6.856 — Randomized Algorithms

Spring Term, 2021

Mar. 3, 2021 — Problem Set 3, Due 3/10

Problem 1. Consider a collection of n random variables X_i drawn independently from the *geometric distribution* with mean 2 – that is, X_i is the number of flips of an unbiased coin up to and including the first occurrence of heads. Let $X = \sum X_i$. Use two different methods to derive bounds on the probability that $X > (1+\epsilon)(2n)$ for any fixed $\epsilon > 0$:

(a) Figure out how to reduce this to a question involving just the sum of independent Bernoulli (i.e. indicator) variables, allowing you to apply the Chernoff bound we already know.

(b) Use the method of the in-class Chernoff bound analysis to derive ab initio an upper bound for deviation of sums of independent, identically distributed geometric random variables. In other words, mimicking the analysis from class, derive a Chernoff-like bound for sums of geometric random variables.

Problem 2—This problem should be done without collaboration. MR 4.14. Show that the Quicksort algorithm of Chapter 1 runs in $O(n \log n)$ time with high probability. Do so by bounding the number of pivots to which each element is compared. **Hint:** for a given item x, call a pivoting round *good* if x ends up in the smaller subproblem. How many good rounds can x be in? How long will that take to happen? You may use the previous problem.

Problem 3. Some problems with bit-fixing:

(a) Based on MR Exercise 4.2. Consider the *transpose* permutation: writing i as the concatenation of two $n/2$-bit strings a_i and b_i, we want to route $a_i b_i$ to $b_i a_i$. Show the bit fixing strategy takes $\Omega(\sqrt{N})$ steps on this permutation.

(b) MR 4.9. Consider the following randomized variant of the bit fixing algorithm. Each packet randomly orders the bit positions in the label of its source and then corrects the mismatched bits in that order. Show that there is a permutation for which with high probability that algorithm uses $2^{\Omega(n)}$ steps to route. An inequality that might be helpful:

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k.$$

Problem 4—This problem should be done without collaboration. In class we showed that n balls in n random bins see a maximum load of $O(\log n/\log\log n)$. Show this bound is tight:

(a) Show there is a $k = \Omega(\log n/\log\log n)$ such that bin 1 has at least k balls with probability at least $1/\sqrt{n}$. You may want to use the inequality:

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k .$$

(b) Argue that conditioning on the first bin not having k balls only increases the probability that the second bin does, and so on. Conclude that with high probability, some bin has $\Omega(\log n/\log\log n)$ balls.

Problem 5—Optional. MR 4.7. Prove that Chernoff bounds hold for arbitrary random variables in the $[0,1]$ interval:

(a) A function f is said to be *convex* if for any x, y, and $0 \leq \lambda \leq 1$, $f(\lambda x+(1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$. Show that $f(x) = e^{tx}$ is convex for any $t > 0$ (you can use the fact that e^{tx} has positive second derivative everywhere). What if $t \leq 0$?

(b) Let Z be a random variable that takes values in the interval $[0,1]$ and let $p = E[Z]$. Define the Bernoulli random variable X such that $\Pr\left([\right) X = 1] = p$. Show that for any convex f, $E[f(Z)] \leq E[f(X)]$.

(c) Let $Y_1, \ldots, Y_n$ be independent identical distributed random variables over $[0,1]$ and define $Y = \sum Y_i$. Derive Chernoff-type upper and lower tail bounds for the random variable Y. In particular, show that for $\delta \leq 1$,

$$\Pr\left([\right) Y - E[Y] > \delta] \leq \exp(-\delta^2/2n).$$

Problem 6—Optional. Variant of MR 4.22. Chernoff bounds are quite powerful, but are limited to sums of *independent* random variables. In the next problem, we will consider ways to apply them to sums of *dependent* random variables by comparing the dependent distributions to independent ones.

Consider the model of n balls tossed randomly in n bins. We derive tight bounds on the number of empty bins. Let X_i be the indicator variable that is 1 if the i-th bin is empty. Let $Z = \sum I_i$ be the number of empty bins. Define $p = E[X_i] = (1-1/n)^n$ and let X_i' be n mutually independent Bernoulli random variables that are 1 with probability p. Note that $Y = \sum X_i'$ has the binomial distribution with parameters n and p.

(a) Show that for all $t \geq 0$, $E[e^{tZ}] \leq E[e^{tY}]$ (hint: think about comparing $E[Y^k]$ and $E[Z^k]$ by expanding them). Conclude that any Chernoff bound on the upper tail of Y's distribution also applies to the upper tail of Z's distribution, even though the Bernoulli variables X_i are not independent. (The point is that their correlation is negative and only helps to reduce the tail probability.) Give a resulting bound on the upper tail of Z.

(b) (This one is very hard) Perform the same sort of analysis for the lower tail.