Analysis of Boolean Functions **(CMU 18-859S, Spring 2007)**

Lecture 24: Russo-Margulis Lemma, Graph threshold properties

April 12, 2007

Lecturer: Ryan O'Donnell *Scribe: Amitabh Basu*

We study threshold phenomena in p-biased boolean functions in this lecture. As an application, this kind of analysis gives us tools to investigate threshold phenomena in random graph structures. This approach to random graph structures has a very different flavor from the techniques developed in the random graph community. Before we begin our discussion, let us recall some of the key results about p-biased boolean functions from the last class. For $f : \{T, F\}_p^n \to \Re$ (recall that this means $\mathbf{Pr}[T] = p$ and $\mathbf{Pr}[F] = 1 - p = q$),

- $f = \sum_{S \subseteq [n]} \tilde{f}(S) \cdot \phi_S$, where ϕ_S is the p-biased mapping defined in the previous lecture.
- $D_i^{(p)} f = \sum_{S \ni i} \tilde{f}(S) \cdot \phi_{S \setminus \{i\}}$. Also, $D_i^{(p)} f = \sqrt{pq}(f(x^{(i=F)}) - f(x^{(i=T)}))$
- $Inf_i(f) = E_p[(D_i f)^2] = \sum_{S \ni i} \tilde{f}(S)^2$
- $\mathbb{I}_p(f) = \sum_{i=1}^n Inf_i(f) = \sum_{S \subseteq [n]} |S| \tilde{f}(S)^2$
- If $f : \{T, F\}_p^n \to \{-1, 1\}$, and is monotone (which means flipping F to T in the inputs can only flip f's value from 1 to -1), then $Inf_i(f) = 2\sqrt{pq}\tilde{f}(i)$

1 Russo-Margulis Lemma

Imagine fixing some monotone function $f : \{T, F\}^n \to \{-1, 1\}$, and varying p. Monotonicity ensures that as p goes from 0 to 1, $\mathbf{Pr}_p[f = -1]$ increases from 0 to 1 (unless f is a constant). See Figure 1.

We want to study the behavior of this "probability-plot". In particular, we want to find out if the function turns "mostly" -1 from "mostly" 1 in a very small interval (which is like a threshold phenomenon); or whether the change is more gradual. For example, think of the function as indicating connectedness in a graph (think of T and F as corresponding to the edge being present or not). If we can prove that this function has a "probability-plot" like that of a step function, we would have shown the threshold phenomena for connectedness in random graphs.

For studying such behavior of p-biased functions, the following theorem is of central importance.

Theorem 1.1 ("Russo-Margulis Lemma") *If $f : \{T, F\} \to \Re$,*

$$\frac{d}{dp}\mathbf{E}_p[f] = -\frac{1}{pq}\sum_{i=1}^{n} \tilde{f}(i)$$

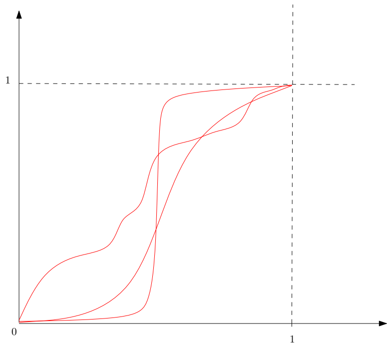

Figure 1: Plots of $\mathbf{Pr}_p[f = -1]$ v/s p

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathbf{Pr}[f = -1] = \frac{\mathrm{d}}{\mathrm{d}p}[\tfrac{1}{2} - \tfrac{1}{2}\mathbf{E}_p[f]] = \frac{1}{2\sqrt{pq}}\sum_{i=1}^{n}\tilde{f}(i)$$

Moreover, if f is monotone,

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathbf{Pr}[f = -1] = \frac{1}{2\sqrt{pq}}\sum_{i=1}^{n}\frac{Inf_i(f)}{2\sqrt{pq}} = \frac{1}{4pq}\mathbb{I}_p(f)$$

Proof: In input, identify T with -1 and F with 1. Now think of $f : \{-1, 1\}^n \to \Re$ in the usual way under the uniform distribution. Recall that

$$T_{1-2p}f(x) = \mathbf{E}_{y\sim_{1-2p}x}[f(y)]$$

Therefore,

$$T_{1-2p}f(\vec{1}) = \mathbf{E}_{y\sim_{1-2p}\vec{1}}[f(y)] = \mathbf{E}_p[f]$$

Also, $T_{1-2p}f = \sum_{S\subseteq[n]}(1-2p)^{|S|}\hat{f}(S)\chi_S$. Therefore,

$$\begin{aligned}
T_{1-2p}f(\vec{1}) &= \sum_{S\subseteq[n]}(1-2p)^{|S|}\hat{f}(S) \\
\Rightarrow \frac{\mathrm{d}}{\mathrm{d}p}\mathbf{E}_p[f] &= \frac{\mathrm{d}}{\mathrm{d}p}(\sum_{S\subseteq[n]}(1-2p)^{|S|}\hat{f}(S)) \\
&= -2\sum_{S\subseteq[n]}|S|(1-2p)^{|S|-1}\hat{f}(S) \\
&= -2\sum_{i=1}^{n}\sum_{S\ni i}(1-2p)^{|S|-1}\hat{f}(S) \qquad (1)
\end{aligned}$$

Now note that $T_{1-2p}D_i f = \sum_{S\ni i}(1-2p)^{|S|-1}\hat{f}(S)\chi_{S\setminus\{i\}}$. Therefore,

$$\begin{aligned}\sum_{S\ni i}(1-2p)^{|S|-1}\hat{f}(S) &= T_{1-2p}D_i f(\vec{1}) \\ &= \mathbf{E}_p[D_i f] \\ &= \mathbf{E}_p[\tfrac{1}{2}(f(x^{(i=F)})-f(x^{(i=T)}))]\end{aligned}$$

Substituting into (1),

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p}\mathbf{E}_p[f] &= -2\sum_{i=1}^{n}\mathbf{E}_p[\tfrac{1}{2}(f(x^{(i=F)})-f(x^{(i=T)}))] \\ &= \frac{-1}{\sqrt{pq}}\sum_{i=1}^{n}\mathbf{E}_p[\sqrt{pq}(f(x^{(i=F)})-f(x^{(i=T)}))] \\ &= \frac{-1}{\sqrt{pq}}\sum_{i=1}^{n}\mathbf{E}_p[D_i^{(p)}f] \\ &= \frac{-1}{\sqrt{pq}}\sum_{i=1}^{n}\tilde{f}(i)\end{aligned}$$

$\square$

2 Threshold phenomena for monotone functions

We will now formalize the notion of threshold phenomena.

Definition 2.1 *Given a non-trivial (non-constant) monotone* $f:\{T,F\}^n\to\{-1,1\}$*, define critical probability* $p_c(f)$ *to be the* p *such that* $\mathbf{Pr}_{p_c}[f=-1]=\frac{1}{2}$

Definition 2.2 f *has a "sharp threshold" if* $\frac{d}{dp}\mathbf{Pr}[f=-1]$ *at the the critical probability* p_c *is "large" compared to min(*p_c, q_c*).*

Now recall that $\frac{\mathrm{d}}{\mathrm{d}p}\mathbf{Pr}[f=-1]=\frac{\mathbb{I}_{p_c}(f)}{4p_cq_c}$. So equivalent to the definition, we want $\mathbb{I}_{p_c}(f)$ to be "large". This would imply that in a small interval around p_c (say $[p_c, p_c(1+\delta)]$), the change in the probability (which would equal $\approx \mathbb{I}_{p_c}(f)\delta$) is huge.

Remark 2.3 *We measure the "largeness" of the derivative with respect to* p_c *because we want it to be "sharp" irrespective of scaling. So if* p_c *happens to be small, the change occurs in a small interval like* $[p_c, p_c(1+\delta)]$*.*

If f does not have a "sharp" threshold, it is said to be "coarse". Usually, "coarse" will mean $\mathbb{I}_{p_c}(f)=O(1)$ and "sharp" will correspond to $\mathbb{I}_{p_c}(f)=\omega(1)$.

We illustrate these concepts with a few examples and then inspect graph threshold properties in this framework.

- $f(x) = x_1$. $p_c = \frac{1}{2}$. $\mathbb{I}_{p_c}(f) = 1$. So this function is "coarse"

- f = Majority. $p_c = \frac{1}{2}$. $\mathbb{I}_{p_c}(f) \approx \sqrt{\frac{2}{\pi}}\sqrt{n}$. So Majority has a "sharp" threshold. In general, let f be T if more than $p_0 n$ of the inputs are T, and F otherwise. Then $p_c = p_0$. $\mathbb{I}_{p_0}(f) \approx 2\sqrt{p_0 q_0}\sqrt{\frac{2}{\pi}}\sqrt{n}$. So it sharp as long as $p_0 \geq \omega(\frac{1}{n})$

As described earlier, it is easy to model random graphs in the framework of p-biased boolean functions. Let the vertex set of the graph be V. Let $n = \binom{|V|}{2}$. We can then identify $x \in \{T, F\}^n$ with a graph on V (T corresponds to the edge being present and F corresponds to the edge being absent). Moreover, if x is drawn from a p-biased distribution, we get the random graph $G(V, p)$. Now any boolean function indicates a collection of graphs. For example, say $f(x) = -1$ is x contains a clique of size $\lfloor 2 \ln v \rfloor$. It is well known that $p_c = \frac{1}{2}$. Also, it can be shown that $\mathbb{I}_{\frac{1}{2}}(f) = \theta(\log^2 n)$. So this property of graphs has a "sharp" threshold : At $p \leq \frac{1}{2}$ we will almost surely *not* have a clique of size $\lfloor 2 \ln v \rfloor$ and at $p \geq \frac{1}{2}$ we will almost surely see a clique of that size. The above discussion motivates another definition.

Definition 2.4 $f : \{T, F\}^{\binom{V}{2}} \to \{T, F\}$ *is a monotone graph property if :*

1. f *is monotone*

2. f *is not constant*

3. f *is invariant under permutation of the vertices.*

Some examples of monotone graph properties are connectedness, contains a k-clique, contains a hamiltonian path, is *NOT* k-colorable, contains at least $\frac{1}{2}\binom{V}{2}$ edges.

Remark 2.5 *Note that monotone graph properties are "weakly symmetric" or "transitive", and therefore all coordinates have the same influence.*

Observation 2.6 *The KKL theorem tells us that for any balanced function, if all the coordinates have the same influence then* $\mathbb{I}_{\frac{1}{2}}(f) \geq \Omega(Var[f]) \cdot \log n$. *Therefore, if* $p_c = \frac{1}{2}$, *we get* $\mathbb{I}_{\frac{1}{2}}(f) \geq \Omega(\log n)$, *and so we have a sharp threshold.*

The only caveat to the above observation is that the KKL/Friedgut theorem relied on the hypercontractivity theorem for $\{T, F\}_{\frac{1}{2}}$. The condition $p_c = \frac{1}{2}$ is therefore crucial. In general, to make such an argument work in the p-biased framework for any arbitrary p_c, we would need a p-biased version of hypercontractivity.

3 p-biased Hypercontractivity and KKL/Friedgut

Recall that the hypercontractivity theorem stated that if $f : \{-1,1\}^n \rightarrow \Re$ is a multilinear polynomial of degree at most d, then $\forall r \geq 2$, $\|f\|_r^2 \leq (r-1)^d \|f\|_2^2$

The p-biased version of the above theorem was proved by Oleszkiewicz.

Theorem 3.1 *Let $f : \{T,F\}_p^n \rightarrow \Re$ of degree $\leq d$ (i.e. $\tilde{f}(S) = 0$ if $|S| \geq d$). Then $\forall r \geq 2$, $\|f\|_r^2 \leq C(r,p)^d \|f\|_2^2$, where $C(r,p) = \frac{\theta^{\frac{1}{r'}} - \theta^{\frac{1}{r'}}}{\theta^{\frac{1}{r}} - \theta^{\frac{1}{r}}}$, $\theta = \frac{p}{q}$, $\frac{1}{r'} + \frac{1}{r} = 1$ and the constant $C(r,p)$ is best possible.*

Note as $p \rightarrow 0$, $C(r,p) \rightarrow (\frac{1}{p}^{1-2/r})$.

Now for the KKL/Friedgut theorems.

Theorem 3.2 *Let $f : \{T,F\}_p^n \rightarrow \{-1,1\}$ and let $\epsilon > 0$. Then f is ϵ-close to a $(\frac{1}{p})^{O(\mathbb{I}_p(f)/\epsilon)}$-junta.*

Theorem 3.3 *Let $f : \{T,F\}_p^n \rightarrow \{T,F\}$. Then $\exists i \in [n]$ with $Inf_i(f) \geq \frac{\Omega(Var[f])}{\log \frac{1}{min(p,q)}} \cdot \frac{\log n}{n}$*

So the bottomline is the following corollary.

Corollary 3.4 *Let $f : \{T,F\}_p^n \rightarrow \{T,F\}$ be a monotone graph property, and p_c be the critical probability. Then KKL implies $\mathbb{I}_p(f) \geq \Omega(\frac{\log n}{\log \frac{1}{min(p,q)}})$. So we have a sharp threshold unless p_c or $q_c \leq \frac{1}{n^{\theta(1)}}$*