Computational Indistinguishability: A Sample Hierarchy

Oded Goldreich[*] Madhu Sudan[†]

March 28, 1998

Abstract

We consider the existence of pairs of probability ensembles which may be efficiently distinguished given k samples but cannot be efficiently distinguished given $k' < k$ samples. It is well known that in any such pair of ensembles it cannot be that both are efficiently computable (and that such phenomena cannot exist for non-uniform classes of distinguishers, say, polynomial-size circuits). It was also known that there exist pairs of ensembles which may be efficiently distinguished based on *two* samples but cannot be efficiently distinguished based on a *single* sample. In contrast, it was not known whether the distinguishing power increases when one moves from two samples to polynomially-many samples.

We show the existence of pairs of ensembles which may be efficiently distinguished given $k+1$ samples but cannot be efficiently distinguished given k samples, where k can be any function bounded above by a polynomial in the security parameter.

In course of establishing the above result, we prove several technical lemmas regarding polynomials and graphs. We believe that these may be of independent interest.

Keywords: Computational Indistinguishability, Polynomials, Extractors, Expanders, Eigenvalues and Rapid Mixing.

[*]Department of Computer Science, Weizmann Institute of Science, Rehovot, ISRAEL. E-mail: oded@wisdom.weizmann.ac.il. Currently visiting the Laboratory for Computer Science of MIT, and partially supported by DARPA grant DABT63-96-C-0018.

[†]Laboratory for Computer Science, MIT, 545 Technology Sq., Cambridge, MA 02139, USA. E-mail: madhu@theory.lcs.mit.edu.

1 Introduction

Computational indistinguishability, introduced by Goldwasser and Micali [7] and defined in full generality by Yao [11], is a central concept of complexity theory. Two probability ensembles, $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$, where both X_n and Y_n range over $\{0,1\}^n$, are said to be **indistinguishable by a complexity class** if for every machine M in the class the difference $d_M(n) \stackrel{\text{def}}{=} |\Pr(M(X_n)=1) - \Pr(M(Y_n)=1)|$ is a negligible function in n (i.e., decreases faster than $1/p(n)$ for any positive polynomial p).

We stress that, in the definition recalled above, the distinguishing machine (i.e., M) obtains a single sample (from either distributions), and casts its "verdict" based on this sample. An important and natural question is what happens when the distinguishing machine is given several samples. It is well known that in several cases (see below), computational indistinguishability is preserved also when many samples are given to the distinguisher. That is, in these cases, if two ensembles are computationally indistinguishable by a single sample then they are also computationally indistinguishable by (polynomially) many samples. Two important cases where this happens are:

1. When the two probability ensembles are polynomial-time computable, and one considers probabilistic polynomial-time distinguishers. (An ensemble $\{Z_n\}_{n\in\mathbb{N}}$ is polynomial-time computable if there exists a probabilistic polynomial-time sampling algorithm, S, such that $S(1^n)$ and Z_n are identically distributed.)

2. When one considers computational indistinguishability with respect to the class of non-uniform polynomial-size circuits.

In both cases the proof amounts to using the multi-sample distinguisher to derive a single-sample distinguisher, by incorporating copies of the two ensembles into the single-sample distinguisher (cf., [5, 4]). This is possible using the fact that the class of distinguishers is able to generate samples from each of the two ensembles.

However, it has been shown that the above may fail in certain other cases (cf., [3, 8, 6]). Specifically, there exists a pair of (non-efficiently computable) ensembles which, on one hand, are computationally indistinguishable by (uniform) probabilistic polynomial-time algorithms which take a *single sample*, while on the other hand, can be distinguished in polynomial-time given *two samples*.

It has been unknown whether separations as above may exists between distinguishability based on, say, 2 samples and 3 samples. Furthermore, it was not known if there is a separation between 2 samples and polynomially many samples.

We show a separation between k samples and $k+1$, for any polynomially-bounded function $k : \mathbb{N} \mapsto \mathbb{N}$. That is, there exist a pair of probability ensembles which are (polynomial-time) indistinguishable based on k samples and yet can be distinguished (in polynomial-time) given $k+1$ samples.

2 Formal Setting

In this paper we call $\mathbf{P} = \{P_n\}_{n\in\mathbb{N}}$ a *probability ensemble* if, for some polynomially-bounded length function $\ell : \mathbb{N} \mapsto \mathbb{N}$, P_n is a distribution on the set of strings of length $\ell(n)$. The corresponding (to the length function ℓ) *uniform ensemble*, denoted $\mathbf{U} = \{U_n\}_{n\in\mathbb{N}}$, has each U_n uniformly distributed over $\{0,1\}^{\ell(n)}$. A function, $\mu : \mathbb{N} \mapsto [0,1]$, is called *negligible* if for every positive polynomial p and all sufficiently large n's, $\mu(n) < 1/p(n)$. The latter definition is naturally coupled with the association of efficient computation with polynomial-time algorithms: An event "occurs negligibly" if it cannot be observed after a feasible (i.e., expected polynomial) number of trials.

Definition 2.1 (indistinguishability by k samples): *Let $k : \mathbb{N} \mapsto \mathbb{N}$ be any polynomially bounded function, and $\mathbf{P} = \{P_n\}_{n\in\mathbb{N}}$ and $\mathbf{Q} = \{Q_n\}_{n\in\mathbb{N}}$ be a pair of probability ensembles. The ensembles $\mathbf{P}$ and $\mathbf{Q}$ are said to be* indistinguishable by k samples *if for every probabilistic polynomial-time machine M the function*

$$d_M(n) \stackrel{\text{def}}{=} |\Pr(M(\overline{P}_n^{k(n)}) = 1) - \Pr(M(\overline{Q}_n^{k(n)}) = 1)|$$

is negligible, where $\overline{P}_n^{k(n)}$ (resp., $\overline{Q}_n^{k(n)}$) represents $k(n)$ independent copies of P_n (resp., Q_n).

A "strong" negation of the notion of indistinguishability is presented by the notion of distinguishability. A function, $\mu : \mathbb{N} \mapsto [0,1]$, is called *noticeable* if there exists a positive polynomial p so that for all sufficiently large n's, $\mu(n) > 1/p(n)$. We stress that the two notions do not complement one another, but rather leave a gap in-between, since the underlying notions of negligible and noticeable are not complementary. Clearly, a negligible function is not noticeable, but there are functions $\mu : \mathbb{N} \mapsto [0,1]$ which are neither negligible nor noticeable (e.g., $\mu(n) = 1$ if n is even and 0 otherwise).

Definition 2.2 (distinguishability by k samples): *Let $k : \mathbb{N} \mapsto \mathbb{N}$, $\mathbf{P} = \{P_n\}_{n\in\mathbb{N}}$ and $\mathbf{Q} = \{Q_n\}_{n\in\mathbb{N}}$ be as in Definition 2.1 above. The ensembles $\mathbf{P}$ and $\mathbf{Q}$ are said to be* distinguishable by k samples *if there exists a probabilistic polynomial-time machine M so that the function d_M, defined as above, is noticeable.*

Theorem 2.3 (main result): *Let $k : \mathbb{N} \mapsto \mathbb{N}$ be any polynomially bounded function. Then, there exists a probability ensemble, $\mathbf{P} = \{P_n\}_{n\in\mathbb{N}}$, where P_n ranges over strings of length $2n$, so that*

1. Indistinguishability by k samples: The ensemble $\{P_n\}_{n\in\mathbb{N}}$ is indistinguishable from the uniform ensemble, $\mathbf{U} = \{U_n\}_{n\in\mathbb{N}}$, by k samples. *Furthermore, for any probabilistic Turing machine M which takes k samples, and for all sufficiently large n's,*

$$|\Pr(M(\overline{P}_n^{k(n)}) = 1) - \Pr(M(\overline{U}_n^{k(n)}) = 1)| \; < \; 2^{-\Omega(n)}$$

where $\overline{P}_n^{k(n)}$ (resp., $\overline{U}_n^{k(n)}$) represents $k(n)$ independent copies of P_n (resp., U_n).

2. Polynomial-time distinguishability by $k+1$ samples: The ensemble $\{P_n\}_{n\in\mathbb{N}}$ is distinguishable from the uniform ensemble $\mathbf{U}$ by $k+1$ samples. *Furthermore, there exists a deterministic polynomial-time machine M such that for all sufficiently large n's,*

$$|\Pr(M(\overline{P}_n^{k(n)+1}) = 1) - \Pr(M(\overline{U}_n^{k(n)+1}) = 1)| \; > \; \frac{1}{3}$$

where $\overline{P}_n^{k(n)+1}$ (resp., $\overline{U}_n^{k(n)+1}$) represents $k(n)+1$ independent copies of P_n (resp., U_n).

Furthermore, P_n can be generated by a probabilistic circuit of size polynomial in n. In case one only wishes to fool probabilistic polynomial-time *distinguishers* (in item 1)*, the n^{th} circuit can be constructed in time $e(n)$, where $e : \mathbb{N} \mapsto \mathbb{N}$ is any function which grows faster than 2^{n^c}, for every $c > 0$.*

Thus, with respect to uniform computations (and general ensembles which may not be polynomial-time computable), the "sample hierarchy" is strict. We comment that one may also construct a pair of probability ensembles, $\mathbf{P} = \{P_n\}_{n\in\mathbb{N}}$ and $\mathbf{Q} = \{Q_n\}_{n\in\mathbb{N}}$ such that both satisfy the above theorem and furthermore

$$|\Pr(M(\overline{P}_n^{k(n)+1}) = 1) - \Pr(M(\overline{Q}_n^{k(n)+1}) = 1)| \ > \ 1 - 2^{-\Omega(n)}$$

where M is as in Item 2 above.

3 Proof of Main Result

We prove Theorem 2.3 by first studying a problem concerning polynomials of low degree over a big finite field.

3.1 Typical Polynomials

Standard Notations: Let F be a finite field. Denote by F_d the set of polynomials of degree at most d over F.

Less Standard Notations: For $\overline{x} = (x_1, ..., x_k) \in \mathrm{F}^k$ (i.e., each x_i in F), we extend the definition of polynomials so that, for any polynomial p, we have $p(\overline{x}) = (p(x_1), ..., p(x_k))$.

Motivating Discussion. Clearly, for every $f : (\mathrm{F}^k)^2 \mapsto [0, 1]$,

$$\mathrm{E}_{\overline{x}\in\mathrm{F}^k, p\in\mathrm{F}_{k-1}}(f(\overline{x}, p(\overline{x}))) \ \approx \ \mathrm{E}_{\overline{x},\overline{y}\in\mathrm{F}^k}(f(\overline{x}, \overline{y}))$$

Equality would hold if $\overline{x}$ was uniformly selected among the set of k-sequences consisting of k distinct elements of F. For such $\overline{x}$'s, the sequence $p(\overline{x})$ is uniformly selected over F^k, given that p is uniformly distributed in F_{k-1}. It is appealing to conjecture that there exists a polynomial $p \in \mathrm{F}_{k-1}$ so that

$$\mathrm{E}_{\overline{x}\in\mathrm{F}^k}(f(\overline{x}, p(\overline{x}))) \ \approx \ \mathrm{E}_{\overline{x},\overline{y}\in\mathrm{F}^k}(f(\overline{x}, \overline{y}))$$

However, as shown below (see Proposition 3.5), this is false. Instead, we consider degree k polynomials which are examined at k arguments (rather than at $k + 1$ arguments). In this case, we show that for every $f : (\mathrm{F}^k)^2 \mapsto [0, 1]$ most polynomials $p \in \mathrm{F}_k$ satisfy

$$\mathrm{E}_{\overline{x}\in\mathrm{F}^k}(f(\overline{x}, p(\overline{x}))) \ \approx \ \mathrm{E}_{\overline{x},\overline{y}\in\mathrm{F}^k}(f(\overline{x}, \overline{y}))$$

We call such polynomials (f, k)-typical. More generally,

Definition 3.1 (typical functions): *Let $k \in \mathbb{N}$, $\epsilon \in [0, 1]$ and $f : \mathrm{F}^k \times \mathrm{F}^k \mapsto [0, 1]$. A function $g : \mathrm{F} \mapsto \mathrm{F}$ is called (f, k, ϵ)-typical if*

$$|\mathrm{E}_{\overline{x}\in\mathrm{F}^k}(f(\overline{x}, g(\overline{x}))) \ - \ \mathrm{E}_{\overline{x},\overline{y}\in\mathrm{F}^k}(f(\overline{x}, \overline{y}))| \ < \ \epsilon$$

Following the above discussion we will consider an arbitrary $f : (\mathrm{F}^k)^2 \mapsto [0, 1]$ and prove

1. For some absolute constant $c > 0$ the following holds. For every finite field F, $k < |\mathrm{F}|^{1/c}$ and every $f : (\mathrm{F}^k)^2 \mapsto [0, 1]$ all but at most an $|\mathrm{F}|^{-c}$ fraction of the degree k polynomials are $(f, k, |\mathrm{F}|^{-c})$-typical. (See Lemma 3.2.)

2. For every finite field F and every $k < \sqrt{|\mathrm{F}|/10}$ there exists a (polynomial-time computable) function $f : (\mathrm{F}^k)^2 \mapsto [0,1]$ so that no degree $k-1$ polynomial is $(f, k, 0.4)$-typical. (See Proposition 3.5.)

Using the above, Theorem 2.3 is proven by standard diagonalization. The high level plan is as follows. Using parameter n, we consider $\mathrm{F} = \mathrm{GF}(2^n)$, and wish to fool the first $t(n)$ (e.g., $t(n) = n$) probabilistic machines which takes $k(n)$ samples. These machines give rise to $t(n)$ functions f_i as above, and by Item 1 there exists a degree $k(n)$ polynomial, denoted p, which is $(f_i, k(n), 2^{-\Omega(n)})$-typical for all i's. Using p, we define the n^{th} distribution, denoted P_n, as $(x, p(x))$ where x is uniformly distributed over F, and infer that none of the above machines can distinguish $k(n)$ samples taken from P_n from $k(n)$ samples taken from the uniform distribution over pairs $\mathrm{F} \times \mathrm{F}$. On the other hand, by Item 2 (substituting k for $k(n)+1$), there exists a polynomial-time algorithm which distinguishes $k(n)+1$ samples from P_n from $k(n)+1$ samples taken from the uniform distribution. For details see Section 3.4.

3.2 Almost all degree k polynomials are k-typical

The most involved technical part of this work is proving that for any $f : (\mathrm{F}^k)^2 \mapsto [0,1]$ most degree k polynomials are $(f, k, |\mathrm{F}|^{-\Omega(1)})$-typical. That is,

Lemma 3.2 *There exists a constant $c > 0$ so that for every $f : (\mathrm{F}^k)^2 \mapsto [0,1]$, setting $\mu \stackrel{\mathrm{def}}{=} \mathrm{E}_{\overline{x},\overline{y} \in \mathrm{F}^k}(f(\overline{x}, \overline{y}))$ and $\epsilon \stackrel{\mathrm{def}}{=} \frac{k^{1/c}}{c \cdot |\mathrm{F}|^c}$ the following holds*

$$\mathrm{Pr}_{p \in \mathrm{F}_k}\left(|\mathrm{E}_{\overline{x} \in \mathrm{F}^k}(f(\overline{x}, p(\overline{x}))) - \mu| \;>\; \epsilon\right) \;<\; \epsilon$$

The lemma is proven in the next section. As a warm-up we prove that for any such f most degree $2k-1$ polynomials are $(f, k, |\mathrm{F}|^{-\Omega(1)})$-typical. This suffices to establish a weaker version of Theorem 2.3 (i.e., separating distinguishability by k samples from distinguishability by $2k$ samples).

Lemma 3.3 *Let $f : (\mathrm{F}^k)^2 \mapsto [0,1]$, and $\mu \stackrel{\mathrm{def}}{=} \mathrm{E}_{\overline{x},\overline{y} \in \mathrm{F}^k}(f(\overline{x}, \overline{y}))$. Then, for any $\epsilon > 0$*

$$\mathrm{Pr}_{p \in \mathrm{F}_{2k-1}}\left(|\mathrm{E}_{\overline{x} \in \mathrm{F}^k}(f(\overline{x}, p(\overline{x}))) - \mu| \;>\; \epsilon\right) \;<\; \frac{k^2}{\epsilon^2 \cdot |\mathrm{F}|}$$

Proof: Consider the probability space of all possible choices of $p \in \mathrm{F}_{2k-1}$ with uniform distribution. Define random variables (over this probability space) so that $\zeta_{\overline{x}} \stackrel{\mathrm{def}}{=} f(\overline{x}, p(\overline{x}))$, for every $\overline{x} \in \mathrm{F}^k$. The claim of the lemma can be rephrased as

$$\mathrm{Pr}_{p \in \mathrm{F}_{2k-1}}\left(\left|\sum_{\overline{x} \in \mathrm{F}^k} \zeta_{\overline{x}} - |\mathrm{F}|^k \cdot \mu\right| > \epsilon \cdot |\mathrm{F}|^k\right) \;<\; \frac{k^2}{\epsilon^2 \cdot |\mathrm{F}|} \qquad (1)$$

This will be established by applying Chebyshev's inequality to the ζ_x's. Specifically, we will show that the expected value of the sum of the $\zeta_{\overline{x}}$'s is approximately $|\mathrm{F}|^k \cdot \mu$, and that with high probability the sum of the ζ_x's is close to its expected value. In showing the latter we will use the fact that the ζ_x's are "almost pairwise independent" (as in [1, Sec. 4.3]).

Fact 3.3.1: $\left|\,|\mathrm{F}|^k \cdot \mu \;-\; \sum_{\overline{x} \in \mathrm{F}^k} \mathrm{E}(\zeta_{\overline{x}})\,\right| \;<\; \frac{k^2}{2 \cdot |\mathrm{F}|}$.

Proof: For every $\overline{x} = (x_1, ..., x_k) \in \mathrm{F}^k$ with $|\{x_1, ..., x_k\}| = k$, we have

$$\begin{aligned} \mathrm{E}(\zeta_{\overline{x}}) &= \mathrm{E}_{p\in \mathrm{F}_{2k-1}}(f(\overline{x}, p(\overline{x}))) \\ &= \mathrm{E}_{\overline{y}\in \mathrm{F}^k}(f(\overline{x}, \overline{y})) \end{aligned}$$

since for such an $\overline{x} = (x_1, ..., x_k)$ the values $p(x_1), ..., p(x_k)$ are uniformly and independently distributed in F. Observe that the fraction of $\overline{x}$'s consisting of k distinct x_i's is at least $1 - \binom{k}{2} \cdot |\mathrm{F}|^{-1}$, and so

$$\begin{aligned} \sum_{\overline{x}\in \mathrm{F}^k} \mathrm{E}(\zeta_{\overline{x}}) &= |\mathrm{F}|^k \cdot \mathrm{E}_{\overline{x},\overline{y}\in \mathrm{F}^k}(f(\overline{x}, \overline{y})) \pm \binom{k}{2} \cdot |\mathrm{F}|^{k-1} \\ &= |\mathrm{F}|^k \cdot \left(\mu \pm \binom{k}{2} \cdot |\mathrm{F}|^{-1}\right) \end{aligned}$$

as claimed. □

Fact 3.3.2:

$$\Pr\left(\left|\sum_{\overline{x}\in \mathrm{F}^k} \zeta_{\overline{x}} - \sum_{\overline{x}\in \mathrm{F}^k} \mathrm{E}(\zeta_{\overline{x}})\right| > \frac{\epsilon}{2} \cdot |\mathrm{F}|^k\right) < \frac{k^2}{\epsilon^2 \cdot |\mathrm{F}|}$$

Proof: We first observe that for every $\overline{x} \in \mathrm{F}^k$, for all but at most a $\binom{k}{2}/|\mathrm{F}|$ fraction of the $\overline{y}$'s in F^k, the random variables $\zeta_{\overline{x}}$ and $\zeta_{\overline{y}}$ are independent. This follows since these random variables are independent whenever the sequences $\overline{x}$ and $\overline{y}$ have no common element. (Here we use the hypothesis that the probability space is uniform over the set of polynomials of degree $2k - 1$ over F. For such a random polynomial, p, the sequence $p(x_1), ..., p(x_k), p(y_1), ..., p(y_k)$ is uniformly distributed over F^{2k}.) Now applying Chebyshev's inequality (cf., [1]), we have

$$\begin{aligned} \Pr\left(\left|\sum_{\overline{x}\in \mathrm{F}^k} \zeta_{\overline{x}} - \sum_{\overline{x}\in \mathrm{F}^k} \mathrm{E}(\zeta_{\overline{x}})\right| > \frac{\epsilon}{2} \cdot |\mathrm{F}|^k\right) &< \frac{\mathrm{VAR}(\sum_{\overline{x}\in \mathrm{F}^k} \zeta_{\overline{x}})}{(\epsilon/2)^2 \cdot |\mathrm{F}|^{2k}} \\ &< \frac{4 \cdot \sum_{\overline{x}} \mathrm{VAR}(\zeta_{\overline{x}})}{\epsilon^2 \cdot |\mathrm{F}|^{2k}} + \frac{4 \cdot \sum_{\overline{x}\neq\overline{y}} \mathrm{COV}(\zeta_{\overline{x}}, \zeta_{\overline{y}})}{\epsilon^2 \cdot |\mathrm{F}|^{2k}} \end{aligned}$$

Now, as usual, the first term is upper bounded by $4 \cdot |\mathrm{F}|^k \cdot \frac{1/4}{\epsilon^2 \cdot |\mathrm{F}|^{2k}} = \frac{1}{\epsilon^2 \cdot |\mathrm{F}|^k} \leq \frac{1}{\epsilon^2 \cdot |\mathrm{F}|}$. As for the second term, let $I_{\overline{x}}$ denote the set of $\overline{y}$'s for which $\zeta_{\overline{x}}$ and $\zeta_{\overline{y}}$ are stochastically independent. By the above observation we have $\frac{|I_{\overline{x}}|}{|\mathrm{F}^k|} > 1 - \frac{k^2-1}{|\mathrm{F}|}$, and by definition $\mathrm{COV}(\zeta_{\overline{x}}, \zeta_{\overline{y}}) = 0$ for every $\overline{y} \in I_{\overline{x}}$. Thus, the second term is bounded by

$$\begin{aligned} 4 \cdot \sum_{\overline{x}\neq\overline{y}} \frac{\mathrm{COV}(\zeta_{\overline{x}}, \zeta_{\overline{y}})}{\epsilon^2 \cdot |\mathrm{F}|^{2k}} &< 4 \cdot \sum_{\overline{x}\in \mathrm{F}^k} \sum_{\overline{y}\in \mathrm{F}^k} \frac{\mathrm{COV}(\zeta_{\overline{x}}, \zeta_{\overline{y}})}{\epsilon^2 \cdot |\mathrm{F}|^{2k}} \\ &< 4 \cdot \sum_{\overline{x}\in \mathrm{F}^k} \frac{|\mathrm{F}^k \setminus I_{\overline{x}}| \cdot (1/4)}{\epsilon^2 \cdot |\mathrm{F}|^{2k}} \\ &< \frac{k^2 - 1}{\epsilon^2 \cdot |\mathrm{F}|} \end{aligned}$$

The claimed bound follows by combining the bounds for the two terms. □

We may assume that $\frac{k^2}{\epsilon^2 \cdot |\mathrm{F}|} \leq 1$ and $\epsilon < 1$ (or else the lemma holds vacuously). It follows that $\frac{k^2}{2|\mathrm{F}|} \leq \frac{\epsilon^2}{2} < \frac{\epsilon}{2}$. Thus, combining the two facts, the lemma follows. Specifically, by Fact 3.3.1 $\left|\, |\mathrm{F}|^k \cdot \mu - \sum_{\overline{x}\in \mathrm{F}^k} \mathrm{E}(\zeta_{\overline{x}}) \right| < \frac{\epsilon}{2}$, and using Fact 3.3.2 – Eq. (1) follows. ■

Instantiating the above lemma (using $\epsilon = |\mathrm{F}|^{-1/3}$), we have

Corollary 3.4 *Let f be as above, and $k \leq \sqrt[6]{|\mathrm{F}|}$. Then for all but a $|\mathrm{F}|^{-1/6}$ fraction of p's in F_{2k-1}*

$$|\mathrm{E}_{\overline{x}\in\mathrm{F}^k}(f(\overline{x},p(\overline{x}))) - \mathrm{E}_{\overline{x},\overline{y}\in\mathrm{F}^k}(f(\overline{x},\overline{y}))| < |\mathrm{F}|^{-1/3}$$

That is, all but a $|\mathrm{F}|^{-1/6}$ fraction of the degree $2k-1$ polynomials over F are $(f,k,|\mathrm{F}|^{-1/3})$-typical.

3.3 No degree $k-1$ polynomial is k-typical

In contrast to Lemma 3.2 (as well as to the weaker Lemma 3.3), we have

Proposition 3.5 *There exists an* (efficiently computable) *function f so that for any polynomial $p \in \mathrm{F}_{k-1}$*

$$|\mathrm{E}_{\overline{x}\in\mathrm{F}^k}(f(\overline{x},p(\overline{x}))) - 0.5| \quad > \quad 0.5 - \frac{k^2}{|\mathrm{F}|} \tag{2}$$

$$\mathrm{E}_{\overline{x},\overline{y}\in\mathrm{F}^k}(f(\overline{x},\overline{y})) \quad = \quad 0.5 \tag{3}$$

Proof: Consider any easily recognizable set, S, containing exactly half the elements of F. Consider the algorithm f, which given k pairs, denoted $(x_1,y_1),...,(x_k,y_k)$, finds a (typically unique) degree $k-1$ polynomial p' satisfying $p'(x_i) = y_i$, for $i = 1,...,k$. (In case there are several possibilities, the algorithm selects p' uniformly among them.) The algorithm outputs 1 if $p'(0) \in S$ and 0 otherwise. (Here is where we use the hypothesis that S is an easily recognizable set.)

Consider any $p \in \mathrm{F}_{k-1}$, and suppose that the algorithm is given k random pairs with $y_i = p(x_i)$. With probability greater than $1 - k^2 \cdot |F|^{-1}$, we have $|\{x_1,...,x_k\}| = k$ and so the extrapolated polynomial (i.e., p') equals p. In such a case the algorithm's output is determined by the predicate $p(0) \in S$ and so is identically zero or identically one. Thus, Eq. (2) follows.

However, when the y_i's are uniformly selected, the value of the extrapolated degree $k-1$ polynomial p' at any fixed point (e.g., $p'(0)$) is uniformly distributed. Thus the algorithm's output is uniformly distributed in $\{0,1\}$, and Eq. (3) follows. ■

3.4 Using Typical Polynomials

Using Lemma 3.3 and Proposition 3.5, we can prove the existence of probability ensembles which are indistinguishable from the uniform ensemble by k samples but distinguishable from it by $2k$ samples. More generally, we have the following lemma.

Lemma 3.6 *Let $t : \mathbb{N} \mapsto \mathbb{N}$ be any non-decreasing and unbounded function, and $k,k' : \mathbb{N} \mapsto \mathbb{N}$ be two polynomially-bounded functions so that $k(n) < k'(n)$ for every n. Suppose that for some $c > 0$ and any function $f : (\mathrm{GF}(2^n)^{k(n)})^2 \mapsto [0,1]$ all but at most a $1/2t(n)$ fraction of the degree $k'(n)-1$ polynomials over $\mathrm{GF}(2^n)$ are $(f,k(n),2^{-cn})$-typical. Then, there exists probability ensembles, $\mathbf{P} = \{P_n\}_{n\in\mathbb{N}}$ and $\mathbf{Q} = \{Q_n\}_{n\in\mathbb{N}}$, where P_n* (resp. *Q_n*) *ranges over strings of length $2n$ and can be generated by a probabilistic circuit of size* poly(n), *so that*

1. *The ensemble $\mathbf{P}$ is indistinguishable from the uniform ensemble, $\mathbf{U} = \{U_n\}_{n\in\mathbb{N}}$, by k samples. Furthermore, for any probabilistic Turing machine M which takes k samples,*

$$|\Pr(M(\overline{P}_n^{k(n)}) = 1) - \Pr(M(\overline{U}_n^{k(n)}) = 1)| \;<\; 2^{-\Omega(n)}$$

where $\overline{P}_n^{k(n)}$ (resp., $\overline{U}_n^{k(n)}$) are as in Theorem 2.3. Same for $\mathbf{Q}$.

2. *The ensemble* **P** *is distinguishable from the uniform ensemble* **U** *by* k' *samples. Furthermore, there exists a deterministic polynomial-time machine* M *such that*

$$|\Pr(M(\overline{P}_n^{k'(n)}) = 1) - \Pr(M(\overline{U}_n^{k'(n)}) = 1)| \;>\; \frac{1}{2} - 2^{-\Omega(n)}$$

Same for **Q**. *Furthermore,*

$$|\Pr(M(\overline{P}_n^{k'(n)}) = 1) - \Pr(M(\overline{Q}_n^{k'(n)}) = 1)| \;>\; 1 - 2^{-\Omega(n)}$$

Theorem 2.3 follows by combining the above lemma (using $k'(n) = k(n) + 1$) with Lemma 3.2, whereas a weaker statement with $k'(n) = 2k(n)$ follows by combining the above lemma with Corollary 3.4. In both cases we may set $t : \mathbb{N} \mapsto \mathbb{N}$ to be any non-decreasing and unbounded function so that $t(n) < 2^{n/O(1)}$ (e.g., $t(n) = n$ or $t(n) = \log n$ will do, alas the hypothesis holds even for $t(n) = 2^{n/O(1)}$).

Proof: We construct P_n by considering the first $t(n)$ machines in an enumeration of probabilistic Turing machines. For each such machine, M, we define $f_M(\alpha, \beta) \stackrel{\rm def}{=} \Pr(M(\alpha, \beta) = 1)$.[1] By the hypothesis, for each such M, all but at most $1/2t(n)$ of the polynomials, p, of degree $k'(n) - 1$ over $\mathrm{F} = \mathrm{GF}(2^n)$ satisfy

$$|\mathrm{E}_{\overline{x}\in \mathrm{F}^{k(n)}}(f_M(\overline{x}, p(\overline{x}))) - \mathrm{E}_{\overline{x},\overline{y}\in \mathrm{F}^{k(n)}}(f_M(\overline{x}, \overline{y}))| \leq 2^{-cn} \tag{4}$$

Thus, for more than half of the polynomials, p, of degree $k'(n) - 1$ over F it holds that for each of the first $t(n)$ machines, M,

$$|\mathrm{E}_{\overline{x}\in \mathrm{F}^{k(n)}}(\Pr(M(\overline{x}, p(\overline{x}))=1)) - \mathrm{E}_{\overline{x},\overline{y}\in \mathrm{F}^{k(n)}}(\Pr(M(\overline{x}, \overline{y})=1))| \leq 2^{-n/3} \tag{5}$$

In particular, let fix an arbitrary polynomial $p \in \mathrm{F}_{k'(n)-1}$ satisfying Eq. (5) (for all these M's) so that $p(0)$ is one of the first 2^{n-1} elements of F (by some standard enumeration). Such a polynomial does exist since exactly half of the polynomials satisfy the latter condition and less than half do not satisfy the former. Similarly, we fix $q \in \mathrm{F}_{k'(n)-1}$ satisfying Eq. (5) so that $q(0)$ is one of the last 2^{n-1} elements of F.

Using this polynomial p, we define P_n to be uniformly distributed over $\{(x, p(x)) : x \in \mathrm{GF}(2^n)\}$. Similarly, Q_n is defined to be uniformly distributed over $\{(x, q(x)) : x \in \mathrm{GF}(2^n)\}$.

By Eq. (5), Item 1 of the lemma holds. To establish Item 2, we use the algorithm of Proposition 3.5: We extrapolate a degree $k'(n) - 1$ polynomial, based on the given $k'(n)$ samples, and test whether its free term is one of the first 2^{n-1} elements of F. Clearly, the answer is almost always YES when given $k'(n)$ samples from P_n, whereas it is almost always NO when given $k'(n)$ samples from Q_n. (Here "almost always" means with probability $1 - 2^{-\Omega(n)}$.) The answer is YES with probability $\frac{1}{2}$ when given $k'(n)$ samples from the uniform distribution over $\{0, 1\}^{2n}$. The lemma follows. ■

4 Proof of Lemma 3.2

Our proof consists of the following four steps:

[1] We slightly abuse notation here. The input to M is a sequence of k pairs, $(\alpha_1, \beta_1), ..., (\alpha_k, \beta_k)$, and so we actually have $f_M(\alpha_1, ..., \alpha_k, \beta_1, ..., \beta_k) = \Pr(M((\alpha_1, \beta_1), ..., (\alpha_k, \beta_k))=1)$.

1. We consider a bipartite graph in which edges link left-side vertices of the form $(\overline{x},\overline{y}) \in (\mathrm{F}^k)^2$ with right-side vertices $p \in \mathrm{F}_k$ iff $p(\overline{x}) = \overline{y}$. We *claim* that for any $f : (\mathrm{F}^k)^2 \mapsto [0,1]$, for almost all $p \in \mathrm{F}_k$ the average of f over the neighbors of p approximates the average of f over all $(\mathrm{F}^k)^2$.

2. We consider an auxiliary multi-graph (having parallel edges and self-loops) over the vertex set F_k with edges representing paths of length 2 in the former graph. We show that a good upper bound on the second eigenvalue of the auxiliary graph implies the former claim.

3. Reversing the well-known connection between eigenvalues and rapid-mixing, we show that the rapid-mixing of a random walk on a graph implies a good upper bound on the second eigenvalue of the graph. (This part has appeared implicitly in many works.)

4. Finally, we show that a random walk on the auxiliary graph is sufficiently rapidly mixing (to yield a good enough bound on the second eigenvalue).

Initial simplification. We assume throughout that $k^2 < |\mathrm{F}|$ (as otherwise Lemma 3.2 holds vacuously). Recall that Lemma 3.2 asserts that for some $\epsilon_0, \delta_0 \stackrel{\mathrm{def}}{=} O(k^{1/c} \cdot |\mathrm{F}|^{-c})$, all but at most an ϵ_0 fraction of the k degree polynomials are (f, k, δ_0)-typical. This statement refers to expectation taken over all $\overline{x}$'s in F^k. As we have seen in the previous section, it is more convenient to consider only $\overline{x} = (x_1, ..., x_k)$'s consisting of distinct x_i's. Let $\mathrm{F}^{(k)}$ denote the set of such sequences, that is

$$\mathrm{F}^{(k)} \stackrel{\mathrm{def}}{=} \{(x_1, ..., x_k) \in \mathrm{F}^k : x_i \neq x_j \ (\forall i \neq j)\} \tag{6}$$

Then, Lemma 3.2 would follow if we establish, for $\epsilon_1 = \epsilon_0$ and $\delta_1 = \delta_0 - \frac{k^2}{|\mathrm{F}|}$, that all but at most a ϵ_1 fraction of the k degree polynomials satisfy

$$|\mathrm{E}_{\overline{x}\in\mathrm{F}^{(k)}}(f(\overline{x}, p(\overline{x}))) - \mathrm{E}_{\overline{x}\in\mathrm{F}^{(k)},\overline{y}\in\mathrm{F}^k}(f(\overline{x},\overline{y}))| \leq \delta_1 \tag{7}$$

(Lemma 3.2 follows since the difference between expectation taken over $\overline{x} \in \mathrm{F}^k$ and expectation taken over $\overline{x} \in \mathrm{F}^{(k)}$ is at most $\binom{k}{2} \cdot |\mathrm{F}|^{-1}$.) From this point on, we consider probability spaces where $\overline{x}$ is uniformly distributed over $\mathrm{F}^{(k)}$.

The bipartite graph $\mathcal{G}_{\mathrm{F},k}$. We consider a bipartite graph, denoted $\mathcal{G}_{\mathrm{F},k}$, with vertex set $\mathcal{U}_{\mathrm{F},k} \cup \mathcal{V}_{\mathrm{F},k}$, where $\mathcal{U}_{\mathrm{F},k} \stackrel{\mathrm{def}}{=} \mathrm{F}^{(k)} \times \mathrm{F}^k$ and $\mathcal{V}_{\mathrm{F},k} \stackrel{\mathrm{def}}{=} \mathrm{F}_k$. The edge set of the graph, denoted $\mathcal{E} \subset \mathcal{U}_{\mathrm{F},k} \times \mathcal{V}_{\mathrm{F},k}$, consists of pairs $((\overline{x},\overline{y}), p)$ where $p(\overline{x}) = \overline{y}$. Clearly, each vertex $p \in \mathcal{V}_{\mathrm{F},k}$ has exactly $|\mathrm{F}^{(k)}|$ neighbours; specifically, its neighbour set, denoted $\Gamma(p)$, equals $\{(\overline{x}, p(\overline{x})) : \overline{x} \in \mathrm{F}^{(k)}\}$. Using the fact that $\overline{x}$ consists of distinct elements, we know that each vertex $(\overline{x},\overline{y}) \in \mathcal{U}_{\mathrm{F},k}$ has exactly $|\mathrm{F}|$ neighbours, corresponding to the $|\mathrm{F}|$ degree k polynomials p's which satisfy $p(\overline{x}) = \overline{y}$. Thus, Eq. (7) can be rephrased as asserting that all but at most an ϵ_1 fraction of $v \in \mathcal{V}_{\mathrm{F},k}$ satisfy

$$\left| \frac{1}{|\Gamma(v)|} \sum_{u\in\Gamma(v)} f(u) - \frac{1}{|\mathcal{U}_{\mathrm{F},k}|} \sum_{u\in\mathcal{U}_{\mathrm{F},k}} f(u) \right| \leq \delta_1 \tag{8}$$

Thus, our aim is to establish Eq. (8).

4.1 It suffices to show that $\mathcal{G}_{\mathrm{F},k}$ is a good extractor

Following Zuckerman [12], we observe that the above holds (i.e., at most an ϵ_1 fraction of $v \in \mathcal{V}_{\mathrm{F},k}$ violate Eq. (8)) in case $\mathcal{G}_{\mathrm{F},k}$ is an (ϵ_2, δ_2)-extractor, with $\epsilon_2 = \epsilon_1/2$ and $\delta_2 = \delta_1$.

Definition 4.1 (extractor): *The regular bipartite graph with edge set $E \subseteq U \times V$ is called an (ϵ, δ)-* extractor *if for every set $V' \subseteq V$ of cardinality $\epsilon \cdot |V|$, the distribution induced on U by uniformly selecting $v \in V'$ and $u \in \Gamma(v)$ is δ-close* (in variation distance) *to the uniform distribution on U.*

Lemma 4.2 [12]: *Suppose that a regular bipartite graph with edge set $E \subseteq U \times V$ is an (ϵ, δ)-extractor. Then, for every $f : U \mapsto [0, 1]$, for all but at most a 2ϵ fraction of $v \in V$*

$$\left| \frac{1}{|\Gamma(v)|} \sum_{u\in\Gamma(v)} f(u) - \frac{1}{|U|} \sum_{u\in U} f(u) \right| \leq \delta$$

Proof: Assuming on the contrary that the conclusion does not hold, we let V' be a set of $\epsilon \cdot |V|$ vertices v's for which, without loss of generality,

$$\frac{1}{|\Gamma(v)|} \sum_{u\in\Gamma(v)} f(u) - \frac{1}{|U|} \sum_{u\in U} f(u) > \delta$$

This implies that $\mathrm{E}_{u\in\Gamma(v)}(f(u)) - \mathrm{E}_{u\in U}(f(u)) > \delta$ holds for every $v \in V'$. Thus,

$$\mathrm{E}_{v\in V', u\in\Gamma(v)}(f(u)) - \mathrm{E}_{u\in U}(f(u)) > \delta$$

Letting X denote the distribution induced on U by uniformly selecting $v \in V'$ and $u \in \Gamma(v)$, and by Y the uniform distribution on U, we have $\mathrm{E}(f(X)) - \mathrm{E}(f(Y)) > \delta$. Defining $S \subset U$ so that $x \in S$ iff $\Pr(X = x) > \Pr(Y = x)$, and using the fact that f ranges over $[0, 1]$, we have

$$\begin{aligned}
\Pr(X \in S) - \Pr(Y \in S) &= \sum_{x\in S} (\Pr(X = x) - \Pr(Y = x)) \\
&\geq \sum_{x\in U} (\Pr(X = x) - \Pr(Y = x)) \cdot f(x) \\
&= \mathrm{E}(f(X)) - \mathrm{E}(f(Y)) > \delta
\end{aligned}$$

However, this contradicts the lemma's hypothesis, which asserts that the distribution X (i.e., uniformly selecting $v \in V'$ and $u \in \Gamma(v)$) is δ-close to Y (i.e., the uniform distribution on U). ■

Corollary 4.3 *Let $\epsilon_2, \delta_2 \stackrel{\mathrm{def}}{=} \frac{k^{1/c}}{2c\cdot|\mathrm{F}|^c}$ and suppose $c \leq 1/2$. If $\mathcal{G}_{\mathrm{F},k}$ is an (ϵ_2, δ_2)-extractor then Lemma 3.2 follows.*

Proof: By Lemma 4.2 and the setting of the parameters, the hypothesis implies Eq. (8), which in turn (by the above discussion) implies Lemma 3.2. (Note, $\epsilon_0 = \epsilon_1 = 2\epsilon_2 \leq k^{1/c} \cdot |\mathrm{F}^{-c}|/c$ and $\delta_0 = \delta_1 + k^2 \cdot |\mathrm{F}|^{-1} \leq k^{1/c} \cdot |\mathrm{F}^{-c}|/c$.) ■

4.2 The auxiliary graph $\mathcal{A}_{\mathrm{F},k}$ and the relevance of its eigenvalues

In order to show that $\mathcal{G}_{\mathrm{F},k}$ is a good extractor, we consider an auxiliary multi-graph with vertex set $\mathcal{V}$ and edge set corresponding to all possible paths of length 2 in $\mathcal{G}_{\mathrm{F},k}$. That is, for every $v, u \in \mathcal{V}$ and every path of length 2 in $\mathcal{G}_{\mathrm{F},k}$ between v and u (passing through a vertex in $\mathcal{U}_{\mathrm{F},k}$), we introduce an edge in the auxiliary multi-graph. We stress that this multi-graph, denoted $\mathcal{A}_{\mathrm{F},k}$, has $|\mathrm{F}^{(k)}|$ self-loops per each vertex, and that it is regular (with degree $|\mathrm{F}^{(k)}| \cdot |\mathrm{F}|$).

Let A denote the normalized adjacency matrix of $\mathcal{A}_{\mathrm{F},k}$ (i.e., $\mathcal{A}_{\mathrm{F},k}$'s adjacency matrix divided by its degree), and let $\lambda_{\mathrm{F},k}$ denote the second largest (in absolute value) eigenvalue of A. Then we have

Lemma 4.4 *Let $\lambda \stackrel{\mathrm{def}}{=} \lambda_{\mathrm{F},k}$ be as above. Then $\mathcal{G}_{\mathrm{F},k}$ is an $(\lambda^{1/3}, \lambda^{1/3})$-extractor.*

Proof: Let $\epsilon \stackrel{\mathrm{def}}{=} \lambda^{1/3}$, and suppose for contradiction that $\mathcal{G}_{\mathrm{F},k}$ is *not* an (ϵ, ϵ)-extractor. Then, there exists a set $V' \subset \mathcal{V}_{\mathrm{F},k}$ of cardinality at least $\epsilon \cdot |\mathcal{V}_{\mathrm{F},k}|$ so that the distribution induced on $\mathcal{U}_{\mathrm{F},k}$ by uniformly selecting $v \in V'$ and $u \in \Gamma(v)$ is ϵ-far (in variation distance) to the uniform distribution on U. Denoting by p_u the probability assigned to vertex $u \in \mathcal{U}_{\mathrm{F},k}$, the contradiction hypothesis yields

$$\sum_{u \in \mathcal{U}_{\mathrm{F},k}} |p_u - |\mathcal{U}_{\mathrm{F},k}|^{-1}| \; > \; 2\epsilon \tag{9}$$

On the other hand, denoting by $\Gamma(x)$ the neighbor set of any vertex x in $\mathcal{G}_{\mathrm{F},k}$, we have

$$p_u \;=\; \frac{1}{|V'|} \cdot \sum_{v \in V'} \frac{|\Gamma(v) \cap \{u\}|}{|\Gamma(v)|} \;=\; \frac{|\Gamma(u) \cap V'|}{|V'| \cdot (|\mathcal{U}_{\mathrm{F},k}| \cdot |\Gamma(u)| / |\mathcal{V}_{\mathrm{F},k}|)} \tag{10}$$

Considering a random walk of length 2 in $\mathcal{G}_{\mathrm{F},k}$, starting at a uniformly selected vertex $v \in V'$, we have

$$\begin{aligned}
\Pr_{v \in V', u \in \Gamma(v), v' \in \Gamma(u)}[v' \in V'] &= \sum_{u \in \mathcal{U}_{\mathrm{F},k}} p_u \cdot \Pr_{v' \in \Gamma(u)}[v' \in V'] \\
&= \sum_{u \in \mathcal{U}_{\mathrm{F},k}} p_u \cdot \frac{|\Gamma(u) \cap V'|}{|\Gamma(u)|} \\
&= \sum_{u \in \mathcal{U}_{\mathrm{F},k}} p_u^2 \cdot \frac{|V'| \cdot |\mathcal{U}_{\mathrm{F},k}|}{|\mathcal{V}_{\mathrm{F},k}|}
\end{aligned}$$

Looking at the same walk as a random edge in $\mathcal{A}_{\mathrm{F},k}$, and denoting by $\Gamma'(v)$ the neighbor multiset of a vertex v in $\mathcal{A}_{\mathrm{F},k}$, we have

$$\begin{aligned}
\Pr_{v \in \mathcal{V}_{\mathrm{F},k}, v' \in \Gamma'(v)}[v, v' \in V'] &= \frac{|V'|}{|\mathcal{V}_{\mathrm{F},k}|} \cdot \Pr_{v \in V', v' \in \Gamma'(v)}[v' \in V'] \\
&= \frac{|V'|^2}{|\mathcal{V}_{\mathrm{F},k}|^2} \cdot |\mathcal{U}_{\mathrm{F},k}| \cdot \sum_{u \in \mathcal{U}_{\mathrm{F},k}} (|\mathcal{U}_{\mathrm{F},k}|^{-1} + (p_u - |\mathcal{U}_{\mathrm{F},k}|^{-1}))^2 \\
&= \frac{|V'|^2}{|\mathcal{V}_{\mathrm{F},k}|^2} \cdot \left(1 + |\mathcal{U}_{\mathrm{F},k}| \cdot \sum_{u \in \mathcal{U}_{\mathrm{F},k}} (p_u - |\mathcal{U}_{\mathrm{F},k}|^{-1})^2\right)
\end{aligned}$$

Thus, using Eq. (9) and setting $N \stackrel{\rm def}{=} |\mathcal{U}_{\mathrm{F},k}|$, we have

$$\begin{aligned}\Pr_{v\in\mathcal{V}_{\mathrm{F},k}, v'\in\Gamma'(v)}[v, v' \in V'] &\geq \frac{|V'|^2}{|\mathcal{V}_{\mathrm{F},k}|^2} \cdot \left(1 + N \cdot \min_{x_i\geq 0,\, \sum_i x_i > 2\epsilon} \left\{\sum_{i=1}^{N} x_i^2\right\}\right) \\ &> \frac{|V'|^2}{|\mathcal{V}_{\mathrm{F},k}|^2} \cdot (1 + (2\epsilon)^2)\end{aligned}$$

However, as we shall shortly see, this contradicts the *Expander Mixing Lemma* (cf., Corollary 2.5 in [1, Chap. 9])[2], by which

$$\left|\Pr_{v\in\mathcal{V}_{\mathrm{F},k}, v'\in\Gamma'(v)}[v, v' \in V'] - \frac{|V'|^2}{|\mathcal{V}_{\mathrm{F},k}|^2}\right| < \lambda \cdot \frac{|V'|}{|\mathcal{V}_{\mathrm{F},k}|}$$

Specifically, we obtain $\frac{|V'|^2}{|\mathcal{V}_{\mathrm{F},k}|^2} \cdot (2\epsilon)^2 < \lambda \cdot \frac{|V'|}{|\mathcal{V}_{\mathrm{F},k}|}$, and so $\epsilon \cdot (2\epsilon)^2 < \lambda$. This, however, contradicts our setting of $\epsilon = \lambda^{1/3}$. The lemma follows. ■

Corollary 4.5 *Suppose that for some constant c, $\lambda_{\mathrm{F},k} \leq \frac{k^{3/c}}{(2c\cdot|\mathrm{F}|^c)^3}$. Then Lemma 3.2 holds with constant c.*

4.3 Reversing the eigenvalue connection

It is well-known that good upper bounds on the second eigenvalue of a (regular) graph yield rapid mixing (i.e., fast convergence of a random walk to the uniform distribution). The converse is less known, holds as well and has been used in various papers. In particular, the fact that the trace of the t^{th} power of the (normalized) adjacency matrix is the sum of the eigenvalues t^{th} powers [2], can be used to derive such a bound (Noga Alon, priv. comm.).[3] For sake of selfcontainment, we provide a proof of the desired result.

Lemma 4.6 *Consider a regular connected graph on N vertices, let A be its normalized adjacency matrix and λ_2 denote the absolute value of the second eigenvalue of A. Let t be an integer and Δ_t denote an upper bound on the maximum, taken over all possible start vertices v, of the difference in Norm2 between the distribution induced by a t-step random walk starting at v and the uniform distribution. Then $\lambda_2 \leq (N \cdot \Delta_t)^{1/t}$.*

Proof: Under the hypothesis all eigenvectors and eigenvalues are reals, and $\vec{e_1} \stackrel{\rm def}{=} \sqrt{N}\cdot(N^{-1}, ..., N^{-1})$ is the (normalized) eigenvector corresponding to the eigenvalue 1. Let $\vec{e_2}$ be the (normalized) eigenvector corresponding to λ_2, and consider the probability vector $\vec{p} \stackrel{\rm def}{=} (N^{-1}, ..., N^{-1}) + N^{-1} \cdot \vec{e_2}$. (The latter is a probability vector since the absolute value of any entry in $\vec{e_2}$ is bounded by 1.) Since $\vec{p}$ is in the convex hull of the probability vectors referred to in the hypothesis, the distance $\|A^t\vec{p} - (N^{-1}, ..., N^{-1})\|$ is bounded above by Δ_t. On the other hand,

$$\begin{aligned}\|A^t\vec{p} - (1/N, ..., 1/N)\| &= \frac{1}{N} \cdot \|A^t\vec{e_2}\| \\ &= \frac{1}{N} \cdot \lambda_2^t\end{aligned}$$

[2]The Expander Mixing Lemma refers to arbitrary sets A, B of vertices in a regular graph $G = (V, E)$ of normalized eigenvalue λ. It asserts that the absolute difference between $\frac{|(A\times B)\cap E|}{|E|}$ and $\frac{|A|}{|V|} \cdot \frac{|B|}{|V|}$ is at most $\lambda \cdot \frac{\sqrt{|A|\cdot|B|}}{|V|}$.

[3]In this case one may use a upper bound on the t-step "return probability" of random walks. Thus, an upper bound on the max-norm deviation of a t-step random walk from any start vertex implies an upper bound on the second eigenvalue. The hypothesis is thus weaker than the one we use below.

and so $\frac{\lambda_2^t}{N} \leq \Delta_t$. The lemma follows. ■

Corollary 4.7 *Suppose that for any vertex v in $\mathcal{A}_{\mathrm{F},k}$, the difference in Norm2 between the distribution induced by a $O(k)$-step random walk starting at v and the uniform distribution is at most $O(k)^{O(k)} \cdot |\mathrm{F}|^{-(2k+1)}$. Then, Lemma 3.2 follows.*

Proof: By Lemma 4.6, we have $\lambda_{\mathrm{F},k} \leq (|\mathrm{F}|^{-k})^{1/O(k)}$, and by Corollary 4.5 we are done. ■

4.4 Showing that the auxiliary graph is rapid-mixing

We conclude the proof of Lemma 3.2 by establishing the hypothesis of Corollary 4.7. That is, we consider an arbitrary fixed polynomial $p_0 \in \mathrm{F}_k = \mathcal{V}_{\mathrm{F},k}$ and a random walk of length $t \stackrel{\mathrm{def}}{=} O(k)$ on $\mathcal{A}_{\mathrm{F},k}$ starting at p_0, and prove that such a walk converges to the uniform distribution. That is,

Lemma 4.8 *Let $p_0 \in \mathrm{F}_k$ be any vertex in $\mathcal{A}_{\mathrm{F},k}$, and $t = 3k+1$. Then, the Norm2 difference between the distribution induced by a t-step random walk starting at v and the uniform distribution is at most $O(k)^{O(k)} \cdot |\mathrm{F}|^{-(2k+1)}$.*

Proof: For $i = 1, ..., t$, we denote by p_i a random variable representing the distribution after i steps of this walk. Note that p_i is derived from p_{i-1} by the following two step random process:

1. Uniformly select $\overline{\alpha_i} = (\alpha_{i,1}, ..., \alpha_{i,k}) \in \mathrm{F}^{(k)}$.
2. Uniformly select a polynomial p_i among the $|\mathrm{F}|$ polynomials p satisfying $p(\overline{\alpha_i}) = p_{i-1}(\overline{\alpha_i})$.

Expressing these degree k polynomials as polynomials in a formal variable x, we have

$$\begin{aligned} p_i(x) \;&=\; p_{i-1}(x) + r_i \cdot \prod_{j=1}^{k} (x - \alpha_{i,j}) \\ &\quad \text{where } r_i \text{ is uniformly selected in F} \end{aligned} \tag{11}$$

Using the symmetric functions

$$\sigma_j(z_1, ..., z_k) \stackrel{\mathrm{def}}{=} (-1)^j \sum_{S \subseteq [k]\ ,\ |S|=j} \prod_{i \in S} z_i,$$

we have

$$p_i(x) \;=\; p_{i-1}(x) + r_i \cdot \sum_{j=0}^{k} \sigma_j(\overline{\alpha_i}) \cdot x^j \tag{12}$$

Switching to vector notation, we write each p_i as a $(k+1)$-dimensional vector of random variables, denoted $\overline{p_i}$, and so have

$$\overline{p_i} \;=\; \overline{p_{i-1}} + r_i \cdot (\sigma_0(\overline{\alpha_i}), \sigma_1(\overline{\alpha_i}), ..., \sigma_k(\overline{\alpha_i}))^\top \tag{13}$$

Denoting $\overline{\sigma}_{\overline{\beta}} \stackrel{\mathrm{def}}{=} (\sigma_0(\overline{\beta}), \sigma_1(\overline{\beta}), ..., \sigma_k(\overline{\beta}))^\top$, we have $\overline{p_i} \;=\; \overline{p_{i-1}} + r_i \cdot \overline{\sigma}_{\overline{\alpha}_i}$, and so

$$\overline{p_t} \;=\; \overline{p_0} + \sum_{i=1}^{t} r_i \cdot \overline{\sigma}_{\overline{\alpha}_i} \tag{14}$$

Finally, we move to matrix notation: Letting $M(\overline{\alpha}_1, ..., \overline{\alpha}_t)$ denote the $(k+1)$-by-t matrix in which $\overline{\sigma}_{\overline{\alpha}_i}$ is the i^{th} column, and $\overline{r} \stackrel{\text{def}}{=} (r_1, ..., r_t)^\top$, we have

$$\overline{p_t} \;=\; \overline{p_0} + M(\overline{\alpha}_1, ..., \overline{\alpha}_t) \cdot \overline{r} \tag{15}$$

Since $t \geq k+1$ and $\overline{r}$ is uniformly distributed in F^t, the random variable $\overline{p_t}$ is uniformly distributed in F_k provided that the matrix $M(\overline{\alpha}_1, ..., \overline{\alpha}_t)$ has full rank. Thus, the Norm2 (as well as any other norm) distance of $\overline{p_t}$ from the uniform probability distribution (over F_k) is bounded above by twice the probability that $M(\overline{\alpha}_1, ..., \overline{\alpha}_t)$ is not of full rank, where the probability is taken over the choices of the $\overline{\alpha}_i$'s. Thus,

Fact 4.8.1: The lemma follows if the probability, over $\overline{\alpha}_i$'s chosen uniformly and independently from $\mathrm{F}^{(k)}$, that the matrix $M(\overline{\alpha}_1, ..., \overline{\alpha}_t)$ does not have full rank is bounded above by $(2k)^{O(k)} \cdot |\mathrm{F}|^{-(2k+1)}$.

On the other hand, the hypothesis of Fact 4.8.1 follows by establishing that with high probability, as long as the matrix does not have full rank, its rank increases with any additional column. Let us establish the latter fact first. That is,

Fact 4.8.2: Let $\overline{\alpha}_1, ..., \overline{\alpha}_i \in \mathrm{F}^{(k)}$ be fixed so that the matrix $M(\overline{\alpha}_1, ..., \overline{\alpha}_i)$ does not have full rank. Then, for uniformly chosen $\overline{\beta} \in \mathrm{F}^{(k)}$, with probability at least $1 - 2k \cdot |\mathrm{F}|^{-1}$, the matrix $M(\overline{\alpha}_1, ..., \overline{\alpha}_i, \overline{\beta})$ has higher rank than the matrix $M(\overline{\alpha}_1, ..., \overline{\alpha}_i)$.

Proof: We use the well know fact by which the rank of a matrix is r if and only if it contains an r-by-r sub-matrix having a non-zero determinant. Suppose that $M(\overline{\alpha}_1, ..., \overline{\alpha}_i)$ has rank $r \leq k$, and let A denote a corresponding r-by-r (non-singular) sub-matrix. Let j be an arbitrary row not included in A (such a row exists as $r < k+1$), and using the formal variables $\overline{z} = (z_1, ..., z_k)$ (with each z_ℓ ranging over F), consider the formal matrix $F(z_1, ..., z_k) \stackrel{\text{def}}{=} M(\overline{\alpha}_1, ..., \overline{\alpha}_i, \overline{z})$. Actually, we consider the $(r+1)$-by-$(r+1)$ sub-matrix, denoted $F'(z_1, ..., z_k)$, of $F(z_1, ..., z_k)$ encompassing the sub-matrix A, the j^{th} row and the last column (of F). Recall that the first r columns of $F'(\overline{z})$ are elements of F, whereas the last column contains $r+1$ distinct symmetric functions $\sigma_\ell(\overline{z})$'s. That is, the elements of the last column are homogeneous polynomials in distinct degrees in the range $\{0, 1, ..., k\}$. Developing the determinant of $F'(z_1, ..., z_k)$ according to the last column we have

1. The determinant of $F'(z_1, ..., z_k)$ is a polynomial in $z_1, ..., z_k$ of total degree at most k.
2. The determinant of $F'(z_1, ..., z_k)$ is not zero. This follows by noting that
   (a) the expression obtained for the determinant contains the term $\det(A) \cdot \sigma_j(\overline{z})$, where $\det(A) \in \mathrm{F} \setminus \{0\}$ denotes the determinant of A);
   (b) whereas the term above is of degree j no other term in the expression has degree j.

Thus, by Schwartz's Lemma [10], the probability that for uniformly chosen $\overline{\beta} \in \mathrm{F}^k$, the determinant of $F'(\overline{\beta})$ is zero is bounded above by $k/|\mathrm{F}|$. However, in our case $\overline{\beta}$ is uniformly chosen in $\mathrm{F}^{(k)}$, and so the bad event occurs with probability at most $1/\mathrm{Pr}_{\overline{\beta} \in \mathrm{F}^k}[\overline{\beta} \in \mathrm{F}^{(k)}] < 2$ times bigger. The current fact follows. □

Using Fact 4.8.2, the probability that the matrix $M(\overline{\alpha}_1, ..., \overline{\alpha}_t)$ does not have full rank is bounded above by

$$\begin{aligned} \sum_{i=0}^{k} \binom{t}{i} \cdot (2k \cdot |\mathrm{F}|^{-1})^{t-i} \;&<\; 2^t \cdot (2k \cdot |\mathrm{F}|^{-1})^{t-k} \\ &=\; 2^{5k+2} \cdot k^{2k+1} \cdot |\mathrm{F}|^{-(2k+1)} \end{aligned}$$

Using Fact 4.8.1, the lemma follows. ■

Acknowledgments

Oded Goldreich is grateful to Bernd Meyer for disputing his false belief that 2 samples are always as powerful as polynomially-many samples. We thank Noga Alon for useful discussions.

References

[1] N. Alon and J.H. Spencer, *The Probabilistic Method*, John Wiley & Sons, Inc., 1992.

[2] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica* 1 (1981), pages 233–241.

[3] M.J. Fischer and S.A. Paleologou. On the Indistinguishability of Probabilistic Ensembles. Unpublished manuscript, 1994. See [9]

[4] O. Goldreich. *Foundation of Cryptography – Fragments of a Book*. February 1995. Available from `http://theory.lcs.mit.edu/~oded/frag.html`

[5] O. Goldreich, S. Goldwasser, and S. Micali. How to Construct Random Functions. *JACM*, Vol. 33, No. 4, pages 792–807, 1986.

[6] O. Goldreich and B. Meyer. Computational Indistinguishability – Algorithms vs. Circuits. *Theoretical Computer Science*, Vol. 191, pages 215–218, 1998.

[7] S. Goldwasser and S. Micali. Probabilistic Encryption. *JCSS*, Vol. 28, No. 2, pages 270–299, 1984. Preliminary version in *14th STOC*, 1982.

[8] B. Meyer. Constructive Separation of Classes of Indistinguishable Ensembles. *Structure in Complexity Theory*, 1994. pages 198–204.

[9] S.A. Paleologou. Probabilistic Decision Making in Games and Cryptographic Protocols. Ph.D. Thesis, Yale University, May 1995.

[10] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM*, 27:701–717, 1980.

[11] A.C. Yao. Theory and Application of Trapdoor Functions. In *23rd FOCS*, pages 80–91, 1982.

[12] D. Zuckerman. Randomness-Optimal Oblivious Sampling. *Journal of Random structures and Algorithms*, Vol. 11, No. 4, December 1997, pages 345–367.