

ON THE CONCENTRATION OF MULTI-VARIATE POLYNOMIALS WITH SMALL EXPECTATION

VAN. H. VU
MICROSOFT RESEARCH
MICROSOFT CORPORATION
ONE MICROSOFT WAY
REDMOND, WA 98052
VANHAVU@ MICROSOFT.COM

ABSTRACT. Let t_1, \dots, t_n be independent, but not necessarily identical, $\{0, 1\}$ random variables. We prove a general large deviation bound for multi-variate polynomials (in t_1, \dots, t_n) with small expectation (order $O(\text{polylog}(n))$). Few applications in random graphs and combinatorial number theory will be discussed.

Our result is closely related to a classical result of Janson [Jan]. Both of them can be applied in similar situations. On the other hand, our result is symmetric, while Janson's inequality only deals with the lower tail probability.

§1 INTRODUCTION

Let t_1, \dots, t_n be independent $\{0, 1\}$ random variables (through the paper we call these r.v's *atom* r.v's) and \mathbb{S} be the product space spanned by them. A common task in probabilistic combinatorics is to show that a certain function $Y = Y(t_1, \dots, t_n)$ from \mathbb{S} to \mathbb{R} is strongly concentrated around its mean (in other words, with high probability $Y \approx \mathbb{E}(Y)$).

Very frequently, a function Y of interest can be written in the form $Y = \sum_{i=1}^m I_i$, where I_i is a product of few t_j 's (see §5, also [AS], chapter 8). The purpose of this paper is to prove a concentration result for such functions Y , when the expectation of Y is small (at most $\text{polylog}(n)$). The case when the expectation of Y is large was studied in two other papers [KV, Vu1] and will be briefly discussed here.

Readers familiar with probabilistic combinatorics would immediately realize that the famous result of Janson [Jan] provides a strong bound for the lower tail probability of functions of the above type. We write $I_i \sim I_j$ if the two monomials share a common atom variable.

Part of this work was done while the author was with the Institute for Advanced Study (Princeton, NJ) and was supported by a grant from NEC and the state of New Jersey

Janson’s inequality. *With Y as above and $\Delta = \sum_{I_i \sim I_j} \mathbb{E}(I_i I_j)$, the following holds*

$$\Pr(Y \leq (1 - \epsilon)\mathbb{E}(Y)) \leq e^{-\frac{(\epsilon\mathbb{E}(Y))^2}{2(\mathbb{E}(Y) + \Delta)}}.$$

The main goal of this paper is to prove a large deviation inequality which can be applied for both upper tail and lower tail. The strength of our bound and that of the bound given by Janson’s inequality are comparable in several examples. By this reason, we believe that our result would have a wide range of applications.

Instead of functions of type $Y = \sum I_j$, we consider a wider class of functions. A polynomial Y is *positive* if it can be written as $Y = \sum_{i=1}^m c_i I_i$, where c_i are positive. We say that a polynomial Y is *normal* if it is positive, its free coefficient is 0, and all other coefficients are at most 1. Our result is proven for the class of normal polynomials. We say that Y is *homogeneous* of degree k if every monomial of Y has degree k .

Let us start by describing our intuition which leads to the main theorem of this paper and several other concentration results proven in [KV] and [Vu1]. For a more detailed discussion, see [Vu1].

The general phenomenon in the theory of concentration is the following: *if a function Y is smooth, then Y is strongly concentrated.* The usual way to quantify “smoothness” is to require that a function has small Lipschitz coefficient, i.e., changing any variable does not change the value of the function by more than a constant (see [Tal] where this phenomenon is discussed in detail). Several classical concentration results such as Azuma’s inequality are based on this definition of smoothness (see [Tal] or [AS], chapter 7).

Restricting ourself to the class of positive polynomials, we propose a new way to define “smoothness”. Consider a positive polynomial Y , we say that Y is “smooth” if the expectation of any partial derivative (of any order of Y) is small. This is, in a sense, equivalent to saying that Y is smooth on average. Our intuition is that, for certain cases, the “average smoothness” would be already enough to guarantee strong concentration. The meta-theorem we have in mind is the following

If a positive polynomial Y of “low” degree is “smooth” (in the new sense), then it is “strongly” concentrated around its mean.

The key advantage of this new notion of smoothness is that its requirement is very mild. It occurs quite frequently that a function we want to study does not have small Lipschitz coefficient, but the expectation of any of its partial derivatives is small. The reader would convince himself about this point through the examples given in the paper.

For any multi-set A , let $\partial_A Y$ denote the partial derivative of Y according to A . For instance, if $Y = t_1^2 t_2 t_3 + t_2^2 t_3 t_4$ and $A = \{1, 1\}$, $B = \{2, 3\}$ then $\partial_A(Y) = 2t_2 t_3$ and $\partial_B(Y) = t_1^2 + 2t_2 t_4$.

The first result based on this intuition was proven in [KV], using a different terminology and can be generalized to the following [Vu1].

Theorem 1.1. *There is a constant c_k depending on k such that the following holds. Let $Y(t_1, \dots, t_n)$ be a positive polynomial of degree k , where t_i can have arbitrary distribution on the interval $[0, 1]$. Assume that*

$$\mathbb{E}(Y) \geq \mathbb{E}'(Y) = \max_{A, 0 < |A| \leq k} \mathbb{E}(\partial_A(Y)).$$

Then for any $\lambda \geq 1$

$$Pr(|Y - \mathbb{E}(Y)| \geq c_k \lambda^k (\mathbb{E}(Y)\mathbb{E}'(Y))^{1/2}) \leq e^{-\lambda + (k-1) \log n}.$$

Theorem 1.1 is useful when the ratio $\mathbb{E}(Y)/\mathbb{E}'(Y)$ is large (in other words, the expectation of Y is much larger than that of any of its partial derivatives). For instance, if this ratio is at least n^γ for some positive constant γ , then for any constant $\epsilon > 0$ there is a constant $\nu = \nu(\epsilon, \gamma)$ such that

$$(1.1) \quad Pr(|Y - \mathbb{E}(Y)| \geq \epsilon \mathbb{E}(Y)) \leq e^{-n^\nu}.$$

The result in [KV], Theorem 1.1 and their variations have several applications, especially when combined with the semi-random method (see [KV2, Vu1, Vu2, Vu3, Vu5]).

Unfortunately, Theorem 1.1 loses its power when $\mathbb{E}(Y)$ is small (of order $O(\log n)$, say). Recall that in combinatorial applications our objective function is frequently of the form $\sum_{i=1}^n I_i$, so $\mathbb{E}'(Y)$ is at least 1. Furthermore, we usually need λ to be at least of order $\Omega(\log n)$. Thus, the tail in Theorem 1.1 is of order at least $\Omega(\log^k n (\mathbb{E}(Y))^{1/2})$, which will be larger than the expectation of Y if $\mathbb{E}(Y) = o(\log^{2k} n)$. Although the theorem still says something, it is not too practical, since in applications we usually require that the tail is small compared to the mean.

The goal of this paper is to prove a concentration result which deals with the case $\mathbb{E}(Y) = o(\log^{2k} n)$. Our favorite range is when $\mathbb{E}(Y) = \Theta(\log n)$ and $\mathbb{E}(Y)/\log n$ is large. The following theorem covers this case.

Theorem 1.2. *For any positive constants $k, \alpha, \beta, \epsilon$ there is a constant $Q = Q(k, \epsilon, \alpha, \beta)$ such that the following holds. If Y is normal and homogeneous of degree k , $n/Q \geq \mathbb{E}(Y) \geq Q \log n$ and $\mathbb{E}(\partial_A(Y)) \leq n^{-\alpha}$ for all non-empty set A of cardinality at most $k - 1$, then*

$$Pr(|Y - \mathbb{E}(Y)| \geq \epsilon \mathbb{E}(Y)) \leq n^{-\beta}.$$

Although the assumption of Theorem 1.2 might seem a bit technical, it is, in fact, quite general. In many combinatorial applications, the function we are interested in

is homogeneous and normal. Moreover, when $\mathbb{E}(Y)$ is of order $\Theta(\log n)$, then very frequently $\mathbb{E}(\partial_A(Y))$ is sufficiently small for all admissible sets A (see the example below and other applications in §5). Moreover, the deviation bound $n^{-\Omega(1)}$ is (in general) the best one may hope for for a function with expectation of order $O(\log n)$.

Example. Let Y be the number of triangles in $G(m, p)$. Set $n = \binom{m}{2}$; we have n random variables t_{ij} , $1 \leq i < j \leq m$. Y can be written as $\sum_{1 \leq i < j < k \leq m} t_{ij} t_{jk} t_{ki}$. If p is chosen so that $\mathbb{E}(Y) = \Theta(\log n)$, then it is easy to check that $\mathbb{E}(\partial_A(Y)) = O(n^{-1+o(1)})$ for all admissible sets A . Thus, Theorem 1.2 applies and implies that if $\mathbb{E}(Y) = Q \log n$ and Q is sufficiently large, then $\Pr(|Y - \mathbb{E}(Y)| > \mathbb{E}(Y)) \leq n^{-g(Q)}$, where $g(Q) \rightarrow \infty$ as $Q \rightarrow \infty$. This implies that $\Pr(Y = 0) \leq n^{-g(Q)}$ and $\Pr(Y \geq 2\mathbb{E}(Y)) \leq n^{-g(Q)}$. The reader who is familiar with the theory of random graphs would recognize that the bound on the lower tail ($\Pr(Y = 0)$) is a special case of the well known result by Janson, Łuczak and Ruciński [LJR] on the probability that $G(n, p)$ does not contain a copy of a fixed graph, which can be proven using Janson's inequality. On the other hand, Janson's inequality do not give any bound for the upper tail.

Theorem 1.2 is a corollary of Theorem 1.3, which is the main theorem of this paper. To state this theorem, we need few more definitions.

Through the paper, $\mathcal{N} = \{1, 2, \dots, n\}$ and I_j denotes a monomial of form $t_{i_1} \dots t_{i_l}$ where $l \leq k$. If i_1, \dots, i_l are different, we say that the monomial is *simple*. In this case, we can think of I_j as both a monomial and a set ($I_j = \{i_1, \dots, i_l\}$). Consequently, if A is a subset of \mathcal{N} , then $I_j^A = I_j \setminus A$ can be interpreted as the monomial $\prod_{i \in I_j \setminus A} t_i$. A polynomial Y is *simplified* if it contains only simple monomials. Since we are dealing with $\{0, 1\}$ random variables, any polynomial Y has a unique simplification, and it is more convenient and natural to deal with simplified polynomials.

Consider a normal simplified polynomial $Y = \sum_{j=1}^m c_j I_j$. For a set $A \subset \mathcal{N}$ (A can be the empty set), let $\mathbb{E}_A(Y) = \mathbb{E}(\sum_{A \subsetneq I_j} c_j I_j^A)$. In particular, if A is the empty set, then $\mathbb{E}_\emptyset(Y) = \mathbb{E}(Y)$. The following definition plays a key role in the paper.

Definition. For any $0 \leq j \leq k-1$: $\mathbb{E}_j(Y) = \max_{A \subset \mathcal{N}, |A| \geq j} \mathbb{E}_A(Y)$.

Set $f(K) = \max\{1, \lceil (K/k!)^{1/k} \rceil - 1\}$, $b(k, n) = \sum_{j=0}^{k-1} \binom{n}{j}$. Furthermore, set

$$r(k, K, n, \delta) = 2 \frac{b(k, n) \delta^{\frac{f(K/2)}{2}}}{f(K/2)!} + (\delta^{1/8}/K)^{\lfloor \frac{1}{8k} \log \frac{1}{\delta} \rfloor}.$$

Given a normal polynomial Y with expectation $\mathbb{E}(Y)$, define $h(k, K, n, \delta)$ recursively as follows:

$$h(1, K, n, \delta) = 0; h(k, K, n, \delta) = h(k-1, K, n + \lceil \mathbb{E}(Y) \rceil, \delta) + nr(k-1, K, n, \delta).$$

If $\mathbb{E}(Y) \leq n/Q$, for a sufficiently large Q , then it follows, by a rough estimate, that $h(k, K, n, \delta) \leq 2knr(k-1, K, n, \delta)$.

Theorem 1.3. (*Main Theorem*) *Let Y be a simplified normal polynomial of degree at most k . Suppose that there are positive numbers δ , λ , and K satisfying $K \geq 2k$, $\mathbb{E}_1(Y) \leq \delta \leq 1$ and $4kK\lambda \leq \mathbb{E}(Y)$. Then*

$$Pr(|Y - \mathbb{E}(Y)| \geq (4kK\lambda\mathbb{E}(Y))^{1/2}) \leq 2ke^{-\lambda/4} + h(k, K, n, \delta).$$

Remark. The number k in this theorem does need to be bounded. However, there is a trade-off between k and K . First, K should be large compared to k to keep $h(k, K, n, \delta)$ small. On the other hand, a very large K may blow up the tail. So the theorem gives a good bound only in the case k tends to infinity sufficiently slowly.

Before showing that Theorem 1.3 implies Theorem 1.2, let us mention the following delicate point. In Theorem 1.2, we do not require Y to be simplified. This non-restriction appears to be convenient in applications (see §5). Although not mentioned explicitly, information about the expectations of partial derivatives is contained in the definition of \mathbb{E}_A .

To see that Main Theorem implies Theorem 1.2, consider a polynomial Y as described in Theorem 1.2 and let Y^{sim} denote the (unique) simplification of Y . We apply Theorem 1.3 to Y^{sim} .

For any non-empty set $A \subset \mathcal{N}$, let \mathcal{A} denote the family of multi-sets of size at most $k-1$ obtained from A by repeating its elements (for instance, if $k=5$ and $A = \{1, 2\}$, then \mathcal{A} contains 6 multi-sets $\{1, 2\}, \{1, 1, 2\}, \{1, 2, 2\}, \{1, 1, 1, 2\}, \{1, 1, 2, 2\}, \{1, 2, 2, 2\}$). It is clear that for any fixed A , $|\mathcal{A}| \leq a(k)$, where $a(k)$ is a number depend only on k . Set $\delta = a(k)n^{-\alpha}$, one can verify that for any non-empty set A ,

$$\mathbb{E}_A(Y^{sim}) \leq \sum_{A' \in \mathcal{A}} \mathbb{E}(\partial_{A'}(Y)) \leq a(k)n^{-\alpha} = \delta.$$

Set $\lambda = (4\beta + 1) \log n$ and choose a constant $K = K(k, \alpha, \beta)$ sufficiently large so that $h(k, K, n, a(k)n^{-\alpha}) \leq n^{-\beta-1}$ (using the fact that $\mathbb{E}(Y)/n$ is sufficiently small). Apply Theorem 1.3 to Y^{sim} , the deviation bound obtained is less than $n^{-\beta}$. On the other hand, if $\mathbb{E}(Y)/\log n \geq Q = Q(K, \epsilon)$, then the tail $(2kK\lambda\mathbb{E}(Y))^{1/2} \leq \epsilon\mathbb{E}(Y)$.

Without any difficulty, one may state a theorem similar to Theorem 1.2 when $\mathbb{E}(Y)$ is of order other than $\Theta(\log n)$. For instance, one may derive the following.

Corollary 1.4. *Assume that Y is a normal homogeneous polynomial of degree k and the expectation of Y is $f \log n$, where $0 < f \leq 1$ can be a function depending on n . Assume furthermore that for all A , $1 \leq |A| \leq k-1$, $\mathbb{E}(\partial_A(Y)) \leq n^{-\alpha}$ for some positive constant α . Then there are positive constants $c = c(\alpha, k)$ and $d = d(\alpha, k)$ such that for any $0 \leq \epsilon \leq 1$,*

$$Pr(|Y - \mathbb{E}(Y)| \geq \epsilon\mathbb{E}(Y)) \leq de^{-c\epsilon^2\mathbb{E}(Y)}.$$

The proof of Theorem 1.3 relies on Main Lemma I and Main Lemma II, which are proved in the next two sections. These lemmas are of independent interest and their proofs require some non-trivial ideas. The proof of Theorem 1.3 follows in §4 and few applications are discussed in §5. These applications focus only on the author recent interest and are, by no mean, exclusive. We encourage the reader to contact us if he or she finds a new application.

Let us finish the current section by posing a question. First notice that Theorem 1.1 holds without the restriction that t_i 's are $\{0, 1\}$. On the other hand, we do use the fact that t_i has only two possible values 0 or 1 in our proof of Theorem 1.3.

Question. *Can one prove Theorem 1.3 (or a comparable statement) without the restriction that the t_i 's are $\{0, 1\}$ random variables ?*

§2 MAIN LEMMA I

Given n atom variables t_1, \dots, t_n , a positive number δ and a positive integer k , we denote by $\mathbf{Poly}_k(\delta)$ the set of all simplified normal polynomials $Y = Y(t_1, \dots, t_n)$ of degree at most k satisfying $\mathbb{E}_0(Y) \leq \delta$.

In this section, we always assume that a monomial is simple. We recall the following definition from the previous section $f(K) = \max\{1, \lceil (K/k!)^{1/k} \rceil - 1\}$.

Main Lemma I. *Assume that $1 \geq \delta > 0$. For any $K > 0$ and any $Y \in \mathbf{Poly}_k(\delta)$*

$$Pr(Y \geq K) \leq 2 \frac{b(k, n) \delta^{\frac{f(K/2)}{2}}}{f(K/2)!} + (\delta^{1/8}/K)^{\lfloor \frac{1}{8k} \log \frac{1}{\delta} \rfloor}.$$

Remark. The upper bound is the function $r(k, K, n, \delta)$ defined in the previous section.

Consider a random variable $X = I_1 + \dots + I_m$. A family $D = \{I_1, \dots, I_r\}$ is *disjoint* if I_i 's (as sets) are pairwise disjoint. Let \mathcal{D} be the collection of all disjoint families in X ; define $Disfam(X) = \max_{D \in \mathcal{D}} \sum_{j \in D} I_j$.

Lemma 2.1. *Suppose that $X = \sum_{j=1}^m I_j$ and $\mathbb{E}(X) = \gamma$, then*

$$Pr(Disfam(X) \geq s) \leq \gamma^s / s!.$$

The proof of this lemma is relatively simple and we leave it to the reader as an exercise. A detailed proof can be found in [AS].

We say that the sets A_1, \dots, A_r form a sun flower if they have pairwise the same intersection. The following lemma was proven by Erdős and Rado [ERa].

Lemma 2.2. *(Sun flower) If H is a hypergraph with edges of size at most k and H has more than $(r-1)^k k!$ edges then there are r edges forming a sun flower.*

We say that a sunflower is *strong* if no petal contains another. A sunflower with r petals must contain a strong sunflower with at least $r-1$ petals. Thus, the previous

lemma implies that if H has $K \geq 1$ edges of size at most k , then one can choose from these edges a strong sun flower of size at least $f(K)$.

Lemma 2.3. *Let $X = I_1 + \dots + I_m$ where I_j are (different) monomials of degree at most k and $\mathbb{E}_0(X) = \gamma$. Then for any positive number K ,*

$$Pr(X \geq K) \leq b(k, n)\gamma^{f(K)}/f(K)!.$$

Proof. Set $q = f(K)$. For any $t = (t_1, \dots, t_n)$ consider the hypergraph H_t whose edges are those I_j where $I_j(t) = 1$. If $X(t) \geq K$, then by Lemma 2.2 H_t contains a strong sunflower J_1, \dots, J_q , where $J_r \in \{I_1, \dots, I_m\}$. Setting $A = \cap_{r=1}^q J_r$, the sets $G_r = J_r \setminus A$ are pairwise disjoint.

Consider the random variable $X_A = \sum_{A \subset I_j} I_j^A$, where $I_j^A = I_j/A$. By the assumption of the lemma $\mathbb{E}(X_A) \leq \mathbb{E}_0(X) = \gamma$, thus by Lemma 2.1,

$$Pr(Disfam(X_A) \geq q) \leq \gamma^q/q!.$$

Since $|A| \leq k - 1$, there are less than $b(k, n)$ (recall that $b(k, n) = \sum_{j=0}^{k-1} \binom{n}{j}$) possibility to choose A , the statement follows. \square

Lemma 2.4. *Assume that $Y \in \mathbf{Poly}_k(\delta)$ with $\delta \leq 1$. Then*

$$Pr(Y \geq M) \leq 2^{ks^2-s}/M^s,$$

for any $M > 0$ and any non-negative integer s .

Proof. We use the moment method. By Markov's inequality, it suffices to prove that

$$\mathbb{E}(Y^s) \leq 2^{ks^2-s}.$$

Let us note that the case $k = 1$ and $s = 2$ is easy to verify. In this case one can show that $\mathbb{E}(Y)^2 \leq 4$ using the fact that Y is a sum of independent variables J_i (since $k = 1$).

The general case requires little more work. Define $G_k(s, \delta) = \max_{Y \in \mathbf{Poly}_k(\delta)} \mathbb{E}(Y^s)$. It is clear that $G_k(0, \delta) = 1$. For two polynomials Z and Z' , we write $Z \prec Z'$ if $Z' - Z$ is positive or 0.

Consider $Y = \sum_{i=1}^m c_i I_i \in \mathbf{Poly}_k(\delta)$. Let $c_i I_i = J_i$ (again we think about J_i as both a monomial and a set; in particular, we say $J_i \subset J_j$ if $I_i \subset I_j$). We have $J_i^2 = c_i^2 I_i^2$ etc. Since $c_i \leq 1$, $J_i^l \prec J_i$ for any $l \geq 2$.

Given $Y = \sum_{i=1}^m J_i$, we denote by Y_s the sum of products of form $J_{i_1} \dots J_{i_s}$, where i_1, \dots, i_s run over the set of all different ordered s tuples (if $s > m$, this set is empty).

Claim. For any s ,

$$Y^s \prec (s-1)^2 Y^{s-1} + Y_s.$$

Proof. We use induction on s . The statement is trivial for $s = 1$. Assuming that it holds for s , it follows that

$$Y^{s+1} = Y^s Y \prec ((s-1)^2 Y^{s-1} + Y_s) Y = (s-1)^2 Y^s + Y_s Y.$$

Consider, for instance, a term $X = J_1 \dots J_s$ of Y_s . If $s \leq m$ then $XY \prec \sum_{l>s} J_1 \dots J_s J_l + s J_1 \dots J_s$. Since the sum (which might be empty) belongs to Y_{s+1} and $J_1 \dots J_s$ is an element of Y^s , we have

$$Y^{s+1} \prec (s-1)^2 Y^s + s Y^s + Y_{s+1} \prec s^2 Y^s + Y_{s+1}.$$

The case $s > m$ is trivial, since $Y_s = 0$. \square

For any $1 \leq i \leq m$, let \mathcal{S}_i be the sum of all ordered products $J_{i_1} \dots J_{i_{s-1}}$, where i_j ($j = 1, \dots, s-1$) are different and $J_{i_j} \not\subseteq J_i$. One can show

$$Y_s \prec s \sum_{i=1}^m J_i \mathcal{S}_i.$$

On the other hand,

$$\mathcal{S}_i \prec \left(\sum_{A \subset J_i} \sum_{J_l \cap J_i = A, J_l \neq A} J_l^A \right)^{s-1} = Z_i^{s-1}.$$

By the definition of $\mathbb{E}_0(\cdot)$, we have $\mathbb{E}_0(\sum_{J_l \cap J_i = A, J_l \neq A} J_l^A) \leq \delta$. Since J_i has at most 2^k subsets (empty set included), it follows that $Z_i/2^k \in \mathbf{Poly}_k(\delta)$. Therefore, $\mathbb{E}(Z_i^{s-1}) \leq 2^{k(s-1)} G(s-1, \delta)$. Thus, we can conclude that

$$\mathbb{E}(Y_s) \leq s 2^{k(s-1)} G(s-1, \delta) \sum_{i=1}^m \mathbb{E}(J_i) \leq s 2^{k(s-1)} G(s-1, \delta),$$

because $\sum_{i=1}^m \mathbb{E}(J_i) \leq \delta \leq 1$. Together with the claim, we have

$$\mathbb{E}(Y^s) \leq G(s-1, \delta) ((s-1)^2 + s 2^{k(s-1)}).$$

Since this holds for any $Y \in \mathbf{Poly}_k(\delta)$, it follows that

$$G(s, \delta) \leq G(s-1, \delta) ((s-1)^2 + s 2^{k(s-1)}).$$

We show that $G(s, \delta) \leq 2^{ks^2-s}$ by induction. If $k = 1$, start from $s = 2$. For $k > 1$, one can start from $s = 1$. The simple details are omitted. \square

In the following let $F(\delta, K) = \sup_{Y \in \mathbf{Poly}_k(\delta)} Pr(Y \geq K)$.

Lemma 2.5.

$$F(\delta, K) \leq \frac{b(k, n)(4\delta)^{f(K/2)}}{f(K/2)!} + F(4\delta, 2K).$$

Proof. Consider $Y = \sum_{i=1}^m c_j I_j \in \mathbf{Poly}_k(\delta)$. Let $Y_1 = \sum_{c_j \geq 1/4} c_j I_j$ and $Y_2 = \sum_{c_j < 1/4} c_j I_j$. It is trivial that

$$Pr(Y \geq K) \leq Pr(Y_1 \geq K/2) + Pr(Y_2 \geq K/2).$$

Set $X_1 = \sum_{j, c_j \geq 1/4} I_j$; it follows that $\mathbb{E}_0(X_1) \leq 4\delta$. By Lemma 2.3,

$$Pr(Y_1 \geq K/2) \leq Pr(X_1 \geq K/2) \leq \frac{b(k, n)(4\delta)^{f(K/2)}}{f(K/2)!}.$$

Furthermore, by definition, $4Y_2 \in \mathbf{Poly}_k(4\delta)$ (all coefficients in Y_2 are less than $1/4$, thus all coefficients in $4Y_2$ are less than 1). This yields,

$$Pr(Y_2 \geq K/2) = Pr(4Y_2 \geq 2K) \leq F(4\delta, 2K).$$

This ends the proof. \square

Now we complete the proof of Main Lemma I. By Lemma 2.5 and induction, we have

$$F(\delta, K) \leq \sum_{i=1}^l \frac{b(k, n)(4^i \delta)^{f(K_i)}}{f(K_i)!} + F(4^l \delta, 2^l K)$$

for any $l = 0, 1, 2, \dots$, where $K_i = 2^{i-2}K \geq K/2$. Choosing l such that $4^l \delta \leq \delta^{1/2} < 4^{l+1} \delta$ gives

$$F(\delta, K) \leq \sum_{i=1}^l \frac{b(k, n)(4^{i-l} \delta^{1/2})^{f(K/2)}}{f(K/2)!} + F(\delta^{1/2}, K\delta^{-1/4}/2).$$

By Lemma 2.4,

$$F(\delta^{1/2}, K\delta^{-1/4}/2) \leq \frac{2^{ks^2-s}}{(K\delta^{-1/4}/2)^s} = \frac{2^{ks^2} \delta^{s/4}}{K^s},$$

for any non-negative integer s . Choosing $s = \lfloor \frac{1}{8k} \log \frac{1}{\delta} \rfloor$ gives

$$F(\delta^{1/2}, K\delta^{-1/4}) \leq (\delta^{1/8}/K)^{\lfloor \frac{1}{8k} \log \frac{1}{\delta} \rfloor}.$$

Notice that $\sum_{i=1}^l 4^{i-l} < 2$, we have

$$F(\delta, K) \leq \frac{2b(k, n)\delta^{f(K/2)/2}}{f(K/2)!} + (\delta^{1/8}/K)^{\lfloor \frac{1}{8k} \log \frac{1}{\delta} \rfloor},$$

completing the proof. \square

Corollary 2.6. *For any $\alpha, \beta > 0$ there is a constant $K = K(\alpha, \beta)$ such that if $\delta \leq n^{-\alpha}$ then $Pr(Y \geq K) < n^{-\beta}$, for any $Y \in \mathbf{Poly}_k(\delta)$.*

§3 MAIN LEMMA II

Given a function $Y = Y(t_1, \dots, t_n)$ from \mathbb{S} to \mathbb{R} , one can determine the value of $Y(t_1, \dots, t_n)$ using a decision tree structure. We consider a decision tree of depth n , and at a node at level i , we ask the question “what is the value of t_i ”. If the answer is 1, then we go to the right hand side child of the recent node, if the answer is 0 then we go to the left and continue until we reach a leaf. There will be 2^n leaves representing the vectors in the space. In general a node at level i will be labeled by a 0, 1 vector of length i , which is the sequence of answer leading to this node. We label the root by the empty set and at each leaf $t = (t_1, \dots, t_n)$ we write the corresponding value of $Y(t)$. Let p_i denote the expectation of t_i and $q_i = 1 - p_i$.

For any leaf t , let $t^i = (t_1, \dots, t_i)$ be the vector formed by the first i coordinates of t . The vectors t^i 's ($t \in \mathbb{S}$) label the nodes at level i^{th} . For a node a at level i^{th} , we let $E(a)$ denote the expected value of the leaves below a , namely:

$$E(a) = \sum_{t, t^i=a} Y(t) \prod_{j=i+1}^n Pr(t_j).$$

By definition we have at the root that $E(\emptyset) = \mathbb{E}(Y)$. If a is a vector of length i and b is a vector of length j , then $\langle a, b \rangle$ denotes the vector of length $i + j$ obtained by writing b behind a . For $z = 0$ or 1, let

$$\mu_{i,z}(t) = E(\langle t^{i-1}, z \rangle) - E(t^{i-1}).$$

It is easy to compute that

$$\mu_{i,1}(t) = q_i(E(\langle t^{i-1}, 1 \rangle) - E(\langle t^{i-1}, 0 \rangle))$$

and

$$\mu_{i,0}(t) = p_i(E(\langle t^{i-1}, 0 \rangle) - E(\langle t^{i-1}, 1 \rangle))$$

Set

$$V_i(t) = p_i \mu_{i,1}(t)^2 + q_i \mu_{i,0}(t)^2.$$

We have

$$V_i(t) = p_i q_i (E(\langle t^{i-1}, 1 \rangle) - E(\langle t^{i-1}, 0 \rangle))^2 \leq p_i C_i(t)^2,$$

where $C_i(t) = |E(\langle t^{i-1}, 1 \rangle) - E(\langle t^{i-1}, 0 \rangle)|$. Denote by $c(t)$ the maximum value of $C_i(t)$ over all possible choice of i . Let $c_Y = \max_t c(t)$. Finally, let

$$V(t) = \sum_{i=1}^n V_i(t) \quad \text{and} \quad V_Y = \max_t V(t).$$

It is apparent that $V(t) \leq c_Y \sum_{i=1}^n p_i C_i(t)$.

The following lemma (Main Lemma II) was proven in [KV]. Since the proof is short, we repeat it here (with a slight modification) for the sake of completeness.

Main Lemma II. *Let \mathbf{V} and \mathbf{c} be two arbitrary positive numbers and $\mathbf{B} = \{t | c(t) \geq \mathbf{c} \text{ or } V(t) \geq \mathbf{V}\}$. If $0 < \lambda \geq \mathbf{V}/\mathbf{c}^2$ then*

$$Pr(|Y - E(Y)| \geq (\lambda \mathbf{V})^{1/2}) \leq 2e^{-\lambda/4} + Pr(\mathbf{B}).$$

Remark. Main Lemma II is not restricted to polynomials.

Proof. It is easy to see that $C_i(t)$ and $V_i(t)$ are invariant under shifting. Thus, without loss of generality, we can assume that $E(Y) = 0$.

For any $t \in \mathbf{B}$, let $i(t)$ be the smallest index i (between 1 and n) such that either $C_i(t) > \mathbf{c}$ or $\sum_{j=1}^i V_j(t) > \mathbf{V}$. Let $\mathbf{B}_t = \{z \in \mathbb{S} | z_i = t_i \text{ for all } i < i(t)\}$. It is clear that

- \mathbf{B}_t is a subhypercube
- $\mathbf{B}_t \subset \mathbf{B}$
- For any $t, t' \in \mathbf{B}$, \mathbf{B}_t and $\mathbf{B}_{t'}$ are either identical or disjoint.

It follows that \mathbf{B} is a disjoint union of finite subhypercubes. Now define a function Y' from \mathbb{S} to \mathbb{R} as follows

$$Y'(z) = Y(z) \text{ if } z \notin \mathbf{B}$$

$Y'(z) = \mathbb{E}_{\mathbf{B}_t}(Y)$ if $z \in \mathbf{B}_t \subset \mathbf{B}$, where $\mathbb{E}_{\mathbf{B}_t}(Y)$ is the expectation of Y in the subhypercube \mathbf{B}_t .

By the definitions of Y' , the following properties hold

$$\begin{aligned} \mathbb{E}(Y') &= \mathbb{E}(Y) = 0 \\ C_{Y'} &\leq \mathbf{c} \\ V_{Y'} &\leq \mathbf{V} \\ Pr(Y \neq Y') &\leq Pr(\mathbf{B}). \end{aligned}$$

It suffices to prove that

Claim 3.1.

$$Pr(|Y'| \geq (\lambda \mathbf{V})^{1/2}) \leq 2e^{-\lambda/4}.$$

Lemma 3.2. *Let Z be a function from \mathbb{S} to \mathbb{R} with mean 0. If $x \leq 1/c_Z$, then*

$$\mathbb{E}(e^{xZ}) \leq e^{x^2 V_Z}.$$

To see that Lemma 3.2 implies Claim 3.1, set $x = (\frac{\lambda}{4\mathbf{V}})^{1/2}$. Since $\lambda \leq \frac{\mathbf{V}}{\mathbf{c}^2}$, $x \leq \frac{1}{\mathbf{c}} \leq \frac{1}{c_{Y'}}$, and the lemma yields

$$\mathbb{E}(e^{xY'}) \leq e^{x^2 V_{Y'}} \leq e^{x^2 \mathbf{V}}.$$

By Markov's inequality

$$\begin{aligned} \Pr(Y' \geq (\lambda \mathbf{V})^{1/2}) &= \Pr(e^{xY'} \geq e^{x(\lambda \mathbf{V})^{1/2}}) \\ &= \Pr(e^{xY'} \geq e^{\lambda/2}) \\ &\leq e^{x^2 \mathbf{V} - \lambda/2} = e^{-\lambda/4}, \end{aligned}$$

and the claim follows by symmetry.

Lemma 3.2 is a special case of a more general statement shown by Grabble in [Gra] and we repeat his proof here. The proof uses induction on n . The statement is trivial when $n = 1$. Consider a generic $n > 1$. Notice that, by definition, $C_1(t)$ and $V_1(t)$ do not depend on t . In the following we set $V_1(t) = V_1$.

Consider the function $Z - \mu_{1,0}$ assigned to the left subtree of depth $n - 1$ of the original tree. This function has expected value 0, so the induction hypothesis gives

$$\mathbb{E}(e^{x(Z - \mu_{1,0})}) < e^{x^2(V_Z - V_1)}.$$

A similar argument on the right subtree gives

$$\mathbb{E}(e^{x(Z - \mu_{1,1})}) < e^{x^2(V_Z - V_1)}.$$

On the other hand,

$$\mathbb{E}(e^{xZ}) = p_1 e^{x\mu_{1,1}} \mathbb{E}(e^{x(V_Z - \mu_{1,1})}) + q_1 e^{x\mu_{1,0}} \mathbb{E}(e^{x(V_Z - \mu_{1,0})}),$$

therefore

$$\mathbb{E}(e^{xZ}) < p_1 e^{x\mu_{1,1}} e^{x^2(V_Z - V_1)} + q_1 e^{x\mu_{1,0}} e^{x^2(V_Z - V_1)}.$$

It remains to show

$$p_1 e^{x\mu_{1,1}} + q_1 e^{x\mu_{1,0}} \leq e^{x^2 V_1}.$$

Consider the Taylor expansion of the left hand side of the above inequality

$$\begin{aligned}
 & p_1(1 + x\mu_{1,1} + x^2\mu_{1,1}^2/2 + \dots) + q_1(1 + x\mu_{1,0} + x^2\mu_{1,0}^2/2 + \dots) \\
 &= 1 + 0 + x^2(p_1\mu_{1,1}^2 + q_1\mu_{1,0}^2)/2 + \dots \\
 &= 1 + V_1x^2 \sum_{i=2}^{\infty} \frac{1}{i!} (q_1(x\mu_{1,1})^{i-2} + p_1(x\mu_{1,0})^{i-2}).
 \end{aligned}$$

Since $x < 1/c_Z$, both $x\mu_{0,1}$ and $x\mu_{1,1}$ have absolute values less than 1. Thus

$$\sum_{i=2}^{\infty} \frac{1}{i!} (q_1(x\mu_{1,1})^{i-2} + p_1(x\mu_{1,0})^{i-2}) < \sum_{i=2}^{\infty} \frac{1}{i!} (p_1 + q_1) = \sum_{i=2}^{\infty} \frac{1}{i!} < 1.$$

The last inequality implies that $p_1e^{x\mu_{1,1}} + q_1e^{x\mu_{1,0}}$ is at most $1 + V_1x^2$. Since $1 + V_1x^2 \leq e^{x^2V_1}$, the proof of Lemma 3.2 is complete. \square

§4 PROOF OF MAIN THEOREM

In this section, all polynomials are simplified. The proof follows the same framework as in [KV] and [Vu1], using induction on k . We start by Main Lemma II, with properly chosen \mathbf{c} and \mathbf{V} . In order to bound $Pr(\mathbf{B})$, first notice that

$$Pr(\mathbf{B}) \leq Pr(W(t) \geq \mathbf{V}/\mathbf{c}) + \sum_{i=1}^n Pr(C_i(t) \geq \mathbf{c}).$$

To bound the right hand side, we apply the induction hypothesis and Main Lemma I. The details now follow.

Let us take a closer look at the function $C_i(t)$ defined in the last section. For any $1 \leq i \leq n$, let $Y_i(t)$ be the sum of those monomials in Y that contain i : $Y_i(t) = \sum_{i \in I_j} c_j I_j$. $Y'_i(t)$ is obtained from $Y_i(t)$ by fixing $t_i = 1$. One can check that C_i is the conditional expectation of Y'_i with respect to t_1, \dots, t_{i-1} : $C_i(t) = \mathbb{E}(Y'_i | t_1, \dots, t_{i-1})$. It is clear that C_i is a positive polynomial with degree at most $k - 1$. Let γ_i denote the free coefficient of C_i and set $Z_i = (C_i - \gamma_i)/2$.

Lemma 4.1. *All coefficients of C_i are at most 2. $\mathbb{E}(Z_i) \leq \mathbb{E}_1(Y)$ and for any $1 \leq j \leq k - 1$ $\mathbb{E}_j(Z_i) = \mathbb{E}(C_i)/2 \leq \mathbb{E}_{j+1}(Y)$.*

Proof. Consider a monomial $I = \prod_{j \in A} t_j$. The coefficient of I in C_i is at most $\alpha + \mathbb{E}_{A \cup i}(Y)$, where α is the coefficient of It_i in Y . Since both α and $\mathbb{E}_{A \cup i}(Y)$ are at most 1, we are done with the first statement.

To show the second statement, notice that $\mathbb{E}(Z_i) \leq \mathbb{E}_{\{i\}}(Y) \leq \mathbb{E}_1(Y)$. The last statement can be proven similarly. \square

Let $W(t) = \sum_{i=1}^n p_i C_i(t)$. By the above, $W(t)$ is also a positive polynomial with degree at most $k - 1$.

Lemma 4.2. *The free coefficient of W is $\mathbb{E}(Y)$ and any other coefficient of W is at most $\mathbb{E}_1(Y)$. For any $1 \leq j \leq k-2$, $\mathbb{E}_j(W) \leq (k-1)\mathbb{E}_j(Y)$. Furthermore $k\mathbb{E}(Y) \geq \mathbb{E}(W)$.*

Proof. Consider a term $\alpha t_{i_1} \dots t_{i_l}$ in Y ($i_1 < i_2 < \dots < i_l$ and $l \leq k$). This term gives rise to the following l terms in W :

$$\alpha p_{i_1} p_{i_2} \dots p_{i_l}, \alpha t_{i_1} p_{i_2} \dots p_{i_l}, \dots, \alpha t_{i_1} \dots t_{i_{l-1}} p_{i_l}.$$

Since each of these $l \leq k$ terms has the same expectation as $t_{i_1} \dots t_{i_l}$, it follows that $\mathbb{E}(W) \leq k\mathbb{E}(Y)$. It is also clear that the free coefficient of W is $\mathbb{E}(Y)$. The coefficient of a monomial $I = \prod_{i \in A} t_i$ in W is at most $\mathbb{E}_A(Y) \leq \mathbb{E}_{|A|}(Y) \leq \mathbb{E}_1(Y)$. The last statement can be verified by a similar argument. \square

We show by induction on k that

$$Pr(|Y - \mathbb{E}(Y)| \geq (4kK\lambda\mathbb{E}(Y))^{1/2}) \leq 2ke^{-\lambda/4} + h(k, K, n, \delta),$$

for any $\lambda > 0, K \geq 2k$ satisfying $4kK\lambda \leq \mathbb{E}(Y)$.

For $k = 1$, set $\mathbf{c} = 1, \mathbf{V} = \mathbb{E}(Y)$. We have $\lambda \leq \mathbf{V}/\mathbf{c}^2$. The set \mathbf{B} as defined in Main Lemma II is empty; thus by Main Lemma II

$$Pr(|Y - \mathbb{E}(Y)| \geq (\lambda\mathbb{E}(Y))^{1/2}) \leq 2e^{-\lambda/4}.$$

Now consider a generic $k > 1$. Set $\mathbf{V} = 4kK\mathbb{E}(Y)$ and $\mathbf{c} = 2(K+1)$. By the assumption on λ and K , \mathbf{V} and \mathbf{c} satisfy $\mathbf{V}/\mathbf{c}^2 \geq \lambda$.

Set $X(t) = (W(t) - \mathbb{E}(Y))/(k-1)$; Lemma 4.2 yields that X is normal and $\mathbb{E}(X) \leq \mathbb{E}(Y)$. Set $q = \lceil \mathbb{E}(Y) \rceil$. Let $X' = X + x_1 + \dots + x_q$, where x_i are dummy i.i.d $\{0, 1\}$ random variables with expectation chosen properly so that $\mathbb{E}(X') = \mathbb{E}(Y)$ (this step is little bit artificial, however we need it to guarantee to condition $\mathbb{E}(X') \geq 4kK\lambda$ in the induction hypothesis). It is clear that X' is normal and $\mathbb{E}_1(X') = \mathbb{E}_1(X) \leq \delta$. Therefore, by the induction hypothesis (notice that X' contains $n+q$ random variables)

$$Pr(|X' - \mathbb{E}(X')| \geq (4(k-1)K\lambda\mathbb{E}(X'))^{1/2}) \leq 2(k-1)e^{-\lambda/4} + h(k-1, K, n+q, \delta).$$

Since $\mathbb{E}(X') = \mathbb{E}(Y) \geq 4kK\lambda$, it follows that

$$Pr(X \geq 2\mathbb{E}(Y)) \leq Pr(X' \geq 2\mathbb{E}(Y)) \leq 2(k-1)e^{-\lambda/4} + h(k-1, K, n+q, \delta).$$

By the definition of X , this yields

$$Pr(W \geq (2k-1)\mathbb{E}(Y)) \leq 2(k-1)e^{-\lambda/4} + h(k-1, K, n+q, \delta).$$

Since $K \geq 2k$, $\mathbf{V}/\mathbf{c} = 2kK\mathbb{E}(Y)/(K+1) \geq (2k-1)\mathbb{E}(Y)$, it follows

$$\Pr(W \geq \mathbf{V}/\mathbf{c}) \leq 2(k-1)e^{-\lambda/4} + h(k-1, K, n+q, \delta).$$

By Lemma 4.1, $Z_i \in \mathbf{Poly}_{k-1}(\delta)$. Therefore, by Main Lemma I,

$$\Pr(C_i \geq 2(K+1)) \leq \Pr(Z_i \geq K) \leq r(k-1, K, n, \delta).$$

Thus,

$$\Pr(\mathbf{B}) \leq 2(k-1)e^{-\lambda/4} + h(k-1, K, n+q, \delta) + nr(k-1, K, n, \delta).$$

Main Lemma II and the definition of $h(k, K, n, \delta)$ complete the proof. \square

Remark. In the proof of Theorem 1.1 and its variations ([KV, Vu1]), Main Lemma I is not needed, and one can prove the statements for general atom variables, without the restriction that they have only two values 0 or 1. On the other hand, the proof of Main Lemma I does rely on this restriction and it is not clear to us how to avoid it, although we do think that Theorem 1.3 still holds with variables with arbitrary distribution in the interval $[0, 1]$.

§5 APPLICATIONS

As mentioned in the beginning of the paper, in probabilistic combinatorics, we frequently have to prove a strong concentration result on multi-variate polynomial of type $Y = \sum_{j=1}^m I_j$, where each I_j is a products of few atom variables. In several cases, such function has very large Lipschitz coefficient which does not allow us apply classical tools such as Azuma or Talagrand's inequality.

In such cases, the proof usually breaks into two separate parts. To bound the probability that $Y \leq (1-\epsilon)\mathbb{E}(Y)$, one can routinely use Janson's inequality. The problem is with bounding the probability that $Y \geq (1+\epsilon)\mathbb{E}(Y)$. There was no general method for this case, and usually one has to work out an ad hoc argument. Finding such ad hoc arguments requires ingenuity and could occasionally be fairly involved.

Theorems 1.1-1.3 provide a universal and simple way to derive a concentration result in situations as mentioned above, provided the polynomials have fixed degrees. The crucial advantage we have here is that these theorems, at one strike, give a strong large deviation bound for both lower tail and upper tail. Beside, using these theorems requires a minimum amount of computation. In most cases, the calculation of the expectations of the partial derivatives is straightforward. In the rest of this section, we provide few examples to illustrate the idea. The problems considered in these examples are also discussed in [AS], chapter 8, and it would be very instructive for the reader to read this chapter and compare the methods.

Random graphs. The calculation in the example present in §1 can be repeated for an arbitrary strictly balanced graph H , instead of the triangle. It is easy to see

that if H is strictly balanced and the expectation of $Y(H)$ (the number of copies of H in $G(n, p)$) is $O(\log n)$, then $\mathbb{E}_1 = O(n^{-\alpha})$ for some positive constant $\alpha = \alpha(H)$. Thus, Corollary 1.4 applies and give the following.

Corollary 5.1. *Let H be a fixed strictly balanced graph on k edges and $Y = Y(H)$ be the number of copies of H in $G(n, p)$. Assume that $\mathbb{E}(Y) \leq \log n$. There are positive constants $c = c(k)$ and $d = d(k)$ such that for any $0 \leq \epsilon \leq 1$.*

$$Pr(|Y - \mathbb{E}(Y)| \geq \epsilon \mathbb{E}(H)) \leq d e^{-c\epsilon^2 \mathbb{E}(Y)}.$$

Theorem 1.1 can be used to study the same question when the expectation of Y is large. Since this is already done elsewhere [KV][Vu1], we omit the details.

By the generality of Theorems 1.2 and 1.3, we can generalize Corollary 5.1 in several directions, without any special effort.

- First, since our theorems do not require the atom variables be i.i.d., one can also consider a more general model of random graphs where the edge probabilities are different.

- Another direction of generalization is to consider a different type of substructures. For instance, instead of the number subgraphs, one can consider the number of rooted subgraphs. In [KV], this problem is actually worked out, giving a short proof of a theorem of Spencer [Spe2] on counting extensions (again we think that it is worth checking both proofs to compare the methods).

- Finally, one could see that random graphs play no particular role and one can easily formalized a similar statement for random hypergraphs or other random structures. Interested readers may try to work out the details as an exercise.

Random sequences. In this section, \mathbb{N} denotes the set of positive integers. For each $x \in \mathbb{N}$, chose x with probability p_x . Let a random variable t_x represent this choice: $t_x = 1$ if x is chosen and $t_x = 0$ other wise. The sequence X of chosen numbers is a random sequence and the probability space is the (infinite dimension) product space spanned by the t_x 's. A common task in the theory of random sequences is to show that with positive probability X satisfies a given property $\mathcal{P}(n)$ for all sufficiently large $n \in \mathbb{N}$.

The general strategy for such problem is the following. For each n , show that $\mathcal{P}(n)$ fails with small probability, say $s(n)$. If $s(n)$ is sufficiently small so that $\sum_{n=1}^{\infty} s(n)$ converges, then by Borel-Cantelli's lemma, $\mathcal{P}(n)$ holds for all sufficiently large every n with probability 1 (see, for instance, [HR, Chapter 3]).

The crucial point of the argument is to show that for each n , $\mathcal{P}(n)$ holds with high probability. In several cases, this is equivalent to showing that a properly defined polynomial Y_n (with variables t_x , $x \leq n$) is close to its expectation with high probability.

Theorem 1.2 supplies a convenient way to deal with the above task. Recall that we need to show that the failure probability $s(n)$ are small enough so that

$\sum_{n=1}^{\infty} s(n)$ converges. So it suffices to show that $s(n) \leq n^{-2}$, and this fits nicely into the range of Theorem 1.2.

To illustrate the idea, let us give a simple proof for the following theorem, proven by Erdős and Tetali [ET]. For a set $X \subset \mathbb{N}$, $R_X^k(n)$ denotes the number of ways to represent n as a sum of k elements in X .

Theorem 5.2. ([ET]) *There is a subset $X \subset \mathbb{N}$ such that $R_X^k(n) = \Theta(\log n)$, for all sufficiently large n .*

The set X is defined randomly. For each $x \in \mathbb{N}$, pick x with probability $p_x = \mathbf{c}x^{1/k-1} \log^{1/k} x$, where \mathbf{c} is a positive constant to be determined later. Let t_x be the characteristic random variable of this choice; thus, t_x is a $\{0, 1\}$ random variable with mean p_x .

The number of representations of n as a sum of k elements from X can be written as a polynomial in the following way

$$Y_n = \sum_{\substack{x_1 \leq \dots \leq x_k \\ x_1 + \dots + x_k = n}} t_{x_1} \dots t_{x_k}.$$

We now show that with probability 1, Y_n is of order $\Theta(\log n)$ for sufficiently large n . Set $a = 0.9$ (one can use any positive constant less than 1 instead of 0.9). We first break Y_n as follows

$$Y_n = Y'_n + Y''_n$$

where

$$Y'_n = \sum_{\substack{n^a \leq x_1 \leq \dots \leq x_k \\ x_1 + \dots + x_k = n}} t_{x_1} \dots t_{x_k}.$$

Y'_n is the main part of Y_n since there are very few solutions which have a small element (in a typical solution of $x_1 + \dots + x_k = n$, all x_i have order $\Theta(n)$). To finish the proof it suffices to show that

(1) There are positive constants $c_1 < c_2$ such that $Pr(c_1 \log n \geq Y'_n) + Pr(Y'_n \geq c_2 \log n) = O(n^{-2})$

(2) For almost every sequence X , there is a finite number $M(X)$ such that $Y''_n < M(X)$ for all sufficiently large n .

The main part of the proof is to show (1), and here we shall apply Theorem 1.2. First, we need the following lemma, which asserts that $\mathbb{E}(\partial_A Y'_n)$'s are small.

Lemma 5.3. *For all non-empty multi-sets A , $\mathbb{E}(\partial_A Y'_n) = O(n^{-a/2k})$*

Proof. Consider a (multi-) set A . Assume that $|A| = k - l$ and $\sum_{x \in A} x = n - m$, there is a constant $c = c(A)$ such that

$$\partial_A Y'_n \leq c \sum_{\substack{n^a \leq x_1 \leq \dots \leq x_l \\ x_1 + \dots + x_l = m}} t_{x_1} \dots t_{x_l}.$$

Notice that $x_l \geq m/l$. Using the fact that $\sum_{x=1}^m x^{1/k-1} \approx \int_1^m z^{1/k-1} \partial z \approx m^{1/k}$, we have

$$\begin{aligned} \mathbb{E}(\partial_A Y'_n) &= O\left(\sum_{\substack{n^a \leq x_1 \leq \dots \leq x_l \\ x_1 + \dots + x_l = m}} p_{x_1} \dots p_{x_l} \right) \\ &= O(\log n) \sum_{\substack{n^a \leq x_1 \leq \dots \leq x_l \\ x_1 + \dots + x_l = m}} x_1^{1/k-1} \dots x_l^{1/k-1} \\ &= O(\log n) O\left(\left(\sum_{x=1}^m x^{1/k-1} \right)^{l-1} (m/l)^{1/k-1} \right) \\ &= O(\log n) O(m^{(l-1)/k} (m/l)^{1/k-1}) \\ &= O(\log n) O(m^{(l-k)/k}) = O(n^{-a/2k}), \end{aligned}$$

since $k-l \geq 1$ and $m \geq x_1 \geq n^a$. This ends the proof of the lemma. \square

The last step in the above calculation explains why we did not apply Theorem 1.2 directly to Y'_n and need to make the restriction $x_1 > n^a$. Without this assumption, there would be some partial derivatives with large expectation.

From the above calculation, it follows immediately (by setting $l = k$ and $m = n$) that $\mathbb{E}(Y'_n) = O(\log n)$ (our calculation is simpler than the one used in [ET], which involves a multiple integral). Moreover, a straightforward argument shows that if $\mathbf{c} \rightarrow \infty$, then $\mathbb{E}(Y'_n)/\log n \rightarrow \infty$. Indeed, there are at least $\frac{n^{k-1}}{(4k)^{k-1}k!}$ tuples $x_1 \leq x_2 \leq \dots \leq x_k$ where $n/4k \leq x_i \leq n/2k$ for all $i < k$; each such tuple contributes at least $\mathbf{c}^k n^{1-k} \log n$ to $\mathbb{E}(Y'_n)$. Thus, by setting \mathbf{c} big, we can assume that $\mathbb{E}(Y'_n)/\log n$ is sufficiently large. Theorem 1.2 then applies and implies (1).

The proof of (2) is simple and relies mainly on Lemma 2.3 and the calculation in the proof of Lemma 5.3. First, for all $l < k$, let $R_l(n)$ be the number of representation of n as the sum of l elements from X . With essentially the same computation as in Lemma 5.3, one can show that $\mathbb{E}(R_l(n)) = O(n^{-1/k} \log n) = O(n^{-1/2k})$. Lemma 2.3 then implies that for a sufficiently large M_1 with probability $1 - O(n^{-2})$, the maximum number of disjoint representations of n in $R_l(n)$ is at most M_1 . By Borel-Cantelli's lemma, we can conclude that a.s., the maximum number of disjoint representations of n as a sum of l elements of X is at most M_1 , for all $l < k$ and n sufficiently large. It follows that almost surely, for each random sequence X there

is a **finite** number $M_1(X)$ such that for any $l < k$ and **all** n , the maximal number of disjoint representations of n as a sum of l elements of X is at most $M_1(X)$.

Using a computation similar to the one in the proof of Lemma 5.3, one can also deduce that $\mathbb{E}(Y'') = O(n^{(a-1)/k} \log n) = O(n^{-1/2k})$ (since $x_1 \leq n^a$, instead of $(\sum_{x=1}^n x^{1/k-1})^{k-1}$, one can write $\sum_{x=1}^{n^a} x^{1/k-1} (\sum_{x=1}^n x^{1/k-1})^{k-2}$ and the bound follows). So, again by Lemma 2.3 (or Corollary 2.6) and Borel-Cantelli's lemma, there is a constant M_2 such that a.s. the maximum number of disjoint representations of n in Y'' is at most M_2 for all large n . It would be useful to think of $Y''(n)$ as a family of sets of size k , each corresponds to a representation of n .

We say that a sequence X is *good* if it satisfies the properties described in the last two paragraphs.

To finish the proof, we need only show that if X is good, then $Y''(n)$ is bounded. Set $M(X) = (\max(M_1(X), M_2))^k k!$. Assume that n is sufficiently large. It is clear that $Y''(n) \geq M(X)$, then by Lemma 2.2, $Y''(n)$ contain a $M_3 = \max(M_1(X), M_2) + 1$ sunflower. If the intersection of this sunflower is empty, then the petals form a family of M_3 disjoint sets of size k , each of them is a representation of n . If the intersection has cardinality $g > 0$ and its elements sum up to f , then the petals minus the intersections form a family of M_3 disjoint sets of size $l = k - g$, each of them is a representation of $m = n - f$. Any of these two events contradicts the fact that X is good. \square

In this proof, the finite number $M(X)$ depends on X . One can avoid this by the following trick. Instead of (2), notice that it is also sufficient to prove

(2') There is a constant M such that with probability at least $1/2$, $Y''(n) \leq M$ for all sufficiently large n .

We know that there is a constant M_1 such that the following holds: The maximum number of disjoint representations of n as a sum of l elements of X is at most M_1 , for all $l < k$ and $n \geq N(X)$, for some finite $N(X)$. Let $p(s)$ be the probability that $N(X) \leq s$. Then $p(1) \leq p(2) \leq p(3) \leq \dots \rightarrow 1$; so there is a number L such that $p(L) \geq 1/2$. Let $M'_1 = \max(M_1, L)$, one can conclude that with probability at least $1/2$, a sequence X satisfies the following: The maximum number of disjoint representations of n as a sum of l elements of X is at most M'_1 , for all $l < k$ and all n . Now repeat the previous proof with $M_1(X)$ replaced by M'_1 .

Theorem 5.2 can be generalized without any difficulty to the following more general theorem. Fix k positive integers a_1, \dots, a_k , where $\gcd(a_1, \dots, a_k) = 1$. Let $Q_X^k(n)$ be the number of representations of the form $n = a_1 x_1 + \dots + a_k x_k$, where $x_i \in X$.

Theorem 5.4. *There is a subset $X \subset \mathbb{N}$ such that $Q_X^k(n) = \Theta(\log n)$, for all sufficiently large n .*

Another type of generalization is to require X consist of special integers. The classical Waring problem (proved first by Hilbert in 1909[Vau]) asserts that for any

fixed r , every positive integer n can be represented as sum of k r^{th} powers (if k is sufficiently large compared to r ; by r^{th} power we mean the r^{th} power of a non-negative integer). For instance, every positive integer is a sum of 4 squares, 9 cubes and so on. Let X be a subset of the set \mathbb{N}^r of all r^{th} powers and define $R_X^k(n)$ as in Theorem 5.3. In [Vu2], the present author prove the following

Theorem 5.5. *Given a fixed positive constant r , there is $k_0 = k_0(r)$ such that the following holds. For all $k \geq k_0$, there is a set X consisting of r^{th} powers such that $R_X^k(n) = \Theta(\log n)$, for all sufficiently large n .*

The proof of Theorem 5.5 uses the above frame work and makes a crucial use of Theorem 1.2. In addition, it requires several sophisticated number theoretic arguments (see [Vu4] for details). Theorem 5.5 improves and generalizes results of many researchers, including Choi, Erdős, Nathanson, Spencer, Zöllner and Wirsing [CEN, EN, Nat, Spe, Zöl1, Zöl2, Wir]. In particular, this theorem (see [Vu4], §1) implies (via the pigeon hole principle) that there is a subset $X \subset \mathbb{N}^r$ with density $n^{1/k+o(1)}$ so that every positive number can be represented as a sum of k elements of X . In other words, one needs only an as small as possible part of \mathbb{N}^r to represent all positive numbers (the density $n^{1/k+o(1)}$ is optimal up to the term $o(1)$, by the pigeon hole principle). This sharpens Waring's assertion and gives a complete answer for an open question of Nathanson posed twenty years ago [Nat].

Several other applications in number theory which use the same frame work will appear in a future paper.

Acknowledgement. We would like to thank the referees for pointing out several errors.

REFERENCES

- [AS] N. Alon and J. Spencer, *The probabilistic method*, Wiley 1992.
- [CEN] S.L.G Choi, P. Erdős and M. Nathanson, *Lagrange's theorem with $N^{1/3}$ squares*, Proc. Am. Math. Soc., 79: 203-2-5, 1980.
- [EN] P. Erdős and M. Nathanson, *Largange's theorem and thin subsequences of squares*. In J.Gani and V.K. Rohatgi, editors, *Contribution to Probability*, p.3-9, Academic Press, New York, 1981.
- [ERa] P. Erdős and R. Rado, *Intersection theorems for systems of sets*, J. London Math. Soc., 35, 85-90 (1960).
- [ETe] P. Erdős and P. Tetali, *Representations of integers as sum of k terms*, Random Structures and Algorithms 1 (1990), 245-261.
- [Gra] D. Grable, *A large deviation inequality for functions of independent, multi-way choices*, Combinatorics, probability and Computing (1998) 7, 57-63.
- [HR] H. Halberstam and K. F. Roth, *Sequences*, Springer-Verlag, New York, 1983.
- [Jan] Janson, S. *Poisson approximation for large deviations*, Random Structures and Algorithms 1, 221-230 (1990).

- [JLR] S. Janson, T. Łuczak and A. Ruciński, *An exponential bound for the probability of nonexistence of a specified subgraph in a random graph*, in: M. Karoński et al. eds., *Random graphs 87* (Wiley, New York, 1990) 73-87.
- [KV1] J.H. Kim and V. H. Vu, *Small complete arcs on projective planes*, submitted.
- [KV2] J. H. Kim and V.H. Vu, *Concentration of polynomials and its applications*, to appear in *Combinatorica*.
- [Nat2] M. Nathanson, *Waring's problem for sets of density zero*, *Analytic Number Theory*, edited by M. Knopp, *Lecture Notes in Mathematics 899*, Springer-Verlag, 1980.
- [Spe] J. Spencer, *Four squares with few squares*, p 295-297, D.V. Chudnovsky et al. editors, *Number Theory, New York Seminar 1991-1995*, Springer.
- [Spe2] J. Spencer, *Counting extensions*, *Journal of Combinatorial Theory, Series A* (1990) 55, 247-255.
- [Tal] M. Talagrand, *A new look at independence*, *The Annals of probability* (1996) Vol 24, No1, 1-34.
- [Vu1] V. H. Vu, *Average smoothness: concentration of multi-variate polynomials and its applications*, manuscript.
- [Vu2] V. H. Vu, *On the list chromatic number of locally sparse graphs*, manuscript.
- [Vu3] V. H. Vu, *On some degree conditions which guarantee the upper bound of chromatic (choice) number of random graphs*, *Journal of Graph Theory*, 31, 1999, 201-226.
- [Vu4] V. H. Vu, *On a refinement of Waring's problem*, to appear in *Duke M. Journal*.
- [Vu5] V. H. Vu, *New bounds on nearly perfect matching in hypergraphs: higher codegrees do help*, to appear in *Random Structures and Algorithms*.
- [Wir] E. Wirsing, *Thin subbases*, *Analysis* 6 (1986), 285-308.
- [Zöl1] J. Zöllner, *Der Vier-Quadrate-Satz und ein Problem von Erdős and Nathanson*, Ph.D thesis, Johannes Gutenberg-Universität, Mainz, 1984.
- [Zöl2] J. Zöllner, *Über eine Vermutung von Choi, Erdős and Nathanson*, *Acta Arith.*, 45: 211-213, 1985.