The Harmonic Analysis of Boolean Functions

Jacob Denson

January 27, 2017

Table Of Contents

Chapter 1

Introduction to Boolean Analysis

This course is about the harmonic analysis of functions on the abelian group $\mathbf{F}_2^n$, and the surprising applications of the field to computing science. It seems that whenever one wants to deal with a combinatorial problem involving the Hamming distance, which takes two strings x and y, and counts the number of indices for which $x_i \neq y_i$,

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

For these problems, harmonic analysis over the additive group $\mathbf{F}_2^n$ seems to be the right tool for the job. It turns out that measuring the similarity of Fourier coeffients of a function tends to give good information about the hamming distance between two functions.

   If you've ever had to write a program, you will probably have made the intuitive correspondence between $\mathbf{F}_2$ and boolean truth values

$$0 \mapsto \bot \quad 1 \mapsto \top$$

The group structure on $\mathbf{F}_2^n$ then induces a group structure on $\{\top, \bot\}$, where addition corresponds to the xor operation $\oplus$, defined by

$$A \oplus B = \begin{cases} \top & A \text{ is } \top, \text{ or } B \text{ is } \top, \text{ but not both.} \\ \bot & \text{otherwise} \end{cases}$$

and multiplication corresponds to the logical operation $\wedge$ of conjunction. Already we see interactions with the operations and the hamming distances, for if X and Y are truth values, $X \oplus Y = \top$ if and only if $X \neq Y$,

so that the output of xor is intimately connected to the Hamming distance between two strings.

We shall also consider a less intuitive, but incredibly important correspondence between the additive structure of $\mathbf{F}_2$ and the multiplicative group $\{-1, 1\} = \{-1, 1\}$ of 2nd roots of unity.bo

$$0 \mapsto 1 \quad 1 \mapsto -1$$

This is initially a very strange correspondence to pick, since this means that $\top$ corresponds to -1, and $\bot$ to 1, and we normally think of 1 as the 'true' truth value. However, the correspondence does give an isomorphism between the two group structures, and fits with the general convention of performing harmonic analysis over the multiplicative= group $\mathbf{T}^n$ where possible. Furthermore, since harmonic analysis considers the representations of functions valued in real or complex numbers by characters, we require a natural representation of $\mathbf{F}_2$ in the complex numbers in order to apply harmonic analysis to functions $f : \mathbf{F}_2^n \to \mathbf{F}_2$.

The characters on the circle group $\mathbf{T}$ take the form of powers z^m, for $m \in \mathbf{Z}$. On $\mathbf{T}^n$, the characters are therefore monomials $z_1^{m_1} \ldots z_n^{m_n}$. The embedding $\mu_2^n \to \mathbf{T}^n$ gives rise to a surjective reduction map $(\mathbf{T}^n)^* \to (\{-1, 1\}^n)^*$, so that every character on $\{-1, 1\}$ can be represented by a monomial, and since $x_i^2 = 1$ for all $x_i \in \mathbf{F}_2$, these monomials can be chosen with $m_i \in \{0, 1\}$. Abstract Harmonic analysis tells us that there are exactly 2^n characters on the group $\mathbf{F}_2^n$, so these are all the identifications we can make. Since this is equivalent to identifying the monomial with the subset S of $[n]$ upon which m_i takes the value one, we shall often let x^S stand for this monomial $\prod_{i \in S} z_i$. Analogously, we define the characters $\chi_S : \mathbf{F}_2^n \to \{-1, 1\}$ by

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

The basic theory of discrete abelian harmonic analysis tells us that the characters x^S form an orthonormal basis for the set of all functions $f : \mathbf{F}_2^n \to \mathbf{R}$, so we have a unique expansion

$$f(x) = \sum_{S \subset [n]} \widehat{f}(S) x^S$$

where we call $\widehat{f}(S) \in \mathbf{R}$ is the Fourier coefficient corresponding to S. An easy way to find these coeffients in the case of $\mathbf{F}_2$, without any harmonic

3

analysis, is to expand the 'conjunctive normal form' of the function

$$f(x) = \sum_y \mathbf{I}(x = y) f(y) = \frac{1}{2^n} \sum_y \left(\prod_{i=1}^n 1 + x_i y_i \right) f(y)$$

remembering that $x_i^2 = 1$.

Example. *The maximum function* $\max(x, y)$ *on* $\{-1, 1\}^2$ *has a Fourier expansion*

$$\max(x, y) = \frac{1}{2}(1 + x + y - xy)$$

which corresponds to the the conjunction function $f(x, y) = x \wedge y$ *on* $\{\top, \bot\}^2$, *or the minimum function on* $\mathbf{F}_2^2$. *To obtain this expansion, and a more general expansion for the maximum function* $\max_n$ *on* n *variables note that we can write*

$$\mathbf{I}(x_1 = -1, \ldots, x_n = -1) = \mathbf{I}(x = -1) = \frac{1}{2^n} \prod_{i=1}^n (1 - x_i) = \frac{1}{2^n} \sum_S (-1)^{|S|} x^S$$

so that

$$\begin{aligned}
\max(x_1, \ldots, x_n) &= \mathbf{I}(x \neq -1) - \mathbf{I}(x = -1) \\
&= 1 - 2\mathbf{I}(x = -1) \\
&= \left(1 - \frac{1}{2^{n-1}}\right) - \frac{1}{2^{n-1}} \sum_{S \neq \varnothing} (-1)^{|S|} x^S
\end{aligned}$$

The general approach of writing a function as a linear combination of functions with well known Fourier expansion is essentially a 'power series' method of Boolean expansions.

Example. *The majority function* $Maj_3(x, y, z)$ *on* $\{-1, 1\}^3$ *returns the boolean value which appears most often in* x, y, z. *Then*

$$Maj_3(x, y, z) = \frac{1}{2}(x + y + z - xyz)$$

Since we are counting elements of the domain, the majority function has the same interpretation on any of the other boolean domains.

Example. *Consider the minimum function* $\min(x) : \{-1,1\}^n \to \{-1,1\}$ *on n bits, which corresponds to the= disjunction operation on $\{\top,\bot\}^n$. Since we have the relationship* $\min(x) = -\max(-x)$, *we have* $\min = -\max \circ (\pi^1,\ldots,\pi^n)$, *where* $\pi \in S_2$ *is the permutation which swaps 1 with -1. In general, if $h = f \circ g$, where* $x^S \circ g = \sum c_{S'}^S x^{S'}$ *for some constants* $c_{S'}^S$, *and if we have expansions*

$$h(x) = \sum a_S x^S \quad f(x) = \sum b_S x^S$$

Then

$$(f \circ g)(x) = \sum b_S c_{S'}^S x^{S'}$$

so

$$a_S = \sum b_{S'} c_S^{S'}$$

In our case, if g is the negation operation $g(x_1,\ldots,x_n) = (-x_1,\ldots,-x_n)$, *then* $(x^i \circ g)(x) = -x^i$, *so* $(x^S \circ g) = (-1)^S x^S$ *and therefore*

$$\min(x) = -\left[\left(1 - \frac{1}{2^{n-1}}\right) - \frac{1}{2^{n-1}} \sum_{S \neq \varnothing} x^S\right] = \left(\frac{1}{2^{n-1}} - 1\right) + \frac{1}{2^{n-1}} \sum_{S \neq \varnothing} x^S$$

Note that the L^1 distances between the functions $\min$ *and* $\max$ *and constant functions tend to zero, so that the L^∞ distance of their Fourier transforms also tend to zero, and therefore the non-constant Fourier coefficients of* $\min$ *and* $\max$ *become very small over time (this becomes rigorous only when we view each $\{-1,1\}^n$ as a subset of the compact abelian group $\{-1,1\}^\infty = \prod_{i=1}^\infty \{-1,1\}$, with the same notion of L_1 distance and L_∞ distance).*

Example. *In general, the indicator functions* $\mathbf{I}(x = a) : \{-1,1\}^n \to \{0,1\}$, *for some* $a \in \{-1,1\}^n$ *can be expressed as*

$$\mathbf{I}(x = a) = \frac{1}{2^n} \prod_{i=1}^n (1 - x_i a_i) = \frac{1}{2^n} \sum_S (-1)^{|S|} \left(\prod_{i \in S} a_i\right) x^S$$

If we only have a $X = \{i_1,\ldots,i_n\}$ subset of indices we want to specify, then

$$\mathbf{I}(x_{i_1} = a_1,\ldots,x_{i_k} = a_k) = \frac{1}{2^n} \sum_S (-1)^{|S|} \left(\prod_{i \in S \cap X} a_i\right) x^S$$

5

Example. *Consider n $\{-1,1\}$-valued Bernoulli distributions with means $\mu_1,\ldots,\mu_n$, and consider the corresponding product distribution of $\{-1,1\}^n$, with density function f. If X is a random variable with this distribution, then*

$$\mathbf{P}(X_i = 1) - \mathbf{P}(X_i = -1) = 2\mathbf{P}(X_i = 1) - 1 = \mu_i$$

so $\mathbf{P}(X_i = 1) = \frac{1}{2}(\mu_i + 1) = P_i \in [0,1]$. Then we can write

$$f(x_1,\ldots,x_n) = \prod_{i=1}^n [P_i \mathbf{I}(x_i = 1) + (1 - P_i)\mathbf{I}(x_i = 1)]$$

$$= \sum_{a \in \mathbf{F}_2^n} \left(\prod_{a_i=1} P_i\right)\left(\prod_{a_i=0}(1 - P_i)\right)\mathbf{I}(x = a)$$

$$= \frac{1}{2^n}\sum_S (-1)^{|S|}\sum_{a \in \mathbf{F}_2^n}\left(\prod_{i \in S} a_i\right)\left(\prod_{a_i=1} P_i\right)\left(\prod_{a_i=0}(1 - P_i)\right)x^S$$

$$= \frac{1}{2^n}\sum_S (-1)^{|S|}\prod_{i \in S}(2P_i - 1)x^S = \frac{1}{2^n}\sum_S (-1)^{|S|}\left(\prod_{i \in S}\mu_i\right)x^S$$

where we obtain the last equation by repeatedly factoring out the coordinates a_i for $a_i = 1$ and $a_i = -1$.

Example. *Consider the 'inner product mod 2' function f on $\mathbf{F}_2^n \times \mathbf{F}_2^n$ defined by*

$$f(x,y) = (-1)^{\langle x,y\rangle} = \prod_{i=1}^n (-1)^{x_i y_i}$$

Then $f(x,y)$ measures the parity of the number of x_i which are equal to y_i. For $n = 1$, we have

$$f(x,y) = (-1)^{xy} = \begin{cases} -1 & x = y = 1 \\ 1 & otherwise \end{cases}$$

Hence, if we view f as a function on $\{-1,1\}^2$, the induced function is just the $\max$ function on two variables, and has a Fourier representation

$$f(x,y) = \frac{1}{2}(1 + x + y - xy)$$

*In general, any set S corresponding to a set of indices on the functions of f can
be associated with a unique pair of $S_1, S_2 \subset [n]$, corresponding to the indices
relating to x on the left side of the function, and the indices relating to y on the
right side of the equation. Let $\alpha(S)$ be the cardinality of $S_1 \cap S_2$. Then we have*

$$f(x,y) = \frac{1}{2^n} \prod_{i=1}^{n} (1 + x_i + y_i - x_i y_i)$$

$$= \frac{1}{2^n} \sum_{S \subset [n]} (-1)^{|S|} \left(\prod_{i \in S} x_i y_i \right) \left(\prod_{i \notin S} (1 + x_i + y_i) \right)$$

$$= \frac{1}{2^n} \sum_{S \subset [n]} (-1)^{|S|} \sum_{S' \subset S^c} \left(\prod_{i \in S} x_i y_i \right) \left(\prod_{i \in S'} (x_i + y_i) \right)$$

$$= \frac{1}{2^n} \sum_{S} (-1)^{\alpha(S)} x^S$$

1.1   Testing Linearity

A function $f : \mathbf{F}_2^n \to \mathbf{F}_2$ is linear if and only if $f(x + y) = f(x) + f(y)$ for all
$x, y \in \mathbf{F}_2^n$, or equivalently, if there is $a \in \mathbf{F}_2^n$ such that $f(x) = \langle a, x \rangle$. Given a
general function f, there are effectively only two ways to explicitly check
that the function is linear, we either check that $f(x+y) = f(x)+f(y)$ for all
choices of the arguments x and y, or check that $f(x) = \langle a, x \rangle$ holds for some
choice of a, over all choices of x. Even assuming that f can be evaluated
in constant time, both methods take exponential time in n to compute.
This is effectively guaranteed, because the description of f is always expo-
nential in n, and if an algorithm determining linearity does not check the
entire description of a linear function f, we can modify f to a non-linear
function g which is not linear, yet indistinguishable to f int he algorithm.
The problem of testing linearity is a method of family of problesm in the
field of **property testing**, which attempts to design efficient algorithms
to determine whether a particular boolean-valued function has a certain
property.

We might not be able to come up with a polynomial time algorithm to
verify linearity, but we can make headway by considering the possiblity
of coming up with an appropriate *randomized* algorithm which can verify
linearity with high probability. The likely solution to the problem, given

some function f, would to perform the linearity test for $f(X+Y) = f(X) + f(Y)$ for a certain set of randomly chosen inputs X and Y. If $f(X+Y) \neq f(X) + f(Y)$ for some particular input, we can guarantee the function is non-linear. Otherwise, we hope for= a high probability that the function is linear, in a manner which becomes more likely as we test more random inputs.

To proceed along these lines, we must introduce the probabilistic interpretation of the Fourier expansion. The Fourier coefficients of $f : \mathbf{F}_2^n \to \mathbf{R}$ can, of course, be calculated as the inner product

$$\widehat{f}(S) = \langle f, x^S \rangle = \frac{1}{2^n} \sum_x f(x) x^S$$

where the 2^n factor is included so that characters are normalized. We can also see this as a probabilistic statement, if we consider the uniform distribution on $\{-1, 1\}^n$, and consider the corresponding random variable X. Then we can write

$$\widehat{f}(S) = \mathbf{E}[f(X)X^S]$$

So that the Fourier coeffient measures the average value of f, relative to the parity over S. Note that this implies that a boolean-function $f : \{-1, 1\}^n \to \{-1, 1\}^n$ is *unbiased*, that is, it has an equal chance of taking -1 and 1, if and only if $\widehat{f}(\varnothing) = 0$. Of course, we have Parseval's equality

$$\mathbf{E}[f^2(X)] = \sum \widehat{f}^2(S)$$

and so

$$\mathbf{V}[f(X)] = \mathbf{E}[f^2(X)] - \mathbf{E}[f(X)]^2 = \sum_{S \neq \varnothing} \widehat{f}^2(S)$$

and similarily,

$$\mathrm{Cov}[f(X), g(X)] = \sum_{S \neq \varnothing} \widehat{f}(S)\widehat{g}(S)$$

So that the Fourier coefficients are efficient measures of probabilistic quantities.

If f and g are *boolean-valued* maps $\mathbf{F}_2^n \to \{-1, 1\}$, then

$$\mathbf{E}[f(X)g(X)] = \mathbf{P}(f(X) = g(X)) - \mathbf{P}(f(X) \neq g(X))$$
$$= 1 - 2\mathbf{P}(f(X) \neq g(X))$$

8

Note that $\mathbf{P}(f(X) \neq g(X))$ differs from the Hamming distance of f and g by a constant factor. We define $\mathbf{P}(f(X) \neq g(X))$ to be the relative hamming distance between f and g, denoted $d(f,g)$. Since $f^2 = 1$, this implies $\|f\|_2 = 1$, and therefore

$$
\begin{aligned}
\mathbf{V}(f) &= \mathbf{E}[f(X)^2] - \mathbf{E}[f(X)]^2 \\
&= 1 - \mathbf{E}[f(X)]^2 \\
&= 1 - (\mathbf{P}(f(X) = 1) - \mathbf{P}(f(X) = -1))^2 \\
&= 4\mathbf{P}(f(X) = -1)\mathbf{P}(f(X) = 1) \in [0,1]
\end{aligned}
$$

so the variance of a boolean-valued function is very closely related to the degree to which the function is constant.

Theorem 1.1. *If $\varepsilon = \min[d(f,1), d(f,-1)]$, then $2\varepsilon \leqslant \mathbf{V}(f) \leqslant 4\varepsilon$.*

Proof. We are effectively proving that for any $x \in [0,1]$, $2\min(x, 1-x) \leqslant 4x(1-x) \leqslant 4\min(x, 1-x)$. Since $4x(1-x) = 4\max(x, 1-x)\min(x, 1-x)$, we may divide by $4\min(x, 1-x)$ to restate the inequality as

$$
1/2 \leqslant \max(x, 1-x) \leqslant 1
$$

which is true, because $x, 1-x \leqslant 1$, and also

$$
\max(x, 1-x) \geqslant \frac{1}{2}[x + (1-x)] = 1/2
$$

This shows that if we have a sequence of functions $f_i : \{-1,1\}^n \to \{-1,1\}$, then $\varepsilon_i \sim x_i$ holds if and only if $\mathbf{V}(f) \sim x_i$. The minimum hamming distance to a constant value is asymptotically the same as the variation of the function. In general domains we intuitively view $\mathbf{V}(f)$ as a measure of how constant the function f is, but for functions $\{-1,1\}$ we have a quantitative estimate. $\square$

Since a function $f : \{-1,1\}^n \to \{-1,1\}$ always has square-norm 1, because $f^2 = 1$, Parsevel's theorem implies that $\sum \widehat{f}(S)^2 = 1$ if we sum over all the Fourier coeffients, so that a boolean-valued function on n variables gives rise to a probability distribution over $[n]$. We call this the spectral sample of f, and denote the distribution by $\mathcal{S}_f$.

Often times, the fourier coeffients of sets over some particular cardinality are more than enough to determine some result. We define the weight of a function $f : \{-1,1\}^n \to \mathbf{R}$ of degree k to be

$$W^k(f) = \sum_{|S|=k} \widehat{f}^2(S)$$

If f is boolean-valued, then we could have also defined

$$W^k(f) = \mathbf{P}(|S| = k)$$

where S is a set randomly drawn from the spectral sample of f. $W^k(f)$ is also the square norm of the function

$$(P^k f)(x) = \sum_{|S|=k} \widehat{f}(S)x^S$$

which can be seen as a certain truncation of f.

The simplest version of linearity testing runs using one random query as a test – it just takes one pair of inputs (X, Y), and tests the linearity of this function against these inputs. This is known as the Blum-Luby-Rosenfeld algorithm, or BLR for short. It turns out that the sucess of this method is directly related to how linear a function is. Define a function to be **approximately linear** if

- $f(x + y) = f(x) + f(y)$ for a large majority of inputs x and y.

- There is a such that $f(x) = \langle a, x \rangle$ for a large number of inputs x.

It is clear that the second property implies the first, but not that the first implies the second in a precise relationship. To be more precise, we say that f is ε-close to being linear if there is a linear function g with $d(f, g) \leqslant \varepsilon$. What the analysis of BLR shows is that $f(X + Y) = f(X) + f(Y)$ holds with probability ε, then f is ε close to a linear function. Since we are applying Hamming distances in measuring how linear a function is, we can guess that the Harmonic analysis of boolean functions will come in handy.

Theorem 1.2. *Given a function $f : \mathbf{F}_2^n \to \mathbf{F}_2$, the BLR algorithm is correct with probability $1 - \varepsilon$ if and only if f is ε-close to being linear.*

Proof. We switch to multiplicative notation, so that $f : \{-1,1\}^n \to \{-1,1\}$. Note that the probability that $f(XY) = f(X)f(Y)$ is the same as the probability that

$$\frac{1}{2}[1 + f(X)f(Y)f(XY)] = 1$$

and if $f(XY) \neq f(X)f(Y)$, then

$$\frac{1}{2}[1 + f(X)f(Y)f(XY)] = 0$$

so that the function $g(x,y) = (1/2)[1 + f(x)f(y)f(xy)]$ is 0-1 valued, and

$$
\begin{aligned}
1 - \varepsilon &= \mathbf{P}[\text{BLR accepts } f] \\
&= \mathbf{E}[g(X,Y)] \\
&= \frac{1}{2} + \frac{1}{2}\mathbf{E}_X[f(X)\mathbf{E}_Y[f(Y)f(XY)]] \\
&= \frac{1}{2} + \frac{1}{2}\mathbf{E}_X[f(X)(f * f)(X)] \\
&= \frac{1}{2} + \frac{1}{2}\sum \widehat{f}(S)^3
\end{aligned}
$$

and therefore

$$1 - 2\varepsilon = \sum \widehat{f}(S)^3 \leqslant \|\widehat{f}\|_\infty \|f\|_2 = \|\widehat{f}\|_\infty$$

Now $\widehat{f}(S) = 1 - 2d(f,x^S)$, and since there is S' with $\widehat{f}(S') \geqslant 1 - 2\varepsilon$, this implies $d(f,x^{S'}) \leqslant \varepsilon$, and therefore f is ε-close to the linear function $x^{S'}$. $\qquad\square$

Note that this algorithm, with only three evaluations of the function f, we can determine with high accuracy that f is not linear, or, if we are wrong with high probability, then f is very similar to a linear function in the first case. Yet given f, we cannot use this algorithm to determine the linear function x^S which f is similar to. Of course, for most x, $f(x) = x^S$. Yet we cannot use this estimate to obtain accurate estimates for a fixed x, in the sense that there is no measure of the likelihood of the estimate being equal to the actual value. Note, however, that $x^S(x) = x^S(x + y)x^S(y)$, and if we let y be a random quantity, then $x^S(x + y)$ and $x^S(y)$ will likely be equal to $f(x + y)$ and $f(y)$ with a probability that is feasible to determine.

Theorem 1.3. *If $f : \mathbf{F}_2^n \to \{-1, 1\}$ is ε-close to x^S, then for any y,*

$$\mathbf{P}_X[f(X + y)f(X) = x^S(y)] \geqslant 1 - 2\varepsilon$$

Proof. The probability that $f(X + y) \neq x^S(X + y)$ is less than ε, as is the probability that $f(X) \neq x^S(X)$. By the union bound, this implies that the probability of either occuring is less than or equal to 2ε. But then the probability that both do not occur is greater than or equal to $1 - 2\varepsilon$, and when this occurs, we can guarantee that $f(X + y)f(X) = x^S(y)$, thus our estimate $f(X + y)f(X)$ is equal to $x^S(y)$ with probability $\geqslant 1 - 2\varepsilon$. $\qquad\square$

Chapter 2

Social Choice Functions

In this chapter, we interpret boolean functions in the context of the theory of social choice. Here we think of a boolean function $f : \{-1,1\}^n \to \{-1,1\}$ as a voting rule; n people choose between two candidates, labelled -1 and 1, and the function f determines the overall outcome of the vote. In this interpretation it is natural to try and find a fair voting rule. This leads to natural specifications of properties of the Boolean functions, and we will find these properties naturally relate to harmonic analysis.

The standard voting rule, which we have already seen, is the Majority function $\mathrm{Maj}_n : \{-1,1\}^n \to \{-1,1\}$, for n odd, which choose the candidate with the majority of votes. But there are other voting rules which give us more examples of Boolean functions.

Example. *The voting rule $\mathrm{And}_n : \{-1,1\}^n \to \{-1,1\}$ corresponds to choosing candidate 1 unless every other person votes for candidate -1. The rule $\mathrm{Or}_n : \{-1,1\}^n \to \{-1,1\}$ votes for candidate -1 unless everyone else votes against him.*

Example. *The dictator voting rule $\chi_i : \{-1,1\}^n \to \{-1,1\}$ defined by $\chi_i(x) = x_i$, which places the power of decision making in a single person.*

Example. *All of these functions can be quantified as a version of the weight majority, or linear threshold function, those functions f which can be defined, for some $a_0, \ldots, a_n \in \mathbf{R}$, as*

$$f(x) = sgn(a_0 + a_1 x_1 + \cdots + a_n x_n)$$

Thus each voter has a certain voting power, specified by the a_i, and there is an additional bias a_0. The United Nations Council of Ministers accepts a decision

if and only if 55% of people agree with the decision. This can be specified naturally as a threshold policy.

Example. *In the United States, a recursive majority voting rule is used. A simply version of the rule, defined on n^d bits, is defined by*

$$Maj_n^{\otimes d}(x^{(1)}, \ldots, x^{(d)}) = Maj_n(Maj_n^{\otimes(d-1)}(x^{(1)}), Maj_n^{\otimes(d-1)}(x^{(n)}))$$

where each $x^{(i)} \in \{-1, 1\}^{n^{d-1}}$.

Example. *The tribes function divides voters into tribes, and the outcome of the vote holds if and only if one tribe is unanimously in favour of the vote. The width w, size s tribe voting rule is defined by*

$$Tribes_{ws} : \{-1, 1\}^{sw} \to \{-1, 1\}$$

$$Tribes_{ws}(x^{(1)}, \ldots, x^{(s)}) = Or_s(And_w(x^{(1)}), \ldots, And_w(x^{(s)}))$$

Any boolean function is a voting rule on two candidates, so the study of voting functions is just a language for talking about general properties of boolean functions. The language of social choice just provides the theory with an intuitive backing.

We can specify the type of election we wish to have by specifying properties of the boolean function which determines the vote. In particular,

- A voting rule f is **monotone** if $x \leqslant y$ implies $f(x) \leqslant f(y)$.

- A voting rule is **odd** if $f(x) = -f(-x)$.

- A rule is **unanimous** if $f(1) = 1$, and $f(-1) = -1$.

- A voting rule is **symmetric** if $f(x^\pi) = f(x)$, where $\pi \in S_n$ acts on $\{-1, 1\}^n$ by permuting coordinates.

- A voting rule is **transitive symmetric** if, for any index i, there is a permutation π taking i to any other index j, such that $f(x^\pi) = f(x)$.

There is only a single function which satisfies all of these properties for odd n, the majority function Maj_n.

Giving a voting rule $f : \{-1, 1\} \to \{-1, 1\}$, we may wish to quantify the power of each voter in changing the outcome of the vote. We call this

14

the **influence** of each voter, with the influence of the i'th voter denoted $\text{Inf}_i(f)$. It is important to note that even if a voting rule is symmetric, it does not imply that each person has equal power over the results of the election, once other's votes have been fixed (In a majority voting system, where there are 9 voters, 4 voting for politician A, and 5 for politician B, the 5 B voters have more power than the A voters, because they could choose to change their decision and 'flip' the election). However, we require a valid number measure the influence a person in all circumstances, and the best way to do this is to calculate expected power. In short, the influence of the i'th voter is the probability that the outcome of the vote would change if the i'th voter flipped his decision. In order to quantify this, we need a probability distribution on the choices of all voters. For simplicity, we assume that all voters choose the outcomes independently, and uniformly randomly, so the independence takes the form

$$\text{Inf}_i(f) = \mathbf{P}(f(X) \neq f(X^{\oplus i}))$$

Where if $x \in \{-1,1\}^n$, $x^{\oplus i}$ is the element of $\{-1,1\}^n$ obtained by 'flipping the i'th bit'. The real choices of people in an election will rarely be validly modelled in this system, but it is a good choice for measure the properties of the abstract voting rule f. Give a boolean-valued function f and input x, we say an index i is **pivotal** at x if inverting the value of x^i changes the output of the function f. In other words, if all the other voters have fixed choices, i has the power to sway the vote. Thus the influence measures the probability that index i is pivotal uniformly across all elements of the domain.

Because the choices of voters is uniformly random, there is a combinatorial interpretation of the influence over the discrete geometry of the Hamming cube. If we colour a node on the cube based on the output of the function f, then the influence of an index i is the fraction of coordinates on the cube whose color changes when we move along the i'th dimension of the cube.

Example. *The dictator function χ_i satisfies $\text{Inf}_i(\chi_i) = 1$, and $\text{Inf}_j(\chi_i) = 0$ otherwise. If f is an arbitrary voting rule for which $\text{Inf}_i(f) = 0$, then f completely ignores index i, we can actually specify f as a function of the remaining $n-1$ variables.*

Example. *The function And_n only changes on index i from 1 and $1^{\oplus i}$, so that*

$$\text{Inf}_i(And_n) = 2/2^n = 2^{1-n}$$

Example. *The function Maj_n only changes on index i from -1 to 1, on inputs x such there are $(n+1)/2$ indices j such that $x^j = -1$, and i is one of the indices. Once i is fixed, the total possible choices of indices in which these circumstances hold is*

$$\binom{n-1}{\frac{n-1}{2}}$$

hence

$$Inf_i(Maj_n) = 2\binom{n-1}{\frac{n-1}{2}}/2^n = \binom{n-1}{\frac{n-1}{2}}2^{1-n}$$

Applying Stirling's approximation, we have

$$Inf_i(Maj_n) = \sqrt{\frac{2}{n\pi}} + O(n^{-3/2})$$

To connect the influence of a boolean function to it's Fourier expansion, we must come up with an analytic expression for the influence. Given an expansion

$$f(x) = \sum a_S x^S$$

define the i'th partial derivative operator D_i by

$$(D_i f)(x) = \sum_{i \in S} a_S x^{S-\{i\}}$$

abstract though it may be, we also have an expression of this operator without using the Fourier coefficients as

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

because both operators are linear, and

$$(D_i x^S)(x) = \frac{(x^S)^{(i \to 1)} - (x^S)^{(i \to -1)}}{2} = \begin{cases} x^{S-\{i\}} & i \in S \\ 0 & i \notin S \end{cases}$$

For Boolean functions f, $D_i f$ is intimately connected to the influence of f, because

$$(D_i f)(x) = \begin{cases} \pm 1 & f(x^i) \neq f(x^{\oplus i}) \\ 0 & f(x^i) = f(x^{\oplus i}) \end{cases}$$

16

Hence $(D_i f)^2(x)$ is the 0-1 indicator of whether i is pivotal at x, and so

$$\text{Inf}_i(f) = \mathbf{E}[(D_i f)^2] = \|D_i f\|_2^2 = \sum_{i \in S} a_S^2$$

We obtain an even better measurement of the influence if f is monotone, in which case $D_i f = (D_i f)^2$, and so

$$\text{Inf}_i(f) = \mathbf{E}[D_i f] = a_{\{i\}}$$

This will allows us to show that all voters have small influence in a voting system which is transitive symmetric and monotone.

Theorem 2.1. *If $f : \{-1,1\}^n \to \{-1,1\}$ is monotone and transitive symmetric, then*

$$Inf_i(f) \leq \frac{1}{\sqrt{n}}$$

Proof. On one hand, we have

$$\text{Inf}_i(f) = \sum_{i \in S} a_S^2 \geq \sum_{i=1}^n a_{\{i\}}^2 = n a_{\{i\}}^2$$

whereas also

$$\text{Inf}_i(f) = a_{\{i\}}$$

hence $a_{\{i\}} \geq n a_{\{i\}}^2$, implying $a_{\{i\}} \leq 1/\sqrt{n}$. $\qquad\qquad\square$

The **total influence** of a boolean function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\mathbf{I}(f)$, is the sum of the influences $\text{Inf}_i(f)$ over each coordinate. It is not a measure of how powerful each voter is, but instead how chaotic the system is with respect to how any of the voters change their coordinates. For instance, constant functions have total influence zero, where no vote change effects the outcome in any way, whereas the function $x^{[n]}$ has total influence n, since every change in a voters choices flips the outcome of the vote.

There is a probabilistic interpretation of this equation. Define $\text{sens}_f(x)$ to be the number of pivotal indices for f at x. Then

$$\mathbf{I}(f) = \mathbf{E}[\text{sens}_f(X)]$$

There is also a combinatorial interpretation. Each vertex has n edges, and if we consider the probability distribution which first picks a point on the cube uniformly, and then an edge uniformly, then the resulting distribution will be the uniform distribution on edges. Since the expected number of **boundary edges** (those (x, y) for which $f(x) \neq f(y)$) eminating from a vertex is $\mathbf{I}(f)$, the chance that the edge we pick will be a boundary edge is $\mathbf{I}(f)/n$.

As with the influence, the total influence of a function has an interpretation via Fourier coefficients. Note that

$$\mathbf{I}(f) = \sum_{i=1}^{n} \mathrm{Inf}_i(f) = \sum_{i=1}^{n} \sum_{i \in S} a_S^2 = \sum_S |S| a_S^2 = \sum_{k=0}^{n} k W^k(f)$$

So that total influence is $\mathbf{E}[|\mathcal{S}|]$, where $\mathcal{S}$ is a random variable taking values in $[n]$ according to the spectral distribution of the Fourier coefficients. Since

$$\mathbf{V}(f) = \sum_{k=0}^{n} W^k(f)$$

We obtain the Poincare inequality

$$\mathbf{V}(f) \leqslant \mathbf{I}(f)$$

with equality if and only if the Fourier coefficients of order 2 or more vanish. For Boolean valued functions, the only such functions are the characters $x^{\{i\}}$ and the constant functions.

Geometrically, we can see this as an edge expansion bound on functions on the Hamming cube. Given a subset A of the n dimensional Hamming cube with m points, and if $\alpha = m/2^n$, then the 'characteristic function' $I_f : \{-1, 1\}^n \to \{-1, 1\}$ has variation $4\alpha(1 - \alpha)$, and therefore the fraction of edges in the cube which form a boundary edge is lower bounded by $4\alpha(1 - \alpha)/n$. In particular, for a set A where $\alpha = 1/2$, 1 in every n edges are boundary edges, and by looking at the dictator functions we see this bound is tight. Using more advanced theory, we can prove that if $\beta = \min(\alpha, 1 - \alpha)$, then the number of boundary edges is bounded below by

$$\frac{2\beta \log(1/\beta)}{n}$$

hence small subsets of the Hamming cube have a large boundary – the Hamming cubes are small set expanders.

Returning to our discussion of voting systems, an additional property to consider is the expected number of voters who agree with the outcome of the vote.