Analysis of Boolean functions

Introduction

A *Boolean function* (Bf) describe an event depending on a number (n, usually large) of variables taking two values. We can define it as a function on $\{0,1\}^n$ taking values in {0,1} (Boolean input & outputs).

Applications.

a) Logic. Bf represent propositional formulas, two formulas are equivalent iff they are associated to the same Bf.

b) Random graph properties. Take a graph with n edges: a Bf (in this case boolean random variable) describe the presence or absence of a certain property in the subgraph obtained by randomly erasing a subset of the edges. E.g. the presence of a path of edges connecting two specific vertices of the graph.

c) Social choice theory. Each variable is the preference of a certain individual and the Bf describe a way to aggregating these preferences to express a choice at the level of the whole society.

d) Cryptography.

It is mathematically convenient to consider Bfs as functions from $\Omega_n = \{\pm 1\}^n$ to $\Omega = \Omega_1$.

Examples. (Names are usually due to social choice interpretation.)

a) Dictatorship. Behaves according to one of the variables (the dictator):

$$\mathrm{DICT}_n(x) = x_1$$

b) Parity. Tells whether there is an odd or even number of -1:

$$\mathrm{PAR}_n(x) = x_1 \cdots x_n$$

c) Majority. The most popular value wins: (assume n odd)

$$\mathrm{MAJ}_n(x) = \mathrm{sign}\left(\sum_i x_i\right)$$

d) Iterated majority. Two(or multi)-level election system: (assume $n = k\, m$ with k, m odd integers)

$$\mathrm{MAJ}_{n,k}(x) = \mathrm{sign}\left(\sum_{\ell=1}^{m} \mathrm{sign}\left(\sum_{i=1}^{k} x_{(\ell-1)m+i}\right)\right)$$

Social choice theory

To motivate our study of Boolean functions we look at a particular application outside "hard sciences": social choice theory.

Social choice theory studies (among other similar questions) the way individual preferences aggregate into the choice of the society. Its origins are in economics and political science. Individuals can also be "computer" or "distributed agents" which is in the interests of computer science.

Condorcet (1743–1794) published an work in 1785 "Essay on the application of Analysis to the Probability of Majority decisions" where he points out two facts which essentially initiated the mathematical study of social choice.

a) **Condorcet's paradox.** It is possible that the majority of the society prefers A to B and B to C while still preferring C to A. Aggregated preferences are not transitive.

b) **Condorcet's "jury theorem".** We have two candidates A and B. A is better than B. Each individual get indications as to whether one of the candidate is better than the other and acts accordingly by voting for this candidate. We assume that the "signal is small" that is that the indications received by each individual are random and independent and that they points to A with a probability $p > 1/2$. Signals tells something on the real state of affairs but there is noise. An elementary probability computation shows that, as the size of the population tends to infinity, the majority will correctly prefer A to B. Aggregation (via majority) filter out the noise and reveal an arbitrarily small signal (smallness is measured as a deviation from $p = 1/2$).

Arrow (1951) shaped modern social choice theory via an axiomatic analytic approach. For a recent perspective on the current status and perspective of social choice theory a very nice reading is given by the Nobel lecture by Amartya Sen, 1998 Nobel price in economics, which has the title "The possibility of social choice".

Arrow's most famous result is an "impossibility theorem" in the theory of social choice [K. Arrow, "A Difficulty in the Concept of Social Welfare", Journal of Political Economy 58(4) (August, 1950), pp. 328-346, http://gatton.uky.edu/Faculty/hoytw/751/articles/arrow.pdf]. He is the corecipient of the 1972 Nobel prize in Economics (with J. Hicks). A way to state his findings is the following

Theorem 1. (Arrow) *Any social choice between at least three alternatives which respect transitivity, independence of irrelevant alternatives and unanimity is a dictatorship.*

Let us explain this statement. A preference relation between a set $\mathbb{A}$ of alternatives is a binary relation aRb beween elements $a, b \in \mathbb{A}$. The relation is transitive if aRb and bRc imply aRc. *Independence of irrelevant alternatives* means that we assume that the social preference aRb is only a function of the set $\{i: aR_i b\}$, the preference of a w.r.t b for each individual in the society. *Unanimity* means that if $aR_i b$ for all $i = 1, \ldots, n$ the also $a Rb$ holds. The social choice is a *dictatorship* if the society choice coincide with the preference of a certain individual whenever this preference is strict: $(aR_i b) \& {\sim}(bR_i a) \Rightarrow a R b$ for some i and all $a, b \in \mathbb{A}$.

Condorcet found that social preference relations can be irrational (i.e. not transitive). Arrow found that they are rational only if they are given by dictatorship. Not a very encouraging beginning for a theory.

Preference relations can be casted into vectors of boolean functions. To keep things simple take three alternatives, $\mathbb{A} = \{A, B, C\}$. If we assume that the preference relation is strict (that is either aRb or bRa but not both) then to specify a preference relation on $\mathbb{A}$ we need three boolean values (for aRb, bRc, cRa). We do not assume rationality for the moment. The social choice can then be seen as a function F: $\{-1, 1\}^{3n} \to \{-1, 1\}^3$. Independence of irrelevant alternatives implies that the social preference aRb is determined uniquely by $aR_i b$ for $i = 1, \ldots, n$. So if we set $x_i = 1$ iff $aR_i b$ and $x_i = -1$ otherwise then whether or not aRb it is given by a Boolean function $f(x_1, \ldots, x_n)$. Similarly we can introduce variables y_i for $bR_i c$, z_i for $cR_i a$ and functions $g(y_1, \ldots, y_n)$ and $h(z_1, \ldots, z_n)$ for bRc and cRa respectively. The three boolean functions f, g, h describe completely the social choice function $F(x, y, z) = (f(x), g(y), h(z))$.

Example 2. Take $n = 3$ and f, g, h given by simple majority functions. Then we have Condorcet's paradox.

Order	Voter 1	Voter 2	Voter 3
1	a	b	c
2	b	c	a
3	c	a	b

Then $x = (1, -1, 1)$, $y = (1, 1, -1)$, $z = (-1, 1, 1)$, $F = (f, g, h) = (1, 1, 1)$. Each of the individual preferences $(x_{i,}\ y_i,\ z_i)$ is rational (transitive) while the social choice F is not rational : $aRbRcRa$.

Rational preferences over 3 alternatives corresponds to vectors (α, β, γ) of boolean values which are not all equal. This property is encoded by the function NAE: $\{\pm 1\}^3 \to \{0, 1\}$ given by

$$\mathrm{NAE}(\alpha, \beta, \gamma) = 1 - \sum_{k=\pm 1} \mathbb{I}_{(\alpha, \beta, \gamma) = (k, k, k)}.$$

Rational individual preferences determine the set $\Psi = \big\{(x, y, z) \in \{\pm 1\}^{3n} : \forall i \in [\![n]\!], \mathrm{NAE}(x_i, y_i, z_i) = 1\big\}$.

A neutral choice function is a function which is invariant by permutation of the alternatives. The choice function is symmetric if it is also invariant by a transitive group of permutations of the individuals. Common voting method are not necessarily invariant over the full group of permutations.

Another way to state Arrows' theorem which is more adapted to the point of view of this course is:

Theorem 3. *No choice function can be rational, independent of irrelevant alternatives, neutral and symmetric.*

Remark 4. Condorcet devised a voting method which give a rational outcome, which is neutral and symmetric but which is not independent over irrelevant alternatives. (Google for Condorcet's voting method).

In his 1788 essay "On the form of decisions made by plurality vote" Condorcet remarked on the subject of the possibility of irrational choice functions: "But after considering the facts, the average values or the results, we still need to determine their probability."

To quantify the "impossibility" in Arrow's result we introduce a way to measure the set of inputs in Ψ which result in an irrational aggregated outcome. Being combinatorial in its essence the most natural measure over the set Ψ is the uniform one. So we let $\mathbb{P}$ be the uniform measure over $\Psi \subseteq \Omega_n^3$. This is often called the *Impartial Culture* (IC) assumption.

Remark 5. The IC assumption is quite unrealistic both from the point of view of independence of different voters and uniformity over the voters preference relations. But we can think to encode biased electors into the properties of the social choice function and model with i.i.d. random variables the "undecided" electors.

Gil Kalai (2002) gave a quantitative version of Arrow's theorem.

Theorem 6. (Kalai) *The probability of a rational outcome for a symmetric neutral social choice on three alternatives is less than* 0.9192*.*

The proof uses insights coming for the theory of boolean functions via Fourier theoretic methods.

Kalai's statement is then that for any triples of boolean random variables $(f, g, h)\colon \Omega_n^3 \to \{\pm 1\}^3$ which are neutral and symmetric we have:

$$\mathbb{P}(\mathrm{NAE}(f, g, h) = 1) \leqslant 0.9192.$$

In the following we will introduce basic material in preparation to the proof of this and related results.

Harmonic analysis

Let $f\colon \Omega_n \to \Omega$ be a character, i.e. $f(x\, y) = f(x)\, f(y)$. Let $\left(x^i\right)_j = (x_j)^{\mathbb{I}_{i=j}}$. Then $x = x^1 \cdots x^n$ and

$$f(x) = f\left(x^1 \cdots x^n\right) = f\left(x^1\right) \cdots f(x^n) = \prod_{i \in S} f\left(x^i\right) = \prod_{i \in S} x_i = x_S$$

where $S = \left\{i \in [\![n]\!] : f\left(x^i\right) = -1\right\}$ and $x_\varnothing = 1$. The function f is the parity function on the subset S of coordinates. It is also evident that any parity function is a character. Characters of Ω_n are in bijection with subsets of $[[n]]$. Harmonic analysis over Ω_n consist in decomposing functions over Ω_n as linear combinations of characters.

The above caracterization of multplicative functions has a "robust" counterpart

Theorem 7. *Assume that $\mathbb{P}[f(x)\ f(y) = f(x\, y)] \geqslant 0.9$. Then f is close to some character, i.e. exists $S \subseteq [\![n]\!]$ such that*

$$\mathbb{P}[f(x) = x_S] \geqslant 0.9.$$

This is the kind of results we are looking at and Kalai's form of Arrow's theorem has this flavor.

Here the probability is given by uniform choice of both inputs x, y over Ω_n independently.

Consider the uniform measure $\mathbb{P}$ over Ω_n and the associated scalar product for real valued functions :

$$\langle f, g\rangle = 2^{-n} \sum_{x \in \Omega_n} f(x)\, g(x)$$

We will use also the corresponding norm $\|f\|_2^2 = \langle f, f\rangle$. Note that if f, g are Boolean then

$$\|f - g\|_2^2 = \langle f - g, f - g\rangle = \|f\|_2^2 + \|g\|_2^2 - 2\langle f, g\rangle = 2 - 2\langle f, g\rangle$$

so that closedness can be measured by the angle between the two corresponding vectors.

Characters are orthogonal for this scalar product

$$\langle x_S, x_T\rangle = 2^{-n} \sum_x x_S\, x_T = 2^{-n} \sum_x x_{S \Delta T} = \begin{cases} 0 & \text{if } S \neq T \\ 1 & \text{if } S = T \end{cases}$$

Fourier coefficients $\hat{f}\colon \mathcal{P}([\![n]\!]) \to \mathbb{R}$ of f are defined as $\hat{f}(S) = \langle f, x_S\rangle$ and

$$f(x) = \sum_{S \subseteq [\![n]\!]} \hat{f}(S)\, x_S\, .$$

Plancherel formula holds

$$\langle f, g\rangle = \sum_{S \subseteq [\![n]\!]} \hat{f}(S)\hat{g}(S)$$

and implies uniqueness of the representation as sum of characters.

Exercise 1. Compute the Fourier transform of these functions:

1. $\mathrm{MAJ}(x_1, x_2, x_3)$
2. $\mathrm{AND}(x_1, x_2) = 1$ si $x_1 = x_2 = 1$ and -1 otherwise.

Example 8. If f is multiplicative,

$$f(x\,y) = \sum_{S \subseteq [\![n]\!]} \hat{f}(S)\,(x\,y)_S = \sum_{S \subseteq [\![n]\!]} \hat{f}(S)\,x_S\,y_S$$

then $f(y)\hat{f}(S) = \langle f(x)\,f(y), x_S\rangle_x = \langle f(x\,y), x_S\rangle_x = \hat{f}(S)y_S$ which implies either $\hat{f}(S) = 0$ or $y_S = f(y)$ for all $y \in \Omega_n$.

Fourier transform over Ω_n applies naturally to all real (or complex) valued functions (even if characters are Boolean). Harmonic analysis of Boolean functions however shows additional peculiar phenomena. For Boolean functions Fourier coefficients must conjure so that all terms in the Fourier decomposition adds up exactly to ±1.

Consider the following problem: we draw at random and independently x, y in Ω_n and check whether $f(x)f(y) = f(x\ y)$ when it happens we set $\mathrm{BLR}(f) = 1$ otherwise we take it to be 0. The distribution of $\mathrm{BLR}(f)$ measures the multiplicativeness of f. It is clear that if f is a character then this random test always succeed (that is $\mathrm{BLR}(f) = 1$ always). Now we want to show that if this random test fail too often then f cannot be too near to a character.

Let us say that two Bf f, g are α-close if $\mathbb{P}(f = g) = 1 - \alpha$. Observe that

$$\mathbb{P}(f = g) = \mathbb{P}(f g = 1) = \mathbb{E}[1 + f g]/2 = \frac{1}{2} + \frac{1}{2}\mathbb{E}[f g]$$

so f, g are ε-close iff $\mathbb{E}[f g] = 1 - 2\,\varepsilon$.

Theorem 9. (Blum, Luby and Rubinfeld, 1990) *If* $\mathbb{P}[\mathrm{BLR}(f) = 1] \geqslant 1 - \varepsilon$ *then* f *is* ε-*close to some character.*

Proof. First let us express the probability using the Fourier expansion of f:

$$\mathbb{P}[\mathrm{BLR}(f) = 1] = \mathbb{E}\big[\mathbb{I}_{f(x)f(y) = f(x y)}\big] = \mathbb{E}\big[\mathbb{I}_{f(x)f(y)f(x y) = 1}\big] = \frac{1}{2}\mathbb{E}[1 + f(x)f(y)f(x\,y)]$$

by Fourier expansion:

$$\mathbb{E}[f(x)f(y)f(x\,y)] = \sum_{S, T, U \subseteq [\![n]\!]} \hat{f}(S)\hat{f}(T)\hat{f}(U)\mathbb{E}[x_S\,y_T x_U\,y_U]$$

by independence and orthogonality of characters

$$= \sum_{S,T,U \subseteq [\![n]\!]} \hat{f}(S)\hat{f}(T)\hat{f}(U)\mathbb{E}[y_T y_U]\mathbb{E}[x_S x_U] = \sum_{S \subseteq [\![n]\!]} \hat{f}(S)^3$$

Then

$$1 - \varepsilon \leqslant \frac{1}{2} + \frac{1}{2} \sum_{S \subseteq [\![n]\!]} \hat{f}(S)^3$$

Using that $\sum_{S \subseteq [\![n]\!]} \hat{f}(S)^2 = \langle f, f \rangle = 1$ we get

$$1 - 2\varepsilon \leqslant \sum_{S \subseteq [\![n]\!]} \hat{f}(S)^3 \leqslant \max_S \hat{f}(S) \sum_{S \subseteq [\![n]\!]} \hat{f}(S)^2 = \max_S \hat{f}(S)$$

That is there exists $T \subseteq [\![n]\!]$ such that $\hat{f}(T) \geqslant 1 - 2\varepsilon$ which means that f is strongly correlated to the character x_T. In particular $\|f - x_T\|_2^2 = 2 - 2\hat{f}(T) \leqslant 2 - 2 + 4\varepsilon = 4\varepsilon$ or that f is ε-close to x_T. □

The above proof is due to Bellare, Coppersmith, Håstad, Kiwi, and Sudan in 1995. The interest of this result is that it allow to test for multiplicativity of the "black box" f with $O(3/\varepsilon)$ tests instead of $O(2^n)$.

Another interesting property: assume f ε-close to the character x_S. Consider a uniform $y \in \Omega_n$ and define the random transformation $Tf(x) = f(y)f(x\,y)$ then for every fixed $x \in \Omega_n$ we have

$$\mathbb{P}[Tf(x) = x_S] \geqslant 1 - 2\varepsilon.$$

Exercise 2. Prove this.

Note that the nontrivial fact is that the random function Tf gives with high probability the correct value for the nearest character.

Another example: a Boolean function concentrated on the first Fourier level is a dictatorship (modulo ±1):

Lemma 10. (Friedgut) *Let f be a Boolean function such that $\hat{f}(S) = 0$ when $\#(S) > 1$ then either $f = \pm 1$ or $f(x) = \pm x_i$ for some i.*

Proof. *f is of the form $f(x) = \hat{f}(\varnothing) + \sum_i \hat{f}(\{i\})x_i$. Since it is Boolean $f(x)^2 = 1$ but by direct computation*

$$f(x)^2 = \hat{f}(\varnothing)^2 + \sum_i \hat{f}(\{i\})^2 + 2\sum_i \hat{f}(\{i\})\hat{f}(\varnothing)x_i + \sum_{i \neq j} \hat{f}(\{i\})\hat{f}(\{j\})x_{\{ij\}}$$

and by uniqueness of Fourier expansion we get $\hat{f}(\{i\})\hat{f}(\varnothing) = 0$, $\hat{f}(\{i\})\hat{f}(\{j\}) = 0$ for all i, j and

$$\hat{f}(\varnothing)^2 + \sum_i \hat{f}(\{i\})^2 = 1.$$

This implies that either $\hat{f}(\varnothing)^2 = 1$ or $\hat{f}(\{i\})^2 = 1$ for some i. □

Note that if we restrict ourselves to balanced functions, i.e. such that $\mathbb{E}[f] = 0$ then only (anti)-dictators are possible in this last result.

To assess the approximate counterpart of the previsous lemma we need to gauge the "spectrum" of the function. Define $f^{\leqslant k}$ as the projection of f over the span of $\{x_S : S \subseteq [\![n]\!], \#(S) \leqslant k\}$ and $f^{>k} = f - f^{\leqslant k}$ the projection on the orthogonal space and finally $f^{=k} = f^{\leqslant k} - f^{\leqslant k}$

$$f^{\leqslant k}(x) = \sum_{S \subseteq [\![n]\!], \#(S) \leqslant k} \hat{f}(S) x_S, \qquad f^{=k}(x) = \sum_{S \subseteq [\![n]\!], \#(S) = k} \hat{f}(S) x_S.$$

We also introduce the weight $W_k(f)$ at level k of the function f as

$$W_k(f) = \|f^{=k}\|_2^2 = \sum_{\#S = k} \hat{f}(S)^2.$$

Note that for a Bf $\sum_{k \geqslant 0} W_k(f) = 1$.

Theorem 11. (Friedgut, Kalai, Naor) *A Boolean function f such that $\|f^{>1}\| \leqslant \varepsilon < \varepsilon_0$ is $O(\varepsilon)$-close to a dictatoriship, i.e. there exists $i \in [\![n]\!]$ such that $\|f - g\| \lesssim \varepsilon$ where $g(x) = \hat{f}(\varnothing) + \hat{f}(\{i\}) x_i$.*

[E. Friedgut, G. Kalai and A. Naor, Boolean functions whose Fourier transform is concentrated on the first two levels, Adv. in Appl. Math., 29(2002), 427-437. http://www.ma.huji.ac.il/~kalai/fkn.pdf.]

We will prove first a small variation.

Theorem 12. *If $W_1(f) = 1 - \varepsilon$ then f is $O(\varepsilon)$-close to a dictator (or anti-dictator).*

Proof. We can assume that f is balanced. By hypothesis $\|f^{>1}\|^2 = \varepsilon$. Moreover

$$1 = f^2 = (f^{=1} + f^{>1})^2 = (f^{=1})^2 + f^{>1}(2f - f^{>1})$$

and

$$(f^{=1})^2 = \sum_i \hat{f}(\{i\})^2 + \underbrace{\sum_{i \neq j} \hat{f}(\{i\}) \hat{f}(\{j\}) x_{\{ij\}}}_{q} = 1 - \varepsilon + q.$$

Then

$$q = \varepsilon - 2 f f^{>1} - (f^{>1})^2$$

By Chebishev inequality

$$\mathbb{P}\left(|f^{>1}| \geqslant 10\, \varepsilon^{1/2}\right) \leqslant \frac{\mathbb{E}[(f^{>1})^2]}{100\, \varepsilon} = \frac{1}{100}$$

so that with probability 0.99 we have

$$|q| \leqslant \varepsilon + 2|f^{>1}| + |f^{>1}|^2 \leqslant \varepsilon + 20\, \varepsilon^{1/2} + 100\, \varepsilon \leqslant 21\, \varepsilon^{1/2}$$

for ε sufficiently small. Now the key point is that a "second level" function like q cannot be small with large probability unless also its second moment be small. We will prove later that the above probability estimate implies

$$\mathbb{E}[q^2] \leqslant O(\varepsilon)$$

for some proportionality constant (∼1000). But now

$$O(\varepsilon) \geqslant \mathbb{E}\left[q^2\right] = \sum_{i \neq j} \hat{f}(\{i\})^2 \hat{f}(\{j\})^2 = \left[\underbrace{\sum_i \hat{f}(\{i\})^2}_{1-\varepsilon}\right]^2 - \sum_i \hat{f}(\{i\})^4$$

$$\Rightarrow \sum_i \hat{f}(\{i\})^4 \geqslant 1 - O(\varepsilon)$$

$$\Rightarrow 1 - O(\varepsilon) \leqslant \max_i \hat{f}(\{i\})^2 \sum_i \hat{f}(\{i\})^2$$

$$\Rightarrow \max_i \hat{f}(\{i\})^2 \geqslant 1 - O(\varepsilon)$$

□

Let us compute the probability that a given Bf allows for a rational outcome

Lemma 13. *Let (x, y, z) be uniform over Ψ then*

$$\mathbb{P}[\mathrm{NAE}(f(x), f(y), f(z)) = 1] = \frac{3}{4} - \frac{3}{4} \sum_{S \subseteq [\![n]\!]} \left(-\frac{1}{3}\right)^{\#(S)} \hat{f}(S)^2.$$

Proof. The NAE function has the following Fourier transform

$$\mathrm{NAE}(\alpha, \beta, \gamma) = \frac{3}{4} - \frac{1}{4}(\alpha\beta + \beta\gamma + \gamma\alpha)$$

Hence

$$\mathbb{P}[\mathrm{NAE}(f(x), f(y), f(z)) = 1] \ = \ \mathbb{E}[\mathrm{NAE}(f(x), f(y), f(z))]$$

$$= \mathbb{E}\left[\frac{3}{4} - \frac{1}{4}(f(x) f(y) + f(y)\, f(z) + f(z)\, f(x))\right]$$

(all the pairs (x, y), (y, z), (z, x) have the same distribution)

$$= \frac{3}{4} - \frac{3}{4}\mathbb{E}[f(x) f(y)] = \frac{3}{4} - \frac{3}{4} \sum_{S, T \subseteq [\![n]\!]} \hat{f}(S) \hat{f}(T) \mathbb{E}[x_S\, y_T]\,.$$

Now assume that $S \neq T$ and for example that $i \in S$ but $i \notin T$. Then given independence of different voters profiles we get

$$\mathbb{E}[x_S\, y_T] = \mathbb{E}\left[x_i\, x_{S \setminus \{i\}} y_T\right] = \mathbb{E}[x_i] \mathbb{E}\left[x_{S \setminus \{i\}} y_T\right] = 0$$

while if $S = T$

$$\mathbb{E}[x_S\, y_S] = \prod_{i \in S} \mathbb{E}[x_i\, y_i] = (\mathbb{E}[x_1 y_1])^{\#(S)} = \left(-\frac{1}{3}\right)^{\#(S)}$$

since the four possibilities $(x_1, y_1) = (\pm 1, \pm 1)$ are not equally likely and

$$2\mathbb{P}[x_1 y_1 = 1] = \mathbb{P}[x_1 y_1 = -1] = \frac{2}{3}, \qquad \mathbb{E}[x_1 y_1] = \frac{1}{3} - \frac{2}{3} = -\frac{1}{3}.$$

□

Corollary 14. *Let f be a balanced Boolean function. If $\mathbb{P}[\mathrm{NAE}(f(x), f(y), f(z))] \geqslant 1-\varepsilon$ then $W_1(f) \geqslant 1-\frac{9}{2}\varepsilon$.*

Proof. Given the above formula for the probability of a rational outcome we have

$$1-\varepsilon \leqslant \frac{3}{4}-\frac{3}{4}\sum_{k\geqslant 1}\left(-\frac{1}{3}\right)^k W_k(f) = \frac{3}{4}+\frac{1}{4}W_1(f)-\frac{3}{4}\sum_{k\geqslant 2}\left(-\frac{1}{3}\right)^k W_k(f)$$

since $W_0(f)=\hat{f}(\varnothing)^2=0$ by the balancedness of f. Then

$$W_1(f) \geqslant 1-4\varepsilon+3\sum_{k\geqslant 2}\left(-\frac{1}{3}\right)^k W_k(f) \geqslant 1-4\varepsilon+3\inf_g \sum_{k\geqslant 2}\left(-\frac{1}{3}\right)^k W_k(g)$$

where the inf it is taken over Boolean functions g such that $W_0(g)=0$ and $W_1(g)=W_1(f)$. Observe that for this class of functions $\sum_{k\geqslant 2} W_k(g)=1-W_1(g)=1-W_1(f)$ so that the inf is attained when $W_3(g)=1-W_1(f)$ and $W_k(g)=0$ for all others values of $k\geqslant 2$. Optimization over g then yields

$$W_1(f) \geqslant 1-4\,\varepsilon+\frac{1-W_1(f)}{9}$$

which gives the claim. □

Putting these result together we can show that if f passes the NAE test with probability $1-\varepsilon$ then it must be $O(\varepsilon)$-close to a dictator.