Boolean Function Analysis — Assignment 2

Yuval Filmus

July 14, 2021

1. The proof of the FKN theorem in Section 3 of the lecture notes uses the estimate

$$\mathbb{E}\left[\left(\sum_{i=1}^n c_i x_i\right)^4\right] \leq 3\left(\sum_{i=1}^n c_i^2\right)^2.$$

Show that this follows from hypercontractivity (up to the constant 3).

Let $f = \sum_i c_i x_i$. Then

$$\mathbb{E}[f^4] = \|f\|_4^4 = \|T_{1/\sqrt{3}} T_{\sqrt{3}} f\|_4^4 \stackrel{(*)}{\leq} \|T_{\sqrt{3}} f\|_2^4,$$

where $(*)$ is hypercontractivity. The right-hand side is the square of

$$\|T_{\sqrt{3}} f\|_2^2 = \mathbb{E}[(T_{\sqrt{3}} f)^2] = 3\sum_{i=1}^n c_i^2.$$

We conclude that

$$\mathbb{E}\left[\left(\sum_{i=1}^n c_i x_i\right)^4\right] \leq 9\left(\sum_{i=1}^n c_i^2\right)^2.$$

2. The goal of this exercise is to present an alternative proof of the FKN theorem using the Berry–Esseen theorem, a quantitative version of the central limit theorem.
A special case of the Berry–Esseen theorem states that if $X = \sum_{i=1}^n c_i x_i$, where $\sum_{i=1}^n c_i^2 = 1$ and $|c_i| \leq \delta$ for all i, then for all $t \in \mathbb{R}$,

$$|\Pr[X < t] - \Pr[N(0,1) < t]| \leq \delta,$$

where $N(0,1)$ is the Gaussian distribution with zero mean and unit variance.

   (a) Use the Berry–Esseen to show that the following holds for some constant $c > 0$.
   If $f = \sum_{i=1}^n a_i x_i$ satisfies $1 - c \leq \sum_{i=1}^n a_i^2 \leq 1$ and $|a_i| \leq c$ for all i then $\mathbb{E}[\operatorname{dist}(f, \{\pm 1\})^2] \geq c$.
   Let $A = \sqrt{\sum_{i=1}^n a_i^2}$, and define $X = f/A$. Thus $X = \sum_{i=1}^n c_i x_i$, where $|c_i| = |a_i|/A \leq c/\sqrt{1-c}$ and $\sum_{i=1}^n c_i^2 = 1$. Applying the Berry–Esseen theorem twice,

$$\Pr[-\tfrac{1}{2} < X < \tfrac{1}{2}] \geq \Pr[-\tfrac{1}{2} < N(0,1) < \tfrac{1}{2}] - \frac{2c}{\sqrt{1-c}}.$$

This shows that

$$\mathbb{E}[\operatorname{dist}(f,\{\pm 1\})^2] \geq \left(\Pr[-\tfrac{1}{2} < N(0,1) < \tfrac{1}{2}] - \frac{2c}{\sqrt{1-c}}\right) \cdot \left(1 - \frac{A}{2}\right)^2$$
$$\geq \frac{1}{4}\left(\Pr[-\tfrac{1}{2} < N(0,1) < \tfrac{1}{2}] - \frac{2c}{\sqrt{1-c}}\right).$$

If $c > 0$ is small enough, then the right-hand side is at least c.

(b) Explain how to deduce the FKN theorem (in its version for Boolean functions) from this, possibly with a worse error bound.

We will prove FKN in the following version: if $F\colon \{\pm 1\}^n \to \{\pm 1\}$ satisfies $\|F^{>1}\|^2 \leq \epsilon$ then F is $O(\epsilon)$-close to a function of the form ± 1 or $\pm x_i$. We can assume that ϵ is small enough, since otherwise FKN is trivial.

Let $f = F^{\leq 1}$. Then $\deg(f) \leq 1$ and $\mathbb{E}[\operatorname{dist}(f,\{\pm 1\})^2 \leq \mathbb{E}[(f-F)^2] \leq \epsilon$.

Let $g(x_0, x_1, \ldots, x_n) = x_0 f(x_0 x_1, \ldots, x_0 x_n)$, and observe that $\deg(g) \leq 1$, $\mathbb{E}[\operatorname{dist}(g,\{\pm 1\})^2] \leq \epsilon$, and $\mathbb{E}[g] = 0$. Moreover, $\|g\|^2 = \|f\|^2 \geq 1 - \epsilon$.

If $a, b \in \{\pm 1\}$ then $a - b \in \{0, \pm 2\}$. As shown in class, this implies that

$$\mathbb{E}[\operatorname{dist}(g(x_0,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n) - g(x_0,\ldots,x_{i-1},-1,x_{i+1},\ldots,x_n), \{0,\pm 2\})^2] = O(\epsilon).$$

Since $g(x_0,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n) - g(x_0,\ldots,x_{i-1},-1,x_{i+1},\ldots,x_n) = 2\hat{g}(\{i\})$, we conclude that $\operatorname{dist}(\hat{g}(\{i\}), \{0, \pm 1\}) = O(\sqrt{\epsilon})$ for all i.

Applying part (a), we see that at least one $\hat{g}(\{i\})$ has to be $O(\sqrt{\epsilon})$-close to $\sigma \in \{\pm 1\}$. It follows that $\mathbb{E}[g\sigma x_i] = 1 - O(\sqrt{\epsilon})$, and so $\mathbb{E}[(g - \sigma x_i)^2] = \mathbb{E}[g^2] + 1 - 2\,\mathbb{E}[g\sigma x_i] = O(\sqrt{\epsilon})$. In other words, g is $O(\sqrt{\epsilon})$-close to $\pm x_i$. Therefore f is $O(\sqrt{\epsilon})$-close to some $h \in \{\pm 1, \pm x_i\}$. Thus $\mathbb{E}[(F-h)^2] \leq 2\,\mathbb{E}[(F-f)^2] + 2\,\mathbb{E}[(f-h)^2] = O(\sqrt{\epsilon})$.

3. The goal of this exercise is to show that Friedgut's junta theorem fails for bounded functions.

(a) Let $f(x_1, \ldots, x_n) = \frac{x_1 + \cdots + x_n}{\sqrt{n}}$. Calculate $\operatorname{Inf}_1[f]$ and $\operatorname{Inf}[f]$.

We first calculate $L_i f$:

$$L_i f = \frac{x_i}{\sqrt{n}}.$$

Therefore

$$\operatorname{Inf}_i[f] = \|L_i f\|^2 = \frac{1}{n}.$$

It follows that $\operatorname{Inf}[f] = 1$.

(b) Let $g(x)$ result from clipping $f(x)$ to $[-1, 1]$, that is, $g(x) = f(x)$ if $f(x) \in [-1, 1]$, $g(x) = -1$ if $f(x) < -1$, and $g(x) = 1$ if $f(x) > 1$. Show that $\operatorname{Inf}[g] = O(1)$.

We will show more generally that clipping can only reduce individual influences.

Let $\operatorname{clip}(x) = x$ if $x \in [-1,1]$, $\operatorname{clip}(x) = -1$ if $x < -1$, and $\operatorname{clip}(x) = 1$ if $x > 1$; we say that x is *clipped* if $x \notin [-1, 1]$; it is *clipped to* $\operatorname{clip}(x)$. Then $4\operatorname{Inf}_i[h]$ is the expected value of $(h(x) - h(x^{\oplus i}))^2$, while $4\operatorname{Inf}_i[\operatorname{clip}(h)]$ is the expected value of $(\operatorname{clip}(h)(x) - \operatorname{clip}(h)(x^{\oplus i}))^2$. Hence it suffices to show that for all $a, b \in \mathbb{R}$,

$$|\operatorname{clip}(a) - \operatorname{clip}(b)| \leq |a - b|.$$

Suppose, without loss of generality, that $a \geq b$. We consider several cases:

i. If both a, b are clipped to the same value then $\mathrm{clip}(a) = \mathrm{clip}(b)$.

ii. If a, b are clipped to different values, or if only one of them is clipped, then clipping has the effect of bringing the two values closer to each other.

iii. If none of a, b are clipped then $\mathrm{clip}(a) - \mathrm{clip}(b) = a - b$.

In all cases, $|\mathrm{clip}(a) - \mathrm{clip}(b)| \leq |a - b|$, and we conclude that $\mathrm{Inf}_i[\mathrm{clip}(h)] \leq \mathrm{Inf}_i[h]$, and so $\mathrm{Inf}[\mathrm{clip}(h)] \leq \mathrm{Inf}[h]$.

Taking $h = f$, this shows that $\mathrm{Inf}[g] \leq \mathrm{Inf}[f] = 1$.

(c) Show that for some constants $\epsilon > 0$ and $N \in \mathbb{N}$, if $n \geq N$ and g is ϵ-close to a function $h\colon \{-1, 1\}^n \to \mathbb{R}$ (that is, $\mathbb{E}[(g - h)^2] \leq \epsilon$) then h depends on at least $n/2$ variables.

We will show that if h depends on fewer than $n/2$ variables, then we reach a contradiction for appropriate ϵ, N. We assume for simplicity that n is even (otherwise, replace $n/2$ with $\lfloor n/2 \rfloor$ throughout).

If h depends on fewer than $n/2$ variables, then we can certainly write it as a function depending on exactly $n/2$ variables. Suppose without loss of generality that h depends on the first $n/2$ variables. Define $S = x_1 + \cdots + x_{n/2}$ and $T = x_{n/2+1} + \cdots + x_n$. The central limit theorem shows that $S/\sqrt{n/2}$ and $T/\sqrt{n/2}$ tend to standard Gaussians, and so each of the following events happens with constant probability, say at least $c > 0$, assuming that n is large enough:

- $|S| \leq \frac{1}{4}\sqrt{n}$.
- $T \geq \frac{3}{4}\sqrt{n}$.
- $T \leq -\frac{3}{4}\sqrt{n}$.

Let $x_1, \ldots, x_{n/2}$ be an input for which $|S| \leq \frac{1}{4}\sqrt{n}$, which happens with probability at least c. If $h(x_1, \ldots, x_{n/2}) \leq 0$ then with probability at least c, we have $T \geq \frac{3}{4}\sqrt{n}$, and on these inputs $(g - h)^2 \geq \frac{1}{4}$. Similarly, if $h(x_1, \ldots, x_{n/2}) \geq 0$ then with probability at least c, we have $T \leq -\frac{3}{4}\sqrt{n}$, and on these inputs $(g - h)^2 \geq \frac{1}{4}$. Therefore

$$\mathbb{E}[(g - h)^2] \geq \frac{c^2}{4}.$$

4. The goal of this exercise is to show that the parameters in Friedgut's junta theorem are tight.

(a) Let $f\colon \{-1, 1\}^{2^m + m} \to \{-1, 1\}$ be the addressing function $f(x, y) = x_y$ (that is, $x \in \{-1, 1\}^{2^m}$, $y \in \{-1, 1\}^m$, and we interpret y as an index into x). Calculate the individual influences and the total influence of f.

The influence of a variable x_i is the probability that flipping x_i flips the output of f. This happens precisely when $y = i$, which happens with probability 2^{-m}. Hence the influence of x_i is 2^{-m}.

The influence of a variable y_i is the probability that flipping y_i flips the output of f. Let $y = j$ and $y^{\oplus i} = k$. Then $x_j \neq x_k$ with probability $1/2$. Hence the influence of y_i is $1/2$.

The total influence of f is thus

$$2^m \cdot 2^{-m} + m \cdot \tfrac{1}{2} = \tfrac{1}{2}m + 1.$$

(b) Let $g\colon \{-1,1\}^{2^m+m+k} \to \{-1,1\}$ be the function given by $g(x,y,z) = f(x,y)$ if $z = \mathbf{1}$ (where $z \in \{-1,1\}^k$), and $g(x,y,z) = 1$ otherwise. Calculate the individual influences and the total influence of g.

To calculate the influence of x_i or y_i, note that when $z = \mathbf{1}$, which happens with probability 2^{-k}, they have the same influence that they have in f, and in contrast, if $z \neq \mathbf{1}$ then they have no influence. Therefore the influence of x_i is 2^{-m-k}, and the influence of y_i is 2^{-k-1}.

To calculate the influence of z_i, notice first that if $z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_k \neq \mathbf{1}$ then flipping z_i has no effect. Otherwise, when $z_i = -1$ the output is 1, and when $z_i = 1$ the output is $f(x,y)$. It is easy to check that $f(x,y)$ is balanced, that is, $\Pr[f(x,y) = 1] = 1/2$, and so the influence of z_i is $2^{-(k-1)} \cdot (1/2) = 2^{-k}$.

The total influence of g is thus

$$2^{-k} \operatorname{Inf}[f] + k \cdot 2^{-k} = \frac{m/2 + k + 1}{2^k}.$$

(c) Let $m = k$ and $\epsilon = 2^{-k}/100$. Show that if $h\colon \{-1,1\}^{2^m+m+k} \to \{-1,1\}$ is ϵ-close to g (that is, $\Pr[g \neq h] \leq \epsilon$) then h depends on $2^{\Omega(\operatorname{Inf}[g]/\epsilon)}$ variables.

First of all, let us notice that when $m = k$ and $\epsilon = 2^{-k}/100$, we have $\operatorname{Inf}[g] = \Theta(k/2^k)$ and so $\operatorname{Inf}[g]/\epsilon = \Theta(k)$. Therefore we need to show that h must depend on $2^{\Omega(k)}$ variables.

If $\Pr[g \neq h] \leq \epsilon$ then $\Pr[g \neq h \mid z = \mathbf{1}] \leq \epsilon / \Pr[z = \mathbf{1}] = 1/100$. When $z = \mathbf{1}$, $g(x,y,z) = f(x,y)$, and so defining $H(x,y) = h(x,y,\mathbf{1})$, it suffices to show that if $\Pr[f \neq H] \leq 1/100$ then H depends on $2^{\Omega(m)}$ variables (recall $m = k$).

Let S be the set of i such that H does not depend on x_i. If $i \in S$ then $\Pr[f \neq H \mid y = i] = 1/2$, and so

$$\Pr[f \neq H] \geq \frac{1}{2} \cdot \frac{|S|}{2^m}.$$

This shows that $|S| \leq 2^m/50$, and in particular, H depends on at least $(1 - 1/50)2^m$ variables.