Themistoklis Gouleakis Winter 2021

Property Testing, Exercise Sheet 5

https://www.mpi-inf.mpg.de/departments/algorithms-complexity/teaching/winter20/property-testing

Total Points: 0 + 20 Due: Wednesday, **February 3**, 2021

You are allowed to collaborate on the exercise sheets, but you have to write down a solution on your own, ***using your own words****. Please indicate the names of your collaborators for each exercise you solve. Further, 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 send your solutions directly to Philip (`wellnitz@mpi-inf.mpg.de`).*

Exercise 1 — **0** points + 3 + 3 + 3 + 1 bonus points

Let $p : [n] \to [0,1]$ denote a probability function. For a given number $m = O(n/\varepsilon^2)$ let $i_1, \ldots, i_m \in [n]$ denote samples that are drawn according to p. Then the empirical distribution $\tilde{p}$ is defined by $\tilde{p}(i) := |\{j \in [m] : i_j = i\}|/m$, that is, $\tilde{p}$ represents the relative frequency of each of the values $i \in [n]$ in the sample $i_1, \ldots, i_m$.

Using the following steps, prove that, with high probability, $\tilde{p}$ is ε-close to p.

a) For every $i \in [n]$, let X_i denote the distribution of the fraction of the number of occurrences of i in the sample. Then, $\mathbb{E}[\, X_i \,] = p(i)$ and $\mathsf{Var}[\, X_i \,] \leq p(i)/m$.

b) Show that $\mathbb{E}[\, |X_i - p(i)| \,] \leq \mathsf{Var}[\, X_i \,]^{1/2}$.

c) Show that $\mathbb{E}\Big[\sum_{i \in [n]} |X_i - p(i)| \Big] \leq \sqrt{n/m}\,$.

d) Conclude that for a suitable choice of m, we have $\mathsf{Pr}\Big[\sum_{i \in [n]} |X_i - p(i)| > \varepsilon \Big] < 1/3$.

Exercise 2 — **0** points + 10 bonus points

Let $F : [n] \to S$ denote a randomized process such that the supports of the different $F(i)$'s are disjoint. Prove that for every two distributions X and X' over $[n]$, the total variation distance between $F(X)$ and $F(X')$ equals the total variation distance between X and X'.

Note that distances may not be preserved if the supports of some $F(i)$'s are not disjoint, and that the level of preservation is related to the relation between the distributions of the various $F(i)$'s.