COMS 4995 (Randomized Algorithms): Exercise Set #3

For the week of September 16–20, 2019

Instructions:

(1) *Do not turn anything in.*

(2) The course staff is happy to discuss the solutions of these exercises with you in office hours or in the course discussion forum.

(3) While these exercises are certainly not trivial, you should be able to complete them on your own (perhaps after consulting with the course staff or a friend for hints).

Exercise 11

Fix a domain U and a range $\{1, 2, \ldots, n\}$. Suppose $\mathcal{H}$ is a *universal* family of hash functions, meaning that for every $x, y \in U$ with $x \neq y$, $\mathbf{Pr}[h(x) = h(y)] \leq \frac{1}{n}$, where the probability is over the uniform distribution on $\mathcal{H}$. Prove that in a hash table with separate chaining, for every fixed data set $S \subseteq U$ and element $x \notin S$, the expected (unsuccessful) search time for x is $O(1+\alpha)$, where $\alpha = \frac{|S|}{n}$ is the load of the hash table. (Again, the expectation is over the uniformly random choice of $h \in \mathcal{H}$.)

[Hint: Define an indicator random variable X_y for each $y \neq x$, indicating whether $h(y) = h(x)$. Use linearity of expectation.]

Exercise 12

Consider repeatedly flipping a coin with bias (i.e., probability of "heads") equal to p. Let X denote the number of flips needed until the first "heads." (This is called a *geometric random variable* with parameter p.) Prove that $\mathbf{E}[X] = \frac{1}{p}$.

[Recall why this came up in Lecture #4: in a hash table with open addressing, if we heuristically treat every probe in a probe sequence as an independent and uniformly random array position, then the number of probes needed to insert a new element is a geometric random variable with parameter $1 - \alpha$, where α is the load of the hash table.]

Exercise 13

Recall from lecture the cartoon plot of the functions $y = e^x$ and $y = 1 + x$. Let's compare the two functions more rigorously.

(a) Deduce that $1 + x \leq e^x$ for all $x \in \mathbb{R}$.

[Hint: one approach is to fiddle with the Taylor expansion of e^x. Another is to notice that e^x is convex (in fact, all of its derivatives are nonnegative) and consider a first-order approximation at $x = 0$.]

(b) Deduce that for x sufficiently close to 0, the following approximate reverse inequalities hold: for $x \in [0, \frac{1}{2}]$, $e^x \leq 1 + 2x$; and for $x \in [-\frac{1}{2}, 0]$, $e^x \leq 1 + \frac{x}{2}$.

[Hint: Fiddle with the Taylor expansion. Recall that this type of approximation came up in our analysis in Lecture #4 of the false positive rate of bloom filters under the random oracle assumption.]

Exercise 14

Let X be a random variable with finite expectation and variance; recall that $\mathrm{Var}[X] = \mathbf{E}\big[(X - \mathbf{E}[X])^2\big]$ and $\mathrm{StdDev}[X] = \sqrt{\mathrm{Var}[X]}$. Prove *Chebyshev's inequality*: for every $t > 1$,

$$\mathbf{Pr}[|X - \mathbf{E}[X]\,| \geq t \cdot \mathrm{StdDev}[X]] \leq \frac{1}{t^2}.$$

[Hint: apply Markov's inequality to the (non-negative!) random variable $(X - \mathbf{E}[X])^2$.]