Randomized Algorithms

Kousha Etessami

Recap: Chernoff Bounds (upper tail)

Poisson trials - sequence of Bernoulli variables X_i with varying p_is.

Theorem (4.4)

Let $X_1, \ldots, X_n$ be independent 0/1 Poisson trials such that $\Pr[X_i = 1] = p_i$ for all $i \in [n]$. Let $X = \sum_{i=1}^{n} X_i$, and $\mu = E[X]$. We have the following Chernoff bounds*:*

1. *For any $\delta > 0$,*

$$\Pr[X \geqslant (1+\delta)\mu] \leqslant \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu};$$

2. *For any $0 < \delta \leqslant 1$,*

$$\Pr[X \geqslant (1+\delta)\mu] \leqslant e^{-\mu\delta^2/3};$$

3. *For $R \geqslant 6\mu$,*

$$\Pr[X \geqslant R] \leqslant 2^{-R}.$$

Recap: Chernoff Bounds (lower tail)

Theorem (4.5)

Let $X_1, \ldots, X_n$ *be independent 0/1 Poisson trials such that* $\Pr[X_i = 1] = p_i$ *for all* $i \in [n]$*. Let* $X = \sum_{i=1}^{n} X_i$*, and* $\mu = E[X]$*. For any* $0 < \delta < 1$*, we have the following* Chernoff bounds*:*

1. $$\Pr[X \leqslant (1-\delta)\mu] \leqslant \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu};$$

2. $$\Pr[X \leqslant (1-\delta)\mu] \leqslant e^{-\mu\delta^2/2};$$

- Proof is similar to Thm 4.4.
- Bound of (2.) is slightly better than the bound for $\geqslant (1+\delta)\mu$.

Recap: Concentration

Corollary (4.6)

Let $X_1, \ldots, X_n$ be independent 0/1 Poisson trials such that $\Pr[X_i = 1] = p_i$ for all $i \in [n]$. Let $X = \sum_{i=1}^n X_i$, and $\mu = E[X] = \sum_{i=1}^n p_i$. Then for any $\delta, 0 < \delta < 1$,

$$\Pr[|X - \mu| \geqslant \delta\mu] \ \leqslant \ 2e^{-\mu\delta^2/3}.$$

- For almost all applications, we will want to work with such a *symmetric* version like the Corollary.
- We "threw away" a bit in moving from the $\left(\frac{e^{\pm\delta}}{(1\pm\delta)^{1\pm\delta}}\right)^{\mu}$ versions, but they are tricky to work with.

Recap: Unbiased $+1/-1$ variables

For unbiased variables, we can do better than $2e^{-\mu\delta^2/3}$ by switching to +1/-1 variables.

Theorem (4.7)

Let $X_1, \ldots, X_n$ be independent random variables with $\Pr[X_i = 1] = 1/2 = \Pr[X_i = -1]$ for all $i \in [n]$. Let $X = \sum_{k=1}^{n} X_k$. Note $\mu = E[X] = 0$. Then for any $a > 0$,

$$\Pr[X \geqslant a] \leqslant e^{-a^2/2n}.$$

Recap: Unbiased 0/1 variables

Consider $Y_1, \ldots, Y_n$ such that $\Pr[Y_i = 0] = \Pr[Y_i = 1] = 1/2$ for all $i \in [n]$. Define $X_i = 2Y_i - 1$ for every $i \in [n]$. Then

$$X_i = \begin{cases} 1 & \text{if } Y_i = 1 \\ -1 & \text{if } Y_i = 0 \end{cases}$$

Corollary (4.9, 4.10)

For $Y = \sum_{i=1}^{n} Y_i$, $X = \sum_{i=1}^{n} X_i$, we have

$$\Pr[Y \geqslant \tfrac{n}{2} + a] = \Pr[X \geqslant 2a] \quad \leqslant \quad e^{-2a^2/n};$$

$$\Pr[Y \leqslant \tfrac{n}{2} - a] = \Pr[X \leqslant -2a] \quad \leqslant \quad e^{-2a^2/n}.$$

i.i.d. Bernoulli variables

For independent identically distributed (i.i.d.) Bernoulli variables X_i with a fixed constant parameter p, Chernoff bounds on their sum $X = \sum_{i=1}^{n} X_i$ yield that, roughly speaking, X has deviation from expectation

- $\Omega(\sqrt{n})$ with probability $O(1)$;
- $\Omega(\sqrt{n \ln n})$ with probability $O(n^{-c})$;
- $\Omega(n)$ with probability $e^{-\Omega(n)}$.

Application: set balancing and "discrepency" minimization

We have an $n \times m$ binary matrix A (entries from $\{0, 1\}$). We consider the value of

$$A \cdot \bar{b} = \bar{c},$$

when $\bar{b} \in \{-1, +1\}^m$ (note $\bar{c}$ will then be n-dimensional).

Goal is to find $\bar{b} \in \{-1, +1\}^m$ such that the value of $\|A \cdot \bar{b}\|_\infty = \max_{j=1}^{n} |c_j|$ is minimized.

Application: set balancing and “discrepency” minimization

We have an $n \times m$ binary matrix A (entries from $\{0, 1\}$). We consider the value of

$$A \cdot \bar{b} = \bar{c},$$

when $\bar{b} \in \{-1, +1\}^m$ (note $\bar{c}$ will then be n-dimensional).

Goal is to find $\bar{b} \in \{-1, +1\}^m$ such that the value of $\|A \cdot \bar{b}\|_\infty = \max_{j=1}^{n} |c_j|$ is minimized.

Exact optimization is NP-hard.

Application: set balancing and "discrepency" minimization

We have an $n \times m$ binary matrix A (entries from $\{0, 1\}$). We consider the value of

$$A \cdot \bar{b} = \bar{c},$$

when $\bar{b} \in \{-1, +1\}^m$ (note $\bar{c}$ will then be n-dimensional).

Goal is to find $\bar{b} \in \{-1, +1\}^m$ such that the value of $\|A \cdot \bar{b}\|_\infty = \max_{j=1}^{n} |c_j|$ is minimized.

Exact optimization is NP-hard.

Randomly choosing b is already pretty good: choose $\bar{b} \in \{-1, +1\}^m$ u.a.r. by generating b_i independently and uniformly from $\{-1, +1\}$. We can show

Theorem (4.11)

For $\bar{b}$ chosen u.a.r. from $\{-1, +1\}^m$,

$$\Pr[\|A\bar{b}\|_\infty \geqslant \sqrt{4m\ln(n)}] \leqslant \frac{2}{n}.$$

Set balancing: proof

- $\|\cdot\|_\infty$ is the absolute value of the largest entry of the tuple. We want to show that with high probability, *every entry* of $A \cdot \bar{b}$ has absolute value $\leqslant \sqrt{4m\ln(n)}$.

Set balancing: proof

- $\|\cdot\|_\infty$ is the absolute value of the largest entry of the tuple. We want to show that with high probability, *every entry* of $A \cdot \bar{b}$ has absolute value $\leqslant \sqrt{4m\ln(n)}$.
- There are n different entries of $\bar{c} = A \cdot \bar{b}$; we will show that for each entry, it is "too large" with probability $\leqslant \frac{2}{n^2}$. It then follows from the Union Bound that the probability that *some* entry is "too large" is $\leqslant n \cdot \frac{2}{n^2} = \frac{2}{n}$.

Set balancing: proof

- $\|\cdot\|_\infty$ is the absolute value of the largest entry of the tuple. We want to show that with high probability, *every entry* of $A \cdot \bar{b}$ has absolute value $\leqslant \sqrt{4m\ln(n)}$.
- There are n different entries of $\bar{c} = A \cdot \bar{b}$; we will show that for each entry, it is "too large" with probability $\leqslant \frac{2}{n^2}$. It then follows from the Union Bound that the probability that *some* entry is "too large" is $\leqslant n \cdot \frac{2}{n^2} = \frac{2}{n}$.
- For row i of A, there are $k_i \leqslant m$ entries that are non-0 (i.e., 1). The absolute value of $A_i \cdot \bar{b}$ is the (absolute) weighted sum of these entries, *randomly* weighted by +1 or -1 ... so we have k_i random trials of unbiased +1/-1. Let $Y_i = |A_i \cdot \bar{b}|$ be the random variable representing this sum. Setting $a = \sqrt{4m\ln(n)}$, the Chernoff bound in Thm 4.7 says

$$\Pr[Y_i \geqslant \sqrt{4m\ln(n)}] \leqslant 2e^{-4m\ln(n)/2k_i} = 2n^{-2m/k_i} \leqslant \frac{2}{n^2},$$

as required. □

More on set balancing

This last result implies that for *most* $\bar{b}$ we have $\|A \cdot \bar{b}\|_\infty = O(\sqrt{m \ln n})$, but better $\bar{b}$ exists, at least if $m = n$:

Theorem (Spencer, 1985)

For a n-by-n $0/1$ matrix A, there exists $\bar{b} \in \{+1, -1\}^n$ such that

$$\|A \cdot \bar{b}\|_\infty \leqslant 6\sqrt{n}.$$

This is tight up to constants. There exists A such that $\|A \cdot \bar{b}\|_\infty = \Omega(\sqrt{n})$ for any $\bar{b}$.

More on set balancing

This last result implies that for *most* $\bar{b}$ we have $\|A \cdot \bar{b}\|_\infty = O(\sqrt{m \ln n})$, but better $\bar{b}$ exists, at least if $m = n$:

Theorem (Spencer, 1985)

For a n-by-n 0/1 *matrix A, there exists* $\bar{b} \in \{+1, -1\}^n$ *such that*

$$\|A \cdot \bar{b}\|_\infty \leqslant 6\sqrt{n}.$$

This is tight up to constants. There exists A such that $\|A \cdot \bar{b}\|_\infty = \Omega(\sqrt{n})$ for any $\bar{b}$.

Spencer's result was non-constructive. Subsequently, efficient randomized polynomial-time algorithms to find such $\bar{b}$ where discovered by Bansal (2010) and by Lovett and Meka (2012). These algorithms have (subsequently) been derandomized.

To learn more on this topic, see Chapter 13 of the following book:
N. Alon and J. Spencer, "The Probabilistic Method", 4th edition, 2016.

Application: Monte Carlo algorithms with 2-sided error

Consider a decision problem, $D : \{0,1\}^* \rightarrow \{\text{“Yes”}, \text{“No”}\}$.
Suppose we have a (Monte Carlo) randomized polynomial time algorithm, M, with 2-sided error, that on input $x \in \{0,1\}^*$ of length $n = |x|$, runs in time $q(n)$, for some polynomial $q(\cdot)$, and such that for all $x \in \{0,1\}^*$,

$$\Pr[M(x) = D(x)] \geqslant \frac{3}{4}.$$

(N.B. here $3/4$ can be replaced with any $p = \frac{1}{2} + \epsilon$, where $\epsilon \in \Omega(\frac{1}{|x|})$.)

Question: Suppose we want to devise a new 2-sided error Monte Carlo randomized polynomial time algorithm, M', such that

$$\Pr[M'(x) = D(x)] \geqslant 1 - \frac{1}{2^n}.$$

Can we do it?

Application: Monte Carlo algorithms with 2-sided error

Consider a decision problem, $D : \{0,1\}^* \rightarrow \{\text{"Yes"}, \text{"No"}\}$.
Suppose we have a (Monte Carlo) randomized polynomial time algorithm, M, with 2-sided error, that on input $x \in \{0,1\}^*$ of length $n = |x|$, runs in time $q(n)$, for some polynomial $q(\cdot)$, and such that for all $x \in \{0,1\}^*$,

$$\Pr[M(x) = D(x)] \geqslant \frac{3}{4}.$$

(N.B. here 3/4 can be replaced with any $p = \frac{1}{2} + \epsilon$, where $\epsilon \in \Omega(\frac{1}{|x|})$.)

Question: Suppose we want to devise a new 2-sided error Monte Carlo randomized polynomial time algorithm, M', such that

$$\Pr[M'(x) = D(x)] \geqslant 1 - \frac{1}{2^n}.$$

Can we do it?
Hint: Yes, we can, with a simple algorithm, and we can prove its correctness using Chernoff bounds.

Application: Monte Carlo algorithms with 2-sided error

Consider a decision problem, $D : \{0, 1\}^* \to \{\text{"Yes"}, \text{"No"}\}$.
Suppose we have a (Monte Carlo) randomized polynomial time algorithm, M, with 2-sided error, that on input $x \in \{0, 1\}^*$ of length $n = |x|$, runs in time $q(n)$, for some polynomial $q(\cdot)$, and such that for all $x \in \{0, 1\}^*$,

$$\Pr[M(x) = D(x)] \geqslant \frac{3}{4}.$$

(N.B. here $3/4$ can be replaced with any $p = \frac{1}{2} + \epsilon$, where $\epsilon \in \Omega(\frac{1}{|x|})$.)

Question: Suppose we want to devise a new 2-sided error Monte Carlo randomized polynomial time algorithm, M', such that

$$\Pr[M'(x) = D(x)] \geqslant 1 - \frac{1}{2^n}.$$

Can we do it?
Hint: Yes, we can, with a simple algorithm, and we can prove its correctness using Chernoff bounds.

Error reduction for 2-sided error algorithms

Algorithm M'**:** On input x, with $n = |x|$, repeatedly run $M(x)$, a total of $20n$ times. Let $y_1, \ldots, y_{20n}$ denote the sequence of outputs of the different (independent) runs of $M(x)$. Our algorithm $M'(x)$ will answer "Yes" if a majority, i.e., $> 10n$, of the $20n$ different runs answered "Yes". Otherwise, it will answer "No".

Let the random variables $X_1, \ldots, X_{20n} \in \{0, 1\}$ be defined as follows:

$$X_i = \begin{cases} 1 & \text{if } y_i = D(x) \\ 0 & \text{otherwise} \end{cases}$$

Note that $X_1, \ldots, X_{20n}$ are mutually independent, and that $\Pr[X_i = 1] = 3/4$, for all $i \in [20n]$.

Let $X = \sum_{i=1}^{20n} X_i$. Note that $\mu = E[X] = \frac{3}{4}(20n) = 15n$.

Note that the new algorithm M' answers incorrectly only if $X \leqslant 10n$.
We want to bound the probability of this bad event.
We will use Chernoff bounds.

Error reduction for 2-sided error algorithms – proof

We will use Chernoff bounds for the lower tail (Theorem 4.5(2.)), which tells us that for any $0 < \delta < 1$,

$$\Pr[X \leqslant (1-\delta)\mu] \leqslant e^{-\mu\delta^2/2}$$

Let $\delta := \frac{1}{3}$. Note that $(1-\delta)\mu = \frac{2}{3} \cdot 15n = 10n$.
Hence we have:

$$\Pr[X \leqslant 10n] \leqslant e^{-15n(1/3)^2/2} = e^{-\frac{15}{18}n} \leqslant 2^{-n}.$$

(The last inequality follows because $e^{\frac{15}{18}} = 2.300975\ldots$.)

This completes the proof that the new algorithm M' has error probability at most $\frac{1}{2^n}$. Note M' has polynomial running time $(20n) \cdot q(n)$.□

Hoeffding's inequality — beyond Bernoulli

Chernoff bounds, as given, only work for sums of Bernoulli r.v.'s. What if allow sums of real-valued r.v's, $X_i \in [a, b]$?

Theorem (4.12, Hoeffding's inequality)

Let $X_1, \ldots, X_n$ be independent r.v.'s with $\mathrm{E}[X_i] = \mu$ and $\Pr[a \leqslant X_i \leqslant b] = 1$. Then,

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geqslant \varepsilon\right] \leqslant 2e^{-2n\varepsilon^2/(b-a)^2}.$$

The proof also goes through the moment generating function $\mathrm{E}[e^{tX}]$.
A slightly more general form of the theorem is:

Theorem (4.14, Hoeffding's inequality)

Let $X_1, \ldots, X_n$ be independent r.v.'s with $\mathrm{E}[X_i] = \mu_i$ and $\Pr[a_i \leqslant X_i \leqslant b_i] = 1$. Then,

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mu_i\right| \geqslant \varepsilon\right] \leqslant 2e^{-\frac{2n^2\varepsilon^2}{\sum_{i=1}^{n}(b_i-a_i)^2}}.$$

Not necessarily independent variables: Martingales and the Azuma-Hoeffding inequality

NOT Examinable. To learn more, see Chap. 13 of [MU] on "Martingales".

A sequence of r.v.'s $Z_0, Z_1, Z_2, \ldots$ such that $\mathrm{E}[|Z_i|] < \infty$ for all $i \geqslant 0$, is called a martingale (respectively, a super-martingale) if $\mathrm{E}[Z_{i+1} \mid Z_0, \ldots, Z_i] = Z_i$ (respectively, if $\mathrm{E}[Z_{i+1} \mid Z_0, \ldots, Z_i] \leqslant Z_i$) with probability 1, for all $i \geqslant 0$.

Example: let $X_1, X_2, X_3, \ldots$ be i.i.d. r.v.'s, $X_i \in \{-1, +1\}$, with $\Pr[X_i = +1] = p$, for all i. Let $q = (1-p)$. Let $S_n := \sum_{i=1}^{n} X_i$; $S_0 := 0$. Let $Z_n := S_n - n(p-q)$. Then $Z_0, Z_1, Z_2, \ldots$ defines a martingale. If $p \leqslant q$, then $S_0, S_1, S_2, \ldots$ defines a super-martingale. (Note $\mathrm{E}[|S_n|] \leqslant n$ and $\mathrm{E}[|Z_n|] \leqslant 2n$.)

Theorem (13.4: Azuma-Hoeffding inequality)

If $Z_0, \ldots, Z_n$ is a (super-)martingale such that for all $k \geqslant 1$ there is some $c_k \geqslant 0$ such that $\Pr[|Z_k - Z_{k-1}| \leqslant c_k] = 1$, then for all $t \geqslant 1$ and any $\lambda > 0$

$$\Pr[Z_t - Z_0 \geqslant \lambda] \leqslant \exp\left[\frac{-\lambda^2}{2(\Sigma_{k=1}^{t} c_k)}\right].$$

Proof is similar to proof of Hoeffding's inequality (see Chap. 13 of [MU]).

Another variation on Hoeffding's inequality

Not Examinable.
There are *many many* variations of Chernoff-Hoeffding bounds.
Here's another useful one (see Chap. 13 of [MU]):

Theorem (13.7: McDiarmid's Inequality)

Let $X_1, \ldots, X_n$ be independent random variables, X_k taking values in $A_k \subseteq \mathbb{R}$, for each $k \in [n]$. Suppose that the (measurable) function $f : \left(\times_{k=1}^{n} A_k\right) \to \mathbb{R}$ satisfies

$$|f(\bar{x}) - f(\bar{x}')| \leqslant c_k$$

whenever $\bar{x}, \bar{x}'$ only differ in their k-th coordinate.
Define the random variable $Y = f[X_1, \ldots, X_n]$. Then for any $t > 0$,

$$\Pr[|Y - E[Y]| \geqslant t] \leqslant 2 \exp\left[\frac{-2t^2}{\sum_{k \in [n]} c_k^2}\right].$$

McDiarmid's inequality can be derived from the Azuma-Hoeffding inequality. (See Chapter 13 of [MU].)

References

- Chapter 4 of [MU] sections 4.1-4.5
- If you want to learn more about the rich subject of Martingales, and the Azuma-Hoeffding inequality, see Chapter 13 of [MU]. (But that chapter and content is not examinable in this course.)