Lecture 1: Introduction

John Sylvester Nicolás Rivera Luca Zanetti Thomas Sauerwald

Lent 2019

UNIVERSITY OF CAMBRIDGE

Outline

Probability and Computation

What? Randomised algorithms utilise random bits to compute their output.

Why? A randomised algorithm often provides an efficient (and elegant!) solution or approximation to a problem that is costly to solve deterministically.

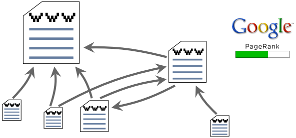
"... If somebody would ask me, what in the last 10 years, what was the most important change in the study of algorithms I would have to say that people getting really familiar with randomized algorithms had to be the winner."
- Donald E. Knuth

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

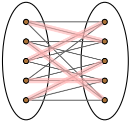
" What if I don't care about randomised algorithms?"
Much of the theory in this course (Markov Chains, Concentration of measure, Spectral theory) is very relevant to current "hot" areas of research and employment such as Data science and Machine learning.

Randomised Algorithms

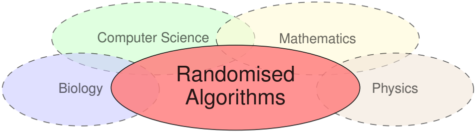
Ranking Websites



Sampling/Counting



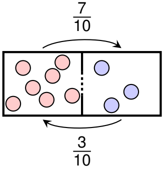
$$A = \begin{pmatrix} 0&0&0&0&0&1&1&0&0&1\\ 0&0&0&0&0&0&1&1&0&1\\ 0&0&0&0&0&1&1&0&1&0\\ 0&0&0&0&0&1&0&1&1&0\\ 0&0&0&0&0&0&0&1&1&1\\ 1&0&1&1&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0&0\\ 0&1&0&1&1&0&0&0&0&0\\ 0&0&1&1&1&0&0&0&0&0\\ 1&1&0&0&1&0&0&0&0&0 \end{pmatrix}$$



. . .

Graph Clustering/Sparsification

Particle Processes



Outline of the Course

Teaching Plan

- Probability and Markov chains (4 lectures) - John Sylvester .
- Concentration and Martingales (4 lectures) - Nicolás Rivera.
- Spectral techniques for MC's and algorithms (4 lectures) - Luca Zanetti.
- Applications to randomised algorithms (4 lectures) - Thomas Sauerwald.

Running along side these lectures will be

- Problem classes (6/7 total) - Hayk Saribekyan and Leran Cai.

Lecture and Problem class times

- Lectures: Tuesdays and Thursdays 2pm-3pm in LT2
- Problem classes: Thursdays 3.30pm-4.30pm in LT2 (Starting 24th Jan)

Assessment

- Recall: There is a "tick style" Homework Assessment to be submitted by 2pm Thursday 24th Jan via moodle and at reception.
- There will also be a 1.5 hour Written Test 9am on Friday 15 March in LT1.

Running Example 1: Max-Cut

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

MAX-CUT Problem

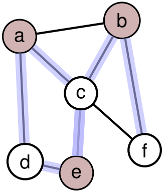
- Given: Undirected graph $G = (V, E)$
- Goal: Find $S \subseteq V$ such that $e(S, S^c) := |E(S, V \setminus S)|$ is maximised.

Applications:

- Semi-supervised learning
- Data mining

Comments:

- Max-Cut is NP-hard
- NP-hard to approximate with ratio $> 16/17 \approx .941$
- This example will be covered repeatedly:
  - Me - "Random guess": poly-time, approx. ratio $= 1/2$.
  - Nicolás - Concentration for max cut of a random graph.
  - Luca - Bi-partition via graph spectrum.
  - Thomas - SDP: poly-time, approximation ratio $\approx .879$.



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

Simple Randomised Algorithm for Max-Cut

Algorithm: RandMaxCut

Input $G = (V, E)$.
-Start with $S = \emptyset$.
-For each $v \in V$ add v to S independently with probability $1/2$.
Return S.

Proposition

In expectation RandMaxCut gives a $1/2$ approximation in linear time

Proof: What is the expected size of $e(S, S^c)$ for S output by RandMaxCut?

$$\mathbf{E}\left[\, e\left(S, S^c\right)\,\right] = \mathbf{E}\left[\sum_{vu \in E} \mathbf{1}_{\{v\in S, u \in S^c\}\cup\{v \in S^c, u\in S\}}\right] = \sum_{vu\in E}\mathbf{E}\left[\mathbf{1}_{\{v\in S, u \in S^c\}\cup\{v \in S^c, u\in S\}}\right]$$

$$= \sum_{vu\in E}\mathbf{P}\left[\{v \in S, u \in S^c\} \cup \{v \in S^c, u \in S\}\right]$$

$$= 2\sum_{vu\in E}\mathbf{P}\left[\, v \in S, u \in S^c\,\right] = 2\sum_{vu\in E}\mathbf{P}[\, v \in S\,]\,\mathbf{P}\left[\, u \in S^c\,\right] = |E|/2.$$

Since for any $S \subseteq V$ we have $e(S, S^c) \leq |E|$ this always gives us at least (in expectation) a $1/2$-approximation to the Max-Cut problem. □

Outline

Introduction

Probability Theory (Review)

First and Second Moment Methods

The Probabilistic Method

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* $\omega_1, \omega_2, \ldots$ 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 $\mathcal{E}_1, \mathcal{E}_2, \ldots \in \Sigma$ are pairwise disjoint ($\mathcal{E}_i \cap \mathcal{E}_j = \emptyset$ for all $i \neq j$) then

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

Probability Spaces from Randomised Algorithms

Running any randomised algorithm induces a probability space.

Algorithm: RandMaxCut

Given $G = (V, E)$ as input we output a cut-set S.
-Start with $S = \emptyset$.
-For each $v \in V$ add v to S independently with probability $1/2$.
Return S.

This is an example of a *Product Space*.

RandMaxCut on G with $|V| = n$ generates a Probability space $(\Omega, \Sigma, \mathbf{P})$ with

- $\Omega = \{0,1\}^n = \{(b_1, \ldots, b_n) : b_i = 1 \text{ if } i \in S, b_i = 0 \text{ if } i \notin S\}$.[1]
- $\Sigma = \mathcal{P}(\{0,1\}^n)$ (powerset of Ω). An example on an event $\mathcal{E} \in \Sigma$ is $\mathcal{E} = \{i \in S\} = \{b_i = 1\} = \bigcup_{j \neq i,\, b_j \in \{0,1\}} \{(b_1, \ldots, b_{i-1}, 1, b_{i+1}, \ldots, b_n)\}$.
- $\mathbf{P}$ is given by $\mathbf{P}[\,\mathcal{E}\,] = \sum_{\omega \in \mathcal{E}} \mathbf{P}[\,\{\omega\}\,] = |\mathcal{E}| 2^{-n}$ for any $|\mathcal{E}| \in \Sigma$. For example the event $\{i \in S\}$ above: $\mathbf{P}[\,i \in S\,] = \left(1 \cdot 2^{n-1}\right) 2^{-n} = 1/2$.

[1] $\{0,1\}^n = \{0,1\} \times \cdots \times \{0,1\}$ is a Cartesian product of sets $\{0,1\}$.

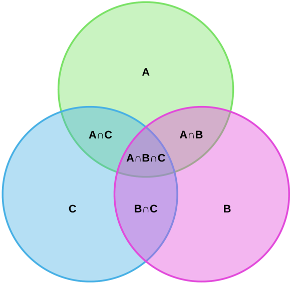
Union Bound

Union Bound/Boole's inequality

For any events $\mathcal{E}_1, \ldots, \mathcal{E}_n \in \Sigma$ the following holds,

$$\mathbf{P}[\,\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_n\,] \leq \mathbf{P}[\,\mathcal{E}_1\,] + \cdots + \mathbf{P}[\,\mathcal{E}_n\,],$$

with equality if events are disjoint.



Thus $\mathbf{P}[\,A \cup B \cup C\,] \leq \mathbf{P}[\,A\,] + \mathbf{P}[\,B\,] + \mathbf{P}[\,C\,]$.

Running Example 2: Balls into Bins

Content delivery problem

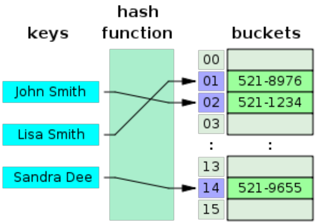
Assign m jobs to n servers as evenly as possible under constraints.

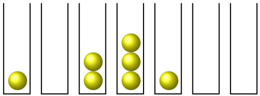
Balls into Bins

Assign balls (jobs) *Uniformly at Random*: a ball is equally likely to be assigned to any of the bins (servers), independently of the other balls.

Settings and Applications:

- Load Balancing: Assign m jobs to n servers as evenly as possible.
- Hash functions: Assign keys efficiently whilst trying to minimise clashes.





Apllication of the Union Bound: Balls into Bins

How many balls ensure there are no empty bins?

Assign m balls uniformly and independently to n bins. If $m = n \log n + Cn$ for $C > 0$ then with probability at least $1 - e^{-C}$ there is no empty bin.

Proof: Let $\mathcal{E}_i$ be the event that bin i is empty after m throws.

Since each ball is thrown independently

$$\mathbf{P}[\, \mathcal{E}_i \,] = \prod_{k=1}^{m} \mathbf{P}[\, \text{ball } k \text{ not in bin } i \,] = (1 - 1/n)^m.$$

Thus we have

$$\begin{aligned}
\mathbf{P}[\, \text{some bin is empty after } m \text{ balls} \,] &= \mathbf{P}\left[\bigcup_{i=1}^{n} \mathcal{E}_i \right] \\
&\overset{\text{union bdd}}{\leq} n \cdot \mathbf{P}[\, \mathcal{E}_i \,] \\
&= n \cdot (1 - 1/n)^m \\
&= n \cdot (1 - 1/n)^{n(\log n + C)} \\
&\leq n e^{-(\log n + C)} = e^{-C}.
\end{aligned}$$

For any real x, $1 + x \leq e^x$.

Random Variables

A *Random Variable* X on $(\Omega, \Sigma, \mathbf{P})$ is a function $X : \Omega \to \mathbb{R}$ mapping each sample "outcome" to a real number.

Intuitively random variables are the "observables" in our experiment.

Examples of random variables

- In RandMaxCut the size of the cut is a random variable given by

$$e(S, V \setminus S) = \sum_{u,v \in V} \mathbf{1}_{\{u \in S, v \in V \setminus S\}}.$$

- 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 $\mathbf{E}[\,\mathbf{1}_{\mathcal{E}}\,] = \mathbf{P}[\,\mathcal{E}\,]$.

Balls into Bins (Random Variables edition)

How many balls ensure there are no empty bins?

Let M be the number of balls we need to assign uniformly at random to occupy all bins. Then $\mathbf{E}[\,M\,] = n \log n + O(n)$.

Proof: During our first treatment of the problem we showed that

$$\mathbf{P}[\,\text{some bin is empty after } n \log n + Cn \text{ balls}\,] \leq e^{-C}$$

However this directly implies that $\mathbf{P}[\,M > n \log n + Cn\,] \leq e^{-C}$.
We can now calculate the expectation of M using the above *Tail Bound*.

a : Q2 of the homework assessment!
b : $\mathbf{P}[\,M > k\,] \geq \mathbf{P}[\,M > k+1\,]$

$$\begin{aligned}
\mathbf{E}[\,M\,] &\overset{a}{=} \sum_{m=0}^{\infty} \mathbf{P}[\,M > m\,] \\
&\leq \sum_{m=0}^{n \log n + n - 1} 1 + \sum_{m = n \log n + n}^{\infty} \mathbf{P}[\,M > m\,] \\
&\overset{b}{\leq} n \log n + n + \sum_{k=1}^{\infty} n \cdot \mathbf{P}[\,M > n \log n + kn\,] \\
&\leq n \log n + n + n \cdot \sum_{k=1}^{\infty} e^{-k} = n \log n + O(n)\,.
\end{aligned}$$

RandMaxCut Revisited

Proposition

For any $C < \infty$ there is an algorithm which runs in time $O(|E|^2)$ and gives a $1/2$ approximation with probability at least $1 - e^{-C}$.

The algorithm: Run RandMaxCut repeatedly until we get a cut $\geq |E|/2$.

Proof: The size of the cut can be checked in time $|E|$ so if we run RandMaxCut t times then the total run time will be $O(t \cdot |E|)$.

Let $p = \mathbf{P}[\, e(S, S^c) \geq |E|/2\,]$ and recall $\mathbf{E}[\, e\,(S, S^c)\,] = |E|/2$. We have

$$\frac{|E|}{2} = \sum_{i=1}^{|E|/2-1} i \cdot \mathbf{P}\left[\, e(S, S^c) = i\,\right] + \sum_{i=|E|/2}^{|E|} i \cdot \mathbf{P}\left[\, e(S, S^c) = i\,\right]$$

$$\leq (1-p)\left(\frac{|E|}{2} - 1\right) + p|E|.$$

This implies that $p \geq \frac{1}{|E|/2+1}$. If we run RandMaxCut $t = C|E|$ times using independent bits the probability of all the cuts being less than $|E|/2$ is at most

$$(1-p)^t \leq \left(1 - \frac{1}{|E|/2+1}\right)^{C|E|} \leq e^{-\frac{1}{|E|/2+1}\cdot C|E|} \leq e^{-C}. \qquad \square$$

Outline

Controlling the Probability Distribution using Moments

For $k \geq 1$ the *k^{th} Moment* of X is denoted $\mathbf{E}\left[X^k \right]$ and given by

$$\mathbf{E}\left[X^k \right] = \sum_{\omega \in \Omega} X(\omega)^k \cdot \mathbf{P}[\, \{\omega\} \,] .$$

Markov Inequality (First Moment Method)

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

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

Proof: Observe that $a \cdot \mathbf{1}_{\{X \geq a\}}(\omega) \leq X(\omega)$ for any $\omega \in \Omega$. Thus we have

$$\mathbf{E}[\, X \,] \geq \mathbf{E}\left[\, a \cdot \mathbf{1}_{\{X \geq a\}} \,\right] \stackrel{\text{linearity}}{=} a \cdot \mathbf{E}\left[\, \mathbf{1}_{\{X \geq a\}} \,\right] = a \cdot \mathbf{P}[\, X \geq a \,] . \qquad \square$$

The *Variance* is the centred second moment and is given by

$$\mathbf{Var}\,[\, X \,] = \mathbf{E}\left[(X - \mathbf{E}[\, X \,])^2 \right] = \mathbf{E}\left[X^2 \right] - \mathbf{E}[\, X \,]^2 .$$

Chebychev Inequality (Second Moment Method)

If X is a random variable and $a > 0$, then

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

MaxCut Revisited (Again)

Proposition

For any $C < \infty$ running RandMaxCut once returns a cut with at least $|E|/2 - \sqrt{C|E|}$ edges with probability at least $1 - 1/C$.

Proof: Let $X_{uv} = 1$ if u, v in different parts of the partition into S and $V \setminus S$. Observe that by independence and symmetry for any distinct vertices x, y, u, v we have $\mathbf{E}[\, X_{xy} X_{uv} \,] = \mathbf{E}[\, X_{xy} X_{xv} \,] = \mathbf{E}[\, X_{xy} X_{uy} \,] = 1/4$. For example

$$\mathbf{E}[\, X_{xy} X_{xv} \,] \overset{\text{sym}}{=} 2\mathbf{P}\left[\, x \in S, y \in S^c, v \in S^c \,\right] \overset{\text{ind}}{=} 2 \cdot (1/2)^3.$$

Notice $e(S, S^c) = \sum_{vu \in E} X_{uv}$. Thus for the second moment we have

$$\mathbf{E}\left[\, e(S, S^c)^2 \,\right] = \mathbf{E}\left[\left(\sum_{vu \in E} X_{uv} \right)^2 \right] = \sum_{xy \in E} \sum_{vu \in E} \mathbf{E}[\, X_{xy} X_{uv} \,]$$

$$= \sum_{xy \in E} \mathbf{E}\left[\, X_{xy}^2 \,\right] + \sum_{xy \in E} \sum_{uv \in E, uv \neq xy} \mathbf{E}[\, X_{xy} X_{uv} \,]$$

$$\leq \mathbf{E}\left[\, e(S, S^c) \,\right] + |E|^2/4.$$

Hence $\mathbf{Var}\left[\, e(S, S^c) \,\right] = \mathbf{E}[e(S, S^c)^2] - \mathbf{E}[\, e(S, S^c) \,]^2 \leq \mathbf{E}[\, e(S, S^c) \,]$ and so

Chebychev : $\mathbf{P}\left[\, e(S, S^c) \leq \mathbf{E}[\, e(S, S^c) \,] - C\sqrt{\mathbf{E}[\, e(S, S^c) \,]} \,\right] \leq 1/C^2.$ □

For Large Graphs, Once is Enough

Proposition

For any $C < \infty$ RandMaxCut returns a cut with at least $|E|/2 - \sqrt{C|E|}$ edges with probability at least $1 - 1/C$ in linear time.

Same Proposition Rehrased

For any $\varepsilon > 0$ there exists M s.t. if $m > M$ then RandMaxCut is a linear time $1/2 - \varepsilon$ approximation to Max-Cut with probability at least $1 - \varepsilon$.

We say an event $\mathcal{E}$ (depending on n) occurs *With High Probability* (w.h.p.) if

- for all $\varepsilon > 0$ there exists N such that for all $n > N$, $\mathbf{P}[\,\mathcal{E}\,] \geq 1 - \varepsilon$.

Outline

The Probabilistic Method - Non-Constructive Existence Result

Broadly speaking the Probabilistic Method is when we use probability to prove results in combinatorics.

Aim : proving a structure with certain desired properties exists.

Method : define a probability measure (typically uniform) on the structures and show the desired properties hold in this space with positive probability.

Example of a non-constructive existence result

Every finite graph has a cut of size at least $\lceil |E|/2 \rceil$

Proof: We saw if S is output by RandMaxCut ran on any graph then $\mathbf{P}\left[e(S, S^c) \geq \frac{|E|}{2} \right] \geq \frac{1}{|E|/2+1}$. Thus there exists a cut of size at least $\lceil \frac{|E|}{2} \rceil$. $\square$

We did not actually need a specific lower bound on $\mathbf{P}\left[e(S, S^c) \geq \frac{|E|}{2} \right]$:

- A discrete random variable X must take at least one value $\leq \mathbf{E}[\,X\,]$ with positive probability and at least one value $\geq \mathbf{E}[\,X\,]$ with positive probability.

Non-assessed - An Inequalitiy from the Probabilistic Method

A family $\mathcal{F}$ of sets is called *intersecting* if $A, B \in \mathcal{F}$ implies $A \cap B \neq \emptyset$.

Example of a non-trivial inequality: Erdős-Ko-Rado Theorem

> Suppose $n > 2k$ and let $\mathcal{F}$ be an intersecting family of k-element subsets of an n-set, then $|\mathcal{F}| \leq \binom{n-1}{k-1}$.

Proof: Let $A_i = \{s, s+1, \ldots, s+k-1\}$ where addition is modulo n. If we fix A_s then the other sets A_j, $j \neq s$, can be partitioned into $k-1$ sets of pairs $(A_{s-\ell}, A_{s+k-\ell})$, $0 \leq \ell \leq k-1$. The members of each pair are disjoint thus:

$$\mathcal{F} \text{ can contain at most } k \text{ of the sets } A_s. \tag{1}$$

Let a bijection $\sigma : [n] \to [n]$ and $i \in [n]$ be chosen uniformly independent from each other. Let $B = \{\sigma(i), \sigma(i+1), \ldots, \sigma(i+k-1)\}$ and observe:

(i) Since σ and i are random B is a uniform k-set, thus $\mathbf{P}[\, B \in \mathcal{F}\,] = |\mathcal{F}|/\binom{n}{k}$.

(ii) Any fixed σ is just a relabelling of the elements of n and $B = \sigma(A_i)$. Thus

$$\mathbf{P}[\, B \in \mathcal{F} \mid \sigma\,] = \mathbf{P}[\, A_i \in \mathcal{F}\,] \leq k/n.$$

Combining the above yields

$$\frac{|\mathcal{F}|}{\binom{n}{k}} = \mathbf{P}[\, B \in \mathcal{F}\,] = \sum_{\sigma} \mathbf{P}[\, B \in \mathcal{F} \mid \sigma\,]\, \mathbf{P}[\,\sigma\,] \leq \frac{k}{n} \sum_{\sigma} \mathbf{P}[\,\sigma\,] = \frac{k}{n}. \qquad \square$$