Lecture 8

More on the influence of variables on boolean functions

March 4, 2005
Lecturer: Nati Linial
Notes: Neva Cherniavsky & Atri Rudra

In this lecture we will look at the following natural question– do there exist balanced boolean functions f on n variables such that for every variable x the influence of x on f is "small" and how small can this bound be made? ("Balanced" means that $Pr(f = 0) = Pr(f = 1) = \frac{1}{2}$ but $\mathbb{E}f = \alpha$ for some α bounded away from 0 and 1 is just as interesting.) In the last lecture we showed that for the "tribes" function (which was defined by Ben-Or and Linial in [1]), every variable has influence $\Theta(\frac{\log n}{n})$. Today, we will prove the result of Kahn, Kalai and Linial [2] which shows that this quantity is indeed the best one can hope for. In the process we will look into the Bonami Beckner Inequality and will also look at threshold phenomena in random graphs.

8.1 The Kahn Kalai Linial Result

Recall the definition of influence. Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function and let $S \subset [n]$. The influence of the set of variables S on the function f, denoted by $Inf_f(S)$, is the probability that f is still undetermined when all variables in $[n] - S$ have been assigned values at random.

We also talk about influences in the case when the function is defined on a solid cube– $f : [0,1]^n \to \{0,1\}$. This formulation has connections to game theory– variables are controlled by the players. Note that in this case we can talk about things like "influence of a subset of variables towards 0".

The situation for the case when $|S| = 1$ is relatively better understood. As we saw in the last class, the situation for calculating $Inf_f(x)$ looks like Figure 8.1. In particular, calculating the influence is same as counting the number of non-solid edges in Figure 8.1.

The situation is much less understood for the more general case, for example when $|S| = 2$. In a nutshell, we are interested in situations other than those in Figure 8.2. This scenario is not well understood and is still a mysterious object. Unlike the case of a single variable, the number of zeroes and ones in a "mixed" 2-dimensional subcube can vary and the whole situation is consequently more complex. As an interesting special case consider the "tribes" example and the case when S is an entire tribe. It is easy to see that the influence of S towards 1 is exactly 1 (as any tribe can force the result to be 1) while it is not hard to see that the influence of S towards 0 is only $O(\frac{\log n}{n})$. As another example consider the case when S consists of one element from each tribe (one can consider each element of S as a "spy" in a tribe). Here the influence of S

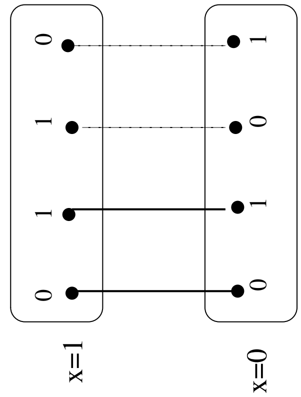

Figure 8.1: Influence of a variable

towards 0 is exactly 1 (as each spy can force its tribe to be 0). Further, its influence towards 1 can be shown to be $\frac{\sqrt{2}-1}{4} + o(1)$.

Let us now turn back to the motivating question for this lecture–

Question 8.1. Find boolean functions f with $\mathbb{E}f = \frac{1}{2}$ for which $Inf_f(x)$ is small for each variable x.

For any variable x_i define $\beta_i = Inf_f(x_i)$ and let $\boldsymbol{\beta} = \langle \beta_1, \cdots, \beta_n \rangle$. Thus, the quantity we are interested in is $||\boldsymbol{\beta}||_\infty$. Note that the edge isoperimetric inequality on the cube implies that $\sum_{i=1}^n \beta_i \geq 1$ which by an averaging argument gives $||\boldsymbol{\beta}||_\infty \geq \frac{1}{n}$. Also note that for the "tribes" example, $||\boldsymbol{\beta}||_\infty = \Theta(\frac{\log n}{n})$. The following result due to Kahn, Kalai and Linial shows that this is the best possible–

Theorem 8.1. *For any* $f : \{-1, 1\} \to \{-1, 1\}$ *with* $\mathbb{E}f = 0$*, there exists a variable* x *such that* $Inf_f(x) \geq \Omega(\frac{\log n}{n})$.

Before we start to prove the theorem, let us collect some facts that we know or follow trivially from what we have covered in previous lectures.

$$\sum_{S \subseteq [n]} (\hat{f}(S))^2 = 1 \tag{8.1}$$

$$\hat{f}(\emptyset) = 0 \tag{8.2}$$

$$\beta_i = 4 \cdot \sum_{S \ni i} (\hat{f}(S))^2 \tag{8.3}$$

$$\sum_{i=1}^n \beta_i = 4 \cdot \sum_{S \subseteq [n]} |S| (\hat{f}(S))^2 \tag{8.4}$$

Equation (8.1) follows from Parseval's identity and the fact that f takes values in $\{-1, 1\}$. Equation (8.2) follows from the fact that $\chi_\emptyset = 1$ which implies $\hat{f}(\emptyset) = \langle f, \chi_\emptyset \rangle = 2^n \mathbb{E}f = 0$. Equation (8.3)was proved in the last lecture. Equation (8.4) follows from summing Equation (8.3) for all i.

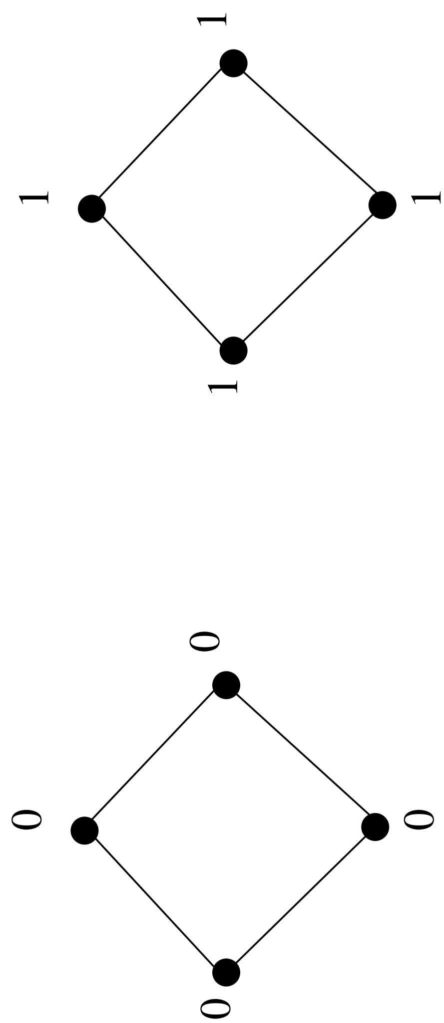

Figure 8.2: Influence of a general subset of variables

We will first show that if “most” of the $\sum_{S\subseteq[n]}(\hat{f}(S))^2$ comes from 'large' sets S then the conclusion of Theorem 8.1 holds. In fact, in this case even the *average* influence is $\Omega(\frac{\log n}{n})$. To be more precise let $T = \frac{\log n}{10}$ and $H = \sum_{|S|\geq T}(\hat{f}(S))^2$. Further assume that $H \geq \frac{1}{10}$. Then by (8.4), $\sum_{i=1}^{n}\beta_i \geq 4\cdot\sum_{|S|\geq T}|S|(\hat{f}(S))^2 \geq 4HT \geq \frac{\log n}{25}$.

It follows that it suffices to prove the theorem under the complementary assumption that $H < \frac{1}{10}$. In view of (8.1) this is the same as showing $\sum_{|S|<T}(\hat{f}(S))^2 > 0.9$. In the proof we will need to estimate the following quantity–

$$\sum_{|S|<T}(\hat{\phi}(S))^2 \equiv \sum_{S} W_T(S)(\hat{\phi}(S))^2$$

for $\phi = f^{(i)}$ (recall from the last lecture that $f^{(i)}(z) = f(z) - f(z\oplus e_i)$). Here $W_T(\cdot)$ is the step function which takes value 1 for any set $S\subseteq[n]$ such that $|S|\leq T$ and 0 otherwise. We use two ideas to solve the problem in hand–

- Try and approximate $W(\cdot)$ by functions which are sufficiently close to the step function $W_T(\cdot)$ and are easy to analyze.
- As the bound $||\boldsymbol{\beta}||_1 \geq 1$ is tight (which is insufficient for our purposes) and since it is difficult to work directly with $||\boldsymbol{\beta}||_\infty$, we could perhaps estimate $||\boldsymbol{\beta}||_p$ for some $\infty > p > 1$. In the proof we will use $p = \frac{4}{3}$ but this is quite arbitrary.

We focus on the second alternative and use the Bonami Beckner inequality which we consider in the next section.

8.2 Bonami Beckner Inequality

Let T_ϵ be a linear operator which maps real functions on $\{-1,1\}^n$ to real functions on $\{-1,1\}^n$. By linear, we mean that the following holds for functions f and g and scalars a and b: $T_\epsilon(a \cdot f + b \cdot g) = a \cdot T_\epsilon(f) + b \cdot T_\epsilon(g)$.

As $T_\epsilon(\cdot)$ is a linear operator, one can fully determine it by just specifying it at the basis of the functions $\{\chi_S\}$. We define the operator as follows

$$T_\epsilon(\chi_S) = \epsilon^{|S|}\chi_S$$

By linearity, $T_\epsilon(f) = \sum_{S\subseteq[n]} \epsilon^{|S|}\hat{f}(S)\chi_S(\cdot)$. Note that $T_1(f) = f$. In other words, T_1 is the identity operator.

We will now state the main result of this section–

Theorem 8.2. *Let $0 < \epsilon < 1$ and consider T_ϵ as an operator from L_p to L_2 where $p = 1 + \epsilon^2$. Then its operator norm is 1.*[1]

Let us first explain the terms used in the statement above. Let $T : (X, ||\cdot||_X) \to (Y, ||\cdot||_Y)$ be a linear operator– here X and Y are normed spaces and $||\cdot||_X$ and $||\cdot||_Y$ are their respective norms. The operator norm of T is defined as

$$||T||_{op} = sup_{x\in X}\frac{||Tx||_Y}{||x||_X}$$

This quantity measures how much the "length" (norm) of an element $x \in X$ can grow by an application of the operator T. We now turn to the proof.

What is the size of f? How expanding is the operator? These are very narrow passages; we have no wiggle room. We can only use Parseval's in L_2, so the norm on the right hand side needs to be L_2. On the left hand side, our norm is L_p, which is usually very difficult to calculate. But because our functions (the $f^{(i)}$) only take on the values $\{-1,0,1\}$, we can calculate the necessary L_p norms.

That the operator norm of T_ϵ is at least 1 is obvious. Let f be identically 1 everywhere. Then $\hat{f}(T) = \sum f(S)(-1)^{|S\cap T|} = 0$ for $T \neq \emptyset$ and $\hat{f}(\emptyset) = 1$. So $||T_\epsilon f||_2 = 1 = ||f||_p$. What the Bonami-Beckner inequality says is that for every $f : \{-1,1\}^n \to \mathbb{R}$, $||T_\epsilon f||_2 \leq ||f||_p$.

We'll do a part of the proof only. The proof is via induction on n, the dimension of the cube. The base case is the main part of the proof; the method for inducting is standard in analysis and we'll skip it.

[1] This is called a *hypercontractive inequality*.

Again, the surprising thing here is that $p < 2$ and still $||T||_{op} \leq 1$.

For $n = 1$, every function f has the form $f = a + bx$ and then $T_\epsilon f = a + \epsilon bx$. (There are only two characters, $\chi_\emptyset$ and $\chi_{\{x\}}$). Then at $x = -1$, $f = a - b$ and $T_\epsilon f = a - \epsilon b$. At $x = 1$, $f = a + b$ and $T_\epsilon f = a + \epsilon b$. So

$$\begin{aligned} ||f||_p &= \left[\frac{|a+b|^p + |a-b|^p}{2}\right]^{\frac{1}{p}} \\ ||T_\epsilon f||_2 &= \sqrt{\frac{(a+\epsilon b)^2 + (a-\epsilon b)^2}{2}} \\ &= \sqrt{a^2 + \epsilon^2 b^2} \end{aligned}$$

We want to prove $||T_\epsilon f||_2 \leq ||f||_p$, i.e., we want to prove

$$\left[\frac{|a+b|^p + |a-b|^p}{2}\right]^{\frac{1}{p}} \geq \sqrt{a^2 + \epsilon^2 b^2}. \tag{8.5}$$

Suppose $|a| \geq |b|$. Let $b = ta$ and divide by $|a|$:

$$\begin{aligned} \left(\frac{|a+b|^p + |a-b|^p}{2}\right)^{\frac{1}{p}} &= \left(\frac{|a+ta|^p + |a-ta|^p}{2}\right)^{\frac{1}{p}} \\ &= |a| \left(\frac{|1+t|^p + |1-t|^p}{2}\right)^{\frac{1}{p}} \\ \sqrt{a^2 + \epsilon^2 b^2} &= \sqrt{a^2 + \epsilon^2 (at)^2} \\ &= |a|\sqrt{1 + \epsilon^2 t^2} \end{aligned}$$

So we will prove

$$\frac{|1+t|^p + |1-t|^p}{2} \geq (1 + \epsilon^2 t^2)^{\frac{p}{2}} \text{ when } |t| \leq 1 \tag{8.6}$$

and (8.5) will follow. Note that if $|a| < |b|$, we'd substitute $a = bt$ and divide by $|b|$, and would want to prove

$$\frac{|t+1|^p + |t-1|^p}{2} \geq (\epsilon^2 + t^2)^{\frac{p}{2}} \text{ when } |t| \leq 1 \tag{8.7}$$

But since

$$(1 + \epsilon^2 t^2) \geq \epsilon^2 + t^2,$$

once we prove equation (8.6), (8.7) will follow immediately.

Proof of (8.6) is via the Taylor expansion. For the left hand side, terms in odd places will cancel out, and terms in even places will double. Recall $p = 1 + \epsilon^2$ and $|t| \leq 1$. The left hand side becomes

$$\sum_{j=0}^{\infty} t^{2j} \left(\begin{array}{c} p \\ 2j \end{array} \right)$$

and the right hand side becomes

$$\sum_{j=0}^{\infty} \epsilon^{2j} t^{2j} \binom{p/2}{j}.$$

Let's examine the first two terms of each side. On both sides, the first term is 1. On the left hand side, the second term is $t^2(p(p-1))/2$ and on the right hand side the second term is $\epsilon^2 t^2 p/2$; since $\epsilon^2 = p-1$, this means the second terms are also equal on both sides. Therefore it is sufficient to compare the terms from $j \geq 2$.

What we discover is that on the left hand side, all terms are positive, whereas on the right hand side, the $j = 2k$ and $j = 2k+1$ terms have a negative sum for all $k \geq 1$. First we show the left hand side is positive. The $(2j)$th coefficient equals $p(p-1)(p-2)\ldots(p-2j+1)$ divided by some positive constant. Note that $p(p-1)$ is positive and all the terms $(p-2)\ldots(p-2j+1)$ are negative. But since there are an even number of these negative terms, the product as a whole is positive. Therefore, on the left hand side, all terms are positive.

Now consider the right hand side. We will show that the $j = 2k$ and $j = 2k+1$ terms have a negative sum for all $k \geq 1$. Consider the sum

$$\epsilon^{4k} t^{4k} \binom{p/2}{2k} + \epsilon^{4k+2} t^{4k+2} \binom{p/2}{2k+1}.$$

We can divide out $\epsilon^{4k} t^{4k}$ without affecting the sign. Since the second term is the positive one, and $|t| \leq 1$, we can replace t^2 by 1 without loss of generality. So now we have

$$\begin{aligned}
\binom{p/2}{2k} + \epsilon^2 \binom{p/2}{2k+1} &= \binom{p/2}{2k} + (p-1)\binom{p/2}{2k+1} \\
&= \frac{p/2(p/2-1)\ldots(p/2-2k+1)}{2k!} + (p-1)\frac{p/2(p/2-1)\ldots(p/2-2k)}{(2k+1)!} \\
&= \left[(2k+1)\frac{p}{2}(\frac{p}{2}-1)\ldots(\frac{p}{2}-2k+1) + (p-1)\frac{p}{2}(\frac{p}{2}-1)\ldots(\frac{p}{2}-2k)\right]/(2k+1)! \\
&= \left[\frac{p}{2}(\frac{p}{2}-1)\ldots(\frac{p}{2}-2k+1)\right]\left[2k+1+(p-1)(\frac{p}{2}-2k)\right]/(2k+1)!
\end{aligned}$$

Notice that the first term in brackets is negative and the second term in brackets is positive. Thus the sum of the kth even and odd term is negative for all k, and we've proved equation (8.6). Equation (8.6) implies equation (8.7); equations (8.6) and (8.7) together imply equation (8.5); and equation (8.5) implies $||f||_p \geq ||T_\epsilon f||_2$ for $p = 1 + \epsilon^2$. Thus we've proved the base case for the Bonami-Beckner inequality.

8.3 Back to Kahn Kalai Linial

In general it is not obvious how to utilize this inequality, since it's hard to compute the p-norm. But we're looking at an easy case. Specifically, if $g : \{0,1\}^n \to \{-1,0,1\}$, the inequality says that

$$\Pr(g \neq 0) = t \Rightarrow t^{\frac{2}{1+\delta}} \geq \sum_S \delta^{|S|}(\hat{g}(S))^2 \quad (8.8)$$

To see this, let $\delta = \epsilon^2$. We know $||g||_p \geq ||T_\epsilon g||_2$.

$$||g||_p = \left(\frac{1}{2^n}\sum |g(x)|^p\right)^{\frac{1}{p}} = \left(\frac{1}{2^n}\sum_{g(x)\neq 0} 1\right)^{\frac{1}{p}} = t^{\frac{1}{p}} = t^{\frac{1}{1+\epsilon^2}}$$

So the Bonami-Beckner inequality tells us

$$t^{\frac{1}{1+\epsilon^2}} \geq \sqrt{\sum_S \epsilon^{2|S|}(\hat{g}(S))^2}$$

and squaring both sides and substituting δ gives us equation (8.8).

We applied this inequality for $g = f^{(i)}$. Then $t = \beta_i$, the influence of the ith variable, which is exactly what we're looking for. Recall that

$$\widehat{f^{(i)}(S)} = \begin{cases} 0 & i \notin S \\ \hat{f}(S) & i \in S \end{cases}$$

We want to prove that $\max \beta_i \geq \Omega(\log n/n)$. By substituting the new values into (8.8), we get

$$\beta_i^{\frac{2}{1+\delta}} \geq \sum_{S \ni i} \delta^{|S|}(\hat{f}(S))^2$$

The $\delta^{|S|}$ terms are the W_S's that we want to behave sufficiently close to the step function. Recall, we also know that $\sum_{0\leq|S|<T}(\hat{f}(S))^2 > 0.9$ where $T = \log n/10$. Since we assume f to be balanced (i.e. $\Pr(f = 1) = 1/2$), we can ignore the 0 term because $\hat{f}(S) = 0$ when $S = \emptyset$, but we cannot ignore it for imbalanced functions. Still, it won't matter, and we'll come back to this point later.

So we know $\sum_{0<|S|<T}(\hat{f}(S))^2 > 0.9$. Let $\delta = 1/2$ (the choice is arbitrary).

$$\begin{aligned}
n \cdot \max \beta_i^{\frac{4}{3}} &\geq \sum_i \beta_i^{\frac{4}{3}} \\
&\geq \sum_i \sum_{S\ni i} (\frac{1}{2})^{|S|}(\hat{f}(S))^2 \\
&= \sum_S \left(\frac{1}{2}\right)^{|S|} |S|(\hat{f}(S))^2 \\
&\geq \sum_{|S|<T} \left(\frac{1}{2}\right)^{|S|} |S|(\hat{f}(S))^2 \\
&\geq \left(\frac{1}{2}\right)^{T} \sum_{|S|<T} (\hat{f}(S))^2 \\
&\geq (n^{-\frac{1}{10}})(0.9)
\end{aligned}$$

Therefore, $n \cdot \max \beta_i^{4/3} \geq 0.9/n^{1/10}$ and so

$$\max \beta_i \geq \left(\frac{c}{n^{11/10}}\right)^{3/4} = \Omega(n^{-33/40}) >> \frac{\log n}{n}$$

There is some progress in understanding what the vector of influences is in general, but we have a long way to go.

We return to an issue we had left open. What happens when we deal with functions that are imbalanced? What if $f : \{0,1\}^n \to \{0,1\}, \mathbb{E}f = p$ (not necessarily $p = 1/2$). Now we cannot ignore $\hat{f}(\emptyset)$ in the previous argument.

Indeed, $f(\emptyset) = p$. $\sum(\hat{f}(S))^2 = p$. We'd have to subtract $(\hat{f}(\emptyset))^2 = p^2$ off of p in general. But this is fine as long as $0 < p < 1$. We mention this because this technique is used often.

8.4 Sharp thresholds in graphs

Theorem 8.3. *Every monotone graph property has a sharp threshold.*

This theorem is also the title of the paper, by Friedgut and Kalai. The background is that in the late 1950s, Erdős and Renyi defined random graphs, which became an important ingredient for many investigations in modern discrete mathematics and theoretical computer science. A random graph $G(n, p)$ is a graph on n vertices in which the probability that (x, y) is an edge in G is p. That is, for each pair of vertices, flip a p-weighted coin, and put an edge in the graph if the coin comes up heads. We do this independently for each pair of vertices. So

$$Pr(G) = p^{e(G)}(1-p)^{\binom{n}{2}-e(G)}$$

Already Erdős and Renyi had noticed that some properties of graphs have special behavior. For example, take the property that G is connected. Write $Pr(G \text{ is connected})$ as $f(p)$. When p is small, the graph is unlikely to be connected; when p is big, it is almost certain to be connected. The shape of $f(p) = Pr(G \text{ is connected})$ is shown in Figure 8.3.

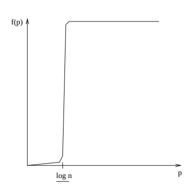

Figure 8.3: Sharp threshold

The transition from disconnected to connected is very rapid. This is related to a class of physical phenomena called phase transition; there is some physical system that is dependent on a single parameter, such as temperature or pressure, and the system goes from one phase to the other very sharply (for instance, the system goes from not magnetized to magnetized).

There are other graph properties that also exhibit this behavior (e.g Hamiltonian cycle, planarity, etc). But there was no satisfactory general theorem until Freidgut-Kalai.

To have a precise form of the theorem, we define the terms:

Definition 8.1. A *graph property* is a class of labeled graphs that is closed under permutations of the vertex labels.

Intuitively, a graph property holds or does not hold regardless of the labeling of the vertices, such as connectedness, "the graph contains a triangle", the graph is 17-colorable, etc.

Definition 8.2. A graph property is called *monotone* if it continues to hold after adding more edges to the graph.

Again, connectedness is monotone; non-planarity is also. An example of a graph property which is non-monotone is "the number of edges in the graph is even".

Let A be a monotone graph property.

$$\mu_p(A) = Pr(A|G \in_R G(n,p))$$

where G is sampled randomly. Clearly $\mu_p(A)$ is an increasing function of p. The theorem says that p_0, where $\mu_{p_0}(A) = \epsilon$, and p_1, where $\mu_{p_1}(A) = 1 - \epsilon$ are very close. Namely,

$$p_1 - p_0 \leq O\left(\frac{\log(1/\epsilon)}{\log n}\right)$$

In the connectivity example, this is not very interesting. The transition from almost surely disconnected to almost surely connected takes place around $p = \log n/n$, and $1/\log n$ is much bigger. In other words, the gap is much bigger than the critical value where the threshold occurs.

Later, Bourgain and Kalai showed that for all monotone graph properties, the same bound holds with $p_1 - p_0 \leq O\left(\frac{1}{\log^{2-\gamma} n}\right)$, for every $\gamma > 0$ and n large enough. This theorem is nearly tight, since there exist examples of monotone graph properties where the width of the gap is $\Theta\left(\frac{1}{\log^2 n}\right)$. For instance, "$G$ contains a k-clique" for specific $k = \Theta(\log n)$ where the critical $p = 1/2$.

So what can we do about the problem in the connectivity example, where the threshold comes at a point much smaller than the gap? We can ask a tougher question: is it true that the transition from $\mu_{p_0} = \epsilon$ to $\mu_{p_1} = 1 - \epsilon$ occurs between $(1 \pm o(1))q$, where q is the critical probability (i.e. $\mu_q(A) = 1/2$)?

If the answer were yes, we'd have a more satisfactory gap in relation to our critical value in the connectivity case. However, the answer is negative for certain graph properties: this is asking too much. For example, suppose the property is that G contains a triangle. Here the critical $p = c/n$. The expected number of triangles in G is

$$\binom{n}{3} p^3 = \frac{c^3}{6}(1 - \frac{1}{n})(1 - \frac{2}{n})$$

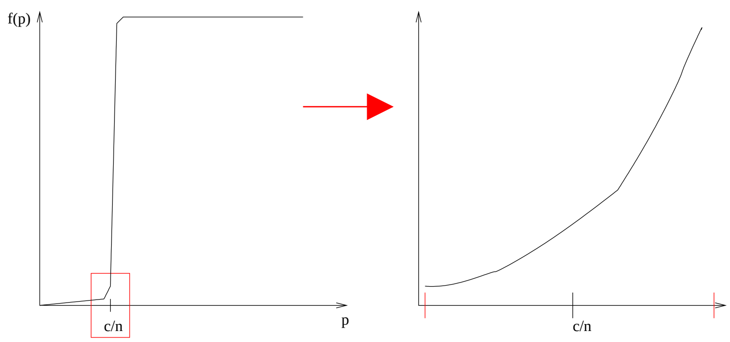

Figure 8.4: G contains a triangle

The picture looks like Figure 8.4. At the smaller scale, the threshold isn't as sharp. As we vary c in $p = c/n$, the probability that G contains no triangle changes from one constant to another constant, both bounded away from zero and from one. The reason behind this picture is the number of triangles is a random variable with a Poisson distribution. Therefore, the probability there is no triangle is $Pr(X = 0) = e^{-\mu}$, where μ is the expectation of X.

One of Freidgut's major contributions was to characterize which graph properties have a sharp threshold and which don't in this stronger sense. It's a little complicated to even state his result precisely, but the spirit of the theorem is this: properties like "G contains a triangle" are considered "local". Friedgut's theorem says that "a graph property has a sharp threshold (in the strong sense)" is roughly equivalent to the property being "non-local".

References

[1] M. Ben-Naor and N. Linial. Collective Coin Flipping. In *Randomness and Computation* S. Micali ed. Academic press, new York, 1990.

[2] J. Kahn, G. Kalai and N. Linial The influence of variables on boolean function. In *Proceedings of the 29th Symposium on the Foundations of Computer Science*, White Planes, 1988, Pages 68–80.