

# 6.856 — Randomized Algorithms

David Karger

Handout #9, October 2, 2002 — Homework 5, Due 10/9

M. R. refers to this text:

Motwani, Rajeez, and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge: Cambridge University Press, 1995.

1. MR Section 3.3. Evaluate the behavior of the randomized selection algorithm using Chebyshev's inequality. Use this to design a slightly changed version of the algorithm that uses  $O(\log n)$  random bits and fails (takes superlinear time) with probability  $O(1/n)$ .
2. MR 5.12. Show how the method of conditional expectations can be used to build a 2-dimensional binary space partition of size  $O(n \log n)$ .
3. MR 5.11. In this problem, we will finish establishing the properties of the pessimistic estimator  $\hat{P}(a)$  used in the set balancing derandomization by conditional probabilities.
  - (a) Show that for a node  $a$  at the  $i$ -th level of the computation tree,  $\hat{P}(a)$  is of the form  $N(a)/2^{n-i}$ , where  $N(a)$  is a sum of binomial coefficients times powers of two.
  - (b) Prove that for any node  $a$ , we can compute  $\hat{P}(a)$  in polynomial time
  - (c) Prove that  $\min\{\hat{P}(b), \hat{P}(c)\} \leq \hat{P}(a)$  if  $a$  has children  $b$  and  $c$ .
  - (d) Give an upper bound on the running time of the deterministic algorithm for either the unit cost or the log-cost RAM.
4. A collection of  $m$  sets  $S_j \subseteq U$  is given. Our goal is to find a small set  $Z$  that intersects every  $S_j$ . Unfortunately this problem is  $NP$ -complete. Consider the following integer linear program:

$$\begin{aligned} \min \sum_{i \in U} z_i \\ \sum_{i \in S_j} z_i &\geq 1 \\ z_i &\in \{0, 1\} \end{aligned}$$

The intent is that  $Z = \{i \mid z_i = 1\}$ . Let  $w$  denote the optimum ILP value.

- (a) Argue that this integer linear program determines the optimum set.
- (b) Show that in polynomial time, you can find a set of size at most  $w$  that intersects a constant fraction of the sets  $S_j$ . **Hint:** use randomized rounding.

(c) Show that in polynomial time, you can find a set of size at most  $O(\log m)$  times optimum that intersects every  $S_j$ .

5. (optional) MR 5.14. An  $(n, m)$  safe-set instance consists of a universe  $U$  of size  $n$ , a safe set  $S \subseteq U$ , and  $m$  target sets  $T_1, \dots, T_m \subseteq U$ . such that:

- $\|S\| = \|T_1\| = \dots = \|T_m\|$ ,
- $S \cap T_i = \emptyset$  for every  $i = 1, \dots, m$

An *isolator* for a safe set instance is a set  $I \subseteq U$  that intersects all the target sets but not the safe set. An  $(n, m)$  *universal isolating family* is a collection of subsets of  $U$  that contains an isolator for *every*  $(n, m)$  safe set instance.

Show that there exists an  $(n, m)$  universal isolating family that contains a number of sets polynomial in  $n$  and  $m$ .

The application of an isolator is to certain optimization problems involving sets, where you need to eliminate certain “distractions”  $T_i$  (by erasing some of their elements) without hurting the “safe” optimal set that you want to find. If you have an isolator for your given input, you can solve the problem easily.

To use a universal isolating family, we simply try each set in it as a candidate isolator for our input, and know that one of them will work. If the family is of polynomial size, it only takes polynomial time to try all the possibilities, which leads to polynomial time deterministic algorithms.

This problem only shows existence. Using some more complicated techniques, one can deterministically construct the isolating family. A step towards this (the case  $m = 1$ ) is the next problem.

6. (optional) MR 3.9. A universe  $U$  contains a (unknown) subset of  $n$  good elements  $T$  and  $n$  bad elements  $S$ . Our goal is to choose a subset of  $U$  containing at least one good elements and no bad ones.

- (a) Let  $S$  and  $T$  be two disjoint size- $n$  subsets of a universe  $U$ . Suppose we select a random subset  $R \subseteq U$  by independently including each element of  $U$  with probability  $p$ . We say the sample is good if it contains an element of  $T$  but no element of  $S$ . Prove that when  $p = 1/n$ , the probability our sample is good is larger than some positive constant independent of  $n$ .
- (b) Suppose now that we sample the elements only pairwise independently. Prove that for a suitable choice of  $p$ , we still get a positive chance of success. **Hint:** Use inclusion exclusion, which says that

$$\sum \Pr[A_i] - \sum \Pr[A_i \cap A_j] \leq \Pr[\cup A_i] \leq \sum \Pr[A_i]$$