CS5330: Assignment for Week 7

Due: Tuesday, 24th Mar 2020.

Please submit your solutions to the "Assignments/Week 7/Submissions" folder on LumiNUS by 24th March, 6:29 pm. I strongly encourage you to write your solutions using LaTeX.

You may discuss the problems with your classmates, though you should write up your solutions on your own. Please note the names of your collaborators in your submission.

1. Chebyshev's inequality shows that when n items are hashed into n bins using a hash function from a 2-universal family, the maximum load is at most $1 + \sqrt{2n}$ with probability at least $1/2$. Generalize this argument to k-universal hash functions. That is, find a value such that the probability that the maximum load is larger than that value is at most $1/2$.

2. Suppose $M = \{0,1\}^m$ and $N = \{0,1\}^n$. Let $\mathcal{M} = \{0,1\}^{(m+1)\times n}$ denote the space of Boolean matrices with $m+1$ rows and n columns. For any $x \in M$, let $x^{(1)}$ denote the $(m+1)$-bit vector obtained by appending a 1 to the end of x. For $A \in \mathcal{M}$, define $h_A(x) = x^{(1)}A \pmod 2$. Show that $H = \{h_A : A \in \mathcal{M}\}$ is a 2-universal hash family. Is it also strongly 2-universal?

3. For any hash function $h : M \to N$, say it is ϵ-*good* for two sets $A \subseteq M$ and $B \subseteq N$ if for x drawn uniformly from M:

$$\left|\Pr[x \in A, h(x) \in B] - \frac{|A|}{|M|}\frac{|B|}{|N|}\right| \le \epsilon$$

   Suppose h is drawn uniformly from a strongly 2-universal hash family $\mathcal{H}$. Show that for any $\epsilon > 0, A \subseteq M, B \subseteq N$, the probability that h is not ϵ-good for A and B is at most:

$$\frac{|A|/|M| \cdot |B|/|N|}{\epsilon^2 |M|}.$$