SMALL DEVIATIONS OF SUMS OF RANDOM VARIABLES

BY BRIAN GARNETT

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Mathematics

Written under the direction of

Swastik Kopparty

and approved by

New Brunswick, New Jersey

May, 2016

ABSTRACT OF THE DISSERTATION

Small Deviations of Sums of Random Variables

by Brian Garnett

Dissertation Director: Swastik Kopparty

In this thesis, we study the probability of a *small* deviation from the mean of a sum of independent or semi-independent random variables. In contrast with the rich history of large deviation inequalities, small deviations have only recently gained attention, and we make contributions to several problems on this topic.

Perhaps the most significant result in this field was an inequality proved by Feige [6]. Let $X_1, \ldots, X_n$ be nonnegative independent random variables, with $\mathbb{E}[X_i] \leq 1 \ \forall i$, and let $X = \sum_{i=1}^{n} X_i$. Then for any n,

$$\Pr[X < \mathbb{E}[X] + 1] \geq \alpha > 0,$$

for some $\alpha \geq 1/13$. This bound was later improved to 1/8 by He, Zhang, and Zhang [7]. Building off their work, we improve the bound to approximately .14. The conjectured true bound is $1/e \simeq .368$, so there is still (possibly) quite a gap left to fill.

We also consider whether or not such small deviation inequalities hold for k-wise independent random variables. We show that for some classes of random variables, 4-wise independence is sufficient for a constant lower bound of $\alpha = 1/6$, which we show to be tight. Furthermore, we present counterexamples showing that 3-wise independence is insufficient for a positive constant lower bound.

For sums of Bernoulli random variables, we can let $\alpha = 1/e$. We also show that k-wise independence can bring us arbitrarily close to that bound for large enough k.

Table of Contents

Chapter 1

Introduction

We will study the problem of bounding the probability that a sum of independent random variables deviates from the mean by some amount. We present several results on this topic, which are useful when the deviation is a small constant. In particular, we improve the best-known bound for Feige's Theorem [6]. We also investigate this question applied to k-wise independent random variables. In this context, we will see that 4-wise independence is much stronger than 3-wise.

1.1 Deviation Inequalities

For a real-valued random variable X, we often want to find an upper bound for

$$\Pr[X \geq \mathbb{E}[X] + \delta] \tag{1.1}$$

where $\delta \geq 0$. This is of course a well-studied problem. If X is nonnegative, then Markov's inequality gives

$$\Pr[X \geq \mathbb{E}[X] + \delta] \leq \frac{\mathbb{E}[X]}{\mathbb{E}[X] + \delta}.$$

If we know that $\mathrm{Var}[X] = \sigma^2$ (and X is not necessarily nonnegative), then Cantelli's–or one-sided Chebyshev's–inequality says that

$$\Pr[X \geq \mathbb{E}[X] + \delta] \leq \frac{\sigma^2}{\sigma^2 + \delta^2}.$$

We will consider the case where δ is a small constant, whereas the mean and variance of X are arbitrarily large. In various situations, we will want to find a constant upper bound for (1.1), which is less than 1. In this case, the classic inequalities above are clearly insufficient, so we will need more information than just the first and second moments.

If X is a sum of independent random variables, we have Chernoff-type bounds for (1.1), but again if we are after an upper bound away from 1, these are ineffective for small deviations. For this problem, Feige [6] made a remarkable discovery. Let $X_1, \ldots, X_n$ be nonnegative independent random variables, with $\mathbb{E}[X_i] \leq 1$ for each i. Let $X = \sum_{i=1}^n X_i$. There exists a constant $\alpha > 0$ such that

$$\Pr[X < \mathbb{E}[X] + 1] \geq \alpha. \tag{1.2}$$

Feige showed that $\alpha \geq 1/13$. In [7], He, Zhang, and Zhang improved the constant to $1/8$. However, it is believed that α can be improved to $1/e$. This bound would be tight, as consider letting all X_i have mean 1 and support $\{0, n+1\}$. Then

$$\Pr[X_1 + \ldots + X_n < n+1] = \left(1 - \frac{1}{n+1}\right)^n \longrightarrow \frac{1}{e}.$$

We point out that raising the deviation of $\delta = 1$ to a higher constant results in the same asymptotic bound in the conjectured tight example. However, as Feige pointed out, we cannot lower δ too much and hope for the same constant bound. Consider X_1 having mean 1 and support $\{0, 1+\delta\}$, and $X_i \equiv 1$ for $i \geq 2$. In this case,

$$\Pr[X_1 + \ldots + X_n < n + \delta] = \frac{1}{1+\delta}. \tag{1.3}$$

Related to Feige's conjecture that $\alpha = 1/e$ in (1.2) is the more general:

Conjecture 1.1.1 (Samuels)**.** *Let $X_1, \ldots, X_n$ be nonnegative independent random variables with $\mathbb{E}[X_i] = \mu_i$ for each i. Assume that $0 \leq \mu_1 \leq \ldots \leq \mu_n$. If $\sum_{i=1}^n \mu_i < 1$, then*

$$\Pr[X_1 \ldots + X_n < 1] \geq \min_{t=0,\ldots,n-1} \prod_{i=t+1}^{n} \left(1 - \frac{\mu_i}{1 - \sum_{j=1}^t \mu_j}\right). \tag{1.4}$$

Note that for a particular t,

$$\Pr[X_1 \ldots + X_n < 1] = \prod_{i=t+1}^{n} \left(1 - \frac{\mu_i}{1 - \sum_{j=1}^t \mu_j}\right),$$

when $X_i \equiv \mu_i$ for $i \leq t$ and X_i has support $\{0, 1 - \sum_{j=1}^t \mu_j\}$ and mean u_i for $i > t$. Samuels established that his conjecture is true for $n \leq 4$ ($n = 1$ is actually Markov's inequality, and $n = 2$ had been shown previously). This implies that for $n \leq 4$, in (1.2) we can let $\alpha = 1/e$. Further work in [12] also immediately implies that if in (1.2), X is a sum of independent Bernoulli random variables, then $\alpha = 1/e$ (we explain in Section 4.1).

1.2 Our Results

In Chapter 3, we improve the constant in Feige's Theorem from the current .125 to .14.

Theorem 1.2.1. *Let $X_1, \ldots, X_n$ be nonnegative independent random variables, with $\mathbb{E}[X_i] \leq 1$ for each i. Let $X = \sum_{i=1}^{n} X_i$. Then*

$$\Pr\left[X < \mathbb{E}[X] + 1\right] \geq \frac{7}{50}. \tag{1.5}$$

In the previous improvement, He, Zhang, and Zhang [7] applied bounds they had developed for the more general (1.1) in terms of the first, second, and fourth moments. The source of our improvement comes from also considering the central third moment, and what happens in the cases where it is positive versus negative. This idea is best illustrated by a (tight) moment bound we prove in Section 2.1:

Theorem 1.2.2. *Let X be a random variable with $\mathbb{E}[X] = 0, \mathbb{E}[X^2] = \sigma^2$, and $\mathbb{E}[X^3] \geq 0$. If $\mathbb{E}[X^4] \leq c\sigma^4$, then*

$$\Pr[X \geq 0] \leq 1 - \frac{1}{2c}.$$

The assumption on the third moment allows for a slightly smaller bound than the one proved in [8], which made no mention of the third moment (but otherwise had an identical hypothesis).

We also consider whether we can obtain similar small deviation bounds if the random variables are only k-wise independent for some $k \geq 2$. Recall that a collection of random variables is k-wise independent if any k-sized subcollection is mutually independent. This is a natural consideration, since calculating up to the kth moment of a sum of independent random variables in fact only uses the assumption that they are k-wise independent. In addition, k-wise independent random variables are of particular interest in the field of computer science. For many randomized algorithms, k-wise independence is just as adequate as full independence, and the benefit of using the former is that it requires much less randomness to generate. In this realm, we show that for certain types of random variables, 4-wise independence is sufficient for a nontrivial small deviation bound. Our most general result of this type, which we prove in Chapter 2, is

Theorem 1.2.3. *Let $X_1, \ldots, X_n$ be a 4-wise independent collection of random variables where for each i, $\mathbb{E}[X_i] = 0$, and $|X_i| \leq 1$. Let $X = \sum_{i=1}^n X_i$. Then if $\delta \geq 1/3$,*

$$\Pr[X < \delta] \geq \frac{1}{6}.$$

In Section 4.4, we show that 1/6 is the best possible constant bound for this and other related theorems. Similar to the conjectured $\alpha = 1/e$ in (1.2), the bound 1/6 cannot be improved by raising δ to a higher constant, but in this case the bound does not hold when $\delta < 1/5$, due to the same example that produces (1.3). Thus, there may be some slight room for improvement to the above theorem, but not much. As we will see in our approach to Theorem 1.2.1, letting δ be as small as possible is a worthwhile endeavor. As an obvious corollary to Theorem 1.2.3,

Corollary 1.2.4. *Let $X_1, \ldots, X_n$ be a collection of 4-wise independent Bernoulli random variables with respective marginal probabilities $p_1, \ldots, p_n \in [0, 1]$. Let $X = \sum_{i=1}^n X_i$. Then*

$$\Pr[X < \mathbb{E}[X] + 1/3] \geq \frac{1}{6}.$$

If the Bernoulli random variables have marginal probabilities at most 1/2, we can drop the deviation to $\delta = 0$, while keeping the same bound. We prove this in Section 4.2.

Theorem 1.2.5. *Let $X_1, \ldots, X_n$ be a collection of 4-wise independent Bernoulli random variables, with respective marginal probabilities $p_1, \ldots, p_n \in (0, 1/2]$. Let $X = \sum_{i=1}^n X_i$. Then*

$$\Pr\left[X < \mathbb{E}[X]\right] \geq \frac{1}{6}. \tag{1.6}$$

In Section 4.5, we present a counterexample to show that 3-wise independence is insufficient for any nontrivial small deviation bound on a sum of random variables. This settles a question in [7], regarding whether or not a nontrivial bound can be obtained from only the first, second, and third moments. In addition, the assumption of pairwise independence does not lead to an improvement on Markov's inequality for a deviation bound on a sum of nonnegative random variables.

Theorem 1.2.6. *Let $\delta > 0$. If $(n+\delta)/(\delta+1) \in \mathbb{Z}$, then there exists a collection of nonnegative pairwise independent random variables $X_1, \ldots, X_n$, each with mean 1 such that*

$$\Pr[X_1 + \ldots + X_n < n + \delta] = \frac{\delta}{n+\delta}$$

Theorem 1.2.7. *Let $\delta > 0$. If $(n+\delta)/(\delta+2) \in \mathbb{Z}$, then there exists a collection of nonnegative 3-wise independent random variables $X_1, \ldots, X_n$, each with mean 1 such that*

$$\Pr[X_1 + \ldots + X_n < n + \delta] = \frac{(\delta+1)^2}{(\delta+2)(n+\delta)}.$$

At the end of Chapter 4, we show that for a sum of Bernoulli random variables, if the marginal probabilities are bounded away from 1, we can get a bound ε-close to $1/e$ with K-wise independence for large enough K. This relies on the bound of $1/e$ already established for mutually independent Bernoullis.

Theorem 1.2.8. *Let $0 < \varepsilon \leq 1/16$ and $0 < a < 1$. There exists an integer $K := K(\varepsilon, a)$ such that if $X_1, \ldots, X_n$ is a K-wise independent collection of Bernoulli random variables with respective means $p_1, \ldots, p_n$, where $p_i \leq 1 - a$ for all i, then*

$$\Pr\left[X_1 + \ldots + X_n < \sum_{i=1}^{n} p_i + 1\right] \geq \frac{1}{e} - \varepsilon.$$

1.3 Other Related Work

In [3], Berger showed that 4-wise independence is sufficient–and 3-wise independence is insufficient–for a good lower bound on $\mathbb{E}[|X_1 + \ldots + X_n|]$, where each $X_i = \pm a_i$ for $a_i > 0$, and $\mathbb{E}[X_i] = 0$. Thus, for more than one reason, 4-wise independence is much stronger than 3-wise when dealing with a sum of random variables.

In a recent paper [11], Peled, Yadin, and Yehudayoff studied k-wise independent Bernoulli random variables with identical marginal probability p, and bounded the probability that all random variables are equal to 1. They mention the existence of a

3-wise independent distribution where this probability is $\omega(\frac{1}{n})$, to be presented in an upcoming paper. Although it was discovered with a different goal in mind, the example given in this paper also happens to have that property. We acknowledge that their work inspired our approach to proving the results on 4-wise independence.

Our proof of Theorem 1.2.8 is modeled off of the work in [5], wherein they showed that bounded independence is sufficient to fool linear threshold functions with random unbiased inputs. The difference is we look at a particular threshold function, and the marginal probabilities may be biased and differ from each other.

Chapter 2

Sums of 4-wise Independent Random Variables

2.1 A Moment Problem and Setup

Many of our results and their proofs will be versions of a standard moment problem. Let X be a real-valued random variable. Given information of the moments of X up to some k, we want to bound the probability that X lies in a set S. This is a well-studied optimization problem that gives rise to an elegant dual problem, first utilized in [9] and [10], and treated extensively in [4]. The setup of the general problem is

$$\begin{aligned} \underset{X}{\text{maximize}} \quad & \Pr[X \in S] \\ \text{subject to} \quad & \mathbb{E}[X^i] = M_i,\ 0 \le i \le k. \end{aligned}$$

Of course, $M_0 = 1$ always. The dual problem is then

$$\begin{aligned} \underset{y}{\text{minimize}} \quad & \sum_{i=0}^{k} y_i M_i \\ \text{subject to} \quad & \sum_{i=0}^{k} y_i x^i \ge \mathbf{1}_{\{x \in S\}} \end{aligned}$$

In other words, this minimizes $\mathbb{E}[Q(X)]$ over all polynomials Q of degree up to k, where $Q \ge \mathbf{1}_S$.

For most of this paper, we let $k = 4$ and $S = \{x : x \ge \mathbb{E}[X] + \delta\}$. Without loss of generality, assume $M_1 = \mathbb{E}[X] = 0$. Thus for the dual problem, we need a polynomial Q of degree at most 4 such that $Q(x) \ge \mathbf{1}_{\{x \ge \delta\}}(x)$ for all x. The polynomial, which we will denote $Q_{\ell,r}$ for $\ell, r > 0$, we use throughout the paper will be uniquely determined by the following properties:

- $Q_{\ell,r}(x)$ has a double root at $x = -\ell$.

- $Q_{\ell,r}(0) = 1$ (we will often just shift by the small value δ when needed).

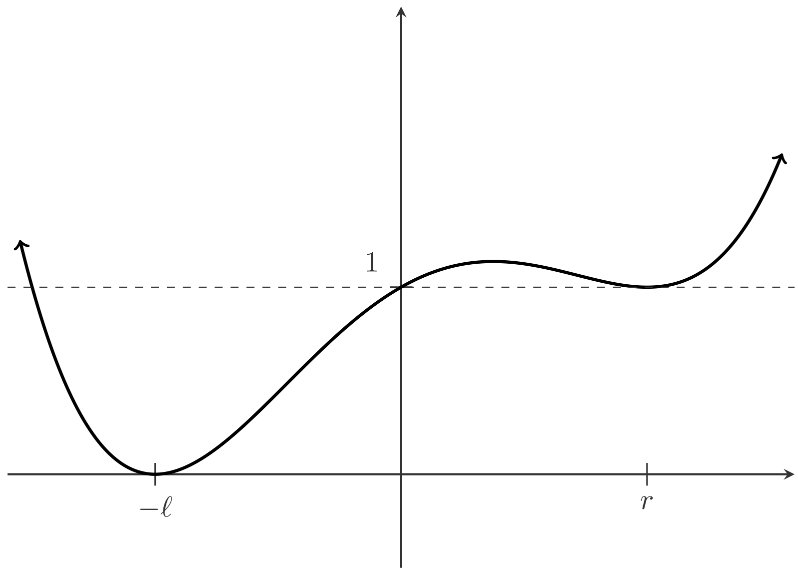

Figure 2.1: $Q_{\ell,r}$

- $Q_{\ell,r}(x) - 1$ has a double root at $x = r$.

Most often, $r = \ell$, in which case we will denote it as Q_r. In that case,

$$Q_r(x) = 1 + \frac{3}{4r}x - \frac{1}{r^2}x^2 - \frac{1}{4r^3}x^3 + \frac{1}{2r^4}x^4. \tag{2.1}$$

We first use this approach to prove Theorem 1.2.2, restated here:

Theorem 2.1.1. *Let X be a random variable with $\mathbb{E}[X] = 0, \mathbb{E}[X^2] = \sigma^2$, and $\mathbb{E}[X^3] \geq 0$. If $\mathbb{E}[X^4] \leq c\sigma^4$, then*

$$\Pr[X \geq 0] \leq 1 - \frac{1}{2c}.$$

Proof. Consider the polynomial

$$Q(x) = Q_{\sqrt{c}\sigma}(x) = 1 + \frac{3}{4\sqrt{c}\sigma}x - \frac{1}{c\sigma^2}x^2 - \frac{1}{4c^{3/2}\sigma^3}x^3 + \frac{1}{2c^2\sigma^4}x^4,$$

which satisfies $Q(x) \geq 1_{\{x\geq 0\}}$ for all x (we prove this for the more general expression of $Q_{\ell,r}$ in the next section). Using the assumptions on the moments, we have

$$\Pr[X \geq 0] = \mathbb{E}[1_{X\geq 0}] \leq \mathbb{E}[Q(X)] \leq 1 - \frac{1}{c\sigma^2}\sigma^2 + \frac{1}{2c^2\sigma^4}c\sigma^4 = 1 - \frac{1}{2c}.$$

□

Note that the bound is tight if we consider, for any $a > 0$ and $p < 1/2$,

$$X = \begin{cases} -a, & \text{with probability } p \\ 0, & \text{with probability } 1-2p \\ a, & \text{with probability } p \end{cases} \tag{2.2}$$

This happens to also be a tight example to Chebyshev's inequality. Without any assumption on the third moment, He et al proved an upper bound of $1 - (2\sqrt{3} - 3)/c$ [8]. Using our $Q_{\ell,r}$, we can choose $\ell = (1 + \sqrt{3})r/2$ (which makes the degree-3 coefficient 0) and optimize over r, to get the same bound.

Many of our proofs will be of the same flavor as Theorem 1.2.2, with $c = 3$ (which, not coincidentally, is the kurtosis of the normal distribution). However, two complications will often arise, which one can predict by examining the idealistic conditions of the previous theorem. Namely, the third moment could be negative, and the fourth moment may be a bit larger than $c\sigma^4$ for the optimal c we are after. Consider, for example, a sum of bounded independent random variables.

Let $\{X_i\}_{i \leq 1 \leq n}$ be independent random variables with $\mathbb{E}[X_i] = 0$ for each i. Let $X = \sum_{i=1}^n X_i$. If $|X_i| \leq 1$ for each i, then

$$\begin{aligned} \left|\mathbb{E}[X^3]\right| &= \left|\sum_{i=1}^n \mathbb{E}[X_i^3]\right| \\ &\leq \sum_{i=1}^n \mathbb{E}[|X_i|^3] \\ &\leq \sum_{i=1}^n \mathbb{E}[X_i^2] \\ &= \mathbb{E}[X^2]. \end{aligned}$$

In addition,

$$\begin{aligned}
\mathbb{E}[X^4] &= \sum_{i=1}^{n} \mathbb{E}[X_i^4] + 6\sum_{i<j} \mathbb{E}[X_i^2]\,\mathbb{E}[X_j^2] \\
&= 3\left(\sum_{i=1}^{n} \mathbb{E}[X_i^2]\right)^2 + \sum_{i=1}^{n} \left(\mathbb{E}[X_i^4] - 3\,\mathbb{E}[X_i^2]^2\right) \\
&= 3\,\mathbb{E}[X^2]^2 + \sum_{i=1}^{n} \left(\mathbb{E}[X_i^4] - 3\,\mathbb{E}[X_i^2]^2\right) \\
&\le 3\,\mathbb{E}[X^2]^2 + \sum_{i=1}^{n} \mathbb{E}[X_i^2] \\
&= 3\,\mathbb{E}[X^2]^2 + \mathbb{E}[X^2].
\end{aligned}$$

So we see in this case that even if $\mathbb{E}[X^3]$ is negative, it can only be as low as $-\mathbb{E}[X^2]$, and $\mathbb{E}[X^4]$ can only exceed $3\,\mathbb{E}[X^2]^2$ by as much as $\mathbb{E}[X^2]$. This will not present much of a problem asymptotically when the variance is large, but it will cause issues for small variances. In that case, we just modify the polynomial. However, in general, we cannot achieve the constant upper bound of $5/6$ unless we allow some deviation $\delta > 0$.

Note that if $\{X_i\}_{i\le 1\le n}$ are only 4-wise independent, then X will have the same moments above. Since we will only use the first four moments of X to prove Theorem 1.2.3 and the related Lemmas in Chapter 3, we can assume the random variables are only 4-wise independent. In each situation, we will use this information on the moments to show there exist $\ell, r > 0$ such that

$$\mathbb{E}[Q_{\ell,r}(X-\delta)] \le \frac{5}{6}. \tag{2.3}$$

Therefore,

$$\Pr[X \ge \delta] = \mathbb{E}[\mathbf{1}_{\{x\ge\delta\}}(X)] \le \mathbb{E}[Q_{\ell,r}(X-\delta)] \le \frac{5}{6}, \tag{2.4}$$

where the first inequality is shown in the next section.

2.2 $Q_{\ell,r}$

The polynomial $Q_{\ell,r}$ described in the previous section is explicitly given as

$$Q_{\ell,r}(x) = \sum_{i=0}^{n} q_i x^i, \tag{2.5}$$

where

$$\begin{aligned} q_0 &= 1, \\ q_1 &= \frac{2r^2(2\ell + r)}{\ell(\ell + r)^3}, \\ q_2 &= \frac{r(-8\ell^2 - \ell r + r^2)}{\ell^2(\ell + r)^3}, \\ q_3 &= \frac{4\ell^2 - 4\ell r - 2r^2}{\ell^2(\ell + r)^3}, \\ q_4 &= \frac{3\ell + r}{\ell^2(\ell + r)^3}. \end{aligned} \tag{2.6}$$

If $\ell = r$, then these coefficients simplify to

$$q_0 = 1,\ q_1 = \frac{3}{4r},\ q_2 = -\frac{1}{r^2},\ q_3 = -\frac{1}{4r^3},\ q_4 = \frac{1}{2r^4}. \tag{2.7}$$

We will show directly that this polynomial satisfies

Lemma 2.2.1. *Let $\ell, r > 0$. For all $x \in \mathbb{R}$, $Q_{\ell,r}(x) \geq 1_{\{x \geq 0\}}$.*

Proof.

$$Q_{\ell,r}(x) = \frac{1}{\ell^2(\ell + r)^3}(\ell + x)^2\big((\ell + r)^3 - 2(\ell^2 + 3\ell r + r^2)x + (3\ell + r)x^2\big),$$

which is zero if $x = -\ell$. Otherwise, since

$$\begin{aligned} \frac{\ell^2(\ell + r)^3 Q_{\ell,r}(x)}{(\ell + x)^2} &= (\ell + r)^3 - 2(\ell^2 + 3\ell r + r^2)x + (3\ell + r)x^2 \\ &\geq (\ell + r)^3 - 2(\ell^2 + 3\ell r + r^2)\left(\frac{\ell^2 + 3\ell r + r^2}{3\ell + r}\right) + (3\ell + r)\left(\frac{\ell^2 + 3\ell r + r^2}{3\ell + r}\right)^2 \\ &= \frac{(3\ell + r)(\ell + r)^3 - (\ell^2 + 3\ell r + r^2)^2}{3\ell + r} \\ &= \frac{2\ell^4 + 4\ell^3 r + \ell^2 r^2}{3\ell + r} \geq 0, \end{aligned}$$

we have $Q_{\ell,r}(x) \geq 0\ \forall x$. On the other hand,

$$\begin{aligned} Q_{\ell,r}(x) - 1 &= \frac{1}{\ell^2(\ell + r)^3}x\big(4\ell^2 + 2\ell r + (3\ell + r)x\big)(-r + x)^2 \\ &\geq 0,\ \text{for all } x \geq 0. \end{aligned}$$

□

Now, let $\mathbb{E}[X]=0$, and $\delta>0$. Then

$$\begin{aligned}\mathbb{E}[Q_{\ell,r}(X-\delta)] &= \sum_{i=0}^{4} q_i\,\mathbb{E}[(X-\delta)^i] \\ &= \sum_{i=0}^{4} q_i\,\mathbb{E}[(X)^i] + \sum_{i=1}^{4}(-1)^i\delta^i q_i + (6\delta^2 q_4 - 3\delta q_3)\,\mathbb{E}[X^2] - 4\delta q_4\,\mathbb{E}[X^3] \\ &= \sum_{i=0}^{4}(-1)^i\delta^i q_i + (q_2 - 3\delta q_3 + 6\delta^2 q_4)\,\mathbb{E}[X^2] + (q_3 - 4\delta q_4)\,\mathbb{E}[X^3] + q_4\,\mathbb{E}[X^4].\end{aligned} \tag{2.8}$$

Looking at the coefficients in (2.6), notice that $q_4>0$ always, and if $\ell \le (1+\sqrt{3})r/2$, then $q_3<0$. In fact, we will always choose $\ell \le r$. Therefore, we will always have

$$q_3 < 0, \text{ and } q_4 > 0. \tag{2.9}$$

Thus, if X satisfies the inequalities (2.13) below, then

$$\mathbb{E}[Q_{\ell,r}(X-\delta)] \le \sum_{i=0}^{4}(-\delta)^i q_i + (q_2 - 3\delta q_3 + 6\delta^2 q_4)\sigma^2 + (q_3 - 4\delta q_4)(-\sigma^2) + q_4(3\sigma^4+\sigma^2). \tag{2.10}$$

We will often let $\ell = r = \sqrt{3}\sigma$. In that case, (2.7) becomes

$$q_0 = 1,\ q_1 = \frac{3}{4\sqrt{3}\sigma},\ q_2 = -\frac{1}{3\sigma^2},\ q_3 = -\frac{1}{12\sqrt{3}\sigma^3},\ q_4 = \frac{1}{18\sigma^4}. \tag{2.11}$$

Plugging these into (2.10) and simplifying yields

$$\mathbb{E}[Q_{\sqrt{3}\sigma}(X-\delta)] \le \frac{5}{6} + \frac{2\delta^4 + \sqrt{3}\delta^3\sigma + (2+8\delta)\sigma^2 + (\sqrt{3} - 6\sqrt{3}\delta)\sigma^3}{36\sigma^4}. \tag{2.12}$$

2.3 Proof of Theorem 1.2.3

We restate it here in an equivalent form.

Theorem 2.3.1. *Let $X_1,\ldots,X_n$ be a 4-wise independent collection of random variables where for each i, $\mathbb{E}[X_i]=0$, and $|X_i|\le 1$. Let $X=\sum_{i=1}^n X_i$. Then*

$$\Pr[X \ge 1/3] \le \frac{5}{6}.$$

Proof. Let $X_1,\ldots,X_n$ be 4-wise independent random variables with $\mathbb{E}[X_i]=0$ and $|X_i|\le 1$ for each i. Let $X=\sum_{i=1}^n X_i$ (so that $\mathbb{E}[X]=0$), and let $\sigma^2=\mathbb{E}[X^2]$. At the end of Section 2.1, we showed that since $|X_i| \leq 1$ for each i,

$$\mathbb{E}[X^3] \geq -\sigma^2,$$
$$\mathbb{E}[X^4] \leq 3\sigma^4 + \sigma^2. \tag{2.13}$$

As explained in the same section, it is sufficient to show that for any such X, there is a choice of ℓ and r such that

$$\mathbb{E}[Q_{\ell,r}(X-\delta)] \leq \frac{5}{6}.$$

For this proof, we can let $\ell = r$ for each case, so we refer to the polynomial as Q_r. First, let $\ell = r = \sqrt{3}\sigma$. Referencing (2.12) with $\delta = 1/3$, we have

$$\mathbb{E}[Q_{\sqrt{3}\sigma}(X-1/3)] \leq \frac{5}{6} + \frac{1}{36\sigma^4}\left(\frac{2}{81} + \frac{\sqrt{3}}{27}\sigma + \frac{14}{3}\sigma^2 - \sqrt{3}\sigma^3\right).$$

If $\sigma \geq 3$, then

$$\begin{aligned}
\mathbb{E}[Q_{\sqrt{3}\sigma}(X-1/3)] &\leq \frac{5}{6} + \frac{1}{36\sigma}\left(\frac{2}{81}\sigma^{-3} + \frac{\sqrt{3}}{27}\sigma^{-2} + \frac{14}{3}\sigma^{-1} - \sqrt{3}\right) \\
&\leq \frac{5}{6} + \frac{1}{36\sigma}\left(\frac{2}{81}3^{-3} + \frac{\sqrt{3}}{27}3^{-2} + \frac{14}{3}3^{-1} - \sqrt{3}\right) \\
&\leq \frac{5}{6}.
\end{aligned}$$

Now If we let $r = a\sigma$ for a constant $a > 0$, and $\delta = 1/3$, plugging the coefficients of Q (2.7) into (2.10) yields

$$\mathbb{E}[Q_{a\sigma}(X-1/3)] \leq \frac{3-2a^2+2a^4}{2a^4} + \frac{2-a^2}{4a^3}\sigma^{-1} + \frac{27-2a^2}{18a^4}\sigma^{-2} + \frac{1}{108a^3}\sigma^{-3} + \frac{1}{162a^4}\sigma^{-4}.$$

Let $B_a(\sigma)$ be the quantity on the righthand side. Examining the coefficients, we see that if $27 - 2a^2 \geq 0$, then B_a is a convex polynomial in the variable σ^{-1}. Thus, for a fixed a, and $\sigma_1 < \sigma_2$, if we show that $B_a(\sigma_1)$ and $B_a(\sigma_2)$ are both bounded above by $5/6$, then $\mathbb{E}[Q_{a\sigma}(X-1/3)] \leq 5/6$ for all $\sigma \in [\sigma_1, \sigma_2]$.

First, let $a = 2$. Then

$$B_2(\sigma) = \frac{27}{32} - \frac{1}{16}\sigma^{-1} + \frac{19}{288}\sigma^{-2} + \frac{1}{864}\sigma^{-3} + \frac{1}{2592}\sigma^{-4}.$$

and it can be easily checked that $B_2(3/2) < 5/6$ and $B_2(3) < 5/6$.

If $a = 9/4$, then

$$B_{9/4}(\sigma) = \frac{1883}{2187} - \frac{49}{729}\sigma^{-1} + \frac{80}{2187}\sigma^{-2} + \frac{16}{19683}\sigma^{-3} + \frac{128}{531441}\sigma^{-4},$$

with $B_{9/4}(1) < 5/6$ and $B_{9/4}(3/2) < 5/6$.

If $a = 5/2$, then

$$B_{5/2}(\sigma) = \frac{549}{625} - \frac{17}{250}\sigma^{-1} + \frac{116}{5625}\sigma^{-2} + \frac{2}{3375}\sigma^{-3} + \frac{8}{50625}\sigma^{-4},$$

with $B_{5/2}(1/2) < 5/6$ and $B_{5/2}(1) < 5/6$.

Thus, we have covered all $\sigma \geq 1/2$. Lastly, we set $r = 3/2$, for which plugging (2.7) into (2.10) gives

$$\mathbb{E}[Q_{3/2}(X - 1/3)] \leq \frac{10339}{13122} + \frac{8}{27}\sigma^4 \leq \frac{5}{6},$$

when $\sigma < 1/2$. □

As we discussed in the introduction, the deviation $\delta = 1/3$ could possibly be lowered, but not to anything below 1/5. However, due to the small variance case, our approach cannot allow for a δ much lower than the one we set.

Chapter 3

Sums of Fully Independent Random Variables

In this chapter, we prove Theorem 1.2.1. First, we will need two modified versions of Theorem 1.2.3. Although we will have full independence when we apply these lemmas, we only assume 4-wise independence for maximal generality. We treat separately the cases of positive and negative third moment.

3.1 Lemmas

Due to the third-degree coefficient q_3 of our polynomial being negative, if we know the central third moment is positive, we can lower the allowed deviation δ from $1/3$, while keeping the same upper bound of $5/6$ on the probability. It will be important to lower δ as much as possible, without having to raise the bound on the probability (which would not be a good tradeoff).

Lemma 3.1.1. *Let $X_1, \ldots, X_n$ be a 4-wise independent collection of random variables where for each i, $\mathbb{E}[X_i] = 0$, and $|X_i| \leq 1$. Let $X = \sum_{i=1}^{n} X_i$. If $\mathbb{E}[X^3] \geq 0$, then*

$$\Pr[X \geq 4/25] \leq \frac{5}{6}.$$

For the next lemma, we will assume each random variable is supported on two points; this will be the case when we apply it in the upcoming proof. Now, if we assume the central third moment of the sum is nonpositive and add one small condition, we can remove the assumption of a universal upper bound (intuitively, a negative central third moment implies the distributions of the random variables are skewed below their means). This will also be a crucial component to the proof of the theorem.

Lemma 3.1.2. *Let $X_1, \ldots, X_n$ be a 4-wise independent collection of random variables where for each i, $\mathbb{E}[X_i] = 0$, and X_i has support $\{-a_i, b_i\}$. Assume that $a_i \leq 1$ for each*

i, $b_1 = \max_i\{b_i\}$, and $a_1 \geq 1/16$. Let $X = \sum_{i=1}^n X_i$. If $\mathbb{E}[X^3] \leq 0$, then

$$\Pr[X < 1] \geq \frac{1}{6}.$$

The allowed deviation of 1 and the 1/16 assumption above can be tinkered with, but we fixed $\delta = 1$ in preparation for the theorem.

3.2 Proof of Theorem 1.2.1

We state it again, this time with a slightly better but also less nice-looking constant:

Theorem 3.2.1. *Let $X_1, \ldots, X_n$ be nonnegative independent random variables with means $\mu_1, \ldots, \mu_n$ such that $\mu_i \leq 1$ for every i. Then*

$$\Pr\left[\sum_{i=1}^n X_i < \sum_{i=1}^n \mu_i + 1\right] \geq \beta, \tag{3.1}$$

where we set $\beta = \frac{46}{279}e^{-4/25}$ $(> \frac{7}{50})$.

In his proof [6] which first established a lower bound on this probability, Feige explained via a linear programming argument that without loss of generality, we may assume that each X_i is non-constant and has support of size two. This was one aspect of his overall strategy, which was to apply a sequence of transformations to the collection of random variables, where each transformation does not increase the probability that we wish to lower bound. The next step is to simply subtract some nonnegative amount from each X_i, so that it has support $\{0, c_i\}$ for some $c_i > 0$. This step may reduce the mean μ_i but leaves the probability in (3.1) unchanged.

The goal of the next transformation, which he called "merge," was to make the means closer to one another. With "merge," we take the two random variables with the smallest means, say X_i and X_j with means μ_i and μ_j, and merge them into the random variable $X' = X_i + X_j$ with mean $\mu' = \mu_i + \mu_j$. Now X' possibly has support of size up to 4, but as before, we may reduce its size to two and align it with 0. For some threshold $t \leq 1/2$, we will apply "merge" (followed by reducing the support and aligning with 0) on the two random variables with smallest means, $\mu_i < \mu_j$, if and only if $\mu_i < t$ and $\mu_j \leq 1 - t$. Thus, we will never create a random variable with a mean larger than 1. Furthermore, when we have finished these transformations, we have at most one random variable with mean below t, in which case all other means are above $1-t$.

Proof. As explained in the precursor to this proof, we may assume that each X_i has support $\{0, c_i\}$ for some $c_i > 0$, so that $\Pr[X_i = c_i] = \mu_i/c_i$. For each i, let $s_i = c_i - \mu_i$, the "surplus" to the mean. We may assume

$$s_1 \geq \ldots \geq s_n.$$

Using a trick from [7], fix $\tau > 0$, and define

$$N = \max\left\{0, \max_{1\leq k\leq n}\{k : s_k \geq \tau(\mu_1 + \ldots \mu_k)\}\right\}.$$

Let $m = \sum_{i=1}^{N} \mu_i$, the mean of the sum of the first N. If $i > N$, then

$$s_i \leq s_{N+1} \leq \tau \sum_{i=1}^{N+1} \mu_i \leq \tau(m + \mu_{N+1}) \leq \tau(m+1). \tag{3.2}$$

Otherwise, if $i \leq N$, $s_i \geq s_N \geq \tau m$. If $N > 0$, then

$$\begin{aligned}
\Pr\left[\sum_{i=1}^{N} X_i = 0\right] &= \prod_{i=1}^{N} \Pr[X_i = 0] \\
&= \prod_{i=1}^{N}\left(1 - \frac{\mu_i}{c_i}\right) \\
&= \prod_{i=1}^{N}\left(1 - \frac{\mu_i}{s_i + \mu_i}\right) \\
&\geq \prod_{i=1}^{N}\left(1 - \frac{\mu_i}{\tau m + \mu_i}\right) \\
&\geq \prod_{i=1}^{N} e^{-\mu_i/(\tau m)} = e^{-1/\tau}.
\end{aligned}$$

The utility of this splitting of the random variables is that conditioning on the sum of first N (which has a Poisson-like distribution with low mean) being 0, the rest are bounded by an amount comparable to the allowed deviation. Here in particular, we are using full (as opposed to just k-wise) independence of the random variables (we also implicitly used full independence during the merge operation described above).

$$\begin{aligned}\Pr\left[\sum_{i=1}^{n} X_i < \sum_{i=1}^{n}\mu_i + 1\right] &\geq \Pr\left[\sum_{i=1}^{N} X_i = 0\right]\cdot \Pr\left[\sum_{i=N+1}^{n} X_i < \sum_{i=1}^{n}\mu_i + 1\right]\\ &\geq e^{-1/\tau}\Pr\left[\sum_{i=N+1}^{n} X_i < \sum_{i=1}^{n}\mu_i + 1\right]\\ &= e^{-1/\tau}\Pr\left[\sum_{i=N+1}^{n} X_i < \sum_{i=N+1}^{n}\mu_i + (m+1)\right],\end{aligned}$$

and we will now focus on the latter probability. We fix $\tau = 25/4$. Assume $N < n$, otherwise we are done. For $1 \leq j \leq n-N$, let $Y_j = X_{N+j} - \mu_{N+j}$, and let $n' = n - N$. Set $Y = \sum_{j=1}^{n'} Y_j$. Each Y_j has mean 0 and support $\{-a_j, b_j\}$ where $0 < a_j \leq 1$ and $0 < b_j \leq 25(m+1)/4$. We break the analysis into two cases, depending on the sign of the third moment of Y.

Case 1: $\mathbb{E}[Y^3] \geq 0$.

In this case, for each j, let $Y_j' = \dfrac{4}{25(m+1)} Y_j$, and $Y' = \sum_{i=1}^{n'} Y_j'$. Note that $\mathbb{E}[(Y')^3] \geq 0$, and for each j, $|Y_j'| \leq 1$. By Lemma 3.1.1,

$$\Pr\left[\sum_{j=1}^{n'} Y_j < (m+1)\right] = \Pr\left[\sum_{j=1}^{n'} Y_j' < 4/25\right] \geq \frac{1}{6}.$$

Thus, we have

$$\Pr\left[\sum_{i=1}^{n} X_i < \sum_{i=1}^{n}\mu_i + 1\right] \geq \frac{e^{-4/25}}{6} > \beta.$$

Although proving this case was immediate, it required the bounding of the latter random variables and drove the choice of $\tau = 25/4$.

Case 2: $\mathbb{E}[Y^3] < 0$.

The major fact about Lemma 3.1.2 we use in this case is that we do not need an upper bound on the Y_j's. Above we had to divide the random variables by some amount in order to apply our positive third moment lemma, which lowered the allowed deviation in our strict application of the statement. This time, we do not have to do so, and the allowed deviation δ remains at least 1.

Now, each Y_j has support $\{-a_j, b_j\}$, where $0 < a_j \leq 1$ for each i. Since $b_j = s_{N+j}$, we have $b_1 \geq \ldots \geq b_{n'}$. If $a_1 \geq 1/16$, we can immediately apply Lemma 3.1.2, and we are done. So we can assume $a_1 < 1/16$. Since $Y_j = X_{N+j} - \mu_{N+j}$, each $a_j = \mu_{N+j}$.

Thus, $\mu_{N+1} < 1/16$. By the stopping condition of the merge process, this means that all other means exceed 15/16.

We may also assume at this point that $b_1 \geq 3(m+1)$. Otherwise, like in Case 1, we can divide by $3(m+1)$, and by Theorem 1.2.3,

$$\Pr\left[\sum_{j=1}^{n'} Y_j < (m+1)\right] = \Pr\left[\sum_{j=1}^{n'} Y'_j < 1/3\right] \geq \frac{1}{6}.$$

Thus, considering

$$\Pr[Y_1 = b_1] = \frac{a_1}{a_1 + b_1} \leq \frac{1}{48(m+1)},$$

this variable being positive is quite unlikely, and in order to discard it, we will also condition on this not occurring. Once we do so, we must take note that the third moment of the remaining sum is also negative, as we have subtracted from it

$$E[Y_1^3] = a_1 b_1 (b_1 - a_1) > a_1 b_1 (3 - 1/16) > 0.$$

Furthermore, for $j \geq 2$, $a_j > 1/16$ (in fact $a_j \geq 15/16$), so we can apply Lemma 3.1.2 to the remaining sum. Now we consider two cases: $N = 0$ and $N \geq 1$.

If $N = 0$, then $m = 0$, and

$$\begin{aligned}
\Pr\left[\sum_{i=1}^{n} X_i < \sum_{i=1}^{n} \mu_i + 1\right] &= \Pr\left[\sum_{j=1}^{n} Y_j < 1\right] \\
&\geq \Pr[Y_1 = 0] \cdot \Pr\left[\sum_{j=2}^{n} Y_j < 1\right] \\
&\geq \left(1 - \frac{1}{48}\right) \cdot \frac{1}{6} \qquad \text{(by Lemma 3.1.2)} \\
&> \beta
\end{aligned}$$

If $N \geq 1$, $m \geq \mathbb{E}[X_1] \geq 15/16$, and

$$\begin{aligned}
\Pr\left[\sum_{i=1}^{n} X_i < \sum_{i=1}^{n} \mu_i + 1\right] &\geq e^{-4/25} \Pr\left[\sum_{j=1}^{n'} Y_j < m+1\right] \\
&\geq e^{-4/25} \Pr[Y_1 = 0] \cdot \Pr\left[\sum_{j=2}^{n'} Y_j < m+1\right] \\
&\geq e^{-4/25} \left(1 - \frac{1}{48(m+1)}\right) \cdot \Pr\left[\sum_{j=2}^{n'} Y_j < 1\right] \\
&\geq e^{-4/25} \left(1 - \frac{1}{48(m+1)}\right) \cdot \frac{1}{6} \quad \text{(by Lemma 3.1.2)} \\
&\geq e^{-4/25} \left(\frac{92}{93}\right) \left(\frac{1}{6}\right) = \beta.
\end{aligned}$$

□

We remark that given the tightness of Theorem 1.2.3 and the lemmas in this section (which we show in Section 4.4), one cannot achieve a constant higher than 1/6 in Theorem 1.2.1 with only the information of the first four moments. The room for improvement in our work lies in the possibility of lowering $\delta = 4/25$ in Lemma 3.1.1. We could not do so (by more than a negligible amount) in our proof below. However, perhaps a deeper analysis could allow it.

Furthermore, we believe a tractable approach to bridging some of the gap between our 7/50 and the conjectured $1/e$ would be to apply a similar $2k$th moment method. An effective dual $2k$-degree polynomial Q may be similarly defined as our $Q_{\ell,r}$ but possibly with more double roots for Q and $Q-1$. In addition, Q could be defined so that many of its odd-degree coefficients are 0, at the benefit of disregarding the odd moments of those orders.

3.3 Proofs of Lemmas

As explained at the end of Section 2.1, we will show that for any X meeting the conditions, there is a choice of ℓ and r such that (2.3) and thus (2.4) hold. We will also refer to properties of the polynomial $Q_{\ell,r}$ laid out in Section 2.2.

3.3.1 Lemma 3.1.1

Proof. Let $X_1, \ldots, X_n$ be 4-wise independent random variables such that for each i, $\mathbb{E}[X_i] = 0$, and $|X_i| \leq 1$. Let $X = \sum_{i=1}^n X_i$. This time, by assumption, we have $\mathbb{E}[X^3] \geq 0$. Otherwise, $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^4] \leq 3\sigma^4 + \sigma^2$, as shown in Section 2.1. Now, from (2.8) and (2.9) we have

$$\mathbb{E}[Q_{\ell,r}(X-\delta)] \leq \sum_{i=0}^{4}(-\delta)^i q_i + (q_2 - 3\delta q_3 + 6\delta^2 q_4)\sigma^2 + q_4(3\sigma^4 + \sigma^2). \tag{3.3}$$

With $\ell = r = \sqrt{3}\sigma$,

$$\mathbb{E}[Q_{\ell,r}(X-\delta)] \leq \frac{5}{6} + \frac{2\delta^4 + \sqrt{3}\delta^3\sigma + 2\sigma^2 - 6\sqrt{3}\delta\sigma^3}{36\sigma^4}.$$

Letting $\delta = 4/25$ and $\sigma > 5/4$,

$$\begin{aligned}
\mathbb{E}[Q_{\sqrt{3}\sigma}(X - 4/25)] &\leq \frac{5}{6} + \frac{1}{36\sigma}\left(2\left(\frac{4}{25}\right)^4\sigma^{-3} + \sqrt{3}\left(\frac{4}{25}\right)^3\sigma^{-2} + 2\sigma^{-1} - \frac{24\sqrt{3}}{25}\right) \\
&\leq \frac{5}{6} + \frac{1}{36\sigma}\left(2\left(\frac{4}{25}\right)^4\left(\frac{4}{5}\right)^3 + \sqrt{3}\left(\frac{4}{25}\right)^3\left(\frac{4}{5}\right)^2 + 2\left(\frac{4}{5}\right) - \frac{24\sqrt{3}}{25}\right) \\
&\leq \frac{5}{6}.
\end{aligned}$$

For $\sigma \in [0, 5/4]$, we will be forced to choose $\ell < r$. In order to mitigate some of the upcoming messiness, we refer to $\delta = 4/25$ as δ.

Let $\ell = 2\sigma$ and $r = 5\sigma/2$. Then using (3.3) and (2.6),

$$\mathbb{E}[Q_{2\sigma,5\sigma/2}(X-\delta)] \leq \frac{835}{972} - \frac{226\delta}{729}\sigma^{-1} + \frac{68 - 207\delta^2}{2916}\sigma^{-2} + \frac{11\delta^3}{243}\sigma^{-3} + \frac{17\delta^4}{729}\sigma^{-4}.$$

Since $68 - 207\delta^2 \geq 0$, the right-hand side is a convex polynomial of the variable $\sigma^{-1} > 0$. One can check that when $\sigma = .68$ and when $\sigma = 1.25$ (and $\delta = .16$), it is less than $5/6$. Therefore,

$$\mathbb{E}[Q_{2\sigma,5\sigma/2}(X - 4/25)] \leq \frac{5}{6}$$

for all $\sigma \in [.68, 1.25]$.

Next, let $\ell = 15\sigma/7$ and $r = 3\sigma$. Then

$$\mathbb{E}[Q_{15\sigma/7,3\sigma}(X-\delta)] \leq \frac{256957}{291600} - \frac{10633\delta}{32400}\sigma^{-1} + \frac{26411 - 128625\delta^2}{1749600}\sigma^{-2} + \frac{7889\delta^3}{194400}\sigma^{-3} + \frac{26411\delta^4}{1749600}\sigma^{-4}.$$

Again, this is a convex polynomial of the variable $\sigma^{-1} > 0$, since the coefficient of σ^{-2} is positive for $\delta = 4/25$. One can check that when $\sigma = .5$ and when $\sigma = .68$, the right-hand side is less than 5/6. Therefore,

$$\mathbb{E}[Q_{(15\sigma/7),3\sigma}(X - 4/25)] \leq \frac{5}{6}$$

for all $\sigma \in [.5, .68]$.

Lastly, let $\ell = 1$ and $r = 2$. Then

$$\mathbb{E}[Q_{1,2}(X - 4/25)] \leq \frac{62573}{78125} - \frac{59}{3375}\sigma^2 + \frac{5}{9}\sigma^4.$$

One can verify with the quadratic formula or by other means that the right-hand side is bounded above by 5/6 when $\sigma \in [0, 1/2]$.

Overall, we have provided a suitable polynomial Q for every $\sigma \geq 0$. □

3.3.2 Lemma 3.1.2

Proof. Let $X_1, \ldots, X_n$ be 4-wise independent mean-zero random variables distributed as

$$X_i = \begin{cases} -a_i, & \text{with probability } \dfrac{b_i}{a_i + b_i} \\ b_i, & \text{with probability } \dfrac{a_i}{a_i + b_i}, \end{cases}$$

where $a_i \leq 1$ for each i, $b_1 = \max_i\{b_i\}$, and $a_1 \geq 1/16$. Then

$$\begin{aligned}
\mathbb{E}[X] &= 0, \\
\mathbb{E}[X^2] := \sigma^2 &= \sum_{i=1}^{n} a_i b_i \ , \\
\mathbb{E}[X^3] &= \sum_{i=1}^{n} a_i b_i (b_i - a_i) \\
&\geq -\sum_{i=1}^{n} a_i^2 b_i \\
&\geq -\sum_{i=1}^{n} a_i b_i = -\sigma^2, \\
\mathbb{E}[X^4] &\leq 3\sigma^4 + \sum_{i=1}^{n} \mathbb{E}[X_i^4] \\
&= 3\sigma^4 + \sum_{i=1}^{n} a_i b_i (a_i^2 + b_i^2) \\
&\leq 3\sigma^4 + \sum_{i=1}^{n} a_i^3 b_i + \sum_{i=1}^{n} a_i b_i^3 \\
&\leq 3\sigma^4 + \sigma^2 + \sum_{i=1}^{n} a_i b_i^3
\end{aligned}$$

Now we will show $\sum_{i=1}^{n} a_i b_i^3 \leq 4\sigma^3$. Despite the lack of an upper bound on the b_i's, the nonpositivity of the third moment, along with the prescribed interval of a_1, brings the sum under control (the latter condition, simply put, prevents an extremely large b_1 being "hidden" by an extremely small a_1). First, note that $\mathbb{E}[X^3] \leq 0$ implies

$$\sum_{i=1}^{n} a_i b_i^2 \leq \sum_{i=1}^{n} a_i^2 b_i \leq \sigma^2.$$

Then

$$\begin{aligned}
\sum_{i=1}^{n} a_i b_i^3 &\leq b_1 \sum_{i=1}^{n} a_i b_i^2 \\
&\leq b_1 \sigma^2 \\
&= \sqrt{b_1^2} \sigma^2 \\
&= \left(\frac{1}{\sqrt{a_1}} \sqrt{a_1 b_1^2} \right) \sigma^2 \\
&\leq \left(4 \sqrt{\sum a_i b_i^2} \right) \sigma^2 \\
&\leq (4\sqrt{\sigma^2})\sigma^2 = 4\sigma^3.
\end{aligned}$$

From (2.8) and (2.9), we have

$$\mathbb{E}[Q_{\ell,r}(X-1)] \leq \sum_{i=0}^{4}(-1)^i q_i + (q_2 - 3q_3 + 6q_4)\sigma^2 + (q_3 - 4q_4)(-\sigma^2) + q_4(3\sigma^4 + 4\sigma^3 + \sigma^2). \tag{3.4}$$

If $\ell = r = \sqrt{3}\sigma$,

$$\mathbb{E}[Q_{\sqrt{3}\sigma}(X-1)] \leq \frac{5}{6} + \frac{2 + \sqrt{3}\sigma + 10\sigma^2 + (8 - 5\sqrt{3})\sigma^3}{36\sigma^4}.$$

If $\sigma \geq 16$,

$$\mathbb{E}[Q_{\sqrt{3}\sigma}] \leq \frac{5}{6} + \frac{1}{36\sigma}(2(16)^{-3} + \sqrt{3}(16)^{-2} + 10(16)^{-1} + 8 - 5\sqrt{3}) \leq \frac{5}{6}.$$

For the rest of this proof, we will still have $\ell = r$. We will find a Q_r for each $\sigma \in [0, 16]$. First, let $r = 19\sigma/10$. From (3.4), we have

$$\mathbb{E}[Q_{(19\sigma/10)}(X-1)] \leq \frac{218442 - 24885\sigma^{-1} + 37800\sigma^{-2} + 9500\sigma^{-3} + 10000\sigma^{-4}}{260642}.$$

As in the cases in the other proofs, the right-hand side is a convex polynomial of the parameter $\sigma^{-1} > 0$. One can check that for $\sigma = 5/2$ and $\sigma = 16$, the right-hand side is less than $5/6$. Therefore,

$$\mathbb{E}[Q_{(19\sigma/10)}(X-1)] \leq \frac{5}{6}$$

when $\sigma \in [5/2, 16]$.

For the remaining σ, we can let $r = 5$. Then (3.4) becomes

$$\begin{aligned}\mathbb{E}[Q_5(X-1)] &\leq \frac{1016 - 29\sigma^2 + 4\sigma^3 + 3\sigma^4}{1250} \\ &\leq \frac{5}{6} \text{ when } \sigma \in [0, 5/2],\end{aligned}$$

and the last inequality can be verified using basic calculus. All cases are covered. □

Chapter 4

Sums of Bernoulli Random Variables

4.1 Full Independence

For $\mu_1, \ldots, \mu_n$ satisfying $0 \leq \mu_1 \leq \ldots \leq \mu_n$, define $\mathcal{S}(\mu_1, \ldots, \mu_n)$ to be the set of all n-tuples of nonnegative independent random variables $(X_1, \ldots, X_n)$ satisfying $\mathbb{E}[X_i] = \mu_i$ $\forall i$. We restate Samuels' conjecture with this notation:

Conjecture 4.1.1 (Samuels)**.** *If $\sum_{i=1}^{n} \mu_i < 1$, then*

$$\inf_{(X_1,\ldots,X_n)\in\mathcal{S}(\mu_1,\ldots,\mu_n)} \Pr[X_1 \ldots + X_n < 1] = \min_{t=0,\ldots,n-1} \prod_{i=t+1}^{n} \left(1 - \frac{\mu_i}{1 - \sum_{j=1}^{t} \mu_j}\right). \quad (4.1)$$

In their discussion of this conjecture, the authors of [1] showed that

Lemma 4.1.1 ([1])**.** *If $\mu_1 = \ldots \mu_n = x$, where $0 < x \leq \frac{1}{n+1}$, then the righthand side of (4.1) is minimized at $t = 0$.*

Now let $\mathcal{B}_k(\mu_1, \ldots, \mu_n)$ be the subclass of $\mathcal{S}(\mu_1, \ldots, \mu_n)$ where for each i, X_i has support $\{0, b_i\}$ for some $b_i > 0$, and for any k-sized subset of $\{b_1, \ldots, b_n\}$, we have

$$\sum_{j=1}^{k} b_{i_j} < 1$$

whereas for any $(k+1)$-sized subset we have

$$\sum_{j=1}^{k+1} b_{i_j} \geq 1.$$

Let $\mathcal{B}(\mu_1, \ldots, \mu_n) = \cup_{k=0}^{n-1} \mathcal{B}_k(\mu_1, \ldots, \mu_n)$. In [12], Samuels proved that his conjecture holds for this subclass.

Lemma 4.1.2 ([12])**.** *Conjecture 5.1 is true for $\mathcal{S}(\mu_1, \ldots, \mu_n)$ replaced with $\mathcal{B}(\mu_1, \ldots, \mu_n)$.*

Together, these show

Theorem 4.1.3. *For any $p > 0$, if $X \sim B(n,p)$, then*

$$\Pr[X < (n+1)p] > \frac{1}{e}. \tag{4.2}$$

Proof. Let $X_1, \ldots, X_n$ be i.i.d. random variables where $X_1 \sim \frac{1}{(n+1)p} Ber(p)$ for $p > 0$, then it is easy to see that $(X_1, \ldots, X_n) \in \mathcal{B}(\frac{1}{n+1}, \ldots, \frac{1}{n+1})$ and that they also satisfy the conditions in Lemma 4.1.1. Therefore

$$\Pr[X_1 + \ldots + X_n < 1] \geq \left(1 - \frac{1}{n+1}\right)^n > 1/e,$$

and (4.2) follows. □

With the benefit of Samuels' lemma, it is also not hard to show

Theorem 4.1.4. *Let $X_1, \ldots, X_n$ be Bernoulli random variables with respective probabilities $p_1, \ldots, p_n$. Let $X = \sum_{i=1}^n X_i$. Then*

$$\Pr\left[X < \sum_{i=1}^n p_i + 1\right] > \frac{1}{e}.$$

Proof. Let $\mu = \sum_{i=1}^n p_i$, $Y_i = X_i/(\mu+1)$, and $Y = \sum_{i=1}^n Y_i$. Now $(Y_1, \ldots, Y_n) \in \mathcal{B}(\frac{p_1}{\mu+1}, \ldots, \frac{p_n}{\mu+1})$, so by Lemma 4.1.2,

$$\begin{aligned}
\Pr[Y_1 \ldots + Y_n] &= \min_{t=0,\ldots,n-1} \prod_{i=t+1}^n \left(1 - \frac{p_i}{\mu + 1 - \sum_{j=1}^t p_j}\right) \\
&= \min_{t=0,\ldots,n-1} \prod_{i=t+1}^n \left(1 - \frac{p_i}{\sum_{j=t+1}^n p_j + 1}\right)
\end{aligned}$$

Let $t = k$ be the integer at which the above attains its minimum. Then

$$\begin{aligned}
\Pr[Y_1 \ldots + Y_n] &= \prod_{i=k+1}^n \left(1 - \frac{p_i}{\sum_{j=k+1}^n p_j + 1}\right) \\
&\geq \prod_{i=k+1}^n \left(1 - \frac{p_i}{\sum_{j=k+1}^n p_j + p_i}\right) \\
&\geq \prod_{i=k+1}^n e^{\frac{-p_i}{\sum_{j=k+1}^n p_j}} \\
&= e^{-1}.
\end{aligned}$$

□

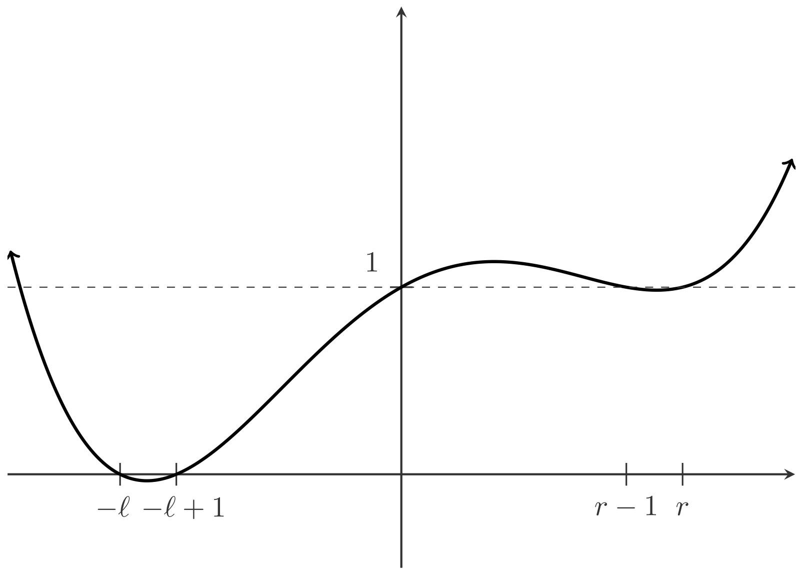

Figure 4.1: $Q'_{\ell,r}$

4.2 Probabilities at most 1/2

For nontrivial Bernoulli random variables with probabilities at most $1/2$, we can strengthen Corollary 1.2.4 by reducing the allowed deviation to 0. We can do this mainly because we now benefit from knowing that the sum can only be a nonnegative integer (and the restriction on the probabilities ensures that the central third moment of every random variable is nonnegative). This means that for the polynomial in the dual problem, we can use a modified version of $Q_{\ell,r}$ described in Section 2.1. Now we let $Q'_{\ell,r}$ be the degree-4 polynomial uniquely defined by interpolating through the following points:

$$(-\ell, 0), (-\ell+1, 0), (0, 1), (r-1, 1), (r, 1),$$

where $\ell, r \geq 2$ are integers. Thus, it will pass through the lines $y = 0$ and $y = 1$, but only at integers. The amount gained by this alteration is negligible for large ℓ and r but will be critical for dealing with the Poisson-like case. We elaborate in the proof.

Theorem 4.2.1. *Let $Y_1, \ldots, Y_n$ be 4-wise independent Bernoulli random variables, with marginal probabilities $(p_1, \ldots, p_n) \in (0, 1/2]^n$. Then*

$$\Pr[Y_1 + \ldots + Y_n \geq \sum_{i=1}^{n} p_i] \leq \frac{5}{6}. \tag{4.3}$$

Proof. For each i, let $X_i = Y_i - p_i$, and let $X = \sum_{i=1}^n X_i$. We have

$$\begin{aligned}
\mathbb{E}[X] &= 0, \\
\mathbb{E}[X^2] := \sigma^2 &= \sum_{i=1}^n p_i(1-p_i), \\
\mathbb{E}[X^3] &= \sum_{i=1}^n p_i(1-p_i)(1-2p_i), \\
\mathbb{E}[X^4] &= 3\sigma^4 + \sum_{i=1}^n p_i(1-p_i)(1-6p_i+6p_i^2)
\end{aligned}$$

Since $0 < p_i \leq 1/2$ for each i,

$$0 \leq \mathbb{E}[X^3] \leq \sigma^2,$$

and

$$\begin{aligned}
\mathbb{E}[X^4] &= 3\sigma^4 + \sum_{i=1}^n p_i(1-p_i)(1-2p_i)^2 - 2\sum_{i=1}^n p_i^2(1-p_i)^2 \\
&\leq 3\sigma^4 + \sum_{i=1}^n p_i(1-p_i)(1-2p_i)^2 \\
&\leq 3\sigma^4 + \sum_{i=1}^n p_i(1-p_i)(1-2p_i) \quad (0 \leq 1-2p_i \leq 1 \ \forall i) \\
&\leq 3\sigma^4 + \mathbb{E}[X^3].
\end{aligned}$$

As in the proofs of Theorems 1.2.2 and 1.2.3, if suffices to show that in each case there exists a suitable polynomial $Q(x)$ in the dual problem such that

$$\mathbb{E}[Q(X)] \leq \frac{5}{6}.$$

First, as usual, we use $Q_{\sqrt{3}\sigma}$ (see Sections 2.1 and 2.2). In this case,

$$\begin{aligned}
Q_{\sqrt{3}\sigma}(X) &= 1 + \frac{3}{4\sqrt{3}\sigma}\,\mathbb{E}[X] - \frac{1}{3\sigma^2}\,\mathbb{E}[X^2] - \frac{1}{12\sqrt{3}\sigma^3}\,\mathbb{E}[X^3] + \frac{1}{18\sigma^4}\,\mathbb{E}[X^4] \\
&\leq \frac{5}{6} + \frac{\mathbb{E}[X^3]}{36\sigma^4}(2-\sqrt{3}\sigma) \\
&\leq \frac{5}{6}, \quad \text{for } \sigma \geq \frac{2}{\sqrt{3}}.
\end{aligned}$$

Now let us take a closer look at the third and fourth moments of X. We can assume that $p_1 \leq \ldots \leq p_n$. Let $c = 1/2 - \sqrt{3}/6$, chosen because

$$1 - 6p_i + 6p_i^2 \leq 0, \quad \text{for } p_i \in [c, 1/2].$$

If $p_i \geq c$ for all i, then $\mathbb{E}[X^4] \leq 3\sigma^4$, and

$$Q_{\sqrt{3}\sigma}(X) \leq 1 - \frac{1}{3\sigma^2}\sigma^2 + \frac{1}{18\sigma^4}\sigma^4 = \frac{5}{6}.$$

Otherwise, let $N = \max_{1 \leq k \leq n}\{k : p_k < c\}$. Then

$$\mathbb{E}[X^3] \geq \sum_{i=1}^{N} p_i(1-p_i)(1-2p_i) \geq \frac{1}{\sqrt{3}} \sum_{i=1}^{N} p_i(1-p_i),$$

and

$$\begin{aligned}\mathbb{E}[X^4] &\leq 3\sigma^4 + \sum_{i=1}^{N} p_i(1-p_i)(1-6p_i+6p_i^2) \\ &\leq 3\sigma^4 + \sum_{i=1}^{N} p_i(1-p_i).\end{aligned}$$

Let us denote this partial variance as $t := \sum_{i=1}^{N} p_i(1-p_i)$. Now we will use $Q'_{\ell,r}$, as defined in the previous section. Let $\mu = \sum_{i=1}^{n} p_i$. Since Y can only take on integer values,

$$\Pr[Y \geq \mu] = \Pr[Y \geq \lceil \mu \rceil].$$

We will use the fact that $Q'_{\ell,r}(z) \geq \mathbf{1}_{\{z \geq 0\}}$ for all $z \in \mathbb{Z}$ (technically we have not proved this property for all ℓ and r but it will be easily verifiable for the specific two cases in which we use this polynomial). Thus,

$$\Pr[Y \geq \lceil \mu \rceil] \leq \mathbb{E}[Q'_{\ell,r}(Y - \lceil \mu \rceil)].$$

Let $\delta = \lceil \mu \rceil - \mu$. Since we are centering about $\lceil \mu \rceil$, we may have an allowed deviation from the mean, $0 \leq \delta < 1$. This will be much needed when $\mu < 1$. However, note that by Theorem 3.1.1, we can assume that $0 \leq \delta \leq 4/25$. For each of our next polynomials, denoted $Q(x) = \sum_{i=1}^{4} q_i x^i$, we have

$$\begin{aligned}\mathbb{E}[Q(Y - \lceil \mu \rceil)] &= \mathbb{E}[Q(X - \delta)] \\ &= \sum_{i=0}^{4} q_i \, \mathbb{E}[X^i] + \mathrm{Err}_Q(\delta),\end{aligned}$$

where

$$\mathrm{Err}_Q(\delta) = (-q_1 - 3q_3\sigma^2 - 4q_4\,\mathbb{E}[X^3])\delta + (q_2 + 6q_4\sigma^2)\delta^2 - q_3\delta^3 + q_4\delta^4.$$

For $5/7 \leq \sigma^2 \leq 4/3$, we let $\ell = 2$ and $r = 3$.

$$Q'_{2,3}(x) = 1 + \frac{17}{20}x - \frac{43}{120}x^2 - \frac{3}{20}x^3 + \frac{7}{120}x^4.$$

First, note that for this polynomial and our range of σ^2,

$$\begin{aligned}
\mathrm{Err}(\delta) &= \frac{27\sigma^2 - 14\,\mathbb{E}[X^3] - 51}{60}\delta + \frac{42\sigma^2 - 43}{120}\delta^2 + \frac{3}{20}\delta^3 + \frac{7}{120}\delta^4 \\
&\leq \frac{27(4/3) - 51}{60}\delta + \frac{42(4/3) - 43}{120}\delta^2 + \frac{3}{20}\delta^3 + \frac{7}{120}\delta^4 \\
&= \delta\left(\frac{-1}{4} + \frac{13}{120}\delta + \frac{3}{20}\delta^2 + \frac{7}{120}\delta^3\right) \\
&\leq \delta\left(\frac{-1}{4} + \frac{13}{120}\left(\frac{4}{25}\right) + \frac{3}{20}\left(\frac{4}{25}\right)^2 + \frac{7}{120}\left(\frac{4}{25}\right)^3\right) \\
&\leq 0.
\end{aligned}$$

Hence, for $5/7 \leq \sigma^2 \leq 4/3$,

$$\begin{aligned}
\mathbb{E}[Q'_{2,3}(X-\delta)] &\leq \mathbb{E}[Q'_{2,3}(X)] \\
&= 1 + \frac{17}{20}\,\mathbb{E}[X] - \frac{43}{120}\,\mathbb{E}[X^2] - \frac{3}{20}\,\mathbb{E}[X^3] + \frac{7}{120}\,\mathbb{E}[X^4] \\
&\leq 1 - \frac{43}{120}\sigma^2 - \frac{3}{20}\left(\frac{t}{\sqrt{3}}\right) + \frac{7}{120}(3\sigma^4 + t) \\
&\leq 1 - \frac{43}{120}\sigma^2 + \frac{7}{120}(3\sigma^4) \\
&\leq \frac{5}{6}.
\end{aligned}$$

Now let $\ell = r = 2$. Then

$$Q'_{2,2}(x) = 1 + \frac{7}{12}x - \frac{5}{8}x^2 - \frac{1}{12}x^3 + \frac{1}{8}x^4.$$

Again, we must examine the error term ($4/9 < \sigma^2 \leq 5/7 < 1$):

$$\begin{aligned}
\mathrm{Err}(\delta) &= \frac{3\sigma^2 - 6\,\mathbb{E}[Y^3] - 7}{12}\delta + \frac{6\sigma^2 - 5}{8}\delta^2 + \frac{1}{12}\delta^3 + \frac{1}{8}\delta^4 \\
&\leq -\frac{1}{4}\delta + \frac{1}{8}\delta^2 + \frac{1}{12}\delta^3 + \frac{1}{8}\delta^4 \\
&= \delta\left(\frac{-1}{4} + \frac{1}{8}\delta + \frac{1}{12}\delta^2 + \frac{1}{8}\delta^3\right) \\
&\leq \delta\left(\frac{-1}{4} + \frac{1}{8}\left(\frac{4}{25}\right) + \frac{1}{12}\left(\frac{4}{25}\right)^2 + \frac{1}{8}\left(\frac{4}{25}\right)^3\right) \\
&\leq 0.
\end{aligned}$$

Now for $4/9 \le \sigma^2 \le 5/7$,

$$\begin{aligned}
\mathbb{E}[Q'_{2,2}(X-\delta)] &\le \mathbb{E}[Q'_{2,2}(X)] \\
&= 1 + \frac{7}{12}\,\mathbb{E}[X] - \frac{5}{8}\,\mathbb{E}[X^2] - \frac{1}{12}\,\mathbb{E}[X^3] + \frac{1}{8}\,\mathbb{E}[X^4] \\
&\le 1 - \frac{5}{8}\sigma^2 - \frac{1}{12}\left(\frac{t}{\sqrt{3}}\right) + \frac{1}{8}(3\sigma^4 + t) \\
&= 1 - \frac{5}{8}\sigma^2 + \frac{3}{8}\sigma^4 + \frac{9-2\sqrt{3}}{72}t \\
&\le 1 - \frac{5}{8}\sigma^2 + \frac{3}{8}\sigma^4 + \frac{9-2\sqrt{3}}{72}\sigma^2 \\
&\le 1 - \frac{13}{24}\sigma^2 + \frac{3}{8}\sigma^4 \\
&\le \frac{5}{6}.
\end{aligned}$$

Now we can assume $\sigma^2 < 4/9$. Here we will use the fact that

$$0 \le \mu = \sum_{i=1}^{n} p_i \le 2\sum_{i=1}^{n} p_i(1-p_i) = 2\sigma^2 < \frac{8}{9},$$

given our range of probabilities. Thus, $\delta = \lceil \mu \rceil - \mu > 1/9$. We will take advantage of this guaranteed deviation. In addition,

$$\sigma^2 < 1/4 \Longrightarrow \mu < 1/2 \Longrightarrow \delta > 1/2,$$

and we would apply Theorem 3.1.1, as previously mentioned. Thus, we can assume that $1/2 \le \sigma^2 < 4/9$, as well as $1/9 < \delta < 4/25$. Now

$$\begin{aligned}
\mathrm{Err}(\delta) &= \frac{3\sigma^2 - 6\,\mathbb{E}[Y^3] - 7}{12}\delta + \frac{6\sigma^2-5}{8}\delta^2 + \frac{1}{12}\delta^3 + \frac{1}{8}\delta^4 \\
&\le \frac{3\sigma^2 - 2\sqrt{3}t - 7}{12}\delta + \frac{6\sigma^2-5}{8}\delta^2 + \frac{1}{12}\delta^3 + \frac{1}{8}\delta^4 \\
&= \left(-\frac{7}{12}\delta - \frac{5}{8}\delta^2 + \frac{1}{12}\delta^3 + \frac{1}{8}\delta^4\right) + \left(\frac{1}{4}\delta + \frac{3}{4}\delta^2\right)\sigma^2 - \frac{\sqrt{3}}{6}\delta t.
\end{aligned}$$

We will need to combine this with the bound on $\mathbb{E}[Q'_{2,2}(X)]$ given above:

$$\begin{aligned}
\mathbb{E}[Q'_{2,2}(X-\delta)] &\le \mathbb{E}[Q'_{2,2}(X)] + \mathrm{Err}(\delta) \\
&\le \left(1 - \frac{7}{12}\delta - \frac{5}{8}\delta^2 + \frac{1}{12}\delta^3 + \frac{1}{8}\delta^4\right) + \frac{6\delta^2+2\delta-5}{8}\sigma^2 + \frac{3}{8}\sigma^4 + \frac{9-2\sqrt{3}-12\sqrt{3}\delta}{72}t \\
&\le \left(1 - \frac{7}{12}\delta - \frac{5}{8}\delta^2 + \frac{1}{12}\delta^3 + \frac{1}{8}\delta^4\right) - \frac{2829}{5000}\sigma^2 + \frac{3}{8}\sigma^4 + \frac{27-10\sqrt{3}}{216}\sigma^2,
\end{aligned}$$

where we used the fact that $1/9 \leq \delta \leq 4/25$, and $t \leq \sigma^2$. It is also not hard to verify that for this range of δ,

$$\left(1 - \frac{7}{12}\delta - \frac{5}{8}\delta^2 + \frac{1}{12}\delta^3 + \frac{1}{8}\delta^4\right) \leq \left(1 - \frac{7}{12}\left(\frac{1}{9}\right) - \frac{5}{8}\left(\frac{1}{9}\right)^2 + \frac{1}{12}\left(\frac{1}{9}\right)^3 + \frac{1}{8}\left(\frac{1}{9}\right)^4\right) = \frac{6086}{6561}.$$

Thus, for $1/4 \leq \sigma^2 \leq 4/9$

$$\begin{aligned}
\mathbb{E}[Q'_{2,2}(X-\delta)] &\leq \mathbb{E}[Q'_{2,2}(X)] + \mathrm{Err}(\delta) \\
&\leq \frac{6086}{6561} - \frac{(29754 + 3125\sqrt{3})}{67500}\sigma^2 + \frac{3}{8}\sigma^4 \\
&\leq \frac{13}{14} - \frac{1}{2}\sigma^2 + \frac{3}{8}\sigma^4 \\
&\leq \frac{5}{6}.
\end{aligned}$$

Since this was the last case to consider, we are done. □

4.3 Equal Marginal Probabilities

Using notation from [11], let $\mathcal{A}(n,k,p)$ be the set of all collections of n k-wise independent Bernoulli random variables with equal marginal probabilities p. We denote

$$Z_P(n,k,p,\delta) = \max_{(X_1,\ldots,X_n)\in\mathcal{A}(n,p,k)} \Pr[X_1 + \ldots X_n \geq np + \delta]. \tag{4.4}$$

We can find the above quantity using linear programming. Let $S = X_1 + \ldots + X_n$. Since we are interested in the symmetric event $\{S \geq np + \delta\}$, there is no loss in assuming that our identically distributed random variables are also symmetric. Hence, our programming problem will be in the $n+1$ variables $p_0, \ldots, p_n$, where

$$p_r := \Pr[S = r]. \tag{4.5}$$

Let $X \sim \mathrm{Bin}(n,p)$. For k-wise independence to hold, it is sufficient for the moments of S and X to be identical up to order k. Thus, we have the constraints

$$\mathbb{E}[X^i] = \mathbb{E}[S^i] = \sum_{r=0}^{n} r^i p_r, \tag{4.6}$$

for $0 \leq i \leq k$. Letting $m = \lceil np + \delta \rceil$, our objective function is $\sum_{r=m}^{n} p_r$. The dual problem is then

$$Z_D(n,k,p,\delta) = \min_{Q\in\mathcal{P}_k} \mathbb{E}_{X\sim Bin(n,p)}[Q(X)], \tag{4.7}$$

where $\mathcal{P}_k$ is the set of univariate polynomials Q of degree at most k, with

$$Q(i) \geq 0 \; \forall i \in \{0, \ldots, m(d) - 1\}, \text{ and} \tag{4.8}$$

$$Q(j) \geq 1 \; \forall j \in \{m(d), \ldots, n\}. \tag{4.9}$$

By linear programming duality, $Z_P = Z_D$ $(:= Z)$. As explained in [11], an optimal Q_0 in (4.7) would give us information about the optimal distribution S in the primal problem (4.4). Assuming optimality of each,

$$\sum_{i=m}^{n} \Pr[S = i] = Z = \mathbb{E}[Q_0(S)] = \sum_{i=0}^{n} Q_0(i) \Pr[S = i].$$

Thus, the support of S contains only integers which are zeros of Q_0 as well as the $i \geq m$ where $Q_0(i) = 1$. With this information, one can simply use the $k+1$ linear constraints to solve for the probabilities.

4.4 4-wise Independence (Tightness of Bound)

4.4.1 Theorem 1.2.5

This and the following example will be a symmetric 4-wise independent collection of Bernoulli random variables $X_1, \ldots, X_n$ with equal marginal probabilities of 1/2. Let $S = \sum_{i=1}^{n} X_i$, and $n = 12k^2$ for some integer k. This choice ensures that $\mu = n/2 = 6k^2$ and $\sqrt{3}\sigma = \sqrt{3n}/2 = 3k$ are integers. We have seen that when S has support on the integers, the polynomial $Q'_{\ell,r}$ can be a better choice than $Q_{\ell,r}$ when it comes to minimizing $\mathbb{E}[Q(S - \mu)]$ (in fact, much better when σ is small). For large variance, we inevitably choose $r = \ell = \sqrt{3}\sigma$ to optimize the bound, so we do so with $Q'_{\ell,r}$. If this were in fact the optimal choice, the discussion at the end of the previous section tells us that S would have support

$$\begin{aligned}
\text{supp}(S) &= \{s_1, s_2, s_3, s_4, s_5\} \\
&= \{\mu - \ell, \mu - \ell + 1, \mu, \mu + r - 1, \mu + r\} \\
&= \{\mu - \sqrt{3}\sigma, \mu - \sqrt{3}\sigma + 1, \mu, \mu + \sqrt{3}\sigma - 1, \mu + \sqrt{3}\sigma\} \\
&= \{6k^2 - 3k, 6k^2 - 3k + 1, 6k^2, 6k^2 + 3k - 1, 6k^2 + 3k\}
\end{aligned}$$

Solving the linear system (4.6) gives

$$\begin{aligned}\Pr[S = s_1] = \Pr[S = s_5] &= \frac{4k-1}{4(6k-1)},\\ \Pr[S = s_2] = \Pr[S = s_4] &= \frac{3k^2}{4(6k-1)(3k-1)^2},\\ \Pr[S = s_3] &= \frac{12k^2-8k+1}{2(3k-1)^2}.\end{aligned}$$

Then

$$\begin{aligned}\Pr[S \geq np] &= \Pr[S = s_3] + \Pr[S = s_4] + \Pr[S = s_5]\\ &= \frac{30k^2-20k+3}{4(3k-1)^2},\end{aligned}$$

which tends to 5/6 as k grows. As explained in Section 4.3, we need only provide the distribution of S. Thus, the bound in Theorem 1.2.5 cannot be improved to a lower constant.

4.4.2 Other Theorems

Notice that in the above example, most of the weight of the distribution lies on the mean itself. Specifically, $\Pr[S = \mu]$ is around 2/3, whereas $\Pr[S > \mu] \leq 1/6$. This begs the question as to whether the 5/6 bound is still tight if we allow some slight deviation. The answer turns out to be in the affirmative.

We show the existence of a symmetric 4-wise independent collection $X_1, \ldots, X_n$ such that, letting $S = \sum_{i=1}^n X_i$ and $\mu = \mathbb{E}[S]$, $\Pr[S \geq \mu+1]$ tends to 5/6. Again, we let $p = 1/2$ and $n = 12k^2$ as in the previous section. Frankly, we had to experiment with a few choices of the support of S until finding one that satisfies the system (4.6) (that is, with nonnegative probabilities). We began with the assumption that, as usual, the dual polynomial $Q'_{\ell,r}$ which minimizes $\mathbb{E}[Q(S-\mu-1)]$ has both ℓ and r near $\sqrt{3}\sigma = 3k$. It turns out that in this case we must raise ℓ and r by at least one (or adjust the definition of $Q'_{\ell,r}$). We let

$$\begin{aligned}\operatorname{supp}(S) &= \{s_1, s_2, s_3, s_4, s_5\}\\ &= \{\mu - \sqrt{3}\sigma, \mu - \sqrt{3}\sigma + 1, \mu + 1, \mu + \sqrt{3}\sigma + 1, \mu + \sqrt{3}\sigma + 2\}\\ &= \{6k^2 - 3k, 6k^2 - 3k + 1, 6k^2 + 1, 6k^2 + 3k + 1, 6k^2 + 3k + 2\}\end{aligned}$$

With this choice, solving (4.6) yields

$$\begin{aligned}
\Pr[S = s_1] &= \frac{15k^2 + 6k + 4}{4(3k+1)^2(6k+1))}, \\
\Pr[S = s_2] &= \frac{12k^2 + 9k + 4}{12k(6k+1)}, \\
\Pr[S = s_3] &= \frac{36k^3 + 24k^2 + 3k - 4}{6k(3k+1)^2}, \\
\Pr[S = s_4] &= \frac{12k^2 - 15k + 4}{12k(6k+1)}, \\
\Pr[S = s_5] &= \frac{3k(5k-2)}{4(3k+1)^2(6k+1))}.
\end{aligned}$$

Then

$$\begin{aligned}
\Pr[S \geq np + 1] &= \Pr[S = s_3] + \Pr[S = s_4] + \Pr[S = s_5] \\
&= \frac{540k^4 + 342k^3 + 24k^2 - 33k - 4}{12k(3k+1)^2(6k+1)},
\end{aligned}$$

which tends to 5/6 as k grows. Thus, the 5/6 bound is tight in all of our theorems involving 4-wise independence. One may ask if this is merely a property of the $p = 1/2$ case, but we believe the bound is tight for any fixed p. However, in other cases, solving the linear system is made much easier if one can first correctly guess the optimal distribution of S.

4.5 2- and 3-wise Independence

We refer to the programming problem in Section 4.3. We will find it convenient to set $\delta = dp$, so that $m = \lceil np + dp \rceil = \lceil (n+d)p \rceil$.

4.5.1 Theorem 1.2.6 ($k = 2$)

For $k = 2$, the optimal solution occurs when

$$\begin{aligned}
p_0 &= \frac{(1-p)(m - np + p)}{m}, \\
p_m &= \frac{p(1-p)n(n-1)}{m(n-m)}, \\
p_n &= \frac{p(np - m + 1 - p)}{n - m},
\end{aligned}$$

valid as long as $m \leq np + 1 - p$. Then

$$\Pr[S \geq m] = \frac{p(n + m - np - 1 + p)}{m}. \tag{4.10}$$

The optimal polynomial in the corresponding dual problem (4.7) is

$$f(x) = \frac{1}{mn}\big((m+n)x - x^2\big).$$

Note that f satisfies the conditions, $f(0) = 0, f(m) = f(n) = 1$, and

$$\begin{aligned}\mathbb{E}[f(X)] &= \frac{(m+n)np - \big(np(1-p) + n^2p^2\big)}{mn} \\ &= \frac{p(n + m - np - 1 + p)}{m}.\end{aligned}$$

Therefore,

$$Z(n, 2, p, dp) = \frac{p(n + m - np - 1 + p)}{m},$$

when $m \leq np + 1 - p$. If $m = np + dp$ (so this number is already an integer), then this is equivalent to $p \leq 1/(d+1)$. In this case,

$$\Pr[S \geq (n+d)p] = \frac{(n + (d+1)p - 1)}{n+d}.$$

Setting $p = 1/(d+1)$ (to maximize the above) and assuming $m = (n+d)p = (n+d)/(d+1) \in \mathbb{Z}$ gives the simple solution of

$$\begin{aligned}p_0 &= \frac{d}{n+d}, \\ p_m &= \frac{n}{n+d}, \\ p_n &= 0,\end{aligned}$$

so that

$$\Pr[X_1 + \dots X_n \geq (n+d)p] = \frac{n}{n+d},$$

which is the same bound given by Markov's inequality.

4.5.2 Theorem 1.2.7 ($k = 3$)

For $k = 3$, the expressions are a little messier, so we will omit some of the details on the way to the punchline. In this case, the support of the optimal solution is

$\{p_0, p_m, p_{n-1}, p_n\}$, with

$$\Pr[S \geq m] = \frac{p\big((n-2)(1-p)^2 + m(2-p)\big)}{m}, \tag{4.11}$$

as long as $m \leq np + 1 - 2p$. The optimal polynomial in the dual problem is

$$g(x) = \frac{1}{n(n-1)m}\big((n^2 + 2mn - n - m)x - (2n + m - 1)x^2 + x^3\big).$$

Note that g satisfies the conditions, $g(0) = 0, g(m) = g(n-1) = g(n) = 1$, and it can be checked that $\mathbb{E}[g(X)]$ equals the quantity in (4.11). Therefore,

$$Z(n, 3, p, dp) = \frac{p\big((n-2)(1-p)^2 + m(2-p)\big)}{m},$$

when $m \leq m \leq np+1-2p$. If $m = (n+d)p \in \mathbb{Z}$, then this is equivalent to $p \leq 1/(d+2)$, and again Z is maximized with p equal that value. With these choices, the solution simplifies to

$$\begin{aligned}
p_0 &= \frac{(d+1)^2}{(d+2)(n+d)}, \\
p_m &= \frac{(d+1)n(n-1)}{(n+d)(n+nd-d)}, \\
p_{n-1} &= 0, \\
p_n &= \frac{1}{(d+2)(n+nd-d)},
\end{aligned}$$

so that

$$\Pr[X_1 + \ldots X_n \geq (n+d)p] = 1 - \frac{(d+1)^2}{(d+2)(n+d)}.$$

4.6 (Large k)-wise Independence

4.6.1 Preliminaries

We state some fundamental results in probability theory that we will need to use in our next result. The first gives a guarantee on how close a sum of independent random variables is to a normal distribution.

Theorem 4.6.1 (Berry-Esseen Theorem)**.** *Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E}[X_i] = 0$ for all i, $\sigma^2 = \sum_{i=1}^n \mathbb{E}[X_i^2]$, and $\rho = \sum_{i=1}^n \mathbb{E}[|X_i|^3]$. Let $Z =$*

$(X_1 + \ldots X_n)/\sigma$ with cumulative distribution function $F(x)$. Then

$$|F(x) - \Phi(x)| \leq \frac{\rho}{\sigma^3} \text{ for all } x \in \mathbb{R},$$

where $\Phi(x)$ refers to the cdf of the standard normal.

We will apply this theorem to what is sometimes called a Poisson binomial random variable. Let $Y_1, \ldots, Y_n$ be independent Bernoulli random variables with respective means $p_1, \ldots, p_n$. Let $Y = \sum_{i=1}^n Y_i$, $\mu = \mathbb{E}[Y]$, and $\sigma^2 = \mathrm{Var}[Y]$.

Corollary 4.6.2. *For Y as above, let $Z = (Y - \mu)/\sigma$ with cumulative distribution function $F(x)$. Then*

$$|F(x) - \Phi(x)| \leq \frac{1}{\sigma} \text{ for all } x \in \mathbb{R}.$$

Proof. This follows immediately from the Berry-Esseen Theorem, given that

$$\sum_{i=1}^n \mathbb{E}[|Y_i - p_i|^3] = \sum_{i=1}^n p_i(1-p_i)|1-2p_i| \leq \sum_{i=1}^n p_i(1-p_i) = \sigma^2.$$

□

We will also use some standard Chernoff bounds, the appendix of [2] being a good reference for these. First a lower tail bound:

Theorem 4.6.3. *Let Y be a sum of independent Bernoulli random variables with $\mathbb{E}[Y] = \mu > 0$. Then for $a > 0$,*

$$\Pr[Y < \mu - a] < e^{-a^2/2\mu}.$$

And we use the following upper tail bound:

Theorem 4.6.4. *Let Y be a sum of independent Bernoulli random variables with $\mathbb{E}[Y] = \mu > 0$. For $\varepsilon > 0$,*

$$\Pr[Y > (1+\varepsilon)\mu] \leq \left(e^{\varepsilon}(1+\varepsilon)^{-(1+\varepsilon)}\right)^{\mu}.$$

We state and prove a couple of simple facts relating to the efficacy of the previous bound. Neither should be suprising.

Lemma 4.6.5. *Let Y be a sum of independent Bernoulli random variables with $\mathbb{E}[Y] = \mu > 0$. For a constant $a > 0$, if we use the above bound, then*

$$\Pr[Y > \mu + a] \leq \frac{e^a}{\left(1 + \frac{a}{\mu}\right)^{\mu+a}}.$$

The right-hand side is nondecreasing in μ.

Proof. It suffices to show that the function

$$f(x) = (1 + ax^{-1})^{x+a}$$

is nonincreasing for $x > 0$. We do this by showing $\big(\log f(x)\big)' \leq 0$ for $x > 0$.

$$\begin{aligned}(\log(f(x)))' &= \big((x+a)\log(1 + ax^{-1})\big)' \\ &= \log(1 + ax^{-1}) - ax^{-1} \leq 0.\end{aligned}$$

□

However, if we scale the deviation by $\sqrt{\mu}$, the bound is better for larger μ. Intuitively, as Y moves from being Poisson-like to more like a normal distribution, this upper tail bound improves.

Lemma 4.6.6. *Let Y be a sum of independent Bernoulli random variables with $\mathbb{E}[Y] = \mu > 0$. By the above bound, we know that for a constant a,*

$$\Pr[Y \geq \mu + a\sqrt{\mu}] \leq \frac{e^{a\sqrt{\mu}}}{\left(1 + a\mu^{-1/2}\right)^{\mu + a\sqrt{\mu}}}.$$

For any $a > 0$, the righthand side is nonincreasing in μ.

Proof. Clearly it suffices to show that the function

$$f(x) = \frac{e^{ax}}{(1 + ax^{-1})^{ax+x^2}}$$

is nonincreasing for $x > 0$. We will do so by showing that $\big(\log f(x)\big)' \leq 0$ for $x > 0$.

$$\begin{aligned}(\log(f(x)))' &= \big(ax - (ax + x^2)\log(1 + ax^{-1})\big)' \\ &= 2a - (a + 2x)\log(1 + ax^{-1}).\end{aligned}$$

By a bound on the logarithm given in [14], since $ax^{-1} > 0$,

$$(a+2x)\log(1+ax^{-1}) \geq (a+2x)\frac{2ax^{-1}}{2+ax^{-1}}$$
$$= 2a,$$

so we are done. □

4.6.2 Proof of Theorem 1.2.8

We state it again for convenience.

Theorem 4.6.7. *Let $0 < \varepsilon \leq 1/16$ and $0 < a < 1$. There exists an integer $K := K(\varepsilon, a)$ such that if $\mathbf{X} = (X_1, \ldots, X_n)$ is a K-wise independent collection of Bernoulli random variables with respective means given by $\mathbf{p} = (p_1, \ldots, p_n)$, where $p_i \leq 1 - a$ for all i, then*

$$\Pr\left[X_1 + \ldots + X_n < \sum_{i=1}^{n} p_i + 1\right] \geq \frac{1}{e} - \varepsilon.$$

Proof. Without loss of generality, assume $\epsilon = 2^{-i}$ where $i \geq 4$, and for this proof, let log refer to $\log_2$. Let K be an integer at least $(1/a)(1/\varepsilon)^{6500/\sqrt{a}}$. We can assume that $n > K$, because otherwise Theorem 4.1.4 would hold. With a fixed $\mathbf{p}$, let $\mathcal{I}$ refer to the fully independent distribution on $\{0,1\}^n$ with marginal probabilities given by $\mathbf{p}$, and let $\mathcal{D}$ refer to any such distribution which is K-wise independent. Let $\mu = \sum_{1=1}^{n} p_i$.

We prove something slightly more general by letting the deviation $\delta \geq 0$ be any small constant. Let $h(t) = \mathbf{1}_{\{t \geq 0\}}(t)$. Let $f = f_{\mathbf{p}} : \{0,1\}^n \to \{0,1\}$ be the linear threshold function

$$f(\mathbf{x}) = f_{\mathbf{p}}(\mathbf{x}) = h\big(x_1 + \ldots + x_n - (\mu + \delta)\big).$$

We point out that

$$h\big(x_1 + \ldots + x_n - (\mu + \delta)\big) = h\big(x_1 + \ldots + x_n - \lceil \mu + \delta \rceil)\big),$$

so we can assume that $(\mu + \delta)$ is a positive integer. Note that for each respective distribution,

$$\mathbb{E}[f(\mathbf{X})] = \Pr[f(\mathbf{X}) = 1] = \Pr[X_1 + \ldots + X_n \geq \mu + \delta].$$

By Theorem 4.1.4, it will be sufficient to show that

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{D}}[f(\mathbf{X})] - \mathbb{E}_{\mathbf{X}\sim\mathcal{I}}[f(\mathbf{X})] \le \varepsilon. \tag{4.12}$$

We show the existence of a polynomial $Q : \{0,1\}^n \to \mathbb{R}$ of degree at most K such that

- $Q(\mathbf{x}) \ge f(\mathbf{x})$ for all $\mathbf{x} \in \{0,1\}^n$ and
- $\mathbb{E}_{\mathcal{I}}[Q(\mathbf{x}) - f(\mathbf{x})] \le \varepsilon$.

Inequality (4.12) then easily follows:

$$\mathbb{E}_{\mathcal{D}}[f(\mathbf{X})] \le \mathbb{E}_{\mathcal{D}}[Q(\mathbf{X})] = \mathbb{E}_{\mathcal{I}}[Q(\mathbf{X})] \le \mathbb{E}_{\mathcal{I}}[f(\mathbf{X})] + \varepsilon.$$

Let $B = B(\varepsilon)$ be a (large) parameter depending on ε to be specified later, and

$$D = 1500 \log\left(\frac{1}{\varepsilon}\right) B + 2 < 1600 \log\left(\frac{1}{\varepsilon}\right) B.$$

The work done in [5] implies the existence of a univariate polynomial P of degree at most D with the following properties (see the remark following this proof for details):

1. $P(t) \ge h(t)$ for all $t \in \mathbb{R}$;
2. $P(t) \in [h(t), h(t) + \varepsilon/2]$ for $t \in [-1/2, -1/B] \cup [0, 1/2]$;
3. $P(t) \in [0, 1 + \varepsilon/2]$ for $t \in (-1/B, 0)$;
4. $|P(t)| \le (4t)^D$ for all $|t| \ge 1/2$,

To summarize properties 2 and 3, the polynomial is close to h within $[-1/2, 1/2]$ except for a small interval in the middle where it must increase from near 0 to near 1.

Let $X = \sum_{i=1}^n X_i$ and $\sigma^2 = \mathrm{Var}[X] = \sum_{i=1}^n p_i(1-p_i)$. Let $s = \max\{\sigma, 8/\varepsilon\}$. Finally, let $t(\mathbf{x}) = \dfrac{\sum_{i=1}^n x_i - (\mu + \delta)}{(\varepsilon/8)sB}$ and $Q(\mathbf{x}) = P(t(\mathbf{x}))$. Note that by property 1 above, $Q(\mathbf{x}) \ge f(\mathbf{x})$ and Q is a polynomial of degree D. As done in [5], to show Q satisfies inequality (1), we will condition on three events:

$$A_1 := t(\mathbf{x}) \in (-1/B, 0)$$

$$A_2 := t(\mathbf{x}) \in [-1/2, -1/B] \cup [0, 1/2]$$

$$A_3 := t(\mathbf{x}) \in (-\infty, -1/2) \cup (1/2, \infty)$$

Then

$$\mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x})] = \sum_{i=1}^{3} \mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid A_i] \Pr[A_i],$$

and we will bound each of the three terms on the righthand side.

Term 1: In this interval where P makes the necessary jump, the difference between the functions could be rather high, so we will need to show that $\Pr[A_1]$ is small. Note that

$$\Pr[A_1] = \Pr\left[X - (\mu + \delta) \in \left(-\frac{\varepsilon s}{8}, 0\right)\right].$$

If $s = 8/\varepsilon$, then we have

$$\Pr[A_1] = \Pr\left[X - (\mu + \delta) \in (-1, 0)\right] = 0,$$

since $X - (\mu + \delta)$ can only be an integer. So we can assume that $s = \sigma$. In that case, we use the Berry-Esseen Theorem (specifically, Corollary 4.6.2). Let $Z = \frac{X-\mu}{\sigma}$ and F be the cdf of Z.

$$\begin{aligned}
\Pr[A_1] &= \Pr\left[Z \in \left(-\frac{\varepsilon}{8} + \frac{\delta}{\sigma}, \frac{\delta}{\sigma}\right)\right] \\
&= F\left(\frac{\delta}{\sigma}\right) - F\left(-\frac{\varepsilon}{8} + \frac{\delta}{\sigma}\right) \\
&\leq \Phi\left(\frac{\delta}{\sigma}\right) - \Phi\left(-\frac{\varepsilon}{8} + \frac{\delta}{\sigma}\right) + \frac{2}{\sigma} \\
&\leq \frac{\varepsilon}{8} + \frac{\varepsilon}{4} = \frac{3\varepsilon}{8},
\end{aligned}$$

since $\sigma \geq 8/\varepsilon$. Thus the first term

$$\mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid A_1] \Pr[A_1] \leq \left(1 + \frac{\epsilon}{2}\right) \frac{3\varepsilon}{8} < \frac{7\varepsilon}{16},$$

given our earlier declaration that $\varepsilon < 1/16$.

Term 2: This is by far the easiest term to bound, as the error in this region is small:

$$\begin{aligned}
\mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid A_2] \Pr[A_2] &\leq \mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid A_2] \\
&= \mathbb{E}[P(t) - h(t) \mid t \in [-1/2, -1/B] \cup [0, 1/2]] \\
&\leq \frac{\varepsilon}{2},
\end{aligned}$$

by property (2) of the polynomial P.

Term 3: Here the large error will be offset by the small probability of the event. For each positive integer j, define

$$I_j^+ = \left[\frac{j}{2}, \frac{j+1}{2}\right) \text{ and}$$
$$I_j^- = \left(-\frac{j+1}{2}, -\frac{j}{2}\right].$$

Then

$$\mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid A_3] \Pr[A_3] = \sum_{j=1}^{\infty} \mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid t(\mathbf{x}) \in I_j^+] \Pr[t(\mathbf{x}) \in I_j^+]$$
$$+ \sum_{k=1}^{\infty} \mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid t(\mathbf{x}) \in I_k^-] \Pr[t(\mathbf{x}) \in I_k^-]$$

Let S_1 and S_2 be the first and second respective sums above. We first focus on S_1. Note that $Q(\mathbf{x}) = P(t(\mathbf{x})) \leq (4t(\mathbf{x}))^D$, so when $t(\mathbf{x}) \in I_j^+$, $t(\mathbf{x}) < (j+1)/2$, and $Q(\mathbf{x}) < (2j+2)^D$. Therefore

$$\mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid t(\mathbf{x}) \in I_j^+] \leq (2j+2)^D - 1 < (2j+2)^D.$$

Now we must bound the probabilities of these events.

$$\begin{aligned}
\Pr[t(\mathbf{x}) \in I_j^+] &\leq \Pr[t(\mathbf{x}) \geq j/2] \\
&= \Pr\left[\frac{X - (\mu + \delta)}{(\varepsilon/8)Bs} \geq \frac{j}{2}\right] \\
&\leq \Pr\left[X \geq \mu + \frac{\varepsilon Bsj}{16}\right] \\
&= \Pr\left[X \geq \mu\left(1 + \frac{\varepsilon Bsj}{16\mu}\right)\right] \\
&\leq e^{(\varepsilon Bsj/16)}\left(1 + \frac{\varepsilon Bsj}{16\mu}\right)^{-(\mu + \varepsilon Bsj/16)}
\end{aligned}$$

First, consider $s = 8/\varepsilon$. Then

$$\Pr[t(\mathbf{x}) \in I_j^+] \leq e^{(Bj/2)}\left(1 + \frac{Bj}{2\mu}\right)^{-(\mu + Bj/2)},$$

and the right-hand side increases as μ increases, by Lemma 4.6.5. We now find an upper bound for μ. By the upper bound on the marginal probabilities,

$$\mu = \sum_{i=1}^{n} p_i \leq \frac{1}{a}\sum_{i=1}^{n} p_i(1 - p_i) = \frac{1}{a}\sigma^2 \leq \frac{1}{a}\left(\frac{8}{\varepsilon}\right)^2.$$

This implies that for $s = 8/\varepsilon$,

$$\begin{aligned}\Pr[t(\mathbf{x}) \in I_j^+] &\leq e^{(Bj/2)} \left(1 + \frac{a\varepsilon^2 Bj}{128}\right)^{-(64/a\varepsilon^2 + Bj/2)} \\ &\leq \left(\frac{128e}{a\varepsilon^2 Bj}\right)^{Bj/2}.\end{aligned}$$

Together,

$$\begin{aligned}S_1 &= \sum_{j=1}^{\infty} \mathbb{E}[Q(\mathbf{x}) - f(x) \mid t(\mathbf{x}) \in I_j^+] \cdot \Pr[t(\mathbf{x}) \in I_j^+] \\ &< \sum_{j=1}^{\infty} (2j+2)^D \left(\frac{128e}{a\varepsilon^2 Bj}\right)^{Bj/2} \\ &\leq \sum_{j=1}^{\infty} (2j+2)^{1600\log(\varepsilon^{-1})B} \left(\frac{128e}{a\varepsilon^2 Bj}\right)^{Bj/2} \\ &= \sum_{j=1}^{\infty} \left(\frac{128e(2j+2)^{3200j^{-1}\log(\varepsilon^{-1})}}{a\varepsilon^2 Bj}\right)^{Bj/2} \\ &\leq \sum_{j=1}^{\infty} \left(\frac{128e \cdot 4^{3200\log(\varepsilon^{-1})}}{a\varepsilon^2 Bj}\right)^{Bj/2} \\ &= \sum_{j=1}^{\infty} \left(\frac{128e \cdot (1/\varepsilon)^{6400}}{a\varepsilon^2 Bj}\right)^{Bj/2}.\end{aligned}$$

If $s = \sigma > 8/\varepsilon$, then since $\mu \leq \sigma^2/a$,

$$\begin{aligned}\Pr[t(\mathbf{x}) \in I_j^+] &\leq \Pr\left[X \geq \mu + \frac{\varepsilon B\sigma j}{16}\right] \\ &\leq \Pr\left[X \geq \mu + \frac{\varepsilon B\sqrt{a}j}{16}\sqrt{\mu}\right]\end{aligned}$$

For the latter bound, we use Lemma 4.6.6 and the fact that

$$\mu = \sum_{i=1}^{n} p_i \geq \sum_{i=1}^{n} p_i(1-p_i) = \sigma^2 > \frac{64}{\epsilon^2}.$$

Then in this case,

$$\begin{aligned}\Pr[t(\mathbf{x}) \in I_j^+] &\leq \Pr\left[X \geq \mu + \frac{\varepsilon B\sqrt{a}j}{16}\sqrt{\mu}\right] \\ &\leq e^{(\sqrt{a}Bj/2)} \left(1 + \frac{\sqrt{a}\varepsilon^2 Bj}{128}\right)^{-(64/\varepsilon^2 + \sqrt{a}Bj/2)} \\ &\leq \left(\frac{128e}{\sqrt{a}\varepsilon^2 Bj}\right)^{\sqrt{a}Bj/2}\end{aligned}$$

Thus for $\sigma > 8/\varepsilon$,

$$\begin{aligned}
S_1 &= \sum_{j=1}^{\infty} \mathbb{E}[Q(\mathbf{x}) - f(x) \mid t(\mathbf{x}) \in I_j^+] \cdot \Pr[t(\mathbf{x}) \in I_j^+] \\
&\le \sum_{j=1}^{\infty} (2j+2)^{1600\log(\varepsilon^{-1})B} \left(\frac{128e}{\sqrt{a}\varepsilon^2 Bj}\right)^{\sqrt{a}Bj/2} \\
&= \sum_{j=1}^{\infty} \left(\frac{128e(2j+2)^{3200j^{-1}\log(\varepsilon^{-1})/\sqrt{a}}}{\sqrt{a}\varepsilon^2 Bj}\right)^{\sqrt{a}Bj/2} \\
&\le \sum_{j=1}^{\infty} \left(\frac{128e \cdot 4^{3200\log(\varepsilon^{-1})/\sqrt{a}}}{\sqrt{a}\varepsilon^2 Bj}\right)^{\sqrt{a}Bj/2} \\
&= \sum_{j=1}^{\infty} \left(\frac{128e \cdot (1/\varepsilon)^{6400/\sqrt{a}}}{\sqrt{a}\varepsilon^2 Bj}\right)^{\sqrt{a}Bj/2}.
\end{aligned}$$

Now we choose B to be an integer greater than $(400/a)(1/\varepsilon)^{6402/\sqrt{a}}$. Then in either case,

$$S_1 < \sum_{j=1}^{\infty} \left(\frac{128e}{400j}\right)^{Bj/2} \ll \frac{\varepsilon}{32}.$$

For the other side, we have

$$\mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid t(\mathbf{x}) \in I_j^-] \le (2j+2)^D \le e^{2jD},$$

and we use a lower tail bound:

$$\begin{aligned}
\Pr[t(\mathbf{x}) \in I_j^-] &\le \Pr[t(\mathbf{x}) \le -j/2] \\
&= \Pr\left[\frac{X - (\mu + \delta)}{(\varepsilon/8)Bs} \le \frac{-j}{2}\right] \\
&= \Pr\left[X \le \mu + \delta - \frac{\varepsilon Bsj}{16}\right] \\
&\le \Pr\left[X \le \mu - \frac{\varepsilon Bsj}{32}\right] \\
&\le e^{-\varepsilon^2 B^2 s^2 j^2 / 64\mu}.
\end{aligned}$$

Together, we have no trouble bounding this sum:

$$\begin{aligned}
S_2 &= \sum_{k=1}^{\infty} \mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid t(\mathbf{x}) \in I_k^-] \Pr[t(\mathbf{x}) \in I_k^-] \\
&\leq \sum_{k=1}^{\infty} e^{2jD} e^{-\varepsilon^2 B^2 s^2 j^2/64\mu} \\
&\leq \sum_{k=1}^{\infty} e^{3200 \log(\varepsilon^{-1}) Bj} e^{-\varepsilon^2 B^2 s^2 j^2/64\mu} \\
&= \sum_{k=1}^{\infty} \exp\left(Bj(3200 \log(\varepsilon^{-1}) - \varepsilon^2 B s^2 j/64\mu)\right) \\
&\leq \sum_{k=1}^{\infty} \exp\left(Bj(3200 \log(\varepsilon^{-1}) - \varepsilon^2 B \sigma^2 j/64\mu)\right) \\
&\leq \sum_{k=1}^{\infty} \exp\left(Bj(3200 \log(\varepsilon^{-1}) - a\varepsilon^2 Bj/64)\right) \\
&\leq \sum_{k=1}^{\infty} \exp\left(Bj(3200 \log(\varepsilon^{-1}) - 6j/\varepsilon^{6400})\right) \ll \frac{\varepsilon}{32}.
\end{aligned}$$

Thus, for term 3,

$$\mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid A_3] \Pr[A_3] = S_1 + S_2 \leq \frac{\varepsilon}{16}.$$

All together,

$$\begin{aligned}
\mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x})] &= \sum_{i=1}^{3} \mathbb{E}[Q(\mathbf{x}) - f(\mathbf{x}) \mid A_i] \Pr[A_i] \\
&\leq \frac{7\varepsilon}{16} + \frac{\varepsilon}{2} + \frac{\varepsilon}{16} = \varepsilon.
\end{aligned}$$

The proof is now complete, since $\varepsilon \leq 1/16$ implies that

$$D \leq 1600 \log(1/\varepsilon) B \leq 1600 \log(1/\varepsilon)(400/a)(1/\varepsilon)^{6402/\sqrt{a}} \leq (1/a)(1/\varepsilon)^{6500/\sqrt{a}} \leq K.$$

□

Remarks on the polynomial P: Letting P' be the polynomial described in [5], our $P = (P' + 1)/2$, since we choose to have our boolean functions map to $\{0, 1\}$ rather than $\{-1, 1\}$. Using their notation, we set $c = 200$. Regarding their value a, we let $B = 1/2a$, and for our purposes we make this parameter much larger. This effectively trades a tighter jump from 0 to 1 at the cost of a larger degree.

References

[1] N. Alon, P. Frankl, H. Huang, V. Rödl, A. Ruciński, B. Sudakov. "Large Matchings in Uniform Hypergraphs and the Conjectures of Erdős and Samuels." In *Journal of Combinatorial Theory, Series A.* vol. 119, pp. 1200-1215, 2012.

[2] N. Alon, J. Spencer. *The Probabilistic Method.* Wiley, New York, 2000.

[3] B. Berger. "The Fourth Moment Method." In *Siam Journal on Computing* vol. 26, no. 4, pp. 11881207, 1997.

[4] D. Bertsimas, I. Popescu. "Optimal Inequalities in Probability Theory: A Convex Optimization Approach." In *Siam Journal on Optimization* vol. 15, no. 3, pp. 780-804, 2005.

[5] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. Servedio, E. Viola. "Bounded Independence Fools Halfspaces. " In Proc. *50th Annual Symposium on Foundations of Computer Science (FOCS),* 2009, pp. 171–180.

[6] U. Feige. "On Sums of Independent Random Variables with Unbounded Variance, and Estimating the Average Degree in a Graph." In *Siam Journal on Computing* vol. 35, no. 4, pp. 964984, 2006.

[7] S. He, J. Zhang, S. Zhang. "Bounding Probability of Small Deviation: A Fourth Moment Approach." In *Mathematics of Operations Research* vol. 35, no. 1, pp. 208-232, 2010.

[8] S. He, Z. Luo, J. Nie, S. Zhang. "Semidefinite Relaxation Bounds for Indefinite Homogeneous Quadratic Optimization. In *Siam Journal on Optimization* vol. 19, no. 2, pp. 503-523, 2008.

[9] K. Isii. "The Extrema of Probability Determined by Generalized Moments (I) Bounded Random Variables." In *Ann. Math Statist.* vol. 12, no. 2, pp. 119-134.

[10] S. Karlin, W. Studden. "Tchebycheff Systems: With Applications in Analysis and Statistics." In *Siam Review* vol. 9, no. 2, pp. 257-258.

[11] R. Peled, A. Yadin, A. Yehudayoff. "The Maximal Probability that k-wise independent bits are all 1." In *Random Structures and Algorithms* vol. 38, no. 4, pp. 502-525, 2011.

[12] S. Samuels. "On a Chebyshev-Type Inequality for Sums of Independent Random Variables." In *Ann. Math. Statist.* vol. 37, pp. 248-259, 1966.

[13] S. Samuels. "More on a Chebyshev-type inequality for sums of independent random variables." Purdue Stat. Dept. Mimeo. Series no. 155, 1968.

[14] F. Tompsøe. "Some bounds for the logarithm function."