THEORY OF COMPUTING, Volume 16 (17), 2020, pp. 1–29
www.theoryofcomputing.org

Concentration for Limited Independence via Inequalities for the Elementary Symmetric Polynomials

Parikshit Gopalan Amir Yehudayoff*

Received October 8, 2015; Revised August 30, 2020; Published December 24, 2020

Abstract. We study the extent of independence needed to approximate the product of bounded random variables in expectation. This natural question has applications in pseudo-randomness and min-wise independent hashing.

For random variables with absolute value bounded by 1, we give an error bound of the form $\sigma^{\Omega(k)}$ when the input is k-wise independent and σ^2 is the variance of their sum. Previously, known bounds only applied in more restricted settings, and were quantitatively weaker.

Our proof relies on a new analytic inequality for the elementary symmetric polynomials $S_k(x)$ for $x \in \mathbb{R}^n$. We show that if $|S_k(x)|, |S_{k+1}(x)|$ are small relative to $|S_{k-1}(x)|$ for some $k > 0$ then $|S_\ell(x)|$ is also small for all $\ell > k$.

We use this to give a simpler and more modular analysis of a construction of min-wise independent hash functions and pseudorandom generators for combinatorial rectangles due to Gopalan et al., which also improves the overall seed-length.

ACM Classification: G.3

AMS Classification: 68Q87

Key words and phrases: pseudorandomness, k-wise independence, hashing, concentration, symmetric polynomials

A preliminary version of this paper appeared as ECCC TR14-019 in 2014.

*Horev fellow—supported by the Taub foundation. Research supported by ISF and BSF.

 DOI: 10.4086/toc.2020.v016a017

1 Introduction

The power of independence in probability and randomized algorithms stems from the fact that it lets us control expectations of products of random variables. If $X_1, \ldots, X_n$ are independent random variables, then $\mathbb{E}\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n \mu_i$ where the μ_i are their respective expectations. (To avoid measurability issues, we assume all random variables have finite support.) However, there are numerous settings in computer science, where true independence either does not hold, or is too expensive (in terms of memory or randomness).

Motivated by this, we explore settings when approximate versions of the product rule for expectations hold even under limited independence. Concretely, let $X_1, \ldots, X_n$ be random variables lying in the range $[-1, 1]$, where X_i has mean μ_i and variance σ_i^2. We are interested in the smallest $k = k(\delta)$ such that whenever the X_is are drawn from a k-wise independent distribution $\mathcal{D}$, it holds that

$$\left| \mathbb{E}_{\mathcal{D}}\left[\prod_{i=1}^n X_i\right] - \prod_{i=1}^n \mu_i \right| \leq \delta. \tag{1.1}$$

As stated, we cannot hope to make do even with $k = n-1$. Consider the case where each X_i is a random $\{\pm 1\}$ bit. If $X_n = \prod_{i=1}^{n-1} X_i$, then the resulting distribution is $(n-1)$-wise independent, but $\mathbb{E}[\prod_i X_i] = 1$, whereas $\prod_i \mu_i = 0$. So, we need some additional assumptions about the random variables.

The main message of this paper is that small total variance is sufficient to ensure that the product rule holds approximately even under k-wise independence.

Theorem 1.1. *Let $X_1, \ldots, X_n$ be random variables each distributed in the range $[-1, 1]$, where X_i has mean μ_i and variance σ_i^2. Let $\sigma^2 = \sum_i \sigma_i^2$. There exist constants $c_1 > 1$ and $1 > c_2 > 0$ such that under any k-wise independent distribution $\mathcal{D}$,*

$$\left| \mathbb{E}_{\mathcal{D}}\left[\prod_{i=1}^n X_i\right] - \prod_{i=1}^n \mu_i \right| \leq (c_1 \sigma)^{c_2 k}. \tag{1.2}$$

Specifically, if $\sigma < 1/(2c_1)$ then $k = O(\log(1/\delta))$-wise independence suffices for Equation (1.1).

An important restriction that naturally arises is positivity, where each X_i lies in the interval $[0, 1]$. This setting of parameters (positive variables, small total variance) is important for the applications considered in this paper: pseudorandom generators (PRGs) for combinatorial rectangles [7, 12] and min-wise independent permutations [4]. The former is an important problem in the theory of unconditional pseudorandomness which has been studied intensively [7, 12, 19, 3, 13, 9]. Min-wise independent hashing was introduced by Broder et al. [4] motivated by similarity estimation, and further studied by [11, 5, 19]. The authors of [19] showed that PRGs for rectangles give min-wise independent hash functions.

The results of [7, 11] tell us that under k-wise independence, positivity and boundedness, the LHS of Equation (1.1) is bounded by $\exp(-\Omega(k))$, hence $k = O(\log(1/\delta))$ suffices for error δ. In contrast, we have seen that such a bound cannot hold in the $[-1, 1]$ case. However, once the variance is smaller than some constant, our bound beats this bound even in the $[0, 1]$ setting. Concretely, when $\sigma^2 < n^{-\varepsilon}$ for some $\varepsilon > 0$, our result says that $O(1)$-wise independence suffices for inverse polynomial error in Equation (1.1), as opposed to $O(\log(n))$-wise independence. This improvement is crucial in analyzing

PRGs and hash functions in the polynomially small error regime. A recent result of [9] achieves near-logarithmic seed-length for both these problems, even in the regime of inverse polynomial error. Their construction is simple, but its analysis is not. Using our results, we give a modular analysis of the pseudorandom generator construction for rectangles of [9], using the viewpoint of hash functions. Our analysis also improves the seed-length of the construction, getting the dependence on the dimension n down to $O(\log\log(n))$ as opposed to $O(\log(n))$, which (nearly) matches a lower bound due to [12].

The main technical ingredient in our work is a new analytic inequality about symmetric polynomials in real variables which we believe is of independent interest. The k-th symmetric polynomial in $a = (a_1, a_2, \ldots, a_n)$ is defined as

$$S_k(a) = \sum_{T \subseteq [n] : |T| = k} \prod_{i \in T} a_i \tag{1.3}$$

(we let $S_0(a) = 1$). We show that for any real vector a, if $|S_k(a)|, |S_{k+1}(a)|$ are small relative to $|S_{k-1}(a)|$ for some $k > 0$, then $|S_\ell(a)|$ is also small for all $\ell > k$. This strengthens and generalizes a result of [9] for the case $k = 1$.

We give an overview of the new inequality, its use in the derivation of bounds under limited independence, and finally the application of these bounds to the construction of pseudorandom generators and hash functions.

1.1 The elementary symmetric polynomials

The elementary symmetric polynomials appear as coefficients of a univariate polynomial with real roots, since $\prod_{i \in [n]}(\xi + a_i) = \sum_{k=0}^{n} \xi^k S_{n-k}(a)$. They have been well studied in mathematics, dating back to classical results of Newton and Maclaurin (see [20] for a survey). This work focuses on their growth rates. Specifically, we study how local information on $S_k(a)$ for two consecutive values of k implies global information for all larger values of k.

It is easy to see that symmetric polynomials over the real numbers have the following property:

Fact 1.2. *Over the real numbers, if $S_1(b) = S_2(b) = 0$ then $b = 0$.*

This is equivalent to saying that if $p(\xi)$ is a real univariate polynomial of degree n with n nonzero roots and $p'(0) = p''(0) = 0$ then $p \equiv 0$. This does not hold over all fields, for example, the polynomial $p(\xi) = \xi^3 + 1$ has three nonzero complex roots and $p'(0) = p''(0) = 0$.

A robust version of Fact 1.2 was recently proved in [9]: For every $a \in \mathbb{R}^n$ and $k \in [n]$,

$$|S_k(a)| \leq \left(S_1^2(a) + 2|S_2(a)|\right)^{k/2}. \tag{1.4}$$

That is, if $S_1(a), S_2(a)$ are small in absolute value, then so is everything that follows. We provide an essentially optimal bound.

Theorem 1.3. *For every $a \in \mathbb{R}^n$ and $k \in [n]$,*

$$|S_k(a)| \leq \left(\frac{6\mathrm{e}(S_1^2(a) + |S_2(a)|)^{1/2}}{k^{1/2}}\right)^k.$$

The parameters promised by Theorem 1.3 are tight up to an exponential in k which is often too small to matter (we do not attempt to optimise the constants). For example, if $a_i = (-1)^i$ for all $i \in [n]$ then $|S_1(a)| \leq 1$ and $|S_2(a)| \leq n+1$ but $S_k(a)$ is roughly $(n/k)^{k/2}$.

A more general statement than Fact 1.2 actually holds (see Section 2.1 for a proof).

Fact 1.4. *Over the reals, if $S_k(a) = S_{k+1}(a) = 0$ for $k > 0$ then $S_\ell(a) = 0$ for all $\ell \geq k$.*

We prove a robust version of this fact as well: A twice-in-a-row bound on the increase of the symmetric functions implies a bound on what follows.

Theorem 1.5. *For every $a \in \mathbb{R}^n$, if $S_k(a) \neq 0$ and*

$$\left| \binom{k+1}{k} \frac{S_{k+1}(a)}{S_k(a)} \right| \leq C \quad \textit{and} \quad \left| \binom{k+2}{k} \frac{S_{k+2}(a)}{S_k(a)} \right| \leq C^2$$

then for every $1 \leq h \leq n-k$,

$$\left| \binom{k+h}{k} \frac{S_{k+h}(a)}{S_k(a)} \right| \leq \left(\frac{8\mathrm{e}C}{h^{1/2}} \right)^h .$$

Theorem 1.5 is proved by reduction to Theorem 1.3. The proof of Theorem 1.3 is analytic and uses the method of Lagrange multipliers, and is different from that of [9] which relied on the Newton–Girard identities. The argument is quite general, and similar bounds may be obtained for functions that are recursively defined.

Stronger bounds are known when the inputs are nonnegative. When $a_i \geq 0$ for all $i \in [n]$, the classical Maclaurin inequalities [20] imply that $S_k(a) \leq (\mathrm{e}/k)^k (S_1(a))^k$. In contrast, when we do not assume non-negativity, one cannot hope for such bounds to hold under the assumption that $|S_1(a)|$ or any single $|S_k(a)|$ is small (see the alternating signs example above).

1.2 Tail bounds

We return to the question alluded to earlier about how much independence is required for the approximate product rule of expectation. This question arises in the context of min-wise hashing [11], PRGs for combinatorial rectangles [7, 9], read-once DNFs [9] and more.

One could derive bounds of similar shape to ours using the work of [9], but with much stronger assumptions on the variables. More precisely, one would require $\mathbb{E}[X_i^{2k}] \leq (2k)^{2k} \sigma_i^{2k}$ for all $i \in [n]$, and get an error bound of roughly $k^{O(k)} \sigma^{\Omega(k)}$. These stronger assumptions limit the settings where their bound can be applied (biased variables typically do not have good moment bounds), and ensuring these conditions hold led to tedious case analysis in analyzing their PRG construction.

We briefly outline our approach. We start from the results of [7, 11] who give an error bound of $\delta \leq \exp(-k)$ in (1.1). To prove this, they consider random variables $Y_i = 1 - X_i$, so that

$$\prod_{i=1}^n X_i = \prod_{i=1}^n (1 - Y_i) = \sum_{j=0}^n (-1)^j S_j(Y_1, \ldots, Y_n). \tag{1.5}$$

By inclusion-exclusion/Bonferroni inequalities, the sum on the right gives alternating upper and lower bounds, and the error incurred by truncating to k terms is bounded by $S_k(Y)$. So we can bound the expected error by $\mathbb{E}[S_k(Y)]$ for which k-wise independence suffices.

Our approach replaces inclusion-exclusion by a Taylor-series style expansion about the mean, as in [9]. Let us assume $\mu_i \neq 0$ and let $X_i = \mu_i(1+Z_i)$. Thus,

$$\prod_{i=1}^{n} X_i = \prod_{i=1}^{n} \mu_i(1+Z_i) = \prod_{i=1}^{n} \mu_i \left(\sum_{j=0}^{n} S_j(Z) \right). \tag{1.6}$$

In this approach, it is usually not sufficient to bound $\mathbb{E}[|S_k(Z)|]$, since Z may have negative entries (even if we start with X_is all positive). So, to argue that the first k terms are a good approximation, we need to bound the tail $\left|\sum_{\ell \geq k} S_\ell(Z)\right|$. At first, this seems problematic, since this involves high degree polynomials, and it seems hard to get their expectations right assuming just k-wise independence.[1] Even though we cannot bound $\mathbb{E}[S_\ell(Z)]$ under k-wise independence once $\ell \gg k$, we use our new inequalities for symmetric polynomials to get strong tail bounds on them. This lets us show that truncating Equation (1.6) after k terms gives error roughly σ^k, and thus $k = O(\log(1/\delta)/\log(1/\sigma))$ suffices for error δ. We next describe these tail bounds in detail.

We assume the following setup: $Z = (Z_1, \ldots, Z_n)$ is a vector of real valued random variables where Z_i has mean 0 and variance σ_i^2, and $\sigma^2 = \sum_i \sigma_i^2 < 1$. Let $\mathcal{U}$ denote the distribution where the coordinates of Z are independent. One can show that $\mathbb{E}_{Z \in \mathcal{U}}[|S_\ell(Z)|] \leq \sigma^\ell/\sqrt{\ell!}$ and hence by Markov's inequality (see Corollary 3.2) when $t > 1$ and $t\sigma \leq 1/2$,

$$\Pr_{Z \in \mathcal{U}}\left[\sum_{\ell=k}^{n} |S_\ell(Z)| \geq 2(t\sigma)^k \right] \leq 2t^{-2k}. \tag{1.7}$$

Although k-wise independence does not suffice to bound $\mathbb{E}[S_\ell(Z)]$ for $\ell \gg k$, we use Theorem 1.5 to show that a similar tail bound holds under limited independence.

Theorem 1.6. *Let $\mathcal{D}$ denote a distribution over $Z = (Z_1, \ldots, Z_n)$ as above where the Z_is are $(2k+2)$-wise independent. For $t > 0$ and*[2] $16et\sigma < 1$,

$$\Pr_{X \in \mathcal{D}}\left[\sum_{\ell=k}^{n} |S_\ell(Z)| \geq 2(8et\sigma)^k \right] \leq 2t^{-2k}. \tag{1.8}$$

Typically proofs of tail bounds under limited independence proceed by bounding the expectation of some suitable low-degree polynomial. The proof of Theorem 1.6 does not follow this route. In Section 3.2, we give an example of Z_is and a $(2k+2)$-wise independent distribution on where $\mathbb{E}[|S_\ell(Z)|]$ for $\ell \in \{2k+3, \ldots, n-2k-3\}$ is much larger than under the uniform distribution. The same example also shows that our tail bounds are close to tight.

[1] We formally show this in Section 3.2.

[2] A weaker but more technical assumption on t, σ, k suffices, see Equation (3.11).

1.3 Applications

The notion of min-wise independent hashing was introduced by Broder et al. [4] motivated by similarity estimation, and independently by Mulmuley [15] motivated by computational geometry. A hash function is a map $h : [n] \to [m]$. Let $\mathcal{U}$ denote the family of all hash functions $h : [n] \to [m]$. Let $\mathcal{H} \subseteq \mathcal{U}$ be a family of hash functions. For $S \subseteq [n]$, let $\min h(S) = \min_{x \in S} h(x)$. The following generalization was introduced by Broder et al. [5]:

Definition 1.7. We say that $\mathcal{H} : [n] \to [m]$ is approximately ℓ-minima-wise independent with error ε if for every $S \subseteq [n]$ of size $|S| \geq \ell$ and for every sequence $T = (t_1, \ldots, t_\ell)$ of ℓ distinct elements of S,

$$\left| \Pr_{h \in \mathcal{H}}[h(t_1) < \cdots < h(t_\ell) < \min h(S \setminus T)] - \Pr_{h \in \mathcal{U}}[h(t_1) < \cdots < h(t_\ell) < \min h(S \setminus T)] \right| \leq \varepsilon.$$

Combinatorial rectangles are a well-studied class of tests in pseudorandomness [7, 12, 19, 3, 13, 9]. In addition to being a natural class of statistical tests, constructing generators for them with optimal seeds (up to constant factors) will improve on Nisan's generator for logspace [3], a long-standing open problem in derandomization.

Definition 1.8. A combinatorial rectangle is a function $f : [m]^n \to \{0,1\}$ which is specified by n coordinate functions $f_i : [m] \to \{0,1\}$ as $f(x_1, \ldots, x_n) = \prod_{i \in m} f_i(x_i)$. A map $\mathcal{G} : \{0,1\}^r \to [m]^n$ is a PRG for combinatorial rectangles with error ε if for every combinatorial rectangle $f : [m]^n \to \{0,1\}$,

$$\left| \mathbb{E}_{x \in \{0,1\}^r}[f(\mathcal{G}(x))] - \mathbb{E}_{x \in [m]^n}[f(x)] \right| \leq \varepsilon.$$

A generator $\mathcal{G} : \{0,1\}^r \to [m]^n$ can naturally be thought of as a collection of 2^r hash functions, one for each seed. For $y \in \{0,1\}^r$, let $\mathcal{G}(y) = (x_1, \ldots, x_n)$. The corresponding hash function is given by $g_y(i) = x_i$. The corresponding hash functions have the property that the probability that they fool all test functions given by combinatorial rectangles. Saks et al. [19] showed that this suffices for ℓ-minima-wise independence. They state their result for $\ell = 1$, but their proof extends to all ℓ.

Constructions of PRGs for rectangles and min-wise hash functions that achieve seed-length $O(\log(mn)\log(1/\varepsilon))$ were given by [7, 11] using limited independence. The first construction $\mathcal{G}_{\mathcal{MR}}$ to achieve seed-length $\tilde{O}(\log(mn/\varepsilon))$ was given recently by [9]. We use our results to give an analysis of their generator which we believe is simpler and more intuitive, and also improves the seed-length, to (nearly) match the lower bound from [12].

We take the view of $\mathcal{G}_{\mathcal{MR}}$ as a collection of hash functions $g : [n] \to [m]$, based on iterative applications of an *alphabet squaring* step. We describe the generator formally in Section 5. We start by observing that fooling rectangles is easy when m is small; $O(\log(1/\delta))$-wise Independence suffices, and this requires $O(\log(1/\delta)\log(m)) = O(\log(1/\delta))$ random bits for $m = O(1)$.

The key insight in [9] is that gradually increasing the alphabet is also easy (in that it requires only logarithmic randomness). Assume that we have a hash function $g_0 : [n] \to [m]$ and from it, we define $g_1 : [n] \to [m^2]$. To do this, we pick a function $g_1' : [n] \times [m] \to [m^2]$ and set $g_1(i) = g_1'(i, g_0(i))$. The key observation is that it suffices to pick g_1' using only $O(\log(1/\delta)/\log(m))$-wise independence, rather than the $O(\log(1/\delta))$-wise independence needed for one shot (the larger m is, the less independence is required).

To see why this is so, fix subsets $S_i \subseteq [m^2]$ for each co-ordinate and pretend that g_0 is truly random. One can show that the random variable $\Pr_{g_0}[g_1(i) \in S_i]$ over the choice of g_1' has variance $1/\mathrm{poly}(m)$. Since we are interested in $\prod_i \Pr_{g_0}[g_1(i) \in S_i]$, which is the product of n small variance random variables, Theorem 1.1 says it suffices to use limited independence.[3]

Theorem 1.9. *Let $\mathcal{G}_{\mathcal{MR}}$ be the family of hash functions from $[n]$ to $[m]$ defined in Section 5.2 with error parameter $\delta > 0$. The seed length is at most $O((\log\log(n) + \log(m/\delta))\log\log(m/\delta))$. Then, for every $S_1, \ldots, S_n \subseteq [m]$,*

$$\left| \Pr_{g \in \mathcal{G}_{\mathcal{MR}}}[\forall\, i \in [n]\ g(i) \in S_i] - \Pr_{h \in \mathcal{U}}[\forall\, i \in [n]\ h(i) \in S_i] \right| \leq \delta.$$

This improves the bound from [9] in the dependence on n and δ; their bound was

$$O(\log(mn/\delta)\log\log(m) + \log(1/\delta)\log\log(1/\delta)\log\log\log(1/\delta)).$$

In particular, the dependence on n reduces from $\log(n)$ to[4] $\log\log(n)$. The authors of [12] showed a lower bound of $\Omega(\log(m) + \log(1/\varepsilon) + \log\log(n))$ even for hitting sets, so our bound is tight upto the $\log\log(m/\delta)$ factor. While [12] constructed hitting-set generators for rectangles with near-optimal seedlength, we are unaware of previous constructions of pseudorandom generators for rectangles where the dependence of the seedlength on n is $o(\log(n))$.

Saks et al. [19] showed how to translate a PRG for combinatorial rectangles to an approximately minima-wise independent family (for completeness, see Section 5.5 for a proof).

Theorem 1.10 ([19]). *Let $\mathcal{G} : \{0,1\}^r \to [m]^n$ be a* PRG *for combinatorial rectangles with error ε. The resulting family $\{g_y : y \in \{0,1\}^r\}$ of hash functions is approximately ℓ-minima-wise independent with error at most $\varepsilon\binom{m}{\ell}$.*

We thus get the following corollary.

Corollary 1.11. *For every ℓ, there is a family of approximately ℓ-minima-wise independent hash functions with error ε and seed length at most $O((\log\log(n) + \log(m^\ell/\varepsilon))(\log\log(m^\ell/\varepsilon)))$.*

1.4 Follow-up work

The basic nature of the questions we consider has led to follow-up work which we now briefly describe. (A preliminary version of this article appeared as [10]).

Gopalan, Kane and Meka [8] constructed the first PRG with seed-length $O((\log(n/\delta)\log\log(n/\delta)^2)$ for several classes of functions, including halfspaces, modular tests and combinatorial shapes. The key technical ingredient of their work is a generalization of Theorem 1.1 to the complex numbers. Their proof, however, is different from ours, and in particular it does not imply the inequalities and tail bounds for symmetric polynomials that are proved here.

[3] To optimize the seed-length, we actually use almost k-wise independence rather than exact k-wise independence. So the analysis does not use Theorem 1.1 as a black-box, but rather it directly uses Theorem 1.6.

[4] The reason $\log\log(n)$ seedlength is possible is because every rectangle can be ε-approximated by one that depends only on $O(m\log(1/\varepsilon))$ co-ordinates. Hence the number of functions to fool grows polynomially in n, rather than exponentially.

Understanding the tradeoff between space and randomness as computational resources is an important problem in computational complexity theory. A central technique for understanding this tradeoff is via PRGs for branching programs [17]. Meka, Reingold and Tal [14] constructed the first PRG for width-3 branching programs with nearly optimal seed length. Their proof relies on the proof strategy describe here.

The *iterated restrictions* approach is one of the few general mechanisms for constructing PRGs. It was suggested by Ajtai and Wigderson [1] and in most applications does not yield truly optimal seed length. Doron, Hatami and Hoza [6] showed that that the iterated restrictions approach can achieve optimal seed length (in a specific scenario). A key step in their proof is an extension of our results on the elementary symmetric polynomials to *subset-wise* symmetric polynomials.

1.5 Organization

We present the proofs of our inequalities for symmetric polynomials (Theorems 1.3 and 1.5) in Section 2 and tail bounds for symmetric polynomials (Theorem 1.6) in Section 3. We use these bounds to prove the bound on products of low-variance variables (Theorem 1.1) in Section 4, and to analyze the generator from [9] in Section 5.

2 Inequalities for symmetric polynomials

The proof of our inequality for the elementary symmetric polynomials is by induction on k, and uses the method of Lagrange multipliers together with the Maclaurin identities.

Proof of Theorem 1.3. It will be convenient to use

$$E_2(a) = \sum_{i \in [n]} a_i^2.$$

By Newton's identity, $E_2 = S_1^2 - 2S_2$ so for all $a \in \mathbb{R}^n$,

$$S_1^2(a) + E_2(a) \leq 2(S_1^2(a) + |S_2(a)|).$$

It therefore suffices to prove that for all $a \in \mathbb{R}^n$ and $k \in [n]$,

$$S_k^2(a) \leq \frac{(16\mathrm{e}^2(S_1^2(a) + E_2(a)))^k}{k^k}.$$

We prove this by induction. For $k \in \{1,2\}$, it indeed holds. Let $k > 2$. Our goal will be upper bounding the maximum of the projectively defined[5] function

$$\phi_k(a) = \frac{S_k^2(a)}{(S_1^2(a) + E_2(a))^k}$$

[5]That is, for every $a \neq 0$ in $\mathbb{R}^n$ and $c \neq 0$ in $\mathbb{R}$, we have $\phi_k(ca) = \phi_k(a)$.

under the constraint that $S_1(a)$ is fixed. Since ϕ_k is projectively defined, its supremum is attained in the (compact) unit sphere, and is therefore a maximum. Choose $a \neq 0$ to be a point that achieves the maximum of ϕ_k. We assume, without loss of generality, that $S_1(a)$ is non-negative (if $S_1(a) < 0$, consider $-a$ instead of a). There are two cases to consider:

The first case is that for all $i \in [n]$,

$$a_i \leq \frac{2k^{1/2}(S_1^2(a)+E_2(a))^{1/2}}{n}. \tag{2.1}$$

In this case we do not need the induction hypothesis and can in fact replace each a_i by its absolute value. Let $P \subseteq [n]$ be the set of $i \in [n]$ so that $a_i \geq 0$. Then by Equation (2.1),

$$\sum_{i \in P} |a_i| \leq 2k^{1/2}(S_1^2(a)+E_2(a))^{1/2}.$$

Note that

$$S_1(a) = \sum_{i \in P} |a_i| - \sum_{i \notin P} |a_i| \geq 0.$$

Hence

$$\sum_{i \notin P} |a_i| \leq \sum_{i \in P} |a_i| \leq 2k^{1/2}(S_1^2(a)+E_2(a))^{1/2}.$$

Overall we have

$$\sum_{i \in [n]} |a_i| \leq 4k^{1/2}(S_1^2(a)+E_2(a))^{1/2}.$$

We then bound

$$\begin{aligned} |S_k(a_1,\ldots,a_n)| &\leq S_k(|a_1|,\ldots,|a_n|) \\ &\leq \left(\frac{\mathrm{e}}{k}\right)^k \left(\sum_{i \in [n]} |a_i|\right)^k \quad \text{by the Maclaurin identities} \\ &\leq \left(\frac{4\mathrm{e}}{\sqrt{k}}\right)^k (S_1^2(a)+E_2(a))^{k/2}. \end{aligned}$$

The second case is that there exists $i_0 \in [n]$ so that

$$a_{i_0} > \frac{2k^{1/2}(S_1^2(a)+E_2(a))^{1/2}}{n}. \tag{2.2}$$

In this case we use induction and Lagrange multipliers. For simplicity of notation, for a function F on $\mathbb{R}^n$ denote

$$F(-i) = F(a_1,a_2,\ldots,a_{i-1},a_{i+1},\ldots,a_n)$$

for $i \in [n]$. For every $\delta \in \mathbb{R}^n$ so that $\sum_i \delta_i = 0$ we have $\phi_k(a+\delta) \le \phi_k(a)$. Hence,[6] for all δ so that $\sum_i \delta_i = 0$,

$$\begin{aligned}\phi_k(a) &\ge \frac{S_k^2(a+\delta)}{(S_1^2(a+\delta)+E_2(a+\delta))^k} \\ &\ge \frac{(S_k(a)+\sum_i \delta_i S_{k-1}(-i)+O(\delta^2))^2}{(S_1^2(a)+E_2(a)+2\sum_i a_i\delta_i+O(\delta^2))^k} \\ &\ge \frac{S_k^2(a)+2S_k(a)\sum_i \delta_i S_{k-1}(-i)+O(\delta^2)}{(S_1^2(a)+E_2(a))^k+2k(S_1^2(a)+E_2(a))^{k-1}\sum_i a_i\delta_i+O(\delta^2)}.\end{aligned}$$

Hence, for all δ close enough to zero so that $\sum_i \delta_i = 0$,

$$\frac{S_k^2(a)}{(S_1^2(a)+E_2(a))^k} \ge \frac{S_k^2(a)+2S_k(a)\sum_i \delta_i S_{k-1}(-i)+O(\delta^2)}{(S_1^2(a)+E_2(a))^k+2k(S_1^2(a)+E_2(a))^{k-1}\sum_i a_i\delta_i+O(\delta^2)},$$

or

$$\sum_i \delta_i \left(a_i S_k(a)k - (S_1^2(a)+E_2(a))S_{k-1}(-i)\right) \ge 0. \tag{2.3}$$

For the above inequality to hold for all such δ, it must be that there is λ so that for all $i \in [n]$,

$$a_i S_k(a)k - (S_1^2(a)+E_2(a))S_{k-1}(-i) = \lambda.$$

To see why this is true, set $\lambda_i = a_i S_k(a)k - (S_1^2(a)+E_2(a))S_{k-1}(-i)$. We now have $\lambda_1,\ldots,\lambda_n$ so that

$$\sum_i \lambda_i \delta_i \ge 0 \tag{2.4}$$

for every $\delta_1,\ldots,\delta_n$ of sufficiently small norm where $\sum_i \delta_i = 0$. We claim that this implies that in fact $\lambda_i = \lambda$ for every i. To see this, assume for contradiction that $\lambda_1 \ne \lambda_2$ and $|\lambda_1| > |\lambda_2|$. Set

$$\delta_1 = -\mu\lambda_1,\ \delta_2 = \mu\lambda_1,\ \delta_3 = \delta_4 = \ldots = \delta_n = 0$$

for $\mu > 0$ sufficiently small. It follows that $\sum_i \delta_i = 0$ and $\sum_i \lambda_i\delta_i = \mu(\lambda_1\lambda_2 - \lambda_1^2) < 0$ so Equation (2.4) is violated.

Sum over i to get

$$\lambda n = S_1(a)S_k(a)k - (S_1^2(a)+E_2(a))(n-(k-1))S_{k-1}(a).$$

Thus, for all $i \in [n]$,

$$\begin{aligned}&a_i S_k(a)k - (S_1^2(a)+E_2(a))S_{k-1}(-i) \\ &\qquad = \frac{1}{n}\left(S_1(a)S_k(a)k - (S_1^2(a)+E_2(a))(n-(k-1))S_{k-1}(a)\right),\end{aligned}$$

[6] Here and below, $O(\delta^2)$ means of absolute value at most $C \cdot \|\delta\|_\infty$ for $C = C(n,k) \ge 0$.

or

$$S_k(a)k\left(a_i - \frac{S_1(a)}{n}\right) = (S_1^2(a)+E_2(a))(S_{k-1}(-i)-S_{k-1}(a)) + \frac{(k-1)}{n}(S_1^2(a)+E_2(a))S_{k-1}(a)).$$

This specifically holds for i_0, so using Equation (2.2) we have

$$\begin{aligned}
&\left|S_k(a)k\frac{a_{i_0}}{2}\right| \\
&< \left|S_k(a)k\left(a_{i_0} - \frac{S_1(a)}{n}\right)\right| \\
&\leq \left|(S_1^2(a)+E_2(a))a_{i_0}S_{k-2}(-i_0)\right| + \left|\frac{(k-1)(S_1^2(a)+E_2(a))S_{k-1}(a)}{n}\right|,
\end{aligned}$$

or

$$\begin{aligned}
&|S_k(a)| \\
&\leq \left|\frac{2(S_1^2(a)+E_2(a))S_{k-2}(-i_0)}{k}\right| + \left|\frac{2(k-1)(S_1^2(a)+E_2(a))S_{k-1}(a)}{nka_{i_0}}\right| \\
&< \left|\frac{2(S_1^2(a)+E_2(a))S_{k-2}(-i_0)}{k}\right| + \left|\frac{(S_1^2(a)+E_2(a))^{1/2}S_{k-1}(a)}{k^{1/2}}\right|.
\end{aligned} \tag{2.5}$$

To apply induction we need to bound $S_1^2(-i_0)+E_2(-i_0)$ from above. Since

$$\begin{aligned}
S_1^2(a)+E_2(a)-S_1^2(-i_0)-E_2(-i_0) &= a_{i_0}^2 + 2a_{i_0}S_1(-i_0) + a_{i_0}^2 \\
&= 2a_{i_0}S_1(a) \geq 0.
\end{aligned}$$

we have the bound

$$S_1^2(-i_0)+E_2(-i_0) \leq S_1^2(a)+E_2(a).$$

Finally, by induction and Equation (2.5),

$$\begin{aligned}
|S_k(a)| &\leq \frac{2(S_1^2(a)+E_2(a))}{k}\frac{(16\mathrm{e}^2(S_1^2(-i_0)+E_2(-i_0)))^{(k-2)/2}}{(k-2)^{(k-2)/2}} \\
&\quad + \frac{(S_1^2(a)+E_2(a))^{1/2}}{k^{1/2}}\frac{(16\mathrm{e}^2(S_1^2(a)+E_2(a)))^{(k-1)/2}}{(k-1)^{(k-1)/2}} \\
&\leq \frac{(16\mathrm{e}^2(S_1^2(a)+E_2(a)))^{k/2}}{k^{k/2}}\left(\frac{2}{16\mathrm{e}^2\left(1-\frac{2}{k}\right)^{(k-2)/2}} + \frac{1}{4\mathrm{e}\left(1-\frac{1}{k}\right)^{(k-1)/2}}\right) \\
&< \frac{(16\mathrm{e}^2(S_1^2(a)+E_2(a)))^{k/2}}{k^{k/2}}.
\end{aligned}$$

□

The proof of the more general inequality (Theorem 1.5) is by reduction to Theorem 1.3, and uses the connection between real polynomials in one variable and the symmetric polynomials.

Proof of Theorem 1.5. Assume $a_1,\dots,a_m$ are nonzero and $a_{m+1},\dots,a_n$ are zero. Denote $a' = (a_1,\dots,a_m)$ and notice that for all[7] $k \in [n]$,

$$S_k(a) = S_k(a').$$

Write

$$p(\xi) = \prod_{i\in[m]} (\xi a_i + 1) = \sum_{k=0}^{m} \xi^k S_k(a).$$

Derive k times to get

$$p^{(k)}(\xi) = S_k(a)k!\left(\binom{m}{k}\frac{S_m(a)}{S_k(a)}\xi^{m-k} + \binom{m-1}{k}\frac{S_{m-1}(a)}{S_k(a)}\xi^{m-k-1} + \dots \right.$$
$$\left. \dots + \binom{k+1}{k}\frac{S_{k+1}(a)}{S_k(a)}\xi + 1\right).$$

Since p has m real roots, $p^{(k)}$ has $m-k$ real roots. Since $p^{(k)}(0) \neq 0$, there is $b \in \mathbb{R}^{m-k}$ so that

$$p^{(k)}(\xi) = S_k(a)k! \prod_{i\in[m-k]} (\xi b_i + 1).$$

For all $h \in [m-k]$,

$$S_h(b) = \binom{k+h}{k}\frac{S_{k+h}(a)}{S_k(a)}.$$

By assumption,

$$|S_1(b)| \le C \text{ and } |S_2(b)| \le C^2.$$

Theorem 1.3 implies

$$|S_h(b)| = \left|\binom{k+h}{k}\frac{S_{k+h}(a)}{S_k(a)}\right| \le \frac{(8\mathrm{e}C)^h}{h^{h/2}}.$$ □

2.1 Zeros of polynomials

We conclude the section with a proof of Fact 1.4 which states that over the reals, if $S_k(a) = S_{k+1}(a) = 0$ for $k > 0$ then $S_\ell(a) = 0$ for all $\ell \ge k$.

For a univariate polynomial $p(\xi)$ and a root $y \in \mathbb{R}$ of p, denote by $\mathsf{mult}(p,y)$ the multiplicity of the root y in p. We use the following property of polynomials $p(\xi)$ with real roots (see, e. g., [18]), which can be proved using the interlacing of the zeroes of $p(\xi)$ and $p'(\xi)$: If $\mathsf{mult}(p',y) \ge 2$ then $\mathsf{mult}(p,y) \ge \mathsf{mult}(p',y)+1$.

[7] For $k > m$ we have $S_k(a) = 0$ so there is nothing to prove.

Proof of Fact 1.4. Let

$$p(\xi) = \prod_{i \in [n]} (\xi + b_i) = \sum_{k=0}^{n} \xi^k S_{n-k}(b).$$

Consider $p^{(n-k-1)}(\xi)$ which is the $(n-k-1)$-th derivative of $p(\xi)$. Since $S_k(b) = S_{k+1}(b) = 0$ for $k > 0$, it follows that ξ^2 divides $p^{(n-k-1)}(\xi)$ and hence $\mathsf{mult}(p^{(n-k-1)}, 0) \geq 2$. Applying the above fact $n-k-1$ times, we get $\mathsf{mult}(p, 0) \geq n-k+1$ so $S_n(b) = \ldots = S_k(b) = 0$. □

3 Tail bounds under limited independence

In this section we work with the following setup. Let $X = (X_1, \ldots, X_n)$ be a vector of real valued random variables so that $\mathbb{E}[X_i] = 0$ for all $i \in [n]$. Let σ_i^2 denote the variance of X_i, and let $\sigma^2 = \sum_{i=1}^n \sigma_i^2$. The goal is proving a tail bound on the behavior of the symmetric functions under limited independence.

We start by obtaining tail estimates, under full independence. Let $\mathcal{U}$ denote the distribution over $X = (X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ are independent.

Lemma 3.1. $\mathbb{E}_{X \in \mathcal{U}}[S_\ell^2(X)] \leq \frac{\sigma^{2\ell}}{\ell!}$.

Proof. Since the expectation of X_i is zero for all $i \in [n]$,

$$\begin{aligned}
\mathbb{E}[S_\ell^2(X)] &= \sum_{T, T' \subseteq [n] : |T| = |T'| = \ell} \mathbb{E}\left[\prod_{t \in T} X_t \prod_{t' \in T'} X_{t'}\right] \\
&= \sum_{T \subseteq [n] : |T| = \ell} \mathbb{E}\left[\prod_{t \in T} X_t^2\right] = \sum_{T \subseteq [n] : |T| = \ell} \prod_{t \in T} \sigma_t^2 \\
&\leq \frac{1}{\ell!} \left(\sum_{i \in [n]} \sigma_i^2\right)^\ell = \frac{\sigma^{2\ell}}{\ell!}.
\end{aligned}$$
□

Corollary 3.2. *For $t > 0$ and $\ell \in [n]$, by Markov's inequality,*

$$\Pr_{X \in \mathcal{U}}\left[|S_\ell(X)| \geq \left(\frac{e^{1/2} t \sigma}{\ell^{1/2}}\right)^\ell \geq \frac{(t\sigma)^\ell}{\sqrt{\ell!}}\right] \leq \frac{1}{t^{2\ell}}. \tag{3.1}$$

If $2e^{1/2} t \sigma \leq k^{1/2}$ then by the union bound

$$\Pr_{X \in \mathcal{U}}\left[\sum_{\ell=k}^{n} |S_\ell(X)| \geq 2\left(\frac{e^{1/2} t \sigma}{k^{1/2}}\right)^k\right] \leq \frac{1}{t^{2k} - t^{2(k-1)}}. \tag{3.2}$$

We now consider limited independence.

Lemma 3.3. *Let $\mathcal{D}$ denote a distribution over $X = (X_1, \dots, X_n)$ where $X_1, \dots, X_n$ are $(2k+2)$-wise independent. Let $t \geq 1$. Except with $\mathcal{D}$-probability at most $2t^{-2k}$, the following bounds hold for all $\ell \in \{k, \dots, n\}$:*

$$|S_\ell(X)| \leq (8\mathrm{e}t\sigma)^\ell \left(\frac{k}{\ell}\right)^{\ell/2}. \tag{3.3}$$

Proof. In the following the underlying probability distribution over X is $\mathcal{D}$. By Lemma 3.1, for $i \in \{k, k+1\}$,

$$\mathbb{E}[S_i^2(X)] \leq \frac{\sigma^{2i}}{i!}.$$

Hence by Markov's inequality,

$$\Pr\left[|S_i(X)| \geq \frac{(t\sigma)^i}{\sqrt{i!}}\right] \leq t^{-2i}.$$

From now on, condition on the event that

$$|S_k(X)| \leq \frac{(t\sigma)^k}{\sqrt{k!}} \text{ and } |S_{k+1}(X)| \leq \frac{(t\sigma)^{k+1}}{\sqrt{(k+1)!}}, \tag{3.4}$$

which occurs with probability at least $1 - 2t^{-2k}$. Fix $x = (x_1, \dots, x_n)$ such that Equation (3.4) holds.

We claim that there must exist $k_0 \in \{0, \dots, k-1\}$ for which the following bounds hold:

$$|S_{k_0}(x)| \geq \frac{(t\sigma)^{k_0}}{\sqrt{k_0!}}, \tag{3.5}$$

$$|S_{k_0+1}(x)| \leq \frac{(t\sigma)^{k_0+1}}{\sqrt{(k_0+1)!}}, \tag{3.6}$$

$$|S_{k_0+2}(x)| \leq \frac{(t\sigma)^{k_0+2}}{\sqrt{(k_0+2)!}}. \tag{3.7}$$

To see this, mark point $j \in \{0, \dots, k+1\}$ as *high* if

$$|S_j(x)| \geq \frac{(t\sigma)^j}{\sqrt{j!}}$$

and *low* if

$$|S_j(x)| \leq \frac{(t\sigma)^j}{\sqrt{j!}}.$$

A point is marked both high and low if equality holds. Observe that 0 is marked high (and low) since $S_0(x) = 1$ and k and $k+1$ are marked low by Equation (3.4). This implies the existence of a triple k_0, k_0+1, k_0+2 where the first point is high and the next two are low.

Let $\gamma > 0$ be the smallest number so that the following inequalities hold:

$$|S_{k_0+1}(x)| \le |S_{k_0}(x)| \frac{\gamma}{\sqrt{k_0+1}}, \tag{3.8}$$

$$|S_{k_0+2}(x)| \le |S_{k_0}(x)| \frac{\gamma^2}{\sqrt{(k_0+1)(k_0+2)}}. \tag{3.9}$$

By definition, one of Equations (3.8) and (3.9) holds with equality so

$$|S_{k_0}(x)| = \max\left\{ \frac{|S_{k_0+1}(x)|\sqrt{k_0+1}}{\gamma}, \frac{|S_{k_0+2}(x)|\sqrt{(k_0+1)(k_0+2)}}{\gamma^2} \right\}.$$

Observe further that $\gamma \le t\sigma$ by Equations (3.5), (3.6) and (3.7). Combining this with the bounds in Equations (3.6) and (3.7)

$$|S_{k_0}(x)| \le \max\left\{ \frac{(t\sigma)^{k_0+1}}{\gamma\sqrt{k_0!}}, \frac{(t\sigma)^{k_0+2}}{\gamma^2\sqrt{k_0!}} \right\} = \frac{(t\sigma)^{k_0+2}}{\gamma^2\sqrt{k_0!}}. \tag{3.10}$$

Equations (3.8) and (3.9) let us apply Theorem 1.5 with $C = \gamma\sqrt{k_0+1}$ and $h \ge 3$ to get

$$\left| \frac{S_{k_0+h}(x)}{S_{k_0}(x)} \right| \le (8\mathrm{e}\gamma)^h \frac{(k_0+1)^{h/2}}{h^{h/2}\binom{k_0+h}{k_0}}.$$

Bounding $|S_{k_0}|$ by Equation 3.10, we get

$$|S_{k_0+h}(x)| \le (8\mathrm{e}\gamma)^h \frac{(k_0+1)^{h/2}}{h^{h/2}\binom{k_0+h}{k_0}} \frac{(t\sigma)^{k_0+2}}{\gamma^2\sqrt{k_0!}} \le (8\mathrm{e}t\sigma)^{k_0+h} \frac{(k_0+1)^{h/2}}{h^{h/2}\sqrt{k_0!}\binom{k_0+h}{h}}.$$

Since

$$\binom{k_0+h}{h} \ge \max\left\{ \left(\frac{k_0+h}{k_0}\right)^{k_0}, \left(\frac{k_0+h}{h}\right)^{h} \right\} \ge \frac{(k_0+h)^{(k_0+h)/2}}{k_0^{k_0/2} h^{h/2}},$$

we have

$$\frac{(k_0+1)^{h/2}}{h^{h/2}\sqrt{k_0!}\binom{k_0+h}{h}} \le \left(\frac{k_0+1}{h}\right)^{h/2} \frac{k_0^{k_0/2} h^{h/2}}{(k_0+h)^{(k_0+h)/2}} \le \left(\frac{k_0+1}{k_0+h}\right)^{(k_0+h)/2}.$$

Therefore, denoting $\ell = k_0 + h$, since $k_0 + 1 \le k$,

$$|S_\ell(x)| \le (8\mathrm{e}t\sigma)^\ell \left(\frac{k}{\ell}\right)^{\ell/2}. \qquad \square$$

3.1 Proof of tail bounds

Proof of Theorem 1.6. As in Lemma 3.3, fix $x = (x_1, \dots, x_n)$ such that Equation 3.4 holds (the random vector X has this property with $\mathcal{D}$-probability at least $1 - 2t^{-2k}$). By the proof of lemma, since by assumption $8\mathrm{e}t\sigma < 1/2$,

$$\sum_{\ell=k}^{n} |S_\ell(x)| \le \frac{(t\sigma)^k}{k!} + \frac{(t\sigma)^{k+1}}{\sqrt{(k+1)!}} + \sum_{\ell=k+2}^{n} (8\mathrm{e}t\sigma)^\ell \left(\frac{k}{\ell}\right)^{\ell/2} \le 2(8\mathrm{e}t\sigma)^k. \tag{3.11}$$

□

3.2 On the tightness of the tail bounds

We conclude by showing that $(2k+2)$-wise independence is insufficient to fool $|S_\ell|$ for $\ell > 2k+2$ in expectation. We use a modification of a simple proof due to Noga Alon of the $\Omega(n^{k/2})$ lower bound on the support size of a k-wise independent distribution on $\{-1,1\}^n$, which was communicated to us by Raghu Meka.

For this section, let $X_1, \dots, X_n$ be so that each X_i is uniform over $\{-1,1\}$. Thus $\sigma^2 = \sum_i \mathsf{Var}[X_i] = n$. By Lemma 3.1, we have

$$\mathbb{E}_{X\in\mathcal{U}}[|S_\ell(X)|] \le \left(\mathbb{E}_{X\in\mathcal{U}}[S_\ell^2(X)]\right)^{1/2} \le \frac{n^{\ell/2}}{\sqrt{\ell!}}. \tag{3.12}$$

In contrast we have the following:

Lemma 3.4. *There is a $(2k+2)$-wise independent distribution on $X = (X_1, X_2, \dots, X_n)$ in $\{-1,1\}^n$ such that for every $\ell \in [n]$,*

$$\Pr_{X\in\mathcal{D}}\left[|S_\ell(X)| \ge \binom{n}{\ell}\right] \ge \frac{1}{3n^{k+1}}.$$

Specifically,

$$\mathbb{E}_{X\in\mathcal{D}}[|S_\ell(X)|] \ge \frac{\binom{n}{\ell}}{3n^{k+1}}. \tag{3.13}$$

Proof. Let $\mathcal{D}$ be a $(2k+2)$-wise independent distribution on $\{-1,1\}^n$ that is uniform over a set D of size $2(n+1)^{k+1} \le 3n^{k+1}$. Such distributions are known to exist [2]. Further, by translating the support by some fixed vector if needed, we may assume that $(1,1,\dots,1) \in D$. It is easy to see that every such translate also induces a $(2k+2)$-wise independent distribution. The claim holds since $S_\ell(1,\dots,1) = \binom{n}{\ell}$. □

When, for example, $k = O(\log n)$, which is often the case of interest, for $2k+3 \le \ell \le n-(2k+3)$, the RHS of (3.13) is much larger than the bound guaranteed by Equation 3.12. The tail bound provided by Lemma 3.3 can not therefore be extended to a satisfactory bound on the expectation. Furthermore, applying Lemma 3.3 with

$$t = \frac{1}{8\mathrm{e}}\sqrt{\frac{n}{\ell k}}$$

implies that for any $(2k+2)$-wise independent distribution,

$$\Pr\left[|S_\ell(X)| \geq \binom{n}{\ell}\right] \leq \Pr\left[|S_\ell(X)| \geq (8\mathrm{e}t\sqrt{n})^\ell \left(\frac{k}{\ell}\right)^{\ell/2}\right] \leq 2\left(\frac{64\mathrm{e}^2 k\ell}{n}\right)^k.$$

When $k\ell = o(n)$, this is at most $O(n^{-k+o(1)})$. Comparing this to the bound given in Lemma 3.4, we see that the bound provided by Lemma 3.3 is nearly tight.

4 Limited independence fools products of variables

In this section we work with the following setup. We have n random variables $X_1, \ldots, X_n$, each distributed in the interval $[-1,1]$. Let μ_i and σ_i^2 denote the mean and variance of X_i, and let $\sigma^2 = \sum_{i=1}^n \sigma_i^2$. The following theorem shows that limited independence fools products of bounded variables with low total variance.

Theorem 4.1. *There exists $C > 0$ such that under any Ck-wise independent distribution $\mathcal{D}$,*

$$\left|\mathbb{E}_{\mathcal{D}}[\prod_{i=1}^n X_i] - \prod_{i=1}^n \mu_i\right| \leq (C\sigma)^k. \tag{4.1}$$

Proof. Denote by $\mathcal{U}$ the distribution on $(X_1, \ldots, X_n)$ in which the X_is are independent with the same marginal distribution as in $\mathcal{D}$. Define $H \subseteq [n]$ to be the set of indices such that $|\mu_i| \leq \sqrt{\sigma}$. There are two cases to consider.

Case one: The first case is that $|H| \geq 2k$. In this case, let H' be a subset of H of size $2k$. Since the variables are bounded in $[-1,1]$, we have

$$\left|\mathbb{E}_{\mathcal{D}}[\prod_{i\in[n]} X_i]\right| \leq \mathbb{E}_{\mathcal{D}}[\left|\prod_{i\in[n]} X_i\right|] \leq \mathbb{E}_{\mathcal{D}}[\left|\prod_{i\in H'} X_i\right|].$$

The $2k$-wise independence implies

$$\mathbb{E}_{\mathcal{D}}[\left|\prod_{i\in H'} X_i\right|] = \prod_{i\in H'} \mathbb{E}[|X_i|] \leq \prod_{i\in H'} \sqrt{\mathbb{E}[X_i^2]} = \prod_{i\in H'} \sqrt{\sigma_i^2 + \mu_i^2} \leq (2\sigma)^k.$$

The same bound also holds under $\mathcal{U}$. Hence,

$$\left|\mathbb{E}_{\mathcal{D}}[\prod_{i\in[n]} X_i] - \mathbb{E}_{\mathcal{U}}[\prod_{i\in[n]} X_i]\right| \leq 2(2\sigma)^k.$$

Case two: The second case is that $|H| < 2k$. Let $T = H \setminus [n]$. For ease of notation, we shall assume that $T = [m]$ for some $m \leq n$. We may assume that $m > 10k$, since otherwise there is nothing to prove. Even after conditioning on the outcome of variables in H, the resulting distribution on $X_1, \ldots, X_m$ is $10k$-wise independent. Since the variables have absolute value at most 1, it suffices to show that for a $10k$-wise independent distribution $\mathcal{D}$,

$$\left|\mathbb{E}_{\mathcal{D}}[\prod_{i\in[m]} X_i] - \mathbb{E}_{\mathcal{U}}[\prod_{i\in[m]} X_i]\right| \le 2\sigma^k.$$

Write $X_i = \mu_i(1+Z_i)$ so that Z_i has mean 0 and variance σ_i^2/μ_i^2. Define the random variables

$$P = \prod_{i\in[m]} X_i = \prod_{i\in[m]} \mu_i \cdot \sum_{\ell=0}^{m} S_\ell(Z),$$
$$P' = \prod_{i\in[m]} \mu_i \cdot \sum_{\ell=0}^{4k} S_\ell(Z),$$

where $Z = (Z_1,\dots,Z_m)$. We will prove the following claim.

Claim 4.2. *For a $4k$-wise independent distribution $\mathcal{D}$,*

$$\left|\mathbb{E}_{\mathcal{D}}[P-P']\right| \le (c\sigma)^k/2.$$

We first show how to finish the proof of Theorem 4.1 with this claim. We have

$$|\mathbb{E}_{\mathcal{D}}[P] - \mathbb{E}_{\mathcal{U}}[P]| \le \left|\mathbb{E}_{\mathcal{D}}[P-P']\right| + \left|\mathbb{E}_{\mathcal{U}}[P-P']\right| + \left|\mathbb{E}_{\mathcal{D}}[P'] - \mathbb{E}_{\mathcal{U}}[P']\right|.$$

The first two terms are bounded from above by $(c\sigma^k)/2$ by the claim, and the last is 0 since $10k$-wise independence fools degree $4k$ polynomials, such as P'.

Proof of Claim 4.2. Denote by $\bar{\sigma}_i^2$ the variance of Z_i. By definition of T, we have $\bar{\sigma}_i^2 = \sigma_i^2/\mu_i^2 \le \sigma_i^2/\sigma$. The variance of the Z can be bounded by

$$\bar{\sigma}^2 = \sum_{i=1}^{m} \bar{\sigma}_i^2 \le \sum_{i\in T} \frac{\sigma_i^2}{\sigma} \le \sigma.$$

Let G denote the event that $|P-P'| \le 2(8\mathrm{e}\sqrt{\bar{\sigma}})^{4k}$, and denote by $\neg G$ the complement of G. Write

$$\mathbb{E}[P-P'] = \mathbb{E}[(P-P')\mathbb{1}(G)] + \mathbb{E}[(P-P')\mathbb{1}(\neg G)].$$

By the definition of G,

$$|\mathbb{E}[(P-P')\mathbb{1}(G)]| \le 2(8\mathrm{e}\sqrt{\bar{\sigma}})^{4k}.$$

Bound the second term as follows. First, since $-1 \le P \le 1$,

$$\begin{aligned} |\mathbb{E}[(P-P')\mathbb{1}(\neg G)]| &\le |\mathbb{E}[P\mathbb{1}(\neg G)]| + |\mathbb{E}[P'\mathbb{1}(\neg G)]| \\ &\le |\mathbb{E}[\mathbb{1}(\neg G)]| + \sqrt{\mathbb{E}[P'^2]\cdot\mathbb{E}[\mathbb{1}(\neg G)]}. \end{aligned} \tag{4.2}$$

Recall that
$$P - P' = \sum_{\ell=4k+1}^{m} S_\ell.$$
Letting $t = 1/\sqrt{\bar{\sigma}}$ and applying Theorem 1.6,
$$\mathbb{E}[\mathbb{1}(\neg G)] \le 2t^{-8k} = 2\bar{\sigma}^{4k}. \tag{4.3}$$
Since $\mathbb{E}[Z_i] = 0$ for all i, and by Lemma 3.1,
$$\mathbb{E}[P'^2] \le \sum_{i=0}^{4k} \mathbb{E}[S_i^2] \le \sum_{i=0}^{4k} \frac{\bar{\sigma}^{2i}}{i!} \le 2. \tag{4.4}$$
So, the RHS of Equation (4.2) is at most $\bar{\sigma}^{4k} + \sqrt{2}\bar{\sigma}^{2k} \le 3\sigma^k$, as required. □

□

5 Analyzing the PRG for rectangles

Gopalan et al. [9] proposed and analyzed a PRG for combinatorial rectangles, which we denote by $\mathcal{G}_{\mathcal{MR}}$. In this section, we provide a different presentation and analysis of their construction, which is based on our results concerning the symmetric polynomials. Our analysis is simpler and follows the intuition that products of low variance events are easy to fool using limited independence. It also improves on their seedlength in the dependence on n, δ, as discussed above.

5.1 Preliminaries

Let $\mathcal{U}$ denote the uniform distribution on $[m]^n$, and let $\mathcal{D}$ be a distribution on $[m]^n$. We denote by $\Pr_{x\in\mathcal{D}}$ the probability distribution induced by choosing x according to $\mathcal{D}$. For $K \subseteq [n]$, denote by $\mathcal{D}_K$ the marginal distribution of $\mathcal{D}$ in co-ordinates in K.

Definition 5.1. A distribution $\mathcal{D}$ on $[m]^n$ is (k,ε)-wise independent if for every $K \subseteq [n]$ of size at most k, the total variation distance between $\mathcal{D}_K$ and $\mathcal{U}_K$ is at most ε.

Naor and Naor [16] showed that such distributions (for m a power of two) can be generated using seed-length $O(\log\log(n) + k\log(m) + \log(1/\varepsilon))$. Indeed, such distributions can be generated by taking a $(k\log(m))$-wise ε-dependent string of length $n\log(m)$. We can also assume that every co-ordinate is uniformly random in $[m]$, by adding the string $(a,a,\dots,a)$ modulo m, where $a \in [m]$ is uniformly random.

The following property holds. Let P be a real linear combination of combinatorial rectangles,
$$P = \sum_S c_S f_S,$$
where $f_S(x) = \prod_{i\in S} f_{S,i}(x_i)$ where $f_{S,i} : [m] \to \{0,1\}$ for all S, i. Let $\mathrm{L}_1(P) = \sum_S |c_S|$. The degree of P is the maximum size of S for which $c_S \neq 0$. Convexity implies that if $\mathcal{D}$ is (k,ε)-wise independent and P has degree at most k then
$$|\mathbb{E}_{x\in\mathcal{D}}[P(x)] - \mathbb{E}_{x\in\mathcal{U}}[P(x)]| \le \mathrm{L}_1(P)\varepsilon.$$

5.2 The generator

We use an alternate view of $\mathcal{G}_{\mathcal{MR}}$ as a collection of hash functions $g : [n] \to [m]$. The generator $\mathcal{G}_{\mathcal{MR}}$ is based on iterative applications of an *alphabet increasing* step. The first alphabet m_0 is chosen to be large enough, and at each step $t > 1$ the size of the alphabet m_t is squared $m_t = m_{t-1}^2$.

There is a constant $C > 0$ so that the following holds. Denote by δ the error parameter of the generator. Let $T \le C \log\log(m)$ be the first integer so that $m_T \ge m$. Let $\delta' = \delta/T$.

Base Case: Let $m_0 \ge C\log(1/\delta)$ be a power of 2. Sample $g_0 : [n] \to [m_0]$ using a (k_0, ε_0)-wise independent distribution on $[m_0]^n$ with

$$k_0 = C\log(1/\delta'),\ \varepsilon_0 = \delta' \cdot m_0^{-Ck_0}. \tag{5.1}$$

This requires seed length $O(\log\log(n) + \log(\log\log(m)/\delta)\log\log(\log\log(m)/\delta))$.

Squaring the alphabet: Pick $g'_t : [m_{t-1}] \times [n] \to [m_t]$ using a (k_t, ε_t)-wise independent distribution over $[m_t]^{[m_{t-1}]\times[n]}$ with

$$k_t = \max\left\{C\frac{\log(1/\delta')}{\log(m_t)}, 2\right\},\ \varepsilon_t \le m_t^{-Ck_t}.$$

Define a hash function $g_t : [n] \to [m_t]$ as

$$g_t(i) = g'_t(g_{t-1}(i), i).$$

This requires seed length $O(\log\log(n) + \log(m_t) + \log(\log\log(m)/\delta))$.

5.3 Two lemmas

We first analyze the base case using the inclusion-exclusion approach of [7]. We need to extend their analysis to the setting where the co-ordinates are only approximately k-wise independent.

Lemma 5.2. *Let $\mathcal{D}$ be a (k,ε)-wise independent distribution on $[m]^n$ with k odd. Then, for every $S_1, \dots, S_n \subseteq [m]$,*

$$\left| \Pr_{g\in\mathcal{D}}[\forall\, i \in [n]\ g(i) \in S_i] - \Pr_{h\in\mathcal{U}}[\forall\, i \in [n]\ h(i) \in S_i] \right| \le \varepsilon m^k + \exp(-\Omega(k)).$$

Proof. Let $p_i = |S_i|/m$, and $q_i = 1 - p_i$. Observe that

$$\Pr_{h\in\mathcal{U}}[\forall\, i \in [n]\ h(i) \in S_i] = \prod_{i=1}^{n} p_i = \prod_{i=1}^{n}(1 - q_i) \le \exp\left(-\sum_{i=1}^{n} q_i\right). \tag{5.2}$$

We consider two cases based on $\sum_i q_i$.

Case 1: When $\sum_i q_i \le k/(2\mathrm{e})$. Since every non-zero q_i is at least $1/m$, there can be at most $mk/(2\mathrm{e})$ indices i so that $q_i > 0$. For i so that $q_i = 0$, we have $S_i = [m]$, so we can drop such indices and assume $n \le mk/(2\mathrm{e})$. By Bonferroni inequality, since k is odd,

$$\left| \Pr_g[\forall\, i \in [n]\ g(i) \in S_i] - \sum_{j=0}^{k-1} (-1)^j \sum_{J \subseteq [n] : |J| = j} \Pr_g[\forall\, i \in J\ g(i) \notin S_i] \right| \\ \le \sum_{J \subseteq [n] : |J| = k} \Pr_g[\forall\, i \in J\ g(i) \notin S_i].$$

A similar bound holds for h. Using the (k,ε)-wise independence, and since $\binom{n}{k} \le (\mathrm{e}n/k)^k$,

$$\left| \Pr_g[\forall\, i \in [n]\ g(i) \in S_i] - \Pr_h[\forall\, i \in [n]\ h(i) \in S_i] \right| \\ \le \varepsilon(\mathrm{e}n/k)^k + 2 \sum_{J \subseteq [n] : |J| = k} \Pr_h[\forall\, i \in J\ h(i) \notin S_i].$$

The second term is twice $S_k(q_1,\ldots,q_n)$, which we can bound by Maclaurin's identity as

$$S_k(q_1,\ldots,q_n) \le (\mathrm{e}/k)^k \left(\sum_{i=1}^{n} q_i \right)^k \le 2^{-k}.$$

The lemma is proved since $n \le mk/(2\mathrm{e})$.

Case 2: When $\sum_i q_i > k/(2\mathrm{e})$. Once again, we drop indices i so that $q_i = 0$. Consider the largest n' such that

$$k/(2\mathrm{e}) - 1 \le \sum_{i=1}^{n'} q_i \le k/(2\mathrm{e}).$$

Repeating the argument from Case 1 for this n', we get

$$\left| \Pr_g[\forall\, i \in [n']\ g(i) \in S_i] - \Pr_h[\forall\, i \in [n']\ h(i) \in S_i] \right| \le \varepsilon(m/2)^k + 2^{-k+1}.$$

Similarly to Equation (5.2),

$$\Pr_h[\forall\, i \in [n']\ h(i) \in S_i] \le \mathrm{e}^{1-k/(2\mathrm{e})}.$$

Finally, since

$$\Pr_g[\forall\, i \in [n]\ g(i) \in S_i] \le \Pr_g[\forall\, i \in [n']\ g(i) \in S_i],$$

the lemma is proved. □

To analyze the iterative steps, we use the following lemma. To simplify notation, for a finite set X, we denote by $\Pr_{x \in X}$ the probability distribution induced by choosing x uniformly in X.

Lemma 5.3. *There is $C > 0$ so that the following holds. Let $0 < \delta < 1/C$. Assume*

$$k > 1,\ \ell \geq \log(1/\delta),\ \ell \geq k,\ \ell^{-k} \leq \delta^C,\ \varepsilon \leq (m\ell)^{-Ck}.$$

Let $\mathcal{D}$ be a (Ck,ε)-wise independent distribution on $g' : [\ell] \times [n] \to [m]$ so that for every $(a,i) \in [\ell] \times [n]$, the distribution of $g'(a,i)$ is uniform on $[m]$. Let $g : [n] \to [m]$ be defined by $g(i) = g'(x_i, i)$. Then,

$$\left| \Pr_{g' \in \mathcal{D}, x \in [\ell]^n}[\forall\, i \in [n]\ g(i) \in S_i] - \Pr_{h \in [m]^n}[\forall\, i \in [n]\ h(i) \in S_i] \right| \leq \delta.$$

Proof. Let $p_i = |S_i|/m$ and $q_i = 1 - p_i$. Partition $[n]$ into a head $H = \{i : p_i < \ell^{-0.1}\}$ and a tail $T = \{i : p_i \geq \ell^{-0.1}\}$. There are two cases to consider.[8]

The head is large. If $|H| \geq k$, we show that both probabilities are small which means that they are close. Indeed, let H' be the first k indices in H. First, by definition of H,

$$\Pr_h[\forall\, i \in [n]\ h(i) \in S_i] \leq \Pr_h[\forall\, i \in H'\ h(i) \in S_i] = \prod_{i \in H'} \Pr[h(i) \in S_i] \leq \ell^{-0.1k}.$$

Second, (k,ε)-wise independence implies

$$\Pr_g[\forall\, i \in [n]\ g(i) \in S_i] \leq \Pr_g[\forall\, i \in H'\ g(i) \in S_i] \leq \ell^{-0.1k} + \varepsilon,$$

so the proof is complete.

The head is small. From now on, we may assume that $|H| < k$. We may also assume that $q_i \geq 1/m$ and $p_i > 0$ for all $i \in T$, since otherwise S_i is trivial and we can drop such an index. As in the proof of Lemma 5.2, by restricting to a subset if necessary, we can also assume that

$$\sum_{i \in T} q_i \leq C \log(1/\delta). \tag{5.3}$$

Therefore, $|T| \leq Cm \log(1/\delta)$.

For $i \in T$, define the random variable

$$Y_{i,a} = \mathbb{1}(g'(a,i) \in S_i) - p_i.$$

Since $g'(a,i)$ is uniform over $[m]$ for all a,i, we have $\mathbb{E}[Y_{i,a}] = 0$ and $\mathsf{Var}[Y_{i,a}] = q_i p_i$. Define the random vector $A = (A_i : i \in T)$ by

$$A_i = \frac{1}{\ell p_i} \sum_{a=1}^{\ell} Y_{i,a}.$$

Define the random variable

$$Q = \Pr_x[\forall\, i \in H\ g(i) \in S_i].$$

[8] Standard arguments imply that if (k,ε)-wise independence fools both $\forall\, i \in H\ g(i) \in S_i$ and $\forall\, i \in T\ g(i) \in S_i$ with error δ then $(O(k), \varepsilon^{O(1)})$-wise independence fools their intersection with error $O(\delta)$. So it suffices to consider each of them separately. However, since we could not find an explicit reference for this statement, we provide a self-contained argument.

So, for every fixed g',

$$\Pr_x[\forall\, i \in [n]\ g(i) \in S_i] = Q \cdot \prod_{i \in T} p_i(1+A_i) = Q \cdot \prod_{i \in T} p_i \cdot \sum_{i=0}^{|T|} S_i(A). \tag{5.4}$$

Define

$$P = Q \cdot \prod_{i \in T} p_i \cdot \sum_{i=0}^{|T|} S_i(A),$$
$$P' = Q \cdot \prod_{i \in T} p_i \cdot \sum_{i=0}^{k} S_i(A).$$

The degree of P' is at most $2k$. We will show that P' is a good approximation to P.

Claim 5.4. $|\mathbb{E}_{\mathcal{D}}[P - P']| \le O(\ell^{-0.2k})$.

The claim completes the proof:

$$|\mathbb{E}_{\mathcal{D}}[P] - \mathbb{E}_{\mathcal{U}}[P]| \le |\mathbb{E}_{\mathcal{D}}[P - P']| + |\mathbb{E}_{\mathcal{D}}[P'] - \mathbb{E}_{\mathcal{U}}[P']| + |\mathbb{E}_{\mathcal{U}}[P - P']|.$$

Bound the first and third terms by $O(\ell^{-0.2k})$ using the claim ($\mathcal{U}$ is also (Ck, ε)-wise independent). Bound the second term as follows. Since $k \ge 2$, for all i,

$$\mathsf{Var}[A_i] = \frac{1}{\ell^2 p_i^2} \sum_{a=1}^{\ell} \mathsf{Var}[Y_{i,a}] = \frac{q_i^2}{\ell p_i} \le \frac{q_i}{\ell^{0.9}},$$
$$\mathrm{L}_1(A_i) \le \frac{1}{\ell p_i} \sum_{a=1}^{\ell} \mathrm{L}_1(Y_{i,a}) \le \frac{2}{p_i} \le \ell.$$

Plugging in the bounds from Equations (5.3):

$$\sum_{i=1}^{|T|} \mathsf{Var}[A_i] \le \frac{C \log(1/\delta)}{\ell^{0.9}} \le \frac{1}{\ell^{0.6}},$$
$$\sum_{i=1}^{|T|} \mathrm{L}_1(A_i) \le C m \log(1/\delta) \ell \le m \ell^{O(1)},$$
$$\mathrm{L}_1(S_k(A)) \le \left(\sum_{i=1}^{n} \mathrm{L}_1(A_i) \right)^k \le m^k \ell^{O(k)}.$$

Thus, since the degree of P' is $2k$,

$$|\mathbb{E}_{\mathcal{D}}[P'] - \mathbb{E}_{\mathcal{U}}[P']| \le \varepsilon \mathrm{L}_1(P') \le \ell^{-k}.$$

Overall,

$$\left| \Pr_g[\forall\, i \in [n]\ g(i) \in S_i] - \Pr_h[\forall\, i \in [n]\ h(i) \in S_i] \right| = |\mathbb{E}_{\mathcal{D}}[P] - \mathbb{E}_{\mathcal{U}}[P]| \le O(\ell^{-0.2k}) \le \delta.$$

Proof of Claim 5.4. We repeat the proof of Lemma 3.3 with $\sigma^2 = \ell^{-0.5}$ and $t = \ell^{0.2}$. The event G defined as

$$G = \left\{ |S_k(A)| \leq \frac{\ell^{-0.05k}}{\sqrt{k!}} \text{ and } |S_{k+1}(A)| \leq \frac{\ell^{-0.05(k+1)}}{\sqrt{(k+1)!}} \right\}$$

occurs with probability at least $1 - 2\ell^{-0.4k}$. Denote by $\neg G$ the complement of G. Write

$$\mathbb{E}[P - P'] = \mathbb{E}[(P - P')\mathbb{1}(G)] + \mathbb{E}[(P - P')\mathbb{1}(\neg G)].$$

Bound the first term as follows. If the co-ordinates of A are $(Ck, 0)$-wise independent, then, by Lemma 3.1,

$$\mathbb{E}[S_k(A)^2] = \frac{(\sum_i \mathsf{Var}[A_i])^k}{k!} \leq \frac{\ell^{-0.6k}}{k!}.$$

Hence, under (Ck, ε)-wise independence,

$$\mathbb{E}[S_k(A)^2] \leq \frac{\ell^{-0.6k}}{k!} + \varepsilon \mathrm{L}_1(S_k) \leq \frac{\ell^{-0.5k}}{k!}. \tag{5.5}$$

As in the proof of Theorem 1.6, conditioned on G,

$$|P - P'| \leq 2(8\mathrm{e}\ell^{-0.05})^k.$$

Thus,

$$|\mathbb{E}[(P - P')\mathbb{1}(G)]| \leq 2(20\ell^{-0.25})^k.$$

It remains to bound the second term from above. Note that $0 \leq P \leq 1$ since it is the probability of an event. Bound

$$|\mathbb{E}[(P - P')\mathbb{1}(\neg G)]| \leq |\mathbb{E}[P\mathbb{1}(\neg G)]| + |\mathbb{E}[P'\mathbb{1}(\neg G)]| \leq |\mathbb{E}[\mathbb{1}(\neg G)]| + \sqrt{\mathbb{E}[P'^2] \cdot \mathbb{E}[\mathbb{1}(\neg G)]}.$$

Since $\mathbb{E}[A_i] = 0$ for all i and $\mathrm{L}_1(S_k) \leq \ell^{O(k)}$, using Equation (5.5), it follows that under (Ck, ε)-wise independence, we have $\mathbb{E}[P'^2] \leq O(1)$. So, we can bound the RHS from above by $O(\ell^{-0.2k})$. □

□

5.4 Completing the analysis

Proof of Theorem 1.9. The proof uses an hybrid argument. The $\mathcal{G}_{\mathcal{MR}}$ generator chooses $g_0 : [n] \to [m_0]$, and then $g'_1, \ldots, g'_T$ where $g'_t = [m_{t-1}] \times [n] \to [m_t]$ defines the map

$$g_t(i) = g'_t(g_{t-1}(i), i).$$

Let $h_0, h'_1, \dots, h'_t$ be truly random hash functions with similar domains and ranges. For $0 \le t, q \le T$, define the hybrid family $\mathcal{G}_t^q = \{f_t^q : [n] \to [m_t]\}$ as follows: for $t = 0$ and every q, define

$$f_0^q = \begin{cases} g_0 & \text{for } q = 0, \\ h_0 & \text{for } q > 0, \end{cases}$$

and for $t > 0$ and every q,

$$f_t^q(i) = \begin{cases} g'_t(f_{t-1}^q(i), i) & \text{for } t \ge q, \\ h'_t(f_{t-1}^q(i), i) & \text{for } t < q. \end{cases}$$

For every q, let $\mathcal{G}^q = \mathcal{G}_T^q$. Thus, $\mathcal{G}^0 = \mathcal{G}_{\mathcal{MR}}$ and $\mathcal{G}^T = \mathcal{U}$. We will show that for every $q \ge 0$,

$$\left| \Pr_{f^{q+1} \in \mathcal{G}^{q+1}}[\forall\, i \in [n]\ f^{q+1}(i) \in S_i] - \Pr_{f^q \in \mathcal{G}^q}[\forall\, i \in [n]\ f^q(i) \in S_i] \right| \le \delta' = \delta/T.$$

The desired bound then follows by the triangle inequality.

In the case $q = 0$, couple $\mathcal{G}^0$ and $\mathcal{G}^1$ by picking the same $g'_1, \dots, g'_T$, and use them to define the function $f' : [m_1] \times [n] \to [m]$ so that

$$f^0(i) = f'(g_0(i), i),\ f^1(i) = f'(h_0(i), i).$$

For $i \in [n]$, define

$$S'_i = \{a \in [m_1] : f'(a, i) \in S_i\}.$$

Thus,

$$\begin{aligned} &\left| \Pr_{f^1 \in \mathcal{G}^1}[\forall\, i \in [n]\ f^1(i) \in S_i] - \Pr_{f^0 \in \mathcal{G}^0}[\forall\, i \in [n]\ f^0(i) \in S_i] \right| \\ &\qquad = \left| \Pr_{h_0}[\forall\, i \in [n]\ h_0(i) \in S'_i] - \Pr_{g_0}[\forall\, i \in [n]\ g_0(i) \in S'_i] \right| \le \delta', \end{aligned}$$

by applying Lemma 5.2 with $k = O(\log(1/\delta'))$ and $\varepsilon = \delta' \cdot m_0^{-O(k)}$.

For the case $q > 0$, couple $\mathcal{G}^{q+1}$ and $\mathcal{G}^q$ by picking the same $g'_{q+1}, \dots, g'_T$, and pick $x \in [m_{q-1}]^n$ uniformly at random. There is a function $f' : [m_q] \times [n] \to [m]$ so that

$$f^q(i) = f'(h'_q(x_i, i), i),\ f^{q-1}(i) = f'(g'_q(x_i, i), i).$$

As before, define

$$S_i = \{a \in [m_q] : f'(a, i) \in S_i\}.$$

Hence,

$$\begin{aligned} &\left| \Pr_{f^{q+1} \in \mathcal{G}^{q+1}}[\forall\, i \in [n]\ f^{q+1}(i) \in S_i] - \Pr_{f^q \in \mathcal{G}^q}[\forall\, i \in [n]\ f^q(i) \in S_i] \right| \\ &\qquad = \left| \Pr_{h'_q, x}[\forall\, i \in [n]\ h'_q(x_i, i) \in S'_i] - \Pr_{g'_q, x}[\forall\, i \in [n]\ g'_q(x_i, i) \in S'_i] \right| \le \delta', \end{aligned}$$

by Lemma 5.3 with

$$k_q > 1,\ m_{q-1} \ge \log(1/\delta'),\ m_{q-1} \ge k_q,\ m_{q-1}^{-k_q} \le \delta'^C,\ \varepsilon_{q-1} \le (m_q m_{q-1})^{-Ck_q}. \qquad \square$$

5.5 Minima-wise independence

Saks et al. [19] showed how to translate a PRG for combinatorial rectangles to an approximately minima-wise independent family. We conclude with a routine extension of their result to large ℓ.

Proof of Theorem 1.10. Fix $S \subseteq [n]$ and a sequence $T = (t_1, \ldots, t_\ell)$ of ℓ distinct elements from S. The event

$$g(t_1) < \cdots < g(t_\ell) < \min g(S \setminus T)$$

can be viewed as the disjoint union of $\binom{m}{\ell}$ events by fixing the set $A = \{a_1 < \ldots < a_\ell\}$ that T maps to. The indicator $\mathbb{1}_A$ of the event

$$g(t_1) = a_1, \ldots,\ g(t_\ell) = a_\ell,\ g(S \setminus T) > a_\ell$$

is a combinatorial rectangle: Define

$$\begin{aligned} f_i(x_i) &= 1 \text{ for } i \notin S \\ f_i(x_i) &= \mathbb{1}(x_i = a_j) \text{ for } i = t_j \in T \\ f_i(x_i) &= \mathbb{1}(x_i > a_\ell) \text{ for } i \in S \setminus T \end{aligned}$$

and

$$f_A(x_1, \ldots, x_n) = \prod_{i \in [n]} f_i(x_i).$$

Since $g(i) = x_i$, it follows that $\mathbb{1}_A(g) = f_A(x)$. Further, choosing $h \in \mathcal{U}$ is equivalent to choosing $x \in [m]^n$ uniformly at random. Hence,

$$\begin{aligned} &\Pr_{g \in \mathcal{G}}[g(t_1) < \cdots < g(t_\ell) < \min g(S \setminus T)] \\ &\qquad = \sum_A \mathbb{E}_{y \in \{0,1\}^r}[f_A(\mathcal{G}(y))] \\ &\qquad = \sum_A (\mathbb{E}_{h \in \mathcal{U}}[\mathbb{1}_A(h)] \pm \varepsilon) \\ &\qquad = \Pr_{h \in \mathcal{U}}[h(t_1) < \cdots < h(t_\ell) < \min h(S \setminus T)] \pm \binom{m}{\ell} \varepsilon. \end{aligned}$$

□

Acknowledgments

We thank Nati Linial, Raghu Meka, Yuval Peres, Dan Spielman, Avi Wigderson and David Zuckerman for helpful discussions. We thank an anonymous referee for pointing out an error in the statement of Theorem 1.6 in a previous version of the paper.

References

[1] MIKLÓS AJTAI AND AVI WIGDERSON: Deterministic simulation of probabilistic constant depth circuits. In SILVIO MICALI AND FRANCO PREPARATA, editors, *Randomness and Computation*, volume 5 of *Adv. in Computing Research*, pp. 199–222. JAI Press, 1989. Available on author's homepage. Preliminary version in FOCS'85. 8

[2] NOGA ALON, LÁSZLÓ BABAI, AND ALON ITAI: A fast and simple randomized parallel algorithm for the maximal independent set problem. *J. Algorithms*, 7(4):567–583, 1986. [doi:10.1016/0196-6774(86)90019-2] 16

[3] ROY ARMONI, MICHAEL E. SAKS, AVI WIGDERSON, AND SHIYU ZHOU: Discrepancy sets and pseudorandom generators for combinatorial rectangles. In *Proc. 37th FOCS*, pp. 412–421. IEEE Comp. Soc., 1996. [doi:10.1109/SFCS.1996.548500] 2, 6

[4] ANDREI Z. BRODER, MOSES CHARIKAR, ALAN M. FRIEZE, AND MICHAEL MITZENMACHER: Min-wise independent permutations. *J. Comput. System Sci.*, 60(3):630–659, 2000. Preliminary version in STOC'98. [doi:10.1006/jcss.1999.1690] 2, 6

[5] ANDREI Z. BRODER, MOSES CHARIKAR, AND MICHAEL MITZENMACHER: A derandomization using min-wise independent permutations. *J. Discr. Algorithms*, 1(1):11–20, 2003. Preliminary version in RANDOM'98. [doi:10.1016/S1570-8667(03)00003-0] 2, 6

[6] DEAN DORON, POOYA HATAMI, AND WILLIAM M. HOZA: Log-seed pseudorandom generators via iterated restrictions. In *Proc. 35th Comput. Complexity Conf. (CCC'20)*, pp. 1–36. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2020. [doi:10.4230/LIPIcs.CCC.2020.6] 8

[7] GUY EVEN, ODED GOLDREICH, MICHAEL LUBY, NOAM NISAN, AND BOBAN VELIČKOVIĆ: Efficient approximation of product distributions. *Random Struct. Algor.*, 13(1):1–16, 1998. Preliminary version in STOC'92. [doi:10.1002/(SICI)1098-2418(199808)13:1<1::AID-RSA1>3.0.CO;2-W] 2, 4, 6, 20

[8] PARIKSHIT GOPALAN, DANIEL M. KANE, AND RAGHU MEKA: Pseudorandomness via the discrete Fourier transform. *SIAM J. Comput.*, 47(6):2451–2487, 2018. Preliminary version in FOCS'15. [doi:10.1137/16M1062132, arXiv:1506.04350] 7

[9] PARIKSHIT GOPALAN, RAGHU MEKA, OMER REINGOLD, LUCA TREVISAN, AND SALIL P. VADHAN: Better pseudorandom generators from milder pseudorandom restrictions. In *Proc. 53rd FOCS*, pp. 120–129. IEEE Comp. Soc., 2012. [doi:10.1109/FOCS.2012.77, arXiv:1210.0049] 2, 3, 4, 5, 6, 7, 8, 19

[10] PARIKSHIT GOPALAN AND AMIR YEHUDAYOFF: Inequalities and tail bounds for elementary symmetric polynomials. *Electron. Colloq. Comput. Complexity*, TR14-019, 2014. [ECCC] 7

[11] PIOTR INDYK: A small approximately min-wise independent family of hash functions. *J. Algorithms*, 38(1):84–90, 2001. Preliminary version in SODA'99. [doi:10.1006/jagm.2000.1131] 2, 4, 6

[12] NATHAN LINIAL, MICHAEL LUBY, MICHAEL E. SAKS, AND DAVID ZUCKERMAN: Efficient construction of a small hitting set for combinatorial rectangles in high dimension. *Combinatorica*, 17(2):215–234, 1997. Preliminary version in STOC'93. [doi:10.1007/BF01200907] 2, 3, 6, 7

[13] CHI-JEN LU: Improved pseudorandom generators for combinatorial rectangles. *Combinatorica*, 22(3):417–434, 2002. Preliminary version in ICALP'98. [doi:10.1007/s004930200021] 2, 6

[14] RAGHU MEKA, OMER REINGOLD, AND AVISHAY TAL: Pseudorandom generators for width-3 branching programs. In *Proc. 51st STOC*, pp. 626–637. ACM Press, 2019. [doi:10.1145/3313276.3316319, arXiv:1806.04256] 8

[15] KETAN MULMULEY: Randomized geometric algorithms and pseudorandom generators. *Algorithmica*, 16(4/5):450–463, 1996. Preliminary version in FOCS'92. [doi:10.1007/BF01940875] 6

[16] JOSEPH NAOR AND MONI NAOR: Small-bias probability spaces: Efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993. Preliminary version in STOC'90. [doi:10.1137/0222053] 19

[17] NOAM NISAN: Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992. Preliminary version in STOC'90. [doi:10.1007/BF01305237] 8

[18] GEORGE PÓLYA AND GEORGE SZEGŐ: *Problems and Theorems in Anaysis II*. Springer, 1976. 12

[19] MICHAEL E. SAKS, ARAVIND SRINIVASAN, SHIYU ZHOU, AND DAVID ZUCKERMAN: Low discrepancy sets yield approximate min-wise independent permutation families. *Information Processing Letters*, 73(1–2):29–32, 2000. Preliminary version in RANDOM'99. [doi:10.1016/S0020-0190(99)00163-5] 2, 6, 7, 26

[20] J. MICHAEL STEELE: *The Cauchy-Schwarz Master Class*. Cambridge Univ. Press, 2004. [doi:10.1017/CBO9780511817106] 3, 4

AUTHORS

Parikshit Gopalan
Senior Researcher
VMware
Palo Alto, CA, US
pgopalan@vmware.com
https://research.vmware.com/researchers/parikshit-gopalan

Amir Yehudayoff
Associate Professor
Department of Mathematics
Technion-IIT, Haifa, Israel
amir.yehudayoff@gmail.com
https://yehudayoff.net.technion.ac.il/

ABOUT THE AUTHORS

PARIKSHIT GOPALAN has been a researcher at Microsoft Research (Silicon Valley and Redmond) and previously a postdoc at UT Austin and the University of Washington. After completing his undergraduate studies at IIT Bombay, he received his Ph. D. in 2006 from the Algorithms, Combinatorics and Optimization program at Georgia Tech under the supervision of Richard Lipton. He has worked on coding theory, erasure coding for distributed storage and computational complexity. He is currently interested in algorithms and systems for big data and machine learning.

AMIR YEHUDAYOFF received his Ph. D. in 2008 from the Weizmann Institute of Science under the supervision of Ran Raz. Subsequently he spent two years at the Institute for Advanced Study in Princeton. He is currently an associate professor in the Department of Mathematics at the Technion–Israel Institute of Technology. His main research area is theoretical computer science, with a recent focus on learning theory and information theory.