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 7: Concentration II

To be handed in by **December 4th, 2018** via e-mail to André Nusser
(CC to Antonios Antoniadis and Marvin Künnemann)

Total points : 40 regular + 5 bonus points

Exercise 1 *(5 Points)* In Lecture 5, we have shown a randomized streaming algorithm for estimating the number of distinct elements d in the input. In particular, using $O(\log n)$ space it returned an estimate $\tilde{d}$ such that

$$\Pr[d \geq 3\tilde{d}] \leq \alpha,$$
$$\Pr[d \leq \tilde{d}/3] \leq \alpha,$$

where $\alpha = \sqrt{2}/3$. Prove that for any $0 < \delta < 1$, there is a streaming algorithm that uses $O(\log(n)\log(1/\delta))$ space and returns an estimate $\tilde{d}$ with

$$\Pr[d \notin [\tilde{d}/3, 3\tilde{d}]] \leq \delta.$$

Exercise 2 *(10 Points)* Often, one only has upper or lower bounds on the expected value of an interesting random variable X. In these cases, the following variant of the Chernoff bound can be useful.

Let $X_1, \dots, X_n$ be independent random variables with values in $\{0,1\}$. Let $X := \sum_{i=1}^{n} X_i$ and denote by $\mu := \mathbf{E}[X]$ its expectation. For any upper bound $\mu_H \geq \mu$ on its expectation, we have

$$\Pr[X \geq (1+\delta)\mu_H] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu_H}.$$

Alternative: For partial credit, show that we have $\Pr[X \geq (1+\delta)\mu_H] \leq e^{-\frac{\delta^2}{3}\mu_H}$.

Exercise 3 *(10 Points)* Let X be the sum of n independent random variables with values in $\{0,1\}$ and $\mathbf{E}[X] = n \cdot p(n)$ for some function $p(n)$ possibly depending on n and let $c \geq 1$.

The aim of this exercise is to determine the smallest interval I such that $X \in I$ with probability $1 - O(n^{-c})$.

Note: try to estimate the size $|I|$ with the leading constant, i.e., write a size such as $|I| = 3n^3 + 2n^2 - n$ as $|I| = 3n^3 + o(n^3)$.

a) *(5 Points)* Use a Chernoff bound to give an upper bound on $|I|$ if $p(n)$ is a fixed constant $0 < p < 1$ (independent of n).

b) *(5 Points)* Use a Chernoff bound to give an upper bound on $|I|$ if $p(n) = 1/n$.

Exercise 4 *(20 Points)* Consider flipping a fair coin until it comes up heads, and let D denote the resulting distribution of the number of coin flips. Consider the independent random variables $C_1, \ldots, C_n$, where each C_i is distributed as D, and observe that $C := \sum_{i=1}^{n} C_i$ has expectation $\mu := 2n$.

a) *(10 Points)* Give an upper bound on $\Pr[C \geq (1+\delta)\mu]$ using a Chernoff bound to a sequence of fair coin tosses.

b) *(8 Points)* Give an upper bound on $\Pr[C \geq (1+\delta)\mu]$ by adapting the proof of the Chernoff bound in class:

(i) *(2 Points)* Determine the moment-generating function of D, i.e., show that for any $t < \ln 2$, we have

$$\mathbf{E}[e^{tC_i}] = \frac{1}{2}\left(\frac{e^t}{1 - e^t/2}\right),$$

for any i.

(ii) *(2 Points)* Show that this yields

$$\mathbf{E}[e^{tC}] \leq \prod_{i=1}^{n} \frac{1}{1-2t},$$

for $t < 1/2$.

(iii) *(2 Points)* Use the above inequality to show that

$$\Pr[C \geq (1+\delta)\mu] \leq e^{-n\ln(1-2t)-(1+\delta)t\mu},$$

for all $0 < t < 1/2$.

(iv) *(2 Points)* Optimize over t to obtain a good bound on $\Pr[C \geq (1+\delta)\mu]$.

c) *(2 Points)* Compare the two bounds you have obtained in a) and b). Which one is stronger?