Randomness and Computation

or, "Randomized Algorithms"

Heng Guo

(Based on slides by M. Cryan)

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 $\delta > 0$,*
$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu};$$
2. *For any $0 < \delta \leq 1$,*
$$\Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3};$$
3. *For $R \geq 6\mu$,*
$$\Pr[X \geq R] \leq 2^{-R}.$$

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

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

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 $\mu = E[X]$. Then for any δ, $0 < \delta < 1$,
$$\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}.$$

- For almost all applications, we will want to work with 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.

Unbiased $+1/-1$ variables

In fact, for the case of unbiased variables, we can do even 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 $a > 0$,

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

Unbiased 0/1 variables

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

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

Note also that for any $t \in \mathbb{Z}$, that

$$\sum_{i=1}^{n} Y_i = t \quad \Leftrightarrow \quad \sum_{i=1}^{n} X_i = 2t - n$$

Corollary (4.9, 4.10)

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

$$\Pr[Y \geq \tfrac{n}{2} + a] = \Pr[X \geq 2a] \leq e^{-2a^2/n};$$
$$\Pr[Y \leq \tfrac{n}{2} - a] = \Pr[X \leq -2a] \leq e^{-2a^2/n}.$$

i.i.d. Bernoulli variables

For independent identically distributed (i.i.d.) Bernoulli variables with parameter p, their sum X satisfies the condition of Chernoff bounds.

Roughly speaking, X has

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

Set balancing

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.

Random choices are already pretty good: choose $\bar{b} \in \{-1, +1\}^m$ by generating b_i independently and uniformly from $\{-1, +1\}$. We can show

Theorem (4.11)

For $\bar{b}$ chosen uar from $\{-1, +1\}^m$,

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

Set balancing

- $\|\cdot\|_\infty$ 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 $\leq \sqrt{4m\ln(n)}$.

Set balancing

- $\|\cdot\|_\infty$ 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 $\leq \sqrt{4m\ln(n)}$.
- 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 $\leq \frac{2}{n^2}$. Then Union Bound shows that one of the entry is "too large" with probability $\leq \frac{2}{n}$.

Set balancing

- $\|\cdot\|_\infty$ 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 $\leq \sqrt{4m\ln(n)}$.
- 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 $\leq \frac{2}{n^2}$. Then Union Bound shows that one of the entry is "too large" with probability $\leq \frac{2}{n}$.
- For row i of A, there are S_i ($|S_i| \leq m$) entries which are non-0 (ie, 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 S_i random trials of unbiased +1/-1. Setting $a = \sqrt{4m\ln(n)}$, Thm 4.7 says the probability we exceed this is at most

$$2e^{-4m\ln(n)/2|S_i|} = 2n^{-2m/|S_i|} \leq \frac{2}{n^2},$$

as required.

Six standard deviations suffice

Last result implies that most $\bar{b}$ 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

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

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

Six standard deviations suffice

Last result implies that most $\bar{b}$ 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*

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

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

There are also efficient algorithms to find such $\bar{b}$ by Bansal (2010) and by Lovett and Meka (2012).

Check out Chapter 13 of "The Probabilistic Method" by Alon and Spencer.

Unbalancing lights

Let A be a n-by-n ± 1 matrix. There exist $\bar{x}, \bar{y} \in \{+1, -1\}^n$ such that

$$\bar{x}^{\mathrm{T}} A \bar{y} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i y_j \ge \left(\sqrt{2/\pi} + o(1)\right) n^{3/2}.$$

If we randomize both $\bar{x}$ and $\bar{y}$, the expectation is 0!

Unbalancing lights

Let A be a n-by-n ± 1 matrix. There exist $\bar{x}, \bar{y} \in \{+1, -1\}^n$ such that

$$\bar{x}^{\mathrm{T}} A \bar{y} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i y_j \ge \left(\sqrt{2/\pi} + o(1)\right) n^{3/2}.$$

If we randomize both $\bar{x}$ and $\bar{y}$, the expectation is 0!

However that is apparently a bad choice. Once $\bar{y}$ is fixed, we can choose $\bar{x}$ so that the signs of $\bar{x}$ and $A\bar{y}$ all match. Thus we are interested in

$$\sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} y_j \right|.$$

Unbalancing lights

Regardless of the value of a_{ij}, $a_{ij} y_j$ is a uar ± 1 rv. Call it s_j. In fact,

$$\begin{aligned} \mathrm{E}\left[\left|\sum_{j=1}^{n} a_{ij} y_j\right|\right] &= \mathrm{E}\left[\left|\sum_{j=1}^{n} s_j\right|\right] \\ &= \frac{2n}{2^n} \binom{n-1}{\lfloor (n-1)/2 \rfloor} \\ &= \left(\sqrt{2/\pi} + o(1)\right) n^{1/2} \end{aligned}$$

(The second equality was a 1974 Putnam competition problem.)

Unbalancing lights

Regardless of the value of a_{ij}, $a_{ij}y_j$ is a uar ± 1 rv. Call it s_j. In fact,

$$\begin{aligned} \mathrm{E}\left[\left|\sum_{j=1}^{n} a_{ij}y_j\right|\right] &= \mathrm{E}\left[\left|\sum_{j=1}^{n} s_j\right|\right] \\ &= \frac{2n}{2^n}\binom{n-1}{\lfloor (n-1)/2 \rfloor} \\ &= \left(\sqrt{2/\pi} + o(1)\right) n^{1/2} \end{aligned}$$

(The second equality was a 1974 Putnam competition problem.)
Thus,

$$\mathrm{E}\left[\sum_{i=1}^{n}\left|\sum_{j=1}^{n} a_{ij}y_j\right|\right] = \sum_{i=1}^{n}\left(\sqrt{2/\pi} + o(1)\right) n^{1/2} = \left(\sqrt{2/\pi} + o(1)\right) n^{3/2}.$$

There exists $\bar{y}$ that beats the expectation. We can use, for example, conditional expectation to find it.

Hoeffding's inequality — beyond Bernoulli

Chernoff bounds only work for Bernoulli rvs.

Theorem (4.12)

Let $X_1, \ldots, X_n$ be independent rvs such that $E[X_i] = \mu$ and $\Pr[a \le X_i \le b] = 1$. Then,

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

The constant is slightly weaker than Chernoff bounds (where $a = 0$ and $b = 1$). However it does not require X_i's to be Bernoulli.

The proof also goes through the moment generating function $\mathrm{E}[e^{tX}]$.

Hoeffding's inequality

Theorem (4.14)

Let $X_1, \ldots, X_n$ be independent rvs such that $E[X_i] = \mu_i$ and $\Pr[a_i \le X_i \le b_i] = 1$. Then,

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

More general Chernoff bounds

Many many variations. One general statement worth mentioning is due to McDiarmid:

Theorem

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

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

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

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

Correlation and concentration

Consider two Bernoulli random variable X and Y with parameter 1/2.

Independent: $\Pr[X = i \wedge Y = j] = 1/4$

$$X + Y = \begin{cases} 0 & \text{w.p. } 0.25 \\ 1 & \text{w.p. } 0.5 \\ 2 & \text{w.p. } 0.25 \end{cases}$$

Positive correlation: $\Pr[X = Y] = 1$

$$X + Y = \begin{cases} 0 & \text{w.p. } 0.5 \\ 2 & \text{w.p. } 0.5 \end{cases}$$

Negative correlation: $\Pr[X = 1 - Y] = 1$

$$X + Y = 1 \quad \text{w.p. } 1$$

Correlation and concentration

Consider two Bernoulli random variable X and Y with parameter 1/2.

Independent: $\Pr[X = i \wedge Y = j] = 1/4$

$$X + Y = \begin{cases} 0 & \text{w.p. } 0.25 \\ 1 & \text{w.p. } 0.5 \\ 2 & \text{w.p. } 0.25 \end{cases}$$

Positive correlation: $\Pr[X = Y] = 1$

$$X + Y = \begin{cases} 0 & \text{w.p. } 0.5 \\ 2 & \text{w.p. } 0.5 \end{cases}$$

Negative correlation: $\Pr[X = 1 - Y] = 1$

$$X + Y = 1 \quad \text{w.p. } 1$$

For more variables, negative correlation gets trickier. For example, Cryan, G., and Mousa (2019) give concentration bounds for rvs under matroid constraints.

RC (2019/20) – Lecture 8 – slide 16

References

- Chapter 4 of [MU]
- Chapter 2 of "The Probabilistic Method" (unbalancing lights) and Chapter 13 (six standard deviations suffice)
- We will not have time to cover the packet routing analysis of 4.5, but it's worth reading (not examinable in the exam).
- Next week: balls into bins, Chapter 5 of [MU]