6.856 — Randomized Algorithms

David Karger

Handout #5, September 18, 2002 — Homework 3, Due 9/25

M. R. refers to this text:
Motwani, Rajeez, and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge: Cambridge University Press, 1995.

1. Based on MR 4.1. Suppose that you wish to estimate the (small) fraction f of Republicans in Massachusetts. Assume that you are able to select a resident uniformly at random and determine their political affiliation. Assume also that you know some lower bound $a < f$. Devise a procedure for estimating f by some $\hat{f}$ such that $\Pr[|f - \hat{f}| > \varepsilon f] < \delta$, for any choice of constants $0 < a, \varepsilon, \delta < 1$. Let N be the number of residents you must query to get the estimate. What is the smallest value of N for which you can give your guarantee?

2. MR 4.14. Show that the Quicksort algorithm of Chapter 1 runs in $O(n \log n)$ time with high probability. **Hint:** bound the number of pivots to which a given item is compared.

3. (Based on MR 3.4). This problem can be thought of as modeling some parallel system in which the solution to contention for a resource is for the contenders to back off and try again.

   Consider the following experiment that proceeds in a sequence of *rounds*. For the first round, we have n balls, which are thrown independently and uniformly at random into n bins. After round i, for $i \geq 1$, we discard every ball that ended up in a bin by itself in round i. The remaining balls are retained for round $i + 1$, in which they are again thrown independently and uniformly at random into the n bins.

   (a) Suppose that in some round we have $k = \varepsilon n$ balls. How many balls should you expect to have in the next round?

   (b) Assuming that everything proceeded according to expectation, prove that we would discard all the balls within $O(\log \log n)$ rounds.

   (c) Convert the previous part into a true proof that with probability $1 - o(1)$, we discard all balls within $O(\log \log n)$ rounds. **Hint:** call a round *good* if the number of balls retained is not much more than expected. What is the probability that a round is good? Show that with probability $1 - o(1)$, we get enough good rounds among the first $O(\log \log n)$ to finish.

4. *(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 $\Pr[X = 1] = p$. 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 $\delta \leq 1$,

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

5. *(Optional)* (variant of MR 4.22). **Chernoff bounds with dependent variables:** 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.

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