CS 225 - Pseudorandomness Prof. Salil Vadhan

Problem Set 5

Harvard SEAS - Spring 2015 *Due: Fri. April 17, 2015*

Your problem set solutions must be typed (in e.g. LaTeX) and submitted electronically to `cs225-hw@seas.harvard.edu`. You are allowed 12 late days for the semester, of which at most 5 can be used on any individual problem set. (1 late day = 24 hours exactly). Please name your file `PS5-`lastname`.*`.

The problem sets may require a lot of thought, so be sure to start them early. You are encouraged to discuss the course material and the homework problems with each other in small groups (2-3 people). Identify your collaborators on your submission. Discussion of homework problems may include brainstorming and verbally walking through possible solutions, but should not include one person telling the others how to solve the problem. In addition, each person must write up their solutions independently, and these write-ups should not be checked against each other or passed around.

Strive for clarity and conciseness in your solutions, emphasizing the main ideas over low-level details. Do not despair if you cannot solve all the problems! Difficult problems are included to stimulate your thinking and for your enjoyment, not to overwork you. *ed problems are extra credit.

Problem 6.1. Min-entropy and Statistical Difference

1. Prove that for every two random variables X and Y,

$$\Delta(X,Y) = \max_f |\operatorname{E}[f(X)] - \operatorname{E}[f(Y)]| = \frac{1}{2} \cdot |X - Y|_1,$$

   where the maximum is over all $[0,1]$-valued functions f. (Hint: first identify the functions f that maximize $|\operatorname{E}[f(X)] - \operatorname{E}[f(Y)]|$.)

2. Suppose that (W,X) are jointly distributed random variables where (W,X) is a k-source and $|\mathrm{Supp}(W)| \leq 2^\ell$. Show that for every $\varepsilon > 0$, with probability at least $1-\varepsilon$ over $w \stackrel{R}{\leftarrow} W$, we have $X|_{W=w}$ is a $(k - \ell - \log(1/\varepsilon))$-source.

3. Suppose that X is an $(n-\Delta)$-source taking values in $\{0,1\}^n$, and we let X_1 consist of the first n_1 bits of X and X_2 the remaining $n_2 = n - n_1$ bits. Show that for every $\varepsilon > 0$, (X_1, X_2) is ε-close to some $(n_1 - \Delta, n_2 - \Delta - \log(1/\varepsilon))$ block source.

Problem 6.5. Extractors vs. Samplers Use the connection between extractors and averaging samplers to do the following:

1. Prove that for all constants $\varepsilon, \alpha > 0$, there is a constant $\beta < 1$ such that for all n, there is an explicit $(\beta n, \varepsilon)$ extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $d \leq \log n$ and $m \geq (1-\alpha)n$.

3. Suppose we are given a constant-error **BPP** algorithm that uses $r = r(n)$ random bits on inputs of length n. Show how, using the explicit extractor of Theorem 6.36, we can reduce its

error probability to $2^{-\ell}$ using $O(r) + \ell$ random bits, for any polynomial $\ell = \ell(n)$. (Note that this improves the $r + O(\ell)$ given by expander walks for $\ell \gg r$.) Conclude that every problem in **BPP** has a randomized polynomial-time algorithm that only errs for $2^{q^{0.01}}$ choices of its $q = q(n)$ random bits.

Problem 6.9. The Building-Block Extractor Prove Lemma 6.37: Show that for every *constant* $t > 0$ and all positive integers $n \geq k$ and all $\varepsilon > 0$, there is an *explicit* (k, ε)-extractor $\mathrm{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m \geq k/2$ and $d = k/t + O(\log(n/\varepsilon))$. (Hint: convert the source into a block source with blocks of length $k/O(t) + O(\log(n/\varepsilon))$.)

Problem 6.10. Encryption and Deterministic Extraction A (one-time) *encryption scheme* with key length n and message length m consists of an encryption function $\mathrm{Enc}\colon \{0,1\}^n \times \{0,1\}^m \to \{0,1\}^\ell$ and a decryption function $\mathrm{Dec}\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ such that $\mathrm{Dec}(k, \mathrm{Enc}(k,u)) = u$ for every $k \in \{0,1\}^n$ and $u \in \{0,1\}^m$. Let K be a random variable taking values in $\{0,1\}^n$. We say that $(\mathrm{Enc}, \mathrm{Dec})$ is *(statistically) ε-secure with respect to K* if for every two messages $u, v \in \{0,1\}^m$, we have $\Delta(\mathrm{Enc}(K,u), \mathrm{Enc}(K,v)) \leq \varepsilon$. For example, the *one-time pad*, where $n = m = \ell$ and $\mathrm{Enc}(k,u) = k \oplus u = \mathrm{Dec}(k,u)$ is 0-secure (aka perfectly secure) with respect to the uniform distribution $K = U_m$. For a class $\mathcal{C}$ of sources on $\{0,1\}^n$, we say that the encryption scheme $(\mathrm{Enc}, \mathrm{Dec})$ is *ε-secure with respect to $\mathcal{C}$* if Enc is ε-secure with respect to every $K \in \mathcal{C}$.

1. Show that if there exists a deterministic ε-extractor $\mathrm{Ext}\colon \{0,1\}^n \to \{0,1\}^m$ for $\mathcal{C}$, then there exists an 2ε-secure encryption scheme with respect to $\mathcal{C}$.

2. Conversely, use the following steps to show that if there exists an ε-secure encryption scheme $(\mathrm{Enc}, \mathrm{Dec})$ with respect to $\mathcal{C}$, where $\mathrm{Enc}\colon \{0,1\}^n \times \{0,1\}^m \to \{0,1\}^\ell$, then there exists a deterministic 2ε-extractor $\mathrm{Ext}\colon \{0,1\}^n \to \{0,1\}^{m - 2\log(1/\varepsilon) - O(1)}$ for $\mathcal{C}$, provided $m \geq \log n + 2\log(1/\varepsilon) + O(1)$.

   (a) For each fixed key $k \in \{0,1\}^n$, define a source X_k on $\{0,1\}^\ell$ by $X_k = \mathrm{Enc}(k, U_m)$, and let $\mathcal{C}'$ be the class of all these sources (i.e., $\mathcal{C}' = \{X_k : k \in \{0,1\}^n\}$). Show that there exists a deterministic ε-extractor $\mathrm{Ext}'\colon \{0,1\}^\ell \to \{0,1\}^{m - 2\log(1/\varepsilon) - O(1)}$ for $\mathcal{C}'$, provided $m \geq \log n + 2\log(1/\varepsilon) + O(1)$.

   (b) Show that if Ext' is a deterministic ε-extractor for $\mathcal{C}'$ and Enc is ε-secure with respect to $\mathcal{C}$, then $\mathrm{Ext}(k) = \mathrm{Ext}'(\mathrm{Enc}(k, 0^m))$ is a deterministic 2ε-extractor for $\mathcal{C}$.

Thus, a class of sources can be used for secure encryption iff it is deterministically extractable.

Problem 6.11. Extracting from Symbol-Fixing Sources* A generalization of a bit-fixing source is a *symbol-fixing source* X taking values in Σ^n for some alphabet Σ, where subset of the coordinates of X are fixed and the rest are uniformly distributed and independent elements of Σ. For $\Sigma = \{0,1,2\}$ and $k \in [0,n]$, give an explicit ε-extractor $\mathrm{Ext} : \Sigma^n \to \{0,1\}^m$ for the class of symbol-fixing sources on Σ^n with min-entropy at least k, with $m = \Omega(k)$ and $\varepsilon = 2^{-\Omega(k)}$. (Hint: use a random walk on a consistently labelled 3-regular expander graph.)