Philip Wellnitz, Barış Can Esmer **Winter 2022/23**

— Randomized Algorithms and Probabilistic Analysis of Algorithms, Exercise Sheet 3 —

https://www.mpi-inf.mpg.de/departments/algorithms-complexity/teaching/winter22/random

Total Points: 100 Due: Wednesday, **December 7**, 2022

You are allowed to collaborate on the exercise sheets. Justify your answers. Cite all external sources that you use (books, websites, research papers, etc.). You need to collect at least 50% of all points on exercise sheets to be admitted to the exam. Please hand in your solutions before the lecture on the day of the deadline.

Exercise 1 — **25** points

Consider the general balls-and-bins experiment, where m balls are thrown into n bins, where $m > n$. Show that the highest loaded bin contains $O(m/n + \ln n)$ balls with high probability.

Exercise 2 — **10 + 10 + 5 + 15 + 10** points

Recall the Coupon Collector's Problem, where we wish to obtain n different coupons by repeatedly and independently pulling for a random type of coupon (where each coupon appears with probability $1/n$). Write X_C for the number of coupons pulled before obtaining (at least) one coupon of every type.

In this exercise, we prove the following *sharp threshold* of X_C: for any constant $c \in \mathbb{R}$, we have

$$\lim_{n\to\infty} \Pr[\, X_C > n \ln n + cn \,] = 1 - e^{-e^{-c}}. \tag{1}$$

In a first step, we reinterpret the Coupon Collector's Problem as a balls-and-bins experiment. To that end, imagine that you have n bins and you throw balls into bins uniformly at random until every bin contains at least one ball.

a) Prove that after throwing $m \coloneqq n \ln n + cn$ balls, with probability at most e^{-c}, there is a bin without any balls. (*Hint: You may use results from the lecture without repeating their proofs.*)

b) Consider n independent Poisson random variables P_i with mean $(\ln n + c)$ each, with the idea that the i-th Poisson random variable approximates the number of balls in the i-th bin. Compute the probabilities that any fixed P_i is zero and prove that the probability that no P_i is zero is approximately $e^{-e^{-c}}$ (for large n). What can you conclude for the balls-and-bins experiment?

In the next step, we show that the Poisson approximation is indeed accurate. To that end, write $\mathcal{E}$ for the event that none of the Poisson random variables P_i is zero and write $X = \sum_{i=1}^{n} P_i$.

c) Argue that $\lim_{n\to\infty} \Pr[\, \mathcal{E} \,] = e^{-e^{-c}}$.

d) Write $\Pr[\, \mathcal{E} \,] = \Pr\Big[\, \mathcal{E} \wedge (|X - m| \le \sqrt{2m \ln m}) \,\Big] + \Pr\Big[\, \mathcal{E} \wedge (|X - m| > \sqrt{2m \ln m}) \,\Big]$ and argue that

$$\Pr\Big[\, |X - m| > \sqrt{2m \ln m} \,\Big] = o(1) \quad \text{and} \quad \Big|\Pr\Big[\, \mathcal{E} \mid |X - m| \le \sqrt{2m \ln m} \,\Big] - \Pr[\, \mathcal{E} \mid X = m \,]\Big| = o(1).$$

(*Hint: You may use the following Chernoff bounds for a Poisson random variable X with mean μ: for $a > \mu$, we have $\Pr[\, X \ge a \,] \le e^{-\mu}(e\mu)^a/a^a$; for $a < \mu$, we have $\Pr[\, X \le a \,] \le e^{-\mu}(e\mu)^a/a^a$.*)

e) Conclude that $\Pr[\, \mathcal{E} \,] = \Pr[\, \mathcal{E} \mid X = m \,]\,(1 - o(1)) + o(1)$.
Finally, argue that $\lim_{n\to\infty} \Pr[\, \mathcal{E} \,] = \lim_{n\to\infty} \Pr[\, \mathcal{E} \mid X = m \,]$ and derive (1).

Exercise 3 — **25** points

For a positive integer n and a constant $c \in \mathbb{R}$, write $N = 1/2 \cdot (n \ln n + cn)$. Prove that any graph G generated in $G_{n,N}$ (that is, starting with n isolated vertices, we pick a random non-edge N times) satisfies

$$\lim_{n\to\infty} \Pr[\, G \text{ has at least one isolated vertex} \,] = e^{-e^{-c}}.$$