Randomness and Computation

or, “Randomized Algorithms”

Mary Cryan

School of Informatics
University of Edinburgh

Bounding deviation

We already have ...

Theorem (3.1, Markov's Inequality)
Let X be any random variable that takes only non-negative values. Then for any $a > 0$,

$$\Pr[X \geq a] \leq \frac{\mathrm{E}[X]}{a}.$$

And also ...

Theorem (3.2, Chebyshev's Inequality)
For every $a > 0$,

$$\Pr[|X - \mathrm{E}[X]| \geq a] \leq \frac{\mathrm{Var}[X]}{a^2}.$$

These are *generic*. Chernoff/Hoeffding bounds (specific) give tighter bounds for *sums of independent 0/1 variables* and related distributions.

Chernoff Bounds

Many many variations. My favourite (general statement) is McDiarmid's presentation:

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

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

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 $t > 0$,

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

Corollaries from previous slide

Corollary
Let $X_1, \ldots, X_n$ be independent Bernoulli variables with parameters $p_1, \ldots, p_n$ respectively, and let $Y = \sum_{k=1}^{n} X_k$. (this case of Bernoulli's with differing ps is often called Poisson trials*). Let $\mu = \mathrm{E}[Y]$. Then for any $\delta > 0$,* $\Pr[|Y - \mathrm{E}[Y]| \geq \delta\mu] \leq 2\exp\left[\frac{-2(\delta\mu)^2}{n}\right]$.

This is not as tight as we'd like ... we can't cancel an n with μ as we don't know the size of the average p_k. **See Corollary 4.6 later**

Corollary
Let $X_1, \ldots, X_n$ be independent fair coin flips ($\Pr[X_k = 1] = 1/2$ *for every k) respectively, and let $Y = \sum_{k=1}^{n} X_k$. Note $\mu = \mathrm{E}[Y] = n/2$. Then for any $\delta > 0$,*

$$\Pr[|Y - \mathrm{E}[Y]| \geq \delta\mu] \leq 2\exp\left[\frac{-n\delta^2}{2}\right].$$

See Theorem 4.7 later (has $\frac{1}{4}$ rather than $\frac{1}{2}$).

Chernoff Bounds from the book

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

Theorem (4.4)

Let $X_1, \ldots, X_n$ be independent Poisson trials such that $\Pr[X_i = 1] = p_i$ *for all $i \in [n]$. Let $X = \sum_{k=1}^{n} X_k$, and $\mu = \mathrm{E}[X]$. 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 from the book

Lemma

For n independent Poisson trials $X_1, \ldots, X_n$ and $X = \sum_{i=1}^{n} X_i$, $\mu = \mathrm{E}[X]$,
$$\mathrm{E}[e^{tX}] \;\leq\; e^{\mu(e^t - 1)}.$$

Proof.

To prove the result, we will consider $\mathrm{E}[e^{tX}]$ for $t > 0$.

This is $\mathrm{E}[e^{t(\sum_{i=1}^{n} X_i)}] = E[\prod_{i=1}^{n} e^{tX_i}]$. The X_i and hence the e^{tX_i} are mutually independent, so by Thm 3.3, $\mathrm{E}[e^{tX}] = \prod_{i=1}^{n} \mathrm{E}[e^{tX_i}]$.

Each e^{tX_i} has expectation

$$\begin{aligned}
\mathrm{E}[e^{tX_i}] &= p_i \cdot e^t + (1 - p_i) \cdot 1 \\
&= 1 + p_i(e^t - 1) \\
&\leq e^{p_i(e^t - 1)} \qquad \text{by } 1 + x \leq e^x \text{ for } x \in \mathbb{R}
\end{aligned}$$

$$\mathrm{E}[e^{tX}] \leq \prod_{i=1}^{n} e^{p_i(e^t - 1)} = e^{\sum_{i=1}^{n} p_i(e^t - 1)} = e^{\mu(e^t - 1)}.$$

Chernoff Bounds from the book

Proof of Thm 4.4 (1.)

Interested in events when $X \geq (1+\delta)\mu$.
Identical to when $e^X \geq e^{(1+\delta)\mu}$, or for any $t > 0$, when $e^{tX} \geq e^{t(1+\delta)\mu}$.

$$\begin{aligned}
\Pr[X \geq (1+\delta)\mu] &= \Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \\
&\leq \frac{\mathrm{E}[e^{tX}]}{e^{t(1+\delta)\mu}} \qquad \text{by Markov's Inequality} \\
&\leq \frac{e^{\mu(e^t - 1)}}{e^{t(1+\delta)\mu}} \qquad \text{by Lemma just proved}
\end{aligned}$$

Now take $t = \ln(1+\delta)$ (and note this is > 0) to see

$$\begin{aligned}
\Pr[X \geq (1+\delta)\mu] &\leq \frac{e^{\mu(e^{\ln(1+\delta)} - 1)}}{e^{\ln(1+\delta)(1+\delta)\mu}} \\
&= \frac{e^{\mu\delta}}{(1+\delta)^{(1+\delta)\mu}} = \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.
\end{aligned}$$

Chernoff Bounds from the book

Proof of Thm 4.4 (2.)

Already have

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

The rhs will be $\leq e^{-\mu\delta^2/3}$ if and only if (taking μ-th root, then ln)

$$\delta - (1+\delta)\ln(1+\delta) < -\delta^2/3$$

We will show the following f is always negative for $\delta \in (0, 1)$

$$f(\delta) \;=_{def}\; \delta - (1+\delta)\ln(1+\delta) + \delta^2/3$$

Differentiating,

$$\begin{aligned}
f'(\delta) &= 1 - \ln(1+\delta) + (1+\delta)\frac{1}{1+\delta} + \frac{2\delta}{3} \\
&= -\ln(1+\delta) + \frac{2\delta}{3}.
\end{aligned}$$

Chernoff Bounds from the book

Proof of Thm 4.4 (2.) cont'd.

$$f'(\delta) = -\ln(1+\delta) + \frac{2\delta}{3}.$$

Differentiating again

$$f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3} = -\frac{1}{1+\delta} + \frac{2}{3}$$

Note

$$f''(\delta) \begin{cases} <0 & \text{for } 0<\delta<1/2 \\ 0 & \delta = 1/2 \\ >0 & \delta > 1/2 \end{cases}$$

Also $f'(0) = 0$, $f'(1) < 0$ (check $\delta = 1$ in top equation), and by f' decreasing first, then increasing from $1/2$) $f'(\delta) < 0$ on $(0,1)$.
By $f(0) = 0$, this implies that $f(\delta) \le 0$ in all of $[0,1]$.
Hence $\delta - (1+\delta)\ln(1+\delta) < -\delta^2/3$, proving 2. □

Chernoff Bounds from the book (other direction)

Theorem (4.5)

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

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

- Proof is similar to Thm 4.4.
- Bound of 2. is slightly *better* than for the $\ge (1+\delta)\mu$ bound.
- No 3. Why?

Concentration

Corollary (4.6)

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

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

- 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.

Analysing a collection of coin flips

We have $p_i = 1/2$ for all $i \in [n]$.
We have $\mu = \mathrm{E}[X] = \frac{n}{2}$, $\mathrm{Var}[X] = \frac{n}{4}$.
Consider the probability of being further than $5\sqrt{n}$ from μ.

Chebyshev $\Pr[|X - \mu| \ge 5\sqrt{n}] \le \frac{\mathrm{Var}[X]}{25n} = \frac{1}{100}$

Chernoff Work out the δ - we need $\mu\delta = 5\sqrt{n}$, so need $\delta = 5\sqrt{n}/\mu = 10\sqrt{n}/n = \frac{10}{\sqrt{n}}$. Then by Chernoff

$$\Pr[|X-\mu| \ge 5\sqrt{n}] \le 2e^{-\mu\delta^2/3} = 2e^{\frac{-10^2\cdot n}{2\cdot 3\cdot \sqrt{n}^2}} = 2e^{-16.6\cdots}.$$

This is much smaller than the Chebyshev bound (though note it doesn't depend on n).

Get much improved bounds because Chernoff uses specialised analysis for sums of independent Bernoulli variables.

Unbiased $+1/-1$ variables

In fact, for the case of unbiased variables, we can do even better than $2e^{-\mu\delta^2/3}$.

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 $\mu = \mathrm{E}[X] = 0$. Then for any $a > 0$,

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

Proof is in the book.

(uses Taylor series expansions for e^t, e^{-t}). $\square$

Constant is just a bit better than with Theorem 4.6. We did the details of this on the BOARD.

Set Balancing for statistical experiments

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.

Solution: 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 for statistical experiments (added post-lecture)

- $\|\cdot\|_\infty$ 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, we are "small enough" with probability $\geq 1 - \frac{2}{n^2}$. Then Union Bound shows that $\|A \cdot \bar{b}\cdot\|_\infty$ is bounded with prob $\geq 1 - \frac{2}{n}$.
- For any row i of A, there are some entries $S_i, |S_i| \leq m$ which are non-0 (ie, 1). The absolute value of $A_i \cdot \bar{b}$ is the (absolute) weighted sum of these 1s, *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.

References

- Chapter 4 of "Probability and Computing"
- We don't have time to cover the packet routing analysis of 4.5. But it's worth reading (but not examinable in the exam).