CMSC 651, Spring 2018, University of Maryland

HW2, due as PDF to the address *cmsc651.umd@gmail.com* by 11:59PM on February 23, 2018

Notes: (i) Please work on this with your group (one writeup per group). Consulting other sources (including the Web) is not allowed. (ii) Write your solutions neatly and *include your names*; if you are able to make partial progress by making some additional assumptions, then **state these assumptions clearly and submit your partial solution**. (iii) **Please make the subject line of your email** "*651 HW2*" followed by your full name, and email a PDF to *cmsc651.umd@gmail.com*: PDF generated from Word or LaTeX strongly encouraged.

Note. Here and in general, we will let $\exp(x)$ denote e^x; also, when we refer to the Williamson-Shmoys book below, please use the *version of the book posted under "Resources"*.

0. Read and understand: (i) Chapter 1 of Williamson-Shmoys; and (ii) the use of the Chernoff-Hoeffding approach in dimension reduction from *http://cseweb.ucsd.edu/~dasgupta/papers/jl.pdf*. (The latter is not used in this homework, but is given to you to understand how the simple Chernoff-Hoeffding approach can help solve a fundamental problem in data science.)

1. Recall that a random variable X has a Poisson distribution with mean μ if X takes on only non-negative integer values, with $\Pr[X = i] = \exp(-\mu)\mu^i/i!$ for $i \geq 0$. (Verify for yourself that $\mathbf{E}[X] = \mu$; you don't need to prove this as part of this problem.)

(i) Use the method we used in developing the Chernoff-Hoeffding bound, to get an upper bound on $\Pr[X \geq \mu(1 + \delta)]$ for $\delta > 0$. Do not express the result as an infinite sum etc.: express it in closed form. Do you see a similarity with the Chernoff-Hoeffding bound we derived in class? **(10 points)**

(ii) Why are we interested in the Poisson distribution? For some $n \gg \mu$ (we will take $n \to \infty$), consider independent Bernoulli random variables[1] $Y_1, Y_2, \ldots, Y_n$, with $\Pr[Y_i = 1] = \mu/n$ for each i. Let $Y = \sum_i Y_i$. For any fixed μ and any fixed non-negative integer i, prove that

$$\lim_{n \to \infty} \Pr[Y = i] = \exp(-\mu)\mu^i/i!;$$

i.e., the sum of a large number of such Y_j's "converges" to a Poisson distribution. (This is why Poisson is a natural choice when aggregating a large number of (almost) independent, and rarely nonzero, Bernoulli random variables.) **(4 points)**

2. We will now prove *lower-tail* bounds for the Coupon Collector problem. For $u = \lfloor 0.5n \ln n \rfloor$, we want good upper-bounds on the probability that the wait-time W in the problem is at most u.

We will use the following very powerful "balls and bins" result due to Dubhashi and Ranjan. Suppose we throw m balls *independently* at random into n bins that are numbered $1, 2, \ldots, n$, where each ball is thrown randomly into a bin with an *arbitrary* given distribution. Let X_i count the number of balls that fall into bin i. (The integers m and n are arbitrary here.) Then, for any $k \leq n$, any set $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$, and any non-negative integers $t_1, t_2, \ldots, t_k$, the following powerful *negative-correlation* result is known:

$$\Pr[(X_{i_1} \geq t_1) \wedge (X_{i_2} \geq t_2) \wedge \cdots (X_{i_k} \geq t_k)] \leq \prod_{\ell=1}^{k} \Pr[X_{i_\ell} \geq t_\ell].$$

(a) Let Y_i be the event that coupon i was drawn within our u draws. Show that $\Pr[Y_i] = 1 - \Theta(1/\sqrt{n})$. **(3 points)**

---

[1] This is just a fancy way of saying "random variables taking on values in $\{0, 1\}$"

(b) Prove that our required lower-tail probability, i.e., $\Pr[W \leq u]$, is at most $\exp(-\Theta(\sqrt{n})$ using the above balls-and-bins result. **(6 points)**

3. Recall that we view computational problems as decision problems, and therefore every problem becomes a *language*: i.e., a set of finite-length strings from a fixed alphabet, say, $\{0, 1\}$. Thus, given a language L, its complement $\overline{L}$ is the set of all finite-length strings over the alphabet $\{0, 1\}$ that *do not* lie in L.

Suppose languages L_1 and L_2 lie in NP. For the problems below, prove your claim if possible, and if the proof looks very hard, just say crisply why you believe your claim (and specifically what unexpected result will follow if your claim is not true).

(a) Is $L_1 \cup L_2$ necessarily in NP? **(5 points)**

(b) Is $L_1 \cup \overline{L_2}$ necessarily in NP? **(5 points)**

4. Do Exercise 1.1 from Williamson-Shmoys. **(7 + 6 points)**

5. Do Exercise 1.2 from Williamson-Shmoys. **(7 points)**