Problem Set 3

- Due Date: **20 Nov, 2017**
- Turn in your problem sets electronically (LaTeX, pdf or text file) by email. If you submit handwritten solutions, start each problem on a fresh page.
- Collaboration is encouraged, but all writeups must be done individually and must include names of all collaborators.
- Refering sources other than the text book and class notes is strongly discouraged. But if you do use an external source (eg., other text books, lecture notes, or any material available online), ACKNOWLEDGE all your sources (including collaborators) in your writeup. This will not affect your grades. However, not acknowledging will be treated as a serious case of academic dishonesty.
- The points for each problem are indicated on the side.
- Problems 1,2,4,5 are from O'Donnell's book/course. Problem 6 is due to Impagliazoo, Moore and Russell while Problem 7 is due to Håstad.
- Be clear in your writing.

1. **[Average vs. Noise sensistivity]**

   In the proof of Peres' theorem, we showed that the statement for any LTF f, $I[f] \leq O(\sqrt{n})$ implies the statement $NS_\delta[f] \leq O(\sqrt{\delta})$ for all LTF f. First show that this proof can be easily generalized to degree k PTFs as follows: if there exists a constant c_k such that for all degree k PTFs f, we have $I[f] \leq c_k\sqrt{n}$, then there exists a constant d_k such that for all degree k PTFs f, we have $NS_\delta(f) \leq d_k\sqrt{\delta}$.

   Show the converse implication. I.e., If there exists a constant d'_k such that for all degree k PTFs f, we have $NS_\delta(f) \leq d'_k\sqrt{\delta}$, then there exists a constant c'_k such that for all degree k PTFs f, $I[f] \leq c'_k\sqrt{n}$

   [Hint: Use Exercise 2.43(a) in O'Donnell's book]

2. **[Parity as PTF]**

   (a) Show that the parity function $\chi_{[n]}$ cannot be written as a PTF of degree $n-1$.

   (b) Show that every function f on n bits which is not parity or its negation, can be written as a PTF of degree at most $n-1$

   [Hint: Consider $f^{\leq n-1}$.]

3. **[Levin's proof of Yao's XOR Lemma]** (2+2+4+4+4+4)

   For any two functions, $f, g : \{0,1\}^n \to \{+1,-1\}$, define the correlation of f and g (w.r.t. uniform distribution) as follows:

$$corr(f,g) \stackrel{\text{def}}{=} |\mathbb{E}[f(X) \cdot g(X)]| .$$

We say that the function $f : \{0,1,\}^n \to \{+1,-1\}$ is (ρ, S)-hard if for all circuits C of size S, $corr(f,C) \leq \rho$. Define the t-wise XOR of f as follows: the function $f^{(t)} : (\{0,1\}^n)^t \to \{+1,-1\}$, where

$$f^{(t)}(X_1, X_2, \ldots, X_t) \stackrel{\text{def}}{=} \prod_{i=1}^{t} b(X_i).$$

Given the above notation, recall the statement of Yao's XOR Lemma.

Yao's XOR Lemma 1 (2 function case)**.** *If $f : \{0,1\}^n \to \{+1,-1\}$ is (ρ, S)-hard then for all $\varepsilon > 0$, $f^{(2)}$ is $(\rho^2 + \varepsilon, \varepsilon^2 S - O(1))$-hard.*

In this problem, we will give a different proof of the XOR lemma due to Levin (the proof discussed in class is due to Impagliazzo).

Let $C(X,Y)$ be a circuit; let S' be its size. We want to show that if S' is small, then it cannot predict $f^{(2)}(X,Y)$ well. Now,

$$\begin{aligned} corr(C, f^{(2)}) &= |\mathbb{E}_{X,Y}[f(X)f(Y)C(X,Y)]| \\ &= |\mathbb{E}_X[f(X)\mathbb{E}_Y[f(Y)C(X,Y)]]| \\ &= \rho \left| \mathbb{E}_X \left[f(X) \frac{\mathbb{E}_Y[f(Y)C(X,Y)]}{\rho} \right] \right| \\ &= \rho |\mathbb{E}_X[f(X)\tilde{g}(X)]| \end{aligned}$$

where

$$\tilde{g}(x) \stackrel{\text{def}}{=} \frac{\mathbb{E}_Y[f(Y)C(x,Y)]}{\rho}$$

The main step in Levin's proof is to show that given such a function $\tilde{g}$ defined as above, there exists a randomized circuit $\tilde{D}$ with the following properties. For all δ, there exists a random variable R and a randomized circuit $\tilde{D}$ (whose inputs are x and r) such that

(I) For all x, $|\mathbb{E}_R[\tilde{D}(x;R)] - \tilde{g}(x)| \leq \frac{\delta}{\rho}$;

(II) The size of $\tilde{D}$ is at most $\frac{1}{\delta^2}S' + O(\frac{1}{\delta})$.

(a) Complete the proof of Yao's XOR Lemma 1 assuming the existence of such a randomized circuit $\tilde{D}$.

In the remaining parts, we will construct such a circuit $\tilde{D}$. The obstacles towards constructing such a circuit are as follows: (i) $\tilde{g}$ is an average of 2^n many terms and (ii) each of these terms involves computing $f(Y)$ which we assumed was a hard function to begin with! We will get around the first obstacle by generating a small sample of values Y and compute the expectation of $f(Y)C(x,Y)$ over this sample instead of the whole set of values. The error will go down exponentially with the size of the sample, so we can approximate $\tilde{g}(X)$ quite efficiently. But, the second obstacle seems even harder. Even if it is possible to generate values for Y at random, we don't know how to compute f efficiently. To circumvent this problem, instead of generating values of Y and computing $f(Y)$ ourselves, we will generate a sample of pairs $\langle Y, f(Y)\rangle$—that

is, the computation of f now becomes the headache of the distribution of R. The claim that $\tilde{D}$ is a randomized circuit should be taken with a pinch of salt—the random bits it uses, admittedly, come from quite a complicated distribution. Our argument above, however, does not suffer on account of this: we nowhere assumed that the distribution of R was easy to compute.)

Fix the sample size $s = 1/\delta^2$, and let $\tilde{R}$ be random variable taking values in $(\{0,1\}^n \times \{1,-1\})^s$ such that

$$\Pr[\tilde{R} = \langle\langle y_1, e_1\rangle, \langle y_2, e_2\rangle, \ldots, \langle y_s, e_s\rangle\rangle] = \begin{cases} \prod_{i=1}^{s} \Pr[Y = y_i] & \text{if } \wedge_{i=1}^{t} f(y_i) = e_i \\ 0 & \text{otherwise} \end{cases}.$$

Now, the circuit $\tilde{D}$ implements the following algorithm. Let the input be x.

i. Pick $\tilde{R} = \langle\langle y_1, e_1\rangle, \langle y_2, e_2\rangle, \ldots, \langle y_s, e_s\rangle\rangle$.

ii. Compute $v = \langle e_1 C(x, y_1), e_2 C(x, y_2), \ldots, e_s C(x, y_s)\rangle$. Thus, v corresponds to values for $f(Y)C(x,Y)$ for a sample of s randomly chosen values for Y. We expect the number of 1's in this list to be between $k_1 = \frac{1-\rho}{2}s$ and $k_2 = \frac{1+\rho}{2}s$. Let the actual number of 1's in v be i.

iii. If $i \leq k_1$ then output -1. If $i \geq k_2$, then output 1. Otherwise, let $i = \frac{1+q}{2}s$ $(-\rho < q < \rho)$; output 1 with probability $\frac{1}{2}(1 + \frac{q}{\rho})$ and -1 with probability $\frac{1}{2}(1 - \frac{q}{\rho})$.

Observe that the random bits R used by $\tilde{D}$ are $\tilde{R}$ used for generating v and the random bits used for deciding the output when the number of 1's in v is between k_1 and k_2.

(b) Verify that the size of circuit $\tilde{D}$ is at most $\frac{1}{\delta^2}S' + O(\frac{1}{\delta})$ as promised in (II).

We will now show that

$$\left|\mathbb{E}_R[\tilde{D}(x)] - \tilde{g}(x)\right| \leq \frac{1}{\rho}\sqrt{\frac{1-\rho^2}{2\pi s}}. \tag{1}$$

Setting $s = 1/\delta^2$ yields (I) completing the proof of the lemma.

(c) Show that

$$\mathbb{E}[\tilde{D}(x) \mid \text{number of 1's in } = i] = \begin{cases} +1 & i \geq \frac{1+p}{2}s \\ -1 & i \leq \frac{1-p}{2}s \\ \frac{2i-s}{sp} & \text{otherwise} \end{cases}.$$

Now if we define $\alpha_x \stackrel{\text{def}}{=} \Pr_Y[b(Y)C(x,Y) = 1]$, show that the above implies the following.

$$\mathbb{E}_R[\tilde{D}(x)] \quad = \quad -1 \cdot \sum_{i=0}^{k_1} \binom{s}{i} \alpha_x^i (1-\alpha_x)^{s-i} + \sum_{k_1 < i < k_2} \binom{s}{i} \alpha_x^i (1-\alpha_x)^{s-i} \frac{2i-s}{s\rho} + 1 \cdot \sum_{i=k_2}^{s} \binom{s}{i} \alpha_x^i (1-\alpha_x)^{s-i}$$

(d) Show that $\tilde{g}(x)$ can be written in terms of α_x as follows:

$$\tilde{g}(x) = \frac{\mathbb{E}_Y[f(Y)C(x,Y)]}{\rho} = \frac{2\alpha_x - 1}{\rho}.$$

We will now derive for $\tilde{g}(x)$ and expression similar to that of $\mathbb{E}_R[\tilde{D}(x)]$. Observe that

$$\sum_{i=0}^{s} \binom{s}{i} \alpha_x^i (1-\alpha_x)^{s-i} \frac{2i-s}{s\rho} = E_I\left[\frac{2I-s}{s\rho}\right] = \frac{2\alpha_x s - s}{s\rho} = \frac{2\alpha_x - 1}{\rho} = \tilde{g}(x),$$

Hence conclude that

$$\begin{aligned}\mathbb{E}_R[\tilde{D}(x)] - \tilde{g}(x) &= \sum_{i=0}^{k_1} \binom{s}{i} \alpha_x^i (1-\alpha_x)^{s-i} \left(-1 - \frac{2i-s}{\rho s}\right) \\ &\quad + \sum_{i=k_2}^{s} \binom{s}{i} \alpha_x^i (1-\alpha_x)^{s-i} \left(1 - \frac{2i-s}{\rho s}\right).\end{aligned}$$

We will now bound this error. In fact, the first and last sums are symmetrical; the first is always positive; the last is always negative. So, it will suffice if we bound the absolute value of one of them. Let us concentrate on the last sum. We write it as

$$-\frac{2}{\rho s} \sum_{i=k_2}^{s} \binom{s}{i} \alpha_x^i (1-\alpha_x)^{s-i} (i-k_2). \tag{2}$$

(e) Show that the above expression is maximized (in absolute value) when $\alpha_x = k_2/s$. Furthermore, show that when $\alpha = k_2/s$, the above expression has the closed form

$$s\binom{s-1}{k_2-1} \alpha_x^{k_2} (1-\alpha_x)^{s-k_2+1}.$$

[

$$\left. \frac{df}{dp} = \sum_{i=1}^{s} \frac{\partial g}{\partial p_i} \right|_{\forall j: p_j = p}.$$

thus,

On the other hand, the derivative of $f(p)$ can be calculated directly. Consider s independent $0-1$ variables where the ith variable takes the value 1 with probability p_i. Let $g(p_1, \ldots, p_s)$ be the probability that the sum of these variables is at least k. Note that $f(p) = g(p, p, \ldots, p)$;

$$\frac{df}{dp} = \sum_{i=k}^{s} \binom{s}{i} p^i (1-p)^{s-i-1} (i - ps).$$

That is, $f(p)$ is the probability that in s independent trials of a 0-1 random variable that takes the value 1 with probability p, we see at least k ones. Then,

$$f(p) = \sum_{i=k}^{s} \binom{s}{i} p^i (1-p)^{s-i}.$$

[Hint: To obtain the above closed form, consider the following combinatorial problem. Let

(f) Use the above closed form expression and the following Stirling's formula due to Robbins:

$$\left(\frac{n}{e}\right)^n \sqrt{2\pi n} \times e^{1/(12n+1)} < n! < \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \times e^{1/12n}.$$

to conclude (1).

This proof is due to Levin, the bound in (1) is due to Boppana and Hirschfeld and the above writeup is due to Radhakrishnan.

4. **[Bonami's Lemma implies (2,4)-hypercontractivity]** Exercise 9.6 in O'Donnell's book

5. **[small ball probabilities around all points]** Exercise 9.12 in O'Donnell's book

6. **[Information theoretic proof of Level-1 inequality]**

   In lecture, we proved the following level-1 inequality (also called Chang's inequality). For any subset $A \subset \{\pm 1\}^n, |A| = \alpha \cdot 2^n$ and $f = \mathbb{1}_A$, we have

   $$\sum_{i=1}^{n} \widehat{f}(i)^2 = O\left(\alpha^2 \ln \frac{1}{\alpha}\right).$$

   In this problem, we will give an alternate proof via information theory.

   Let $X = (X_1, \ldots, X_n) \in \{\pm 1\}^n$ be the n-bit random variable obtained by picking uniformly at random an element X in the set A. Clearly, we have that the entropy of X, given by $H(X) = \lg(\alpha \cdot 2^n) = n - \log_2 \frac{1}{\alpha}$.

   (a) Let $p_i = \Pr[X_i = +1]$. Show that

   $$p_i = \frac{1}{2}\left(1 + \mathbb{E}_{x \in A} x_i\right) = \frac{1}{2}\left(1 + \frac{\widehat{f}(i)}{\alpha}\right).$$

   (b) Show that $H(X_i) = h_2(p_i)$ where $h_2(p)$ is the binary entropy given

   $$h_2(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p} = (\log_2 e) \cdot \left(p \ln \frac{1}{p} + (1-p) \ln \frac{1}{1-p}\right).$$

   Use the Taylor series around $p = 1/2$ to conclude that $h(\frac{1+x}{2}) \leq 1 - (\log_2 e) \cdot \frac{x^2}{2}$.

   Hence, $H(X_i) \leq 1 - (\log_2 e) \frac{\widehat{f}(i)^2}{\alpha^2}$.

   (c) Use subadditivity of entropy $H(X) \leq \sum H(X_i)$ to conclue that $\sum \widehat{f}(i)^2 \leq 2\alpha^2 \ln \frac{1}{\alpha}$.