Randomized Algorithms assignment 5

Rasmus Stavenuiter, Jacob Harder

The probabilistic method

Two main ideas:

- Any random variable assumes at least one value no less and one value no greater than its expectation with positive probability
- If an object chosen according to some distribution from a universe satisfies a property with positive probability then at least one object from this universe satisifies said property.

Usefull observations for existence proofs: construct thought experiments and show that the output of such experiments satisfy certain properties with positive probability or proof statements about the expected outcome of the experiment.

Example

For an undirected graph $G(E, V)$ with n vertices and m edges, there is a partition of the vertices into sets A and B such that

$$\#\{(u, v) \in E \mid u \in A \text{ and } v \in B\} \geq \frac{m}{2}$$

To see this consider the experiment of choosing A and B by independently and equiprobably assigning each vertex to either A or B and compute the expected size of the above set.

Theorem For n large enough there exists a bipartite graph $G(L, R, E)$ with $\#L = n$ and $\#R = 2^{\log^2 n}$ such that

- Every subset of $n/2$ vertices of L has at least $2^{\log^2 n} - n$ neighbors in R
- No vertex of R has more thatn $12 \log^2 n$ neighbors

Can be proven by considering the experiment of letting each vertex in L choose $2^{\log^2 n} 4 \log^2 n/n$ neighbors in R independently at random and considering the probability of the thus obtained graph satisfying the above requirements (its strictly positive).

Such expanding graphs can be used for *probability amplification*: in the case where we have an RP algorithm, A, for deciding membership of a language L with witnesses from $\mathbb{Z}_n$ we have seen how two point sampling can be used to obtain an error probability of less than $1/t$ with t trials, using $2 \log n$ random bits to sample two random numbers from $\mathbb{Z}_n$.
Using $\log^2 n$ bits of randomness we can distinguish a vertex of R in such a graph and consider its (at most $12 \log^2 n$) neighbors in L, $r_1, \cdots, r_k$ using these as potential witnesses.
By the above theorem the error probability obtained by using these neighbors in stead of random numbers is at most $n/n^{\log n}$.

We observe that the above theorem only asserts the existence of a graph with the stated properties and does not construct it, but at least the theorem guarantees that a search for such a graph will not be in vain, and once one such graph is found it can be used for probability amplification over and over again.

Problem 5.3

(a)

Each vertice survives the deletion process with probability $\frac{1}{d}$ so the number of surviving vertices is binomially distributed with trial parameter n and success parameter $\frac{1}{d}$ and hence the expected number of surviving vertices is $\frac{n}{d}$.

Now observe that an edge is deleted if and only if one of its end points is. Hence it survives if and only if both its end points do. Since the deletions are performed independently, this happens with probability $\frac{1}{d^2}$, so the number of surviving edges is binomially distributed with trial parameter $\frac{nd}{2}$ and success parameter $\frac{1}{d^2}$ and as before we expect $\frac{nd}{2} \cdot \frac{1}{d^2} = \frac{n}{2d}$ surviving edges.

(b)

Consider the following extention of the experiment:
After the deletion process iterate through the remaining edges deleting one of its end points (in any fashion - deterministically or probabilistically) and its incident edges.

This will yield an independent subset (which might be empty!) since any vertice still remaining after this procedure will have all its neighbors deleted.

If v_0 and e_0 denote the number of vertices resp. edges remaining after the first deletion round then the resulting independent subset after the second deletion process will have at least $\max\{0, v_0 - e_0\}$ vertices.

By subproblem (a) the expected number of vertices in the independent set resulting from this experiment run on a graph with n vertices and $\frac{nd}{2}$ edges is at least $\max\left\{0, \frac{n}{d} - \frac{n}{2d}\right\} = \max\left\{0, \frac{n}{2d}\right\} = \frac{n}{2d}$.
Since a random variable assumes at least one value no less than its expectation with positive probability this yields the existence of at least one independent set of at least $\frac{n}{2d}$ vertices.

1 Problem 5.4

1.1 (a)

Suppose $g(x) = ax + b$. Let $x \in [0, 1]$. Then

$$\begin{aligned} f(x) &= f((1-x)\cdot 0 + (1-(1-x))\cdot 1) \\ &\geq (1-x)f(0) + xf(1) \\ &\geq (1-x)g(0) + xg(1) \\ &= (1-x)b + x(a+b) \\ &= b + ax = g(x) \end{aligned}$$

1.2 (b)

Assuming $k \in \mathbb{N}_0$ and that f is defined on $[0, 1]$. When $k = 0, 1$ f is constant (or undefined) or linear so concave. Otherwise f is smooth on $[0, 1]$ with second derivative

$$f''(x) = -\frac{k-1}{k}\left(1 - \frac{x}{k}\right)^{k-2} < 0$$

and so concave.

For $k \in -\mathbb{N}$ f is not defined on 0, but smooth on $(0, 1]$ with the same second derivative as before, so concave again.

1.3 (c)

Since $f(0) = 0 \geq 0 = g(0)$, $g(1) = (1 - (1/k))^k \leq (1 - (1/k))^k = f(1)$, and g is linear, we are done by (a) and (b).

Problem 5.5

Say each $\ell \in L$ chooses $n^{3/4}$ R-neighbors at random *without* replacement. Thus for fixed $r \in R, \ell \in L$

$$P(\ell \sim r) = \frac{n^{3/4}}{n} = n^{-1/4}$$

So (for r still fixed)

$$\begin{aligned} P(d(r) > 3n^{3/4}) &= \binom{n}{3n^{3/4}} P(\ell \sim r)^{3n^{3/4}} \\ &= \left(\frac{ne}{3n^{3/4}}\right)^{3n^{3/4}} \left(n^{-1/4}\right)^{3n^{3/4}} \\ &= \left(\frac{ne}{3n}\right)^{3n^{3/4}} \\ &= \left(\frac{e}{3}\right)^{3n^{3/4}} \end{aligned}$$

So

$$P(\exists r \in R \mid d(r) > 3n^{3/4}) \leq n \left(\frac{e}{3}\right)^{3n^{3/4}} \overset{n\to\infty}{\longrightarrow} 0$$

In particular, at some point this probability is less that $1/2$. This ensures the condition that every vertex in R has degree at most $3n^{3/4}$.

Lets estimate the probability that some subset of $n^{3/4}$ vertices in L does *not* have $n - n^{3/4}$ R-neighbors. This is

$$\begin{aligned} &P(\exists \ell_1, \cdots, \ell_{n^{3/4}} \in L, r_1, \cdots, r_{n^{3/4}} \in R \mid \ell_i \not\sim r_j \forall i, j) \\ &\leq \binom{n}{n^{3/4}} \binom{n}{n^{3/4}} P(\ell_1, \ldots, \ell_{n^{3/4}} \not\sim r_1, \ldots, r_{n^{3/4}}) \\ &\approx \left(\frac{ne}{n^{3/4}}\right)^{2n^{3/4}} P(r \not\sim \ell_1, \ldots, r \not\sim \ell_{n^{3/4}}) \\ &= e^{(2+\ln(n)/2)n^{3/4}} \left(\frac{n - n^{3/4}}{n}\right)^{n^{3/4}} \\ &\approx e^{(2+\ln(n)/2)n^{3/4} - n^{5/4}} \overset{n\to\infty}{\longrightarrow} 0 \end{aligned}$$

Thus for n large enough, by union bounding, the probability that our conditions is not satisfied is less than 1.