Cambridge University Press
978-1-107-03832-5 - Analysis of Boolean Functions
Ryan O'Donnell
Excerpt
More information

1

Boolean Functions and the Fourier Expansion

In this chapter we describe the basics of analysis of Boolean functions. We emphasize viewing the Fourier expansion of a Boolean function as its representation as a real multilinear polynomial. The viewpoint based on harmonic analysis over $\mathbb{F}_2^n$ is mostly deferred to Chapter 3. We illustrate the use of basic Fourier formulas through the analysis of the Blum–Luby–Rubinfeld linearity test.

1.1. On Analysis of Boolean Functions

This is a book about Boolean functions,

$$f : \{0, 1\}^n \to \{0, 1\}.$$

Here f maps each length-n binary vector, or *string*, into a single binary value, or *bit*. Boolean functions arise in many areas of computer science and mathematics. Here are some examples:

- In circuit design, a Boolean function may represent the desired behavior of a circuit with n inputs and one output.
- In graph theory, one can identify v-vertex graphs G with length-$\binom{v}{2}$ strings indicating which edges are present. Then f may represent a property of such graphs; e.g., $f(G) = 1$ if and only if G is connected.
- In extremal combinatorics, a Boolean function f can be identified with a "set system" $\mathscr{F}$ on $[n] = \{1, 2, \ldots, n\}$, where sets $X \subseteq [n]$ are identified with their 0-1 indicators and $X \in \mathscr{F}$ if and only if $f(X) = 1$.
- In coding theory, a Boolean function might be the indicator function for the set of messages in a binary error-correcting code of length n.

- In learning theory, a Boolean function may represent a "concept" with n binary attributes.
- In social choice theory, a Boolean function can be identified with a "voting rule" for an election with two candidates named 0 and 1.

We will be quite flexible about how bits are represented. Sometimes we will use True and False; sometimes we will use -1 and 1, thought of as real numbers. Other times we will use 0 and 1, and these might be thought of as real numbers, as elements of the field $\mathbb{F}_2$ of size 2, or just as symbols. Most frequently we will use -1 and 1, so a Boolean function will look like

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

But we won't be dogmatic about the issue.

We refer to the domain of a Boolean function, $\{-1, 1\}^n$, as the *Hamming cube* (or hypercube, n-cube, Boolean cube, or discrete cube). The name "Hamming cube" emphasizes that we are often interested in the *Hamming distance* between strings $x, y \in \{-1, 1\}^n$, defined by

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

Here we've used notation that will arise constantly: x denotes a bit string, and x_i denotes its ith coordinate.

Suppose we have a problem involving Boolean functions with the following two characteristics:

- the Hamming distance is relevant;
- you are *counting* strings, or the uniform probability distribution on $\{-1, 1\}^n$ is involved.

These are the hallmarks of a problem for which *analysis of Boolean functions* may help. Roughly speaking, this means deriving information about Boolean functions by analyzing their *Fourier expansion*.

1.2. The "Fourier Expansion": Functions as Multilinear Polynomials

The *Fourier expansion* of a Boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ is simply its representation as a real, multilinear polynomial. (*Multilinear* means that no variable x_i appears squared, cubed, etc.) For example, suppose $n = 2$ and

$f = \max_2$, the "maximum" function on 2 bits:

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

Then $\max_2$ can be expressed as a multilinear polynomial,

$$\max_2(x_1, x_2) = \tfrac{1}{2} + \tfrac{1}{2}x_1 + \tfrac{1}{2}x_2 - \tfrac{1}{2}x_1x_2; \tag{1.1}$$

this is the "Fourier expansion" of $\max_2$. As another example, consider the *majority function* on 3 bits, $\mathrm{Maj}_3 : \{-1,1\}^3 \to \{-1,1\}$, which outputs the ± 1 bit occurring more frequently in its input. Then it's easy to verify the Fourier expansion

$$\mathrm{Maj}_3(x_1, x_2, x_3) = \tfrac{1}{2}x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{2}x_3 - \tfrac{1}{2}x_1x_2x_3. \tag{1.2}$$

The functions $\max_2$ and Maj_3 will serve as running examples in this chapter.

Let's see how to obtain such multilinear polynomial representations in general. Given an arbitrary Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ there is a familiar method for finding a polynomial that interpolates the 2^n values that f assigns to the points $\{-1,1\}^n \subset \mathbb{R}^n$. For each point $a = (a_1, \dots, a_n) \in \{-1,1\}^n$ the *indicator polynomial*

$$1_{\{a\}}(x) = \left(\tfrac{1+a_1x_1}{2}\right)\left(\tfrac{1+a_2x_2}{2}\right)\cdots\left(\tfrac{1+a_nx_n}{2}\right)$$

takes value 1 when $x = a$ and value 0 when $x \in \{-1,1\}^n \setminus \{a\}$. Thus f has the polynomial representation

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

Illustrating with the $f = \max_2$ example again, we have

$$\begin{aligned} \max_2(x) &= (+1)\left(\tfrac{1+x_1}{2}\right)\left(\tfrac{1+x_2}{2}\right) \\ &+ (+1)\left(\tfrac{1-x_1}{2}\right)\left(\tfrac{1+x_2}{2}\right) \\ &+ (+1)\left(\tfrac{1+x_1}{2}\right)\left(\tfrac{1-x_2}{2}\right) \\ &+ (-1)\left(\tfrac{1-x_1}{2}\right)\left(\tfrac{1-x_2}{2}\right) \quad = \quad \tfrac{1}{2} + \tfrac{1}{2}x_1 + \tfrac{1}{2}x_2 - \tfrac{1}{2}x_1x_2. \end{aligned} \tag{1.3}$$

Let us make two remarks about this interpolation procedure. First, it works equally well in the more general case of *real-valued Boolean functions*, $f : \{-1,1\}^n \to \mathbb{R}$. Second, since the indicator polynomials are multilinear when expanded out, the interpolation always produces a multilinear polynomial.

Indeed, it makes sense that we can represent functions $f : \{-1,1\}^n \to \mathbb{R}$ with multilinear polynomials: since we only care about inputs x where $x_i = \pm 1$, any factor of x_i^2 can be replaced by 1.

We have illustrated that every $f : \{-1,1\}^n \to \mathbb{R}$ can be represented by a real multilinear polynomial; as we will see in Section 1.3, this representation is unique. The multilinear polynomial for f may have up to 2^n terms, corresponding to the subsets $S \subseteq [n]$. We write the monomial corresponding to S as

$$x^S = \prod_{i \in S} x_i \qquad (\text{with } x^\emptyset = 1 \text{ by convention}),$$

and we use the following notation for its coefficient:

$$\widehat{f}(S) = \text{coefficient on monomial } x^S \text{ in the multilinear representation of } f.$$

This discussion is summarized by the *Fourier expansion theorem*:

Theorem 1.1. *Every function $f : \{-1,1\}^n \to \mathbb{R}$ can be uniquely expressed as a multilinear polynomial,*

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\, x^S. \tag{1.4}$$

This expression is called the Fourier expansion *of f, and the real number $\widehat{f}(S)$ is called the* Fourier coefficient *of f on S. Collectively, the coefficients are called the* Fourier spectrum *of f.*

As examples, from (1.1) and (1.2) we obtain:

$$\widehat{\max_2}(\emptyset) = \tfrac{1}{2}, \quad \widehat{\max_2}(\{1\}) = \tfrac{1}{2}, \quad \widehat{\max_2}(\{2\}) = \tfrac{1}{2}, \quad \widehat{\max_2}(\{1,2\}) = -\tfrac{1}{2};$$

$$\widehat{\mathrm{Maj}_3}(\{1\}),\ \widehat{\mathrm{Maj}_3}(\{2\}),\ \widehat{\mathrm{Maj}_3}(\{3\}) = \tfrac{1}{2}, \quad \widehat{\mathrm{Maj}_3}(\{1,2,3\}) = -\tfrac{1}{2},$$

$$\widehat{\mathrm{Maj}_3}(S) = 0 \text{ else}.$$

We finish this section with some notation. It is convenient to think of the monomial x^S as a function on $x = (x_1, \dots, x_n) \in \mathbb{R}^n$; we write it as

$$\chi_S(x) = \prod_{i \in S} x_i.$$

Thus we sometimes write the Fourier expansion of $f : \{-1,1\}^n \to \mathbb{R}$ as

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\, \chi_S(x).$$

So far our notation makes sense only when representing the Hamming cube by $\{-1, 1\}^n \subseteq \mathbb{R}^n$. The other frequent representation we will use for the cube is $\mathbb{F}_2^n$. We can define the Fourier expansion for functions $f : \mathbb{F}_2^n \to \mathbb{R}$ by "encoding" input bits $0, 1 \in \mathbb{F}_2$ by the real numbers $-1, 1 \in \mathbb{R}$. We choose the encoding $\chi : \mathbb{F}_2 \to \mathbb{R}$ defined by

$$\chi(0_{\mathbb{F}_2}) = +1, \quad \chi(1_{\mathbb{F}_2}) = -1.$$

This encoding is not so natural from the perspective of Boolean logic; e.g., it means the function $\max_2$ we have discussed represents logical AND. But it's mathematically natural because for $b \in \mathbb{F}_2$ we have the formula $\chi(b) = (-1)^b$. We now extend the χ_S notation:

Definition 1.2. For $S \subseteq [n]$ we define $\chi_S : \mathbb{F}_2^n \to \mathbb{R}$ by

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

which satisfies

$$\chi_S(x + y) = \chi_S(x)\chi_S(y). \tag{1.5}$$

In this way, given any function $f : \mathbb{F}_2^n \to \mathbb{R}$ it makes sense to write its Fourier expansion as

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\, \chi_S(x).$$

In fact, if we are really thinking of $\mathbb{F}_2^n$ the n-dimensional vector space over $\mathbb{F}_2$, it makes sense to identify subsets $S \subseteq [n]$ with vectors $\gamma \in \mathbb{F}_2^n$. This will be discussed in Chapter 3.2.

1.3. The Orthonormal Basis of Parity Functions

For $x \in \{-1, 1\}^n$, the number $\chi_S(x) = \prod_{i \in S} x_i$ is in $\{-1, 1\}$. Thus $\chi_S : \{-1, 1\}^n \to \{-1, 1\}$ is a Boolean function; it computes the logical *parity*, or *exclusive-or* (XOR), of the bits $(x_i)_{i \in S}$. The parity functions play a special role in the analysis of Boolean functions: the Fourier expansion

$$f = \sum_{S \subseteq [n]} \widehat{f}(S)\, \chi_S \tag{1.6}$$

shows that any f can be represented as a linear combination of parity functions (over the reals).

It's useful to explore this idea further from the perspective of linear algebra. The set of all functions $f : \{-1, 1\}^n \to \mathbb{R}$ forms a vector space V, since we can add two functions (pointwise) and we can multiply a function by a real scalar. The vector space V is 2^n-dimensional: if we like we can think of the functions in this vector space as vectors in $\mathbb{R}^{2^n}$, where we stack the 2^n values $f(x)$ into a tall column vector (in some fixed order). Here we illustrate the Fourier expansion (1.1) of the $\max_2$ function from this perspective:

$$\max_2 = \begin{bmatrix} +1 \\ +1 \\ +1 \\ -1 \end{bmatrix} = (1/2)\begin{bmatrix} +1 \\ +1 \\ +1 \\ +1 \end{bmatrix} + (1/2)\begin{bmatrix} +1 \\ -1 \\ +1 \\ -1 \end{bmatrix} + (1/2)\begin{bmatrix} +1 \\ +1 \\ -1 \\ -1 \end{bmatrix} + (-1/2)\begin{bmatrix} +1 \\ -1 \\ -1 \\ +1 \end{bmatrix}. \tag{1.7}$$

More generally, the Fourier expansion (1.6) shows that every function $f : \{-1, 1\}^n \to \mathbb{R}$ in V is a linear combination of the parity functions; i.e., the parity functions are a *spanning set* for V. Since the number of parity functions is $2^n = \dim V$, we can deduce that they are in fact a *linearly independent basis* for V. In particular this justifies the uniqueness of the Fourier expansion stated in Theorem 1.1.

We can also introduce an inner product on pairs of function $f, g : \{-1, 1\}^n \to \mathbb{R}$ in V. The usual inner product on $\mathbb{R}^{2^n}$ would correspond to $\sum_{x \in \{-1,1\}^n} f(x)g(x)$, but it's more convenient to scale this by a factor of 2^{-n}, making it an average rather than a sum. In this way, a Boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ will have $\langle f, f \rangle = 1$, i.e., be a "unit vector".

Definition 1.3. We define an inner product $\langle \cdot, \cdot \rangle$ on pairs of function $f, g : \{-1, 1\}^n \to \mathbb{R}$ by

$$\langle f, g \rangle = 2^{-n} \sum_{x \in \{-1,1\}^n} f(x)g(x) = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \{-1,1\}^n} [f(\boldsymbol{x})g(\boldsymbol{x})]. \tag{1.8}$$

We also use the notation $\|f\|_2 = \sqrt{\langle f, f \rangle}$, and more generally,

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

Here we have introduced probabilistic notation that will be used heavily throughout the book:

Notation 1.4. We write $\boldsymbol{x} \sim \{-1,1\}^n$ to denote that $\boldsymbol{x}$ is a uniformly chosen random string from $\{-1,1\}^n$. Equivalently, the n coordinates $\boldsymbol{x}_i$ are independently chosen to be $+1$ with probability $1/2$ and -1 with probability $1/2$. We always write random variables in **boldface**. Probabilities **Pr** and expectations **E** will always be with respect to a uniformly random $\boldsymbol{x} \sim \{-1,1\}^n$ unless otherwise specified. Thus we might write the expectation in (1.8) as $\mathbf{E}_{\boldsymbol{x}}[f(\boldsymbol{x})g(\boldsymbol{x})]$ or $\mathbf{E}[f(\boldsymbol{x})g(\boldsymbol{x})]$ or even $\mathbf{E}[fg]$.

Returning to the basis of parity functions for V, the crucial fact underlying all analysis of Boolean functions is that this is an *orthonormal basis*.

Theorem 1.5. *The 2^n parity functions $\chi_S : \{-1,1\}^n \to \{-1,1\}$ form an orthonormal basis for the vector space V of functions $\{-1,1\}^n \to \mathbb{R}$; i.e.,*

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

Recalling the definition $\langle \chi_S, \chi_T \rangle = \mathbf{E}[\chi_S(\boldsymbol{x})\chi_T(\boldsymbol{x})]$, Theorem 1.5 follows immediately from two facts:

Fact 1.6. *For $x \in \{-1,1\}^n$ it holds that $\chi_S(x)\chi_T(x) = \chi_{S \triangle T}(x)$, where $S \triangle T$ denotes symmetric difference.*

Proof. $$\chi_S(x)\chi_T(x) = \prod_{i \in S} x_i \prod_{i \in T} x_i = \prod_{i \in S \triangle T} x_i \prod_{i \in S \cap T} x_i^2 = \prod_{i \in S \triangle T} x_i = \chi_{S \triangle T}(x).$$ □

Fact 1.7. $$\mathbf{E}[\chi_S(\boldsymbol{x})] = \mathbf{E}\Big[\prod_{i \in S} \boldsymbol{x}_i\Big] = \begin{cases} 1 & \text{if } S = \emptyset, \\ 0 & \text{if } S \neq \emptyset. \end{cases}$$

Proof. If $S = \emptyset$ then $\mathbf{E}[\chi_S(\boldsymbol{x})] = \mathbf{E}[1] = 1$. Otherwise,

$$\mathbf{E}\Big[\prod_{i \in S} \boldsymbol{x}_i\Big] = \prod_{i \in S} \mathbf{E}[\boldsymbol{x}_i]$$

because the random bits $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are independent. But each of the factors $\mathbf{E}[\boldsymbol{x}_i]$ in the above (nonempty) product is $(1/2)(+1) + (1/2)(-1) = 0$. □

1.4. Basic Fourier Formulas

As we have seen, the Fourier expansion of $f : \{-1,1\}^n \to \mathbb{R}$ can be thought of as the representation of f over the orthonormal basis of parity functions $(\chi_S)_{S \subseteq [n]}$. In this basis, f has 2^n "coordinates", and these are precisely the

Fourier coefficients of f. The "coordinate" of f in the χ_S "direction" is $\langle f, \chi_S \rangle$; i.e., we have the following formula for Fourier coefficients:

Proposition 1.8. *For $f : \{-1,1\}^n \to \mathbb{R}$ and $S \subseteq [n]$, the Fourier coefficient of f on S is given by*

$$\widehat{f}(S) = \langle f, \chi_S \rangle = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \{-1,1\}^n} [f(\boldsymbol{x})\chi_S(\boldsymbol{x})].$$

We can verify this formula explicitly:

$$\langle f, \chi_S \rangle = \left\langle \sum_{T \subseteq [n]} \widehat{f}(T)\, \chi_T, \chi_S \right\rangle = \sum_{T \subseteq [n]} \widehat{f}(T) \langle \chi_T, \chi_S \rangle = \widehat{f}(S), \qquad (1.9)$$

where we used the Fourier expansion of f, the linearity of $\langle \cdot, \cdot \rangle$, and finally Theorem 1.5. This formula is the simplest way to calculate the Fourier coefficients of a given function; it can also be viewed as a streamlined version of the interpolation method illustrated in (1.3). Alternatively, this formula can be taken as the *definition* of Fourier coefficients.

The orthonormal basis of parities also lets us measure the squared "length" (2-norm) of $f : \{-1,1\}^n \to \mathbb{R}$ efficiently: it's just the sum of the squares of f's "coordinates" – i.e., Fourier coefficients. This simple but crucial fact is called *Parseval's Theorem*.

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

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

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

$$\sum_{S \subseteq [n]} \widehat{f}(S)^2 = 1.$$

As examples we can recall the Fourier expansions of $\max_2$ and Maj_3:

$$\max_2(x) = \tfrac{1}{2} + \tfrac{1}{2}x_1 + \tfrac{1}{2}x_2 - \tfrac{1}{2}x_1x_2, \quad \mathrm{Maj}_3(x) = \tfrac{1}{2}x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{2}x_3 - \tfrac{1}{2}x_1x_2x_3.$$

In both cases the sum of squares of Fourier coefficients is $4 \times (1/4) = 1$.

More generally, given two functions $f, g : \{-1,1\}^n \to \mathbb{R}$, we can compute $\langle f, g \rangle$ by taking the "dot product" of their coordinates in the orthonormal basis of parities. The resulting formula is called *Plancherel's Theorem*.

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

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

We can verify this formula explicitly as we did in (1.9):

$$\langle f, g\rangle = \Big\langle \sum_{S\subseteq[n]} \widehat{f}(S)\,\chi_S, \sum_{T\subseteq[n]} \widehat{g}(T)\,\chi_T \Big\rangle = \sum_{S,T\subseteq[n]} \widehat{f}(S)\widehat{g}(T)\langle\chi_S, \chi_T\rangle$$
$$= \sum_{S\subseteq[n]} \widehat{f}(S)\widehat{g}(S).$$

Now is a good time to remark that for Boolean-valued functions $f, g : \{-1, 1\}^n \to \{-1, 1\}$, the inner product $\langle f, g\rangle$ can be interpreted as a kind of "correlation" between f and g, measuring how similar they are. Since $f(x)g(x) = 1$ if $f(x) = g(x)$ and $f(x)g(x) = -1$ if $f(x) \neq g(x)$, we have:

Proposition 1.9. *If $f, g : \{-1, 1\}^n \to \{-1, 1\}$,*

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

Here we are using the following definition:

Definition 1.10. Given $f, g : \{-1, 1\}^n \to \{-1, 1\}$, we define their *(relative Hamming) distance* to be

$$\mathrm{dist}(f, g) = \mathop{\mathbf{Pr}}_{\boldsymbol{x}}[f(\boldsymbol{x}) \neq g(\boldsymbol{x})],$$

the fraction of inputs on which they disagree.

With a number of Fourier formulas now in hand we can begin to illustrate a basic theme in the analysis of Boolean functions: interesting combinatorial properties of a Boolean function f can be "read off" from its Fourier coefficients. Let's start by looking at one way to measure the "bias" of f:

Definition 1.11. The *mean* of $f : \{-1, 1\}^n \to \mathbb{R}$ is $\mathbf{E}[f]$. When f has mean 0 we say that it is *unbiased*, or *balanced*. In the particular case that $f : \{-1, 1\}^n \to \{-1, 1\}$ 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.

Fact 1.12. *If $f : \{-1, 1\}^n \to \mathbb{R}$ then $\mathbf{E}[f] = \widehat{f}(\emptyset)$.*

This formula holds simply because $\mathbf{E}[f] = \langle f, 1\rangle = \widehat{f}(\emptyset)$ (taking $S = \emptyset$ in Proposition 1.8). In particular, a Boolean function is unbiased if and only if its empty-set Fourier coefficient is 0.

Next we obtain a formula for the *variance* of a real-valued Boolean function (thinking of $f(\boldsymbol{x})$ as a real-valued random variable):

Proposition 1.13. *The* variance *of* $f : \{-1, 1\}^n \to \mathbb{R}$ *is*

$$\mathbf{Var}[f] = \langle f - \mathbf{E}[f], f - \mathbf{E}[f] \rangle = \mathbf{E}[f^2] - \mathbf{E}[f]^2 = \sum_{S \neq \emptyset} \widehat{f}(S)^2.$$

This Fourier formula follows immediately from Parseval's Theorem and Fact 1.12.

Fact 1.14. *For* $f : \{-1, 1\}^n \to \{-1, 1\}$,

$$\mathbf{Var}[f] = 1 - \mathbf{E}[f]^2 = 4\,\mathbf{Pr}[f(\boldsymbol{x}) = 1]\,\mathbf{Pr}[f(\boldsymbol{x}) = -1] \in [0, 1].$$

In particular, a Boolean-valued function f has variance 1 if it's unbiased and variance 0 if it's constant. More generally, the variance of a Boolean-valued function is proportional to its "distance from being constant".

Proposition 1.15. *Let* $f : \{-1, 1\}^n \to \{-1, 1\}$. *Then* $2\epsilon \leq \mathbf{Var}[f] \leq 4\epsilon$, *where*

$$\epsilon = \min\{\mathrm{dist}(f, 1), \mathrm{dist}(f, -1)\}.$$

The proof of Proposition 1.15 is an exercise. See also Exercise 1.17.

By using Plancherel in place of Parseval, we get a generalization of Proposition 1.13 for *covariance*:

Proposition 1.16. *The* covariance *of* $f, g : \{-1, 1\}^n \to \mathbb{R}$ *is*

$$\mathbf{Cov}[f, g] = \langle f - \mathbf{E}[f], g - \mathbf{E}[g] \rangle = \mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g] = \sum_{S \neq \emptyset} \widehat{f}(S)\widehat{g}(S).$$

We end this section by discussing the *Fourier weight distribution* of Boolean functions.

Definition 1.17. The *(Fourier) weight* of $f : \{-1, 1\}^n \to \mathbb{R}$ on set S is defined to be the squared Fourier coefficient, $\widehat{f}(S)^2$.

Although we lose some information about the Fourier coefficients when we square them, many Fourier formulas only depend on the weights of f. For example, Proposition 1.13 says that the variance of f equals its Fourier weight on nonempty sets. Studying Fourier weights is particularly pleasant for Boolean-valued functions $f : \{-1, 1\}^n \to \{-1, 1\}$ since Parseval's Theorem says that they always have total weight 1. In particular, they define a *probability distribution* on subsets of $[n]$.

Definition 1.18. Given $f : \{-1, 1\}^n \to \{-1, 1\}$, the *spectral sample* for f, denoted $\mathcal{S}_f$, is the probability distribution on subsets of $[n]$ in which the set S has probability $\widehat{f}(S)^2$. We write $\boldsymbol{S} \sim \mathcal{S}_f$ for a draw from this distribution.