Institut for Matematik og Datalogi
Syddansk Universitet, Odense

13 November 2023
JBJ

DM551-MM851 – 2. Exam assignment

Hand in by Monday December 4 at 09:00.

Rules

This is the second of two sets of problems which together with the oral exam in January constitute the exam in DM551/MM851. **This second set of problems must be solved individually.** Any collaboration with other students will be considered as exam fraud. Thus you are not allowed to show your solutions to fellow students. On the other hand, you can learn a lot from discussing the problems with each other so you may do this to some extend, such as which methods can be used or similar problems from the book or exercise classes.

Remember that this counts as part of your exam, so do a good job and try to answer all questions carefully. It is important that you **argue so that the reader can follow your calculations and explanations**.

How to hand in your report

Your report, which should be written in english, must be handed in on itslearning by Monday December 4, 2023 at 9:00.

On the first page you must write your **name** and the first 6 digits of your **CPR-number**. Do not write the last 4 digits!

Institut for Matematik og Datalogi
Syddansk Universitet, Odense

13 November 2023
JBJ

Problems

Solve the following problems and **Remember to justify all answers:**

Problem 1 (15 points)

This problem is about random permutations of a set with n distinct elements. For simplicity we assume that $n = 2^k$ for some $k > 0$. Suppose we have an algorithm $\mathcal{A}$ which given a set X of $2r$ elements generates a random subset X_1 of size r. We want to use this to make an algorithm RANDPERM which takes a set S of 2^p vertices and returns a random permutation of S. RANDPERM will do this as follows:

- If $p > 0$ then first use $\mathcal{A}$ to find a random subset X_1 of S of size 2^{p-1}. Then call RANDPERM recursively on X_1 and $X_2 = S \setminus X_1$ respectively to get permutations π_1 and π_2 of X_1 and X_2 respectively and return the permutation $\pi = \pi_1\pi_2$.
- If $p = 0$, that is $S = \{x\}$ for some x, just return x (the permutation with just that element).

For example, if $n = 4$ and $X = \{a, b, c, d\}$ the algorithm $\mathcal{A}$ may select $\{a, c\}$ as X_1 (so $X_2 = \{b, d\}$) and the two recursive calls to RANDPERM may result in the permutations $\pi_1 = ca$ and $\pi_2 = bd$, respectively so the algorithm would return the permutation $\pi = \pi_1\pi_2 = cabd$.

Question a:

Prove by induction on k that if $\mathcal{A}$ works correctly, then RANDPERM returns a random permutation of any given set S of $n = 2^k$ elements.

RANDPERM clearly works when there is only one element. Suppose it works when there are at most $r = 2^{k-1}$ elements and consider a set of $2r$ elements. The probability that RANDPERM returns a fixed permutation $a_1 \dots a_r, a_{r+1} \dots a_{2r}$ is the product of the probabilities that $\mathcal{A}$ returns the set $X_1 = \{a_1, \dots, a_r\}$, the probability that the call of RANDPERM on X_1 returns the permutation $a_1 \dots a_r$ and the probability that the call of RANDPERM on on $X_2 = \{a_{r+1}, \dots, a_{2r}\}$ returns the permutation $a_{r+1} \dots a_{2r}$. Since we have assumed that $\mathcal{A}$ works correctly, the first probability is $\frac{1}{\binom{2r}{r}} = \frac{r!r!}{(2r)!}$. By induction, the probability that RANDPERM returns each of the desired permutations is $\frac{1}{r!}$. Combining this we get that the probability that RANDPERM returns the permutation $a_1 \dots a_r, a_{r+1} \dots a_{2r}$ is $\frac{r!r!}{(2r)!}(\frac{1}{r!})^2 = \frac{1}{(2r)!}$ as desired.

Question b:

Describe an implementation of the algorithm $\mathcal{A}$ and prove that this will return a random subset of size $\frac{S}{2}$ of S when the input is a set S with an even number of elements.

To obtain an implementation of $\mathcal{A}$ we can take the first $r = \frac{n}{2}$ iterations of the algorithm RANDOMIZE-IN-PLACE from Cormen. We proved that after these r steps the first r elements of the array form in fact a random r-permutation of the full set on $2r$ elements. Now we can return the subset consisting of the first r elements and since there are $r!$ permutations of a fixed r-set X_1 of S, the probability that we get X_1 as output from $\mathcal{A}$ is $r!\frac{r!}{(2r)!} = \frac{r!r!}{(2r)!} = \frac{1}{\binom{2r}{r}}$ as desired.

Problem 2 (10%)

Solve the recurrence equation $a_n = a_{n-1} + 2a_{n-2}$ with initial conditions $a_0 = 4$ and $a_1 = 6$.

The characteristic equation is $x^2 - x - 2 = 0$. This has roots -1 and 2 so the solutions the the recurrence equation are of the form $\alpha_1(-1)^n + \alpha_2 2^n$. Now using the initial conditions we get 2 equations with 2 unknown (α_1, α_2):

$$\begin{aligned} 4 = a_0 &= \alpha_1 + \alpha_2 \\ 6 = a_1 &= -\alpha_1 + 2\alpha_2 \end{aligned}$$

This implies (adding the equations) that $3\alpha_2 = 10$ and hence $\alpha_2 = \frac{10}{3}$. Inserting this in the equation for a_0 we get that $4 = \alpha_0 + \frac{10}{4}$ so $\alpha_1 = \frac{2}{3}$. Thus the solution to the recurrence equation is $a_n = \frac{2}{3}(-1)^n + \frac{10}{3}2^n$.

Problem 3 (15 %)

Consider the following variant of the coupon collector problem where the number of distinct coupons n is an even number: Each time you collect two coupons. If they are both new (not among those you already have) you can keep them. If at most of them is new, you cannot keep any of them and hence make no progress towards the goal which is still to collect all n different coupons.

Question a:

Suppose you have i of the coupons already for some even number $0 \leq i < n$. What is the probability of having $i + 2$ coupons after the next step?

There are $n - i$ coupons that we still do not have so we can choose two of these in $\binom{n-i}{2}$ ways. There are $\binom{n}{2}$ ways of choosing two coupons so the probability of getting a good pair is $\frac{\binom{n-i}{2}}{\binom{n}{2}}$

Question b:

Prove that when you have exactly $i < n$ different coupons, for some even integer i, the expected number of times you have to collect two coupons before you will have $i + 2$ coupons is $\frac{\binom{n}{2}}{\binom{n-i}{2}}$.

The success probability p (getting two new cards) was obtained above. The expected number of pairs we must choose before success is $\frac{1}{p}$ which is $\frac{\binom{n}{2}}{\binom{n-i}{2}}$.

Question c:

Show that the expected number of coupons you need to collect before you have all coupons is $O(n^2)$. You may use that $\frac{1}{q(q-1)} = \frac{1}{q-1} - \frac{1}{q}$.

Let $k = \frac{n}{2}$ and define phases $F_0, F_1, \ldots, F_{k-1}$ so that in phase F_i we have $2i$ distinct cards. Let $X_0, X_1, \ldots, X_{k-1}$ denote the number of pairs we collect in each of the phases before we go to the next one. Then $X_0 = 1$, $X = \sum_{i=0}^{k-1} X_i$ is the total number of pairs collected and it follows from (b) that $E(X_i) = \frac{\binom{n}{2}}{\binom{n-i}{2}}$. Hence

$$\begin{aligned}
E(X) &= E(\sum_{i=0}^{k-1} X_i) \\
&= \sum_{i=0}^{k-1} E(X_i) \\
&= \sum_{i=0}^{k-1} \frac{\binom{n}{2}}{\binom{n-2i}{2}} \\
&= \binom{n}{2} \sum_{i=0}^{k-1} \frac{1}{\binom{n-2i}{2}} \\
&= n(n-1) \sum_{i=0}^{k-1} \frac{1}{(n-2i)(n-2i-1)} \\
&= n(n-1) \sum_{i=0}^{k-1} (\frac{1}{n-2i-1} - \frac{1}{n-2i}) \\
&= n(n-1)(1 - \frac{1}{n}) = O(n^2)
\end{aligned}$$

Problem 4 (15 points)

In Kleinberg and Tardós section 13.4 you saw a randomized approximation algorithm $\mathcal{B}$ for MAX-3-SAT whose expected number of satisfied clauses is within a factor $\frac{7}{8}$ of optimal. In fact we showed that if there are m clauses in the input, then the expected number of clauses that will be satisfied by $\mathcal{B}$ is $\frac{7 \cdot m}{8}$ and from this we proved (using the probabilistic method) that every instance of 3-SAT on m clauses has a truth assignment which satisfies at least $\frac{7 \cdot m}{8}$ clauses. We also saw that the expected number of times we need to guess a truth assignment (run $\mathcal{B}$) before obtaining a truth assignment which satisfies at least $\frac{7 \cdot m}{8}$ clauses is less than $8m$.

Below we want to use the same idea to find a randomized approximation algorithm for MAX-k-SAT, where $k \geq 3$.

Question a:

What is the expected number R of clauses we will satisfy if there are m clauses? You must show how to obtain your answer.

By the technique as for $k = 3$, using an indicator random variable X_j to denote whether clause j will be satisfied by a random truth assignment, we get that $p(X_j = 1) = \frac{2^k - 1}{2^k}$ so the expected number of clauses satisfied out of m is $\frac{(2^k - 1) \cdot m}{2^k}$

Question b:

Prove that there exists a truth assignment that makes at least R of the clauses true.

Question C;

Describe a Las Vegas algorithm $\mathcal{A}$ which given a k-SAT formula $\mathcal{F}$ will find a truth assignment that makes at least R of the clauses in $\mathcal{F}$ true.
Generate random truth assignments until we reach one where at least R clauses are satisfied.

Question D:

Give an upper bound on the expected running time of $\mathcal{A}$. Hint: follow the idea in Kleinberg and Tardós section 13.4.
For 3-SAT we saw that the expected number of repetitions is at most $2^3 \cdot m$ and very similar calculations show that for k-SAT, the expected number of repetitions is at most $2^k \cdot m$.

Problem 5 (20 %)

Consider an experiment where we randomly distribute n balls into m distinct boxes labelled $B_1, B_2, \ldots, B_n$. Assume below that $n = m^2$.

Question a:

Prove that the expected number of balls in a fixed box B_i is m.

Let X_{ij} be the indicator variable that is 1 precisely if ball i lands in box B_j. Then $p(X_{ij} = 1) = \frac{1}{m}$. Let $Y_j = \sum_{i=1}^{m^2} X_{ij}$. Then, using linearity of expectation, we get $E(Y_j) = \frac{m^2}{m} = m$.

Question b:

Use Chebyshev's inequality to bound the probability that the number of balls in box B_i is more than 50% away from its expected value.

When we study the balls each landing in a fixed box i this is a Bernouilli experiment with success probability $p = \frac{1}{m}$ and m^2 experiments.Hence the expected value is (as above) $m^2 p = m^2 \frac{1}{m} = m$ and the variance is $m^2 p(1-p) = m^2 \frac{1}{m} \frac{m-1}{m} = m-1$. Now we can apply Chebyshev's inequality: let X be the number of balls landing in box i. Then the inequality says

$$p(|X - m| \geq \frac{m}{2}) \leq \frac{m-1}{(\frac{m}{2})^2} = \frac{4m-4}{m^2}$$

Question c:

Use Chernoff bounds to bound the probability that the number of balls in box B_i is more than 50% away from its expected value.

We will use (13.42) and (13.43) from KT:

$$p(X > 1.5m) < \left[\frac{e^{\frac{1}{2}}}{1.5^{1.5}} \right]^m < e^{-\frac{m}{12}}$$

$$p(X < \frac{m}{2}) < e^{-\frac{1}{2}(\frac{1}{2})^2 m} = e^{-\frac{m}{8}}$$

.

combining these two we get that the probability that X is more than 50% away from m is at most $e^{-\frac{m}{12}} + e^{-\frac{m}{8}} < 2e^{-\frac{m}{12}}$.

Question d:

Compare the two bounds above and explain the difference. Which bound is best as m gets large?

The second bound decreases exponentially in m whereas the first is roughly $\frac{4}{m}$ so as m gets larger the second bound is much tighter.

Question e:

Use your bound in Question c and the union bound to show that when $m = 100$ the probability that there is any box which has less than 50 or more than 150 balls is less than 1%.

We found $p(|X - m| > \frac{m}{2}) < 2e^{-\frac{m}{12}}$. Inserting $m = 100$ we get a bound of 0.00024037 on the probability that a particular box has less than 50 or more than 150 elements. Hence, by the union bound, the probability that there is any box with that property is at most 100 times this which is much less than 1%.