

Randomised Algorithms

Lecture 1: Introduction to Course & Introduction to Chernoff Bounds

Thomas Sauerwald (tms41@cam.ac.uk)

Lent 2024



UNIVERSITY OF
CAMBRIDGE

Outline

Introduction

Topics and Syllabus

A (Very) Brief Reminder of Probability Theory

Basic Examples

Introduction to Chernoff Bounds

Randomised Algorithms

What? **Randomised Algorithms** utilise random bits to compute their output.

Why? **Randomised Algorithms** often provide an efficient (and elegant!) solution or approximation to a problem that is costly (or impossible) to solve deterministically.

But often: **simple** algorithm at the cost of a **sophisticated** analysis!

“... If somebody would ask me, what in the last 10 years, what was the most important change in the study of algorithms I would have to say that people getting really familiar with randomised algorithms had to be the winner.”

- Donald E. Knuth (in *Randomization and Religion*)



How? This course aims to strengthen your knowledge of **probability theory** and apply this to analyse examples of **randomised algorithms**.

What if I (initially) don't care about randomised algorithms?

Many of the techniques in this course (**Markov Chains**, **Concentration of Measure**, **Spectral Theory**) are very relevant to other popular areas of research and employment such as **Data Science** and **Machine Learning**.

Some stuff you should know...

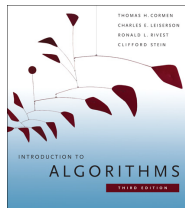
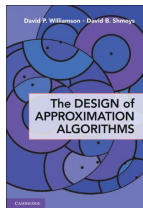
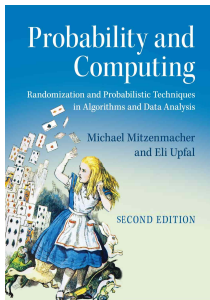
In this course we will assume some basic knowledge of **probability**:

- random variable
- computing expectations and variances
- notions of independence
- “general” idea of how to compute probabilities (manipulating, counting and **estimating**)



You should also be familiar with basic **computer science**, **mathematics** knowledge such as:

- graphs
- basic algorithms (sorting, graph algorithms etc.)
- matrices, norms and vectors



- (★) **Michael Mitzenmacher and Eli Upfal. Probability and Computing: Randomized Algorithms and Probabilistic Analysis, Cambridge University Press, 2nd edition, 2017**
- David P. Williamson and David B. Shmoys. The Design of Approximation Algorithms, Cambridge University Press, 2011
- Cormen, T.H., Leiserson, C.D., Rivest, R.L. and Stein, C. Introduction to Algorithms. MIT Press (3rd ed.), 2009
(We will adopt some of the labels (e.g., Theorem 35.6) from this book in Lectures 6-10)

Outline

Introduction

Topics and Syllabus

A (Very) Brief Reminder of Probability Theory

Basic Examples

Introduction to Chernoff Bounds

1 Introduction (Lecture)

- Intro to Randomised Algorithms; Logistics; Recap of Probability; Examples.

Lectures 2-5 focus on probabilistic tools and techniques.

2–3 Concentration (Lectures)

- Concept of Concentration; Recap of Markov and Chebyshev; Chernoff Bounds and Applications; Extensions: Hoeffding's Inequality and Method of Bounded Differences; Applications.

4 Markov Chains and Mixing Times (Lecture)

- Recap; Stopping and Hitting Times; Properties of Markov Chains; Convergence to Stationary Distribution; Variation Distance and Mixing Time

5 Hitting Times and Application to 2-SAT (Lecture)

- Reversible Markov Chains and Random Walks on Graphs; Cover Times and Hitting Times on Graphs (Example: Paths and Grids); A Randomised Algorithm for 2-SAT Algorithm

Lectures 6-8 introduce linear programming, a (mostly) deterministic but very powerful technique to solve various optimisation problems.

6–7 Linear Programming (Lectures)

- Introduction to Linear Programming, Applications, Standard and Slack Forms, Simplex Algorithm, Finding an Initial Solution, Fundamental Theorem of Linear Programming

8 Travelling Salesman Problem (Interactive Demo)

- Hardness of the general TSP problem, Formulating TSP as an integer program; Classical TSP instance from 1954; Branch & Bound Technique to solve integer programs using linear programs

We then see how we can efficiently combine linear programming with randomised techniques, in particular, rounding:

9–10 Randomised Approximation Algorithms (Lectures)

- MAX-3-CNF and Guessing, Vertex-Cover and Deterministic Rounding of Linear Program, Set-Cover and Randomised Rounding, Concluding Example: MAX-CNF and Hybrid Algorithm

Lectures 11-12 cover a more advanced topic with ML flavour:

11–12 Spectral Graph Theory and Spectral Clustering (Lectures)

- Eigenvalues, Eigenvectors and Spectrum; Visualising Graphs; Expansion; Cheeger's Inequality; Clustering and Examples; Analysing Mixing Times

Outline

Introduction

Topics and Syllabus

A (Very) Brief Reminder of Probability Theory

Basic Examples

Introduction to Chernoff Bounds

Recap: Probability Space

In probability theory we wish to evaluate the likelihood of certain results from an experiment. The setting of this is the **probability space** $(\Omega, \Sigma, \mathbf{P})$.

Components of the Probability Space $(\Omega, \Sigma, \mathbf{P})$

- The **Sample Space** Ω contains all the possible **outcomes** $\omega_1, \omega_2, \dots$ of the experiment.
- The **Event Space** Σ is the power-set of Ω containing **events**, which are combinations of outcomes (subsets of Ω including \emptyset and Ω).
- The **Probability Measure** \mathbf{P} is a function from Σ to \mathbb{R} satisfying
 - (i) $0 \leq \mathbf{P}[\mathcal{E}] \leq 1$, for all $\mathcal{E} \in \Sigma$
 - (ii) $\mathbf{P}[\Omega] = 1$
 - (iii) If $\mathcal{E}_1, \mathcal{E}_2, \dots \in \Sigma$ are pairwise disjoint ($\mathcal{E}_i \cap \mathcal{E}_j = \emptyset$ for all $i \neq j$) then

$$\mathbf{P} \left[\bigcup_{i=1}^{\infty} \mathcal{E}_i \right] = \sum_{i=1}^{\infty} \mathbf{P}[\mathcal{E}_i].$$

Recap: Random Variables

A **random variable** X on $(\Omega, \Sigma, \mathbf{P})$ is a function $X : \Omega \rightarrow \mathbb{R}$ mapping each sample “outcome” to a real number.

Intuitively, random variables are the “**observables**” in our experiment.

Examples of random variables

- The **number of heads** in three coin flips $X_1, X_2, X_3 \in \{0, 1\}$ is:

$$X_1 + X_2 + X_3$$

- The **indicator random variable** $\mathbf{1}_{\mathcal{E}}$ of an event $\mathcal{E} \in \Sigma$ given by

$$\mathbf{1}_{\mathcal{E}}(\omega) = \begin{cases} 1 & \text{if } \omega \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases}$$

For the indicator random variable $\mathbf{1}_{\mathcal{E}}$ we have $\mathbf{E}[\mathbf{1}_{\mathcal{E}}] = \mathbf{P}[\mathcal{E}]$.

- The **number of sixes** of two dice throws $X_1, X_2 \in \{1, 2, \dots, 6\}$ is

$$\mathbf{1}_{X_1=6} + \mathbf{1}_{X_2=6}$$

Recap: Boole's Inequality (Union Bound)

Union Bound is one of the most basic probability inequalities, yet it is extremely useful and easy to apply!

Union Bound

Let $\mathcal{E}_1, \dots, \mathcal{E}_n$ be a collection of events in Σ . Then

$$\mathbf{P} \left[\bigcup_{i=1}^n \mathcal{E}_i \right] \leq \sum_{i=1}^n \mathbf{P}[\mathcal{E}_i].$$

A Proof using Indicator Random Variables:

1. Let $\mathbf{1}_{\mathcal{E}_i}$ be the random variable that takes value 1 if \mathcal{E}_i holds, 0 otherwise
2. $\mathbf{E}[\mathbf{1}_{\mathcal{E}_i}] = \mathbf{P}[\mathcal{E}_i]$ (**Check this**)
3. It is clear that $\mathbf{1}_{\bigcup_{i=1}^n \mathcal{E}_i} \leq \sum_{i=1}^n \mathbf{1}_{\mathcal{E}_i}$ (**Check this**)
4. Taking expectation completes the proof.

Outline

Introduction

Topics and Syllabus

A (Very) Brief Reminder of Probability Theory

Basic Examples

Introduction to Chernoff Bounds

A Randomised Algorithm for MAX-CUT (1/2)

$E(A, B)$: set of edges with one endpoint in $A \subseteq V$ and the other in $B \subseteq V$.

MAX-CUT Problem

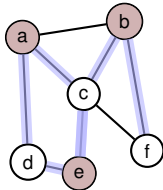
- **Given:** Undirected graph $G = (V, E)$
- **Goal:** Find $S \subseteq V$ such that $e(S, S^c) := |E(S, S^c)|$ is maximised.

Applications:

- network design, VLSI design
- clustering, statistical physics

Comments:

- This problem will appear again in the course
- MAX-CUT is NP-hard
- It is different from the **clustering** problem, where we want to find a **sparse cut**
- Note that the **MIN-CUT** problem is solvable in polynomial time!



$$S = \{a, b, e\}$$
$$e(S, S^c) = 6$$

A Randomised Algorithm for MAX-CUT (2/2)

RANDOMMAXCUT(G)

- 1: Start with $S \leftarrow \emptyset$
- 2: **For** each $v \in V$, add v to S with probability $1/2$
- 3: **Return** S

This kind of “random guessing” will appear often in this course!

Proposition

More details on [approximation algorithms](#) from Lecture 9 onwards!

RANDOMMAXCUT(G) gives a 2-approximation using time $O(n)$.

Later: learn stronger tools that imply concentration around the expectation!

Proof:

- We need to analyse the [expectation](#) of $e(S, S^c)$:

$$\begin{aligned} \mathbf{E}[e(S, S^c)] &= \mathbf{E}\left[\sum_{\{u,v\} \in E} \mathbf{1}_{\{u \in S, v \in S^c\} \cup \{u \in S^c, v \in S\}}\right] \\ &= \sum_{\{u,v\} \in E} \mathbf{E}[\mathbf{1}_{\{u \in S, v \in S^c\} \cup \{u \in S^c, v \in S\}}] \\ &= \sum_{\{u,v\} \in E} \mathbf{P}[\{u \in S, v \in S^c\} \cup \{u \in S^c, v \in S\}] \\ &= 2 \sum_{\{u,v\} \in E} \mathbf{P}[u \in S, v \in S^c] = 2 \sum_{\{u,v\} \in E} \mathbf{P}[u \in S] \cdot \mathbf{P}[v \in S^c] = |E|/2. \end{aligned}$$

- Since for any $S \subseteq V$, we have $e(S, S^c) \leq |E|$, the proof is complete.

Example: Coupon Collector



Source: <https://www.express.co.uk/life-style/life/567954/Discount-codes-money-saving-vouchers-coupons-mum>

This is a very important example in the design and analysis of randomised algorithms.

Coupon Collector Problem

Suppose that there are n coupons to be collected from the cereal box. Every morning you open a new cereal box and get one coupon. We assume that each coupon appears with the same probability in the box.

Example Sequence for $n = 8$: 7, 6, 3, 3, 3, 2, 5, 4, 2, 4, 1, 4, 2, 1, 4, 3, 1, 4, 8 ✓

Exercise ([Ex. 1.11])

In this course: $\log n = \ln n$

1. Prove it takes $n \sum_{k=1}^n \frac{1}{k} \approx n \log n$ expected boxes to collect all coupons
2. Use **Union Bound** to prove that the probability it takes more than $n \log n + cn$ boxes to collect all n coupons is $\leq e^{-c}$.

Hint: It is useful to remember that $1 - x \leq e^{-x}$ for all x

Outline

Introduction

Topics and Syllabus

A (Very) Brief Reminder of Probability Theory

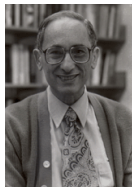
Basic Examples

Introduction to Chernoff Bounds

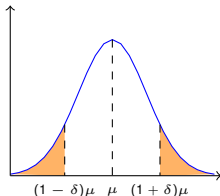
- **Concentration** refers to the phenomena where random variables are very close to their mean
- This is very useful in randomised algorithms as it ensures an **almost** deterministic behaviour
- It gives us the best of two worlds:
 1. **Randomised Algorithms:** Easy to Design and Implement
 2. **Deterministic Algorithms:** They do what they claim

Chernoff Bounds: A Tool for Concentration (1952)

- Chernoff's bounds are “strong” bounds on the tail probabilities of sums of independent random variables
- random variables can be discrete (or continuous)
- usually these bounds decrease exponentially as opposed to a polynomial decrease in Markov's or Chebyshev's inequality (see example)
- easy to apply, but requires independence
- have found various applications in:
 - Randomised Algorithms
 - Statistics
 - Random Projections and Dimensionality Reduction
 - Learning Theory (e.g., PAC-learning)
-
-



Hermann Chernoff (1923-)



Recap: Markov and Chebyshev

Markov's Inequality

If X is a non-negative random variable, then for any $a > 0$,

$$\mathbf{P}[X \geq a] \leq \mathbf{E}[X]/a.$$

Chebyshev's Inequality

If X is a random variable, then for any $a > 0$,

$$\mathbf{P}[|X - \mathbf{E}[X]| \geq a] \leq \mathbf{V}[X]/a^2.$$

- Let $f : \mathbb{R} \rightarrow [0, \infty)$ and **increasing**, then $f(X) \geq 0$, and thus

$$\mathbf{P}[X \geq a] \leq \mathbf{P}[f(X) \geq f(a)] \leq \mathbf{E}[f(X)]/f(a).$$

- Similarly, if $g : \mathbb{R} \rightarrow [0, \infty)$ and **decreasing**, then $g(X) \geq 0$, and thus

$$\mathbf{P}[X \leq a] \leq \mathbf{P}[g(X) \geq g(a)] \leq \mathbf{E}[g(X)]/g(a).$$

Chebyshev's inequality (or Markov) can be obtained by choosing $f(X) := (X - \mu)^2$ (or $f(X) := X$, respectively).

Markov and Chebyshev use the **first and second moment** of the random variable. Can we keep going?

- **Yes!**

We can consider the first, second, **third and more** moments! That is the basic idea behind the **Chernoff Bounds**

Our First Chernoff Bound

Chernoff Bounds (General Form, Upper Tail)

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

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

This implies that for any $t > \mu$,

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

While (\star) is one of the easiest (and most generic) Chernoff bounds to derive, the bound is complicated and hard to apply...

Example: Coin Flips (1/3)

- Consider throwing a **fair coin** n times and count the **total number of heads**
- $X_i \in \{0, 1\}$, $X = \sum_{i=1}^n X_i$ and $\mathbf{E}[X] = n \cdot 1/2 = n/2$
- The **Chernoff Bound** gives for any $\delta > 0$,

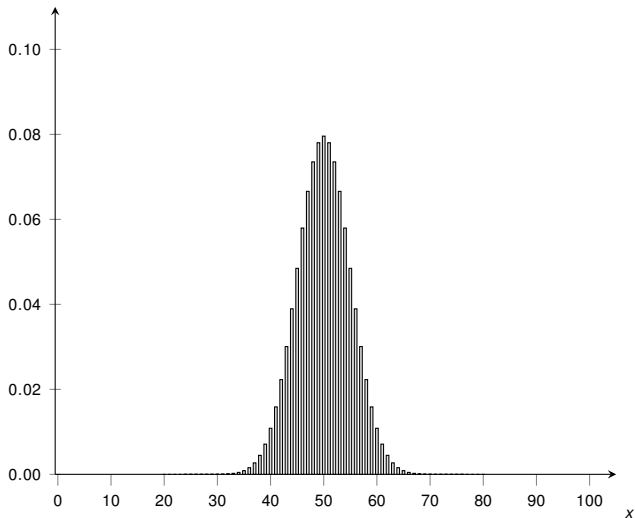
$$\mathbf{P}[X \geq (1 + \delta)(n/2)] \leq \left[\frac{e^\delta}{(1 + \delta)^{(1+\delta)}} \right]^{n/2}.$$

- The above expression equals 1 only for $\delta = 0$, and then it gives a value strictly less than 1 (check this!)
- ⇒ The inequality is **exponential in n** , (for fixed δ) which is much better than Chebyshev's inequality.

What about a **concrete value** of n , say $n = 100$?

Example: Coin Flips (2/3)

$P[\text{Bin}(100, 1/2) = x]$



Example: Coin Flips (3/3)

Consider $n = 100$ independent coin flips. We wish to find an upper bound on the probability that the number of heads is greater or equal than 75.

- Markov's inequality: $\mathbf{E}[X] = 100/2 = 50$.

$$\mathbf{P}[X \geq 3/2 \cdot \mathbf{E}[X]] \leq 2/3 = \mathbf{0.666}.$$

- Chebyshev's inequality: $\mathbf{V}[X] = \sum_{i=1}^{100} \mathbf{V}[X_i] = 100 \cdot (1/2)^2 = 25$.

$$\mathbf{P}[|X - \mu| \geq t] \leq \frac{\mathbf{V}[X]}{t^2},$$

and plugging in $t = 25$ gives an upper bound of $25/25^2 = 1/25 = \mathbf{0.04}$, much better than what we obtained by Markov's inequality.

- Chernoff bound: setting $\delta = 1/2$ gives

$$\mathbf{P}[X \geq 3/2 \cdot \mathbf{E}[X]] \leq \left(\frac{e^{1/2}}{(3/2)^{3/2}} \right)^{50} = \mathbf{0.004472}.$$

- Remark: The exact probability is $\mathbf{0.00000028 \dots}$

Chernoff bound yields a much better result (but needs independence!)