CSCI 6220/4030: Homework 2

Assigned Thursday September 19 2019. Due at beginning of class Thursday October 10 2019.

Remember to typeset your submission, and label it with your name. Please start early so you have ample time to see me during office hours. Provide mathematically convincing arguments for the following problems. Ask me if you are unclear whether your arguments are acceptable.

1. Consider the following randomized algorithm for computing the smallest element in an array of n numbers.

   min $\leftarrow \infty$
   for $i \leftarrow 1$ to n in random order **do**
       if $A[i] < \min$ **then**
           min $\leftarrow A[i]$

   (a) What is the running time of this algorithm, including the time to generate a random permutation of $1, \ldots, n$ (using, e.g. the Knuth shuffle algorithm you looked at in HW1, problem 6)?

   (b) Assume the numbers in A are unique. What is the expected number of times that min is overwritten?

2. Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, let $\mathbf{x}^\star \in \mathbb{R}^n$ be the sole solution to the system $\mathbf{Ax} = \mathbf{b}$. Consider iteratively solving the system using the following modified Kaczmarz method that samples a row at each step with a probability proportional to its length:

   $\mathbf{x}_0$ is given an initial value
   for $i \leftarrow 1$ to T **do**
       $r_i \leftarrow$ a value j in $1 \ldots, n$ with probability $p_j = \|\mathbf{a}_j\|_2^2 / \|\mathbf{A}\|_F^2$
       $\mathbf{x}_i \leftarrow \mathbf{x}_{i-1} - \frac{\mathbf{a}_{r_i}(\mathbf{a}_{r_i}^T \mathbf{x}_{i-1} - b_{r_i})}{\|\mathbf{a}_{r_i}\|_2^2}$

   (a) Prove this algorithm converges exponentially fast: $\mathbb{E}\|\mathbf{x}^\star - \mathbf{x}_T\|_2^2 \leq \rho^T \|\mathbf{x}^\star - \mathbf{x}_0\|_2^2$, and give a simple expression for ρ in terms of $\sigma_{\min}(\mathbf{A})$ and $\|\mathbf{A}\|_F$.

   (b) By comparing their convergence rates, determine which variant of the Kaczmarz algorithm—this, or the one that samples the rows uniformly at random—should be expected to converge faster.

3. The family of Poisson distributions is widely used to model the number of occurrences of events in time and space. A discrete Poisson random variable $X \sim \text{Pois}(\mu)$ with rate parameter μ has the following law:

   $$\mathbb{P}(X = j) = \frac{\mathrm{e}^{-\mu}\mu^j}{j!} \quad \text{for } j = 0, 1, \ldots$$

   (a) Let $X \sim \text{Pois}(\mu)$ and $Y \sim \text{Pois}(\lambda)$ be independent Poisson random variables, and define $Z = X + Y$. Compute the probability mass function $\mathbb{P}(Z = k)$ for $k = 0, 1, \ldots$. What is the distribution of Z?

   A multinomial random vector models the process of randomly sampling n objects of k distinct types. The random vector $X = (X_1, \ldots, X_k)$ is distributed Multinomial$(n; \alpha_1, \ldots, \alpha_k)$, where all the α_i are non-negative and $\sum_i \alpha_i = 1$, if it has law

   $$\mathbb{P}(X_1 = j_1, \ldots, X_k = j_k) = \frac{n!}{j_1! \cdots j_k!} \alpha_1^{j_1} \cdots \alpha_k^{j_k},$$

   when $j_1, \ldots, j_k$ are non-negative integers summing to n, and otherwise $\mathbb{P}(X_1 = j_1, \ldots, X_k = j_k) = 0$.

(b) Assume $X_i \sim \text{Pois}(\lambda_i)$ for $i = 1, \ldots, k$ are independent. Show that the distribution of $(X_1, \ldots, X_k)$ conditioned on the event $\sum_{i=1}^k X_i = n$ is multinomial and identify the parameters $\alpha_1, \ldots, \alpha_k$. You will find the multinomial theorem useful:

$$\left(\sum_{i=1}^k \lambda_i\right)^n = \sum_{\substack{\ell_1+\cdots+\ell_k=n \\ \ell_1,\ldots,\ell_k \geq 0}} \binom{n}{\ell_1, \ell_2, \ldots, \ell_k} \lambda_1^{\ell_1} \cdots \lambda_k^{\ell_k}$$
$$= \sum_{\substack{\ell_1+\cdots+\ell_k=n \\ \ell_1,\ldots,\ell_k \geq 0}} \frac{n!}{\ell_1! \cdots \ell_k!} \lambda_1^{\ell_1} \cdots \lambda_k^{\ell_k}.$$

4. Recall that a union bound argument shows that taking at least $n \ln n + cn$ draws ensures that all n unique types of coupons have been observed with probability at least $1 - \exp(-c)$. This gives an upper bound on the number of draws we need to encounter all the types of coupons. What about a lower bound?

   Use Chebyshev's inequality to show that the probability that it takes between $nH_n - cn$ and $nH_n + cn$ draws to see all n unique types of coupons is at least $1 - \frac{\pi^2}{6c^2}$. You may find the identity $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$ useful in bounding the variance of X, the number of draws it takes to see all n types of coupons, from above with $\frac{\pi^2}{6} n^2$.

   In words, this shows that $\Theta(n \ln n)$ draws are both sufficient **and necessary** to see examples of all n types of coupons. These probability bounds can be sharpened.

5. Recall the strings in a box problem: starting with n strings in a box, at each time step, we uniformly randomly pick two of the free ends in the box and tie the edges together. After n steps, the number of loops in the box is $L = \sum_{i=1}^n (2i-1)^{-1}$ in expectation [1].

   (a) Compute the variance of the number of loops in the box after n steps.

   (b) Use Chebyshev's inequality to argue that the probability that the number of loops is greater than $4L$ is smaller than $\frac{1}{9L}$. How does this compare to the bound you can obtain using Markov's inequality?

   (c) Use Chebyshev's inequality to argue that the probability that the number of loops is smaller than $\frac{L}{4}$ is smaller than $\frac{16}{9L}$.

6. [**CSCI6220 students**] Suppose that we can obtain independent samples $X_1, X_2, \ldots$ of a random variable X and that we want to use these samples to estimate $\mathbb{E}[X]$. Given t samples, we use the sample average $\overline{X} = \left(\sum_{i=1}^t X_i\right)/t$ for our estimate of $\mathbb{E}[X]$. We want the estimate to be within distance $\varepsilon\mathbb{E}[X]$ of the true value $\mathbb{E}[X]$ with probability at least $1 - \delta$. That is, we want

   $$\mathbb{P}\left(\left|\overline{X} - \mathbb{E}[X]\right| \geq \varepsilon\mathbb{E}[X]\right) \leq \delta,$$

   where the probability is taken with respect to the t random samples. How many samples must we take in order to satisfy this requirement?

   This is a ubiquitous question that naturally comes up in polling (where X_i could be binary or numerical preferences), quality control (where X_i could be measures of quality of a product), and many other applications. Due to the cost of obtaining samples of X, it is generally desirable to use as few samples as possible to obtain the desired confidence in our estimate of $\mathbb{E}[X]$.

   In this problem, we develop two approaches that require only knowing a bound on the variance of X. Let $r = \sqrt{\text{Var}(X)}/\mathbb{E}[X]$.

[1] By comparing this sum to an integral, we can show that $L \leq 1 + \frac{1}{2}\ln(2n-3)$.

(a) Show using Chebyshev's inequality that $O(r^2/(\varepsilon^2\delta))$ samples are sufficient to solve the problem.

(b) Suppose that we need only a weak estimate that is within distance $\varepsilon\mathbb{E}[X]$ of $\mathbb{E}[X]$ with probability at least $3/4$. Argue that $O(r^2/\epsilon^2)$ samples are enough for this weak estimate.

(c) Show that, by taking the median of $O(\ln(1/\delta))$ weak estimates, we can obtain an estimate within $\varepsilon\mathbb{E}[X]$ of $\mathbb{E}[X]$ with probability at least $1-\delta$. Conclude that we need only $O(r^2\ln(1/\delta)/\varepsilon^2)$ samples.

(d) Assuming the constant terms in 6.a and 6.c are equal, how many fewer samples do you need using approach 6.c than 6.a to upper bound the failure probability by $\delta = 10^{-6}$?