

Note: Your TA may not get to all the problems. This is totally fine, the discussion worksheets are not designed to be finished in an hour. The discussion worksheet is also a resource you can use to practice, reinforce, and build upon concepts discussed in lecture, readings, and the homework.

Philosophy of analyzing randomized algorithms. The first step is to always identify a *bad event*. I.e. identify when your randomness makes your algorithm fail. We will review some techniques from class using the following problem as our “test bed”.

Let G be a bipartite graph with n left vertices, and n right vertices on $n^2 - n + 1$ edges.

- Prove that G always has a perfect matching.
- Give a polynomial in n time algorithm to find this perfect matching.

We will analyze the following algorithm **BlindMatching**:

- Let π and σ be independent and uniformly random permutations of $[n]$.
- If $\{\pi(1), \sigma(1)\}, \{\pi(2), \sigma(2)\}, \dots, \{\pi(n), \sigma(n)\}$ is a valid matching output it.
- Else output **failed**.

Union Bound. Suppose $\mathbf{X}_1, \dots, \mathbf{X}_n$ are (not necessarily independent) $\{0, 1\}$ valued random variables, then

$$\Pr[\mathbf{X}_1 + \dots + \mathbf{X}_n \geq 1] \leq \Pr[\mathbf{X}_1 = 1] + \Pr[\mathbf{X}_2 = 1] + \dots + \Pr[\mathbf{X}_n = 1].$$

Now we analyze our algorithm using union bound. An output $M = (\{\pi(1), \sigma(1)\}, \dots, \{\pi(n), \sigma(n)\})$ is a valid perfect matching exactly when all edges of the form $\{\pi(i), \sigma(i)\}$ are present in G . A “bad event” happens if any of those pairs are not edges in G .

Let \mathbf{X}_i be the indicator of the event that $\{\pi(i), \sigma(i)\}$ is *not* present in our graph.

1. What is the probability that $\mathbf{X}_i = 1$?
2. Use the union bound to upper bound the probability that M is *not* a valid perfect matching.
3. Conclude that G has a valid perfect matching.

The upper bound obtained on the probability of our bad event, i.e. of M not being a valid perfect matching, is fairly high. In light of this, we introduce the technique of *amplification*.

Amplification. The philosophy of amplification is that if we have a randomized algorithm that fails with probability p , we can repeat the algorithm many times and aggregate the output of all the runs to produce a new output such that the failure probability of the randomized algorithm is significantly smaller. Now consider the following algorithm **SpamBlindMatching**.

- Run **BlindMatching** independently T times.
 - If at least one of the runs outputted a valid perfect matching, return the output of such a run.
 - Else output **failed**.
1. What is the failure probability of **SpamBlindMatching**?
 2. How large should we set T if we want a failure probability of δ ?

Now we switch gears and turn our attention to concentration phenomena and its usefulness in analyzing randomized algorithms.

Markov's inequality. Let \mathbf{X} be a *nonnegative valued* random variable, then for every $t \geq 0$:

$$\Pr[\mathbf{X} \geq t\mathbf{E}[\mathbf{X}]] \leq \frac{1}{t}.$$

1. Markov's inequality is *false* for random variables that can take on negative values! Give an example.
2. Give a tight example for Markov's inequality. In particular, given μ and t , construct a random variable \mathbf{X} such that $\mu = \mathbf{E}[\mathbf{X}]$ and $\Pr[\mathbf{X} \geq t\mu] = \frac{1}{t}$.

Chebyshev's inequality. Let \mathbf{X} be any random variable with well-defined variance¹, then

$$\Pr\left[|\mathbf{X} - \mathbf{E}[\mathbf{X}]| > t\sqrt{\mathbf{Var}[\mathbf{X}]}\right] \leq \frac{1}{t^2}.$$

To see the above inequality in action, consider the following problem:

Let B be a bag with n balls, k of which are red and $n-k$ of which are blue. We do not have knowledge of k and wish to estimate k from ℓ independent samples (with replacement) drawn from B .

Let \mathbf{X} be the number of red balls sampled.

1. What is $\mathbf{E}[\mathbf{X}]$?
2. What is $\mathbf{Var}[\mathbf{X}]$?
3. Choose a value for ℓ and give an algorithm that takes in n and \mathbf{X} and outputs a number \tilde{k} such that $\tilde{k} \in [k - \varepsilon\sqrt{k}, k + \varepsilon\sqrt{k}]$ with probability at least $1 - \delta$.

¹In this course, all random variables will have well-defined variance