Lecture 8: “Chernoff Bounds”

Sotiris Nikoletseas
Professor

CEID - ETY Course
2017 - 2018

Overview

A1. Chernoff Bounds

A2. A Randomized Algorithm for Dominating Sets

A1. The Chernoffs Bounds

- Tail Bounds on the probability of deviation from the expected value of a random variable.

- Focus on
  sums of independent, indicator (Bernoulli) random variables

- Such sums abstract quantitative results of random choices in randomized algorithms (e.g. the number of comparisons in the randomized quicksort algorithm, where for each two elements i, j the r.v. X_{ij} is 1 if $S_{(i)}, S_{(j)}$ are compared (and 0 otherwise).

- The Chernoff bounds are exponential i.e. much sharper than the Markov, Chebyshev bounds.

Poisson and Bernoulli trials

- Poisson trials:
  - repeated, independent trials
  - two possible outcomes in each trial ("success", "failure")
  - potentially different success probability p_i in each trial i.
  - we take the sum of the corresponding indicator variables X_i for each trial i (it measures the total number of successes).

i.e. for $1 \leq i \leq n$:

$$X_i = \begin{cases} 1 & \text{(success), with probability } p_i \\ 0 & \text{(failure), with probability } q_i = 1 - p_i \end{cases}$$

$$X = X_1 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

- Bernoulli trials: Poisson trials when $p_i = p, \forall i$. (in that case $X \sim B(n,p)$ i.e. it follows the binomial distribution)

The tail probability

- Clearly, the expected value (or mean) of X is:

$$\mu = E(X) = \sum_{i=1}^{n} p_i$$

- Two important questions:

  (1) For a real number $\beta > 0$ what is the probability that $\underline{X \text{ exceeds } (1+\beta)\mu}$?
  (i.e. we seek a bound on the tail probability)
  $\rightarrow$ this is useful in the analysis of randomized algorithms (i.e., to show that the probability of failure to achieve a certain expected performance is small).

  (2) How large must β be so that the tail probability is less than a certain value ϵ ? (this is relevant in the design of the algorithm).

The Chernoff bound (for exceeding the mean) (I)

Theorem 1. Let X_i be n independent Poisson(p_i) trials,

$$X = \sum_{i=1}^{n} X_i, \quad \mu = E(X) = \sum_{i=1}^{n} p_i.$$

Then, for any $\beta > 0$, it is :

$$\Pr\{X > (1+\beta)\mu\} < \left[\frac{e^{\beta}}{(1+\beta)^{(1+\beta)}}\right]^{\mu}$$

Proof: For any positive t:

$$\Pr\{X > (1+\beta)\mu\} = \Pr\{e^{tX} > e^{t(1+\beta)\mu}\}$$

By the Markov inequality: $\Pr\{e^{tX} \geq e^{t(1+\beta)\mu}\} < \frac{E[e^{tX}]}{e^{t(1+\beta)\mu}}$ (1)

But $E[e^{tX}] = E\left[e^{t(X_1+\cdots+X_n)}\right] = E\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} E[e^{tX_i}]$ (2)

Now $E[e^{tX_i}] = e^t p_i + 1 \cdot (1-p_i) = 1 + p_i(e^t - 1) \leq e^{p_i(e^t-1)}$ (3)

The Chernoff bound (for exceeding the mean) (II)

$$\textcircled{2} \wedge \textcircled{3} \Rightarrow E[e^{tX}] = \prod_{i=1}^{n} e^{p_i(e^t-1)} = e^{\sum_{i=1}^{n} p_i(e^t-1)} = e^{(e^t-1)\mu}$$

$$\textcircled{1} \Rightarrow \Pr\{e^{tX} \geq e^{t(1+\beta)\mu}\} < \frac{e^{(e^t-1)\mu}}{e^{t(1+\beta)\mu}}$$

The right part is minimized when $t = \ln(1+\beta)$
(<u>note that $t > 0 \Leftrightarrow \beta > 0$</u>).

$$t = \ln(1+\beta) \Rightarrow \Pr\{X > (1+\beta)\mu\} < \frac{e^{(e^{\ln(1+\beta)}-1)\mu}}{e^{\ln(1+\beta)(1+\beta)\mu}} =$$
$$= \frac{e^{\beta\mu}}{(1+\beta)^{(1+\beta)\mu}} = \left[\frac{e^{\beta}}{(1+\beta)^{(1+\beta)}}\right]^{\mu}$$

The various Chernoff bounds (I)

Theorem 2. (Chernoff bound for exceeding the mean)

For X_i, X, μ as in Theorem 1 it is:

$$\forall \beta \in [0,1] : \Pr\{X \geq (1+\beta)\mu\} \leq e^{-\frac{\beta^2 \mu}{3}}$$

Proof: $\forall \beta \in [0,1] : \left(\frac{e^{\beta}}{(1+\beta)^{(1+\beta)}}\right)^{\mu} \leq e^{-\frac{\beta^2 \mu}{3}}$

Theorem 3. (Chernoff bound for the left tail)

For X_i, X, μ as in Theorem 1 it is:

$$\forall \beta \in [0,1] : \Pr\{X \leq (1-\beta)\mu\} \leq e^{-\frac{\beta^2 \mu}{2}}$$

Proof: See the MR book.

The various Chernoff bounds (II)

Theorem 4. (The combined Chernoff bound)

For X_i, X, μ as in Theorem 1 it is:

$$\forall\beta \in [0,1] : \Pr\{X \in (1 \pm \beta)\mu\} \geq 1 - 2e^{-\frac{\beta^2\mu}{3}}$$

Proof: Follows easily from Theorems 2, 3.

- Important remark:
  - $\mu \to \infty$ (at an arbitrarily slow rate) $\Rightarrow$
    $\Pr\{X \in (1 \pm \beta)\mu\} \to 1$
  - $\mu = \Omega(\log n) \Rightarrow \Pr\{X \in (1 \pm \beta)\mu\} \geq 1 - 2n^{-\gamma}$, where $\gamma > 1$
    i.e. a logarithmic mean guarantees a fast convergence of the concentration probability to 1.

A2. A Randomized Algorithm for Dominating Sets

- The dominating set problem: Find a subset of the vertices of a graph such that every vertex not in this set is adjacent to at least one vertex in it. Formally:
  $V' \leq V(G)$ dominating set in $G(V, E)$ iff
  $\forall u \notin V', \exists v \in V' : (u, v) \in E(G)$
- The problem is important and well motivated from real networks, since the dominating set plays a "central" role in the graph.
- Obviously we want to find a dominating set which is as small as possible.

The complexity of the problem / using randomness

- Finding a minimum dominating set is an NP-hard problem.

- Randomness can be used to "attack" the problem:
  - It has been shown that in $G_{n,p}(p = \frac{1}{2})$ random graphs, $\nexists$ dominating set of size $< \log n$ w.h.p. (technique: Linearity of expectation + Markov inequality)
  - Also, w.h.p. $\exists$ dominating set of size $\lceil \log n \rceil$ (technique: the second moment method)
  - We will here present a randomized algorithm that w.h.p. finds a d.s. of size $(1 + \epsilon) \log n$ ($\epsilon > 0$ arbitrarily small), in polynomial time (thus, this algorithm is near-optimal).

The GREEDY-DS algorithm - basic idea

- choose a random vertex and put it in a dominating set under construction
- remove from graph all vertices adjacent to that vertex (they are "covered" by the vertex)
- repeat until the number of vertices left becomes small
- explicitly add those vertices to the dominating set under construction (obviously, the resulting set is a dominating set)

The algorithm - pseudo code

ALGORITHM "GREEDY-DS"

Input: a random instance $G(V,E)$ of $G_{n,p}$ $(p = \frac{1}{2})$

(1) $i \leftarrow 0; V_i \leftarrow V; D \leftarrow \varnothing$

(2) **until** $|V_i| \leq \epsilon \log n$ **do**

  begin

    choose a random vertex $u_i \in V_i$

    $V' \leftarrow V_i - N_i$ (N_i: neighor vertices of u_i)

    $D \leftarrow D \cup \{u_i\}$ (add the vertex to the evolving d.s.)

    $i \leftarrow i + 1$

    $V_i \leftarrow V'$

  end

(3) $D \leftarrow D \cup V_i$ (add the vertices left)

(4) **return** D

Analysis - informal

- in each repetition about half (because of $p = \frac{1}{2}$) of the vertices left are "covered" by the random vertex and are removed from the graph
- thus, after $\log n$ repetitions the entire graph is almost covered and the number of vertices left drops to less than $\epsilon \log n$
- Then, these $\epsilon \log n$ vertices are explicity added to the dominating set under construction
- The resulting set has size $\log n + \epsilon \log n = (1 + \epsilon) \log n$ and is obviously a dominating set.

Two comments

(1)

- when we choose a vertex, we "expose" its neighbours, so they can not be assumed "random" anymore (e.g. we know exactly their number, which is not a random variable anymore)
- however these exposed vertices are anyway removed from the graph, so the rest graph (the vertices left) remains random and the randomized analysis remains valid.

(2)

- the Chernoff bound is used to show concentration of the number of "covered" vertices in each repetition around the expected number (which is easy to compute)
- as said, the bound is polynomially fast approaching 1 as long as the mean (of the vertices left) is logarithmic! This is why when this number gets logarithmic, the vertices left are explicitly added to the dominating set.

Analysis - Chernoff bound

Lemma 1 Let $\beta \in [0, 1]$. Then for any constant $\epsilon > 0$ and given that $|V_i| \geq \epsilon \log n$ it is:

$$\Pr\left\{|N_i| \geq (1-\beta)\frac{|V_i|}{2}\right\} \geq 1 - n^{-\frac{\beta^2}{4}\epsilon}$$

Proof: The vertices "covered" at repetition i (set N_i) follow a Binomial distribution with parameters $|V_i|, \frac{1}{2}$. In other words, we have $|V_i|$ Bernoulli trials each one with success probability $\frac{1}{2}$. Thus the mean μ_i of their sum ($|N_i|$) is bounded by the Chernoff bound as follows:

$$\Pr\{|N_i| \geq (1-\beta)\mu_i\} \geq 1 - e^{-\frac{\beta^2}{2}\mu_i} = 1 - e^{-\frac{\beta^2}{2}\frac{|V_i|}{2}} \geq$$

$$\geq 1 - e^{-\frac{\beta^2}{2}\frac{\epsilon \log n}{2}} = 1 - n^{-\frac{\beta^2}{4}\epsilon} \qquad \square$$

Analysis - time complexity

- Let the event $\mathcal{E}_i$= "at the i-th random repetition it is $|N_i| \geq (1-\beta)\frac{|V_i|}{2}$"
- Let the event $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \cdots \cap \mathcal{E}_t$ until $|V_t| < \epsilon \log n$
- Lemma 2: Assuming (probabilistically) event $\mathcal{E}$ the number t of repetitions of the random loop of the algorithm is at most $(1+\epsilon')\log n$, where $\epsilon' > 0$ constant.

  Proof: After repetition i:
  $|V_{i+1}| = |V_i| - |N_i| \leq |V_i| - \frac{(1-\beta)}{2}|V_i| = (1-\gamma)|V_i|$ where $\gamma = \frac{1}{1-\beta}$.
  Recursively: $|V_t| \leq (1-\gamma)^t n$, so for $|V_t| \leq \epsilon \log n$ we need $t \geq \frac{\log n}{\log(\frac{1}{1-\gamma})} + \Theta(\log\log n)$. We note that $\frac{1}{1-\gamma} = \frac{2}{1+\beta}$ so by choosing (for any $\epsilon' > 0$) $\beta = 2^{\frac{\epsilon'}{1+\epsilon'}} - 1$ we finally get $t \leq (1+\epsilon')\log n$ $\square$

Analysis - continued

- Lemma 3: For any constants $\beta \in [0,1], \epsilon > 0$ it is:
$$\Pr\{\mathcal{E}\} \geq 1 - n^{-\frac{\beta^2}{8}\epsilon}$$

  Proof: From previous lemmata, it is
  $\Pr\{\overline{\mathcal{E}}\} \leq \sum_i \Pr\{\overline{\mathcal{E}_j}\} \leq t \cdot n^{-\frac{\beta^2}{4}\epsilon} \leq (1+\epsilon')\log n \cdot n^{-\frac{\beta^2}{4}\epsilon} \leq n^{-\frac{\beta^2}{8}\epsilon}$

- Time complexity: Clearly, the algorithm needs time $O((1+\epsilon')n\log n)$ with probability at least $1 - n^{-\frac{\beta^2}{8}\epsilon}$.

- The constructed dominating set has size at least $(1+\epsilon'+\epsilon)\log n$

Summary (I)

- The algorithm constructs a near optimal dominating set (i.e. the approximation ratio to the optimal $\log n$ size is $1+\epsilon$, where $\epsilon > 0$ arbitrarily small).
- The time complexity is polynomial (both in expectation and w.h.p.)
- However the probability of good performance, although tending to 1, is not "polynomially large" since in the $1 - n^{-\frac{\beta^2}{8}\epsilon}$ bound the $\frac{\beta^2}{8}\epsilon$ constant is just positive, (not necessarily larger than 1)

Summary (II)

- In the full paper (S. Nikoletseas and P. Spirakis, "Near-Optimal Dominating Sets in Dense Random Graphs in Polynomial Expected Time", in the Proceedings of the 19th International Workshop on Graph-Theoretic Concepts in Computer Science (WG)) includes an enhanced algorithm (of repetitive trials) boosting this probability to $1 - n^{-a}$, where $a > 1$.