Pseudorandomness

Problem Set 1

Due date: August 31st, 2018

Instructions

1. The problem set has **4 questions** with a total score of **90 points**.
2. You are welcome to collaborate with other classmates. But if you do, please mention who all you collaborated with.

   I'd suggest you discuss only after you have spent enough time thinking about the problems independently.
3. Solutions are expected as a LaTeX document. You may use this very file by obtaining the source files from course webpage.
4. The deadline is **31st August 2018, 2359 hrs IST**. For each day of delay, you lose **7 points** of your total score in this assignment. So if you plan to delay, be smart about it.

Questions

Question 1. [15 points] Use the method of conditional expectation to give a deterministic polynomial time algorithm that achieves a $(7/8)$-approximation to Max-Exact3CNF.

Question 2. (Problem 3.2 in Vadhan's manuscript) A combinatorial design is a collection of subsets of some universe that has small pairwise intersection. Formally, an (ℓ, k, r, m)-design is a collection of subsets $S_1, \ldots, S_m$ that satisfy the following properties:

- $S_i \subseteq [\ell]$,
- $|S_i| = k$,
- For any $i \neq j$, we have $\left|S_i \cap S_j\right| \leq r$.

Normally, given the parameter k, we would like to find designs where m is large and ℓ, r are small so that we have a lot of k-subsets with small pairwise intersection in a universe that is not-too-large.

1. **[10 points]** Using the probabilistic method, prove that if $m \leq \frac{\binom{\ell}{r}}{\binom{k}{r}^2}$, then there exists (ℓ, k, r, m)-designs.

2. **[5 points]** Show that, for every constant $\gamma > 0$, there are (ℓ, k, r, m)-designs for $\ell = O\left(\frac{k^2}{r}\right)$, and $r = \gamma \log m$.

You might want to look up some standard approximations for binomial coefficients.

3. **[10 points]** Use the method of conditional expectations to deterministically construct such designs in $\text{poly}(m, k)$ time.

Question 3. (Problem 3.5 in Vadhan's manuscript) Let $\mathcal{H} = \{h : [N] \to [M]\}$ be a pairwise independent hash family.

1. **[5 points]** If $N \geq 2$, show that $|\mathcal{H}| \geq M^2$.
2. **[10 points]** If $M = 2$, show that $|\mathcal{H}| \geq N + 1$.

   (Hint: Based on $\mathcal{H}$, try to construct some orthogonal vectors in $\{\pm 1\}^{|\mathcal{H}|}$.)

3. **[10 points]** More generally, prove that for arbitrary M, we have $|\mathcal{H}| \geq N \cdot (M-1) + 1$.

   (Hint: For each $x \in [N]$, construct $M-1$ linearly independent vectors $v_{x,y} \in \mathbb{R}^{|\mathcal{H}|}$ such that $v_{x,y} \perp v_{x',y'}$ if $x \neq x'$.)

Question 4. (Based on Problem 3.8 in Vadhan's manuscript) Suppose $X_1, \ldots, X_t$ are 4-wise independent random variables with $X_i \in [0, 1]$. Let $\mu_i = \mathbb{E}[X_i]$, and $\mu = \mathbb{E}[X]$ where $X = X_1 + \cdots + X_t$. We would like to prove a better concentration bound for X than what we get out of Chebychev's inequality.

1. **[5 points]** Prove that $\mathbb{E}[(X - \mu)^4] \leq O(t + t^2)$.

   (Hint: Note that $(X - \mu)^4 = (\sum(X_i - \mu_i))^4$. When you expand this out and take the expectation of each term, what happens to terms that involve three or more X_i's?)

2. **[10 points]** Conclude that

$$\Pr[|X - \mu| \geq t\epsilon] \leq O\left(\frac{1}{t\epsilon^2}\right)^2.$$

3. **[10 points]** Generalise this to k-wise independent random variables, where k is some large constant that is an even number, and show that

$$\Pr[|X - \mu| \geq t\epsilon] \leq O\left(\frac{1}{t\epsilon^2}\right)^{k/2}$$

   (the constant in the $O(\cdot)$ notation would depend on k and you can ignore that).