Homework 0

Algorithms for Big Data
CS498ABD Spring 2019

Exercise 1: Balls and bins Consider the standard balls and bins process. A collection of m identical balls are thrown into n bins. Each ball is thrown independently into a bin chosen uniformly at random.

(a) What is the (precise) probability that a particular bin i contains exactly k balls at the end of the experiment.

(b) Let X be the number of bins that contain exactly k balls. What is the expected value of X?

(c) What is the variance of X?

Exercise 2: Randomized max cut. In the **max cut** problem, the input is a graph $G = (V, E)$ with $m = |E|$ edges and $n = |V|$ vertices, and the goal is to partition V into two sets (A, B) (where $B = V \setminus A$) maximizing the number of edges $\{e = (u, v) \in E : u \in A, v \in B\}$ with endpoints in different sets. (Such an edge is said to be **cut** by the partition (A, B)). This problem is known to be NP-Hard, but we will show that it is very easy to get a constant factor approximation.

(a) Consider the following randomized algorithm.

```
random-partition(G = (V, E))
  1. A, B ← ∅
  2. for each v ∈ V
       A. with probability 1/2
          i. A ← A ∪ {v}
       B. else B ← B ∪ {v}
  3. return (A, B)
```

`random-partition` randomly partitions the vertices by assigning each vertex to A or B independently with equal probability. Show that this algorithm cuts $m/2$ edges in expectation.

(b) Let $k \in \mathbb{N}$. In the **max k-cut** problem, we want to partition V into k sets $(A_1, \ldots, A_k)$ maximizing the number of edges with endpoints in different parts. Consider the following randomized algorithm.

```
random-k-partition(G = (V, E))

1. A_1, ..., A_k ← ∅

2. for each v ∈ V

     // [k] = {1, ..., k}
     A. sample i ∈ [k] uniformly at random
     B. A_i ← A_i ∪ {v}

3. return (A_1, ..., A_k)
```

`random-k-partition` randomly partitions the vertices into k sets analogously to `random-partition`. Show that this algorithm cuts $(1 - 1/k)m$ edges in expectation.

Exercise 3: Coupon Collectors. In the coupon collectors problem, there are n coupons, and each round we are given one of the coupons uniformly at random. Coupons can repeat. We want to collect all n coupons, and in particular, we want to analyze the expected number of rounds before collecting all n coupons.

1. Suppose a coin flips heads with probability p. Show that the expected number of coin tosses until flipping heads is $1/p$.

2. For $i \in [n]$, show that the expected number of iterations between collecting the $(i-1)$th coupon and the ith coupon is $\frac{n}{n+1-i}$.

3. Show that the expected number of iterations until collecting all n coupons is nH_n, where $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$ is the **nth harmonic number** (and approximately $\ln(n)$).