Harmonic Analysis of Boolean Functions, and applications in CS

Lecture 2

February 25, 2008

Lecturer: Guy Kindler | Scribe by: Or Meir | Updated: March 9, 2008

In the main part of this lecture we will see an example for how one can reason about influences of boolean functions without using harmonic analysis. This example demonstrates some useful tricks in probability, and also demonstrates how much sweat one needs to shed in order to solve such questions without using harmonic analysis. Specifically, we prove the following:

Theorem 1 *Let $f : \{-1,1\}^n \to \{-1,1\}$ be an unbiased (i.e., balanced) boolean function. Then $I(f) \geq 1$.*

We will prove Theorem 1 using tools from probability theory. In order to use those tools, we will first reformulate the theorem in terms of probability theory.

Notation As common in Theoretical Computer Science, given a function f over two random variables X, Y, we denote by $\mathbb{E}_X f(X,Y)$ the expectation of $f(X,Y)$ where Y is *fixed* and the probability is taken only over X. In other words, we denote

$$\mathbb{E}_X f(X,Y) = \mathbb{E}\left[f(X,Y)|Y\right]$$

A similiar notation will be used for the variance $\mathbb{V}_X f(X,Y)$. We warn that this notation might be dangerous to use when X and Y depend on each other, since it may cause the reader to forget that the distribution of X depends on the choice of Y. Thus, we will mostly use it for independent random variables.

Notation Given a vector $x \in \{-1,1\}^n$ and a coordinate $i \in [n]$, we denote by $x\backslash i$ the set of all x's coordinates except for i.

Observe that using the foregoing notation, we can define the influence $I_i(f)$ of a boolean function f as follows:

$$I_i(f) = \mathbb{E}_{x\backslash i}\mathbb{V}_{x_i} f(x)$$

This new definition of $I_i(f)$ has two advantages:

1. This definition is more comfortable to work with when using probabilistic tools, since it is phrased using "probabilistic terms" such as expectation and variance.
2. This definition does not assume that the function f is boolean. Thus, we can use this definition to define the influence $I_i(f)$ of non-boolean functions $f : \{-1,1\}^n \to \mathbb{R}$, or even the influence of functions of the more general form $f : \prod_{i=1}^n X_i \to \mathbb{R}$, where the X_i's denote probability spaces. Indeed, most of the arguments that will be presented throughout this course work for functions of this general form, and not just for boolean functions.

Before we can prove Theorem 1, we first need to prove some general lemmata from probability theory.

Lemma 2 *Let X and X' be i.i.d (independent and identically distributed) random variables. Then*

$$\mathbb{V}X = \frac{1}{2}\mathbb{E}\left(X - X'\right)^2$$

Proof It holds that

$$\begin{aligned}
\mathbb{E}\left(X - X'\right)^2 &= \mathbb{E}\left[X^2 - 2 \cdot X \cdot X' + X'^2\right] \\
\text{(By linearity of expectation)} &= \mathbb{E}X^2 - \mathbb{E}\left[2 \cdot X \cdot X'\right] + \mathbb{E}X'^2 \\
\text{(By independence of } X \text{ and } X'\text{)} &= \mathbb{E}X^2 - 2 \cdot \mathbb{E}\left[X\right]\mathbb{E}\left[X'\right] + \mathbb{E}X'^2 \\
\text{(Since } X \text{ and } X' \text{ are identically distributed)} &= \mathbb{E}X^2 - 2\mathbb{E}\left[X\right]\mathbb{E}\left[X\right] + \mathbb{E}X^2 \\
&= 2EX^2 - 2\mathbb{E}^2X \\
&= 2\mathbb{V}X
\end{aligned}$$

as required. ■

Remark Lemma 2 reformulates the definition of the variance of a random variable X by adding an additional auxiliary random variable X'. In general, adding auxiliary random variables in order to simplify probabilistic expressions is a very useful trick.

Lemma 3 (Special case of the law of total variance) *Let X_1 and X_2 be independent random variables, and let f be a function over two variables. Then*

$$\mathbb{V}_{X_1,X_2}f(X_1,X_2) = \mathbb{V}_{X_2}\mathbb{E}_{X_1}f(X_1,X_2) + \mathbb{E}_{X_2}\mathbb{V}_{X_1}f(X_1,X_2)$$

Proof It holds that

$$\begin{aligned}
\text{r.h.s.} &= \mathbb{V}_{X_2}\mathbb{E}_{X_1}f(X_1,X_2) + \mathbb{E}_{X_2}\mathbb{V}_{X_1}f(X_1,X_2) \\
&= \mathbb{E}_{X_2}\mathbb{E}^2_{X_1}f(X_1,X_2) - \mathbb{E}^2_{X_2}\mathbb{E}_{X_1}f(X_1,X_2) \\
&\quad +\mathbb{E}_{X_2}\left[\mathbb{E}_{X_1}f^2(X_1,X_2) - \mathbb{E}^2_{X_1}f(X_1,X_2)\right] \\
\text{(By the linearity of expectation)} &= \mathbb{E}_{X_2}\mathbb{E}^2_{X_1}f(X_1,X_2) - \mathbb{E}^2_{X_2}\mathbb{E}_{X_1}f(X_1,X_2) \\
&\quad +\mathbb{E}_{X_2}\mathbb{E}_{X_1}f^2(X_1,X_2) - \mathbb{E}_{X_2}\mathbb{E}^2_{X_1}f(X_1,X_2) \\
&= \mathbb{E}_{X_2}\mathbb{E}_{X_1}f^2(X_1,X_2) - \mathbb{E}^2_{X_2}\mathbb{E}_{X_1}f(X_1,X_2) \\
&= \mathbb{E}_{X_1,X_2}f^2(X_1,X_2) - \mathbb{E}^2_{X_1,X_2}f(X_1,X_2) \\
&= \mathbb{V}_{X_1,X_2}f(X_1,X_2) \\
&= \text{l.h.s.}
\end{aligned}$$

as required. ■

Remark Note that the proof of Lemma 3 did not use the assumption that X_1 and X_2 are independent. We assumed it merely for notational convinience. However, if the proof is formulated more carefully, it can be made to hold for general variables.

The following simple fact will be useful in the proof of the next lemma.

Fact 4 *Let X be a random variable. Then $\mathbb{E}\left[X^2\right] \geq \mathbb{E}^2[X]$.*

Proof Follows immediately from the fact that the variance $\mathbb{V}X = \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X]$ is always non-negative. ■

Lemma 5 *Let X_1 and X_2 be independent random variables, and let f be a function over two variables. Then*

$$\mathbb{V}_{X_2}\mathbb{E}_{X_1} f(X_1, X_2) \leq \mathbb{E}_{X_1}\mathbb{V}_{X_2} f(X_1, X_2)$$

where equality holds if and only if there exist functions g, h over one variable each such that

$$f(X_1, X_2) = g(X_1) + h(X_2)$$

Remark While Lemma 5 might seem odd at first look, it is actually very intuitive, and reflects the fact that the operation of averaging reduces the variance: Usually, when doing many experiments, the variance of the average of the experiments is smaller than the variance of each of the experiments separately. Thus, we expect the variance of the average of the experiments to be at most the average of the variances of the separate experiments. The latter assertion is equivalent to the inequality stated in Lemma 5.

Proof of Lemma 5 Let X_2' be a random variable that is independent from and identically distributed as X_2. By Lemma 2, it holds that

$$\begin{aligned}
\mathbb{V}_{X_2}\mathbb{E}_{X_1} f(X_1, X_2) &= \frac{1}{2}\mathbb{E}_{X_2, X_2'}\left(\mathbb{E}_{X_1} f(X_1, X_2) - \mathbb{E}_{X_1} f(X_1, X_2')\right)^2 \\
\text{(By the linearity of expectation)} &= \frac{1}{2}\mathbb{E}_{X_2, X_2'}\mathbb{E}^2_{X_1}\left[f(X_1, X_2) - f(X_1, X_2')\right] \\
\text{(By Fact 4)} &\leq \frac{1}{2}\mathbb{E}_{X_2, X_2'}\mathbb{E}_{X_1}\left(f(X_1, X_2) - f(X_1, X_2')\right)^2 \\
&= \mathbb{E}_{X_1}\left[\frac{1}{2}\mathbb{E}_{X_2, X_2'}\left(f(X_1, X_2) - f(X_1, X_2')\right)^2\right] \\
\text{(By Lemma 2)} &= \mathbb{E}_{X_1}\mathbb{V}_{X_2} f(X_1, X_2)
\end{aligned}$$

as required. The claim about equality is left as an exercise. ■

We are now ready to prove Theorem 1. In fact, we will prove a more general theorem:

Theorem 6 *Let $f : \{-1, 1\}^n \to \mathbb{R}$. Then $\mathbb{V}_x f(x) \leq I(f)$, where $I(f) = \sum_{i=1}^n \mathbb{E}_{x\setminus i}\mathbb{V}_{x_i} f(x)$.*

Note that Theorem 1 indeed follows as a corollary from Theorem 6, since if f is an unbiased boolean function then $\mathbb{V}_x f(x) = 1$.

Remark Note that Theorem 6 does not imply that if f is unbiased and $I(f) = 1$ then f is a dictatorship. The reason is that this is not necessarily true for non-boolean functions, yet our proof holds for any real-valued function.

Exercise 1 *Show an unbiased real valued function $f : \{-1, 1\}^n \to \mathbb{R}$ such that $I(f) = 1$ and f is far from a dictatorship.*

Exercise 2 *Show that for an unbiased* boolean *function f it holds that $I(f) = 1$ if and only if f is a dictatorship.*

Proof of Theorem 6: The proof goes by induction on n. For $n = 1$ we have that

$$I(f) = I_1(f) = \mathbb{V}_x f(x)$$

We now assume that the theorem holds for some $n - 1$ and prove it for n. By Lemma 3, it holds that

$$\begin{aligned}
\mathbb{V}_x f(x) &= \mathbb{E}_{x\setminus 1}\mathbb{V}_{x_1} f(x) + \mathbb{V}_{x\setminus 1}\mathbb{E}_{x_1} f(x) \\
\text{(By definition of } I_1\text{)} &= I_1(f) + \mathbb{V}_{x\setminus 1}\mathbb{E}_{x_1} f(x) \\
\text{(By Lemma 5)} &\leq I_1(f) + \mathbb{E}_{x_1}\mathbb{V}_{x\setminus 1} f(x) \\
\text{(By the induction Hypothesis)} &\leq I_1(f) + \mathbb{E}_{x_1}\left[\sum_{i=2}^{n} \mathbb{E}_{x\setminus\{i,1\}}\mathbb{V}_{x_i} f(x)\right] \\
&= I_1(f) + \sum_{i=2}^{n} \mathbb{E}_{x\setminus i}\mathbb{V}_{x_i} f(x) \\
&= I(f)
\end{aligned}$$

as required. ■

Remark We mention that if one only wants to prove Theorem 1, without its generalization in Theorem 6, then he can use somewhat simpler proofs. However, even those proofs are not that simple.

Concluding remark Recall that one of our motivations to show this proof of Theorem 1 was to show that how much sweat one needs to shed in order to solve such questions without using harmonic analysis. One might claim that this proof is not very difficult if one already knows the general probabilitic lemmata we used. Indeed, the actual proof of Theorem 6 is quite a short one. However, even though this proof is short, it is not very straightforward, that is, it is not clear how can one come up with such proof on the first place. Using harmonic analysis, on the other hand, one obtains a proof that is very straightforward.

Fourier Transform

We can view functions of the form $f : \{-1, 1\}^n \to \mathbb{R}$ as vectors in $\mathbb{R}^{\{-1,1\}^n}$, and the space of real valued functions over $\{-1, 1\}^n$ as a linear space of dimension 2^n. We can then consider bases for this linear space. For example, we can consider the standard basis $\{\delta_y\}_{y\in\{-1,1\}^n}$, where δ_y is defined as

$$\delta_y(x) = \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases}$$

For this basis we have for every function f:

$$f = \sum_{y \in \{-1,1\}^n} f(y) \cdot \delta_y$$

We will now discuss another basis.

Definition 7 *Let $\emptyset \neq S \subseteq [n]$. Define $\chi_S(x) = \prod_{i \in S} x_i$. We also define $\chi_\emptyset(x) = 1$.*

We show $\{\chi_S\}_{S \subseteq [n]}$ is a basis. The set $\{\chi_S\}_{S \subseteq [n]}$ contains 2^n functions, so in order to show that it is a basis, it suffices to show that it spans the space.

Claim 8 *$\{\chi_S\}_{S \subseteq [n]}$ spans the space of real valued functions over $\{-1, 1\}$.*

Proof It suffices to show that the set $\{\chi_S\}_{S \subseteq [n]}$ spans the elements of the standard basis. For every $y \in \{-1, 1\}^n$ it holds that

$$\begin{aligned} \delta_y(x) &= \prod_{i=1}^{n} \left(\frac{1 + x_i \cdot y_i}{2} \right) \\ &= \frac{1}{2^n} \sum_{S \subseteq [n]} \prod_{i \in S} x_i \cdot y_i \\ &= \frac{1}{2^n} \sum_{S \subseteq [n]} \left(\prod_{i \in S} y_i \right) \chi_S(x) \end{aligned}$$

Thus, δ_y can be viewed as a linear combination of the functions in $\{\chi_S\}_{S \subseteq [n]}$,
as required. ■

The basis $\{\chi_S\}_{S \subseteq [n]}$ is called the "Fourier Basis", and will solve all of our problems.