With high probability

So far we have been mainly concerned with expected behaviour: expected running times, expected competitive ratio's. But it would often be much more interesting if we would be able to show that realisations that are negatively far away from expected behaviour does not occur too frequently. It would be much more useful if we would know that the running time of an algorithm is almost always small, i.e., with high probability. Techniques for deriving such statements is the subject of Chapters 3 and 4 of [MR].

Many theorems in probability theory exist that describe deviation from expectations, so-called tail probabilities, like Markov's Inequality and Chebyshev's Inequality. I assume these inequalities to be known to you, but will cite them only when used.

The Coupon Collectors Problem and its application to Stable Marriages

Another very nice application of bounding probabilities is the Stable Marriage Problem. We will see that it we can find an efficient algorithm for this problem, the analysis of which is based on analysis of a fundamental stochastic process, which in the computer science literature has become known under the name *Coupon Collectors Problem*. I will explain it in the more traditional bins-and-balls terms.

Coupon Collectors Problem

Given n bins and m balls, each of which is thrown independently in one of the bins according to a uniform distribution over the n bins (each bin has probability $1/n$ of receiving the ball). We are interested in the relation between m and the probability that every bin has received at least one ball. What we will see is that there is a surprisingly sharp bound on m such that if m is slightly smaller than this bound then the probability goes to 0 and if m is slightly larger than the bound, the probability is going to 1. Such a bound we call a *threshold*. Such thresholds are common in stochastic process theory in general and in random graph theory in particular.

To be precise, what we will see is that for $m = n \log n + cn$,

$$\lim_{n\to\infty} Pr\{\text{more than } m \text{ balls needed}\} = 1 - e^{-e^{-c}}$$

and for $m = n \log n - cn$,

$$\lim_{n\to\infty} Pr\{\text{less than } m \text{ balls needed}\} = e^{-e^{c}}.$$

Let X be the random variable that defines the number of balls needed. Let ξ_i^r be the event that bin i has no ball after r trials. Thus,

$$Pr\{X > m\} = Pr\{\cup_{i=1}^{n} \xi_i^m\}.$$

We apply the so-called *inclusion-exclusion principle* for dependent events $\xi_1, \ldots, \xi_n$ from probability theory:

$$\begin{aligned} Pr\{\cup_{i=1}^{n} \xi_i\} &= \sum_{i=1}^{n} Pr\{\xi_i\} - \sum\sum_{i<j} Pr\{\xi_i \cap \xi_j\} + \sum\sum\sum_{i<j<\ell} Pr\{\xi_i \cap \xi_j \cap \xi_\ell\} \ldots \\ &= \sum_{k=1}^{n} (-1)^{k+1} \sum \cdots \sum_{i_1 \leq \ldots \leq i_k} Pr\{\cap_{j=1}^{k} \xi_{i_j}\}. \end{aligned}$$

Denote

$$P_k^n = \sum \cdots \sum_{i_1 \leq \ldots \leq i_k} Pr\{\cap_{j=1}^{k} \xi_{i_j}\}$$

and

$$S_k^n = P_1^n - P_2^n + \ldots + (-1)^{k+1} P_k^n.$$

Since $P_{k+1}^n \leq P_k^n$ we have the following sandwich property:

$$S_{2k}^n \leq Pr\{\cup_{i=1}^{n} \xi_i\} \leq S_{2k+1}^n$$

and

$$S_{2k+1}^n - S_{2k}^n = P_{2k+1}^n.$$

Notice that $Pr\{\cap_{j=1}^{k} \xi_{i_1}^m\}$ is independent of the particular choice of $i_1, \ldots, i_k$. Hence:

$$P_k^n = \binom{n}{k} Pr\{\cap_{i=1}^{k} \xi_i^m\},$$

with $Pr\{\cap_{i=1}^{k} \xi_i^m\}$ being the probability that none of the first m balls landed in any of the first k bins. That is

$$Pr\{\cap_{i=1}^{k} \xi_i^m\} = (1 - \frac{k}{n})^m$$

and therefore

$$P_k^n = \binom{n}{k} (1 - \frac{k}{n})^m.$$

Taking limits and using Lemma 3.7 in [RM] yields

$$\lim_{n\to\infty} P_k^n = \lim_{n\to\infty} \binom{n}{k} (1 - \frac{k}{n})^{n \log n + cn} = \frac{e^{-ck}}{k!}.$$

Hence for finite k,

$$\lim_{n\to\infty} S_k^n = \sum_{j=1}^{k} (-1)^{j+1} \frac{e^{-cj}}{j!} := S_k.$$

These are the first k terms of the power expansion of the function $f(c) = 1 - e^{-e^{-c}}$ and elementary function approximation theory tells us that for k large enough $|f(c) - S_k| < \epsilon$ and since $|S_k^n - S_k| < \epsilon$ for k large enough, we have that for k large enough that

$$|Pr\{\cup_{i=1}^n \xi_i^m\} - f(c)| < \epsilon.$$

I.e.,

$$\lim_{n \to \infty} Pr\{\cup_{i=1}^n \xi_i^m\} = lim_{n\to\infty} Pr\{X > n \log n + cn\} = 1 - e^{-e^{-c}}.$$

A similar proof holds for

$$lim_{n\to\infty} Pr\{X \leq n \log n - cn\} = e^{-e^{c}}.$$

Hence,

$$lim_{n\to\infty} Pr\{n \log n - cn \leq X \leq n \log n + cn\} = e^{-e^{-c}} - e^{-e^{c}},$$

which goes to 1 rapidly with growing c.

The Stable Marriage

The nature of the result that we will see here is different from what we have seen so far. We will study a deterministic algorithm in its average case performance behavior, i.o. worst case behavior of a randomized algorithm.

The stable marriage problem is a problem that can be formulated on a bipartite graph. It reflects relations in the occidental world, where official relations are predominantly monogamous.

Assume we have a set of of n men $\{A, B, C, \ldots\}$ and a set of n women $\{a, b, v, \ldots\}$. Each man has a complete preference list of all women and each woman has a preference list of all man. A stable marriage is a perfect matching of men to women such that the following situation does not occur:

- woman x is matched to man Y but prefers man X to man Y *and*
- man X is matched to woman z but prefers woman x to woman z.

In this situation X and x would start a liaison which endangers the stability of the matching. It can be proved that for any instance of the problem a stable marriage always exists. We are interested in finding one. Therefore, consider the following algorithm:
Deterministic Algorithm. Fix a list of men (in arbitrary order). Start from the top of this list and let the first man propose to his most preferred woman. A woman will only say NO if she is already matched and prefers her current match to the proposer. At any step of the algorithm the man, X say, currently on top of the list proposes to the woman, x say, who is on top of his list and who has

not already rejected him. If x says YES, then either she was unmatched and now X and x are matched, or x had already man Y but she preferred X to Y and therefore dismissed Y, i.e., the couple Y-x is replaced by X-x, and Y is put back on top of the list of men. At a proposal of a man to a woman he deletes her from his list, independent if she accepts him or not. Thus, a particular man will make a proposal to a particular woman only once. Hence the algorithm will need at most n^2 proposals to find a stable marriage.

You may verify for yourself that this algorithm indeed always finds a stable marriage.

One would expect that this algorithm will typically require less than n^2 proposals. We will show that indeed the number of proposals needed is significantly less on randomly drawn instances. More specifically, we will show that the probability that this number is higher than $n \log n + cn$ is asymptotically $1 - e^{-e^{-c}}$.

As random instance assume that each man selects his preference list independently and uniformly distributed over all possible preference lists. The women have arbitrary but fixed preference lists. Let T_P be the random variable denoting the number of proposals required.

Studying T_P directly from the algorithm is difficult because of the enormous amount of dependencies that emerge. An alternative way of looking at this model is that a man does not choose his list in advance, but each time it is his turn to propose he selects a woman uniformly at random from amongst the ones who have not rejected him yet, and proposes her.

We make even a further simplification to make the analysis amenable. By it we arrive at the Amnesiac Algorithm: A man, when it is his turn to propose, picks a woman uniformly from amongst all women. He forgets it if she he has rejected him already before. If she did, then she will definitely reject him again and the outcome will be exactly the same as the original algorithm. However, clearly an extra number of superfluous proposals will occur in the Amnesiac Algorithm.

Let us denote by T_A the number of proposals needed by the Amnesiac Algorithm. Then T_A dominates stochastically T_P:

$$Pr\{T_A > m\} \geq Pr\{T_P > m\}.$$

But T_A is fairly simple to analyse. Notice that the algorithm stops as soon as the last woman on the list has received a proposal. Thus is we see the woman as bins and a proposal as the throwing of a ball then T_A is exactly the number of balls needed to be thrown independently, uniformly over the bins such that each bin has received at least one ball. Thus, we can directly apply the result of the COUPON COLLECTORS PROBLEM to conclude

$$lim_{n\to\infty} Pr\{T_A > n \log n + cn\} = 1 - e^{-e^{-c}}.$$

Randomized rounding

The probability bounds that we used so far were supposing no more than finite second moments. If we may assume more we can also derive stronger bounds on tail probabilities. One such bound is Chernoff's bound. This is a very popular bound for deriving strong results in randomised approximation algorithms for hard problems. [MR]. We will show an example of that.

Theorem 4.1, 4.2, 4.3 [MR]. *Chernoff Bounds.* Let $X_1, \ldots, X_n$ be independent random variables with $Pr\{X_i = 1\} = p_i$, $Pr\{X_i = 0\} = 1 - p_i$, $0 < p_i < 1$, $i = 1, \ldots, n$. Then for $X = \sum_{i=1}^n X_i$, we have $E[X] = \mu = \sum_{i=1}^n p_i$ and for $\delta > 0$,

$$Pr\{X > (1+\delta)\mu\} < \left(\frac{e^\delta}{(\delta+1)^{\delta+1}}\right)^\mu.$$

For $0 < \delta \leq 2e - 1$,

$$Pr\{X > (1+\delta)\mu\} < e^{-\mu\delta^2/4}.$$

For $0 < \delta \leq 1$,

$$Pr\{X < (1-\delta)\mu\} < e^{-\mu\delta^2/2}.$$

The proof uses the moment generating function $E[e^{tX}]$ and you can read it in the book. We will show an application that is interesting on its own since it uses a technique that has proven to be very powerful in algorithmic approximation theory: *randomised rounding.* One starts by formulating the combinatorial optimisation problem to be solved by a $\{0,1\}$ integer linear program, solves its LP-relaxation and use the optimal solution as probabilities to round it to an integer solution.

The Global Wiring Problem

Consider a $\sqrt{n} \times \sqrt{n}$ array of what we call n gates. We number them $1, \ldots, n$. Pairs of gates are to be connected by wires, that run parallel to the axes. An example is drawn on the blackboard. We are not concerned about the exact location of the wires in the gates, but just with the gates that a wire passes from its one end-point to its other end-point.

In the boundary of each pair of adjacent gates there is only a limited number of holes, each hole can accommodate one wire. This imposes a constraint on the number of wires that can pass the boundary between two gates.

Global Wiring Problem. Given a set of gate-pairs, put wires to connect the pairs, such that the number of wires that pass through a boundary between two gates does not exceed the number of holes in the boundary.

We assume that each boundary has the same number W of holes and solve the feasibility problem by the following optimisation version. Let $W_S(b)$ be the number of wires that pass boundary b in wiring solution S then we wish to find

$$W^{OPT} = \min_S W_S = \min_S\{\max_b W_S(b)\}.$$

Thus, if $W^{OPT} \leq W$ then the optimal solution is feasible. We simplify the problem even further for didactic purposes. We will allow only zero- and one-bend wires. I.e., each wire has at most one piece that runs vertically and at most one piece that runs horizontally. (In the figure it will be made clear which wires do and which don't satisfy this restriction.) Think of wiring a pair of gates as going from its leftmost gate to rightmost gate. Then for a one-bend wire we may choose to go first horizontally and then vertically or the other way round. If the gate-pair is in the same column or row, then there is no choice and they are connected by a zero-bend wire.

Build on this insight we can now formulate a $\{0,1\}$ integer linear programming problem. For wire (gate-pair) i, set $x_{i0} = 1$ if it goes first horizontally and $x_{i1} = 1$ if it goes first vertically; x_{i0} and x_{i1} are 0 otherwise. Let

$$T_{b0} = \{i \mid \text{wire } i \text{ passes through } b \text{ if } x_{i0} = 1\}$$

and

$$T_{b1} = \{i \mid \text{wire } i \text{ passes through } b \text{ if } x_{i1} = 1\}.$$

It is crucial to notice that $T_{b0} \cap T_{b1} = \emptyset$. Then we seek

$$\begin{array}{rll} W^{OPT} = & \min\ w & \\ \text{s.t.} & x_{i0} + x_{i1} = 1 & \forall\ i \\ & \sum_{i \in T_{b0}} x_{i0} + \sum_{i \in T_{b1}} x_{i1} \leq w & \forall\ b \\ & x_{i0}, x_{i1} \in \{0,1\} & \forall\ i \end{array}$$

This problem is known to be NP-hard. We solve the LP-relaxation obtained by replacing the constraints $x_{i0}, x_{i1} \in \{0,1\}$ by $0 \leq x_{i0}, x_{i1} \leq 1, \forall\ i$. Obviously the optimal solution value will be a lower bound on the value of the optimal integer solution. Let $\hat{x}_{i0}$, $\hat{x}_{i1}$ denote the optimal solution of the LP-relaxation and $\hat{w}$ its value. Thus $W^{OPT} \geq \hat{w}$.

If $\hat{x}_{i0}, \hat{x}_{i1}$ are fractional and therefore not feasible for the integer problem, we propose to round them randomly, by letting $\hat{x}_{i0}, \hat{x}_{i1}$ dictate the rounding probabilities: we set $\bar{x}_{i0}$ equal to 1 with probability $\hat{x}_{i0}$. So,

$$Pr\{\bar{x}_{i0} = 1\} = \hat{x}_{i0} \text{ andhence } Pr\{\bar{x}_{i1} = 1\} = \hat{x}_{i1} = 1 - \hat{x}_{i0}.$$

We do this independently for all the wires. Clearly this gives a feasible solution to the integer problem. So for wire i we flip a coin with TAIL probability $\hat{x}_{i0}$ and if it turns up TAIL we set $\bar{x}_{i0} = 1$ and $\bar{x}_{i1} = 0$, otherwise we set $\bar{x}_{i1} = 1$ and

$\bar{x}_{i0} = 0$. This yield a random solution S.

Let us study the random variable $W_S(b)$ that now emerges from the random solution S, for a particular but arbitrary boundary b. We know because of LP-feasibility that

$$\sum_{i \in T_{b0}} \hat{x}_{i0} + \sum_{i \in T_{b1}} \hat{x}_{i1} \leq \hat{w}.$$

Moreover, we have

$$W_S(b) = \sum_{i \in T_{b0}} \bar{x}_{i0} + \sum_{i \in T_{b1}} \bar{x}_{i1}.$$

Therefore,

$$\begin{aligned} E[W_S(b)] &= \sum_{i \in T_{b0}} E[\bar{x}_{i0}] + \sum_{i \in T_{b1}} E[\bar{x}_{i1}]. \\ &= \sum_{i \in T_{b0}} \hat{x}_{i0} + \sum_{i \in T_{b1}} \hat{x}_{i1} \leq \hat{w}. \end{aligned}$$

Since $\bar{x}_{i0}, \bar{x}_{i1}$ are independent from $\bar{x}_{j0}, \bar{x}_{j1}$ for $i \neq j$, we may apply the Chernoff bound.

$$Pr\{W_S(b) = \sum_{i \in T_{b0}} \bar{x}_{i0} + \sum_{i \in T_{b1}} \bar{x}_{i1} > \hat{w}(1+\delta)\} \leq \left(\frac{e^{\delta}}{(\delta+1)^{\delta+1}} \right)^{\hat{w}}.$$

Notice that here we need that T_{b0} and T_{b1} are disjoint. Let $\Delta(\hat{w}, \epsilon)$ be the smallest value of δ for which $\left(\frac{e^{\delta}}{(\delta+1)^{\delta+1}} \right)^{\hat{w}} \leq \frac{\epsilon}{2n}$. Then, using W_S to denote the random solution value obtained by this randomised rounding algorithm, we have

$$Pr\left\{\{W_S > \hat{w}(1 + \Delta(\hat{w}, \epsilon))\}\right\} \leq \sum_{b} Pr\{W_S(b) > \hat{w}(1 + \Delta(\hat{w}, \epsilon))\} \leq \epsilon,$$

since the number of boundaries is less than $2n$.

Thus, what we found is that with probability $1 - \epsilon$

$$W_S \leq \hat{w}(1 + \Delta(\hat{w}, \epsilon)) \leq W_{OPT}(1 + \Delta(\hat{w}, \epsilon)).$$

Hence, the quality of the result depends (through $\Delta(\hat{w}, \epsilon)$) on the optimal LP-value $\hat{w}$. If e.g. $\hat{w} = n^{\gamma}$ then with probability $1 - \epsilon$

$$W_S \leq W_{OPT} \left(1 + \sqrt{\frac{4 \ln(2n/\epsilon)}{n^{\gamma}}} \right)$$

and we find a solution that is very close to optimal for large n.

If $\hat{w} = c$ for some constant c then the ratio obtained is not good at all. However, we always have the alternative to round deterministically such that the worst-case ratio is at most 2, as you will be asked to find out in Exercise 4.7 in [MR].

Exercises

Exercises Week 3: From Chapter 3 in [MR]: Exercises 3.5 and 3.6 and Problems 3.13 to be proven $1 - e^{-2e^{-c}}$.
Open Problems are 3.11, 3.15

Material

During the fourth lecture I treated:
Sections 3.1, 3.2, 3.5 and 3.6 of [MR]