Homework for “Sublinear-Time Algorithms”

Please select **four** of the following six questions and send solutions to `phdopen@onak.pl`.

1. **(Detection by Sampling)** Consider an urn with n balls, out of which k are red and $n-k$ are black. You repeat the following process $5 * n/k$ times: you draw a ball uniformly at random from the urn and place it back inside the urn (i.e., you sample with replacement). Prove that the probability of drawing a red ball at least once is at least 99/100.

   (Hint: What is the probability of never drawing a red ball?)

2. **(Birthday Paradox Upper Bound)** You select independently, uniformly at random $6\lceil\sqrt{n}\rceil$ numbers in $\{1, 2, \ldots, n\}$. Show that you will draw some number at least twice with probability at least 99/100.

   (Hint: What happens if you don't find a pair of identical elements with the first $\lceil\sqrt{n}\rceil$ samples?)

3. **(Birthday Paradox Lower Bound)** You select independently, uniformly at random at most $\sqrt{n}/10$ numbers in $\{1, 2, \ldots, n\}$. Prove that all selected numbers are distinct with probability at least 99/100.

   (Hint: What is the expected number of pairs of elements that are identical? Can you apply Markov's inequality?)

4. **(Estimating Distance between Distributions)** Consider a model in which algorithms have access to two probability distributions $\mathcal{P} = (p_1, \ldots, p_n)$ and $\mathcal{Q} = (q_1, \ldots, q_n)$ on $\{1, \ldots, n\}$. They can:

   - draw independent samples from $\mathcal{P}$ and $\mathcal{Q}$, and
   - they can query the probability p_i or q_i, $1 \le i \le n$, of an arbitrary element.

   Show that the complexity of estimating $\|\mathcal{P} - \mathcal{Q}\|_1$ up to an additive ε is $O(1/\varepsilon^2)$.

   (Hint: Use the following algorithm:

   - Draw k independent samples a_1, a_2, …, a_k from $\mathcal{P}$.
   - Output $\frac{1}{k}\sum_{i=1}^{k} \frac{|p_{a_i} - q_{a_i}|}{p_{a_i}}$.

   and the Hoeffding Bound.)

5. **(Approximating the Diameter of a Metric Space)** Consider a model in which algorithms can query the distance between an arbitrary pair of points of a metric space $\mathcal{M}$ on n points. Show that approximating the diameter of $\mathcal{M}$ up to a factor less than 2 requires $\Omega(n^2)$ queries.

6. **(Auxiliary Lemma for Disjointness Testing in the External Memory Model)** Consider a sequence of n elements partitioned into blocks of size B. Suppose that it contains εn disjoint pairs of identical elements such that the elements of each pair lie in different blocks. Show that there are $\Omega(\varepsilon n/B)$ disjoint pairs of blocks such that the blocks in each pair share a common element.