CS 880: Advanced Complexity Theory 3/24/2008

Lecture 22: Biased Harmonic Analysis

Instructor: Dieter van Melkebeek Scribe: Baris Aydinlioglu

In the next two lectures we will discuss threshold phenomena, the study of which originated in statistical physics and later turned out fruitful in computer science. Today we lay the groundwork for our study, namely Biased Harmonic Analysis. In the next lecture we will use today's results to prove that every non-trivial monotone graph property has a sharp threshold.

Our study of Harmonic Analysis up to this point has been with respect to the uniform distribution. Today we move to the more general setting of the p-biased distribution, where each bit of a function's input is set to -1 independently with probability p (hence the word "Biased"). Most of the results that we've obtained in the uniform setting can also be obtained more generally in the biased setting. We go over these first. Then we obtain an important result that we will use in the next lecture.

1 Setting

Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be a monotone function. Define

$$\mu_p(f) = \Pr_p[f(x) = -1],$$

where p in the subscript denotes the fact that each bit x_i of the input x to f is independently set to -1 with probability p. Note that for fixed f, $\mu_p(f)$ is a continuous function of p, since it can be written as the sum of at most 2^n polynomials in p of degree at most n.

Our study of threshold phenomena can be summed up in one question:

How does $\mu_p(f)$ behave as p changes from 0 to 1?

In particular, we are interested in the case where the input x to f represents the characteristic string of a graph. For a graph $G = (V, E)$ its characteristic is the string x of length $\binom{|V|}{2}$ where $x_i = -1$ iff the ith edge is present in G. A *monotone graph property* is then a property which is invariant under graph isomorphism and which, once satisfied by G, continues to be satisfied as we add more edges to G. Hamiltonicity and connectivity are two examples.

Proposition 1. *If f is monotone then $\mu_p(f)$ is monotone (in p).*

Proof. By induction on n, the length of f's argument, x. The base case, $n = 1$, is clear. In the inductive step we condition on the value of x_1 and use independence of the bits of x:

$$\mu_p(f) = p\,\mu_p(f|_{x_1=-1}) + (1-p)\,\mu_p(f|_{x_1=1})$$

taking derivatives w.r.t. p we obtain

$$\begin{aligned}\frac{d\mu_p(f)}{dp} &= p\,\frac{d\mu_p(f|_{x_1=-1})}{dp} + (1-p)\,\frac{d\mu_p(f|_{x_1=1})}{dp} + \mu_p(f|_{x_1=-1}) - \mu_p(f|_{x_1=1}) \\ &\geq \mu_p(f|_{x_1=-1}) - \mu_p(f|_{x_1=1}) \\ &\geq 0.\end{aligned}$$

We obtain the first inequality by applying the induction hypothesis to the first two terms, which we can do because any restriction of f is also monotone. The second inequality follows from the monotonicity of f: the event $f|_{x_1=1} = -1$ implies the event $f|_{x_1=-1} = -1$ (where we view $f|_{x_1=1}$ and $f|_{x_1=-1}$ as random variables on the sample space of values for $x_2, \dots, x_n$.) □

Remark: If the domain of f is infinite, $\mu_p(f)$ need not be differentiable. Even in this case Proposition 1 holds, although we need a different argument. We don't have to worry about this, however; in our finite setting $\mu_p(f)$ is always differentiable.

Proposition 1 and continuity imply that for a monotone f that is not constant, $\mu_p(f)$ is zero for $p = 0$ and one for $p = 1$ and takes all values in the range $[0, 1]$ in between. In fact, the proof of Proposition 1 then shows that μ_p is strictly increasing. This leads to the following definition.

Definition 1. *For a non-constant and monotone function f, its* critical probability p_c *is defined as the unique value of $p \in (0, 1)$ for which $\mu_p(f) = \frac{1}{2}$.*

The study of threshold phenomena concerns the behavior $\mu_p(f)$ around p_c for various f. If $\mu_p(f)$ grows rapidly around p_c then f is said to have a *sharp threshold*. Note that for a balanced function f we have $p_c = \frac{1}{2}$. In that case, we can use the standard harmonic analysis we have developed so far to analyze the threshold behavior. There are many interesting functions, however, for which $p_c \neq \frac{1}{2}$. To analyze their threshold behavior we need to develop biased harmonic analysis.

2 Biased Harmonic Analysis

In the uniform setting, many of the results that we developed hinged on the existence of a basis χ_S, $S \subseteq [n]$, for the space of all functions from $\{-1, 1\}$ to $\mathbb{R}$ with the following properties.

(i) The basis functions are orthonormal with respect to the inner product $\langle f, g \rangle = \mathrm{E}[f(x)g(x)]$.

(ii) The function χ_S can be expressed as the product $\chi_S(x) = \prod_{i \in S} \chi(x_i)$, where $\chi(b) = b$ for $b \in \{-1, 1\}$.

We want to maintain generalized versions of these properties in the p-biased setting so that many of the results from the uniform setting carry over. For that reason, we would like there to be a basis ϕ_S, $S \subseteq [n]$, for the space of functions from $\{-1, 1\}^n$ to $\mathbb{R}$ such that the following hold.

(p-i) The basis functions are orthonormal with respect to the inner product $\langle f, g \rangle_p = \mathrm{E}_p[f(x)g(x)]$.

(p-ii) The functions ϕ_S can be expressed as the product $\phi_S(x) = \prod_{i \in S} \phi(x_i)$ for some fixed function $\phi : \{-1, 1\} \to \mathbb{R}$.

We now derive what the function ϕ must be.

Claim 1. *The following two conditions on ϕ are necessary and sufficient:*

$$\mathrm{E}_p[\phi] = 0, \tag{1}$$
$$\mathrm{E}_p\left[\phi^2\right] = 1. \tag{2}$$

Proof. (1) is necessary: Property (p-ii) implies $\phi_\emptyset = 1$, because for $S = \emptyset$ the (empty) product $\prod_{i\in S} \phi(x_i)$ is just 1. Property (p-i) implies $\langle \phi_S, \phi_\emptyset \rangle_p = 0$ for non-empty S, in particular for S a singleton, say $S = \{i\}$. Therefore $0 = \langle \phi_{\{i\}}, \phi_\emptyset \rangle_p = \mathrm{E}_p\big[\phi_{\{i\}} \cdot 1\big] = \mathrm{E}_p[\phi]$. (Note that the last expectation is over 1 bit while the preceding expectation is over n bits.)

(2) is necessary: $\mathrm{E}_p\big[\phi^2\big] = \mathrm{E}_p[\phi \cdot \phi] = \mathrm{E}_p\big[\phi_{\{i\}} \cdot \phi_{\{i\}}\big] = \langle \phi_{\{i\}}, \phi_{\{i\}} \rangle_p = 1$. (The first and second expectation is over 1 bit and the third is over n bits.)

(1) and (2) are sufficient:

$$\begin{aligned}\langle \phi_S, \phi_T \rangle &= \mathrm{E}_p\Big[\textstyle\prod_{i\in S\triangle T} \phi(x_i) \cdot \prod_{i \in S\cap T} (\phi(x_i))^2\Big] \\ &= \Big(\prod_{i\in S\triangle T} \mathrm{E}_p[\phi(x_i)]\Big)\Big(\prod_{i\in S\cap T} \mathrm{E}_p\big[(\phi(x_i))^2\big]\Big),\end{aligned}$$

which follows by using the independence of bits. Now, by (1), the first product in the last expression is zero whenever it is non-empty, while the second product is always one, by (2). Therefore we have orthonormality. ☐

Solving (1) and (2) we obtain two solutions for ϕ, namely $(\phi(-1), \phi(1)) = (\pm\sqrt{\frac{q}{p}}, \mp\sqrt{\frac{p}{q}})$. To be consistent with the uniform setting, where we have $\chi(x_i) = x_i$, we set $\phi(-1)$ negative.

Definition 2. *For any fixed $p \in (0,1)$ and every $S \subseteq [n]$, we define $\phi_S(x) = \prod_{i\in S}(x_i)$, where $\phi : \{-1,1\} \to \mathbb{R}$ is given by $\phi(-1) = -\sqrt{\frac{q}{p}}$ and $\phi(1) = \sqrt{\frac{p}{q}}$, with $q = 1-p$.*

The functions ϕ_S in Definition 2 satisfy the properties (p-i) and (p-ii). Now that we have attained those two crucial properties of the uniform setting in the general p-biased setting, we revisit the results from the uniform setting that carry over.

2.1 Fourier Expansion and Elementary Properties

Since we have an orthonormal basis ϕ_S, every function f can be written as

$$f = \sum_{S\subseteq[n]} \widetilde{f}(S)\, \phi_S, \text{ where } \widetilde{f}(S) = \mathrm{E}_p[f \cdot \phi_S].$$

We denote the Fourier coefficient of f in the p-biased setting by $\widetilde{f}$ to distinguish it from its uniform version $\widehat{f}$.

As an immediate consequence, all of the following elementary properties from the uniform setting generalize to the p-biased setting.

- Plancherel: $\langle f, g \rangle_p = \sum_S \widetilde{f}(S)\widetilde{g}(S)$
- Parseval: $\langle f, f \rangle_p = \sum_S (\widetilde{f}(S))^2$
- $\mathrm{E}_p[f] = \widetilde{f}(\emptyset)$
- $\sigma_p^2[f] = \sum_{S\neq\emptyset} (\widetilde{f}(S))^2$

The proofs are left as exercises.

2.2 Influence

Influence is generalized in the natural way:

$$I_i^{(p)}(f) = \Pr_p\left[f(x) \neq f(x^{(i)})\right], \text{ where } x^{(i)} \text{ is } x \text{ with its } i\text{-th bit flipped.}$$

$$I^{(p)}(f) = \mathrm{E}_p[\#\text{ neighbors with different value under } f] = \sum_{i=1}^{n} I_i^{(p)}(f).$$

Proposition 2. *For any* f

$$I_i^{(p)}(f) = \tfrac{1}{4pq}\sum_{S\ni i}(\widetilde{f}(S))^2.$$

Moreover, if f *is monotone then*

$$I_i^{(p)}(f) = \tfrac{1}{2\sqrt{pq}}\,\widetilde{f}(\{i\}).$$

Proof. As in the uniform setting we make use of the difference operator, $D_i f$:

$$(D_i f)(x) = \frac{f(x^{(i=1)}) - f(x^{(i=-1)})}{2}.$$

For any $S \subseteq [n]$, if $i \notin S$ then $D_i\phi_S = 0$ since $\phi_S(x^{(i=1)}) = \phi_S(x^{(i=-1)})$. If $i \in S$, we can write $\phi_S(x) = \phi_{\{i\}}\phi_{S\setminus\{i\}}(x) = \phi(x_i)\phi_{S\setminus\{i\}}(x)$, so we get $D_i\phi_S = \big(\sqrt{\frac{p}{q}} + \sqrt{\frac{q}{p}}\big)\phi_{S\setminus\{i\}} = \frac{1}{2\sqrt{pq}}\phi_{S\setminus\{i\}}$. By linearity it follows that

$$D_i f = \tfrac{1}{2\sqrt{pq}}\sum_{S\ni i}\widetilde{f}(S)\phi_{S\setminus\{i\}}. \tag{3}$$

As a sanity check, note that $p = q = 1/2$ gives the original expression we obtained for $D_i f$ in the uniform setting.

We conclude that

$$\begin{aligned} I_i^{(p)}(f) &= \Pr_p\left[f(x) \neq f(x^{(i)})\right] \\ &= \mathrm{E}_p\left[(D_i f)^2\right] \\ &= \tfrac{1}{4pq}\sum_{S\ni i}(\widetilde{f}(S))^2. \end{aligned}$$

If f is monotone, then $D_i(f)$ is nonnegative and thus

$$\begin{aligned} I_i^{(p)}(f) &= \Pr_p\left[f(x) \neq f(x^{(i)})\right] \\ &= \mathrm{E}_p[D_i f] \\ &= \tfrac{1}{2\sqrt{pq}}\sum_{S\ni i}\widetilde{f}(S)\,\langle\phi_{S\setminus\{i\}}, \phi_\emptyset\rangle_p \\ &= \tfrac{1}{2\sqrt{pq}}\,\widetilde{f}(\{i\}). \end{aligned} \tag{4}$$

□

2.3 More Advanced Properties

We discuss the generalizations of some of the more advanced results we derived in the uniform settin. We omit the proofs of the generalizations to the p-biased setting.

- Hypercontractivity: A precise formulation of this property in the p-biased setting is rather involved, and we will not need it in that detail. We just mention that hypercontractivity carries over to the p-biased setting, which allows the next two properties to be also carried over.

- Approximability by juntas: Every f differs on a set of measure $\leq \epsilon$ from some r-junta, where $r = \left(\frac{1}{pq}\right)^{cpqI^{(p)}(f)/\epsilon}$ and c is a universal constant. The measure refers to the p-biased distribution.

- Influential variables: Every f has an $i \in [n]$ such that

$$I_i^{(p)}(f) \geq \frac{d\,\sigma_p^2(f)}{pq\,\log\frac{1}{pq}}\,\frac{\log n}{n},$$

  where d is a universal constant. Note that the variance in the nominator prevents the first fraction from growing unboundedly for small p or q.

3 Threshold Behavior and Influence

We now state and prove a key result relating the slope of μ_p to the influence of the underlying function. We will use it next lecture in our study of threshold phenomena.

Theorem 1. *For any f we have*

$$\frac{d\mu_p(f)}{dp} = \frac{1}{2\sqrt{pq}}\sum_{i=1}^{n}\widetilde{f}(\{i\})$$

Before we prove Theorem 1 we draw a corollary and give some examples.

Corollary 1. *For any monotone f and $p \in (0,1)$ we have*

$$\frac{d\mu_p(f)}{dp} = I^{(p)}(f).$$

If f is a dictator then $I^{(p)}(f) = 1$ for every $p \in (0,1)$, in particular for p_c, and therefore f does not have a sharp threshold. Indeed, the graph of $\mu_p(f)$ is just a line.

If f is majority then $I^{(p)}(f) = \Theta(\sqrt{n})$ for $p = p_c = 1/2$, which asymptotically grows very large. Therefore, the majority function has a sharp threshold.

If f is a tribes function then $I^{(p)}(f) = \Theta(\log n)$ for $p = p_c$, which grows slower than the case of majority but still grows very large asymptotically. Therefore, the tribes function also has a sharp threshold.

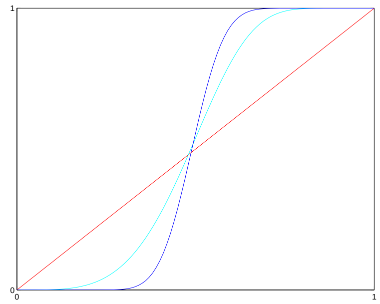

Figure 1: Graphs of μ_p for dictator, tribes, and majority functions.

Proof of Theorem 1. By using $\mathrm{E}_p[f] = \mathrm{Pr}_p[f = 1] - \mathrm{Pr}_p[f = -1]$, we can write

$$\mu_p(f) = \frac{1 - \mathrm{E}_p[f]}{2},$$

and work with $\mathrm{E}_p[f]$ instead $\mu_p(f)$.

One way to obtain a p-biased distribution is to start from the all-ones vector and flip each bit with probability p. Therefore we can write

$$\begin{aligned} \mathrm{E}_p[f] &= (T_\alpha f)(1, 1, \ldots, 1), \qquad \text{where } p = \frac{1-\alpha}{2} \qquad (5) \\ &= \sum_{S \subseteq [n]} \alpha^{|S|} \widehat{f}(S) \chi_S(1, 1, \ldots, 1) \\ &= \sum_{S \subseteq [n]} \alpha^{|S|} \widehat{f}(S). \end{aligned}$$

Note that we used original Fourier coefficients and characters.

Differentiating,

$$\begin{aligned} \frac{d\mathrm{E}_p[f]}{dp} &= -2 \sum_{S \subseteq [n]} |S| \alpha^{|S|-1} \widehat{f}(S) && (p \text{ is hidden in } \alpha) \\ &= -2 \sum_{i=1}^{n} \sum_{S \ni i} \alpha^{|S|-1} \widehat{f}(S) && \text{(regrouping terms)} \\ &= -2 \sum_{i=1}^{n} \underbrace{\sum_{S \ni i} \alpha^{|S|-1} \widehat{f}(S) \chi_{S \setminus \{i\}}(1, 1, \ldots, 1)}_{(*)} \\ &= -2 \sum_{i=1}^{n} \big(T_\alpha(D_i f)\big)(1, 1, \ldots, 1) && \big(\text{claim: } (*) = \big(T_\alpha(D_i f)\big)(1, 1, \ldots, 1)\big) \\ &= -2 \sum_{i=1}^{n} \mathrm{E}_p[D_i f] && \text{(same trick as in (5))} \\ &= \frac{-1}{\sqrt{pq}} \sum_{i=1}^{n} \widetilde{f}(\{i\}) && \text{(same as in (4))} \end{aligned}$$

In order to establish (∗), it suffices to observe the uniform equivalent of (3), namely

$$D_i f = \sum_{S \ni i} \hat{f}(S) \chi_{S \setminus \{i\}},$$

and apply the linearity of the operator T_α. □

4 Next Lecture

We will prove that every non-trivial monotone graph property has a sharp threshold.