Randomised Algorithms

Lecture 2: Concentration Inequalities, Application to Balls-into-Bins

Thomas Sauerwald (tms41@cam.ac.uk)

Lent 2024

UNIVERSITY OF CAMBRIDGE

Outline

How to Derive Chernoff Bounds

Application 1: Balls into Bins

General Recipe for Deriving Chernoff Bounds

Recipe

The three main steps in deriving Chernoff bounds for sums of independent random variables $X = X_1 + \cdots + X_n$ are:

1. Instead of working with X, we switch to the **moment generating function** $e^{\lambda X}$, $\lambda > 0$ and apply Markov's inequality $\rightsquigarrow$ $\mathbf{E}\left[e^{\lambda X} \right]$
2. Compute an upper bound for $\mathbf{E}\left[e^{\lambda X} \right]$ (using independence)
3. Optimise value of λ to obtain best tail bound

Chernoff Bound: Proof

Chernoff Bound (General Form, Upper Tail)

Suppose $X_1, \ldots, X_n$ are independent Bernoulli random variables with parameter p_i. Let $X = X_1 + \ldots + X_n$ and $\mu = \mathbf{E}[X] = \sum_{i=1}^n p_i$. Then, for any $\delta > 0$ it holds that

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

Proof:

1. For $\lambda > 0$,

$$\mathbf{P}[X \geq (1+\delta)\mu] \underset{e^{\lambda x}\text{ is incr}}{\leq} \mathbf{P}\left[e^{\lambda X} \geq e^{\lambda(1+\delta)\mu}\right] \underset{\text{Markov}}{\leq} e^{-\lambda(1+\delta)\mu}\mathbf{E}\left[e^{\lambda X}\right]$$

2. $\mathbf{E}\left[e^{\lambda X}\right] = \mathbf{E}\left[e^{\lambda \sum_{i=1}^n X_i}\right] \underset{\text{indep}}{=} \prod_{i=1}^n \mathbf{E}\left[e^{\lambda X_i}\right]$

3. $$\mathbf{E}\left[e^{\lambda X_i}\right] = e^{\lambda}p_i + (1-p_i) = 1 + p_i(e^{\lambda}-1) \underset{1+x\leq e^x}{\leq} e^{p_i(e^{\lambda}-1)}$$

Chernoff Bound: Proof

1. For $\lambda > 0$,

$$\mathbf{P}\left[\, X \geq (1+\delta)\mu \,\right] \underset{e^{\lambda x} \text{ is incr}}{=} \mathbf{P}\left[\, e^{\lambda X} \geq e^{\lambda(1+\delta)\mu} \,\right] \underset{\text{Markov}}{\leq} e^{-\lambda(1+\delta)\mu}\, \mathbf{E}\left[\, e^{\lambda X} \,\right]$$

2. $\mathbf{E}\left[\, e^{\lambda X} \,\right] = \mathbf{E}\left[\, e^{\lambda \sum_{i=1}^{n} X_i} \,\right] \underset{\text{indep}}{=} \prod_{i=1}^{n} \mathbf{E}\left[\, e^{\lambda X_i} \,\right]$

3.
$$\mathbf{E}\left[\, e^{\lambda X_i} \,\right] = e^{\lambda} p_i + (1 - p_i) = 1 + p_i(e^{\lambda} - 1) \underset{1+x \leq e^{x}}{\leq} e^{p_i(e^{\lambda} - 1)}$$

4. Putting all together

$$\mathbf{P}\left[\, X \geq (1+\delta)\mu \,\right] \leq e^{-\lambda(1+\delta)\mu} \prod_{i=1}^{n} e^{p_i(e^{\lambda}-1)} = e^{-\lambda(1+\delta)\mu} e^{\mu(e^{\lambda}-1)}$$

5. Choose $\lambda = \log(1+\delta) > 0$ to get the result.

Chernoff Bounds: Lower Tails

We can also use Chernoff Bounds to show a random variable is **not too small** compared to its mean:

Chernoff Bounds (General Form, Lower Tail)

Suppose $X_1, \ldots, X_n$ are independent Bernoulli random variables with parameter p_i. Let $X = X_1 + \ldots + X_n$ and $\mu = \mathbf{E}[X] = \sum_{i=1}^{n} p_i$. Then, for any $0 < \delta < 1$ it holds that

$$\mathbf{P}[X \leq (1-\delta)\mu] \leq \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu},$$

and thus, by substitution, for any $t < \mu$,

$$\mathbf{P}[X \leq t] \leq e^{-\mu} \left(\frac{e\mu}{t}\right)^{t}.$$

Exercise on Supervision Sheet

Hint: multiply both sides by -1 and repeat the proof of the Chernoff Bound

Nicer Chernoff Bounds

"Nicer" Chernoff Bounds

Suppose $X_1, \ldots, X_n$ are independent Bernoulli random variables with parameter p_i. Let $X = X_1 + \ldots + X_n$ and $\mu = \mathbf{E}[X] = \sum_{i=1}^{n} p_i$. Then,

- For all $t > 0$,

$$\mathbf{P}[X \geq \mathbf{E}[X] + t] \leq e^{-2t^2/n}$$

$$\mathbf{P}[X \leq \mathbf{E}[X] - t] \leq e^{-2t^2/n}$$

- For $0 < \delta < 1$,

$$\mathbf{P}[X \geq (1+\delta)\mathbf{E}[X]] \leq \exp\left(-\frac{\delta^2 \mathbf{E}[X]}{3}\right)$$

$$\mathbf{P}[X \leq (1-\delta)\mathbf{E}[X]] \leq \exp\left(-\frac{\delta^2 \mathbf{E}[X]}{2}\right)$$

All upper tail bounds hold even under a relaxed independence assumption: For all $1 \leq i \leq n$ and $x_1, x_2, \ldots, x_{i-1} \in \{0, 1\}$,

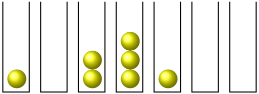
$$\mathbf{P}[X_i = 1 \mid X_1 = x_1, \ldots, X_{i-1} = x_{i-1}] \leq p_i.$$

Outline

How to Derive Chernoff Bounds

Application 1: Balls into Bins

Balls into Bins



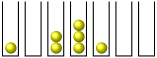
Balls into Bins Model

You have m balls and n bins. Each ball is allocated in a bin picked independently and uniformly at random.

- A very natural but also rich mathematical model
- In computer science, there are several interpretations:
  1. Bins are a hash table, balls are items
  2. Bins are processors and balls are jobs
  3. Bins are data servers and balls are queries

Exercise: Think about the relation between the Balls into Bins Model and the Coupon Collector Problem.

Balls into Bins: Bounding the Maximum Load (1/4)



Balls into Bins Model

You have m balls and n bins. Each ball is allocated in a bin picked independently and uniformly at random.

Question 1: How large is the maximum load if $m = 2n \log n$?

- Focus on an arbitrary single bin. Let X_i the indicator variable which is 1 iff ball i is assigned to this bin. Note that $p_i = \mathbf{P}[X_i = 1] = 1/n$.
- The total balls in the bin is given by $X := \sum_{i=1}^{n} X_i$.
- Since $m = 2n \log n$, then $\mu = \mathbf{E}[X] = 2 \log n$

here we could have used the "nicer" bounds as well!

$\mathbf{P}[X \geq t] \leq e^{-\mu}(e\mu/t)^t$

- By the Chernoff Bound,

$$\mathbf{P}[X \geq 6 \log n] \leq e^{-2 \log n} \left(\frac{2e \log n}{6 \log n}\right)^{6 \log n} \leq e^{-2 \log n} = n^{-2}$$

Balls into Bins: Bounding the Maximum Load (2/4)

- Let $\mathcal{E}_j := \{X(j) \geq 6 \log n\}$, that is, bin j receives at least $6 \log n$ balls.
- We are interested in the probability that at least one bin receives at least $6 \log n$ balls $\Rightarrow$ this is the event $\bigcup_{j=1}^{n} \mathcal{E}_j$
- By the Union Bound,

$$\mathbf{P}\left[\bigcup_{j=1}^{n} \mathcal{E}_j\right] \leq \sum_{j=1}^{n} \mathbf{P}\left[\, \mathcal{E}_j \,\right] \leq n \cdot n^{-2} = n^{-1}.$$

- Therefore whp, no bin receives at least $6 \log n$ balls
- By pigeonhole principle, the max loaded bin receives at least $2 \log n$ balls. Hence our bound is pretty sharp.

> whp stands for *with high probability*:
> An event $\mathcal{E}$ (that implicitly depends on an input parameter n) occurs whp if
> $\mathbf{P}\left[\, \mathcal{E} \,\right] \to 1$ as $n \to \infty$.
> This is a very standard notation in randomised algorithms
> but it may vary from author to author. Be careful!

Balls into Bins: Bounding the Maximum Load (3/4)

Question 2: How large is the maximum load if $m = n$?

- Using the Chernoff Bound: $\mathbf{P}[X \geq t] \leq e^{-\mu}(e\mu/t)^t$

$$\mathbf{P}[X \geq t] \leq e^{-1}\left(\frac{e}{t}\right)^t \leq \left(\frac{e}{t}\right)^t$$

- By setting $t = 4\log n / \log\log n$, we claim to obtain $\mathbf{P}[X \geq t] \leq n^{-2}$.
- Indeed:

$$\left(\frac{e\log\log n}{4\log n}\right)^{4\log n/\log\log n} = \exp\left(\frac{4\log n}{\log\log n}\cdot\log\left(\frac{e\log\log n}{4\log n}\right)\right)$$

- The term inside the exponential is

$$\frac{4\log n}{\log\log n}\cdot(\log(e/4) + \log\log\log n - \log\log n) \leq \frac{4\log n}{\log\log n}\left(-\frac{1}{2}\log\log n\right),$$

obtaining that $\mathbf{P}[X \geq t] \leq n^{-4/2} = n^{-2}$.

This inequality only works for large enough n.

Balls into Bins: Bounding the Maximum Load (4/4)

We just proved that

$$\mathbf{P}[X \geq 4 \log n / \log \log n] \leq n^{-2},$$

thus by the Union Bound, no bin receives more than $\Omega(\log n / \log \log n)$ balls with probability at least $1 - 1/n$. $\square$

- As mentioned on the to prove that whp at least one bin receives at least $c \log n / \log \log n$ balls, for some constant $c > 0$.

Conclusions

- If the number of balls is $2\log n$ times n (the number of bins), then to distribute balls at random is a good algorithm
  - This is because the worst case maximum load is whp. $6\log n$, while the average load is $2\log n$
- For the case $m = n$, the algorithm is not good, since the maximum load is whp. $\Theta(\log n/\log\log n)$, while the average load is 1.

A Better Load Balancing Approach

For any $m \geq n$, we can improve this by sampling two bins in each step and then assign the ball into the bin with lesser load.
$\Rightarrow$ for $m = n$ this gives a maximum load of $\log_2 \log n + \Theta(1)$ w.p. $1 - 1/n$.

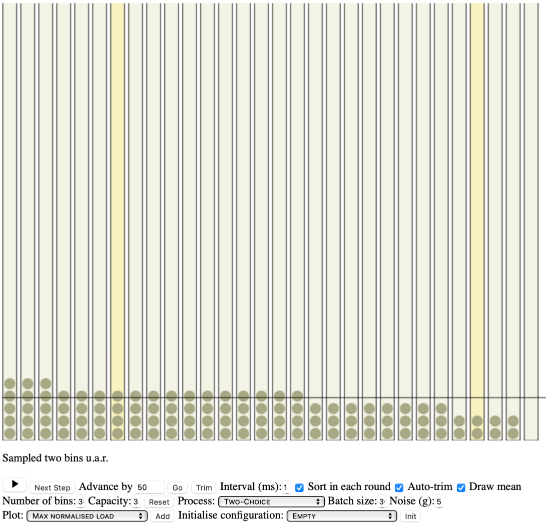
This is called the **power of two choices**: It is a common technique to improve the performance of randomised algorithms (covered in Chapter 17 of the textbook by Mitzenmacher and Upfal)

ACM Paris Kanellakis Theory and Practice Award 2020

For "the discovery and analysis of balanced allocations, known as the power of two choices, and their extensive applications to practice."

"These include i-Google's web index, Akamai's overlay routing network, and highly reliable distributed data storage systems used by Microsoft and Dropbox, which are all based on variants of the power of two choices paradigm. There are many other software systems that use balanced allocations as an important ingredient."

Simulation



https://www.dimitrioslos.com/balls_and_bins/visualiser.html