Probability and Computing

Lecturers
Samuel Baguley, Andreas Goebel, Panagiotis Aivasiliotis

Notes
Simon Cyrani

Version
git: ef46b4f-*
compiled: Wednesday 17th July, 2024 04:01

Abstract

The following lecture notes are my personal (and therefore unofficial) write-up for 'Probability and Computing' aka 'ProbComp', which took place in summer semester 2024 at Hasso-Plattner-Institut. I do not guarantee correctness, completeness, or anything else. Importantly, note that I willfully changed some specific notations, reordered some material, and left out parts that I didn't found worth typing down.

If you miss something, feel free to contribute in the repository!

Contents

Summary of lectures

Part I

Lecture notes

Lecture 1
Mo 15 Apr 2024

1 Probabilistic method

This section will introduce how probability can be used to solve problems that at first glance do not have anything to do with probability. Consider following combinatorial problem.

Definition 1.1 (Ramsay numbers)**.** We define R_k as the smallest integer n such that any graph with n vertices must contain either a clique or an independent set of size K. We call R_k the (symmetric) **Ramsey numbers**.

Let us first look at some examples to get a feel for what this section is about.

Example 1.2. The first few Ramsay numbers are given as

- $R_2 = 2$: Obviously, $R_k \geq k$. Furthermore, there are only two different graphs with two nodes, i.e. with or without an edge. In the former case, both nodes form a clique, in the latter they form an independet set of size 2.
- $R_3 = 6$: First, let us show $R_3 \geq 6$. Consider a cycle of 5 nodes. Then, there is no clique of size 3 (since there is no 3-cycle). Also, among any three nodes two nodes are connected by an edge, so there is no independent set. Therefore, this is a counterexample. Now, consider a graph with 6 nodes. Suppose there is no clique of size 3. Then it suffices to show that there is an independent set of 3 nodes. Indeed, with an ugly case distinction this is possible: If there are no cycles, the graph is bipartite, so there is an independent set of at least 3. Otherwise, there exists at least a 4-cycle, but no 3-cycle (i.e. chord-free). We can then select at least two independent nodes from the cycle, and if needed the missing third node from the non-cycle nodes such that they form an independent set.
- $R_4 = 18$. Trust me, we do not want the proof here.
- $R_5 \in [43, 48]$. The exact value is indeed still unknown!

As we can see, even for small k, it is not trivial to determine their Ramsay number. Instead, let us try to at least find some bound for their value.

Theorem 1.3. For every $k \geq 1$ holds $R_k > 2^{k/2}$.

Proof. Consider a uniform distribution over all graphs with n vertices (i.e. the Erdős–Rényi random graph $\mathcal{G}(n, \frac{1}{2})$). Each edge in particular exists with probability $\frac{1}{2}$. Now, have

a look at $p \coloneqq P(G \sim \mathcal{G}(n, \frac{1}{2})$ has a k-clique or k independent set). If we can show that this probability p is less than 1, then this means there is a graph with n vertices such that the property is *not* satisfied, and therefore $R_k > n$.

Let S be a k-tuple of vertices. S is per definition a k-clique if *all* or *none* of its edges is existent. Therefore, its probability of being either one is given as

$$P(S \text{ is } k\text{-clique or } k\text{-independent set}) = 2 \cdot \frac{1}{2^{\binom{k}{2}}}. \tag{1}$$

The total number of k-subsets given n vertices is given as $\binom{n}{k}$, so by basic properties of probability and binomial coefficients we see

$$p \leq \binom{n}{k} \cdot \frac{1}{2^{\binom{k-1}{2}}} \leq \frac{n^k}{k!} 2^{1-\frac{k^2-k}{2}} \tag{2}$$

For $n = 2^{k/2}$ the right-hand side reduces to $\frac{2^{k+2}}{k!}$, which can be shown easily to be smaller than 1 for $k \geq 3$. □

Notice how we suddenly imposed a probabilistic view on this presumably deterministic problem! This technique of using a suitable random model to demonstrate the existence and/or non-existence of certain properties is known as **Probabilistic Method**.

Before we have a look at another example, we need following lemma.

Lemma 1.4. Let X be a discrete random variable over a set Ω with $\mathbb{E}[X] = \mu$. Then, $P(X \geq \mu) > 0$ and $P(X \leq \mu) > 0$.

Proof. Assume $P(X \geq \mu) = 0$. Then $P(X = x) = 0$ for $x \geq \mu$. By definition and assumption therefore

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} xP(X = x) = \sum_{x < \mu} xP(X = x) \tag{3}$$

$$< \sum_{x < X(\Omega)} \mu P(X = x) = \mu \sum_{x \in X(\Omega)} P(X = x) = \mu. \tag{4}$$

This is a contradiction! □

Consider following problem.

Definition 1.5 (Max-Cut)**.** Given a graph $G = (V, E)$, find a set A maximizing the number of edges between A and $V \setminus A$. We call this the **Maximum Cut Problem**.

As it is usual in lectures of this kind, its canonical decision variant is indeed a NP-complete problem. Again, let us try instead to find a "good" cut.

Theorem 1.6 (Minimal Max-Cut)**.** Given a graph $G = (V, E)$ with $|E| = m$. There exists $A \subseteq V$ with at least $\frac{m}{2}$ cut size.

Proof. Choose A uniformly over $\mathcal{P}(V)$, i.e. every node is chosen with probability $\frac{1}{2}$. Then, every edge is with probability $\frac{1}{2}$ included in the cut, which happens iff exactly one of the nodes of the edge is in A. Let X be the number of cut edges, and X_e be the indicator variable for e being in the cut. By linearity of expectancy,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{e \in E} X_e\right] = \sum_{e \in E} \mathbb{E}[X_e] = m \cdot \frac{1}{2}. \tag{5}$$

Using Lemma 1.4 there is positive probability for a choice of A having cut size at least $m/2$. □

You might have noticed that these are non-constructive results, so naturally following question emerges: Can we use these results to construct a solution? Indeed, there is a simple answer!

Definition 1.7 (Las-Vegas Algorithm)**.** Let T be the run-time of an algorithm and n its input size. We call a **Las-Vegas Algorithm** an algorithm which

1. always returns the correct answer with $\mathbb{E}[T] \in \mathcal{O}(n^k)$ for some k, or
2. has runtime always in $\mathcal{O}(n^k)$ for some k, and returns a correct answer with probability at least $\delta > 0$ (and otherwise returns no answer).

Remark 1.8. Both definitions are equivalent, which stems from the fact the second variant can be seen as a geometric process: Consider the number T of attempts until the algorithm returns a correct result. Then, $\mathbb{E}[T] = \frac{1}{1-\delta}$, so the total runtime is still $\mathcal{O}(n^k)$ in expectancy.

Turning back to the Max-Cut problem, let us transform our result of Theorem 1.6 into a Las-Vegas algorithm. Simply generate A randomly as constructed in the proof, and check if it has enough cut edges in polynomial time. The probability that this works is $P(X_A \geq \frac{m}{2}) = p$. However, this is just an abstract value - let us try to find a more meaningful way of expressing p:

$$\frac{m}{2} = \mathbb{E}[X_A] = \sum_{i < \frac{m}{2}} i \cdot P(X_A = i) + \sum_{i \geq \frac{m}{2}} i \cdot P(X_A = i) \tag{6}$$

$$\leq \left(\frac{m}{2} - 1\right) \sum_{i < \frac{m}{2}} P(X_A = i) + m \sum_{i \geq \frac{m}{2}} P(X_A = i) = \left(\frac{m}{2} - 1\right)(1 - p) + mp \tag{7}$$

Here we just upper-bound the corresponding first factors in each sum, and then use $P(X_A \geq \frac{m}{2}) = p$. Using some easy algebra, we deduce $p \geq \frac{1}{\frac{m}{2}+1}$, so our Las-Vegas approach seems to get gradually worse the more edges our graph has.

Interestingly, we do not even need a randomized algorithm to find a big cut. Instead, using probabilistic arguments we can construct a deterministic algorithm still running in polynomial time. This technique is known as **Derandomization**.

Algorithm 1: Find Big-Cut of $G = (V, E)$

$A \leftarrow \emptyset, B \leftarrow \emptyset$
for $k = 1, \dots, n$ **do**
  if $|\{(v_k, u) \in E \mid u \in A\}| \leq |\{(v_k, u) \in E \mid u \in B\}|$ **then**
    $A \leftarrow A + v_k$
  end
  else
    $B \leftarrow B + v_k$
  end
end
return A

The idea is to greedily decide for each vertex if we want it in our cut set or not based on a case distinction using conditional expectation.

Theorem 1.9. Given a graph $G = (V, E)$. Let $A \sim \mathcal{U}_{\mathcal{P}(V)}$, C_A the amount of cuts, and $x_1, \dots, x_n$ indicate if the vertices $v_1, \dots, v_n \in A$. Then Algorithm 1 satisfies following statements:

1. $\mathbb{E}[C_A \mid x_1, \dots, x_k] \leq \mathbb{E}[C_A \mid x_1, \dots, x_k, x_{k+1}]$ after iteration $k+1$ of the for-loop (such that x_i is fixed by the algorithm).
2. The algorithm runs in $\mathcal{O}(n+m)$ and returns a big cut of size at least $m/2$.

Proof. Notice for $1 \leq k < n$ by definition of conditional expectancy

$$\mathbb{E}[C_A \mid x_1, \dots, x_k] = \frac{1}{2}\mathbb{E}[C_A \mid x_1, \dots, x_k, x_{k+1} = 1] + \frac{1}{2}\mathbb{E}[C_A \mid x_1, \dots, x_k, x_{k+1} = 0].$$

So, at least one of the choices for x_{k+1} satisfy the required lower bound for the conditional expectancy. It remains to prove that our algorithm also chooses this value, i.e. the choice with larger condtional expectancy. Let us observe how the expected number of cut edges can change if we fix the position of vertex v_{k+1}. Consider following cases for an edge $e \in E$:

- e does not contain v_{k+1}. Then, by indepence, the expected value of its Is-Cut-Edge indicator X_e conditioned on all fixed vertices does not change by fixing v_{k+1} (i.e 1 or 0 if both vertices are determined, else $\frac{1}{2}$).
- e contains v_{k+1}, but the other vertex is not fixed. Again, aforementioned expected value does not change since there is still a $\frac{1}{2}$ probability for the other vertex being in A.
- e contains v_{k+1}, but the other vertex *is* fixed. Then we suddenly fix the value of X_e depending on the choice of x_{k+1}. In particular, this changes the conditional expectancy of X_e, increasing it to 1 or decreasing it to 0, and therefore the conditional expectancy on the number of cut edges changes in the same way.

In summary, the change in conditional expectancy by fixing x_{k+1} only depends on the neighborhood of fixed vertices of v_{k+1}. If there are more neighbors that are fixed to be A than not in A, then

$$\mathbb{E}\left[C_A \mid x_1, \ldots, x_k, x_{k+1} = 0\right] \geq \mathbb{E}\left[C_A \mid x_1, \ldots, x_k, x_{k+1} = 1\right],$$

otherwise the other way around, adhering to our algorithm design.

In fact, by induction it immediately follows that $\frac{m}{2} \leq \mathbb{E}\left[X_A\right] \leq \mathbb{E}\left[X_A \mid x_1, \ldots, x_n\right]$ for $x_1, \ldots, x_n$ being chosen according to the algorithm. So, our algorithm works, and has mentioned runtime. □

Lecture 2
We 17 Apr 2024

Definition 1.10 (SAT problem)**.** Given a boolean formula in k-CNF (conjunctive normal form). The decision problem if there is an assignment that the formula is true is called k-SAT.

Assume we have a k-SAT instance φ, and assume each variable appears in exactly one clause. Then it is possible to show $\varphi \in \mathsf{SAT}$.

Example 1.11. For 3-SAT consider $(x_1 \vee x_2 \vee \neg x_3) \wedge (x_4 \vee \neg x_5 \vee x_6)$. Then, a valid assignment is $x_1 = x_2 = x_3 = 1, x_4 = x_5 = x_6 = 0$.

It is rather intuitive this statement must hold, since we enforce some form of indepence. This makes it interesting to look at from the lense of the probability, and its notion of indepence.

So, let us assume a uniform random assignment α for such a φ (with n clauses). Let E_i be the event that clause i is not satisfied by α. Then, $\bigcup E_i$ is the event that φ is not satisfied by α, and thus $\bigcap \overline{E_i}$ denotes φ being satisfied by α. Similarly to previous applications of the probabilistic method we see that

$$P\left(\bigcap_{i=1}^{n} \overline{E_i}\right) = \prod_{i=1}^{n} P(\overline{E_i}) = \prod_{i=1}^{n} 1 - 2^{-k} > 0 \tag{8}$$

shows that there *are* assignments that satisfy α. However, we were now able to use indepence of E_i in this case.

While this example does not seem too spectacular at first, we can now try to leviate our conditions and only assume limited independence. Let us introduce a new construct.

Definition 1.12 (Dependency graph)**.** Consider a set of events $E_1, \ldots, E_n$. We call $G = (V, E)$ with $V = \{1, \ldots, n\}$ the **Dependency Graph** if E satisfies that every E_i is mutually independent of the set $\{E_j \mid (i, j) \neq E_i\}$.

Lemma 1.13. If events $E_1, \dots, E_n$ are mutually independent, then its counterparts $\overline{E_1}, \dots, \overline{E_n}$ are mutually independent.

Using these notions, we can now introduce a rather strong tool!

Theorem 1.14 (Symmetric **Lovász Local Lemma**)**.** Let $E_1, \dots, E_n$ be a set of events and assume

1. $P(E_i) \leq p$ for some fixed p,
2. the maximum degree of the dependency graph of these events is d, and
3. $4dp \leq 1$.

Then $P(\bigcap_{i=1}^{n} \overline{E_i}) > 0$.

The proof is rather technical, but the main idea is to use two nested inductions, and rewrite our result as a fancy product using conditional probabilities.

Proof. Let $S \subseteq \{1, \dots, n\}$. We will show by induction on the size s of S that for all $k \notin S$ it holds

$$P(E_k \mid \bigcup_{j \in S} \overline{E_j}) \leq 2p. \tag{9}$$

For the base case $s = 0$ by assumption $P(E_k) \leq p \leq 2p$.

For the induction step we first need to show $P(\bigcap_{j \in S} \overline{E_j}) > 0$. Again, we use another induction. For $s = 1$ this is clear by assumption. (Notice $p < 1$). Now, w.l.o.g. let $S = \{1, \dots, s\}$. By definition of conditional expectancy,

$$\begin{aligned} P(\bigcap_{j \in S} \overline{E_j}) &= P(\overline{E_s} \mid \bigcap_{j=1}^{s-1} \overline{E_j}) \cdot P(\bigcap_{i=1}^{s-1} \overline{E_j}) \\ &= \underbrace{\left(1 - P(E_s \mid \bigcap_{j=1}^{s-1} \overline{E_j})\right)}_{>0 \text{ by outer induction}} \cdot \underbrace{P(\bigcap_{i=1}^{s-1} \overline{E_j})}_{>0 \text{ by inner induction}} > 0 \end{aligned}$$

This concludes the inner induction, so let us continue with the outer induction step. Let us split S into $S_1 = \{j \in S \mid (k, j) \in E\}, S_2 = S \setminus S_1$. Denote with $F_k = \bigcup_{i=1}^{k} \overline{E_k}$. Then

$$P(E_k \mid F_s) = \frac{P(E_k \cap F_s)}{P(F_s)} = \frac{P(E_k \cap F_{S_1} \cap F_{S_2})}{P(F_{S_1} \cap F_{S_2})} = \frac{P(E_k \cap F_{S_1} \mid F_{S_2})}{P(F_{S_1} \mid F_{S_2})}$$

Consider both parts of the final fraction, and let us bound them:

$$\begin{aligned} P(E_k \cap F_{S_1} \mid F_{S_2}) &\leq P(E_k \mid F_{S_2}) \leq p \\ P(F_{S_1} \mid F_{S_2}) &= 1 - P(\overline{F_{S_1}} \mid F_{S_2}) = 1 - P(\bigcup_{i \in S_1} E_i \mid F_{S_2}) \\ &\geq 1 - \sum_{i \in S_1} P(E_i \mid \bigcap_{j \in S_2} \overline{E_j}) \geq 1 - |S_1| \cdot 2p \geq 1 - 2pd \geq \frac{1}{2} \end{aligned}$$

Applying these results yields $2p$ as an upper bound for $P(E_k \mid F_s)$, which also concludes the outer induction.

Continuing with the actual statement, we get

$$P(\bigcap_{i=1}^{n} \overline{E_i}) = P(\overline{E_n} \mid \bigcap_{i=1}^{n-1} \overline{E_i}) \cdot P(\bigcap_{i=1}^{n-1} \overline{E_i}) = \prod_{i=1}^{n} P(\overline{E_i} \mid \bigcap_{j=1}^{i-1} \overline{E_i}) \geq (1-2p)^n > 0.$$

□

Turning back to our k-SAT problem, we are now able to show a stronger version.

Theorem 1.15. Given a k-SAT instance φ (with n clauses), and assume that no variable appears in more than $\frac{2^k}{4k}$ clauses. Then, $\varphi \in \mathsf{SAT}$.

Proof. We motivate Theorem 1.14. Firstly, we see $P(E_i) \leq 2^{-k}$. For each vertex of the dependency graph its degree is at most $\frac{2^k}{4k} \cdot k$. Therefore, $4 \cdot 2^{-k} \cdot \frac{2^k}{4k} k \leq 1$, and $P(\bigcap_{i=1}^{n} \overline{E_i}) > 0$ implies the existence of a valid assignment. □

Some more applications.

Theorem 1.16. Assume n pairs of vertices need to be connected using n disjoint paths on a given network E. Each pair i can choose from a collection F_i of m paths. If any path in F_i shares edges by at most k paths in F_j and $\frac{8nk}{m} \leq 1$, then we can always choose an edge-disjoint collection of paths.

Proof. Consider a probability space where every pair i chooses a path in F_i uniformly distributed, i.e. with probability $1/m$. We define $E_{i,j}$ as the "bad" event that paths i, j share any edges, which occurs with probability k/m. The degree of the corresponding dependency graph then is $2n$. By Theorem 1.14, we are done.

□ Lecture 3
Mo 22 Apr 2024

2 Random graphs

In this section we want to try answer following question.

Question 2.1. What does the "average" graph look like?

We can introduce two notions for distributions over graphs.

Definition 2.2.
1. The $\mathcal{G}_{n,m}$ model is a probability distribution over $\mathcal{G}$ given by a uniform distribution over all graphs with n nodes and m edges.
2. The $\mathcal{G}_{n,p}$ model is a probability distribution over $\mathcal{G}$ for graphs with n nodes such that the existence of every edge is drawn with probability p.

However, for our goals it is easier to work with the second notion. If $G \sim \mathcal{G}_{n,p}$, then $\mathbb{E}[|E(G)|] = \binom{n}{2}p$.

Lemma 2.3. For all $G \sim \mathcal{G}_{n,p}, G' \sim \mathcal{G}_{n,m}$ it holds that for any graph H

$$P(H = G \mid |E(H)| = m) = P(H = G') \tag{10}$$

Proof. Using some simple transformations and thinking about the probabilities of our graphs we see

$$\begin{aligned} P(H = G \mid E(H) = m) &= \frac{P(H = G \cap E(H) = m)}{P(E(H) = m)} = \frac{P(H = G)}{P(E(H)) = m} \\ &= \frac{p^m(1-p)^{\binom{n}{2}-m}}{\binom{\binom{n}{2}}{m}} p^m(1-p)^{\binom{n}{2}-m} \\ &= \frac{1}{\binom{\binom{n}{2}}{m}} = P(H = G') \end{aligned}$$

□

This gives us the ability to try prove well-known graph problems on random graphs. For example, we can ask outselves if $G \sim \mathcal{G}_{n,p}$ contains K_4. Consider $C \subseteq V(G)$ such that $|C| = 4$, and let X_c be the indicator variable if $G[C] = K_4$. Furthermore, let X be the numbe of 4-cliques in G.

We easily see that $P(X_C = 1) = p^6 = \mathbb{E}[X_C]$, and $\mathbb{E}[X] = \binom{n}{4}p^6 \in \theta(n^4p^6)$. Therefore

- $p << n^{-2/3}$ implies $\mathbb{E}[X] \longrightarrow_{n\to\infty} 0$, and
- $p >> n^{-2/3}$ implies $\mathbb{E}[X] \longrightarrow_{n\to\infty} \infty$.

What happens though in the case of equality, i.e $p(n) = n^{-2/3}$?

Definition 2.4. We call $f(n)$ a **Threshold** for a property Q in $\mathcal{G}_{n,p}$ if

- $p >> f(n)$ implies $P(G \sim \mathcal{G}_{n,p} \text{ has } Q) \longrightarrow_{n\to\infty} 1$,
- $p << f(n)$ implies $P(G \sim \mathcal{G}_{n,p} \text{ has } Q) \longrightarrow_{n\to\infty} 0$.

Let us show $p(n) = n^{-2/3}$ is a threshold for the existence of a 4-clique.

1. Case $p << n^{-2/3}$: Then using Markov's inequality we immediately see

$$P(X > 0) = P(X \geq 1) \leq \frac{\mathbb{E}[X]}{1} \longrightarrow_{n\to\infty} 0. \tag{11}$$

2. Case $p >> n^{-2/3}$: Then using Tschebychev's inequality we see

$$P(X = 0) \leq P(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathbb{V}[X]}{\mathbb{E}[X]^2}. \tag{12}$$

Having a closer at the variance by smart reordering, we conclude

$$\begin{aligned}
\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 &= \mathbb{E}\left[\left(\sum_C X_C\right)^2\right] - \mathbb{E}\left[\sum_C X_C\right]^2 \\
&= \sum_C \left(\mathbb{E}\left[X_C^2\right] - \mathbb{E}[X_C]^2\right) + \sum_{C\neq D} \left(\mathbb{E}[X_C X_D] - \mathbb{E}[X_C]\,\mathbb{E}[X_D]\right) \\
&= \sum_C \mathbb{V}[X_C] + \sum_{C\neq D} \mathbf{Cov}[X_C, X_D]
\end{aligned}$$

It suffices to show that both sums independently tend to 0 for $n \to \infty$ if divided by $\mathbb{E}[X]^2$ as seen in (12).

Notice $\mathbb{V}[X_C] = \mathbb{E}\left[X_C^2\right] - \mathbb{E}[X_C]^2 \leq \mathbb{E}\left[X_C^2\right] = p^6$, so taken over all $\binom{n}{4}$ instances of C the first sum has its upper bound in $\Theta(n^4 p^6) \subseteq \Theta(n^8)$. Since $\mathbb{E}[X]^2 = p^{12} >> n^{-8}$, the first sum converges indeed to 0 for $n \to \infty$.

For the second sum, we need a case distinction over the overlap of C and D. Notice that we only need to consider $\mathbb{E}[X_C X_D] \geq \mathbf{Cov}[X_C, X_D]$.

- $|C \cap D| \leq 1$: Then $\mathbf{Cov}[X_C, X_D] = 0$.
- $|C \cap D| = 2$: Then $\mathbb{E}[X_C X_D] = P(X_C X_D = 1) = p^{11}$. This happens $\binom{n}{6}\binom{6}{4}\binom{4}{2} \in \Theta(n^6)$-times.
- $|C \cap D| = 3$: Then $\mathbb{E}[X_C X_D] = p^9$. This happens $\binom{n}{5}\binom{5}{4}\binom{4}{3} \in \Theta(n^5)$-times.

Analoguously, this concludes the convergence.

Another interesting property is the largest connected component of a random graph.

Theorem 2.5. For $G \sim \mathcal{G}_{n,p}$ it holds that $f(n) \coloneqq \frac{1}{n}$ is a threshold for G having a connected component of size $\Theta(n)$. Furthermore, for $p = \frac{c}{n}$, it holds that

$$\text{Largest connected component is } \begin{cases} \Theta(\log n), & c < 1 \\ \Theta(n^{\frac{2}{3}}), & c = 1 \\ \Theta(n), & c > 1 \end{cases}$$

Lecture 4
We 24 Apr 2024

Theorem 2.6. In almost every $G \sim \mathcal{G}_{n,\frac{1}{2}}$, the largest clique has size approximately $2\log_2(n)$.

Proof sketch. Let X_k be the number of k-cliques in $G \sim \mathcal{G}_{n,\frac{1}{2}}$. As previously shown for 4-cliques, we easily generalize

$$g(k) \coloneqq \mathbb{E}[X_k] = \binom{n}{k} 2^{-\binom{k}{2}}.$$

Let $K_0(n)$ be the largest k such that $g(k) \geq 1$. If we show $K_0(n) \approx 2\log n$, then $g(k) \approx 2^{k\log n - k^2/2}$. Furthermore, let c be a constant (to be determined later), and

$$K_1(n) \coloneqq K_0(n) - c, \qquad K_2(n) \coloneqq K_0(n) + c. \tag{13}$$

We will show that

$$P(X_{K_1(n)} > 0) \xrightarrow{n\to\infty} 1, \tag{14}$$
$$P(X_{K_2(n)} > 0) \xrightarrow{n\to\infty} 0. \tag{15}$$

One can show (we just believe it) that

$$\mathbb{E}[X_{K_1}] \xrightarrow{n\to\infty} \infty, \tag{16}$$
$$\mathbb{E}[X_{K_2}] \xrightarrow{n\to\infty} 0. \tag{17}$$

Apparently following part gives us some intuition for that?

$$\frac{g(k+1)}{g(k)} = \frac{\binom{n}{k+1}2^{-\binom{k+1}{2}}}{\binom{n}{k}2^{-\binom{k}{2}}} = \frac{n-k}{k+1}\cdot 2^{-k}, \tag{18}$$

so for $k \approx 2\log n$ this is approximately

$$\frac{g(k+1)}{g(k)} \approx \frac{n}{2\log n} n^{-2} \xrightarrow{n\to\infty} 0. \tag{19}$$

Using Markov's inequality (**First Moment Method**) we conclude

$$P(X_{K_2(n)} \geq 1) \leq \frac{\mathbb{E}[X_{K_2}]}{1} \xrightarrow{n\to\infty} 0. \tag{20}$$

which shows (15). Using Tschebychev's inequality (**Second Moment Method**) we can show

$$P(X_{K_1(n)} = 0) \leq P(|X_{K_1} - \mathbb{E}[X_{K_1}]| \geq \mathbb{E}[X_{K_1}]) \leq \frac{\mathbb{V}[X_{K_1}]}{\mathbb{E}[X_{K_1}]^2} \xrightarrow{n\to\infty} 1. \tag{21}$$

To show the limit actually holds, we use the same decomposition trick as in : ref

$$\mathbb{V}[X_{K_1}] = \mathbb{V}\left[\sum_S X_S\right] = \sum_S \mathbb{V}[X_S] + \sum_{S\neq D} \mathbf{Cov}[X_S, X_D] \tag{22}$$

Let us define $S \vDash D$ if $|S \cap D| \geq 2$. We simplify further

$$\begin{aligned}
\sum_S \mathbb{V}[X_S] + \sum_{S\neq D} \mathbf{Cov}[X_S, X_D] &= \sum_S \mathbb{V}[X_S] + \sum_{S\vDash D} \mathbf{Cov}[X_S, X_D] \\
&\leq \sum_S \mathbb{E}\left[X_S^2\right] + \sum_{S\vDash D} \mathbb{E}[X_S X_D] \\
&= \sum_S \mathbb{E}[X_S] + \sum_{S\vDash D} \mathbb{E}[X_S X_D] \\
&= \mathbb{E}[X_{K_1}] + \sum_{S\vDash D} \mathbb{E}[X_S X_D].
\end{aligned}$$

Notice that we divide this term by $\mathbb{E}[X_{K_1}]^2$ in (21), so the first summand of previous result reduces to $1/\mathbb{E}[X_{K_1}]$ which tends to 0 by (16).

For the second sum, using even more reordering and conditional probabilities

$$\begin{aligned}
\sum_{S\vDash D} \mathbb{E}[X_S X_D] &= \sum_{S\vDash D} P(X_S = 1 \cap X_D = 1) = \sum_{S\vDash D} P(X_S = 1 \mid X_D = 1) P(X_D = 1) \\
&= \sum_D P(X_D = 1) \left(\sum_{S: S\vDash D} P(X_S = 1 \mid X_D = 1) \right) \\
&= \sum_D P(X_D = 1) \left(\sum_{S: S\vDash D_0} P(X_S = 1 \mid X_{D_0} = 1) \right) \text{ for fixed } D_0 \\
&= \sum_{S: S\vDash D_0} P(X_S = 1 \mid X_{D_0} = 1) \cdot \sum_D P(X_D = 1) \\
&= \sum_{S: S\vDash D_0} P(X_S = 1 \mid X_{D_0} = 1) \cdot \mathbb{E}[X_{K_1}].
\end{aligned}$$

The factor vanishes after dividing by $\mathbb{E}[X_{K_1}]$, so we only remain with one last sum.

$$\sum_{S: S\vDash D_0} P(X_S = 1 \mid X_{D_0} = 1) = \sum_{i=2}^{K_1 - 1} 2^{-\binom{K_1}{2} + \binom{i}{2}} \binom{n - K_1}{K_1 - i} \binom{K_1}{i}$$

Dividing by the second $\mathbb{E}[X_{K_1}]$ using some binomial magic we conclude

$$\frac{\sum_{S:S\vDash D_0} P(X_S = 1 \mid X_{D_0} = 1)}{\mathbb{E}[X_{K_1}]} = \sum_{i=2}^{K_1-1} \underbrace{2^{\binom{i}{2}} \frac{\binom{K_1}{i}\binom{n-K_1}{K_1-i}}{\binom{n}{K_1}}}_{:=f(i)} \leq K_1 \max_{2\leq i\leq K_1-1} f(i) \tag{23}$$

For our previously introduced c, if chosen large enough, then the maximum is reached for $i = 2$, thus

$$f(2) = \frac{K_1!}{(K_1-2)!} \cdot \frac{K_1!}{(K_1-2)!} \cdot \frac{(n-K_1)!(n-K_1)!}{n!(n-2K_1-2)!} \approx \frac{K_1^2 K_1^2}{n^2}$$

and we upper bound (23) by

$$\frac{K_1^5}{n^2} \approx \frac{\log n^5}{n^2} \xrightarrow{n\to\infty} 0$$

showing (14) and concluding this wonderful proof. □

Let us find a notion for randomizing $k-\mathsf{SAT}$.

Conjecture 2.7. For all k there exists a threshold value $r_k^\star \in \mathbb{R}$ such that

$$P(\varphi_k(n,m) \in \mathsf{SAT}) \xrightarrow{n\to\infty} \begin{cases} 0, & r > r_k^\star \\ 1, & r < r_k^\star \end{cases}. \tag{24}$$

Note. We know that $3.52 \leq r_3^\star \leq 4.51$.

Lecture 5
Mo 29 Apr 2024

3 Martingales and Concentration

In this section we study "bounds" of random variables.

Definition 3.1 (**Bounds**)**.** We introduce several notions.

- A **Tail Bound** is an upper bound for $P(X \geq a)$ given that X is non-negative.
- A **Deviation Inequality** is an upper bound on $P(|X-a| \geq \varepsilon)$ where ε usually denotes $\mathbb{E}[X]$.

Let us introduce an important inequality.

Theorem 3.2 (**Azuma's Inequality**)**.** Let X be a martingale with $X_0 = 0$. Suppose that $|X_{n+1} - X_n| \leq 1$ for all n. Then, for any $\lambda > 0$, it holds that

$$P(X_n \geq \lambda) \leq \exp \frac{-\lambda^2}{2n}, \tag{25}$$

and equivalently

$$P(X_n \geq \lambda\sqrt{n}) \leq \exp \frac{-\lambda^2}{2}, \tag{26}$$

Proof sketch. By properties of a Martingale,

$$P(X_n \geq \lambda) \leq P(\exp X_n t \geq \exp t\lambda) \leq \mathbb{E}\left[\exp tX_n\right] \exp -t\lambda. \tag{27}$$

Consider the telescopic augmentation $X_n = \sum_{i=1}^{n}(X_i - X_{i-1})$ and let us apply it to the previous expected value,

$$\mathbb{E}\left[\exp tX_n\right] = \mathbb{E}\left[\underbrace{\exp t \sum_{i=1}^{n}(X_i - X_{i-1})}_{:=F_n}\right] = \mathbb{E}\left[\prod_{i=1}^{n} \exp t(X_i - X_{i-1})\right] \tag{28}$$

$$= \mathbb{E}\left[\prod_{i=1}^{n-1} F_{i-1}\right] \mathbb{E}\left[\exp t(X_i - X_{i-1}) \mid F_{i-1}\right] \tag{29}$$

Notice that $0 \leq \mathbb{E}\left[\exp t(X_i - X_{i-1}) \mid F_{i-1}\right] \leq \exp \frac{t^2}{2}$, so using Lemma 3.3 on this conditional expected value yields

$$\mathbb{E}\left[\exp t(X_i - X_{i-1}) \mid F_{i-1}\right] \leq \exp\left(t\mathbb{E}\left[X_n - X_{n-1} \mid F_{n-1} + \frac{t^2}{2}\right]\right) = \exp\left(\frac{t^2}{2}\right). \tag{30}$$

Notice that we used the Martingale properties in the last step. □

Lemma 3.3 (**Hoeffding's Lemma**)**.** If Y is a random variable bounded with $a \leq Y \leq b$, then

$$\mathbb{E}\left[\exp tY\right] \leq \exp t\mathbb{E}\left[Y\right] + \frac{t^2(b-a)^2}{8}. \tag{31}$$

We omit the proof for brevity. Roughly, the idea is to work with hyperbolic functions.

Notice that this version of Azuma's inequality uses quite restrictive requirements, which we in fact do not need. Instead, let us formulate following generalization which now allows for more dynamic step sizes and arbitrary start values.

Theorem 3.4. If X is a martingale and $|X_{n+1} - X_n| \leq a_n$ for all n, then

$$P(|X_n - X_0| \geq \lambda) \leq 2\exp\left(\frac{-\lambda^2}{2\sum_{i=1}^{n} a_i^2}\right) \tag{32}$$

We omit the proof, but it is quite analoguos.

Let us introduce following notion for continuity.

Definition 3.5 (Lipschitz condition)**.** Given a function $f : \mathbb{R}^n \to \mathbb{R}$. If there exists a $C \in \mathbb{R}$ such that for any $x, y \in \mathbb{R}^n$ whose values only differ in one place holds that

$$|f(x) - f(y)| \leq C \tag{33}$$

we say f satisfies the **Lipschitz Condition**.

The Lipschitz condition intuitively tries to limit the derivative in every point of f. In fact, one can show that this is an even stronger condition than "regular" continuity.

Theorem 3.6 (**McDiarmid's Inequality**)**.** Let X be a stochastic process, and $f : \mathbb{R}^n \to \mathbb{R}$ a function satisfying the Lipschitz condition with Lipschitz constant c. Then, for the **Doob Martingale**[a] defined as

$$Z_k = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid \bigotimes_{i=1}^{k} X_i \right] \tag{34}$$

we have

$$P(|Z_k - Z_0| \geq \lambda) \leq 2 \exp\left(\frac{-2\lambda^2}{kc^2} \right) \tag{35}$$

[a] The fact that Z is indeed a martingale is a classical result.

Again, we omit the proof, since it does not contain any exciting ideas.

Remark 3.7. Most of the time, only the special case $k = n$ is interesting. (Conditional expectancies are hard to calculate!) In this case, the statement (35) simplifies to

$$P(f(X_1, \ldots, X_n) - \mathbb{E}\left[f(X_1, \ldots, X_n)\right] \geq \lambda) \leq \exp\left(\frac{-2\lambda^2}{nc^2} \right). \tag{36}$$

Notice that the absence of the absolute value gets rid of the factor 2 (if needed).

We now turn to some applications of these results. In the context of bioinformatics, DNA sequences are the most important research topic. It might be useful to treat such sequences as random strings, and find interesting patterns in it.

Problem (Pattern matching)**.** Find *interesting* substrings of some given random string following an arbitrary unknown distribution. We say a substring is *interesting* if it occurs more often than it would if the random string followed a uniform distribution.

Let N be the number of occurences of a fixed substring $B = (B_1, \ldots, B_k)$ in the random string $X = (X_1, \ldots, X_n)$ given the underlying alphabet η with $s \coloneqq |\eta|$. If X is drawn

uniformly, we can easily calculate the expected number of occurences of B as

$$\mathbb{E}[N] = \mathbb{E}\left[\sum_{i=1}^{n-k+1} \mathbb{1}(X_i = B_i, \dots, X_{i+k} = B_k)\right] \tag{37}$$

$$= P(X_i = B_i, \dots, X_{i+k} = B_k) = (n-k+1)\frac{1}{s^k}. \tag{38}$$

Let us now treat $N : \mathbb{R}^n \to \mathbb{R}$ as a function. Then $|N(x) - N(y)| \leq k$, and define the Doob martingale $Z_k \coloneqq \mathbb{E}\left[N \mid \bigotimes_{i=1}^{k} X_i\right]$. This allows us to apply Theorem 3.6, and deduce

$$P(|Z_n - Z_0| \geq \lambda) = P(|N - \mathbb{E}[N]| \geq \lambda) \leq 2\exp\left(\frac{-2\lambda}{nk^2}\right). \tag{39}$$

Lecture 6
Mo 06 May 2024

Problem (Balls-into-bins)**.** Suppose we throw m balls into n bins. What is the (expected) number of empty bins?

Let X_i be the bin into which the ith ball falls. Then, for $f : [n]^m \to [n]$ such that f counts the number of empty bins, we see $|f(x) - f(y)| \leq 1$, and thus

$$P(|f(X) - \mathbb{E}[f(X)]| \geq \lambda) \leq 2\exp\left(\frac{-2\lambda^2}{m}\right).$$

Problem (Bin packing)**.** Let $X_1, \dots, X_n$ be random variables with values in $[0, 1]$. What is the (expected) number of bins of capacity 1 needed to pack all X_i?

Define $f : [0,1]^n \to [n]$ such that f maps into the number of bins needed. We see $|f(x) - f(y)| \leq 1$, and thus

$$P(|f(X) - \mathbb{E}[f(X)]| \geq \lambda) \leq 2\exp\left(\frac{-2\lambda^2}{m}\right).$$

Problem (Random knapsack)**.** Let $X_1, \dots, X_n$ be random variables with values in $[0, 1]$. Consider a knapsack of capacity b. A packing of X is a subset of some X_i such that their sum is at most b. We are interested in the *maximal* packing $K : [0,1]^n \to \mathbb{R}$, i.e. the largest sum of X_i's that is still at most b.

McDiarmid applies as above.

Lemma 3.8. Let K_n be the complete graph on n vertices. Let $(A_1, \dots, A_m)$ be a partition of the edges in K_n. Furthermore, let $f : \mathcal{G}_{\setminus} \to \mathbb{R}$ a function defined on the space of all graphs with n vertices, and assume $|f(G) - f(H)| \leq C$ if the symmetric difference $E_G \triangle E_H$ is contained in *one* A_i. Then, for a random graph G on n vertices

$$P(f(G) - \mathbb{E}[f(G)] \geq \lambda) \leq \exp(\frac{-2\lambda^2}{mC^2}).$$

Proof. Trivial by Theorem 3.6. □

Corollary 3.9. If $|f(G) - f(H)| \leq C$ when G, H only differ by one edge, then

$$P(f(G) - \mathbb{E}[f(G)] \geq \lambda) \leq \exp(\frac{-4\lambda^2}{n(n-1)C^2}).$$

Proof. Trivial. □

Corollary 3.10. If $|f(G) - f(H)| \leq C$ when G, H only differ at edges connected to *one* vertex, then we can take $m = n$.

Proof. Use $A_i = \{\{j, i\} \in E \mid j < i\}$. □

Problem. Let G be a graph on n vertices. A coloring of G is a map $c : V_G \to [n]$ such that neighbors are not colored the same. The chromatic number $\chi(G)$ of G is the minimal amount of different colors needed for a valid coloring of G. We are interested in the concentration of the chromatic number.

Corollary 3.10 tells us that

$$P(|\chi(G) - \mathbb{E}[\chi(G)]| \geq \lambda) \leq 2\exp\left(\frac{-2\lambda^2}{n}\right).$$

Theorem 3.11. Let $G \sim \mathcal{G}_{n,\frac{1}{2}}$. Then, following property for the probability of the concentration for the chromatic number holds for every $\varepsilon > 0$:

$$P\left(2\log_2(n)(1-\varepsilon) \leq \chi(G) \leq \frac{n}{2\log_2(n)}(1+\varepsilon)\right) \xrightarrow{n\to\infty} 1.$$

Let us outline our core idea first: Suppose we can find some $s(n)$ such that the probability that $G_{n,p}$ has no independent set of size $s(n)$ is small. Then, we will iteratively color this set, remove it, and apply again $s(n - s(n))$. In fact, we will upper bound the mentioned probability by $q(n) \coloneqq \exp(-n^{4/3})$.

Proof. More detailed, let $\tilde{n} \coloneqq \frac{n}{\log(n)^2}$. We call a set $W \subseteq V_G$ with $|W| \geq \tilde{n}$ *bad*, if it contains no independent set of size $s(\tilde{n})$. We show that the probability that there exists a bad set in $\mathcal{G}_{n,p}$ is upper-bounded by $2^n q(\tilde{n})$, which tends to 0 for $n \to \infty$. The procedure above then lets us iteratively

- color the set of size $s(\tilde{n})$,
- remove those $s(\tilde{n})$ vertices,

until we are left at most $\tilde{n}$ vertices and color them uniquely. This process uses $\frac{n}{s(\tilde{n})} + \tilde{n}$ colors. *Proof continues next lecture.* □

Lecture 7
We 08 May 2024

Proof continued. We claim that $s(n) = (2-\varepsilon)\log(n)$. Then, this would yield

$$\begin{aligned}
\frac{n}{s(\tilde{n})} + \tilde{n} &= \frac{n}{(2-\varepsilon)\log(\frac{n}{\log(n)^2})} + \frac{n}{\log(n)^2} \\
&= \frac{n}{(2-\varepsilon)(\log(n) - 2\log\log(n))} + \frac{n}{\log(n)^2} \\
&= \frac{n}{\log(n)} \left(\frac{1}{\log(n)} + \frac{1}{(2-\varepsilon)(1 - \frac{2\log\log(n)}{\log(n)})} \right) \\
&\xrightarrow{n\to\infty} \frac{n}{\log(n)} \left(0 + \frac{1}{2-\varepsilon} \right) \sim \frac{n}{\log(n)} \left(\frac{1}{2} + \varepsilon \right).
\end{aligned}$$

Let $f(G)$ be the maximum number of independent sets of size at least $s(n)$ in G that (pairwise) share at most one common vertex. Then, $|f(G) - f(H)| \leq 1$ if G, H only differ by 1 edge, since an edge belongs to at most one such independent set, and by Corollary 3.10 we deduce

$$P(f(G) = 0) = P(f(G) - \mathbb{E}[f(G)] \leq -\mathbb{E}[f(G)]) \leq \exp\left(-\frac{4\mathbb{E}[f(G)]^2}{n^2} \right).$$

It remains to show $\mathbb{E}[f(G)] \geq n^{4/3}$. □

Definition 3.12 (**Hamming distance**)**.** Let $X_1, \ldots, X_n$ be random variables in some space E. We call

$$d_H(X, Y) \coloneqq \sum_{i=1}^{n} \mathbf{1}_{X_i \neq Y_i}$$

the Hamming distance of $X, Y \in E^n$. Furthermore, for $A \subseteq E^n$, we define

$$d_H(X, A) \coloneqq \min_{y \in A} d_H(X, Y).$$

Theorem 3.13. For any $\lambda \geq 0$ and $A \subseteq E^n$,

$$P(X \in A) \cdot P(d_H(X, A) \geq \lambda) \leq \exp\left(\frac{-\lambda^2}{2n} \right).$$

Proof. Let $\mu = \mathbb{E}[d_H(X, A)]$. Also, $d_H(\cdot, A) \to \mathbb{N}_0$ has a bounded (element-wise) differ-

ence of 1. By Theorem 3.6, we see

$$P(d_H(X,A) - \mu \geq \lambda) \leq \exp \frac{-2\lambda^2}{n},$$
$$P(d_H(X,A) - \mu \leq -\lambda) \leq \exp \frac{-2\lambda^2}{n}.$$

The latter inequality yields

$$P(X \in A) = P(d_H(X,A) = 0) = P(d_H(X,A) - \mu \leq -\mu) \leq \exp \frac{-2\lambda^2}{n},$$

the former inequality yields

$$P(d_H(X,A) \geq \lambda) = P(d_H(X,A) - \mu \geq \lambda - \mu) \leq \exp \frac{-2(\lambda-\mu)^2}{n}.$$

Now, consider the minimum of these two probabilities. By our upper-bounds,

$$\min P(X \in A), P(d_H(X,A) \geq \lambda) \leq \exp\left(-\frac{2}{n}\max\{\mu, \lambda - \mu\}^2\right).$$

By a case distinction if $\mu \geq \lambda/2$, simple algebra both times yields the wanted upper-bound $\exp(\frac{-\lambda^2}{2n})$. □

This theorem allows us to upper-bound the probability that X is either contained in A or "far away" from A. One "famous" application is the case $E = \{0,1\}$, i.e. X_i adheres to a Bernoulli distribution.

Remark 3.14. The proof also works pretty analogously if we use

$$d_\alpha(X,Y) \coloneqq \sum_{i=1}^{n} \alpha_i \cdot \mathbf{1}_{X_i \neq Y_i}$$

such that $\max_i \alpha_i \leq 1$.

Definition 3.15 (Median)**.** A median of a random variable X is defined as any $m \in \mathbb{R}$ such that

$$P(X \geq m) \geq \frac{1}{2}, \quad P(X \leq m) \geq \frac{1}{2}.$$

Notice that saying $|f(x) - f(y)| \leq C$ for x, y differing in at most one place is equivalent to saying $|f(x) - f(y)| \leq d_\alpha(x,y)$ using $\alpha = (C, \ldots, C)$. We can use the previous result to show a version of McDiarmid on the median.

Theorem 3.16. Let $|f(x) - f(y)| \leq d_\alpha(x, y)$ for some $\alpha = (C, \ldots, C)$. Then, for a random variable X with a median m,

$$P(|f(X) - m| \geq \lambda) \leq 2 \exp\left(\frac{-\lambda^2}{2C^2}\right).$$

Lecture 8
We 15 May 2024

(Proof omitted) We are interested in the number of triangles $f(G)$ of a random graph $G \sim \mathcal{G}_{n,p}$. Its Lipschitz condition number is given as $n - 2$ as one can check easily.

In the following, we want to generalize McDiarmid's theorem to use knowledge of our probabilistic notions. First, suppose some random variable X with values in $\mathbb{R}^n$. We can relax the Lipschitz condition to only hold for $x, y \in A \subseteq \mathbb{R}^n$ such that A satisfies $P(X \in A) = 1$ (but still differing in at most 1 coordinate). Then, $g(x) \coloneqq f(x)\mathbb{1}_{x \in A} + c_0$ (for suitable c_0) still satisfies the original McDiarmid conditions and we can conclude

$$P(f(X) - \mathbb{E}[f(X)] \geq \lambda) = P(g(X) - \mathbb{E}[g(X)] \geq \lambda) \leq \exp\left(\frac{-\lambda^2}{nc^2}\right) \tag{40}$$

by utilizing our distributional knowledge of X.

We can generalize even more and choose $A \subseteq \mathbb{R}^n$ arbitrarily instead. This enables us to split X into two cases depending on if it is included in A, and we can write

$$\begin{aligned} &P(f(X) - \mathbb{E}[f(X)] \geq \lambda) \\ &\quad = P(f(X) - \mathbb{E}[f(X)] \geq \lambda \mid X \in A) P(X \in A) \\ &\qquad + P(f(X) - \mathbb{E}[f(X)] \geq \lambda \mid X \notin A) P(X \notin A) \\ &\quad \leq \exp\left(\frac{-\lambda^2}{nc^2}\right) + P(X \notin A). \end{aligned} \tag{41}$$

Now, let us return to counting triangles. Define for vertices $u, v \in V$ the random variables $X_{u,v}$ denoting the number of common neighbors of u and v, which is binomially distributed according to $Bin(n-2, p^2)$. The Chernoff nound for binomial random variables tells us for $\delta > 0$ that

$$P(X_{u,v} \geq (1 + \delta)\mathbb{E}[X_{u,v}]) \leq \exp\left(\frac{-\mathbb{E}[X_{u,v}]\,\delta^2}{3}\right).$$

This implies in our context of triangle counting that

$$P(X_{u,v} - \mathbb{E}[X_{u,v}] \geq \delta(n-2)p^2) \leq \exp\left(\frac{-(n-2)p^2\delta^2}{3}\right),$$

which, using $\delta = \frac{\lambda}{(n-2)p^2}$ reduces to the more familiar form

$$P(X_{u,v} - \mathbb{E}[X_{u,v}] \geq \lambda) \leq \exp\left(\frac{-\lambda^2}{3(n-2)p^2}\right). \tag{42}$$

We now want to find a non-trivial set A in the sense that (41) gives us a meaningful upper-bound, i.e. the righthand-side should be smaller than 1. Define A as the set of graphs such that all pairs of vertices u, v satisfy $X_{u,v} < \mathbb{E}[X_{u,v}] + \lambda$. Consider the case $G \notin A$. Using (42) and union bounding, we see

$$\begin{aligned} P(G \notin A) &= P(\exists u, v : X_{u,v} \geq (n-2)p^2 + \lambda) \\ &= P\left(\bigcup_{u \neq v} X_{u,v} \geq (n-2)p^2 + \lambda\right) \leq n^2 \exp\left(\frac{-\lambda^2}{3np^2}\right). \end{aligned}$$

If we choose $\lambda = pn^{2+\varepsilon}$ for some small $\varepsilon > 0$, the righthand-side converges to a value smaller 1. Notice our choice of λ implies $\lambda + np^2 \leq \lambda^2 + \lambda$ as a suitable Lipschitz constant. Adding the case $P(G \in A)$, we conclude according to (41) for $\mu > 0$

$$P(f(G) - \mathbb{E}[f(G)] \geq \mu) \leq \exp\left(\frac{-\mu^2}{n(\lambda + \lambda^2)^2}\right) + n^2 \exp\left(\frac{-\lambda^2}{3np^2}\right).$$

Using some confusing blackboard magic and oral remarks we concluded for $\varepsilon > 0$

$$P(f(G) - \mathbb{E}[f(G)] \geq n^{\frac{1}{2}+\varepsilon}) \xrightarrow{n\to\infty} 0, \tag{43}$$

which is a rather strong concentration result for the number of triangles.

Let us summarize the technique we used here as a more general theorem.

Theorem 3.17. Let $X_1, \ldots, X_n$ be i.i.d. random variables and let $\Lambda \subseteq \mathbb{R}$ be such that

1. $P(X_i \notin \Lambda) \leq \exp(-\omega(n^\varepsilon))$ for some $\varepsilon > 0$, and
2. $|f(x) - f(y)| \leq c$ for all $x, y \in \Lambda^n$ that differ in at most one coordinate.

Then

$$P(f(X) - \mathbb{E}[(f(X))] \geq \omega(\sqrt{n})) \xrightarrow{n\to\infty} 0.$$

Lecture 9
We 22 May 2024

Finally, we introduce one more important concentration result. Let us introduce a more generalized distance function first.

Definition 3.18 (**Talagrand's Convex Distance**)**.** We define Talagrand's Convex Distance as the distance function

$$d_T(x, A) \coloneqq \sup_{||\alpha||=1} d_\alpha(x, A).$$

In particular, it holds that

$$d_T(x, A) \geq \frac{1}{\sqrt{n}} d_H(x, A). \tag{44}$$

This effectively means any upper bound on d_T is also an upper bound for Hamming distances. As previously shown, Theorem 3.13 (together with its accompanying remark) yielded some concentration bounds. Building on this, we can generalize to show following result:

Theorem 3.19 (**Talagrand's inequality**)**.** Let $x = (X_1, \dots, X_n)$ be a vector of i.i.d. random variable in $\mathbb{R}$ and $A \subseteq \mathbb{R}^n$. Then for $\lambda \geq 0$

$$P(X \in A)P(d_T(X, A) \geq \lambda) \leq \exp\left(\frac{-\lambda^2}{4}\right).$$

The proof uses similar techniques as for the Hamming distance version, but is mainly just technical, which is why we omit it here. Instead, let us focus on the application of this result: Generally, Talagrand's inequality is effective at concentrating functions h for which $h(x) \geq s$ can be verified by looking at some small subset of $\{X_i\}$.

Definition 3.20. We call a function $h : \mathbb{R}^n \to \mathbb{R}$ is called f-certifiable with regard to a function $f : [n] \to [n]$ if for all $x \in \mathbb{R}^n$ with $h(x) \geq s$ it holds that there is an index set $\mathcal{I} \subseteq [n]$ such that

1. $|\mathcal{I}| \leq f(s)$, and
2. if $y \in \mathbb{R}^n$ is equal to x on $\mathcal{I}$, also $h(y) \geq s$.

Let us first build some intuition for this rather intricate definition:

Example 3.21. Let $h(G)$ be the number of triangles in $G \sim \mathcal{G}_{n,p}$. Then, h is f-certifiable using $f(s) = 3s$.

Proof. If $h(G) \geq s$, then there exist (at least) s triangles, which have at most $3s$ edges in total. We use every such edge to define $\mathcal{I}$ (note that we deliberately not use *every* triangle to still satisfy the bound). Any H which has those same edges also has at least s triangles. □

In fact, this definition allows us prove a more deliberate concentration result.

Theorem 3.22. If h is f-certifiable and 1-Lipschitz continuous, then it holds for any λ, μ

$$P\left(h(X) \leq \mu - \lambda\sqrt{f(\mu)}\right) P(h(X) \geq \mu) \leq \exp\left(-\frac{\lambda^2}{4}\right).$$

Proof. Begin by choosing $A \coloneqq \{x \mid h(x) \leq \mu - \lambda\sqrt{f(\mu)}\}$ as the "bad" set. Suppose $h(y) \geq \mu$. We prove by contradiction that $d_T(y, A) \geq \lambda$. Let $\mathcal{I}$ be the witness set for y, i.e. $|\mathcal{I}| \leq f(s)$, and define $\alpha = (\alpha_1, \dots \alpha_n)$ using $\alpha_i = \frac{\mathbf{1}(i \in \mathcal{I})}{\sqrt{|\mathcal{I}|}}$ (i.e. evenly distributed over

all used values of $\mathcal{I}$). If $d_T(y, A) < \lambda$, then there is a $z \in A$ with $d_\alpha(y, z) < \lambda$. Also, y, z can differ at at most $\lambda\sqrt{|\mathcal{I}|} \leq \lambda\sqrt{f(\mu)}$ coordinates of $\mathcal{I}$. Let the vector w equal to y on indices in $\mathcal{I}$, otherwise equal to z. Notice $h(w) \geq \mu$, and $d_H(w, z) \leq \lambda\sqrt{f(\mu)}$. Therefore, using our Lipschitz condition, it holds that

$$h(z) \leq h(w) - \lambda\sqrt{f(\mu)} \leq \mu - \lambda\sqrt{f(\mu)},$$

which contradicts our assumption and concludes $d_T(y, A) \geq \lambda$. □

Remark 3.23. In applications, it is often useful to choose μ as the median of $h(X)$, because in this case the second factor can be lower-bounded by $\frac{1}{2}$.

Let us apply this result to deduce some concentration bounds:

Problem. Let $X_1, \ldots, X_n \sim \mathcal{U}([0, 1])$, and let $h(X)$ be the length of the longest increasing subsequence[1] of X. Then, h is 1-Lipschitz.

McDiarmid would simply yield

$$P(|h(X) - \mathbb{E}[h(X)]| \leq \sqrt{n} + \omega(1)),$$

which is not good because $\mathbb{E}[h(X)] \in \Theta(\sqrt{n})$ (exercise!), resulting in

$$P(h(X) \geq \mathbb{E}[h(X)] + \omega(1) + \sqrt{n}) \leq \frac{\mathbb{E}[h(x)]}{\mathbb{E}[h(x)] + \omega(1)} \xrightarrow{n\to\infty} 0,$$

i.e. our concentration bound converges to 1. However, we see that h is f-certifiable with $f(s) = s$ by choosing $\mathcal{I}$ as the indices of the largest increasing subsequence. Let m be the median of $h(X)$, then by Talagrand

$$P(h(X) \leq m - \lambda\sqrt{m})P(h(X) \geq m) \leq \exp\left(\frac{-\lambda^2}{4}\right),$$

and as stated in Remark 3.23 plus by choosing $\mu := m + \lambda\sqrt{m}$, we conclude

$$P(|h(X) - m| \geq \lambda\sqrt{m}) \leq 4\exp\left(\frac{-\lambda^2}{4}\right).$$

One can show $m \in \Theta(\sqrt{n})$, which implies

$$P(|h(X) - m| \geq \omega(n^{1/4})) \xrightarrow{n\to\infty} 0.$$

Lecture 10
Mo 27 May 2024

[1] remember that subsequence do not need to be consecutive

4 Monte Carlo Sampling

We start this section with the problem of approximating π using probabilistic methods. Following algorithm yields a simple Monte Carlo-based approach by checking if a uniformly generated point lies inside a quarter of a circle:

Algorithm 2: Approximate π

for $i = 1, \dots, m$ **do**
  $(X_i, Y_i) \sim [0,1]^2$
  $A_i \leftarrow \mathbf{1}(X_i^2 + Y_i^2 < 1)$
end
return $\tilde{\pi} \leftarrow \frac{4}{m} \sum_{i=1}^{m} A_i$

While it is clear that this approach can approximate π infinitely good, it is more interesting *how fast* it will approximate π up to a certain precision for a large enough probability. Following notion formalizes the preceding statement.

Definition 4.1. We say that $X \in \mathbb{R}$ is an (ε, δ)-approximation for $V \in \mathbb{R}$ if $P(|X - V| \leq \varepsilon V) \geq 1 - \delta$

Since we work in the context of algorithms, we also introduce a new complexity class that is useful in this domain.

Definition 4.2 (**Fully Polynomial Randomized Approximation Scheme**)**.** A fully polynomial randomized approximation scheme (short: **FPRAS**) for a function f on an input x and parameters $0 < \varepsilon, \delta < 1$ is an algorithm that returns in time $poly(\frac{1}{\varepsilon}, \ln(\frac{1}{\delta}), |x|)$ an (ε, δ)-approximation for $f(x)$.

Then, for our approximation of π, we can conclude using Theorem 3.6

$$\begin{aligned} P\left(|\pi - \tilde{\pi}| \geq \varepsilon\pi\right) &= P\left(\left|\pi - \frac{4\sum_i X_i}{m}\right| \geq \varepsilon\pi\right) \\ &= P\left(\left|\frac{m\pi}{4} - \sum_i X_i\right| \geq \varepsilon\frac{m\pi}{4}\right) \\ &= P\left(\left|\mathbb{E}\left[\sum_i X_i\right] - \sum X_i\right| \geq \varepsilon\mathbb{E}\left[\sum_i X_i\right]\right) \leq 2\exp\left(\frac{-\mu\varepsilon^2}{3}\right) =: \delta, \end{aligned}$$

which yields for corresponding choice of μ

$$\exp\left(\frac{-m\pi\varepsilon^2}{12}\right) \leq \frac{\delta}{2} \iff m \geq \frac{12\ln(\frac{2}{\delta})}{\pi\varepsilon^2}. \tag{45}$$

This is polynomial in $\varepsilon^{-1}, \ln(\delta^{-1}), |x|$, and therefore in **FPRAS**!

.. Monte Carlo approaches also enables us to approximate values for (difficult) counting problems. In general, we will use following approach:

counting stuff

1. Define a set $|S| = V$ for the value V we want to approximate.
2. Define a sample space Ω from which we sample.
3. Estimate $\frac{|S|}{|\Omega|} =: \hat{r}$ by generating samples.
4. Return $\hat{r}|\Omega|$.

A famous example where we can apply this approach is #SAT, i.e. determing the number of valid assignments for a boolean CNF φ with n variables. Denote with $C(\varphi)$ said count of valid assignments, and notice that

$$C(\neg\varphi) = 2^n - C(\varphi), \tag{46}$$

i.e. $\neg\varphi$ is the equivalent problem stated in DNF. We will continue to work with DNF.

Consider following helpful result first.

Lemma 4.3. Let $X_1, \dots, X_m$ be i.i.d. $0-1$ random variables with $\mu = \mathbb{E}[X_i]$. If $m \geq \frac{3\ln(\frac{2}{\delta})}{\varepsilon^2\mu}$, then

$$P\left(\left|\frac{1}{m}\sum_i X_i - m\mu\right| \geq \varepsilon\mu\right) \leq \delta.$$

Proof. It holds that

$$P\left(\left|\sum_i X_i - m\mu\right| \geq \varepsilon m\mu\right) \leq 2\exp\left(\frac{-m\mu\varepsilon^2}{3}\right) \leq \delta,$$

which implies the lemma. □

We can state following algorithm for DNF-based #SAT.

Algorithm 3: Approximate $C(\varphi)$

```
for i = 1, ..., t do
|   SC_i ← {(i, a) : a satisfies clause i}
end
Ω ← {(i, a) : a satisfies clause i}
|Ω| ← Σ_i |SC_i|
C(φ) = |⋃ SC_i|
return ??
```

We can show that $\frac{1}{t}$ is a lower bound for $\mathbb{E}[X_i]$, which implies the need for

$$m \geq \frac{3t\ln(\frac{2}{\delta})}{\varepsilon^2} \tag{47}$$

samples to get a (ε, δ)-approximation. Again, this is a **FPRAS** algorithm!

Lecture 11
We 29 May 2024

Definition 4.4 (**Almost Uniform Sampler**)**.** An almost uniform sampler is an algorithm for a set S that outputs a random element of S with probability distribution A such that

$$||A(S) - \mathcal{U}(S)||_V \leq \delta$$

for $\delta \in (0, 1)$.

We call it a **Fully Polynomial Almost Uniform Sampler** (**FPAUS**) if its runtime is in $poly(|x|, \log(1/\delta))$ (using $|x|$ as decoded size of x).

Define with $\mathcal{M} : \mathcal{G} \to \mathcal{P}(E(G))$ a function that outputs all matchings of a graph G. We want to estimate $\mathcal{M}(G)$ for general graphs. Consider following trick, which uses $G_i \coloneqq (V(G), \{e_1, \ldots, e_i\})$ as incrementally growing subgraphs (with G_0 as the graph with no edges). Then, we can rephrase our desired value as a telescopic product

$$|\mathcal{M}(G_m)| = \underbrace{\frac{|\mathcal{M}(G)_m|}{|\mathcal{M}(G)_{m-1}|}}_{\coloneqq p_m^{-1}} \cdot \ldots \cdot \underbrace{\frac{|\mathcal{M}(G)_1|}{|\mathcal{M}(G)_0|}}_{\coloneqq p_1^{-1}} \cdot \underbrace{|\mathcal{M}(G)_0|}_{=1}. \tag{48}$$

In particular, we introduce the ratios p_i. Notice that every added edge accounts for up to two new (distinct) matchings per old matching, therefore

$$1 \geq p_i \geq \frac{1}{2}. \tag{49}$$

Since we now try to apply Monte Carlo Sampling for p_i, this ensures us that the sample space size does not explode, which in the case for #SAT restricted us from using a simple sampling approach.

Lemma 4.5. Let G be a graph and consider its matchings $\mathcal{M}(G)$. Then there is an **FPAUS**-algorithm for $\mathcal{M}(G)$ exactly iff there is an **FPRAS**-algorithm for $|\mathcal{M}(G)|$.

Proof of $\to$. Assume we use **FPAUS** to estimate p_i. We will show that

$$P\left(\left|\prod_i \overline{Z}_i - \prod_i p_i\right| \leq e^{\varepsilon} \prod_i p_i\right) \geq \frac{3}{4}. \tag{50}$$

Set $\delta = \frac{\varepsilon}{6m}$ and let $Z_i^{(j)}$ be the indicator variable for the case if the jth sample is in $\mathcal{M}(G_{i-1})$. Then, $\overline{Z}_i = \sum_{i=1}^{s} \frac{Z_i^{(j)}}{s}$. Since $Z_i^{(j)}$ is sampled almost uniformly by assumption, we know that in the worst case

$$p_i - \frac{\varepsilon}{6m} \leq \mu_i \leq p_i + \frac{\varepsilon}{6m}.$$

Applying $p_i \geq \frac{1}{2}$ gives the estimates

$$p_i(1 - \frac{\varepsilon}{3m}) \leq \mu_i \leq p_i(1 + \frac{\varepsilon}{3m}).$$

Using $e^x \geq 1+x$ and $e^{-\frac{x}{k+1}} \leq 1 - \frac{x}{k}$ for $0 \leq x \leq 1$ yields

$$\exp\left(-\frac{\varepsilon}{3m}\right) p_i \leq \mu_i \leq \exp\left(\frac{\varepsilon}{3m}\right) p_i.$$

By Chernoff, we can show that $s \geq 1296 \frac{m^2}{\varepsilon^2} \ln(m) c$. However, we can show a better lower bound of $s > \frac{75m}{\varepsilon^2}$:

$$\mathbb{V}[Z_i] = \mathbb{E}\left[(Z_i - \mu_i)^2\right] = P(Z_1 = 1)(1-\mu_i)^2 + P(Z_i = 0)\mu_i^2 = \mu_i(1-\mu_i)$$

$$\frac{\mathbb{V}\left[\overline{Z}_i\right]}{\mu_i^2} \leq \frac{\sum_j \mathbb{V}\left[Z_i^{(j)}\right]}{\mu_i^2} \leq \frac{2}{5} \leq \frac{\varepsilon^2}{37m}$$
$$\frac{\mathbb{V}\left[1-\overline{Z}_i\right]}{(1-\mu_i)^2} \leq \frac{\sum_j \mathbb{V}\left[Z_i^{(j)}\right]}{\mu_i^2} \leq \frac{2}{5} \leq \frac{\varepsilon^2}{37m}$$

In summary,

$$\begin{aligned}\frac{\mathbb{V}\left[\prod_i \overline{Z}_i\right]}{\prod_i \mu_i^2} &= \frac{\prod_i \mathbb{E}\left[\overline{Z}_i^2\right]}{\prod_i \mu_i^2} - \frac{\prod_i \mathbb{E}\left[\overline{Z}_i\right]^2}{\prod_i \mu_i^2} = \prod_i \left(1 + \frac{\mathbb{V}\left[\overline{Z}_i\right]}{\mu_i^2}\right) - 1 \\ &\leq \left(1 + \frac{\varepsilon^2}{37m}\right)^m - 1 \leq \exp\left(\frac{\varepsilon^2}{37m}\right) - 1 \leq \frac{\varepsilon^2}{36}.\end{aligned}$$

Finally, using Chebyshev's inequality,

$$P\left(\left|\prod_i \overline{Z}_i - \prod_i \mu_i\right| \geq \frac{\varepsilon}{3} \prod_i \mu_i\right) \leq \frac{\varepsilon^2}{36} \frac{3^2}{\varepsilon^2} = \frac{1}{4}.$$

□

Lecture 12
Mo 03 June 2024
sick...
Lecture 13
We 05 June 2024
sick...
Lecture 14
Mo 10 June 2024

Theorem 4.6 (Path Coupling)**.** Let Z be a Markov chain on a graph G that can only move between adjacent states. If (X, Y) is a coupling of Z (for X, Y adjacent) with $\mathbb{E}\left[d(X_i, Y_i)\right]_{X,Y} \leq p$ where d is the graph distance and $p < 1$, then

$$\tau(\varepsilon) \leq \frac{1}{1-p} \ln\left(\frac{\operatorname{diam}(G)}{\varepsilon}\right).$$

Proof. Use $D \coloneqq \operatorname{diam}(G)$. The key fact of the proof is that

$$(x, y) \mapsto \mathbb{E}\left[d(X, Y)\right]_{X,Y} \tag{51}$$

is a *metric* for X, Y following any distribution on V. Let X, Y be any vertices on V, and let $x = v_0, \ldots, v_k, v_{k+1} = y$ be any path from x to y in G. Then by assumption

$$\mathbb{E}\left[d(x_1, y_1)\right]_{X,Y} \leq \sum_{i=0}^{k} \mathbb{E}\left[d(x_i, v_i)\right]_{v_i, v_{i+1}} \leq \sum_{i=0}^{k} p = p(k+1).$$

Since the path from x to y was arbitrary, we just take the shortest. Then,

$$\mathbb{E}\left[d(X_1, Y_1)\right]_{x,y} \leq pd(x, y), \tag{52}$$

and by Markov's inequality

$$P_{X,Y}(X_1 \neq Y_1) = P_{X,Y}(d(X_1, Y_1) > 0) \leq \mathbb{E}\left[d(X_1, Y_1)\right]_{X,Y} \leq pd(X, Y).$$

Thus by Markov property

$$\begin{aligned} P_{x,y}(X_t \neq Y_t) \leq \mathbb{E}\left[d(X_t, Y_t)\right]_{x,y} &= \mathbb{E}\left[\mathbb{E}\left[X_t \neq Y_t \mid X_{t-1}, Y_{t-1}\right]_{x,y}\right]_{x,y} \\ &= \mathbb{E}\left[\mathbb{E}\left[d(X_1, Y_1)\right]_{X_{t-1}, Y_{t-1}}\right]_{X,Y} \\ &\overset{(52)}{\leq} p \cdot \mathbb{E}\left[d(X_{t-1}, Y_{t-1})\right]_{X,Y} \leq \cdots \leq p^t d(x, y) \leq p^t D. \end{aligned}$$

If $p^t D \leq \varepsilon$, then $\tau(\varepsilon) \leq t$, so using $e^x \geq 1 + x$

$$\exp\left(-(1-p)t\right) \leq (1-(1-p))^t = p^t \leq \frac{\varepsilon}{D}$$

and

$$t \geq \frac{1}{1-p} \ln\left(\frac{D}{\varepsilon}\right).$$

□

Example 4.7. Let us consider some coupled Markov chains.

1. Let Z be a simple random walk on K_n. Then (X, Y) always land on the same state except if X chose the position of Y (which happens with probability $1/n$). So, the hitting time behaves like a geometric distribution with parameter $1/n$.
2. Let Z be a simple random walk on a complete bipartite graph with self-loops. Then X, Y behave independently, except if they lie on the same side of the bipartite graph. Only then they start to choose the same vertex.

Lecture 15
We 12 June 2024

5 Sources of randomness

Sources of randomness are crucial for the implementation of *random algorithms*. Typical applications are

- Markov Chain Monte Carlo (MCMC)
- simulation of distributions (i.e. by using a uniform distribution)
- cryptographic applications

Theoretic analysis of these assumes a *perfectly random* source (usually uniform). In reality, surces of randomness are *weak* or highly imperfect. **Randomness Extractors** aim to shrink the gap between theory and practice; take as input

1. a weak source of randomness satisfying as few assumptions as possible,
2. a small *truly random* (i.e. truly uniform) seed,

and output an (almost) uniform random variable.

Entropy is a way of measuring *how random* a distribution is. We shall use it to characterize *weakness* of random sources. There are different established definitions of entropy. Notice that for the remainder of this section we denote with log the binary logarithm $\log_2$ as this is the usual notion in the context of information theory.

Definition 5.1 (**Shannon Entropy**)**.** Let μ be a probability distribution on a countable set E. Then we define

$$H(\mu) \coloneqq \sum_{x \in E} \mu(x) \cdot \log\left(\frac{1}{\mu(x)}\right) = -\mathbb{E}\left[\log(\mu(X))\right]$$

(assuming $X \sim \mu$ for the latter version) as the *Shannon entropy*.

Shannon entropy was originally established in the context of information theory and tries to quantify information density. In fact, randomness and information are tightly coupled; generally, the more *random* an information source is, the less *useful* information it contains; this is reflected in H: the larger its value, the less information (or *structure*) is encoded in μ.

Example 5.2. Consider $E = \{0, 1\}$. Intuition says $\mu \sim Ber(\frac{1}{2})$ is the *most random* distribution on E. If we use Shannon entropy to quantify this intuition, we get

$$H(\mu) = \frac{1}{2}\log(2) + \frac{1}{2}\log(2) = \log(2).$$

More generally, for any $E, |E| < \infty$ and $\mu \sim \mathcal{U}(E)$ we can deduce

$$H(\mu) = \log(|E|)$$

In fact, we can prove that the entropy is maximized in such cases, which adheres to our notion of *most random*. First, we introduce a helpful result.

Definition 5.3 (**Relative Entropy**)**.** Let μ, η be two probability distributions on a countable set E. Then we define

$$H(\mu||\eta) = \sum_{x \in E} \mu(x) \log\left(\frac{\mu(x)}{\eta(x)}\right) = -\mathbb{E}\left[\log\left(\frac{\eta(X)}{\mu(X)}\right)\right]$$

(assuming $X \sim \mu$ for the latter version) as the *relatie entropy*. Literature also refers to it as the **Kullback-Leibler Divergence**.

Lemma 5.4. The relative entropy satisfies $H(\mu||\eta) \geq 0$ with equality iff $\mu = \eta$.

Proof. First, using $X \sim \mu$ and Jensen's inequality (notice log is strictly concave)

$$H(\mu||\eta) = \sum_{x \in E} \mu(x) \log\left(\frac{\mu(x)}{\eta(x)}\right) \geq -\log\left(\sum_{x \in E} \mu(x) \frac{\eta(x)}{\mu(x)}\right) = -\log(1) = 0.$$

Equality holds if and only if $\frac{\eta(x)}{\mu(x)}$ are equal for all $x \in E$, i.e. $\eta = \mu$. □

Now following result is quite easy to deduce

Theorem 5.5. Let E be finite.

(i) $H(\mu)$ is *maximal* if $\mu \sim \mathcal{U}(E)$. In particular, $H(\mu) = \log(|E|)$.

(ii) $H(\mu)$ is *minimal* if $\mu = \delta_x$ for any $x \in E$. In particular, $H(\mu) = 0$.

Proof. The second part is trivial since $H(\mu) = 1 \cdot \log(1) = 0$ and $H(\mu) \geq 0$, since any finite probability measure μ' satisfies $0 \leq \mu' \leq 1$.

For the first part we use relative entropy. Let μ be any distribution over E and $\eta \sim \mathcal{U}(E)$. It holds by Lemma 5.4 that

$$\begin{aligned} 0 \leq H(\mu||\eta) &= \sum_{x \in X} \mu(x) \log\left(\frac{\mu(x)}{\eta(x)}\right) \\ &= -\sum_{x \in X} \mu(x) \log(\mu(x)^{-1}) + \sum_{x \in X} \mu(x) \log(\eta(x)^{-1}) \\ &= -H(\mu) + \log(|E|). \end{aligned}$$

Therefore, $H(\mu) \leq \log(|E|)$ with equality iff $\mu = \eta$. □

Shannon entropy measures the amount of randomess *on average* of a distribution. However, sometimes we want to consider the *worst case* randomness.

Definition 5.6 (**Min-Entropy**)**.** Let μ be a probability distribution on a countable set E. We define the *min-entropy* of μ as

$$H_\infty(\mu) := \min_{x \in E} \log\left(\frac{1}{\mu(x)}\right).$$

Note that $H_\infty(\mu) \geq k$ iff $\mu(x) \leq 2^{-k}$ for all $x \in E$. So, if $E = \{0,1\}^k$, the uniform distribution clearly has the largest min-entropy.

Turning back to our random number generation, suppose we want to build a seed-less extractor that takes as input any random variable on $\{0,1\}^n$ and should output $\mathcal{U}(\{0,1\})$. Assume that our input X satisfies the rather strict condition

$$H_\infty(X) \geq n - 1, \tag{53}$$

i.e. no bitstring occurs with probability higher than 2^{n-1}. Therefore, any generator function $f : \{0,1\}^n \to \{0,1\}$ must satisfy one of

$$|\{x \in \{0,1\}^n \mid f(x) = b\}| \geq 2^{n-1}$$

for $b \in \{0,1\}$. We now try to construct an adversary input variable: Let A_b be the set of inputs mapping to the same output that satisfies previous requirement, and let $Y \sim \mathcal{U}(A_b)$. Then $H_\infty(Y) \geq n - 1$ since $f(A_b) = b$ and $|A_b| \geq 2^{n-1}$, which shows that even with a quite strict condition we cannot construct a seedless generator. We will now try to introduce a seed in order to improve our number generation.

Lecture 16
Mo 17 June 2024

Definition 5.7 (**Random Extractor**)**.** An algorithm $\xi : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is called an *extractor* of error $\varepsilon > 0$ if for all random variables X in some class it holds (using binary uniform variables U_x of dimension x) that

$$||\xi(X, U_d) - U_m||_{TV} < \varepsilon.$$

We call the first input part the *weak source* and the second part the *seed*. If the class of inputs is X such that

$$H_\infty(X) \geq k,$$

then ξ is called a (k, ε)-*extractor*.

We can show that there always exist a random extractor under certain conditions.

Lemma 5.8. Let $d = \log \frac{n}{\varepsilon^2}, m = k + \log(n)$. Then there exists a (k, ε)-extractor.

The proof is using the probabilistic method, however notice that this means it is non-constructive and does not yield a method to generate an actual extractor. We skip the proof for reasons of brevity.

Let us take a look at the seed length: The case $d \in O(\log\left(\frac{n}{\eta}\right))$ is important for a few reasons.

Definition 5.9 (**Bounded-error Probabilistic Polynomial Time Algorithm**)**.** We call a random algorithm $A : W \times Z \to N$ that is allowed to generate random bit sequences Z a *BPP-algorithm* if it computes a goal function $f : W \to N$ such that

1. it runs in polynomial time,
2. outputs the correct solution $A(w, z) = f(z)$ with probability over $2/3$.

Consider a BPP algorithm A for a goal f. Extractors allow us to simulate such algorithms in the following way:

(i) Take input w and instance of weak source $X \in \{0,1\}^n$.

(ii) Compute $\xi(x, y)$ for all $y \in \{0,1\}^d$.

(iii) Compute $A(w, \xi(x, y))$ for all $y \in \{0,1\}^d$.

This has a runtime of 2^d, therefore we can guarantee polynomial runtime if $d \in O(\log n)$.

Another observation arises in following construct: For a random variable $X \in \{0,1\}^n$ and any function $f : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ it can be shown that $||\xi(X, U_d) - U_m||_{TV} < \varepsilon$ implies that $||X - Y||_{TV} < \varepsilon$ for some Y with $H_\infty(Y) \geq m - d - 1$. If we assume that ξ is a (k, ε)-extractor, then $H_\infty(X) \geq k \approx H_\infty(Y)$ implies $d \gtrsim m - k - 1$. If $m = k + \log(n)$, then even $d \gtrsim \log(n)$, once again yielding a logarithmic bound. We will show how we can construct an averaging sampler using an extractor.

Lecture 17
We 19 June 2024
sick...
Lecture 18
Mo 24 June 2024

Theorem 5.10. Let $\xi : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be an (ε, k)-extractor. Also, let $G : \{0,1\}^n \to (\{0,1\}^m)^{2^d}$ with $G(x) = \xi(x, u)$ for $u \in \{0,1\}^d$. Then, G is an averaging sampler with accuracy ε and error $\delta = 2^{1+k-n}$.

Proof. Suppose for contradiction that there exists a function $f : \{0,1\}^m \to [0,1]$ such that with probability larger δ it holds that

$$\left| \frac{1}{2^d} \sum_{i=1}^{2^d} f(\xi(X, u_i)) - \mathbb{E}\left[f(U_m)\right] \right| > \varepsilon \tag{54}$$

for some random variable X with $H_\infty(X) \geq k$. Then, there are $\delta \cdot 2^n$ strings $x \in \{0,1\}^n$ for which (54) holds (replacing X by x).

Without loss of generality, for at least half of these x it holds that[2]

$$\sum_{i=1}^{2^d} f(\xi(x, u_i)) > 2^d(\mathbb{E}\left[f(U_m)\right] + \varepsilon). \tag{55}$$

Let B be the set of such x. It has size $|B| \geq \delta \cdot 2^{n-1} = 2^k$. Then, the uniform distribution Y on $|B|$ has min-entropy k. Since ξ is an extractor, by definition

$$||\xi(Y, U_d) - U_m||_{TV} < \varepsilon,$$

thus as a natural property of the total variational distance

$$||f(\xi(Y, U_d)) - f(U_m)||_{TV} < \varepsilon,$$

[2] otherwise consider the other direction of the inequality

and by some measure theory even

$$|\mathbb{E}\left[f(\xi(Y, U_d))\right] - \mathbb{E}\left[f(U_m)\right]| < \varepsilon.$$

But, for said choice of Y,

$$\mathbb{E}\left[f(\xi(Y, U_d))\right] > \mathbb{E}\left[f(U_m)\right] + \varepsilon.$$

So, ξ is *not* a (k, ε)-extractor - contradiction, G must be an averaging sampler! □

What remains to demonstrate is a way to explicitly construct an extractor, which we will do using an averaging sampler. First, let us deepen our understanding on what the *seed* is supposed to do. Recall that we initially gave some intuition why a meaningful seedless extractor cannot exist in general. We wanted to solve this problem by introducing a random seed as additional input; however, let us now think of the seed as a way of randomly choosing a function f to evaluate.

Lemma 5.11. Suppose that we can randomly draw an f from a set F of functions $f : \{0,1\}^{\log(n)} \to \{0,1\}^k$ such that $||f - U_F||_{TV} < \varepsilon$. Then, using bitstrings of length $\log(n)$,

$$\xi : \{0,1\}^{n \log n} \times F \to \{0,1\}^{nk},$$
$$X \coloneqq [X_1, \ldots, X_n], f \mapsto [f(X_1), \ldots, f(X_n)]$$

is an extractor with error $\varepsilon \cdot n$.

Proof. We use induction to show $[f(X_i)]$ is close to $[U_k]$. Let

$$Z_i \coloneqq [f(X_1), \ldots, f(X_i), V_{i+1}, \ldots, V_n]$$

using independent $V_i \sim \mathcal{U}(\{0,1\}^k)$. Then

$$||Z_0 - U_{nk}||_{TV} = 0.$$

As inductive assumption, suppose that

$$||Z_i - U_{nk}||_{TV} \leq i \cdot \varepsilon.$$

Then, by triangle inequality

$$||Z_{i+1} - U_{nk}||_{TV} \leq ||Z_{i+1} - Z_i||_{TV} + \underbrace{||Z_i - U_{nk}||_{TV}}_{< i \cdot \varepsilon}.$$

For the remaining part, the only source of variation stems from

$$||f(X_i) - V_i||_{TV} = ||f(X_i) - f_n(X_i)||_{TV} \leq \varepsilon,$$

which finishes the proof. □

If now we can construct a way to choose a random element of F, we can use this lemma to finish our extractor construction. Notice that any $f : \{0,1\}^{\log(n)} \to \{0,1\}^k$ can be represented as a bitstring in $\{0,1\}^{2^{\log(n)} \cdot k} = \{0,1\}^{nk}$. We saw that as long as $n \leq 2^k$ we can approximately sample elements of F with only (roughly) $O(k)$ independent coin flips. This "approximation" is a problem, since pairwise independence is *not* close enough. However, there is a way to explicitly draw f close to uniform using $O(k)$ true bit flips. It uses the averaging sampler and "refines" it via so-called *expander graphs*.

Thus, the choice of $k = \log(n)$ allows us to extract an (almost) uniform random variable of the same order as our weak input, using significantly fewer true random bit flips.

Part II
Appendix

A Exercise sheets

1. Exercise Sheet

Exercise 1.1. Given a balanced graph H with $v \coloneqq |V(H)|, e \coloneqq |E(H)|$. Show that $n^{v/e}$ is a threshold for the property if H is contained in a random graph $G \sim \mathcal{G}_{n,p}$.

Proof. For every subset $S \subseteq V(G)$ with $|S| = v$ let X_S be the number of containments of H in S, and X be the total number of containments. The probability for a specific containment is given as p^e. However, for every H independent of n, p there is a specific number of times c_H it can appear in such an subset S. Therefore,

$$\mathbb{E}[X] = \sum_{S \subseteq V(G)} \mathbb{E}[X_S] = \binom{n}{v} c_H p^e.$$

By Markov's inequality,

$$P(X > 0) = P(X \geq 1) \leq \frac{\mathbb{E}[X]}{1} = \binom{n}{v} c_H p^e \in O(n^v p^e)$$

If $p << n^{-v/e}$, then $\binom{n}{v} c_H p^e \xrightarrow{n \to \infty} 0$, proving one part of the threshold property.

For the other direction, consider the variance of X. Let H_i be the indicator for every possible containment of H in G. Then, the sum over all H_i is also X, i.e. it suffices to consider all pairs H_i and H_j for their covariance. Notice the covariance only depends on the number of overlapping (existing) edges. Let $\hat{H} = H_i \cap H_j$ be the cross-sectional graph with edges only existing if they are in both graphs. Then,

$$\begin{aligned} \mathbf{Cov}[H_i, H_j] &= \mathbb{E}[H_i H_j] - \mathbb{E}[H_i]\,\mathbb{E}[H_j] \\ &= p^{e+e-|E(\hat{H})|} - p^{2e}. \end{aligned}$$

Also, consider how often each specific overlap $\hat{H}$ could occur. The number of involved nodes is by inclusion-exclusion principle $v+v-|V(\hat{H})|$, therefore the pattern resulting in $\hat{H}$ could occur $\binom{n}{2v-|V(\hat{H})|}$-times. Finally, we again introduce a constant $c_{H,\hat{H}}$ denoting the number of ways two copies of H intersect in $\hat{H}$. Thus, combining everything we conclude

$$\mathbb{V}[X] = \sum_i \mathbb{V}[H_i] + \sum_{i,j} \mathbf{Cov}[H_i, H_j] \tag{56}$$

$$= \sum_{\hat{H} \subseteq H} c_{H,\hat{H}} n^{2v-|V(\hat{H})|} p^{2e-|E(\hat{H})|}. \tag{57}$$

Notice that variance is just a special form of covariance. Applying Chebyshev's

inequality we get

$$P(X=0) = P(|X - E(X)| = E(X)) \leq \frac{\mathbb{V}[X]}{\mathbb{E}[X]^2}$$
$$= \sum_{\hat{H} \subseteq H} c_{H,\hat{H}} n^{-|V(\hat{H})|} p^{-|E(\hat{H})|}.$$

Since H is balanced, every subgraph $\hat{H}$ must satisfy $e/v \geq |E(\hat{H})|/|V(\hat{H})|$ by definition. Therefore, if $p >> n^{-v/e}$, then also $p >> n^{-|V(\hat{H})|/|E(\hat{H})|}$, and $p^{-|E(\hat{H})|} << n^{|V(\hat{H})|}$, and the previous upper bound tends to 0 for $n \to \infty$. This proves the other condition for the threshold. □

Exercise 1.2. Given an undirected graph $G = (V, E)$. Let $d \in \mathbb{N}$ and $U \subseteq V$ the set of all vertices with degree at least d. Then there exists a dominating set for U of size at most

$$\left\lfloor n \frac{\log(d+1)+1}{d+1} \right\rfloor.$$

Proof. Define $p \coloneqq \frac{\log(d+1)}{d+1}$ and randomly choose a subset $S \subseteq V$ such that every node is chosen independently with probability p. Then, add a set $|T|$ that contains every node $u \in U$ if u is not already dominated by S. Clearly, $S \cup T$ is a dominating set then. Let T_u be the indicator variable if $u \in T$. Consider the expected number of nodes in our final set

$$\mathbb{E}[|S \cup T|] = \mathbb{E}[|S|] + \mathbb{E}\left[\sum_u T_u\right] = \mathbb{E}[|S|] + |U| P(T_u = 1).$$

Notice $|S|$ follows a binomial distribution $\mathcal{B}_{n,p}$. Also, $P(T_u = 0)$ denotes the probability that u is not dominated by S, i.e. none of its $\deg(u) \geq d$ neighbors or itself was chosen, resulting in the probability $P(T_u = 0) = (1-p)^{d+1}$. In summary,

$$\begin{aligned}
\mathbb{E}[|S|] + |U| P(T_u = 1) &= np + |U|(1-p)^{d+1} \\
&\leq n(p + (1-p)^{d+1}) = n\left(\frac{\log(d+1)}{d+1} + \left(1 - \frac{\log(d+1)}{d+1}\right)^{d+1}\right) \\
&\leq n\left(\frac{\log(d+1)}{d+1} + e^{-\log(d+1)}\right) = n\frac{\log(d+1)+1}{d+1}.
\end{aligned}$$

Notice that we utilized Euler's inequality $e^x > (1 + \frac{x}{n})^n$ in the last line using $x = -\log(d+1)$. In particular, since we are working with discrete random variables, there is a probability of larger zero that $|S \cup T|$ is at most the floored value of our final upper bound for the expected value (also see Lemma 1.4). □

Exercise 1.3. Show that if

$$4\binom{k}{2}\binom{n}{k-2}2^{1-\binom{k}{2}} \leq 1$$

then the kth symmetric Ramsay number satisfies $R_k > n$.

Proof. Let $n, k \in \mathbb{N}$ such that the inequality is satisfied, and $G \sim \mathcal{G}_{n,p}$. Define for every subset $S \subseteq V(G)$ with $|S| = k$ an event E_S that S is a k-clique or k-independent set, i.e. induces a complete or empty graph. As shown in the lecture (see (1)), $P(E_S) = 2^{1-\binom{k}{2}}$. Consider the dependency graph for these events. For $S, T \subseteq V(G)$ of size k, it holds that E_S, E_T are independent iff $|S \cap T| \leq 1$, since only in this case they do not share an edge. In particular, E_S is mutually independent of all E_T with $|S \cap T| \leq 1$ since edges are existing independently from each other.

Therefore, a trivial upper bound of the degree of E_S is given as $\binom{k}{2}\binom{n}{k-2}$, i.e. choose at least 2 vertices from S to guarantee $|S \cap T| \geq 2$, and choose any $k-2$ vertices from all the other vertices (in fact, we could reduce n to $n-2$ to get a tighter upper bound, but that's not what the exercise wants). Since $4 \cdot \binom{k}{2}\binom{n}{k-2} \cdot P(E_S) \leq 1$ by assumption, we can apply Lovász Local Lemma (Theorem 1.14) and deduce that $P(\bigcap_S \overline{E}_S) > 0$, i.e. with non-zero probability there exists a graph with n nodes that does not contain any k-cliques or k-independent sets, proving $R_k > n$. □

Exercise 1.4. Consider the general form of Lovász Local Lemma:

Theorem A.1. Let $E_1, \ldots, E_n$ be a set of events in some probability space and $G = (V, E)$ their dependency graph. If there exist $x_1, \ldots, x_n \in [0, 1]$ such that

$$P(E_i) \leq x_i \prod_{(i,j) \in E} (1 - x_j),$$

then it holds that

$$P\left(\bigcap_{i=1}^{n} \overline{E}_i\right) \geq \prod_{i=1}^{n} (1 - x_i).$$

Use this to prove that you can replace the condition $4dp \leq 1$ in the symmetric version (Theorem 1.14) by the weaker condition $ep(d+1) \leq 1$ (where e denotes Euler's number).

Proof. Assume $P(E_i) \leq p$ such that the maximum degree of the dependency event

graph $G = (V, E)$ is d, and $ep(d+1) \leq 1$. Let $x_i \coloneqq \frac{1}{d+1}$ for every i. Then

$$\begin{aligned} x_i \prod_{(i,j)\in E} (1 - x_j) &= \frac{1}{d+1} \prod_{(i,j)\in E} \left(1 - \frac{1}{d+1}\right) && \mid \max_{v\in V} \deg(v) \leq d \\ &\geq \frac{1}{d+1} \left(1 - \frac{1}{d+1}\right)^d && \mid \text{Euler's inequality } e^x > \left(1 + \frac{x}{n+1}\right)^n \\ &> \frac{1}{d+1} \frac{1}{e} \geq p \geq P(E_i). \end{aligned}$$

By the general Local Lemma, $P(E_i) \geq \prod_{i=1}^{n} \frac{d}{d+1} > 0$. □

2. Exercise Sheet

Exercise 2.1. Let $X_1, \dots X_k \sim Pois(m)$ independent random variables and $c \in \mathbb{N}$. Find $h \in \mathbb{N}$ such that the probability of at least one X_i satisfying $X_i + c > h$ is at most 0.01.

For the proof we use following Chernoff bound which can be found in the accompanying book (Theorem 5.4).

> **Fact A.2.** For a Poisson-distributed random variable X with mean m it holds for $x > m$ that
>
> $$P(X \geq x) \leq \frac{e^{-m}(em)^x}{x^x}.$$

Proof. We can assume w.l.o.g. $c = 0$, since it has no influence on the random variable itself. For $c > 0$, we can simply add c to the corresponding value of h if c would be 0. Furthermore, we can formalize the exercise quite elegantly using the inverse event that no X_i satisfies the inequality, i.e. find h such that

$$\prod_{i=1}^{k} P(X_i \leq h) \overset{!}{\geq} 0.99.$$

Using our Chernoff bound, it holds for $h > m$ that

$$\prod_{i=1}^{k} P(X_i \leq h) = \prod_{i=1}^{k} 1 - P(X_i > h) \geq \prod_{i=1}^{k} 1 - P(X_i \geq h) \geq \left(1 - \frac{e^{-m}(em)^h}{h^h}\right)^k.$$

At this point, we need to solve the last term equaling 0.99 for h to find a valid value for h. According to WolframAlpha we need to use the Lambert W function and deduce

$$h = m \exp\left(W\left(\frac{-b}{me}\right) + 1\right), \quad b \coloneqq m + \log(1 - \sqrt[k]{0.99}) < m.$$

I cannot be bothered to write down if this indeed yields the wanted inequality. However, we can easily verify $h > m$: Notice that $\frac{-b}{me} > -\frac{1}{e}$, so we use W on the upper branch only, on which it is strictly monotonically increasing. Since $W(e^{-1}) = -1$, the exponent of the upper term is always greater than 0, and thus $h > m \exp(0) = m$. □

Exercise 2.2. Let $X = (X_1, \dots, X_n), Y = (Y_1, \dots, Y_n)$ be two random strings of i.i.d characters drawn from some finite set. The longest common subsequence is definied as the longest pair of sequences $i_1, i_2, \dots$ and $j_1, j_2, \dots$ such that $X_{i_1} = Y_{j_1}$ etc. Also, define Z as the singular long sequence of $2n$ i.i.d. random variables by

concatenating X to Y.

(a) The expected length l of the longest common subsequence for alphabet size Γ satisfies $\frac{1}{\Gamma}n < l$. There is also an upper bound $l/n < c_2 < 1$ whose discovery we leave as an exercise to the reader.

Proof. Consider following problem: Find the longest common subsequence in Y that is a substring starting in position 0 in X. In other words, we fix the subsequence in X and try to find matching characters in Y. This can be done easily using a greedy approach, i.e. if you are not yet at the end of Y and already found a subsequence of length k, from your current position on find the next possible index in Y that contains the character X_{k+1}. The number of steps until we find a matching character in Y is geometrically distributed with parameter $1/\Gamma$, so we expect to need k steps. Since every character in Y has an independent probability of $1/k$ of matching with the respective value in X, the total number of expected matches is n/k by linearity.

Since our problem yields a lower bound of the expected length, we have shown $\frac{n}{k} \leq l$. Strict inequality follows by repeating the process for another starting position in X and taking the maximum of the respective output length. The sole existence of instances such that this increases our lower bound suffices to disprove equality, and we do not need to do any exact math with highly correlated processes. □

(b) Let h be the function that maps Z to the length of the longest common subsequence. Then, following concentration result holds for any $\lambda > 0$:

$$P(|h(Z) - \mathbb{E}\left[h(Z)\right]| \geq \lambda) \leq 2\exp\left(\frac{-\lambda^2}{n}\right).$$

Proof. Clearly, h is 1-Lipschitz continuous, since changing exactly one letter at worst increases or decreases the longest common subsequence by 1. Applying McDiarmid (Theorem 3.6) immediately yields our desired result (notice Z uses $2n$ variables). □

(c) Again, use h as before. Following concentration result holds for the mean m of $h(Z)$ and any $\lambda > 0$:

$$P(|h(Z) - m| \geq \lambda\sqrt{2m}) \leq 4\exp\left(-\frac{\lambda^2}{4}\right).$$

Proof. We already established h to be 1-Lipschitz continuous. Now, use $f(s) \coloneqq 2s$. We show h is f-certifiable. Assume $f(Z) \geq s$, and choose $\mathcal{I}$ to be the indices of Z that belong to a common subsequence of length s (which exists by definition). In particular, $|\mathcal{I}| = 2s \leq f(s)$, since we choose one index of X and Y each per length unit. Also, if $Z' =_{\mathcal{I}} Z$, then by choice of $\mathcal{I}$ it holds that Z' contains a common subsequence of length at least s. Therefore, h is f-certifiable.

By Talagrand (**??**), we can now deduce using m as the median of $h(Z)$, and $\lambda > 0$

$$P(h(Z) \leq m - \lambda\sqrt{2m})P(h(X) \geq m) \leq \exp\left(-\frac{\lambda^2}{4}\right),$$
$$P(h(Z) \leq m)P(h(X) \geq m + \lambda\sqrt{2m}) \leq \exp\left(-\frac{\lambda^2}{4}\right).$$

Using $P(h(Z) \leq m), P(h(Z) \geq m) \geq \frac{1}{2}$ and adding up both cases yields

$$P(|h(Z) - m| \geq \lambda\sqrt{2m}) \leq 4\exp\left(-\frac{\lambda^2}{4}\right).$$

□

Exercise 2.3. Let X be the simple random walk on $\mathbb{Z}$ with $X_0 = 0$ and $M_n = \max_{0\leq i\leq n} X_i$ the running maximum. For $1 \leq i \leq n$ we define $E_i := X_i - X_{i-1}$. Note that $X_k = \sum_{i=1}^{k} E_k$.

(a) Since X_n is the sum of n independent random variables with variance $Var[E_i] = 1$, we know that $Var[X_n] = n$. By Kolmogorov's inequality, it holds for $\lambda > 0$ that

$$P(M_n \geq \lambda) \leq \frac{Var[X_n]}{\lambda^2}.$$

For $\lambda \in \omega(\sqrt{n})$ the upper bound goes to 0 for $n \to \infty$.

(b) We define $h(E_1, ..., E_n) := \frac{M_n}{2}$. It clearly holds that h is 1-Lipschitz continuous. Note that for h to be f-certifiable, f must be in $\Theta(n)$, since we have to choose all indices i with $E_i = 1$ and even for $s = 0$ we can still have $\frac{n}{2}$ indices with $E_i = 1$. Therefore we choose $f(s) = n$, as the results only differ in constant factors.

Using Talagrand's inequality we get

$$P(h(E) \leq \mu - \lambda\sqrt{f(\mu)})P(h(E) \geq \mu) \leq \exp\left(\frac{-\lambda^2}{4}\right)$$

Using $f(s) = n$ it follows

$$P(h(E) \leq \mu - \lambda\sqrt{n})P(h(E) \geq \mu) \leq \exp\left(\frac{-\lambda^2}{4}\right)$$

Choosing $\mu = m + \lambda\sqrt{n}$ where m is the median of $h(E)$ and using $h(E) = \frac{M_n}{2}$

gives

$$P(h(E) \geq m + \lambda\sqrt{n}) = P(M_n \geq 2(m + \lambda\sqrt{n})) \leq 2\exp\left(\frac{\lambda^2}{4}\right)$$

With $\lambda' = 2(m + \lambda\sqrt{n})$ we get

$$P(M_n \geq \lambda') \leq 2\exp\left(\frac{(\lambda' - 2m)^2}{16n}\right)$$

This bound does not improve the bound from (a) as for $\lambda' \in O(\sqrt{n})$ the term $\exp\left(\frac{(\lambda'-2m)^2}{16n}\right)$ does not go to 0.

Exercise 2.4. Consider the one-dimensional random geometric graph G using n vertices with independently drawn positions $X_1, \ldots, X_n \sim \mathcal{U}([0,1])$ and distance cutoff $\varepsilon \in O(n^{-\frac{1}{2}})$. Then, the maximum degree $\Delta \coloneqq \max_i d_i$ of G, with d_i as degree of vertex i, has sub-linear growth (i.e. $o(n)$) with probability converging to 1.

Proof. Clearly, $\Delta \in O(n)$. Therefore, it suffices to show that any linear bound converges to zero probability, i.e. for every $c > 0$

$$P(\Delta \geq cn) \xrightarrow{n\to\infty} 0.$$

By definition, $\Delta = \max_{i=1,\ldots,n} d_i$, which enables us to use a union bound approach:

$$\begin{aligned} P(\Delta \geq cn) &= P\left(\bigcup_{i=1,\ldots,n} \{d_i \geq cn\}\right) \leq \sum_{i=1}^{n} P(d_i \geq cn) \\ &= nP(d_i - \mathbb{E}[d_i] \geq cn - \mathbb{E}[d_i]) \leq nP(|d_i - \mathbb{E}[d_i]| \geq cn - \mathbb{E}[d_i]). \end{aligned}$$

Notice that $d_i \sim Bin(n-1, O(\varepsilon))$, since every vertex has an independent probability of $p \in O(\varepsilon)$ to be in range ε of X_i. Therefore, $\mathbb{E}[d_i] \in O(n\varepsilon) \subseteq O(\sqrt{n})$, and $\mathbb{V}[d_i] \in O(n\varepsilon(1-\varepsilon) \subseteq O(\sqrt{n})$. By Tschebychev, we conclude

$$\begin{aligned} nP(|d_i - \mathbb{E}[d_i]| \geq cn - \mathbb{E}[d_i]) &\leq \frac{\mathbb{V}[d_i]}{(cn - \mathbb{E}[d_i])^2} \\ &\in O\left(n\frac{\sqrt{n}}{(n-\sqrt{n})^2}\right) = O(n^{-\frac{1}{2}}) \xrightarrow{n\to\infty} 0. \end{aligned}$$

□

Our proof also works completely analogous for any $k > 2$ such that $\varepsilon \in O(n^{-\frac{1}{k}})$, which leads to the final bound $O(n^{-\frac{1}{k}}) \xrightarrow{n\to\infty} 0$. It also seems to work for any

$\varepsilon \in o(1)$, which would imply that as long as ε tends to 0, we will observe sublinearity of the maximum degree. (Which is surprisingly strong?)

3. Exercise Sheet

Definition A.3 (Randomized Polynomial-time)**.** A decision problem is part of **RP** iff there is a randomized algorithm running in polynomial time that

1. for yes-instances returns correctly "Yes" $\frac{1}{2}$ of the time,
2. for no-instances always returns correctly "No".

Assuming #SAT $\in$ **FPRAS**. Show that **NP** = **BPP** = **RP**.

Proof. We will show that SAT $\in$ **BPP**. Since SAT $\in$ **NPC**, we can construct an algorithm for *any* decision problem $P \in$ **NP** by (deterministically) transforming the instance in polynomial time and space to a SAT instance. Then, solving the SAT instance with the **RP** algorithm still yields polynomial running time and keeps the required probabilities.

Consider an instance $\mathcal{I}$, and let C be its solution count (i.e. the solution to the equivalent #SAT instance). We construct following algorithm: Since #SAT $\in$ **FPRAS**, there is an algorithm polynomial in input size with parameters $\varepsilon = \delta = \frac{1}{3}$ such that following error bounds hold for the approximate output count X:

1. If $\mathcal{I}$ is a yes-instance, then $C \geq 1$, and
$$\frac{2}{3} \leq P(|X - C| \leq \frac{1}{3}C) \leq P(C - X \leq \frac{1}{3}C) = P(X \geq \frac{2}{3}C) \leq P(X \geq \frac{2}{3}).$$
2. If $\mathcal{I}$ is a no-instance, then $C = 0$, and
$$\frac{2}{3} \leq P(|X - C| \leq \frac{1}{3}C) \leq P(|X| \leq 0) = P(X = 0).$$

We will guess $\mathcal{I}$ by running the **FPRAS** algorithm to estimate C as X, and say it is a yes-instance if $\lceil X \rceil \geq 1$, otherwise it is a no-instance ($X = 0$). Our error bounds then show that all properties of a **BPP** algorithm are satisfied.

Analogous, it suffices to show SAT $\in$ **RP** to imply **NP** = **RP**. The general idea now is that we only declare a yes-instance as yes if we can construct a certificate (here: admissible assignment) to ensure that a no-instance is always declared as no correctly. Consider following algorithm for a SAT instance $\mathcal{I}$ with n variables x_i:

Algorithm 4: RP algorithm for SAT

```
for k = 1, ..., n do
    x_k ←getRandomBit()
    I_k ← simplify I by fixing x_1, ... x_k
    X ←csatFprasSolver(I, ε, δ_n)
    if X = 0 then
        x_k ← ¬x_k
    end
end
return isSatSolution(x, I)
```

Obviously, for a no-instance we always return no, since it never has valid assignments. For a yes-instance, notice that we try to generate the certificate by fixing a variable one after another and checking if the corresponding #SAT instance is estimated to have a solution. W.l.o.g. assume there is only one valid assignment. If I_{k-1} is still a yes-instance, we keep the initial correct assignment of the new variable x_k with probability $1 - \delta_n$ (derived by same calculations as above), and reject a wrong assignment in favor of the correct assignment with probability $1-\delta_n$. Therefore, our final assignment has a probability of at least $(1-\delta_n)^n$ to be valid. We were not able to choose δ_n in such way that showed that the FPRAS algorithm is still polynomial in n. We tried $\delta_n = \sqrt[n]{1/2}$.

□

Exercise 3.2. Consider a deck of cards and following shuffling process: Choose two cards (by rank) uniformly independent and swap them. Choosing the same card twice results in no swap. We will show that the expected shuffling time needed is $O(n^2)$.

First, we show that this process is equivalent to choosing a fixed position in the deck and a card (by rank) uniformly indepent.

Proof. There is a bijection from rank to position, so uniformity is conserved. □

Using this variant, we consider their coupling where its two copies choose the same position and card. Denote with A_t the number of cards with different positions in the two decks at timestamp t. Then, A_t is non-increasing.

Proof. Consider following case distinction:

1. The cards chosen by rank are already in the same position. Then, they get swapped to the same position, as well as the cards chosen by position, resulting in no change.
2. The card chosen by rank has different positions in the deck, and the card chosen by position have same rank. Then, the cards chosen by rank get swapped into

the same position, but the match of the cards chosen by position is broken, resulting in no change.

3. The cards chosen by rank has different positions in the deck, and the cards chosen by position have different rank. Then, the cards chosen by rank get swapped into the same position, resulting in a new match, and the cards chosen by position might match by chance, resulting in a decrease of at least 1 (up to 3) of A_t.

So, case 1 and 2 result in no change, and 3 results in a decrease. □

It holds that

$$P(A_{t+1} \leq A_t - 1 \mid A_t \geq 1) \geq \left(\frac{A_t}{n}\right)^2$$

Proof. As shown previously, only case 3 results in a decrease, so we only need to calculate the probability of case 3 happening. If there are A_t cards with differing positions, there also A_t positions with differing cards. Therefore, the probability that the cards chosen by position are different has probability A_t/n. The same holds for the cards chosen by rank not being in the same position. Since the two events are independent, their total probability is their product. □

The expected time $\mathbb{E}[T]$ until $A_t = 0$ is $O(n^2)$ regardless of the initial coupling state.

Proof. By using the previous result, the number of steps until A_t decreases follows a geometric distribution with mean $(n/A_t)^2$. In the worst case, $A_t = n$, and we need to decrease n times until we hit $A_t = 0$, i.e. the expected number of steps until $A_t = 0$ is upper-bounded by the expected number of steps of n geometric processes $G_k \sim Geo\left(\frac{k}{n}\right)^2$. Then, we get a partial harmonic series

$$\mathbb{E}[T] \leq \mathbb{E}\left[\sum_{k=1}^{n} G_k\right] = \sum_{k=1}^{n} \frac{n^2}{k^2} \leq n^2 \sum_{k=1}^{\infty} \frac{1}{k^2} = n^2 \frac{\pi^2}{6} \in O(n^2).$$

□

Exercise 3.3. Consider a card shuffling process for n cards where we choose a position i from $0, \ldots n-1$ and swap the cards i and $i+1$, with 0 implying no shuffle.

This process converges to the uniform distribution of cards.

Proof. The process is clearly aperiodic (every state has a self-loop) and irreducible (there is a swap pattern to reach every shuffling state from each other). Since the process is symmetric (every transition is reversible with same probability), the inbound probabilities (i.e. the columns of the state transition matrix) sum to 1 for

every state. By the Perron–Frobenius theorem for Markov chains, the process converges to its unique stationary distribution, which clearly is the uniform distribution over all possible shufflings. □

Consider the coupling of two deck copies X, Y:
Let $S = \{j_0, \dots, j_k\}, 0 = j_0 < \cdots < j_k$ denote the set of positions i where the cards at position j *and* $j+1$ match up. First, let X choose a position j and change according to above process. Then, Y swaps the same cards if $j \notin S$. Otherwise $j = j_i \in S$ for some i and we swap j_{i+1} and its successor (i.e. we "rotate" within S). Its resulting mixing time $\tau(\varepsilon)$ is $O(n^4)$ for a fixed ε.

Proof. First, we show that the coupling never destroys matches, i.e. S is always a subset of the previous timestamp. Consider following case distinction:

1. $j \notin S$: Then we swap the same positions, clearly resulting in no change.
2. $j = j_i \in S$: Then the cards at $j_i, j_i + 1$ as well as $j_{i+1}, j_{i+1} + 1$ do not match up by definition, so there is nothing to destroy.

Also, the relative position of cards between decks stays consistent, i.e. if $i \leq j$ for a specific card with positions i, j at one point in X, Y respectively, then at any point of the process $i \leq j$ for their positions. Only the cases where the cards are at consecutive positions are interesting as this is the only time the relative position could change (since matches do not get destroyed, $i = j$ is clear). In particular, we would need to swap i to the left and j to the right (or vice versa, depending on $i > j$ or $i < j$). Notice that $\min(i, j) = j_k \in S$ since both positions of the card in question contain non-matching cards.

1. Case $i+1 = j$: For a rightwards swap in X, suppose we choose position $i = j_k$. Since we rotate within S for Y, the card at position j at worst can only get swapped further to the right.
2. Case $i = j+1$: For a leftwards swap in X, suppose we choose position $j = j_k$. Then, the card at j in Y does not get swapped anyway since we rotate in S.

Now, for a card with fixed rank, if the righter card in its corresponding deck at one point reached the beginning of the deck (position 1), the previous results guarantee that the cards will be matched up from that point in time on. Thus, as soon as *all* cards reached the beginning once, we can guarantee that X, Y are synchronized, yielding us an upper bound for mixing time. We are interested in the expected hitting time T_k of a card initially in position k reaching position 1. Notice this number is given by following recurrent relation, gained by considering all alternatives:

$$\mathbb{E}[T_k] = \begin{cases} 1 + \frac{1}{n}(\mathbb{E}[T_{k-1}] + \mathbb{E}[T_{k+1}]) + \frac{n-2}{n}\mathbb{E}[T_k], & 1 < k < n \\ 1 + \frac{1}{n}\mathbb{E}[T_{k-1}] + \frac{n-1}{n}\mathbb{E}[T_k], & k = n \\ 0, & k = 1 \end{cases}.$$

By induction from n to 2, we can show

$$\mathbb{E}[T_k] = (n-k+1)n + \mathbb{E}[T_{k-1}].$$

For $k = n$ a simple reordering yields $\mathbb{E}[T_n] = n + \mathbb{E}[T_{n-1}]$. Now, assuming the identity holds for all $k+1$ to n. By assumption and again some reordering

$$\begin{aligned} & \mathbb{E}[T_k] = 1 + \frac{1}{n}(\mathbb{E}[T_{k-1}] + (n-k)n + \mathbb{E}[T_k]) + \frac{n-2}{n}\mathbb{E}[T_k] \\ \iff \quad & \mathbb{E}[T_k] = \mathbb{E}[T_{k-1}] + n(n-k+1). \end{aligned}$$

Using another straightforward induction we see

$$\mathbb{E}[T_k] = n\sum_{i=2}^{k}(n-i+1) \leq n\frac{n(n-1)}{2} \leq n^3$$

Let $T_{\max}$ be the hitting time that every card reached the beginning at least once. As an upper bound, n cards need in expectancy at most n^4 steps until every card reached the beginning (i.e. observe one card until it reaches the beginning, then the next etc), so $\mathbb{E}[T_{\max}] \leq n^4$. Let $t(\varepsilon) = \varepsilon n^4$. By monotony of probability and Markov bounds we conclude

$$P_{X,Y}(X_{t(\varepsilon)} \neq Y_{t(\varepsilon)}) \leq P(T_{\max} \geq t(\varepsilon)) \leq P(T_{\max} \geq \frac{1}{\varepsilon}\mathbb{E}[T_{\max}]) \leq \varepsilon,$$

and by applying the mixing time lemma from the lecture $\tau(\varepsilon) \leq t(\varepsilon) \in O(n^4)$. □

Index