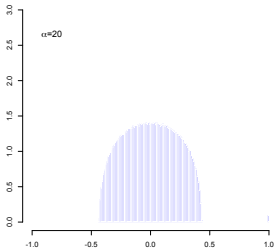
$$S^{-1}D^{\frac{1}{2}}\boldsymbol{P}D^{-\frac{1}{2}}S = \begin{bmatrix} 1 & 0 \\ 0 & \Lambda \end{bmatrix} \tag{5.13}$$

where the symmetric matrix Λ entails all the randomness (and dependence of the entries as can be seen in picture f of figure ??) of the initial matrix. The constraint of being stochastic has thus the effect of fixing the eigenvector corresponding to the maximum eigenvalue 1, and leaving the random component in the $n-1$ dimensional subspace orthogonal to it.

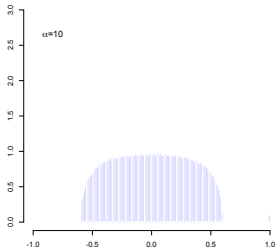
[10] Robustness, has many definition in network theory depending on the event we want to study. In general is the property of a network of maintaining its characteristics despite small perturbation, like the removal of nodes or edges.



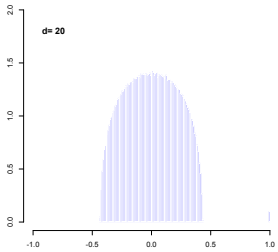
(a) $G(n,p)$ *with* $\alpha = 50$*; the second largest eigenvalue is* $\lambda_2 = 0.2758$ *while the expected numerical value is* $\overline{\lambda} = 0.2756$.



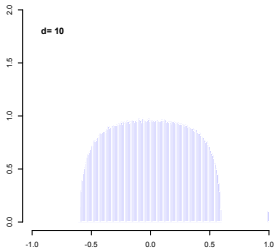
(b) $G(n,p)$ *with* $\alpha = 20$*; the second largest eigenvalue is* $\lambda_2 = 0.4375$ *while the expected numerical value is* $\overline{\lambda} = 0.4427$.



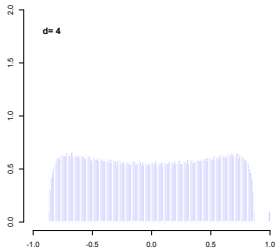
(c) $G(n,p)$ *with* $\alpha = 10$*; the second largest eigenvalue is* $\lambda_2 = 0.5990$ *while the expected numerical value is* $\overline{\lambda} = 0.6292$.



(d) $G_{n,d}$ *with* $d = 20$*; the second largest eigenvalue is* $\lambda_2 = 0.4344$ *while the expected numerical value is* $\overline{\lambda} = 0.4427$.



(e) $G_{n,d}$ *with* $d = 10$*; the second largest eigenvalue is* $\lambda_2 = 0.6112$ *while the expected numerical value is* $\overline{\lambda} = 0.6292$.



(f) $G_{n,d}$ *with* $d = 4$*; the second largest eigenvalue is* $\lambda_2 = 0.8700$ *while the expected numerical value is* $\overline{\lambda} = 0.9979$.

Figure 5.4: Spectral distribution of the transition matrix $\boldsymbol{P}$ in the $G(n,p)$ and $G_{n,d}$ cases. We took $n = 1000$ and a population of 100 random graphs.

Conclusions

The mathematical tools developed in chapters 2 and 3 let us analyze $k-$regular graphs $G_{n,d}$ and Erdős-Renyi graphs $G(n,p)$ and the related random walks given by the Wigner matrix constructed on the adjacency matrix. The results founded, are well summed up by the pictures of the spectra we have calculated: while at great p and d the results are ruled by the non abelian CLT given by Wigner semicircular law, at low p and d the spectra are highly different. In $k-$regular sparse graphs, thank to their generation process, connectedness is favored, while the same is not true for large sparse Erdős-Renyi graphs. This has profound consequences on the respective random walks, which have a spectral behavior similar to the linked adjacency matrices. These results are in accordance with the trending methods used by physicist, as found in [42]. Future developments could look with more deepness to the linkage between these different approaches that give the same results, trying also to enlarge the RMT study to different kinds of graphs with different laws that are being well studied in statistical physics and less in mathematical ambients.

Bibliography

[1] Elihu Abrahams. *50 years of Anderson Localization*. world scientific, 2010.

[2] David Aldous and James Allen Fill. Reversible markov chains and random walks on graphs, 2002. Unfinished monograph, recompiled 2014, available at http://www.stat.berkeley.edu/$\sim$aldous/RWG/book.html.

[3] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[4] B Aneva. Symmetry of the riemann operator. *Physics Letters B*, 450(4):388 – 396, 1999.

[5] Søren Asmussen. *Applied probability and queues*, volume 51. Springer Science & Business Media, 2008.

[6] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. part i. wigner matrices. *Ann. Probab.*, 21(2):625–648, 04 1993.

[7] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. part ii. sample covariance matrices. *Ann. Probab.*, 21(2):649–672, 04 1993.

[8] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.

[9] Zhidong D Bai. Methodologies in spectral analysis of large dimensional random matrices, a review. In *Advances In Statistics*, pages 174–240. World Scientific, 2008.

[10] M. Bauer and O. Golinelli. Random incidence matrices: Moments of the spectral density. *Journal of Statistical Physics*, 103(1):301–337, Apr 2001.

[11] Oriol Bohigas, Marie-Joya Giannoni, and Charles Schmit. Characterization of chaotic quantum spectra and universality of level fluctuation laws. *Physical Review Letters*, 52(1):1, 1984.

[12] B. Bollobás and B. Béla. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001.

[13] Jean-Philippe Bouchaud and Marc Potters. *Theory of financial risks*, volume 4.

[14] Tómas A Brody, J Flores, J Bruce French, PA Mello, A Pandey, and Samuel SM Wong. Random-matrix physics: spectrum and strength fluctuations. *Reviews of Modern Physics*, 53(3):385, 1981.

[15] Fan Chung, Linyuan Lu, and Van Vu. The spectra of random graphs with given expected degrees. *Internet Math.*, 1(3):257–275, 2003.

[16] Fan RK Chung and Fan Chung Graham. *Spectral graph theory*. Number 92 in CBMS Regional Conference Series in Mathematics. American Mathematical Soc., 1997.

[17] Kai Lai Chung. *Lectures from Markov processes to Brownian motion*, volume 249. Springer Science & Business Media, 2013.

[18] Alain Connes. *Noncommutative Geometry Year 2000*, pages 481–559. Birkhäuser Basel, Basel, 2010.

[19] Philippe Di Francesco, Paul Ginsparg, and Jean Zinn-Justin. 2d gravity and random matrices. *Physics Reports*, 254(1-2):1–133, 1995.

[20] J.L. Doob. *Stochastic processes*. Wiley publications in statistics. Wiley, 1990.

[21] Richard Durrett. *Random graph dynamics*, volume 200 of *Cambridge series in statistical and probabilistic mathematics*. Cambridge university press, 2007.

[22] Rick Durrett. *Probability: theory and examples*. Cambridge university press, 2010.

[23] S F Edwards and P W Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965–974, may 1975.

[24] S F Edwards and R C Jones. The eigenvalue spectrum of a large symmetric random matrix. *Journal of Physics A: Mathematical and General*, 9(10):1595–1603, oct 1976.

[25] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Semicircle law on short scales and delocalization of eigenvectors for wigner random matrices. *Ann. Probab.*, 37(3):815–852, 05 2009.

[26] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[27] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.

[28] J F. Lutzeyer and A T. Walden. Comparing graph spectra of adjacency and laplacian matrices. 12 2017.

[29] J. Garnett. *Bounded Analytic Functions*. Graduate Texts in Mathematics. Springer New York, 2007.

[30] Subhankar Ghosh, Larry Goldstein, and Martin Raič. Concentration of measure for the number of isolated vertices in the erdős–rényi random graph by size bias couplings. *Statistics & Probability Letters*, 81(11):1565 – 1570, 2011.

[31] Friedrich Götze, A Naumov, and A Tikhomirov. Semicircle law for a class of random matrices with dependent entries. *arXiv preprint arXiv:1211.0389*, 2012.

[32] Thomas Guhr, Axel Muller-Groeling, and Hans A. Weidenmuller. Random matrix theories in quantum physics: Common concepts. *Phys. Rept.*, 299:189–425, 1998.

[33] Harish-Chandra. Differential operators on a semisimple lie algebra. *American Journal of Mathematics*, 79(1):87–120, 1957.

[34] Jiaoyang Huang, Benjamin Landon, and Horng-Tzer Yau. Bulk universality of sparse random matrices. *Journal of Mathematical Physics*, 56(12):123301, 2015.

[35] C. Itzykson and J. B. Zuber. The planar approximation. ii. *Journal of Mathematical Physics*, 21(3):411–421, 1980.

[36] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2011.

[37] Kurt Johansson. Non-intersecting paths, random tilings and random matrices. *Probability theory and related fields*, 123(2):225–280, 2002.

[38] Olav Kallenberg. *Foundations of modern probability*. Springer Science & Business Media, 2006.

[39] Scott Kirkpatrick and David Sherrington. Infinite-ranged models of spin-glasses. *Phys. Rev. B*, 17:4384–4403, Jun 1978.

[40] Ivan K. Kostov. Conformal field theory techniques in random matrix models. In *Workshop on Random Matrices and Integrable Systems Warwick, Coventry, England, November 2-4, 1998*, 1999.

[41] Reimer Kühn. Spectra of sparse random matrices. *Journal of Physics A: Mathematical and Theoretical*, 41(29):295002, 2008.

[42] Reimer Kühn. Random matrix spectra and relaxation in complex networks. *Acta Phys. Polon. B*, 46:1653–1682, 2015.

[43] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Phys. Rev. Lett.*, 83:1467–1470, Aug 1999.

[44] M. Ledoux. *The Concentration of Measure Phenomenon*. Mathematical surveys and monographs. American Mathematical Society, 2001.

[45] Brendan D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40:203 – 216, 1981.

[46] Madan Lal Mehta. *Random matrices*, volume 142. Elsevier, 2004.

[47] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, 2000.

[48] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[49] James A Mingo and Roland Speicher. *Free probability and random matrices*, volume 35. Springer.

[50] Bojan Mohar, Y Alavi, G Chartrand, and OR Oellermann. The laplacian spectrum of graphs. *Graph theory, combinatorics, and applications*, 2(871-898):12, 1991.

[51] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.

[52] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[53] Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*, volume 13. Cambridge University Press, 2006.

[54] Athanasios Papoulis. Random variables and stochastic processes. *McGraw Hill*, 1976.

[55] G. J. Rodgers and A. J. Bray. Density of states of a sparse random matrix. *Phys. Rev. B*, 37:3557–3562, Mar 1988.

[56] Tim Rogers, Isaac Pérez Castillo, Reimer Kühn, and Koujin Takeda. Cavity approach to the spectral density of sparse symmetric random matrices. *Phys. Rev. E*, 78:031116, Sep 2008.

[57] Timothy Rogers. *New results on the spectral density of random matrices*. PhD thesis.

[58] Opiyo Samuel, Yudi Soeharyadi, and Marcus Wono Setyabudhi. The first two largest eigenvalues of laplacian, spectral gap problem and cheeger constant of graphs. *AIP Conference Proceedings*, 1913(1):020014, 2017.

[59] S. Sharifi, M. Crane, A. Shamaie, and H. Ruskin. Random matrix theory for portfolio optimization: a stability approach. *Physica A: Statistical Mechanics and its Applications*, 335(3):629 – 643, 2004.

[60] AN Spiridonov. Spectra of sparse graphs and matrices. *Preprint*, 2005.

[61] M.D. Springer. *The algebra of random variables*. Probability and Statistics Series. Wiley, 1979.

[62] Michel Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 81(1):73–205, Dec 1995.

[63] Toshiyuki Tanaka. Moment problem in replica method. *Interdisciplinary Information Sciences*, 13(1):17–23, 2007.

[64] Terence Tao. *An introduction to measure theory*. American Mathematical Society Providence, RI, 2011.

[65] Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

[66] Linh V Tran, Van H Vu, and Ke Wang. Sparse random graphs: Eigenvalues and eigenvectors. *Random Structures & Algorithms*, 42(1):110–134, 2013.

[67] GIRKO V., KIRSCH W., and KUTZELNIGG A. A necessary and sufficient conditions for the semicircle law. 2:195, 2019-03-01T03:57:16.823+01:00 1994.

[68] J L van Hemmen and R G Palmer. The replica method and solvable spin glass model. *Journal of Physics A: Mathematical and General*, 12(4):563–580, apr 1979.

[69] J J M Verbaarschot and M R Zirnbauer. Critique of the replica trick. *Journal of Physics A: Mathematical and General*, 18(7):1093–1109, may 1985.

[70] Dan Voiculescu. Limit laws for random matrices and free products. *Inventiones mathematicae*, 104(1):201–220, Dec 1991.

[71] Herbert S Wilf. *generatingfunctionology*. AK Peters/CRC Press, 2005.

[72] Yufei Zhao. Spectral distributions of random graphs. 2012.