max planck institut informatik

SAARLAND UNIVERSITY

COMPUTER SCIENCE

Antonios Antoniadis and Marvin Künnemann **Winter 2018/19**

Exercises for Randomized and Approximation Algorithms

www.mpi-inf.mpg.de/departments/algorithms-complexity/teaching/winter18/rand-apx-algo/

Exercise Sheet 5: Concentration I

To be handed in by **November 20th, 2018** via e-mail to André Nusser (CC to Antonios Antoniadis and Marvin Künnemann)

Exercise 1 *(5 Points)* Consider a fair die showing the numbers $\{1, \ldots, D\}$. Let X be the sum of the numbers obtained after rolling it N times. Use Chebychev's inequality to give an upper bound on

$$\Pr[|X - \mathbf{E}[X]| \geq \alpha\, \mathbf{E}[X]],$$

for any $\alpha > 0$.

Exercise 2 *(10 Points)* Let x, y be length-n strings. We define their *Hamming distance* as $\mathrm{Ham}(x, y) := \#\{1 \leq i \leq n \mid x[i] \neq y[i]\}$, i.e., the number of positions where x and y disagree.

Consider the following algorithm approximating $\mathrm{Ham}(x, y)$ by means of "alphabet reduction": (here, for any function $h : \Sigma \to \mathbb{N}$ and string $x = x[1] \ldots x[n]$, we write $h(x) = h(x[1]) \ldots h(x[n])$.)

function ApproxHam(x, y, ε)
  for $i = 1, \ldots, \lceil c \log n \rceil$ **do**
    pick h u.a.r. from the set of all functions $\Sigma \to \{1, \ldots, \lceil 2/\varepsilon \rceil\}$
    $d_i \leftarrow \mathrm{Ham}(h(x), h(y))$
  return $\max_{1 \leq i \leq \lceil c \log n \rceil} d_i$

Show that this algorithm computes an estimate $\tilde{d}$ satisfying $(1 - \varepsilon)\mathrm{Ham}(x, y) \leq \tilde{d} \leq \mathrm{Ham}(x, y)$ with probability at least $1 - n^{-c}$.
(Hint: Use Markov!)

Exercise 3 *(12 Points)* We say that a hash family $\mathcal{H}$ from X to Y is k-universal (in the strong sense) if for all pairwise distinct $x_1, \ldots, x_k \in X$ and all $y_1, \ldots, y_k \in Y$, we have

$$\Pr_{h \leftarrow \mathcal{H}}[h(x_1) = y_1 \text{ and } \cdots \text{ and } h(x_k) = y_k] = \frac{1}{|Y|^k}.$$

Let p be a prime number and recall that computation modulo p yields a field (which we write as $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$). Define the hash family $\mathcal{H}_{\text{simple}}$ from $\mathbb{F}_p$ to $\mathbb{F}_p$ as the set of functions $h_{a,b}$ with $h_{a,b}(x) = ax + b \pmod{p}$ for $a, b \in \mathbb{F}_p$.

a) *(7 Points)* Prove that $\mathcal{H}_{\text{simple}}$ is 2-universal and that any $h_{a,b} \in \mathcal{H}_{\text{simple}}$ can be stored using $O(\log p)$ bits.

b) *(2 Points)* Show that $\mathcal{H}_{\text{simple}}$ is in general not 3-universal.

c) *(3 Points)* The construction of $\mathcal{H}_{\text{simple}}$ does not (immediately) yield a 2-universal hash family from $[n]$ to $[n]$ for arbitrary (non-prime) n. Why can we still make the algorithm for estimating the number of distinct elements in a stream (given in the lecture) work?

Exercise 4 *(13 Points)* Let X be a (discrete) random variable and recall that $\sigma[X] = \sqrt{\mathrm{Var}[x]}$ denotes its standard deviation.

a) *(10 Points)* Prove the following inequality: For any $t > 0$ we have

$$\Pr\left[X - \mu \geq t\sigma[X]\right] \leq \frac{1}{1+t^2}.$$

(Hint: Note that $X - \mu \geq \alpha$ if and only if $X - \mu + u \geq \alpha + u$. Optimize over u!)

b) *(3 Points)* Prove the following two-sided variant of the above inequality: For any $t > 0$, we have

$$\Pr\left[|X - \mu| \geq t\sigma[X]\right] \leq \frac{2}{1+t^2}.$$

In which situations does this provide a better bound than Chebychev's inequality?
(Note: You may make use of a) even if you did not prove it.)