University of Edinburgh
INFR11156: Algorithmic Foundations of Data Science (2019)
Homework 5

Problem 1: Let $H = \{h : [m] \to \{0,1\}^n\}$ be a family of pairwise independent hash functions. Let $I \subseteq [m]$ and $\mu := \frac{|I|}{2^n}$. Then, it holds for every $y \in \{0,1\}^n$ that

$$\mathbb{P}_{h\sim H}\Big[\big||\{i \in I : h(i) = y\}| - \mu\big| > \varepsilon\mu\Big] < \frac{1}{\varepsilon^2\mu},$$

where $h \sim H$ stands for the fact that h is chosen uniformly at random from H.

Problem 2: Let $Y_1, \ldots, Y_n$ be independent random variables with $\mathbb{P}[Y_i = 0] = \mathbb{P}[Y_i = 1] = 1/2$. Let $Y := \sum_{i=1}^{n} Y_i$ and $\mu := \mathbb{E}[Y] = n/2$. Apply the uniform Chernoff Bound to prove it holds for any $0 < \lambda < \mu$ that

$$\mathbb{P}[Y \geq \mu + \lambda] \leq \mathrm{e}^{-2\lambda^2/n}.$$

Problem 3: Prove that the median of the returned values from $\Theta(\log(1/\delta))$ independent copies of the BJKST algorithm gives an (ε, δ)-approximation of F_0.