Pseudorandom Generators for Combinatorial Shapes

Parikshit Gopalan
MSR-SVC
parik@microsoft.com

Raghu Meka[*]
UT Austin
raghu@cs.utexas.edu

Omer Reingold
MSR-SVC
omereing@microsoft.com

David Zuckerman[†]
UT Austin
diz@cs.utexas.edu

Abstract

We construct pseudorandom generators for *combinatorial shapes*, which substantially generalize combinatorial rectangles, ϵ-biased spaces, 0/1 halfspaces, and 0/1 modular sums. A function $f : [m]^n \to \{0,1\}$ is an (m,n)-combinatorial shape if there exist sets $A_1, \ldots, A_n \subseteq [m]$ and a symmetric function $h : \{0,1\}^n \to \{0,1\}$ such that $f(x_1, \ldots, x_n) = h(1_{A_1}(x_1), \ldots, 1_{A_n}(x_n))$. Our generator uses seed length $O(\log m + \log n + \log^2(1/\varepsilon))$ to get error ε. When $m = 2$, this gives the first generator of seed length $O(\log n)$ which fools all weight-based tests, meaning that the distribution of the weight of any subset is ε-close to the appropriate binomial distribution in statistical distance. Along the way, we give a generator for combinatorial rectangles with seed length $O(\log^{3/2} n)$ and error $1/\text{poly}(n)$, matching Lu's bound [ICALP 1998].

For our proof we give a simple lemma which allows us to convert closeness in Kolmogorov (cdf) distance to closeness in statistical distance. As a corollary of our technique, we give an alternative proof of a powerful variant of the classical central limit theorem showing convergence in statistical distance, instead of the usual Kolmogorov distance.

[*] Work done while an intern at Microsoft Research, Silicon Valley.
[†] Work done while visiting Microsoft Research, Silicon Valley.

1 Introduction

Pseudorandom generators are of fundamental importance in complexity theory, cryptography, and beyond. A pseudorandom generator (PRG) takes as input a short random seed and outputs a long string which appears random to a class of functions.

Definition 1.1. *A function $G : \{0,1\}^s \to [m]^n$ is a pseudorandom generator (PRG) with seed length s and error ε for a class of functions $\mathcal{F} : [m]^n \to \{0,1\}$ – or more succinctly, G ε-fools $\mathcal{F}$ with seed length s – if for all $f \in \mathcal{F}$,*

$$\left| \Pr_{x \in_u \{0,1\}^s} [f(G(x)) = 1] - \Pr_{y \in_u [m]^n} [f(y) = 1] \right| \leq \varepsilon.$$

While we know very strong PRGs under computational assumptions, constructing provably-good PRGs without assumptions is a major challenge. Some of the most powerful unconditional constructions are PRGs for space-bounded computations. In particular, the PRGs of Nisan [Nis92] and Impagliazzo, Nisan, and Wigderson [INW94] use a seed of length $O(\log^2 n)$ to fool polynomial-width branching programs. These generators have played a central role in studying the relative strength of randomness vs. memory. In particular, reducing their seed length to $O(\log n)$-bit would show that RL=L, namely every randomized algorithm can be derandomized with only a multiplicative constant blow-up in its memory. Improving [Nis92, INW94] is a central open question, not only for the possibility of proving RL=L, but also for other important applications [Ind00, Siv02, KNR05, HHR06]. Despite much effort, the above seed lengths have not been improved in nearly two decades.

While PRGs with logarithmic-seed that fool polynomial-width branching programs are still not known, logarithmic-seed PRGs for weaker classes of distinguishers have been previously constructed and found many applications. In this paper we define a natural common generalization and significant extension of many of these distinguisher classes, which we name *combinatorial shapes*. Combinatorial shapes look at their inputs in consecutive chunks of $\log m$ bits (usually m would be at most polynomial in n). On each chunk of bits the combinatorial shape may apply an arbitrary boolean function. Nevertheless, these Boolean functions are combined into a single output by a symmetric (i.e., order independent) function. Combinatorial shapes generalize combinatorial rectangles, halfspaces with 0/1 coefficients, and modular sums. Our main result is a construction of PRGs with seed length $O(\log n)$ that fools combinatorial shapes.

Definition 1.2. *A function $f : [m]^n \to \{0,1\}$ is an (m,n)-combinatorial shape if there exist sets $A_1, \ldots, A_n \subseteq [m]$ and a symmetric function $h : \{0,1\}^n \to \{0,1\}$ such that $f(x_1, \ldots, x_n) = h(1_{A_1}(x_1), \ldots, 1_{A_n}(x_n))$. We denote the class of such functions by $\mathsf{CShape}(m,n)$.*

We call them *combinatorial shapes* because they generalize combinatorial rectangles, which are simply the subset of $\mathsf{CShape}(m,n)$ where the symmetric function h is the AND function. PRGs for combinatorial rectangles have received considerable attention [EGL$^+$92, ASWZ96, Lu02], and have applications to numerical integration.

The class $\mathsf{CShape}(2,n)$ is interesting in its own right, as it comprises all Boolean functions $f : \{0,1\}^n \to \{0,1\}$ that are symmetric functions of a subset $S \subseteq [n]$ of variables. In order to fool $\mathsf{CShape}(2,n)$, the distribution of $\sum_{i \in S} x_i$ needs to be ε-close to $\mathsf{BIN}(|S|, \frac{1}{2})$ in *statistical* distance for every $S \subseteq [n]$. [1] Prior to our work, the best known generator for this problem was Nisan's generator [Nis92] which gives seed-length $O(\log^2 n)$. Similarly, PRGs for $\mathsf{CShape}(m,n)$ imply generators that

[1] For $n > 0$, $p \in [0,1]$, $\mathsf{BIN}(n,p)$ denotes the binomial distribution of order n and bias p.

can fool such tests under multinomial distributions, by choosing the set A_i so that $1_{A_i}(x_i) = 1$ with probability p_i.

Parities of subsets are a special case of $\mathsf{CShape}(2, n)$; hence PRGs that fool $\mathsf{CShape}(2, n)$ are a strengthening of the ever so versatile ε-biased generators [NN93]. Recently, a different strengthening of ε-biased generators was considered, where bit-generators were given that fool sums modulo larger primes or even composites [LRTV09, MZ09a]. The seed-length of these constructions is super-logarithmic unless the moduli is constant. It is easy to argue that a generator that fools $\mathsf{CShape}(2, n)$ also fools sums modulo an arbitrary moduli, or even non-modular sums.[2]

Note that in the above examples of combinatorial shapes, the symmetric function h could be computed by a constant width branching program. In this sense, combinatorial shapes seem significantly more powerful. Halfspaces with $0/1$ coefficients are also special cases of $\mathsf{CShape}(2, n)$, where the symmetric function cannot be evaluated by a constant width branching program. PRGs which fool halfspaces were recently given in [DGJ$^+$10, MZ10]; the latter will be a useful tool in our construction. Note however that these results only guarantee that $\sum_{i \in S} x_i$ is close to $\mathsf{BIN}(|S|, \frac{1}{2})$ in Kolmogorov distance, whereas our goal is to get closeness in statistical distance. (For definitions of these distances, see Section 2.)

1.1 Main Results

Our main result is a PRG construction which fools $\mathsf{CShape}(m, n)$.

Theorem 1.3 (Main)**.** *For every $\varepsilon > 0$, there exists a PRG that ε-fools $\mathsf{CShape}(m, n)$ with seed-length $O(\log m + \log n + \log^2(1/\varepsilon))$.*

When m is polynomial in n, these PRGs have seed length $O(\log n + \log^2(1/\varepsilon))$. Previously, the best known PRGs had seed length $O(\log^2 n)$, even for $m = 2$; these were the PRGs for space-bounded computation by Nisan and Impagliazzo, Nisan and Wigderson.

Along the way we also give a new PRG for combinatorial rectangles with seed-length $O(\log^{3/2} n)$ and error $1/\mathrm{poly}(n)$. This matches the parameters of the previous best generator due to Lu [Lu02] for polynomially small ε.

Theorem 1.4. *For every $\varepsilon > 0$, there exists a generator that ε-fools (m, n)-combinatorial rectangles with seed-length $O(\log n \sqrt{\log(1/\varepsilon)})$.*

Our constructions are based on a simple lemma about the convolution of two real-valued distributions. This lemma enables us to amplify closeness in Kolmogorov distance to closeness in statistical distance. We further use this lemma to give a new proof of a powerful variant of the classical Central Limit Theorem which guarantees convergence to the appropriate binomial distribution in statistical distance, as opposed to Kolmogorov distance.

The classical Central Limit Theorem (CLT) says that a sum of independent random variables should be close, in Kolmogorov distance, to the corresponding Gaussian or Binomial random variable. The Kolmogorov distance is weaker than statistical (total variation) distance $\mathsf{d_{TV}}$, since Kolmogorov distance allows only special types of statistical tests, namely threshold functions. Nevertheless, if the random variables are integer-valued, then under some reasonable conditions it is known that a sum of independent variables approaches the appropriate binomial distribution in statistical distance. Such theorems are called *discrete central limit theorems.*

For clarity, in the introduction we only state our discrete central limit theorem for the case of multinomial distributions.

[2] Note that [LRTV09, MZ09a] gives generators that fool sums with arbitrary coefficients. Generators that fool $\mathsf{CShape}(2, n)$ also fool modular (and non-modular) sums with $0/1$ coefficients.

Theorem 1.5. *Let $X_1, \ldots, X_n$ be independent indicator random variables with $\mathsf{Pr}[X_i = 1] = p_i$. Let $X = \sum_i X_i$, $\mathbb{E}[X] = \mu, \mathrm{Var}(X) = \sum_i p_i(1-p_i) = \sigma^2$. Then, for $Z \leftarrow \mathsf{BIN}(m,q)$, where $m = \mu^2/(\mu - \sigma^2)$, $q = (\mu - \sigma^2)/\mu$, $\mathsf{d}_{\mathsf{TV}}(X,Z) = O\left(\sqrt{\log(\sigma)/\sigma}\right)$.*

The parameters m, q above are chosen so that $\mathbb{E}[Z] = \mathbb{E}[X]$ and $\mathrm{Var}[Z] = \mathrm{Var}[X]$. Limit theorems as above with almost optimal error estimates $(\Theta(1/\sigma))$ are known in the probability literature (see [BX99, BC02] and references therein). However, most previous results use Fourier techniques or Stein's method and appear significantly more complicated, at least to us. In contrast our proof is elementary, relying only on the classical Berry-Esséen theorem and few simple properties of the binomial distribution. We also obtain a more general *invariance principle*, Theorem 4.2, for the case of sums of integer-valued random variables.

Discrete central limit theorems as above have, at least implicitly, been used before in computer science. Two prominent instances are the works of Daskalakis and Papadimitriou [DP07, DP08]. A main technical result in these works can be viewed as a discrete limit theorem and roughly says the following: given a multinomial distribution (or more generally, a multivariate-multinomial distribution), the probabilities of each of the indicator variables can be rounded to multiples of a parameter $1/\varepsilon$, so as to not incur too much of a loss in statistical distance. Their arguments for showing the discrete CLT are quite involved and use a variety of sampling and Poisson approximation techniques. Given the generality of our argument for proving Theorem 1.5, it is conceivable that a similar argument can be extended to the more nuanced discrete limit theorems of [DP07, DP08].

1.2 Outline of Constructions

We say that a random variable Y is α-shift invariant if $\mathsf{d}_{\mathsf{TV}}(Y, Y+1) \leq \alpha$. Several common distributions, such as binomial, Gaussian, and multinomial distributions, are all shift-invariant, roughly, inversely proportional to their standard deviation.

The starting point for our results is the following lemma, which says that two distributions that are close in Kolmogorov distance when convolved with a shift-invariant distribution become close in statistical distance.

Lemma 1.6 (Main Convolution Lemma). *Let X be a α-shift invariant distribution and let Y, Z be integer-valued distributions with support contained in $[a, a+b]$ for some $a \in \mathbb{R}, b > 0 \in \mathbb{R}$. Then,*

$$\mathsf{d}_{\mathsf{TV}}(X+Y, X+Z) \leq 4\sqrt{\alpha b \mathsf{d}_{\mathsf{cdf}}(Y,Z)}.$$

We next sketch the proof of the discrete central limit theorem Theorem 1.5, as similar (though somewhat more involved) ideas underlie our PRG for fooling combinatorial shapes. We partition the variables into two sets S and T such that $X_S = \sum_{i \in S} X_i$ and $X_T = \sum_{j \in T} X_i$ have approximately the same mean and variance. We introduce variables Y_S and Y_T which are two independent copies of $\mathsf{BIN}(m/2, q)$. Then, the Berry-Esséen theorem, which is a quantitative form of the classical central limit theorem, guarantees the closeness of X_S, Y_S and X_T, Y_T in Kolmogorov distance. Secondly, multinomial distributions are shift-invariant. Hence we bound the statistical distance between $X_S + X_T$ and $Y_S + Y_T$, by using our Convolution lemma to show that each of them is close to $X_S + Y_T$ in statistical distance.

1.2.1 PRG for Combinatorial Shapes

For intuition, it is easier to work with the equivalent goal of fooling *combinatorial sums* in statistical distance.

Definition 1.7. *A function $f : [m]^n \to [n]$ is an (m,n)-combinatorial sum if there exist sets $A_1, \ldots, A_n \subseteq [m]$ such that $f(x_1, \ldots, x_n) \equiv 1_{A_1}(x_1) + 1_{A_2}(x_2) + \cdots + 1_{A_n}(x_n)$. We denote this class of functions by* $\mathsf{CSum}(m,n)$.

It is straightforward to verify that fooling combinatorial shapes is equivalent to fooling combinatorial sums in the stronger, statistical distance.

The basic building block for our constructions is a natural extension, $G_{\mathcal{H},k,t}$, of the main generator for fooling halfspaces over $\{0,1\}^n$ of Meka and Zuckerman [MZ10] (see Equation 5.1 for the exact definition), which in turn is a simplified version of a hitting set generator due to Rabani and Shpilka [RS10]. The generator $G_{\mathcal{H},k,t}$ uses a random hash function from $\mathcal{H}$ to map variables to t buckets pairwise independently and then uses $k = O(1)$-wise independence within each bucket.

Our high level approach to fooling combinatorial sums is as follows:

1. We first show that $G_{\mathcal{H},k,t}$ fools combinatorial sums with small variance in statistical distance. We show that since the combinatorial sum restricted to each bucket has very small variance, bounded independence fools the sum restricted to a bucket in statistical distance. We then take a union bound across the different buckets. A weak bound for fooling the sum in each bucket is easy; however to apply the union bound requires a much stronger bound, which we prove using the "sandwiching polynomials" technique introduced by Bazzi [Baz09].
2. We then show that $G_{\mathcal{H},k,t}$ fools combinatorial sums with high variance in Kolmogorov distance. We use the pairwise independence of $\mathcal{H}$ to argue that the total variance is well spread among the t buckets and then apply the Berry-Esséen theorem to show that the distribution is close to the right distribution in Kolmogorov distance. The analysis for this case is similar to the argument of Meka and Zuckerman [MZ10] for regular halfspaces.
3. We construct a generator $H_{m,n}$ fooling n dimensional combinatorial sums in statistical distance by recursively combining a generator fooling $n/2$ dimensional sums in Kolmogorov distance with a generator fooling $n/2$ dimensional sums in statistical distance. Unfolding this recursion, the generator $H_{m,n}$ hashes variables into $\log n$ buckets of geometrically increasing sizes and applies the generator $G_{\mathcal{H},k,t}$ to each bucket. We analyze this generator by exploiting the recursive construction to apply Lemma 1.6 at every step. We view this recursive construction and analysis of the $H_{m,n}$ as the most novel part of our PRG construction. The analysis, while similar in spirit to our proof of the discrete central limit theorem Theorem 1.5 is more involved.
4. Finally, we show that one can generate the seeds for each bucket using the PRGs for small-space sources of [INW94], [NZ96] rather than independently. This is done by constructing small-width sandwiching branching programs for combinatorial sums.

We obtain our result on fooling combinatorial rectangles by setting the parameters of $G_{\mathcal{H},k,t}$ appropriately and then derandomizing the construction using [Nis92, INW94] as above. The analysis however is different and uses a simple application of the principle of inclusion-exclusion and few properties of k-wise independent hash functions.

1.3 Prior Work

Independently and simultaneously, Watson [Wat11] studied the special case of combinatorial shapes where the symmetric function h is the parity function which are called *combinatorial checkerboards* by Watson. Watson obtains a seed-length of $O(\log m + \log n \log\log n + \log^{3/2}(1/\epsilon))$ which is better than the seed-length we get for small ϵ.

As indicated earlier, PRGs for several special cases of combinatorial shapes have been studied previously. There was a lot of classical work on low-discrepancy sets for axis-parallel rectangles in low dimension; see for example [Mat99]. Even, Goldreich, Luby, Nisan, and Velickovic [EGL+92] were the first to give good constructions in high dimension; they gave PRGs for combinatorial rectangles which used an $O(\log^2 n)$ bit seed to achieve error $1/\mathrm{poly}(n)$ when $m = \mathrm{poly}(n)$. Armoni, Saks, Wigderson, and Zhou [ASWZ96] improved the parameters to achieve a seed of length $O(\log m + \log n + \log^2(1/\varepsilon))$. The best construction is by Lu [Lu02], who achieved a seed length of $O(\log m + \log n + \log^{3/2}(1/\varepsilon))$.

Diakonikolas, Gopalan, Jaiswal, Servedio, and Viola [DGJ+10] showed that $O(\log^2(1/\varepsilon)/\varepsilon^2)$-wise independence ε-fools halfspaces, which gives a seed of length $O((\log n)\log^2(1/\varepsilon)/\varepsilon^2)$. Meka and Zuckerman [MZ10] gave a different PRG with seed length $O(\log n + \log^2(1/\varepsilon))$.

The notion of ε-biased spaces was introduced by Naor and Naor [NN93], who gave a PRG using $O(\log n + \log(1/\varepsilon))$ bits. Alon, Goldreich, Hastad, and Peralta [AGHP92] gave alternate constructions matching this bound. Lovett, Reingold, Trevisan, and Vadhan [LRTV09] gave a PRG over bits that fools sums modulo m, requiring a seed of length $O(\log n + \log(m/\varepsilon)\log(m\log(1/\varepsilon)))$. A similar, somewhat weaker construction was found independently by Meka and Zuckerman [MZ09a].

2 Notation and Preliminaries

We use the following notation.

- Most upper case letters $X, Y, Z, \ldots$ denote real-valued random variables.
- For integer-valued random variables X, Y, the statistical distance $\mathsf{d}_{\mathsf{TV}}(X, Y)$ between X, Y is defined as follows:
$$\mathsf{d}_{\mathsf{TV}}(X,Y) \equiv \sup_{A \subseteq \mathbb{Z}} |\,\mathsf{Pr}[X \in A] - \mathsf{Pr}[Y \in A]| = \frac{1}{2}\sum_i |\,\mathsf{Pr}[X = i] - \mathsf{Pr}[Y = i]|.$$
- For real-valued random variables X, Y, the Kolmogorov distance (or cdf distance) $\mathsf{d}_{\mathsf{cdf}}(X, Y)$ between X, Y is defined by $\mathsf{d}_{\mathsf{cdf}}(X,Y) \equiv \sup_{\theta \in \mathbb{R}} |\,\mathsf{Pr}[X < \theta] - \mathsf{Pr}[Y < \theta]|$.
- For a real-valued random variable X, we let $\mu(X), \sigma(X), \mathrm{Var}(X)$ denote the expectation, standard deviation and variance of X respectively. For $a, b > 0$, $\mathcal{N}(a, b)$ denotes the Gaussian distribution with mean a and variance b.

We use the following formulation of the Berry-Esséen theorem:

Theorem 2.1 ([Fel71], [She07])**.** *For $Y = \sum_i Y_i$ a sum of independent random variables and $Z \leftarrow \mathcal{N}(0,1)$,*
$$\mathsf{d}_{\mathsf{cdf}}\left(\frac{Y - \mathbb{E}[Y]}{\sigma(Y)}, Z\right) \leq \frac{(\,\sum_i \mathbb{E}[\,|Y_i - \mathbb{E}[Y_i]|^3\,]\,)}{\sigma(Y)^3} \leq \frac{(\,\sum_i \mathbb{E}[\,|Y_i - \mathbb{E}[Y_i]|^4\,]\,)^{1/2}}{\sigma(Y)^2}.$$

The proofs of the following simple facts are presented in Appendix A

Corollary 2.2 (Berry-Esséen for Multinomials)**.** *For $Y = \sum_i Y_i$ a sum of independent indicator variables, $Z \leftarrow \mathcal{N}(0,1)$,*
$$\mathsf{d}_{\mathsf{cdf}}\,(\,(Y - \mathbb{E}(Y))/\sigma(Y), Z\,) \leq 1/\sigma(Y).$$

Proof. Follows from Theorem 2.1, as for $0,1$ valued Y_i, $\sum_i \mathbb{E}[|Y_i - \mathbb{E}[Y_i]|^4] \leq \sum_i \mathbb{E}[|Y_i - \mathbb{E}[Y_i]|^2]$. □

Fact 2.3. *For $Z_1 \leftarrow \mathcal{N}(\mu_1, \sigma_1)$, $Z_2 \leftarrow \mathcal{N}(\mu_2, \sigma_2)$, for $\sigma_1 \geq 1$,*

$$\mathsf{d}_{\mathsf{cdf}}(Z_1, Z_2) = O\left(\frac{|\mu_1 - \mu_2|}{\sigma_1} + \frac{\sqrt{|\sigma_1^2 - \sigma_2^2| \log(\sigma_1)}}{\sigma_1}\right).$$

Fact 2.4. *Any multinomial distribution X with $\mathrm{Var}(X) = \sigma^2$ is $(2/\sigma)$-shift invariant.*

Fact 2.5. *For any multinomial distribution X, and $\delta > 0$, $\mathsf{Pr}[\, |X - \mathbb{E}[X]| \geq 3\sigma(X)\sqrt{\log(1/\delta)}\,] \leq \delta$.*

Below we define some of the standard tools in derandomization that we use.

Definition 2.6 (Hash Families)**.** *A family of hash functions $\mathcal{H} = \{h : [n] \to [t]\}$ is k-wise independent if for all distinct $i_1, \ldots, i_k \in [n]$ and $\ell_1, \ldots, \ell_k \in [t]$,*

$$\mathsf{Pr}_{h \in_u \mathcal{H}}[\, h(i_1) = \ell_1 \,\wedge\, h(i_2) = \ell_2 \,\wedge \cdots \wedge h(i_k) = \ell_k \,] = \frac{1}{t^k}.$$

Efficient constructions of $\mathcal{H}$ as above with $|\mathcal{H}| = O(n^k)$ are known. A family of Pairwise-independent permutations $\mathcal{H} = \{h : [n] \to [n]\}$ is defined similarly, with the additional requirement that the hash functions $h : [n] \to [n]$ be permutations.

Definition 2.7 (k-wise independent spaces)**.** *A generator $G : \{0,1\}^r \to [m]^n$ is said to generate a k-wise independent space if for $y \in_u \{0,1\}^r$, for all distinct $i_1, \ldots, i_k \in [n]$, $b_1, \ldots, b_k \in [m]$,*

$$\mathsf{Pr}[\, (G(y))_{i_1} = b_1 \,\wedge\, (G(y))_{i_2} = b_2 \,\wedge \cdots \wedge (G(y))_{i_k} = b_k \,] = \frac{1}{m^k}.$$

Efficient constructions of generators G as above with $r = O(k(\log m + \log n))$ are known. We also use the following generalization of k-wise independence to arbitrary non-uniform distributions.

Definition 2.8. *A collection of random variables $(X_1, \ldots, X_n)$ over a universe U is k-wise independent if for all $i_1, \ldots, i_k \in [n]$, $u_1, \ldots, u_k \in U$,*

$$\mathsf{Pr}[\, X_{i_1} = u_1 \,\wedge\, X_{i_2} = u_2 \,\wedge \cdots \wedge X_{i_k} = u_k \,] = \mathsf{Pr}[\, X_{i_1} = u_1 \,] \cdot \mathsf{Pr}[\, X_{i_2} = u_2 \,] \cdots \mathsf{Pr}[\, X_{i_k} = u_k \,].$$

Finally, we describe the pseudorandom generators for small-width read-once branching programs (ROBPs) of [Nis92, INW94, NZ96] which play a crucial role in reducing the seed length of our constructions. We remark that we only use these results in a black-box fashion.

Definition 2.9 (ROBP)**.** *A (S, D, T)-ROBP (read-once branching program) M is a layered directed multi-graph with $T+1$ layers and at most 2^S vertices in each layer. The first layer has a single start vertex v_0 and the vertices in the last layer are labeled 0 (accepting) or 1 (rejecting). For $0 \leq i < T$, a vertex v in layer i of M has at most 2^D outgoing edges labeled with distinct elements of $\{0,1\}^D$, all leading to a vertex in layer $i+1$.*

A ROBP M as above defines a natural function $M : (\{0,1\}^D)^T \to \{0,1\}$, where on input $(z^1, \ldots, z^T)$ we traverse the graph according to the edge labels $z^1, \ldots, z^T$ and output the label of the final vertex reached.

Definition 2.10 (PRGs for ROBPs)**.** *A generator $G : \{0,1\}^r \to (\{0,1\}^D)^T$ is said to ε-fool (S, D, T)-ROBPs if for all (S, D, T)-ROBPs M,*

$$\left| \mathsf{Pr}_{y \in_u \{0,1\}^r}[\, M(G(y)) = 1 \,] - \mathsf{Pr}_{x \in_u (\{0,1\}^D)^T}[\, M(x) = 1 \,] \right| \leq \varepsilon.$$

Nisan [Nis92] gave a PRG that ε-fools (S, D, T)-ROBPs with seed length $O((S+D+\log(T/\varepsilon))\log T)$. We use the PRG of Impagliazzo et al. [INW94] who gave a slightly better PRG with seed length $O(D + (S + \log(T/\varepsilon))\log T)$ for fooling (S, D, T)-ROBPs with error ε. We also use the result of Nisan and Zuckerman [NZ96] who obtained a better PRG for the case when $T = \mathrm{poly}(S, D)$. In particular, they gave a PRG with seed length $O(S + D)$ for fooling (S, D, T)-ROBPs with error ε, when $T = \mathrm{poly}(S, D)$ and $\varepsilon \geq 2^{\log^{1-\gamma}(S+D)}$ for arbitrary $\gamma > 0$.

3 Main Convolution Lemma

We now prove Lemma 1.6. Recall that it enables us to translate closeness in Kolmogorov distance to closeness in statistical distance, and hence plays a key role in our results. The lemma says that if we consider two distributions Y, Z that are close in cdf distance and bounded by b, and convolve them with a distribution which is $(1/b)$-shift invariant, then the resulting distributions are statistically close.

Proof of Lemma 1.6. Without loss of generality suppose that Y, Z are supported in $[0, b)$. For $d \in \mathbb{Z}_+$ to be chosen later, let Y_d be the integer random variable with support over $S_d = \{id : i \in \mathbb{Z}_+, i \leq \lfloor b/d \rfloor\}$, with pdf p_d defined by, $p_d(id) = \mathsf{Pr}[Y \in [id, (i + 1)d)]$. We first show that

$$\mathsf{d}_{\mathsf{TV}}(X + Y, X + Y_d) \leq \alpha d. \tag{3.1}$$

There is a natural coupling of Y and Y_d: we set $Y_d = id$ with probability $p_d(id)$ and then sample $Y = Y_d + \bar{Y}$ from the interval $[id, (i + 1)d)$ according to the marginal distribution of Y conditioned on the event that $Y \in [id, (i + 1)d)$. Note that $\bar{Y} \in \{0, 1, \ldots, d - 1\}$ and it is an integer. We have

$$\mathsf{d}_{\mathsf{TV}}(X + Y, X + Y_d) = \mathsf{d}_{\mathsf{TV}}(X + Y_d + \bar{Y}, X + Y_d).$$

Further, conditioned on a particular value of $Y_d = id$,

$$\mathsf{d}_{\mathsf{TV}}(X + Y_d + \bar{Y}, X + Y_d) = \mathsf{d}_{\mathsf{TV}}(X + \bar{Y}, X) \leq \alpha d,$$

where the last inequality follows from the shift invariance of X and the fact that $\bar{Y} \in \{0, \ldots, d-1\}$. Therefore,

$$\mathsf{d}_{\mathsf{TV}}(X + Y, X + Y_d) = \mathsf{d}_{\mathsf{TV}}(X + Y_d + \bar{Y}, X + Y_d) \leq \alpha d.$$

We define Z_d similarly. It follows that $\mathsf{d}_{\mathsf{TV}}(X + Z, X + Z_d) \leq \alpha d$. Next we bound $\mathsf{d}_{\mathsf{TV}}(Y_d, Z_d)$.

Observe that Y_d, Z_d both have supports of size at most b/d. For any i,

$$|\,\mathsf{Pr}[Y_d = id] - \mathsf{Pr}[Z_d = id]\,| = |\,\mathsf{Pr}[Y \in [id, (i + 1)d)\,] - \mathsf{Pr}[Z \in [id, (i + 1)d)\,]\,| \leq 2\mathsf{d}_{\mathsf{cdf}}(Y, Z).$$

Hence $\mathsf{d}_{\mathsf{TV}}(Y_d, Z_d) \leq (2b/d)\mathsf{d}_{\mathsf{cdf}}(Y, Z)$. Combining the above equations,

$$\begin{aligned}\mathsf{d}_{\mathsf{TV}}(X + Y, X + Z) &\leq \mathsf{d}_{\mathsf{TV}}(X + Y, X + Y_d) + \mathsf{d}_{\mathsf{TV}}(X + Y_d, X + Z_d) + \mathsf{d}_{\mathsf{TV}}(X + Z_d, X + Z)\\ &\leq 2\alpha d + \frac{2b\mathsf{d}_{\mathsf{cdf}}(Y, Z)}{d}.\end{aligned}$$

The lemma now follows by setting $d = \lceil \sqrt{b\mathsf{d}_{\mathsf{cdf}}(Y, Z)/\alpha} \rceil$. □

One can weaken the boundedness requirement to say that Y and Z rarely exceed b. We record the following easy corollary without proof.

Corollary 3.1. *Let X be a α-shift invariant distribution and let Y, Z be two integer-valued distributions. Then, for $a \in R$ and $b \in \mathbb{R}^+$*

$$\mathsf{d}_{\mathsf{TV}}(X + Y, X + Z) \leq 4\sqrt{\alpha b \mathsf{d}_{\mathsf{cdf}}(Y, Z)} + \mathsf{Pr}[Y \notin [a, a + b)] + \mathsf{Pr}[Z \notin [a, a + b)].$$

4 Discrete Central Limit Theorems

We now prove the discrete central limit theorem Theorem 1.5. As outlined in the introduction, the proof proceeds by partitioning the variables appropriately and using the convolution lemma. The following easy fact (whose proof we omit) is used to partition the variables.

Fact 4.1. *Let $0 \le a_1 \le \cdots \le a_n \le 1$. Let $S \subset [n]$ consist of all odd indices. Then $|\sum_{i \in S} a_i - (\sum_j a_j)/2| \le a_n/2$.*

Proof of Theorem 1.5. Without loss of generality suppose that $\sigma_1 \le \sigma_2 \le \cdots \le \sigma_n$, where $\sigma_i = \sigma(X_i)$. Let S and T consist of odd and even indices respectively. Let $X_S = \sum_{i \in S}(X_i - \mathbb{E}[X_i])$ and $X_T = \sum_{i \in T}(X_i - \mathbb{E}[X_i])$. Let $\sigma_S^2 = \mathrm{Var}(X_S)$. Then, from Fact 4.1 $|\,\sigma_S^2 - \sigma^2/2\,| \le 1/2$.

Let Y_S, Y_T denote two independent copies of $(\mathsf{BIN}(m/2, q) - \mu/2)$ for m, q as in the theorem statement. Note that $Y_S + Y_T$ has distribution $\mathsf{BIN}(m, q) - \mu$ and that $\mathbb{E}[Y_S] = \mathbb{E}[Y_T] = 0$ and $\mathrm{Var}(Y_S) = \mathrm{Var}(Y_T) = \sigma^2/2$.

We proceed to bound the various quantities (α, B and $\mathsf{d_{cdf}}$) required to apply the convolution lemma. By Fact 2.4, X_S, Y_S, X_T, Y_T are all $\alpha = (2/\sigma)$-shift invariant. By Theorem 2.1 and Fact 2.3,

$$\begin{aligned}\mathsf{d_{cdf}}(X_S, Y_S) &\le \mathsf{d_{cdf}}(X_S, \mathcal{N}(0, \sigma_S^2)) + \mathsf{d_{cdf}}(Y_S, \mathcal{N}(0, \sigma^2/2)) + \mathsf{d_{cdf}}(\mathcal{N}(0, \sigma_S^2), \mathcal{N}(0, \sigma^2/2)) \\ &\le \frac{1}{\sigma} + \frac{1}{\sigma} + O\left(\frac{\sqrt{\log(\sigma)}}{\sigma}\right) = O\left(\frac{\sqrt{\log(\sigma)}}{\sigma}\right). \qquad (4.1)\end{aligned}$$

A similar bound holds for $\mathsf{d_{cdf}}(X_T, Y_T)$.

Next we show that X_S, X_T, Y_S, Y_T are bounded in a range $[-B, B]$ with probability $(1 - 1/\sigma)$. By Fact 2.5, for $B = 12(\sigma\sqrt{\log\sigma})$, $\mathsf{Pr}[\,|X_S| > B\,] \le 1/4\sigma$, and a similar statement holds for X_T, Y_S, Y_T. We then apply the union bound. Therefore, applying Corollary 3.1,

$$\begin{aligned}\mathsf{d_{TV}}(X_S + X_T, Y_S + Y_T) &\le \mathsf{d_{TV}}(X_S + X_T, X_S + Y_T) + \mathsf{d_{TV}}(X_S + Y_T, Y_S + Y_T) \\ &\le 4\sqrt{\alpha B \mathsf{d_{cdf}}(X_T, Y_T)} + 4\sqrt{\alpha B \mathsf{d_{cdf}}(X_S, Y_S)} + \frac{1}{\sigma} \\ &= O\left(\sqrt{\log(\sigma)/\sigma}\right). \quad \text{(By Equation 4.1)}\end{aligned}$$

□

We next generalize Theorem 1.5 to sums of independent integer-valued variables (as opposed to indicator random variables). The error term in the statistical distance guarantee we get depends on the Kolmogorov distance guarantee given by the Berry-Esséen theorem and on the shift invariance of the individual random variables. The dependence on these terms is in some sense unavoidable (as explained below). As for the case of indicator random variables our bound is weaker those those of the more fine-grained results of [BX99, BC02]. However, the arguments and exact technical conditions of [BX99, BC02] are complicated and the parameters we get are comparable up to $\Omega(1)$ factors in the exponents.

Theorem 4.2. *Let $\bar{X} = (X_1, \ldots, X_n), \bar{Y} = (Y_1, \ldots, Y_m)$ be two sets of independent integer-valued variables. Let $X = \sum_i X_i, Y = \sum_i Y_i$ and let $\mathbb{E}[X] = \mathbb{E}[Y]$, $\sigma^2 = \mathrm{Var}(X) = \mathrm{Var}(Y)$. Further, let*

$$\max_i\{\mathrm{Var}(X_i), \mathrm{Var}(Y_i)\} \le \sigma^2/2, \quad \max(\sum_i \mathbb{E}[\,|X_i - \mathbb{E}[X_i]|^3\,], \sum_i \mathbb{E}[\,|Y_i - \mathbb{E}[Y_i]|^3\,]\,) \le \rho,$$

$$4 \le U = \min(\sum_i (1 - \mathsf{d_{TV}}(X_i, X_i + 1)), \sum_j (1 - \mathsf{d_{TV}}(Y_j, Y_j + 1))\,).$$

Then,

$$\mathsf{d_{TV}}(X,Y) = O\left(\left(\frac{\rho \log(1/\sigma)}{\sigma^2 U^{1/2}} \right)^{1/2} + \frac{\rho}{\sigma^3} + \frac{1}{\sigma} \right).$$

Note that for a limit theorem as above to hold, we need assumptions on X, Y stronger than matching means and variances which was enough for the Berry-Esséen theorem. For instance, the X_i's could be supported on even integers and Y_i's on odd integers with X, Y having the same mean and variances. In this case the statistical distance between X, Y is 1, whereas the Kolmogorov distance could still be small. Thus, the additional assumption that X_i's, Y_i's have some shift-invariance is a natural restriction to have.

We use the following tricky generalization of Fact 4.1 whose proof uses Hall's theorem.

Lemma 4.3. *Given $a_1, \ldots, a_n > 0$ and $b_1, \ldots, b_n > 0$, there exists a set $S \subseteq [n]$ such that*

$$|\sum_{i \in S} a_i - \frac{\sum_j a_j}{2}| \leq \frac{\max_i a_i - \min_i a_i}{2}, \quad |\sum_{i \in S} b_i - \frac{\sum_j b_j}{2}| \leq \frac{\max_i b_i - \min_i b_i}{2}.$$

Proof. Let n be even, the case of n odd is similar. Let $A = \sum_i a_i, B = \sum_i b_i$. Suppose that $a_1 \leq a_2 \leq \cdots a_n$ and let $\pi : [n] \to [n]$ be such that $b_{\pi(1)} \leq b_{\pi(2)} \leq \cdots \leq b_{\pi(n)}$. Form a bipartite graph $G = (L, R, E)$, where $|L| = |R| = [n/2]$ with vertices on left corresponding to pairs $\{(a_1, a_2), (a_3, a_4), \ldots, (a_{n-1}, a_n)\}$ and vertices on right corresponding to $\{(b_{\pi(2i-1)}, b_{\pi(2i)}) : i \in [n/2]\}$. Finally, add an edge in G between vertices (a_i, a_{i+1}) and $(b_{\pi(j)}, b_{\pi(j+1)})$ if and only if $\{i, i+1\} \cap \{\pi(j), \pi(j+1)\} \neq \emptyset$.

Observe that G is a 2-regular graph and hence by Hall's theorem there exists perfect matching M in G. For each $i \in [n/2]$, let M connect vertex $(a_{2i-1}, a_{2i}) \in L$ to a vertex $(b_j, b_{j'}) \in R$ so that index $r_i \in \{2i-1, 2i\} \cap \{j, j'\}$. Let $S = \{r_i : i \in [n/2]\}$. We claim that S satisfies the required properties. Note that

$$A_o = \sum_{i \in [n/2]} a_{2i-1} \leq \sum_{i \in [n/2]} a_{r_i} \leq \sum_{i \in [n/2]} a_{2i} = A_e.$$

Further, $A_e - A_o \leq a_n - a_1$. Thus,

$$\frac{A - (a_n - a_1)}{2} \leq A_o \leq \sum_i a_{r_i} \leq A_e \leq \frac{A + (a_n - a_1)}{2}.$$

The lemma now follows by a similar argument applied to b_{r_i} for $i \in [n/2]$. □

We also use the following elegant lemma of Barbour and Xia [BX99] which they show using an elementary coupling argument. Intuitively, the lemma says that shift-invariance *amplifies* when taking sums of independent shift-invariant variables.

Lemma 4.4 (Barbour and Xia, Proposition 4.6)**.** *Let $Z_1, \ldots, Z_n$ be integer valued random variables, $Z = \sum_i Z_i$ and $U_Z = \sum_i (1 - \mathsf{d_{TV}}(Z_i, Z_i + 1))$. Then $\mathsf{d_{TV}}(Z, Z+1) \leq 2/\sqrt{U_Z}$.*

Proof of Theorem 4.2. Let $\nu = \max_i(\mathrm{Var}(X_i), \mathrm{Var}(Y_i))$. Let $U_X = \sum_i (1 - \mathsf{d_{TV}}(X_i, X_i + 1))$ and let U_Y be defined similarly. Now, by Lemma 4.3 applied to $Var(X_1), \ldots, Var(X_n)$ and $(1 - \mathsf{d_{TV}}(X_1, X_1 + 1)), \ldots, (1 - \mathsf{d_{TV}}(X_n, X_n + 1))$, there exists a subset $S \subseteq [n]$ such that

$$|\sum_{i \in S} Var(X_i) - \frac{\sigma^2}{2}| \leq \frac{\nu}{2}, \quad |\sum_{i \in S} (1 - \mathsf{d_{TV}}(X_i, X_i + 1)) - \frac{U_X}{2}| \leq \frac{1}{2}.$$

Similarly, there exists a subset $T \subseteq [n]$ such that

$$|\sum_{i \in T} Var(Y_i) - \frac{\sigma^2}{2}| \leq \frac{\nu}{2}, \quad |\sum_{i \in T}(1 - \mathsf{d}_{\mathsf{TV}}(Y_i, Y_i + 1)) - \frac{U_Y}{2}| \leq \frac{1}{2}.$$

Let $X_S = \sum_{i \in S} X_i$, $X'_S = \sum_{i \notin S} X_i$ and let Y_T, Y'_T be defined similarly. Without loss of generality suppose that $\mathbb{E}[X_S] = \mathbb{E}[Y_T] = \mathbb{E}[X'_S] = \mathbb{E}[Y'_T] = 0$ (if not, we can translate the variables accordingly). Then, by the above equations and Lemma 4.4 it follows that X_S, X'_S, Y_T, Y'_T are α-shift invariant for $\alpha = 4/\sqrt{U}$.

Let $\delta = \rho/(\sigma^2 - \nu)^{3/2}$. Now, by an argument similar to that of Equation 4.1 and the Berry-Esséen theorem,

$$\begin{aligned}
\mathsf{d}_{\mathsf{cdf}}(X_S, Y_T - \mathbb{E}[Y_T]) &\leq \mathsf{d}_{\mathsf{cdf}}(X_S, \mathcal{N}(0, Var(X_S))) + \mathsf{d}_{\mathsf{cdf}}(Y_T, \mathcal{N}(0, Var(Y_T))) \\
&\quad + \mathsf{d}_{\mathsf{cdf}}(\mathcal{N}(0, Var(X_S)), \mathcal{N}(0, Var(Y_T))) \\
&\leq \frac{2\rho}{(\sigma^2 - \nu)^{3/2}} + \frac{2\rho}{(\sigma^2 - \nu)^{3/2}} + O\left(\frac{\sqrt{\log \sigma}}{\sigma}\right). \\
&\leq 4\delta + O\left(\frac{\sqrt{\log \sigma}}{\sigma}\right).
\end{aligned}$$

Now, by the Berry-Esséen theorem, for $B = O(\sigma\sqrt{\log(\sigma)})$,

$$\mathsf{Pr}[|X_S - \mathbb{E}[X_S]| > B] \leq 2\delta + 1/\sigma, \quad \mathsf{Pr}[|Y_T - \mathbb{E}[Y_T| > B] \leq 2\delta + 1/\sigma.$$

Further, similar inequalities hold for X'_S, Y'_T as well. Therefore, by Corollary 3.1, and the above inequalities,

$$\begin{aligned}
\mathsf{d}_{\mathsf{TV}}(X_S + X'_S, Y_T + Y'_T) &\leq \mathsf{d}_{\mathsf{TV}}(X_S + X'_S, X_S + Y'_T) + \mathsf{d}_{\mathsf{TV}}(X_S + Y'_T, Y_T + Y'_T) \\
&\leq 4\sqrt{\alpha B \mathsf{d}_{\mathsf{cdf}}(X'_S, Y'_T)} + 4\sqrt{\alpha B \mathsf{d}_{\mathsf{cdf}}(X_S, Y_T)} + O(\delta) + O(1/\sigma) \\
&= O\left(\frac{\sigma \log(1/\sigma)\rho}{(\sigma^2 - \nu)^{3/2} U^{1/2}}\right)^{1/2} + O(\delta) + O(1/\sigma).
\end{aligned}$$

The theorem now follows as $\rho \leq \sigma^2/2$. □

5 PRGs for Combinatorial Shapes

We use the following extension of the main generator for fooling halfspaces over $\{0,1\}^n$ of Meka and Zuckerman [MZ10]. Fix $k, t > 0$ and let $d = n/t$. Let $\mathcal{H} = \{h : [n] \to [t]\}$ be a pairwise independent family of hash functions. Let $G_k : \{0,1\}^{r_k} \to [m]^d$ generate a k-wise independent space over $[m]^d$. Efficient constructions of $\mathcal{H}$ with $|\mathcal{H}| = \text{poly}(n)$ and G_k with $r_k = O(k(\log m + \log d))$ are known. The generator $G_{\mathcal{H},k,t} : \mathcal{H} \times (\{0,1\}^{r_k})^t \to [m]^n$ is defined as follows:

$$G_{\mathcal{H},k,t}(h, z^1, \ldots, z^t) = x, \text{ where } x_{h^{-1}(i)} = G_k(z^i) \text{ for } i = 1, \ldots, t. \tag{5.1}$$

As sketched in the introduction we work with fooling combinatorial sums in statistical distance and first study the case of combinatorial sums with small variance.

Definition 5.1. *A generator $G : \{0,1\}^r \to [m]^n$ ε-fools $\mathsf{CSum}(m,n)$ in statistical distance if for any $f \in \mathsf{CSum}(m,n)$, the random variables $X = f(G(x)), x \in_u \{0,1\}^r$ and $Y = f(y), y \in_u [m]^n$ satisfy $\mathsf{d}_{\mathsf{TV}}(X,Y) \leq \varepsilon$. Similarly, we say that G ε-fools $\mathsf{CSum}(m,n)$ in Kolmogorov (cdf) distance if X and Y satisfy $\mathsf{d}_{\mathsf{cdf}}(X,Y) \leq \varepsilon$.*

We first set up some notation to be used henceforth. Let $f : [m]^n \to [n]$ be an (m, n)-combinatorial sum with $f(x) = \sum_{i=1}^{n} 1_{A_i}(x_i)$ for $A_i \subseteq [m]$. For $x_i \in_u [m]$, define the indicator variable $X_i = 1_{A_i}(x_i)$. Let

$$p_i = \mathbb{E}[X_i],\ \sigma_i^2 = \mathrm{Var}[X_i] = p_i(1 - p_i),\ \mu = \sum_{i=1}^{n} p_i,\ \sigma^2 = \sum_{i=1}^{n} \sigma_i^2$$

Let $X = \sum_{i=1}^{n} X_i$, so $\mathbb{E}[X] = \mu$ and $\sigma^2(X) = \sigma^2$ provided the X_i's are pairwise independent.

5.1 Fooling Small Combinatorial Sums

We now study the case of combinatorial sums with small variance. The strategy is as follows: since $\mathrm{Var}[f]$ is small, there is a small set $L \subseteq [n]$ of *large* variance variables, such that all other indicator random variables $X_i = 1_{A_i}(x_i)$, $i \notin L$, have small variance. To handle variables in L, we argue that they will each be hashed into a different bucket. Thus the distribution on these variables is truly uniform, and moreover, conditioned on their values, the distribution of the output of the generator in each bucket is $(k - 1)$-wise independent. We then use the fact that the combinatorial sum restricted to each bucket has very small total variance and show that bounded independence fools the sum restricted to a bucket in statistical distance. Finally we take a union bound across the different buckets to show the desired claim. As mentioned in the introduction, we use the "sandwiching polynomials" technique introduced by Bazzi to show a sufficiently strong bound for fooling the sum in each bucket so as to apply a union bound.

Theorem 5.2 (Fooling Small Combinatorial Sums)**.** *Let $f \in \mathsf{CSum}(m, n)$ with $\mathrm{Var}[f] \le 6/\varepsilon^2$. For $k = 35$ and $t = C/\varepsilon^{15}$, the generator $G_{\mathcal{H},k,t}$ $O(\varepsilon)$-fools f in statistical distance.*

Fix a $f \in \mathsf{CSum}(m, n)$ with $\sigma^2 \le 6/\varepsilon^2$ and let k, t be as above. Let $L = \{i : \sigma_i^2 \ge \varepsilon^5\}$. Since $\sigma^2 = \sum_i \sigma_i^2 \le 6/\varepsilon^2$, we have $|L| \le 6/\varepsilon^7$. For each bucket B_j we define the variable $T_j = \sum_{i \in B_j \setminus L} \sigma_i^2$. We say a hash function $h \in \mathcal{H}$ is *good* if the following conditions hold:

1. All variables in L are mapped to distinct buckets.

2. For every bucket B_j, $T_j \le \varepsilon$.

Lemma 5.3. *A random hash function $h \in_u \mathcal{H}$ is good with probability at least* $1 - 2\varepsilon$.

Proof. By the pairwise independence of $\mathcal{H}$, each pair of variables $i \neq j \in L$ maps to the same bucket with probability $\frac{1}{t}$. By a union bound, the probability that condition (1) fails is at most $|L|^2/2t \le \varepsilon$.

Fix $j \in [t]$ and for $i \in L^c$, let I_i be the indicator of the event $h(i) = j$. Then $T_j = \sum_{i \in L^c} \sigma_i^2 I_i$,

$$\mathbb{E}[T_j^2] = \mathbb{E}[\,(\sum_{i \in L^c} \sigma_i^2 I_i)^2\,] \ \le \sum_{i \in L^c} \frac{\sigma_i^4}{t} + \sum_{i \neq l \in L^c} \frac{\sigma_i^2 \sigma_l^2}{t^2}$$

$$\le (\max_{i \in L^c} \sigma_i^2) \sum_{i \in L^c} \frac{\sigma_i^2}{t} + \frac{1}{t^2}\left(\sum_{i \in L^c} \sigma_i^2\right)^2 \ \le \frac{\varepsilon^5 \sigma^2}{t} + \frac{\sigma^4}{t^2} \le \frac{12\varepsilon^3}{t}.$$

Therefore, by Markov's inequality

$$\mathsf{Pr}[T_j > \varepsilon] < \frac{\mathbb{E}[T_j^2]}{\varepsilon^2} \le \frac{\varepsilon}{t}$$

By a union bound, $T_j \le \varepsilon$ holds for all $j \in [t]$ except with probability ε.

Thus overall h is good with probability $1 - 2\varepsilon$. □

The above lemma essentially reduces us to the case where all the indicator random variables in each bucket have very small variance, and thus have bias very close to 0 or 1. The following lemma lets us handle such variables.

Lemma 5.4. *Let $X = \sum_{i=1}^n X_i$ and $Y = \sum_{j=1}^n Y_j$ be sums of independent indicator random variables such that $\mathbb{E}[X], \mathbb{E}[Y] \le \varepsilon$. Let D be a $(2d+2)$-wise independent distribution over $\{0,1\}^{2n}$ with the same coordinate-wise marginals as $(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$. Then, for $(X'_1, \ldots, X'_n, Y'_1, \ldots, Y'_n) \leftarrow D$, $(\sum_i X'_i, \sum_i Y'_i)$ is $O_d(\varepsilon^d)$-close in statistical distance to (X, Y).*

We note that a bound of $O(\varepsilon)$ is trivial for the lemma above: each of X and Y are non-zero with probability at most ε under a pairwise independent distribution. However we need a stronger $O(\varepsilon^d)$ bound so that we can use the union bound over all buckets, and this requires more work. We first prove Theorem 5.2 assuming the above lemma.

Proof of Theorem 5.2. Let $x \in [m]^n$ be the string generated by $G_{\mathcal{H},k,t}$ and let $y \in_u [m]^n$. Let $X_i = 1_{A_i}(x_i)$ and $Y_i = 1_{A_i}(y_i)$ be the indicator variables on each co-ordinate. Assume that the hash function h is good in the sense of Lemma 5.3. Then, each variable in L is mapped to a distinct bucket, so the values of $\{x_i\}_{i \in L}$ are uniform and independent. By coupling the variables x_i and y_i for $i \in L$, it suffices to show that $\sum_{i \in L^c} X_i$ and $\sum_{i \in L^c} Y_i$ are close in statistical distance when the distribution within each bucket B_j is $(k-1)$-wise independent, and the buckets are independent. To simplify our notation, we henceforth assume that $L = \varphi$ and $L^c = [n]$.

Fix a bucket B_j. We can partition B_j into $B_j^0 = \{i \in B_j : p_i < \frac{1}{2}\}$ and $B_j^1 = \{i \in B_j : p_i \ge \frac{1}{2}\}$. Let $\bar{X}_i = 1 - X_i$ for $i \in B_j^1$, so that $\Pr[\bar{X}_i = 1] = 1 - p_i$. Define variables $Z_j = \sum_{i \in B_j^0} X_i$ and $Z'_j = \sum_{i \in B_j^1} \bar{X}_i$.

$$\sum_{i \in B_j} X_i = \sum_{i \in B_j^0} X_i + \sum_{i \in B_j^1} (1 - \bar{X}_i) = Z_j - Z'_j + |B_j^1|.$$

Now, since h is good, $T_j \le \varepsilon$, and $\mathbb{E}[Z_j], \mathbb{E}[Z'_j] \le 2\varepsilon$. Since the distribution in each bucket is $k-1 \ge 34$-wise independent, we can apply Lemma 5.4 to the collections $\{X_i : i \in B_j^0\}$, $\{1 - X_i : i \in B_j^1\}$ with $d = 16$ to conclude that (Z_j, Z'_j) is $O(\varepsilon^{16})$-close in statistical distance to the distribution when the variables $X_i \in B_j$ are truly independent.

This implies that $\sum_{i \in B_j} X_i$ is $O(\varepsilon^{16})$ close in statistical distance to $\sum_{i \in B_j} Y_i$. Since variables across buckets are independent of one another, we conclude by a union bound that $\sum_{i \in [n]} X_i = \sum_{j \in [t]} \sum_{i \in B_j} X_i$ is $O(t\varepsilon^{16}) = O(\varepsilon)$ close in statistical distance to $\sum_{i \in [n]} Y_i$. □

5.1.1 Proof of Lemma 5.4

We start with a simple concentration bound for k-wise independent variables.

Lemma 5.5. *Let $X_1, \ldots, X_n$ be k-wise independent $\{0,1\}$ variables such that $\sum_{i=1}^n \mathbb{E}(X_i) \le \beta$. Then for all $\ell \ge k$,*

$$\Pr[\sum_{i=1}^n X_i \ge \ell] \le \left(\frac{e\beta}{\ell}\right)^k.$$

Proof. Let $S_k(X_1, \ldots, X_n) = \sum_{J \subseteq [n]; |J| = k} \prod_{j \in J} X_j$. By the k-wise independence of $X_1, \ldots, X_n$,

$$\mathbb{E}[S_k(X_1, \ldots, X_n)] = \sum_{J \subseteq [n]; |J| = k} \prod_{j \in n} \mathbb{E}[X_j].$$

But since $\sum_i \mathbb{E}[X_i] \le \beta$, it follows that

$$\mathbb{E}[S_k(X_1,\ldots,X_n)] \le \binom{n}{k}\frac{\beta^k}{n^k}.$$

This can be proved by the power-mean inequality, or a weight-shifting argument.

Note that if $\sum_i X_i \ge \ell$, then $S_k(X_1,\ldots,X_n) \ge \binom{\ell}{k}$. Hence by Markov's inequality,

$$\Pr[\sum_i X_i \ge \ell] \le \frac{\mathbb{E}[S_k(X_1,\ldots,X_n)]}{\binom{\ell}{k}} \le \frac{\binom{n}{k}\beta^k}{n^k\binom{\ell}{k}} \le \left(\frac{e\beta}{\ell}\right)^k.$$

□

The following easy corollary follows by taking $k=\ell$:

Corollary 5.6. *If indicator random variables $X_1,\ldots,X_n$ are (fully) independent with $\sum_i \mathbb{E}[X_i] \le \varepsilon$, then for $\ell \in [n]$,*

$$\Pr[\sum_{i=1}^n X_i \ge \ell] \le \left(\frac{e\varepsilon}{\ell}\right)^\ell.$$

Let $X = \sum_i X_i$. Let $I_r(X)$ be the indicator random variable for the the event $X = r$. Let U denote the distribution where each X_i is drawn independently with $\mathbb{E}[X_i] = p_i$. We show that there exist constant degree *sandwiching polynomials* for $I_r(X)$.

Lemma 5.7. *Let $\mathbb{E}[X] \le \varepsilon$. For $d \ge 2$ and every $r \le d$, there exist univariate polynomials $P_r, Q_r : \mathbb{Z} \to \mathbb{Z}$ with $\deg(P_r), deg(Q_r) \le d+1$ such that $P_r(i) \le I_r(i) \le Q_r(i)$, for all $i \in \mathbb{Z}_+$, and*

$$\mathop{\mathbb{E}}_U[Q_r(X) - P_r(X)] = O(\varepsilon^d).$$

Proof. Assume that $d-r$ is even. Let

$$Q_r(x) = \frac{1}{r!(d-r)!}\prod_{i\in\{0,\ldots,d\}\setminus\{r\}}(x-i), \quad P_r(x) = Q_r(x)\cdot\frac{d+1-x}{d+1-r}.$$

Clearly $P_r(\ell) = I_r(\ell) = Q_r(\ell) = 0$ for $\ell \in \{0,\ldots,d\}\setminus\{r\}$. Further, since $d-r$ is even, we have

$$P_r(r) = Q_r(r) = \frac{1}{r!(d-r)!}\prod_{i\in\{0,\ldots,d\}\setminus\{r\}}(r-i) = (-1)^{d-r} = 1.$$

Thus $P_r(\ell) = I_r(\ell) = Q_r(\ell)$ for $\ell \in \{0,\ldots,d\}$. For $\ell \ge d+1$ we have $I_r(\ell) = 0$ whereas

$$\frac{-\ell^d}{r!(d-r)!} \le P_r(\ell) \le 0, \quad \le 0 \le Q_r(\ell) \le \frac{\ell^d}{r!(d-r)!}.$$

Hence $P_r(\ell) \le I_r(\ell) \le Q_r(\ell)$ as claimed. Further, using Corollary 5.6

$$\begin{aligned}\mathop{\mathbb{E}}_U[Q_r(X) - P_r(X)] &\le \sum_{\ell\ge d+1}(Q_r(\ell) - P_r(\ell))\Pr[X \ge \ell] \le \sum_{\ell \ge d+1}\frac{2\ell^d}{r!(d-r)!}\left(\frac{e\varepsilon}{\ell}\right)^\ell \\ &\le \frac{2(e\varepsilon)^{d+1}}{r!(d-r)!}\sum_{\ell\ge d+1}\frac{1}{\ell^{\ell-d}} \\ &= O(\varepsilon^{d+1}).\end{aligned}$$

In the case when $d-r$ is odd, it holds that $r \le d-1$ and $d-1-r$ is even. So we repeat the above argument with d replaced by $d-1$ to get an error bound of $O(\varepsilon^d)$. □

Next we consider the setting where we have two sets $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ of $\{0,1\}$ variables. Let $X = \sum_i X_i$ and $Y = \sum_j Y_j$ and $\mathbb{E}[X], \mathbb{E}[Y] \leq \varepsilon$. Let U^2 denote the distribution where all $2n$ variables are independent.

Corollary 5.8. *For any $d \geq 2$ and $r, s \in \{0, \ldots, d\}$ there are polynomials $P_{r,s}(X,Y)$ and $Q_{r,s}(X,Y)$ where $\deg(P_{r,s}), \deg(Q_{r,s}) \leq 2d+2$, $P_{r,s}(X,Y) \leq I_r(X)I_s(Y) \leq Q_{r,s}(X,Y)$ and*

$$\mathbb{E}_{U^2}[Q_{r,s}(X,Y) - P_{r,s}(X,Y)] = O(\varepsilon^d).$$

Proof. Let $P_{r,s}(X,Y) = P_r(X)Q_s(Y)$ and $Q_{r,s}(X,Y) = Q_r(X)Q_s(Y)$. Then, it follows from the calculations of the previous lemma that $P_{r,s}(X,Y) \leq I_r(X)I_s(Y) \leq Q_{r,s}(X,Y)$. Further,

$$\mathbb{E}[Q_{r,s}(X,Y) - P_{r,s}(X,Y)] = \mathbb{E}_U[Q_r(X) - P_r(X)] \cdot \mathbb{E}_U[Q_s(Y)] \leq O(\varepsilon^d(1+\varepsilon^d)) = O(\varepsilon^d),$$

where we used Lemma 5.7 to bound the error between $P_r(X), Q_r(X)$ and also to bound $\mathbb{E}[Q_s(Y)$ using

$$\mathbb{E}_U[Q_s(Y)] \leq \mathbb{E}_U[I_s(Y)] + \mathbb{E}_U[Q_s(Y) - I_s(Y)] \leq 1 + O(\varepsilon^d).$$

□

We now show that $(2d+2)$-wise independence on $(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ suffices to fool (X,Y) in statistical distance. To do this we shall use the following observation due to Bazzi.

Lemma 5.9. *Let $f, g, h : V \to \{0,1\}$ be functions on a universe V such that $f \leq g \leq h$. Further, let D, D' be two distributions on V such that $\mathbb{E}_{u \leftarrow D}[h(u) - f(u)] \leq \epsilon$ and*

$$|\mathbb{E}_{v \leftarrow D'}[f(v)] - \mathbb{E}_{u \leftarrow D}[f(u)]| \leq \delta, \quad |\mathbb{E}_{v \leftarrow D'}[h(v)] - \mathbb{E}_{u \leftarrow D}[h(u)]| \leq \delta.$$

Then,

$$|\mathbb{E}_{v \leftarrow D'}[g(v)] - \mathbb{E}_{v \leftarrow D}[g(v)]| \leq \varepsilon + \delta.$$

Proof. Let $u \leftarrow D, v \leftarrow D'$. Then,

$$\mathbb{E}[g(v)] \leq \mathbb{E}[h(v)] \leq \mathbb{E}[h(u)] + \delta \leq \mathbb{E}[f(u)] + \epsilon + \delta \leq \mathbb{E}[g(u)] + \epsilon + \delta.$$

A similar chain starting with f instead of h shows the lower bound and the lemma. □

Proof of Lemma 5.4. Let $X' = \sum_i X'_i, Y' = \sum_i Y'_i$. Fix $r, s \in \{0, 1 \ldots, d\}$. Then, as $(2d+2)$-wise independence fools degree $(2d+2)$ polynomials, by Corollary 5.8 and Lemma 5.9 we get that

$$|\Pr[(X,Y) = (r,s)] - \Pr[(X',Y') = (r,s)]| = |\mathbb{E}[I_r(X)I_s(Y) = 1] - \mathbb{E}[I_r(X')I_s(Y')] = 1| = O(\varepsilon^d).$$

Further, by Lemma 5.5, $\Pr[X' \geq d+1 \vee Y' \geq d+1] \leq O(\varepsilon^{d+1})$. Therefore,

$$\begin{aligned} \mathsf{d_{TV}}((X,Y),(X',Y')) &= \sum_{0 \leq r,s \leq n} |\Pr[(X,Y) = (r,s)] - \Pr[(X',Y') = (r,s)]| \\ &\leq \sum_{0 \leq r,s \leq d} |\Pr[(X,Y) = (r,s)] - \Pr[(X',Y') = (r,s)]| + \\ &\qquad \Pr[X \geq d+1 \vee Y \geq d+1] + \Pr[X' \geq d+1 \vee Y' \geq d+1] \\ &\leq d^2 O(\varepsilon^d) + O(\varepsilon^d) = O(d^2\varepsilon^d). \end{aligned}$$

□

5.2 Fooling Large Combinatorial Sums in Kolmogorov Distance

We next show that the generator $G_{\mathcal{H},k,t}$ fools combinatorial sums in Kolmogorov distance when the variance σ^2 of the sum is large.

Theorem 5.10 (Fooling Large Combinatorial Sums). *Let $f \in \mathsf{CSum}(m,n)$ with* $\mathrm{Var}[f] \geq 1/\varepsilon^2$. *Then for $k \geq 4$ and $t \geq 1/\varepsilon^2$, the generator $G_{\mathcal{H},k,t}$ $O(\varepsilon)$-fools f in Kolmogorov distance.*

We use the following property of pairwise independent hash functions. For a hash function $h \in_u \mathcal{H}$, Let $B_j = \{i : h(i) = j\}$ denote the j^{th} bucket of variables. Let $P_j = \sum_{i \in B_j} p_i$ and $S_j = \sum_{i \in B_j} \sigma_i^2$. Finally, let $S_h = (\sum_{j=1}^t S_j^2)^{\frac{1}{2}}$.

Lemma 5.11. *We have $\mathbb{E}_h[S_h] \leq \sigma + \sigma^2/\sqrt{t}$.*

Proof of Lemma 5.11. Fix $j \in [t]$. For each $i \in [n]$, let I_i be the indicator of the event $h(i) = j$ where $h \in_R \mathcal{H}$. Then, $\mathbb{E}_h[I_i] = 1/t$ and for $l \neq i$, $\mathbb{E}_h[I_i I_l] = 1/t^2$ by pairwise independence. As $S_j = \sum_{i=1}^n I_i \sigma_i^2$,

$$\begin{aligned}\mathop{\mathbb{E}}_h[S_j^2] &= \sum_{i=1}^n \sigma_i^4 \mathop{\mathbb{E}}_h[I_i] + 2\sum_{i \neq j} \sigma_i^2 \sigma_j^2 \,\mathbb{E}[I_i I_j] \\ &\leq \frac{1}{t}\sum_{i=1}^n \sigma_i^2 + \frac{2}{t^2}\sum_{i\neq j}\sigma_i^2\sigma_j^2 \qquad \text{since } \sigma_i^4 \leq \sigma_i^2 \\ &\leq \frac{\sigma^2}{t} + \frac{\sigma^4}{t^2}.\end{aligned}$$

Since $S_h^2 = \sum_{j=1}^t S_j^2$, using linearity of expectation we get

$$\mathop{\mathbb{E}}_h[S_h^2] \leq \sum_{j=1}^t \mathop{\mathbb{E}}_h[S_j^2] \leq \sigma^2 + \frac{\sigma^4}{t}.$$

The claim now follows using $\mathbb{E}_h[S_h] \leq \sqrt{\mathbb{E}_h[S_h^2]}$. □

Proof of Theorem 5.10. Let random variable $Y = f(y)$ for $y \in_u [m]^n$. Then, Y has a multinomial distribution with variance $\sigma^2 = \sum_i p_i(1-p_i) > 1/\varepsilon^2$. Therefore, by Corollary 2.2,

$$\mathsf{d}_{\mathsf{cdf}}\left(\frac{Y-\mu}{\sigma}, \mathcal{N}(0,1)\right) \leq \frac{1}{\sigma} = \varepsilon. \tag{5.2}$$

Let $x \in [m]^n$ be generated according to the generator $G_{\mathcal{H},k,t}$ with parameters as in the theorem and let indicator random variables $X_i = 1_{A_i}(x_i)$ and let $X = \sum_i X_i$. We shall show that $(X-\mu)/\sigma$ is also close to $\mathcal{N}(0,1)$. Fix a hash function $h \in \mathcal{H}$. Let $Z_j = \sum_{i \in B(j)} X_i$. Since the X_is are 4-wise independent, $\mathbb{E}[Z_j] = P_j$, $\mathrm{Var}[Z_j] = \sum_{i \in B_j} \sigma_i^2 = S_j$. Further, we have

$$\begin{aligned}\mathbb{E}[(Z_j - P_j)^4] &= \mathbb{E}[(\sum_{i \in B_j}(X_i - p_i))^4] \\ &= \sum_{i \in B_j} \mathbb{E}[(X_i - p_i)^4] + 3\sum_{i \neq l \in B_j} \mathbb{E}[(X_i - p_i)^2]\,\mathbb{E}[(X_l - p_l)^2] \\ &\leq \sum_{i \in B_j} \sigma_i^2 + 3\sum_{i \neq l \in B_j} \sigma_i^2 \sigma_l^2 \qquad \text{since} (X_i - p_i)^4 \leq (X_i - p_i)^2 \\ &= S_j + 3S_j^2.\end{aligned}$$

Therefore, summing over all j we get

$$\sum_{j=1}^{t} \mathbb{E}[(Z_j - P_j)^4] \leq \sum_{j=1}^{t} S_j + 3\sum_{j=1}^{t} S_j^2 = \sigma^2 + 3S_h^2.$$

Using the Berry-Esséen theorem applied to independent random variables $Z_1, \dots, Z_t$, for a fixed hash function h,

$$\mathsf{d_{cdf}}\left(\frac{X-\mu}{\sigma}, \mathcal{N}(0,1)\right) \leq \frac{(\sigma^2 + 3S_h^2)^{1/2}}{\sigma^2} \leq 2\left(\frac{1}{\sigma} + \frac{S_h}{\sigma^2}\right).$$

Further, as $\mathsf{d_{cdf}}$ is a convex function, using Lemma 5.11,

$$\mathsf{d_{cdf}}\left(\frac{X-\mu}{\sigma}, \mathcal{N}(0,1)\right) \leq 2\left(\frac{1}{\sigma} + \frac{\mathbb{E}_h[S_h]}{\sigma^2}\right) \leq 2\left(\frac{2}{\sigma} + \frac{1}{\sqrt{t}}\right) \leq 6\varepsilon.$$

By Equation (5.2) we get $\mathsf{d_{cdf}}((X-\mu)/\sigma, (Y-\mu)/\sigma) = O(\varepsilon)$ which implies $\mathsf{d_{cdf}}(X,Y) = O(\varepsilon)$. □

5.3 Reducing the seed-length via INW

We now derandomize $G_{\mathcal{H},k,t}$ using PRGs for small space sources of Impagliazzo, Nisan, and Wigderson [INW94], which we call the INW PRG. The derandomization follows from Theorems 5.2, 5.10 and replacing the independent seeds $z^1, \dots, z^t$ in Equation 5.1 with the output of the INW PRG.

Theorem 5.12 (Derandomizing $G_{\mathcal{H},k,t}$)**.** *There exists a generator $G \equiv G_{m,n,\varepsilon} : \{0,1\}^{r_{m,n}} \to [m]^n$ with seed-length $r_{m,n} = O(\log m + \log n + \log^2(1/\varepsilon))$ with the following properties:*

1. *G $O(\varepsilon)$-fools all $f \in \mathsf{CSum}(m,n)$ with $\mathrm{Var}[f] < 6/\varepsilon^2$ in statistical distance.*
2. *G $O(\varepsilon)$-fools all $f \in \mathsf{CSum}(m,n)$ with $\mathrm{Var}[f] > 1/\varepsilon^2$ in Kolmogorov distance.*

Consider $G_{\mathcal{H},k,t}$ with parameters set so as to satisfy the conditions of Theorems 5.2, 5.10. Note that the seed length of $G_{\mathcal{H},k,t}$ is $O((\log n)\mathrm{poly}(1/\varepsilon))$. We will reduce the seed length by choosing the seeds $z^1, \dots, z^t$ from the output of the INW PRG (instead of independently as before). The analysis proceeds roughly by arguing that for any (m,n)-combinatorial sum f and hash function $h \in \mathcal{H}$, $f(G_{\mathcal{H},k,t}(h, z^1, \dots, z^t)) \equiv g_h(z^1, \dots, z^t)$ is computable by a small-space machine when viewed as a function of $z^1, \dots, z^t$.

Let $\mathsf{INW} : \{0,1\}^r \to (\{0,1\}^{r_k})^t$ be the INW generator that ε-fools $(10\log(1/\varepsilon), r_k, t)$, read-once branching programs. Define

$$G : \mathcal{H} \times \{0,1\}^r \to [m]^n \text{ by } G(h,y) = G_{\mathcal{H},k,t}(h, \mathsf{INW}(y)).$$

We claim that G satisfies the conditions of Theorem 5.12.

Proof of Theorem 5.12. The claim on the seed length of G follows from the seed length of the INW generator which uses $r = O(r_k + (\log(1/\varepsilon) + \log(t/\varepsilon))\log t) = O(\log m + \log n + \log^2(1/\varepsilon))$ bits (see the discussion in Section 2). We next show that G satisfies properties (1), (2).

Fix an (m,n)-combinatorial sum f and let x be the output of generator $G_{\mathcal{H},k,t}$ with parameters as above. Fix a hash function $h \in \mathcal{H}$ and define $g_h : (\{0,1\}^{r_k})^t \to [n]$ by $g_h(z^1, \dots, z^t) = f(G_{\mathcal{H},k,t}(h, z^1, \dots, z^t))$. For $\ell \in [t]$, let $B_\ell = \{i : h(i) = \ell\}$ and let random variable $Y_\ell = \sum_{j: j \in B_\ell} 1_{A_j}(x_j)$. Then, Y_ℓ depends only on z^ℓ and $g_h(z^1, \dots, z^t) = \sum_\ell Y_\ell$.

There is a natural $(\log n, r_k, t)$-ROBP M for computing g_h: the vertices of M are labeled $\{1, \ldots, n\}$ with states in layer ℓ corresponding to the possible values of the partial sum $\sum_{i\le\ell} Y_i$ and the edges out of layer ℓ are drawn according to the change in the value of the partial sum. However, using M directly to do the derandomization is problematic as G_S only fools $O(\log(1/\varepsilon))$ space ROBPs. We get over this hurdle by appropriately sandwiching M between smaller-width branching programs.

Case 1: $\mathrm{Var}[f] < 6/\varepsilon^2$. Observe that $x_1, \ldots, x_n$ are k-wise independent. Therefore, by an argument similar to that of Lemma 5.5, it follows that for $\ell \in [t]$,

$$\mathsf{Pr}[\,|\sum_{j\le\ell}(Y_j - \mu(Y_j))| > 6e/\varepsilon^4\,] \le \varepsilon^{2k}. \tag{5.3}$$

We exploit this fact by ignoring all states of M corresponding to partial sums not in $I = [-6e/\varepsilon^4, 6e/\varepsilon^4]$.

Fix a statistical test function $F : [n] \to \{0,1\}$. Let $\bar{z} = (z^1, \ldots, z^t) \in_u (\{0,1\}^{r_k})^t$. Observe that $F(\bar{z}) \equiv F(g_h(\bar{z})) = F(M(z))$ is computable by a $(\log n, r_k, t)$-ROBP, say M'. We now sandwich F between two small-width branching programs. Let M_u be a ROBP that works the same as M' except that it accepts all strings $\bar{z}$ that lead to a partial sum $\sum_{i\le\ell}(Y_i - \mu(Y_i)) \notin I$. Similarly, let M_l be a machine a ROBP that works the same as M' except that it rejects all strings $\bar{z}$ that lead to a partial sum $\sum_{i\le\ell}(Y_i - \mu(Y_i)) \notin I$. Then, $M_l \le M' \le M_u$ and M_l, M_u are computable by $((\log|I|) + 1, r_k, t)$-ROBPs. Further, from Equation 5.3 and a union bound over $\ell \in [t]$,

$$\mathsf{Pr}[M_u(z) = 1] - \mathsf{Pr}[M_l(z) = 1] \le t\varepsilon^{2k} = O(\varepsilon).$$

Now, as G_S fools M_u, M_l with error at most ε, it follows from the above equation and the sandwiching property (Lemma 5.9) that G_S fools M' with error at most $O(\varepsilon)$. The theorem now follows from the above fact and Theorem 5.2.

Case 2: $\mathrm{Var}[f] > 1/\varepsilon^2$. This case follows straightforwardly from Theorem 5.10 and the *monotone trick* argument of Meka and Zuckerman [MZ10]. We refer to Section 4.3 of the full version of their paper [MZ09b] for details. □

5.4 Fooling Combinatorial Sums

We now combine the generators from the previous section to get our final generator fooling combinatorial sums in statistical distance. The basic idea is as follows: we partition the n variables into two subsets L, R with $|L| \sim n/2$, and then use $G_{m,n/2}$ for the variables in L and an independent $G_{m,n/2}$ on the variables in R. We analyze the construction by induction and considering two cases. If the variance of the combinatorial sum is small, we invoke Theorem 5.12 (1). So now assume that the variance is large.

Let f be a combinatorial sum with $\mathrm{Var}[f] > 6/\varepsilon^2$ and write $f = f_L + f_R$, where f_L, f_R are the combinatorial sums obtained by restricting to variables in L, R respectively. We use the induction hypothesis to get a statistical distance guarantee for f_L and use Theorem 5.12 (2) to get a Kolmogorov distance guarantee for f_R. We then argue that the combinatorial sum f_L has high variance and hence is shift invariant. We then apply Lemma 1.6 and get a statistical distance guarantee for $f = f_L + f_R$.

Fix $\varepsilon \in [1/\sqrt{n}, 1/\log n]$ and let $s = \log(n+1)$. Let $\mathcal{H}_1 = \{\pi : [n] \to [n]\}$ be a family of pairwise independent permutations. Efficient constructions of $\mathcal{H}_1$ with $\mathcal{H}_1 = \mathrm{poly}(n)$ are known. We pick $\pi \in_u \mathcal{H}_1$ and use it to partition $[n]$ into s buckets of geometrically increasing sizes. We define sets $B_1, \ldots, B_s$ where $B_j = \{\pi(2^{j-1}), \ldots, \pi(2^j - 1)\}$, thus $|B_j| = 2^{j-1}$. Let r_j be the seed-length of the generator $G_{m,2^{j-1},\varepsilon}$ from Theorem 5.12. Our main generator $H_{m,n} : \mathcal{H}_1 \times \{0,1\}^{r_1} \times \cdots \times \{0,1\}^{r_s} \to [m]^n$ uses an independent sample from $G_{m,2^{j-1},\varepsilon}$ for each bucket B_j:

$$H_{m,n}(\pi, z^1, \dots, z^s) = x, \text{ where } x_{B_j} = G_{m,2^{j-1},\varepsilon}(z^j). \tag{5.4}$$

As before, let $f(x_1, \dots, x_n) = \sum_{i=1}^n X_i$ where $X_i = 1_{A_i}(x_i)$ has mean p_i and variance σ_i^2. For each bucket B_j, let $S_j = \sum_{i \in B_j} \sigma_i^2$. Let $q \in \{1, \dots, s\}$ be the least index such that $\mathbb{E}[S_q] > 3/\varepsilon^2$.

Call a permutation π *bad* if one of the following conditions holds and *good* otherwise:

1. There exists an index $j \in \{q, \dots, s\}$ such that $S_j \notin [0.5\,\mathbb{E}[S_j], 1.5\,\mathbb{E}[S_j]]$.
2. There exists $j \in \{1, \dots, q-1\}$ such that $S_j \geq 6/\varepsilon^2$.

Note that the sequence $\{\mathbb{E}[S_j]\}_{j=1}^s$ is in geometric progression. If π is good, then $\{S_j\}_{j=q}^s$ is roughly geometric, and none of $\{S_j\}_{j \leq q}$ are too large.

Claim 5.13. $\Pr_{\pi \in_u \mathcal{H}_1}[\pi \text{ is bad}] \leq 2\epsilon$.

Proof. Fix $j \in \{q, \dots, s\}$. Let Z_i be the indicator of the event $\pi^{-1}(i) \in \{2^{j-1}, \dots, 2^j - 1\}$ and hence $i \in B_j$. Then

$$S_j = \sum_{i=1}^n \sigma_i^2 Z_j \;\Rightarrow\; \mathbb{E}[S_j] = \frac{\sigma^2 2^{j-1}}{n}.$$

By the pairwise-independence of π,

$$\begin{aligned}\mathbb{E}[S_j^2] &= \sum_i \sigma_i^2\,\mathbb{E}[Z_i] + \sum_{i \neq l} 2\sigma_i^2\sigma_l^2\,\mathbb{E}[Z_i Z_l] \leq \frac{\sigma^2 2^{j-1}}{n} + \frac{\sigma^2 2^{j-1}(2^{j-1}-1)}{n(n-1)} \\ &\leq \frac{\sigma^2 2^{j-1}}{n} + \frac{\sigma^4 2^{2(j-1)}}{n^2},\end{aligned}$$

hence, $\mathrm{Var}[S_j] \leq \mathbb{E}[S_j^2] - \mathbb{E}[S_j]^2 \leq \sigma^2 2^{j-1}/n = \mathbb{E}[S_j]$.

We now bound the probability of bad event (1). Fix $j \in \{q, \dots, s\}$ so that $\mathbb{E}[S_j] \geq \frac{3}{\varepsilon^2}$. By Chebychev's inequality

$$\mathsf{Pr}\left[|S_j - \mathbb{E}[S_j]| > \frac{\mathbb{E}[S_j]}{2}\right] \leq \frac{4\,\mathrm{Var}[S_j]}{(\mathbb{E}[S_j]^2)} \leq \frac{4}{\mathbb{E}[S_j]} \leq 2\epsilon^2.$$

Similarly, to bound bad event (2), we observe that $\mathbb{E}[S_j] \leq 3/\varepsilon^2$ for $j \leq q-1$, hence

$$\mathsf{Pr}[S_j \geq 6/\varepsilon^2] \leq \Pr[|S_j - \mathbb{E}[S_j]| > 3/\varepsilon^2] \leq \varepsilon^4\,\mathrm{Var}[S_j]/9 \leq \varepsilon^2.$$

Since $\varepsilon < 1/\log n$, the claim follows by a union bound over $i \in \{1, \dots \log n\}$. □

Theorem 5.14. *The Generator $H_{m,n}$ fools $\mathsf{CSum}(m,n)$ with error $O(\log n\sqrt{\varepsilon\log(1/\varepsilon)})$.*

Proof. Let $x \in [m]^n$ be sampled from $H_{m,n}$, while $y \in_u [m]^n$. Let $X_i = 1_{A_i}(x_i), Y_i = 1_{A_i}(y_i)$ and

$$X^j = \sum_{i \in B_j} X_i, \quad Y^j = \sum_{i \in B_j} Y_i, \qquad X^{\leq j} = \sum_{l \leq j} X^l, \quad Y^{\leq j} = \sum_{l \leq j} Y^l.$$

We assume from now on we condition on the chosen permutation π being good. Observe that $\mathbb{E}[X^j] = \mathbb{E}[Y^j]$ and

$$\mathrm{Var}[X^j] = \mathrm{Var}[Y^j] = \sum_{i \in B_j} \mathrm{Var}[X_i] = \sum_{i \in B_j} \sigma_i^2 = S_j.$$

We claim that there is a constant C such that for $j \in [s]$,

$$\mathsf{d}_{\mathsf{TV}}(X^{\leq j}, Y^{\leq j}) \leq Cj\sqrt{\varepsilon(\log(1/\varepsilon))}. \tag{5.5}$$

The proof is by induction on j. It is easy to prove for $j \leq q$. Since $\mathrm{Var}[X^l] = \mathrm{Var}[Y^l] = S_l < 6/\varepsilon^2$ for all $l \leq j$, by Theorem 5.12 (1), $\mathsf{d}_{\mathsf{TV}}(X^l, Y^l) \leq \varepsilon$. As $X^1, \ldots, X^j$ are independent of one another, we have $\mathsf{d}_{\mathsf{TV}}(X^{\leq j}, Y^{\leq j}) \leq j\varepsilon$. Now consider $j \in \{q+1, \ldots, s\}$. We have

$$\mathsf{d}_{\mathsf{TV}}(X^{\leq j-1} + X^j, Y^{\leq j-1} + Y^j) \leq \mathsf{d}_{\mathsf{TV}}(X^{\leq j-1} + X^j, Y^{\leq j-1} + X^j) + \mathsf{d}_{\mathsf{TV}}(Y^{\leq j-1} + X^j, Y^{\leq j-1} + Y^j). \tag{5.6}$$

The first term can be bounded using the induction hypothesis:

$$\mathsf{d}_{\mathsf{TV}}(X^{\leq j-1} + X^j, Y^{\leq j-1} + X^j) \leq \mathsf{d}_{\mathsf{TV}}(X^{\leq j-1}, Y^{\leq j-1}) \leq C(j-1)\sqrt{\varepsilon(\log(1/\varepsilon))}. \tag{5.7}$$

To bound the second term, we will apply Corollary 3.1. As π is good and $j > q$, $\mathrm{Var}[X^j] = \mathrm{Var}[Y^j] = S_j \geq \mathbb{E}[S_j]/2 > 1/\varepsilon^2$. Thus the variance is sufficiently large to apply Theorem 5.12 (2), which gives $\mathsf{d}_{\mathsf{cdf}}(X^j, Y^j) < \varepsilon$. Moreover, by Fact 2.5,

$$\mathsf{Pr}\left[|Y^j - \mathbb{E}[Y^j]| > 3\sqrt{S_j \log(1/\varepsilon)}\right] \leq \varepsilon.$$

Since X^j and Y^j have the same mean and $\mathsf{d}_{\mathsf{cdf}}(X^j, Y^j) < \varepsilon$, we get similar concentration for X^j:

$$\mathsf{Pr}\left[|X^j - \mathbb{E}[X^j]| > 3\sqrt{S_j \log(1/\varepsilon)}\right] \leq 3\varepsilon.$$

Thus, with probability $1 - 4\varepsilon$, we have $X^j, Y^j \in [\mathbb{E}[X^j] - b, \mathbb{E}[X^j] + b]$, where $b = 3\sqrt{S_j \log(1/\varepsilon)}$. Further, since π is good, we have

$$\mathrm{Var}[Y^{\leq j-1}] \geq \mathrm{Var}[Y^{j-1}] = S_{j-1} > \mathbb{E}[S_{j-1}]/2 \geq \mathbb{E}[S_j]/4 > S_j/6.$$

Hence by Fact 2.4, $Y^{\leq j-1}$ is $\alpha = (6/\sqrt{S_j})$-shift invariant.

We can now apply Corollary 3.1 with $\alpha = 6/\sqrt{S_j}$ and $b = 6\sqrt{S_j \log(1/\varepsilon)}$ to get

$$\mathsf{d}_{\mathsf{TV}}(Y^{\leq j-1} + X^j, Y^{\leq j-1} + Y^j) \leq 24\sqrt{\varepsilon \log(1/\varepsilon)} + 4\varepsilon. \tag{5.8}$$

Substituting the bounds from Equations (5.7) and (5.8) back into Equation (5.6) gives

$$\mathsf{d}_{\mathsf{TV}}(X^{\leq j}, Y^{\leq j}) \leq C(j-1)\sqrt{\varepsilon \log(1/\varepsilon)} + 24\sqrt{\varepsilon \log(1/\varepsilon)} + 4\varepsilon \leq Cj\sqrt{\varepsilon \log(1/\varepsilon)},$$

where $C = 30$. □

We now derandomize the generator of Theorem 5.14 to get our main result for fooling combinatorial shapes.

Proof of Theorem 1.3. We derandomize the generator $H_{m,n}$ of Equation 5.4 as was done in Theorem 5.12 by choosing the seeds $z^1, \ldots, z^s$ from the output of PRGs for ROBPs. Fix $\delta > 0$ and set the parameters of $H_{m,n}$ as in Theorem 5.14 with $\varepsilon = \delta/(\log(1/\delta) \cdot \log n)$. Fix a (m,n)-combinatorial shape f and note that for a hash function $g \in \mathcal{H}_1$, $f(H_{m,n}(g, z^1, \ldots, z^s))$ when viewed as a function of $z^1, \ldots, z^s$ is computable by a (S, D, T)-ROBP, where $S = \log n$, $D = O(\log m + \log n + \log^2(1/\varepsilon))$, and $T = s = O(\log n)$. Further, as $T = O(S + D)$, such ROBPs can be fooled with error ε and seed length $O(\log m + \log n + \log^2(1/\varepsilon))$ by using the PRG of [NZ96].

Let G be the generator obtained from $H_{m,n}$ by using the PRG of [NZ96] with parameters as above to generate the seeds $z^1, \ldots, z^s$ of Equation 5.4 instead of independently as before. Then, by Theorem 5.14, G $O(\delta)$-fools (m,n)-combinatorial sums with seed length $O(\log m + \log n + \log^2(1/\varepsilon)) = O(\log m + \log n + \log^2(1/\delta))$. □

6 PRGs for Combinatorial Rectangles

We prove that the generator $G_{\mathcal{H},k,t}$ of Equation 5.1 with $k = O(\sqrt{\log(1/\varepsilon)})$ and $t = \exp(O(\sqrt{\log n}))$ and $\mathcal{H}$ k-wise independent fools combinatorial rectangles. We then derandomize the generator using the INW generator as in the proofs of Theorems 5.12 and 1.3 to get our final PRG for combinatorial rectangles. As mentioned before, though our result is weaker than Lu's generator, our construction is perhaps simpler than Lu's and our analysis is different from Lu's. Moreover, we match Lu's parameters for the important case when the desired error $\varepsilon = \text{poly}(n)$.

Theorem 6.1. *The generator $G_{\mathcal{H},k,t}$ with $k = 5\sqrt{\log(1/\varepsilon)}$, $t = \exp(5\sqrt{\log(1/\varepsilon)})$ and $\mathcal{H}$ a k-wise independent family of hash functions, fools combinatorial rectangles with error at most $O(\varepsilon)$.*

We use the following properties of a k-wise independent family of hash functions.

Lemma 6.2. *For $\mathcal{H} = \{h : [n] \to [t]\}$, k-wise independent, the following properties hold.*

1. *For any $L \subseteq [n]$, $|L| \leq r$, $\mathsf{Pr}[\, \exists \ell,\ |h^{-1}(\ell) \cap L| \geq k/2\,] \leq t \cdot (2re/kt)^{k/2}$.*
2. *Let $q_1, \ldots, q_n \in [0,1]$, $\sum_i q_i = Q$ and $\max_i q_i \leq \beta Q$. Then, for any $\ell \in [t]$,*

$$\mathsf{Pr}[\sum_{i:h(i)=\ell} q_i \geq Q/t + \beta^{1/4} Q\,] \leq 2(k\beta^{1/2}\log(1/\beta))^{k/2}.$$

Proof. (1). Without loss of generality, let $L = \{1, \ldots, r\}$. Fix $\ell \in [t]$ and let $X_1, \ldots, X_n$ be indicator random variables with $X_i = 1$ if $h(i) = \ell$ and 0 else. Then, $X_1, \ldots, X_r$ are k-wise independent and

$$\mathsf{Pr}[\sum_i X_i \geq k/2\,] \leq \mathbb{E}[\sum_{J \subseteq [r], |J| = k/2} \prod_{j \in J} X_j\,] = \binom{r}{k/2}\frac{1}{t^{k/2}} \leq \left(\frac{2re}{kt}\right)^{k/2}.$$

The claim now follows by taking a union bound over $\ell \in [t]$.

(2). Fix $\ell \in [t]$ and let $X_1, \ldots, X_n$ be as above. Then, $X = \sum_{i:h(i)=\ell} q_i = \sum_i q_i X_i$, where the X_i are k-wise independent with $\mathsf{Pr}[X_i = 1] = 1/t$. Let $Y_1, \ldots, Y_n$ be independent random variables with $\mathsf{Pr}[Y_i = 1] = 1/t$ and $Y = \sum_i q_i Y_i$. Then, by Hoeffding's inequality, for all $\gamma > 0$,

$$\mathsf{Pr}[\,|Y - Q/t| \geq \gamma\,] \leq 2\exp(-2\gamma^2/\sum_i q_i^2) \leq 2\exp(-2\gamma^2/\beta Q^2).$$

Let k be even and fix $\gamma > 0$ to be chosen later. Then, as $Y \leq Q$,

$$\mathbb{E}[\,(Y - Q/t)^k\,] \leq \gamma^k + Q^k\,\mathsf{Pr}[\,|Y - Q/t| \geq \gamma\,] \leq \gamma^k + Q^k 2\exp(-2\gamma^2/\beta Q^2).$$

Since $\mathbb{E}[(X - Q/t)^k] = \mathbb{E}[(Y - Q/t)^k]$, it follows from Markov's inequality that for any $\theta > 0$,

$$\mathsf{Pr}[\,|X - Q/t| > \theta\,] \leq \frac{\gamma^k + Q^k 2\exp(-2\gamma^2/\beta Q^2)}{\theta^k}.$$

Setting $\theta = \beta^{1/4} \cdot Q$, $\gamma = (2k\beta\log(1/\beta))^{1/2}Q$, we get

$$\mathsf{Pr}[\,|X - Q/t| > \beta^{1/4}Q\,] \leq 2(k\beta^{1/2}\log(1/\beta))^{k/2}.$$

□

Proof of Theorem 6.1. Fix an (m,n)-combinatorial rectangle $f:[m]^n \to \{0,1\}$ with $f(x_1,\ldots,x_n) = 1_{A_1}(x_1) \wedge 1_{A_2}(x_2) \cdots 1_{A_n}(x_n)$. Let $y \in_u [m]^n$ and $Y_i = 1_{A_i}(y_i)$, $q_i = 1 - \mathbb{E}[Y_i]$. Let x be the output of the generator with parameters as in the statement. Let $X_i = 1_{A_i}(x_i)$ and $X = \sum_i X_i$. Note that

$$\mathsf{Pr}[f(y)=1] = (1-q_1)(1-q_2)\cdots(1-q)_n \leq \exp(-\sum_i q_i).$$

Therefore, if $\sum_i q_i > \log(1/\epsilon)$, then $\mathsf{Pr}[f(y)=1] < \epsilon$. We accordingly consider two cases to analyze our generator.

Case 1: $Q = \sum_i q_i \leq 3\log(1/\epsilon)$. Let $L = \{i : q_i > Q/\sqrt{t}\}$, $L^c = [n]/L$. Then, $|L| < \sqrt{t}$ and by Lemma 6.2 (1) it follows that for $h \in_u \mathcal{H}$, $\max_\ell |h^{-1}(\ell) \cap L| \leq k/2$ with probability at least $1 - 1/t^{\Omega(k)} = 1 - \epsilon$. Consequently, for a random h we can assume that the variables in L are truly independent of one another. Moreover, when conditioned on the variables in L, the variables from L^c in each bucket, $\{x_i : i \in B_\ell = h^{-1}(\ell), \wedge i \notin L^c\}$ for $\ell \in [t]$, are $(k/2)$-wise independent. To simplify notation we assume that $L = \emptyset$ and analyze the case where the X_i's in a single bucket are $(k/2)$-wise independent.

Now, for $\beta = 1/\sqrt{t}$, $\max_i q_i < \beta Q$. Therefore, by Lemma 6.2 (2), for $h \in_u \mathcal{H}$ with probability at least $1-\epsilon$, $Q^\ell = \sum_{i:h(i)=\ell} q^i < 6\log(1/\epsilon)/t^{1/8}$ for all $\ell \in [t]$. Further, by the principle of inclusion-exclusion and $(k/2)$-wise independence of $X_i, i \in B_\ell$,

$$\begin{aligned}
|\mathsf{Pr}[\wedge_{i\in B_\ell} X_i] - \mathsf{Pr}[\wedge_{i \in B_\ell} Y_i]| &\leq \sum_{J \subseteq B_\ell, |J| = k/2} \mathsf{Pr}[\wedge_{i\in J} X_i] \\
&\leq \binom{|B_\ell|}{k/2} \left(\frac{Q^\ell}{|B_\ell|}\right)^{k/2} \qquad \text{(power-mean inequality)} \\
&\leq \left(\frac{2eQ^\ell}{k}\right)^{k/2} \\
&= \left(\frac{O(\sqrt{\log(1/\epsilon)})}{t^{1/8}}\right)^{k/2} = O(\epsilon/t).
\end{aligned}$$

Therefore, as the X_i's in different buckets are independent of one another, by a union bound over $\ell \in [t]$ it follows that $|\mathsf{Pr}[\wedge_i X_i = 1] - \mathsf{Pr}[\wedge_i Y_i = 1]| = O(\epsilon)$.

Case 2: $\sum_i q_i > 3\log(1/\epsilon)$. Let $j \in [n]$ be the maximum index such that $\sum_i q_i \leq 3\log(1/\epsilon)$. Then, $\sum_{i \leq j} q_i \geq 3\log(1/\epsilon) - 1 > 2\log(1/\epsilon)$. Therefore, $\mathsf{Pr}[\wedge_{i\leq j} Y_i = 1] \leq \exp(-\sum_{i\leq j} q_i) \leq \epsilon$. Now, by applying the argument of the previous case to the collection of variables $X_1,\ldots,X_j$ it follows that $\mathsf{Pr}[\wedge_{i \leq j} X_i = 1] = O(\epsilon)$. Therefore, $\mathsf{Pr}[\wedge_i X_i = 1] = O(\epsilon)$ from which the claim follows. □

Proof of Theorem 1.4. The theorem follows by derandomizing $G_{\mathcal{H},k,t}$ with parameters as above by using the INW PRG to generate $z^1,\ldots,z^t$ of Equation 5.1 instead of independently as before. □

References

[AGHP92] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple constructions of almost k–wise independent random variables. *Random Structures and Algorithms*, 3(3):289–303, 1992.

[ASWZ96] R. Armoni, M. Saks, A. Wigderson, and S. Zhou. Discrepancy sets and pseudorandom generators for combinatorial rectangles. In *FOCS*, pages 412–421, 1996.

[Baz09] Louay M. J. Bazzi. Polylogarithmic independence can fool dnf formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009.

[BC02] A. D. Barbour and V. Cekanavičius. Total variation asymptotics for sums of independent integer random variables. *The Annals of Probability*, 30(2):509–545, 2002.

[BX99] Andrew D. Barbour and Aihua Xia. Poisson perturbations. *ESAIM*, 3:131–150, october 1999.

[DGJ$^+$10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM J. Comput.*, 39(8):3441–3462, 2010.

[DP07] Constantinos Daskalakis and Christos H. Papadimitriou. Computing equilibria in anonymous games. In *FOCS*, 2007.

[DP08] Constantinos Daskalakis and Christos H. Papadimitriou. Discretized multinomial distributions and nash equilibria in anonymous games. In *FOCS*, 2008.

[EGL$^+$92] G. Even, O. Goldreich, M. Luby, N. Nisan, and B. Veličković. Approximations of general independent distributions. In *STOC*, pages 10–16, 1992.

[Fel71] William Feller. *An Introduction to Probability Theory and Its Applications, Vol. 2 (Volume 2).* Wiley, 2 edition, January 1971.

[HHR06] Iftach Haitner, Danny Harnik, and Omer Reingold. On the power of the randomized iterate. In *CRYPTO*, pages 22–40, 2006.

[Ind00] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings and data stream computation. In *41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000)*, pages 189–197. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *STOC*, pages 356–364, 1994.

[KNR05] Eyal Kaplan, Moni Naor, and Omer Reingold. Derandomized constructions of k-wise (almost) independent permutations. In *Proceedings of the 8th International Workshop on Randomization and Computation (RANDOM '05)*, number 3624 in Lecture Notes in Computer Science, pages 354 – 365, Berkeley, CA, August 2005. Springer.

[LRTV09] Shachar Lovett, Omer Reingold, Luca Trevisan, and Salil P. Vadhan. Pseudorandom bit generators that fool modular sums. In *APPROX-RANDOM*, pages 615–630, 2009.

[Lu02] Chi-Jen Lu. Improved pseudorandom generators for combinatorial rectangles. *Combinatorica*, 22:417–434, 2002.

[Mat99] J. Matousek. *Geometric Discrepancy.* Springer, 1999.

[MZ09a] R. Meka and D. Zuckerman. Small-bias spaces over finite groups. In *APPROX-RANDOM*, 2009.

[MZ09b] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions, 2009. arXiv: 0910.4122.

[MZ10] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. In *STOC*, pages 427–436, 2010.

[Nis92] Noam Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992.

[NN93] J. Naor and M. Naor. Small–bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *J. Comput. Syst. Sci.*, 52(1):43–52, 1996.

[RS10] Yuval Rabani and Amir Shpilka. Explicit construction of a small epsilon-net for linear threshold functions. *SIAM J. Comput.*, 39(8):3501–3520, 2010.

[She07] I. G. Shevtsova. Sharpening of the upper bound of the absolute constant in the berry–esseen inequality. *Theory of Probability and its Applications*, 51(3):549–553, 2007.

[Siv02] D. Sivakumar. Algorithmic derandomization via complexity theory. In *STOC '02: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 619–626, New York, NY, USA, 2002. ACM.

[Wat11] Thomas Watson. Pseudorandom generators for combinatorial checkerboards. In *IEEE Conference on Computational Complexity*, 2011. To Appear.

A Missing Proofs from Section 2

Proof of Fact 2.3. We'll use the following anti-concentration property of Gaussians: for $Z \leftarrow \mathcal{N}(0,\sigma)$, $\delta > 0$, $\mathsf{Pr}[Z \in [\theta, \theta+\delta]] = O(\delta/\sigma)$. Suppose that $\mu_2 > \mu_1$. Then,

$$\mathsf{d_{cdf}}(Z_1, \mathcal{N}(\mu_2,\sigma_1)) \le \mathsf{Pr}[Z_1 \in [\mu_1,\mu_2]] = O(|\mu_2-\mu_1|/\sigma_1).$$

Thus, it suffices to study the case when $\mu_1 = \mu_2 = 0$. Let $\sigma_2 > \sigma_1$ and $\lambda = \sqrt{\sigma_2^2 - \sigma_1^2}$. Observe that Z_2 can be generated as $Z_2 = Z_1 + Z'$, where Z' is an independent $\mathcal{N}(0,\lambda)$ random variable. Now, $\mathsf{Pr}[|Z'| > 3\lambda\sqrt{\log \sigma_1}] \le 1/\sigma_1$. Therefore, for any $\theta \in \mathbb{R}$,

$$\begin{aligned}
\mathsf{Pr}[Z_2 < \theta] &= \mathsf{Pr}[Z_1 + Z' < \theta] \\
&\le \mathsf{Pr}[Z_1 < \theta + 3\lambda\sqrt{\log\sigma_1}] + \mathsf{Pr}[|Z'| > 3\lambda\sqrt{\log\sigma_1}] \\
&\le \mathsf{Pr}[Z_1 < \theta] + \mathsf{Pr}[Z_1 \in [\theta, \theta + 3\lambda\sqrt{\log\sigma_1}] + 1/\sigma_1 \\
&\le \mathsf{Pr}[Z_1 < \theta] + O(3\lambda\sqrt{\log\sigma_1}/\sigma_1) + 1/\sigma_1.
\end{aligned}$$

The claim now follows from a similar argument by starting from Z_1 instead of Z_2. □

Proof of Fact 2.4. A simple induction shows that multinomial distributions are unimodal, with the density function being maximized either at a unique value j or at j and $j+1$. For this value j, it holds that $\mathsf{d}_{\mathsf{TV}}(X, X+1) = \mathsf{Pr}[X \leq j] - \mathsf{Pr}[X+1 \leq j] = \mathsf{Pr}[X = j]$. We now use the anti-concentration of X, which follows from the Berry-Esséen theorem. Indeed by Theorem 2.1, if $Z \leftarrow \mathcal{N}(0,1)$, then $\mathsf{Pr}[X = j] \leq \mathsf{Pr}\left[Z = (j-\mu)/\sigma\right] + 2/\sigma = 2/\sigma$. □