1 Basic definitions

We can view a boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ as a means to aggregate votes in a 2-outcome election. Common examples are:

- The AND and OR functions
- The MAJORITY function, $\text{MAJ}(x_1, \ldots, x_n) = \text{sign}(x_1 + \ldots + x_n)$ (here we assume $n$ is odd; otherwise we need to break a tie arbitrarily).
- More generally, weighted majority: for weights $a_1, \ldots, a_n \in \mathbb{R}$ define $f(x) = \text{sign}(a_1 x_1 + \ldots + a_n x_n)$.
- Dictator: $\chi_i(x_1, \ldots, x_n) = x_i$.
- More general, a $k$-junta is a function $f(x_1, \ldots, x_n)$ that depends only on $k$ inputs.
- The TRIBES function: let $n = ws$, then $\text{TRIBES}_{w,s}(x) = \text{AND}(\text{OR}(x_1, \ldots, x_w), \text{OR}(x_{w+1}, \ldots, x_{2w}), \ldots, \text{OR}(x_{n-w+1}, \ldots, x_n))$.

Some natural properties that we may want from a voting function are:

- Monotone: $f(x) \leq f(y)$ whenever $x \leq y$ coordinate-wise.
- Unanimous: $f(1, \ldots, 1) = 1$ and $f(-1, \ldots, -1) = -1$.
- Odd: $f(-x) = -f(x)$.
- Symmetric: $f(x^\pi) = f(x)$ for all permutations $\pi$ of $[n]$ (we denote $x^\pi = (x_{\pi(1)}, \ldots, x_{\pi(n)})$. That is, $f(x)$ depends only on the number of 1’s in $x$.

For example, the MAJORITY function is monotone, unanimous, odd and symmetric. In fact, it’s the only such function. The TRIBES function is monotone, unanimous, but not odd or symmetric. In order to allow for more general symmetries, we define the symmetry group of a function.
Definition 1.1 (Symmetry group). Let $f: \{-1, 1\}^n \to \mathbb{R}$. A permutation $\pi$ of $[n]$ is a symmetry of $f$ if $f(x^\pi) = f(x)$ for all $x$. The set of all symmetries of $f$ is a group (verify this!), called the symmetry group of $f$.

Definition 1.2 (Transitive symmetric). A function $f: \{-1, 1\}^n \to \mathbb{R}$ is transitive symmetric if its symmetry group $G$ is transitive. Namely, if for any $i, j \in [n]$ there is $\pi \in G$ such that $\pi(i) = j$.

The condition that a function is symmetric is equivalent to that it’s symmetry group is the full symmetry group of $[n]$. Note that while the TRIBES function, while not being symmetric, is transitive symmetric.

Example 1.3 (Graph property). Let $n = \binom{m}{2}$, and identify $\{0, 1\}^{\binom{m}{2}}$ with undirected graphs on $m$ vertices. A graph property is a function $P: \{0, 1\}^{\binom{m}{2}} \to \{0, 1\}$ which is invariant under permuting the vertices. That is, if $\tau \in S_m$ is permutation of the vertices, let $\pi \in S_n$ be its induced permutations on unordered pairs (edges in the graph). We require that $P(x^\pi) = P(x)$ for all $x$.

Examples of graph properties are: being 3-colorable; having an hamiltonian cycle; having a clique of size $k$, and many more. Note that graph properties are always transitive symmetric (we can map any potential edge $(i, j)$ to any other potential edge $(i', j')$ by a permutation on the vertices, mapping $\tau(i) = i'$ and $\tau(j) = j'$).

2 Influences

Definition 2.1 (Pivotal coordinate for an input). Let $f: \{-1, 1\}^n \to \{-1, 1\}$ be a boolean function. Let $x \in \{-1, 1\}^n$ and $i \in [n]$. We say that coordinate $i$ is pivotal for $x$ if $f(x^\oplus i) \neq f(x)$, where $x^\oplus i = x \oplus e_i$ is the input obtained by flipping the $i$-th bit of $x$.

Definition 2.2 (Influence). Let $f: \{-1, 1\}^n \to \{-1, 1\}$ be a boolean function. The influence of coordinate $i$ of $f$ is the probability that for a uniform input, the $i$-th coordinate is pivotal. Namely,

$$\text{Inf}_i(f) = \Pr_{x \in \{-1, 1\}^n}[f(x) \neq f(x^\oplus i)].$$

Equivalently, the $i$-th influence of $f$ is the number of edges in the hypercube in direction $i$ which are sensitive to $f$ ($f$ assigns the endpoints different values).

For example, the $i$-th influence of the $i$-th dictator $\chi_i$ is 1, while the $j$-th influence of $\chi_i$ for $j \neq i$ is 0. The $i$-th influence of MAJORITY is $\approx 1/\sqrt{n}$. It will be useful to derive an analytic expression of influence, using derivatives.
**Definition 2.3** (Derivative). Let \( f : \{-1, 1\}^n \to \mathbb{R} \). The \( i \)-th discrete derivative is
\[
D_if(x) = \frac{f(x^{(i \rightarrow 1)}) - f(x^{(i \rightarrow -1)})}{2}.
\]
Here we use the notation \( x^{(i \rightarrow b)} = (x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_n) \).

Observe that \( D_if(x) \) does not depend on \( x_i \), and that \( D_i \) is a linear operator: \( D_i(f + g) = D_if + D_ig \).

Observe that if \( f : \{-1, 1\}^n \to \{-1, 1\} \) is a boolean function then
\[
D_if(x) = \begin{cases} 
\pm 1 & \text{if direction } i \text{ is pivotal for } x \\
0 & \text{otherwise}
\end{cases}
\]

Thus, \( D_if(x)^2 \) is the 0-1 indicator for direction \( i \) being pivotal for \( x \). We will take this as the definition of the influence for real-valued functions.

**Definition 2.4** (Influence of real valued functions). Let \( f : \{-1, 1\}^n \to \mathbb{R} \) be a boolean function. The influence of coordinate \( i \) of \( f \) is defined as
\[
\text{Inf}_i(f) = \mathbb{E}_{x \in \{-1, 1\}^n}[D_if(x)^2].
\]

It will be useful to view influences in the Fourier basis.

**Lemma 2.5.** Let \( f : \{-1, 1\}^n \to \mathbb{R} \). Then
\[
\text{Inf}_i(f) = \sum_{S \ni i} \hat{f}(S)^2.
\]

**Proof.** Recall that \( f(x) = \sum_S \hat{f}(S)x^S \). We will compute the Fourier expansion of \( D_if \). In order to do so, it suffices to compute it for monomials \( x^S \) and then use linearity. It is easy to check that
\[
D_ix^S = \begin{cases} 
x^{S \setminus \{i\}} & \text{if } i \in S \\
0 & \text{otherwise}
\end{cases}
\]

Thus
\[
D_if(x) = \sum_{S \ni i} \hat{f}(S)x^{S \setminus \{i\}}.
\]

So by Parseval’s identity we have that
\[
\text{Inf}_i(f) = \mathbb{E}[D_if(x)^2] = \sum_S \hat{D}_if(S)^2 = \sum_{S \ni i} \hat{f}(S)^2.
\]

Note that the expression for \( D_if \) as a multilinear polynomial gives another justification why we call \( D_i \) a derivative operator: \( D_if(x) \) can be obtained by deriving \( f(x) \) on the \( x_i \) variable.

Monotone functions have especially nice influences. Here and below we shorthand \( \hat{f}(i) = \hat{f}(\{i\}) \).
Claim 2.6. Let \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \). If \( f \) is monotone then \( \inf_i f_i = \hat{f}(i) \).

Proof. If \( f \) is monotone then \( D_i f(x) = 1 \) if coordinate \( i \) is pivotal for \( x \) and 0 otherwise. So \( \inf_i f_i = \mathbb{E}[D_i f(x)] = \hat{D}_i f(\emptyset) = \hat{f}(i) \).

\[
\square
\]

Claim 2.7. Let \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \). If \( f \) is monotone and transitive symmetric then \( \inf_i f_i \leq 1/\sqrt{n} \) for all \( i \in [n] \).

Proof. As \( f \) is transitive symmetric, \( \inf_i f_i = \inf_j f_j \) for all \( i, j \). So it suffices to upper bound \( \inf_i f_i \). By Parseval

\[
1 = \mathbb{E}[f^2] = \sum_S \hat{f}(S)^2 \geq \sum_i \hat{f}(i)^2 = nf(1)^2.
\]

\[
\square
\]

3 Total influence

Definition 3.1 (Total influence). The total influence of \( f \) is \( I(f) = \sum_i \inf_i f_i \).

The influence of \( f \) is equivalently the number of edges in the hypercube which are sensitive for \( f \) (\( f \) assigns the endpoints different values). So, if we define \( \text{sens}(f, x) = \#\{i : f(x) \neq f(x^i)\} \) to be the number of sensitive edges touching node \( x \), then the following claim holds.

Claim 3.2. Let \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \). Then \( I(f) = \mathbb{E}_x \text{sens}(f, x) \).

Proof. This is just a double count: \( \inf_i f_i \) is the fraction of edges in direction \( i \) which is sensitive to \( f \), and \( I(f) \) is the sum over all \( i \). On the other hand, \( \mathbb{E}_x \text{sens}(f, x) \) first sums over directions \( i \), then averages over \( x \).

Its instructive to look at a few examples. The total influence of a character \( x^S \) is \( |S| \), as \( \inf_i f_i = 1 \) if \( i \in S \) and 0 otherwise. The total influence of a dictator is 1, and of a \( k \)-junta is at most \( k \). The total influence of MAJORITY is \( \approx \sqrt{n} \). Recall that we saw that \( \inf_i f_i \leq 1/\sqrt{n} \) for any monotone and transitive symmetric function; this turns out to be pretty tight for MAJORITY. The accurate value is

\[
I(\text{MAJORITY}) = (1 + o(1))\sqrt{2/\pi} \sqrt{n}.
\]

It will be convenient to relate the total influence to the Fourier coefficients.

Lemma 3.3. Let \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \). Then \( I(f) = \sum_S \hat{f}(S)^2 |S| \).

Proof. We have seen that \( \inf_i f_i = \sum_{S : i \in S} \hat{f}(S)^2 \). The expression for \( I(f) \) is obtained by summing over \( i \).

Among all monotone functions, MAJORITY obtains the maximal influence.
Lemma 3.4. Let $f : \{-1,1\}^n \to \{-1,1\}$ be monotone. Then $I(f) \leq I(MAJORITY) = (1 + o(1))\sqrt{2n/\pi}$.

Proof. We already saw that for monotone functions $\text{Inf}_i(f) = \hat{f}(i)$. So

$$I(f) = \sum_i \hat{f}(i) = \sum_i \mathbb{E}[f(x)x_i] = \mathbb{E}[f(x)(x_1 + \ldots + x_n)].$$

The expression is maximized by choosing $f(x) = \text{sign}(x_1 + \ldots + x_n)$, e.g. the MAJORITY function.

Returning to the view of boolean functions as voting rules, Rousseau suggested that an ideal voting rule is one where most voters agree with the outcome. We analyze his suggestion under the assumption of uniform votes, and show that MAJORITY is the unique maximizer.

Lemma 3.5. For any function $f : \{-1,1\}^n \to \{-1,1\}$ denote

$$\text{agree}(f) = \sum_{i=1}^n \Pr[x_i = f(x)].$$

Then $\text{agree}(f) \leq \text{agree}(MAJORITY)$.

Proof. We can express

$$\Pr[x_i = f(x)] = \frac{1}{2} \left(1 + \mathbb{E}[x_i f(x)]\right).$$

Thus

$$\text{agree}(f) = \frac{1}{2} \left(1 + \mathbb{E}[f(x)(x_1 + \ldots + x_n)]\right)$$

which is maximized by taking $f(x) = \text{sign}(x_1 + \ldots + x_n)$.

Let $A \subset \{0,1\}^n$ of size $|A| = \alpha 2^n$, and let $f(x) = 1$ if $x \in A$, and $f(x) = -1$ if $x \notin A$. The influence of $f$ can be interpreted as an isoperimetric parameter for $A$: $I(f)/n$ is the fraction of edges in the boolean cube which cross between $A$ and its complement. A simple lower bound on this number is the Poincaré inequality. The variance of $f$ is $\text{Var}(f) = \mathbb{E}[f^2] - \mathbb{E}[f]^2$. If $\Pr[f = 1] = \alpha$ then a simple calculation gives

$$\text{Var}(f) = 1 - (2\alpha - 1)^2 = 4\alpha(1 - \alpha).$$

Lemma 3.6 (Poincaré inequality). Let $f : \{-1,1\}^n \to \{-1,1\}$. Then $I(f) \geq \text{Var}(f) = 4\alpha(1 - \alpha)$.

Proof. We compute a Fourier expression of the variance. Recall that $\mathbb{E}[f] = \hat{f}(\emptyset)$ and $\mathbb{E}[f^2] = 1 = \sum_S \hat{f}(S)^2$, so

$$\text{Var}(f) = \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \sum_{S \neq \emptyset} \hat{f}(S)^2.$$
It turns out that the Poincaré inequality is not tight. The following is a stronger isoperimetric inequality, which is tight for $\alpha = 2^{-k}, k = 1, \ldots, n$.

**Theorem 3.7** (Edge isoperimetric inequality). Let $f: \{-1,1\}^n \to \{-1,1\}$ and assume that $\alpha = \Pr[f = 1] \leq 1/2$. Then

$$I(f) \geq 2\alpha \log(1/\alpha).$$

What is the function with the lowest possible influence? It makes sense to require this function to be somewhat unbiased (so $\text{Var}(f) \geq \Omega(1)$). In such a case, one can take a dictator, say $f(x) = x_1$, for which $I(f) = 1$.

However, what if we require the function to depend on all $n$ inputs? We can still use essentially a dictator, say take $f(x) = x_1 \oplus \text{AND}(x_2, \ldots, x_n)$. However, in this example there is a variable ($x_1$) whose influence is very high. Does this always have to be the case? What is the minimal maximal influence?

Consider the TRIBES function, with tribes of size $w \approx \log n$ to get a constant variance. One can compute that each variable in this function has influence $\Theta\left(\frac{\log n}{n}\right)$. Ben-Or and Linial conjectures that this is optimal. This was proved by Kahn, Kalai and Linial.

**Theorem 3.8** (KKL). Let $f: \{-1,1\}^n \to \{-1,1\}$. There exists a variable $i \in [n]$ such that $\text{Inf}_i(f) \geq \text{Var}(f) \cdot \Omega\left(\frac{\log n}{n}\right)$.

The proof utilizes the study of how a boolean function behaves under noise, which we begin to explore next. The technical tool needed to prove KKL is hypercontractivity, which we will only address later in the course.

## 4 Noise stability

Let $f: \{-1,1\}^n \to \{-1,1\}$ be a boolean function, viewed as a voting rule. Assume that we wish to compute $f(x)$, but each vote has some probability of being flipped, independently from the rest. The noise stability of $f$ is the ability of $f$ to be resilient to such errors. This concept will play an important role later on.

**Definition 4.1** (Noisy distribution). Let $x \in \{-1,1\}^n$ and $\rho \in [0,1]$. We define a distribution $N_\rho(x)$ on $\{-1,1\}^n$ as follows: sample $y \in \{-1,1\}^n$, where each $y_i$ equals $x_i$ with probability $\rho$, and otherwise is a random bit, independently for all $i$. That is,

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

Note that the latter definition extends to $\rho \in [-1,1]$.

A $\rho$-correlated pair $(x,y)$ is sampled as follows: first sample $x \in \{-1,1\}^n$ and then sample $y_i \in N_\rho(x_i)$. This distribution is symmetric between $x$ and $y$, and can be equivalently defined as follows: for any $i \in [n]$ independently, sample $x_i, y_i$ such that $E[x_i] = E[y_i] = 0$ and $E[x_i y_i] = \rho$. 


Definition 4.2 (Noise stability). Let $f : \{-1, 1\}^n \to \{-1, 1\}$. The noise stability of $f$ at rate $\rho$ is

$$Stab_\rho(f) = \mathbb{E}_{(x,y) \rho\text{-correlated}}[f(x)f(y)].$$

An equivalent definition is

$$Stab_\rho(f) = \Pr_{(x,y) \rho\text{-correlated}}[f(x) = f(y)] - \Pr_{(x,y) \rho\text{-correlated}}[f(x) \neq f(y)].$$

A related expression is the noise sensitivity of a function.

Definition 4.3 (Noise sensitivity). Let $f : \{-1, 1\}^n \to \{-1, 1\}$ and let $0 \leq \delta \leq 1/2$. The noise sensitivity of $f$ at rate $\delta$ is

$$NS_\delta(f) = \Pr[f(x) \neq f(y)]$$

where $x \in \{-1, 1\}^n$ is chosen uniformly, and each $y_i$ is chosen independently to be equal to $x_i$ with probability $1 - \delta$, and otherwise be the opposite value.

Observe that noise stability and noise sensitivity are related by

$$NS_\delta(f) = \frac{1 - Stab_{1-2\delta}(f)}{2}.$$

This is since if $\Pr[x_i = y_i] = 1 - \delta$ then $\mathbb{E}[x_i y_i] = 1 - 2\delta$, so $x, y$ in the definition of noise sensitivity are $\rho = 1 - 2\delta$ correlated, and since $Stab_\rho(f) = 1 - 2\Pr_{(x,y) \rho\text{-correlated}}[f(x) \neq f(y)]$.

It might not be so easy to compute the noise stability of various functions. However, we can nicely relate it to the Fourier expansion of $f$.

Lemma 4.4. Let $f : \{-1, 1\}^n \to \{-1, 1\}$. Then

$$Stab_\rho(f) = \sum_S \hat{f}(S)^2 \rho^{|S|}$$

and

$$NS_\delta(f) = \frac{1}{2} \sum_S \hat{f}(S)^2 (1 - (1 - 2\delta)^{|S|}).$$

Proof. The expression for $NS_\delta(f)$ just follows from the relation between noise sensitivity and noise stability, and by Parseval. We focus on proving the expression for $Stab_\rho(f)$.

Let $f(x) = \sum_S \hat{f}(S)x^S$. Let $x, y$ be a $\rho$-correlated sampled pair. Then

$$Stab_\rho(f) = \mathbb{E}[f(x)f(y)] = \mathbb{E}[\sum_{S,S'} \hat{f}(S)\hat{f}(S')x^Sy^{S'}] = \sum_{S,S'} \hat{f}(S)\hat{f}(S')\mathbb{E}[x^Sy^{S'}]$$

7
So, it suffices to study $E[x^S y^{S'}]$. We have

$$E[x^S y^{S'}] = E\left( \prod_{i \in S \cap S'} x_i y_i \cdot \prod_{i \in S \setminus S'} x_i \cdot \prod_{i \in S' \setminus S} y_i \right) = \left( \prod_{i \in S \cap S'} E[x_i y_i] \cdot \prod_{i \in S \setminus S'} E[x_i] \cdot \prod_{i \in S' \setminus S} E[y_i] \right),$$

where we used the fact that the pairs $(x_i, y_i)$ are independent of each other. In particular, if $S \neq S'$ then $E[x^S y^{S'}] = 0$ as $E[x_i] = E[y_i] = 0$. When $S = S'$ we get that $E[x^S y^{S}] = \rho^{|S|}$ as $E[x_i y_i] = \rho$. The lemma follows by linearity.

In particular, observe that when $\rho = 1$ (no noise) then indeed $\text{Stab}_1(f) = 1$ as needed. For $\rho < 1$ it gives a nonzero weight to the larger Fourier coefficients, but one which decays exponentially with the hamming weight of $S$. Thus, it is a very useful tool to isolate the contribution of the Fourier coefficients supported on small hamming weights.

An operator related to the the noise stability is the noise operator, which averages a noisy evaluation of a function.

**Definition 4.5.** Let $\rho \in [-1, 1]$. The noise operator $T_\rho$ is a linear operator on functions $f : \{-1, 1\}^n \to \mathbb{R}$, defined as

$$T_\rho f(x) = E_{y \sim N_\rho(x)}[f(y)].$$

**Claim 4.6.** The Fourier expansion of $T_\rho f$ is

$$T_\rho f(x) = \sum_S \hat{f}(S) \rho^{|S|}. $$

**Proof.** As $T_\rho$ is a linear operator, its enough to prove this for monomials $x^S$. We have

$$(T_\rho x^S)(x) = E_{y \sim N_\rho(x)}[y^S] = \prod_{i \in S} E[y_i] = \rho^{|S|} x^S,$$

since $E[y_i] = \rho x_i$ for each $i$ independently. \qed

The following corollary follows by Parseval’s identity.

**Corollary 4.7.** $\text{Stab}_\rho(f) = \langle f, T_\rho f \rangle$.

Next, we investigate what is the function which is the most noise insensitive (that is, with the largest noise stability). As it turns out, these are the dictator functions.

**Lemma 4.8.** Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be an unbiased function (namely $E[f] = 0$). Then $\text{Stab}_\rho(f) \leq \rho$. 

8
Proof. We know that $\text{Stab}_\rho(f) = \sum_S \hat{f}(S)^2 \rho^{|S|}$. As $\hat{f}(\emptyset) = \mathbb{E}[f] = 0$ we obtain that
$$\text{Stab}_\rho(f) \leq \sum_S \hat{f}(S)^2 \rho = \rho.$$Extremal functions are those with $\text{Stab}(f) = \rho$. As we just saw, it must be the case that $\hat{f}(S) = 0$ whenever $|S| > 1$. That is, all the nonzero Fourier coefficients are supported on the first level of the hamming cube. The next lemma shows that such functions are only the dictators $x_i$ or their negation $-x_i$.

**Lemma 4.9.** Let $f : \{-1,1\}^n \to \{-1,1\}$ be a function where $\hat{f}(S) = 0$ if $|S| > 1$. Then $f$ is either a constant function, a dictator or a negated dictator.

**Proof.** Let $a_0 = \mathbb{E}[f]$ and $a_i = \hat{f}(i)$. Then we have $f(x) = a_0 + \sum_{i=1}^n a_i x_i$. Assume that $f$ is not constant, so $a_j \neq 0$ for some $j \geq 1$. If we flip the value of $x_j$ then we get
$$f(x^{(j\to1)}) - f(x^{(j\to-1)}) = 2a_j.$$On the other hand, as $f$ is boolean, $f(x^{(j\to1)}) - f(x^{(j\to-1)}) \in \{-2,0,2\}$. As $a_j \neq 0$ then $a_j \in \{-1,1\}$. But we know that $1 = \sum \hat{f}(S)^2 = \sum a_i^2$, which means that $a_i = 0$ for all $i \neq j$. Thus $f(x) = a_j x_j$ which means that either $f(x) = x_j$ (dictator) or $f(x) = -x_j$ (negated dictator).

## 5 Arrow’s theorem

We saw that MAJORITY is the best possible voting rule for 2-outcome elections, when we want to maximize agreement. It is also a good approximation to many voting rules used in practice. However, as we shall soon see, things become more complicated when there are more than 2 outcomes.

To be concrete, we will consider 3-outcome elections. Assume there are 3 outcomes, $a, b, c$, and each voter votes on their relative order, say $b > a > c$. Condorcet in 1785 suggested an approach to define a winner based on a 2-outcome voting rule (i.e. a boolean function), where we apply the boolean function to any pair of potential outcomes.

<table>
<thead>
<tr>
<th>Voter preference</th>
<th>Voter preference</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#1 #2 #3 . . .</td>
<td></td>
</tr>
<tr>
<td>$a\ (+1)$ vs $b\ (-1)$</td>
<td>+1 +1 -1 . . . = x</td>
<td>$f(x)$</td>
</tr>
<tr>
<td>$b\ (+1)$ vs $c\ (-1)$</td>
<td>+1 -1 -1 . . . = y</td>
<td>$f(y)$</td>
</tr>
<tr>
<td>$c\ (+1)$ vs $a\ (-1)$</td>
<td>-1 +1 +1 . . . = z</td>
<td>$f(z)$</td>
</tr>
</tbody>
</table>

The condition that the $i$-th voter submits an ordering of $a, b, c$ is equivalent to the assumption that $x_i, y_i, z_i$ are not all equal. That is, if we define a boolean predicate $\text{NAE} : \{-1,1\}^3 \to \{0,1\}$ such that $\text{NAE}(w_1, w_2, w_3) = 1$ when $w_1, w_2, w_3$ are not all equal,
then we have that $\text{NAE}(x_i, y_i, z_i) = 1$ for all $i$. Observe that $|\text{NAE}^{-1}(1)| = 6$, which corresponds to the 6 possible permutations of $a, b, c$.

Condorcet defined a winner if we can infer an ordering of $a, b, c$ from $f(x), f(y), f(z)$; namely if $\text{NAE}(f(x), f(y), f(z)) = 1$.

**Definition 5.1** (Condorcet winner). Let $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$, and let $x, y, z \in \{-1, 1\}^n$ such that $\text{NAE}(x_i, y_i, z_i) = 1$ for all $i$. The outcome $f(x), f(y), f(z)$ is a Condorcet winner if $\text{NAE}(f(x), f(y), f(z)) = 1$.

Condorcet asked which voting systems $f$ guarantee that there is always a Condorcet winner. Arrow in 1950 proved that only dictators achieve that. Kalai proved a robust version of this, given uniform choices for each voter. The proof is based on Fourier analysis.

**Theorem 5.2.** Let $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$. Sample $(x, y, z) \in (\{-1, 1\}^n)^3$ by sampling for each $i$ independently $x_i, y_i, z_i \in \text{NAE}^{-1}(1)$. Then

$$\Pr[\exists \text{ Condorcet winner}] = \Pr[\text{NAE}(f(x), f(y), f(z)) = 1] = 3/4 - 3/4 \text{Stab}_{-1/3}(f).$$

In particular, the only functions $f$ for which this is 1 are dictators.

**Proof.** The Fourier expansion of NAE is

$$\text{NAE}(w_1, w_2, w_3) = \frac{3}{4} - \frac{1}{4} w_1 w_2 - \frac{1}{4} w_1 w_3 - \frac{1}{4} w_2 w_3.$$ 

Thus

$$\Pr[\exists \text{ Condorcet winner}] = \frac{3}{4} - \frac{1}{4} \mathbb{E}[f(x)f(y)] - \frac{1}{4} \mathbb{E}[f(x)f(z)] - \frac{1}{4} \mathbb{E}[f(y)f(z)].$$

By symmetry $\mathbb{E}[f(x)f(y)] = \mathbb{E}[f(x)f(z)] = \mathbb{E}[f(y)f(z)]$, so let's consider $\mathbb{E}[f(x)f(y)]$. We claim that the marginal distribution of $(x, y)$ is that of a $\rho$-correlated pair, for $\rho = -1/3$. In order to see that, note that the pairs $(x_i, y_i)$ are independent of each other, that $\mathbb{E}[x_i] = \mathbb{E}[y_i] = 0$ and that $\Pr[x_i = y_i] = 1/3$, so $\mathbb{E}[x_i y_i] = -1/3$. Thus

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

Let's assume that $f$ is a function such that this evaluates to 1. Applying the Fourier expansion of noise stability, we get that

$$1 = \frac{3}{4} - \frac{3}{4} \sum \hat{f}(S)^2 \left(-\frac{1}{3}\right)^{|S|}. $$

Rearranging gives

$$\sum \hat{f}(S)^2 \left(-\frac{1}{3}\right)^{|S|} = -\frac{1}{3}. $$

As $\sum \hat{f}(S)^2 = 1$, it must be that all the Fourier weight is supported on Fourier coefficients $S$ with $|S| = 1$. But we saw that this implies that $f$ is either a dictator or a negated dictator. \qed