Lecturer: Mihai Pătraşcu
May 9, 2006

Stability, Influence, and Fourier Analysis on the Cube

In this lecture, we will develop a theory of boolean functions concerned with issues like stability under random noise, and influence of variables. An increasingly important player in this research is Fourier analysis on the Hamming cube; we introduce this important notion and the connections to stability and influence.

This theory of boolean functions can be motivated through reference to an analysis of social choice (voting schemes, and quantitative versions of Arrow's theorem) or simply through mathematical elegance. However, on the practical side, we will use it in the next lecture to give an optimal (conditional) inapproximability result for MaxCut.

Normally, we think of the boolean domain as $\{0,1\}$. However, to simplify expressions, we will now use $\{-1,+1\}$. The domain $\{\pm 1\}^n$ is called the n-dimensional hypercube. We will be studying n-variable boolean functions, $f : \{\pm 1\}^n \to \{\pm 1\}$, or sometimes real functions defined on the cube, $f : \{\pm 1\}^n \to \Re$.

1 Influence

Given a n-variable function, one thing we might look at is how "influent" each variable is, i.e. how much the variable influences the output of the function in the average case.

Definition 1. *Let $f : \{\pm 1\}^n \to \Re$. Then the influence of x_i on f is defined as:*

$$\mathrm{Inf}_i(f) = \mathop{\mathbb{E}}_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n}\left[\mathop{\mathbf{var}}_{x_i}[f]\right]$$

This definition should already be very intuitive, but for boolean functions we can simplify it further. Note that for a function $g : \{\pm 1\} \to \{\pm 1\}$, $\mathbf{var}[g] = 1$ if $g(1) \neq g(-1)$, and $\mathbf{var}[g] = 0$ otherwise. Then, if f is defined with an output in $\{\pm 1\}$, we can restate the definition as:

$$\mathrm{Inf}_i(f) = \Pr_{x \in \{\pm 1\}^n}[f(x) \neq f(x_1,\ldots,x_{i-1},-x_i,x_{i+1},\ldots,x_n)]$$

Common metaphors for the influence interpret f as a voting scheme. The n voters express a binary opinion, and the function decides the outcome of the vote by aggregating all opinions. Let us consider a few possible schemes:

Predetermined output. If f is constant, $\mathrm{Inf}_i(f) = 0$ for any i. In general, functions biased towards $+1$ or -1 are not too interesting, so we generally work with *balanced functions* (satisfying $\mathbb{E}_x[f(x)] = 0$).

Dictatorship. This is the extreme case when $f(x) = x_i$, for some i. In this case $\mathrm{Inf}_i(f) = 1$ and $\mathrm{Inf}_j(f) = 0$ for any $j \neq i$. A generalization of this notion is a *junta*, where f only depends on the choices of a few voters.

Majority. The output is the most popular opinion among the n voters, i.e. f is the majority function. Let us assume n is odd, so we do not have to deal with tie-breaking (extending to n even is a simple computation). Since majority is symmetric, we must have $\mathrm{Inf}_i(f) = \mathrm{Inf}_j(f)$ for any i and j. Thus, let us compute $\mathrm{Inf}_n(f)$. Note that the probability x_n influences the result at all is very small: this only happens when among $x_1, \dots, x_{n-1}$ there is an equal number of positive and negative votes. The probability of this event is $\binom{(n-1)/2}{n-1}/2^{n-1} = \Theta(\frac{1}{\sqrt{n}})$. When x_n is relevant, this contributes 1 to the influence, so we have $\mathrm{Inf}_n(f) = \Theta(\frac{1}{\sqrt{n}})$.

Parity. The output is the XOR of the voters' choices; with a $\{\pm 1\}$ domain, $f(x) = \prod_{i=1}^n x_i$. Note that $\mathrm{Inf}_i(f) = 1$ for all i. This is an extreme example where the influence of every variable is maximal, yet if votes are announced simultaneously, nobody has any control of the output.

Tribes. The smallest influences we have seen so far (for balanced functions) were those of majority. Can we do better? Consider grouping the variables into n/b "tribes" of b variables. In each tribe, we take the AND of the choices, and then take an OR of the choices of each tribe. We can imagine a bunch of tribes pondering whether to build a temple on a hill top. The construction requires the joint efforts of b workers, but workers from different tribes speak different languages. Then, the temple will only be built if (at least) one tribe agrees unanimously to do it.

For $b = \log n - \lg\lg n + \Theta(1)$, the outcome will be balanced. Furthermore, a variable can influence the outcome only when the remaining $b-1$ members of the tribe vote YES, so the influence if at most $\frac{1}{2^{b-1}} = O(\frac{\lg n}{n})$. This is complemented by a famous result of Kahn, Kalai and Linial [1], showing that any balanced function has a variable with influence $\Omega(\frac{\lg n}{n})$.

2 Stability under Noise

In interpreting influence as a measure of the quality of a voting scheme, we are faced with a couple of problems. On the one hand, even if some voter has a high influence, it does not necessarily mean he has any control over the outcome (parity). However, if our voter is allowed to delay his decision a bit, he can often decide the outcome, which is clearly bad. On the other hand, the tribes function can achieve minimal influences, but still it is not appealing for general purpose voting. We can justify this through a new notion, stability.

Definition 2. *Let $f : \{\pm 1\}^n \to \Re$ and $-1 \le \rho \le 1$. Let x be uniformly random in $\{\pm 1\}^n$ and let y be ρ-correlated with x, i.e. each y_i is chosen independently so that $\mathbb{E}[x_i y_i] = \rho$.*

Then the noise stability of f at ρ is defined as:

$$\mathbb{S}_\rho(f) = \mathop{\mathbb{E}}_{x,y}[f(x)f(y)]$$

In standard probability theory, $\mathbb{E}[xy]$ is called the *correlation* of x and y (assuming $\mathbb{E}[x] = \mathbb{E}[y] = 0$, which is the case for the quantities we are looking at). Thus, we can imagine evaluating f at an input which is x, plus some random noise. The measure for the noise is ρ, the correlation between each coordinate of x and the perturbed coordinate. Stability measures the correlation between f at the original input and f at a noisy input.

Back on the realm of voting, it is not hard to convince ourselves that voting is a noisy process. Some x_i might not represent the true choice of voter i, e.g. due to error in expressing the vote, or simply because the voter has a bad day and he makes a wrong decision that he will later regret. Then, high stability is a desirable property of voting schemes.

A dictator function has stability ρ (immediate from the definition), and it can be shown that this is the highest possible. However, this is not satisfying because we want variables to have small influences.

The tribes scheme should intuitively have small stability: a unanimity reached by a tribe is overthrown if any one vote is corrupted by noise, which happens with a probability decaying exponentially in b. On the other hand, majority should have fairly good stability, since the vote is often won by a larger margin, and corrupting a few variables doesn't matter.

3 A Mathematician's Defense of Democracy

We later show that the stability of the majority is the rather peculiar quantity $1 - \frac{2}{\pi}\arccos\rho$ (plus terms vanishing with n). However, let us now understand the importance of this quantity. The following theorem, due to Mossel, O'Donnell and Oleszkiewicz [2] is usually known as "majority is stablest":

Theorem 3. *For any $\rho \in [0,1)$ and $\varepsilon > 0$, there is a small enough $\delta > 0$ such that for any balanced $f : \{-1,1\}^n \to [-1,1]$ at least one of the following holds:*

$$\mathbb{S}_\rho(f) \le 1 - \frac{2}{\pi}\arccos\rho + \varepsilon \qquad \textit{or} \qquad (\exists) i : \mathrm{Inf}_i(f) \ge \delta$$

This theorem tells us that one cannot beat the stability of majority (by any $\varepsilon > 0$), without sacrificing a lot: at least one variable has a large (constant) influence. Thus, if we want to avoid dictator- or junta-like behavior, majority is as stable as you can be. It should be noted, however, that majority is not the only function achieving this stability. For example, weighted majorities achieve the same stability, so with regards to this theory it is Okay to have democratic schemes in which members have different weights (e.g. according to wealth or IQ).

4 Stability of Majority

Let us now try to show how one obtains a stability of $1 - \frac{2}{\pi}\arccos\rho + o(1)$ for majority. First, we sketch a formal proof based on the Central Limit Theorem. Then, we give a semiformal geometric argument, which is very similar to the geometric analysis in the Goemans-Williamson algorithm for MaxCut.

The random experiment defining the stability chooses x uniformly, and y to be ρ-correlated with x. This means each coordinate of y differs from the corresponding coordinate in x with probability $p = \frac{1-\rho}{2}$. Since coordinates are independent, a standard concentration argument shows that with $1 - o(1)$ probability, the fraction of differing coordinates is $p + o(1)$. Ignoring $o(1)$ factors in the stability, we can simply condition on the number of differing coordinated to be some $d = (p + o(1)) \cdot n$.

Now we can express the random experiment as follows. First choose d coordinates where x and y will differ. By symmetry, this choice is irrelevant, so let's say coordinates $1, \dots, d$ differ. Then, choose common values for coordinates $d+1, \dots, n$. Finally, choose values for x on coordinates $1, \dots, d$; the values for y are the negations. The majority of x will be different from the majority of y iff $\left|\sum_{i=1}^{d} x_i\right| \geq \left|\sum_{i=d+1}^{n} x_i\right|$. But by the Central Limit Theorem, both of these sums behave are normals (within $o(1)$ error). The remaining task is computational (integrating normal distributions), and we ignore the details.

Now let us switch to the geometric view. Viewing x and y as vectors, we can compute the angle between them. This is $\arccos\frac{\langle x,y\rangle}{\|x\|\cdot\|y\|} = \arccos\frac{n-2d}{\sqrt{n}\cdot\sqrt{n}} = \arccos\rho + o(1)$; here $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ denote the standard Euclidean inner product and norm. The majority function (and in general, any weighted majority function) can be described by a hyperplane going through the origin. Points on one side receive an answer of $+1$, and points on the other side an answer of -1. Projecting on the 2-dimensional plane defined x and y, this separation hyperplane becomes a line. The majorities of x and y are different iff the line falls inside the angle between x and y.

In the Goemans-Williamson algorithm, the separation hyperplane is random, while the two vectors are fixed. In our case, the hyperplane is fixed, but x and y are chosen randomly (at a fixed angle from each other). If they were chosen from the sphere in n dimensions, we would be done: the probability they are separated would be $\frac{\arccos\rho}{\pi} + o(1)$, translating into the desired stability. Intuitively, the n-dimensional cube behaves very much like the sphere in n dimensions with regard to projections in 2 dimensions, so the same result should hold.

5 Fourier Analysis on the Hypercube

Consider the space of all functions $\{f : \{\pm 1\}^n \to \Re\}$. We can look at this as a linear space with the operation $(f \cdot g)(x) = f(x) \cdot g(x)$. Furthermore, we can define an inner product on this space by $\langle f, g\rangle = \mathbb{E}_x[f(x)g(x)]$. In Euclidean space, the inner product describes how close the two vectors are (i.e. how small an angle they make). This intuition extends to functions, since our inner product describes the correlation between f and g (at least for balanced f and g).

Now for any $S \subseteq [n]$, define the *character* $\chi_S(x) = \prod_{i \in S} x_i$. By computation, it can be shown that the χ_S's form an orthonormal basis of our linear space. This means that we can express any function as a weighted sum of the characters: $f = \sum_S \hat{f}(S)\chi_S$. The weights $\hat{f}(S)$ are called the *Fourier coefficients* of f. By orthonormality of the basis, a Fourier coefficient can be computed by $\hat{f}(S) = \langle f, \chi_S \rangle$. Furthermore, the inner product can be computed as $\langle f, g \rangle = \sum_S \hat{f}(S)\hat{g}(S)$.

To draw the connection to Fourier analysis over the reals, note that the characters there are sinusoidals at different frequencies. In the appropriate setup, these also form an orthonormal basis, and functions can be expressed in this basis.

6 Connections to Influence and Stability

To see why Fourier analysis over the cube comes in, let us first draw the connection to stability. For this, we need the following definition:

Definition 4. *The* Bonami-Beckner operator T_ρ *is a linear map on the space of functions taking* f *to* $T_\rho f$ *satisfying* $T_\rho f(x) = \mathbb{E}[f(y)]$*, where* y *is a noisy* ρ*-correlated version of* x*.*

We can see T_ρ as "error-correcting" or "antialiasing" f: each value is decided by averaging values from the neighbors. By inspection of the definition of stability, we see $\mathbb{S}_\rho(f) = \langle f, T_\rho f \rangle$. To understand this intuitively, remember that stability is measuring the correlation (a.k.a. inner product) between f at some input x and at a noisy version of x. The average of the noisy copies is centralized at x by T_ρ. Finally, we have to connect the Bonami-Beckner operator to the Fourier coefficients:

Proposition 5. $T_\rho f = \sum_S \rho^{|S|} \cdot \hat{f}(S)\chi_S$.

Proof. $T_\rho f(x) = \mathbb{E}_y[f(y)] = \mathbb{E}_y[\sum_S \hat{f}(S)\chi_S(y)] = \sum_S \hat{f}(S)\,\mathbb{E}_y[\chi_S(y)] = \sum_S \hat{f}(S)\,\mathbb{E}_y[\prod_{i \in S} y_i] = \sum_S \hat{f}(S) \prod_{i \in S} E[y_i] = \sum_S \hat{f}(S) \prod_{i \in S}(\rho x_i) = \sum_S \rho^{|S|} \cdot \hat{f}(S)\chi_S(x)$. $\square$

Having established a connection to stability, let us consider influence:

Proposition 6. $\mathrm{Inf}_i(f) = \sum_{S: i \in S} \hat{f}(S)^2$.

Proof. For ease of notation, assume $i = 1$. Then

$$\mathrm{Inf}_1(f) = \mathop{\mathbb{E}}_{x_2,\ldots,x_n} [\mathop{\mathbf{var}}_{x_1}[f(x)]] = \mathop{\mathbb{E}}_{x_2,\ldots,x_n} \left[\mathop{\mathbb{E}}_{x_1}[f^2(x)] - \mathop{\mathbb{E}}_{x_1}[f(x)]^2 \right] = \mathop{\mathbb{E}}_{x}[f^2(x)] - \mathop{\mathbb{E}}_{x_2,\ldots,x_n} [\mathop{\mathbb{E}}_{x_1}[f(x)]^2]$$

To evaluate the first term, note $\mathbb{E}_x[f^2(x)] = \langle f, f \rangle = \sum_S \hat{f}(S)^2$. For the second term, note $\mathbb{E}_{x_1}[f(x)] = \mathbb{E}_{x_1}[\sum_S \hat{f}(S)\chi_S(x)] = \sum_{S: 1 \notin S} \hat{f}(S)\chi_S(x)$, because when $1 \in S$, the expectation of $\chi_S(x)$ is zero. Then

$$\mathop{\mathbb{E}}_{x_2,\ldots,x_n} [\mathop{\mathbb{E}}_{x_1}[f(x)]^2] = \mathop{\mathbb{E}}_{x_2,\ldots,x_n} \left[\left(\sum_{S: 1 \notin S} \hat{f}(S)\chi_S(x) \right)^2 \right] = \mathop{\mathbb{E}}_{x} \left[\left(\sum_{S: 1 \notin S} \hat{f}(S)\chi_S(x) \right)^2 \right]$$

because x_1 does not appear inside the expectation. Finally, this is the inner product of $\sum_{S:1\notin S}\hat{f}(S)\chi_S(x)$ with itself, so it is $\sum_{S:1\notin S}\hat{f}(S)^2$. We conclude that $\mathrm{Inf}_1(f)=\sum_S\hat{f}(S)^2-\sum_{S:1\notin S}\hat{f}(S)^2=\sum_{S:1\in S}\hat{f}(S)^2$. □

7 Proving Something Useful

Having developed all this theory, let us finally prove a technical result that we shall use in the next lecture. Remember that for functions which are noticeably more stable than majority, the majority-is-stablest theorem guarantees a variable with constant influence. But on the upper-bound side, there could be as many as n variables with constant influence (remember parity). To get a better upper bound, we slightly change the notion of influence:

Definition 7. *The k-degree influence of coordinate i on f is* $\mathrm{Inf}_i^{\le k}(f)=\sum_{S:i\in S,|S|\le k}\hat{f}(S)^2$.

This definition should be contrasted to Proposition 6.

For f valued in $\{\pm 1\}$, observe that $\sum_i \mathrm{Inf}_i^{\le k}(f)\le k$. Indeed, the sum includes each Fourier coefficient at most k times, so it is bounded by $k\sum_S\hat{f}(S)^2=k\langle f,f\rangle=k$. Then, there can be at most $O(k)$ variables having a k-degree influence of $\Omega(1)$.

To make k-degrees influences useful, we must show that majority-is-stablest continues to hold when replacing $\mathrm{Inf}_i(f)$ by $\mathrm{Inf}_i^{\le k}(f)$, for some constant k:

Proposition 8. *For any $\rho\in[0,1)$ and $\varepsilon>0$, there exist constants $\delta>0$ and k such that for any balanced $f:\{-1,1\}^n\to[-1,1]$ at least one of the following holds:*

$$\mathbb{S}_\rho(f)\le 1-\frac{2}{\pi}\arccos\rho+\varepsilon \qquad \text{or} \qquad (\exists)i:\mathrm{Inf}_i^{\le k}(f)\ge\delta$$

Proof. We apply Theorem 3 to $g=T_{1-\gamma}f$, where $\gamma>0$ is a small enough constant. The intuition is that this makes a Fourier coefficient $\hat{f}(S)$ decrease by $(1-\gamma)^{|S|}$, so for large enough $|S|$, it becomes negligible, and hence we can only look at influence up to some degree. At the same time, we can show that stability doesn't change too much if γ is small enough.

Note that $\mathbb{E}_x[g(x)]=\mathbb{E}_x[\mathbb{E}_y[f(y)]]=\mathbb{E}_y[f(y)]=0$. In other words, g is unbalanced because a noisy ρ-correlated version of a uniformly random vector is uniformly random.

We use Theorem 3 with $\varepsilon'=\varepsilon/4$. Let δ' be the constant it gives for influences, and define $\delta=\delta'/2$. Note that:

$$\mathrm{Inf}_i(g)=\sum_{S:i\in S}(1-\gamma)^{2|S|}\hat{f}(S)^2\le\sum_{S:i\in S,|S|\le k}\hat{f}(S)^2+(1-\gamma)^{2k}\sum_{S:i\in S,|S|\le k}\hat{f}(S)^2\le\mathrm{Inf}_i^{\le k}(f)+(1-\gamma)^{2k}$$

By choosing k large enough (depending on γ), the second term can be at most $\delta'/2$. Then, $\mathrm{Inf}_i(g)\ge\delta'$ implies $\mathrm{Inf}_i^{\le k}(f)\ge\delta'/2=\delta$.

The other possibility is that $\mathbb{S}_\rho(g)\le 1-\frac{2}{\pi}\arccos\rho+\frac{\varepsilon}{4}$. Note that:

$$\mathbb{S}_\rho(g)=\langle g,T_\rho g\rangle=\sum_S\rho^{|S|}\hat{g}(S)^2=\sum_S\rho^{|S|}\left((1-\gamma)^{|S|}\hat{f}(S)\right)^2$$

But now we have:

$$\mathbb{S}_\rho(f) = \sum_S \rho^{|S|} \hat{f}(S)^2 = \mathbb{S}_\rho(g) + \sum_S \rho^{|S|} \hat{f}(S)^2 \cdot \left(1 - (1-\gamma)^{2|S|}\right)$$

The first term is at most $1 - \frac{2}{\pi} \arccos \rho + \frac{\varepsilon}{4}$. By making γ a small enough constant (depending on ρ and ε), we can make $\rho^t(1-(1-\gamma)^{2t}) \le \frac{3}{4}\varepsilon$, for any $t \ge 1$. Then, the second term is at most $\frac{3}{4}\varepsilon$, and we have $\mathbb{S}_\rho(f) \le 1 - \frac{2}{\pi} \arccos \rho + \varepsilon$. □

References

[1] J. Kahn, G. Kalai and N. Linial: *The influence of variables on boolean functions.* FOCS'88, pp. 68–80

[2] Elchanan Mossel, Ryan O'Donnell, Krzysztof Oleszkiewicz: *Noise stability of functions with low influences invariance and optimality.* FOCS'05, pp. 21–30