Lecture 1: Introduction

John Sylvester   Nicolás Rivera   Luca Zanetti   Thomas Sauerwald

Lent 1920

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.

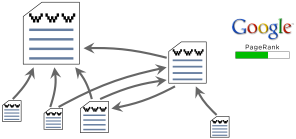
"... If somebody would ask me, what in the last 10 years, what was the most important change in the study of algorithms I would have to say that people getting really familiar with randomized algorithms had to be the winner."
- Donald E. Knuth

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

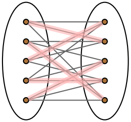
" What if I don't care about randomised algorithms?"
Much of the theory in this course (Markov Chains, Concentration of measure, Spectral theory) is very relevant to current "hot" areas of research and employment such as Data science and Machine learning.

Randomised Algorithms

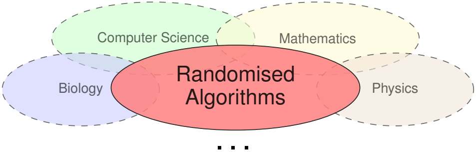
Ranking Websites



Sampling/Counting

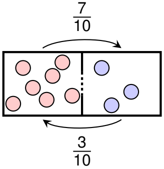


$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0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



Graph Clustering/Sparsification

Particle Processes



Outline of the Course

Teaching Plan

- Probability and Concentration (4 lectures) - Nicolas Rivera .
- Markov Chains (4 lectures) - John Sylvester.
- 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: Monday and Wednesday 2pm-3pm in LT2
- Problem class: Friday 2pm-3pm in LT2 (Starting 24th Jan)

Assessment

- Recall: There is a "tick style" Homework Assessment to be submitted by 2pm Monday 27th Jan via moodle and at reception.
- There will also be a 1.5 hour Written Test at the end of the term (Friday 13th March)

Outline

Introduction

Probability Reviews

Our first randomised Algorithm

Some old stuff you should know

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

- random variable
- computation of mean and variance
- notions of independence
- general ideas of how to compute probabilities (manipulation and some counting)

You should also be familiar with some basic computer science/discrete mathematics knowledge like

- graphs
- basic algorithms (sorting, some graph algorithm, etc)

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

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

To ground the previous definition. Consider the outcome of a die.

- Sample Space: $\{1, 2, 3, 4, 5, 6\}$
- Event Space: All subsets of $\{1, 2, 3, 4, 5, 6\}$.
  Example of event: “The outcome is even" =$\{2, 4, 6\}$
- Probability measure: $\mathbf{P}[\,\mathcal{E}\,] = |\mathcal{E}|/6$.
  Example: $\mathbf{P}[\,$“The outcome is even"$\,] = \mathbf{P}[\,\{2, 4, 6\}\,] = 3/6 = 1/2$.

We can do the same with two dice.

- Sample Space: $\{(1,1),(1,2),\ldots,(6,6)\}$
- Event Space: All subsets of $\{(1,1),(1,2),\ldots,(6,6)\}$.
  Example of event: " The sum of the dice is 10" =$\{(4,6),(5,5),(6,4)\}$
- Probability measure: $\mathbf{P}[\,\mathcal{E}\,] = |\mathcal{E}|/36$.
  Example: $\mathbf{P}[$" The sum of the dice is 10"$] = 3/36 = 1/12$

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.
Examples of Random variables: Consider our example with two dice.

- The canonical random variables are X_1 and X_2 that output the value of the first and second die, respectively
  E.g. $X_1(2, 3) = 2$, and $X_2(2, 3) = 3$
- Y takes value 1 if the second die is greater than the first one, otherwise it is 0.
  E.g. $Y(2, 3) = 1$, $Y(6, 1) = 0$
- Operations between random variables give new random variables.
  $Z = X_1 + X_2$ represent the sum.
  E.g. $Z(2, 3) = X_1(2, 3) + X_2(2, 3) = 2 + 3 = 5$.

Usually, we do not write the $X(\omega)$ and we just write X, since we do not really care about evaluating X, and it help us to see the idea of random behind it.

Random Variables

Given a random variable X. We write the events $\{X = k\}$ and $\{X \leq k\}$ as shortcut of

$$\{X = k\} = \{\omega \in \Omega : X(\omega) = k\}$$
$$\{X \leq k\} = \{\omega \in \Omega : X(\omega) \leq k\}$$

E.g. with the two dice

$$\{X_1 + X_2 = 10\} = \{\omega \in \Omega : X_1(\omega) + X_2(\omega) = 10\} = \{(4,6),(5,5),(6,4)\}$$

Moments of a Random Variable

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

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

The *variance* of X, denoted $\mathbf{Var}[X]$, is given by

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

The variance measures how much a random variable moves from it means.

E.g. Consider the two dice example, and let X_1 and X_2 the value of the first and second die respectively.

$$\mathbf{E}[X_1] = \sum_{i=1}^{6} \sum_{j=1}^{6} X_1((i,j)) \mathbf{P}[(i,j)] = \sum_{i=1}^{6} \sum_{j=1}^{6} \frac{i}{36} = \sum_{i=1}^{6} \frac{i}{6} = 3.5$$

Exercise: Let Z_1 and Z_2 be two random variables, and let a, b be constants. Prove that $\mathbf{E}[aZ_1 + bZ_2] = a\mathbf{E}[Z_1] + b\mathbf{E}[Z_2]$.
Exercise: Prove that $\mathbf{Var}[X] = \mathbf{E}\left[X^2\right] - (\mathbf{E}[X])^2$

Probability Mass Function

For a random variable X taking values in $x_1, \ldots, x_n$, we define its probability mass function $p_X : \mathbb{R} \to \mathbb{R}$ as

$$p_X(x) = \mathbf{P}[\, X = x\,] \,.$$

The cumulative distribution function $F_X : \mathbb{R} \to \mathbb{R}$ of the random variable X is

$$F_X(x) = \mathbf{P}[\, X \leq x\,] \,.$$

Recall that p_X or F_X are enough to compute moments of random variables.

Most of the random variables in this course are of discrete nature, meaning that they take a countable number of values. We will not discuss a lot about continuous random variables, but it is good to be comfortable with continuous distributions such as the Normal or Exponential distributions.

Probability Mass Function

Most random variables are named after the shape of their pmf or cdf.

$$p_X(x) = \mathbf{P}[\, X = x \,] \,.$$

- X is Bernoulli (also, X has Bernoulli distribution) of parameter $p \in (0, 1)$ if $p_X(1) = p$ and $p_X(0) = 1 - p$
- X is Binomial of parameter (n, p) with $n \in \mathbb{N}$, $p \in (0, 1)$ if for $k \in \{0, 1, \ldots, n\}$

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- X is Geometric of parameter p if for $k \in \{1, 2, \ldots, \}$, $p_X(k) = p(1 - p)^{k-1}$
- X is Poisson of parameter λ if for $k \in \{0, 1, \ldots, \}$, $p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$

Exercise: Compute the mean and variance of random variables with the distributions above.

Independency

- We say that to events $\mathcal{E}_1, \mathcal{E}_2 \in \Sigma$ are independent if and only if

$$\mathbf{P}[\,\mathcal{E}_1 \cap \mathcal{E}_2\,] = \mathbf{P}[\,\mathcal{E}_1\,]\,\mathbf{P}[\,\mathcal{E}_2\,]\,.$$

- In general, $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are (mutually) independent if and only if for any subset A of $\{1, \ldots, n\}$ we have

$$\mathbf{P}[\,\cap_{i \in A} \mathcal{E}_i\,] = \prod_{i \in A} \mathbf{P}[\,\mathcal{E}_i\,]\,.$$

- Two random variables X, Y are independent if for every $x, y \in \mathbb{R}$ the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent.
- A collection of random variables $X_1, \ldots, X_m$ are (mutually) independent if for every $x_1, \ldots, x_n \in \mathbb{R}$ the events $\{X_1 \leq x_1\}, \ldots, \{X_n \leq x_n\}$ are (mutually) independent.

Notation: Sometimes it is easier to write $\mathbf{P}[\,A, B\,]$ instead of $\mathbf{P}[\,A \cap B\,]$

Independency

Consider again the two dice examples. Let X_1 and X_2 the random variables giving the outcome of the first and second die, respectively.
We check that X_1 and X_2 are independent.

- $\mathbf{P}[\, X_1 \leq k_1 \,] = \mathbf{P}[\, \{(i,j) : i \leq k, j \in \{1,2,3,4,5,6\}\} \,] = k_1 \times 6/(36) = k/6$
- $\mathbf{P}[\, X_1 \leq k_1, X_2 \leq k_2 \,] = \mathbf{P}[\, \{(i,j) : i \leq k_1, j \leq k_2\} \,] = k_1 \times k_2/36$

Exercise. Let X and Y be two independent random variables with pmf p_X and p_Y. Suppose that X takes values $x_1, \ldots, x_n$ and that Y takes values $y_1, \ldots, y_m$.

1. Prove that $\mathbf{E}[\, X \,] = \sum_{i=1}^{n} x_i p_X(x_i)$ and $\mathbf{E}[\, Y \,] = \sum_{i=1}^{m} y_i P_Y(x_j)$.
2. Prove that if X and Y are independent. Then $f(X)$ and $g(Y)$ are also independent for any function f and g.
3. Prove that $\mathbf{E}[\, XY \,] = \mathbf{E}[\, X \,] \, \mathbf{E}[\, Y \,]$
4. Prove that $\mathbf{Var}\,[\, X + Y \,] = \mathbf{Var}\,[\, X \,] + \mathbf{Var}\,[\, Y \,]$

Conditional Probability

Given two events A and B with $\mathbf{P}[\,B\,] > 0$. We denote the conditional probability of A given B as $\mathbf{P}[\,A|B\,]$ and it is defined as

$$\mathbf{P}[\,A|B\,] = \frac{\mathbf{P}[\,A \cap B\,]}{\mathbf{P}[\,B\,]}.$$

Outline

Introduction

Probability Reviews

Our first randomised Algorithm

Max-Cut Problem

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

MAX-CUT Problem

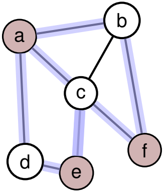
- Given: Undirected graph $G = (V, E)$
- Goal: Find $S \subseteq V$ such that $e(S, S^c) := |E(S, V \setminus S)|$ is maximised.

Applications:

- Semi-supervised learning
- Data mining

Comments:

- Max-Cut is NP-hard
- NP-hard to approximate with ratio $> 16/17 \approx .941$



$S = \{a, e, f\}$
$e(S, S^c) = 7$

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.

Simple Randomised Algorithm for Max-Cut

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

- $\Omega = \{0,1\}^n = \{(\omega_1, \ldots, \omega_n), \omega_i \in \{0,1\} \forall i\}$.[1]
- $\Sigma = \mathcal{P}(\{0,1\}^n)$ (the family of all subsets of Ω)
- $\mathbf{P}$ is given by $\mathbf{P}[\,\omega\,] = \frac{1}{2^n}$
- For example the event $\{i \in S\}$ above:
  $\mathbf{P}[\,i \in S\,] = |\{\omega \in \Omega : \omega_i = 1\}|/2^{-n} = 1/2$.

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

Simple Randomised Algorithm for Max-Cut

In practice, we don't need to define the probability space but it is good to know it is there.

Simple Randomised Algorithm for Max-Cut

In this algorithm, we are interested in the size of the cut created from our (random) set S, i.e. $e(S, S^c)$.

- In the analysis of randomised algorithm it is very important to have simple description of random variables
- Lets find a simple way of writting $e(S, S^c)$.
- Define X_i as 1 if vertex $i \in S$, 0 otherwise. X_i is what we call an **indicator random variable**
- For $i \neq j$, The event $\{X_i(1 - X_j) + (1 - X_i)X_j = 1\} = \{i \in S, j \in S^c\} \cup \{i \in S^c, j \in S\}$, that is $Y_{ij} = X_i(1 - X_j) + (1 - X_i)X_j$ indicates whether vertices i and j are in different sides of the partition.
- $e(S, S^c) = \sum_{\{i,j\} \in E} Y_{ij}$

Simple Randomised Algorithm for Max-Cut

- Define X_i as 1 if vertex $i \in S$, 0 otherwise
- $Y_{ij} = X_i(1 - X_j) + (1 - X_i)X_j$
- $e(S, S^c) = \sum_{\{i,j\} \in E} Y_{ij}$

We can analyse the size of the cut. Let's compute its expectation

$$\begin{aligned}
\mathbf{E}\left[\, e(S, S^c) \,\right] &= \sum_{\{i,j\} \in E} \mathbf{E}[\, Y_{ij} \,] \\
&= \sum_{\{i,j\} \in E} \mathbf{E}[\, X_i(1 - X_j) + (1 - X_i)X_j \,] \\
&= \sum_{\{i,j\} \in E} \mathbf{E}[\, X_i \,]\, \mathbf{E}[\, 1 - X_j \,] + \mathbf{E}[\, 1 - X_i \,]\, \mathbf{E}[\, X_j \,] = (1/2)|E|
\end{aligned}$$

Hence, we conclude that

$$\frac{1}{2}|E| = \mathbf{E}\left[\, e(S, S^c) \,\right] \leq \max_S e(S, S^c) \leq |E| \leq 2\mathbf{E}\left[\, e(S, S^c) \,\right]$$

Thus, in expectation, our algorithm gives us a 2-approximation of the Max-Cut.

Simple Randomised Algorithm for Max-Cut

Let's measure how much the random variable $e(S, S^c)$ moves from its mean. We would like to compute the variance but it is a bit hard.

$$\begin{aligned}
\mathbf{E}\left[\, e(S,S^c)^2 \,\right] &= \sum_{\{i,j\}\in E} \sum_{\{k,\ell\}\in E} \mathbf{E}[\, Y_{ij} Y_{k\ell} \,] \\
&= \sum_{\{i,j\}\in E} \mathbf{E}[\, Y_{ij} \,] + \sum_{\{i,j\}\in E} \sum_{\{k,\ell\}\neq\{i,j\}} \mathbf{E}[\, Y_{ij} Y_{k\ell} \,] \\
&\leq \sum_{\{i,j\}\in E} \mathbf{E}[\, Y_{ij} \,] + \sum_{\{i,j\}\in E} \sum_{\{k,\ell\}\in E} \mathbf{E}[\, Y_{ij} \,] \, \mathbf{E}[\, Y_{k\ell} \,] \\
&= \mathbf{E}\left[\, e(S,S^c) \,\right] + \mathbf{E}\left[\, e(S,S^c) \,\right]^2
\end{aligned}$$

In this course you don't have to be afraid of inequalities.

Simple Randomised Algorithm for Max-Cut

We have

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

Then using that $\mathbf{Var}\left[\, X \,\right] = \mathbf{E}\left[\, X^2 \,\right] - \mathbf{E}[\, X \,]^2$, we get

$$\mathbf{Var}\left[\, e(S, S^c) \,\right] \leq \mathbf{E}\left[\, e(S, S^c) \,\right].$$

Next class we will see that this upper bound on the variance implies that the random variable is very close to its mean.