Sublinear Algorithms for Big Datasets

Exam Problems

Cristian Rosa　　Jose Luis Diaz

August 20, 2014

Problem 1: Modified Chernoff Bound

Let $X_1, \ldots, X_t$ be independent and identically distributed r.vs with range $[0, c]$ and expectation μ. Then if $X = \frac{1}{t}\sum_i X_i$ and $1 > \delta > 0$,

$$Pr[|X - \mu| \geq \delta\mu] \leq 2e^{\left(-\frac{t\mu\delta^2}{3c}\right)}$$

Proof.

$$0 \leq X_i \leq c, i \in 1 \ldots t \Rightarrow 0 \leq \frac{X_i}{c} \leq 1$$

Let $Z = \frac{X}{c}$, then by Chernoff Bound 1

$$Pr[|Z - \mathbb{E}[Z]| \geq \delta\mathbb{E}[Z]] \leq 2e^{\left(-\frac{t\mathbb{E}[Z]\delta^2}{3}\right)}$$

definition of Z

$$Pr[|\tfrac{X}{c} - \mathbb{E}[\tfrac{X}{c}]| \geq \delta\mathbb{E}[\tfrac{X}{c}]] \leq 2e^{\left(-\frac{t\mathbb{E}[\frac{X}{c}]\delta^2}{3}\right)}$$

linearity of expectation

$$Pr[|\tfrac{X}{c} - \tfrac{1}{c}\mathbb{E}[X]| \geq \delta\tfrac{1}{c}\mathbb{E}[X]] \leq 2e^{\left(-\frac{t\frac{1}{c}\mathbb{E}[X]\delta^2}{3}\right)}$$

simplification and definition of $\mu = \mathbb{E}[X]$

$$Pr[|X - \mu| \geq \delta\mu] \leq 2e^{\left(-\frac{t\mu\delta^2}{3c}\right)}$$

□

Problem 2: Modified Chebychev

Let X be a random variable with finite expectation $\mathbb{E}[X]$. For every $c' > 0$,

$$Pr[|X - \mathbb{E}[X]| \geq c'\mathbb{E}[X]] \leq \frac{Var[X]}{(c'\mathbb{E}[X])^2}$$

Proof. Let X be a random variable with finite expectation $\mathbb{E}[X]$ and $c > 0$, then by Chebychev bound

$$Pr[|X - \mathbb{E}[X]| \geq c\sqrt{Var[X]}] \leq \frac{1}{c^2}$$

Let $c = \dfrac{c'\mathbb{E}[X]}{\sqrt{Var[X]}}$ with $c' > 0$, then

$$Pr[|X - \mathbb{E}[X]| \geq \frac{c'\mathbb{E}[X]}{\sqrt{Var[X]}}\sqrt{Var[X]}] \leq \frac{1}{\left(\frac{c'\mathbb{E}[X]}{\sqrt{Var[X]}}\right)^2}$$

Simplification

$$Pr[|X - \mathbb{E}[X]| \geq c'\mathbb{E}[X]] \leq \frac{Var[X]}{(c'\mathbb{E}[X])^2}$$

□