

# Bernstein-like Concentration and Moment Inequalities for Polynomials of Independent Random Variables: Multilinear Case

Warren Schudy\*

Maxim Sviridenko<sup>†</sup>

## Abstract

Polynomials of independent random variables arise in a variety of fields such as Machine Learning, Analysis of Boolean Functions, Additive Combinatorics, Random Graphs Theory, Stochastic Partial Differential Equations etc. They naturally model the expected value of objective function (or lefthand side of constraints) for randomized rounding algorithms for non-linear optimization problems where one finds a solution of an "easy" continuous problem and rounds it to a solution of a "hard" integral problem (one such example is Convex Integer Programming [6]). To measure the performance guarantee of such algorithms one needs analogously to the analysis employed by Raghavan and Thompson [17] for boolean integer programming problems an analog of Chernoff Bounds for polynomials of independent random variables. There are many known forms and variations of Chernoff Bounds. One of the tightest ones is based on a variance of a sum of random variables known as Bernstein inequality. Another popular albeit a weaker version is using an estimate of a variance through the expectation. The later versions of concentration inequalities for polynomials of independent random variables are known [12, 18]. In this paper we derive an analog of Bernstein Inequality for multilinear polynomials of independent random variables.

We show that the probability that a multilinear polynomial  $f$  of independent random variables exceeds its mean by  $\lambda$  is at most  $e^{-\lambda^2/(R^a \text{Var}(f))}$  for sufficiently small  $\lambda$ , where  $R$  is an absolute constant. This matches (up to constants in the exponent) what one would expect from the central limit theorem. Our methods handle a variety of types of random variables including Gaussian, Boolean, exponential, and Poisson. Previous work by Kim-Vu and Schudy-Sviridenko gave bounds of the same form that involved less natural parameters in place of the variance.

## 1 Introduction

Polynomials of independent random variables arise in a variety of fields such as Machine Learning, Analysis of Boolean Functions, Additive Combinatorics, Random Graphs Theory, Stochastic Partial Differential Equations etc. They naturally model the expected value of objective function (or lefthand side of constraints) for randomized rounding algorithms for non-linear optimization problems where one finds a solution of an "easy" continuous problem and rounds it to a solution of a "hard" integral problem (one such example is Convex Integer Programming [6]). To measure the performance guarantee of such algorithms one needs analogously to the analysis employed by Raghavan and Thompson [17] for boolean integer programming problems an analog of Chernoff Bounds for polynomials of independent random variables. There are many known forms and variations of Chernoff Bounds. One of the tightest ones is based on a variance of a sum of random variables known as Bernstein Inequality [4, 3, 19]. Another popular albeit a weaker version is using

---

\*IBM T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598. [wjschudy@us.ibm.com](mailto:wjschudy@us.ibm.com)

<sup>†</sup>University of Warwick, [sviri@dcs.warwick.ac.uk](mailto:sviri@dcs.warwick.ac.uk), Work supported by EPSRC grant EP/J021814/1, FP7 Marie Curie Career Integration Grant and Royal Society Wolfson Merit Award.

an estimate of a variance through the expectation. The later versions of concentration inequalities for polynomials of independent random variables are known [12, 18]. In this paper we derive an analog of Bernstein Inequality for multilinear polynomials of independent random variables.

Perhaps the most celebrated theorem in statistics is the central limit theorem. This theorem (actually family of theorems) states conditions under which a sum of  $n$  independent random variables converges to being normally (i.e. Gaussian) distributed as  $n \rightarrow \infty$ . Let  $Y_1, Y_2, \dots$  be a sequence of independent random variables. Let  $\text{Var}[Z] = \mathbb{E}[(Z - \mathbb{E}[Z])^2] = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2$  be the variance of the random variable  $Z$ . Various central limit theorems state various conditions on the  $Y_i$  under which

$$\lim_{n \rightarrow \infty} \Pr \left[ \sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]) \geq \lambda \sqrt{\text{Var} \left[ \sum_{i=1}^n Y_i \right]} \right] = \int_{-\infty}^{\lambda} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \quad (1.1)$$

for any  $\lambda \in \mathbb{R}$ . One sufficient condition is that the  $Y_i$  are identically distributed with finite variance [9]. Another set of sufficient conditions is that there exists  $M, \epsilon > 0$  such that all  $\mathbb{E}[|Y_i|^{2+\epsilon}] \leq M$  and  $\lim_{n \rightarrow \infty} \text{Var}[\sum_{i=1}^n Y_i] = \infty$  [9].

The rate of convergence of the limit is often of interest. The Berry-Esseen theorem [9] states that when the  $Y_i$  are identically distributed with finite  $\mathbb{E}[|Y_1|^3]$  and  $\mathbb{E}[Y_1^2] = \sigma^2$  then

$$\left| \Pr \left[ \sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]) \geq \lambda \sqrt{\text{Var} \left[ \sum_{i=1}^n Y_i \right]} \right] - \int_{-\infty}^{\lambda} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right| \leq \frac{0.77 \mathbb{E}[|Y_1|^3]}{\sigma^3 \sqrt{n}}. \quad (1.2)$$

Many applications require upper bounds on the probability of large deviations for finite  $n$ . The Berry-Esseen bound (1.2) is exponentially far from tight for many such applications, for example the probability that at least three-quarters of a sequence of coin flips are heads is  $2^{-\Theta(n)}$  but the Berry-Esseen bound is  $O(1/\sqrt{n})$ . Fortunately it is possible to do much better in many cases. For example if the  $Y_i$  are independent random variables with  $0 \leq Y_i \leq 1$  then a standard Bernstein inequality (e.g. Theorem 2.3 (b) in [15]) states that

$$\Pr \left[ \sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]) \geq \lambda \right] \leq \exp \left( -\frac{\lambda^2}{2\mu + 2\lambda/3} \right) \quad (1.3)$$

for any  $\lambda > 0$  where  $\mu = \mathbb{E}[\sum_{i=1}^n Y_i]$ . Note that the small- $\lambda$  probability bound is roughly  $\exp \left( -\frac{\lambda^2}{2\mu} \right)$ , which matches the Gaussian behavior suggested by the central limit theorem except for the use of the upper-bound for the variance  $\mu$  in place of the variance. This discrepancy can be remedied, yielding another variant of the Bernstein inequality (see Theorem 2.7 in [15])

$$\Pr \left[ \sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]) \geq \lambda \right] \leq \exp \left( -\frac{\lambda^2}{2V + 2\lambda/3} \right) \quad (1.4)$$

where  $V = \text{Var}(\sum_{i=1}^n Y_i)$ . For  $\lambda \leq V$  this matches (up to constants in the exponent) what the central limit theorem suggests.

Kim and Vu introduced variants of Chernoff bound (1.3) for polynomials of independent Boolean random variables [12]. Vu [20] tightened and generalized the bounds to handle independent random variables with arbitrary distributions in the interval  $[0, 1]$ . Schudy and Sviridenko [18] proved a

stronger concentration inequality for polynomials of independent random variables satisfying a general condition (see Definition 1.1). Note that [20] contains one extension not handled in [18] and this paper, namely using less than  $q$  (the degree of the polynomial) smoothness parameters. These bounds share the Gaussian-like behavior for small  $\lambda$  with (1.3), but they use an upper bound on the variance that is more complicated than the  $\mu$  used in (1.3). The behavior for large  $\lambda$  is also different. Our main contribution is an analog of (1.4) for polynomial  $f(Y)$  of power  $q$ :

$$\Pr[|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] \leq e^2 \exp\left(-\frac{\lambda^2}{\text{Var}[f(Y)]R^q}\right) \quad (1.5)$$

for all sufficiently small  $\lambda$  (see Theorem 1.3 for the precise statement), where  $R$  is an absolute constant. What values of  $\lambda$  are “sufficiently small” depends on parameters  $\mu_1, \mu_2, \dots, \mu_q$  defined in the next section. For example in the setting of (1.4) we reproduce that bound up to constants in the exponent: Gaussian-like tails for  $\lambda \leq V$  and exponential tails for larger  $\lambda$ . Some polynomials require  $\lambda$  to be so small that  $e^2 e^{-\lambda^2/R^q}$  always exceeds 1 and hence (1.5) is vacuous. We expect that most applications will involve  $\lambda$  sufficiently small for (1.5) to apply.

The improvement of (1.5) compared to the concentration inequalities in Schudy-Sviridenko [18] is analogous to the improvement of (1.4) compared to (1.3). There are countless applications of Bernstein Inequality (or its variants known as Chernoff or Hoeffding Bounds) and its martingale versions [3, 8]. Recent algorithmic applications of the martingale version of Bernstein Inequality that require dependence on variance instead of expectation are [7] and [14]. Analogously, we expect that in the future there will be many applications (e.g. counting in random graphs) where one would necessarily need a stronger inequality of Theorem 1.3 (analog of (1.4) for polynomials) instead of Theorem 1.4 (analog of (1.3) for polynomials). Note that before our work such statements were not even known for boolean random variables.

## 1.1 Our Results

We are given a hypergraph  $H = (\mathcal{V}(H), \mathcal{H}(H))$  consisting of a set  $\mathcal{V}(H) = \{1, 2, \dots, n\} = [n]$  of vertices and a set  $\mathcal{H}(H)$  of hyperedges. A hyperedge  $h$  consists of a set  $\mathcal{V}(h) \subseteq \mathcal{V}(H)$  of  $|\mathcal{V}(h)| \leq q$  vertices. We are also given a weight  $w_h$  for each  $h \in \mathcal{H}(H)$ . For each such weighted hypergraph and real-valued weight  $w_h$  for its hyperedges, we define a multilinear polynomial

$$f(x) = \sum_{h \in \mathcal{H}(H)} w_h \prod_{v \in \mathcal{V}(h)} x_v. \quad (1.6)$$

We call the maximum hyperedge cardinality  $q$  the *power* of the polynomial  $f$ .

We use essentially the same smoothness parameters as Kim and Vu [12, 20] in our previous work [18]. For a given collection of independent random variables  $Y = (Y_1, \dots, Y_n)$ , hypergraph  $H$ , weights  $w$  and integer parameter  $r \geq 0$ , we define

$$\mu_r = \mu_r(Y, H, w) = \max_{S \subseteq [n]: |S|=r} \left( \sum_{h \in \mathcal{H}[\mathcal{V}(h) \supseteq S]} |w_h| \prod_{v \in \mathcal{V}(h) \setminus S} \mathbb{E}[|Y_v|] \right).$$

Note that  $S$  need not be a vertex set of some hyperedge of  $H$  and may even be the empty set. Sometimes we will also use the notation  $\mu_r(f) = \mu_r(Y, H, w)$  to emphasize the dependence on polynomial  $f$ .

In the previous work [18], we proved moment and concentration inequalities that could be viewed as an extension of (1.3) to polynomials of random variables satisfying the following condition.

**Definition 1.1** [18] A random variable  $Z$  is called **moment bounded** with real parameter  $L > 0$ , if for any integer  $i \geq 1$  we have

$$\mathbb{E} \left[ |Z|^i \right] \leq i \cdot L \cdot \mathbb{E} \left[ |Z|^{i-1} \right].$$

That work [18] showed that three large classes of random variables are moment bounded: bounded, continuous log-concave [2, 1] and discrete log-concave [1]. The results of the current paper apply to a related type of random variable.

**Definition 1.2** A random variable  $Z$  is called **central moment bounded** with real parameter  $L > 0$ , if for any integer  $i \geq 1$  we have

$$\mathbb{E} \left[ |Z - \mathbb{E}[Z]|^i \right] \leq i \cdot L \cdot \mathbb{E} \left[ |Z - \mathbb{E}[Z]|^{i-1} \right].$$

In Section 7 we show that the three classes of random variables that are known to be moment bounded (i.e. bounded, continuous log-concave and discrete log-concave) are also central moment bounded. For example Poisson, geometric, normal (i.e. Gaussian), and exponential distributions are all central moment bounded.

We prove the following:

**Theorem 1.3** We are given  $n$  independent central moment bounded random variables  $Y = (Y_1, \dots, Y_n)$  with the same parameter  $L$ . We are given a multilinear polynomial  $f(y)$  of power  $q$ . Let  $f(Y) = f(Y_1, \dots, Y_n)$  then

$$\Pr [|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] \leq e^2 \cdot \max \left\{ e^{-\frac{\lambda^2}{\text{Var}[f(Y)] \cdot R^q}}, \max_{r \in [q]} e^{-\left(\frac{\lambda}{\mu_r L^r R^q}\right)^{1/r}} \right\}, \quad (1.7)$$

where  $R$  is some absolute constant.

Quite often in the applications  $\mu_r$  for  $r = 1, \dots, q-1$  are negligibly small and  $\mu_q = 1$  (e.g. [21]). In this case, the right hand side of (1.7) becomes

$$e^2 \cdot \max \left\{ e^{-\frac{\lambda^2}{\text{Var}[f(Y)] \cdot R^q}}, e^{-\frac{\lambda^{1/q}}{L \cdot R}} \right\}$$

and our Theorem 1.3 implies that tails of multilinear polynomials of central moment bounded random variables have Gaussian-like distribution for  $\lambda < \text{Var}[f(Y)]^{q/(2q-1)}$  and constants  $L, R$ . Previous work [18] proved a similar theorem:

**Theorem 1.4** [18] We are given  $n$  independent moment bounded random variables  $Y = (Y_1, \dots, Y_n)$  with the same parameter  $L$ . We are given a multilinear polynomial  $f(x)$  with nonnegative coefficients of total power  $q$ . Let  $f(Y) = f(Y_1, \dots, Y_n)$  then

$$\Pr [|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] \leq e^2 \cdot \max \left\{ e^{-\frac{\lambda^2}{\max_{r \in [q]} (\mu_0 \mu_r L^r \cdot R^q)}}, \max_{r \in [q]} e^{-\left(\frac{\lambda}{\mu_r L^r R^q}\right)^{1/r}} \right\},$$

where  $R$  is some absolute constant.

We show that the parameter  $\text{Var}[f(Y)]$  in Theorem 1.3 is always at least as good as the  $\max_{r \in [q]} (\mu_0 \mu_r L^r)$  in Theorem 1.4:

**Lemma 1.5** *For a multilinear polynomial  $f$  as in Theorem 1.3 we have*

$$\text{Var}[f(Y)] \leq 2q4^q \max_{r \in [q]} (\mu_0(f, Y) \mu_r(f, Y) 4^r L^r).$$

Lemma 1.5 implies that our Theorem 1.3 dominates Theorem 1.4 from [18] in the common case when the central moment boundedness parameter  $L$  is of the same order as the moment boundedness parameter.

Previous work [18] showed that Theorem 1.4 has a tight dependence on the parameters  $\mu_1, \dots, \mu_q$  up to factors of logarithms and  $q^{O(q)}$  in the exponent. That lower bound only applies to bounds that depend only on those parameters and hence Theorem 1.3, which additionally depends on  $\text{Var}[f(Y)]$ , does not contradict it.

## 1.2 Comparing with Hypercontractivity concentration inequality and other results

It is well known that considering sums of centered (i.e.  $\mathbb{E}[Y] = 0$ ) and subgaussian (i.e.  $\mathbb{E}[|Y|^k] \leq L^k k^{k/2} \mathbb{E}[|X|]$ ) random variables improves the concentration bounds. Namely the concentration around its mean stays gaussian even for large values of  $\lambda$  unlike the case of the sum of non-centered (even boolean) random variables where the concentration bounds start to behave like the ones of exponential random variable. Therefore, we can expect a similar phenomenon for the polynomials of independent centered subgaussian random variables, i.e. the concentration bounds for polynomials of independent centered subgaussians should have tighter concentration around the mean for larger values of  $\lambda$ .

Two specific examples of such variables are centered Gaussian and Rademacher (+1 or -1 with probability 1/2) random variables. There are two concentration inequalities known in the literature specific for that setting

**Theorem 1.6 (Hypercontractivity Concentration Inequality)** *Consider a multilinear degree  $q$  polynomial  $f(Y) = f(Y_1, \dots, Y_n)$  of independent Normal or Rademacher random variables  $Y_1, \dots, Y_n$ . Then*

$$\Pr[|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] \leq e^2 \cdot e^{-\left(\frac{\lambda^2}{R \cdot \text{Var}[f(Y)]}\right)^{1/q}},$$

where  $\text{Var}[f(Y)]$  is the variance of the random variable  $f(Y)$  and  $R > 0$  is an absolute constant.

The history of these concentration and corresponding moment inequalities is quite rich see S. Janson [10] (Sections V and VI). Latala [13] tightened these inequalities for Normal random variables using smoothness parameters similar but incomparable to ours (see the next Section).

Unfortunately, the Hypercontractivity and even Latala Concentration Inequalities do not strictly dominate our concentration inequality (Theorem 1.3). Our concentration behaves better for small values of  $\lambda$  with respect to Hypercontractivity Concentration Inequality and for some polynomials we beat the Latala bounds for large values of  $\lambda$  since our smoothness parameters are incomparable.

The conclusion is that it is likely that there exists a yet to be discovered concentration inequality for polynomials of independent centered subgaussian random variables that dominates ours (Theorem 1.3), Hypercontractivity (Theorem 1.6) and Latala's [13] concentration inequalities in this setting. Deriving such an inequality is a challenging open problem.

In our previous work [18], we provide an extensive comparison of Theorem 1.4 and its analog for general polynomials with various known concentration inequalities for polynomials. Mossel,

O'Donnell and Oleszkiewicz [16] showed that the distribution of a multilinear polynomial of independent random variables is approximately invariant with respect to the distribution of the random variables as long as the random variables have mean 0 and variance 1. In particular they bound

$$|\Pr[f(X_1, \dots, X_n) \geq \lambda] - \Pr[f(G_1, \dots, G_n) \geq \lambda]|$$

where  $f$  is a multilinear polynomial,  $X_1, \dots, X_n, G_1, \dots, G_n$  are independent random variables with mean 0 and variance 1, and  $G_1, \dots, G_n$  have a Gaussian distribution. Such bounds can be considered to be a generalization of the Berry-Esseen type bounds because in the linear case the sum of Gaussians  $f(G_1, \dots, G_n)$  has a Gaussian distribution. Note that as usual central limit theorem or invariance principle type of results have very wide range of applicable random variables but weaker concentration bounds (polynomial instead of exponential).

### 1.3 Our Techniques

Our work follows the same general scheme of the moment computation method developed in the proof of Theorem 1.4 in [18] but there are many subtle differences in the proofs since we basically want to replace each term  $\mu_0 \mu_q$  in the proof of the **Initial Moment Lemma** from [18] with variance. For example, our Section 2.2 and the analogous Section 2.2 in [18] are devoted to bounding a certain sum (the sum over  $\pi$  in (2.12)). Previous work [18] gets a factor  $\mu_t$  for some  $0 \leq t \leq q$  for each hyperedge in a certain hypergraph  $G'$ . Each connected component of  $G'$  includes two hyperedge weights, call them  $w_1$  and  $w_2$ , which contribute to a  $\mu_0$  and a  $\mu_q$  factor respectively in [18]. Instead of having  $w_1 w_2$  contribute to a  $\mu_0 \mu_q$  we do the following. We bound  $w_1 w_2 \leq (w_1^2 + w_2^2)/2$  and then  $w_1^2$  and  $w_2^2$  each contribute to a factor of variance. Our Ordering Lemma in Section 4 is different from the analogous statement in [18]. To transition from Initial Moment Lemma to the General Even Moment Lemma we use certain orthogonality properties of multi-linear polynomials which do not seem to hold for general polynomials. Our key property of random variables (central moment boundness) is different from the moment boundness in [18] which forced us to re-prove that the classical classes of discrete and continuous random variables satisfy that property.

While we were able to extend the Theorem 1.4 in [18] to the case of general polynomials we were not able to prove a similar extension of Theorem 1.3. While we believe that such a statement is true, it seems it would require another property of random variables different from moment boundness or central moment boundness.

### 1.4 Outline

The high-level organization of our analysis follows [18]. Sections 2 and 3 state and prove key lemmas on the moments of polynomials of variables with zero expectation. Section 4 proves various technical lemmas that are omitted from the main flow. Section 5 states and proves bounds on the moments of polynomials with arbitrary expectation. Section 6 uses those bounds and Markov's inequality to prove Theorem 1.3. Section 7 shows that a wide variety of classical random variables are central moment bounded.

## 2 Moment Lemma for Centered Multilinear Polynomials

The proof of the Theorem 1.3 will follow from the application of the Markov's inequality to the upper bound on the  $k$ -th moment of the polynomial in question. The first step is to look at moments of "centered" multilinear polynomials that replace  $Y_v$  with  $Y_v - \mathbb{E}[Y_v]$ .



**Lemma 2.1 (Initial Moment Lemma)** *We are given a hypergraph  $H = ([n], \mathcal{H})$ ,  $n$  independent central moment bounded random variables  $Y = (Y_1, \dots, Y_n)$  with the same parameter  $L > 0$  and a polynomial  $g(x)$  with nonnegative coefficients  $w_h \geq 0$  such that every monomial (or hyperedge)  $h \in \mathcal{H}$  has power (or cardinality) exactly  $q$ . We define random variables  $X_v = Y_v - \mathbb{E}[Y_v]$  for  $v \in [n]$ . Then for any integer  $k \geq 1$  we have*

$$\left| \mathbb{E} \left[ g(X)^k \right] \right| \leq \max_{\bar{\sigma}} \left\{ R_3^{qk} L^{qk - q\sigma_0 - \ell} \cdot k^{qk - (q-1)\sigma_0 - \ell} \cdot \text{Var}[g(X)]^{\sigma_0} \cdot \left( \prod_{t=1}^q \mu_t(g, Y)^{\sigma_t} \right) \right\} \quad (2.8)$$

where  $R_3 \geq 1$  is some absolute constant,  $\ell = \sum_{t=0}^q (q-t)\sigma_t$ , and  $\bar{\sigma} = (\sigma_0, \dots, \sigma_q)$ . The maximum is over all non-negative integers  $\sigma_t$ ,  $0 \leq t \leq q$  satisfying  $2\sigma_0 + \sum_{t=1}^q \sigma_t = k$  and  $\ell \leq qk/2$ .

Note that the constraint  $2\sigma_0 + \sum_{t=1}^q \sigma_t = k$  in Lemma 2.1 implies  $\sigma_0 \leq k/2$  hence the powers of  $L$  and  $k$  in (2.8) are non-negative.

**Proof.** Fix hypergraph  $H = ([n], \mathcal{H})$ , random variables  $Y = (Y_1, \dots, Y_n)$ , non-negative weights  $\{w_h\}_{h \in \mathcal{H}}$ , an integer  $k$  and total power  $q$ . Without loss of generality we assume that  $\mathcal{H}$  is the complete uniform hypergraph (setting additional edge weights to 0 as needed), i.e.  $\mathcal{H}$  includes every possible hyperedge over vertex set  $[n]$  with  $q$  vertices. Note that the cardinality of the hyperedge is equal to the total power of the corresponding monomial in the polynomial  $g(x)$ .

A *labeled* hypergraph  $G = (\mathcal{V}(G), \mathcal{H}(G))$  consists of a set of vertices  $\mathcal{V}(G)$  and a *sequence* of  $k$  (not necessarily distinct) hyperedges  $\mathcal{H}(G) = \langle h_1, \dots, h_k \rangle$ . In other words a labeled hypergraph is a hypergraph whose  $k$  hyperedges are given unique labels from  $[k]$ . We write e.g.  $\prod_{h \in \mathcal{H}(G)} w_h$  as a shorthand for  $\prod_{i=1}^k w_{h_i}$  where  $\mathcal{H}(G) = \langle h_1, \dots, h_k \rangle$ ; in particular duplicate hyperedges count multiple times in such a product.

Consider the sequence of hyperedges  $h_1, \dots, h_k \in \mathcal{H}$  from our original hypergraph  $H$ . These hyperedges define a labeled hypergraph  $H(h_1, \dots, h_k)$  with vertex set  $\cup_{i=1}^k \mathcal{V}(h_i)$  and hyperedge sequence  $h_1, \dots, h_k$ . Note that the vertices of  $H(h_1, \dots, h_k)$  are labeled by the indices from  $[n]$  and the edges are labeled by the indices from  $[k]$ . Note also that some hyperedges in  $H(h_1, \dots, h_k)$  could span the same set of vertices, i.e. they are multiple copies of the same hyperedge in the original hypergraph  $H$ . Let  $\mathcal{P}(H, k)$  be the set of all such edge and vertex labeled hypergraphs that can be generated by any  $k$  hyperedges from  $H$ . We say that the *degree* of a vertex (in a hypergraph) is the number of hyperedges it appears in. Let  $\mathcal{P}_2(H, k) \subseteq \mathcal{P}(H, k)$  be the set of such labeled hypergraphs where each vertex has degree at least two. We split the whole proof into more digestible pieces by subsections.

## 2.1 Changing the vertex labeling

In this section we will show how to transform the formula for the  $k$ -th moment to have the summation over the hypergraphs that have its own set of labels instead of being labeled by the set  $[n]$ . Let  $X_v = Y_v - \mathbb{E}[Y_v]$  for  $v \in h$ . By linearity of expectation, independence of random variables  $X_v$

for different vertices  $v \in \mathcal{V}$  and definition of  $\mathcal{P}(H, k)$  we obtain

$$\begin{aligned}
\left| \mathbb{E} \left[ g(X)^k \right] \right| &= \left| \sum_{h_1, \dots, h_k \in \mathcal{H}} \mathbb{E} \left[ \prod_{i=1}^k \left( w_{h_i} \prod_{v \in \mathcal{V}(h_i)} X_v \right) \right] \right| \\
&= \left| \sum_{G \in \mathcal{P}(H, k)} \mathbb{E} \left[ \prod_{h \in \mathcal{H}(G)} \left( w_h \prod_{v \in \mathcal{V}(h)} X_v \right) \right] \right| \\
&= \left| \sum_{G \in \mathcal{P}(H, k)} \left( \prod_{h \in \mathcal{H}(G)} w_h \right) \left( \prod_{v \in \mathcal{V}(G)} \mathbb{E} \left[ \prod_{h \in \mathcal{H}(G) | v \in \mathcal{V}(h)} X_v \right] \right) \right| \\
&= \left| \sum_{G \in \mathcal{P}_2(H, k)} \left( \prod_{h \in \mathcal{H}(G)} w_h \right) \left( \prod_{v \in \mathcal{V}(G)} \mathbb{E} \left[ \prod_{h \in \mathcal{H}(G) | v \in \mathcal{V}(h)} X_v \right] \right) \right| \quad (2.9) \\
&\leq \sum_{G \in \mathcal{P}_2(H, k)} \left( \prod_{h \in \mathcal{H}(G)} w_h \right) \left( \prod_{v \in \mathcal{V}(G)} \left| \mathbb{E} \left[ \prod_{h \in \mathcal{H}(G) | v \in \mathcal{V}(h)} X_v \right] \right| \right) \\
&\leq \sum_{G \in \mathcal{P}_2(H, k)} \left( \prod_{h \in \mathcal{H}(G)} w_h \right) \left( \prod_{v \in \mathcal{V}(G)} \mathbb{E} \left[ |X_v|^{d_v(G)} \right] \right), \quad (2.10)
\end{aligned}$$

where the equality (2.9) follows from the fact that  $\mathbb{E}[X_v] = 0$  for all  $v \in \mathcal{V}(h)$  and  $d_v(G)$  is the degree of vertex  $v$  in the hypergraph  $G$ .

Note that a labeled hypergraph  $G \in \mathcal{P}_2(H, k)$  could have the number of vertices ranging from  $q$  up to  $kq/2$  since every vertex has degree at least two. For  $k$  and  $q$  clear from context, let  $\mathcal{S}_2(\ell)$  be the set of labeled hypergraphs with vertex set  $[\ell]$  having  $k$  hyperedges such that each hyperedge has cardinality exactly  $q$  and every vertex has degree at least 2. For each hypergraph  $G \in \mathcal{S}_2(\ell)$  the vertices are labeled by the indices from the set  $[\ell]$  and the hyperedges are labeled by the indices from the set  $[k]$ . Let  $M(S)$  for  $S \subseteq [\ell]$  be the set of all possible injective functions  $\pi : S \rightarrow [n]$ , in particular  $M([\ell])$  is the set of all possible injective functions  $\pi : [\ell] \rightarrow [n]$ . We will use the notation  $\pi(h)$  for a copy of hyperedge  $h \in \mathcal{H}(G)$  with its vertices relabeled by injective function  $\pi$ , i.e.  $\mathcal{V}(\pi(h)) = \{\pi(v) : v \in \mathcal{V}(h)\}$ . We claim that

$$\begin{aligned}
&\sum_{G \in \mathcal{P}_2(H, k)} \left( \prod_{h \in \mathcal{H}(G)} w_h \right) \left( \prod_{v \in \mathcal{V}(G)} \mathbb{E} \left[ |X_v|^{d_v(G)} \right] \right) \\
&= \sum_{\ell=q}^{kq/2} \frac{1}{\ell!} \sum_{G' \in \mathcal{S}_2(\ell)} \sum_{\pi \in M([\ell])} \left( \prod_{h \in \mathcal{H}(G')} w_{\pi(h)} \right) \left( \prod_{u \in \mathcal{V}(G')} \mathbb{E} \left[ |X_{\pi(u)}|^{d_u(G')} \right] \right). \quad (2.11)
\end{aligned}$$

Indeed, every labeled hypergraph  $G = (\mathcal{V}(G), \mathcal{H}(G)) \in \mathcal{P}_2(H, k)$  on  $\ell$  vertices has  $\ell!$  labeled hypergraphs  $G' = (\mathcal{V}(G'), \mathcal{H}(G')) \in \mathcal{S}_2(\ell)$  that differ from  $G$  by vertex labellings only. Each of those hypergraphs has one corresponding mapping  $\pi$  that maps its  $\ell$  vertex labels into vertex labels of hypergraph  $G \in \mathcal{P}_2(H, k)$ .

Then, combining (2.10) and (2.11) we obtain

$$\left| \mathbb{E} \left[ g(Y)^k \right] \right| \leq \sum_{\ell=q}^{kq/2} \frac{1}{\ell!} \sum_{G' \in \mathcal{S}_2(\ell)} \sum_{\pi \in M([\ell])} \left( \prod_{h \in \mathcal{H}(G')} w_{\pi(h)} \right) \left( \prod_{u \in \mathcal{V}(G')} \mathbb{E} \left[ |X_{\pi(u)}|^{d_u(G')} \right] \right). \quad (2.12)$$



## 2.2 Estimating the term for each hypergraph $G'$

We now fix integer  $\ell$  and labeled hypergraph  $G' \in \mathcal{S}_2(\ell)$ . Let  $c$  be the number of *connected components* in  $G'$ , i.e.  $c$  is a maximal number such that the vertex set  $\mathcal{V}(G')$  can be partitioned into  $c$  parts  $\mathcal{V}_1, \dots, \mathcal{V}_c$  such that for each hyperedge  $h \in \mathcal{H}(G')$  and any  $j \in [c]$  if  $\mathcal{V}(h) \cap \mathcal{V}_j \neq \emptyset$  then  $\mathcal{V}(h) \subseteq \mathcal{V}_j$ . Intuitively, we can split the vertex set of  $G'$  into  $c$  components such that there are no hyperedges that have vertices in two or more components. By definition of degree  $\sum_{v \in \mathcal{V}(G')} d_v = qk$  and  $d_v \geq 2$  for all  $v \in \mathcal{V}(G')$ .

We use a canonical ordering  $h^{(1)}, \dots, h^{(k)}$  of the hyperedges in  $\mathcal{H}(G')$  that will be specified later in Lemma 4.3. (This canonical ordering is distinct from and should not be confused with the ordering of the hyperedges inherent in a labeled hypergraph.) We iteratively remove hyperedges from the hypergraph  $G'$  in this order. Let  $G'_s = (\mathcal{V}'_s, \mathcal{H}'_s)$  be the hypergraph defined by the hyperedges  $\mathcal{H}'_s = h^{(s)}, \dots, h^{(k)}$  and vertex set  $\mathcal{V}'_s = \cup_{h \in \mathcal{H}'_s} \mathcal{V}(h)$ . In particular  $G'_1$  is identical to  $G'$  except for the order of the hyperedges. Let  $V_s$  be the vertices of the hyperedge  $h^{(s)}$  that have degree one in the hypergraph  $G'_s$ , i.e.  $\mathcal{V}'_{s+1} = \mathcal{V}'_s \setminus V_s$ . By definition,  $0 \leq |V_s| \leq q$ . Moreover,  $0 \leq |V_s| \leq q - 1$  for  $s = 1, \dots, k - c$  by Lemma 4.3 since the hyperedge  $h^{(s)}$  must be connected with at least one of remaining hyperedges. By the properties of the canonical ordering  $h^{(1)}, \dots, h^{(k)}$  from Lemma 4.3 we know that the first  $c$  edges (set  $S_2$  of hyperedges) in that ordering belong to different connected components. Since degree of each node is at least two we obtain that  $V_1 = \dots = V_c = \emptyset$ .

Analogously, we consider the second canonical ordering  $\tilde{h}^{(1)}, \dots, \tilde{h}^{(k)}$  from Lemma 4.3 and define analogous notions. Let  $\tilde{G}'_s = (\tilde{\mathcal{V}}'_s, \tilde{\mathcal{H}}'_s)$  be the hypergraph defined by the hyperedges  $\tilde{\mathcal{H}}'_s = \tilde{h}^{(s)}, \dots, \tilde{h}^{(k)}$  and vertex set  $\tilde{\mathcal{V}}'_s = \cup_{h \in \tilde{\mathcal{H}}'_s} \mathcal{V}(h)$ . Let  $\tilde{V}_s$  be the vertices of the hyperedge  $\tilde{h}^{(s)}$  that have degree one in the hypergraph  $\tilde{G}'_s$ , i.e.  $\tilde{\mathcal{V}}'_{s+1} = \tilde{\mathcal{V}}'_s \setminus \tilde{V}_s$ . The first  $c$  edges in the second canonical ordering  $\tilde{h}^{(1)}, \dots, \tilde{h}^{(k)}$  define the set  $S_1$  of hyperedges that belong to different connected components. Therefore,  $\tilde{V}_1 = \dots = \tilde{V}_c = \emptyset$ .

Let  $S_1$  and  $S_2$  be the sets of special hyperedges defined in Lemma 4.3. Each set  $S_i$  contains exactly one hyperedge per connected component of hypergraph  $G'$  and therefore, hyperedges belonging to the same  $S_i$  are disjoint. Let  $W'_1 = \cup_{h \in S_1} \mathcal{V}(h)$  be the set of vertices incident to hyperedges in  $S_1$  and  $W'_2 = \cup_{h \in S_2} \mathcal{V}(h)$  be the set of vertices incident to hyperedges in  $S_2$ . Note that  $|W'_1| = |W'_2| = qc$ .

We apply the standard fact  $ab \leq \frac{a^2+b^2}{2}$  to the  $\prod_{h \in S_1 \cup S_2} w_{\pi(h)}$  in the last term of the inequality (2.12) and obtain

$$\left( \prod_{h \in S_1 \cup S_2} w_{\pi(h)} \right) \leq \frac{1}{2} \left( \prod_{h \in S_1} w_{\pi(h)}^2 \right) + \frac{1}{2} \left( \prod_{h \in S_2} w_{\pi(h)}^2 \right). \quad (2.13)$$

We will use the notation  $d_u$  instead of  $d_u(G')$ . By central moment boundness, we estimate

$$\begin{aligned}
\prod_{u \in \mathcal{V}(G')} \mathbb{E} [|X_{\pi(u)}|^{d_u}] &\leq \left( \prod_{u \in W'_1} \frac{d_u!}{2} L^{d_u-2} \mathbb{E} [X_{\pi(u)}^2] \right) \left( \prod_{u \in \mathcal{V}(G') \setminus W'_1} d_u! L^{d_u-1} \mathbb{E} [|X_{\pi(u)}|] \right) \\
&\leq 2^{-qc} L^{qk-qc-\ell} \left( \prod_{u \in \mathcal{V}(G')} d_u! \right) \left( \prod_{u \in W'_1} \mathbb{E} [X_{\pi(u)}^2] \right) \prod_{u \in \mathcal{V}(G') \setminus W'_1} \mathbb{E} [|X_{\pi(u)}|] \\
&\leq 2^{\ell-2qc} L^{qk-qc-\ell} \left( \prod_{u \in \mathcal{V}(G')} d_u! \right) \left( \prod_{u \in W'_1} \mathbb{E} [X_{\pi(u)}^2] \right) \prod_{u \in \mathcal{V}(G') \setminus W'_1} \mathbb{E} [|Y_{\pi(u)}|]
\end{aligned} \tag{2.14}$$

where the last inequality uses the inequality  $\mathbb{E} [|X_{\pi(u)}|] = \mathbb{E} [|Y_{\pi(u)} - \mathbb{E} [Y_{\pi(u)}]|] \leq \mathbb{E} [|Y_{\pi(u)}|] + |\mathbb{E} [Y_{\pi(u)}]| \leq 2\mathbb{E} [|Y_{\pi(u)}|]$ .

Analogously,

$$\prod_{u \in \mathcal{V}(G')} \mathbb{E} [|X_{\pi(u)}|^{d_u}] \leq 2^{\ell-2qc} L^{qk-qc-\ell} \left( \prod_{u \in \mathcal{V}(G')} d_u! \right) \left( \prod_{u \in W'_2} \mathbb{E} [X_{\pi(u)}^2] \right) \prod_{u \in \mathcal{V}(G') \setminus W'_2} \mathbb{E} [|X_{\pi(u)}|] \tag{2.15}$$

Recall  $[\ell] = \mathcal{V}(G')$ ,  $\mathcal{V}'_s = \mathcal{V}(G') \setminus \cup_{t=1}^{s-1} V_t$  for  $s = 1, \dots, k$  and  $V_1 = \dots = V_c = \emptyset$ . Analogously,  $\tilde{\mathcal{V}}'_s = \mathcal{V}(G') \setminus \cup_{t=1}^{s-1} \tilde{V}_t$  for  $s = 1, \dots, k$  and  $\tilde{V}_1 = \dots = \tilde{V}_c = \emptyset$ . For each  $s = c+1, \dots, k-c$ , we will use the notations

$$\begin{aligned}
\Upsilon_s(\pi) &= \left( \prod_{u \in W'_1} \mathbb{E} [X_{\pi(u)}^2] \right) \prod_{u \in \mathcal{V}'_s \setminus W'_1} \mathbb{E} [|Y_{\pi(u)}|], \\
\tilde{\Upsilon}_s(\pi) &= \left( \prod_{u \in W'_2} \mathbb{E} [X_{\pi(u)}^2] \right) \prod_{u \in \tilde{\mathcal{V}}'_s \setminus W'_2} \mathbb{E} [|Y_{\pi(u)}|].
\end{aligned} \tag{2.16}$$

Therefore, combining inequalities (2.13), (2.14), (2.15) and notations (2.16), for each graph  $G' \in \mathcal{S}_2(\ell)$ , we obtain

$$\begin{aligned}
&\sum_{\pi \in M([\ell])} \left( \prod_{h \in \mathcal{H}(G')} w_{\pi(h)} \right) \left( \prod_{u \in \mathcal{V}(G')} \mathbb{E} [|X_{\pi(u)}|^{d_u(G')}] \right) \\
&\leq 2^{\ell-2qc-1} L^{qk-qc-\ell} \left( \prod_{u \in \mathcal{V}(G')} d_u! \right) \cdot \\
&\quad \left( \sum_{\pi \in M([\ell])} \left( \prod_{h \in S_1} w_{\pi(h)}^2 \right) \cdot \left( \prod_{h=h^{(c+1)}, \dots, h^{(k-c)}} w_{\pi(h)} \right) \Upsilon_1(\pi) \right. \\
&\quad \left. + \sum_{\pi \in M([\ell])} \left( \prod_{h \in S_2} w_{\pi(h)}^2 \right) \cdot \left( \prod_{h=\tilde{h}^{(c+1)}, \dots, \tilde{h}^{(k-c)}} w_{\pi(h)} \right) \tilde{\Upsilon}_1(\pi) \right)
\end{aligned} \tag{2.17}$$

We now analyze two terms in the inequality (2.17) separately using different canonical orderings of the hyperedges from Lemma 4.3. We consider  $\sum_{\pi \in M([\ell])} \left( \prod_{h \in S_1} w_{\pi(h)}^2 \right) \cdot \left( \prod_{h=h^{(c+1)}, \dots, h^{(k-c)}} w_{\pi(h)} \right) \Upsilon_1(\pi)$  and the corresponding canonical ordering of the hyperedges  $h^{(c+1)}, \dots, h^{(k)}$ . For each  $s = c + 1, \dots, k - c$  we obtain

$$\begin{aligned}
& \sum_{\pi \in M(\mathcal{V}'_s)} \left( \prod_{h \in S_1} w_{\pi(h)}^2 \right) \left( \prod_{h=h^{(s)}, \dots, h^{(k-c)}} w_{\pi(h)} \right) \Upsilon_s(\pi) \\
&= \sum_{\pi' \in M(\mathcal{V}'_{s+1})} \sum_{\substack{\pi \in M(\mathcal{V}'_s) \\ \text{s.t. } \pi \text{ extends } \pi'}} \left( \prod_{h \in S_1} w_{\pi(h)}^2 \right) \left( \prod_{h=h^{(s+1)}, \dots, h^{(k-c)}} w_{\pi(h)} \right) \Upsilon_{s+1}(\pi) \left( w_{\pi(h^{(s)})} \prod_{v \in V_s} \mathbb{E}[|Y_{\pi(v)}|] \right) \\
&= \sum_{\pi' \in M(\mathcal{V}'_{s+1})} \left( \prod_{h \in S_1} w_{\pi'(h)}^2 \right) \left( \prod_{h=h^{(s+1)}, \dots, h^{(k-c)}} w_{\pi'(h)} \right) \Upsilon_{s+1}(\pi') \left[ \sum_{\substack{\pi \in M(\mathcal{V}'_s) \\ \text{s.t. } \pi \text{ extends } \pi'}} \left( w_{\pi(h^{(s)})} \prod_{v \in V_s} \mathbb{E}[|Y_{\pi(v)}|] \right) \right]
\end{aligned}$$

where we say that  $\pi$  *extends*  $\pi'$  if  $\pi(v) = \pi'(v)$  for every  $v$  in the domain of  $\pi'$ .

We now group the sum over  $\pi$  by the value of  $\pi(h^{(s)}) \equiv h \in \mathcal{H}$ . Note that for any fixed mapping  $\pi' \in M(\mathcal{V}'_{s+1})$  there are exactly  $|V_s|!$  possible mappings  $\pi \in M(\mathcal{V}'_s)$  that extend  $\pi'$  and map the vertex labels of hyperedge  $h^{(s)} \in G'$  into vertex labels of the hyperedge  $h \in \mathcal{H}$ . Let  $S' = \{\pi'(v) : v \in \mathcal{V}(h^{(s)}) \setminus V_s\}$ , which is the portion of  $\pi(\mathcal{V}(h^{(s)}))$  that is fixed by  $\pi'$ . Then

$$\begin{aligned}
\sum_{\substack{\pi \in M(\mathcal{V}'_s) \\ \text{s.t. } \pi \text{ extends } \pi'}} w_{\pi(h^{(s)})} \prod_{v \in V_s} \mathbb{E}[|Y_{\pi(v)}|] &= |V_s|! \sum_{h \in \mathcal{H}: \mathcal{V}(h) \supseteq S'} w_h \prod_{u \in \mathcal{V}(h) \setminus S'} \mathbb{E}[|Y_u|] \\
&\leq |V_s|! \max_{S: |S|=q-|V_s|} \left\{ \sum_{h \in \mathcal{H}: \mathcal{V}(h) \supseteq S} w_h \prod_{u \in \mathcal{V}(h) \setminus S} \mathbb{E}[|Y_u|] \right\} \\
&= |V_s|! \mu_{q-|V_s|}(g, Y).
\end{aligned}$$

For  $s = k - c + 1, \dots, k$ ,  $\mathcal{H}'_s \subseteq S_1$ ,  $|V_s| = q$  and we have a similar argument (note that  $|M(\mathcal{V}'_{k+1})| =$

$$|M(\emptyset)| = 1)$$

$$\begin{aligned}
& \sum_{\pi \in M(\mathcal{V}'_s)} \left( \prod_{h \in \mathcal{H}'_s} w_{\pi(h)}^2 \right) \left( \prod_{u \in \mathcal{V}'_s} \mathbb{E} [X_{\pi(u)}^2] \right) \\
&= \sum_{\pi' \in M(\mathcal{V}'_{s+1})} \sum_{\substack{\pi \in M(\mathcal{V}'_s) \\ \text{s.t. } \pi \text{ extends } \pi'}} \left( \prod_{h \in \mathcal{H}'_{s+1}} w_{\pi(h)}^2 \right) \left( \prod_{u \in \mathcal{V}'_{s+1}} \mathbb{E} [X_{\pi(u)}^2] \right) \left( w_{\pi(h^{(s)})}^2 \prod_{u \in \mathcal{V}_s} \mathbb{E} [X_{\pi(u)}^2] \right) \\
&= \sum_{\pi' \in M(\mathcal{V}'_{s+1})} \left[ \left( \prod_{h \in \mathcal{H}'_{s+1}} w_{\pi'(h)}^2 \right) \left( \prod_{u \in \mathcal{V}'_{s+1}} \mathbb{E} [X_{\pi'(u)}^2] \right) \sum_{\substack{\pi \in M(\mathcal{V}'_s) \\ \text{s.t. } \pi \text{ extends } \pi'}} \left( w_{\pi(h^{(s)})}^2 \prod_{u \in \mathcal{V}_s} \mathbb{E} [X_{\pi(u)}^2] \right) \right] \\
&= |V_s|! \cdot \sum_{\pi' \in M(\mathcal{V}'_{s+1})} \left[ \left( \prod_{h \in \mathcal{H}'_{s+1}} w_{\pi'(h)}^2 \right) \left( \prod_{u \in \mathcal{V}'_{s+1}} \mathbb{E} [X_{\pi'(u)}^2] \right) \right] \cdot \left( \sum_{h \in \mathcal{H}} w_h^2 \prod_{u \in \mathcal{V}(h)} \mathbb{E} [X_u^2] \right) \\
&= |V_s|! \cdot \sum_{\pi' \in M(\mathcal{V}'_{s+1})} \left[ \left( \prod_{h \in \mathcal{H}'_{s+1}} w_{\pi'(h)}^2 \right) \left( \prod_{u \in \mathcal{V}'_{s+1}} \mathbb{E} [X_{\pi'(u)}^2] \right) \right] \cdot \text{Var}[g(X)],
\end{aligned}$$

where in the last equality we used Lemma 4.1.

In the end we bound the first term of (2.17) as follows:

$$\begin{aligned}
& 2^{\ell-2qc-1} L^{qk-qc-\ell} \left( \prod_{u \in \mathcal{V}(G')} d_u! \right) \sum_{\pi \in M([\ell])} \left( \prod_{h \in S_1} w_{\pi(h)}^2 \right) \cdot \left( \prod_{h=h^{(c+1)}, \dots, h^{(k-c)}} w_{\pi(h)} \right) \Upsilon_1(\pi) \\
&\leq 2^{\ell-2qc-1} L^{qk-qc-\ell} \left( \prod_{v \in \mathcal{V}(G')} d_v! \right) \left( \prod_{s=k-c+1}^k |V_s|! \text{Var}[g(X)] \right) \prod_{s=c+1}^{k-c} (|V_s|! \mu_{q-|V_s|}) \\
&= 2^{\ell-2q\sigma_0-1} L^{qk-q\sigma_0-\ell} \left( \prod_{v \in \mathcal{V}(G')} d_v! \right) \left( \prod_{s=1}^k |V_s|! \right) \text{Var}[g(X)]^{\sigma_0} \prod_{t=1}^q \mu_t^{\sigma_t} \\
&\leq 2^{\ell-2q\sigma_0-1} L^{qk-q\sigma_0-\ell} \left( \prod_{v \in \mathcal{V}(G')} d_v! \right) \text{Var}[g(X)]^{\sigma_0} q^\ell \cdot \prod_{t=1}^q \mu_t^{\sigma_t} \tag{2.18}
\end{aligned}$$

where  $\sigma_0 = c$ ,  $\sigma_t$  for  $t \geq 1$  is the number of indices  $s = c+1, \dots, k-c$  with  $q - |V_s| = t$  and  $\mu_t = \mu_t(w, Y)$ . In the last inequality we used the fact that  $\sum_{s=1}^k |V_s| = \ell$  and  $|V_s| \leq q$ . The quantities  $\sigma_t$  must satisfy the equalities  $\sum_{t=0}^q (q-t)\sigma_t = \ell$  and  $2\sigma_0 + \sum_{t=1}^q \sigma_t = k$ .

Using analogous argument for the canonical ordering  $\tilde{h}^{(1)}, \dots, \tilde{h}^{(k)}$  we show

$$\begin{aligned}
& 2^{\ell-2qc-1} L^{qk-qc-\ell} \left( \prod_{u \in \mathcal{V}(G')} d_u! \right) \sum_{\pi \in M([\ell])} \left( \prod_{h \in S_2} w_{\pi(h)}^2 \right) \cdot \left( \prod_{h=\tilde{h}^{(c+1)}, \dots, \tilde{h}^{(k-c)}} w_{\pi(h)} \right) \tilde{\Upsilon}_1(\pi) \\
&\leq 2^{\ell-2q\sigma'_0-1} L^{qk-q\sigma'_0-\ell} \left( \prod_{v \in \mathcal{V}(G')} d_v! \right) \text{Var}[g(X)]^{\sigma'_0} q^\ell \cdot \prod_{t=1}^q \mu_t^{\sigma'_t}, \tag{2.19}
\end{aligned}$$

where  $\sigma'_0 = c$  and  $\sigma'_1, \dots, \sigma'_q$  is a different collection of powers satisfying conditions of the Lemma. Combining the inequalities (2.17), (2.18) and (2.19) we derive

$$\begin{aligned} & \sum_{\pi \in M([\ell])} \left( \prod_{h \in \mathcal{H}(G')} w_{\pi(h)} \right) \left( \prod_{u \in \mathcal{V}(G')} \mathbb{E} \left[ |X_{\pi(u)}|^{d_u(G')} \right] \right) \\ & \leq \max_{\bar{\sigma}} \left\{ 2^{\ell-2q\sigma_0} L^{qk-q\sigma_0-\ell} \left( \prod_{v \in \mathcal{V}(G')} d_v! \right) \text{Var}[g(X)]^{\sigma_0} q^\ell \cdot \prod_{t=1}^q \mu_t^{\sigma_t} \right\} \end{aligned} \quad (2.20)$$

where the maximum is over all non-negative integers  $\sigma_0, \sigma_1, \dots, \sigma_q$  satisfying  $2\sigma_0 + \sum_{t=1}^q \sigma_t = k$ ,  $\sigma_0 = c$  and  $\ell = \sum_{t=0}^q (q-t)\sigma_t$ .

### 2.3 Using the Counting Lemma

We decompose  $\mathcal{S}_2(\ell)$  as  $\mathcal{S}_2(\ell) = \bigcup_{c, \bar{d} \geq \bar{2}} \mathcal{S}(\ell, c, \bar{d})$  where  $\bar{2}$  is a vector of  $\ell$  twos and  $\mathcal{S}(\ell, c, \bar{d})$  is the number of vertex and hyperedge labeled hypergraphs with vertex set  $[\ell]$  and  $k$  labeled hyperedges such that each hyperedge has cardinality  $q$ , the number of connected components is  $c$ , the degree vector is  $\bar{d}$ . (Note that  $\mathcal{S}(\ell, c, \bar{d})$  depends on  $k$  and  $q$  as well.) Let  $\bar{\sigma} = (\sigma_0, \dots, \sigma_q)$ . Combining, (2.12) and (2.20) we obtain

$$\begin{aligned} \left| \mathbb{E} [g(X)^k] \right| & \leq \sum_{\ell=q}^{kq/2} \frac{1}{\ell!} \sum_{c=1}^{\ell/q} \sum_{\bar{d} \geq \bar{2}} \left( \sum_{G' \in \mathcal{S}(\ell, c, \bar{d})} \max_{\bar{\sigma}} \left\{ 2^{\ell-2q\sigma_0} L^{qk-q\sigma_0-\ell} \text{Var}[g(X)]^{\sigma_0} \left( \prod_{t=1}^q \mu_t^{\sigma_t} \right) q^\ell \prod_{v \in [\ell]} d_v! \right\} \right) \\ & \leq \max_{\ell, c, \bar{d} \geq \bar{2}, \bar{\sigma}} \left\{ \frac{kq}{2} \cdot \frac{1}{\ell!} \cdot \frac{\ell}{q} \cdot 2^{qk+\ell} \cdot |\mathcal{S}(\ell, c, \bar{d})| \cdot 2^{\ell-2q\sigma_0} L^{qk-q\sigma_0-\ell} \text{Var}[g(X)]^{\sigma_0} \left( \prod_{t=1}^q \mu_t^{\sigma_t} \right) q^\ell \prod_v d_v! \right\} \\ & \leq \max_{\ell, c, \bar{d} \geq \bar{2}, \bar{\sigma}} \left\{ \frac{k\ell}{2 \cdot \ell!} \cdot 2^{q(k-2\sigma_0)+2\ell} \cdot L^{qk-q\sigma_0-\ell} \text{Var}[g(X)]^{\sigma_0} \cdot q^\ell \cdot \underbrace{R_0^{qk} \cdot k^{qk-(q-1)c}}_{|\mathcal{S}(\ell, c, \bar{d})| \prod_v d_v! \leq \text{this by counting Lemma 4.4}} \cdot \left( \prod_{t=1}^q \mu_t^{\sigma_t} \right) \right\} \end{aligned}$$

where the sum is over  $\bar{d} \geq \bar{2}$  with  $\sum_{v \in [\ell]} d_v = qk$  and the maximum over  $\bar{\sigma}$  has the same constraints as in (2.20). The maximums over  $\ell$ ,  $c$ , and  $\bar{d}$  are over the same sets that those quantities were previously summed over. The second inequality follows from the fact that the total number of feasible degree vectors  $\bar{d}$  is at most  $2^{qk+\ell}$  ( $qk$  is the sum of all the degrees and we need to compute the total number of partitions of the array with  $qk$  entries into  $\ell$  possible groups of consecutive entries which is  $\binom{qk+\ell-1}{\ell-1}$ ).

We now substitute  $\sigma_0$  for  $c$ , and remove the unreferenced variables  $c$  and  $\bar{d}$  from the maximum. We

also remove  $\ell$  from the maximum since it is completely defined by the vector  $\bar{\sigma}$ . We continue

$$\begin{aligned}
\left| \mathbb{E} \left[ g(X)^k \right] \right| &\leq \max_{\bar{\sigma}} \left\{ \frac{k\ell}{\ell!} \cdot L^{qk-q\sigma_0-\ell} \text{Var}[g(X)]^{\sigma_0} \cdot q^\ell \cdot R_1^{qk} \cdot k^{qk-(q-1)\sigma_0} \cdot \left( \prod_{t=1}^q \mu_t^{\sigma_t} \right) \right\} \\
&\leq \max_{\bar{\sigma}} \left\{ R_2^{qk} \cdot \left( \frac{q}{\ell} \right)^\ell \cdot L^{qk-q\sigma_0-\ell} \text{Var}[g(X)]^{\sigma_0} \cdot k^{qk-(q-1)\sigma_0} \cdot \left( \prod_{t=1}^q \mu_t^{\sigma_t} \right) \right\} \\
&= \max_{\bar{\sigma}} \left\{ R_2^{qk} \cdot \left( \frac{qk}{\ell} \right)^\ell \cdot L^{qk-q\sigma_0-\ell} \text{Var}[g(X)]^{\sigma_0} \cdot k^{qk-(q-1)\sigma_0-\ell} \cdot \left( \prod_{t=1}^q \mu_t^{\sigma_t} \right) \right\} \\
&\leq \max_{\bar{\sigma}} \left\{ R_3^{qk} L^{qk-q\sigma_0-\ell} \text{Var}[g(X)]^{\sigma_0} \cdot k^{qk-(q-1)\sigma_0-\ell} \cdot \left( \prod_{t=1}^q \mu_t^{\sigma_t} \right) \right\} \tag{2.21}
\end{aligned}$$

where  $R_0 < R_1 < R_2 < R_3$  are some absolute constants, the second inequality uses the fact that  $\ell! \geq (\ell/e)^\ell$ , and the last inequality is implied by the fact that

$$\left( \frac{qk}{\ell} \right)^\ell \leq \max_{x>0} \left( \frac{kq}{x} \right)^x = e^{qk/e}.$$

Inequality (2.21) is precisely the inequality (2.8) that we needed to prove. ■

### 3 Intermediate moment lemma

**Lemma 3.1 (Intermediate Moment Lemma)** *We are given  $n$  independent central moment bounded random variables  $Y = (Y_1, \dots, Y_n)$  with the same parameter  $L > 0$  and a general polynomial  $g(x)$  with nonnegative coefficients such that every monomial (or hyperedge)  $h \in \mathcal{H}$  has power exactly  $q$ . Let  $X_v = Y_v - \mathbb{E}[Y_v]$  then*

$$\left| \mathbb{E} \left[ g(X)^k \right] \right| \leq \max \left\{ \left( \sqrt{k R_3^q \text{Var}[g(X)]} \right)^k, \max_{t \in [q]} (k^t R_3^q L^t \mu_t(g, Y))^k \right\}. \tag{3.22}$$

where  $R_3 \geq 1$  is some absolute constant and  $X$  is the vector of centered random variables defined in the Lemma 2.1.

**Proof.** First we note that  $2\sigma_0 + \sum_{t=1}^q \sigma_t = k$  and  $\sum_{t=0}^q (q-t)\sigma_t = \ell$  imply,

$$\begin{aligned}
\sum_{t=1}^q t\sigma_t &= qk - q(2\sigma_0 + \sum_{t=1}^q \sigma_t) + \sum_{t=1}^q t\sigma_t \\
&= qk - q\sigma_0 - \sum_{t=0}^q (q-t)\sigma_t \\
&= qk - q\sigma_0 - \ell.
\end{aligned}$$

Therefore,  $\sigma_0 + \sum_{t=1}^q t\sigma_t = qk - (q-1)\sigma_0 - \ell$ . Combining these facts with Lemma 2.1, we derive

$$\begin{aligned}
\left| \mathbb{E} \left[ g(X)^k \right] \right| &\leq \max_{\vec{\sigma}} \left\{ R_3^{qk} L^{qk-q\sigma_0-\ell} \text{Var}[g(X)]^{\sigma_0} \cdot k^{qk-(q-1)\sigma_0-\ell} \cdot \left( \prod_{t=1}^q \mu_t^{\sigma_t} \right) \right\} \\
&\leq \max_{\vec{\sigma}} \left\{ (k R_3^q \text{Var}[g(X)])^{\sigma_0} \cdot \prod_{t=1}^q (k^t R_3^q L^t \mu_t)^{\sigma_t} \right\} \\
&= \max_{\vec{\sigma}} \left\{ \left( \sqrt{k R_3^q \text{Var}[g(X)]} \right)^{2\sigma_0} \cdot \prod_{t=1}^q (k^t R_3^q L^t \mu_t)^{\sigma_t} \right\} \\
&\leq \max \left\{ \left( \sqrt{k R_3^q \text{Var}[g(X)]} \right)^k, \max_{t \in [q]} (k^t R_3^q L^t \mu_t)^k \right\},
\end{aligned}$$

where the last inequality is based on the fact that  $2\sigma_0 + \sum_{t=1}^q \sigma_t = k$ . ■

## 4 Technical Lemmas

**Lemma 4.1** *We are given  $n$  independent random variables  $X_1, \dots, X_n$  such that  $\mathbb{E}[X_i] = 0$  for all  $i \in [n]$ . We are also given a multilinear polynomial  $f(x) = \sum_{h \in \mathcal{H}(H)} w_h \prod_{v \in \mathcal{V}(h)} x_v$  with  $|\mathcal{V}(h)| \geq 1$  for any  $h \in \mathcal{H}(H)$ . Then*

$$\text{Var}[f(X)] = \sum_{h \in \mathcal{H}(H)} w_h^2 \prod_{v \in \mathcal{V}(h)} \mathbb{E}[X_v^2].$$

**Proof.** Let  $\mathcal{H} = \mathcal{H}(H)$ . Clearly  $\mathbb{E}[f(X)] = 0$  hence

$$\begin{aligned}
\text{Var}[f(X)] &= \mathbb{E}[(f(X) - \mathbb{E}[f(X)])^2] = \mathbb{E}[f(X)^2] = \mathbb{E} \left[ \left( \sum_{h \in \mathcal{H}} w_h \prod_{v \in h} X_v \right)^2 \right] \\
&= \sum_{h \in \mathcal{H}} \sum_{h' \in \mathcal{H}} w_h w_{h'} \mathbb{E} \left[ \left( \prod_{v \in h} X_v \right) \left( \prod_{v \in h'} X_v \right) \right] \\
&= \sum_{h \in \mathcal{H}} \sum_{h' \in \mathcal{H}} w_h w_{h'} \left( \prod_{v \in h \cap h'} \mathbb{E}[X_v^2] \right) \prod_{v \in (h \setminus h') \cup (h' \setminus h)} \mathbb{E}[X_v] \\
&= \sum_{h \in \mathcal{H}} w_h^2 \prod_{v \in h} \mathbb{E}[X_v^2]
\end{aligned}$$

where the last equality follows because  $\mathbb{E}[X_v] = 0$  by assumption. ■

**Lemma 4.2** *We are given  $n$  independent random variables  $X_1, \dots, X_n$  such that  $\mathbb{E}[X_i] = 0$  for all  $i \in [n]$ . We are also given two multilinear polynomials  $g_1(x) = \sum_{h \in \mathcal{H}(H)} w_h \prod_{v \in h} x_v$  and  $g_2(x) = \sum_{h \in \mathcal{H}(H)} w'_h \prod_{v \in h} x_v$  such that  $w_h w'_h = 0$  for any hyperedge  $h$ . Then  $\mathbb{E}[g_1(X)g_2(X)] = 0$ .*

**Proof.** The proof is similar to the proof of Lemma 4.1:

$$\begin{aligned}
\mathbb{E}[g_1(X)g_2(X)] &= \sum_{h \in \mathcal{H}(H)} \sum_{h' \in \mathcal{H}(H)} w_h w'_{h'} \mathbb{E} \left[ \left( \prod_{v \in h} X_v \right) \left( \prod_{v \in h'} X_v \right) \right] \\
&= \sum_{h \in \mathcal{H}(H)} \sum_{h' \in \mathcal{H}(H)} w_h w'_{h'} \left( \prod_{v \in h \cap h'} \mathbb{E}[X_v^2] \right) \left( \prod_{v \in (h \setminus h') \cup (h' \setminus h)} \mathbb{E}[X_v] \right) \\
&= 0
\end{aligned}$$



where the final equality follows because each  $h, h'$  term either has  $w_h w_{h'}' = 0$  or else  $h \neq h'$  and hence at least one  $\mathbb{E}[X_v] = 0$  factor.  $\blacksquare$

**Lemma 4.3 (Ordering Lemma)** *We are given a hypergraph  $G' = (\mathcal{V}, \mathcal{H})$  with  $c$  connected components and degree of each vertex  $\geq 2$ . We can define two disjoint sets of hyperedges  $S_1$  and  $S_2$  each containing exactly one hyperedge per connected component of  $G'$ , i.e.  $|S_1| = |S_2| = c$ , such that there exist two canonical orderings  $h^{(1)}, \dots, h^{(k)}$  and  $\tilde{h}^{(1)}, \dots, \tilde{h}^{(k)}$  of the hyperedges  $\mathcal{H}$  with the following properties:*

1.  $S_2 = \{h^{(1)}, \dots, h^{(c)}\} = \{\tilde{h}^{(k-c+1)}, \dots, \tilde{h}^{(k)}\}$  and  $S_1 = \{h^{(k-c+1)}, \dots, h^{(k)}\} = \{\tilde{h}^{(1)}, \dots, \tilde{h}^{(c)}\}$ , i.e. the hyperedges from  $S_2$  appear first in the canonical ordering  $h^{(1)}, \dots, h^{(k)}$  and last in the canonical ordering  $\tilde{h}^{(1)}, \dots, \tilde{h}^{(k)}$ , while the hyperedges from  $S_1$  appear last in the canonical ordering  $h^{(1)}, \dots, h^{(k)}$  and first in the canonical ordering  $\tilde{h}^{(1)}, \dots, \tilde{h}^{(k)}$ ;
2. for any  $s = 1, \dots, k - c$  the hypergraph  $G_s$  induced by the hyperedges  $h^{(s)}, \dots, h^{(k)}$  has exactly  $c$  connected components;
3. Analogously, for any  $s = 1, \dots, k - c$  the hypergraph  $\tilde{G}_s$  induced by the hyperedges  $\tilde{h}^{(s)}, \dots, \tilde{h}^{(k)}$  has exactly  $c$  connected components.

**Proof.** Let  $\mathcal{L}$  be the line graph of  $G'$ , i.e. an undirected graph with one vertex for each of the  $k$  hyperedges of  $G'$  and an edge connecting every pair of vertices that correspond to hyperedges with intersecting vertex sets. Pick an arbitrary spanning forest  $\mathcal{F}$  of  $\mathcal{L}$ . Pick two leaves arbitrarily from each connected component of  $\mathcal{F}$  and arbitrarily put one from each component in  $S_1$  and the others in  $S_2$ . The existence of at least two leaves in each component follows because all vertices of  $G'$  have degrees at least 2 and hence each connected component has at least two hyperedges. It is easy to see that any tree with at least two vertices has at least two leaves.<sup>1</sup>

We show the construction of  $h^{(1)}, \dots, h^{(k)}$  only; the construction of  $\tilde{h}^{(1)}, \dots, \tilde{h}^{(k)}$  is analogous with the roles of  $S_1$  and  $S_2$  swapped.

We pick  $h^{(1)}, \dots, h^{(k)}$  iteratively (in that order) as follows. Let  $\mathcal{F}_i$  denote the subforest of  $\mathcal{F}$  induced by vertices  $\mathcal{H} \setminus \{h^{(1)}, \dots, h^{(i-1)}\}$ . We pick  $h^{(i)}$  to be an arbitrary leaf of  $\mathcal{F}_i$  subject to the constraint that  $h^{(i)} \in S_2$  if  $1 \leq i \leq c$ ,  $h^{(i)} \notin S_1 \cup S_2$  if  $c + 1 \leq i \leq k - c$ , and  $h^{(i)} \in S_1$  if  $k - c + 1 \leq i \leq k$ .

For any  $1 \leq i \leq k$  we assert that:

1. there is a leaf satisfying the desired constraint available to be  $h^{(i)}$  and
2. if  $i \leq k - c$  there are  $c$  connected components of  $\mathcal{F}_{i+1}$  and each contains a vertex (hyperedge of  $G'$ ) in  $S_1$ .

The second property follows because we always choose a leaf and never choose a vertex from  $S_1$ .

For  $1 \leq i \leq c$  the first property follows because removing a vertex from a graph cannot make a leaf into a non-leaf and every vertex of  $S_2$  is a leaf in  $\mathcal{F}_1$ . For  $c + 1 \leq i \leq k - c$  the first property follows because the second property implies that there is a connected component of  $\mathcal{F}_i$  with at least two vertices and hence leaves and at most one can be from  $S_1$  and none from  $S_2$ .  $\blacksquare$

The next lemma was proven in [18] in the setting of general polynomials. We state below a special case corresponding to multilinear polynomials.

---

<sup>1</sup>Indeed root each tree arbitrarily; if the root has degree two or more pick arbitrary leaf descendants (in the rooted sense) of two neighbors of the root; otherwise pick the root and an arbitrary leaf descendant of the root.

**Lemma 4.4 (Main Counting Lemma [18])** For any  $k, q \geq 1, \ell, c$  and  $\bar{d} \geq \bar{2}$  we have

$$|\mathcal{S}(\ell, c, \bar{d})| \left( \prod_{v \in [\ell]} d_v! \right) \leq R_0^{qk} k^{qk - (q-1)c},$$

for some universal constant  $R_0 > 1$ .

**Lemma 4.5 (Hölder's Inequality)** Let  $p_1, \dots, p_k \in (1, +\infty)$  such that  $\sum_{i=1}^k \frac{1}{p_i} = 1$  then for arbitrary collection  $X_1, \dots, X_k$  of random variables on the same probability space the following inequality holds

$$\mathbb{E} \left[ \left| \prod_{i=1}^k X_i \right| \right] \leq \prod_{i=1}^k \mathbb{E} [|X_i|^{p_i}]^{1/p_i}.$$

We will use the following corollary of Hölder's inequality.

**Corollary 4.6 (Minkowski's Inequality)** Let  $k$  be a positive integer and  $Z_1, Z_2, \dots, Z_m$  be (potentially dependent) random variables with  $\mathbb{E}[|Z_i|^k] \leq z_i^k$  for  $z_i \in R_+$ . It follows that

$$\mathbb{E} \left[ \left( \sum_{i=1}^m |Z_i| \right)^k \right] \leq \left( \sum_{i=1}^m z_i \right)^k. \quad (4.23)$$

## 5 General Even Moment Lemma

**Lemma 5.1 (General Even Moment Lemma)** We are given  $n$  independent central moment bounded random variables  $Y = (Y_1, \dots, Y_n)$  with the same parameter  $L > 0$  and a general power  $q$  polynomial  $f(x)$ . Let  $k \geq 2$  be an even integer then

$$\mathbb{E} [|f(Y) - \mathbb{E}[f(Y)]|^k] \leq \max \left\{ \left( \sqrt{k R_4^q \text{Var}[f(Y)]} \right)^k, \max_{t \in [q]} (k^t R_4^q L^t \mu_t(f, Y))^k \right\}. \quad (5.24)$$

where  $R_4 \geq 1$  is some absolute constant.

**Proof.** Let weight function  $w$  and hypergraph  $H = ([n], \mathcal{H})$  be such that  $f(Y) = \sum_{h \in \mathcal{H}} w_h \prod_{v \in \mathcal{V}(h)} Y_v$ . Let  $X_v = Y_v - \mathbb{E}[Y_v]$ . Let  $\mathcal{H}'$  denote the set of all possible hyperedges (including the empty hyper-edge) with at most  $q$  vertices (from  $\mathcal{V}(H) = [n]$ ). First we note that

$$\begin{aligned} f(Y) &= \sum_{h \in \mathcal{H}} w_h \prod_{v \in \mathcal{V}(h)} (X_v + \mathbb{E}[Y_v]) \\ &= \sum_{h' \in \mathcal{H}'} \sum_{h \in \mathcal{H}: \mathcal{V}(h) \supseteq \mathcal{V}(h')} w_h \left( \prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h')} \mathbb{E}[Y_v] \right) \left( \prod_{v \in \mathcal{V}(h')} X_v \right) \\ &= \sum_{h' \in \mathcal{H}'} w'_{h'} \prod_{v \in \mathcal{V}(h')} X_v \end{aligned} \quad (5.25)$$

where

$$w'_{h'} = \sum_{h \in \mathcal{H}: \mathcal{V}(h) \supseteq \mathcal{V}(h')} w_h \left( \prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h')} \mathbb{E}[Y_v] \right).$$

We next group the monomials on the right hand side of (5.25) by cardinality and sign of coefficient, yielding  $m \leq 2q$  polynomials  $g^{(1)}, \dots, g^{(m)}$  with corresponding weight functions for all monomials  $w^{(1)}, \dots, w^{(m)}$  and powers  $q_1, \dots, q_m$ . That is,

$$\begin{aligned} f(Y) &= w'_{\{\}} + \sum_{i=1}^m \sum_{h': |h'| \geq 1} w_{h'}^{(i)} \prod_{v \in \mathcal{V}(h')} X_v \\ &= \mathbb{E}[f(Y)] + \sum_{i=1}^m g^{(i)}(X) \end{aligned} \quad (5.26)$$

where  $\{\}$  is the empty hyperedge. We have

$$\begin{aligned} \mu_r(w^{(i)}, Y) &\leq \mu_r(w', Y) = \max_{S: |S|=r} \sum_{h': \mathcal{V}(h') \supseteq S} |w'_{h'}| \prod_{v \in \mathcal{V}(h') \setminus S} \mathbb{E}[|Y_v|] \\ &\leq \max_{S: |S|=r} \sum_{h': \mathcal{V}(h') \supseteq S} \sum_{h \in \mathcal{H}: \mathcal{V}(h) \supseteq \mathcal{V}(h')} |w_h| \left( \prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h')} \mathbb{E}[|Y_v|] \right) \prod_{v \in \mathcal{V}(h') \setminus \mathcal{V}(h_0)} \mathbb{E}[|Y_v|] \\ &\leq 2^q \max_{S: |S|=r} \sum_{h: \mathcal{V}(h) \supseteq S} |w_h| \left( \prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h_0)} \mathbb{E}[|Y_v|] \right) = 2^q \mu_r(w, Y) = 2^q \mu_r, \end{aligned} \quad (5.27)$$

where the last inequality follows from the fact that for any  $h \in \mathcal{H}$  the number of different  $h' \in \mathcal{H}'$  such that  $\mathcal{V}(h) \supseteq \mathcal{V}(h') \supseteq S$  is at most  $2^q$ . In addition by Lemma 4.2 we derive

$$\text{Var}[f(Y)] = \sum_{i=1}^m \text{Var}[g^{(i)}(X)]. \quad (5.28)$$

For even  $k \geq 2$  Lemma 3.1 implies that

$$\mathbb{E} \left[ \left| g^{(i)}(X) \right|^k \right] = \left| \mathbb{E} \left[ g^{(i)}(X)^k \right] \right| \leq \max \left\{ \left( \sqrt{k R_3^{q_i} \text{Var}[g^{(i)}(X)]} \right)^k, \max_{t \in [q_i]} (k^t R_3^{q_i} L^t \mu_t(w^{(i)}, Y))^k \right\} = z_i^k.$$

Applying Corollary 4.6 together with (5.27) and (5.28) yields

$$\begin{aligned} \mathbb{E} \left[ |f(Y) - \mathbb{E}[f(Y)]|^k \right] &\leq \mathbb{E} \left[ \left( \sum_{i=1}^m |g^{(i)}(X)| \right)^k \right] \leq \left( \sum_{i=1}^m z_i \right)^k \leq m^k \max_i z_i^k \\ &\leq \max \left\{ \left( \sqrt{k R_4^q \text{Var}[f(Y)]} \right)^k, \max_{t \in [q]} (k^t R_4^q L^t \mu_t)^k \right\} \end{aligned}$$

where we choose  $R_4 \geq 1$  such that  $m^{2k} R_3^{qk} 2^{qk} \leq R_4^{qk}$ . ■

**Proof.** [Proof of Lemma 1.5] As in the proof of Lemma 5.1 we write  $f(Y) = \mathbb{E}[f(Y)] + \sum_i g^{(i)}(X)$  where  $X_v = Y_v - \mathbb{E}[Y_v]$ . Let  $\mathcal{H}'$ ,  $m$ ,  $g^{(i)}$ ,  $w^{(i)}$  and  $q_i$  be defined as in that proof. Using Lemma 4.1,

the inequality (5.27) with  $r = q_i$  and the central moment boundness we get

$$\begin{aligned}
\text{Var}[g^{(i)}(X)] &= \sum_{h \in \mathcal{H}'} (w_h^{(i)})^2 \prod_{v \in \mathcal{V}(h)} \mathbb{E}[X_v^2] \\
&\leq \sum_{h \in \mathcal{H}'} \mu_{q_i}(w^{(i)}, Y) \cdot |w_h^{(i)}| \cdot \prod_{v \in \mathcal{V}(h)} (2L \mathbb{E}[|X_v|]) \\
&\leq \sum_{h \in \mathcal{H}'} \mu_{q_i}(w^{(i)}, Y) \cdot |w_h^{(i)}| \cdot \prod_{v \in \mathcal{V}(h)} (4L \mathbb{E}[|Y_v|]) \\
&= (4L)^{q_i} \mu_{q_i}(w^{(i)}, Y) \mu_0(w^{(i)}, Y).
\end{aligned} \tag{5.29}$$

Combining Lemma 4.2 and the inequality (5.29) we get

$$\begin{aligned}
\text{Var}[f(Y)] &= \sum_{i=1}^m \text{Var}[g^{(i)}(X)] \\
&\leq 2q \max_{r \in [q]} (4L)^r \mu_r(w^{(i)}, Y) \mu_0(w^{(i)}, Y) \\
&\leq 2q 4^q \max_{r \in [q]} 4^r L^r \mu_r(w, Y) \mu_0(w, Y)
\end{aligned}$$

where the last inequality uses (5.27). ■

## 6 Proof of the Theorem 1.3

Now we prove Theorem 1.3 by applying Markov's inequality.

**Proof.** By Markov's inequality we derive

$$Pr[|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] = Pr[|f(Y) - \mathbb{E}[f(Y)]|^k \geq \lambda^k] \leq \frac{\mathbb{E}[|f(Y) - \mathbb{E}[f(Y)]|^k]}{\lambda^k}.$$

Choosing  $k^* \geq 0$  to be the even integer such that  $k^* \in (K - 2, K]$  for

$$K = \min \left\{ \frac{\lambda^2}{e^2 R_4^q \text{Var}[f(Y)]}, \min_{t \in [q]} \left( \frac{\lambda}{e R_4^q L^t \mu_t} \right)^{1/t} \right\}$$

i.e.

$$\frac{\sqrt{k^* R_4^q \text{Var}[f(Y)]}}{\lambda} \leq 1/e \text{ and } \frac{(k^*)^t R_4^q L^t \mu_t}{\lambda} \leq 1/e$$

for all  $t \in [q]$ . Using inequality (5.24) from Lemma 5.1 we derive

$$\begin{aligned}
Pr[|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] &\leq \frac{\mathbb{E}[|f(Y) - \mathbb{E}[f(Y)]|^{k^*}]}{\lambda^{k^*}} \\
&\leq \max \left\{ e^{k^* \ln \frac{\sqrt{k^* R_4^q \text{Var}[f(Y)]}}{\lambda}}, \max_{t \in [q]} e^{k^* \ln \frac{(k^*)^t R_4^q L^t \mu_t}{\lambda}} \right\} \\
&\leq e^{-k^*} \leq e^{-K+2} \\
&\leq e^2 \cdot \max \left\{ e^{-\frac{\lambda^2}{R^q \text{Var}[f(Y)]}}, \max_{t \in [q]} e^{-\left( \frac{\lambda}{R^q L^t \mu_t} \right)^{1/t}} \right\},
\end{aligned}$$

for some universal constant  $R > R_4$ . This implies the statement of the Theorem. ■

## 7 Examples of Central Moment Bounded Random Variables

### 7.1 Bounded Random variables

**Lemma 7.1** *Any random variable  $Z$  with  $|Z - \mathbb{E}[Z]| \leq L$  is central moment bounded with parameter  $L$ .*

**Proof.** For any  $i \geq 1$  we clearly have  $|Z - \mathbb{E}[Z]|^i \leq L|Z - \mathbb{E}[Z]|^{i-1}$  hence  $\mathbb{E}[|Z - \mathbb{E}[Z]|^i] \leq L\mathbb{E}[|Z - \mathbb{E}[Z]|^{i-1}] \leq iL\mathbb{E}[|Z - \mathbb{E}[Z]|^{i-1}]$ . ■

### 7.2 Continuous log-concave random variables

We say that non-negative function  $f(x)$  is *log-concave* if  $f(\lambda x + (1 - \lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}$  for any  $0 \leq \lambda \leq 1$  and  $x, y \in \mathbb{R}$  (see [5] Section 3.5). Equivalently  $f$  is log concave if  $\ln f(x)$  is concave on the set  $\{x : f(x) > 0\}$  where  $\ln f(x)$  is defined and this set is a convex set (i.e. an interval). A continuous random variable (or a continuous distribution) with density  $f$  is *log-concave* if  $f$  is a log-concave function. See [1, 2, 5] for introductions to log-concavity.

Schudy and Sviridenko [18] proved:

**Lemma 7.2** [18] *Any log-concave random variable  $X$  with density  $f$  is moment bounded with parameter  $L = \frac{1}{\ln 2}\mathbb{E}[|X|] \approx 1.44\mathbb{E}[|X|]$ .*

If  $X$  is log-concave with density  $f$  then  $X - \mathbb{E}[X]$  clearly has density  $\tilde{f}(x) = f(x + \mathbb{E}[X])$ , which is evidently log-concave. Therefore:

**Corollary 7.3** *Any log-concave random variable  $X$  with density  $f$  is central moment bounded with parameter  $L = \frac{1}{\ln 2}\mathbb{E}[|X - \mathbb{E}[X]|] \approx 1.44\mathbb{E}[|X - \mathbb{E}[X]|] \leq 2.88\mathbb{E}[|X|]$ .*

### 7.3 Discrete log-concave random variables

A distribution over the integers  $\dots, p_{-2}, p_{-1}, p_0, p_1, p_2, \dots$  is said to be *log-concave* [1, 11] if  $p_{i+1}^2 \geq p_i p_{i+2}$  for all  $i$ . An integer-valued random variable  $X$  is log-concave if its distribution  $p_x = \Pr[X = x]$  is.

The discrete case is a bit trickier than the continuous case since  $X - \mathbb{E}[X]$  might take non-integer values even if  $X$  takes integer ones. We therefore can only get inspiration from the proof in [18] that discrete log-concave random variables are moment bounded rather than using it as we did in the continuous case.

**Lemma 7.4** *Let  $X$  be a log-concave integer-valued random variable with  $\Pr[X \geq \ell] = 1$  and  $\Pr[X = \ell] > 0$  for some  $\ell \in \mathbb{Z}$ . Let  $a \in \mathbb{R}$  be an arbitrary real such that  $a \leq \ell \leq a + 1$ . Then*

$$\mathbb{E}[|X - a|^k] \leq kL\mathbb{E}[|X - a|^{k-1}]$$

where  $L = 1 + \mathbb{E}[|X - a|]$ .

**Proof.** Let  $u$  be the largest index  $i$  such that  $p_i > 0$  or infinity if there is no such index. By log-concavity we have  $p_i > 0$  for all  $i \in \mathbb{Z}$  with  $\ell \leq i \leq u$ . Let  $r_i = \Pr[X \geq i]$ . Note that  $r_\ell = 1$

and when  $u$  is finite  $r_{i+1} = 0$  for all  $i \geq u$ . We bound

$$\begin{aligned}
\mathbb{E} \left[ |X - a|^k \right] &= \sum_{x=\ell}^{\infty} (x - a)^k (r_x - r_{x+1}) \\
&= (\ell - a)^k + \sum_{x=\ell+1}^{\infty} r_x [(x - a)^k - (x - 1 - a)^k] \\
&\leq (\ell - a)^k + \sum_{x=\ell+1}^{\infty} r_x k (x - a)^{k-1} \\
&\leq \sum_{x=\ell}^{\infty} r_x k (x - a)^{k-1} \\
&= \sum_{x=\ell}^u \frac{r_x}{p_x} k (x - a)^{k-1} p_x \\
&\leq \left( \sum_{x=\ell}^u \frac{r_x}{p_x} p_x \right) \left( \sum_{x=\ell}^u k (x - a)^{k-1} p_x \right) \\
&= \mathbb{E} [1 + |X - \ell|] \mathbb{E} \left[ k |X - a|^{k-1} \right] \\
&\leq (1 + \mathbb{E} [|X - a|]) k \mathbb{E} [|X - a|^{k-1}]
\end{aligned}$$

where the second inequality uses the fact that  $\ell - a \leq 1 \leq k$  and the third inequality follows from Chebychev's summation inequality, which applies because  $r_x/p_x$  is a non-increasing sequence (Proposition 10 in [1]) and  $k(x - a)^{k-1}$  is a non-decreasing sequence.  $\blacksquare$

**Lemma 7.5** *Any log-concave integer valued random variable  $X$  such that  $\Pr[X = \mathbb{E}[X]] < 1$  is central moment bounded with parameter*

$$L = 1 + \max(\mathbb{E} [|X - \mathbb{E}[X]| \mid X \geq \mathbb{E}[X]], \mathbb{E} [|X - \mathbb{E}[X]| \mid X < \mathbb{E}[X]]).$$

**Proof.** It follows that  $\Pr[X \geq \mathbb{E}[X]]$  and  $\Pr[X < \mathbb{E}[X]]$  are both strictly positive, hence by log-concavity  $\Pr[X = \lceil \mathbb{E}[X] \rceil]$  and  $\Pr[X = \lceil \mathbb{E}[X] \rceil - 1]$  are both strictly positive.

Write

$$\begin{aligned}
\mathbb{E} \left[ |X - \mathbb{E}[X]|^k \right] &= \Pr[X \geq \mathbb{E}[X]] \mathbb{E} \left[ (X - \mathbb{E}[X])^k \mid X \geq \mathbb{E}[X] \right] + \Pr[X < \mathbb{E}[X]] \mathbb{E} \left[ (\mathbb{E}[X] - X)^k \mid X < \mathbb{E}[X] \right] \\
&= \Pr[X \geq \mathbb{E}[X]] \mathbb{E} \left[ (X_+ - \mathbb{E}[X])^k \right] + \Pr[X < \mathbb{E}[X]] \mathbb{E} \left[ (X_- + \mathbb{E}[X])^k \right]
\end{aligned}$$

where  $X_+$  is a random variable with  $\Pr[X_+ = x] = \Pr[X = x \mid X \geq \mathbb{E}[X]]$  and  $X_-$  is a random variable with  $\Pr[X_- = x] = \Pr[X = -x \mid X < \mathbb{E}[X]]$  for integer  $x$ . Clearly  $X_+$  and  $X_-$  inherit log-concavity from  $X$ . We also have  $X_+ \geq \mathbb{E}[X]$  and  $X_- \geq -\mathbb{E}[X]$ .

We apply Lemma 7.4 twice, first to  $X_+$  with  $a = \mathbb{E}[X]$  and  $\ell = \lceil \mathbb{E}[X] \rceil$  and second to  $X_-$  with

$a = -\mathbb{E}[X]$  and  $\ell = 1 - \lceil \mathbb{E}[X] \rceil$ , yielding

$$\begin{aligned}
& \mathbb{E} \left[ |X - \mathbb{E}[X]|^k \right] \\
&= \Pr[X \geq \mathbb{E}[X]] \mathbb{E} \left[ (X_+ - \mathbb{E}[X])^k \right] + \Pr[X < \mathbb{E}[X]] \mathbb{E} \left[ (X_- + \mathbb{E}[X])^k \right] \\
&\leq \Pr[X \geq \mathbb{E}[X]] k(1 + \mathbb{E}[|X_+ - \mathbb{E}[X]|]) \mathbb{E} \left[ |X_+ - \mathbb{E}[X]|^{k-1} \right] + \\
&\quad + \Pr[X < \mathbb{E}[X]] k(1 + \mathbb{E}[|X_- + \mathbb{E}[X]|]) \mathbb{E} \left[ |X_- + \mathbb{E}[X]|^{k-1} \right] \\
&\leq k(1 + \max(\mathbb{E}[|X_+ - \mathbb{E}[X]|], \mathbb{E}[|X_- + \mathbb{E}[X]|])) \cdot \\
&\quad \cdot \left( \Pr[X \geq \mathbb{E}[X]] \mathbb{E} \left[ |X - \mathbb{E}[X]|^{k-1} | X \geq 0 \right] + \Pr[X < \mathbb{E}[X]] \mathbb{E} \left[ |X - \mathbb{E}[X]|^{k-1} | X < \mathbb{E}[X] \right] \right) \\
&= k(1 + \max(\mathbb{E}[|X - \mathbb{E}[X]| | X \geq \mathbb{E}[X]], \mathbb{E}[|X - \mathbb{E}[X]| | X < \mathbb{E}[X]])) \mathbb{E} \left[ |X - \mathbb{E}[X]|^{k-1} \right].
\end{aligned}$$

■

## References

- [1] M. An, Log-concave Probability Distributions: Theory and Statistical Testing, in Game Theory and Information EconWPA 9611002 (1996).
- [2] M. Bagnoli and T. Bergstrom, Log-concave probability and its applications, *Economic Theory* 26(2) (2005), pp. 445-469.
- [3] S.N. Bernstein, *Theory of Probability*, (Russian), Moscow, 1927.
- [4] S.N. Bernstein, On certain modifications of Chebyshev's inequality. *Doklady Akademii Nauk SSSR* 17 (6) (1937), pp. 275-277.
- [5] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.
- [6] C. Buchheim, A. Caprara and A. Lodi, An Effective Branch-and-Bound Algorithm for Convex Quadratic Integer Programming, in *Proceedings of IPCO 2010*, pp. 285-298.
- [7] M. Dudik, D. Hsu, S. Kale, N. Karampatziakis, J. Langford, L. Reyzin and T. Zhang, Efficient Optimal Learning for Contextual Bandits, in *Proceedings of UAI 2011*.
- [8] D. Freedman, On tail probabilities for martingales. *Ann. Probability* 3 (1975), pp. 100-118.
- [9] J. Jacod and P. Protter. *Probability Essentials*, Springer-Verlag (2004).
- [10] S. Janson, *Gaussian Hilbert spaces*. Cambridge Tracts in Mathematics. 129. Cambridge: Cambridge University Press, (1997).
- [11] N. Johnson and S. Kotz, *Discrete Distributions*, John Wiley, New York (1969).
- [12] J. Kim and V. Vu, Concentration of multivariate polynomials and its applications, *Combinatorica* 20 (2000), no. 3, 417-434.
- [13] R. Latała, Estimates of moments and tails of Gaussian chaoses, *Ann. Probab.* 34 (2006), pp. 2315-2331.



- [14] K. Makarychev, W. Schudy and M. Sviridenko, Concentration Inequalities for Nonlinear Matroid Intersection, to appear in SODA2012.
- [15] C. McDiarmid, Concentration, In Probabilistic Methods for Algorithmic Discrete Mathematics, M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed editors, pp. 195–248, Springer, 1998.
- [16] E. Mossel, R. O’Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics* 171(1), pp. 295-341 (2010).
- [17] R. Prabhakar Raghavan and C. Thompson, Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica* 7(4): 365-374 (1987).
- [18] W. Schudy and M. Sviridenko, Concentration and Moment Inequalities for Polynomials of Independent Random Variables, submitted for publication, extended abstract appeared in Proceedings of SODA2012.
- [19] J. V. Uspensky, Introduction to Mathematical Probability, McGraw-Hill Book Company, 1937.
- [20] V. Vu, Concentration of non-Lipschitz functions and applications, Probabilistic methods in combinatorial optimization. *Random Structures Algorithms* 20 (2002), no. 3, 262-316.
- [21] V. Vu, On the concentration of multivariate polynomials with small expectation, *Random Structures Algorithms* 16 (2000), no. 4, 344-363.