2015 IEEE 56th Annual Symposium on Foundations of Computer Science

# Beyond the central limit theorem: Asymptotic expansions and pseudorandomness for combinatorial sums

Anindya De
*Northwestern University*
*Evanston, IL 60201*
*anindya@eecs.northwestern.edu*

**Abstract**

We prove a new asymptotic expansion in the central limit theorem for sums of discrete independent random variables. The classical central limit theorem asserts that if $\{X_i\}_{i=1}^n$ is a sequence of $n$ i.i.d. random variables, then $S = \sum_{i=1}^n X_i$ converges to a Gaussian whose first two moments match those of $S$. Further, the rate of convergence is $O(n^{-1/2})$. Roughly speaking, asymptotic expansions of the central limit theorem show that by considering a family of limiting distributions specified by $k \geq 2$ moments ($k = 2$ corresponds to Gaussians) and matching the first $k$ moments of $S$ to such a limiting distribution, one can achieve a convergence of $n^{-(k-1)/2}$. While such asymptotic expansions have been known since Cramér [1], they did not apply to discrete and non-identical random variables. Further, the error bounds in nearly all cases was non-explicit (in their dependence on $\{X_i\}$), thus limiting their applicability. In this work, we prove a new asymptotic expansions of the central limit theorem which applies to discrete and non-identical random variables and the error bounds are fully explicit.

Given the wide applicability of the central limit theorem in probability theory and theoretical computer science, we believe that this new asymptotic expansion theorem will be applicable in several settings. As a main application in this paper, we give an application in derandomization: Namely, we construct PRGs for the class of *combinatorial sums*, a class of functions first studied by [2] and which generalize many previously studied classes such as combinatorial rectangles [3], small-biased spaces [4] and modular sums [5] among others. A function $f : [m]^n \to \{0, 1\}$ is said to be a combinatorial sum if there exists functions $f_1, \ldots, f_n : [m] \to \{0, 1\}$ such that $f(x_1, \ldots, x_n) = f_1(x_1) + \ldots + f_n(x_n)$. For this class, we give a seed length of $O(\log m + \log^{3/2}(n/\epsilon))$, thus improving upon [2] whenever $\epsilon \leq 2^{-(\log n)^{3/4}}$.

**Keywords**

Central limit theorem; asymptotic expansions; derandomization

## I. Introduction

The central limit theorem (CLT) and its variants are aptly regarded as some of the most central results of probability theory and statistics. In theoretical computer science, CLTs have been used extensively in Boolean function analysis [6], hardness of approximation [7], stochastic optimization [8], property testing [9] and game theory [10] among other areas.

In classical probability theory, CLTs usually state that if $\{X_i\}_{i=1}^n$ is a sequence of independent random variables, then under *mild* conditions (such as all $X_i$ having finite absolute third moments), the random variable $S = \sum X_i$ converges to $\mathcal{N}(\mu, \sigma^2)$ where $\mu = \mathbf{E}[S]$ and $\sigma^2 = \mathrm{Var}(S)$. The convergence is typically weak convergence or convergence in the Kolmogorov distance as in the Berry-Esséen theorem [11] ( which also gives quantitative bounds).

However, in many applications, one requires convergence in stronger metrics such as the $\ell_1$ distance. One thing to be noticed right away is that for convergence in $\ell_1$ distance, there is a need to distinguish between the cases when $\{X_i\}$ are discrete and $\{X_i\}$ are continuous, as otherwise, no non-trivial convergence is possible. In this paper, we will focus on the discrete case. Whenever we say $\{X_i\}$ are discrete, we will mean that there exists $r > 0$ and $\{\alpha_i\}$ such that $X_i$ is supported on $\alpha_i + r \cdot \mathbb{Z}$. Thus, all $\{X_i\}$ are supported on arithmetic progressions (APs) where the common difference is the same for all $i \in [n]$. Thus, a case where $X_1$ is supported on $\mathbb{Z}$ whereas $X_2$ is supported on $\sqrt{2} \cdot \mathbb{Z}$ is not allowed. We remark that this restriction on discrete random variables is very standard in the literature on limit theorems. Discrete CLTs have been at the technical core of many recent works in theoretical computer science such as [10], [12], [9], [13], [2] among others.

To discuss discrete CLTs, we make the following definitions. A discretized normal $\mathcal{Z}(\mu, \sigma^2)$ is the integer valued random variable obtained by sampling from $\mathcal{N}(\mu, \sigma^2)$ and rounding to the nearest integer. Also, the shift-invariance of a random variable $X$, $d_{\mathsf{shift}}(X)$ equals $d_{\ell_1}(X, X+1)$[1]. When $X$ is supported on an AP with common difference 1, $d_{\mathsf{shift}}(X)$ measures the *smoothness* of the random variable $X$. We now state a discrete CLT from the book by Chen-Goldstein-Shao [14] (Theorem 7.4).

The work was mainly done while the author was a postdoc at IAS and DIMACS and supported by NSF CCF-1319788, CCF-0872397 and subcontract No. 00001583.

[1] $d_{\ell_1}$ is also sometimes referred to as $d_{TV}$.

0272-5428/15 $31.00 © 2015 IEEE
DOI 10.1109/FOCS.2015.59

*Theorem 1:* Let $\{X_i\}$ be $n$ independent integer valued random variables such that for $i \in [n]$, $\mathbf{E}[X_i] = \mu_i$, $\mathrm{Var}(X_i) = \sigma_i^2$ and $\mathbf{E}[|X_i - \mu_i|^3] = \gamma_i$. Let $\mu = \sum \mu_i$, $\sigma^2 = \sum \sigma_i^2$ and $\gamma = \sum \gamma_i$. Assume that for all $i \in [n]$, $d_{\mathsf{shift}}(S - X_i) \leq \delta$. Then,

$$d_{\ell_1}(S, \mathcal{Z}(\mu, \sigma^2)) \leq O(\delta) + O\left(\frac{1}{\sigma}\right) + O\left(\frac{\beta}{\sigma^3}\right) + O\left(\frac{\delta\beta}{\sigma^2}\right).$$

We would like the reader to observe two things about this theorem. The first is that for the above bound to be meaningful, we require $d_{\mathsf{shift}}(S - X_i)$ to be small for every $i \in [n]$. Roughly, it means that we require $S$ to be a *smooth* distribution (of course, $S$ is supported on integers, so the smoothness is in a discrete sense). As we will see later, in our asymptotic expansions theorem, we will also explicitly include a term in our error bounds to ensure smoothness of $S$. However, our error term will be expressed in terms of the Fourier spectrum of $S$.

To observe the next limitation, consider the case when $\{X_i\}_{i=1}^n$ are identically distributed with the common distribution $X_0$. If $d_{\mathsf{shift}}(X_0) < 1$ [2], then $d_{\ell_1}(S, \mathcal{Z}(\mu, \sigma^2))$ simplifies to $O(n^{-1/2})$ (where the $O(1)$ factor is dependent on $X_0$). It is a natural question to ask that given we approximate $S$ here in terms of its first two moments (since a normal is specified by its first two moments) and achieve an error rate of $n^{-1/2}$, might it be possible to get a better error bound by considering a larger number of moments? In other words, let $\{X_i\}_{i=1}^n$ be a family of i.i.d. random variables with common distribution $X_0$ and likewise, $\{Y_i\}_{i=1}^n$ is a family of i.i.d. random variables with common distribution $Y_0$. If the first two moments of $X_0$ and $Y_0$ match and $d_{\mathsf{shift}}(X_0), d_{\mathsf{shift}}(Y_0) < 1$, then $S_X = \sum X_i$ and $S_Y = \sum Y_i$ are $O(n^{-1/2})$ close to each other in $\ell_1$ distance. Is it possible that if the first $k > 2$ moments of $X_0$ and $Y_0$ match, one can achieve an even better closeness (potentially $O(n^{-(k-1)/2})$)?

### A. Asymptotic expansions

As a first step, we consider the above question with a weaker metric, namely the Kolmogorov distance. On one hand, it is a significantly weaker metric than $\ell_1$ distance. On the other hand, consider the case when $X_0$ and $Y_0$ are both supported on integers in an interval of size $O(1)$. In this case, note that since $\mathrm{Var}(S_X) = \mathrm{Var}(S_Y) = O(\sqrt{n}) \cdot \mathrm{Var}(X_0) = O(\sqrt{n})$. This means that $99\%$ of the mass of both $S_X$ and $S_Y$ is supported on a set of size $O(\sqrt{n})$ and thus $o(n^{-1/2})$ closeness in Kolmogorov metric implies $o(1)$ closeness in $\ell_1$ distance. Hence, for $o(n^{-1/2})$ error, answering the moment matching question in Kolmogorov distance is a meaningful step to answering the question for $\ell_1$ distance (and indeed, that is what we do in this paper). Also, based on the preceding discussion, for $o(n^{-1/2})$ error, we will need to distinguish between cases where $\{X_i\}_{i=1}^n$ are discrete versus continuous.

The first answer in the affirmative to this line of enquiry was provided by Cramér [1]. To describe this and further results, we will recall the definition of characteristic function. For a $\mathbb{R}$-valued random variable $X$, we define $\widehat{X} : \mathbb{R} \to \mathbb{C}$ as $\widehat{X}(\xi) = \mathbf{E}_{x \sim X}[e^{i \cdot x \cdot \xi}]$. Also, for a random variable $X$, $\beta_{k,X} = \mathbf{E}[|X - \mu_X|]^k$. The following theorem was proven by Cramér.

*Theorem 2:* ***[Cramér]*** Let $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ be two families of i.i.d. random variables with common distribution $X_0$ and $Y_0$ respectively. Further, $\mathbf{E}[X_0^i] = \mathbf{E}[Y_0^i]$ for $1 \leq i \leq k$ and $\limsup_{|\xi| \to \infty} |\widehat{X_0}(\xi)|, |\widehat{Y_0}(\xi)| < 1$. Then,

$$d_K(S_X, S_Y) = O\left(\frac{1}{n^{(k-1)/2}}\right).$$

Note that $O(1)$ hides dependence on $X_0$, $Y_0$ and $k$. Here $d_K$ represents the Kolmogorov or the cdf distance.

Begin by observing that if $X_0$ (resp. $Y_0$) were supported on AP with common difference $r$, then $\widehat{X_0}$ (resp. $\widehat{Y_0}(\cdot)$) is periodic with period $2\pi/r$. In that case, $\limsup |\widehat{X_0}(\xi)| = 1$ (resp. $\limsup |\widehat{Y_0}(\xi)| = 1$). Thus, for this theorem to be meaningful, we require $X_0$ and $Y_0$ to have an absolutely continuous component which will imply that $\limsup_{|\xi| \to \infty} |\widehat{X_0}(\xi)|, |\widehat{Y_0}(\xi)| < 1$.

Note that unlike CLTs (i.e. $k = 2$), here we explictly do not mention the *interpolating distribution* (for $k = 2$, it is the Gaussian). This is because it's somewhat more tedious to describe but we do remark that [1] does provide the interpolating family of distributions (where each distribution is defined only in terms of its first $k$ moments). In fact, these measures are obtained by doing an expansion of the characteristic function, which is the reason such results are known as asymptotic expansions of the central limit theorem.

The two main limitations of this result are that it applies only to i.i.d. continuous random variables. Further, the error bound is not explicit as in the dependence of the $O(1)$ factor on $X_0$, $Y_0$ and $k$ are not specified. Without the knowledge of this factor, it is not possible to get any meaningful error bound in applications where $X_0$ and $Y_0$ depend on $n$. Towards this, later works (such as Petrov [15], Theorem 5.1) makes the error bound explicit when $X_0$ and $Y_0$ have an absolutely continuous

[2]From this condition, it is not difficult to derive that $d_{\mathsf{shift}}(S - X_j) = O(n^{-1/2})$ for all $j \in [n]$.

component.

For discrete i.i.d. random variables, Ibragimov and Linnik [16] showed the following theorem.

*Theorem 3:* ***[Igrabimov-Linnik]*** Let $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ be two families of i.i.d. random variables with common distribution $X_0$ and $Y_0$ respectively. Further, $\mathbf{E}[X_0^i] = \mathbf{E}[Y_0^i]$ for $1 \leq i \leq k$ and both $X_0$, $Y_0$ are supported on APs with common difference 1. If $d_{\mathsf{shift}}(X_0), d_{\mathsf{shift}}(Y_0) < 1$, then,

$$d_K(S_X, S_Y) = O\left(\frac{1}{n^{(k-1)/2}}\right).$$

As with Cramér's result, the error bounds in the Ibragimov-Linnik result are again non-explicit, thus limiting their applicability. Note that as with Theorem 1, one needs to assume that non-trivial bounds on the shift invariance of $X_0$ and $Y_0$ (which in turn implies bounds on the shift-invariance of $S_X$ and $S_Y$). Unfortunately, prior to our work, no explicit error bounds had been obtained for the case of i.i.d. discrete random variables [17], [18].

## II. Our results

In the main result of this paper, we obtain asymptotic expansion theorems for sums of independent (not necessarily identically distributed) discrete random variables. Before we explain the most general version of our theorem, we state the following useful corollary. We hope that this will let the reader ease into the general theorem.

*Corollary 1:* Let $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ are two families of integer-valued independent random variables such that $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ are supported on the interval $[-b, b]$. If the first $k$ moments of $S_X$ and $S_Y$ are identical and $\sigma^2 = \mathrm{Var}(S_X) = \mathrm{Var}(S_Y)$,

$$\max_z \left|\Pr[S_X = z] - \Pr[S_Y = z]\right| \leq \left(\frac{b}{\sigma}\right)^{(k-1)} + e^{-\Theta(\sigma^2/b^2)} + \prod_{i=1}^n \sup_{|\zeta| \in [1/(10\cdot b), \pi]} \left|\mathbf{E}[\widehat{X_i}(\zeta)]\right| + \prod_{i=1}^n \sup_{|\zeta| \in [1/(10\cdot b), \pi]} \left|\mathbf{E}[\widehat{Y_i}(\zeta)]\right|.$$

To appreciate the strength of the above corollary, consider the case when $b = O(1)$. In this case, we get

$$\|S_X - S_Y\|_\infty = O\left(\frac{1}{\sigma}\right)^{(k-1)} + e^{-\Theta(\sigma^2)} + \prod_{i=1}^n \sup_{|\zeta| \in [c, \pi]} \left|\mathbf{E}[\widehat{X_i}(\zeta)]\right| + \prod_{i=1}^n \sup_{|\zeta| \in [c, \pi]} \left|\mathbf{E}[\widehat{Y_i}(\zeta)]\right|.$$

where $c > 0$ is an absolute constant. It is useful to compare this corollary with the error bounds from Theorem 1 in the same setting. By applying Theorem 1, one gets that $\|S_X - S_Y\|_1 = O(1/\sigma + \delta + \eta)$ where $d_{\mathsf{shift}}(S_X) = \delta$ and $d_{\mathsf{shift}}(S_Y) = \eta$. By fairly elementary considerations [19], it is possible to show that $\delta, \eta = \Omega(1/\sigma)$ and thus, this method cannot yield bounds better than $O(1/\sigma)$.

On the other hand, assume that for $i \in [n]$, $d_{\mathsf{shift}}(X_i) = 1 - \delta_i$ and $d_{\mathsf{shift}}(X_i) = 1 - \eta_i$. Then, one can easily deduce that $\sup_{|\zeta| \in [c,\pi]} \left|\mathbf{E}[\widehat{X_i}(\zeta)]\right| \leq 1 - \Omega(\delta_i)$ and $\sup_{|\zeta| \in [c,\pi]} \left|\mathbf{E}[\widehat{Y_i}(\zeta)]\right| \leq 1 - \Omega(\eta_i)$ and thus,

$$\|S_X - S_Y\|_\infty = O\left(\frac{1}{\sigma}\right)^{(k-1)} + e^{-\Theta(\sigma^2)} + \prod_{i=1}^n (1 - \Omega(\delta_i)) + \prod_{i=1}^n (1 - \Omega(\eta_i)).$$

The last two terms typically decay as $\exp(-\Omega(\sigma^2))$ and thus the error is essentially dominated by the first term $O(\sigma^{-(k-1)})$. While Corollary 1 provides us with a bound on $\|S_X - S_Y\|_\infty$, since the size of the effective support of $S_X$ and $S_Y$ is $O(\sigma)$, we can get bounds on $\|S_X - S_Y\|_1$ by multiplying the error bounds by $O(\sigma)$. Thus, we will end up with an error bound of $O(\sigma^{-(k-2)})$. However, if $k > 3$, we still beat the bounds from Theorem 1.

To summarize the discussion, at a qualitative level, the Fourier term is playing the same role as the shift-invariance terms in Theorem 1, that of being a proxy for the smoothness of $S_X$, $S_Y$. However, what works to our advantage is that it is a significantly finer measure of smoothness compared to the shift invariance and in many applications, significantly smaller than shift-invariance.

### A. Asymptotic expansions theorem

We now state the main asymptotic expansion theorem of our paper. To state the main theorem, we will require a few definitions. We will assume that $\{X_i\}_{i=1}^n$ are independent centered (i.e. mean is 0) random variables such that each $X_i$ is supported on an AP of common difference 1. Note that variables can be centered by simply shifting them. Next $\alpha_{i,k}$, $\beta_{i,k}$ and $\gamma_{i,k}$ respectively denote the $k^{th}$ moment, $k^{th}$ absolute moment and $k^{th}$ cumulant of $X_i$. Define $\beta_k = \sum_{i=1}^n \beta_{i,k}$ and

$\sigma_i^2 = \alpha_{i,2}$ and $\sigma^2 = \sum_{i=1}^n \sigma_i^2$. We set $C > 0$ to be a large positive constant (choosing $C = 10^6$ suffices for our purposes). With this, we define

$$L_k = \left(\frac{\beta_k}{\sigma^k}\right)^{1/(k-2)} \quad \text{and} \quad I_k = \frac{1}{C}\min\left\{\min\frac{\sigma}{\sigma_i}, \frac{1}{L_{3k}}\right\}.$$

We note that typically, $I_k \gg 1$ and $L_k \ll 1$. Let $Z = (X_1 + \ldots + X_n)/\sigma$. Note that $Z$ is supported on an AP of common difference $1/\sigma$. Call this AP $\mathcal{L}_Z$. For this $Z$, we also define the family of polynomials $\{P_\nu\}_{\nu\in\mathbb{N}}$ as follows: For the moment, assume that all moments of $Z$ exist and hence all cumulants of $Z$ also exist. Then, the polynomial $P_\nu : \mathbb{C} \to \mathbb{C}$ is defined as follows: $P_\nu(\xi)$ is the coefficient of $w^{\nu+2}$ in the formal expansion of

$$\exp\left(\sum_{j=1}^{\infty} \frac{\lambda_{j+2}\cdot\xi^{j+2}\cdot w^j}{(j+2)!}\right),$$

where $\lambda_j$ is the $j^{th}$ cumulant of $Z$. Note that it is immediately clear that to define $P_\nu$, we only require the existence of moments (or equivalently, cumulants) of $Z$ up to order $\nu+2$. With the definitions in place, we now state the main theorem:

*Theorem 4:* Let $\{X_i\}_{i=1}^n$ and $Z$ be as defined above. If $I_s \geq 2\sqrt{s\cdot\log s}$, then for any $z \in \mathcal{L}_z$,

$$\left|\Pr[Z=z] - \frac{1}{2\pi\sigma}\cdot\int_{-\pi\sigma}^{\pi\sigma} e^{-i\xi z}\cdot e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3}P_\nu(i\xi)\right)d\xi\right| \leq \eta_{\mathsf{low}} + \eta_{\mathsf{med}} + \eta_{\mathsf{high}},$$

where

$$\eta_{\mathsf{low}} = s^{O(s)}\cdot L_{3s}^{s-2}, \quad \eta_{\mathsf{med}} = e^{-\frac{I_s^2}{6}} + s^{O(s)}\cdot e^{-\frac{I_s^2}{4}} \text{ and } \eta_{\mathsf{high}} = \sup_{|\zeta|\in[\sigma^2/(10\cdot\beta_3),\pi]}\left|\prod_{i=1}^n \widehat{X_i}(\zeta)\right| + s^{O(s)}\cdot e^{-\frac{I_s^2}{4}}.$$

We begin by observing that in the above theorem, the expression $\frac{1}{2\pi\sigma}\cdot\int_{-\pi\sigma}^{\pi\sigma} e^{-i\xi z}\cdot e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3}P_\nu(i\xi)\right)$ is dependent only on the first $s-1$ cumulants (or equivalently moments) of $Z$. Thus, if $S_X$ and $S_Y$ (from the earlier discussion) have their first $s-1$ moments equal, then we can apply Theorem 4 to obtain a corresponding *moment matching* theorem. Of course, to get meaningful error bounds, we need to control $\eta_{\mathsf{low}}$, $\eta_{\mathsf{med}}$ and $\eta_{\mathsf{high}}$.

Before we elaborate a little further on the error terms, we note that Theorem 4 gives $\ell_\infty$ bounds on the difference between $Z$ and the interpolating measure as opposed to a $\ell_1$ bound. However, as we will see very soon, the error on the right hand side will typically grow like $O(\sigma^{-(s-2)})$. Thus, as with Corollary 1, since the size of the effective support of $Z$ is $O(\sigma)$, we get a $\ell_1$ error bound of $O(\sigma^{-(s-3)})$.

We elaborate a little further on the error terms. The first error term we focus on is $s^{O(s)}\cdot L_{3s}^{s-2} = s^{O(s)}\cdot\frac{\beta_s}{\sigma^s}$. We note that for $s=3$ (which is the case corresponding to the standard CLT), this is the same term which appears in the Berry-Esséen theorem. To gauge the other error terms, consider the setting where $\{X_i\}_{i=1}^n$ are hypercontractive. In other words, there is some fixed function $c(\cdot) : \mathbb{N} \to \mathbb{R}$ such that $\beta_{i,k} \leq c(k)^{(k-2)}\cdot\beta_{i,2}^{k/2}$. Also, $\max\sigma_i \leq \epsilon\cdot\sigma$ (for some small $\epsilon$). Both these conditions are mild and hold in most applications. In fact, for many applications, the above holds with $c(s) = O(s)$. Under these assumption, $L_s \leq \epsilon\cdot c(s)$ and we get that

$$s^{O(s)}\cdot L_{3s}^{s-2} + s^{O(s)}\cdot e^{\Omega(I_s^2)} = (s\cdot c(3s))^{O(s)}\cdot\epsilon^{s-2} + e^{-\frac{c(3s)}{\epsilon^2}}.$$

This leaves us with the "Fourier error term" namely, $\sup_{|\zeta|\in[\sigma^2/\beta_3,\pi]}\left|\prod_{i=1}^n\widehat{X_i}(\zeta)\right|$. As before, this is a measure of smoothness of the distribution $Z$. For instance, if for $i\in[n]$, $d_{\mathsf{shift}}(X_i) = 1-\delta_i$, then following the same calculation after Corollary 1, one can show that

$$\sup_{|\zeta|\in[\sigma^2/(10\cdot\beta_3),\pi]}\left|\prod_{i=1}^n\widehat{X_i}(\zeta)\right| \leq \prod_{i=1}^n\left(1-\frac{\delta_i}{\Omega(c(3))}\right).$$

For many instances, such as the case when $\{X_i\}_{i=1}^n$ are Bernoullis and in our application for PRGs, this term decays as $\epsilon^{-\Omega(1/\epsilon^2)}$ which means that for any $s \leq (1/\epsilon)^\kappa$ for a sufficiently small constant $\kappa > 0$, the dominant error term is $(s\cdot c(3s))^{O(s)}\cdot\epsilon^{s-2}$. Thus, assuming that $\epsilon$ is sufficiently small and taking $s > 3$, we beat bounds from the central limit theorem.

### B. Related work

As we have mentioned before, there has been a lot of work in classical probability theory on obtaining asymptotic expansions. The name Edgeworth expansion is associated with many of these results. However, most of the work has been in the setting of i.i.d. random variables. Also, apart from the case of continuous i.i.d. random variables [15], error bounds have been non-explicit. A good reference for the existing work are the books by Petrov [15] and Bhattacharya and Rao [20]. Unlike previous works, our estimates do not require the variables to be identically distributed and we obtain explicit error estimates. However, we rely on many techniques and ideas from [15] and Esséen's thesis [21].

As far as explicit estimates are concerned, two of the most related works are Barbour and Čekanvičius [22] and Roos [23]. Barbour and Čekanvičius approximate sums of independent integer valued random variables using Signed Compound Poisson (SCP) measure. SCP measures are a generalization of Poisson measure. For any $k$, $\mathrm{SCP}_k$ is a class of measures which is specified by its first $k$ moments. The error estimates in [22] are fully explicit and for sums of $n$ i.i.d. random variables, the error is roughly of the same quality which is obtained (non-explicitly) using the asymptotic expansion in [16]. In fact, the non-explicit nature of the error in [16] was cited as the main motivation for their work. Unfortunately, the quality of the error rapidly degrades rapidly as the variance of the individual random variables increases makes it unsuitable for our applications. There does not seem to be a fix for this [17]. In terms of techniques, it is substantially different from ours. Namely, Barbour and Čekanvičius use Stein's method for SCP measures while ours is based on Fourier analysis.

Roos also obtains approximations for sums of independent Bernoulli random variables in terms of their first $k$ moments. The error bounds are fully explicit and in general, incomparable to our bounds. When the variance of the sum is small, then these bounds beat our error bounds and in the high-variance regime, our bounds are better. Unfortunately, the scope of applicability of Roos's results is limited by the fact that it applies only to sums of Bernoulli random variables whereas ours are applicable to arbitrary discrete random variables (supported on an arithmetic progression).

The astute reader will also notice that the overall theme of our work bears resemblance to the "moment problem" in probability theory where the overall question is: *How different can two distributions be, if all / many of their moments are identical?* We remark that while the questions are similar, the answers and techniques are very different. For example, most work for the "moment problem" can show closeness of distributions in very weak metrics such as the Lévy metric [11] whereas we show convergence in $\ell_1$ distance.

In theoretical computer science, the work most closely related to ours is the work of Daskalakis and Papadimtriou [12] who used results from [23] to show the following: Let $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ be sequence of independent Bernoullis and for $S_X = \sum X_i$ and $S_Y = \sum Y_i$, the first $k$ moments of $S_X$ and $S_Y$ are equal. Then, $d_{\ell_1}(S_X, S_Y) = 2^{-\Omega(k)}$. We remark that their main application for this result was to obtain sparse covers for Poisson Binomial distributions (i.e. sums of independent indicators). By using Roos's results [23] and our asymptotic expansion theorems, we improve on their results to obtain (nearly) optimal cover size for Poisson Binomial distributions. We elaborate on this a little later. Before that, we give a brief description of our techniques.

### C. Our techniques

Our proof goes through one of the most usual techniques in probability theory to prove limit theorems, namely the characteristic function method. To prove Theorem 4, we observe that $Z$ is supported on an AP with common difference $1/\sigma$. By standard Fourier analysis, this means that to prove Theorem 4, it suffices to prove for $\xi \in [-\pi \cdot \sigma, \pi \cdot \sigma]$,

$$\mathsf{err}_\xi = \left| \widehat{Z}(\xi) - e^{-\frac{\xi^2}{2}} \cdot \left(1 + \sum_{\nu=1}^{s-3} P_\nu(i\xi)\right) \right| \leq \eta_{\mathsf{low}} + \eta_{\mathsf{med}} + \eta_{\mathsf{high}}.$$

To prove this bound, we divide $[-\pi \cdot \sigma, \pi \cdot \sigma]$ into three regions: $I_{\mathsf{low}}$, $I_{\mathsf{med}}$ and $I_{\mathsf{high}}$. Here $I_{\mathsf{low}} = [-I_s, I_s]$, $I_{\mathsf{med}} = [-\frac{1}{10L_3}, \frac{1}{10L_3}] \setminus I_{\mathsf{low}}$ and $I_{\mathsf{high}} = [-\pi \cdot \sigma, \pi \cdot \sigma] \setminus (I_{\mathsf{low}} \cup I_{\mathsf{med}})$. As the names suggest, we show that for $\xi \in I_{\mathsf{low}}$ (resp. $I_{\mathsf{med}}$ and $I_{\mathsf{high}}$ ), $\mathsf{err}_\xi$ is bounded by $\eta_{\mathsf{low}}$ (resp. $\eta_{\mathsf{med}}$, $\eta_{\mathsf{high}}$).

Bulk of the technical work goes in showing the bound in $I_{\mathsf{low}}$. The bound in $I_{\mathsf{med}}$ and $I_{\mathsf{high}}$ are relatively simple. We start by giving some idea of the bound in $I_{\mathsf{med}}$ and $I_{\mathsf{high}}$. For both $I_{\mathsf{med}}$ and $I_{\mathsf{high}}$, we simply rely on the fact that $\mathsf{err}_\xi \leq |\widehat{Z}(\xi)| + |e^{-\frac{\xi^2}{2}} \cdot \left(1 + \sum_{\nu=1}^{s-3} P_\nu(i\xi)\right)|$. Thus, it suffices to individually bound these quantities in $I_{\mathsf{med}}$ and $I_{\mathsf{high}}$.

By fairly easy calculations (based on inequalities from functional analysis on moment comparison), it is possible to show that $e^{-\frac{\xi^2}{2}} \cdot \left(1 + \sum_{\nu=1}^{s-3} P_\nu(i\xi)\right) \leq 2^{O(s)} \cdot e^{-\xi^2/2} \cdot (|\xi|^3 + |\xi|^{3s})$. This implies that if $\xi \in I_{\mathsf{med}} \cup I_{\mathsf{high}}$, then $e^{-\frac{\xi^2}{2}} \cdot \left(1 + \sum_{\nu=1}^{s-3} P_\nu(i\xi)\right) \leq s^{O(s)} \cdot e^{-I_s^2/4}$. See Corollary 12 for details.

To bound $\mathsf{err}_\xi \in I_{\mathsf{med}} \cup I_{\mathsf{high}}$, it remains to bound $\widehat{Z}(\xi) \in I_{\mathsf{med}} \cup I_{\mathsf{high}}$. However, by definition, $\xi \in I_{\mathsf{high}}$, $|\widehat{Z}(\xi)| \leq \sup_{|\zeta| \in [\sigma^2/(10 \cdot \beta_3), \pi]} \left| \prod_{i=1}^n \widehat{X_i}(\zeta) \right|$. This uses the fact that $\widehat{Z}(\xi) = \prod_{i=1}^n \widehat{X_i}(\xi/\sigma)$. On the other hand, for $\xi \in I_{\mathsf{med}}$, $|\widehat{Z}(\xi)| \leq$

$e^{-\xi^2/6}$. While we give the proof of this fact in our paper, it is not new and was used in the Fourier analytic proof of Berry-Esséen theorem [11]. This again implies that $\xi \in I_{\mathsf{med}}$, $|\widehat{Z}(\xi)| \leq e^{-I_s^2/6}$. The details for $\xi \in I_{\mathsf{med}}$ can be found Section IV-E2 and $\xi \in I_{\mathsf{high}}$ can be found Section IV-E3.

Thus all that remains to be done is to bound $\mathsf{err}_\xi$ for $\xi \in I_{\mathsf{low}}$. This is technically the hardest part of the paper. The overall idea is to expand $\widehat{Z}(\xi)$ in terms of powers of $\xi$ and truncating the expansion at $(s-1)^{th}$ power. However, a naive Taylor's expansion of $\widehat{Z}(\xi)$ around $\xi = 0$ does not give meaningful error bounds. Rather, what we do here is to consider the function $V(\xi) = \xi^2/2 + \log \widehat{Z}(\xi)$. Note that the definition of $V(\xi)$ requires us to take logarithms of complex numbers. In itself, it is not problematic but one needs to ensure that all points where this function is considered, $\widehat{Z}(\xi) \neq 0$. Note that $\widehat{Z}(\xi) = e^{-\xi^2/2} \cdot e^{V(\xi)}$. Now, to expand $\widehat{Z}(\xi)$ in terms of powers of $\xi$, we expand $e^{V(\xi)}$ in terms of powers of $\xi$. There are two steps in the argument: the first step shows that in the expansion of $e^{V(\xi)}$, if all powers of $V(\xi)$ beyond the first $s-3$ powers are ignored, the total error is small (i.e. bounded by $\eta_{\mathsf{low}}$). For $V(\xi)^j$ where $j \leq s-3$, we do a Taylor's expansion in powers of $\xi$ and truncate all powers of $\xi$ beyond $s-1$.

Both these calculations are somewhat intricate and essentially rely on combining functional analytic estimates with some calculus. As the calculations are rather technical, we refrain from giving further details. The full details for $\xi \in I_{\mathsf{low}}$ are in Section IV-D.

At a high level, the proof follows a structure which is a combination of those in [15] and [21]. However, because we are dealing with discrete non-identical random variables, our proofs are more intricate and we need to use several new ideas. For example, dealing with i.i.d. variables naturally results in several cancellations, whereas our proofs have to replace those with functional analytic estimates. Similarly, since our theorems are written with an eye towards applications, we optimize on several factors. For example, the dependence of the error on $s$ in Theorem 4 grows like $s^{O(s)}$. However, a straightforward adaptation of the arguments from [15] and [21] gives a factor of $s^{O(s^2)}$. This optimization is critical for the application in [24]. At the level of implementation, our work also draws on some tricks used in contemporary works such as the pathbreaking work of Mossel et. al. [25].

### *D. Applications of the asymptotic expansion theorem*

Given the utility of central limit theorem in theoretical computer science, we expect the new asymptotic expansion theorem to have several applications. So far, we have identified two applications which are discussed below. The first is towards construction of better pseudorandom generators for combinatorial shapes, an important class of functions in space bounded derandomization. We discuss this application in detail in the next section.

The next application is towards construction of sparse covers for Poisson Binomial distributions. As the result requires a lot of ideas on top of what we prove here, we present the results in a separate paper [24]. However, we briefly discuss the results of that paper here. A Poisson Binomial distribution (PBD) of order $n$ consists of the class of random variables $Z = X_1 + \ldots + X_n$ where $\{X_i\}_{i=1}^n$ are independent Bernoulli random variables. Given a class of distributions $\mathcal{S}$, a set $\mathcal{S}_\epsilon$ is said to be a $\epsilon$ cover for $\mathcal{S}$ (under $\ell_1$ metric) if for any $\mathcal{D} \in \mathcal{S}$, there exists $\mathcal{D}' \in \mathcal{S}_\epsilon$ such that $d_{\ell_1}(\mathcal{D}, \mathcal{D}') \leq \epsilon$.

Despite its basic nature and several results concerning the limiting behavior of such distributions [14], [23], [26], [19], [27], until a few years ago, no non-trivial bounds were known for the cover size of such distributions. Motivated by its application in approximating Nash equilibrium in anonymous games [10], [28], [12], Daskalakis and Papadimtriou [29] showed the following theorem:

*Theorem 5:* For the class of PBDs of order $n$ and $\epsilon > 0$, there exists an $\epsilon$-cover $\mathcal{S}_{n,\epsilon}$ of size $O(n^2 + n \cdot (1/\epsilon)^{O(\log^2(1/\epsilon))})$. Further, this set $\mathcal{S}_{n,\epsilon}$ can be constructed in time $\mathsf{poly}(|\mathcal{S}_{n,\epsilon}|)$.

While it is clear that the dependence of the size of $\mathcal{S}_{n,\epsilon}$ on $n$ is optimal up to polynomial factors, it is not clear whether the same holds true for the dependence on $\epsilon$. An improved cover size will imply faster approximation algorithms for Nash equilibrium in anonymous games. In [24], we prove the following theorem.

*Theorem 6:* For the class of PBDs of order $n$ and $\epsilon > 0$, there exists an $\epsilon$-cover $\mathcal{S}_{n,\epsilon}$ of size $O(n^2 + n \cdot (1/\epsilon)^{\tilde{O}(\log(1/\epsilon))})$. Further, this set $\mathcal{S}_{n,\epsilon}$ can be constructed in time $\mathsf{poly}(|\mathcal{S}_{n,\epsilon}|)$.

This result was also proven independently but chronologically earlier by Diakonikolas, Kane and Stewart [30].

## III. Overview of constructing PRGs for combinatorial sums

We begin by recalling the definition read-once branching programs.

*Definition 1:* A $(S, D, T)$ read-once branching program (ROBP) $M$ is a layered directed multi-graph with $T+1$ layers and at most $2^S$ vertices in each layer. For $0 < i < T$, 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 $i+1$.

We can associate a function $g_M : (\{0,1\}^D)^T \to [2^S]$ with $M$. Namely, we label each vertex in the $T^{th}$ layer with a number

in $[2^S]$ (not necessarily distinct). On input $x = (x_1, \ldots, x_T)$ where $x_i \in \{0,1\}^D$ for all $i \in [T]$, $g_M(x)$ is the label of the vertex obtained by starting at $v_0$ and walking along the edges $x_1, \ldots, x_T$ (in order).

For the reader familiar with branching programs, we remark that typically the vertices of the last layer are labeled with an element in $\{0,1\}$, thus the resulting $g_M(\cdot)$ becomes a Boolean function. In this paper, we go for a more general definition as it is more convenient for us. Derandomization of ROBP is one of the central challenges in modern complexity theory. In particular, derandomization of $(S, D, T)$ ROBP for $S = O(\log n)$, $T = n$ and $D = 1$ suffices for derandomization of RL.

Next, we define the notion of pseudorandomness for a complexity class $\mathcal{C}$. Towards this, let us adopt the notation that $U_m$ denotes the uniform distribution on $m$-bit strings. We recall the definition of pseudorandomness for ROBPs. A function $g : \{0,1\}^t \to (\{0,1\}^D)^T$ is said to be an $\epsilon$-PRG for $(S, D, T)$-ROBPs if for every $(S, D, T)$-ROBP $h$, the distributions are $h(U_D^T)$ and $h(g(U_t))$ are $\epsilon$-close in $\ell_1$ distance. Note that for this paper, whenever we say PRG, we would want $g$ to be poly-time computable. Also, $t$ is referred to as the seed length of the PRG. Currently, the best known PRG for ROBP is due to Impagliazzo, Nisan and Wigderson [31] which builds upon the seminal work of Nisan [32].

*Theorem 7:* ***[Impagliazzo-Nisan-Wigderson]*** There is an explicit PRG $G_{\mathsf{INW}} : \{0,1\}^t \to (\{0,1\}^D)^n$ which is $\epsilon$-pseudorandom for $(S, D, T)$-ROBP where $t = O(D + (S + \log(T/\epsilon)) \cdot \log(T))$.

The problem of beating [31] for general ROBPs has so far resisted all attacks and in fact, even for $S = 3$, $D = 1$ and $\epsilon < 1/2$, the best known PRG has seed length $O(\log^2 T)$. On the other hand, it is known that if one foregoes efficiency for the PRG (i.e. $g$ need not be efficiently computable), then the seed length can be $O(S + D + \log(T/\epsilon))$. To get around this, researchers have looked at restricted class of ROBPs: One line of research has dealt with structural restrictions on ROBPs (cf.[33], [34], [35], [36], [37], [38], [39]). The second line of research, which is also the focus of this paper (and predates the first line of research) has dealt with ROBPs computing *semantically restricted* classes of functions. These include small biased spaces [4], combinatorial rectangles [3], [40], [41], [42], combinatorial checkerboards [43] and modular sums [5] among others. A common generalization of many classes in the second line of research is the class of combinatorial sums which was introduced in the work of Gopalan *et al.* [2]. They are defined as follows.

*Definition 2:* The class of combinatorial sums (denoted by $\mathsf{Csum}(m, n)$) consists of $f : [m]^n \to \mathbb{Z}$ of the form

$$f(x_1, \ldots, x_n) = f_1(x_1) + \ldots + f_n(x_n) \text{ where for all } i \in [n],\ f_i : [m] \to \{0,1\}.$$

Sometimes we will use the tuple $(f_1, \ldots, f_n)$ to refer to the combinatorial sum $f$.

The main result of [2] is an $\epsilon$-PRG with seed length $O(\log m + \log n + \log^2(1/\epsilon))$. In this paper, we prove the following theorem.

*Theorem 8:* There is a polynomial time computable PRG $G_{csum} : \{0,1\}^{t_{\mathsf{csum}}} \to [m]^n$ which $\epsilon$-fools $\mathsf{Csum}(m, n)$. Here $t_{\mathsf{csum}} = O(\log m + \log^{3/2}(n/\epsilon))$.

When $\epsilon = n^{-\Theta(1)}$, this provides the first improvement over [31] for the class $\mathsf{Csum}(m, n)$. At the time, the first preprint of this paper appeared, even for $m = 2$, the best known PRG had a seed length of $O(\log^2 n)$ for $\epsilon = n^{-\Theta(1)}$. However, concurrently but independently, Gopalan, Kane and Meka [44] constructed PRGs for the case $m = 2$ (they construct PRGs for $\{0, \pm 1\}$ weight halfspaces) with seed length $\tilde{O}(\log n + \log(1/\epsilon))$. Their techniques seem to be rather different and disjoint from our techniques. Also, on one hand, their seed length is optimal for $m = 2$ but on the other hand, the construction and techniques in [44] only apply to $m = 2$.

We now give intuition for the proof of Theorem 8 and how Theorem 4 is helpful in proving this theorem. To do this, we begin with a high level description of the result of [2]. Let $U_{[m]^n}$ be the uniform distribution on $[m]^n$. Then, note that for $(y_1, \ldots, y_n) \sim U_{[m]^n}$, each $f_i(y_i)$ is an independent $\{0,1\}$ random variable. Let us define $Z = \sum f_i(y_i)$ and $\mathrm{Var}(Z) = \sigma^2$. The key technical component in their result is the use of the following consequence of Theorem 1.

*Corollary 2:* For $Z$ as defined above, let $Z' = \mathcal{Z}(\mathbf{E}[Z], \mathrm{Var}(Z))$ (Recall, that $\mathcal{Z}$ denotes a disretized normal). Then, $d_{\ell_1}(Z, Z') = O(1/\sigma)$.

We now explain the key idea behind using Corollary 2 for derandomization of combinatorial sums: The derandomization problem is split into two cases:

- low variance case: $\mathrm{Var}(Z) \leq \Theta(1/\epsilon^2)$ and
- high variance case: $\mathrm{Var}(Z) \geq \Theta(1/\epsilon^2)$.

We now describe the PRG for these cases. In case (i), we set $t = \mathsf{poly}(1/\epsilon)$, so that $\mathrm{Var}(Z)/t \leq \epsilon$. Let $\mathcal{H}_{2,n,t}$ be a family of pairwise independent hash functions. The PRG first samples $h \sim \mathcal{H}_{2,n,t}$ to partition $[n]$ into $t$ *buckets*. Within each bucket, the PRG uses $O(1)$-wise independence (the distribution across the buckets is independent). [2] use the notion of sandwiching polynomials (cf. [45]) to show that this PRG construction indeed fools $f$ (in case (i)) up to an error $\epsilon$. However, as the seed across the buckets are independent, hence the seed required grows as $t \cdot (\log m + \log n)$. This has a

poor dependence on $\epsilon$. To get around this, the authors observe that $f$ composed with our PRG is still a ROBP except that its length is much smaller than $n$. Thus, they derandomize this using Theorem 7 to bring down the seed length requirement to $O(\log m + \log n + \log^2(1/\epsilon))$.

For case (ii), note that by Corollary 2, $Z$ is $\epsilon$-close to an appropriate discretized normal in $\ell_1$ distance. Thus, it suffices to produce a pseudorandom distribution $(y'_1, \ldots, y'_n)$ so that the induced distribution $Z' = \sum f_i(y'_i)$ is $O(\epsilon)$-close to same discretized normal. In other words, it suffices to *fool* the proof of Corollary 2. The standard proofs of Corollary 2 (based on Stein's method and Fourier analysis) are somewhat less amenable to derandomization. So, the authors in [2] come up with a somewhat simpler proof (of a slightly weaker statement) which is amenable to derandomization. This gives a very high level overview of the utility of Corollary 2 for derandomization of combinatorial shapes (ignoring several details and complications).

Next, we observe that for $\epsilon = n^{-\Theta(1)}$, [2] gets $O(\log^2 n)$ seed. Thus, for this regime of error, it does not beat [31]. The main conceptual barrier in extending their techniques to get an error $o(n^{-1/2})$ with $o(\log^2 n)$ seed is as follows: The main idea is to derandomize the proof of Corollary 2 up to error $\epsilon$. As long as the discrete CLT has error $O(\epsilon)$, this automatically guarantees fooling the combinatorial shape to error $O(\epsilon)$. However, note that the optimal error rate for discrete CLTs with variance $\sigma^2$ is $O(\sigma^{-1})$. Since, $\sigma \leq n^{1/2}$, this approach is not useful for getting error rate $o(n^{-1/2})$.

Given our discussion about asymptotic expansions in the central limit theorem, it seems that it naturally fits into the framework of [2] but should be useful in getting smaller error.

### A. Application to fooling combinatorial sums

We recall our main result concerning derandomization of combinatorial sums.

*Theorem 1:* There is a efficiently computable PRG $G_{csum} : \{0,1\}^t \to [m]^n$ which $\epsilon$-fools every $f \in \mathsf{Csum}(m,n)$ and $t_{\mathsf{csum}} = O(\log m + \log^{3/2}(n/\epsilon))$.

The approach is best demonstrated by considering the setting when $\epsilon = n^{-\Theta(1)}$. Similar to [2], let us assume that the underlying combinatorial sum is specified by the tuple $(f_1, \ldots, f_n)$. Let $(y_1, \ldots, y_n) \sim U_{[m]^n}$ and $Z = \sum f_i(y_i)$.

- Low variance case: $\mathrm{Var}(Z) \leq 2^{c \cdot \sqrt{\log n}}$.
- High variance case: $\mathrm{Var}(Z) > 2^{c \cdot \sqrt{\log n}}$.

Here $c$ is a fixed positive constant. We remark that if $\epsilon = n^{-\omega(1)}$, then we do an *alphabet reduction* step which shows that it suffices to treat the case when $m = \mathsf{poly}(n/\epsilon)$. The seed length required for this step is $O(\log(m \cdot n/\epsilon))$. This is a fairly easy step and hence, we do not describe it here.

For the purposes of the intuition, let $\epsilon = n^{-\Theta(1)}$ and consider case (i). In this case, we set $t = 2^{C\sqrt{\log n}}$ where $C > c$ is a large constant and $k = \Theta(\sqrt{\log n})$. Let $H_{k,n,t}$ be a family of $k$-wise independent hash functions mapping $[n]$ to $[t]$. Note that the seed required to sample from $H_{k,n,t}$ is $t_{\mathsf{hash}} = O(\log^{3/2} n)$. The PRG first samples $h \sim H_{k,n,t}$ to partition $[n]$ into $t$ *buckets*. Within each bucket, the PRG uses $k$-wise independence and independent seeds are used for each bucket. Adapting the machinery for the low-variance case from [2] and combining it concentration bounds for sums of $k$-wise independent random variables, we obtain that the PRG described here $\epsilon$-fools $f \in \mathsf{Csum}(m,n)$ with error $\epsilon$. Unfortunately, as in [2], since the seed across the buckets are independent, the seed length is prohibitively large. However, applying the same trick as in [2], after fixing a choice of $h$, the function obtained by composing $f$ with our current PRG is a $(S', D', T')$ ROBP where $S' = n$, $T' = t$ and $D'$ is the seed required to sample $k$-wise independent distribution for each bucket. Note that $D' = O(\log^{3/2}(n))$. Thus, instead of using independent seeds, we can sample the seed using Theorem 7 which will imply a total seed length of $O(\log m + \log^{3/2}(n/\epsilon))$. Since the seed required to sample $h$ is $O(\log^{3/2} n)$, the total seed length remains $O(\log m + \log^{3/2}(n/\epsilon))$.

We next describe the approach for the high variance case. As the reader can guess, the high variance case is where the power of asymptotic expansions is used. Here we set $t = 2^{C\sqrt{\log n}}$ where $C < c$ is a constant. We also set $k = \Theta(\sqrt{\log n})$. The PRG first samples $h \sim H_{k,n,t}$ to partition $[n]$ into $t$ buckets. Note that the seed required for this step is $t_{\mathsf{hash}} = O(\log^{3/2}(n))$. Let $G_{k,n/t,m} : \{0,1\}^{t_1} \to [m]^{n/t}$ be a $k$-wise independent generator for the domain $[m]^{n/t}$. Further, let $G^{(cs)}_{m,n/t} : \{0,1\}^{t_2} \to [m]^{n/t}$ be the PRG from [2] which $\delta$-fools $\mathsf{Csum}(m, n/t)$ with $\delta$ set to a very small constant. Let $G^{(cs)}_{m,n/t} \otimes G_{k,n/t,m} : \{0,1\}^{t_1+t_2} \to [m]^{n/t}$ be defined as

$$G^{(cs)}_{m,n/t} \otimes G_{k,n/t,m} : (z_1, z_2) \mapsto G_{k,n/t,m}(z_1) \oplus_m G^{(cs)}_{m,n/t}(z_2).$$

Here $\oplus_m$ denotes the addition in the space $\mathbb{Z}_m^{n/t}$. The final PRG applies $t$ independent copies of $G^{(cs)}_{m,n/t} \otimes G_{k,n/t,m}$ across the different buckets. To do the analysis, note that for any specific choice of $h$, the function obtained by composing $f$ with our current PRG is a $(S', D', T')$ ROBP where $S' = n$, $T' = t$ and $D'$ is the seed required to sample from $G^{(cs)}_{m,n/t} \otimes G_{k,n/t,m}$.

Thus, we can use Theorem 7 to sample the seed for the different buckets and since $D = O(\log^{3/2} n)$, the total seed length is $O(\log m + \log^{3/2}(n/\epsilon))$. Additionally, the cost of sampling $h$ is $O(\log^{3/2} n)$, thus the final seed length remains $O(\log m + \log^{3/2}(n/\epsilon))$. What remains to be proven is that the simpler PRG described here (i.e. one before applying Theorem 7) $\epsilon$-fools $f \in \mathsf{Csum}(m,n)$ when $\mathrm{Var}(Z) > 2^{c\cdot\sqrt{\log n}}$.

Define $Z_{f,h^{-1}(i)} = \sum_{j\in h^{-1}(i)} f_j(y_j)$ and $Z'_{f,h^{-1}(i)} = \sum_{j\in h^{-1}(i)} f_j(y'_j)$ where $y'_1,\ldots,y'_n$ are sampled from the output of the PRG. Observe that the random variables $\{Z'_{f,h^{-1}(i)}\}_{i=1}^t$ and $\{Z_{f,h^{-1}(i)}\}_{i=1}^t$ are sequences of independent random variables. Further, by our construction, we have that for any $j \in [1,\ldots,k]$, the $j^{th}$ moment of $Z'_{f,h^{-1}(i)}$ is identical to that of $Z_{f,h^{-1}(i)}$ for $i \in [t]$. In fact, because our PRG *contains an independent copy of* $G^{(cs)}_{m,n/t}$, it ensures the Fourier spectrum of $Z'_{f,h^{-1}(i)}$ is also $\delta$-close to the Fourier spectrum $Z_{f,h^{-1}(i)}$ (for $i \in [t]$) in $\ell_\infty$ distance. An application of concentration bounds for sums of $k$-wise independent random variables implies that with probability $1 - n^{-\theta(1)}$ for $h \sim H_{k,n,t}$, for all $i \in [t]$,

$$\frac{\mathrm{Var}(Z)}{2\cdot t} \le \mathrm{Var}\big(Z_{f,h^{-1}(i)}\big) \le \frac{3\cdot\mathrm{Var}(Z)}{2\cdot t}.$$

It turns out that the above conditions are sufficient to apply the asymptotic expansions theorem and imply that

$$\left\| \sum_{i=1}^{t} Z'_{f,h^{-1}(i)} - \sum_{i=1}^{t} Z_{f,h^{-1}(i)} \right\|_1 \approx t^{-\Omega(k)} \le \epsilon.$$

This concludes our informal description of our PRG. We refer the reader to the full paper for the details.

We note that subsequent to our work, Gopalan, Kane and Meka [46], gave nearly optimal PRGs for combinatorial shapes.

## IV. Asymptotic expansion for sums of independent random variables

The plan for this section is as follows. We will start by setting up some notation for stating the main theorem followed by stating the main theorem. Then, we prove the main technical lemma required in subsection IV-D. This technical lemma shall be used to complete the proof in subsection IV-E.

### A. Preliminaries for the asymptotic expansion

Even though some of the notation was already used in the introduction to state the main theorem, we repeat it here to refresh the memory of the readers. For any random variable $X$ supported on $\mathbb{R}$,

- $\widehat{X} : \mathbb{R} \to \mathbb{C}$ denote its characteristic function. In other words, $\widehat{X}(\xi) = \mathbf{E}_{x\sim X}[e^{i\cdot\xi\cdot x}]$.
- $\alpha_{X,k} = \mathbf{E}[X^k]$, $\beta_{X,k} = \mathbf{E}[|X|^k]$. $\gamma_{X,k}$ denotes the $k^{th}$ cumulant of $X$.

We now let $X_1,\ldots,X_n$ be $n$ independent random variables. We will use the following shorthands.

- The variables $X_1,\ldots,X_n$ are centered.
- $\alpha_{i,k}$ will denote $\alpha_{X_i,k}$. Likewise for $\beta_{i,k}$ and $\gamma_{i,k}$.
- Also, $\beta_k = \sum_{i=1}^n \beta_{i,k}$.
- $\sigma^2 = \sum_{i=1}^n \alpha_{i,2}$ and $Z = (X_1 + \ldots + X_n)/\sigma$.

Let us also define $L_k = (\beta_k/\sigma^k)^{1/(k-2)}$. For $k \in \mathbb{N}$, we define $I_k$ to be

$$I_k = \frac{1}{C}\cdot\min\left\{\min_i \frac{\sigma}{\sigma_i}, \frac{1}{L_{3k}}\right\}, \tag{1}$$

for a sufficiently large constant $C > 0$. Choosing $C = 10^6$ suffices for our purposes. Let us also define $Z = (X_1+\ldots+X_n)/\sigma$. Having defined $Z$, we recall the definition of the family of polynomials $\{P_\nu(\xi)\}_{\nu\in\mathbb{N}}$. $P_\nu(\xi)$ is defined to be the coefficient of $w^{\nu+2}$ in the formal expansion of

$$\exp\left(\sum_{j=1}^{\infty} \frac{\lambda_{j+2}\cdot\xi^{j+2}\cdot w^j}{(j+2)!}\right),$$

where $\lambda_{j+2}$ is the $(j+2)^{th}$ cumulant of $Z$. As we said before, the description of $P_\nu$ requires the existence of only the first $\nu+2$ moments, or equivalently, cumulants of $Z$. We now restate Theorem 4.

*Theorem 2:* Let $X_1,\ldots,X_n$ be independent centered random variables supported on lattices of span 1. Further, let $Z = (X_1+\ldots+X_n)/\sigma$. Note that $Z$ is supported on a lattice with span $1/\sigma$. Call the lattice $L_Z$. Let us assume that $I \ge 2\sqrt{s\cdot\log s}$. Then,

$$\left|\Pr[Z=z] - \frac{1}{2\pi\sigma}\cdot\int_{-\pi\sigma}^{\pi\sigma} e^{-i\xi z}\cdot e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3} P_\nu(i\xi)\right)d\xi\right| \le \eta_{\mathsf{low}} + \eta_{\mathsf{med}} + \eta_{\mathsf{high}}$$

where

$$\eta_{\mathsf{low}} = s^{O(s)} \cdot L_{3s}^{s-2}, \quad \eta_{\mathsf{med}} = e^{-\frac{I^2}{6}} + s^{O(s)} \cdot e^{-\frac{I^2}{4}} \text{ and } \eta_{\mathsf{high}} = \sup_{|\zeta| \in [\sigma^2/(10 \cdot \beta_3), \pi]} \left| \prod_{i=1}^{n} \widehat{X_i}(\zeta) \right| + s^{O(s)} \cdot e^{-\frac{I^2}{4}}.$$

*B. Description of* $\{P_\nu\}$

While we have already described $P_\nu$, we would like to provide some intuition for the definition of $P_\nu$. To do this, we recall the definition of the cumulant generating function $\widetilde{X}$ defined as $\widetilde{X}(\xi) = \log \widehat{X}(\xi)$. Note that this definition means that $\widetilde{X}$ may not be defined on the entire $\mathbb{R}$. However, it is easy to see that for every $X$, there exists $c > 0$, such that $\widetilde{X}$ is defined on $[-c, c]$. We recall the following fact connecting the $k^{th}$ derivative of $\widetilde{X}$ and the $\gamma_{X,k}$.

*Fact 3:* The $k^{th}$ derivative of $\widetilde{X}(\xi)$ exists at $\xi = 0$ if and only if $\gamma_{X,k}$ is finite. Further,

$$\gamma_{X,k} = (-i)^k \cdot \frac{d^k \widetilde{X}(\xi)}{d\xi^k}\bigg|_{\xi=0}$$

Now by definition, we have that $\widetilde{Z}(\xi) = \sum_{i=1}^{n} \widetilde{X_i}(\xi/\sigma)$. For the moment, assume that the Maclaurin expansion of $\widetilde{X_i}$ exists in interval $[-c_i, c_i]$ for some $c_i > 0$. If $c = \min c_i$, then we see that the Maclaurin series expansion of $Z$ exists in the interval $[-\sigma c, \sigma c]$. Further, for $\xi \in [-\sigma c, \sigma c]$, we have

$$\widetilde{Z}(\xi) = \sum_{j=1}^{n} \widetilde{X_j}(\xi/\sigma) = \sum_{\nu=1}^{\infty} \sum_{j=1}^{n} \frac{\gamma_{j,\nu} \cdot i^\nu \cdot \xi^\nu}{\sigma^\nu \cdot \nu!}$$

Next, we notice that $\gamma_{j,1} = 0$ for all $j \in [n]$. Also, $\sum_{j=2}^{n} \gamma_{j,2} = \sigma^2$. This result in the simplified expression

$$\widetilde{Z}(\xi) = -\frac{\xi^2}{2} + \sum_{\nu=1}^{\infty} \frac{\gamma_{\nu+2} \cdot (i\xi)^{\nu+2}}{(\nu+2)! \cdot \sigma^{\nu+2}},$$

where $\gamma_{\nu+2} = \sum_{j=1}^{n} \gamma_{j,\nu+2}$. This implies

$$\widehat{Z}(\xi) = e^{-\frac{\xi^2}{2}} \cdot \exp\left( \sum_{\nu=1}^{\infty} \frac{\gamma_{\nu+2} \cdot (i\xi)^{\nu+2}}{(\nu+2)! \cdot \sigma^{\nu+2}} \right) = e^{-\frac{\xi^2}{2}} \cdot \exp\left( \sum_{\nu=1}^{\infty} \frac{\lambda_{\nu+2} \cdot (i\xi)^{\nu+2}}{(\nu+2)!} \right),$$

where $\lambda_{\nu+2}$ is $(\nu+2)^{th}$ cumulant of $Z$. By defining $P_\nu(i\xi)$ to denote the coefficient of $w^\nu$ in the formal expansion of

$$\exp\left( \sum_{j=1}^{\infty} \frac{\lambda_{j+2} \cdot (i\xi)^{j+2}}{(j+2)!} \cdot w^j \right),$$

we see that we get

$$\widehat{Z}(\xi) = e^{-\frac{\xi^2}{2}} \cdot \left( 1 + \sum_{\nu=1}^{\infty} P_\nu(i\xi) \right).$$

Thus, intuitively truncating the series $\sum_{\nu=1}^{\infty} P_\nu(i\xi)$ is likely to give us a good approximation to $\widehat{Z}(\xi)$.

*C. Alternate formulation for* $\{P_\nu\}$

We now seek an alternate formulation of the polynomials $P_\nu$ which will be useful to us later on. Towards this, we make the following definition:

$$V_w(\xi) = \frac{\xi^2}{2} + \frac{1}{w^2} \cdot \log \widehat{Z}(\xi \cdot w) \tag{2}$$

Note that in the introduction, we had set $w = 1$, for the ease of description to describe $V(\xi)$. With this, we make the following claim.

*Claim 4:* For $V_w(\xi)$ defined as above, the following holds:

$$\sum_{\nu=1}^{s-3} P_\nu(i\xi) \cdot w^\nu = \sum_{j=1}^{s-3} \sum_{k=0}^{2j+s-3} \frac{d^k V_w^j(\xi)}{d\xi^k}\bigg|_{\xi=0} \cdot \frac{\xi^k}{k!} \tag{3}$$

We begin with a few elementary claims about $V_w$. First, note that by definition, we have

$$\left.\frac{d^k \log \widehat{Z}(\xi \cdot w)}{d\xi^k}\right|_{\xi=0} = \sum_{j=1}^{n} \left.\frac{d^k \log \widehat{X_j}(\xi \cdot w/\sigma)}{d\xi^k}\right|_{\xi=0} = \sum_{j=1}^{n} \frac{w^k \cdot \gamma_{j,k} \cdot i^k}{\sigma^k} = \frac{w^k \cdot \lambda_k \cdot i^k}{\sigma^k}$$

The last but one equality follows by definition of cumulants. Using this, we easily have the following equations.

$$\text{For } k \leq 2, \quad V^{(k)}(0) = 0. \tag{4}$$

$$\text{For } k > 2, \quad V^{(k)}(0) = \frac{w^{k-2} \cdot \lambda_k \cdot i^k}{\sigma^k}. \tag{5}$$

Towards achieving the characterization of $P_\nu(\xi)$, we note that,

$$P_\nu(i\xi) \cdot w^\nu = S_{1,\nu} + \ldots + S_{\nu,\nu}, \tag{6}$$

where

$$S_{\ell,\nu} = \sum_{\nu_1+\ldots+\nu_\ell=\nu:\, \prod \nu_j > 0} \prod_{j=1}^{\ell} \frac{\lambda_{\nu_j+2} \cdot (i\xi)^{\nu_j+2}}{(\nu_j+2)! \cdot \sigma^{\nu_j+2}} \cdot w^{\nu_j}$$

Note that in the above, different permutations of the tuple $(\nu_1, \ldots, \nu_i)$ are counted as distinct. We then have the following claim.

*Claim 5:*

$$S_{i,\nu} = \left.\frac{d^{\nu+2i} V_w^i(\xi)}{d\xi^{\nu+2i}}\right|_{\xi=0} \cdot \frac{\xi^{\nu+2i}}{(\nu+2i)!}$$

### *D. Approximation of the Fourier spectrum using moments*

We now state the main lemma of this section which is also the main workhorse for proving Theorem 4.

*Lemma 6:* For $|\xi| \leq I_s$,

$$\left|\widehat{Z}(\xi) - e^{-\frac{\xi^2}{2}} \cdot \left(1 + \sum_{\nu=1}^{s-3} P_\nu(i\xi)\right)\right| \leq c(s) \cdot L_{3s}^{s-2} \cdot \left(|\xi|^s + |\xi|^{3s-8)}\right) \cdot e^{-2\xi^2/5}$$

where $c(s) = 2^{O(s)}$.

We will need a few preliminaries before we start with the proof of Lemma 6. The following inequality follows easily from monotonicity of norms.

*Claim 7:* Let $X$ be a real-valued random variable. Assuming that $\beta_{X,k_1+k_2}$ exists, $\beta_{X,k_1} \cdot \beta_{X,k_2} \leq \beta_{X,k_1+k_2}$.

The following lemma states that $\{L_k\}$ is a monotonically non-decreasing sequence.

*Claim 8:* $\{L_k\}$ is a monotonically non-decreasing sequence.

We will also need a bound on $P_\nu(i\xi)$. Towards that, we establish the following simple bounds.

*Claim 9:* Let $X$ be any centered random variable. Then, for any $k$, (with the notations as before), $|\gamma_{X,k}| \leq 2^k \cdot \beta_{X,k}$.

*Claim 10:* For $Z$ defined as above, $\frac{|\lambda_j|}{\sigma^j} \leq 2^j \cdot j^j$.

*Proof:* We consider the random variable $Z' = \sigma \cdot Z = X_1 + \ldots + X_n$. Note that $\lambda_j = \gamma_{Z',j}$. Thus, using Claim 9, it suffices to bound $\beta_{Z',j}$. Assume for the moment that $j$ is even. For an integer $j$, let $P(j)$ denote the set of all partitions of $j$ none of which are 1. Then, we have

$$\begin{aligned}
\mathbf{E}[|X_1 + \ldots + X_n|^j] &\leq \sum_{(a_1,\ldots,a_t)\in P(j)} \binom{j}{a_1\ a_2\ \ldots\ a_t} \cdot \prod_{i=1}^{t} \mathbf{E}\left[\sum_j |X_j|^{a_i}\right] \\
&= \sum_{(a_1,\ldots,a_t)\in P(j)} \binom{j}{a_1\ a_2\ \ldots\ a_t} \cdot \prod_{i=1}^{t} \sigma^{a_i} \cdot L_{a_i}^{a_i-2} \\
&\leq \sigma^j \cdot j^j.
\end{aligned}$$

Note that the last inequality uses that $L_t \leq 1$ for all $t > 2$. This finishes the proof for even $j$. For odd $j$, it follows from monotonicity of norms. ∎

We next make the following claim which bounds the value of $P_\nu(i\xi)$.

*Claim 11:*

$$|P_\nu(i\xi)| \le 2^{O(\nu)} \cdot (|\xi|^{\nu+2} + |\xi|^{3\nu})$$

*Proof:* Begin by noting that

$$P_\nu(i\xi) = \sum_{\ell=1}^{\nu} \sum_{\nu_1+\ldots+\nu_\ell=\nu:\prod \nu_j>0} \prod_{j=1}^{\ell} \frac{\lambda_{\nu_j+2} \cdot (i\xi)^{\nu_j+2}}{(\nu_j+2)! \cdot \sigma^{\nu_j+2}}$$

Thus,

$$|P_\nu(i\xi)| \le \sum_{\ell=1}^{\nu} \sum_{\nu_1+\ldots+\nu_\ell=\nu:\prod \nu_j>0} \prod_{j=1}^{\ell} \frac{|\lambda_{\nu_j+2}| \cdot |\xi|^{\nu_j+2}}{(\nu_j+2)! \cdot \sigma^{\nu_j+2}} \le \sum_{\ell=1}^{\nu} \sum_{\nu_1+\ldots+\nu_\ell=\nu:\prod \nu_j>0} \prod_{j=1}^{\ell} (2e)^{\nu_j+2} \cdot |\xi|^{\nu_j+2},$$

where the last inequality uses Claim 10. The final term can be easily bound by $2^{O(\nu)} \cdot (|\xi|^{\nu+2} + |\xi|^{3\nu})$. ■

As a result, we also get the following corollary.

*Corollary 12:*

$$\left| e^{-\frac{\xi^2}{2}} \cdot \left(1 + \sum_{\nu=1}^{s-3} P_\nu(i\xi)\right) \right| \le 2^{O(s)} \cdot e^{-\xi^2/2} \cdot (|\xi|^3 + |\xi|^{3s}).$$

*Proof:* of Lemma 6 We start by showing that for $|\xi| \le I_s$, the function $V_w(\xi)$ is well-defined. Note that $V_w(\xi)$ is well defined as long as $\widehat{Z}(\xi \cdot w) \neq 0$. The following claim states the interval in which this indeed holds and thus $V_w(\xi)$ is well defined.

*Claim 13:* For $|\xi| \le \frac{\sigma}{\sigma_i}$, $|\widehat{X_i}(\xi \cdot w)| \ge 1/2$.

*Proof:* Using Taylor's theorem (for complex valued functions), we have

$$\left| \widehat{X}_i\left(\frac{\xi \cdot w}{\sigma}\right) - 1 \right| = \left| \mathbf{E}\left[ \exp\left(\frac{i \cdot w \cdot \xi \cdot X_i}{\sigma}\right) - 1 - \frac{i \cdot \xi \cdot w \cdot X_i}{\sigma} \right] \right| \le \frac{\xi^2 w^2 \cdot \mathbf{E}[X_i^2]}{2 \cdot \sigma^2}.$$

For $|\xi| \le \frac{\sigma}{\sigma_i}$, the right hand side is at most $1/2$, finishing the proof. ■

*Claim 14:* For $|\xi| \le \min_i \frac{\sigma}{\sigma_i}$, $\widehat{Z}(\xi \cdot w) \neq 0$.

*Proof:* Note that $\widehat{Z}(\xi \cdot w) = \prod_{i=1}^{n} \widehat{X_i}((\xi \cdot w)/\sigma)$. Using Claim 13, we get the claim. ■

We now establish an upper bound on $V_w(\xi)$. For this, our strategy is to use Taylor's theorem around $\xi = 0$. In particular, we will show the $V_w$ and its first two derivatives are zero and then establish a bound on the third order derivative. Using Taylor's theorem, this will lead us to an upper bound on $V_w(\xi)$. We start with the following simple claims. Having established that the function $V_w(\xi)$ as well as its first and second order derivatives vanish at 0, we now establish a bound on the supremum of the third order derivative. To do this, we first prove the following simple claim.

*Claim 15:* For any $\xi \in \mathbb{R}$, $\left|\frac{\partial^k \widehat{X_i}(\xi)}{\partial \xi^k}\right| \le \beta_{i,k}$.

*Proof:* Note that $\widehat{X_i}(\xi) = \mathbf{E}_{x \in X_i}[\exp(ix\xi)]$. As a consequence, we get that

$$\frac{\partial^k \widehat{X_i}(\xi)}{\partial \xi^k} = \mathbf{E}_{x \in X_i}[(i \cdot x)^k \exp(ix\xi)].$$

This immediately implies the claim. ■

To establish the upper bound on the third order derivative of $V_w(\xi)$, we establish the more general bound on the $s^{th}$ order derivative for $s > 2$. This will be useful later on. For this, we observe the following simple fact.

$$\frac{d^k \log u}{dx^k} = \sum_{i_1,\ldots,i_k:\sum j \cdot i_j = k} c_{i_1,\ldots,i_k} \cdot \frac{1}{u^{\|(i_1,\ldots,i_k)\|_1}} \prod_{\ell=1}^{k} \left(\frac{d^\ell u}{dx^\ell}\right)^{i_\ell} \tag{7}$$

where a simple induction can be used to show that

$$\sum_{i_1,\ldots,i_k:\sum j \cdot i_j = k} |c_{i_1,\ldots,i_k}| \le 2^k \cdot k!. \tag{8}$$

*Claim 16:* For $s > 2$ and $|\xi| \le I_s$,

$$\left|\frac{d^s V_w(\xi)}{d\xi^s}\right| \le 4^s \cdot s! \cdot w^{s-2} \cdot L_s^{s-2}.$$

*Proof:*

$$\begin{aligned}\frac{d^s V_w(\xi)}{d\xi^s} &= \frac{1}{w^2}\cdot\left(\sum_{\ell=1}^{n}\frac{d^s \log \widehat{X_\ell}(\xi\cdot w/\sigma)}{d\xi^s}\right)\\ &= \frac{1}{w^2}\cdot\left(\sum_{\ell=1}^{n}\sum_{i_1,\ldots,i_s:\sum j\cdot i_j=s} c_{i_1,\ldots,i_s}\cdot\frac{1}{\widehat{X_\ell}(\xi\cdot w/\sigma)^{\|(i_1,\ldots,i_s)\|_1}}\cdot\prod_{j=1}^{s}\left(\frac{d^j\widehat{X_\ell}(\xi\cdot w/\sigma)}{d\xi^j}\right)^{i_j}\right)\end{aligned}$$

The last equality uses (7) and (8). Using Claim 15 and Claim 13, we get

$$\left|\frac{d^s V_w(\xi)}{d\xi^s}\right| \le \frac{1}{w^2}\cdot\left(\sum_{\ell=1}^{n}\sum_{i_1,\ldots,i_s:\sum j\cdot i_j=s}|c_{i_1,\ldots,i_s}|\cdot\frac{1}{|\widehat{X_\ell}(\xi\cdot w/\sigma)|^{\|(i_1,\ldots,i_s)\|_1}}\cdot\prod_{j=1}^{s}\left|\left(\frac{d^j\widehat{X_\ell}(\xi\cdot w/\sigma)}{d\xi^j}\right)\right|^{i_j}\right)$$

Note that $|\widehat{X_\ell}(\xi\cdot w/\sigma)|\ge 1/2$ for $|\xi|\le I_s$ (Claim 13), we get

$$\begin{aligned}\left|\frac{d^s V_w(\xi)}{d\xi^s}\right| &\le \frac{1}{w^2}\cdot\left(\sum_{\ell=1}^{n}\sum_{i_1,\ldots,i_s:\sum j\cdot i_j=s}|c_{i_1,\ldots,i_s}|\cdot 2^{\|(i_1,\ldots,i_s)\|_1}\cdot\prod_{j=1}^{s}\left|\left(\frac{d^j\widehat{X_\ell}(\xi\cdot w/\sigma)}{d\xi^j}\right)\right|^{i_j}\right)\\ &\le \frac{1}{w^2}\cdot\left(\sum_{\ell=1}^{n}\sum_{i_1,\ldots,i_s:\sum j\cdot i_j=s}|c_{i_1,\ldots,i_s}|\cdot 2^{\|(i_1,\ldots,i_s)\|_1}\cdot\prod_{j=1}^{s}\frac{\beta_{\ell,j}^{i_j}\cdot w^{j\cdot i_j}}{\sigma^{j\cdot i_j}}\right)\ \text{(Claim 15)}\\ &\le \frac{1}{w^2}\cdot\frac{w^s}{\sigma^s}\left(\sum_{\ell=1}^{n}\sum_{i_1,\ldots,i_s:\sum j\cdot i_j=s}|c_{i_1,\ldots,i_s}|\cdot 2^{\|(i_1,\ldots,i_s)\|_1}\cdot\prod_{j=1}^{s}\beta_{\ell,j}^{i_j}\right)\\ &\le \frac{w^{s-2}}{\sigma^s}\left(\sum_{\ell=1}^{n}\sum_{i_1,\ldots,i_s:\sum j\cdot i_j=s}|c_{i_1,\ldots,i_s}|\cdot 2^{\|(i_1,\ldots,i_s)\|_1}\cdot\beta_{\ell,s}\right)\ \text{(Claim 7)}\\ &\le \frac{w^{s-2}\cdot\beta_s}{\sigma^s}\cdot 2^s\cdot 2^s\cdot s!\le 4^s\cdot s!\cdot w^{s-2}\cdot L_s^{s-2}.\end{aligned}$$

■

Using this, we have the following corollary.

$$\text{For } |\xi|\le I_s,\ \left|\frac{d^3V_w(\xi)}{d\xi^3}\right|\le 64\cdot 6\cdot w\cdot L_3. \tag{9}$$

Also, repating nearly the same calculation as Claim 16, we also have

$$\text{For } |\xi|\le I_s,\ \left|\frac{d^2V_w(\xi)}{d\xi^2}\right|\le 8! \tag{10}$$

Equipped with this inequality, we now establish an upper bound $V_w(\xi)$. Using Taylor's theorem and (4), we have that for $|\xi|\le I_s$,

$$|V_w(\xi)|\le\frac{1}{6}|\xi^3|\sup_{|\xi'|\le|\xi|}\left|\frac{d^3V_w(\xi')}{d\xi'^3}\right|\le\Theta(1)\cdot w\cdot|\xi|^3\cdot L_3 \tag{11}$$

Likewise, we have,

$$\left|\frac{dV_w(\xi)}{d\xi}\right|\le\Theta(1)\cdot w\cdot|\xi|^2\cdot L_3\quad\text{and}\quad\left|\frac{d^2V_w(\xi)}{d\xi^2}\right|\le\Theta(1)\cdot w\cdot|\xi|\cdot L_3. \tag{12}$$

By choosing $C>0$ sufficiently large in (1) and using $L_{3s}>L_3$, we also get that

$$|V_w(\xi)|\le\frac{w\cdot\xi^2}{10},\quad\left|\frac{dV_w(\xi)}{d\xi}\right|\le\frac{3w\cdot\xi}{10}\quad\text{and}\quad\left|\frac{d^2V_w(\xi)}{d\xi^2}\right|\le\frac{3w}{5}. \tag{13}$$

Note that by definition, we have

$$\left(\widehat{X}(\xi\cdot w)\right)^{1/w^2}=e^{-\xi^2/2}\cdot e^{V_w(\xi)}.$$

We seek to control the quantity

$$\left| e^{-\frac{\xi^2}{2}} \cdot e^{V_w(\xi)} - e^{-\frac{\xi^2}{2}} \cdot \left( 1 + \sum_{\nu=1}^{s-3} w^\nu \cdot P_\nu(i\xi) \right) \right|$$

and then finally put $w = 1$. To control this quantity, we break this the difference into two parts.

$$R_1(w,\xi) = \sum_{\nu=s-2}^{\infty} \frac{e^{-\frac{\xi^2}{2}} \cdot V_w^\nu(\xi)}{\nu!} \qquad R_2(w,\xi) = e^{-\frac{\xi^2}{2}} \cdot \left( \sum_{\nu=1}^{s-3} \left( \frac{V_w^\nu(\xi)}{\nu!} - \frac{w^\nu \cdot P_\nu(i\xi)}{\nu!} \right) \right)$$

It is easier to control $R_1(w,\xi)$, so we begin with that.

$$\begin{aligned} |R_1(w,\xi)| &\leq \sum_{\nu=s-2}^{\infty} \frac{e^{-\frac{\xi^2}{2}} \cdot |V_w(\xi)|^\nu}{\nu!} \leq \frac{e^{-\frac{\xi^2}{2}} \cdot |V_w(\xi)|^{s-2}}{(s-2)!} \cdot e^{|V_w(\xi)|} \\ &\leq \frac{|V_w(\xi)|^{s-2}}{(s-2)!} \cdot e^{-\frac{\xi^2}{2} + \frac{w\xi^2}{10}} \leq \frac{2^{O(s)} \cdot w^{s-2} \cdot \xi^{3(s-2)} \cdot L_3^{s-2}}{(s-2)!} \cdot e^{-\frac{\xi^2}{2} + \frac{w\xi^2}{10}}. \end{aligned} \tag{14}$$

The penultimate inequality uses (13) and the last inequality uses (11). We will now control $R_2(w,\xi)$ which is slightly more tricky to control. Our calculation for this part is somewhat different from those in Petrov [15] or Bhattacharya and Rao [20]. In particular, these two works truncate after $\nu = s-3$ (i.e. do an approximation in terms of the first $s-1$ moments). However, naively going through their calculations, it seems one needs to pay a factor of $s^{O(s^2)}$. On the other hand, we pay a factor of $2^{O(s)}$ but instead need to assume that the first $3s+2$ moments exists and the error bound is in terms of $L_{3s}$. Towards bounding $R_2(w,\xi)$, we have

$$\begin{aligned} R_2(w,\xi) &= e^{-\frac{\xi^2}{2}} \cdot \sum_{\nu=1}^{s-3} \left( \frac{V_w^\nu(\xi)}{\nu!} - \frac{w^\nu \cdot P_\nu(i\xi)}{\nu!} \right) \\ &= e^{-\frac{\xi^2}{2}} \cdot \sum_{\nu=1}^{s-3} \frac{1}{\nu!} \cdot \left( V_w^\nu(\xi) - \sum_{k=0}^{2\nu+s-3} \left. \frac{d^k V_w^\nu(\xi)}{d\xi^k} \right|_{\xi=0} \cdot \frac{\xi^k}{k!} \right) \end{aligned}$$

To get the second equality, we use (3). We next use Taylor's theorem to get

$$|R_2(w,\xi)| \leq \sum_{\nu=1}^{s-3} \frac{e^{-\frac{\xi^2}{2}}}{\nu!} \cdot \frac{|\xi|^{2\nu+s-2}}{(2\nu+s-2)!} \kappa_{\nu,2\nu+s-2}(\xi), \tag{15}$$

where

$$\kappa_{\nu,2\nu+s-2}(\xi) = \sup_{|\xi| \leq I_s} \left| \frac{d^{2\nu+s-2} V_w^\nu(\xi)}{d\xi^{2\nu+s-2}} \right|.$$

Thus our task reduces to bounding $\kappa_{\nu,2\nu+s-2}(\xi)$ for $\nu \in [1,\ldots,s-3]$. To bound this number, we recall the following basic fact about higher order derivatives of products of functions. Let $\mathbb{Z}^+ = \mathbb{N} \cup \{0\}$. Using the Leibniz rule, we have the following:

*Lemma 17:*

$$\frac{d^k \prod_{i=1}^{\ell} u_i}{dx^k} = \sum_{a \in \mathbb{Z}^{+\ell} : \|a\|_1 = k} \binom{k}{a_1,\ldots,a_\ell} \prod_{i=1}^{\ell} \frac{d^{a_i} u_i}{dx^{a_i}}$$

Using Lemma 17, we have

$$\begin{aligned} \left| \frac{d^{2\nu+s-2} V_w^\nu(\xi)}{d\xi^{2\nu+s-2}} \right| &= \left| \sum_{a \in \mathbb{Z}^{+\nu} : \|a\|_1 = 2\nu+s-2} \binom{2\nu+s-2}{a_1,\ldots,a_\nu} \prod_{i=1}^{\nu} \frac{d^{a_i} V_w(\xi)}{d\xi^{a_i}} \right|, \\ &\leq \sum_{a \in \mathbb{Z}^{+\nu} : \|a\|_1 = 2\nu+s-2} \binom{2\nu+s-2}{a_1,\ldots,a_\nu} \prod_{i=1}^{\nu} \left| \frac{d^{a_i} V_w(\xi)}{d\xi^{a_i}} \right|, \\ &= \sum_{a \in \mathbb{Z}^{+\nu} : \|a\|_1 = 2\nu+s-2} \binom{2\nu+s-2}{a_1,\ldots,a_\nu} \prod_{i : a_i \in \{0,1\}} \left| \frac{d^{a_i} V_w(\xi)}{d\xi^{a_i}} \right| \prod_{i : a_i \geq 2} \left| \frac{d^{a_i} V_w(\xi)}{d\xi^{a_i}} \right|. \end{aligned} \tag{16}$$

Using Claim 16 and (10), we recall that for $a_i > 1$ and $|\xi| \le I_s$,

$$\left|\frac{d^{a_i} V_w(\xi)}{d\xi^{a_i}}\right| \le 2^{a_i} \cdot (a_i)! \cdot L_{a_i}^{a_i-2} \cdot w^{a_i-2}.$$

On the other hand, for $a_i \in \{0,1\}$ and $|\xi| \le I_s$, using (13) we have,

$$\left|\frac{d^{a_i} V_w(\xi)}{d\xi^{a_i}}\right| \le \frac{w \cdot \xi^{2-a_i}}{10}.$$

Applying the last two inequalities to (16),

$$\begin{aligned}\left|\frac{d^{2\nu+s-2} V_w^{\nu}(\xi)}{d\xi^{2\nu+s-2}}\right| &\le \sum_{a \in \mathbb{Z}^{+\nu}: \|a\|_1 = 2\nu+s-2} \binom{2\nu+s-2}{a_1,\ldots,a_\nu} \prod_{i:a_i \in \{0,1\}} \frac{w \cdot \xi^{2-a_i}}{10} \cdot \prod_{i:a_i \ge 2} 2^{a_i} \cdot (a_i)! \cdot L_{a_i}^{a_i-2} \cdot w^{a_i-2}, \\ &\le \sum_{a \in \mathbb{Z}^{+\nu}: \|a\|_1 = 2\nu+s-2} \binom{2\nu+s-2}{a_1,\ldots,a_\nu} \prod_{i:a_i \in \{0,1\}} \frac{w \cdot \xi^{2-a_i}}{10} \cdot \prod_{i:a_i \ge 2} 2^{a_i} \cdot (a_i)! \cdot L_{2\nu+s-2}^{a_i-2} \cdot w^{a_i-2}\end{aligned}$$

The last inequality uses that $L_{a_i} \le L_{\|a\|_\infty} \le L_{\|a\|_1}$ (using Claim 8). For any given $a \in \mathbb{Z}^{+\nu}$, use $\#a(0)$ to denote the number of zero entries in $a$ and $\#a(1)$ to denote the one entries. Note that for any term in the above summation, we have

- The exponent of $w$ is $\|a\|_1 - 2\|a\|_0 + 2\#a(1) + \#a(0)$.
- The exponent of $L_{2\nu+s-2}$ is $\|a\|_1 - 2\|a\|_0 + \#a(1)$.
- The exponent of $\xi$ is $2\#a(0) + \#a(1)$.

Note that $|\xi| \le L_{3s}^{-1}$, using Claim 7, we get that $|\xi| \le L_{2\nu+s-2}^{-1}$. Thus,

$$L_{2\nu+s-2}^{\|a\|_1 - 2\|a\|_0 + \#a(1)} \cdot |\xi|^{2\#a(0)+\#a(1)} \le L_{2\nu+s-2}^{\|a\|_1 - 2\|a\|_0 + \#a(1) - 2\#a(0) - \#a(1)} \le L_{2\nu+s-2}^{s-2} \le L_{3s}^{s-2}.$$

Using the above and that $|w| \le 1$, we get

$$\left|\frac{d^{2\nu+s-2} V_w^{\nu}(\xi)}{d\xi^{2\nu+s-2}}\right| \le (2\nu+s-2)! \cdot 2^{O(s)} \cdot L_{3s}^{s-2} \cdot w^{2\nu+s-2}.$$

Applying this bound in (15), we get

$$\begin{aligned}|R_2(w,\xi)| &\le e^{-\frac{\xi^2}{2}} \cdot \sum_{\nu=1}^{s-3} \frac{1}{\nu!} \cdot \frac{|\xi|^{2\nu+s-2}}{(2\nu+s-2)!} \cdot (2\nu+s-2)! \cdot 2^{O(s)} \cdot L_{3s}^{s-2} \cdot w^{2\nu+s-2}, \\ &\le e^{-\frac{\xi^2}{2}} \cdot w^s \cdot 2^{O(s)} \cdot L_{3s}^{s-2} \cdot (|\xi|^s + |\xi|^{3s-8}).\end{aligned}$$

Using the bound on $|R_1(w,\xi)|$ (and using that $|w| \le 1$), we get that

$$|R_1(w,\xi)| + |R_2(w,\xi)| \le 2^{O(s)} \cdot L_{3s}^{s-2} \cdot \left(e^{-\frac{\xi^2}{2}} \cdot (|\xi|^s + |\xi|^{3s-8}) + e^{-\frac{2\xi^2}{5}} \cdot |\xi|^{3s-6}.\right)$$

Finally, plugging in $w = 1$, in the above, we complete the proof. ∎

### *E. From Fourier closeness to $\ell_1$ closeness*

We start with some basics of Fourier analysis.

*Definition 3:* A distribution $\mathbf{p}$ is said to be supported on the lattice $\mathcal{L} = \{a + b \cdot h\}_{b \in \mathbb{Z}}$ if $\mathsf{supp}(\mathbf{p}) \subseteq \mathcal{L}$. If $h$ is the maximum possible number such that there exists a lattice $\mathcal{L}$ and $\mathsf{supp}(\mathbf{p}) \subseteq \mathcal{L}$, then $h$ is said to be the maximal span of the lattice and $a$ its offset.

We need a couple of more facts about Fourier transform of distributions.

*Fact 18:* Shifting the distribution by a quantity $\lambda$ multiplies the Fourier transform at point $\xi$ by $e^{i\xi\lambda}$.

Since our distributions will be supported on lattices, we first recall the Fourier inversion formula for probability distributions on lattices. In particular, let $\mathbf{p}$ be any distribution over $\mathcal{L}$. Then,

$$\widehat{\mathbf{p}}(\xi) = \int e^{i\xi t} d\mathbf{p}(t) = \sum_{\nu=-\infty}^{\infty} \mathbf{p}(\nu \cdot h + a) \cdot e^{i\xi(\nu \cdot h + a)}$$

As a consequence, we get

$$\frac{h}{2\pi}\cdot\int_{-\pi/h}^{\pi/h}\widehat{\mathbf{p}}(\xi)\cdot e^{-i\xi(a+\nu\cdot h)} \;=\; \frac{h}{2\pi}\cdot\sum_{\nu'=-\infty}^{\infty}\mathbf{p}(\nu'\cdot h+a)\cdot\int_{-\pi/h}^{\pi/h}e^{i\xi(\nu'-\nu)h}=\mathbf{p}(\nu\cdot h+a). \tag{17}$$

We now move to stating the main theorem of this section. For this, we assume that $X_1,\ldots,X_n$ are independent centered random variables such that for $i=1,\ldots,n$ supported on lattices of span 1. Further, let $Z=(X_1+\ldots+X_n)/\sigma$. Note that $Z$ is supported on a lattice with span $1/\sigma$. Call the lattice $\mathcal{L}_Z$. We now state our main theorem.

*Theorem 9:* Let $X_1,\ldots,X_n$ and $Z$ be as defined above and let $z\in\mathcal{L}_Z$. Let us assume that $I_s\geq 2\sqrt{s\cdot\log s}$. Then,

$$\left|\Pr[Z=z]-\frac{1}{2\pi\sigma}\cdot\int_{-\pi\sigma}^{\pi\sigma}e^{-i\xi z}\cdot e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3}P_\nu(i\xi)\right)d\xi\right|\leq\eta_{\mathsf{low}}+\eta_{\mathsf{med}}+\eta_{\mathsf{high}}$$

where

$$\eta_{\mathsf{low}}=s^{O(s)}\cdot L_{3s}^{s-2},\quad \eta_{\mathsf{med}}=e^{-\frac{I_s^2}{6}}+s^{O(s)}\cdot e^{-\frac{I_s^2}{4}}\text{ and }\eta_{\mathsf{high}}=\sup_{|\zeta|\in[\sigma^2/\beta_3,\pi]}\left|\prod_{i=1}^{n}\widehat{X_i}(\zeta)\right|+s^{O(s)}\cdot e^{-\frac{I_s^2}{4}}.$$

Now, consider the function

$$\widehat{P}(\xi)=e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3}P_\nu(i\xi)\right).$$

Given (17), it suffices to show that $\widehat{P}(\xi)$ is appropriately close to $\widehat{Z}(\xi)$ in the interval $[-\pi/h,\pi/h]$. We do this in the following proof.

*Proof:* of Theorem 9 We begin by noting that

$$\begin{aligned}&\left|\Pr[Z=z]-\frac{1}{2\pi\sigma}\cdot\int_{-\pi\sigma}^{\pi\sigma}e^{-i\xi z}\cdot e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3}P_\nu(i\xi)\right)d\xi\right|\\ =\;&\left|\frac{1}{2\pi\sigma}\cdot\int_{-\pi\sigma}^{\pi\sigma}e^{-i\xi z}\cdot\widehat{Z}(\xi)\cdot d\xi-\frac{1}{2\pi\sigma}\cdot\int_{-\pi\sigma}^{\pi\sigma}e^{-i\xi z}\cdot e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3}P_\nu(i\xi)\right)d\xi\right|\end{aligned}$$

Towards bounding the quantity on the right hand side, let us divide the interval $[-\pi\sigma,\pi\sigma]$ into three parts. Define $I_{\mathsf{low}}=[-I_s,I_s]$, $I_{\mathsf{med}}=[-\frac{1}{10\cdot L_3},\frac{1}{10\cdot L_3}]\setminus I_{\mathsf{low}}$, $I_{\mathsf{high}}=[-\pi\sigma,\pi\sigma]\setminus(I_{\mathsf{low}}\cup I_{\mathsf{med}})$. We control the errors in these regions separately. We define

$$\eta_{\mathsf{low}}=\left|\frac{1}{2\pi\sigma}\cdot\int_{\xi\in I_{\mathsf{low}}}e^{-i\xi z}\cdot\widehat{Z}(\xi)\cdot d\xi-\frac{1}{2\pi\sigma}\cdot\int_{\xi\in I_{\mathsf{low}}}e^{-i\xi z}\cdot e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3}P_\nu(i\xi)\right)d\xi\right|.$$

$\eta_{\mathsf{med}}$ and $\eta_{\mathsf{high}}$ are defined in an analogous manner.

*1) Bounding in $I_{\mathsf{low}}$: :* We begin by observing that for any $s\geq 2$,

$$\max_{\xi\in\mathbb{R}}\left(|\xi|^s+|\xi|^{(3s-8)}\right)\cdot e^{-\frac{2\xi^2}{5}}\leq s^{2s}. \tag{18}$$

This can be easily deduced from considering two cases: $|\xi|\leq 3\sqrt{s}$ and $|\xi|>3\sqrt{s}$. In the former case,

$$\left(|\xi|^s+|\xi|^{(3s-8)}\right)\cdot e^{-\frac{2\xi^2}{5}}\leq\left(|\xi|^s+|\xi|^{(3s-8)}\right)\leq s^{2s}.$$

In the latter case, both $|\xi|^s$ and $|\xi|^{(3s-8)}$ are bounded by $e^{\frac{2\xi^2}{5}}$ and hence

$$\left(|\xi|^s+|\xi|^{(3s-8)}\right)\cdot e^{-\frac{2\xi^2}{5}}\leq 2.$$

Armed with (18) and applying Lemma 6,

$$\begin{aligned}\left|\frac{1}{2\pi\sigma}\cdot\int_{-I_s}^{I_s}e^{-i\xi z}\cdot\widehat{Z}(\xi)\cdot d\xi-\frac{1}{2\pi\sigma}\cdot\int_{-I_s}^{I_s}e^{-i\xi z}\cdot e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3}P_\nu(i\xi)\right)d\xi\right| &\leq s^{O(s)}\cdot L_{3s}^{s-2}\cdot\frac{|I|}{\sigma}\\ &\leq s^{O(s)}\cdot L_{3s}^{s-2}.\end{aligned}$$

*2) Bounding in $I_{\mathsf{med}}$: :* For this, we begin by proving the following useful estimate.

*Lemma 19:* Let $X$ be a centered random variable with variance $\sigma^2 > 0$ and $\mathbf{E}[|X|^3] = \beta$. Then, for all $t$,

$$\left|\mathbf{E}[e^{iXt}]\right| \leq \exp\left(\frac{-t^2 \cdot \sigma^2}{2} + \frac{|t|^3 \cdot \beta}{3}\right).$$

*Proof:* Consider the random variable $Z = X - X'$ where $X'$ is an independent copy of $X$. Note that

$$\mathbf{E}[e^{iZt}] = \mathbf{E}[e^{iXt}] \cdot \mathbf{E}[e^{-iXt}] = |\mathbf{E}[e^{iXt}]|^2$$

Thus, for our purposes, it suffices to bound $\mathbf{E}[e^{iZt}]$. Also, the above shows that $0 \leq \mathbf{E}[e^{iZt}] \leq 1$. Using the elementary inequality,

$$\text{for all } 0 \leq x \leq 1, \quad \log x \leq x - 1,$$

we get that

$$\log \mathbf{E}[e^{iZt}] \leq \mathbf{E}[e^{iZt}] - 1. \tag{19}$$

Further, using that $\mathbf{E}[e^{iZt}]$ is real and the Taylor's theorem, we have

$$\mathbf{E}[e^{iZt}] - 1 = \mathbf{E}[\cos(Zt) - 1] \leq -\frac{t^2\mathbf{E}[Z^2]}{2} + \frac{1}{6}|t|^3\mathbf{E}[|Z|^3].$$

Combining this with (19), we have

$$\log \mathbf{E}[e^{iZt}] \leq -\frac{t^2\mathbf{E}[Z^2]}{2} + \frac{1}{6}|t|^3\mathbf{E}[|Z|^3] = -t^2\sigma^2 + \frac{1}{6}|t|^3\mathbf{E}[|Z|^3].$$

Thus, it remains to bound $\mathbf{E}[|Z|^3]$. To do this, note that

$$\mathbf{E}[|Z|^3] = \mathbf{E}[|X - X'|^3] \leq 2\mathbf{E}[|X| \cdot (X - X')^2] \leq 4\mathbf{E}[|X|^3].$$

Thus, we have

$$\log \mathbf{E}[e^{iZt}] \leq -t^2\sigma^2 + \frac{2}{3} \cdot |t|^3 \cdot \beta,$$

which finishes the proof. ∎

Thus, we get that that for the random variable $Z' = X_1 + \ldots + X_n$, we have

$$|\mathbf{E}[e^{iZ'\xi}]| \leq e^{-\frac{\xi^2\sigma^2}{2} + \frac{|\xi|^3\beta_3}{3}},$$

and hence

$$|\mathbf{E}[e^{iZ\xi}]| \leq e^{-\frac{\xi^2}{2} + \frac{|\xi|^3 \cdot L_3}{3}}.$$

Thus, for $|\xi| \leq L_3^{-1}$, $|\mathbf{E}[e^{iZ\xi}]| \leq e^{-\frac{\xi^2}{6}}$. Thus, combining this and Corollary 12,

$$\begin{aligned}
& \left|\frac{1}{2\pi\sigma} \cdot \int_{I_{\mathsf{med}}} e^{-i\xi z} \cdot \widehat{Z}(\xi) \cdot d\xi - \frac{1}{2\pi\sigma} \cdot \int_{-I}^{I} e^{-i\xi z} \cdot e^{-\frac{\xi^2}{2}} \cdot \left(1 + \sum_{\nu=1}^{s-3} P_\nu(i\xi)\right) d\xi\right| \\
\leq \; & \frac{1}{2\pi\sigma} \cdot \int_{I_{\mathsf{med}}} e^{\frac{-\xi^2}{6}} \cdot d\xi + \frac{1}{2\pi\sigma} \cdot \int_{I_{\mathsf{med}}} s^{O(s)} \cdot e^{-\frac{\xi^2}{2}} \cdot (|\xi|^3 + |\xi|^{3s}) \cdot d\xi \\
\leq \; & \frac{1}{2\pi\sigma} \cdot \int_{I_{\mathsf{med}}} e^{\frac{-\xi^2}{6}} \cdot d\xi + \frac{1}{2\pi\sigma} \cdot \int_{I_{\mathsf{med}}} s^{O(s)} \cdot e^{\frac{-\xi^2}{4}} \cdot d\xi.
\end{aligned}$$

The last inequality uses $I_s \geq 2\sqrt{s \cdot \log s}$ and hence for all $\xi \in I_{\mathsf{med}}$, $|\xi| \geq 2\sqrt{s \cdot \log s}$. This easily implies that $\eta_{\mathsf{med}} \leq e^{-\frac{I_s^2}{6}} + s^{O(s)} \cdot e^{-\frac{I_s^2}{4}}$. This leaves us with bounding $\eta_{\mathsf{high}}$.

*3) Bounding in $I_{\mathsf{high}}$: :*

$$\begin{aligned}&\left|\frac{1}{2\pi\sigma}\cdot\int_{I_{\mathsf{high}}}e^{-i\xi z}\cdot\widehat{Z}(\xi)\cdot d\xi-\frac{1}{2\pi\sigma}\cdot\int_{I_{\mathsf{high}}}e^{-i\xi z}\cdot e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3}P_\nu(i\xi)\right)d\xi\right|\\ \leq\ &\left|\frac{1}{2\pi\sigma}\cdot\int_{I_{\mathsf{high}}}e^{-i\xi z}\cdot\widehat{Z}(\xi)\cdot d\xi\right|+\left|\frac{1}{2\pi\sigma}\cdot\int_{I_{\mathsf{high}}}e^{-i\xi z}\cdot e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3}P_\nu(i\xi)\right)d\xi\right|.\end{aligned}$$

The latter summand (similar to the case for $\eta_{\mathsf{high}}$) can be bounded by $s^{O(s)}\cdot e^{-\frac{I_s^2}{4}}$. This leaves us with bounding the first summand. Let us call the first summand as $\mathsf{err}_{\mathsf{high}}$. Towards this, we observe that $\widehat{Z}(\xi)=\prod_{i=1}^n\widehat{X_i}(\xi/\sigma)$. As a result,

$$\mathsf{err}_{\mathsf{high}}\leq\sup_{\xi\in I_{\mathsf{high}}}|\widehat{Z}(\xi)|=\sup_{|\zeta|\in[\sigma^2/\beta_3,\pi]}\left|\prod_{i=1}^n\widehat{X_i}(\zeta)\right|.$$

This finishes our proof. ∎

It is instructive to see the corollary of this theorem in the setting of i.i.d. centered lattice valued random variables. Towards this, assume that $X_1,\ldots,X_n$ are i.i.d. lattice valued random variables of maximal span 1 (call the common distribution $X$). Further, let us use $\beta_{(k)}$ to denote $\mathbf{E}[|X|^k]$ and $\alpha_{(k)}$ to denote $\mathbf{E}[X^k]$. Further, let us assume that $X$ is $c(k)$-hypercontractive i.e. $\beta_{(k)}^{1/k}\leq c(k)\cdot\alpha_{(2)}^{1/2}$.

Using this notation,

$$L_k=\left(\frac{\beta_k}{\sigma^k}\right)^{\frac{1}{k-2}}=\left(\frac{n\cdot\beta_{(k)}}{n^{\frac{k}{2}}\cdot\alpha_{(2)}^{k/2}}\right)^{\frac{1}{k-2}}=\frac{1}{\sqrt{n}}\cdot\left(\frac{\beta_{(k)}}{\alpha_{(2)}^{k/2}}\right)^{\frac{1}{k-2}}\leq\frac{1}{\sqrt{n}}\cdot c(k)^{\frac{k}{(k-2)}}.$$

Now, let us define $Z=(X_1+\ldots+X_n)/\sigma$. Note that $Z$ lies on a lattice on a lattice with span $1/\sigma$. Let us call this lattice $\mathcal{L}$. In this setting, we have the following theorem.

*Theorem 10:* Let $X_1,\ldots,X_n$ be as defined above. Finally, let us define $I=\frac{\sqrt{n}}{c(3s)^{\frac{3s}{3s-2}}}$. Then, for any $z\in\mathcal{L}$,

$$\begin{aligned}\left|\Pr[Z=z]-\frac{1}{2\pi\sigma}\cdot\int_{-\pi\sigma}^{\pi\sigma}e^{-i\xi z}\cdot e^{-\frac{\xi^2}{2}}\cdot\left(1+\sum_{\nu=1}^{s-3}P_\nu(i\xi)\right)d\xi\right|\ &\leq\ \frac{s^{O(s)}\cdot c(3s)^{\frac{3s^2}{s-2}}}{n^{\frac{s-2}{2}}}+e^{-\frac{I^2}{6}}+s^{O(s)}\cdot e^{-\frac{I^2}{4}}\\ &+\ \sup_{|\zeta|\in\left[\frac{\alpha_{(2)}}{\beta_{(3)}},\pi\right]}\left|\widehat{X}(\zeta)\right|^n\end{aligned}$$

*Proof:* We simply apply Theorem 9 and evaluate the error terms. First, we evaluate $\eta_{\mathsf{low}}$ to get

$$\eta_{\mathsf{low}}=\frac{s^{O(s)}\cdot c(3s)^{\frac{3s^2}{s-2}}}{n^{\frac{s-2}{2}}}.$$

Next, we note that $I_s=\frac{\sqrt{n}}{c(3s)^{\frac{3s}{3s-2}}}$. Thus, we get

$$\eta_{\mathsf{med}}=e^{-\frac{I_s^2}{6}}+s^{O(s)}\cdot e^{-\frac{I_s^2}{4}}.$$

Finally, noting that

$$\sup_{|\zeta|\in[\sigma^2/\beta_3,\pi]}\left|\prod_{i=1}^n\widehat{X_i}(\zeta)\right|=\sup_{|\zeta|\in\left[\frac{\alpha_{(2)}}{\beta_{(3)}},\pi\right]}\left|\widehat{X}(\zeta)\right|^n,$$

implies that $\eta_{\mathsf{high}}=\sup_{|\zeta|\in\left[\frac{\alpha_{(2)}}{\beta_{(3)}},\pi\right]}\left|\widehat{X}(\zeta)\right|^n+s^{O(s)}\cdot e^{-\frac{I_s^2}{4}}$, which finishes the proof. ∎

## REFERENCES


[1] H. Cramér, "On the composition of elementary errors, I," *Skand. Aktuarietidskr*, vol. 1, pp. 13–74, 1928.

[2] P. Gopalan, R. Meka, O. Reingold, and D. Zuckerman, "Pseudorandom generators for combinatorial shapes," *SIAM Journal of Computing*, vol. 42, pp. 1051–1076, 2013.

[3] G. Even, O. Goldreich, M. Luby, N. Nisan, and B. Velickovic, "Approximations of General Independent Distributions," in *Proc. 24th Annual ACM Symposium on Theory of Computing (STOC)*, 1992, pp. 10–16.

[4] J. Naor and M. Naor, "Small-bias probability spaces: efficient constructions and applications," *SIAM J. on Comput.*, vol. 22(4), pp. 838–856, 1993, earlier version in STOC'90.

[5] S. Lovett, O. Reingold, L. Trevisan, and S. Vadhan, "Pseudorandom bit generators that fool modular sums," in *13th International Workshop on Randomization and Computation*, 2009, pp. 615–630.

[6] R. Servedio, "Every linear threshold function has a low-weight approximator," *Comput. Complexity*, vol. 16, no. 2, pp. 180–209, 2007.

[7] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell, "Optimal inapproximability results for Max-Cut and other 2-variable CSPs?" *SIAM Journal on Computing*, vol. 37, no. 1, pp. 319–357, 2007.

[8] C. Daskalakis, A. De, I. Diakonikolas, A. Moitra, and R. A. Servedio, "A Polynomial-time Approximation Scheme for Fault-tolerant Distributed Storage," in *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, 2014, pp. 628–644.

[9] G. Valiant and P. Valiant, "Estimating the unseen: an $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs," in *Proc. 43rd Annual ACM Symposium on Theory of Computing (STOC)*, 2011, pp. 685–694.

[10] C. Daskalakis and C. H. Papadimitriou, "Computing Equilibria in Anonymous Games," in *48th Annual IEEE Symposium on Foundations of Computer Science*, 2007, pp. 83–93.

[11] W. Feller, *An introduction to probability theory and its applications*. John Wiley & Sons, 1968.

[12] C. Daskalakis and C. H. Papadimitriou, "On oblivious PTAS's for Nash equilibrium," in *Proc. 41st Annual ACM Symposium on Theory of Computing (STOC)*, 2009, pp. 75–84.

[13] C. Daskalakis, I. Diakonikolas, R. O'Donnell, R. A. Servedio, and L. Tan, "Learning sums of independent integer random variables," in *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, 2013, pp. 217–226. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/FOCS.2013.31

[14] L. H. Chen, L. Goldstein, and Q.-M. Shao, "Normal approximations by Stein's method," *Springer*, 2010.

[15] V. Petrov, *Sums of Independent Random Variables*. Springer, 1975.

[16] I. A. Ibragimov and Y. V. Linnik, *Independent and Stationary Sequences of Random Variables*. Woolters Noordhoff, Groningen, 1971.

[17] A. Barbour, 2014, personal communication.

[18] O. Zeitouni, 2014, personal communication.

[19] A. D. Barbour and A. Xia, "Poisson perturbations," *ESIAM: Probability and Statistics*, pp. 131–150, 1999.

[20] R. N. Bhattacharya and R. R. Rao, *Normal approximation and asymptotic expansions*. Robert E. Krieger Publishing Company, 1986.

[21] C.-G. Esseen, "Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law," *Acta Mathematica*, vol. 77, no. 1, pp. 1–125, 1945.

[22] A. D. Barbour and V. Čekanvičius, "Total variation asymptotics for sums of independent integer random variables," *Annals of Probability*, vol. 30, pp. 509–545, 2002.

[23] B. Roos, "Binomial approximation to the Poisson binomial distribution: The Krawtchouk expansion," *Theory Probab. Appl.*, vol. 45, pp. 328–344, 2000.

[24] A. De, "Near optimal sparse-covers for poisson binomial distributions," 2015, available at http://www.math.ias.edu/∼anindya/sparse-cover.pdf.

[25] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, "Noise stability of functions with low influences: Invariance and optimality," *Ann. Math.*, vol. 171(1), pp. 295–341, 2010.

[26] A. D. Barbour and T. Lindvall, "Translated poisson approximation for markov chains," *Journal of Theoretical Probability*, vol. 19, 2006.

[27] A. Röllin, "Translated Poisson Approximation Using Exchangeable Pair Couplings," *Annals of Applied Probability*, vol. 17, no. 5/6, pp. 1596–1614, 2007.

[28] C. Daskalakis and C. H. Papadimitriou, "Discretized Multinomial Distributions and Nash Equilibria in Anonymous Games," in *49th Annual IEEE Symposium on Foundations of Computer Science*, 2008, pp. 25–34.

[29] C. Daskalakis and C. Papadimtriou, "Sparse covers for sums of indicators," *Probability Theory and Related Fields*, pp. 1–27, 2014.

[30] I. Diakonikolas, D. Kane, and A. Stewart, "Nearly optimal learning and sparse covers for sums of independent integer random variables," 2015, available at http://arxiv.org/abs/1505.00662.

[31] R. Impagliazzo, N. Nisan, and A. Wigderson, "Pseudorandomness for network algorithms," in *Proc. 26th Annual ACM Symposium on Theory of Computing (STOC)*, 1994, pp. 356–364.

[32] N. Nisan, "Pseudorandom generators for space-bounded computations," *Combinatorica*, vol. 12, no. 4, pp. 449–461, 1992.

[33] J. Brody and E. Verbin, "The Coin Problem and Pseudorandomness for Branching Programs," in *Proc. 51st IEEE Symposium on Foundations of Computer Science (FOCS)*, 2010, pp. 30–39.

[34] M. Braverman, A. Rao, R. Raz, and A. Yehudayoff, "Pseudorandom Generators for Regular Branching Programs," in *Proc. 51st IEEE Symposium on Foundations of Computer Science (FOCS)*, 2010, pp. 40–47.

[35] M. Koucký, P. Nimbhorkar, and P. Pudlák, "Pseudorandom generators for group products," in *Proc. 43rd Annual ACM Symposium on Theory of Computing (STOC)*, 2011, pp. 263–272.

[36] A. De, "Pseudorandomness for Permutation and Regular Branching Programs," in *IEEE Conference on Computational Complexity*, 2011, pp. 221–231.

[37] T. Steinke, "Pseudorandomness for Permutation Branching Programs Without the Group Theory," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 19, p. 83, 2012.

[38] O. Reingold, T. Steinke, and S. P. Vadhan, "Pseudorandomness for Regular Branching Programs via Fourier Analysis," in *17th International Workshop on Randomization and Computation*, 2013, pp. 655–670.

[39] T. Steinke, S. P. Vadhan, and A. Wan, "Pseudorandomness and Fourier Growth Bounds for Width 3 Branching Programs," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 21, p. 76, 2014.

[40] R. Armoni, M. E. Saks, A. Wigderson, and S. Zhou, "Discrepancy Sets and Pseudorandom Generators for Combinatorial Rectangles," in *Proc. 37th IEEE Symposium on Foundations of Computer Science (FOCS)*, 1996, pp. 412–421.

[41] C.-J. Lu, "Improved Pseudorandom Generators for Combinatorial Rectangles," *Combinatorica*, vol. 22, no. 3, pp. 417–434, 2002.

[42] P. Gopalan, R. Meka, O. Reingold, L. Trevisan, and S. P. Vadhan, "Better Pseudorandom Generators from Milder Pseudorandom Restrictions," in *Proc. 53rd IEEE Symposium on Foundations of Computer Science (FOCS)*, 2012, pp. 120–129.

[43] T. Watson, "Pseudorandom generators for combinatorial checkerboards," *Computational Complexity*, vol. 22, no. 4, pp. 727–769, 2013.

[44] P. Gopalan, D. M. Kane, and R. Meka, "Pseudorandomness for concentration bounds and signed majorities," *CoRR*, vol. abs/1411.4584, 2014.

[45] L. Bazzi, "Polylogarithmic independence can fool DNF formulas," in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007)*. IEEE Computer Society, 2007, pp. 63–73.

[46] P. Gopalan, D. Kane, and R. Meka, "Pseudorandomness via the discrete fourier transform," 2015, available at http://arxiv.org/abs/1506.04350.