ETH Zürich FS 2025
Institute of Theoretical Computer Science
Prof. Rasmus Kyng
Prof. Johannes Lengler
Prof. Vera Traub

Algorithmen und Wahrscheinlichkeit
Theoretical Exercise 3

SUBMIT BY MOODLE () UNTIL 16:00 ON 03.04.2025.

Exercise 1 – *Probability*

(a) Consider a medical diagnostic test for a rare disease. The false negative rate is very small: if a person has a disease, with probability 0.999 the test will turn out positive. Similarly, the false-positive rate is small: if a person does not have a disease, the probability that test is positive is 0.005. Assume morover that 1% of population has the disease in question. If a person chosen uniformly at random is tested and has positive test result, what is the probability that this person has the disease?

(b) Let us consider a variant of a coupon collector problem. There are $2n$ possible coupons, which are organized into n pairs — coupons 1 and 2 are a pair, similarly 3 and 4, and so on. In each round we get a single coupon independently and uniformly from $\{1, \dots 2n\}$, but to win the prize we need only to collect one coupon out of each pair. What is the expected number of rounds, until we win the prize (collect at least one coupon from each pair)?

(c) Generalize the problem above to the scenario where we have kn coupons, organized into groups of size k each — to win the prize we need only to collect one coupon from each group. What is the expected number of rounds in this game until we win the prize?

Solution 1

(a) Let A be the event "test result is positive" and B the event "a person has a disease". We have $\Pr[A|B] = 0.999$, $\Pr[A|\neg B] = 0.005$ and $\Pr[B] = 0.01$. By Bayes rule, we have

$$\Pr[B|A] = \frac{\Pr[A|B]\Pr[B]}{\Pr[A]} = \frac{\Pr[A|B]\Pr[B]}{\Pr[A|B]\Pr[B] + \Pr[A|\neg B]\Pr[\neg B]} \approx 0.67.$$

(b) Let us divide the process into n phases — phase i for $i \in \{0, \dots n-1\}$ consists of all rounds between the first round we have collected k unique pairs, up to the first round we have collected $i+1$ unique pairs. Let T_i for $i \in \{0, \dots n-1\}$ be a random variable denoting the number of rounds in phase i. We have $\Pr[T_i = j] = (\frac{2i}{2n})^{j-1}(1 - \frac{2i}{2n})$, so T_i has distribution $\text{Geo}(1 - \frac{i}{n})$. Hence $\mathbb{E}T_i = \frac{n}{n-i}$, and if T is the total number of rounds to collect n unique pairs, we have

$$\mathbb{E}[T] = \mathbb{E}\left[\sum_{0 \le i < n} T_i\right] = \sum_{0 \le i < n} \mathbb{E}[T_i] = \sum_{1 \le i \le n} \frac{n}{i} = nH_n = n \log n + O(n). \tag{1}$$

(c) The solution of the general variant is the same as above. Again T_i is the number of rounds in the i-th phase, where a phase consist of all rounds in between of the i-th and the $(i+1)$ group of coupons have been collected. The distribution of T_i is given as $\Pr[T_i = j] = (\frac{ki}{kn})^{j-1}(1 - \frac{ki}{kn}) = (\frac{i}{n})^{j-1}(1 - \frac{i}{n})$, hence T_i is distributed according to $\text{Geo}(1 - \frac{i}{n})$ — importantly, this distribution does not depend on k. The same calculation as (1) gives that

$$\mathbb{E}[T] = nH_n = n \log n + O(n).$$