

Tight Chernoff-Like Bounds Under Limited Independence

Maciej Skorski   

University of Luxembourg, Luxembourg

Abstract

This paper develops sharp bounds on moments of sums of k -wise independent bounded random variables, under constrained average variance. The result closes the problem addressed in part in the previous works of Schmidt et al. and Bellare, Rompel. The work also discusses other applications of independent interests, such as asymptotically sharp bounds on binomial moments.

2012 ACM Subject Classification Mathematics of computing → Statistical paradigms; Mathematics of computing → Probabilistic inference problems; Theory of computation → Randomness, geometry and discrete structures

Keywords and phrases concentration inequalities, tail bounds, limited independence, k -wise independence

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2022.15

Category RANDOM

Related Version *Full Version:* https://www.researchgate.net/publication/362034749_Tight_Chernoff-Like_Bounds_under_Limited_Independence

Funding The research supported by the FNR grant C17/IS/11613923.

Acknowledgements The author thanks the reviewers of RANDOM'22 for insightful comments.

1 Introduction

1.1 Motivation

Consider sums of random variables, possibly differently distributed. What can be said about the probability distribution, particularly the tails, if we only assume *k -wise independence*, that is any k of the n summands are independent? Such a dependency condition is an appealing concept studied in a broad class of problems related to pseudorandomness, including hashing, random graphs, random projections and circuits [34, 3, 28, 6]; specifically, concentration results under k -wise independence find important applications including (but not limited to) constructions of pseudorandom generators [25], load balancing [29, 50], derandomization [49, 23, 14], streaming algorithms [36] and cryptography [7, 5, 17, 4].

At first glance, the problem seems well addressed by concentration inequalities, such as the classical bounds due to Bernstein, Chernoff, Hoeffding, Bennet [10, 15, 24, 8] or their modern sub-gaussian or sub-gamma generalizations [13] (obtained from moment generating functions); at the very least one may hope to utilize more exotic moment bounds such as Rosenthal-type inequalities [47, 12] or more general frameworks [32]. However, the exponential moment methods are inadequate for limited dependence, whereas moment methods are hard to apply for sums of heterogenic terms. The state-of-the-art is held by the two influential works [50, 7] which resort to direct moment calculations, offering bounds for certain parameter regimes.

The goal of the current paper is to establish *sharp moment bounds* for sums of bounded k -wise independent variables, strengthening the state-of-art results [50, 7]. As in prior work, we assume that the summands are bounded and the sum variance is known. Formally:



© Maciej Skorski;

licensed under Creative Commons License CC-BY 4.0

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2022).

Editors: Amit Chakrabarti and Chaitanya Swamy; Article No. 15; pp. 15:1–15:14



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Let $S = \sum_{i=1}^n X_i$ be a sum of k -wise independent random variables, possibly differently distributed. Suppose that a) $X_i \in [-1, 1]$ and b) the average variance is $\frac{1}{n} \sum_{i=1}^n \mathbb{V}[X_i] = \sigma^2$. What is the *best* bound on moments of $S - \mathbb{E}S$?

Answering this question obviously gives desired Chernoff-like tail bounds, via an application of Markov’s inequality. This approach, called the moment method, is the state-of-art technique of establishing tail bounds [13, 32], even superior to the exponential moment method [44], so it should give us as much as we can get.

1.2 Our Contribution

The novelty of this work has the following aspects

- sharp bounds are found for *all* parameter regimes, which improves upon prior works
- some elegant techniques, novel in this context, are demonstrated; particularly the powerful method of *symmetrization* from high-dimensional probability [51] and *extreme inequalities* for symmetric polynomials [48] ¹
- other applications, in particular sharp bounds for moments of binomial distribution

We now move to present our results, adopting the following notation: for two expressions A, B we write $A \lesssim B$ when $A \leq K \cdot B$ for some absolute constant K , and $A \simeq B$ when the inequality holds in both direction. By $\|Z\|_d = (\mathbb{E}|Z|^d)^{1/d}$ we denote the d -th norm of a random variable Z . By $\mathbb{E}Z$ and $\mathbb{V}[Z]$ we denote, respectively, the mean and variance of Z .

1.2.1 Sharp Bounds for k -wise Independent Sums

The theorem below gives the complete answer to the posed problem.

► **Theorem 1** (Moments of k -wise Independent Sums). *Consider random variables $(X_i)_{i=1}^n$ satisfying the following conditions*

- (a) $(X_i)_i$ are k -wise independent ($k \geq 2$) and $|X_i - \mathbb{E}X_i| \leq 1$
- (b) $\sum_{i=1}^n \mathbb{V}[X_i] \leq n\sigma^2$ (the sum variance bounded)

Then for $S = \sum_{i=1}^n X_i$ and any positive even integer $d \leq k$ we have:

$$\max_{(X_i)} \|S - \mathbb{E}S\|_d \simeq M(n, \sigma^2, d) = \begin{cases} \sqrt{dn\sigma^2} & \log(d/n\sigma^2) < \max(d/n, 2) \\ \frac{d}{\log(d/n\sigma^2)} & \max(d/n, 2) \leq \log(d/n\sigma^2) \leq d, \\ (n\sigma^2)^{1/d} & d < \log(d/n\sigma^2) \end{cases} \quad (1)$$

where the maximum is over all r.v.s. $(X_i)_i$ satisfying the conditions (a) and (b).

Moreover, the maximal value is realized (up to a constant) when

$$X_i \sim B - B', \quad B, B' \sim^{iid} \text{Bern}(p), p = \frac{1}{2}(1 - \sqrt{1 - 2\sigma^2}). \quad (2)$$

► **Remark 2** (Value of k). In applications value of k should be possibly big, so that we can use as high moments d as possible. For example, some cryptographic applications use $k = 80$ [17].

¹ The prior work [50] actually recognized usefulness of symmetry, but was not able to exploit it in the case of general $[-1, 1]$ -valued random variables.

► **Remark 3 (Formula Regimes).** The formulas may look a little exotic, especially to a reader familiar with previous works. The reason is that the novel bounds above are sharp and capture some non-standard behaviors. The branch with $\sqrt{dn\sigma^2}$ should be familiar, as this is the d -th moment of gaussian distribution with variance $n\sigma^2$; in this regime it would produce the tail $\Pr[|S - \mathbb{E}S| > t] \leq e^{-\Omega(t^2/n\sigma^2)}$. The branch with $d/\log(d/n\sigma^2)$ gives the behavior slightly faster than this of the exponential distribution; it would produce the tail $\Pr[|S - \mathbb{E}S| > t] \leq e^{-\omega(t)}$. Finally, the branch with $(n\sigma^2)^{1/d}$ resembles the behavior of a distribution bounded by σ^2 .

► **Remark 4 (Odd values of d).** It can be shown that the same upper bounds apply when $d > 2$ is odd, due to interpolation inequalities [9].

► **Remark 5 (Explicit Tail Bounds).** For M as in Equation (1) we obtain the following tail bound (which depends on the parameters regime), for any $t > 0$

$$\Pr[|S - \mathbb{E}S| > t] \leq O(M(n, \sigma^2, d)/t)^d. \tag{3}$$

For $t = cM(n, \sigma^2, d)$ with an appropriate constant c , we obtain the tail of $2^{-\Omega(d)}$.

1.2.2 Techniques

We show that for even $d \leq k$ we can assume in addition c) full independence and d) full symmetry of the summands, leveraging *symmetrization* [51]. Then we proceed in two steps.

1.2.2.1 Characterization of Extreme Distribution

First, we characterize “worst-case” distributions X_i that maximize $\|\sum_i X_i\|_d$. This result is the core of our approach and of broader interest, we thus present it as the standalone lemma.

► **Lemma 6 (IID Majorization of Symmetric Sums).** *Let $(Z_i)_{i=1}^n$ be independent symmetric random variables with values in $[-1, 1]$ with average variance $\sigma^2 = \frac{1}{n} \sum_{i=1}^n \mathbb{V}[Z_i]$. Then for any positive even integer d we have that*

$$\|\sum_i Z_i\|_d \leq \|\sum_i Z'_i\|_d \tag{4}$$

where Z'_i are independent and identically distributed as

$$Z'_i \sim \begin{cases} +1 & w.p. \sigma^2/2 \\ 0 & w.p. 1 - \sigma^2 \\ -1 & w.p. \sigma^2/2 \end{cases}. \tag{5}$$

► **Remark 7 (Interpretation).** Observe that $\mathbb{V}[Z'_i] = 2 \cdot \sigma^2/2 = \sigma^2$, thus the theorem essentially says that moments of the sum $\sum_i Z_i$ are maximized for Z_i that are iid with the distribution (5). This three-point distribution is *extreme*, in the sense that it pushes as much mass as possible towards the edge of the interval constraint. This behavior may look intuitive, but we should beware of such intuitions as even for simple symmetric problems whether the maximizer’s behavior is “push to boundary” or “pull to the middle” may not be that intuitive, depending on Schur convexity properties of the optimized expression [21]; to be specific the problems of maximization of $\sum_{i \neq j} p_i p_j$ and $\sum_i p_i^2$ have quite different behavior. Our proof requires some non-trivial facts about multivariate symmetric polynomials.

► **Remark 8 (Proof Techniques).** The symmetry assumption is crucial, and makes it possible to greatly simplify the multinomial expansion of the moment formula. We manage to regroup expressions and see them as positive combinations of elementary symmetry polynomials; then specialized inequalities from the theory of symmetric functions [39] come then to the rescue, allowing for proving that our extreme distribution is indeed the maximizer.

1.2.2.2 Closed-Form Bounds for Extreme Distributions

In addition to characterizing the worst-case behavior, we give the closed-form formula for the bound in Lemma 6. As we will see later, this is also a fact of broader interest; for example, we use it to derive bounds for binomial moments which are sharp in all parameter regimes.

► **Corollary 9 (Best bounds for IID).** *For independent (d -wise independent) symmetric Z_i with values $\{-1, 0, 1\}$ and variance σ^2 and positive even integer d we have*

$$\left\| \sum_{i=1}^n Z_i \right\|_d \simeq \begin{cases} \sqrt{dn\sigma^2} & \log(d/n\sigma^2) < \max(d/n, 2) \\ \frac{d}{\log(d/n\sigma^2)} & \max(d/n, 2) \leq \log(d/n\sigma^2) \leq d \\ (n\sigma^2)^{1/d} & d < \log(d/n\sigma^2) \end{cases} \quad (6)$$

► **Remark 10 (Proof techniques).** The proof requires some effort to compute moments for Equation (5). Loosely speaking, we leverage the specific form of the distribution to obtain regular combinatorial patterns in multinomial expansions. We then obtain an explicit formula, being a weighted sum of binomial-like expressions which involve n , d and σ . Establishing the order of growth, with the help of some calculus, completes the proof.

1.3 Related Work

There are many bounds which cover different models of dependencies among random variables, for example Janson’s correlation inequality [27] which has become very popular in analyses of random graphs [18], or the theory of negative dependence [11] best known from applications to various “balls and bins” problems [19]. However the focus of this paper is on *k-wise independence*, in which the state-of-art bounds are due to Schmidt et al. [50] and Bellare and Rompel [7], derived in the essentially same setup as ours (moment bounds under the variance constraint). These bounds, although useful for many applications, hold only in certain regimes and are not sharp in general; when discussing our applications we will show that in most cases they are inferior to Theorem 1.

It in our work we consider the most natural variance constraint, following prior works [50, 7]. However, one might consider more exotic structural assumptions; recently there has been an attempt, limited only to $k = 2$, to characterize worst bounds by exploring the whole sequence (rather than the sum variance) of Bernoulli parameters of X_i [45].

Regarding the established concentration bounds, we will see that known inequalities actually imply stronger bounds than those developed by Schmidt et al. and Bellare, Rompel [50, 7]. However, even with the use of state-of-art moment inequalities [13] we were not able to recover the sharp bounds from our main result and the characterization from Lemma 6. The key challenge is to precisely characterize the worst-case behavior, while allowing *differently distributed* random variables.

An interesting way of attacking the problem, possibly working for much more exotic constraints, may be to formally follow the presented idea of majorization and establish formally some Schur-convexity properties; this approach has been successfully applied in the past to many other problems (see [20] and follow-up works).

■ **Table 1** Bounds for moments of sum $S = \sum_{i=1}^n X_i$ of d -wise independent random variables, where $d \geq 2$. As discussed, our bounds are strictly better than those of Schmidt at al., which in turn are strictly better than those of Bellare and Rompel.

Bound on $\ S - \mathbb{E}S\ _d$	Author	Assumptions
$\max(\sqrt{dn\sigma^2}, d/\log(d/n\sigma^2), (n\sigma^2)^{1/d})$	this paper	$n\sigma^2 = \mathbb{V}[S], X_i - \mathbb{E}X_i \leq 1$
$\max(\sqrt{dn\sigma^2}, d)$	Schmidt at al., optimized	$n\sigma^2 = \mathbb{V}[S], X_i - \mathbb{E}X_i \leq 1$
$\min(\sqrt{dn}, \sqrt{dn\mu + d^2})$	Bellare and Rompel	$n\mu = \mathbb{E}[S], 0 \leq X_i \leq 1$

1.4 Applications

1.4.1 Limited Independence: Clarifying the State-of-Art

We will demonstrate how our bounds improve on those of Schmidt at al. [50] and Bellare and Rompel [7], clarifying this way the state-of-the-art. In what follows we assume, as in our theorem, that $\|X_i - \mathbb{E}X_i\| \leq 1$, X_i are k -wise independent, and $d \leq k$ for positive even d . The best bound due to Schmidt at al. reads as (cf Eq. 10 in [50])

$$\|S - \mathbb{E}S\|_d^d \leq \sqrt{2} \cdot \cosh(\sqrt{d^3/36C}) \cdot (dC/e)^{d/2}, \quad C \geq n\sigma^2, \sigma^2 = \mathbb{V}[S]/n.$$

The authors did not fully optimize the choice of C , offering a bunch of weaker corollaries instead. In order to clarify the state-of-the-art, we do this effort (see Section 3.4) obtaining

$$\|S - \mathbb{E}S\|_d \lesssim \max(\sqrt{dn\sigma^2}, d). \tag{7}$$

When $dn\sigma^2 \geq 1$ the formula matches ours, but otherwise it is much worse: by a factor of $\log(d/n\sigma^2)$ in the regime $\max(2, d/n) \leq \log(d/n\sigma^2) \leq d$, and by a factor of d in the regime $d < \log(d/n\sigma^2)$ (which necessarily means $n\sigma^2 < 1$). In applications, these factors can be a big constant or more, so derived tail bounds are worse by a big constant in the exponent.

In turn, the bound due to Bellare and Rompel [7] states that when $X_i \in [0, 1]$

$$\|S - \mathbb{E}S\|_d \lesssim \min(\sqrt{dn}, \sqrt{dn\mu + d^2}), \quad \mu = \frac{1}{n}\mathbb{E}S. \tag{8}$$

We claim this is worse than our optimized version of Schmidt at al., in all regimes. Namely,

$$\max(\sqrt{dn\sigma^2}, d) \lesssim \min(\sqrt{dn}, \sqrt{dn\mu + d^2}).$$

Indeed, when $d > n$ the left-hand side is at most n , while the right-hand side is at least n . When $d \leq n$, due to $\mu \leq 1$ (a consequence of $X_i \leq 1$) we see that $\min(\sqrt{dn}, \sqrt{dn\mu + d^2}) \simeq \sqrt{dn\mu + d^2}$. But we have $\mathbb{V}[X_i] \leq \mathbb{E}X_i$, as the consequence of $0 \leq X_i \leq 1$, and thus $\sigma^2 \leq \mu$; this shows $\min(\sqrt{dn}, \sqrt{dn\mu + d^2}) \gtrsim \sqrt{dn\sigma^2}$ and the claim follows.

This discussion should be of broader interest to the TCS community, as it seems that no rigorous comparison between [7] and [50] has been done before (the surveys such as [38] and application works credit both exchangeably). In Table 1 we give a readable summary.

1.4.2 Obtaining Previous Results from Classical Inequalities

Our literature search shows, perhaps surprisingly, that the optimized bounds of Schmidt at al. are actually a *simple consequence of classical inequalities*; we note that the prior works [7, 50] do not discuss the related literature on concentration bounds. The intent of this discussion is to bring those inequalities to the awareness of the wider TCS audience, particularly given the huge interest and the citation credit given to the bounds in [7, 50].

Assume that X_i are k -wide independent; recall that the event moment of order $d \leq k$ can be calculated as if the summands were independent. More precisely, we have $\sum_i \|X_i - \mathbb{E}X_i\|_d = \sum_i \|X'_i - \mathbb{E}X'_i\|_d$ where X'_i are distributed as X_i and independent. The tail bounds due to a century old (!) *Bernstein's inequality* [10] imply that the tail of $S' = \sum_i X'_i$ satisfies $\Pr[|S' - \mathbb{E}S'| > t] \leq \exp(-\Theta(\min(t^2/n\sigma^2, t)))$ for any positive t , if X_i are bounded, and $n\sigma^2 = \sum_i \mathbb{V}[X'_i] = \sum_i \mathbb{V}[X_i]$. By the standard tail integration formula, we find that the moments of the IID sum are $\|S' - \mathbb{E}S'\|_d \lesssim \max(\sqrt{nd\sigma^2}, d)$. As remarked, this matches $\|S - \mathbb{E}S\|_d$ when $d < k$, so we recover the optimized (!) bounds of Schmidt at al., and implies the bounds of Bellare and Rompel. Another argument can be given by the use of *Rosenthal's inequality*, a version of which [22] implies $\|S - \mathbb{E}S\|_d \lesssim d \cdot (\sum_{i=1}^n \mathbb{E}|X_i - \mathbb{E}X_i|^d)^{1/d} + d^{1/2} \cdot (\sum_{i=1}^n \mathbb{E}|X_i - \mathbb{E}X_i|^2)^{1/2}$. This can be further bounded by $\max(\sqrt{dn\sigma^2}, d)$.

1.4.3 Sharp Explicit Bounds on Binomial Moments

Somewhat surprisingly, to the best of author's knowledge, there are no good closed-form estimates on moments of the binomial distribution, despite the clear demand from applications (such as the analysis of random projections [2, 26]). The sharp (up to an $o(1)$ relative error term) tail bounds due to Littlewood [33, 37] in theory imply sharp moment estimates, but calculations lead to very difficult integrals with Kullback-Leibler divergence in the exponent. We obtain closed-form bounds for even binomial moments as a byproduct of our analysis, which are sharp in all parameter regimes. More precisely, we have

► **Corollary 11.** *Let $S \sim \text{Binom}(n, p)$ where $p \leq 1/2$ and d be a positive even integer. Then*

$$\|S - \mathbb{E}S\|_d \simeq \begin{cases} \sqrt{dnp} & \log(d/np) < \max(d/n, 2) \\ \frac{d}{\log(d/np)} & \max(d/n, 2) \leq \log(d/np) \leq d \\ (np)^{1/d} & d < \log(d/np) \end{cases} \quad (9)$$

This follows from the fact that the extreme variables Z'_i in our main result can be expressed as symmetrized Bernoulli distributions, namely $Z'_i \sim B - B'$ where $B, B' \stackrel{iid}{\sim} \text{Bern}(p)$ with $p = \frac{1}{2}(1 - \sqrt{1 - 2\sigma^2})$. Let $S \sim \text{Binom}(n, p)$. By symmetrization $\|S - \mathbb{E}S\|_d \simeq \|S - S'\|_d$ and thus $\|S - \mathbb{E}S\|_d \simeq \|\sum_i (B_i - B'_i)\|_d \simeq \|\sum_i Z'_i\|_d$, thus by our result $\|S - \mathbb{E}S\|_d$ obeys the bound as above with p replaced by σ^2 . It remains to observe that $\sigma^2 = 2p(1 - p)$ so $p \leq \sigma^2 \leq 2p$, and that the bounds above do not change by a more than a constant when p is replaced by $p' \in [p, 2p]$.

1.4.4 Exact Binomial Moments

Binomial moments can be evaluated by means of combinatorics, which yields somewhat complicated recursions [30]. Interestingly, a byproduct of Corollary 9 gives an exact formula for *symmetrized* binomials which has very simple form.

1.4.5 Estimating binomial-like moments

The line of research focused on estimating Renyi entropy of unknown probability distributions ([1, 43]) faces the problem of estimating moments of sum of random variables in “small variance” regime, that is when $n\sigma^2 \ll 1$. For example, the collision estimator requires bounds on the 4-th sum moment. This has been previously done by exploiting somewhat tedious combinatorial identities, but follows easier from Corollary 9 and Lemma 6.

2 Preliminaries

2.1 Multinomial Expansion

The *multinomial coefficient* is defined as

$$\binom{d}{\mathbf{j}} = d! / \prod_{j \in \mathbf{j}} j! \quad (10)$$

when all components of \mathbf{j} are non-negative and $\sum_{j \in \mathbf{j}} j = d$. We also extend this to $\binom{d}{\mathbf{j}} = 0$ when $\min\{j : j \in \mathbf{j}\} < 0$ or $\sum_{j \in \mathbf{j}} j \neq d$; this allows for concise notation. The multinomial formula takes the form $(\sum_i x_i)^d = \sum_{\mathbf{i}} \prod_{i \in \mathbf{i}} x_i^{i_i}$.

► **Remark 12.** Factorials, and therefore binomial coefficients can be formally extended to negative numbers by means of Gamma function. Then indeed multinomial coefficients are zero when negative integers appear as downward arguments [31].

We will occasionally use the Stirling's formula in estimation of multinomial coefficients [40, 46]

$$(d/\ell)! \simeq \sqrt{d/\ell} \cdot (d/e\ell)^{d/\ell}. \quad (11)$$

2.2 Symmetrization

We will need the following facts about symmetrization (cf [51]).

► **Proposition 13** (Convex Symmetrization). *For zero-mean iid X, X' and convex f*

$$\mathbb{E}f(X) \leq \mathbb{E}f(X - X'). \quad (12)$$

Proof of Proposition 13. By independence $\mathbb{E}f(X - X') = \mathbb{E}_X \mathbb{E}_{X'}[f(X - X')|X]$ and by Jensen's inequality $\mathbb{E}_{X'}[f(X - X')|X] \geq f(X - \mathbb{E}X')$. By the zero-mean assumption $f(X - \mathbb{E}X') = f(X)$. The inequality follows by chaining these three bounds. ◀

► **Proposition 14** (Moments are robust under symmetrization). *For any iid random variables $\|X\|_d \leq \|X - X'\|_d \leq 2\|X\|_d$*

Proof of Proposition 14. Since $\|X\|_d = (\mathbb{E}|X|^d)^{1/d}$, the left-hand side follows by applying Proposition 13 to $f(u) = |u|^d$. The right-hand side is due to the triangle inequality (Minkovski's inequality for L_p spaces). ◀

2.3 Symmetric Functions

The ℓ -th elementary symmetric polynomial in variables $u = (u_i)_i$ is defined as

$$\Pi_\ell(u) = \sum_{i_1 < \dots < i_\ell} u_{i_1} u_{i_2} \cdot \dots \cdot u_{i_\ell}. \quad (13)$$

The fundamental theorem on symmetric polynomials states that they generate all other symmetric polynomials (in a sense of the algebraic ring) [16]. We will need some facts about their extreme properties, which we recall below.

► **Proposition 15** (Newton Inequalities [41]). *For $u = (u_i)_{i=1}^n$ let $S_\ell(u) \triangleq \Pi_\ell(u) / \binom{n}{\ell}$ be the ℓ -th elementary symmetric mean. Then $S_{\ell-1}(u)S_{\ell+1}(u) \leq S_\ell(u)^2$.*

This implies the useful inequality due to Maclaurin

► **Proposition 16** (Maclaurin's Inequality [35, 42]). *For $u = (u_i)_{i=1}^n$ we have the inequality $S_\ell(u)^{1/\ell} \geq S_{\ell'}(u)^{1/\ell'}$ when $1 \leq \ell < \ell' \leq n$ (with the equality when u_i are equal).*

3 Proofs

3.1 Proof of Lemma 6

We use the fact that d is an even integer and the multinomial formula to expand

$$\mathbb{E} \left| \sum_{i=1}^n Z_i \right|^d = \mathbb{E} \left(\sum_{i=1}^n Z_i \right)^d = \sum_{\mathbf{j}} \binom{d}{\mathbf{j}} \mathbb{E}[Z_1^{j_1} \cdots Z_n^{j_n}] \quad (14)$$

The summation is over integer tuples $\mathbf{j} \in \mathbb{Z}^n$ called also multiindices. Utilizing the independence assumption, we obtain

$$\mathbb{E} \left| \sum_{i=1}^n Z_i \right|^d = \sum_{\mathbf{j}} \binom{d}{\mathbf{j}} \mathbb{E}[Z_1^{j_1}] \cdots \mathbb{E}[Z_n^{j_n}] \quad (15)$$

Since Z_i are symmetric, all odd moment vanish. Thus, we can write

$$\mathbb{E} \left| \sum_{i=1}^n Z_i \right|^d = \sum_{\mathbf{j}} \binom{d}{2\mathbf{j}} \mathbb{E}[Z_1^{2j_1}] \cdots \mathbb{E}[Z_n^{2j_n}]. \quad (16)$$

Since Z_i are absolutely bounded by 1 and symmetric, we have $\mathbb{E}|Z_i|^{2j} \leq \mathbb{E}|Z_i|^2 \leq \mathbb{V}[Z_i]$ for $j \geq 1$. Denoting $\sigma_i^2 = \mathbb{V}[Z_i]$ we can write

$$\mathbb{E} \left| \sum_{i=1}^n Z_i \right|^d \leq \sum_{\mathbf{j}} \binom{d}{2\mathbf{j}} \prod_{i:j_i \neq 0} \sigma_i^2. \quad (17)$$

► **Remark 17.** The equality is met when Z_i are symmetric with values in the set $\{-1, 0, 1\}$, as this implies $\mathbb{E}|Z_i|^j = \mathbb{E}|Z_i|^2$.

Let $\|\mathbf{j}\|_0 = \#\{i : j_i \neq 0\}$ be the number of non-zero indices in the multiindex \mathbf{j} . Clearly $\ell = \|\mathbf{j}\|_0$ can take values from 1 to $\frac{d}{2}$ and thus

$$\mathbb{E} \left| \sum_{i=1}^n Z_i \right|^d \leq \sum_{\ell=1}^{d/2} \underbrace{\sum_{\mathbf{j}:\|\mathbf{j}\|_0=\ell} \binom{d}{2\mathbf{j}} \prod_{i:j_i \neq 0} \sigma_i^2}_{S_\ell}. \quad (18)$$

Note that S_ℓ is multilinear of order ℓ in $u_i = \sigma_i^2$. We claim that it equals the elementary symmetric polynomial, up to a constant multiplier (this is not clear a-priori as different weights could break the symmetry).

▷ **Claim 18.** The polynomial S_ℓ is a (non-negative) multiplicity of the ℓ -th elementary symmetric polynomial Π_ℓ in variables σ_i^2 .

Proof of Claim. Indeed, consider S_ℓ as the weighted sum of monomials $\prod_{i \in I} \sigma_i^2$, where $\|I\| = \ell$. Every such a monomial appears with the coefficient $c_I \triangleq \sum_{\mathbf{j}:\mathbf{j}_i \neq 0 \Leftrightarrow i \in I} \binom{d}{2\mathbf{j}}$. Due to the symmetry of the multinomial coefficient $\binom{d}{2\mathbf{j}}$, namely the invariance under permuting \mathbf{j} , we claim that c_I is the same for every set I . Indeed, if $\rho(I') = I$ for a bijection ρ then

$$c_{I'} = \sum_{\mathbf{j}:\mathbf{j}_{i'} \neq 0 \Leftrightarrow i' \in I'} \binom{d}{2\mathbf{j}} = \sum_{\mathbf{j}:\mathbf{j}_{\rho(i)} \neq 0 \Leftrightarrow \rho(i) \in I} \binom{d}{2\mathbf{j}} = \sum_{\mathbf{j}:\mathbf{j}_{\rho(i)} \neq 0 \Leftrightarrow \rho(i) \in I} \binom{d}{2\rho(\mathbf{j})} = c_I \quad (19)$$

It follows that S_ℓ is a multiplicity of the ℓ -th elementary symmetric polynomial (as it contains all monomials of order ℓ with equal coefficients). This proves the claim. ◁

We now establish extreme properties of S_ℓ . Namely

▷ **Claim 19.** The expression S_ℓ is maximized, subject to the constraint that $\sum \sigma_i^2$ is kept constant, when all σ_i are equal.

Proof of Claim. This follows by Maclaurin's Inequality in Proposition 16. ◁

Let $\sigma^2 = \frac{1}{n} \sum_i \sigma_i^2$, from the claim we obtain

$$\mathbb{E} \left| \sum_{i=1}^n Z_i \right|^d \leq \sum_{\ell=1}^{d/2} \sum_{\mathbf{j}: \|\mathbf{j}\|_0 = \ell} \binom{d}{2\mathbf{j}} \cdot \sigma^{2\ell} \tag{20}$$

The right-hand side is like in Equation (18) with all σ_i equal, and by the remark we know that it equals $\mathbb{E} \left| \sum_{i=1}^n Z'_i \right|^d$ if Z'_i is symmetric with values $\{-1, 0, 1\}$ and has variance σ^2 .

3.2 Proof of Corollary 9

For Z_i as in Lemma 6 the previous section derives the identity

$$\mathbb{E} \left| \sum_{i=1}^n Z_i \right|^d = \sum_{\ell=1}^{d/2} \sum_{\mathbf{j}: \|\mathbf{j}\|_0 = \ell} \binom{d}{2\mathbf{j}} \cdot \sigma^{2\ell}. \tag{21}$$

We will further simplify this expression. Considering positive components of \mathbf{j} we obtain

$$\sum_{\mathbf{j}: \|\mathbf{j}\|_0 = \ell} \binom{d}{2\mathbf{j}} = \sum_{\ell=1}^{d/2} \sum_{i_1 < \dots < i_\ell} \sum_{\mathbf{j}: j_i \geq 1 \Leftrightarrow i \in \{i_1, \dots, i_\ell\}} \binom{d}{2\mathbf{j}} \cdot \sigma^{2\ell}. \tag{22}$$

Since the expression is invariant under permutations of \mathbf{j} we obtain

$$\sum_{\mathbf{j}: \|\mathbf{j}\|_0 = \ell} \binom{d}{2\mathbf{j}} = \sum_{\ell=1}^{d/2} \sum_{\mathbf{j}=(j_1, \dots, j_\ell) \geq 1} \binom{d}{2\mathbf{j}} \cdot \binom{n}{\ell} \tag{23}$$

where $\binom{n}{\ell}$ counts the number of choices for i_1, \dots, i_ℓ . Therefore

$$\mathbb{E} \left| \sum_{i=1}^n Z_i \right|^d = \sum_{\ell=1}^{d/2} \sum_{\mathbf{j}=(j_1, \dots, j_\ell) \geq 1} \binom{d}{2\mathbf{j}} \binom{n}{\ell} \sigma^{2\ell}. \tag{24}$$

We now estimate the moment up to constants. Since for any positive a_i we have $(\sum_{i=1}^d a_i)^{1/d} \simeq (\max_i a_i)^{1/d}$ up to some absolute constants (in fact, constants are 1 and $d^{1/d} \leq 2$), we obtain

$$\left\| \sum_{i=1}^n Z_i \right\|_d \simeq \max_{\ell=1, \dots, d/2} \left[\underbrace{\sum_{\mathbf{j}=(j_1, \dots, j_\ell) \geq 1} \binom{d}{2\mathbf{j}} \cdot \binom{n}{\ell} \cdot \sigma^{2\ell}}_{F(\ell)} \right]^{1/d} \tag{25}$$

In the next step we estimate $F(\ell)^{1/d}$.

▷ **Claim 20.** We have $F(\ell)^{1/d} \simeq \ell$.

15:10 Tight Chernoff-Like Bounds Under Limited Independence

Proof. For the lower bound we can assume that ℓ (and hence d) are sufficiently big (otherwise the bound is trivial). We can also assume that ℓ divides d and that $r = \lfloor d/\ell \rfloor$ is even; otherwise we replace ℓ with ℓ' between ℓ and $\ell/2$ which satisfies this, use $F(\ell) \geq F(\ell')$ and prove for ℓ' . Consider the term $2j_1 = \dots = 2j_{\ell-1} = d/\ell$, we have

$$F(\ell) \geq \binom{d}{r, \dots, r} = \frac{d!}{(r!)^\ell}.$$

Observe that $(d!)^{1/d} \simeq d$, $(r!)^{1/r} \simeq r$ and $(q!)^{1/q} \simeq q$ (Stirling's formula). Since $r \cdot \frac{\ell}{d} \leq 1$ we get $(r!)^{\frac{\ell}{d}} \simeq r^{r \cdot \frac{\ell}{d}} = r$ (the relation \simeq can be raised to a bounded power). This gives

$$F(\ell)^{1/d} \gtrsim \frac{d}{r} = \ell.$$

As for the upper bound, we simply note that

$$F(\ell)^{1/d} < \left(\sum_{\mathbf{j}=(j_1, \dots, j_\ell)} \binom{d}{\mathbf{j}} \right)^{1/d} \leq (\ell^d)^{1/d} = \ell.$$

These two bounds completes the proof. \triangleleft

By Claim 20 and Equation (25) we obtain the following, much simpler bound

$$\left\| \sum_{i=1}^n Z_i \right\|_d \simeq \max_{\ell=1, \dots, d/2} \left[\ell^d \cdot \binom{n}{\ell} \cdot \sigma^{2\ell} \right]^{1/d}. \quad (26)$$

With some more effort we simplify even further. Namely, we can assume $\ell \leq n$ as for $\ell > n$ we have $\binom{n}{\ell} = 0$. By the elementary inequality $(n/\ell)^\ell \leq \binom{n}{\ell} \leq (ne/\ell)^\ell$ we have $\binom{n}{\ell}^{1/d} \simeq (n/\ell)^{\ell/d}$, for $\ell = 1 \dots d/2$. Thus

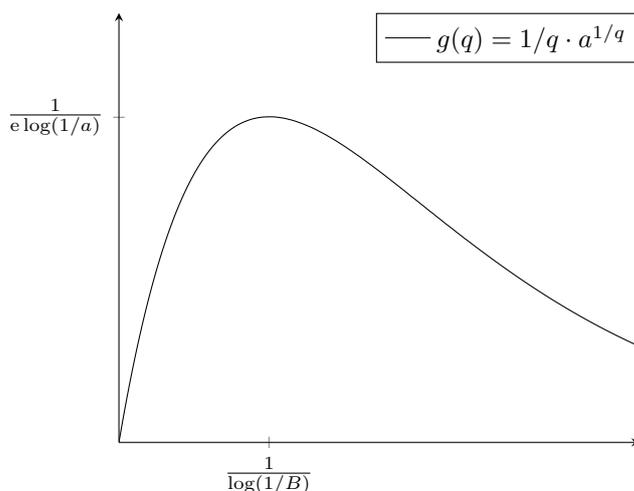
$$\begin{aligned} \left\| \sum_{i=1}^n Z_i \right\|_d &\simeq \max_{\ell=1, \dots, \min(d/2, n)} \left[\ell^d \cdot \binom{n}{\ell} \cdot \sigma^{2\ell} \right]^{1/d} \\ &\simeq \max_{\ell=1, \dots, \min(d/2, n)} \left[\ell \cdot (n/\ell)^{\ell/d} \cdot \sigma^{2\ell/d} \right]. \end{aligned} \quad (27)$$

Losing not more than a constant factor, we can extend the maximum to the continuous interval (the expression under maximum differs by at most a constant factor between two values of ℓ that differ by one or less). Let $q = d/\ell$, we have the equivalent constraint $\max(2, dn/n) \leq q \leq d$ and the maximum of $d/q \cdot (n\sigma^2/d)^{1/q} \cdot q^{1/q}$. Since $q^{1/q} \simeq 1$ when $q \geq 1$

$$\left\| \sum_{i=1}^n Z_i \right\|_d \simeq \max_{q: \max(2, d/n) \leq q \leq d} \left[d/q \cdot (n\sigma^2/d)^{1/q} \right]. \quad (28)$$

It now suffices to analyze the auxiliary function $g(q) \triangleq 1/q \cdot a^{1/q}$ for $q > 0$. The derivative test shows that it is decreasing when $a > 1$ and has the global maximum at $q = \log(1/a)$ with the value $1/e \log(1/a)$ when $a < 1$. This behavior is illustrated on Figure 1.

Applying this fact to Equation (28), with $a = n\sigma^2/d$, and comparing $\log(1/a)$ with the interval boundaries finishes the proof.



■ **Figure 1** Auxiliary function g which determines the moment behavior (for $a < 1$).

3.3 Proof of Theorem 1

We recall the folklore fact that the moment of order $d \leq k$ of the sum of k -wise independent r.v.s. can be computed as if they were independent. That is, let X'_i be distributed as X_i but independent. For even d we have

$$\mathbb{E} \left| \sum_i X_i \right|^d = \mathbb{E} \left| \sum_i X'_i \right|^d \tag{29}$$

which follows by applying the multinomial expansion on both sides and observing that the obtained formulas depend only on products of at most d of random variables X_i (respectively X'_i). Without loss of generality, we can also assume that X_i are centered. Now the result follows if X_i are symmetric, by Lemma 6 and Corollary 9. If they are not symmetric, we can use the general reduction as in Proposition 14; namely, we apply the proof to $X'_i - X''_i$ where $X'_i, X''_i \sim^{iid} X_i$. The moments differ by at most a factor of two. Particularly, the variance changes by a factor of 2, which has no impact on the asymptotic bounds in Equation (1). More precisely, we use the fact that $M(n, \sigma^2, d) \simeq M(n, \sigma'^2, d)$ where $\sigma^2/2 \leq \sigma'^2 \leq 2\sigma^2$.

3.4 Optimized moment bound of Schmidt et al.

Up to a constant factor, their bound is equivalent to

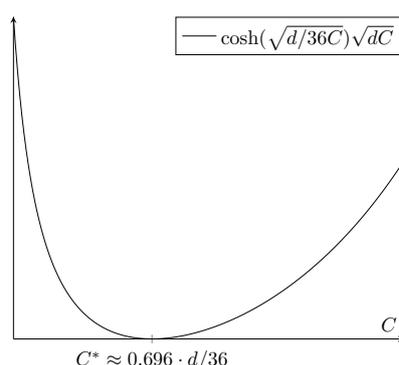
$$\|S - \mathbb{E}S\|_d \lesssim \cosh(\sqrt{d/36C})\sqrt{dC}, \quad \text{for any } C \geq \mathbb{V}[S]. \tag{30}$$

Let $t = \sqrt{d/36C}$, the upper bound is equivalent to $c \cdot d \cdot \cosh(t)/t$ where c is an absolute constant. We use this to function understand the behavior of Equation (30) on C , which is as illustrated on Figure 2. Since C is constrained by $C \geq \mathbb{V}[S]$, the best bound is obtained for

$$C = \max(C^*, \mathbb{V}[S]), \tag{31}$$

which gives the claimed bound of

$$\|S - \mathbb{E}S\|_d \lesssim \max(\sqrt{dn\sigma^2}, d). \tag{32}$$



■ **Figure 2** The moment bound of Schmidt et al., dependency on C .

4 Conclusion

We have developed sharp estimates on the moments of (non necessarily identically distributed) sums of random variables, assuming the variance is constrained. This essentially closes the problem of establishing good concentration bounds, discussed in prior works. Our approach demonstrates the power of *symmetrization* technique, and is of independent interest. We also showed applications, not limited to k -wise independence.

References

- 1 Jayadev Acharya, Alon Orlitsky, Ananda Theertha Suresh, and Himanshu Tyagi. Estimating rényi entropy of discrete distributions. *IEEE Transactions on Information Theory*, 63(1):38–56, 2016.
- 2 Zeyuan Allen-Zhu, Rati Gelashvili, Silvio Micali, and Nir Shavit. Sparse sign-consistent johnson–lindenstrauss matrices: Compression with neuroscience-based constraints. *Proceedings of the National Academy of Sciences*, 111(47):16872–16876, 2014.
- 3 Noga Alon and Asaf Nussboim. K -wise independent random graphs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 813–822. IEEE, 2008.
- 4 Marshall Ball, Dana Dachman-Soled, Siyao Guo, Tal Malkin, and Li-Yang Tan. Non-malleable codes for small-depth circuits. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 826–837. IEEE, 2018.
- 5 Boaz Barak, Ronen Shaltiel, and Eran Tromer. True random number generators secure in a changing environment. In *International Workshop on Cryptographic Hardware and Embedded Systems*, pages 166–180. Springer, 2003.
- 6 Louay MJ Bazzi. Polylogarithmic independence can fool dnf formulas. *SIAM Journal on Computing*, 38(6):2220–2272, 2009.
- 7 Mihir Bellare and John Rompel. Randomness-efficient oblivious sampling. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 276–287. IEEE, 1994.
- 8 George Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.
- 9 Jöran Bergh and Jörgen Löfström. *Interpolation spaces: an introduction*, volume 223. Springer Science & Business Media, 2012.
- 10 SN Bernstein. *Probability theory, ogiz, moscow–leningrad* (1946).
- 11 Henry W Block, Thomas H Savits, Moshe Shaked, et al. Some concepts of negative dependence. *The Annals of Probability*, 10(3):765–772, 1982.

- 12 Stéphane Boucheron, Olivier Bousquet, Gábor Lugosi, Pascal Massart, et al. Moment inequalities for functions of independent random variables. *The Annals of Probability*, 33(2):514–560, 2005.
- 13 Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.
- 14 Keren Censor-Hillel, Merav Parter, and Gregory Schwartzman. Derandomizing local distributed algorithms under bandwidth restrictions. *Distributed Computing*, 33(3):349–366, 2020.
- 15 Herman Chernoff et al. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952.
- 16 Hamza ES Daoub. The fundamental theorem on symmetric polynomials. *The Teaching of Mathematics*, 28:55–59, 2012.
- 17 Yevgeniy Dodis, Krzysztof Pietrzak, and Daniel Wichs. Key derivation without entropy waste. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 93–110. Springer, 2014.
- 18 Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.
- 19 Devdatt P Dubhashi and Desh Ranjan. Balls and bins: A study in negative dependence. *BRICS Report Series*, 3(25), 1996.
- 20 Morris L Eaton. A note on symmetric bernoulli random variables. *The annals of mathematical statistics*, 41(4):1223–1226, 1970.
- 21 Morris L Eaton. A review of selected topics in multivariate probability inequalities. *The Annals of Statistics*, pages 11–43, 1982.
- 22 T Figiel, P Hitczenko, W Johnson, G Schechtman, and J Zinn. Extremal properties of rademacher functions with applications to the khintchine and rosenthal inequalities. *Transactions of the American Mathematical Society*, 349(3):997–1027, 1997.
- 23 Mohsen Ghaffari and Fabian Kuhn. Derandomizing distributed algorithms with small messages: Spanners and dominating set. In *32nd International Symposium on Distributed Computing (DISC 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.
- 24 Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.
- 25 Russell Impagliazzo, Raghu Meka, and David Zuckerman. Pseudorandomness from shrinkage. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 111–119. IEEE, 2012.
- 26 Meena Jagadeesan. Simple analysis of sparse, sign-consistent jl. *arXiv preprint*, 2017. [arXiv:1708.02966](https://arxiv.org/abs/1708.02966).
- 27 Svante Janson. New versions of suen’s correlation inequality. *Random Structures and Algorithms*, 13(3-4):467–483, 1998.
- 28 Daniel M Kane and Jelani Nelson. A derandomized sparse johnson-lindenstrauss transform. *arXiv preprint*, 2010. [arXiv:1006.3585](https://arxiv.org/abs/1006.3585).
- 29 David R Karger and Matthias Ruhl. Simple efficient load balancing algorithms for peer-to-peer systems. In *Proceedings of the sixteenth annual ACM symposium on Parallelism in algorithms and architectures*, pages 36–43, 2004.
- 30 Andreas Knoblauch. Closed-form expressions for the moments of the binomial probability distribution. *SIAM Journal on Applied Mathematics*, 69(1):197–204, 2008.
- 31 MJ Kronenburg. The binomial coefficient for negative arguments. *arXiv preprint*, 2011. [arXiv:1105.3689](https://arxiv.org/abs/1105.3689).
- 32 Rafał Latała et al. Estimation of moments of sums of independent real random variables. *The Annals of Probability*, 25(3):1502–1513, 1997. URL: https://projecteuclid.org/download/pdf_1/euclid.aop/1024404522.
- 33 John E Littlewood. On the probability in the tail of a binomial distribution. *Advances in Applied Probability*, 1(1):43–72, 1969.

- 34 Michael Luby, Michael George Luby, and Avi Wigderson. *Pairwise independence and derandomization*, volume 4. Now Publishers Inc, 2006.
- 35 Colin Maclaurin. A second letter to martin folkes, esq.; concerning the roots of equations, with demonstration of other rules of algebra. *Philos. Trans. Roy. Soc. London Ser. A*, 36:59–96, 1729.
- 36 Andrew McGregor and Hoa T Vu. Better streaming algorithms for the maximum coverage problem. *Theory of Computing Systems*, 63(7):1595–1619, 2019.
- 37 Brendan D McKay. On littlewood’s estimate for the binomial distribution. *Advances in Applied Probability*, 21(2):475–478, 1989.
- 38 Abbas Mehrabian. Summary of concentration inequalities for the sum of k-wise independent random variables, 2011. URL: https://www.cs.mcgill.ca/~amehra13/Articles/kwise_independent_concentration_summary.pdf.
- 39 Dragoslav S. Mitrinović. Certain inequalities for elementary symmetric functions. *Publikacije Elektrotehničkog fakulteta. Serija Matematika i fizika*, (181/196):17–20, 1967. URL: <http://www.jstor.org/stable/43667273>.
- 40 Gergő Nemes. On the coefficients of the asymptotic expansion of $n!$ *arXiv preprint*, 2010. [arXiv:1003.2907](https://arxiv.org/abs/1003.2907).
- 41 Isaac Newton. *Arithmetica universalis sive de compositione et resolutione arithmetica liber*, volume 1. apud Marcum Michaellem Rey, 1761.
- 42 Constantin Niculescu and Lars-Erik Persson. *Convex functions and their applications*. Springer, 2006.
- 43 Maciej Obremski and Maciej Skorski. Renyi entropy estimation revisited. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.
- 44 Thomas K Philips and Randolph Nelson. The moment bound is tighter than chernoff’s bound for positive tail probabilities. *The American Statistician*, 49(2):175–178, 1995.
- 45 Arjun Ramachandra and Karthik Natarajan. Tight probability bounds with pairwise independence. *arXiv preprint*, 2020. [arXiv:2006.00516](https://arxiv.org/abs/2006.00516).
- 46 Herbert Robbins. A remark on stirling’s formula. *The American mathematical monthly*, 62(1):26–29, 1955.
- 47 Haskell P Rosenthal. On the subspaces of ℓ_p ($p > 2$) spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8(3):273–303, 1970.
- 48 Shmuel Rosset. Normalized symmetric functions, newton’s inequalities, and a new set of stronger inequalities. *The American Mathematical Monthly*, 96(9):815–819, 1989.
- 49 Jörg-Rüdiger Sack and Jorge Urrutia. *Handbook of computational geometry*. Elsevier, 1999.
- 50 Jeanette P Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff–hoeffding bounds for applications with limited independence. *SIAM Journal on Discrete Mathematics*, 8(2):223–250, 1995.
- 51 Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.