03683170: Expanders, Pseudorandomness and Derandomization 13/03/16

Lecture 2 – Deterministic Amplification

Amnon Ta-Shma and Dean Doron

1 A quick review of concentration bounds

Theorem 1 (Markov's inequality). *If X is a nonnegative random variable then for every $a > 0$, $\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$.*

Theorem 2 (Chebyshev's inequality). *If X is a random variable, then for every $a > 0$,*

$$\Pr[|X - \mathbb{E}[X]| \geq a] \leq \frac{\mathrm{Var}[X]}{a^2}.$$

Theorem 3 (The Chernoff bound, [4, 2]). *Suppose $Y_1, \ldots, Y_n$ are i.i.d. boolean random variables with expectation μ. Then for every $\varepsilon > 0$,*

$$\Pr\left[\sum_{i=1}^{n} Y_i > (\mu + \varepsilon)n\right] \;\leq\; e^{-2\varepsilon^2 n}.$$

If the Y_i-s are not necessarily boolean, we have:

Theorem 4 (The Chernoff-Hoeffding bound, [4]). *Suppose $Y_1, \ldots, Y_n$ are independent random variables with expectations $\mu_1, \ldots, \mu_n$ such that $Y_i \in [a, b]$ for every $i \in [n]$. Then for every $\varepsilon > 0$,*

$$\Pr\left[\sum_{i=1}^{n} (Y_i - \mu_i) > \varepsilon n\right] \;\leq\; e^{-\frac{2\varepsilon^2 n}{(b-a)^2}}.$$

2 k-wise independence

Definition 5. *Let $X_1, \ldots, X_n$ be a sequence of random variables. We say they are k-wise independent if for all $1 \leq i_1 < \ldots < i_k \leq n$, $X_{i_1}, \ldots, X_{i_k}$ are independent. That is, for every $\alpha_1, \ldots, \alpha_k$ in their support, $\Pr[X_{i_1} = \alpha_1 \wedge \ldots X_{i_1} = \alpha_k] = \Pr[X_{i_1} = \alpha_1] \cdot \ldots \cdot \Pr[X_{i_k} = \alpha_k]$. We will also assume that each X_i by itself is uniform.*

We shall now construct a small pairwise-independent sample space. Namely, $X_1, \ldots, X_n$ where each X_i is uniform over $[n]$ and the support size is n^2 (this is tight! explain why). Assume that n is a power of 2 and consider the field $\mathbb{F} = \mathrm{GF}(n)$.

The sample space is $\mathbb{F} \times \mathbb{F}$ and the distribution on the sample points is uniform. For every $i \in [n]$, we set $X_i(a, b) = a \cdot i + b$, where i is considered as an element from the field $\mathbb{F}$ and addition and multiplication are in $\mathbb{F}$. First, note that every X_i is uniform over $\mathbb{F}$. Now, for every distinct $i, j \in [n]$ and $\alpha_1, \alpha_2 \in \mathbb{F}$,

$$\Pr_{a,b \in \mathbb{F}}[X_i = \alpha_1 \wedge X_j = \alpha_2] = \Pr_{a,b \in \mathbb{F}}\left[\begin{pmatrix} 1 & i \\ 1 & j \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}\right].$$

As the determinant of $\begin{pmatrix} 1 & i \\ 1 & j \end{pmatrix}$ is nonzero,

$$\Pr_{a,b\in\mathbb{F}}[X_i = \alpha_1 \wedge X_j = \alpha_2] = \frac{1}{|\mathbb{F}|^2} = \Pr_{a,b\in\mathbb{F}}[X_i = \alpha_1] \cdot \Pr_{a,b\in\mathbb{F}}[X_j = \alpha_2].$$

To generalize the above construction for k-wise, the sample space is $(a_0, \ldots, a_{k-1}) \in \mathbb{F}^k$ and x_i for $i \in \mathbb{F}$ is $X_i = \sum_{t=0}^{k-1} a_t i^t$. It is not hard to see that this is indeed a k-wise independent sample space of size n^k.

What if we need $X_1, \ldots, X_n$ to be boolean and k-wise independent? One way is to use the previous construction and truncate every element X_i to, say, its least significant bit. We thus have:

Claim 6. *There exists an explicit distribution that is k-wise independent over $\{0,1\}^n$ and has support size n^k.*

Proof. Let $\mathbb{F} = \{a_0, \ldots, a_{n-1}\}$ be a field of size $n = 2^q$. It follows from the above discussion that the sample space $D = \left\{Ay \mid y \in \mathbb{F}^k\right\} \subseteq \mathbb{F}^n$ is k-wise independent over $\mathbb{F}$, where A is the $n \times k$ matrix for which $A_{i,j} = a_{i-1}^{j-1}$ (why?). Note that A is the generator matrix of a Reed-Solomon code, and also known as the Vandermonde matrix of the field elements.

Consider the canonical representation of every field element $a \in \mathbb{F}$ as a vector in $\mathbb{F}_2^q$. Addition in $\mathbb{F}$ is thus a simple addition over $\mathbb{F}_2^q$, whereas multiplication in $\mathbb{F}$ is a linear transformation. Namely, $y \mapsto \alpha \cdot y$ in $\mathbb{F}$ corresponds to $x \mapsto M_\alpha \cdot x$ in $\mathbb{F}_2^q$, where $M_\alpha \in \mathbb{F}_2^{q\times q}$. Under this representation, $Ay \in \mathbb{F}^n$ in mapped to $\bar{A}x \in \mathbb{F}_2^{nq}$ such that $x \in \mathbb{F}_2^{kq}$ encodes y_i in its i-th block and $\bar{A} \in \mathbb{F}_2^{nq\times kq}$ has $M_{A_{i,j}}$ as its (i,j)-th sub-matrix.

Our new sample space, $D' \subseteq \mathbb{F}_2^k$, is obtained by restricting every vector in $\left\{\bar{A}x \mid x \in \mathbb{F}_2^{kq}\right\}$ to n coordinates, e.g., by taking every other q coordinates. This specific construction corresponds to truncating every element of D to its least significant bit.

Take $I \subseteq [nq]$ of size k that fits our restriction. As the corresponding rows in $\bar{A}$ are independent, verify to yourself that indeed $(\bar{A}x)_I$ is uniform where x ranges over $\mathbb{F}_2^{kq}$. D' is of size $2^{kq} = n^k$, as desired. □

In fact we can do better. We will see that for pairwise independence. The sample space is $\{0,1\}^{\log n}$ and the distribution on the sample points is uniform. For every $i \in \{0,1\}^{\log n}$, we set $X_i(a) = \langle a, i\rangle \bmod 2$. The sample space is of size n. We will prove in the exercise that this is indeed a pairwise independent sample space. In fact, this bound is also tight:

Claim 7. *If $X_1, \ldots, X_n$ are boolean random variables that are pairwise independent then the support size is at least n.*

Proof. Consider the $S \times n$ matrix describing the distribution. Consider every column as some $v_i \in \mathbb{R}^S$, where we map every $b \in \{0,1\}$ to $(-1)^b$. We will show that the v_i-s are orthogonal and therefore independent, and this implies $S \geq n$.

For every $i \neq j$,

$$\begin{aligned} \langle v_i, v_j\rangle &= |\{k \in [S] \mid (v_i)_k = (v_j)_k\}| - |\{k \in [S] \mid (v_i)_k \neq (v_j)_k\}| \\ &= |S| \cdot \Pr[v_i = v_j] - |S| \cdot \Pr[v_i \neq v_j] \;=\; 2|S|\left(\Pr[v_i = v_j] - \frac{1}{2}\right) \;=\; 0. \end{aligned}$$

□

In fact, a more general lower bound can be given:

Theorem 8. *If $X_1, \ldots, X_n$ are boolean random variables that are k-wise independent then the support size is at least $\sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \binom{n}{i} \approx n^{\frac{k}{2}}$.*

Proof. As an exercise. □

3 Deterministic amplification

Most of the material in this section (and a lot that is not in this section) is covered in a survey of Goldreich [3] and a monograph of Luby and Wigderson [5].

BPP is the class of decision problems solvable by a probabilistic Turing machine in polynomial time with a two-sided bounded error. RP and coRP are its one-sided variants. Formally:

Definition 9. *For $a < b$, a language $L \in \mathsf{BPP}[a, b]$ if there exists a polynomial-time probabilistic TM $M(x, y)$, where:*

- *If $x \in L$ then $\Pr_y[M(x, y) = 1] \geq b$.*
- *If $x \notin L$ then $\Pr_y[M(x, y) = 1] \leq a$.*

We denote $\mathsf{BPP} = \mathsf{BPP}[\frac{1}{3}, \frac{2}{3}]$, $\mathsf{RP} = \mathsf{BPP}[0, \frac{1}{2}]$ and $\mathsf{coRP} = \mathsf{BPP}[\frac{1}{2}, 1]$.

Suppose we have $L \in \mathsf{BPP}[a - \varepsilon, a + \varepsilon]$, for some constant a and $\varepsilon = \varepsilon(n)$, accepted by a TM M that on input of length n uses $t(n)$ random bits. If we run M k times, each time with fresh, independent, random bits and eventually output according to whether the average of k answers exceeded a, the error probability should decrease exponentially.

If we denote X_i as the answer in the i-th run, when $x \in L$ we err if $\frac{1}{k} \sum_{i=1}^{k} X_i < a$. By Chernoff, the probability for this to happen is bounded by $e^{-\Omega(\varepsilon^2 k)}$. Likewise for $x \notin L$. Thus, to bring the error to δ, we can take $k = O(\frac{\log \frac{1}{\delta}}{\varepsilon^2})$. Thus, we can amplify any polynomially large gap $\varepsilon = n^{-\alpha}$ to an exponentially small error $\delta = 2^{-n^c}$ in polynomial time, and therefore also using polynomially many random bits. The question we ask is whether we can re-use random bits and reduce the error without using too many additional random bits.

Throughout, we are given x and a black-box access to $M(x, y)$. We are allowed to pick $y_1, \ldots, y_T$ in some way, and answer according to $M(x, y_1), \ldots, M(x, y_T)$. Denote $m = |y|$. So far we have seen that with independent trials, with T queries and mT random coins we can amplify $(\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon)$ to $(\delta, 1 - \delta)$ error with $T = O(\frac{1}{\varepsilon^2} \log \frac{1}{\delta})$.

3.1 Via pair-wise independence

Let us start with $k = 2$. Pick $y_1, \ldots, y_T$ from a pairwise independent distribution where each y_i is uniform over $\Sigma = \{0, 1\}^m$. For every $i \in [T]$, let Y_i be the boolean random variable that is 1 iff

$M(x, y_i)$ answered correctly. Denote $\mu_i = \mathbb{E}[Y_i] \geq \frac{1}{2} + \varepsilon$. We answer according to the median of the T trials. By Chebyshev and pairwise independence,

$$\begin{aligned}\Pr[\text{we are wrong}] &\leq \Pr\left[\left|\sum_{i=1}^{T} Y_i - \mu_i\right| \geq \varepsilon T\right] \\ &\leq \frac{\mathrm{Var}[\sum_i Y_i]}{\varepsilon^2 T^2} \leq \frac{(\frac{1}{2} - \varepsilon)(\frac{1}{2} + \varepsilon)}{\varepsilon^2 T} \leq \frac{1}{\varepsilon^2 T} = \delta.\end{aligned}$$

We thus choose $T = \frac{1}{\varepsilon^2 \delta}$. The sample space is of size at most 2^{2m} so overall $2m$ random coins are used. If we want to amplify a non-negligible gap to a constant gap, it is sufficient to use pairwise independence.

3.2 Via k-wise independence

We proceed with $k = 4$. For every $i \in [T]$, let X_i be the output of the i-th run and let $X = \sum_i X_i$, $\mu_i = \mathbb{E}[X_i]$ and $\mu = \sum_i \mu_i$. By Markov,

$$\Pr[|X - \mu| \geq A] \leq \Pr[(X - \mu)^4 \geq A^4] \leq \frac{\mathbb{E}[(X - \mu)^4]}{A^4}.$$

Denote $Z_i = X_i - \mu_i$, $\mathbb{E}[Z_i] = 0$. By linearity,

$$\mathbb{E}[(X - \mu)^4] = \mathbb{E}[(\sum_i Z_i)^4] = \sum_{i_1, i_2, i_3, i_4} \mathbb{E}[Z_{i_1} Z_{i_2} Z_{i_3} Z_{i_4}].$$

By four-wise independence, whenever all i_1, i_2, i_3, i_4 are different, $\mathbb{E}[Z_{i_1} Z_{i_2} Z_{i_3} Z_{i_4}] = E[Z_{i_1}] \cdot E[Z_{i_2}] \cdot E[Z_{i_3}] \cdot E[Z_{i_3}]$. However, for every i, $E[Z_i] = 0$, and so the term vanishes. In fact, this is true for every term i_1, i_2, i_3, i_4 in which some term appears with an odd power. Thus, the only terms that survive are those where every term appears an even number of times. Thus,

$$\begin{aligned}\mathbb{E}[(X - \mu)^4] &= \sum_a \mathbb{E}[Z_a^4] + \binom{4}{2} \sum_{1 \leq a < b \leq T} \mathbb{E}[Z_a^2]\,\mathbb{E}[Z_b^2] \\ &= \sum_a \mathbb{E}[Z_a^4] + \binom{4}{2} \sum_{1 \leq a < b \leq T} \mathrm{Var}[Z_a]\,\mathrm{Var}[Z_b].\end{aligned}$$

As for every i, $\mathrm{Var}[Z_i] = \mu_i(1 - \mu_i) \leq 1$,

$$\mathbb{E}[(X - \mu)^4] \leq T + \binom{4}{2}\binom{T}{2} \leq 4T^2.$$

We then obtain:

$$\begin{aligned}\Pr[\text{we are wrong}] &\leq \Pr\left[\left|\sum_{i=1}^{T} Y_i - \mu_i\right| \geq \varepsilon T\right] \\ &\leq \frac{\mathbb{E}[(X - \mu)^4]}{\varepsilon^4 T^4} \leq \frac{4T^2}{\varepsilon^4 T^4} = \frac{4}{\varepsilon^4 T^2} = \delta.\end{aligned}$$

So, with four-wise independence, we get an error of $O(T^{-2})$. Specifically, we take $T = \frac{2}{\varepsilon^2}\sqrt{\frac{1}{\delta}}$. For arbitrary $2k$-independence, similar analysis shows that the error decreases like $O(T^{-k})$.

Lemma 10. *Let X be the average of T k-wise independent random variables for an even integer k, and let $\mu = \mathbb{E}[X]$. Then,*

$$\Pr[|X-\mu| \geq \varepsilon] \;\leq\; \left(\frac{k^2}{4T\varepsilon^2}\right)^{\frac{k}{2}}.$$

The situation we have so far:

Table 1: Amplifying $(\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon)$ to $(\delta, 1-\delta)$ if r random bits are initially required

	Number of samples	Number of random bits
Truly random	$O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$	$r \cdot O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$
k-wise independence	$O(\frac{1}{\varepsilon^2}\frac{k^2}{\delta^{\frac{2}{k}}})$	$O(kr + k\log\frac{1}{\varepsilon} + \log\frac{1}{\delta})$
Pairwise independence	$O(\frac{1}{\varepsilon^2}\frac{1}{\delta})$	$O(r + \log\frac{1}{\delta\varepsilon})$

3.3 Via expanders

We start with a one-sided error $(0,\alpha)$ algorithm. With full independence, $O(\frac{1}{\alpha}\log\frac{1}{\delta})$ trials are sufficient (Check, and compare to the two sided error). Now, consider an expander $G = (V = \{0,1\}^m, E)$ with a constant degree D and a constant $\lambda = \min\left\{\lambda_2(G), -\lambda_{|V|}(G)\right\} < 1$.

The construction: Choose y_1 uniformly at random and take a random walk on G of length $T-1$ to obtain $y_2, \ldots, y_T$. Accept iff one of $M(x, y_i)$ accepted. Fix $x \in \{0,1\}^n$. If $x \notin L$ then we always reject, so we assume from now on that $x \in L$. Let $Bad \subseteq \{0,1\}^m$ be the set of strings that are bad for x. That is, $Bad = \{y \in \{0,1\}^m \mid M(x,y) = 0\}$. Thus,

$$\Pr[\text{we are wrong}] \;=\; \Pr\left[\bigwedge_{i=1}^{T}(y_i \in Bad)\right].$$

Then:

Theorem 11. *Using our above notations,*

$$\Pr\left[\bigwedge_{i=1}^{T}(y_i \in Bad)\right] \;\leq\; (\beta + (1-\beta)\lambda)^T,$$

where $\beta = \frac{|Bad|}{|V|}$.

In our case, $\beta \leq \alpha$ and $(\beta + (1-\beta)\lambda) = 1-(1-\lambda)(1-\beta) < 1$. Thus, with $m + \log D \cdot (T-1) = m + O(T)$ random coins we can amplify, say, $(0, \frac{1}{2})$ to $(0, 1-2^{-\Omega(T)})$.

Proof. The proof has two main components. First, we need to translate the condition $\bigwedge_{i=1}^{T}(y_i \in Bad)$ to an algebraic terminology, and then we analyze it.

The translation to algebraic terminology. Let M be the transition matrix of G and denote $|V| = 2^m = N$. Pick $y_1 \in V$ uniformly at random. That is, the initial distribution over the vertices is $u = \frac{1}{N}\mathbf{1}_N$. Define an $N \times N$ diagonal matrix B with $B[y,y] = 1$ if $y \in Bad$ and

0 otherwise. In this terminology, $|\langle \mathbf{1}, Bu\rangle|$ is the probability a random element belongs to BAD (and so is β). $|\langle \mathbf{1}, BMBu\rangle|$ is the probability in a random walk of length two, both samples belong to BAD. Similarly, $|\langle \mathbf{1}, (BM)^k Bu\rangle|$ is the probability in a random walk of length $k+1$ the walk is confined to the set BAD, i.e., all samples belong to BAD.

Reducing the analysis to understanding a single step : As B is a projection, $B^2 = B$, and so $(BM)^k Bu = (BMB)^k Bu$. Also, the vector is supported only on coordinates from Bad, Cauchy-Schwartz implies

$$|\langle \mathbf{1}, (BMB)^T Bu\rangle| \quad \leq \quad \sqrt{\beta N} \, \| (BMB)^T Bu \|_2$$

and since $\| AB \|_2 \leq \| A \|_2 \| B \|_2$,

$$\begin{aligned} |\langle \mathbf{1}, (BMB)^T Bu\rangle| &\leq \sqrt{\beta N} \, \| BMB \|_2^T \, \| Bu \|_2 \\ &= \sqrt{\beta N} \sqrt{\frac{\beta}{N}} \, \| BMB \|_2^T \\ &= \beta \, \| BMB \|_2^T \leq \| BMB \|_2^T . \end{aligned}$$

Summing up, it is enough to show $\| BMB \|_2 < 1$, i.e., it is enough to analyze a single step.

Thus, we are left with analyzing a single step. We will show, $\| BMB \|_2 \leq \beta + (1-\beta)\lambda$.

Claim 12 ([6], Proposition 3.2)**.** *Let G be an undirected regular graph on n vertices, with $\lambda = \min\left\{\lambda_2(G), -\lambda_{|V|}(G)\right\}$ and its transition matrix is B. Then, $B = (1-\lambda)J + \lambda E$ for some E with $\| E \|_2 \leq 1$ and J that is the normalized all-ones matrix. I.e., B is a convex combination of J (that corresponds to a completely random walk) and E (that is some arbitrary error matrix).*

Proof. The first eigenvector of B is u the all one vector (possibly normalized) with eigenvalue 1. u is also an eigenvector of J with eigenvalue 1. We conclude that u is a common eigenvector of B, J and E and with eigenvalue 1 for all of them (Check!).

What about vectors in the orthogonal complement? Let $W^{\perp}$ denote all vectors perpendicular to x, i.e., all x such that $\langle x, u\rangle = 0$. Then $Jx = 0$ (Why?). Also, $W^{\perp}$ is invariant under B (Why?). Thus, $W^{\perp}$ is invariant also under E (Why?).

Thus, to bound the norm of E, it is enough to limit attention to $W^{\perp}$. For $v \in W^{\perp}$, $\| Ev \| = \frac{1}{\lambda} \| Av \| \leq \frac{\lambda}{\lambda} \| v \| = \| v \|$. Thus, $\| E \|_2 \leq 1$. $\square$

Now, let us express BMB in this decomposition. We get

$$BMB = B((1-\lambda)J + \lambda E)B \; = \; (1-\lambda)BJB + \lambda BEB$$

The BJB part is the part corresponding to a true random walk step, the other part is "junk", and indeed we easily see that $\| BEB \|_2 \leq \| B \|_2 \| E \|_2 \| B \|_2 \leq 1$. Thus, we are now reduced to analyzing BJB, i.e., one true random walk step. For any $x \neq 0$, $x = \sum_i x_i e_i$. Then, $(BJBx)[i] = \frac{1}{N}\sum_{i\in Bad} x_i$ if $i \in Bad$ and 0 otherwise (check!). Thus, by Cauchy-Schwarz,

$$\| BJBx \|_2 = \sqrt{\beta N\left(\frac{1}{N}\sum_{i\in Bad} x_i\right)^2} = \sqrt{\frac{\beta}{N}}\sum_{i\in Bad} x_i \leq \sqrt{\frac{\beta}{N}}\sqrt{\beta N}\,\| x \|_2 = \beta,$$

which completes the proof. □

The two-sided error case is along the same ideas, but a bit more complicated. The analysis may use the useful *expander Chernoff bound.*

Theorem 13. *Let G be an undirected D-regular graph with $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n$ and spectral gap $1-\bar{\lambda}$ and let $f_i : V \to [0,1]$ for $i \in [T]$. Take a random walk $v_1, \ldots, v_T$ and let X_i be the random variable $f_i(v_i)$. Denote $\mu_i = \mathbb{E}[X_i]$ and $\bar{\mu} = \frac{1}{T}\sum_i \mu_i$. Then,*

$$\Pr\left[\left|\frac{1}{T}\sum_i X_i - \bar{\mu}\right| \geq \varepsilon\right] \;\leq\; 2e^{-\frac{1}{4}(1-\bar{\lambda})\varepsilon^2 T}.$$

We can then add the expander walk technique to our table, obtaining:

Table 2: Amplifying $(\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon)$ to $(\delta, 1-\delta)$ if r random bits are initially required

	Number of samples	Number of random bits
Truly random	$O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$	$r \cdot O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$
Expander walk	$O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$	$r + O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$
k-wise independence	$O(\frac{1}{\varepsilon^2}\frac{k^2}{\delta^{\frac{2}{k}}})$	$O(kr + k\log\frac{1}{\varepsilon} + \log\frac{1}{\delta})$
Pairwise independence	$O(\frac{1}{\varepsilon^2}\frac{1}{\delta})$	$O(r + \log\frac{1}{\delta\varepsilon})$

3.4 Via dispersers

We continue with the one-sided error. Let $E : [N] \times [T] \to [M]$ be a (K, α) seeded disperser. The construction: Pick $\bar{y} \in [N]$ uniformly at random and for every $i \in [T]$ choose $y_i = E(\bar{y}, i)$. As usual, accept if and only if some $M(x, y_i)$ accepts.

Suppose we start with a $(0, \alpha)$ error algorithm. If $x \notin L$ then we always reject. If $x \in L$ let $Good = \{y \in [M] \mid M(x,y) = 1\}$, so $|Good| \geq \alpha \cdot 2^m$. Let B be the set

$$B = \{\bar{y} \mid \Gamma(\bar{y}) \cap Good = \emptyset\}.$$

By the disperser property $|B| < K$ (Why?? This is the central point of the proof, so if you don't see it, insist on it until you see it). We reject iff we sampled $\bar{y} \in B$. Thus,

$$\Pr[\text{we reject}] \;\leq\; \frac{K}{N}.$$

The number of random coins used is $\log N$. Say $\alpha = \frac{1}{2}$. An optimal disperser exists with $T = O(\ln\frac{N}{K})$, so $O(\log\frac{1}{\delta})$ samples are sufficient to amplify the error to $(0, 1-\delta)$.

The comparison for one-sided error is given by:

Table 3: Amplifying $(0, \varepsilon)$ to $(\delta, 1-\delta)$ if r random bits are initially required

	Number of samples	Number of random bits
Truly random	$O(\frac{1}{\varepsilon}\log\frac{1}{\delta})$	$r \cdot O(\frac{1}{\varepsilon}\log\frac{1}{\delta})$
Expander walk	$O(\frac{1}{\varepsilon}\log\frac{1}{\delta})$	$r + O(\frac{1}{\varepsilon}\log\frac{1}{\delta})$
Disperser (optimal)	$O(\frac{1}{\varepsilon}\log\frac{1}{\delta})$	$r + O(\log\frac{\varepsilon}{\delta})$

3.5 Via extractors

We return to the two-sided case, and assume that we start with an $(\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon)$ error algorithm. Let $E : [N] \times [T] \to [M]$ be a (k, ε) extractor. The construction: Pick $\bar{y} \in [N]$ uniformly at random and for every $i \in [T]$ choose $y_i = E(\bar{y}, i)$. Accept if and only if the majority of the $M(x, y_i)$ accepted.

Fix x and let $Good = \{y \in [M] \mid M(x, y) \text{ answers correctly}\}$. We know that $\mu(Good) \geq \frac{1}{2} + \varepsilon$. Let $Bad = \left\{\bar{y} \in [N] \mid \Pr_{i\in[T]}[E(\bar{y}, i) \in Good] < \frac{1}{2}\right\}$. That is, $\bar{y} \in Bad$ if and only if the majority is incorrect and we err. Assume to the contrary that $|Bad| \geq 2^k = K$ and let X_B be the uniform distribution over Bad, so $H_\infty(X_B) \geq k$. On one hand, we have $|E(X_B, U_t) - U_m| \leq \varepsilon$. On the other hand, note that

$$\Pr_{\bar{y}\in Bad, i\in[T]}[E(\bar{y}, i) \in Good] \;<\; \frac{1}{2},$$

and as $\mu(Good) \geq \frac{1}{2} + \varepsilon$, we have that $|E(X_B, U_T) - U_M| < \varepsilon$, in contradiction.

Thus, $|Bad| < K$ so the probability that we pick a bad $\bar{y}$ is again at most $\frac{K}{N} = \delta$. The number of random coins used is $\log N$.

Say $\varepsilon = \frac{1}{6}$. An optimal extractor exists with $T = O(\ln\frac{N}{K})$, so $O(\log\frac{1}{\delta})$ samples are sufficient to amplify the error to $(\delta, 1-\delta)$, assuming $M = O(KT)$. Observe our final comparison:

Table 4: Amplifying $(\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon)$ to $(\delta, 1-\delta)$ if r random bits are initially required

	Number of samples	Number of random bits
Truly random	$O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$	$r \cdot O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$
Extractor (optimal)	$O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$	$r + O(\log\frac{1}{\delta})$
Expander walk	$O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$	$r + O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$
k-wise independence	$O(\frac{1}{\varepsilon^2}\frac{k^2}{\delta^{\frac{2}{k}}})$	$O(kr + k\log\frac{1}{\varepsilon} + \log\frac{1}{\delta})$
Pairwise independence	$O(\frac{1}{\varepsilon^2}\frac{1}{\delta})$	$O(r + \log\frac{1}{\delta\varepsilon})$

4 Approximating frequency moments in small space

Definition 14. *A family $\mathcal{H} \subseteq [n] \to \Sigma$ is a k-universal family of hash functions if for any $1 \leq i_1 < \ldots < i_k \leq n$, for all $\sigma_1, \ldots, \sigma_k \in \Sigma$,*

$$\Pr_{h\in\mathcal{H}}[h(i_1) = \sigma_1 \wedge \ldots h(i_k) = \sigma_k] \;=\; \frac{1}{|\Sigma|^k}.$$

Equivalently, if we define random variables n random varibles $X_1, \dots, X_n$ defined by uniformly sampling $h \in \mathcal{H}$ and setting $X_i = h(i)$, then $X_1, \dots, X_n$ are k-wise independent.

Consider a "stream" of inputs $x_1, \dots, x_n \in \Sigma$. For every $a \in \Sigma$, let m_a denote the number of times a occurs. We want to approximate $F_2 = \sum_a m_a^2$ by allowing only a single pass over the inputs. We will achieve an arbitrary constant accuracy using $O(\log(n|\Sigma|))$ space. The result is due to Alon, Matias and Szegedy [1].

The algorithm is as follows:

1. Fix a 4-universal family of hash functions $\mathcal{H} \subseteq \Sigma \to \{-1, 1\}$.
2. Pick $h_1, \dots, h_T \in H$ for some T that we shall soon determine.
3. For each $t = 1$ to T, compute $s_t = \sum_{i=1}^n h_t(x_i)$.
4. Output $\frac{1}{T} \sum_{i=1}^T s_t^2$.

The space complexity is easy. We need T counters. Each counter counts up to n, with $O(\log n)$ bits. Each $h \in H$ is represented by $O(\log|\Sigma|)$ bits (why?).

We now turn to estimating the accuracy (and confidence) of this approximation method. Before we start we notice that $s_t = \sum_{i=1}^n h_t(x_i) = \sum_a m_a h_t(a)$. Thus, if an element appears many times the values $h_t(x_i)$ are more correlated than the case where, say, each element appears once. Now,

$$\mathbb{E}[s_t] \; = \; \sum_{i=1}^{n} \mathbb{E}[h(x_i)] \; = \; 0$$

and due to pairwise independence,

$$\begin{aligned} \mathbb{E}[s_t^2] \; &= \; \sum_{a,b} m_a m_b \, \mathbb{E}[h(a)h(b)] \\ &= \; \sum_a m_a^2 \, \mathbb{E}[h^2(a)] + \sum_{a \neq b} m_a m_b \, \mathbb{E}[h(a)] \, \mathbb{E}[h(b)] \; = \; \sum_a m_a^2 \; = \; F_2. \end{aligned}$$

This means that we use an *unbiased estimator* for F_2, .i.e., a random variable whose average is correct. We are now left with estimating how concentrated is the random variable s_t^2 around its mean.

Note that $s_1, \dots, s_T$ are independent. Let $Y_i = s_i^2$, and we know that $\mathbb{E}[Y_i] = F_2$. We want to say that $\Pr\left[\left|\frac{1}{T}\sum_{i=1}^T Y_i - F_2\right| \geq \varepsilon F_2\right]$ is small. By Chebyshev's inequality,

$$\Pr\left[\left|\frac{1}{T}\sum_{i=1}^{T} Y_i - F_2\right| \geq \varepsilon T F_2\right] \; \leq \; \frac{\operatorname{Var}\left[\sum_{i=1}^T Y_i\right]}{\varepsilon^2 T^2 F_2^2} \; = \; \frac{T \operatorname{Var}[Y_1]}{\varepsilon^2 T^2 F_2^2}.$$

We are back to a single hash function. Computing the variance, we have

$$\operatorname{Var}[Y_1] \; = \; \mathbb{E}[s_1^4] - \left(\mathbb{E}[s_1^2]\right)^2 \; = \; \mathbb{E}[s_1^4] - F_2^2.$$

We compute the fourth moment using 4-wise independence:

$$\begin{aligned}\mathbb{E}[s_1^4] &= \sum_{a,b,c,d\in\Sigma} m_a m_b m_c m_d \,\mathbb{E}[h(a)h(b)h(c)h(d)] \\ &= \sum_a m_a^4 \,\mathbb{E}[h^4(a)] + 3\sum_{a\neq b} m_a^2 m_b^2 \,\mathbb{E}[h^2(a)h^2(b)] \\ &= 3\sum_{a,b} m_a^2 m_b^2 - 2\sum_a m_a^4 \;=\; 3F_2^2 - 2F_4,\end{aligned}$$

so $\mathrm{Var}[Y_1] = 2(F_2^2 - F_4) \leq 2F_2^2$. Hence:

$$\Pr\left[\left|\frac{1}{T}\sum_{i=1}^{T} Y_i - F_2\right| \geq \varepsilon F_2\right] \leq \frac{2T^2F_2^2}{\varepsilon^2 T^2 F_2^2} = \frac{2}{\varepsilon^2 T} \leq \frac{1}{3},$$

for $T \geq \frac{6}{\varepsilon^2}$.

So far, with $O(\frac{1}{\varepsilon^2}\log(n|\Sigma|))$ space, we have a confidence of $\frac{1}{3}$. So far (and if we are only interested in constant confidence) we could have worked with $h_1, \ldots, h_T$ that are chosen in a pairwise independent manner.

If we want to improve the confidence to an arbitrary δ we can repeat the above procedure K independent times and take the median. Trial i succeeds if the answer is within ε from F_2. By Chernoff, the probability that $\frac{1}{2}$ of the trials are unsuccessful is at most $2^{-\Omega(K)} = \delta$. If half are successful, the median is also good (why?).

References

[1] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 20–29. ACM, 1996.

[2] Vasek Chvátal. The tail of the hypergeometric distribution. *Discrete Mathematics*, 25(3):285–287, 1979.

[3] Oded Goldreich. A sample of samplers: A computational perspective on sampling. *def*, 1:2n, 1997.

[4] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[5] Michael Luby and Avi Wigderson. *Pairwise independence and derandomization.* Citeseer, 1995.

[6] Eyal Rozenman and Salil Vadhan. Derandomized squaring of graphs. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 436–447. Springer, 2005.