Homework 0

Algorithms for Big Data: CS498 ABD/ABG, Spring 2020

Instructions and Policy: This homework is not for submission. It's purpose is to provide a refresher and some practice on probabilistic concepts that will be used later in the course. Some of you may find these elementary while others may need to invest some time. Those who need the refresher should consult the provided notes and write down the solutions and if necessary obtain feedback from the course staff.

Problem 1. Suppose $X_1, X_2, \ldots, X_n$ are independent real-valued random variables over the same finite probability space (Ω, p). Let $X = \sum_{i=1}^{n} X_i$. Prove that $Var[X] = \sum_{i=1}^{n} Var[X_i]$.

Problem 2. 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.

- What is the (precise) probability that a particular bin i contains exactly k balls at the end of the experiment.
- Let X be the number of bins that contain exactly k balls. What is the expected value of X?
- What is the variance of X? *Hint:* Write $X = \sum_i Y_i$ where Y_i is indicator random variable for bin i having exactly k balls.

Problem 3. Office hours the day before an exam are somewhat stressful for instructors/TAs because one worries about a student asking a question that is on the exam. Let us play a probability game with this situation. Suppose the number of students that attend the office hour is distributed according to a Poisson random variable with mean λ. Each student that arrives has an independent probability p of asking a question from the already fixed exam. What is the probability that none of the exam questions are asked by any of the students? Express this as a function of p and λ. If you do not recall the Poisson distribution you can look it up on Wikipedia or the notes.

Problem 4. Your company must make a sealed bid for a construction project. If you succeed in winning the contract (by having the lowest bid), then you plan to pay another firm \$100000 to do the work. If you believe that the maximum bid (in thousands of dollars) of the other participating companies can be modeled as being the value of a random variable that is uniformly distributed on $(70, 140)$, how much should you bid to maximize your expected profit?

Problem 5. 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 $\{e = (u, v) \in E \mid u \in A, v \in B\}$ 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.

- 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.

- Let k be a non-negative integer. 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. Show that this algorithm cuts $(1 - 1/k)m$ edges in expectation.

Problem 6. Coupon collectors. 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.

- Suppose a coin flips heads with probability p. Show that the expected number of coin tosses until flipping heads is $1/p$.
- For $i \in [n]$, show that the expected number of iterations between collecting the $(i-1)$th coupon and the ith coupon is $\frac{n}{n+1-i}$.
- 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 $\ln n$).