CS6845 - Pseudorandomness
Assignment 1

Due date: Jan 29, 11.59pm

Instructions

- The solutions must be LaTeX-ed and the pdf file as well as the tex file must be submitted at `https://www.dropbox.com/request/BIlkYswGticDyce660XO`. Please submit both the pdf and tex in a *single* zip file with your roll number as the zip file name.
- For each day's delay in submission, you will lose 20% marks. A delay of 5 days will result in zero marks.
- You are expected to work individually on the assignment. Collaboration on the assignment or checking online sources for answers is *strongly discouraged*. If you need any clarification, you are welcome to discuss with me or Dinesh. Any academic dishonesty will result in zero marks for the assignment.
- Please be concise in stating your answers/proofs.

---

1. (2 points) Let $X_1, \ldots, X_n$ denote unbiased $0-1$ random variables. Let X be the random variable defined as the sum $\sum_{i=1}^{n} \frac{X_i}{2^i}$. Show that X is uniformly distributed over the set $\{0, \frac{1}{2^n}, \frac{2}{2^n}, \ldots, \frac{2^n-1}{2^n}\}$.

2. (2 points) Suppose $n \geq 4$ and $\mathcal{H}$ is a collection of r sets, each of size n from a universe $\mathcal{U}$ where $r \leq 4^{n-1}/3^n$. Show that it is possible to color the elements of $\mathcal{U}$ with 4 colors such that for each set in $\mathcal{H}$ all four colors are present.

3. (3 points) Let F be a finite collection of binary strings of finite length and assume that no member of F is a prefix of another one. Let N_i denote the number of strings of length i in F. Prove that $\sum_i \frac{N_i}{2^i} \leq 1$.

   Hint: If you take a random binary string of large enough length (how large?), what is the probability that an element of F occurs as a prefix of it. Why is the assumption that F is prefix-free necessary?

4. (5 points) For $x \in \{0,1\}^n$, let $x^{(i)}$ be the i^{th} bit of x. Define the Hamming distance between $x, y \in \{0,1\}^n$, denoted by $\Delta(x,y)$, as the number of indices in which x and y differ. I.e. $\Delta(x,y) = |\{i \mid x^{(i)} \neq y^{(i)}, i \in \{1,2,\ldots,n\}\}|$. Show that there exists a set of $t \leq 2^{n/100}$ strings $x_1, x_2, \ldots, x_t \in \{0,1\}^n$ such that

$$Pr\left[\forall\, i,j \in [t] \text{ with } i \neq j,\ \Delta(x_i, x_j) > n/10\right] \geq 1 - \frac{1}{2^{\Omega(n)}}$$

   Hint: Following identity may be useful : for any n and $k \leq n/4$, $\sum_{i=0}^{k} \binom{n}{i} \leq 2\binom{n}{k}$.

5. Let μ be a probability distribution over the set $\{1, 2, \ldots, n\}$. Suppose that you have access to a function $\mathcal{M}$ which returns an element $i \in \{1, 2, \ldots, n\}$ with probability $\mu(i)$ each time you query it. In this exercise we will design a randomized algorithm that will construct a distribution μ' over $\{1, 2, \ldots, n\}$ that is close to μ using samples from $\mathcal{M}$. The algorithm proceeds as follows: Take t samples from $\mathcal{M}$. Suppose s_i many samples are i, then set $\mu'(i) = s_i/t$.

   (a) (1 point) Show that μ' is a probability distribution.

(b) (2 points) What should be the value of t such that with probability at least $1-\delta$ each $i \in \{1, 2, \dots, n\}$ satisfies $|\mu(i) - \mu'(i)| \leq \epsilon$.

6. (5 points) A set of n balls is drawn by sampling with replacement from an urn containing N balls, M of which are red. Give a sharp concentration result for the number of red balls in the sample drawn.

7. (5 points) In this exercise we will prove a Chernoff-like concentration inequality even in the case where the random variables are not independent. Let $X_1, X_2, \dots, X_n$ be identically distributed binary random variables such that for all $S \subseteq [n]$, $\Pr[\wedge_{i \in S} X_i = 1] \leq \prod_{i \in S} \Pr[X_i = 1]$. Show that even under this assumption, Chernoff bounds hold. Assume that $\Pr[X_i = 1] = p$ for all $i \in [n]$.

8. In this exercise, we will prove a different form of the Chernoff bound with an alternate proof that does not use the moment generating functions. Let $X_1, X_2, \dots, X_n$ be independent identically distributed binary random variables such that $\Pr[X_i = 1] = p$. For $X = \sum_i X_i$, we will prove the following inequality:

$$\Pr[X \geq (p+t)n] \leq \left[\left(\frac{p}{p+t} \right)^{p+t} \left(\frac{1-p}{1-p-t} \right)^{1-p-t} \right]^n \tag{1}$$

(a) (2 points) First show that for every $x \geq 1$, $\Pr[X \geq k] \leq \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} x^{i-k}$.

(b) (3 points) Now use the binomial theorem and calculus to optimize for x, and obtain the bound in Equation 1 by substituting $k = (p+t)n$.