Institut for Matematik og Datalogi September 20, 2023
Syddansk Universitet JBJ

DM551/MM851 – Fall 2024– Weekly Note 4

Stuff covered in week 38

- Rosen 7.2.6-7.4 except Section 7.4.4 which is left for self study and the part on spam filters which is not pensum. We will come back to analysis of the running time of algorithms later in the course.
- Indicator random variables (see notes on weekly note 3.
- I also covered the notes on the probabilistic method on Weekly note 3.

Note that last week I gave an example of a **Monte Carlo algorithm** which is not in the book (for majority element). Let $S = [s_1, s_2, \dots, s_n]$ be a collection of n integers. An element s_i of S is a **majority element** of S if $|\{j : s_j = s_i\}| > n/2$. Note that if S has a majority element, then it has only one such (in many copies). It is easy to check whether S has a majority element: just sort the elements of S and make one scan of the sorted set to see if one value occours more than $n/2$ times. Assuming we use comparison based sorting, this takes time $\theta(n \log n)$ since (as you will learn in DM553/MM853) we cannot sort faster than $\theta(n \log n)$. We now describe a very simple randomized algorithm whose running time is $O(n)$ and whose running time can be adjusted to achieve as small (but still non-zero) error probability we wish. Let $\mathcal{A}$ be the algorithm that takes input S, m, where S is as above and m is a non-negative constant and then runs as follows:

1. Repeat m times
   (a) Pick a random element $s \in S$;
   (b) Check whether s is a majority element of S and if so return the value 'true' (and stop);
2. Return 'false'

Let us analyze the performance of $\mathcal{A}$. Note that if $\mathcal{A}$ returns 'true', then S does have a majority element, namely the element s which made the algorithm return 'true'. However, if $\mathcal{A}$ returns the value 'false', there may still be a majority element in S, in which case the answer given by $\mathcal{A}$ is wrong. First assume that S has no majority element. Then $\mathcal{A}$ will correctly return the answer 'false'. Suppose now that S has a majority element but still $\mathcal{A}$ returned 'false'. This means that the value of each of the m random elements picked by $\mathcal{A}$ is different from the value of the majority element. The probability that the value of picked element s is **not** equal to the value of the majority element is less that $1/2$ and since we make independent random choices in each of the m rounds, the probability that they are all different from the majority is at most $(1/2)^m$. By taking $m = 20$ this probability is less that $\frac{1}{1000000}$.

In the lecture on September 13th I showed that the Ramsey number $r(k, k)$ is at least $2^{k/2}$. For this we used the probabilistic method. What we showed is that if n is less than $2^{k/2}$, then at least one 2-coloring (by colours 'red' and 'blue') of the edges of K_n (there are $2^{\binom{n}{2}}$ of these by the product rule) will have the property that for **every** complete subgraph $G = K_k$ of k vertices there is at least one edge of color 'red' inside G and at least one edge of color 'blue' inside G. We proved this by showing that if we take a **random** 2-coloring of K_n (that is, we choose the colour of each edge randomly with probability 1/2 for each of the two colours), then the probability that this colouring has the property above is greater than 0, implying that at least one of the $2^{\binom{n}{2}}$ 2-colorings of K_n has no monochromatic K_k. By the definition of the Ramsey number $r(k, k)$ (minimum n such that **every** 2-coloring of K_n will have a monochromatic K_k), this shows that $r(k, k) \geq 2^{k/2}$

Lecture in week 39, 2022

- Rosen 8.5-8.6

Exercises in week 39,2022

- Rosen exercise 15 page 502. This is the so called Monty Hall puzzle. See also the notes at the bottom of the home page on this famous problem.
- Rosen Section 7.3: 3,7,10, 13,16
- Rosen Page 522: 16. This important equality will be used several times in the course.
- Rosen pages 522-524: 14,24,26,30,32
- Rosen Section 7.4: 37 This important inequality, called Markov's inequality will be used several times in the course.
- Rosen Section 7.4: 8, 12,28,38, 48.
- DM528 Exam 2010, Problem 2
- If there is more time you may discuss the part of Rosen Section 7.3 on Bayesian spam filters (it is not pensum).