Randomized Methods in Computation*
Lecture 2: Laws of large numbers

Notes taken by Shai Mor and Erez Waisbard

Spring 2001

Summary: The focus of the lecture is various laws of large numbers. These laws bound the probability that a random variable deviates far from its expectation. Motivated by the problem of approximating the average of a function, we present stronger bounds than we had in the previous lecture. Moving from pairwise independent sampling to $2k$-wise independent sampling and then to totally independent sampling makes the the error probability exponentially vanishing (instead of linearly vanishing). We also define families of graphs called expanders, and prove the existence of a family of 3-regular expanders (using the probabilistic method).

1 Introduction

1.1 A motivating application

Loosely speaking the law of large numbers tells us that we can give a good estimation for the average of a bounded function by sampling enough points. Lets say we have a bounded function $f : \{0,1\}^n \to \{0,1\}$. It is infeasible for us to compute

$$\bar{f} \stackrel{\text{def}}{=} \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)$$

in time which is polynomial in n, however if we sample enough points we can give a good estimation that will be correct with high probability.

With respect to such sampling, we can associate a random variable X_i withe the value of f obtained at each sample point. Clearly each X_i has expected value $\bar{f}$, and if we take enough samples then the average of the corresponding X_i's is very likely to be close to $\bar{f}$. Our focus here is on quantitative statements of the above intuition.

In the previous lecture we gave the following bound for a set of pairwise independent random variables $\{X_1 \ldots X_m\}$:

$$Prob\left[\left|\left(\frac{1}{m}\sum_{i=1}^{m} X_i\right) - \left(\frac{1}{m}\sum_{i=1}^{m} E(X_i)\right)\right| \geq \delta\right] \leq \frac{max_i\{Var(X_i)\}}{m\delta^2} \tag{1}$$

We see that the more X_i's we have, the better our estimation gets. This is really what the law of large numbers is all about. It says that if you give me more random variables of the same

* Lecture Notes for a course given by Oded Goldreich at the Weizmann Institute of Science, Israel.

distribution then I can give a better estimation for the average of their expectation. A crucial point is that they have to be at least pairwise independent. This property enables us to get a good error bound. We make two observations looking at the above inequality:

1. There is a tradeoff between the deviation we want to bound and the error probability. Naturally we would like to be able to improve the tradeoff by both reducing the deviation and dropping the error probability.
2. The above bound gives us an error probability that is linearly dropping in the number of samples (i.e., m).

By imposing a stronger than pairwise independence restriction on the random variables, we can get a stronger dependence of the error-bound on the number of samples. Loosely speaking, the error probability is exponentially vanishing with the level of independence (between the random variables).

1.2 Independence of random variables

We now define independence between random variables. In the previous lecture we defined pairwise independence. We now define total independence between random variables.

Definition 1 *Let $X_1 \ldots X_m$ be random variables. We say that they are* totally independent *if for every m values $v_1 \ldots v_m$, the following holds*

$$Prob[X_1 = v_1 \wedge \ldots \wedge X_m = v_m] = \prod_{i=1}^{m} Prob[X_j = v_j]$$

Total independence and pairwise independence are two types of notions of independence of random variables. The notion of independence can be extended to express independence of subsets of certain size of random variables.

Definition 2 *Let $X_1 \ldots X_m$ be random variables. We say that they are* t-wise independent *if for every subset of t of them, $X_{i_1} \ldots X_{i_t}$, and for every values $v_1 \ldots v_t$, the following holds*

$$Prob[X_{i_1} = v_1 \wedge \ldots \wedge X_{i_t} = v_t] = \prod_{i=1}^{t} Prob[X_{i_j} = v_j]$$

In other words, $X_1 \ldots X_m$ are t-wise independent, if every t of them are totally independent. Another (equivalent) way to define independence between random variables is by saying that the probability for a random variable to have some value is not changed even if we know the values of the other random variables.

Definition 3 *Let $X_1 \ldots X_m$ be random variables. We say that they are* t-wise independent *if for every subset of t of them, $X_{i_1} \ldots X_{i_t}$, and for every values $v_1 \ldots v_t$, the following holds*

$$Prob[X_{i_1} = v_1 | X_{i_2} = v_2 \wedge \ldots \wedge X_{i_t} = v_t] = Prob[X_{i_1} = v_1]$$

We now see that pairwise independence is simply $t-$wise independence for $t = 2$ and total independence is $t-$wise independence for $t = m$. Clearly it holds that $(i+1)$-independent random variables are also i-independent random variables. The question one can ask at this point is whether it is really a stronger notion? The following example shows that it is.

Example 1.1 *Let X and Y be two independent random variables and let $Z = X \oplus Y$ be a third random variable. Clearly every two of these random variables are independent, however the three of them are dependent - knowing the value of any two of them immediately implies value of the third.*

2 Stronger laws of large numbers

In the previous lecture we initiated the study of techniques for bounding the probability that a random variable deviates far from its expectation. In this lecture we focus on a technique for obtaining considerably sharper bound on such probability. We now come to define the moments of a random variable as a measurement for its concentration. Taking advantage on the fact that the deviation of totaly independent random variables grows much slower than the deviation of dependent random variables we present the Chernoff bound who's error probability is exponentially vanishing. In addition, we present probability bound for the sum of k-wise independent random variables, where this bound vanish exponentially with k.

2.1 Moments

Loosely speaking the moments of a random are a measurement of its concentration. One type of moment is the variance, which was already defined in the last lecture. Other (higher) moments gives us a better measure for its concentration. We will be interested mostly with the moments of a normalized random variable.

Definition 4 *The t-centralized moment of a random variable X is*

$$M_t(X) = E((X - E(X))^t)$$

We can immediately see that $Var(X) = M_2(X)$.
We now come to use the bound on the moments to give a better error probability on the deviation of random variables.

2.2 Law for $2k$-independent trials

Theorem 5 *For $2k$-wise independent random variables $X_1 \ldots X_m$*

$$Prob\left[\left|\left(\frac{1}{m}\sum_{i=1}^{m} X_i\right) - \left(\frac{1}{m}\sum_{i=1}^{m} E(X_i)\right)\right| \geq \delta\right] \leq \frac{\max_i \left(\max_{t\geq 2}(M_t(X_i))^{2k/t}\right)}{\left(\frac{m\delta^2}{k^2}\right)^k}$$

Note that for the case that $k = 1$ this is exactly what we had in Eq. (1).
Proof: In order to prove the above theorem we first normalize the random variables. That is, we define

$$\bar{X}_i = X_i - E(X_i)$$

This technicality will play a crucial part in the proof because the expectation of a normalized random variable is 0.
We rewrite the probability in the claim in terms of the normalized random variable, using the fact that we raise it to an even power to get rid of the absolute value:

$$Prob\left[\left|\left(\frac{1}{m}\sum_{i=1}^{m} \bar{X}_i\right)\right| \geq \delta\right] = Prob\left[\left(\frac{1}{m}\sum_{i=1}^{m} \bar{X}_i\right)^{2k} \geq \delta^{2k}\right]$$

Applying the Markov inequality we get

$$
\begin{aligned}
Prob\left[\left(\frac{1}{m}\sum_{i=1}^{m}\bar{X}_i\right)^{2k} \geq \delta^{2k}\right] &\leq \frac{E\left((\frac{1}{m}\sum_{i=1}^{m}\bar{X}_i)^{2k}\right)}{\delta^{2k}} \\
&= \frac{E((\sum_{i=1}^{m}\bar{X}_i)^{2k})}{m^{2k}\delta^{2k}}
\end{aligned}
$$

We now focus on the numerator. Using the linearity of expectations we get

$$
E\left((\sum_{i=1}^{m}\bar{X}_i)^{2k}\right) = \sum_{(i_1\ldots i_{2k})\in[m]} E(\bar{X}_{i_1}\cdots\bar{X}_{i_{2k}})
$$

where $[m]=\{1,\ldots,m\}$.
We now split the right sum to the sum over two disjoint subsets A and B, where A is a set of $2k-$long sequences in which some integer appears uniquely, and B is a set of $2k-$long sequences in which each integer appears with multiplicity at least 2. That is $(i_1\ldots i_{2k})\in A$ if there exist i such that a unique j satisfies $i_j=i$ and $(i_1\ldots i_{2k})\in B$ if for every j there exist $j'\neq j$ such that $i_j=i_{j'}$.

(Note that if we think of m as being much greater than k, we can expect most of the terms to be from the subset A. As we will see next, these terms will have no contribution to the total sum.)

$$
\begin{aligned}
\sum_{(i_1\ldots i_{2k})\in[m]} E(\bar{X}_{i_1}\cdots\bar{X}_{i_{2k}}) &= \sum_{(i_1\ldots i_{2k})\in A} E(\bar{X}_{i_1}\cdots\bar{X}_{i_{2k}}) + \sum_{(i_1\ldots i_{2k})\in B} E(\bar{X}_{i_1}\cdots\bar{X}_{i_{2k}}) \\
&= 0 + \sum_{(i_1\ldots i_{2k})\in B} E(\bar{X}_{i_1}\cdots\bar{X}_{i_{2k}})
\end{aligned}
$$

The reason that the sum over A equals zero follows from the independence of the $\bar{X}_i$'s and using $E(\bar{X}_{i_j})=0$ for i_j which appears with multiplicity 1. That is, for $(i_1\ldots i_{2k})\in A$ so that $i_j\neq i_{j'}$ for every $j\neq j'$, we have

$$
E(\bar{X}_{i_j}\prod_{j\neq j'}\bar{X}_{i_{j'}}) = E(\bar{X}_{i_j})E(\prod_{j\neq j'}\bar{X}_{i_{j'}}) = 0\cdot E(\prod_{j\neq j'}\bar{X}_{i_{j'}}) = 0
$$

where the first equality uses the fact that $\bar{X}_{i_j}$ is independent of $\prod_{j\neq j'}\bar{X}_{i_{j'}}$ (since all $i_{j'}$ are different from i_j).
We now need to bound the sum over B. We do it by counting the number of sequences of length $2k$ over $[m]$ in which each occurrence is with multiplicity at least 2 (which clearly upper bounds $|B|$). Let j denote the number of integers occurring in the sequence. Since each such integer appears at least twice, j can not be larger than k, thus an upper bound on the number of such sequences is:

$$
\sum_{j=1}^{k}\binom{m}{j}j^{2k} \ll m^k k^{2k}
$$

Each of these sequences contributes

$$
E(\bar{X}_{i_1}^{e_1}\cdots\bar{X}_{i_j}^{e_j}) = E(\bar{X}_{i_1}^{e_1})\cdots E(\bar{X}_{i_j}^{e_j}) = M_{e_1}(\bar{X}_{i_1})\cdots M_{e_j}(\bar{X}_{i_j})
$$

where the e_j's are the number of occurrences of each i_j (and so $\sum_j e_j = 2k$). So we get

$$E((\sum_{i=1}^{m} \bar{X}_i)^{2k}) \quad \leq \quad m^k k^{2k} \cdot \max_i(\max_{e\geq 2}(M_e(\bar{X}_i)^{\frac{2k}{e}})$$

which in turn implies

$$\frac{E((\sum_{i=1}^{m} \bar{X}_i)^{2k})}{m^{2k}\delta^{2k}} \quad \leq \quad \frac{\max_i(\max_{e\geq 2}(M_e(X_i))^{\frac{2k}{e}})}{\left(\frac{m\delta^2}{k^2}\right)^k}$$

and the theorem follows. ■

This is clearly a much stronger bound than what we had before. Intuitively, if we consider k as a constant and consider X_i as a Bernoulli random variable, we get a very similar bound to the one we had before only now we get the k exponent in the denominator.

2.3 The Chernoff/Hoefding bound

Up until now the only restriction on the random variables was for them to be fully or partially independent. We now restrict ourselves to bounded random variables and come up with a tighter bound. The following theorem, which is called the Chernoff bound or the Hoefding bound, is presented here without proof.

Theorem 6 *For totaly independent random variables $X_1 \ldots X_m$, where $X_i \in [a,b]$*

$$Prob\left[\left|\left(\frac{1}{m}\sum_{i=1}^{m} X_i\right) - \left(\frac{1}{m}\sum_{i=1}^{m} E(X_i)\right)\right| \geq \delta\right] \leq 2e^{-\frac{2\delta^2}{(b-a)^2}\cdot m}$$

In the special case where $X_i \in [0,1]$, we get

$$Prob\left[\left|\left(\frac{1}{m}\sum_{i=1}^{m} X_i\right) - \left(\frac{1}{m}\sum_{i=1}^{m} E(X_i)\right)\right| \geq \delta\right] \leq 2e^{-2\delta^2\cdot m} \tag{2}$$

Clearly, Theorem 6 follows from Eq. (2); e.g., for X_i's as in Theorem 6, define $Y_i = (X_i - a)/(b-a)$ and apply Eq. (2).

> Oded's Note: *Following are some remarks regarding the proof of Eq. (2). Firstly, we argue that it suffices to consider 0-1 random variables (intuitively, because they have the largest variance among random variables with range $[0,1]$). For 0-1 X_i's, denote $p_i = \Pr[X_i = 1]$ and $\bar{X}_i = X_i - p_i$. Without loss of generality we bound the probability that $\frac{1}{m}\sum_{i=1}^{m} \bar{X}_i$ is greater from δ. For any $\lambda > 0$, we have*
>
> $$\Pr\left[\frac{1}{m}\sum_{i=1}^{m} \bar{X}_i \geq \delta\right] \quad = \quad \Pr\left[e^{\lambda \sum_{i=1}^{m} \bar{X}_i} \geq e^{\lambda\delta m}\right]$$
> $$< \quad \frac{E(e^{\lambda\sum_{i=1}^{m} \bar{X}_i})}{e^{\lambda\delta m}}$$
>
> *Using the total independence of the X_i's we have $E(e^{\lambda\sum_{i=1}^{m}\bar{X}_i}) = \prod_{i=1}^{m} E(e^{\lambda\bar{X}_i})$. On the other hand, $E(e^{\lambda\bar{X}_i}) = p_i e^{\lambda\cdot(1-p_i)} + (1-p_i)e^{\lambda\cdot(0-p_i)} \leq e^{\lambda^2/8}$, where the last inequality*

holds for any $p_i \in [0,1]$ (and has elementary proofs unrelated to probability theory). *Combining all the above, we get*

$$\begin{aligned}\Pr\left[\frac{1}{m}\sum_{i=1}^{m}\bar{X}_i \;\geq\; \delta\right] &< \frac{(e^{\lambda^2/8})^m}{e^{\lambda\delta m}} \\ &= e^{\left(\frac{\lambda^2}{8}-\lambda\delta\right)\cdot m}\end{aligned}$$

Setting $\lambda = 4\delta$, we get $\Pr\left[\frac{1}{m}\sum_{i=1}^{m}\bar{X}_i \;\geq\; \delta\right] \;<\; e^{2\delta^2\cdot m}$, and the claim follows.

2.4 Multiplicative Chernoff Bounds

Oded's Note: *Unfortunately, I did not manage to say the following in class.*

In some cases it is more useful to use "Multiplicative Chernoff Bounds" of the following form. For independent and indentically distributed X_i's in $[0,1]$ with $p = \Pr[X_i = 1]$, and for every $\gamma > 0$, it holds that

$$\Pr\left[\frac{1}{m}\sum_{i=1}^{m}X_i \;>, (1+\gamma)\cdot p\right] \;<\; e^{\gamma^2 p\cdot m/3} \tag{3}$$

$$\Pr\left[\frac{1}{m}\sum_{i=1}^{m}X_i \;<, (1-\gamma)\cdot p\right] \;<\; e^{\gamma^2 p\cdot m/2} \tag{4}$$

Note that by applying the standard (i.e., additive) Chernoff bound we get a probability bound of $e^{2(\gamma p)^2\cdot m}$ in both cases. The Multiplicative forms are thus preferable for small values of p (e.g., for $p < 0.1$).

3 Application: Estimating the mean of a function

In this section we discuss probabilistic polynomial time computations that will do approximations to problems that require by their inherent nature exponential-time deterministic computations. Note that we regard computations as feasible only if their time of execution is at most of polynomial order of magnitude and infeasible if their time is of exponential order.

These approximations will rely heavily on the "Laws of Large Numbers" presented before. We will use *samples* from the problem's (i.e., function's) domain.

Given a boolean function $f : \{0,1\}^n \to [0,1]$, we denote f's mean value by $\bar{f}$. That is

$$\bar{f} \stackrel{\text{def}}{=} \frac{1}{2^n}\sum_{x\in\{0,1\}^n} f(x).$$

In order to exactly compute $\bar{f}$ one is required to take 2^n evaluations of f (i.e. sample f's domain in its entirety). Thus, a feasible estimation of $\bar{f}$ can not precisely follow its strict definition because the domain of function f is of exponential size in n. Fortunately there are "shortcuts" but inevitably these come with a "price tag". Here we shall introduce the notion of estimating (or approximating) $\bar{f}$ in (probabilistic) polynomial time.

The problem we face is how to give a satisfactory probabilistic approximation of $\bar{f}$ by means of a polynomial computation.

Definition 7 *An* (ϵ, δ)-sampler, *S, is a probabilistic polynomial-time oracle machine that when given oracle access to the function f satisfies*

$$Prob[|S^f(1^n, \epsilon, \delta) - \bar{f}| > \delta] < \epsilon.$$

In the above, $S^f(1^n, \epsilon, \delta)$ denotes the output distribution of the oracle machine S on input 1^n, ϵ and δ, and oracle access to the function f. The above definition means that the probability that the output of the sampler deviates from the actual value, $\bar{f}$, by more than δ is no more than ϵ.

The **question** we now face is "How many oracle access are needed in order to give a (ϵ, δ) approximation to $\bar{f}$"?. That is, what is the query complexity of (ϵ, δ) samplers.

Theorem 8 (a constructive upper bound) *The query complexity of* (ϵ, δ) *samplers is at most* $O(\frac{1}{\delta^2} log(\frac{1}{\epsilon}))$.

Proof: In order to assert the above bound on the number of queries we suggest the following

- Take $m \stackrel{\text{def}}{=} O(\frac{1}{\delta^2} log(\frac{1}{\epsilon}))$ independent uniform samples on f's domain, denoted by $s_1, s_2, ..., s_m \in \{0,1\}^n$.
- Output $\frac{1}{m}\sum_{i=1}^m f(s_i)$ which will be implement as follows:
  For i=1 to m do: Make a oracle call to f with input s_i, obtain $v_i = f(s_i)$, compute $\frac{1}{m}\sum_{i=1}^m v_i$ and output it.

We will prove that the estimation we gave on the number of samples is indeed tight from bellow (i.e., m samples are necessary for an (ϵ, δ) approximation).

> Oded's Note: *The probability space consists of m-long sequences of n-bit strings; that is, of all possible choices of* $\overline{s} = (s_1, ..., s_m) \in (\{0,1\}^n)^m$. *We define m random variables* $X_1, X_2, ..., X_m$ *so that* $X_i(\overline{s}) = f(s_i)$. *Thus,* X_i *actually depends only on* s_i*, and so these random varibales are clearly independent.*

Let $X_1, X_2, ..., X_m$ be m random variables, such that $X_i = f(s_i)$. These random variables are clearly independent of one another, and each ranges in $[0,1]$ and has expectation that equals $\bar{f}$. Applying Chernoff's Bound, we get

$$Prob\left[\left|S^f(1^n, \epsilon, \delta) - \bar{f}\right| > \delta\right] = Prob\left[\left|\frac{1}{m}\sum_{i=1}^{m} X_i - \bar{f}\right| > \delta\right] < 2e^{-2\delta^2 \cdot m}$$

Thus, the bound is exponentially vanishing in $\delta^2 m$. We compare this bound to ϵ and derive an bound on the number of samples, m, which are required by our (ϵ, δ) approximation algorithm; that is, $2\exp[-2\delta^2 \cdot m] = \epsilon$ which implies

$$m = \frac{1}{2\delta^2} \cdot \log(\frac{2}{\epsilon}).$$

as required. ■

Notice that the bound on the number of sufficient samples does NOT dependent on n, the function f's domain size. This means that the bound does not distinguish between functions of large domain and functions of relatively smaller domain.

The upper bound (on the number of samples that are sufficient for an (ϵ, δ) approximation) given in Theorem 2 is tight up to a multiplicative factor. That is:

Theorem 9 (a lower bound) *The query complexity of* (ϵ, δ) *samplers is at least* $\Omega(\frac{1}{\delta^2} log(\frac{1}{\epsilon}))$.

4 Existence of Expanders via the Probabilistic Method

In this section we will show the existence of expanders via the "The Probabilistic Method", which was introduced in previous lecture. This is a combinatorial tool for demonstrating the existence of combinatorial objects. The example we will focus on is a *probabilistic proof of existence* for an expander (i.e., a family of graphs with certain properties). By a probabilistic proof we mean that we do not know how to explicitly find the expander we are looking for and the proof of its existence is implicit. We merely argue that a random object that we construct is not an expander is strictly smaller than 1. This is enough to obtain the theorem.

4.1 Expander Graphs

In order to start speaking about expanders, we will need some definitions from graph-theory to assist us in what follows.

Definition 10 *In a graph $G(V, E)$ a* neighborhood, Γ, *of a vertex $v \in V$ is the set of all adjacent vertices to v (vertices that are connected to v by an edge). A neighborhood is described by the following notation :*

$$\Gamma(v) \stackrel{\text{def}}{=} \{u : (v, u) \in E\}.$$

Note also that one can also extend the above definition for an arbitrary subset $V' \subseteq V$ (referring to the neighborhoods of a subset of vertices) by

$$\Gamma(V') \stackrel{\text{def}}{=} \{v \in V \mid \exists (v', v) \in E \ for\ some\ v' \in V'\}.$$

Thus,

$$\Gamma(V') = \bigcup_{v' \in V'} \Gamma(v').$$

Continuing with the spirit of the above definition we shall now proceed with defining the outer section of a vertices set

Definition 11 *For a set S,* the boundary *of the set, $\Gamma'(S)$, is the set of vertices that are adjacent to some vertex in S and are not in S themselves. That is:*

$$\Gamma'(S) \stackrel{\text{def}}{=} \Gamma(S) \setminus S.$$

> Oded's Note: *The following two motivational paragraphs are due to the scribes themselves. I did not motivate the definition of expander graph beyond saying that they will play an important role in some future lectures.*

A very import class of sparse graphs are expander graphs. Among other things, they can be considered as a model for a network with certain desirable properties. Basically an expander has the property that every subset of its vertices has a large set of neighbors. That implies that any pair of vertices are connected by a short path. Furthermore, removing random edges in the graph (simulating local connection failures) does not reduce the property by much, so a network which is an expander can be considered as a fault-tolerant. We focus on expander graphs that are regular.

Definition 12 *A graph is d-*regular *if and only if every vertex has degree d.*

Now let us proceed to the definition of expander

Definition 13 *A d-regular graph $G(V,E)$ is a (d,c)-*expander *or has a c-*expansion *(for some $c > 0$) if and only if for every subset $V' \subseteq V$ of size $|V'| \leq |V|/2$ it holds that*

$$|\Gamma'(V')| \geq c|V'|.$$

To see that an expander makes a good network, suppose that you want to route a message from a vertex a to a vertex b in G. Every vertex has d neighbors. By the properties of an expander, there are at least $(1+d)(1+c)$ vertices at distance no more than 2 from a. Going further away, there will be at least $(1+d)(1+c)^k$ vertices at distance no more than $k+1$ from a. We can continue expanding from a until the reachable set of vertices V_a has more than $|V|/2$ vertices. The vertex b may not be among them, but if we expand from b in the same way, we eventually obtain a set V_b of more than $|V|/2$ vertices reachable from b. The sets V_a and V_b both contain more than $|V|/2$ so they must overlap. The overlap contains vertices on the path from a to b. In this way, we have shown that for any pair of vertices a to b, there is a path of length no more than $2(k+1)$ from a to b, where $k = \log_{1+c}|V|/2$. As one can observe from the expression for k, the larger the value of c the shorter the expected path between any two vertices.

The notion of an expander will usually be employed in the context of taking c as a small constant and of V as a huge vertex set (because we are interested in cases where we face very large graphs). We will generalize the above definition to a collection of graphs, referred to as a "family", each of which have an expansion.

The above definition of a (d,c)-expander extends naturally a family, F, of graphs $F = \{G_N\}_{N \in S}$, where G_N is an N-vertex graph and S is an infinite subset of $\mathbb{N}$. That is, for each N in S there exists a single graph in the family that has N vertices and all the graphs in the family are (d,c)-expanders.

4.2 On suitable Sets of graph sizes

Let us define a function $next_S$ so that for every integer m:

$$next_S(m) \stackrel{\text{def}}{=} min_{n \in S}\{n \geq m\}.$$

That is, $next_S$ returns the smallest integer that is at least as big as m. The growth rate of $next_S$ determines the suitability of S to be used as graph sizes in families of graphs. The criterion that we will adopt in order to examine the suitability of a given set to be used in this context is that the output of the $next_S$ function is bound by a polynomial on its input. Our reason for adopting such a criterion is because upon given an m, our purpose is to construct a graph, G, with $n \geq m$ vertices. Typically, we will need to embed m objects as values of G. The difficulty is that we don't want the graph, G, to be too big because our measure of computational complexity for any given task/algorithm on a graph is based on the size of that graph.

Some typical examples of sets S that are practical and useful in various applications are the following:

- $S = \{2n : n \in \mathbb{N}\}$. Here $next_S(m) \leq m+1$.
- $S = \{Primes\}$. Here $next_S(m)$ is quite complex but certainly at most $2m$ (because each interval $[m, 2m]$ must contain a prime).
- $S = \{n^2 : n \in \mathbb{N}\}$. Here

$$\begin{aligned} next_S(m) &= \lceil\sqrt{m}\rceil^2 \\ &< m + 2\sqrt{m} + 1 \\ &< m + o(m). \end{aligned}$$

Even for $S = \{2^n : n \in \mathbb{N}\}$, We have

$$\begin{aligned} next_S(m) &= 2^{\lceil \log_2(m) \rceil} \\ &\leq 2m. \end{aligned}$$

On the other hand, bad examples correspond to very-sparse sets such as $S = \{2^{2^{2^n}} : n \in \mathbb{N}\}$. Here, for some m,

$$\begin{aligned} next_S(m) &= 2^{2^{2^{\log\log\log(m)+1}}} \\ &= 2^{2^{2\log\log m}} \\ &= 2^{(\log m)^2} \\ &= m^{log(m)} \end{aligned}$$

and that is more than any polynomial in m.

4.3 On the existence of 3-regular expanders

Now, we proceed to providing a probabilistic proof of the existence of expanders. We will use the following graph-algorithmic definition :

Definition 14 *Given a graph $G(V,E)$ a* matching *of G is a subset of the edges of G such that every vertex appears in at most one edge of the subset. A* perfect matching *is a matching where all vertices in G appear.*

Now we get to the main theorem of our discussion about expanders.

Theorem 15 *There exists a constant $c > 0$ for which, for all sufficiently large even N, there exists an N-vertex, 3-regular graph with expansion c.*

Proof: We use the probability method. We will show that for sufficiently small $c > 0$ and all sufficiently large (even) N, with high probability, a random 3-regular N-vertex graph has expansion c.

We will generate a random 3-regular graph G by selecting three random perfect matchings, denoted G_1, G_2 and G_3. To generate a random undirected graph G_1 with $N = 2K$ vertices and K edges, we will select edges by *selecting uniformly at random* a pair of distinct vertices, and removing them from the pile of available vertices (so our selection of a pair is a probabilistic event which is dependable on previous selections!). We will continue this selection of pairs of vertices until we exhaust the entire set of N vertices. When we are done, we actually have a perfect matching on G (i.e., K edges where created, every vertex is covered once). We will independently repeat this process for three times, thus getting three graphs $\{G_1, G_2, G_3\}$ and three very-possibly disjoint sets of edges, matchings, (no edge will repeat itself unless the same pair was selected more than once, which is highly unlikely as N becomes very large). Then we combine our three graphs into one possibly multi-graph G which is 3-regular (every vertex has no more then three neighbors, if there are less than three then there exists a neighbor to which it has a multi edge (two or three parallel edges)). We want to prove that *the probability that graph G has expansion c is bigger than zero (and moreover it's close to 1).* We will prove this by showing that the probability of the complement event (that there is no such expansion c of G) is strictly less than 1 (and actually close to 0). That is, the probability that there exists a subset of vertices that has too small boundary and so does not permit the existence of an expansion c of G is strictly less than 1. We shall bound the probability of the union of all these "bad events" by means of the Union Bound argument:

Proposition 4.1 (Union Bound): *For every series of sets $\{A_i\}$ and every random variable X:*

$$Prob[X \in \bigcup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} Prob[X \in A_i].$$

We now get to the details of the actual proof. Let $N = 2K$ be the number of vertices of graph G_N. We will evaluate the probability of the complement event to the event that we are interested in; that is, the probability that G_N does not have expansion c. That is, we look at the probability of the following (complementary) event:

$$\text{there exists a subset } V' \text{ s.t. } |V'| \leq K \text{ and } |\Gamma'(V')| < c|V'|. \tag{5}$$

By the union bound, the probability that the event in Eq. (5) occurs is at most

$$\sum_{V':|V'|\leq K} Prob[|\Gamma'(V')| < c|V'|] \tag{6}$$

Again, by the union bound, Eq. (6) is bounded from above by

$$\sum_{V':|V'|\leq K} \sum_{V'':|V''|=c|V'|} Prob[\Gamma'(V') \subset V''] \tag{7}$$

We give an upper bound on Eq. (7) by considering all possible V' of size i and all V'' of size ci, where i goes from 1 to K (i.e., we are enumerating over all possible sizes of the two sets). Clearly, it suffices to consider in the second sum only the sets V'' that are disjoint of V'. Thus, we get the bound

$$\sum_{i=1}^{K} \binom{N}{i} \cdot \binom{N-i}{ci} \cdot Prob[\Gamma'(V'_0) \subset V''_0]$$

where V'_0 and V''_0 are arbitrary fixed disjoint sets of sizes i and ci, respectively. Equivalently, we may bound

$$\sum_{i=1}^{K} \binom{N}{i} \cdot \binom{N-i}{ci} \cdot Prob[\Gamma(V'_0) \subset V'_0 \cup V''_0]$$

which, using $V'_0 = [i] \stackrel{\text{def}}{=} \{1, ..., i\}$ and $V''_0 = \{i+1, ..., i+ci\}$, is upper bounded by

$$\sum_{i=1}^{K} \binom{N}{i} \cdot \binom{N}{ci} \cdot Prob\,[\Gamma([i]) \subseteq [i+ci]] \tag{8}$$

We now focus on bounding the probability that $\Gamma([i]) \subseteq [i+ci]$. That is, we bound the probability that in all three perfect matchings, the vertices in $[i]$ are matched to vertices in $[i+ci]$. Focusing on each of these three perfect matchings, the probability that in a single perfect matching all the vertices in $[i]$ are matched to vertices in $[i+ci]$ is at most

$$\frac{i+ci-1}{N-1} \cdot \frac{i+ci-3}{N-3} \cdots \frac{i+ci-i+1}{N-i+1} \;=\; \prod_{j=1}^{i/2} \frac{i+ci-2j+1}{N-2j+1}$$

where the bound is obtained by considering a process of iteratively matching $i/2$ vertices in $[i]$, where in each step we match the next unmmatched vertex and consider the probability that it is matched to one of the remaining unmatched vertices in $[i+ci]$. By algebraic manipulations we get:

$$\begin{aligned}
&\prod_{j=1}^{i/2} \frac{i+ci-2j+1}{N-2j+1} \\
&< \prod_{j=0}^{(i/2)-1} \frac{i+ci-2j}{N-2j} \\
&= \frac{2^{i/2} \cdot \prod_{j=0}^{(i/2)-1} \frac{i+ci}{2} - j}{2^{i/2} \cdot \prod_{j=0}^{(i/2)-1} \frac{N}{2} - j} \\
&= \frac{\binom{(i+ci)/2}{i/2}}{\binom{N/2}{i/2}}
\end{aligned}$$

Recall that the above line provides an upper bound on the probability that a single random perfect matching matches all vertices in $[i]$ to vertices in $[i+ci]$. Thus, the probability that is happens in all three perfect matchings used to construct the graph is the cube power of the above, and we get:

$$Prob\left[\Gamma([i]) \subseteq [i+ci]\right] < \left(\frac{\binom{(i+ci)/2}{i/2}}{\binom{N/2}{i/2}}\right)^3$$

Plugging this into Eq. (8), we obatin the bound

$$\sum_{i=1}^{K} \frac{\binom{N}{i} \cdot \binom{N}{ci} \cdot \binom{(i+ci)/2}{i/2}^3}{\binom{N/2}{i/2}^3} \tag{9}$$

Using the approximatiom $\binom{n}{\alpha n} = 2^{H_2(\alpha)\cdot n}$, where $H_2(\alpha) = \alpha \log_2(1/\alpha) + (1-\alpha)\log_2(1/(1-\alpha))$ is the binary entropy function, we get

$$\begin{aligned}
&\log_2\left(\frac{\binom{N}{i} \cdot \binom{N}{ci} \cdot \binom{(i+ci)/2}{i/2}^3}{\binom{N/2}{i/2}^3}\right) \\
&\approx H_2(i/N)\cdot N + H_2(ci/N)\cdot N + 3H_2(i/(i+ci)) \cdot \frac{i+ci}{2} - 3H_2(i/N)\cdot\frac{N}{2} \\
&= -(H_2(i/N) - 2H_2(ci/N)) \cdot \frac{N}{2} + 3H_2(1/(1+c)) \cdot \frac{i+ci}{2}
\end{aligned}$$

For a sufficiently small $c > 0$ (and $i \leq N/2$), we have $H_2(i/N) < c' \cdot H_2(ci/N)$ for $c' \approx c\log(1/c)$, and $H_2((1/(1+c)) \approx H_2(c) \approx c\log_2(1/c)$. Plugging all this into Eq. (9), we obatin the bound

$$\sum_{i=1}^{K} 2^{-(1-c')H_2(i/2K)\cdot K + 2c'\cdot i}$$

which approaches zero when c does. The theorem follows. ■

4.4 Non-existence of 2-regular expanders

Contrary to the previous theorem, the following theorem asserts exactly the opposite for 2-regular graphs

Theorem 16 *For all $c > 0$ and for all sufficiently large N, there exists NO $(2, c)$-expander with N vertices.*

Proof: Fix some arbitrary $c > 0$. Consider an N-vertex graph which is 2-regular. Without loss of generality consider a connected graph [if the graph is build out of components then we can choose all vertices in the smallest component (which contains at most $N/2$ vertices) and get that their boundary is null!]. Such a connected graph must be a closed cycle. Let us consider a connected subset of $N/2$ vertices. The boundary of such a subset is of size 2. In order to contradict the requirement for the existence of an expander, choose the number of vertices in the graph, N, such that $c \cdot N/2 > 2$ (i.e. $N > 4/c$ will do). ■