Degree in Mathematics

BACHELOR'S THESIS

Title: Probabilistic and analytic aspects of Boolean functions

Author: Enric Rabasseda i Raventós

Advisor: Juan José Rué Perna

Department: Mathematics

Academic year: 2021-2022

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Degree in Mathematics

Bachelor's Degree Thesis

Probabilistic and analytic aspects of Boolean functions

Enric Rabasseda i Raventós

Supervised by Juan José Rué Perna

January, 2022

First of all, I want to thank Juanjo Rué for being by my side throughout the research to prepare this thesis. I really appreciate his patience and advice to improve my work.

I would also like to thank everyone who have trusted me and supported me during this path to become a mathematician.

Enric Rabasseda i Raventós

Abstract

This thesis will focus on the study of Boolean functions. In point of fact, they can be represented with a Fourier expansion and many of the definitions and results for these functions can be rewritten in terms of the Fourier coefficients. The definition of Boolean functions is simple, this implies that they have a natural interpretation and hence have applications in many areas of scientific research. Specifically, in this thesis we will see applications in Social Choice Theory, Theoretical Computer Science and Combinatorics. For the first area, we will see and prove with Fourier analysis Arrow's theorem and KKL theorem in order to show that it is not possible to define a perfect voting election system from an ethical standpoint. Additionally, we will translate the proof presented by Arrow for his own theorem in terms of mathematical language. The work for the second application will follow the steps to prove Sensitivity Conjecture which, although it has remained unsolved for 30 years, Huang has presented a brilliant short proof in a paper published at Annals of Mathematics very recently (2019). For the last area we will present the strange phenomena of thresholds in Random Graph properties and we will show Margulis-Russo Formula to study this event in terms of Boolean functions Fourier analysis.

Keywords

Boolean functions, Social Choice Theory, Arrow's Theorem, KKL Theorem, Sensitivity Conjecture, Random Graphs, Sharp Thresholds, Margulis-Russo Formula.

MSC 2020 Codes

60B15, 91B14, 68Q15, 05C80.

Contents

1. Introduction to Boolean functions

This chapter follows the first chapter of Ryan O'Donell's reference book about the analysis of Boolean Functions [1], we will see that all the analysis on Boolean functions is based on a Fourier expansion that exists for every function. This part of the book presents the principal definitions and results from Boolean Functions that are needed to understand some applications and theorems that will be presented in the next chapters.

1.1 Basic definitions

Definition 1.1. We define a *Boolean function* f as

$$f : \{-1,1\}^n \longrightarrow \{-1,1\}.$$

where f maps each length-n binary vector called *string* into a single binary value named *bit*. More precisely, we can classify these functions in two types depending on its range. So we call *real-valued* a Boolean function with range $\mathbb{R}$ and we name *boolean-valued* a Boolean function with range $\{-1,1\}$.

Sometimes other domains will be used. For example $\{0,1\}$ as "False" and "True" or elements of the field $\mathbb{F}_2$ of size 2. All the following results are true for every mentioned domain. But in this thesis we will focus mainly on the domain of the Definition 1.1.

Definition 1.2. We refer to the domain of a Boolean function as the *Hamming cube*. In this domain we can also define a distance between two strings $\mathbf{x}$, $\mathbf{y}$.

$$\triangle(\mathbf{x},\mathbf{y}) = \#\{i : x_i \neq y_i\}.$$

These are the basic definitions that allow the reader to understand how the Boolean functions work and what metric can be used in the Hamming cube. But it also will be interesting to write the Boolean functions by a expression that will give us more information.

1.2 Fourier expansion

Given a boolean-valued Boolean function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ there is an easy method for finding a polynomial that interpolates this function by its 2^n strings of the Hamming cube. First, it is necessary to define a characteristic function explicitly to verify, once a given string of the Hamming cube is given, if an input of the Hamming cube is exactly this string.

Definition 1.3. For each point $a = (a_1,\dots,a_n) \in \{-1,1\}^n$ the *indicator polynomial for a* is:

$$\mathbf{1}_{\{a\}}(\mathbf{x}) = \left(\frac{1+a_1x_1}{2}\right)\cdots\left(\frac{1+a_nx_n}{2}\right) = \begin{cases} 1 & \text{if } \mathbf{x} = a, \\ 0 & \text{if } \mathbf{x} \in \{-1,1\}^n \setminus \{a\}. \end{cases}$$

Thus f can be interpolated using the indicator polynomial as

$$f(\mathbf{x}) = \sum_{a\in\{-1,1\}^n} f(\mathbf{x})\mathbf{1}_{\{a\}}(\mathbf{x}). \tag{1.1}$$

We observe that the definition of the indicator polynomial works because of the values in the Hamming cube: -1 and 1. The next possible step is to extend the product of the factors in the indicator polynomial. This product will finish in a linear combination of products of x_i. This also works well for real-valued Boolean functions since we only care about inputs x where $x_i = \pm 1$, so x_i^2 can be replaced by 1. Hence, what we have proved here is that every function f can be written as a multilinear polynomial.

Example 1.4. Here we study the multilinear polynomial representation of the function $max_2 : \{-1,1\}^2 \longrightarrow \{-1,1\}$ where $max_2(\mathbf{x}) = max_{i\in\{1,2\}} x_i$. First we get the image of all the strings in the Hamming cube:

- $max_2\big((-1,-1)\big) = -1$
- $max_2\big((-1,+1)\big) = +1$
- $max_2\big((+1,-1)\big) = +1$
- $max_2\big((+1,+1)\big) = +1$

And once we have the values of the function on all the inputs we can develop (1.1) to obtain the multilinear polynomial representation of max_2:

$$\begin{aligned} max_2(\mathbf{x}) = (+1)\Big(\frac{1+x_1}{2}\Big)\Big(\frac{1+x_2}{2}\Big) + (+1)\Big(\frac{1-x_1}{2}\Big)\Big(\frac{1+x_2}{2}\Big) + (+1)\Big(\frac{1+x_1}{2}\Big)\Big(\frac{1-x_2}{2}\Big)\\ + (-1)\Big(\frac{1-x_1}{2}\Big)\Big(\frac{1-x_2}{2}\Big) = \frac{1}{4} + \frac{x_1}{4} + \frac{x_2}{4} + \frac{x_1x_2}{4} + \frac{1}{4} - \frac{x_1}{4} + \frac{x_2}{4} - \frac{x_1x_2}{4} + \frac{1}{4}\\ + \frac{x_1}{4} - \frac{x_2}{4} - \frac{x_1x_2}{4} - \frac{1}{4} + \frac{x_1}{4} + \frac{x_2}{4} - \frac{x_1x_2}{4}\\ = \frac{1}{2} + \frac{1}{2}x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_1x_2. \end{aligned}$$

Before we can talk about the Fourier expansion of a Boolean function we have to clarify some notation. The multilinear polynomial representation of the Boolean function may have up to 2^n terms corresponding to the subsets $S \subseteq [n]$. So we will write the monomial corresponding to S as $\mathbf{x}^S = \prod_{i\in S} x_i$ with $\mathbf{x}^{\emptyset} = 1$ by convention.

Theorem 1.5. *Every function $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ can be uniquely expressed as a multilinear polynomial:*

$$f(\mathbf{x}) = \sum_{S\subseteq[n]} \hat{f}(S)\mathbf{x}^S.$$

This expression is called the *Fourier expansion of f* and the real numbers $\hat{f}(S)$ are named the *Fourier coefficients of f on S*. Collectively, the coefficients are called the *Fourier spectrum of f*. Later we will explain how to prove this theorem.

Observation 1.6. The monomial $\mathbf{x}^S$ can also be thought as a function on $\mathbf{x} = (x_1,\dots,x_n) \in \mathbb{R}^n$ defined like $\chi_S(\mathbf{x}) = \prod_{i\in S} x_i$. Thus we sometimes write the Fourier expansion of $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ like

$$f(\mathbf{x}) = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S(\mathbf{x}). \tag{1.2}$$

It is important to note that all these representations work for the Hamming cube $\{-1,1\}^n$ but they can not be defined like this in $\mathbb{F}_2^n$.

We define the Fourier expansion for functions $f : \mathbb{F}_2^n \longrightarrow \mathbb{R}$ by "encoding" its input bits $0,1 \in \mathbb{F}_2$ by $-1,1 \in \mathbb{R}$. The idea behind this encoding is to label the elements of the field $\mathbb{F}_2$ by the numbers of the Hamming cube with which a Fourier expansion has been developed. The encoding function $\chi : \mathbb{F}_2 \longrightarrow \mathbb{R}$ is defined as $\chi(0) = +1$ and $\chi(1) = -1$. It's mathematically natural because all elements b in $\mathbb{F}_2$ satisfy $\chi(b) = (-1)^b$. This result can generalize the definition of the encoding function for strings in $\mathbb{F}_2^n$.

Definition 1.7. For $S \subseteq [n]$ we define $\chi_S : \mathbb{F}_2^n \longrightarrow \mathbb{R}$ by:

$$\chi_S(\mathbf{x}) = \prod_{i \in S} \chi_i(x_i) = (-1)^{\sum_{i \in S} x_i}.$$

which satisfies $\chi_S(\mathbf{x} + \mathbf{y}) = \chi_S(\mathbf{x})\chi_S(\mathbf{y})$.

It is good to interpret $\mathbb{F}_2^n$ as the n-dimensional space $\mathbb{F}_2$ and identify subsets $S \subseteq [n]$ with vectors $\gamma \in \mathbb{F}_2^n$. Since we have defined this *encoding function* we can write the Fourier expansion for functions $f : \mathbb{F}_2^n \longrightarrow \mathbb{R}$ equally as (1.2).

1.3 The Hilbert space of Boolean functions

The functions defined before as $\chi_S(\mathbf{x}) = \mathbf{x}^S$ have an own name and are called *parity functions.* They are called like that because of the definition of the functions, since they multiply all the bits identified with the set S and give us the parity.

Looking at the set of all functions $f : \{-1, 1\}^n \longrightarrow \mathbb{R}$ it is possible to check that we can add two functions (pointwise) and we can multiply a function by a real scalar. So this set can be thought as a vector space V. V is 2^n-dimensional, since we can express a Boolean function by stacking the 2^n values $f(\mathbf{x})$ into a tall column vector.

In fact, the Fourier expansion (1.2) states that every real-valued Boolean function, hence in V, is a linear combination of the parity functions. Then parity functions are a spanning set for V. But also the number of parity functions is 2^n, the same as the dimension of V, so we can deduce that they are in fact a linearly independent basis for V. This result allows us to prove the uniqueness of the Fourier expansion for every function as it is remarked in Theorem 1.5.

Additionally, it is also possible to define inner products between two functions living in this space V. Although there are some possible products to choose, the one for Boolean functions is defined as it follows.

Definition 1.8. We define an *inner product* $\langle \cdot, \cdot \rangle$ on pairs of functions $f, g : \{-1, 1\}^n \longrightarrow \mathbb{R}$ by:

$$\langle f, g \rangle = 2^{-n} \sum_{x \in \{-1,1\}^n} f(\mathbf{x})g(\mathbf{x}) = \mathop{\mathbf{E}}_{\mathbf{x} \sim \{-1,1\}^n} [f(\mathbf{x})g(\mathbf{x})].$$

It is trivial to prove that the previous definition is a dot product. We also use the following notation to define **p-norms**:

$$\|f\|_p = \mathbf{E}[|f(\mathbf{x})|^p]^{\frac{1}{p}}.$$

In this definition and later is used the notation $\mathbf{E}$ to denote the expectation. This expression will be used many times along this thesis. Additionally, we write $\mathbf{x} \sim \{-1, 1\}^n$ to denote that $\mathbf{x}$ is a uniformly chosen random string from $\{-1, 1\}^n$. When we do not write $\mathbf{x} \sim \{-1, 1\}^n$ under the expectation $\mathbf{E}$ is because we take it for granted. The n coordinates x_i are independently chosen to be $+1$ and -1 both with probability $1/2$. This inner product can be interpreted as an average rather than a sum. Additionally, for boolean-valued Boolean functions, since $f(\mathbf{x})^2 = 1$, the square of the 2-norm equals 1 always.

Returning now to the basis of parity functions for V the crucial fact is that it is an orthonormal basis with the inner product.

Lemma 1.9. *The 2^n parity functions $\chi : \{-1,1\}^n \longrightarrow \{-1,1\}$ form an orthonormal basis for the vector space V of functions $f : \{-1,1\}^n \longrightarrow \mathbb{R}$, i.e.*

$$\langle \chi_S, \chi_T \rangle = \begin{cases} 1 & \text{if } S = T, \\ 0 & \text{if } S \neq T. \end{cases}$$

Proof. This lemma is proved by two steps that are supported strongly by the fact that the inputs of these functions are boolean strings with bits chosen uniformly random from ± 1, so $x_i^2 = 1$. Then, by developing the parity functions of two subsets S and T we obtain $\chi_S(\mathbf{x})\chi_T(\mathbf{x}) = \chi_{S\triangle T}(\mathbf{x})$ where $S\triangle T$ denotes the symmetric difference $S \cup T \setminus S \cap T$. Afterwards, by the independence of all the bits x_i in a string $\mathbf{x}$ from the Hamming cube, the expectation $\mathbf{E}[\mathbf{x}^S]$ can be computed as a product of all the expectations of the bits from the subset S and, since they are uniformly random chosen, $\mathbf{E}[x_i]$ is null. Therefore, the product of all of them $\mathbf{E}[\chi_S(\mathbf{x})]$ will be null unless S is the empty set because $\chi_\emptyset(\mathbf{x}) = 1$ by convention. □

Summarizing, we have seen that the Fourier expansion of $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ can be thought as the representation of f over the orthonormal basis of parity functions $\{\chi_S\}_{S\subseteq[n]}$ and its coordinates are the Fourier coefficients of f. There is an alternative way to compute the Fourier coefficients of f without simplifying the interpolation method (1.1).

Proposition 1.10. *For $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ and $S \subseteq [n]$ the Fourier coefficient of f on S is given by $\hat{f}(S) = \langle f, \chi_S \rangle$.*

Now that we have a directly way to compute the Fourier representation of f and an orthonormal basis generated by parity functions we can introduce two important theorems. The first one, *Parseval's Theorem*, lets us measure the squared "length" (2-norm) of f efficiently.

Theorem 1.11 (Parseval's Theorem)**.** *For any $f : \{-1,1\}^n \longrightarrow \mathbb{R}$,*

$$\langle f, f \rangle = \mathop{\mathbf{E}}_{\mathbf{x}\sim\{-1,1\}^n}[f(\mathbf{x})^2] = \sum_{S\subseteq[n]} \hat{f}(S)^2.$$

In particular, if $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ is boolean-valued, then

$$\langle f, f \rangle = \sum_{S\subseteq[n]} \hat{f}(S)^2 = 1.$$

The second theorem named *Plancherel's Theorem* generalizes Parseval's Theorem giving a similar result but for two different functions f and g.

Theorem 1.12 (Plancherel's Theorem)**.** *For any $f, g : \{-1,1\}^n \longrightarrow \mathbb{R}$,*

$$\langle f, g \rangle = \mathop{\mathbf{E}}_{\mathbf{x}\sim\{-1,1\}^n}[f(\mathbf{x})g(\mathbf{x})] = \sum_{S\subseteq[n]} \hat{f}(S)\hat{g}(S).$$

The proof of this two theorems is straight forward by computing the inner product substituting each function by its Fourier expansion and using the orthogonality of the basis.

Boolean-valued functions are easier to interpret than real-valued since there are only two possibilities for its images. In mathematical terms it is possible to see if two images are the same by multiplying them: if $f(\mathbf{x})g(\mathbf{x}) = 1$ then $f(\mathbf{x}) = g(\mathbf{x})$ otherwise $f(\mathbf{x}) \neq g(\mathbf{x})$. We can also define a distance between boolean-valued functions.

Definition 1.13. Given $f, g : \{-1,1\}^n \longrightarrow \{-1,1\}$, we define their (relative Hamming) *distance* to be the fraction of inputs on which they disagree:

$$dist(f,g) = \mathop{\mathbf{Pr}}_{\mathbf{x}\sim\{-1,1\}^n}[f(\mathbf{x}) \neq g(\mathbf{x})].$$

So the inner product between two functions can be interpreted as some type of correlation between f and g. And for boolean-valued functions it can be written in terms of the Hamming distance.

Proposition 1.14. *If $f, g : \{-1,1\}^n \longrightarrow \{-1,1\}$:*

$$\langle f, g\rangle = \mathbf{Pr}[f(\mathbf{x}) = g(\mathbf{x})] - \mathbf{Pr}[f(\mathbf{x}) \neq g(\mathbf{x})] = 1 - 2dist(f,g).$$

Finally, we can start studying some interesting combinatorial properties of a Boolean function f from its Fourier coefficients. But first some definitions, related to Probability Theory on f, are given.

Definition 1.15. The *mean* of $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ is $\mathbf{E}[f]$. When f has mean equally 0 we say that it is *unbiased* or *balanced*. In the particular case that f is boolean-valued, its mean is:

$$\mathbf{E}[f] = \mathbf{Pr}[f = 1] - \mathbf{Pr}[f = -1].$$

Thus f is unbiased if and only if it takes value 1 on exactly half of the points of the Hamming cube. Next, we can also define a variance for a real-valued Boolean function:

Definition 1.16. The *variance* of $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ is:

$$\mathbf{Var}[f] = \langle f - \mathbf{E}[f], f - \mathbf{E}[f]\rangle = \mathbf{E}[f^2] - \mathbf{E}[f]^2.$$

Definition 1.17. The *covariance* of $f, g : \{-1,1\}^n \longrightarrow \{-1,1\}$ is:

$$\mathbf{Cov}[f,g] = \langle f - \mathbf{E}[f], g - \mathbf{E}[g]\rangle = \mathbf{E}[fg] - \mathbf{E}[f]\mathbf{E}[g].$$

The Definitions 1.15, 1.16 and 1.17 can also be expressed in terms of the Fourier coefficients.

Proposition 1.18. *If $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ then*

(1) $\mathbf{E}[f] = \hat{f}(\emptyset)$.

(2) $\mathbf{Var}[f] = \sum_{S\neq\emptyset} \hat{f}(S)^2$.

(3) $\mathbf{Cov}[f,g] = \sum_{S\neq\emptyset} \hat{f}(S)\hat{g}(S)$.

Proof. The first one is proved multiplying f by 1 inside the expectation and substituting 1 for the identical function that is $\chi_\emptyset$. Meanwhile the second is proved using the definition of the second equality in Definition 1.16 and the Parseval's Theorem 1.11. Finally Plancherel's Theorem 1.12 shows the third. ☐

Another interesting approach is to study how the variance works on boolean-valued functions using the fact that they only have two possible images: ± 1.

Proposition 1.19. *For $f : \{-1,1\}^n \longrightarrow \{-1,1\}$*

$$\mathbf{Var}[f] = 1 - \mathbf{E}[f]^2 = 4\mathbf{Pr}[f(\mathbf{x}) = 1]\mathbf{Pr}[f(\mathbf{x}) = -1].$$

Hence the variance of a boolean-valued function it is always between 0 and 1. In particular, f has variance 1 if it is unbiased and variance 0 if it is constant. Then we can think that the variance is proportional to its distance from being constant.

1.4 Boolean functions generating probability distributions

This chapter will talk about how a boolean-valued function can create a probability distribution. But first a more general definition is given for real-valued Boolean functions:

Definition 1.20. The *Fourier weight* of $f : \{-1, 1\}^n \longrightarrow \mathbb{R}$ on set S is defined to be the squared Fourier coefficient $\hat{f}(S)^2$.

Surprisingly, many formulas depend on the weights of f. For example, the Proposition 1.18 states that the variance of f depends on its Fourier weights on non empty sets. In boolean-valued functions, Parseval's Theorem 1.11 assures that the total weight of f sums 1. Therefore we can think that they define a probability distribution on subsets of $[n]$.

Definition 1.21. Given $f : \{-1, 1\}^n \longrightarrow \{-1, 1\}$, the *spectral sample* for f, denoted $\mathcal{S}_f$, is the probability distribution on subsets of $[n]$ in which the set S has probability $\hat{f}(S)^2$. We write $S \sim \mathcal{S}_f$ for a draw from this distribution.

We often stratify the subsets $S \subseteq [n]$ according to their cardinal. Equivalently this is the degree associated to the monomial $\mathbf{x}^S$.

Definition 1.22. For $f : \{-1, 1\}^n \longrightarrow \mathbb{R}$ and $0 \leq k \leq n$, the *Fourier weight of f at degree k* is:

$$\mathbf{W}^k[f] = \sum_{\substack{S \subseteq [n] \\ |S|=k}} \hat{f}(S)^2.$$

If $f : \{-1, 1\}^n \longrightarrow \{-1, 1\}$ is boolean-valued, an equivalent definition is:

$$\mathbf{W}^k[f] = \mathop{\mathbf{Pr}}_{S \sim \mathcal{S}_f} [|S| = k].$$

By Parseval's Theorem 1.11 $\mathbf{W}^k[f] = \|f^{=k}\|_2^2$, where

$$f^{=k} = \sum_{|S|=k} \hat{f}(S) \chi_S$$

is called the *degree k part of f*. We will also sometimes use notation like $\mathbf{W}^{>k}[f]$ and $f^{\leq k}$ for similar significance.

2. Social Choice Theory

This chapter introduces the application of Boolean functions in Social Choice Theory. They will be useful tools to approach voting election systems problems. It will follow the structure of the second chapter of O'Donnell's reference book [1] with the objective to define Arrow's Theorem. The theorem will be proved rewriting Arrow's work on [2] with mathematical objects. Some extended results from Kalai presented in [3] will give a mathematical proof of the theorem and a stability result.

2.1 Social choice functions and initial concepts

The fundamental question in Social Choice Theory is how to aggregate the options of many agents. A boolean-valued function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ can be thought as a *voting rule* or *social choice function* for an election with two candidates and n voters mapping the votes of every individual to a winner. Now some social choice functions, which define an election system, will be presented.

Definition 2.1. For n odd: the *majority function* Maj_n is defined by:

$$\begin{aligned} Maj_n \colon \{-1,1\}^n &\to \{-1,1\} \\ \mathbf{x} &\mapsto sgn(x_1 + \cdots + x_n). \end{aligned}$$

Note that it must be defined for n odd. But for n even we say that $f(\mathbf{x})$ is a majority function if $sgn(x_1 + \cdots + x_n)$ equals $f(\mathbf{x})$ whenever is non-zero.

Remember the identification we made before between -1 and "True" and $+1$ and "False". The following two systems elect a winner if all voters agree with "True" or elect a winner if at least one voter agrees with "True", respectively.

Definition 2.2. The function AND_n is defined by:

$$\begin{aligned} AND_n \colon \{-1,1\}^n &\to \{-1,1\} \\ \mathbf{x} &\mapsto \begin{cases} -1 & \text{if } \mathbf{x} = (-1,\dots,-1), \\ +1 & \text{otherwise.} \end{cases} \end{aligned}$$

The function OR_n is defined by:

$$\begin{aligned} OR_n \colon \{-1,1\}^n &\to \{-1,1\} \\ \mathbf{x} &\mapsto \begin{cases} +1 & \text{if } \mathbf{x} = (+1,\dots,+1), \\ -1 & \text{otherwise.} \end{cases} \end{aligned}$$

Another voting rule, simple but really important for the analysis of Boolean functions and Arrow's Theorem, is the *dictator function*.

Definition 2.3. The *ith dictator function* χ_i is defined by:

$$\begin{aligned} \chi_i \colon \{-1,1\}^n &\to \{-1,1\} \\ \mathbf{x} &\mapsto x_i. \end{aligned}$$

We can think this like a projection mapping. A generalization of this notion is defined as it follows.

Definition 2.4. A function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ is called a *k-junta* for $k \in \mathbb{N}$ if it depends on at most k of its input coordinates. In mathematical terms: $f(\mathbf{x}) = g(x_{i_1}, \dots, x_{i_k})$ for some $g : \{-1,1\}^k \longrightarrow \{-1,1\}$ and $i_1, \dots, i_k \in [n]$.

A voting rule can also be defined as a weighted majority rule like the following one.

Definition 2.5. A function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ is called a *weighted majority* or *threshold function* if there exist $a_0, a_1, \dots, a_n \in \mathbb{R}$ such that:

$$\begin{aligned} f \colon \{-1,1\}^n &\to \{-1,1\} \\ \mathbf{x} &\mapsto sgn(a_0 + a_1x_1 + \cdots + a_nx_n). \end{aligned}$$

It is possible to see that the previous Definitions 2.1, 2.2, 2.3 are weighted majority functions. For example, the Dictator function 2.3 is a threshold function with all terms equal to 0 except a_i, which will be equal to 1.

Finally, we will introduce a voting system that will be really important later for the study of KKL Theorem 3.14. Now the voters are divided into tribes (groups of voters) and if a tribe elects a candidate unanimously it will win the election.

Definition 2.6. The *tribes function of width w and size s*, is defined by:

$$\begin{aligned} Tribes_{w,s} \colon \{-1,1\}^{ws} &\to \{-1,1\} \\ (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) &\mapsto OR_s(AND_w(\mathbf{x}^{(1)}), \dots, AND_w(\mathbf{x}^{(s)})). \end{aligned}$$

Where $\mathbf{x}^{(i)} \in \{-1,1\}^w$.

So these are some examples of Boolean functions that can provide a voting system perhaps intuitive for us. Now we can define, like with real functions, some properties that Boolean functions can satisfy. These properties will be reasonable hypothesis to ensure a fair voting system.

Definition 2.7. We say that a Boolean function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ is:

(1) *Monotone*: if $\mathbf{x} \le \mathbf{y}$ coordinate-wise, then $f(\mathbf{x}) \le f(\mathbf{y})$.

(2) *Odd*: if $f(-\mathbf{x}) = -f(\mathbf{x})$.

(3) *Unanimous*: if $f(1, \dots, 1) = 1$ and $f(-1, \dots, -1) = -1$.

(4) *Symmetric*: if $f(\mathbf{x}^\pi) = f(\mathbf{x})$ for all permutations $\pi \in \mathcal{S}_n$.

(5) *Transitive-symmetric*: if for all $i, i' \in [n]$ there exists a permutation $\pi \in \mathcal{S}_n$ exchanging i and i' such that $f(\mathbf{x}^\pi) = f(\mathbf{x})$ for all $\mathbf{x} \in \{-1,1\}^n$.

Note that the symmetric functions depend only on how many 1's the input has. A really important result presented by May in [4] is that many functions satisfy some of these properties but there is only one that satisfy all of them.

Theorem 2.8 (May's Theorem)**.** *Let $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ be monotone, odd, unanimous and symmetric. Then f must be the majority function Maj_n.*

Proof. First, note that the function f is symmetric thus $f(\mathbf{x}^{\pi}) = f(\mathbf{x})$ for every permutation $\pi \in \mathcal{S}_n$. In other words, the value of $f(\mathbf{x})$ only depends on the number of 1's of the input $\mathbf{x}$. So we can pack all the inputs in sets depending on the number of ones, we denote them $\mathbf{N}(k) = \{\mathbf{x} \in \{-1,1\}^n : \#\{x_i : x_i = +1\} = k\}$.

Second, since the function is unanimous $f(\mathbf{N}(0)) = -1$ and $f(\mathbf{N}(n)) = 1$. Additionally $f(\mathbf{N}(i)) \leq f(\mathbf{N}(j))$ for $i \leq j$ because of the monotonicity. Tu sum up, all these conditions imply that f starts being worth -1 and ends being $+1$, however it can not jump on different values -1,+1,-1 because of the monotonicity so we have to find with how many 1's f changes its value from -1 to $+1$.

Third, observe that the input $-\mathbf{x}$, in terms of the defined packages $\mathbf{N}(i)$, is $-\mathbf{N}(i) = \mathbf{N}(n-i)$. Then if f is odd for every $i \in [n]$ we have $f(-\mathbf{N}(i)) = -f(\mathbf{N}(n-i))$. If we start computing this result for the basic cases it makes sense: $+1 = f(\mathbf{N}(n)) = f(-\mathbf{N}(0)) = -f(\mathbf{N}(0)) = -(-1)$. In fact, it is necessary for n to be odd because there will be an even number of packages $\mathbf{N}(i)$ and hence an exact number of pairs $(\mathbf{N}(i), \mathbf{N}(n-i))$ for $i = 0, \dots, \frac{n-1}{2}$.

In conclusion, the four conditions on the function f forced us to define a function for inputs of odd size which take value -1 through the first half of the inputs and at the middle $\frac{n+1}{2}$ starts taking value $+1$. This is the definition of the majority function Maj_n. □

Before finishing this part we can be pessimistic and think that it is impossible that the n voters are not related. At least this should be the objective from a wise point of view. Although it seems unrealistic, we can define a vote in which the n voters preference are independent and it will be a good basis to study the undecided or party-independent voters. It also offers an advantage for studying functions from the point of view of Probability Theory.

Definition 2.9. The *impartial culture assumption* is that the n voters' preferences are independent and chosen uniformly at random.

2.2 Influence, operators and noise stability

Once we have begun to think Boolean functions as voting rules it is natural to question how can we measure the influence or power of a voter. It can be interpreted as the probability that the voter preference affects the outcome.

Definition 2.10. We say that coordinate $i \in [n]$ is *pivotal* for $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ on input $\mathbf{x}$ if $f(\mathbf{x}) \neq f(\mathbf{x}^{\oplus i})$, where $\mathbf{x}^{\oplus i} = (x_1, \dots, x_{i-1}, -x_i, x_{i+1}, \dots, x_n)$. The *influence of coordinate* i on the boolean-valued function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ is defined to be:

$$\mathbf{Inf}_i[f] = \mathop{\mathbf{Pr}}_{\mathbf{x}\sim\{-1,1\}^n}[f(\mathbf{x}) \neq f(\mathbf{x}^{\oplus i})].$$

Influence measures how a variable "influence" the output of the Boolean function f. In another words, how determinant is a voter on the result of an election between two candidates.

Example 2.11. Here we compute the influences of some functions which have been defined in Section 2.1.

(1) For the *ith dictator function* χ_i the coordinate i is pivotal for every input, hence $\mathbf{Inf}_i[\chi_i] = 1$. On the other hand if $j \neq i$ then coordinate j is never pivotal and the influence is $\mathbf{Inf}_j[\chi_i] = 0$.

(2) If $f = \pm 1$ (constant function) then any coordinate $i \in [n]$ is not pivotal and $\mathbf{Inf}_i[f] = 0$.

(3) For the OR_n function, coordinate 1 is pivotal for exactly two inputs: $(-1, 1, \dots, 1)$ and $(1, 1, \dots, 1)$, hence

$$\mathbf{Inf}_1[OR_n] = \frac{1}{2^n} + \frac{1}{2^n} = \frac{1}{2^{n-1}} = 2^{1-n}.$$

Observation 2.12. Another way to treat the Hamming cube is by using geometry. This will help in other applications like the one presented in Sensitivity Conjecture in Chapter 4. A n-dimensional cube can be drawn with vertices as the inputs $\mathbf{x} \in \{-1, 1\}^n$. For boolean-valued functions $f : \{-1, 1\}^n \longrightarrow \{-1, 1\}$ we can paint the vertices, for example, in grey if $f(\mathbf{x}) = -1$ and black if $f(\mathbf{x}) = +1$.

We will define $(\mathbf{x}, \mathbf{y})$ as a *dimension-i edge* if $\mathbf{y} = \mathbf{x}^{\oplus i}$. And this dimension-i edge will be a *boundary edge* if $f(\mathbf{x}) \neq f(\mathbf{y})$. Hence, influence $\mathbf{Inf}_i[f]$ equals the fraction of dimension-i edges which are boundary edges.

Let us give an example by studying the majority function Maj_3. In the Figure 1 it can be seen that there are 2 boundary edges, which are drawn in a thicker edge, in each of the three dimensions. Since there is a total of 4 edges of each dimension i we have $\mathbf{Inf}_i[Maj_3] = 2/4 = 1/2$ for every i in [3].

Now it is possible to think how affected is the function by every coordinate by summing all the influences.

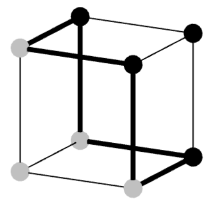

Figure 1: Hamming cube of Maj_n

Definition 2.13. The *total influence* of $f : \{-1, 1\}^n \longrightarrow \mathbb{R}$ is defined to be:

$$\mathbf{I}[f] = \sum_{i=1}^{n} \mathbf{Inf}_i[f].$$

Example 2.14. We see now the total influence of the functions given in Example 2.11.

(1) For the *ith dictator function* we have seen that $\mathbf{Inf}_i[f] = \delta_{i,j}$ where $\delta_{i,j}$ is the Kronecker's delta. So the total influence will be:

$$\mathbf{I}[\chi_i] = \sum_{j=1}^{n} \mathbf{Inf}_j[\chi_i] = \sum_{j=1}^{n} \delta_{i,j} = 1.$$

(2) If $f = \pm 1$ then $\mathbf{I}[f] = 0$ since the influences of all coordinates are zero.

(3) For the OR_n function, we have seen that every coordinate is pivotal for exactly two inputs, thus:

$$\mathbf{I}[OR_n] = \sum_{i=1}^{n} \mathbf{Inf}_i[OR_n] = \sum_{i=1}^{n} 2^{1-n} = n \cdot 2^{1-n}.$$

(4) If we study the total influence of Maj_3 we have:

$$\mathbf{I}[Maj_3] = \sum_{i=1}^{3} \mathbf{Inf}_i[Maj_3] = \sum_{i=1}^{3} 1/2 = 3/2.$$

Definition 2.15. The *sensitivity of f at* $\mathbf{x}$ is defined to be the number of pivotal coordinates for f at input $\mathbf{x}$. Hence:

$$sens_f(\mathbf{x}) = \#\{i \in [n] : f(\mathbf{x}^{(i \to +1)}) \neq f(\mathbf{x}^{(i \to -1)})\}.$$

This definition will be really important for the next application in Chapter 4. But it also gives a natural form to compute the total influence:

Proposition 2.16. *For $f : \{-1,1\}^n \longrightarrow \{-1,1\}$:*

$$\mathbf{I}[f] = \mathbf{E}[sens_f(\mathbf{x})].$$

By now we have been able to compute the influences of many functions. But this was possible because we knew the image for every input and have a intuitive idea of the function. We will introduce an adaptation of an analytical tool that is used in real functions with the aim to obtain the influence more analytically.

Definition 2.17. The *ith (discrete) derivative operator* D_i maps the function $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ to the function $D_i f : \{-1,1\}^n \longrightarrow \mathbb{R}$ like:

$$D_i f(\mathbf{x}) = \frac{f(x^{(i\to+1)}) - f(x^{(i\to-1)})}{2}.$$

The *(discrete) gradient operator* ∇ is defined to map the function $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ to the function $\nabla f : \{-1,1\}^n \longrightarrow \mathbb{R}^n$ like:

$$\nabla f(\mathbf{x}) = (D_1 f(\mathbf{x}), D_2 f(\mathbf{x}), \dots, D_n f(\mathbf{x})).$$

Realize that $D_i f(\mathbf{x})$ does not actually depend on x_i of the concrete input $\mathbf{x}$. Additionally, if $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ is boolean-valued, $D_i f(\mathbf{x})$ is ± 1 if coordinate i is pivotal for $\mathbf{x}$ and is 0 otherwise. Thus $D_i f(\mathbf{x})^2$ is the 0-1 indicator for whether i is pivotal for $\mathbf{x}$ so we can expand the definition of influence given for boolean-valued functions 2.10 to real-valued functions using the fact that the derivative operator is defined for this range of f.

Definition 2.18. We define the *influence of coordinate i on* $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ to be:

$$\mathbf{Inf}_i[f] = \mathbf{E}\left[D_i f(\mathbf{x})^2\right] = \|D_i f(\mathbf{x})\|_2^2.$$

In fact, now that the influence is defined for all the functions that we will work with, we might want to identify the coordinates that have some influence on the function $f : \{-1,1\}^n \longrightarrow \mathbb{R}$. So we will say that coordinate $i \in [n]$ is *relevant* if and only if $\mathbf{Inf}_i[f] > 0$. Thus, a coordinate will be relevant if $f(x^{(i\to+1)}) \neq f(x^{(i\to-1)})$ for some $\mathbf{x} \in \{-1,1\}^n$.

As every Boolean function, $D_i f(\mathbf{x})$ have a unique Fourier expansion like Theorem 1.5 states. The study of this Fourier expansion and the application of Parseval's Theorem 1.11 gives us the influence in terms of the Fourier coefficients of $D_i f(\mathbf{x})$.

Proposition 2.19. *Let $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ have the multilinear expansion $f(\mathbf{x}) = \sum_{S\subseteq[n]} \hat{f}(S)\mathbf{x}^S$.*

(1) Then:

$$D_i f(\mathbf{x}) = \sum_{\substack{S\subseteq[n] \\ i\in S}} \hat{f}(S)\mathbf{x}^{S\setminus\{i\}}.$$

(2) For $i \in [n]$:

$$\mathbf{Inf}_i[f] = \sum_{i\ni S} \hat{f}(S)^2.$$

Proof. The first statement is given by the linearity of the ith discrete derivative operator D_i.

$$D_i f(\mathbf{x}) = D_i\Big(\sum_{S\subseteq[n]} \hat{f}(S)\mathbf{x}^S\Big) = \sum_{S\subseteq[n]} \hat{f}(S) D_i(\mathbf{x}^S). \tag{2.1}$$

There can be two cases computing the ith derivative operator D_i on parity functions:

- $i \notin S$.
$$D_i(\mathbf{x}^S) = \frac{\mathbf{x}^S - \mathbf{x}^S}{2} = 0.$$
- $i \in S$.
$$D_i(\mathbf{x}^S) = \frac{x_1 \cdots (+1) \cdots x_{|S|} - x_1 \cdots (-1) \cdots x_{|S|}}{2} = \mathbf{x}^{S\setminus\{i\}}.$$

Therefore, the sum (2.1) can be rewritten like in the statement.

The second statement is a consequence derived from the first and from Parseval's Theorem 1.11.

$$\mathbf{Inf}_i[f] = \|D_i f(\mathbf{x})\|_2^2 = \sum_{S\ni i} \hat{f}(S)^2.$$

□

It is possible to generalize the last results given. Note that for $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ we have $\|\nabla f(\mathbf{x})\|_2^2 = sens_f(\mathbf{x})$. And we can expand this idea to real-valued functions $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ and obtain a similar proposition like the last one but for the total influence.

Proposition 2.20. *Let $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ have the multilinear expansion $f(\mathbf{x}) = \sum_{S\subseteq[n]} \hat{f}(S)\mathbf{x}^S$.*

(1) Total influence is defined like:
$$\mathbf{I}[f] = \mathbf{E}[\|\nabla f(\mathbf{x})\|_2^2].$$

(2) It can be computed from Fourier coefficients like
$$\mathbf{I}[f] = \sum_{S\subseteq[n]} |S| \cdot \hat{f}(S)^2 = \sum_{k=0}^{n} k \cdot \mathbf{W}^k[f].$$

(3) For a boolean-valued function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ the last result can be rewritten using the spectral sample of f (see Definition 1.21):
$$\mathbf{I}[f] = \mathop{\mathbf{E}}_{S\sim\mathcal{S}_f}[|S|].$$

Proof. First statement is derived from the definition of the gradient 2.17 and the definition of influence 2.18. The next two statements are proved straight forward with Proposition 2.19. □

A really important inequality can be derived from the total influence expressed in terms of the Fourier coefficients. This inequality is so useful that it has its own name.

Theorem 2.21 (Poincaré Inequality)**.** *For any $f : \{-1,1\}^n \longrightarrow \mathbb{R}$, $\mathbf{Var}[f] \leq \mathbf{I}[f]$.*

Proof. The proof expresses the variance and the total influence in terms of the Fourier coefficients as it is presented in Proposition 1.18 and Proposition 2.20, respectively.

$$\mathbf{Var}[f] = \sum_{k>0} \mathbf{W}^k[f] \leq \sum_{k\geq 0} k \cdot \mathbf{W}^k[f] = \mathbf{I}[f].$$

$\square$

Another useful operator can be defined to measure the mean of the value of $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ for a given input $\mathbf{x}$ without caring about some coordinate.

Definition 2.22. The *ith expectation operator* E_i is the linear operator on functions $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ defined by:

$$E_i f(\mathbf{x}) = \mathop{\mathbf{E}}_{x_i}[f(x_1, \dots, x_i, \dots, x_n)] = \frac{f(\mathbf{x}^{(i \to +1)}) + f(\mathbf{x}^{(i \to -1)})}{2}.$$

This operator has a similar behaviour to the differential operator since it also does not depend on the coordinate x_i. The operator applied to a real-valued function $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ is another real-valued function, hence the Theorem 1.5 states that it can be expressed as a Fourier expansion. The next proposition assures this and also provides a meaningful decomposition of f in terms of the differential and expectation operators applied to this function.

Proposition 2.23. *Let $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ have the multilinear expansion $f(\mathbf{x}) = \sum_{S\subseteq[n]} \hat{f}(S)\mathbf{x}^S$.*

(1) The Fourier expansion for the ith expectation operator on f where $i \in [n]$ is:

$$E_i f(\mathbf{x}) = \sum_{S \not\ni i} \hat{f}(S)\mathbf{x}^S.$$

(2) For every $i \in [n]$ f can be expressed as:

$$f(\mathbf{x}) = x_i D_i f(\mathbf{x}) + E_i f(\mathbf{x}).$$

Proof. First statement is given by the linearity of the ith expectation operator E_i.

$$E_i f(\mathbf{x}) = E_i\Big(\sum_{S\subseteq[n]} \hat{f}(S)\mathbf{x}^S\Big) = \sum_{S\subseteq[n]} \hat{f}(S)E_i(\mathbf{x}^S). \tag{2.2}$$

There can be two cases computing the ith expectation operator E_i on parity functions:

- $i \notin S$.

$$E_i(\mathbf{x}^S) = \frac{\mathbf{x}^S - \mathbf{x}^S}{2} = \mathbf{x}^S$$

- $i \in S$.

$$E_i(\mathbf{x}^S) = \frac{x_1 \cdots (+1) \cdots x_{|S|} + x_1 \cdots (-1) \cdots x_{|S|}}{2} = 0.$$

Plugging this result on the sum (2.2) we obtain the desired result.

Second statement is shown simply by differentiating terms on the sum. Recalling the statement above and the first statement of Proposition 2.19 we write f in terms of the Fourier expansion given by Theorem 1.5:

$$f(\mathbf{x}) = \sum_{S\subseteq[n]} f(\mathbf{x}) = x_i \cdot \sum_{\substack{S\subseteq[n] \\ i\in S}} \hat{f}(S)\mathbf{x}^{S\setminus\{i\}} + \sum_{\substack{S\subseteq[n] \\ i\notin S}} \mathbf{x}^S = x_i \cdot D_i f(\mathbf{x}) + E_i f(\mathbf{x}).$$

□

Finally we can define the last operator which lets us have an idea about how different is the value of f on an input $\mathbf{x}$ compared to what is expected to be.

Definition 2.24. The *ith coordinate Laplacian operator* L_i on $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ is defined by:

$$L_i f(\mathbf{x}) = f(\mathbf{x}) - E_i f(\mathbf{x}) = \frac{f(\mathbf{x}) - f(\mathbf{x}^{\oplus i})}{2}.$$

More generally, the *Laplacian operator* L can be defined as a linear operator by:

$$Lf(\mathbf{x}) = \sum_{i=1}^{n} L_i f(\mathbf{x}).$$

The concept behind the interpretation of the Laplacian can be helpful to make assumptions about f. If this operator is defined just behind the definition of influence and the other operators is because some results can be provided by exploiting the previous propositions. In fact we will also see that the Laplacian is closely related to the sensitivity defined in Definition 2.15.

Proposition 2.25. *Let* $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ *have the multilinear expansion* $f(\mathbf{x}) = \sum_{S\subseteq[n]} \hat{f}(S)\mathbf{x}^S$.

(1) *Suppose for this result that* $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ *so we can get the sensitivity in terms of the Laplacian:*

$$Lf(\mathbf{x}) = f(\mathbf{x})sens_f(\mathbf{x}).$$

(2) *The Fourier expansion for the ith coordinate Laplacian operator on* f *with* i *in* $[n]$ *is:*

$$L_i f(\mathbf{x}) = x_i D_i f(\mathbf{x}) = \sum_{i\ni S} \hat{f}(S)\mathbf{x}^S.$$

(3) *The Fourier expansion for the Laplacian operator on* f *is:*

$$Lf(\mathbf{x}) = \sum_{S\subseteq[n]} |S|\hat{f}(S)\chi_S(\mathbf{x}).$$

(4) *Influence on coordinate* i *in* $[n]$ *can be obtained in terms of the ith coordinate Laplacian operator on* f *as:*

$$\langle f, L_i f\rangle = \|L_i f\|_2^2 = \mathbf{Inf}_i[f].$$

(5) *Total influence can be derived also from Laplacian operator on* f*:*

$$\langle f, Lf\rangle = \mathbf{I}[f].$$

Proof. All the results are proved straight forward by the decomposition of f presented in the second statement of Proposition 2.23. □

Until now it was supposed that the votes are always well interpreted, but there exists a possibility that some vote is not correctly recorded. The following definitions approach the study of how garbled votes can affect the result of an election choice.

Definition 2.26. Let $\rho \in [0,1]$. For fixed $\mathbf{x} \in \{-1,1\}^n$ we write $\mathbf{y} \sim N_\rho(\mathbf{x})$ to denote that the random string $\mathbf{y}$ is drawn as it follows: for every i in $[n]$ independently,

$$y_i = \begin{cases} x_i & \text{with probability } \rho, \\ \text{uniformly random} & \text{with probability } 1-\rho. \end{cases}$$

But we extend the notation to all $\rho \in [-1,1]$ as it follows,

$$y_i = \begin{cases} x_i & \text{with probability } 1/2 + 1/2\rho, \\ -x_i & \text{with probability } 1/2 - 1/2\rho. \end{cases}$$

and we say that $\mathbf{y}$ *is ρ-correlated to* $\mathbf{x}$.

Definition 2.27. Given $\mathbf{x} \sim \{-1,1\}^n$ drawn uniformly random and $\mathbf{y} \sim N_\rho(\mathbf{x})$, we say that $(\mathbf{x},\mathbf{y})$ is a *ρ-correlated pair*. Sometimes we will abbreviate it as *ρ-corr*.

Note that it is equivalent to saying that, independently for each i in $[n]$, every pair of random bits (x_i, y_i) satisfies $\mathbf{E}[x_i] = \mathbf{E}[y_i] = 0$ and $\mathbf{E}[x_i y_i] = \rho$.

To sum up we define a new concept and later we will talk about how can we interpret it.

Definition 2.28. For $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ and $\rho \in [-1,1]$, the *noise stability of f at ρ* is:

$$\mathbf{Stab}_\rho[f] = \mathop{\mathbf{E}}_{\substack{(\mathbf{x},\mathbf{y}) \\ \rho-corr}}[f(\mathbf{x})f(\mathbf{y})].$$

In fact, we are computing the expected value of the function f valued on every $\mathbf{x}$ of the Hamming cube multiplied by f valued on an input derived from $\mathbf{x}$ but where the coordinates can be garbled with probability ρ. But this concept does not allow yet to perceive how affected is the value of the outcome if there is an affected input. However, for boolean-valued functions it is really interpretative.

Observation 2.29. Let $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ then, since $f(\mathbf{x})$ and $f(\mathbf{y})$ can be only two options and the sign of the product between them can identify if they are equal or not, we can interpret:

$$\mathbf{Stab}_\rho[f] = \mathop{\mathbf{Pr}}_{\substack{(\mathbf{x},\mathbf{y}) \\ \rho-corr}}[f(\mathbf{x}) = f(\mathbf{y})] - \mathop{\mathbf{Pr}}_{\substack{(\mathbf{x},\mathbf{y}) \\ \rho-corr}}[f(\mathbf{x}) \neq f(\mathbf{y})] = 2 \cdot \mathop{\mathbf{Pr}}_{\substack{(\mathbf{x},\mathbf{y}) \\ \rho-corr}}[f(\mathbf{x}) = f(\mathbf{y})] - 1.$$

So for boolean-valued functions it is possible to compute the probability that the missrecorded votes do not affect the outcome:

$$\mathop{\mathbf{Pr}}_{\substack{(\mathbf{x},\mathbf{y}) \\ \rho-corr}}[f(\mathbf{x}) = f(\mathbf{y})] = \frac{1}{2}\mathbf{Stab}_\rho[f] + \frac{1}{2}.$$

Example 2.30. Here we compute the noise stability of some social choice functions defined previously:

(1) The constant functions $f = \pm 1$ have the same noise stability for every ρ:

$$\mathbf{Stab}_\rho[f] = 2 \cdot \Pr_{\substack{(\mathbf{x},\mathbf{y}) \\ \rho-corr}} [f(\mathbf{x}) = f(\mathbf{y})] - 1 = 2 - 1 = 1.$$

(2) The parity functions' χ_S noise stability only depends on ρ:

$$\begin{aligned}\mathbf{Stab}_\rho[\chi_S] = \mathop{\mathbf{E}}_{\substack{(\mathbf{x},\mathbf{y}) \\ \rho-corr}} [\mathbf{x}^S \mathbf{y}^S] &= \mathop{\mathbf{E}}_{\substack{(\mathbf{x},\mathbf{y}) \\ \rho-corr}} [\prod_{i \in S} x_i y_i] \\ &= \prod_{i \in S} \mathop{\mathbf{E}}_{\substack{(\mathbf{x},\mathbf{y}) \\ \rho-corr}} [x_i y_i] = \prod_{i \in S} \rho = \rho^{|S|}.\end{aligned}$$

Finally, we introduce the most important operator of the chapter. This operator will provide practical results for the noise stability but also good behaviour in terms of hypercontractivity as it will be shown in Section 3.1.

Definition 2.31. For $\rho \in [-1, 1]$, the *noise operator with parameter* ρ is the linear operator T_ρ on functions $f : \{-1, 1\}^n \longrightarrow \mathbb{R}$ defined by:

$$T_\rho f(\mathbf{x}) = \mathop{\mathbf{E}}_{\mathbf{y} \sim N_\rho(\mathbf{x})} [f(\mathbf{y})].$$

So for every input $\mathbf{x}$ the function $T_\rho f$ maps it to the expectation of the value that f will have for all the missrecorded inputs $\mathbf{y}$ derived from $\mathbf{x}$. This operator also has a fancy Fourier expansion and can be applied to compute the noise stability.

Proposition 2.32. *Let $f : \{-1, 1\}^n \longrightarrow \mathbb{R}$ have the multilinear expansion $f(\mathbf{x}) = \sum_{S \subseteq [n]} \hat{f}(S) \mathbf{x}^S$.*

(1) Given $\rho \in [-1, 1]$ we can relate T_ρ and noise stability:

$$\mathbf{Stab}_\rho[f] = \mathop{\mathbf{E}}_{\mathbf{x}}[f(\mathbf{x}) T_\rho f(\mathbf{x})] = \langle f, T_\rho f \rangle.$$

(2) The Fourier expansion of $T_\rho f$ is:

$$T_\rho f = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) \chi_S = \sum_{k=0}^{n} \rho^k f^{=k}.$$

(3) Noise stability can be also provided by the Fourier coefficients:

$$\mathbf{Stab}_\rho[f] = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S)^2 = \sum_{k=0}^{n} \rho^k \mathbf{W}^k[f].$$

Proof. The first statement is proved by writing the definition of noise stability and looking how the noise operator interferes. The second has only one key step and is the proof that $T_\rho \chi_S = \rho^{|S|} \chi_S$. The last result is proved using the previous two equalities and the Plancherel's Theorem 1.12. ☐

Now that we have introduced the noise stability term we can define a new influence. It has not any natural interpretation but we will see how profitable it is in the Section 3.3.

Definition 2.33. For $f : \{-1, 1\}^n \longrightarrow \mathbb{R}$, $\rho \in [0, 1]$ and i in $[n]$ the *ρ-stable influence* of i on f is:

$$\mathbf{Inf}_i^{(\rho)}[f] = \mathbf{Stab}_\rho[D_i f] = \sum_{i \ni S} \rho^{|S|-1} \hat{f}(S)^2.$$

We also define $\mathbf{I}^{(\rho)}[f] = \sum_{i=1}^n \mathbf{Inf}_i^{(\rho)}[f]$.

If we look closer to the representation of the ρ-stable influence of i on f given by the Fourier coefficients we can compare it to Proposition 2.19 and see that it is closer to the natural influence defined in Definition 2.18 when ρ is closer to 1.

2.3 Arrow's Theorem proof using ordered sets

Kenneth Arrow was an American economist that worked in a theory to study which was the better way to take decisions in a society. This work, along with others, awarded him the Nobel prize in 1972. For m alternatives the objective was to take a collective decision that represented, as best as possible, the society's preference given the decisions of n individuals. All this study was presented in [2].

To simplify the theorem he worked with 3 alternatives A, B, C and n individuals. An example will be given now to clarify the procedure. Condorcet created a system to order society preference's. If we use the majority function to decide in every pair-wise competition we obtain a result like this:

	Voter's preference			
	1	**2**	**3**	**Society choice**
a_{+1} vs b_{-1}	+1	-1	+1	+1
b_{+1} vs c_{-1}	+1	+1	-1	+1
c_{+1} vs a_{-1}	-1	+1	+1	+1

Table 1: Condorcet Election with 3 alternatives and 3 candidates

If we look closer to society's choice we see that it gives a non-rational outcome although the voting preferences of all the individuals were rational. This is called the Condorcet paradox and Arrow's objective was to study whether it would be possible to define a voting system where this paradox is avoided.

Before giving any solution or result some notation is defined but the intention of this thesis is to transform the exposition of the concepts, theorem and proof in mathematical terms. Alternatives are denoted as small letters x, y, z and every chooser must make one and only one of three decisions: x is preferred to y, x is indifferent to y and y preferred to x. This decisions are assumed to be consistent: if x is preferred to y and y indifferent to z then x is preferred to z. We denote x preferred or indifferent to y as xRy.

From the definition of preference and indifference there are two axioms that all decisions satisfy:

- *Axiom I.* Every alternative is comparable. For all x and y, either xRy or yRx.
- *Axiom II.* Preferences of the electors are rational. For all x, y and z if xRy and yRz this implies xRz.

However a relation R satisfying these 2 axioms is termed a weak ordering because it does not differ if the alternatives are preferred or indifferent. So a stronger order is created by defining the notation xPy to mean not yRx and to be read like x is preferred to y. It is also defined xIy to mean xRy and yRx and it will be read x indifferent to y.

Lemma 2.34. *All these properties are satisfied by relationships between alternatives x, y, z:*

(a) For all x, xRx.

(b) If xPy, then xRy.

(c) If xPy and yPz, then xPz.

(d) If xIy and yIz then xIz.

(e) For all x and y, either xRy or yPx.

(f) If xPy and yRz, then xPz.

Since we are saying that the R relation between two alternatives satisfy reflexivity, antisymmetry and transitivity properties we can relate this to mathematical order theory. We can think that every elector defines a partially ordered set from the alternatives set S. Hence, if x is preferred to y we have $y \leq x$.

As for notation, let R_i be the ordering relation for alternative social states from the standpoint of individual i. Sometimes when different ordering relations are being considered for the same individual, the symbols will be distinguished by adding a supperscript. For (strict) preference relation we write P_i and for indifference I_i. In mathematical terms is used $\leq_i$ for relation R_i, $<_i$ for preference P_i and $=_i$ for indifference I_i. Similarly, society will be considered to have a social ordering relation for alternative social states designed by R ($\leq$), sometimes with a prime or second. Social preference and indifference will be denoted by P ($<$) and I ($=$), respectively.

For the next analysis, it will be assumed that individuals are rational, by which is meant that the ordering relations R_i satisfy the two axioms. Hence, every individual i will give a partially ordered set from the alternatives set that will be denoted $(S)_{\leq_i}$. The problem will be to obtain also a rational choice-making from the society.

Arrow searched for a *social welfare function* defined as a process which, for each partially ordered set $(S)_{\leq_i}$ given the set of alternatives S, states a partially ordered set for the society's preferences $(S)_{\leq}$. Five conditions were proposed on the social welfare function. They were premeditated and represent a fair way to elect a winner. For simplicity, the work is done with two individuals and 3 alternatives.

1. *Well defined*: The social welfare function is defined for every admissible pair of individuals' partially ordered sets $(S)_{\leq_i}$. Although it is trivial, this should be emphasized to avoid problems with the solutions.

2. *Positive association of social and individual values:* If an alternative x rises or does not fall in each of the partially ordered sets $(S)_{\leq_i}$ without any other change in those orderings then, although there are changes in the orders between other alternatives, x remains at the same position in the society's partially ordered set $(S)_{\leq}$. So if $x \leq_i y$ before the change in individual's orderings, then still $x \leq_i y$.

3. *Independence of irrelevant alternatives*: Given the 2 individuals' preferences $(S)_{\leq_1}$, $(S)_{\leq_2}$ and another option of preferences $(S)'_{\leq_1}$, $(S)'_{\leq_2}$ if there is a pair of alternatives with the same order in the ordered

sets $(S)_{\leq_i}$ and $(S)'_{\leq_i}$ then the order given by the society's ordered set $(S)_{\leq}$ is the same either we choose one or the other individuals' preferences.

4. *Citizen's sovereignty*: The social welfare function is not imposed. There is not any pair of alternatives for any individual's ordered set $(S)_{\leq_i}$ that has an imposed order. Every individual is independent and free to choose his preferences.

5. *Non-dictatorship*: The social welfare function is not a dictatorship function. We have defined this function in Definition 2.3, but in the current terms it means that the society's ordered set $(S)_{\leq}$ is not equally one corresponding to an individual $(S)_{\leq_i}$.

A group of consequences are drawn from these 5 conditions. The combination of them allows to prove the *possibility theorem* or also known as *Arrow's Theorem*:

Theorem 2.35 (Arrow's Theorem, v1)**.** *If there are at least three alternatives among individuals can order in any way, then every social welfare function satisfying conditions 2 and 3 and respecting axioms I and II must be imposed or dictatorial.*

As it was said before, the proof is a construction of some consequences given the previous conditions. It is supposed that there exists a social welfare function satisfying these conditions and it will lead to a contradiction. As notation, x' and y' will be written as variables that represent possible alternatives on the values x, y, z.

The first consequence is a lemma that shows that the social welfare function is unanimous:

Lemma 2.36. *If $y' <_1 x'$ and $y' <_2 x'$, then $y' < x'$.*

Proof. By the condition 4 there exists ordered sets $(S)'_{\leq_1}$ and $(S)'_{\leq_2}$ with corresponding social preference $y' < x'$. Let us form an ordered set $(S)''_{\leq_1}$ from $(S)'_{\leq_1}$ by raising x' to the top while leaving y' and z' positions alone. We form $(S)''_{\leq_2}$ from $(S)'_{\leq_2}$ in the same way.

Since all we have done is raise x' to the top of every ordered set while leaving the others alone, in accordance to the condition 2: $y', z' < x'$. But, by construction, both individuals order $y' <_i x'$ in the ordered sets $(S)''_{\leq_i}$ $i = 1, 2$, and society's ordered set $(S)''_{\leq}$ satisfy $y' < x'$.

By condition 3, the order between x' and y' depends only on $(S)'_{\leq_i}$ for $i = 1, 2$, it follows that whenever both individuals order $y' <_i x'$, regardless of the rank of z', society's order is the same for $(S)'_{\leq_i}$ and $(S)''_{\leq_i}$. So society will order $y' < x'$. $\square$

This second lemma assures that once the order of the first individual prevails against the opposition of the second, then individual 1's order will prevail if 2 is indifferent or if he agrees with 1.

Lemma 2.37. *Suppose that for some fixed alternatives x' and y', when $y' <_1 x'$ and $x' <_2 y'$, $y' < x'$. Then, whenever $y' <_1 x'$, $y' < x'$.*

Proof. Let $(S)_{\leq_1}$ be an ordered set which orders $y' <_1 x'$ and let $(S)_{\leq_2}$ be any ordered set. Let $(S)'_{\leq_1}$ be the same as $(S)_{\leq_1}$ but now let $(S)'_{\leq_2}$ be $(S)_{\leq_2}$ but with x' being the lowest element and prevailing the order of the two other alternatives given in S. By construction $y' <'_1 x'$ and $x' <'_2 y'$. By hypothesis, $y' < x'$, where $(S)'_{\leq}$ is the ordered set derived from $(S)'_{\leq_i}$ for $i = 1, 2$.

Now the only difference between the ordered sets $(S)_{\leq_i}$ and $(S)'_{\leq_i}$ for $i = 1, 2$ is that x' is raised at the scale of order in the latter compared with the former. Hence, by condition 2 it follows from $y' <' x'$ that $y' < x'$. $\square$

The third lemma proves that if two ordered sets have opposite orders on two variables, then society ordered set has an equality in that comparison.

Lemma 2.38. *If $y' <_1 x'$ and $x' <_2 y'$, then $x' = y'$.*

Proof. Let us suppose it is false. Let $(S)_{\leq_i}$ for $i = 1, 2$ be ordered sets and for some alternatives x', y' with opposed order let the society ordered set $(S)_{\leq}$ state $y' < x'$ without loss of generalization. It will be shown that this assumption leads to a contradiction.

Again, without loss of generalization, let us say x', y' are fixed alternatives like $x' = x$ and $y' = y$. Collecting everything, the condition 3 states that society's order depends exactly on the orders given by the individuals:

$$\text{If } y <_1 x \text{ and } x <_2 y, \text{ then } y < x. \tag{2.3}$$

It will be shown that (2.3) leads to a contradiction.

Suppose $y <_1 x$ and $z <_1 y$, while $x <_2 z$ and $z <_2 y$ so individual 2 orders $x <_2 y$. By condition (2.3) society orders $y < x$. Also both order $z <_i y$ so by unanimity proved in Lemma 2.36 society orders $z < y$. By the transitive property of the society's partially ordered set $(S)_{\leq}$ we must have $z < x$. Therefore, there have been exhibited ordered sets $(S)_{\leq_i}$ such that $z <_1 x$ and $x <_2 z$, but $z < x$. But, by the condition 3 again, the society's order between x and z depends only on the individuals' preferences for x and z:

$$\text{If } z <_1 x \text{ and } x <_2 z, \text{ then } z < x. \tag{2.4}$$

Now we will consider all possible order cases and apply the conditions (2.3),(2.4) on both individual's ordered sets $(S)_{\leq_i}$.

Suppose $(S)_{\leq_1}$ has the order $z <_1 x <_1 y$ and $(S)_{\leq_2}$ has $x <_2 y <_2 z$. By unanimity shown at Lemma 2.36 society's order must be $x < y$, and by (2.4) $z < x$, so the society's ordered set $(S)_{\leq}$ must satisfy $z < y$. By the same reasoning as before with the condition 3:

$$\text{If } z <_1 y \text{ and } y <_2 z, \text{ then } z < y. \tag{2.5}$$

In other case if $(S)_{\leq_1}$ has the order $x <_1 z <_1 y$ and $(S)_{\leq_2}$ has $y <_2 x <_2 z$. By unanimity shown at Lemma 2.36 society's order must be $x < z$ and from (2.5) $z < y$, so the society's ordered set $(S)_{\leq}$ must satisfy $x < y$. By the same reasoning as before with the condition 3:

$$\text{If } x <_1 y \text{ and } y <_2 x, \text{ then } x < y. \tag{2.6}$$

Now if $(S)_{\leq_1}$ has the order $x <_1 y <_1 z$ and $(S)_{\leq_2}$ has $y <_2 z <_2 x$. By unanimity shown at Lemma 2.36 society's order must be $y < z$ and from (2.6) $x < y$, so the society's ordered set $(S)_{\leq}$ must satisfy $x < z$. By the same reasoning as before with the condition 3:

$$\text{If } x <_1 z \text{ and } z <_2 x, \text{ then } x < z. \tag{2.7}$$

Finally if $(S)_{\leq_1}$ has the order $y <_1 x <_1 z$ and $(S)_{\leq_2}$ has $z <_2 y <_2 x$. By unanimity shown at Lemma 2.36 society's order must be $y < x$ and from (2.7) $x < z$, so the society's ordered set $(S)_{\leq}$ must satisfy $y < z$. By the same reasoning as before with the condition 3:

$$\text{If } y <_1 z \text{ and } z <_2 y, \text{ then } y < z. \tag{2.8}$$

From (2.3)-(2.8) it follows that for any pair of alternatives x', y', whenever $x' <_1 y'$ then $x' < y'$. Therefore the order in the society's ordered set $(S)_\leq$ is given by the order on $(S)_{\leq_1}$. This implies that the individual 1 is a dictator. But we suppose from condition 5 that our social welfare function can not be a dictatorship. So we have come to a contradiction and we have proved the lemma. □

Now it is possible to prove Arrow's Theorem 2.35.

Proof. (Arrow's Theorem, v1) Suppose $(S)_{\leq_1}$ has the order $z <_1 y <_1 x$, while $(S)_{\leq_2}$ has the order $y <_2 x <_2 z$. By unanimity shown at Lemma 2.36 $y < x$. Since $z <_1 y$ and $y <_2 z$, it follows from Lemma 2.38 that $y = z$. Then by definition of the ordered set $(S)_\leq$ we have $z < x$.

But also $z <_1 x$ and $x <_2 z$, which implies $x = z$ by Lemma 2.38. It cannot be possible that $z < x$ but also $z = x$. Thus the assumption that there is a social welfare function compatible with conditions $1 - 5$ comes to a contradiction. Or in another way, to satisfy conditions 2-3 so Lemma 2.36 holds then either condition 4 or 5 must be violated so we must have an imposed or dictator social welfare function. □

2.4 Kalai's work on Arrow's Theorem

We have seen in the Section 2.3 that although the voters provide rational votes, the social choice function can derive to a non-rational outcome. This anomaly has an own name in honor to the man who presented the system of voting.

Definition 2.39. Given rational votes by individuals, we say that an alternative is a *Condorcet winner* if it wins all the pair-wise voting in which it participates.

Hence it is possible to define Arrow's Theorem 2.35 in another words:

Theorem 2.40 (Arrow's Theorem, v2)**.** *Consider a Condorcet election system for 3 candidates using an unanimous voting rule $f : \{-1,1\}^n \longrightarrow \mathbb{R}$. If there is always a Condorcet winner, then f must be a dictatorship.*

Gil Kalai proved the theorem in mathematical terms (see [3]) instead through results derived from ordered sets like Arrow. He started searching if there was a closed formula to compute the probability of a Condorcet winner.

To find this probability he proposed an easy mathematical way to identify a Condorcet winner. Given an input from $\{-1,1\}^3$, that will provide the preferences of a voter on 3 alternatives, we can identify if this vote is rational by assuring that all the bits are not equal. This can be seen in Table 1, because if the input was for example all 1's, the voter would prefer a to b, b to c and c to a, and this result leads to a contradiction. So first of all it is defined a function to identify non-rational outcomes.

Lemma 2.41. *Let $NAE_3 : \{-1,1\}^n \longrightarrow \{0,1\}$ be the following indicator function:*

$$NAE_3 \colon \{-1,1\}^n \to \{0,1\}$$
$$\mathbf{x} \mapsto \begin{cases} 0 & \text{if } \mathbf{x} = (-1,-1,-1) \text{ or } \mathbf{x} = (1,1,1), \\ +1 & \text{otherwise.} \end{cases}$$

Then the Fourier expansion of this function NAE_3 is:

$$NAE_3(\mathbf{x}) = \frac{3}{4} - \frac{1}{4}\Big(\chi_{\{1,2\}}(\mathbf{x}) + \chi_{\{1,3\}}(\mathbf{x}) + \chi_{\{2,3\}}(\mathbf{x})\Big).$$

This lemma will not be proved since it is easy to do as an exercise applying the method to obtain the Fourier coefficients presented in 1.10. But this is a powerful result that helped Kalai to manage to obtain a closed formula for the probability of having a Condorcet winner.

Lemma 2.42. *Consider a 3-candidate Condorcet election with pair-wise confrontations a vs. b, b vs. c and c vs. a. Suppose that every election is given by the outcome of the functions* $f, g, h : \{-1,1\}^n \longrightarrow \{-1,1\}$*, respectively. Then, under the impartial culture assumption, the probability of a Condorcet winner is:*

$$\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}}[NAE_3(f(\mathbf{x}), g(\mathbf{y}), h(\mathbf{z}))] = \frac{3}{4} - \frac{1}{4}\Big(\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y}}[f(\mathbf{x})g(\mathbf{y})] + \mathop{\mathbf{E}}_{\mathbf{y},\mathbf{z}}[g(\mathbf{y})h(\mathbf{z})] + \mathop{\mathbf{E}}_{\mathbf{x},\mathbf{z}}[f(\mathbf{x})h(\mathbf{z})]\Big).$$

Proof. Every individual participating in the Condorcet election will give votes x_i, y_i, z_i for the preferences on the confrontations a vs. b, b vs. c and c vs. a, respectively. Since the preferences must be rational, $NAE_3(x_i, y_i, z_i) = 1$. So society's election for every pair-wise confrontation will be $f(\mathbf{x}), g(\mathbf{y}), h(\mathbf{z})$ where $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{y} = (y_1, \dots, y_n)$ and $\mathbf{z} = (z_1, \dots, z_n)$ where n is the number of voters. Then we can compute the probability of a Condorcet winner identifying them with the NAE_3 function:

$$Pr[\exists \text{ Condorcet winner}] = \mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}}[NAE_3(f(\mathbf{x}), g(\mathbf{y}), h(\mathbf{z})].$$

Which is equal, by using Lemma 2.41, to:

$$\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}}[\frac{3}{4} - \frac{1}{4}\Big(\chi_{\{1,2\}}(f(\mathbf{x}), g(\mathbf{y}), h(\mathbf{z})) + \chi_{\{1,3\}}(f(\mathbf{x}), g(\mathbf{y}), h(\mathbf{z})) + \chi_{\{2,3\}}(f(\mathbf{x}), g(\mathbf{y}), h(\mathbf{z}))\Big)].$$

Developing the expression above we set:

$$\frac{3}{4} - \frac{1}{4}\Big(\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y}}[f(\mathbf{x})g(\mathbf{y})] + \mathop{\mathbf{E}}_{\mathbf{y},\mathbf{z}}[g(\mathbf{y})h(\mathbf{z})] + \mathop{\mathbf{E}}_{\mathbf{x},\mathbf{z}}[f(\mathbf{x})h(\mathbf{z})]\Big).$$

□

This result is really general since it works for different functions for every pair-wise election. But usually in a Condorcet-election system, also in the hypothesis of Arrow's Theorem, every pair-wise confrontation winner is given by the same social choice function. If we try to compute the last probability in this condition, we can relate it with the noise stability defined in 2.28.

Lemma 2.43. *Consider a 3-candidate Condorcet election using* $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ *as the social choice function. Then, under the impartial culture assumption, the probability of a Condorcet winner is:*

$$Pr[\exists \text{ Condorcet winner}] = \frac{3}{4} - \frac{3}{4}\mathbf{Stab}_{-1/3}[f].$$

Proof. In the joint distribution of $(\mathbf{x}, \mathbf{y})$ the n bit pairs (x_i, y_i) are independent since impartial culture assumption is assumed. The same observation is given for $(\mathbf{y}, \mathbf{z})$ and $(\mathbf{x}, \mathbf{z})$ but we will prove it without loss of generalization for $(\mathbf{x}, \mathbf{y})$.

First, we can compute by inspection the following expectations reminding the only two values the bits can take:

$$\mathbf{E}[x_i] = \mathbf{E}[y_i] = \mathbf{Pr}[x_i = +1] - \mathbf{Pr}[x_i = -1] = \frac{1}{2} - \frac{1}{2} = 0. \tag{2.9}$$

Second, we can compute the probability of $\mathbf{E}[x_i y_i]$ bearing in mind that $NAE_3(x_i, y_i, z_i) = 1$:

$$\mathbf{E}[x_i y_i] = \mathbf{Pr}[x_i y_i = +1] - \mathbf{Pr}[x_i y_i = -1] = \frac{2}{6} - \frac{4}{6} = -\frac{1}{3}. \tag{2.10}$$

Hence, we observed in Definition 2.27 that the results of the expectations given in (2.9)-(2.10) are equivalent to saying that $(\mathbf{x},\mathbf{y})$ is a $-1/3$-correlated pair. Then, by the definition of noise stability given in 2.28, we can rewrite the terms: $\mathbf{E}[f(\mathbf{x})g(\mathbf{y})] = \mathbf{Stab}_{-1/3}[f]$.

Finally we can substitute the expression for the probability of a Condorcet winner found in the Lemma 2.42 by this last result:

$$\begin{aligned}
\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}}[NAE_3(f(\mathbf{x}),f(\mathbf{y}),f(\mathbf{z}))] &= \frac{3}{4} - \frac{1}{4}\Big(\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y}}[f(\mathbf{x})f(\mathbf{y})] + \mathop{\mathbf{E}}_{\mathbf{y},\mathbf{z}}[f(\mathbf{y})f(\mathbf{z})] + \mathop{\mathbf{E}}_{\mathbf{x},\mathbf{z}}[f(\mathbf{x})f(\mathbf{z})]\Big) \\
&= \frac{3}{4} - \frac{1}{4}\Big(\mathbf{Stab}_{-1/3}[f] + \mathbf{Stab}_{-1/3}[f] + \mathbf{Stab}_{-1/3}[f]\Big) \\
&= \frac{3}{4} - \frac{3}{4}\mathbf{Stab}_{-1/3}[f].
\end{aligned}$$

□

Furthermore, we need a result from Fourier coefficients before we can prove the main theorem. As it was said in the part of Fourier expansions 1.2, the Fourier coefficients give a lot of information about a Boolean function.

Lemma 2.44. *Let $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ with the Fourier expansion $f = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S$. Suppose that $\mathbf{W}^1[f] = 1$. Then $f(\mathbf{x}) = \pm x_i$ for some $i \in [n]$.*

Proof. First, note that Parseval's Theorem assures $\langle f, f\rangle = 1$. Then if $\mathbf{W}^1[f] = 1$, all other degree k parts of f $\mathbf{W}^k[f]$ with $k \neq 1$ must be 0, thus $\hat{f}(S) = 0$ for every set such that $|S| \neq 1$.

Second, since we are working with a boolean-valued function, $f(\mathbf{x}) = \pm 1$ for every input from the Hamming cube. So the boolean-valued function squared can be also defined as $f^2 = 1$. In terms of Fourier expansions:

$$1 = \Big(\sum_{i=1}^{n} \hat{f}(i)x_i\Big)^2 = \sum_{i=1}^{n} \hat{f}(i)^2 + \sum_{i,j=1}^{n} \hat{f}(i)\hat{f}(j)x_ix_j = 1 + \sum_{i,j=1}^{n} \hat{f}(i)\hat{f}(j)x_ix_j.$$

Where in the second equality we introduced the hypothesis $\mathbf{W}^1[f] = 1$. So all rests to prove that the second term in the sum is equal to zero. Note that:

$$\sum_{i,j=1}^{n} \hat{f}(i)\hat{f}(j)x_ix_j = \sum_{i,j=1}^{n} \hat{f}(i)\hat{f}(j)\chi_{\{i,j\}}.$$

But for every set S such that $|S| \neq 1$ its Fourier coefficient must be zero. Therefore, by the uniqueness of Fourier expansion, $\hat{f}(\{i,j\}) = \hat{f}(i)\hat{f}(j) = 0$. These Fourier coefficients will be null if the product $\hat{f}(i)\hat{f}(j) = 0$ for $i \neq j$. However, the equality $\mathbf{W}^1[f] = 1$ restricts that there is only one $i \in [n]$ such that $\hat{f}(i) \neq 0$ and indeed it must be ± 1.

Consequently, writing the function by its Fourier expansion, $f(\mathbf{x}) = \pm x_i$. □

Finally, the proof of Arrow's Theorem 2.40 is now deduced from these previous lemmas.

Proof. (Arrow's Theorem, v2) By assumption, the probability of a Condorcet winner is 1 and writing it as the expression given in Lemma 2.43:

$$1 = \frac{3}{4} - \frac{3}{4}\mathbf{Stab}_{-1/3} = \frac{3}{4} - \frac{3}{4}\sum_{k=0}^{n}\Big(-\frac{1}{3}\Big)^k \mathbf{W}^k[f].$$

Where in the second equality the noise stability is expressed in terms of its Fourier coefficients as it is showed in Proposition 2.32. Since $(-1/3)^k \geq -1/3$ for all k, the equality above only occurs when all Fourier weights are on degree 1; i.e. $\mathbf{W}^1[f] = 1$. By the Lemma 2.44 this implies that $f = \pm x_i$. In other words f is a dictator or a negated dictator. Since f is unanimous, it must in fact be a dictator. $\square$

The expression found by Kalai can also be used to bound the probability of a Condorcet winner.

Proposition 2.45. *In a 3-candidate Condorcet election using* $f : \{-1,1\}^n \longrightarrow \{-1,1\}$, *the probability of a Condorcet winner is at most* $7/9 + 2/9 \cdot \mathbf{W}^1[f]$.

Proof. The probability given in Lemma 2.43 is:

$$\begin{aligned}
\frac{3}{4} - \frac{3}{4}\mathbf{Stab}_{-1/3}[f] &= \frac{3}{4} - \frac{3}{4}\sum_{k=0}^{n}\Big(-\frac{1}{3}\Big)^k \mathbf{W}^k[f] \\
&\leq \frac{3}{4} + \frac{1}{4}\mathbf{W}^1[f] + \frac{1}{36}\mathbf{W}^3[f] + \frac{1}{324}\mathbf{W}^5[f] + \cdots \\
&\leq \frac{3}{4} + \frac{1}{4}\mathbf{W}^1[f] + \frac{1}{36}(\mathbf{W}^3[f] + \mathbf{W}^5[f] + \cdots) \\
&\leq \frac{3}{4} + \frac{1}{4}\mathbf{W}^1[f] + \frac{1}{36}(1 - \mathbf{W}^1[f]) \\
&= \frac{7}{9} + \frac{2}{9}\mathbf{W}^1[f].
\end{aligned}$$

$\square$

3. Hypercontractivity and stability

This chapter will introduce the concept of hypercontractivity in Boolean functions. We will follow the results given in chapters 4 and 9 of O'Donnell's reference book [1]. The capability of giving a stability result for Arrow's Theorem 2.40 and prove the KKL Theorem 3.14 depends on some new concepts about hypercontractivity on Boolean functions. When we talk about stability of some statement we refer that being near the hypothesis implies something near the result.

3.1 Hypercontractivity results

In 1979 Aline Bonami introduced and proved in [5] a surprising theorem which states that the noise operator has really good properties applied to the functions that we are working on.

Theorem 3.1 ((p, q)-Hypercontractivity Theorem)**.** *Let $f : \{-1, 1\}^n \longrightarrow \mathbb{R}$ and let $1 \leq p \leq q \leq \infty$. Then $\|T_\rho f\|_q \leq \|f\|_p$ for $0 \leq \rho \leq \sqrt{\frac{p-1}{q-1}}$.*

This result is stronger than Hölder's inequality because somehow we have flipped p and q. We will prove a less general lemma called Bonami Lemma 3.5. But first of all, we define an aspect of Boolean functions that will favor them on having good properties.

Definition 3.2. The *degree* of a Boolean function $f : \{-1, 1\}^n \longrightarrow \mathbb{R}$ is the degree of its Fourier expansion.

Random variables can have sometimes unreasonable behaviour. This is a really general concept but one example is that the random variable is never close to its expectation. Bonami Lemma assures that low degree functions have reasonable behaviour. Let us define now a mathematical condition which secures good behaviour.

Definition 3.3. For a real number $B \geq 1$, we say that the real random variable $\mathbf{X}$ is *B-reasonable* if $\mathbf{E}[\mathbf{X}^4] \leq B\mathbf{E}[\mathbf{X}^2]^2$.

Doing some work this condition gives good properties to the random variable. For example, we can obtain better tail bounds than what we would get out of the Chebychev inequality. For discrete random variables a simple condition to be reasonable is to take on each of its values with nonnegligible probability. The property that we will see shows that we can have an upper bound for the probability of these random variables being close to zero.

Proposition 3.4. *Let $\mathbf{X} \neq 0$ be B-reasonable. Then $\mathbf{Pr}[|\mathbf{X}| > t\|\mathbf{X}\|_2] \geq \frac{(1-t^2)^2}{B}$ for all $t \in [0, 1]$.*

Proof.

$$\mathbf{Pr}[|\mathbf{X}| > t\|\mathbf{X}\|_2] = \mathbf{Pr}[\mathbf{X}^2 > t^2\|\mathbf{X}\|_2^2] = \mathbf{Pr}[\mathbf{X}^2 > t^2\mathbf{E}[\mathbf{X}^2]] \geq (1-t^2)^2\frac{\mathbf{E}[\mathbf{X}^2]^2}{\mathbf{E}[\mathbf{X}^4]}$$
$$\geq (1-t^2)^2\frac{\mathbf{E}[\mathbf{X}^2]^2}{B\mathbf{E}[\mathbf{X}^2]^2} = \frac{(1-t^2)^2}{B}.$$

Where the first inequality comes from the Paley-Zygmund inequality. □

Finally, it is introduced the Bonami Lemma. It will confirm the suspicion that a high degree function is required to have non-reasonable behaviour.

Lemma 3.5 (Bonami Lemma). *For each k, if $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ has degree at most k and the coordinates are uniformly random bits ± 1 i.i.d. then the random variable $f(\mathbf{x})$ is 9^k-reasonable, i.e.:*

$$\mathbf{E}[f^4] \leq 9^k \mathbf{E}[f^2]^2, \textit{ if and only if, } \|f\|_4 \leq \sqrt{3^k}\|f\|_2.$$

Proof. This lemma is proved by induction on n. We use $k \geq 1$, as otherwise f is constant and the claim is trivial. Again, if $n = 0$, then f is also constant. For $n \geq 1$, we will start decomposing f in terms of $D_n f$ and $E_n f$ as it was shown in Proposition 2.23. For brevity we write $f = f(\mathbf{x})$, $d = D_n f(\mathbf{x})$ and $e = E_n f(\mathbf{x})$. Now:

$$\begin{aligned} \mathbf{E}[f^4] &= \mathbf{E}[(x_n d + e)^4] = \mathbf{E}[x_n^4 d^4] + 4\mathbf{E}[x_n^3 d^3 e] + 6\mathbf{E}[x_n^2 d^2 e^2] + 4\mathbf{E}[x_n d e^3] + \mathbf{E}[e^4] \\ &= \mathbf{E}[x_n^4] \cdot \mathbf{E}[d^4] + 4\mathbf{E}[x_n^3] \cdot \mathbf{E}[d^3 e] + 6\mathbf{E}[x_n^2] \cdot \mathbf{E}[d^2 e^2] + 4\mathbf{E}[x_n] \cdot \mathbf{E}[d e^3] + \mathbf{E}[e^4]. \end{aligned} \tag{3.1}$$

The last equality is due to the independence of x_n from d and e since they do not depend on x_n.

We compute the expectations $\mathbf{E}[x_n] = \mathbf{E}[x_n^3] = 0$ and $\mathbf{E}[x_n^2] = \mathbf{E}[x_n^4] = 1$. Then we can simplify the Expression (3.1) like:

$$\mathbf{E}[f^4] = \mathbf{E}[d^4] + \mathbf{E}[d^2 e^2] + \mathbf{E}[e^4]. \tag{3.2}$$

Doing the same procedure but now for f^2:

$$\begin{aligned} \mathbf{E}[f^2] &= \mathbf{E}[(x_n d + e)^2] = \mathbf{E}[x_n^2 d^2] + 2\mathbf{E}[x_n d e] + \mathbf{E}[e^2] \\ &= \mathbf{E}[x_n]^2 \cdot \mathbf{E}[d^2] + 2\mathbf{E}[x_n] \cdot \mathbf{E}[ed] + \mathbf{E}[e^2] \\ &= \mathbf{E}[d^2] + \mathbf{E}[e^2]. \end{aligned} \tag{3.3}$$

To get an upper-bound on (3.2) recall that we have seen in Proposition 2.19 that $D_i f$ has degree $\leq k-1$ and depends on $n-1$ variables so we can apply the induction hypothesis to deduce $\mathbf{E}[d^4] \leq 9^{k-1}\mathbf{E}[d^2]^2$. Similarly, Proposition 2.23 shows that $E_n f$ is a function of degree $\leq k$ and depends on $n-1$ variables so $\mathbf{E}[e^4] \leq 9^k \mathbf{E}[e^2]^2$. To bound $\mathbf{E}[e^2 d^2]$ we apply the Cauchy-Schwarz inequality:

$$\mathbf{E}[e^2 d^2] = \langle e^2, d^2 \rangle \leq \sqrt{\mathbf{E}[d^4]}\sqrt{\mathbf{E}[e^4]}.$$

. We can apply the induction hypothesis for the terms inside the square roots. Thus we have:

$$\begin{aligned} \mathbf{E}[f^4] &\leq 9^{k-1}\mathbf{E}[d^2]^2 + 6\sqrt{9^{k-1}\mathbf{E}[d^2]^2}\sqrt{9^k \mathbf{E}[e^2]^2} + 9^k \mathbf{E}[e^2]^2 \\ &\leq 9^k(\mathbf{E}[d^2]^2 + 2\mathbf{E}[d^2] \cdot \mathbf{E}[e^2] + \mathbf{E}[e^2]^2) \\ &= 9^k(\mathbf{E}[d^2] + \mathbf{E}[e^2])^2 = 9^k \mathbf{E}[f^2]^2. \end{aligned}$$

Where the second inequality is valid since $k \geq 1$. In the last equality we used the result from (3.3). $\square$

Many important results that we will see later use the Hypercontractivity Theorem 3.1 on some fixed pairs (p, q). This concrete results can be proved from the Bonami Lemma.

3.2 Stability of Arrow's Theorem

Once Arrow's Theorem 2.40 is understood, one natural question is if the result is stable. By stable we mean that being really close to the hypothesis almost implies being close to the result of the statement.

We can be curious if a Condorcet winner probability near one also states that the function is "close" to be a dictatorship. Later we will see that it is just like that. But before we define in mathematical terms how can we measure the closeness of a function to being a certain function.

Definition 3.6. If f and g are boolean-valued functions we say that they are ϵ*-close* if $dist(f, g) \leq \epsilon$, where $dist(\cdot, \cdot)$ is the Hamming distance from Definition 1.13.

Therefore it is possible to introduce the *Friedgut-Kalai-Naor Theorem* presented in [6].

Theorem 3.7 (FKN Theorem)**.** *Suppose $\delta > 0$ and $f : \{-1, 1\}^n \longrightarrow \{-1, 1\}$ has $\mathbf{W}^1[f] \geq 1 - \delta$. Then f is $O(\delta)$-close to $\pm\chi_i$ for some $i \in [n]$.*

Note that this theorem states that if almost all the weights $\mathbf{W}[f]$ are in the first level there is one weight in $\mathbf{W}^1[f]$ which is big enough to be decisive on valuing the function. Before we can show this theorem we need a lemma provided from Bonami Lemma.

Lemma 3.8. *Let $f : \{-1, 1\}^n \longrightarrow \mathbb{R}$ be a nonconstant function of degree at most k; write $\mu = \mathbf{E}[f]$ and $\sigma = \sqrt{\mathbf{Var}[f]}$. Then:*

$$\mathop{\mathbf{Pr}}_{\mathbf{x}\sim\{-1,1\}^n}[|f(\mathbf{x}) - \mu| > \frac{1}{2}\sigma] \geq \frac{1}{16} \cdot 9^{1-k}.$$

Proof. Let $g = 1/\sigma(f - \mu)$. a function of degree at most k standardized so $\|g\|_2 = 1$. By the Bonami Lemma 3.5, g is 9^k-reasonable. The result now follows by applying the Proposition 3.4 for reasonable variables with $t = 1/2$:

$$\mathbf{Pr}[|g| > t\|g\|_2] = \mathbf{Pr}[|f - \mu| > \frac{1}{2}\sigma] \geq \frac{(1 - (1/2)^2)^2}{9^k} = \frac{1}{16} \cdot 9^{1-k}.$$

□

Finally it is possible to prove the FKN theorem.

Proof. (FKN Theorem). We write $\ell = f^{=1}$, so $\mathbf{E}[\ell^2] = 1 - \delta$ by assumption. We can assume without loss of generality that $\delta \leq \frac{1}{1600}$. The goal is to show that $\mathbf{Var}[\ell^2]$ is small, in particular we will show $\mathbf{Var}[\ell^2] \leq 6400\delta$.

To bound $\mathbf{Var}[\ell^2]$ we first apply the Lemma 3.8 to the degree two function ℓ^2:

$$\mathbf{Pr}\left[|\ell^2 - (1 - \delta)| \geq \frac{1}{2}\sqrt{\mathbf{Var}[\ell^2]}\right] \geq \frac{1}{16} \cdot 9^{1-2} = \frac{1}{144}.$$

Now suppose by way of contradiction that $\mathbf{Var}[\ell^2] > 6400\delta$, then the inequality above implies:

$$\frac{1}{144} \leq \mathbf{Pr}[|\ell^2 - (1 - \delta)| > 40\sqrt{\delta}] \leq \mathbf{Pr}[|\ell^2 - (1 - \delta)| > 39\sqrt{\delta}]. \tag{3.4}$$

This is saying that $|\ell|$ has some probability of being far from 1. Since $|f| = 1$ always we can deduce that $|f-\ell|^2$ is frequently large. In fact, a precise calculation can be done and see that $(f-\ell)^2 \geq 169\delta$ if $|\ell^2-1| > 39\delta$. But using Markov's inequality on (3.4) derives to $\mathbf{E}[(f-\ell)^2] \geq \frac{1}{144}169\delta > \delta$, a contradiction since it is computable and $\mathbf{E}[(f-\ell)^2] = 1 - \mathbf{W}^1[f] = \delta$ by assumption.

Therefore we prove that $\mathbf{Var}[\ell^2]$ small gives us the desired result. By the expression of variance in terms of the Fourier coefficients seen in Proposition 1.18 we can write:

$$\frac{1}{2}\mathbf{Var}[\ell^2] = \sum_{i\neq j}\hat{f}(i)^2\hat{f}(j)^2 = \left(\sum_{i=1}^{n}\hat{f}(i)^2\right)^2 - \sum_{i=1}^{n}\hat{f}(i)^4 = (1-\delta)^2 - \sum_{i=1}^{n}\hat{f}(i)^4 \geq (1-2\delta) - \sum_{i=1}^{n}\hat{f}(i)^4.$$

And hence using the small variance proved above, $\mathbf{Var}[\ell^2] \leq 6400\delta$, we have:

$$1 - 3202\delta \leq \sum_{i=1}^{n}\hat{f}(i)^4 \leq \max_{i\in[n]}\{\hat{f}(i)^2\}\sum_{i=1}^{n}\hat{f}(i)^2 \leq \max_{i\in[n]}\{\hat{f}(i)^2\} \leq \max_{i\in[n]}\{|\hat{f}(i)|\}.$$

Hence there is a Fourier coefficient with absolute value greater than some threshold and the greatness depends on δ. This is the definition of being $O(\delta)$-close to some dictatorship. □

At the end, as it was suspected, we can introduce a corollary of Arrow's Theorem that states stability on the theorem. It is an immediate consequence of FKN Theorem 3.7.

Corollary 3.9. *Suppose that in a 3-candidate Condorcet election using $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ the probability of a Condorcet winner is $1-\epsilon$. Then f is $O(\epsilon)$-close to $\pm\chi_i$ for some $i \in [n]$.*

Proof. In Proposition 2.45 it is presented an upper-bound for the probability of a Condorcet winner. In the hypothesis the probability is $1-\epsilon$, hence:

$$1-\epsilon \leq \frac{7}{9} + \frac{2}{9}\mathbf{W}^1[f], \text{ then: } \mathbf{W}^1[f] \geq 1 - \frac{9}{2}\epsilon.$$

The conclusion follows immediately from the FKN Theorem. □

3.3 Kahn–Kalai–Linial Theorem

By the moment, boolean-valued functions have proved us that they have a really good behaviour and interpretations. In fact, we can represent them in terms of a decision tree or another more compact representation called DNF formula.

Definition 3.10. A *DNF (Disjunctive Normal Form)* formula over Boolean variables $x_1, \dots, x_n$ is a logical *OR* of terms, each of which is a logical *AND* of *literals*. A literal is either a variable x_i or its negation $\overline{x_i}$. The number of terms is called the *size* and the number of literals in every term is defined as the *width*.

We have seen before a function defined like this. Remember the definition of the Tribes function 2.6. $Tribes_{w,s}$ is a width-w and size-s DNF formula.

One can easily see that

$$\mathbf{Pr}[Tribes_{w,s}(\mathbf{x}) = -1] = 1 - (1-2^{-w})^s.$$

Hence Ben-Or and Linial tried in [7] to construct a nearly unbiased function with the following definition.

Definition 3.11. For $w \in \mathbb{N}^+$ let $s = s_w$ be the largest integer such that $1 - (1 - 2^{-w})^s \leq 1/2$. Then for $n = n_w = sw$ we define $Tribes_n : \{-1, 1\}^n \longrightarrow \{-1, 1\}$ as $Tribes_{w,s}$.

Hereinafter we will denote $log\,(x)$ as $log_2\,(x)$. If we do the computations we can see that $s \approx ln(2)2^w$, hence $n = sw \approx ln(2)w2^w$ and therefore $w \approx log\,(n) - log\,(ln\,(n))$ and $s \approx n/\,log\,(n)$. Before giving some results about this asymptotic terms, let us recall, for the sake of completeness, some expressions in *Big O notation*.

- $f(n) = o(g(n))$ if for every $k > 0$ there exists a n_0 such that $|f(n)| < kg(n)$, $\forall n \geq n_0$.
- $f(n) = O(g(n))$ if there exists a $k > 0$ and n_0 such that $|f(n)| \leq kg(n)$, $\forall n \geq n_0$.
- $f(n) = \Omega(g(n))$ if there exists a $k > 0$ and n_0 such that $f(n) \geq kg(n)$, $n \geq n_0$.
- $f(n) = \tilde{\Omega}(g(n))$ if there exists a $k \neq 0$ such that $f(n) = \Omega\big(g(n)\,log^k\,(g(n))\big)$.
- $f(n) = \Theta(g(n))$ if there exists a $k_1 > 0$, $k_2 > 0$ and n_0 such that $k_1 g(n) \leq f(n) \leq k_2 g(n)$, $\forall n \geq n_0$.

Doing the computations it is possible to see the following asymptotic results for the parameters in the $Tribes_n$ function.

Proposition 3.12. *For the $Tribes_n$ function given in Definition 3.11:*

- $s = ln\,(2)2^w - \Theta(1)$.
- $n = ln\,(2)w2^w - \Theta(w)$.
- $w = log\,(n) - log\,(ln\,(n)) + o(1)$.
- $\mathbf{Pr}[Tribes_n(\mathbf{x}) = -1] = \frac{1}{2} - O(log\,(n)/n)$.

Note that the last result is stating that with this setting of parameters $Tribes_n$ is essentially unbiased. We can try to see if this gives a good (small) influence.

Proposition 3.13.

$$\mathbf{Inf}_i[Tribes_n] = \frac{ln\,(n)}{n}\Big(1 \pm o(1)\Big), \ \textit{for each } i \in [n].$$

And hence, $\mathbf{I}[Tribes_n] = ln\,(n)(1 \pm o(1))$.

Proof. In $Tribes_n$ voter i is pivotal if and only if all the voter's in i's "tribe" vote -1 (True) and all the other tribes produce the outcome $+1$ (False). The probability of this is:

$$\begin{aligned} 2^{-(w-1)} \cdot (1 - 2^{-w})^{s-1} &= \frac{2^{-(w-1)}}{1 - 2^{-w}}(1 - 2^{-w})^s = \frac{2^{1-w}}{1 - 2^{-w}}\mathbf{Pr}[Tribes_n = +1] \\ &= \frac{2}{2^w - 1}\mathbf{Pr}[Tribes_n = +1] = \frac{2}{2^w - 1}\left(\frac{1}{2} + O\Big(\frac{log(n)}{n}\Big)\right) \\ &= \frac{ln\,(n)}{n}\Big(1 \pm o(1)\Big). \end{aligned}$$

Where we used the Proposition 3.12 in the two last equalities. □

Thus, if we are interested in unbiased voting rules in which every voter has small influence, $Tribes_n$ is much better than Maj_n where each voter has influence $\Theta(1/\sqrt{n})$. We can wonder if there exists any other unbiased voting rule such that the maximum influence is smaller than $\Theta(ln(n))/n$. The Kahn-Kalai-Linial Theorem was presented in [8] and shows that the $Tribes_n$ example is tight to constants.

Theorem 3.14 (KKL Theorem). *For any* $f : \{-1,1\}^n \longrightarrow \{-1,1\}$*:*

$$\mathbf{MaxInf}[f] = \max_{i\in[n]} \mathbf{Inf}_i[f] \geq \mathbf{Var}[f]\Omega\Big(\frac{log\,(n)}{n}\Big).$$

We can interpret this theorem like: "The maximum influence is greater than something that depends on the function ($\mathbf{Var}[f]$) and a function that increases with $log\,(n)/n$". The proof of this theorem requires a bunch of lemmas and a derived theorem that will be presented next.

As it was previously said in Section 3.1 the Bonami Lemma 3.5 is a really useful tool to prove the Hypercontractivity Theorem 3.1 for specific pairs (p,q).

Observation 3.15. An immediate consequence of the Bonami Lemma is that for any $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ and $k \in \mathbb{N}$,

$$\|T_{1/\sqrt{3}}f^{=k}\|_4 = \|\Big(\frac{1}{\sqrt{3}}\Big)^k f^{=k}\|_4 = \frac{1}{\sqrt{3^k}}\|f^{=k}\|_4 \leq \frac{\sqrt{3^k}}{\sqrt{3^k}}\|f^{=k}\|_2 = \|f^{=k}\|_2.$$

But there is a generalization of this (2,4)-Hypercontractivity Theorem which says that the assumption of having degree-k f is not necessary.

Lemma 3.16 ((2,4)-Hypercontractivity Theorem). *Let* $f : \{-1,1\}^n \longrightarrow \mathbb{R}$*. Then* $\|T_{1/\sqrt{3}}f\|_4 \leq \|f\|_2$.

If we look close, this is saying that $T_{1/\sqrt{3}}$ is a contraction when is viewed as an operator from $L^2(\{-1,1\}^n)$ to $L^4(\{-1,1\}^n)$. Hence we should think of hypercontractivity like quantifying the extent to which the noise operator T_ρ is a smoothing operator.

Proof. We will prove the identical result $\mathbf{E}[T_{1/\sqrt{3}}f(x)^4] \leq \mathbf{E}[f(x)^2]^2$ using the same induction as in the proof of the Bonami Lemma. Retaining the notation e and d, and now denoting $T = T_{1/\sqrt{3}}$, since the noise operator is lineal:

$$Tf = T(x_n \cdot d + e) = x_n \cdot \frac{1}{\sqrt{3}} Td + Te.$$

Similar computations as with the Bonami lemma, like the independence of x_n on the functions d and e and the expectations on $\mathbf{E}[x_i^k]$ for k in $\{1,2,3,4\}$, yield:

$$\begin{aligned}
\mathbf{E}[(Tf)^4] &= \Big(\frac{1}{\sqrt{3}}\Big)^4 \mathbf{E}[(Td)^4] + 6\Big(\frac{1}{\sqrt{3}}\Big)^2 \mathbf{E}[(Td)^2(Te)^2] + \mathbf{E}[(Te)^4] \\
&\leq \mathbf{E}[(Td)^4] + 2\mathbf{E}[(Td)^2(Te)^2] + \mathbf{E}[(Te)^4] \\
&\leq \mathbf{E}[(Td)^4] + 2\sqrt{\mathbf{E}[(Td)^4]}\sqrt{\mathbf{E}[(Te)^4]} + \mathbf{E}[(Te)^4] \\
&\leq \mathbf{E}[d^2]^2 + 2\mathbf{E}[d^2]\mathbf{E}[e^2] + \mathbf{E}[e^2]^2 = (\mathbf{E}[d^2] + \mathbf{E}[e^2])^2 \\
&= \mathbf{E}[f^2]^2.
\end{aligned}$$

Where in the second inequality we used the Cauchy-Schwarz inequality and the following inequality is the induction step because the number of variables from which d and e depend are less than n. Finally, in the last equality, we write $\mathbf{E}[f^2]$ like it was computed in (3.3). □

This precise hypercontractivity theorem allows to introduce and prove another hypercontractivity theorem for a concrete pair (p, q):

Lemma 3.17 ((4/3,2)-Hypercontractivity Theorem). *Let $f : \{-1,1\}^n \longrightarrow \mathbb{R}$. Then $\|T_{1/\sqrt{3}}f\|_2 \leq \|f\|_{4/3}$. Observing that $\|T_{1/\sqrt{3}}f\|_2 = \sqrt{\mathbf{Stab}_{1/3}[f]}$ we can rewrite this as:*

$$\mathbf{Stab}_{1/3}[f] \leq \|f\|_{4/3}^2. \tag{3.5}$$

Proof.

$$\begin{aligned}\|T_{1/\sqrt{3}}f\|_2^2 &= \langle T_{1/\sqrt{3}}f, T_{1/\sqrt{3}}f\rangle = \langle f, T_{1/\sqrt{3}}T_{1/\sqrt{3}}f\rangle \\ &\leq \|f\|_{4/3}\|T_{1/\sqrt{3}}T_{1/\sqrt{3}}f\|_4 \leq \|f\|_{4/3}\|T_{1/\sqrt{3}}f\|_2.\end{aligned}$$

Where the first inequality is due to Hölder's inequality with $p = 4/3$ and $q = 4$. In the second inequality we applied the (2,4)-Hypercontractivity Theorem 3.16. Dividing now by $\|T_{1/\sqrt{3}}f\|_2$ we have:

$$\|T_{1/\sqrt{3}}f\|_2 \leq \|f\|_{4/3}.$$

□

The left-hand side of the Equation (3.5) is a natural quantity. The right-hand side is always 1 for boolean-valued functions $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ because of the definition of the norm $\|\cdot\|_{4/3}$. It is interesting to treat other Boolean functions g such that $g = |g| = g^2$ for example $g : \{-1,1\}^n \longrightarrow \{0,1\}$ or $g : \{-1,1\}^n \longrightarrow \{-1,0,1\}$.

Corollary 3.18. *Let $f : \{-1,1\}^n \longrightarrow \{-1,1\}$. Then $\mathbf{Inf}_i^{(1/3)}[f] \leq \mathbf{Inf}_i[f]^{3/2}$ for all i in $[n]$.*

Proof. By definition $\mathbf{Stab}_{1/3}[D_if] = \mathbf{Inf}_i^{(1/3)}$ and $\|D_if\|_2^2 = \mathbf{Inf}_i[f] = \mathbf{E}[D_if^2]$. So:

$$\mathbf{Stab}_{1/3}[D_if] \leq \|D_if\|_{4/3}^2 = \left(\mathbf{E}[D_if(\mathbf{x})^{4/3}]^{3/4}\right)^2 = \mathbf{E}[D_if^2]^{3/2}.$$

The first inequality provides from (3.5). The last equality happens only because D_if's range $\{-1,0,1\}$, so $D_if^{4/3} = D_if^2$. □

Hence if the influence of i is small, then its 1/3-stable influence is much smaller.

Before we can show the KKL Theorem 3.14 we will see a variation of the theorem. This result was also presented by Kahn, Kalai and Lineal in [8].

Theorem 3.19 (KKL Edge-Isoperimetric Theorem). *Let $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ be non-constant and let $\tilde{\mathbf{I}}[f] = \mathbf{I}[f]/\mathbf{Var}[f] \geq 1$ (which will be just 1 if f is unbiased). Then*

$$\mathbf{MaxInf}[f] \geq \left(\frac{9}{\tilde{\mathbf{I}}[f]^2}\right) \cdot 9^{-\tilde{\mathbf{I}}[f]}.$$

This statement is different from the statement of KKL Theorem 3.14. Here we can interpret that the maximal influence will be greater than a expression that depends on the total influence and the variance. Remember that the variance measures how far is a Boolean function from being constant. Additionally, note that now maximal influence does not increase with a function depending on the size of the Hamming cube n.

The idea of the proof is to look at the contrapositive: supposing that all of f's influences are small, we want to show that its total influence must be large.

Proof. (KKL Edge-Isoperimetric Theorem). The proof treats only the case when f is unbiased, i.e. $\mathbf{Var}[f] = 1$ so $\tilde{\mathbf{I}}[f] = \mathbf{I}[f]$. Before we can prove the theorem, we will introduce 3 inequalities that will be used in the proof.

The first one is a consequence of the Fourier expansion for the noise stability:

$$3 \cdot \mathbf{Stab}_{1/3}[f] = 3 \sum_{S \subseteq [n]} \left(\frac{1}{3}\right)^{|S|} \hat{f}(S)^2 = 3 \cdot \mathop{\mathbf{E}}_{S \sim \mathcal{S}_f} [3^{-|S|}] \geq 3 \cdot 3^{-\mathbf{E}[|S|]} = 3 \cdot 3^{-\mathbf{I}[f]}. \tag{3.6}$$

Where the second equality comes from the definition of the written expectation. The inequality is derived from the convexity of the function $t \mapsto 3^{-t}$.

The second useful inequality is given by reasoning on Fourier formulas and the fact that $\hat{f}(\emptyset) = 0$ since f is unbiased:

$$\mathbf{I}^{(1/3)}[f] = \sum_{|S|>1} |S| \left(\frac{1}{3}\right)^{|S|-1} \hat{f}(S)^2 \geq 3 \cdot \sum_{|S| \geq 1} \left(\frac{1}{3}\right)^{|S|} \hat{f}(S)^2 = 3 \cdot \mathbf{Stab}_{1/3}[f]. \tag{3.7}$$

The third inequality is immediate since f is a boolean-valued function so $\mathbf{Inf}_i[f] = \sum_{i \in S} \hat{f}(S)^2 \leq 1$.

$$\sum_{i=1}^{n} \mathbf{Inf}_i[f]^{3/2} \leq \sum_{i=1}^{n} \mathbf{MaxInf}[f]^{1/2} \mathbf{Inf}_i[f] = \mathbf{MaxInf}[f]^{1/2} \mathbf{I}[f]. \tag{3.8}$$

Grouping all the inequalities and recalling the Corollary 3.18 we finally have:

$$3 \cdot 3^{-\mathbf{I}[f]} \underset{(3.6)}{\leq} 3 \cdot \mathbf{Stab}_{1/3}[f] \underset{(3.7)}{\leq} \mathbf{I}^{(1/3)}[f] \underset{3.18}{\leq} \sum_{i=1}^{n} [\mathbf{Inf}_i[f]^{3/2} \underset{(3.8)}{\leq} \mathbf{MaxInf}[f]^{1/2} \mathbf{I}[f].$$

So getting the power of two on both sides and reordering the terms we have the desired result. □

With the Edge-Isoperimetric version of the Kahn-Kalai-Linial Theorem we can prove immediately the original KKL Theorem 3.14:

Proof. (KKL Theorem). We may assume f is non-constant. We can have two cases for the previous defined influence $\tilde{\mathbf{I}}[f]$. On one hand, if $\tilde{\mathbf{I}}[f] = \mathbf{I}[f]/\mathbf{Var}[f] \geq 0.1\, log\,(n)$ then we are done, the total influence is at least $\mathbf{I}[f] \geq 0.1 \mathbf{Var}[f]\, log\,(n)$, and hence $\mathbf{MaxInf}[f] \geq 0.1 \mathbf{Var}[f]\, log\,(n)/n$ as the theorem states.

On the other hand, if $\tilde{\mathbf{I}}[f] \leq 0.1\, log\,(n)$, the KKL Edge-Isoperimetric Theorem 3.19 implies:

$$\begin{aligned} \mathbf{MaxInf}[f] &\geq \Omega\left(\frac{1}{(0.1)^2\, log^2\,(n)}\right) 9^{-0.1\, log\,(n)} \geq \Omega\left(\frac{1}{log^2\,(n)}\right) 9^{-0.1\, log\,(n)} \\ &= \Omega\left(\frac{1}{log^2\,(n)}\right) \cdot \left(2^{log\,(9)}\right)^{-0.1\, log\,(n)} \\ &= \Omega\left(\frac{1}{log^2\,(n)} \cdot n^{-0.1\, log\,(9)}\right) = \tilde{\Omega}\left(n^{-0.1\, log\,(9)}\right) = \Omega\left(n^{-0.317}\right). \end{aligned} \tag{3.9}$$

We only need to see that $\Omega(n^{-0.317})$ is greater than the right-side of the inequality of the KKL Theorem. Watch in 3.14 that there is a $\mathbf{Var}[f]$ multiplying and it can take really small values, therefore we should see that $\Omega(n^{-0.317})$ is much greater than $\Omega(log\,(n)/n)$. To prove this we will use limits to see that the first function takes much greater values than the second.

Let $f(n)$ be in $\Omega(n^{-0.317})$, by definition there exists a $k_1 > 0$ and n_1 such that $f(n) \geq k_1 n^{-0.317}$ for every $n \geq n_1$. We can also define a function $g(n)$ in $\Omega(log\,(n))/n$, so there exists a $k_2 > 0$ and n_2 such that $g(n) \geq k_2\, log\,(n)/n$ for every $n \geq n_2$. If now we take limits with n tending to infinity we must note that n will be greater than $max\,\{n_1, n_2\}$ so we can write the limit between $f(n)$ and $g(n)$ as it follows:

$$\begin{aligned}\lim_{n\to+\infty}\frac{f(n)}{g(n)} &\geq \lim_{n\to+\infty}\frac{k_1 n^{-0.317}}{k_2\frac{log\,(n)}{n}} = \frac{k_1}{k_2}\cdot\lim_{n\to+\infty}\frac{n^{0.683}}{\frac{ln\,(n)}{ln\,(2)}} = \frac{k_1\cdot ln\,(2)}{k_2}\cdot\lim_{n\to+\infty}\frac{0,683\cdot n^{-0.317}}{\frac{1}{n}}\\ &= \frac{k_1\cdot ln\,(2)\cdot 0,683}{k_2}\cdot\lim_{n\to+\infty} n^{0,683} = +\infty\end{aligned}$$

Where in the first equality we applied the rule of change of basis of a logarithm. In the second it is necessary to use Hôpital's rule since the limit is a division between infinities so we differentiate with respect to n in the numerator and the denominator. Finally we can plug this result of limits in the Equation (3.9) to show the final result:

$$\mathbf{MaxInf}[f] \geq \Omega(n^{-0.317}) >> \mathbf{Var}[f]\Omega\left(\frac{log\,(n)}{n}\right).$$

$\square$

4. Path to the proof of the Sensitivity Conjecture

This chapter starts presenting what is the Computational Complexity Theory. It will locate the reader in a position to understand some of the most important complexities defined until now and how they are related. The aim of the chapter is to present the relation between two specific complexities. It was first conjectured by Nisan and Szegedy in [9] but proved in [10] by Huang almost 30 years later with an equivalence to the problem presented by Gotsman and Kalai in [11].

4.1 Foundations of Computational Complexity Theory

The Computational Complexity Theory is the branch of Computer Science that defines sensitivity measures and studies how they are related in order to classify computational problems. One type of relation is in polynomial ways.

Definition 4.1. Given the measures of complexity s_1 and s_2 they are said to be *equivalent* if they are polynomially related; i.e. there exist polynomials $p_1(x)$ and $p_2(x)$ such that for every f:

$$s_1(f) \leq p_2(s_2(f)), \; s_2(f) \leq p_1(s_1(f)).$$

Once it is said how measures of complexity are related it is possible to introduce and define two basic measures.

Definition 4.2. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. A *decision tree model* is an algorithm which repeatedly queries input variables until it can determine the value of the function. The cost of the model is the number of queries. The *decision tree complexity*, $D(f)$, is defined to be the cost of the best tree model algorithm for f.

Definition 4.3. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function and $\mathbf{x}$ any input string. A 1-*certificate* for f is an assignment to some subset of variables that forces the value of f to be 1. Same definition can be given but for value 0. The *certificate complexity of f on* $\mathbf{x}$, $C_{\mathbf{x}}(f)$, is the size of the smallest certificate that agrees with $\mathbf{x}$. The *certificate complexity of f*, $C(f)$, is defined like $max_{\mathbf{x}\in\{0,1\}^n}\{C_{\mathbf{x}}(f)\}$.

At the beginnings of the development of Computational Complexity Theory it was proved that these two measures of complexity were related.

Theorem 4.4.

$$C(f) \leq D(f) \leq \big(C(f)\big)^2.$$

Since this two measures were equivalent Nisan tried to search for other measures that could be also in this equivalence group. Now we adapt the definition of sensitivity in 2.15 for actual terms. For $\mathbf{x} \in \{0,1\}^n$ and a subset $S \subseteq [n]$ we denote $\mathbf{x}^S$ the binary vector obtained from $\mathbf{x}$ by flipping all indices in S.

Definition 4.5. For $f : \{0,1\}^n \longrightarrow \{0,1\}$ the *local sensitivity on the input* $\mathbf{x}$, $s(f,\mathbf{x})$, is defined as the number of indices i in $[n]$ such that $f(\mathbf{x}) \neq f(\mathbf{x}^{\{i\}})$. The *sensitivity of f*, $s(f)$, is $max_{\mathbf{x}\in\{0,1\}^n}\, s(f,\mathbf{x})$.

Note that sensitivity measures local changing behaviour of a Boolean function with respect to the Hamming cube. So it can be viewed as a way to measure smoothness in discrete functions.

Nisan studied if sensitivity was also polynomially related to certificate complexity and decision tree complexity. However he could not achieve any result with sensitivity so in [12] he proceeded to search an equivalence with a more relaxed condition rather than sensitivity called block sensitivity.

Definition 4.6. For $f : \{0,1\}^n \longrightarrow \{0,1\}$ the *local block sensitivity on the input* $\mathbf{x}$, $bs(f,\mathbf{x})$, is the maximum number of disjoint blocks $B_1, \dots, B_k$ of $[n]$ such that for each B_i, $f(\mathbf{x}) \neq f(\mathbf{x}^{B_i})$. The *block sensitivity of f*, $bs(f)$, is $max_{\mathbf{x}\in\{0,1\}^n}\, bs(f,\mathbf{x})$.

Then he presented two lemmas to prove the mentioned statement of equivalence.

Lemma 4.7. *For any* $f : \{0,1\}^n \longrightarrow \{0,1\}$*:*

$$s(f) \leq bs(f) \leq C(f).$$

Proof. First inequality follows from the definitions. The singletons which f is sensitive in $s(f)$ will be blocks in $bs(f)$. But it is possible that there exist blocks with more than one element which f is sensitive too. Hence $bs(f) \geq s(f)$.

Observe that for every input $\mathbf{x}$ any certificate for $\mathbf{x}$ must include at least one variable from each set f is sensitive to on $\mathbf{x}$. If it is not like that, there will be sensitive blocks that are not controlled and the certificate for $\mathbf{x}$ will not assure the value of $f(\mathbf{x})$. Thus $C(f) \geq bs(f)$. □

Lemma 4.8. *For any* $f : \{0,1\}^n \longrightarrow \{0,1\}$*:*

$$bs(f) \geq \sqrt{C(f)}.$$

Proof. Let $\mathbf{x}$ be an input achieving certificate complexity so every certificate for $\mathbf{x}$ is of length at least $C(f)$. Let S_1 be some minimal set (with no proper subsets), such that $f(\mathbf{x}) \neq f(\mathbf{x}^{S_1})$. In general we pick S_i to be a minimal disjoint set from the previous created such that $f(\mathbf{x}) \neq f(\mathbf{x}^{S_i})$ until at a certain point no such set exists, say the last set was S_t.

The union of all this sets is a certificate for $\mathbf{x}$, since otherwise it could have been possible to pick another set that changes the value of the function when is flipped. Thus we get:

$$\sum_{i=1}^{t} |S_i| \geq C(f). \tag{4.1}$$

1. Suppose $t \geq \sqrt{C(f)}$. Since f is sensitive to each S_i on $\mathbf{x}$:

$$bs(f) \geq bs(f,\mathbf{x}) \geq t \geq \sqrt{C(f)}.$$

2. Suppose $t < \sqrt{C(f)}$. From the Equation (4.1) it follows:

$$C(f) \leq \sum_{i=1}^{t} |S_i| < \sum_{i=1}^{\sqrt{C(f)}} |S_i| \leq \max_{i\in[t]} \{|S_i|\} \cdot \sqrt{C(f)}.$$

Thus at least one of the sets has to be of size larger than $\sqrt{C(f)}$. For each i, S_i is minimal, then on $\mathbf{x}^{S_i}$ f is sensitive to each element in S_i. Hence,

$$bs(f) \geq bs(f,\mathbf{x}) \geq bs(f,\mathbf{x}^{S_i}) \geq \max_{i\in[t]} \{|S_i|\} > \sqrt{C(f)}.$$

□

Therefore it is given the first result of equivalence between sensitivity measures. The proof is immediate by Lemmas 4.7-4.8 and Theorem 4.4.

Theorem 4.9. *For any $f : \{0,1\}^n \longrightarrow \{0,1\}$:*

$$bs(f) \leq D(f) \leq bs(f)^4.$$

After this statement Nisan co-worked with Szegedy to find other sensitivity measures to add to this equivalence group. They proved that the degree defined in 3.2 and the degree of a polynomial approximating the function are also equivalent. The result is presented in [9] and the paper begins introducing the degree of this polynomial approximation of the Boolean function on L_1.

Definition 4.10. Let f be a boolean function and let p be a real polynomial. We say that p *approximates* f if for every $\mathbf{x} \in \{0,1\}^n$ we have that $|p(\mathbf{x}) - f(\mathbf{x})| \leq 1/3$. The approximate degree of f, $\widetilde{\deg}(f)$, is defined to be the degree of the lower degree polynomial p that approximates f.

For the proof of the required equivalences it will be used the method of symmetrization.

Definition 4.11. Let $p : \mathbb{R}^n \to \mathbb{R}$ be a multivariate polynomial, then the symmetrization of p is

$$p^{sym}(x_1, \dots, x_n) = \frac{\sum_{\pi \in \mathcal{S}_n} p(x_{\pi(1)}, \dots, x_{\pi(n)})}{n!}.$$

As we have observed with symmetric functions in Definition 2.7, since the polynomial's inputs are in $\{0,1\}^n$, truly the symmetrization only depend upon $\sum_{i=1}^n x_i$. Thus it can be represented as an univariate polynomial of $\sum_{i=1}^n x_i$.

Lemma 4.12. *If $p : \mathbb{R}^n \to \mathbb{R}$ is a multivariate polynomial, then there exists a unique univariate polynomial $\tilde{p} : \mathbb{R} \to \mathbb{R}$ of degree at most n such that for all $\mathbf{x} \in \{0,1\}^n$ we have*

$$p^{sym}(\mathbf{x}) = \tilde{p}\Big(\sum_{i=1}^n x_i\Big).$$

Moreover, $deg(\tilde{p}) \leq deg(p)$.

Before the first lemma necessary to prove the next equivalence, it is presented, for completeness, a classical theorem that will be used in the proof of the lemma.

Lemma 4.13. *Let $p : \mathbb{R} \to \mathbb{R}$ be a univariate polynomial of degree d such that any real number x in $[a_1, a_2]$ satisfies $b_1 \leq p(x) \leq b_2$. Then for all $a_1 \leq x \leq a_2$, the derivative of p satisfies $|p'(x)| \leq \frac{d^2 \cdot (b_2 - b_1)}{a_2 - a_1}$.*

Now it is presented the first lemma of the paper and the proof is a consequence of this last result.

Lemma 4.14. *Let $p : \mathbb{R} \to \mathbb{R}$ be a polynomial with the following properties:*

1. *For any integer $0 \leq i \leq n$ we have $b_1 \leq p(i) \leq b_2$.*

2. *For some real $0 \leq x \leq n$ the derivative of p satisfies $|p'(x)| \geq c$.*

Then,

$$deg(p) \geq \sqrt{\frac{c \cdot n}{c + b_2 - b_1}}.$$

Proof. Let $c' = max_{0 \leq x \leq n} |p'(x)| \geq c$. Then it can be seen by Taylor of first order that for all real $0 \leq x \leq n$:

$$b_1 - \frac{c'}{2} \leq p(x) \leq b_2 + \frac{c'}{2}.$$

With a bound on the polynomial evaluated on every real number $a_1 \leq x \leq a_2$ Lemma 4.13 derives that:

$$c' \leq \frac{deg(p)^2(c' + b_2 - b_1)}{n}.$$

Thus,

$$deg(p) \geq \sqrt{\frac{c'n}{c' + b_2 - b_1}} \geq \sqrt{\frac{cn}{c + b_2 - b_1}}.$$

□

Second and main lemma is presented as it follows.

Lemma 4.15. *Let $f : \{0,1\}^n \longrightarrow \{0,1\}$ be an unanimous Boolean function. Then,*

$$deg(f) \geq \sqrt{\frac{n}{2}} \text{ and } \widetilde{deg}(f) \geq \sqrt{\frac{n}{6}}.$$

Proof. Here we give the proof for the inequality on $\widetilde{\deg}(f)$. Let p be a polynomial approximating f and consider $\tilde{p}$ the univariate polynomial giving its symmetrization. This polynomial $\tilde{p}$ satisfies the following properties:

(1) For every integer $0 \leq i \leq n$, $-1/3 \leq \tilde{p}(i) \leq 4/3$ (since for every input $\mathbf{x}$, $p(\mathbf{x})$ is within $1/3$ of a boolean value by the definition of approximation).

(2) $\tilde{p}(0) \leq 1/3$ (by unanimity of the function).

(3) $\tilde{p}(1) \geq 2/3$ (again by unanimity of the function).

Properties (2) and (3) say that for some real $0 \leq z \leq 1$, the derivative of the polynomial $\tilde{p}$ satisfy $|\tilde{p}'(z)| \geq 1/3$. This is because the least value will be $1/3$ if p was a line from 0 to 1 with slope $1/3$. Finally by the condition (1) and the last result $|\tilde{p}'(z)| \geq 1/3$ we can apply Lemma 4.14.

$$deg(p) \geq \deg(\tilde{p}) \geq \sqrt{\frac{1/3 \cdot n}{1/3 + 4/3 - (-1/3)}} = \sqrt{\frac{n}{6}}.$$

The first inequality is mentioned in Lemma 4.12.

The proof for the inequality on $deg(f)$ can be seen analogously by rewriting the three properties and deriving $0 \leq p(z) \leq 1$ and $|p'(z)| \geq 1$. □

Note that the last lemma concerns very special types of Boolean functions. But it turns out that it is enough to give good bounds for all Boolean functions. This can be done by relating the degree to other combinatorial properties of Boolean functions. The following result takes a huge step relating block sensitivity and the degree.

Lemma 4.16. *For every Boolean function $f : \{0,1\}^n \longrightarrow \{0,1\}$:*

$$deg(f) \geq \sqrt{\frac{bs(f)}{2}} \quad and \quad \widetilde{deg}(f) \geq \sqrt{\frac{bs(f)}{6}}.$$

Proof. Suppose $\mathbf{x}$ is the input such that $bs(f) = bs(f, \mathbf{x})$. Let $S_1, \ldots, S_t$ be the sets achieving the block sensitivity. It will be assumed without loss of generality that $f(\mathbf{x}) = 0$. Then it is defined the function $f' : \{0,1\}^t \rightarrow \{0,1\}$ like:

$$f'(y_1, \ldots, y_t) = f(\mathbf{x} \oplus y_1 S_1 \oplus \cdots \oplus y_t S_t).$$

The sign $\oplus$ adds bits in modulo 2. So the j'th bit fed to f is $x_j \oplus y_i$ if $j \in S_i$ for some $1 \leq i \leq t$ and is x_j if j is not in any of the blocks S_i. It is immediate to see that f' is unanimous.

1. $f'(0, \ldots, 0) = f(\mathbf{x}) = 0$ by assumption.
2. $f'(1, \ldots, 1) = 1$ because the fact of feeding any block in the input $(\mathbf{x} \oplus S_i)$ flips the output of $f(\mathbf{x}) = 0$.

Hence this function f' satisfies the hypothesis of Lemma 4.15 and since $deg(f) \geq deg(f')$ because bits x_j are not variables in the definition of f':

$$deg(f) \geq deg(f') \geq \sqrt{\frac{t}{2}} = \sqrt{\frac{bs(f)}{2}}.$$

and

$$\widetilde{\deg}(f) \geq \widetilde{\deg}(f') \geq \sqrt{\frac{t}{6}} = \sqrt{\frac{bs(f)}{6}}.$$

□

Finally it is a direct result from Lemma 4.16 and Theorem 4.9 the equivalence between decision tree complexity and the two degrees defined in 3.2 and 4.10.

Theorem 4.17. *For every $f : \{0,1\}^n \longrightarrow \{0,1\}$:*

$$deg(f) \leq D(f) \leq 16 deg(f)^8.$$

and

$$\widetilde{deg}(f) \leq deg(f) \leq D(f) \leq 1296 \widetilde{deg}(f)^8.$$

At the end of the paper Nisan and Szedegy meditated about the actual group of equivalence of sensitivity measures. They missed having sensitivity on this group remembering that $bs(f) \geq s(f)$. Hence they were not discouraged and conjectured the possible polynomially relation between $bs(f)$ and $s(f)$.

Conjecture 4.18 (Sensitivity Conjecture)**.** *For every $f : \{0,1\}^n \longrightarrow \{0,1\}$ and some c:*

$$bs(f) \leq s(f)^c.$$

From now on, the paper will explain the historical timeline to finally give a proof for the conjecture. It took almost 30 years for it to happen.

There were many reasons to achieve the proof of this conjecture. For example, a parallel RAM is a collection of synchronous parallel processors sharing a global memory. The sensitivity complexity measure $s(f)$ lower bounds $T(f)$ - the time needed by a parallel RAM to compute f. Relating $s(f)$ to other complexities would facilitate the lower bound of $T(f)$ in terms of these measures.

4.2 The Equivalence Theorem

When the Sensitivity Conjecture was presented, there were many researchers investigating how to approach the problem and what was necessary to achieve the proof. Gotsman and Linial presented in [11] a theorem that makes equivalent proving the Sensitivity Conjecture to proving a property on a concrete graph.

Before we can present this theorem, we must introduce some notations and definitions in graphs. We will denote by Q_n the graph on the n-dimensional Hamming cube $\{-1,1\}^n$ where any two vertices are adjacent if and only if they differ in exactly one component. For an induced subgraph G on Q_n we denote the *maximal degree* of G by $\Delta(G)$, i. e. $\Delta(G) = max_{\mathbf{x}\in V(G)} deg_G(\mathbf{x})$. We denote $\Gamma(G) = max\{\Delta(G), \Delta(Q_n \setminus G)\}$.

Given the last two notations it is possible to state and prove the theorem that connects Sensitivity Conjecture problem with a problem in combinatorics.

Theorem 4.19 (Equivalence Theorem). *The following are equivalent for any monotone function $h : \mathbb{N} \to \mathbb{R}$.*

1. *For any induced subgraph G of Q_n such that $|V(G)| \neq 2^{n-1}$, $\Gamma(G) \geq h(n)$.*
2. *For any Boolean function f, $deg(f) < h^{-1}(s(f))$.*

Proof. We begin by transforming the two statements making them depend only on boolean functions. First we associate the subgraph G with a Boolean function $g : \{-1,1\}^n \longrightarrow \{-1,1\}$ defined as it follows:

$$g(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in V(G), \\ -1 & \text{if } \mathbf{x} \notin V(G). \end{cases}$$

Note that $deg_G(\mathbf{x}) = n - s(g,\mathbf{x})$ for $\mathbf{x}$ in $V(G)$ because local sensitivity of g on $\mathbf{x}$ is the number of neighbours of x that do not have the same output of g, i.e. they are not in $V(G)$. The same holds in $Q_n \setminus G$ for $\mathbf{x} \notin V(G)$. We denote by $\mathbf{E}[g]$ the expectation of g and it is possible to compute it explicitly:

$$\mathbf{E}[g] = \frac{1}{2^n} \cdot \sum_{\mathbf{x}\in\{-1,1\}^n} g(\mathbf{x}) = \frac{1}{2^n} \cdot |V(G)| - \frac{1}{2^n} \cdot |V(Q_n \setminus G)|.$$

Now we can transform the two statements. For the first, the condition of a subgraph G such that $|V(G)| \neq 2^{n-1}$ can be stated also as $\mathbf{E}[g] \neq 0$ as it can be seen in the last computation on the expectation of g. The condition $\Gamma(G) \geq h(n)$ means that there exists some vertex such that $deg(\mathbf{x}) = \Gamma(G)$ and this input $\mathbf{x}$ can be in G or either in $Q_n \setminus G$ but for the two subgraphs we have noted that $deg_{\{G, Q_n \setminus G\}}(\mathbf{x}) = n - s(g,\mathbf{x})$ so the condition can be written as $n - s(g,x) \geq h(n)$ and hence $s(g,x) \leq n - h(n)$.

For the second condition we impose now $s(f) < h(n)$ and this will imply $deg(f) < h^{-1}(s(f)) < h^{-1}(h(n)) = n$ since h is monotone.

Thus statements 1 and 2 are clearly equivalent to the following:

1'. For any boolean function g, $\mathbf{E}[g] \neq 0$ implies that there exists some $\mathbf{x}$ such that $s(g, \mathbf{x}) \leq n - h(n)$.

2'. For any boolean function f, $s(f) < h(n)$ implies $deg(f) < n$.

And the proof of the theorem is based in proving the equivalence of this two statements. First it is defined g as it follows:

$$g(\mathbf{x}) = f(\mathbf{x}) \cdot \chi_{[n]}(\mathbf{x}).$$

Remember that $\chi_{[n]}$ denotes the parity function on all indices in $[n]$: $\prod_{i=1}^{n} x_i$. The definition of g given above is universal since every function g can be written like this for some function f and the same for every function f and some function g.

First observe that for all $\mathbf{x}$ in the Hamming cube, since $\chi_{[n]}(\mathbf{x}^i) = -\chi_{[n]}(\mathbf{x})$,

$$s(g, \mathbf{x}) = \#\{i \in [n] : f(\mathbf{x}^i) \cdot \chi_{[n]}(\mathbf{x}^i) \neq f(\mathbf{x}) \cdot \chi_{[n]}(\mathbf{x})\} = \#\{i \in [n] : f(\mathbf{x}^i) = f(\mathbf{x})\} = n - s(f, \mathbf{x}). \quad (4.2)$$

Second the Fourier coefficients of g, $\hat{g}(I)$ where $I \subseteq [n]$, are derived from the Fourier coefficients of f:

$$\hat{g}(I) = \langle g, \chi_I \rangle = \langle f \cdot \chi_{[n]}, \chi_I \rangle = \langle f, \chi_{[n]} \cdot \chi_I \rangle = \langle f, \chi_{[n]-I} \rangle = \hat{f}([n] - I).$$

Therefore, by Proposition 1.18, $\mathbf{E}[g] = \hat{g}(\emptyset) = \hat{f}([n] - \emptyset) = \hat{f}([n])$, i. e. $E[g] \neq 0$ if and only if the highest order coefficient in the representation of f as a polynomial is not null so $deg(f) = n$. The equivalence between $1'$ and $2'$ is shown by proving the contraposition of every statement.

First we prove $1'$ implies $2'$. Suppose $\mathbf{E}[g] \neq 0$ so $deg(f) = n$. By $1'$ there exists an input $\mathbf{x}$ such that $s(g, \mathbf{x}) \leq n - h(n)$. Therefore, by (4.2), $n - s(f, \mathbf{x}) \leq n - h(n)$, then by definition of the sensitivity of f, $s(f) \geq s(f, \mathbf{x}) \geq h(n)$.

Second we prove $2'$ implies $1'$. Suppose that for all $\mathbf{x}$, $s(g, \mathbf{x}) > n - h(n)$. By (4.2) this implies $n - s(f, \mathbf{x}) > n - h(n)$ and again by definition of the sensitivity of f follows $s(f) < h(n)$. Then by $2'$ this implies that $deg(f) < n$, that is equivalent to $\mathbf{E}[g] = 0$, contradicting the premise. □

This theorem was really meaningful since proving $h(n) = \sqrt{n}$ in statement 1 would prove the Sensitivity Conjecture 4.18. Other approaches on the proof of the conjecture were proposed but Equivalence Theorem 4.19 was the one that led to success.

4.3 Huang's proof from the Book

Almost 30 years after the presentation of the conjecture, Huang's proved in a paper published at Annals of Mathematics (see [10]) the statement 1 of Equivalence Theorem 4.19. The proof is so simple that Aaronson and O'Donnell called Huang's paper the "Book" proof of the Sensitivity Conjecture, referring to Paul Erdös' notion of a celestial book in which God writes the perfect proof of every theorem.

Since the condition of the statement 1 in Equivalence Theorem 4.19 is $|V(G)| \neq 2^{n-1}$ Huang worked on a proof for a $(2^{n-1}+1)$-vertex induced subgraph G of Q_n. The key elements of the proof are a theorem presented by Cauchy many years ago and an understandable property of the adjacency matrix of Q_n.

Lemma 4.20 (Cauchy's Interlace Theorem)**.** *Let A be a symmetric $n \times n$ matrix, and B be a $m \times m$ principal submatrix of A, for some $m < n$. If the eigenvalues of A are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and the eigenvalues of B are $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m$, then for all $1 \leq i \leq m$,*

$$\lambda_i \geq \mu_i \geq \lambda_{i+n-m}.$$

A short proof of Cauchy's Interlace Theorem can be found in [13].

Proof. Given the polynomials $f(x)$ and $g(x)$ which have all real roots $r_1 \le r_2 \le \cdots \le r_n$ and $s_1 \le s_2 \le \cdots \le s_{n-1}$ respectively. We say that f and g interlace if and only if

$$r_1 \le s_1 \le r_2 \le s_2 \le \cdots \le s_{n-1} \le r_n.$$

It is known that the roots of polynomials f, g interlace if and only if the linear combinations $f + \alpha g$ have all real roots for every real number α. Using this we can see that if A is a Hermitian matrix and B is a principle submatrix of A, then the eigenvalues of B interlace the eigenvalues of A.

Given a real number α, since the principal submatrix B is the remain of deleting the same number of rows and columns from matrix A, we select the following partition of A without loss of generality

$$\left[\begin{array}{c|c} B & c \\ \hline c^* & d \end{array}\right]$$

Then we write the characteristic polynomial of a Hermitian matrix and consider the following equation that follows from linearity of the determinant:

$$\left|\begin{array}{c|c} B - x\cdot I & c \\ \hline c^* & d - x + \alpha \end{array}\right| = \left|\begin{array}{c|c} B - x\cdot I & c \\ \hline c^* & d - x \end{array}\right| + \left|\begin{array}{c|c} B - x\cdot I & c \\ \hline 0 & \alpha \end{array}\right|$$

From this equation it can be seen that $|A - x \cdot I| + \alpha \cdot |B - x \cdot I|$ is the characteristic polynomial of the following matrix

$$\left[\begin{array}{c|c} B & c \\ \hline c^* & d + 2\cdot\alpha \end{array}\right]$$

Since this matrix is also a Hermitian matrix all its eigenvalues are real. Hence the characteristic polynomial $|A - x \cdot I| + \alpha \cdot |B - x \cdot I|$ has all real roots for any α. Therefore, the roots of the characteristic polynomials of A and B interfere, i.e. the eigenvalues of A and B interfere. □

Lemma 4.21. *We define a sequence of symmetric square matrices iteratively as follows,*

$$A_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \; A_n = \begin{bmatrix} A_{n-1} & I \\ I & -A_{n-1} \end{bmatrix}$$

Then A_n is a $2^n \times 2^n$ matrix whose eigenvalues are $\sqrt{n}$ of multiplicity 2^{n-1}, and -$\sqrt{n}$ of multiplicity 2^{n-1}.

Proof. It is proved by induction that $A_n^2 = n \cdot I$. For $n = 1$, $A_1^2 = I$. Suppose the statement holds for $n-1$, then

$$A_n^2 = \begin{bmatrix} A_{n-1}^2 + I & 0 \\ 0 & A_{n-1}^2 + I \end{bmatrix} = \begin{bmatrix} (n-1)I + I & 0 \\ 0 & (n-1)I + I \end{bmatrix} = nI.$$

Therefore, $A_n^2 - n \cdot I$ is a characteristic polynomial and the eigenvalues of A_n are either $\sqrt{n}$ or $-\sqrt{n}$. Since $Tr(A_n) = 0$, A_n has exactly half of the eigenvalues being $\sqrt{n}$ and the rest being $-\sqrt{n}$. □

Lemma 4.22. *Suppose H is an m-vertex undirected graph, and A is a symmetric matrix whose entries are in $\{-1, 0, 1\}$ whose rows and columns are indexed by $V(H)$, and whenever u and v are non-adjacent in H, $A_{u,v} = 0$. Then,*

$$\Delta(H) \ge \lambda_1 = \lambda_1(A).$$

Proof. Suppose $\mathbf{v}$ is the eigenvector corresponding to λ_1. Then $\lambda_1 \cdot \mathbf{v} = A \cdot \mathbf{v}$. Without loss of generality, assume v_1 is the coordinate of $\mathbf{v}$ that has the largest absolute value. Then,

$$|\lambda_1 \cdot v_1| = |(A \cdot \mathbf{v})_1| = \left| \sum_{j=1}^{m} A_{1,j} \cdot v_j \right| \leq \sum_{j=1}^{m} |A_{1,j}| \cdot |v_1| \leq \Delta(H) \cdot |v_1|.$$

Hence $|\lambda_1| \leq \Delta(H)$. □

Note that the matrix A is almost the adjacency matrix of H. It is not exactly the adjacency matrix because H is an undirected graph but A can take -1 values. Despite this, A conserves the property of valuing zero when two vertices are not adjacent and not zero if they are adjacent.

With the lemmas above it is possible to present and prove the main theorem of the paper which proves Sensitivity Conjecture.

Theorem 4.23 (Huang's Theorem)**.** *For every integer $n \geq 1$, let H be an arbitrary $(2^{n-1}+1)$-vertex induced subgraph of Q_n, then*

$$\Delta(H) \geq \sqrt{n}.$$

Moreover this inequality is tight when n is a perfect square.

Proof. Let A_n be the sequence of matrices defined in Lemma 4.21. As we have said the entries of A_n are in $\{-1, 0, 1\}$. By the iterative construction of A_n, it can be seen that changing every (-1)-entry of A_n to 1 gives exactly the adjacency matrix of Q_n. Thus A_n and Q_n satisfy the conditions in Lemma 4.22. Therefore, a $(2^{n-1}+1)$-vertex induced subgraph H of Q_n and the principal submatrix A_H of A_n naturally induced by H also satisfy the conditions of Lemma 4.22. Then,

$$\Delta(H) \geq \lambda_1(A_H). \tag{4.3}$$

On the other hand, from Lemma 4.21, the eigenvalues of A_n are known to be $\sqrt{n}$ with multiplicity 2^{n-1} and $-\sqrt{n}$ also with multiplicity 2^{n-1}. Note that A_H is a $(2^{n-1}+1) \times (2^{n-1}+1)$ submatrix of the $2^n \times 2^n$ matrix A_n. Then by Cauchy's Interlace Theorem 4.20,

$$\lambda_1(A_H) \geq \lambda_{1+2^n-(2^{n-1}+1)}(A_n) = \lambda_{2^{n-1}}(A_n) = \sqrt{n}. \tag{4.4}$$

Combining the inequality (4.3) and the inequality (4.4) we have the desired result:

$$\Delta(H) \geq \lambda_1(A_H) \geq \sqrt{n}.$$

□

Finally by Lemma 4.16, Equivalence Theorem 4.19 and Huang's Theorem 4.23 it is possible to give a value for c in the Sensitivity Conjecture 4.18.

Corollary 4.24. *For every $f : \{0,1\}^n \longrightarrow \{0,1\}$:*

$$bs(f) \leq s(f)^4.$$

5. Thresholds on graph properties for Random Graphs

Graph Theory is the subject of mathematics that relates various graph properties. Once we work and study this area it is natural to imagine and have the notion of a random graph. This chapter is privileged to move on to the Graph Theory and Random Graphs only because we will see that we can write all the definitions given in this context with the language of Boolean functions. The final objective is to show Margulis-Russo Formula 5.23 to study random graph properties taking advantage of Fourier analysis on Boolean functions.

5.1 Random graphs and graph properties

The study of random graphs was first presented by Erdös and Rényi in [14]. These graphs are attractive to study because of its definition.

First model starts with an empty graph of n vertices and inserts m edges in such a way that all possible $\binom{\binom{n}{2}}{m}$ choices are equally likely.

Definition 5.1 (Uniform random graph)**.** Let $\mathcal{G}_{n,m}$ be the family of all labeled graphs with vertex set $V = [n]$ and exactly m edges, with $0 \leq m \leq \binom{n}{2}$. To every graph $G \in \mathcal{G}_{n,m}$, we assign a probability

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

We will call this a *uniform random graph* and it will be denoted by $\mathbb{G}_{n,m} = ([n], E_{n,m})$.

Second model needs to fix a probability $0 \leq p \leq 1$. Same as before, we will start with an empty graph but now we will perform $\binom{n}{2}$ Bernoulli experiments inserting edges independently with probability p.

Definition 5.2 (Binomial random graph)**.** Let $0 \leq p \leq 1$ be a fixed probability and let $\mathcal{G}_{n,m}$ be the family of all labeled graphs with vertex set $V = [n]$ and exactly m edges, with $0 \leq m \leq \binom{n}{2}$. To every graph $G \in \mathcal{G}_{n,m}$, we assign a probability

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

We will call this a *binomial random graph* and it will be denoted by $\mathbb{G}_{n,p} = ([n], E_{n,p})$. This model of random graph was introduced by Gilbert in [15].

From a practical standpoint, if we have to choose from which type of random graph we want to study a determined graph property we will select binomial random graphs before uniform random graphs. The former are handier because the edges are independent and we only need results from Probability Theory, while the latter are more combinatorial and need counting.

The two models of random graphs presented can be related once we fix the number of edges of the graph. The following proposition states and proves this relationship.

Proposition 5.3. *Given a number of edges m, the random graph $\mathbb{G}_{n,p}$ is equally likely to be one of the $\binom{\binom{n}{2}}{m}$ graphs that have m edges.*

Proof. Let G_0 be any labeled graph with m edges. Since the event $\{\mathbb{G}_{n,p} = G_0\} \subseteq \{|E_{n,p}| = m\}$ we have:

$$\begin{aligned}
p(\mathbb{G}_{n,p} = G_0 \big| \, |E_{n,p}| = m) &= \frac{p(\mathbb{G}_{n,p} = G_0 \cap |E_{n,p}| = m)}{p(|E_{n,p}| = m)} \\
&= \frac{p(\mathbb{G}_{n,p} = G_0)}{p(|E_{n,p}| = m)} \\
&= \frac{p^m(1-p)^{\binom{n}{2}-m}}{\binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}} \\
&= \binom{\binom{n}{2}}{m}^{-1}.
\end{aligned}$$

□

Note that the number of edges for an uniform random graph is fixed. On the other hand, for a binomial random graph it is unknown but we can compute the expected number of edges. Since $\mathbb{G}_{n,p}$ is a binomial random variable of $\binom{n}{2}$ events with probability p the expectation is immediate to compute and we will have an expected number of edges $m = \binom{n}{2}p$.

Hence, for n large random graphs $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,p}$ should behave in a similar fashion when the number of edges in $\mathbb{G}_{n,m}$ equals or is close to the expected number of edges in $\mathbb{G}_{n,p}$. We can achieve this when

$$m = \binom{n}{2} p \approx \frac{n^2 p}{2}.$$

Or, in terms of some fixed number of edges, when the edge probability in $\mathbb{G}_{n,p}$ is

$$p \approx \frac{2m}{n^2}.$$

Here we used the notation $f \approx g$ to denote that $f = (1 + o(1))g$.

We introduce now a useful tool called *coupling technique* that generates the random graph $\mathbb{G}_{n,p}$ in two different steps. The same procedure that will be presented for binomial random graphs can be extended to uniform random graphs. The idea is to obtain a binomial random graph $\mathbb{G}_{n,p}$ in terms of two independent random graphs:

$$\mathbb{G}_{n,p} = \mathbb{G}_{n,p_1} \cup \mathbb{G}_{n,p_2}, \text{ where, } 1 - p = (1 - p_1)(1 - p_2).$$

Thus an edge is not included in $\mathbb{G}_{n,p}$ if it is not included in either of $\mathbb{G}_{n,p_1}$ or $\mathbb{G}_{n,p_2}$.

Definition 5.4. We will say that two graphs are *coupled* and denote this by $\mathbb{G}_{n,p_1} \subseteq \mathbb{G}_{n,p}$ if the random graph $\mathbb{G}_{n,p}$ is obtained by superimposing a random graph G_{n,p_1} with other random graph G_{n,p_2} and replacing eventual double edges by one.

To study graph properties in random graphs we will think about a property $\mathcal{P}$ as a subset of the set of all labeled graphs on vertex set $[n]$, i.e., $\mathcal{P} \subseteq 2^{\binom{n}{2}}$. Thus we say that a fixed graph G satisfies some graph property $\mathcal{P}_0$ if it belongs to the subset $\mathcal{P}_0$, i.e., $G \in \mathcal{P}_0$.

Now we present various definitions about graph properties. A graph property $\mathcal{P}$ is *monotone increasing* if adding an edge to the graph does not destroy the property, i.e., if $G \in \mathcal{P}$ then $G + e \in \mathcal{P}$. A monotone increasing property is *non-trivial* if the empty graph $\emptyset$ does not satisfy the property but the complete graph K_n does. A graph property $\mathcal{P}$ is *monotone decreasing* if removing an edge from a graph does not destroy the property. Note that a graph property $\mathcal{P}$ is monotone increasing if and only if its complement is monotone decreasing. Observe also that not all graph properties are monotone.

From the coupling argument it follows immediately that if $\mathcal{P}$ is a monotone increasing property then, whenever $p < p'$:

$$p(\mathbb{G}_{n,p} \in \mathcal{P}) \leq p(\mathbb{G}_{n,p'} \in \mathcal{P}). \tag{5.1}$$

It is possible to upper-bound the probability of a uniform random graph satisfying a certain graph property in terms of the probability of a binomial random graph satisfying this graph property and the number of edges. There is a more accurate upper-bound that does not depend on the number of edges if the graph property is monotone increasing. For more details see [16].

5.2 Thresholds and sharp thresholds

One of the main reasons of why random graphs are interesting to study is because of the abrupt nature of the appearance and disappearance of certain graph properties. Here we introduce more formally this phenomena.

Definition 5.5 (Threshold v1)**.** A function $p^* = p^*(n)$ is a *threshold* for a monotone increasing property $\mathcal{P}$ in the random graph $\mathbb{G}_{n,p}$ if

$$\lim_{n\to\infty} p(\mathbb{G}_{n,p} \in \mathcal{P}) = \begin{cases} 0 & \text{if } p/p^* \to 0, \\ 1 & \text{if } p/p^* \to \infty. \end{cases}$$

as $n \to \infty$.

This definition can be adapted easily for monotone decreasing graph properties. To short many statements of theorems and proofs of this chapter we will say that a sequence of events A_n occurs *with high probability* (w.h.p.) if

$$\lim_{n\to\infty} p(A_n) = 1.$$

Thus the previous Definition 5.5 for thresholds can be rewritten in another way.

Definition 5.6 (Threshold v2)**.** A function $p^* = p^*(n)$ is a *threshold* for a monotone increasing property $\mathcal{P}$ in the random graph $\mathbb{G}_{n,p}$ if $\mathbb{G}_{n,p} \notin \mathcal{P}$ w.h.p. if $p \ll p^*$, while $\mathbb{G}_{n,p} \in \mathcal{P}$ w.h.p. if $p \gg p^*$.

This definition may seem specific for certain monotone increasing graph properties. However, Bollobás and Thomason presented in [17] a theorem that states that in fact all non-trivial monotone graph properties have a threshold.

Theorem 5.7. *Every non-trivial monotone graph property has a threshold.*

Proof. We will prove the statement constructing a probability that will depend on our non-trivial monotone graph property $\mathcal{P}$ in order to have a threshold in $p^* = p(1/2)$.

Without loss of generality, we define $p : (0,1) \to (0,1)$ such that given any ϵ in $(0,1)$ we define $p(\epsilon)$ to satisfy $p(\mathbb{G}_{n,p(\epsilon)} \in \mathcal{P}) = \epsilon$. Note that we are able to create p because $p(\epsilon)$ exists since

$$p(\mathbb{G}_{n,p} \in \mathcal{P}) = \sum_{G\in\mathcal{P}} p^{|E(G)|}(1-p)^{\binom{n}{2}-|E(G)|}$$

is a monotone increasing function from 0 to 1 because $\mathcal{P}$ is a monotone increasing property. Thus there exists an inverse function that will be our $p(\epsilon)$.

Let $G_1, \dots, G_k$ be independent copies of $\mathbb{G}_{n,p}$. The probability of having an edge in $G_1 \cup \cdots \cup G_k$ is $1-(1-p)^k$. Hence, the graph $G_1 \cup \cdots \cup G_k$ is distributed as $\mathbb{G}_{n,1-(1-p)^k}$. Now, since $0 < p < 1$, $1-(1-p)^k \le kp$ and therefore by the coupling argument of Definition 5.4:

$$\mathbb{G}_{n,1-(1-p)^k} \subseteq \mathbb{G}_{n,kp}.$$

Therefore if $\mathbb{G}_{n,kp} \notin \mathcal{P}$ then $G_1 \cup \cdots \cup G_k$ and this implies $G_1, \dots, G_k \notin \mathcal{P}$. Thus

$$p(\mathbb{G}_{n,kp} \notin \mathcal{P}) \le \left[p(\mathbb{G}_{n,p} \notin \mathcal{P})\right]^k. \tag{5.2}$$

This inequality is derived from (5.1) and the fact that the complement of a monotone increasing property is a monotone decreasing property.

Now let $\omega(n)$ be a function such that $\omega(n) \to \infty$ as $n \to \infty$. This function will increase slowly to infinity so suppose, without loss of generality, that $\omega(n) \ll \log\log n$, i.e., $\omega(n) = o(\log\log n)$. Suppose also that $p = p^* = p(1/2)$ and $k = \omega$ so the graph is being obtained as a big number of random graphs superimposed. Then

$$p(\mathbb{G}_{n,\omega p^*} \notin \mathcal{P}) \le \left[p(\mathbb{G}_{n,p^*}) \notin \mathcal{P}\right]^\omega = \left[p(\mathbb{G}_{n,p(\frac{1}{2})}) \notin \mathcal{P}\right]^\omega = \left(\frac{1}{2}\right)^\omega = o(1). \tag{5.3}$$

The inequality is given by Equation (5.2). On the other hand, for $p = p^*/\omega$:

$$\frac{1}{2} = p(\mathbb{G}_{n,p(\frac{1}{2})} \notin \mathcal{P}) = p(\mathbb{G}_{n,p^*} \notin \mathcal{P}) = p(\mathbb{G}_{n,\omega p} \notin \mathcal{P}) \le \left[p(\mathbb{G}_{n,p} \notin \mathcal{P})\right]^\omega = \left[p(\mathbb{G}_{n,\frac{p^*}{\omega}} \notin \mathcal{P})\right]^\omega.$$

Again, the inequality is given by Equation (5.2). Next, from the equation above we get

$$p(\mathbb{G}_{n,\frac{p^*}{\omega}} \notin \mathcal{P}) \ge \left(\frac{1}{2}\right)^\omega = 1 - o(1). \tag{5.4}$$

Finally, note that we have proved the statement if we follow the definition of a threshold given in Definition 5.5. First, if $p = \omega p^*$ so $p/p^* = \omega \to \infty$ as $n \to \infty$, then (5.3) yields that $p(\mathbb{G}_{n,p} \notin \mathcal{P}) \le o(1)$. Hence,

$$p(\mathbb{G}_{n,p} \in \mathcal{P}) = 1 - p(\mathbb{G}_{n,p} \notin \mathcal{P}) \ge 1 - o(1).$$

Second, if $p = p^*/\omega$ so $p/p^* = 1/\omega \to 0$ as $n \to \infty$, then (5.4) yields that $p(\mathbb{G}_{n,p} \notin \mathcal{P}) \ge 1 - o(1)$. Thus,

$$p(\mathbb{G}_{n,p} \in \mathcal{P}) = 1 - p(\mathbb{G}_{n,p} \notin \mathcal{P}) \le 1 - (1 - o(1)) = o(1).$$

□

In many monotone graph properties it can be observed a more delicate threshold that increases faster that the one defined in Definition 5.5 and 5.6.

Definition 5.8. A function $p^* = p^*(n)$ is a *sharp threshold* for a monotone increasing property $\mathcal{P}$ in the random graph $\mathbb{G}_{n,p}$ if for every $\epsilon > 0$

$$\lim_{n\to\infty} p(\mathbb{G}_{n,p} \in \mathcal{P}) = \begin{cases} 0 & \text{if } p/p^* \leq 1-\epsilon, \\ 1 & \text{if } p/p^* \geq 1+\epsilon. \end{cases}$$

For this reason, the sharp threshold is called "sharp". If we analyze Definition 5.8 we can see that the probability of having a graph property for the random graph will be zero until it arrives to the threshold, where the probability will increase in the interval $[p^* - \epsilon, p^* + \epsilon]$ (almost vertically) from 0 to 1. On the other hand, a normal threshold from Definition 5.5 increases from 0 to 1 but in a wider interval around p^*.

Finally we present two different lemmas, derived directly from Markov's inequality, that will be useful tools to prove thresholds.

Lemma 5.9 (First moment method)**.** *Let $X_n \geq 0$ be an integer valued random variable. If $\mathbf{E}[X_n] \to 0$ then $X_n = 0$ w.h.p. as $n \to \infty$.*

Proof. First we put $t = 1$ in Markov's inequality $p(X_n \geq t) \leq \frac{\mathbf{E}[X_n]}{t}$. Thus

$$p(X_n > 0) = p(X_n \geq 1) \leq \mathbf{E}[X_n].$$

Consequently, if $\mathbf{E}[X_n] \to 0$ as $n \to \infty$ then $p(X_n > 0) \to 0$ so $p(X_n = 0) \to 1$ as $n \to \infty$ and $X_n = 0$ with high probability. □

Lemma 5.10 (Second moment method)**.** *Let $X_n \geq 0$ be an integer valued random variable. If $\mathbf{E}[X_n] > 0$ for n large and $Var[X_n]/\mathbf{E}[X_n]^2 \to 0$ then $X_n > 0$ w.h.p. as $n \to \infty$.*

Proof. We set $t = \mathbf{E}[X_n]$ in Chebyshev's inequality $p(|X_n - \mathbf{E}[X_n]| > t) \leq \frac{Var[X_n]}{t^2}$:

$$p(|X_n - \mathbf{E}[X_n]| > \mathbf{E}[X_n]) \leq \frac{Var[X_n]}{(\mathbf{E}[X_n])^2}.$$

Recall that $X_n \geq 0$ so $p(|X_n - \mathbf{E}[X_n]| > \mathbf{E}[X_n]) = p(X_n = 0)$. Therefore, if $Var[X_n]/\mathbf{E}[X_n]^2 \to 0$ then $p(X_n > 0) \to 1$ as $n \to \infty$ and $X_n > 0$ with high probability. □

5.3 Threshold on cycles in a random graph

We start with the first example of a threshold for a graph property. After taking a look on the proof of a threshold for the graph property for having a triangle in a random graph presented by Novozhilov in [18], a natural question is if the proof can be generalized for k-cycles instead of triangles. Here we proof that the threshold presented is conserved for every k-cycle. From now on, we will refer to a cycle of k vertices by k-cycle.

We will denote $C_{k,n}$ the random variable on the space $\mathbb{G}_{n,p}$ which is equal to the number of k-cycles in a random graph. Then we can introduce the first theorem.

Theorem 5.11. *Let $k \geq 3$ be the number of vertices of the cycle and $\alpha : \mathbb{N} \to \mathbb{R}$ be a function such that $\alpha(n) \to 0$ as $n \to \infty$; let $p(n) = \frac{\alpha(n)}{n}$ for each $n \in \mathbb{N}$. Then $C_{k,n} = 0$ w.h.p..*

Proof. The goal is to show that $p(C_{k,n} = 0) \to 1$ as $n \to \infty$. With First Moment Method 5.9 this is equivalent to proving $\mathbf{E}[C_{k,n}] \to 0$. For each fixed n and k we can represent $C_{k,n}$ as

$$C_{k,n} = \mathbf{1}_{\tau_1} + \cdots + \mathbf{1}_{\tau_s}, s = \binom{n}{k},$$

where τ_i is the event that the ith tuple of k vertices from the set of all vertices of $\mathbb{G}_{n,p}$ forms a k-cycle. Here we assume that all possible k-tuples are ordered and labeled. Using the linearity of expectation:

$$\mathbf{E}[C_{k,n}] = \sum_{i=1}^{s} \mathbf{E}[\mathbf{1}_{\tau_i}] = \sum_{i=1}^{s} p(\tau_i) = \binom{n}{k} p^k,$$

since $p(\tau_i) = p^k$ in the Erdös-Rényi random graphs $\mathbb{G}_{n,p}$. Finally, we have

$$\mathbf{E}[C_{k,n}] = \binom{n}{k} p^k = \frac{n!}{(n-k)!k!} \frac{\alpha^k(n)}{n^k} = \frac{n(n-1)\dots(n-k+1)\alpha^k(n)}{k!n^k} \approx \frac{\alpha^k(n)}{k!} \to 0.$$

$\square$

Now we establish the conditions for Erdös-Rényi random graphs to have a k-cycle almost always.

Theorem 5.12. *Let $k \geq 3$ be the number of vertices of the cycle and $\omega : \mathbb{N} \to \mathbb{R}$ be a function such that $\omega(n) \to \infty$ as $n \to \infty$. Let $p(n) = \omega(n)/n$ for each $n \in \mathbb{N}$. Then $C_{k,n} \geq 1$ w.h.p..*

Proof. From Second Moment Method 5.10 proving this theorem is equivalent to proving $Var[C_{k,n}]/\mathbf{E}[C_{k,n}]^2 \to 0$ as $n \to \infty$. It is a well known definition that $Var[C_{k,n}] = \mathbf{E}[C_{k,n}^2] - \mathbf{E}[C_{k,n}]^2$ and we already know that $\mathbf{E}[C_{k,n}] \approx \omega^k(n)/k!$ from the proof of Theorem 5.11. Hence we only need to find out $\mathbf{E}[C_{k,n}^2]$. Using the same notation as before

$$\begin{aligned}
\mathbf{E}[C_{k,n}^2] &= \mathbf{E}[(\mathbf{1}_{\tau_1} + \cdots + \mathbf{1}_{\tau_s})^2] \\
&= \sum_{i=1}^{s} \mathbf{E}[\mathbf{1}_{\tau_i}^2] + \sum_{i \neq j} \mathbf{E}[\mathbf{1}_{\tau_i}\mathbf{1}_{\tau_j}] \\
&= \sum_{i=1}^{s} \mathbf{E}[\mathbf{1}_{\tau_i}] + \sum_{i \neq j} \mathbf{E}[\mathbf{1}_{\tau_i}\mathbf{1}_{\tau_j}].
\end{aligned}$$

The first term of the sum is $\omega^k(n)/k!$. Note that $\mathbf{E}[\mathbf{1}_{\tau_i}\mathbf{1}_{\tau_j}] = p(\tau_i \cap \tau_j)$ which is the probability that both k-tuples of the vertices subset with indices i and j belong to $\mathbb{G}_{n,p}$.

It is easy to imagine that cycles can intersect, if they do, in a vertex or rather in an edge (two vertices). However it is possible to intersect in at most $k-1$ vertices. On the contrary, if they intersect in k vertices then they must be the same cycle.

Therefore, there are different possible values for $p(\tau_i \cap \tau_j)$. If the two cycles do not intersect or they do in a vertex the probability will be $p(\tau_i \cap \tau_j) = p^{2k}$. On the other hand, if the two cycles intersect in more than two vertices, let m be the number of vertices in common, the probability will be $p(\tau_i \cap \tau_j) = p^{2k-(m-1)}$.

The total number of the pairs of k-tuples i and j with no common vertices is $\binom{n}{k}\binom{n-k}{k}$. On the other hand, the total number of the pairs of k-tuples with m common vertices is $k\binom{n}{k}\binom{n-k}{k-m}$. Summing everything we have:

$$\begin{aligned}\sum_{i\neq j}\mathbf{E}[\mathbf{1}_{\tau_i}\mathbf{1}_{\tau_j}] &= \sum_{i\neq j}p(\tau_i\cap\tau_j)\\ &= \binom{n}{k}\binom{n-k}{k}p^{2k}+\sum_{m=1}^{k-1}k\binom{n}{k}\binom{n-k}{k-m}p^{2k-(m-1)}.\end{aligned} \tag{5.5}$$

Now we will use the following facts:

$$\binom{n}{k}\approx\binom{n-k}{k-m}\approx\frac{n^{k-m}}{(k-m)!},\text{ for } 0\leq m\leq k-1.$$

Plugging the argument above into the equation (5.5) we have the following result:

$$\begin{aligned}\sum_{i\neq j}\mathbf{E}[\mathbf{1}_{\tau_i}\mathbf{1}_{\tau_j}] &\approx \frac{n^k}{k!}\frac{n^k}{k!}p^{2k}+k\frac{n^k}{k!}\sum_{m=1}^{k-1}\frac{n^{k-m}}{(k-m)!}p^{2k-(m-1)}\\ &= \frac{n^{2k}}{(k!)^2}\frac{\omega^{2k}(n)}{n^{2k}}+k\sum_{m=1}^{k-1}\frac{n^{2k-m}}{k!(k-m)!}\frac{\omega^{2k-(m-1)}(n)}{n^{2k-(m-1)}}\\ &= \left(\frac{\omega^k(n)}{k!}\right)^2+\sum_{m=1}^{k-1}\frac{\omega^{2k-(m-1)}(n)}{(k-1)!(k-m)!n}\\ &= \mathbf{E}[C_{k,n}]^2(1+o(1)).\end{aligned}$$

Remember that $\mathbf{E}[C_{k,n}]=\omega^k(n)/k!\to\infty$ as $n\to\infty$. Hence,

$$\begin{aligned}\frac{Var[P_{n,k}]^2}{\mathbf{E}[P_{n,k}]} &= \frac{\mathbf{E}[C_{k,n}^2]-\mathbf{E}[C_{k,n}]^2}{\mathbf{E}[C_{k,n}]^2}\\ &= \frac{\mathbf{E}[C_{k,n}]+\sum_{i\neq j}\mathbf{E}[\mathbf{1}_{\tau_i}\mathbf{1}_{\tau_j}]-\mathbf{E}[C_{k,n}]^2}{\mathbf{E}[C_{k,n}]^2}\\ &= \frac{1}{\mathbf{E}[C_{k,n}]}+\frac{\sum_{i\neq j}\mathbf{E}[\mathbf{1}_{\tau_i}\mathbf{1}_{\tau_j}]-\mathbf{E}[C_{k,n}]^2}{\mathbf{E}[C_{k,n}]^2}\\ &= \frac{1}{\mathbf{E}[C_{k,n}]}+\frac{\mathbf{E}[C_{k,n}]^2(1+o(1))-\mathbf{E}[C_{k,n}]^2}{\mathbf{E}[C_{k,n}]^2}\to 0.\end{aligned}$$

□

Therefore, if we have in mind the definition of a threshold given in Definition 5.6, Theorem 5.11 and Theorem 5.12 prove that $p^*=p^*(n)=1/n$ is a threshold for having k-cycles on a random graph.

5.4 Sharp threshold on connectivity

This section follows with the second example of a threshold for a graph property. In this case we will follow the result given in [19] to give a sharp threshold on connectivity in Erdös-Rényi random graphs. A graph is connected when there is a path between every pair of vertices. One easy way to approach if a graph is connected is by considering the degree of every vertex.

From now on we denote the degree of an arbitrary vertex k of $\mathbb{G}_{n,p}$ by D_k. This will be a random variable where each of the $n-1$ edges connecting the other vertices to the vertex k is present independently with probability p. So D_k will follow a binomial distribution, i.e., $D \sim Bin(n-1, p)$. Hence the expected degree of every vertex will be the same: $\mathbf{E}[D_k] = (n-1)p$ for all k in $[n]$.

Now we present the principal theorem that proves the sharp threshold for connectivity.

Theorem 5.13. *Let*

$$p(n) = \lambda \frac{\ln n}{n}$$

for a constant $\lambda > 0$. *Then*

- *If* $\lambda < 1$, *then* $p(\mathbb{G}_{n,p(n)}$ *is connected*$) \to 0$.
- *If* $\lambda > 1$, *then* $p(\mathbb{G}_{n,p(n)}$ *is connected*$) \to 1$.

Proof. First we treat the case $\lambda < 1$. Let X_n denote the number of isolated vertices in $\mathbb{G}_{n,p}$. Then we need to show that $X_n > 0$ w.h.p..

Next we compute the expected number of isolated vertices. We define I_i to be the indicator random variable which indicates that the *ith* vertex is isolated. Hence, using the linearity of the expectation:

$$\mathbf{E}[X_n] = \sum_{i=1}^{n} \mathbf{E}[I_i] = n \cdot p(\text{node } i \text{ is isolated}) = n \cdot (1 - p(n))^{n-1} = n \cdot q(n). \tag{5.6}$$

From now on we denote the probability of an isolated vertex by $q(n) = (1 - p(n))^{n-1}$. The limit of $\mathbf{E}[X_n]$ as $n \to \infty$ is better computed if we look at the limit of $\ln \mathbf{E}[X_n]$ as $n \to \infty$:

$$\ln \mathbf{E}[X_n] = \ln n + (n-1) \ln (1 - p(n)) \approx \ln n - \frac{n-1}{n} \lambda \ln n \to \infty. \tag{5.7}$$

Remember that the notation $f \approx g$ means that f and g are asymptotically the same. In the last expression we have used the first-order Taylor expansion $\ln (1-x) \approx -x$. Observe that this limit tends to ∞ only because of the hypothesis that $\lambda < 1$.

Thus $\mathbf{E}[X_n] \to \infty$ as $n \to \infty$. This indicates that the statement we want to prove if on its way of being true because we want to prove that $p(X_n > 0) \to 1$. However this is not enough because we will get the proof derived from Second Moment Method 5.10. So we need to compute $Var[X_n]$ and prove that the ratio $Var[X_n]/\mathbf{E}[X_n]^2 \to 0$.

In fact this step is where the hardness of the problem appears because $I_1, \dots, I_n$ are not independent. Consequently, we must use the following formula for the variance of the sum of non-independent identically distributed random variables:

$$Var[X_n] = Var[\sum_{i=1}^{n} I_i] = n \cdot Var[I_1] + n \cdot (n-1) \cdot Cov(I_1, I_2). \tag{5.8}$$

We remember here the definition of the covariance:

$$Cov(I_1, I_2) = \mathbf{E}[I_1 I_2] - \mathbf{E}[I_1]\mathbf{E}[I_2]. \tag{5.9}$$

So now we compute all the necessary in (5.8). First, since I_1 is a Bernoulli random variable, i.e., $I_1 \sim Bernoulli(q(n))$:

$$Var[I_1] = q(n)(1 - q(n)) \tag{5.10}$$

Second, note that $\mathbf{E}[I_1 I_2] = p(I_1 \cap I_2) = p(\text{vertex 1 and vertex 2 are isolated})$ and to have two isolated vertices there must be $2n - 3$ absent edges:

$$\mathbf{E}[I_1 I_2] = p(I_1 \cap I_2) = (1 - p(n))^{2n-3} = \frac{\left((1 - p(n))^{n-1}\right)^2}{1 - p(n)} = \frac{q(n)^2}{1 - p(n)}. \tag{5.11}$$

Therefore we can compute $Cov(I_1, I_2)$ plugging (5.6) and (5.11) in (5.9):

$$Cov(I_1, I_2) = \frac{q(n)^2}{1 - p(n)} - q(n)^2 = \frac{p(n)q(n)^2}{1 - p(n)}. \tag{5.12}$$

Finally we can plug the results from (5.10) and (5.12) in the definition of the variance given in (5.8) to compute the ratio in which depends Second Order Method 5.10:

$$\frac{Var[X_n]}{\mathbf{E}[X_n]^2} = \frac{nq(n)(1 - q(n)) + n(n-1)\frac{p(n)q(n)^2}{1-p(n)}}{n^2 q(n)^2} = \frac{1 - q(n)}{nq(n)} + \frac{n-1}{n}\frac{p(n)}{1 - p(n)}.$$

First term of the sum tends to 0 as $n \to \infty$ because we have seen previously that $nq(n) = \mathbf{E}[X_n] \to \infty$. Second term of the sum also tends to 0 as $n \to \infty$ because $p(n) \to 0$. Consequently, by Second Moment Method, $X_n > 0$ with high probability so $\mathbb{G}_{n,p}$ is not connected w.h.p..

Next, let $\lambda > 1$. This statement will be approached without First Moment or Second Moment Methods. The key idea is that a graph is disconnected if and only if there exists a set of k vertices $k \in \{1, \dots, \lfloor n/2 \rfloor\}$, such that there is no edge connecting these k vertices with the other $n - k$ vertices in the graph. Note that this idea does not only treat the case where we have an isolated vertex since it also treats the case of having some connected subgraphs but a non-connected graph. In short, we study all the possible cases of a non-connected graph.

$$
\begin{aligned}
p(\mathbb{G}_{n,p} \text{ is disconnected}) &= p\Big(\bigcup_{k=1}^{\lfloor n/2 \rfloor} \{\text{some set of } k \text{ vertices is disconnected}\}\Big) \\
&\leq \sum_{k=1}^{\lfloor n/2 \rfloor} p(\{\text{some set of } k \text{ vertices is disconnected}\}) \\
&\leq \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} p(\{\text{a specific set of } k \text{ vertices is disconnected}\}) \\
&= \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} [1-p(n)]^{k(n-k)}.
\end{aligned}
$$

First inequality can be directly seen by taking all positive terms of principle of inclusion-exclusion. Second inequality is the result of computing all possible k vertices being disconnected from the other $n-k$ remaining. The equality is straight forward by the definition of having k vertices that will not have any edge with the other $n-k$ vertices.

To finish the proof it only remains to show that the expression obtained in the equation above tends to 0 as $n \to \infty$ because $\lambda > 1$. This calculations are overlong and use non-trivial bounds thus we will not write them here. However you can take a look in [19, p.5] too see the complete proof. $\square$

All in all, Theorem 5.13, states that $p^* = p^*(n) = \ln(n)/n$ is a sharp threshold for connectivity. Observe that if we move lightly around this threshold (getting λ close to 1 but assuring that $\lambda > 1$ or rather $\lambda < 1$), we can pass from not having a connected graph to having a connected graph with high probability.

5.5 Margulis-Russo Formula

An understandable question that the reader may have is why we talk about these thresholds on this thesis about Boolean functions. In this section we will relate random graphs, graph properties, thresholds and boolean functions following the chapter 8 of O'Donnell's book [1]. Additionally, we present Margulis-Russo Formula 5.23 which tells how "steep" is the threshold in terms of the Fourier coefficients of some Boolean function.

Until now we have worked with Boolean functions $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ where a random input $\mathbf{x}$ in $\{-1,1\}^n$ has each bit x_i independently chosen from $\{-1,1\}$ uniformly random. However, now we will work with $p-$ *biased* hypercubes. In this space each bit follows the probability distribution on $\{-1,1\}$:

$$
\begin{cases} \pi_p(-1) = p, \\ \pi_p(+1) = q = 1-p. \end{cases} \tag{5.13}
$$

In fact, for every probability distribution π_p, we are able to define an *inner product space of functions* $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ with the inner product

$$
\langle f, g \rangle_p = \mathop{\mathbf{E}}_{\mathbf{x} \sim \pi_p^{\otimes n}} [f(\mathbf{x}) \cdot g(\mathbf{x})],
$$

where $\pi_p^{\otimes n}$ denotes the product probability distribution (5.13) on $\{-1,1\}^n$. We will denote this inner product space by $L^2(\{-1,1\}^n, \pi_p^{\otimes n})$. The first difference that appears when using this more generalized space is that the expectation of every bit is not necessary equal to 0:

$$\mu = \mathop{\mathbf{E}}_{x_i \sim \pi_p}[x_i] = p(x_i = +1) - p(x_i = -1) = q - p = 1 - 2p.$$

Next thing we can search is a basis for the space $L^2(\{-1,1\}^n, \pi_p^{\otimes n})$. We have talked in the beginning of Section 1.3 why parity functions are a basis for $L^2(\{-1,1\}^n, \pi_{1/2}^{\otimes n})$. However, we can find another basis ϕ_0, ϕ_1 for the inner space $L^2(\{-1,1\}, \pi_p)$. Once this basis is found, we identify every subset $S \subseteq [n]$ with its $0-1$ indicator vector and write

$$\psi_S(\mathbf{x}) = \prod_{i=1}^{n} \phi_{S_i}(x_i).$$

Then, the set of all products $\phi_{i_1}, \dots, \phi_{i_n}$ forms a basis for the inner product space $L^2(\{-1,1\}^n, \pi_p^{\otimes n})$.

First, observe that it is convenient that the basis ϕ_0, ϕ_1 for $L^2(\{-1,1\}, \pi_p)$ is orthonormal because then it will be orthonormal for $L^2(\{-1,1\}^n, \pi_p^{\otimes n})$ and this makes Parseval's Theorem 1.11 and Plancherel's Theorem 1.12 hold. We can reach orthonormality by defining an orthonormal basis or by defining some basis and submitting it later to the Gramm-Schmidt algorithm.

Second, it is also useful for the basis to contain the constant function $\phi_0 = 1$. Hence,

$$0 = \langle \phi_0, \phi_1 \rangle_p = \langle 1, \phi_1 \rangle_p = \mathop{\mathbf{E}}_{x \sim \pi_p}[\phi_1(x)].$$

Thus, we can obtain the expectation, variance and covariance of f in terms of its Fourier coefficients because Proposition 1.18 holds.

Now, in the context of p-biased Fourier analysis, we define the following basis. We can interpret this basis as the normalization of every bit.

Definition 5.14. We define the *basis function* $\phi : \{-1,1\} \to \mathbb{R}$ by:

$$\phi(x_i) = \frac{x_i - \mu}{\sigma}.$$

Where,

$$\mu = \mathop{\mathbf{E}}_{x_i \sim \pi_p}[x_i] = q - p = 1 - 2p,$$

$$\sigma = \sqrt{\mathop{Var}_{x_i \sim \pi_p}[x_i]} = \sqrt{1 - (1-2p)^2} = \sqrt{4p(1-p)} = 2\sqrt{pq}.$$

Remark 5.15. Note that $\sigma^2 = 1 - \mu^2$ and we can compute directly the following results:

$$\phi(+1) = \frac{1-\mu}{\sigma} = \frac{2p}{2\sqrt{pq}} = \sqrt{\frac{p}{q}}.$$

$$\phi(-1) = \frac{-1-\mu}{\sigma} = \frac{-2-2p}{2\sqrt{pq}} = \frac{-2q}{2\sqrt{pq}} = -\sqrt{\frac{q}{p}}.$$

It is clear that $\{1, \phi\}$ is indeed a Fourier basis for $L^2(\{-1,1\}, \pi_p)$ because is orthonormal and contains the constant function 1. Now we can extend this basis to a product space:

Definition 5.16. In the context of $L^2(\{-1,1\}^n, \pi_p^{\otimes n})$ we define the *product Fourier basis functions* $(\phi_S)_{S\subseteq[n]}$ by:

$$\phi_S(\mathbf{x}) = \prod_{i\in S} \phi(x_i).$$

Given $f \in L^2(\{-1,1\}^n, \pi_p^{\otimes n})$ we will use the same notation for the Fourier coefficients given before in Theorem 1.5. Thus we have the biased Fourier expansion

$$f(\mathbf{x}) = \sum_{S\subseteq[n]} \hat{f}(S)\phi_S(\mathbf{x}).$$

Remark 5.17. By the definition of the basis function given in 5.14 we can go from the usual Fourier expansion (given in Theorem 1.5) to the biased Fourier expansion simply by plugging $\phi(x_i)$ in terms of x_i on the later. This is because $x_i = \mu + \sigma\phi(x_i)$ and usual Fourier expansion comes in terms of parity functions which are products of bits x_i.

Example 5.18. The simplest example of this conversion is given by the dictator function $\psi_i(\mathbf{x}) = x_i$. The Fourier coefficients of the usual Fourier expansion are equal to 0 except $\hat{\psi}(\{i\})$ which is equal to 1. However, if we write this function in terms of the basis function we have $\psi_i(\mathbf{x}) = \mu + \sigma\phi(x_i)$. Consequently, the Fourier coefficients are now:

$$\hat{\psi}_i(S) = \begin{cases} \mu & \text{if } S = \emptyset, \\ \sigma & \text{if } S = \{i\}, \\ 0 & \text{otherwise.} \end{cases}$$

Notation 5.19. Let $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ be any Boolean function and let $p \in (0,1)$. With the tools that we have presented we can now study this function with Fourier analysis no matter what Hamming Cube is. So we will write $f^{(p)}$ for the function when it is viewed as an element of $L^2(\{-1,1\}^n, \pi_p^{\otimes n})$.

Next, we would like to define a derivative operator D_i on $L^2(\{-1,1\}^n, \pi_p^{\otimes n})$ that acts like the one presented in Definition 2.17. In Section 2 we have differentiated with respect to x_i (parity functions) but now we would like to differentiate with respect to $\phi(x_i)$. This can be done with basic calculus:

$$\frac{\partial}{\partial\phi_i} = \frac{\partial x_i}{\partial\phi_i}\frac{\partial}{\partial x_i} = \sigma\frac{\partial}{\partial x_i}$$

First equality is given by the Chain rule. The second is the result from differentiating with respect to ϕ_i the expression of x_i in terms of the basis function. Therefore, recognizing the partial derivative with respect to x_i as the usual *ith* derivative operator D_i we are led to the following:

Definition 5.20. For i in $[n]$, the *ith (discrete) derivative operator* on $L^2(\{-1,1\}^n, \pi_p^{\otimes n})$ is defined by:

$$D_i f(\mathbf{x}) = \sigma \cdot \frac{f(\mathbf{x}^{(i\to 1)}) - f(\mathbf{x}^{(i\to -1)})}{2}.$$

It is important to keep in mind that the last definition depends on σ so the ith derivative operator depends always on p and we will have a different derivative operator in every space $L^2(\{-1,1\}^n, \pi_p^{\otimes n})$ for every $0 \leq p \leq 1$. Now we can introduce the influence in these spaces.

Proposition 5.21. *Suppose $f \in L^2(\{-1,1\}^n, \pi_p^{\otimes n})$ is a boolean-valued function. Then*

$$\mathbf{Inf}_i[f] = \sigma^2 \Pr_{\mathbf{x}\sim\pi_p^{\otimes n}}[f(\mathbf{x}) \neq f(\mathbf{x}^{\oplus i})].$$

Furthermore, if f is monotone then $\mathbf{Inf}_i[f] = \sigma \hat{f}(i)$.

Proof. For this proof we will use D_{x_i} to denote the derivative operator from Definition 2.17 and D_{ϕ_i} to denote the derivative operator from Definition 5.20. Recall that $D_{\phi_i} = \sigma D_{x_i}$. Then

$$\mathbf{Inf}_i[f] = \mathop{\mathbf{E}}_{\pi_p}[(D_{\phi_i} f)^2] = \sigma^2 \mathop{\mathbf{E}}_{\pi_p}[(D_{x_i})^2] = \sigma^2 \cdot \Pr_{\mathbf{x}\sim\pi_p^{\otimes n}}[f(\mathbf{x}) \neq f(\mathbf{x}^{\oplus i})].$$

Moreover, if f is monotone, by definition of the derivative operator D_{x_i}, we have that $(D_{x_i})^2 = D_{x_i}$ because it is equal to 0 or either 1. Hence

$$\mathbf{Inf}_i[f] = \sigma^2 \mathop{\mathbf{E}}_{\pi_p}[(D_{x_i})^2] = \sigma^2 \mathop{\mathbf{E}}_{\pi_p}[D_{x_i}] = \sigma \mathop{\mathbf{E}}_{\pi_p}[D_{\phi_i}] = \sigma \cdot \hat{f}(i).$$

Last equality comes from the fact that the operator D_i satisfies $D_i f = \sum_{S \ni i} \hat{f}(S)\phi_{S\setminus\{i\}}$ (see Proposition 2.19) and $\mathbf{E}[\phi_S] = 0$ for every set S except for the empty set $S = \emptyset$. $\square$

Finally let us relate everything we have presented about Boolean functions with random graphs. For every undirected graph G with n vertices we identify it with the string in $\{\text{True,False}\}^{\binom{n}{2}}$ that indicates which edges are present. We write $\mathbb{G}_{n,p}$ for the distribution $\pi_p^{\otimes\binom{n}{2}}$. Note that if we permute the n vertices of a graph this induces a permutation on the $\binom{n}{2}$ edges.

Every graph property can be represented as a Boolean function $f : \{\text{True,False}\}^{\binom{n}{2}} \to \{\text{True,False}\}$. Thus, given a graph identified with some string in $\{\text{True,False}\}^{\binom{n}{2}}$ this function outputs True if it satisfies the graph property or False otherwise. However this function must not depend on the names of the vertices. Hence, the Boolean function must be invariant under all $n!$ permutations of its input and therefore it is a transitive-symmetric function.

Example 5.22. The following Boolean functions represent all n-vertex graph properties:

- $Conn(G)$ = True if G is connected.
- $Maj_n(G)$ = True if G has at least $\binom{n}{2}/2$ edges.
- $\chi_{[n]}(G)$ = True if G has an odd number of edges.
- $Tri(G)$ = True if G has at least one triangle.

Each of these properties defines a family of Boolean functions, one for each value of n. However, note that this is only notation. We can always think that there exists a Boolean function for every number of vertices n that represents a graph property although it is not always possible or easy to find an explicit Boolean function for this property. For example, identifying True with -1 and False with $+1$, we can define $\chi_{[n]} = \prod_{i=1}^{\binom{n}{2}} x_i$. On the other hand, it is not that easy to define an explicit function for $Conn(G)$ or $Tri(G)$.

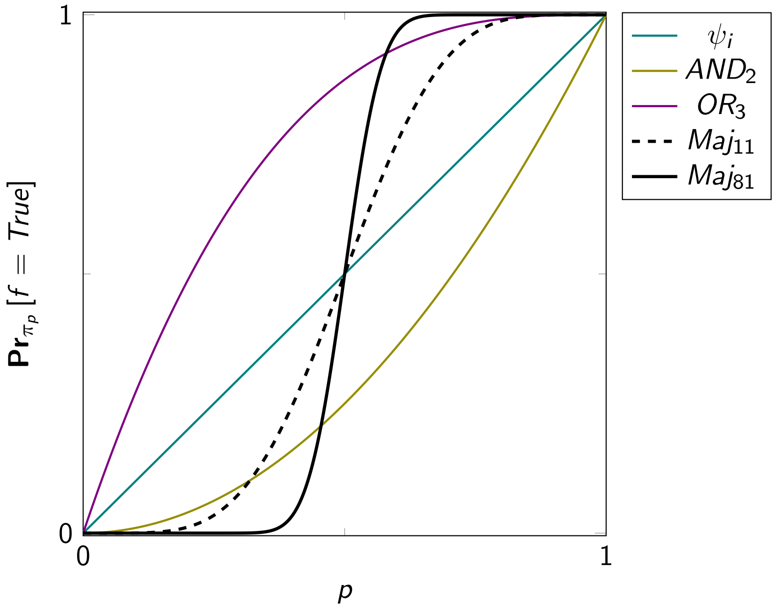

Figure 2: Plot of $\mathbf{Pr}_{\pi_p}[f = True]$ versus p on ψ_i,AND_2,OR_3,Maj_{11} and Maj_{81}.

Graph properties which are monotone are nice to study. A typical question is: "How many edges does a graph need to have before it is likely to satisfy the graph property?". It is intuitively clear that, since p is the probability of having an edge, when p increases from 0 to 1 this causes $\mathbf{Pr}_{\pi_p}[\text{Graph property}(\mathbf{x}) = \text{True}]$ to increase from 0 to 1.

In Figure 2 we plot some examples to sketch how the probability increases in some of the Boolean functions that have been presented in Section 2. The most outstanding function is Maj_n. Observe that the probability increases from 0 to 1 steeper as n increases. This outstanding behaviour is the usual threshold that appears on every property in a random graph.

The Margulis-Russo Formula quantifies how increases the probability of having a graph property in terms of p. Specifically, it relates the slope of the curve $\mathbf{Pr}_{\pi_p}[\text{Graph property}(\mathbf{x}) = \text{True}]$ to the total influence of f under $\pi_p^{\otimes n}$. To prove this formula we switch to the Hamming cube $\{-1,1\}^n$ identifying True with -1 and False with $+1$.

Theorem 5.23 (Margulis-Russo Formula)**.** *Let $f : \{-1,1\}^n \longrightarrow \mathbb{R}$. Working with the relation $\mu = 1 - 2p$ we have:*

$$\frac{d}{d\mu}\mathbf{E}[f^{(p)}] = \frac{1}{\sigma}\sum_{i=1}^{n}\widehat{f^{(p)}}(i).$$

In particular, if $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ is monotone, then:

$$\frac{d}{dp}\mathop{\mathbf{Pr}}_{\mathbf{x}\sim\pi_p^{\otimes n}}[f(\mathbf{x}) = -1] = \frac{d}{d\mu}\mathbf{E}[f^{(p)}] = \frac{1}{\sigma^2}\mathbf{I}[f^{(p)}].$$

Before we can start the proof we need a simple lemma to obtain the expectation of a Boolean function in a new way.

Lemma 5.24. *Let $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ have Fourier expansion $f(\mathbf{x}) = \sum_{S\subseteq[n]} \hat{f}(S)\mathbf{x}^S$. Let $F : \mathbb{R}^n \to \mathbb{R}$ be the extension of f which is also defined by $F(\mathbf{x}) = \sum_{S\subseteq[n]} \hat{f}(S)\mathbf{x}^S$. If $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in [-1,1]^n$ then*

$$F(\boldsymbol{\mu}) = \mathop{\mathbf{E}}_{\mathbf{y}}[f(\mathbf{y})].$$

Proof. (Lemma 5.24). Here we will use the linearity of the expectation and the Fourier expansion of f:

$$\mathop{\mathbf{E}}_{\mathbf{y}}[f(\mathbf{y})] = \mathop{\mathbf{E}}_{\mathbf{y}}\Big[\sum_{S\subseteq[n]} \hat{f}(S)\mathbf{y}^S\Big] = \sum_{S\subseteq[n]} \hat{f}(S)\prod_{i\in S}\mathbf{E}[y_i] = \sum_{S\subseteq[n]} \hat{f}(S)\prod_{i\in S}\mu_i = \sum_{S\subseteq[n]} \hat{f}(S)\boldsymbol{\mu}^S = F(\boldsymbol{\mu}).$$

□

Proof. (Margulis-Russo Formula 5.23). Treating f as a multilinear polynomial over $x_1, \dots, x_n$ we have by Lemma 5.24 that

$$\mathbf{E}[f^{(p)}] = f(\mu, \dots, \mu).$$

Applying some basic calculus rules we can differentiate the expression above with respect to μ:

$$\frac{d}{d\mu} f(\mu, \dots, \mu) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial \mu}\bigg|_{(\mu,\dots,\mu)} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\bigg|_{(\mu,\dots,\mu)} = \sum_{i=1}^{n} D_{x_i} f(\mu, \dots, \mu).$$

However, by Lemma 5.24 again,

$$D_{x_i} f(\mu, \dots, \mu) = \mathbf{E}[D_{x_i} f^{(p)}] = \frac{1}{\sigma}\mathbf{E}[D_{\phi_i} f^{(p)}] = \frac{1}{\sigma}\hat{f}^{(p)}(i).$$

Where the second inequality comes from the definition of the derivative operator 5.20 $D_{\phi_i} = \sigma D_{x_i}$. Last equality is due to the monotonicity of f which derives that $D_{\phi_i} = (D_{\phi_i})^2$. This completes the proof for the general statement.

$$\frac{d}{d\mu}\mathbf{E}[f^{(p)}] = \sum_{i=1}^{n} D_{x_i} f(\mu, \dots, \mu) = \frac{1}{\sigma}\sum_{i=1}^{n} \widehat{f^{(p)}}(i).$$

Next, we give the proof for the statement given for the concrete case of monotone boolean-valued functions. First equality holds because $\mu = 1 - 2p$ and $\mathbf{E}[f] = 1 - 2\mathbf{Pr}[f = -1]$. Then,

$$\begin{aligned}
\frac{d}{d\mu}\Big[\mathbf{E}[f^{(p)}]\Big] &= \frac{dp}{d\mu}\frac{d}{dp}\Big[1 - 2\mathop{\mathbf{Pr}}_{\mathbf{x}\sim\pi_p}[f^{(p)}(\mathbf{x}) = -1]\Big] \\
&= -\frac{1}{2}\frac{d}{dp}\Big[-2\mathop{\mathbf{Pr}}_{\mathbf{x}\sim\pi_p}[f^{(p)}(\mathbf{x}) = -1]\Big] \\
&= \frac{d}{dp}\Big(\mathop{\mathbf{Pr}}_{\mathbf{x}\sim\pi_p}[f^{(p)}(\mathbf{x}) = -1]\Big).
\end{aligned}$$

Since the Boolean function is monotone we recall now Proposition 5.21 . We plug the last result above into the general statement that has been already proved and then we prove the second statement of this proposition.

$$\frac{d}{dp}\left(\Pr_{\mathbf{x}\sim\pi_p}[f^{(p)}(\mathbf{x})=-1]\right)=\frac{d}{d\mu}\left[\mathbf{E}[f^{(p)}]\right]=\frac{1}{\sigma}\sum_{i=1}^{n}\hat{f^{(p)}}(i)=\frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{Inf}_i[f^{(p)}]=\frac{1}{\sigma^2}\mathbf{I}[f^{(p)}].$$

□

In fact, Margulis-Russo Formula says that if we want sharp thresholds we need to find functions with high influence. This influence will give the slope of the function $\mathbf{Pr}_{\mathbf{x}\sim\pi_p}[f^{(p)}(\mathbf{x})=-1]$ and therefore we will have a threshold or sharp threshold depending on how steep is the function around the threshold.

The study of threshold phenomena is a well studied topic in mathematics. There exist more sophisticated theorems that generalize how can we find them. In fact, Friedgut and Kalai showed in [20] that any monotone graph property has a sharp threshold. Later Bourgain proved in [21] a weaker characterization for random graphs that works in every product probability space and thus does not need monotonicity. These two cited papers are a nice way to delve into the subject but we will not present the theorems here because they are out of reach of this thesis.

6. Conclusions

In this thesis we have first presented the basic definitions on Boolean functions. The most important result of the first chapter is the Fourier expansion, not also because it gives an alternative form to write any function, but because many results on Boolean functions can be given by Fourier coefficients. These tools have allowed us to study applications of Boolean functions in diverse scientific research areas in terms of this Fourier analysis.

First application shows how to relate Social Choice Theory and Boolean functions. In fact, we have seen that habitual voting systems can be written as a Boolean function. Moreover, we have defined mathematical operators for Boolean functions that also have a natural interpretation in politics. In addition, we have read the paper where Arrow presented his theorem and proof. This theorem has a really nice interpretation since it says that we can never have a fair and ethical voting system unless the method to decide a winner is a dictatorship or the winner is imposed, thus we will never have a fair and ethical voting system. We have rewritten the proof given by Arrow for this theorem but in terms of Boolean functions and Set Theory.

Furthermore, we have delved into hypercontractivity to get stronger results for Boolean functions rather than with usual real analysis in order to present more advanced applications in Social Choice Theory. On the one hand, we have proved that Arrow's theorem is stable. On the other hand, we have introduced KKL theorem to show that in every voting system it is not possible to avoid that all voters will have a minimum influence, which expression is given in terms of the total number of voters. Moreover, we have showed more extensions of this theorem to obtain the proof of the original one.

The second application has been focused differently, since we do not have information of this subject in O'Donnell's reference book. Thus, we have first read and understood the paper published by Huang at Annals of Mathematics and we have looked at the cited papers by him referring to the origins of the Sensitivity Conjecture. Consequently, one of the main difficulties of this section has been to completely understand various mathematical papers. Nevertheless, we have been able to write a chapter where it is possible to follow the proofs timeline from basic Complexity Theory measures equivalences to the Sensitivity Conjecture.

The third application has been an example of Boolean functions in Combinatorics and Probability Theory. The strange phenomena of thresholds on properties for random graphs is illustratable but not analytically treatable. However, we have written all Random Graphs definitions in terms of Boolean functions and moved to this area with the certainty that we have Margulis-Russo Formula to study thresholds with Fourier analysis on Boolean functions. Notwithstanding, there exist further results from Friedgut, Bourgain and Kalai that deepen the study of this subject.

Hence, a natural way to continue this work on Boolean functions is to take a look at the further work given by Friedgut, Bourgain and Kalai with the aim of having a broader view for thresholds on properties for Random Graphs. Alternatively, another option is to delve into other applications of Boolean functions in scientific research areas.

References

[1] R. O'Donnell, *Analysis of Boolean functions*. Cambridge University Press, New York, 2014.

[2] K. J. Arrow, "A Difficulty in the Concept of Social Welfare," *Journal of Political Economy*, vol. 58, no. 4, pp. 328–346, 1950.

[3] G. Kalai, "A Fourier-theoretic perspective on the Condorcet paradox and Arrow's theorem," *Advances in Applied Mathematics*, vol. 29, no. 3, pp. 412–426, 2002.

[4] K. O. May, "A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision," *Econometrica*, vol. 20, no. 4, pp. 680–684, 1952.

[5] A. Bonami, "Étude des coefficients de Fourier des fonctions de $L^p(G)$," *Ann. Inst. Fourier (Grenoble)*, vol. 20, no. fasc. 2, pp. 335–402, 1970.

[6] E. Friedgut, G. Kalai, and A. Naor, "Boolean functions whose Fourier transform is concentrated on the first two levels," *Advances in Applied Mathematics*, vol. 29, no. 3, pp. 427–437, 2002.

[7] M. Ben-Or and N. Linial, "Collective Coin Flipping," *Randomness and Computation (S. Micali ed.) Academic Press*, vol. 5, pp. 91–115, 1989.

[8] J. Kahn, G. Kalai, and N. Linial, "The influence of variables on boolean functions," in *29th Annual Symposium on Foundations of Computer Science*, 1988, pp. 68–80.

[9] N. Nisan and M. Szegedy, "On the degree of Boolean functions as real polynomials," *Computational Complexity*, vol. 4, no. 4, pp. 301–313, 1994, special issue on circuit complexity (Barbados, 1992).

[10] H. Huang, "Induced subgraphs of hypercubes and a proof of the Sensitivity Conjecture," *Annals of Mathematics. Second Series*, vol. 190, no. 3, pp. 949–955, 2019.

[11] C. Gotsman and N. Linial, "The equivalence of two problems on the cube," *Journal of Combinatorial Theory, Series A*, vol. 61, no. 1, pp. 142–146, 1992.

[12] N. Nisan, "CREW PRAMs and decision trees," *SIAM Journal on Computing*, vol. 20, no. 6, pp. 999–1007, 1991.

[13] S. Fisk, "A very short proof of Cauchy's interlace theorem for eigenvalues of Hermitian matrices," *The American Mathematical Monthly*, vol. 112, no. 2, p. 118, 2005.

[14] P. Erdős and A. Rényi, "On the evolution of random graphs," *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, vol. 5, pp. 17–61, 1960.

[15] E. N. Gilbert, "Random graphs," *Annals of Mathematical Statistics*, vol. 30, pp. 1141–1144, 1959.

[16] A. Frieze and M. Karoński, "Introduction to Random Graphs." Cambridge University Press, 2015, pp. 5–8.

[17] B. Bollobás and A. Thomason, "Threshold functions," *Combinatorica. An International Journal of the János Bolyai Mathematical Society*, vol. 7, no. 1, pp. 35–38, 1987.

[18] A. S. Novozhilov, "Chapter 3: Erdős-Rényi random graphs," lecture notes, Mathematics of Networks, North Dakota State University, delivered Summer 2015.

[19] K. Ramchandra and A. Parekh, "Random Graphs," lecture notes, Probability and Random Processes, University of California, Berkeley, delivered Fall 2018.

[20] E. Friedgut and G. Kalai, "Every monotone graph property has a sharp threshold," *Proceedings of the American Mathematical Society*, vol. 124, no. 10, pp. 2993–3002, 1996.

[21] J. Bourgain, "On sharp thresholds of monotone properties," *Journal of the American Mathematical Society*, vol. 4, no. 12, p. 1046–1053, 1999, Appendix to the main paper, *Sharp thresholds of graph properties, and the k-sat problem* by Ehud Friedgut.