COMP 360 - Fall 2015 - Assignment 5

Due: 6pm Dec 8th.

Guildenstern: The law of probability, as it has been oddly asserted, is something to do with the proposition that if six monkeys (he has surprised himself)... if six monkeys were...
Guildenstern: [...] The law of averages, if I have got this right, means that if six monkeys were thrown up in the air for long enough they would land on their tails about as often as they would land on their -
Rosencrantz: Heads. (He picks up the coin.)
Guildenstern: Which at first glance does not strike one as a particularly rewarding speculation, in either sense, even without the monkeys. I mean you wouldn't bet on it. I mean I would, but you wouldn't...

Tom Stoppard. Rosencrantz and Guildenstern Are Dead

General rules: In solving these questions you may collaborate with other students but each student has to write his/her own solution. There are in total 110 points, but your grade will be considered out of 100. You should drop your solutions in the assignment drop-off box located in the Trottier Building.

1. (20 Points) Given a set P of n points on the plane, consider the problem of finding the smallest circle containing all the points in P. Design a PTAS algorithm for this problem. In other words, given any fixed $\epsilon > 0$, design an algorithm whose running time is polynomial in n, and its output is at most $1 + \epsilon$ times the optimal output.

   Solution 1: Let r be the maximum distance between two points of P. Obviously the radius of the optimal circle is between $r/2$ and r. Divide the part of the plane that contains the points into a gird whose cells are $\frac{r\epsilon}{4} \times \frac{r\epsilon}{4}$. Since the largest distance between two points in P is r, the grid will be of dimensions at most $\frac{4}{\epsilon} \times \frac{4}{\epsilon}$, and hence contains at most $16/\epsilon^2$ intersection points. Our algorithm will check all the $16/\epsilon^2$ intersection points of the grid as possible centers and outputs the best one. Since the optimal center is somewhere inside the grid, and that the diameter of every cell of the grid is at most $\sqrt{2}\frac{r\epsilon}{4} \leq \frac{r\epsilon}{2}$, there is an intersection point which is in distance at most $\frac{r\epsilon}{2}$ from the optimal solution, and thus the output of the algorithm is at most $\mathrm{OPT} + \frac{r\epsilon}{2} \leq (1+\epsilon)\mathrm{OPT}$ (here we used $\frac{r}{2} \leq \mathrm{OPT}$).

Solution 2: One can also give a polynomial algorithm for the problem. Note that either the optimal circle contains two points from P on its perimeter who are antipodals, or there are at least three points from P on its perimeter (Otherwise we can shrink the circle). Note further that given three points, there is a unique circle that passes through them. Hence to find the optimal center we can pick all the possible 3 points and check the unique circle that passes through them, and also pick all the set of the 2 points check the center which is exactly in the middle of the line segment that connects them. The algorithm then outputs the best of these points.

2. (10 Points) Consider the matrix multiplication verification problem (whether $AB = C$ or not). Consider a modification of the Freivalds' algorithm where now at each round we pick a random vector $b \in \{-1, 0, 1\}^n$ and compare $A(Bb)$ and Cb. What is the probability of error for this algorithm in a single round?

   We proceed with analysis as in Freivalds' algorithm, the only difference is that the size of the set of possible values of b has increased by one.

   So suppose $AB \neq C$ but $A(Bb) = Cb$. Let $D = AB - C$, by our assumption $D \neq 0$, thus there exists some entry of it, say $d_{ij} \neq 0$. On the other hand, $Db = 0$, thus

$$\sum_k d_{i,k} b_k = 0.$$

   So we obtain that

$$\mathbb{P}[Db = 0] = \mathbb{P}[b_j = \frac{\sum_{k \neq j} d_{i,k} b_k}{d_{i,j}}] \leq \frac{1}{3}.$$

   The last inequality holds because every coordinate in vector b uniformly gets one of the three possible values, so the probability that this is exactly the desired value (i.e. $\frac{\sum_{k \neq j} d_{i,k} b_k}{d_{i,j}}$) is at most $1/3$. Thus, if we run the modified algorithm k times, the probability of error is at most 3^{-k}.

3. (20 Points) In the MAX-k-COL problem we are given a graph G and we want to color the vertices of the graph with k-colors so as to maximize the number of edges whose endpoints are in different colors.

Give an efficient randomized $(1-\frac{1}{k})$-factor approximation algorithm for the MAX-k-COL problem. (That is the expected value of the output of the algorithm is at least $\frac{k-1}{k}$ times the optimal solution.)

Solution: We color every vertex uniformly at random from the set of colors $\{1,2,\ldots,k\}$. We say that the edge $e=(u,v)$ is *good* if u and v get different colors, otherwise we call it *bad*. The probability that the edge e is bad is $1/k$. Indeed,

$$\mathbb{P}[e \textit{ is bad}] = \sum_{x\in\{1,\ldots,k\}} \mathbb{P}[c(u)=x]\cdot\mathbb{P}[c(v)=x] = \sum_{x\in\{1,\ldots,k\}} \frac{1}{k}\cdot\frac{1}{k} = \frac{1}{k}.$$

Note that in the first equality above we use the fact that colors are picked independently. Hence the probability that e is good is $(k-1)/k$. Let X_e be the indicator variable for edge e being good and let X be the total number of good edges. Then by linearity of expectation we have

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_e X_e\right] = \sum_e \mathbb{E}[X_e] = \frac{k-1}{k}|E| \geq \frac{k-1}{k}OPT^*,$$

where in the last inequality we use the trivial fact that at best all the edges are good in the optimal solution.

4. (20 Points) Consider a k-CNF (each clause contains k terms) with $m < 2^{k-1}$ clauses. Suppose further that the variables involved in any given clause are distinct. Show that the probability that the following randomized algorithm does <u>not</u> succeed in satisfying ϕ is at most 2^{-1000}.

   - For $i=1,\ldots,1000$:
   - Pick a random truth assignment σ.
   - If σ satisfies the CNF, then output σ and terminate.

   Solution: Pick a random truth assignment σ. Let X_i be the indicator variable for ith clause being false and let X be the number of clauses that are false with σ. We want to evaluate the probability that $X>0$. Using the linearity of expectation we have

$$\mathbb{E}[X] = \sum_{i=1}^{m}\mathbb{E}[X_i] = \sum_{i=1}^{m}\mathbb{P}[X_i=1],$$

and also note that obviously

$$\mathbb{E}[X] = \sum_{i=0}^{m} \mathbb{P}[X = i] \times i \geq \sum_{i=1}^{m} \mathbb{P}[X = i] = \Pr[X > 0]$$

So let us evaluate the probability that a fixed clause is not satisfied with σ. It means that all the terms in this clause are FALSE and since by assumption, they are all distinct, this happens with probability $1/2^k$. Thus,

$$\mathbb{P}[X > 0] \leq \frac{m}{2^k} < \frac{1}{2}.$$

We run the same algorithm 1000 times, picking σ's independently, thus the probability of failure after all runs is strictly less than 2^{-1000}.

5. (20 Points) Consider a 4-regular graph (i.e. each vertex is incident to 4 edges) on n vertices. The following algorithm produces a set S such that every vertex either belongs to S or has at least one neighbour in S:

   - Include every vertex in S independently with probability p.
   - Add to S all the vertices that have no neighbours in S.
   - Output S.

   What is the expected size of the outputted S? What value of p minimizes this expected value?

   Solution: The probability that the vertex v is in S is $p+(1-p)^5$, that is, either it was chosen at the first step with probability p or if not, none of the neighbours of v are in S (with probability $(1-p)\cdot(1-p)^4$). Let X_v be the indicator variable for vertex v being in the set S and let X be the random variable corresponding to the size of S. Then by linearity of expecation we have

$$\mathbb{E}[X] = \sum_{v} \mathbb{E}[X_v] = (p + (1-p)^5)n.$$

   Taking the derivative of $p + (1-p)^5$ we find the value of p for which the expected size of S is minimized. It corresponds to the $0 < p < 1$ solution of the equality $1 - 5(1-p)^4 = 0$, that is, $p = 1 - 5^{-1/4}$.

6. (20 Points) Let $v_1, \dots, v_n$ be *unit* vectors in $\mathbb{R}^n$. Prove that there exists $\epsilon_1, \dots, \epsilon_n = \pm 1$ such that

$$|\epsilon_1 v_1 + \ldots + \epsilon_n v_n| \le \sqrt{n}.$$

Hint:

$$|\epsilon_1 v_1 + \ldots + \epsilon_n v_n|^2 = \langle \epsilon_1 v_1 + \ldots + \epsilon_n v_n, \epsilon_1 v_1 + \ldots + \epsilon_n v_n \rangle.$$

Solution: Let X be the random variable $|\epsilon_1 v_1 + \ldots + \epsilon_n v_n|^2$. By the hint we know that

$$X = \sum_{i,j=1}^{n} \epsilon_i \epsilon_j v_i \cdot v_j.$$

Every ϵ_i is uniformly either -1 or 1, hence its expectation is simply zero. On the other hand, ϵ_i^2 is constant one. Thus, using the independence of ϵ_i's and the linearity of expectation we obtain the following.

$$\mathbb{E}[X] = \sum_{i,j=1}^{n} \mathbb{E}[\epsilon_i \epsilon_j] v_i \cdot v_j = \sum_{i \ne j} \mathbb{E}[\epsilon_i \epsilon_j] v_i \cdot v_j + \sum_{i=1}^{n} \mathbb{E}[\epsilon_i^2] v_i \cdot v_i = \sum_{i=1}^{n} v_i \cdot v_i = n.$$

Thus, there will be a choice of ϵ_i's for which $X \le n$, as desired.