Bounding Probability of Small Deviation: A Fourth Moment Approach

Simai He*, Jiawei Zhang†, and Shuzhong Zhang‡

December 13, 2007

Abstract

In this paper we study the problem of bounding the value of the probability distribution function of a random variable X at $\mathsf{E}[X] + a$ where a is a small quantity in comparison with $\mathsf{E}[X]$, by means of the second and the fourth moments of X. In this particular context, many classical inequalities yield only trivial bounds. By studying the primal-dual moments-generating conic optimization problems, we obtain upper bounds for $\text{Prob}\,\{X \geq \mathsf{E}[X] + a\}$, $\text{Prob}\,\{X \geq 0\}$, and $\text{Prob}\,\{X \geq a\}$ respectively, where we assume the knowledge of the first, second and fourth moments of X. These bounds are proved to be tightest possible. As application, we demonstrate that the new probability bounds lead to a substantial sharpening and simplification of a recent result and its analysis by Feige ([7], 2006); also, they lead to new properties of the distribution of the cut values for the max-cut problem. We expect the new probability bounds to be useful in many other applications.

Keywords: probability of small deviation, fourth moment of a random variable, sum of random variables.

MSC subject classification: 60E15, 78M05, 60G50.

*Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong. Email: smhe@se.cuhk.edu.hk.

†Department of Information, Operations, and Management Sciences, Stern School of Business, New York University, New York, USA. Email: jzhang@stern.nyu.edu.

‡Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong. Email: zhang@se.cuhk.edu.hk. Research supported by Hong Kong RGC Earmarked Grants CUHK418505 and CUHK418406.

1 Introduction

For a random variable $X \in \mathbb{R}$, we consider the problem of upper bounding

$$\text{Prob}\{X \geq \mathsf{E}[X] + a\} \tag{1.1}$$

for a given real a. This problem has been studied extensively in the literature. Based on available information about the distribution of X, various inequalities have been developed, including, the well-known Markov inequality and Chebyshev inequality. Such inequalities (1.3) and (1.4) have been extremely useful. However, these two inequalities by themselves could sometimes be too weak to yield useful results especially when a is small or zero. This motivates us to develop stronger probability inequalities that can handle small deviations.

1.1 Our results

In Section 2, we start our investigation by developing upper bounds for $\text{Prob}\{X \geq 0\}$ that are relatively simple functions of the first, second, and fourth moments of X. In particular, we prove that, for any $v > 0$

$$\text{Prob}\{X \geq 0\} \leq 1 - \frac{4}{9}(2\sqrt{3} - 3)\left(-\frac{2M_1}{v} + \frac{3M_2}{v^2} - \frac{M_4}{v^4}\right). \tag{1.2}$$

Here and throughout the paper, we denote $M_m = \mathsf{E}[X^m]$. The above result is in Theorem 2.3 of the current paper. The bound provided by (1.2) has a relatively simple closed-form expression, and we have the freedom to choose any $v > 0$ in the bound. Therefore, it is quite convenient to use this bound as long as the information about M_1, M_2, and M_4 is available. The study of this type of probability bounds is motivated by a lemma used in He *et al.* [9], which is a special case of (1.2) when $M_1 = 0$ with a specific choice of v.

The assumptions on the inequality (1.2) is minimal. In fact, we do not even require the assumption that $\mathsf{E}[X] \leq 0$. That is to say, we can estimate the probability that $X \geq 0$ even when $\mathsf{E}[X] \geq 0$. In fact, inequality (1.2) is non-trivial i.e., the right hand side is less than 1, as long as $\mathsf{E}[X] < 0$ or $\mathsf{E}[X]^2\mathsf{E}[X^4] \leq \mathsf{E}[X^2]^3$. This is in contrast to many other probability inequalities in the literature, as we shall see in the next subsection.

The bound provided by (1.2) however, is not necessarily tight. It is of interest to know whether or not the bound can be further improved. We settle this issue by presenting in Theorem 2.8 a tight upper bound, which is thus the best possible bound, given the moments information. As it turns out, the bound (1.2) is a very good one, in view of the tight bound; it is even tight under a certain condition.

After settling the issue of the probability bound for $\text{Prob}\{X \geq 0\}$, it is natural to consider the bound for $\text{Prob}\{X \geq a\}$, using the same moments information. This extension is useful and is nontrivial to establish. When $\mathsf{E}[X] = 0$, we are able to provide a tight bound for $\text{Prob}\{X \geq a\}$, using the information of M_2 and M_4. The result is presented in Theorem 2.11.

Of course, inequality (1.2) may not be immediately applicable if the second and the fourth moments of X is not directly available. Fortunately, in many applications, as we shall demonstrate in this paper, it is relatively straightforward to compute or bound the second and the fourth moments.

In Section 3 and Section 4, we provide several examples to demonstrate the applicability of Theorem 2.3. Our first example regards the sum of n independent random variables. In particular, given n independent random variables, $X_1, X_2, \cdots, X_n$, we provide upper bounds on

$$\text{Prob}\left\{\sum_{i=1}^{n} X_i \geq \mathsf{E}\left[\sum_{i=1}^{n} X_i\right] + a\right\},\ i = 1, ..., m.$$

The bounds are particularly useful when a is a relatively small non-negative real. Here the random variables X_i could be bounded from both sides, or from below only. As a special case of this result, we obtain the following bound. If X_i is non-negative with expectation 1, then

$$\text{Prob}\left\{\sum_{i=1}^{n} X_i \geq n+1\right\} \leq \frac{7}{8}.$$

This strengthens the main result of a recent paper by Feige [7]. In [7], a weaker upper bound of 12/13 is proved by using a completely different approach, and the proof is considerably more involved and lengthy.

In Section 4 we also apply Theorem 2.3 to the well-known weighted maximum-cut problem. Given an undirected graph $G = (V, E)$ where each edge (u, v) has a weight w_{uv}, we wish to partition the vertices of G into two sets S_1 and S_2 so as to maximize the total weight of the edges (u, v) such that $u \in S_1$ and $v \in S_2$. A simple solution to this problem is to independently and equiprobably assign each vertex of G to either S_1 or S_2. We denote the total weight of edges with end-points in different sets by W. It is clear that the expected value of the W is exactly $\frac{1}{2}\sum_{(u,v)\in E} w_{u,v}$. By applying Theorem 2.3, we can show that

$$\text{Prob}\left\{W \geq \frac{1}{2}\sum_{(u,v)\in E} w_{u,v}\right\} > \frac{2\sqrt{3}-3}{15}$$

and

$$\text{Prob}\left\{W \geq \left(\frac{1}{2} + \frac{0.0036}{|V|}\right)\sum_{(u,v)\in E} w_{u,v}\right\} > 1.2\%.$$

Both bounds seem to be new. Furthermore, the second bound implies that for any graph, there exists a cut so that the total weight of edges in the cut is at least $\left(\frac{1}{2} + \frac{0.0036}{|V|}\right)\sum_{(u,v)\in E} w_{u,v}$.

1.2 Related Literature

In the literature, there are several probability inequalities based on moment information. For example, if X assumes only non-negative values, then

$$\text{Prob}\ \{X \geq a\} \leq \frac{\mathsf{E}[X]}{a}. \tag{1.3}$$

This is the well-known Markov inequality and gives the tightest possible bound when we know only that X is non-negative and has a given expectation. If the standard deviation of X, denoted by σ, is also available, and $t > 0$, then we have

$$\text{Prob}\ \{X \geq \mathsf{E}[X] + t\sigma\} \leq \frac{1}{1+t^2}. \tag{1.4}$$

This inequality is often referred to as the (one-sided) Chebyshev inequality. Both inequalities (1.3) and (1.4) have been extremely useful.

If we know the first three moments of X, it is shown in Bertsimas and Popescu [4] that,

$$\text{Prob}\,\{X > (1+\delta)\mathsf{E}[X]\} \leq \begin{cases} \min\left(\frac{C_M^2}{C_M^2+\delta^2}, \frac{1}{1+\delta}\cdot\frac{D_M^2}{D_M^2+(C_M^2-\delta)^2}\right), & \text{if } \delta > C_M^2, \\ \\ \frac{1}{1+\delta}\cdot\frac{D_M^2+(1+\delta)(C_M^2-\delta)}{D_M^2+(1+C_M^2)(C_M^2-\delta)}, & \text{if } \delta \leq C_M^2, \end{cases} \tag{1.5}$$

where $C_M^2 = \frac{M_2-M_1^2}{M_1^2}$ and $D_M^2 = \frac{M_1M_3-M_2^2}{M_1^4}$, and this bound is tight. Tight bounds on $\text{Prob}\,\{X < (1-\delta)M_1\}$ and $\text{Prob}\,\{|X-M_1| > \delta M_1\}$ are also provided in [4]. These inequalities are potentially useful to bound small deviation probability as well, i.e., when δ is small. However, we noticed that in several applications that we consider in this paper, it is harder to estimate M_3 than M_4. Furthermore, the bound provided by inequality (1.5) could be as weak as Markov's and Chebyshev's bounds, for instance, for the problem considered by Feige [7]. We shall discuss this in more details later.

Zelen [17] showed that, if the first four moments of X are known, then

$$\text{Prob}\ \{X \geq \mathsf{E}[X] + t\sigma\} \leq \left(1 + t^2 + \frac{(t^2 - t\kappa_3 - 1)^2}{\kappa_4 - \kappa_3^2 - 1}\right)^{-1} \quad \text{for} \quad t \geq \frac{\kappa_3 + \sqrt{\kappa_3^2+4}}{2}, \tag{1.6}$$

where $\kappa_m = \frac{M_m}{\sigma^m}$.

There are also probability inequalities that uses absolute moments of the random variable X. Let

$$\nu_m = \mathsf{E}\left[(X - \mathsf{E}[X])\right]^m.$$

Cantelli [3] showed that for $m > 0$,

$$\text{Prob}\ \{|X - \mathsf{E}[X]| \geq a\} \leq \frac{\nu_{2m} - \nu_m^2}{\nu_{2m} - \nu_m^2 + (a^m - \nu_m)^2} \quad \text{for} \quad a \geq \left(\frac{\nu_{2m}}{\nu_m}\right)^{1/m}.$$

When $m = 2$, the above inequality is reduced to the well-known (two-sided) Chebyshev inequality. Von Mises [13] proved that, for $m > k > 0$,

$$\text{Prob}\,\{|X| \geq a\} \leq \frac{J^m - \nu_m}{J^m - a^m} \quad \text{for} \quad a \geq \left(\frac{\nu_m}{\nu_k}\right)^{1/(m-k)},$$

where J is the root, different from a, of the equation

$$(J^m - a^m)/(J^k - a^k) = (\nu_m - a^m)/(\nu_k - a^k).$$

Unfortunately, it is clear from the conditions provided in the above inequalities proved by Zelen [17], Cantelli [3], and Von Mises [13], that none of them is applicable for bounding probabilities when the deviation is very small.

In a recent paper, He *et al.* [9] studied SDP relaxations for certain quadratic optimization problems. The main results are to establish the gap between the SDP relaxations and the quadratic optimization problems. As a key to their main results, they established the following inequality:

$$\text{Prob}\,\{X \geq \mathsf{E}[X]\} \leq 1 - \frac{9}{20} \cdot \frac{\sigma^4}{\mathsf{E}[(X - \mathsf{E}[X])^4]}.$$

This inequality is a special case of Theorem 2.3. The current paper is partly motivated by [9].

Our paper is also related to Berger [2], which uses the fourth moment information to bound the absolute value of a random variable. More specifically, it is shown in [2] that, for all $q > 0$,

$$\mathsf{E}[|X|] \geq \frac{3\sqrt{3}}{2\sqrt{q}} \left(\mathsf{E}[X^2] - \frac{\mathsf{E}[X^4]}{q}\right).$$

This result has been used by Berger to bound the absolute value of a weighted sum of $\{+1, -1\}$ unbiased random variables, and achieve tight bounds for the total discrepancy of a set system.

Our results can be viewed as solutions to a special class of moment problems. Moment problems concern about deriving bounds on the probability that a certain random variable belongs in a given set, given information on some of its moments. The study on moment problems has a long history; see Bertsimas and Popescu [4] for a brief review of this area. The tight bounds derived in our paper use an optimization method and duality theory. This duality approach was proposed independently and simultaneously by Isii [10] and Karlin [11]. Bertsimas and Popescu [4] show that, for univariate random variables, the dual of the moment problem can be formulated as a semidefinite program (SDP). This result is important because SDP problem can be solved in polynomial time within any prescribed accuracy. They also discuss the complexity of solving the dual moment problem for multivariate random variables. Recent results on moment problems can also be found in [5], [15], and [12].

The work by Bertsimas and Popescu [4] seems to have settled the moment problems for univariate random variables, i.e., for the given information of the moments, one can compute the desired probability bound efficiently by solving an SDP. However, such bounds may not be conveniently used because of the lack of simple closed-form expressions.

2 The Moment Problem: Duality Approach

Let us start our discussion by considering the problem:

$$\begin{array}{rcll} Z_P^1 & = & \max & \text{Prob}\{X \geq 0\} \\ & & \text{s.t.} & \mathsf{E}[X] = M_1 \\ & & & \mathsf{E}[X^2] = M_2 \\ & & & \mathsf{E}[X^4] = M_4, \end{array}$$

or equivalently

$$\begin{array}{rcll} Z_P^1 & = & \max_{F(\cdot)} & \int_{x\geq 0} 1 \cdot dF(x) \\ & & \text{s.t.} & \int_{x\in\mathbb{R}} 1 \cdot dF(x) = 1 \\ & & & \int_{x\in\mathbb{R}} x \cdot dF(x) = M_1 \\ & & & \int_{x\in\mathbb{R}} x^2 \cdot dF(x) = M_2 \\ & & & \int_{x\in\mathbb{R}} x^4 \cdot dF(x) = M_4, \end{array} \tag{2.7}$$

where the variable of this infinite dimensional optimization problem is the probability measure $F(\cdot)$.

The dual problem of (2.7) is given as follows:

$$\begin{array}{rcll} Z_D^1 & = & \min & y_0 + M_1 \cdot y_1 + M_2 \cdot y_2 + M_4 \cdot y_4 \\ & & \text{s.t.} & g(x) := y_0 + y_1 \cdot x + y_2 \cdot x^2 + y_4 \cdot x^4 \geq \mathbf{1}_{\{x\geq 0\}}, \quad \forall x \in \mathbb{R}. \end{array} \tag{2.8}$$

We first define a feasible solution to the dual problem (2.8).

Lemma 2.1. *For any $u > v > 0$, let $c = \frac{1}{(u+v)^3(u-v)} > 0$ and $d = \frac{2v}{(u+v)^3} > 0$. Define*

$$\begin{aligned} y_0 &= cu^4 + du^2, \\ y_1 &= 2du, \\ y_2 &= d - 2cu^2, \\ y_4 &= c. \end{aligned} \tag{2.9}$$

Then (y_0, y_1, y_2, y_4) is feasible to problem (2.8) if $u \leq \frac{1+\sqrt{3}}{2}v$.

Proof. If (y_0, y_1, y_2, y_3) is defined as in (2.9), then

$$g(x) = cx^4 + (d - 2cu^2)x^2 + 2dux + cu^4 + du^2 = c(x^2 - u^2)^2 + d(x+u)^2.$$

It is clear that $g(x) \geq 0$ for all $x \in \mathbb{R}$. It is left to verify that $g(x) \geq 1$ for all $x \geq 0$. We first observe that

$$g(0) = \frac{u^4 + 2u^3v - 2u^2v^2}{(u+v)^3(u-v)}.$$

Thus, $g(0) \geq 1$ is reduced to $v^2 + 2uv - 2u^2 \geq 0$, which is true by the assumption that $u \leq \frac{1+\sqrt{3}}{2}v$. Therefore, $g(0) \geq 1$.

Notice that,

$$g(x) = (x+u)^2 \cdot \left(c(x-u)^2 + d\right).$$

Since $c(x-u)^2 + d > 0$, we have $g(x) = 0$ if and only if $x = -u < 0$. Thus $x = -u < 0$ is the only global minimum solution of $g(x)$, and thus one of the local minimum solutions. Since $g(x)$ is a polynomial with order four, it has at most two local minimum solutions, including $x = -u$. We denote the other local minimum solution by z. If $z < 0$, then we must have that $g(x)$ is increasing for $x \geq 0$, and thus $g(x) \geq g(0) \geq 1$. Therefore, we assume that $z > 0 > -u$. If follows that z must be the largest root to $g'(x) = 0$. But

$$g'(x) = 4cx(x^2 - u^2) + 2d(x+u)$$

and the largest root to $g'(x) = 0$ is $\frac{u}{2} + \sqrt{\frac{u^2}{4} - \frac{d}{2c}}$. Therefore,

$$z = \frac{u}{2} + \sqrt{\frac{u^2}{4} - \frac{d}{2c}} = \frac{u}{2} + \sqrt{\frac{u^2}{4} - v(u-v)} = \frac{u}{2} + \left|v - \frac{u}{2}\right| = v.$$

The last equality holds since $u \leq \frac{1+\sqrt{3}}{2}v$. Now it is straightforward to verify that $g(z) = g(v) = 1$.

Finally, we observe that the global minimum solution to $g(x)$ in $[0, \infty)$ is either $x = 0$ or $x = z$. Therefore, $g(x) \geq g(0) = g(z) = 1$ for all $x \geq 0$. This completes the proof. □

In Lemma 2.1, if we choose $u = \frac{1+\sqrt{3}}{2}v$, then we have

Corollary 2.2. *For any $v > 0$, define*

$$\begin{aligned} y_0 &= 1, \\ y_1 &= \tfrac{8}{9}(2\sqrt{3}-3)v^{-1}, \\ y_2 &= \tfrac{4}{3}(3-2\sqrt{3})v^{-2}, \\ y_4 &= \tfrac{4}{9}(2\sqrt{3}-3)v^{-4}. \end{aligned} \tag{2.10}$$

Then (y_0, y_1, y_2, y_4) is feasible to problem (2.8) with an objective value

$$1 - \frac{4}{9}(2\sqrt{3}-3)\left(-\frac{2M_1}{v} + 3\frac{M_2}{v^2} - \frac{M_4}{v^4}\right).$$

Corollary 2.1 immediately leads to our first main result.

Theorem 2.3. *For any $v > 0$,*

$$\text{Prob}\,\{X \geq 0\} \leq 1 - \frac{4}{9}(2\sqrt{3}-3)\left(-\frac{2\mathsf{E}[X]}{v} + 3\frac{\mathsf{E}[X^2]}{v^2} - \frac{\mathsf{E}[X^4]}{v^4}\right). \tag{2.11}$$

In Theorem 2.3, we have the freedom to choose any $v > 0$. In particular, we could choose v that maximizes the function

$$-\frac{2M_1}{v} + 3\frac{M_2}{v^2} - \frac{M_4}{v^4}.$$

Such a v can be obtained by solving the equation

$$-M_1v^3 + 3M_2v^2 - 2M_4 = 0,$$

if a solution exists.

Notice that, even if we choose the best v, the bound provided in Theorem 2.3 is not necessarily tight. In what follows, we develop a tight bound for Prob $\{X \geq 0\}$, given the first, second, and the fourth moments of X. We begin with the case where the bound in Theorem 2.3 is tight. Since when $M_2 = M_1^2$ or $M_4 = M_2^2$ the distribution X can be easily identified by the first two moments, and the result in Theorem 2.8 easily follows, therefore we assume $M_2 > M_1^2$ and $M_4 > M_2^2$ for the remaining part of this section.

Let

$$\alpha = \sqrt{\frac{M_4 - M_2^2}{M_2 - M_1^2}} > 0$$

and

$$V_{\min} = \sqrt{(\sqrt{3}-1)M_2 + \frac{7-4\sqrt{3}}{4}M_1^2} + \frac{2-\sqrt{3}}{2}M_1 \geq 0. \tag{2.12}$$

Lemma 2.4. *If $M_2/M_1^2 \geq M_4/M_2^2$ and $\alpha \geq \frac{\sqrt{3}-1}{2}V_{\min}$, then*

$$\text{Prob}\,\{X \geq 0\} \leq 1 - \frac{4}{9}(2\sqrt{3}-3)\sup_{v>0}\left(-\frac{2\mathsf{E}[X]}{v} + 3\frac{\mathsf{E}[X^2]}{v^2} - \frac{\mathsf{E}[X^4]}{v^4}\right).$$

And the bound is tight.

Proof. The inequality has been established in Theorem 2.3. We need only to show that the bound is tight. In view of Lemma 2.1, it is sufficient to find a feasible solution to problem (2.7) with an objective value that is equal to the right hand side of the bound.

If $M_1 < 0$, define $f(x) = -M_1x^3 + 3M_2x^2 - 2M_4$. Then it can be verified that $f(x)$ is strictly increasing when $x \geq 0$. By assumption, $M_2/M_1^2 \geq M_4/M_2^2$, and thus,

$$f\left((\sqrt{3}-1)\frac{M_2}{-M_1}\right) = 2(M_2^3/M_1^2 - M_4) \geq 0.$$

On the other hand, by (2.12), $V_{\min}$ is a solution of the equation $x^2-(2-\sqrt{3})M_1x-(\sqrt{3}-1)M_2 = 0$. Thus

$$\begin{aligned} f(V_{\min}) &= (\sqrt{3}-2)M_1^2V_{\min}^2 - (4\sqrt{3}-7)M_1M_2V_{\min} + (3\sqrt{3}-3)M_2^2 - 2M_4 \\ &= (\sqrt{3}-2)M_1^2V_{\min}^2 + (\sqrt{3}-2)M_2\left((\sqrt{3}-1)M_2 - V_{\min}^2\right) + (3\sqrt{3}-3)M_2^2 - 2M_4 \\ &= 2M_2^2 - 2M_4 + (2-\sqrt{3})(D - M_1^2)V_{\min}^2 \\ &\leq 0 \end{aligned}$$

where the last inequality holds because of the assumption that $V_{\min} \leq (\sqrt{3}+1)\alpha$. By the monotonicity of $f(x)$ when $x \geq 0$, we must have $V_{\min} \leq (\sqrt{3}-1)\frac{M_2}{-M_1}$. Furthermore, there must exist a unique $v \in [V_{\min}, (\sqrt{3}-1)\frac{M_2}{-M_1}]$ such that $f(v) = 0$. For simplicity, in what follows, we assume v satisfy such a condition. Also, let $u = \frac{1+\sqrt{3}}{2}v$.

We now define a random variable

$$X = \begin{cases} -u, & \text{with probability } q := \frac{4}{9}(2\sqrt{3}-3)(2\frac{-M_1}{v} + 3\frac{M_2}{v^2} - \frac{M_4}{v^4}); \\ 0, & \text{with probability } 1-p-q; \\ v, & \text{with probability } p := \frac{6-2\sqrt{3}}{3}\frac{M_2}{v^2} - \frac{6-2\sqrt{3}}{9}\frac{M_4}{v^4} + \frac{4\sqrt{3}-3}{9}\frac{M_1}{v}. \end{cases}$$

We show X defines a feasible solution to problem (2.7).

First of all, by the fact $f(v) = 0$, or $M_4 = \frac{1}{2}\left(-M_1 v^3 + 3M_2 v^2\right)$, we have

$$q = \frac{2}{3}(2\sqrt{3}-3)\left(-\frac{M_1}{v} + \frac{M_2}{v^2}\right)$$

and

$$p = \frac{3-\sqrt{3}}{3}\frac{M_2}{v^2} + \frac{\sqrt{3}}{3}\frac{M_1}{v}.$$

Therefore $q \geq 0$ and $p \geq 0$ since $v \leq (\sqrt{3}-1)\frac{M_2}{-M_1}$. Furthermore,

$$p + q = (2-\sqrt{3})\frac{M_1}{v} + (\sqrt{3}-1)\frac{M_2}{v^2} \leq 1$$

since $v \geq V_{\min}$. Therefore, X is indeed a well-defined random variable.

It is easy to check that

$$\begin{aligned}
\mathsf{E}[X] &= -qu + pv = -\left(\frac{2}{3}(2\sqrt{3}-3)(-\frac{M_1}{v} + \frac{M_2}{v^2})\right)\frac{1+\sqrt{3}}{2}v + \left(\frac{3-\sqrt{3}}{3}\frac{M_2}{v^2} + \frac{\sqrt{3}}{3}\frac{M_1}{v}\right)v \\
&= M_1, \\
\mathsf{E}[X^2] &= \left(\frac{2}{3}(2\sqrt{3}-3)(-\frac{M_1}{v} + \frac{M_2}{v^2})\right)(\frac{1+\sqrt{3}}{2})^2 v^2 + \left(\frac{3-\sqrt{3}}{3}\frac{M_2}{v^2} + \frac{\sqrt{3}}{3}\frac{M_1}{v}\right)v^2 \\
&= M_2, \text{ and} \\
\mathsf{E}[X^4] &= \left(\frac{2}{3}(2\sqrt{3}-3)(-\frac{M_1}{v} + \frac{M_2}{v^2})\right)(\frac{1+\sqrt{3}}{2})^4 v^4 + \left(\frac{3-\sqrt{3}}{3}\frac{M_2}{v^2} + \frac{\sqrt{3}}{3}\frac{M_1}{v}\right)v^4 \\
&= -\frac{1}{2}M_1 v^3 + \frac{3}{2}M_2 v^2 = M_4.
\end{aligned}$$

Therefore, X is feasible to problem (2.7).

Finally, since $u \geq v > 0$, we have

$$\text{Prob}\{X \geq 0\} = \text{Prob}\{X \neq -u\} = 1 - q = 1 - \frac{4(2\sqrt{3}-3)}{9}\left(\frac{-2M_1}{v} + \frac{3M_2}{v^2} - \frac{M_4}{v^4}\right).$$

This completes the proof of the lemma for the case $M_1 < 0$.

For the case $M_1 \geq 0$ the proof is completely parallel, except that the solution for $f(v) = 0$ exists in range $v \in [V_{\min}, \frac{M_2}{M_1}]$. The details are omitted here. □

In order to get a tight bound for cases that are not covered in Lemma 2.4, we need to define different primal and dual variables, which are summarized in the following three lemmas.

Lemma 2.5. *If $M_2/M_1^2 \geq M_4/M_2^2$ and $\alpha \leq \frac{\sqrt{3}-1}{2} V_{\min}$, then*

$$\text{Prob}\{X \geq 0\} \leq \frac{1}{2} + \frac{1}{2} \frac{\alpha + 2M_1}{\sqrt{4M_2 + \alpha^2 + 4M_1\alpha}}.$$

And the bound is tight.

Proof. Define

$$\begin{cases} s &= \sqrt{4M_2 + \alpha^2 + 4\alpha M_1} \\ z &= \alpha + 2M_1 \\ u &= \frac{s+\alpha}{2} \\ v &= \frac{s-\alpha}{2}. \end{cases}$$

From the assumption $\alpha \leq \frac{\sqrt{3}-1}{2} V_{\min}$, we have

$$(5 + 3\sqrt{3})\alpha^2 - M_1\alpha - M_2 \leq 0.$$

It follows that

$$s = \sqrt{4M_2 + \alpha^2 + 4\alpha M_1} \geq (3 + 2\sqrt{3})\alpha$$

and thus

$$u = \frac{s+\alpha}{2} \leq \frac{1+\sqrt{3}}{2} \cdot \frac{s-\alpha}{2} = \frac{1+\sqrt{3}}{2} v,$$

which also implies that $u > v > 0$. Thus, by Lemma 2.1, the function

$$g(x) = c(x^2 - u^2)^2 + d(x+u)^2 = cx^4 + (d - 2cu^2)x^2 + 2dux + cu^4 + du^2,$$

with $c = \frac{1}{(u+v)^3(u-v)} > 0$ and $d = \frac{2v}{(u+v)^3} > 0$, (implicitly) defines a feasible solution to problem (2.8). The corresponding dual objective value is

$$cM_4 + (d - 2cu^2)M_2 + 2duM_1 + cu^4 + du^2 = \frac{s+z}{2s},$$

where we use the fact that

$$\begin{cases} M_2 &= \frac{s^2-\alpha^2}{4} - M_1\alpha \\ M_4 &= (M_2 - M_1^2)\alpha^2 + M_2^2 \\ c &= s^{-3}\alpha^{-1} \\ d &= s^{-2} - s^{-3}\alpha. \end{cases}$$

On the other hand, we define

$$X = \begin{cases} -u\,(< 0), & \text{with probability } q := \frac{s-z}{2s}; \\ v\,(> 0), & \text{with probability } p := \frac{s+z}{2s}. \end{cases}$$

We shall show that X is a feasible solution to problem (2.7).

It is obvious that $p + q = 1$. Also, by the fact that $M_2 \geq M_1^2$, we have

$$s = \sqrt{4M_2 + \alpha^2 + 4\alpha M_1} \geq |\alpha + 2M_1| = |z|.$$

Therefore, $p, q \geq 0$. Thus, X is a well-defined random variable.

Furthermore,

$$\begin{aligned}
\mathsf{E}[X] &= -uq + vp = -\frac{s+\alpha}{2}\frac{s-\alpha-2M_1}{2s} + \frac{s-\alpha}{2}\frac{s+\alpha+2M_1}{2s} \\
&= M_1 \\
\mathsf{E}[X^2] &= u^2q + v^2p = \frac{(s+\alpha)^2}{4}\frac{s-\alpha-2M_1}{2s} + \frac{(s-\alpha)^2}{4}\frac{s+\alpha+2M_1}{2s} = \frac{s^2-\alpha^2}{4} - M_1\alpha \\
&= M_2, \text{ and} \\
\mathsf{E}[X^4] &= u^2q + v^2p = \frac{(s+\alpha)^4}{16}\frac{s-\alpha-2M_1}{2s} + \frac{(s-\alpha)^4}{16}\frac{s+\alpha+2M_1}{2s} \\
&= (M_2 + \alpha M_1)(M_2 + \alpha^2 + \alpha M_1) - \alpha M_1(2M_2 + \alpha^2 + 2\alpha M_1) \\
&= (M_2^2 - M_1^2)\alpha^2 + M_2^2 \\
&= M_4.
\end{aligned}$$

Finally,

$$\text{Prob}\{X \geq 0\} = \text{Prob}\{X = v\} = \frac{s+z}{2s},$$

which is equal to the dual objective value. This completes the proof of the Lemma. □

Lemma 2.6. *If $M_2/M_1^2 \leq M_4/M_2^2$ and $M_1 < 0$, then*

$$\text{Prob}\{X \geq 0\} \leq 1 - \frac{M_1^2}{M_2}.$$

And the bound is tight.

Proof. The inequality is the well-known Chebyshev inequality, which is known to be tight; see, for example, Bertsimas and Popescu [4]. □

Lemma 2.7. *If $M_2/M_1^2 \leq M_4/M_2^2$ and $M_1 > 0$, then the trivial bound* $\text{Prob}\{X \geq 0\} \leq 1$ *is actually tight.*

Proof. The primal solution X with objective value 1 can be constructed this way: For any $t \geq M_1 > 0$, define random variable X_t as a two point distribution as follows:

$$X_t = \begin{cases} 0, & \text{with probability } 1 - \frac{M_1}{t} \\ t, & \text{with probability } \frac{M_1}{t}. \end{cases}$$

We have that $\mathsf{E}X_t = M_1$, $\mathsf{E}X_t^2 = tM_1$ and $\mathsf{E}X_t^4 = t^3M_1$.

Consider function $f(x) = x^3/M_1^2$, which is convex when $x > 0$. Notice that $M_2 \geq M_1^2$, $f(M_1^2) = M_1^4$ and $f(M_2) \leq M_4$, the line passing through (M_1^2, M_1^4) and (M_2, M_4) intersect with function $f(x)$ at some $t \geq M_2$. Thus there exists a $p \in [0, 1]$, such that

$$p(M_1^2, M_1^4) + (1-p)(t, f(t)) = (M_2, M_4).$$

Let Y be the Bernoulli trial which takes the value 1 with probability p, and let

$$X = YX_{M_1} + (1-Y)X_{t/M_1},$$

where Y is independent to X_{t/M_1} and X_{M_1}, then

$$\begin{aligned}(\mathsf{E}X, \mathsf{E}X^2, \mathsf{E}X^4) &= p(\mathsf{E}X_{M_1}, \mathsf{E}X_{M_1}^2, \mathsf{E}X_{M_1}^4) + (1-p)(\mathsf{E}X_{t/M_1}, \mathsf{E}X_{t/M_1}^2, \mathsf{E}X_{t/M_1}^4) \\ &= p(M_1, M_1^2, M_1^4) + (1-p)(M_1, t, f(t)) \\ &= (M_1, M_2, M_4).\end{aligned}$$

Because $X \geq 0$, this gives a feasible solution of the primal problem 2.7 with objective value 1. Since 1 is an upper bound for Z_P^1, we conclude that $Z_P^1 = 1$ and that X is an optimal primal solution.

For the dual problem (2.8), $y_0 - 1 = y_1 = y_2 = y_4 = 0$ is obviously a feasible solution with objective value 1. Because $Z_D^1 \geq Z_P^1 = 1$, this is an optimal dual solution. □

Lemma 2.6 and Lemma 2.7 indicate that when the fourth moment of X, i.e., M_4, becomes sufficiently large, then the information will not be useful anymore in bounding the probability that $X \geq 0$.

The following theorem summarizes the results we obtained above.

Theorem 2.8.

$$\begin{aligned}&\text{Prob}\{X \geq 0\} \\ &\leq \begin{cases} 1 - \frac{M_1^2}{M_2}, & \text{if } \frac{M_4}{M_2^2} \geq \frac{M_2}{M_1^2} \text{ and } M_1 < 0; \\ 1, & \text{if } \frac{M_4}{M_2^2} \geq \frac{M_2}{M_1^2} \text{ and } M_1 > 0; \\ 1 - \frac{4(2\sqrt{3}-3)}{9} \sup_{v>0}\left(-\frac{2M_1}{v} + 3\frac{M_2}{v^2} - \frac{M_4}{v^4}\right), & \text{if } \frac{M_4}{M_2^2} < \frac{M_2}{M_1^2} \text{ and } \alpha \geq \frac{\sqrt{3}-1}{2}V_{\min}; \\ \frac{1}{2} + \frac{1}{2}\frac{\alpha+2M_1}{\sqrt{4M_2+\alpha^2+4M_1\alpha}}, & \text{if } \frac{M_4}{M_2^2} < \frac{M_2}{M_1^2} \text{ and } \alpha \leq \frac{\sqrt{3}-1}{2}V_{\min}, \end{cases} \end{aligned} \tag{2.13}$$

where $\alpha \triangleq \sqrt{\frac{M_4 - M_2^2}{M_2 - M_1^2}}$, $V_{\min} \triangleq \sqrt{(\sqrt{3}-1)M_2 + \frac{7-4\sqrt{3}}{4}M_1^2} + \frac{2-\sqrt{3}}{2}M_1$. *Furthermore, the bound is tight, i.e., there exists an* X *such that the inequality (2.13) holds as an equality.*

Now we consider a special case where $\mathsf{E}[X] = 0$. As we have mentioned in the introduction, He *et al.* [9] established the following inequality

$$\text{Prob}\{X \geq \mathsf{E}[X]\} \leq 1 - \frac{9}{20} \cdot \frac{\sigma^4}{\mathsf{E}[(X - \mathsf{E}[X])^4]},$$

which has been a key to study an SDP relaxation for certain class of quadratic optimization problems. Here we show that this inequlality can be strengthened by using Theorem 2.3.

Corollary 2.9. *If* $\mathsf{E}[X] = 0$*, then*

$$\sup_X \{\text{Prob } (X \geq 0)\} = \begin{cases} 1 - (2\sqrt{3}-3)\frac{M_2^2}{M_4}, & \text{if } \frac{M_4}{M_2^2} \geq \frac{3\sqrt{3}-3}{2}; \\ \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{3+M_4/M_2^2}}, & \text{if } \frac{M_4}{M_2^2} \leq \frac{3\sqrt{3}-3}{2}. \end{cases}$$

Proof. Notice that if $M_1 = \mathsf{E}[X] = 0$, then $V_{\min} = \sqrt{(\sqrt{3}-1)M_2}$ and $\alpha = \frac{M_4 - M_2^2}{M_2}$. Therefore, The condition $\frac{3\sqrt{3}-3}{2} \le \frac{M_4}{M_2^2}$ is equivalent to $\alpha \ge \frac{\sqrt{3}-1}{2} V_{\min}$. The corollary follows by noting that

$$\max_{v>0}\left(\frac{3M_2}{v^2} - \frac{M_4}{v^4}\right) = \frac{9}{4}\cdot\frac{M_2^2}{M_4}.$$

□

By applying Corollary 2.9, we can obtain a non-trivial bound for the probability $X \ge a$ when $\mathsf{E}[X] = 0$, given the information on M_2 and M_4.

Corollary 2.10. *If* $\mathsf{E}[X] = 0$ *and* $a \ge 0$*, then*

$$\text{Prob}\{X \ge a\} \le 1 - (2\sqrt{3}-3)\frac{(M_2 + a^2)^2}{M_4 + 6a^2 M_2 + a^4}.$$

Proof. Let Y be a random random variable independent to X, and X takes only one of the two values, a or $-a$, each with probability half. Let $Z = X + Y$. Then

$$\begin{aligned} \mathsf{E}[Z] &= 0 \\ \mathsf{E}[Z^2] &= \mathsf{E}[X^2] + \mathsf{E}[Y^2] = M_2 + a^2 \\ \mathsf{E}[Z^4] &= \mathsf{E}[X^4] + 6\mathsf{E}[X^2]a^2 + a^4 = M_4 + 6a^2 M_2 + a^4. \end{aligned}$$

Then, by Corollary 2.9,

$$\text{Prob}\{Z \ge 0\} \le 1 - (2\sqrt{3}-3)\frac{(M_2 + a^2)^2}{M_4 + 6a^2 M_2 + a^4}.$$

However,

$$\text{Prob}\{Z \ge 0\} = \frac{\text{Prob}\{X \ge a\} + \text{Prob}\{X \ge -a\}}{2} \ge \text{Prob}\{X \ge a\}.$$

The desired inequality follows. □

The bound proved in Corollary 2.10 is not tight in general. A tight bound is summarized in the following theorem. Its proof, which is quite technical and similar to the proof of Theorem 2.8, is provided in the appendix.

Theorem 2.11. *Let* $K = M_4/M_2^2$ *and* $L = M_2/a^2$*. If* $\mathsf{E}[X] = 0$ *and* $a \ge 0$*, Then*

$$\text{Prob}\{X \ge a\} \le \begin{cases} \frac{M_2}{M_2 + a^2}, & \text{if } K \ge L + \frac{1}{L} - 1; \\ \frac{M_4 - M_2^2}{M_4 - 2M_2 a^2 + a^2}, & \text{if } K \le L + \frac{1}{L} - 1 \text{ and } L < 1; \\ \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{3 + M_4/M_2^2}}, & \text{if } K \le L + \frac{1}{L} - 1, L \ge 1 \text{ and } \\ & \frac{1}{\sqrt{L}} \ge \sqrt{K - 1 + \sqrt{K^2 + 2K - 3}} - \frac{\sqrt{K+3} - \sqrt{K-1}}{2}; \\ \min\{P(v) \mid v \ge a\}, & \text{otherwise} \end{cases}$$

where

$$P(v) = 1 - \frac{-M_4 + M_2(3v^2 + 2av + a^2) + a^2v^2 + 2av^3}{\frac{a^4}{4} + a^3v + 4a^2v^2 + 6av^3 + \frac{9}{4}v^4 + (3v^3 + 4av^2 + 2a^2v)\sqrt{\frac{v^2}{4} + \frac{(a+v)^2}{2}}}.$$

And the bounds are tight.

3 Small Deviation Bound for Sum of Independent Random Variables

In this section, we consider the problem of bounding the probability of small deviations for sum of independent random variables. In particular, consider n independent random variables $X_1, X_2, \cdots, X_n$ each with a mean of zero. Let $S = \sum_{i=1}^n X_i$. We are interested in the probability that $S < \Delta$ for some given constant Δ. For this purpose, we may directly apply Theorem 2.11. Then we need to estimate $\mathsf{E}[S^2]$ and $\mathsf{E}[S^4]$. We may also apply Theorem 2.8. In this case, we need to estimate $\mathsf{E}[(S-\Delta)^2]$ and $\mathsf{E}[(S-\Delta)^4]$. We demonstrate below how this could be done.

We consider two cases. In the first case, the random variables X_i are uniformly bounded from both sides. In the second case, we assume that the random variables are uniformly bounded only from below.

Given two nonnegative constants c_1 and c_2, define

$$s(c_1, c_2) := \max\{c_1^2 + 4c_1, c_2^2 - 4c_2, c_1^2 + c_2^2 - 4c_1c_2 - 4(c_2 - c_1)\}.$$

Our first result is summarized below.

Theorem 3.1. *Consider n independent random variables $X_1, X_2, \cdots, X_n$. Assume that $\Delta > 0$ is a given constant. Also assume that $\mathsf{E}[X_i] = 0$ and there exists two nonnegative constants c_1 and c_2 such that $-c_1\Delta \le X_i \le c_2\Delta$. Let $S = \sum_{i=1}^n X_i$, then*

$$\text{Prob}\{S < \Delta\} \ge F_1(\Delta, c_1, c_2) \ge F_2(c_1, c_2), \tag{3.14}$$

where

$$F_1(\Delta, c_1, c_2) = \frac{4(2\sqrt{3}-3)}{9} \inf_{D>0} \left(\sqrt{\frac{6(D\Delta^2 + \Delta^4)}{3D^2 + (6 + s(c_1,c_2))D\Delta^2 + \Delta^4}} + \frac{\frac{9}{4}(D + \Delta^2)^2}{3D^2 + (6 + s(c_1,c_2))D\Delta^2 + \Delta^4} \right)$$

and

$$F_2(c_1, c_2) = 4(2\sqrt{3}-3)\frac{s(c_1,c_2) + 2}{s(c_1,c_2)^2 + 12s(c_1,c_2) + 24}.$$

Proof. First of all, we can assume without loss of generality that X_i follows a two point distribution for every $i = 1, 2, \cdots, n$. In particular, given that $\mathsf{E}[X_i] = 0$, we assume that there exists $a_i, b_i \geq 0$, such that

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

It follows that $\mathsf{E}[X_i^2] = a_i b_i$ and $\mathsf{E}[X_i^4] = a_i b_i (a_i^2 - a_i b_i + b_i^2)$. Let denote the variance of S by D, i.e., $D = \sum_{i=1}^n \mathsf{E}[X_i^2] = \sum_{i=1}^n a_i b_i$. Therefore, $\mathsf{E}[(S-\Delta)^2] = D + \Delta^2$. Furthermore,

$$\begin{aligned} \mathsf{E}[(S-\Delta)^4] &= \mathsf{E}[S^4] - 4\Delta \mathsf{E}[S^3] + 6\Delta^2 \mathsf{E}[S^2] + \Delta^4 \\ &= \sum_{i=1}^n \mathsf{E}[X_i^4] + 6 \sum_{i<j} \mathsf{E}[X_i^2]\mathsf{E}[X_j^2] - 4\Delta \sum_{i=1}^n \mathsf{E}[X_i^3] + 6\Delta^2 \sum_{i=1}^n \mathsf{E}[X_i^2] + \Delta^4 \\ &= \sum_{i=1}^n \left(\mathsf{E}[X_i^4] - 4\Delta \mathsf{E}[X_i^3] - 3(\mathsf{E}[X_i^2])^2 \right) + 3D^2 + 6\Delta^2 D + M^4 \\ &= 3D^2 + 6\Delta^2 D + \Delta^4 + \sum_{i=1}^n a_i b_i (a_i^2 + b_i^2 - 4a_i b_i - 4\Delta(b_i - a_i)). \end{aligned}$$

Notice that $a_i^2 + b_i^2 - 4a_i b_i - 4\Delta(b_i - a_i)$ is a convex function of a_i when b_i is fixed, and is convex in b_i when a_i is fixed. Therefore, an optimal solution to the optimization problem

$$\max_{0 \leq a_i \leq c_1, 0 \leq b_i \leq c_2} (a_i^2 + b_i^2 - 4a_i b_i - 4\Delta(b_i - a_i))$$

is in the set $\{(0,0), (0,c_2), (c_1,0), (c_1,c_2)\}$. Thus, we conclude that

$$a_i^2 + b_i^2 - 4a_i b_i - 4\Delta(b_i - a_i) \leq s(c_1, c_2).$$

It then follows that

$$\mathsf{E}[(S-\Delta)^4] \leq 3D^2 + 6\Delta^2 D + \Delta^4 + s(c_1, c_2) D \Delta^2.$$

Thus by Theorem 2.8, we have for any $v > 0$ that

$$\mathrm{Prob}\{S - \Delta < 0\} \geq \frac{4}{9}(2\sqrt{3} - 3) \left(\frac{2\Delta}{v} + \frac{3(D + \Delta^2)}{v^2} - \frac{3D^2 + 6\Delta^2 D + \Delta^4 + s(c_1, c_2) D \Delta^2}{v^4} \right).$$

In particular, we choose v such that

$$v^{-2} = \frac{\frac{3}{2}\left(D + \Delta^2\right)}{3D^2 + (6 + s(c_1, c_2)) D \Delta^2 + \Delta^4}.$$

Then we must have

$$\begin{aligned} & \mathrm{Prob}\{S - \Delta < 0\} \\ \geq\ & \frac{4}{9}(2\sqrt{3} - 3) \left(\sqrt{\frac{6(D\Delta^2 + \Delta^4)}{3D^2 + (6 + s(c_1, c_2)) D \Delta^2 + \Delta^4}} + \frac{\frac{9}{4}(D + \Delta^2)^2}{3D^2 + (6 + s(c_1, c_2)) D \Delta^2 + \Delta^4} \right) \\ \geq\ & F_1(\Delta, c_1, c_2). \end{aligned}$$

Furthermore, it is clear that

$$
\begin{aligned}
&F_1(\Delta, c_1, c_2) \\
\geq\ & \inf_{D>0} (2\sqrt{3}-3)\left(\frac{(D+\Delta^2)^2}{3D^2+(6+s(c_1,c_2))D\Delta^2+\Delta^4}\right) \\
\geq\ & (2\sqrt{3}-3)\frac{4(s(c_1,c_2)+2)}{s(c_1,c_2)^2+12s(c_1,c_2)+24} \\
=\ & F_2(c_1,c_2),
\end{aligned}
$$

where the last inequality uses the fact that if we let $t = \frac{\Delta^2}{D+\Delta^2}$, then

$$
\frac{3D^2+(6+x)D\Delta^2+\Delta^4}{(D+\Delta^2)^2} = 3 + xt(1-t) - 2t^2 \leq 3 + \frac{x^2}{4(x+2)},
$$

for any $x > -2$. This completes the proof of the theorem. □

Now we consider the case where the random variables X_i are bounded from below only. We obtain a similar result as Theorem 3.1.

Theorem 3.2. *Consider n independent random variables $X_1, X_2, \cdots, X_n$. Assume that $\Delta > 0$ is a given constant. Also assume that $\mathsf{E}[X_i] = 0$ and there exists a constant $c > 0$ such that $X_i \geq -c\Delta$ for every i. Let $S = \sum_{i=1}^n X_i$, then for any $\tau > 0$,*

$$
\text{Prob}\{X < \Delta\} \geq e^{-1/\tau} F_1(\Delta, c, \tau \max(1,c)) \geq e^{-1/\tau} F_2(c, \tau\max(1,c)). \tag{3.15}
$$

Proof. Once again, we assume without loss of generality that there exist $a_i, b_i \geq 0$, such that

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

By assumption $a_i \leq c\Delta$. We also assume that without loss of generality that $b_1 \geq b_2 \geq \ldots \geq b_n$. We consider a fixed $\tau > 0$ and define

$$
N = \max\{0, \max\{k \mid b_k \geq \tau(a_1 + a_2 + \ldots + a_k), 1 \leq k \leq n\}\}.
$$

Let $a = \sum_{i=1}^N a_i$; if $N = 0$, then let $a = 0$. If $N < n$, then for every $i > N$,

$$
b_i \leq b_{N+1} \leq \tau \sum_{i=1}^{N+1} a_i \leq \tau(a + a_{N+1}) \leq \tau(a + c\Delta).
$$

For any $i \leq N$, $b_i \geq b_N \geq \tau a$. Thus, if $N > 0$, then

$$
\begin{aligned}
\text{Prob}\left\{\sum_{i=1}^N X_i = -a\right\} &= \prod_{i=1}^N \text{Prob}\{X_i = -a_i\} \\
&= \prod_{i=1}^N \left(1 - \frac{a_i}{a_i+b_i}\right) \geq \prod_{i=1}^N \left(1 - \frac{a_i}{a_i+\tau a}\right) \\
&\geq \prod_{i=1}^N e^{-a_i/(\tau a)} = e^{-1/\tau}.
\end{aligned}
$$

Let $Y = \sum_{i=N+1}^{n} X_i$. Because for each $i > N$, $a_i \le c\Delta \le c(a+\Delta)$ and $b_i \le \max(1,c)\tau(a+\Delta)$, by Theorem 3.1, we know that

$$\begin{aligned}
& \text{Prob}\{S < \Delta\} \\
\ge\ & \text{Prob}\left\{\sum_{i<N} X_i = -a\right\} \cdot \text{Prob}\{Y < a+\Delta\} \\
\ge\ & e^{-1/\tau} F_1(\Delta, c, \tau\max\{1,c\}) \\
\ge\ & e^{-1/\tau} F_2(c, \tau\max\{1,c\}).
\end{aligned}$$

The proof is completed. □

Theorem 3.2 generalizes an inequality that was proved by Feige [7]. In particular, if every X_i is non-negative with expectation 1, then Feige proved that

$$\text{Prob}\left\{\sum_{i=1}^{n} X_i \ge n+1\right\} \le \frac{12}{13}.$$

For this special case, Theorem 3.2 implies a stronger result than the above inequality.

Corollary 3.3. *Consider n independent random variables $X_1, X_2, \cdots, X_n$ each with mean zero. If $X_i \ge -1$ for all $i = 1, 2, \cdots, n$, then we have that*

$$\text{Prob}\left\{\sum_{i=1}^{n} X_i < 1\right\} \ge e^{-1/5}\frac{1}{3(2\sqrt{3}-3)} \ge \frac{1}{8}.$$

Proof. We can apply Theorem 3.2 with $c = 1$ and $\Delta = 1$. We choose $\tau = 5$ and thus $s(c,\tau) = 5$. In this case,

$$F_1(\Delta, c, \tau) = \inf_{D>0} \frac{4}{9}(2\sqrt{3}-3)\left(\sqrt{\frac{6(D+1)}{3D^2+11D+1}} + \frac{\frac{9}{4}(D+1)^2}{3D^2+11D+1}\right).$$

However,

$$\sqrt{\frac{6(D+1)}{3D^2+11D+1}} + \frac{\frac{9}{4}(D+1)^2}{3D^2+11D+1}$$

is a decreasing function of D when $D \ge 0$. By letting D go to infinity, we have that

$$F_1(\Delta, c_1, c_2) \ge \frac{4}{9}(2\sqrt{3}-3)\cdot\frac{3}{4} = \frac{2\sqrt{3}-3}{3}.$$

By Theorem 3.2, we have

$$\text{Prob}\{X < \Delta\} \ge e^{-1/c_2} F_1(\Delta, c_1, c_2) = e^{-1/5}\frac{(2\sqrt{3}-3)}{3},$$

which completes the proof for the corollary. □

It would be interesting to see how strong a bound can be obtained if we apply Markov, Chebyshev, and Bertsimas-Popescu's (three-moments) inequality to the problem considered in Corollary 3.3. Consider the following example, where all the X_is are i.i.d distribution which take value 0 and 2 with probability 1/2 of each. Then for the random variable $X = \sum_{i=1}^n X_i$, and $\delta = \frac{1}{n}$, we have $M_1 = n$, $M_2 = n^2 + n$, and $M_3 = n^3 + 3n^2$. Since $C_M^2 = \frac{1}{n} = \delta$, when $n \to \infty$, the value $f_1(C_M^2, D_M^2, \delta) = \frac{n}{n+1}$ approaches 1. Therefore the three moments inequality alone is not good enough to yield a good bound for the problem.

4 Applications

In many applications and rounding algorithms, the Chernoff type bounds or other similar inequalities can be applied to yield claims of the following spirit: If $n > N(\delta, \epsilon)$, then for n independent samples X_i ($1 \le i \le n$) of a random variable X it follows that $\text{Prob}\left\{\max_{1\le i\le n} X_i \le (1-\delta)\mathsf{E}X\right\} \le \epsilon$.

However it follows from our analysis, the δ can be dropped and we can claim the following:

Lemma 4.1. *If a random variable X has kurtosis $\kappa = \frac{E(X-EX)^4}{(E(X-EX)^2)^2} - 3$, then with $n = \frac{2\sqrt{3}+3}{3}(\kappa + 3)\log(\frac{1}{\epsilon})$ many samples, by Theorem 2.3,*

$$\text{Prob}\left\{\max_{1\le i\le n} X_i \le \mathsf{E}X\right\} \le \epsilon$$

and

$$\text{Prob}\left\{\min_{1\le i\le n} X_i \ge \mathsf{E}X\right\} \le \epsilon.$$

This is to say that, when a distribution's Kurtosis κ can be estimated or upper bounded, then $\Theta(\kappa \log(1/\epsilon))$ many samples would guarantee that we are able to draw one whose value is at least as good as the expected value of the distribution, with high probability $1 - \epsilon$.

Proof. Because for each i, $\text{Prob}\{X_i \le \mathsf{E}X\} \le 1 - (2\sqrt{3}-3)\frac{1}{\kappa+3}$, we have

$$\text{Prob}\left\{\max_{1\le i\le n} X_i \le \mathsf{E}X\right\} \le \left(1 - (2\sqrt{3}-3)\frac{1}{\kappa+3}\right)^n \le \exp\left(-n(2\sqrt{3}-3)\frac{1}{\kappa+3}\right).$$

Thus if $n \ge \frac{2\sqrt{3}+3}{3}(\kappa+3)\log(\frac{1}{\epsilon})$, we have $\text{Prob}\left\{\max_{1\le i\le n} X_i \le \mathsf{E}X\right\} \le \epsilon$. The other inequality is symmetric. □

Also, for sums of independent random variables we have the following:

Lemma 4.2. *If X_i are independent random variables with $\mathsf{E}X_i = 0$, $\mathsf{E}X_i^2 = D$, $\mathsf{E}X_i^4 \le (\kappa + 3)(\mathsf{E}X_i^2)^2$, then we have that*

$$\frac{2\sqrt{3}-3}{3+\frac{\kappa}{n}} \le \text{Prob}\left\{\sum_{i=1}^n X_i \ge 0\right\} \le 1 - \frac{2\sqrt{3}-3}{3+\frac{\kappa}{n}}.$$

Proof. Let $X = \sum_{i=1}^n X_i$, $D_X = \mathrm{Var}(X) = n\mathrm{Var}(X_i) = nD$, $\tau = (\kappa+3)D^2$. Then

$$\tau_X = \mathsf{E}X^4 = \sum_{i=1}^n \mathsf{E}X_i^4 + 6\sum_{i<j}\mathsf{E}X_i^2\mathsf{E}X_j^2 \le n\tau + 3D_X^2 - 3nD^2 = (3n^2 - 3n + n(\kappa+3))D^2.$$

Thus

$$\mathrm{Prob}\,\{X \ge \mathsf{E}X\} \le 1 - (2\sqrt{3}-3)\frac{D_X^2}{\tau_X} \le 1 - \frac{2\sqrt{3}-3}{3+\frac{\kappa}{n}}.$$

The other inequality follows by symmetry. □

Now we consider the weighted maximum cut problem. In this problem, we are given an undirected graph $G = (V, E)$ where each edge (u, v) has a weight $w_{u,v}$, and the goal is to partition the vertices of G into two sets S_1 and S_2 so as to maximize the total weight of the edges (u, v) such that $u \in S_1$ and $v \in S_2$. This problem is NP-hard, but admits a polynomial time 0.878-approximation algorithm; see Goemans and Williamson [6]. Prior to the celebrated result of Goemans and Williamson, the best known approximation ratio for the maximum cut problem was 1/2 for the weighted version, and $\frac{1}{2} + \frac{1}{2\delta}$ for the unweighted version, where δ denotes the maximum degree of a vertex.

It is well-known that, a simple 1/2-approximation algorithm can be obtained by independently and equiprobably assigning each vertex of G to either S_1 or S_2. Indeed, if we denote the total weight of edges with end-points in different sets by W, then it is clear that

$$\mathsf{E}[W] = \frac{1}{2}\sum_{(u,v)\in E} w_{u,v} := \frac{1}{2}W_{tot} \tag{4.16}$$

which of course is no less than half of the maximum weight.

Equation (4.16) has a stronger implication. That is, for any graph, there exists a cut so that weight of the cut is at least half of the total weight of the edges of the graph. However, two interesting questions remain:

- There are $O(2^{|V|})$ many cuts for a graph. Among all the possible cuts, how many of them have a weight larger than $W_{tot}/2$?
- Is it possible to show that there always exists a cut with a weight higher than αW_{tot} for some $\alpha > 1/2$?

When the graph is unweighted, the answer to the second question is "yes" with an $\alpha = \frac{1}{2} + \frac{1}{2n}$ and this bound is the best possible; see Haglin and Venkatesan [8]. The result is obtained by proving the existence of a matching of certain size, which also gives a linear time algorithm to find a cut with a weight larger than $(\frac{1}{2} + \frac{1}{2n})W_{tot}$.

Now we shall answer the above two questions for a general weighted graph by using the simple randomized algorithm described earlier, together with the moment bound developed in this paper.

We slightly formalizes the randomized algorithm as follows. We define $|V|$ independent random binary variables $X_1, \cdots, X_{|V|}$, so that for each node $i \in V$, X_i takes value 1 or -1 with probability half. Thus, $X_i = 1$ indicate node i is assigned to the set S_1, and vice versa. Then we have

$$W = \sum_{i<j} w_{i,j} \frac{1 - X_i X_j}{2}.$$

For convenience, we also define

$$Y = W - \frac{1}{2} W_{tot}$$

so that $\mathsf{E}[Y] = 0$. We now estimate the second and the fourth moments of random variable Y.

$$\mathsf{E}[Y^2] = \frac{1}{4}\mathsf{E}\left[\left(\sum_{i<j} w_{i,j} X_i X_j\right)^2\right] = \frac{1}{4}\sum_{i<j} w_{i,j}^2 \geq \frac{1}{2|V|^2} W_{tot}^2. \tag{4.17}$$

It can also be shown that

$$\mathsf{E}[Y^4] \leq 15(\mathsf{E}[Y^2])^2. \tag{4.18}$$

Therefore, it follows immediately from Corollary 2.9 that

$$\text{Prob}\left\{W \geq \frac{1}{2}W_{tot}\right\} = \text{Prob}\left\{-Y \leq 0\right\} \geq (2\sqrt{3} - 3)\frac{(\mathsf{E}[Y^2])^2}{\mathsf{E}[Y^4]} \geq \frac{2\sqrt{3} - 3}{15}.$$

Denote $\Delta = \left(\mathsf{E}[Y^2]/\sqrt{15}\right)^{1/2} > \sqrt{\mathsf{E}[Y^2]}/2$ and let $Z = t\Delta - Y$ with $t \geq 0$. Then we have $\mathsf{E}[Z] = t\Delta$, $\mathsf{E}[Z^2] = \mathsf{E}[Y^2] + t^2\Delta^2$, and

$$\begin{aligned}
\mathsf{E}[Z^4] &= \mathsf{E}[Y^4] - 4t\Delta\mathsf{E}[Y^3] + 6t^2\Delta^2\mathsf{E}[Y^2] + t^4\Delta^4 \\
&\leq \mathsf{E}[Y^4] + 4t\Delta(\mathsf{E}[Y^2])^{1/2}(E[Y^4])^{1/2} + 6t^2\Delta^2\mathsf{E}[Y^2] + t^4\Delta^4 \\
&\leq \mathsf{E}[Y^4] + 4\sqrt{15}t\Delta(\mathsf{E}[Y^2])^{3/2} + 6t^2\Delta^2\mathsf{E}[Y^2] + t^4\Delta^4 \\
&\leq 15\mathsf{E}[Y^2]^2 + 4\sqrt{15}t\Delta(\mathsf{E}[Y^2])^{3/2} + 6t^2\Delta^2\mathsf{E}[Y^2] + t^4\Delta^4 \\
&= \left(15 + 4(15)^{1/4}t + \frac{6}{\sqrt{15}}t^2 + \frac{1}{15}t^4\right)(\mathsf{E}[Y^2])^2.
\end{aligned}$$

Thus, by Theorem 2.8, we have, for any $v > 0$,

$$\text{Prob}\left\{Y \geq t\Delta\right\} = \text{Prob}\left\{Z \leq 0\right\} \geq \frac{4}{9}(2\sqrt{3} - 3)\left(-\frac{2\mathsf{E}[Z]}{v} + \frac{3\mathsf{E}[Z^2]}{v^2} - \frac{\mathsf{E}[Z^4]}{v^4}\right).$$

In particular, if we choose $v = 10\Delta$ and $t = 0.01$, then

$$\text{Prob}\left\{Y \geq t\Delta\right\} > 1.2\%.$$

It follows that

$$\text{Prob}\left\{W \geq \left(\frac{1}{2} + \frac{0.0036}{|V|}\right)W_{tot}\right\} > 1.2\%.$$

To summarize, we have proved the following:

Theorem 4.3. *For any weighted graph, the following two statements are true.*

1). Among all possible cuts of the graph, at least $\frac{2\sqrt{3}-3}{15} > 3\%$ *of them will have a cut value larger than half of the total weight of the edges of the graph.*

2). There exists a cut whose weight is at least $\left(\frac{1}{2} + \frac{0.0036}{|V|}\right)$ *times the total weight of the edges of the graph.*

References

[1] A. Ben-Tal and A. Nemirovski. Lectures on Modern Convex Optimization: Analysis, Algorithm, and Engineering Applications. *MPS/SIAM Ser. Optim.2, SIAM, Philadelphia*, 2001.

[2] B. Berger. The Fourth Moment Method. *SIAM Journal on Computing* 26, pp. 1188 – 1207, 1999.

[3] F.P. Cantelli. Intorno ad un teorema fundamentale della teoria del rischio. *Boll. Assoc. Attuar. Ital. (Milan)*, pp. 1 – 23, 1910.

[4] D. Bertsimas and I. Popescu. Optimal Inequality in Probability Theory: A Convex Optimization Approach. *SIAM Journal on Optimzation*, 2005.

[5] D. Bertsimas and I. Popescu. On the Relation Between Option and Stock prices: A Convex Optimization Approach. *Operation Research 50 No.2* , pp. 358 – 374, 2002.

[6] M.X. Goemans and D.P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems using Semidefinite Programming. *Journal of the ACM* 42, pp. 1115 – 1145, 1995.

[7] U. Feige. On Sums of Independent Random Variables with Unbounded Variances, and Estimating the Average Degree in a Graph. *SICOMP*, (2006).

[8] D.J. Haglin and S.M. Venkatesan. Approximation and Intractability Results for the Maximum Cut Problem and Its Variants. *IEEE Transactions on Computers* 40, pp. 110 – 113, 1991.

[9] S. He, Z.Q. Luo, J. Nie, and S. Zhang. Semidefnite Relaxation Bounds for Indefinite Homogeneous Quadratic Optimization. *Working Paper*, 2007.

[10] K. Isii. The extrema of probaiblity determined by generalized moments. I. Bounded random variables. *Ann. Insti. Statist. Math.*, 12, pp. 164 – 168, 1960.

[11] S. Karlin and W.J. Studden. Tchebysheff Systems: With Applications in Analysis and Statistics. *Pure Appl. Math. 15*, Interscience, John Wiley and Sons, New York, 1966.

[12] J.B. Lasserre. Bounds on measures satisfying moment conditions. *Ann. Appl. Probab.* 12, pp. 1114 – 1137, 2002.

[13] R. Von Mises. The Limits of a Distribution Function if Two Expected Values Are Given. *Ann. Math. Statist.* 10, pp. 99 – 104, 1939.

[14] Yu. Nesterov. Structure of Non-Negative Polynomial and Optimization Problems. *Preprint DP 9749, Louvain-la-Neuve, Belgium*, 1997.

[15] I. Popescu. A Semidefinite Programming Approach to Optimal Moment Bounds for Convex Classes of Distributions. *Mathematics of Operation Research*, 2005.

[16] J. Smith. Generalized Chebyshev Inequalities: Theory and Applications in Decision Analysis. *Operations Research*, 43, 807 – 825, 1995.

[17] M. Zelen. Bounds on a Distribution Function That Are Functions of Moments to Order Four. *J. Res. Nat. Bur. Stand.*, 53, pp. 377 – 381, 1954.

A Proof of Theorem 2.11

For Theorem 2.11, the primal problem is

$$\begin{aligned} Z_P^2 \;=\; \max \;\; & \text{Prob}\{X \geq a\} \\ \text{s.t.} \;\; & \mathsf{E}[X] = 0 \\ & \mathsf{E}[X^2] = M_2 \\ & \mathsf{E}[X^4] = M_4, \end{aligned}$$

or equivalently

$$\begin{aligned} Z_P^2 \;=\; \max_{F(\cdot)} \;\; & \int_{x\geq a} 1 \cdot dF(x) \\ \text{s.t.} \;\; & \int_{x\in\mathbb{R}} 1 \cdot dF(x) = 0 \\ & \int_{x\in\mathbb{R}} x \cdot dF(x) = M_1 \\ & \int_{x\in\mathbb{R}} x^2 \cdot dF(x)\mu = M_2 \\ & \int_{x\in\mathbb{R}} x^4 \cdot dF(x) = M_4. \end{aligned} \tag{A.19}$$

Its dual problem in this setting can be written as

$$\begin{aligned} Z_D^2 \;=\; \min \;\; & y_0 + 0 \cdot y_1 + M_2 \cdot y_2 + M_4 \cdot y_4 \\ \text{s.t.} \;\; & g(x) := y_0 + y_1 \cdot x + y_2 \cdot x^2 + y_4 \cdot x^4 \geq \mathbf{1}_{\{x\geq a\}}, \quad \forall x \in \mathbb{R}. \end{aligned} \tag{A.20}$$

Lemma A.1. *For any $u > v > a$, let $c = \frac{1}{(u+v)^3(u-v)} > 0$ and $d = \frac{2v}{(u+v)^3} > 0$. Define*

$$\begin{aligned} y_0 &= cu^4 + du^2, \\ y_1 &= 2du, \\ y_2 &= d - 2cu^2, \\ y_4 &= c. \end{aligned} \tag{A.21}$$

Then (y_0, y_1, y_2, y_4) is feasible to problem (A.20) if $u \leq \frac{v}{2} + \sqrt{\frac{v^2}{4} + \frac{(a+v)^2}{2}}$.

Proof. If (y_0, y_1, y_2, y_3) is defined by (A.21), then

$$g(x) = cx^4 + (d - 2cu^2)x^2 + 2dux + cu^4 + du^2 = c(x^2 - u^2)^2 + d(x+u)^2.$$

It is clear that $g(x) \geq 0$ for all $x \in \mathbb{R}$. It is left to verify that $g(x) \geq 1$ for all $x \geq a$. The condition $u \leq \frac{v}{2} + \sqrt{\frac{v^2}{4} + \frac{(a+v)^2}{2}}$ implies that

$$2u^2 - 2uv \leq (v+a)^2.$$

We first observe that

$$\begin{aligned} & \frac{g(a) - 1}{c} \\ = \quad & (a^2 - u^2)^2 + (a+u)^2(2uv - 2v^2) - (u+v)^3(u-v) \\ = \quad & a^4 - 2a^2(u^2 - uv + v^2) + 4av(u^2 - uv) + v^2(v^2 + 2uv - 2u^2) \\ = \quad & (a^2 - v^2)^2 - 2(u^2 - uv)(v-a)^2 \\ \geq \quad & (a^2 - v^2)^2 - (v+a)^2(v-a)^2 \\ = \quad & 0. \end{aligned}$$

Therefore $g(a) \geq 1$. Notice that,

$$g(x) = (x+u)^2 \cdot \left(c(x-u)^2 + d\right).$$

Since $c(x-u)^2 + d > 0$, we have $g(x) = 0$ if and only if $x = -u < 0$. Thus $x = -u < 0$ is the only global minimum solution of $g(x)$, and thus one of the local minimum solutions. Since $g(x)$ is a polynomial with order four, it has at most two local minimum solutions, including $x = -u$. We denote the other local minimum solution by z. If $z < a$, then we must have that $g(x)$ is increasing for $x \geq a$, and thus $g(x) \geq g(a) \geq 1$. Therefore, we assume that $z > 0 > -u$, and z must be the largest root to $g'(x) = 0$. But

$$g'(x) = 4cx(x^2 - u^2) + 2d(x+u)$$

and the largest root to $g'(x) = 0$ is $\frac{u}{2} + \sqrt{\frac{u^2}{4} - \frac{d}{2c}}$. Therefore,

$$z = \frac{u}{2} + \sqrt{\frac{u^2}{4} - \frac{d}{2c}} = \frac{u}{2} + \sqrt{\frac{u^2}{4} - v(u-v)} = \frac{u}{2} + \left|v - \frac{u}{2}\right| = v.$$

The last equality holds since

$$u \leq \frac{v}{2} + \sqrt{\frac{v^2}{4} + \frac{(a+v)^2}{2}} \leq 2v.$$

Now it is straightforward to verify that $g(z) = g(v) = 1$.

Finally, we observe that the global minimum solution to $g(x)$ in $[0, \infty)$ is either $x = 0$ or $x = z$. Therefore, $g(x) \geq g(0) = g(z) = 1$ for all $x \geq 0$. This completes the proof. □

Proof of Theorem 2.11. When $M_4 = M_2^4$, the distribution has to be

$$X = \begin{cases} -\sqrt{M_2}, & \text{with probability } \frac{1}{2}; \\ \sqrt{M_2}, & \text{with probability } \frac{1}{2}, \end{cases}$$

Since

$$\text{Prob}\{X \geq a\} = \begin{cases} 0, & \text{if } L < 1; \\ \frac{1}{2}, & \text{if } L \geq 1, \end{cases}$$

Theorem 2.11 holds under this condition.

If $M_4 > M_2^2$, the given condition of the moment information implies that the strong duality holds. Thus problem (A.19) is equivalent to

$$\begin{array}{rcll} Z_P^2 & = & \max_{F(\cdot)} & \int_{x \geq a} 1 \cdot dF(x) \\ & & \text{s.t.} & \int_{x \in \mathbb{R}} 1 \cdot dF(x) = 0 \\ & & & \int_{x \in \mathbb{R}} x \cdot dF(x) = M_1 \\ & & & \int_{x \in \mathbb{R}} x^2 \cdot dF(x) = M_2 \\ & & & \int_{x \in \mathbb{R}} x^4 \cdot dF(x) \leq M_4. \end{array} \tag{A.22}$$

Case 1: When $K \geq L + \frac{1}{L} - 1$, we define

$$X = \begin{cases} -\frac{M_2}{a}, & \text{with probability } \frac{1}{L+1}; \\ a, & \text{with probability } \frac{L}{L+1}, \end{cases}$$

which is always well defined. Furthermore,

$$\begin{array}{rcl} \mathsf{E}[X] & = & -\frac{M_2}{a}\frac{1}{L+1} + a\frac{L}{L+1} = -aL\frac{1}{L+1} + a\frac{L}{L+1} = 0, \\ \mathsf{E}[X^2] & = & \frac{M_2^2}{a^2}\frac{1}{L+1} + a^2\frac{L}{L+1} = a^2L^2\frac{1}{L+1} + a^2\frac{L}{L+1} = a^2L = M_2, \\ \mathsf{E}[X^4] & = & a^4L^4\frac{1}{L+1} + a^4\frac{L}{L+1} = a^4L(L^2 - L + 1) = \frac{M_2^2}{L}(L^2 - L + 1) \leq M_2^2K = M_4. \end{array}$$

Therefore X is feasible to problem (A.22) with objective value $\text{Prob}\{X \geq a\} = L/(L+1)$.

We now define

$$g(x) = \left(\frac{ax + M_2}{a^2 + M_2}\right)^2,$$

which is feasible to problem (A.20). The corresponding dual objective value is

$$\left(\frac{a}{a^2 + M_2}\right)^2 M_2 + \left(\frac{M_2}{a^2 + M_2}\right)^2 = \frac{L}{(L+1)^2} + \frac{L^2}{(L+1)^2} = \frac{L}{L+1}.$$

Therefore, when $K \geq L + \frac{1}{L} - 1$ the inequality in Theorem 2.11 holds and is tight.

Case 2: When $L < 1$ and $K \leq L + \frac{1}{L} - 1$, define

$$g(x) = \left[\frac{a^2 - M_2}{a^4 - 2a^2M_2 + M_4}(x^2 - a^2) + 1\right]^2,$$

which is a feasible solution of problem (A.20). The corresponding dual objective value is

$$
\begin{aligned}
& \left(\frac{a^2-M_2}{a^4-2a^2M_2+M_4}\right)^2 M_4 + 2\frac{(a^2-M_2)(M_4-a^2M_2)}{(a^4-2a^2M_2+M_4)^2}M_2 + \frac{(M_4-a^2M_2)^2}{(a^4-2a^2M_2+M_4)^2} \\
= & \left(\frac{a^2-M_2}{a^4-2a^2M_2+M_4}\right)^2 (M_4-M_2^2) + \left(\frac{a^2-M_2}{a^4-2a^2M_2+M_4}M_2 + \frac{M_4-a^2M_2}{a^4-2a^2M_2+M_4}\right)^2 \\
= & \frac{M_4-M_2^2}{(a^4-2a^2M_2+M_4)^2}\left((M_4-M_2^2+(a^2-M_2)^2\right) \\
= & \frac{M_4-M_2^2}{a^4-2a^2M_2+M_4}.
\end{aligned}
$$

Now we define $v=\sqrt{\frac{a^2M_2-M_4}{a^2-M_2}}$, $p=\frac{M_4-M_2^2}{M_4-2M_2a^2+a^4}$ and

$$
X = \begin{cases} v, & \text{with probability } q=\frac{1-p(1+a/v)}{2}; \\ -v, & \text{with probability } r=\frac{1-p(1-a/v)}{2}; \\ a, & \text{with probability } p. \end{cases}
$$

Since

$$
a^2M_2-M_4=\frac{1}{L}M_2^2-KM_2^2\geq(1-L)M_2^2>0,
$$

the value v is well defined. We observe that

$$
p=\frac{M_4-M_2^2}{M_4-2M_2a^2+a^4}=\frac{M_4-M_2^2}{M_4-M_2^2+(M_2-a^2)^2},
$$

thus $0<p<1$. The condition $L<1$ and $K\leq L+\frac{1}{L}-1$ implies that

$$
\begin{aligned}
& a^2(M_4-M_2^2)^2+(a^2-M_2)^3M_4 \\
\leq & M_2^5\left(\frac{(K-1)^2}{L}+K\frac{(1-L)^3}{L^3}\right) \\
\leq & M_2^5\left(\frac{(1-L)^4}{L^3}+\frac{(1-L)^3}{L^4}(L^2-L+1)\right) \\
= & M_2^5\frac{(1-L)^3}{L^4} \\
= & a^2M_2(a^2-M_2)^3,
\end{aligned}
$$

which is equivalent to

$$
\frac{a^2}{v^2}=\frac{a^2(a^2-M_2)}{a^2M_2-M_4}\leq\left(\frac{(a^2-M_2)^2}{M_4-M_2^2}\right)^2=\left(\frac{1}{p}-1\right)^2.
$$

Since

$$
1-\frac{a}{v}\leq 1+\frac{a}{v}\leq\frac{1}{p},
$$

we have $q, r \geq 0$. Notice that $p+q+r=1$, the distribution X is well defined. Furthermore,

$$\begin{aligned}
\mathsf{E}[X] &= (q-r)v+pa = -pa+pa = 0, \\
\mathsf{E}[X^2] &= (q+r)v^2+pa^2 = (1-p)v^2+pa^2 = \frac{(a^2-M_2)(a^2M_2-M_4)+a^2(M_4-M_2^2)}{a^4-2a^2M_2+M_4} = M_2, \\
\mathsf{E}[X^4] &= (q+r)v^4+pa^4 = \frac{(a^2M_2-M_4)^2+a^4(M_4-M_2^2)}{a^4-2a^2M_2+M_4} = M_4.
\end{aligned}$$

Thus X is feasible to problem (A.19). Since

$$v = \sqrt{a^2 - \frac{a^4+M_4-2a^2M_2}{a^2-M_2}} < a,$$

the corresponding dual objective value is

$$\text{Prob}\{X \geq a\} = p = \frac{M_4 - M_2^2}{M_4 - 2M_2a^2 + a^4}.$$

Therefore, the inequality in Theorem 2.11 is tight when $L < 1$ and $K \leq L + \frac{1}{L} - 1$.

Case 3: When $L \geq 1$, $K \leq L + \frac{1}{L} - 1$ and $\frac{1}{\sqrt{L}} \geq \sqrt{K-1+\sqrt{K^2+2K-3}} - \frac{\sqrt{K+3}-\sqrt{K-1}}{2}$, define

$$\begin{cases}
u &= \frac{\sqrt{\frac{M_4}{M_2}+3M_2}+\sqrt{\frac{M_4}{M_2}-M_2}}{2} = \sqrt{M_2}\frac{\sqrt{K+3}+\sqrt{K-1}}{2} \\
v &= \frac{\sqrt{\frac{M_4}{M_2}+3M_2}-\sqrt{\frac{M_4}{M_2}-M_2}}{2} = \sqrt{M_2}\frac{\sqrt{K+3}-\sqrt{K-1}}{2} \\
p &= \frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{3+K}} = \frac{\sqrt{K+3}+\sqrt{K-1}}{2\sqrt{K+3}} \\
q &= \frac{1}{2}-\sqrt{\frac{1}{4}-\frac{1}{3+K}} = \frac{\sqrt{K+3}-\sqrt{K-1}}{2\sqrt{K+3}}.
\end{cases}$$

Because $M_4 \geq M_2^2$ for any distribution, these values are well defined. It follows from definition that $u > v > 0$. From the assumption $K \leq L + \frac{1}{L} - 1$ and $L \geq 1$, we have

$$\begin{aligned}
v &= \frac{\sqrt{K+3}-\sqrt{K-1}}{2}\sqrt{M_2} \\
&= \frac{2}{\sqrt{K+3}+\sqrt{K-1}}\sqrt{M_2} \\
&\geq \frac{2}{\frac{L+1}{\sqrt{L}}+\frac{L-1}{\sqrt{L}}}\sqrt{M_2} \\
&= \frac{\sqrt{M_2}}{\sqrt{L}} \\
&= a.
\end{aligned}$$

The assumption $\frac{1}{\sqrt{L}} \geq \sqrt{K-1+\sqrt{K^2+2K-3}} - \frac{\sqrt{K+3}-\sqrt{K-1}}{2}$ implies that

$$\left(\frac{\sqrt{K+3}-\sqrt{K-1}}{2}+\frac{1}{\sqrt{L}}\right)^2 \geq K-1+\sqrt{K^2+2K-3} = (\sqrt{K+3}+\sqrt{K-1})\sqrt{K-1}.$$

Therefore

$$2u(u-v) \le (a+v)^2,$$

which implies that $u \le \frac{v}{2} + \sqrt{\frac{v^2}{4} + \frac{(a+v)^2}{2}}$. It follows from Lemma A.1 that the function

$$g(x) = c(x^2 - u^2)^2 + d(x+u)^2 = cx^4 + (d - 2cu^2)x^2 + 2dux + cu^4 + du^2$$

with

$$c = \frac{1}{(u+v)^3(u-v)} = \frac{1}{(K+3)M_2^2\sqrt{(K+3)(K-1)}} > 0$$

and

$$d = \frac{2v}{(u+v)^3} = \frac{\sqrt{K+3} - \sqrt{K-1}}{M_2(K+3)\sqrt{K+3}} > 0$$

defines a feasible solution to problem (A.20). Denote $t = \sqrt{K^2 + 2K - 3}$, then $d = cM_2(t - K + 1)$ and $u^2 = \frac{K+1+t}{2}$. The corresponding dual objective value is

$$\begin{aligned}
& cM_4 + (d - 2cu^2)M_2 + cu^2 + du^2 \\
= \;& cM_2^2 \left(K + t - K + 1 - (K+1+t) + \frac{(K+1+t)^2}{4} + \frac{K+1+t}{2}(t - K + 1) \right) \\
= \;& cM_2^2 \left(\frac{3}{4}t^2 + \frac{K+3}{2}t - \frac{K^2 + 2K - 3}{4} \right) \\
= \;& cM_2^2 t \frac{t + K + 3}{2} \\
= \;& \frac{1}{2} + \frac{t}{2(K+3)} \\
= \;& p.
\end{aligned}$$

Now we define

$$X = \begin{cases} -u, & \text{with probability } q; \\ v, & \text{with probability } p, \end{cases}$$

which is always feasible since $K \ge 1 > 0$. Furthermore,

$$\begin{aligned}
\mathsf{E}[X] &= pv - qu = 0, \\
\mathsf{E}[X^2] &= pv^2 + qu^2 = pv(u+v) = M_2, \\
\mathsf{E}[X^4] &= pv^4 + qu^4 = pv(u+v)(u^2 - uv + v^2) = M_2^2 K = M_4.
\end{aligned}$$

Therefore, the inequality in Theorem 2.11 is tight when $L \ge 1$, $K \le L + \frac{1}{L} - 1$ and

$$\frac{1}{\sqrt{L}} \ge \sqrt{K - 1 + \sqrt{K^2 + 2K - 3}} - \frac{\sqrt{K+3} - \sqrt{K-1}}{2}.$$

Case 4: Now we consider the case when $L \ge 1$, $K \le L + \frac{1}{L} - 1$ and

$$\frac{1}{\sqrt{L}} < \sqrt{K - 1 + \sqrt{K^2 + 2K - 3}} - \frac{\sqrt{K+3} - \sqrt{K-1}}{2}.$$

Our main goal is to prove there exists a $\hat{v} \geq a$ and the corresponding $\hat{u}$, so they can generate a feasible dual solution, and also satisfies the following conditions (which are the crucial conditions for the feasibility of the primal solution, and for the primal objective to match the dual objective value):

1.
$$a\hat{u} \leq M_2 \leq \hat{v}\hat{u};$$

2.
$$-M_4 + M_2(3\hat{v}^2 + 2a\hat{v} + a^2) + a^2\hat{v}^2 + 2a\hat{v}^3 = \frac{(\hat{u}+\hat{v})^2(\hat{u}-\hat{v})(M_2 + a\hat{v})}{a+\hat{u}}.$$

Define

$$\begin{cases} W(K) &= \sqrt{K-1+\sqrt{K^2+2K-3}} - \frac{\sqrt{K+3}-\sqrt{K-1}}{2}; \\ S(K) &= \sqrt{K-1+\sqrt{K^2+2K-3}}; \\ V(K) &= \frac{\sqrt{K+3}-\sqrt{K-1}}{2}; \\ U(K) &= \frac{\sqrt{K+3}+\sqrt{K-1}}{2}; \\ u(v) &= \frac{v}{2} + \sqrt{\frac{v^2}{4} + \frac{(a+v)^2}{2}}; \\ t(v) &= \sqrt{\frac{v^2}{4} + \frac{(a+v)^2}{2}}. \end{cases}$$

Because $K \geq 1$, there exists a unique $b \geq 1$ such that $K = b^2 + \frac{1}{b^2} - 1$, and

$$\sqrt{K-1+\sqrt{K^2+2K-3}} - \frac{2}{\sqrt{K+3}+\sqrt{K-1}} = \sqrt{2(b^2-1)} - \frac{1}{b},$$

which is a monotonically increasing function of b. Since $L + \frac{1}{L} - 1 \geq K = b^2 + \frac{1}{L^2} - 1$, and $L, b^2 \geq 1$, we can conclude that $L \geq b^2$. If $b^2 \geq 2$, then $KL \geq b^2(b^2 + \frac{1}{b^2} - 1) \geq 3$. If $b^2 < 2$, then

$$KL \geq \frac{b^2 + \frac{1}{b^2} - 1}{(\sqrt{2(b^2-1)} - \frac{1}{b})^2} \geq 3.$$

Therefore, the assumptions $L \geq 1$, $K \leq L + \frac{1}{L} - 1$ and $\frac{1}{\sqrt{L}} < W(K)$ guarantees that $KL \geq 3$, and $L \geq 2$. Also, since b is monotonically increasing to K, we have that $W(K)$ is a monotonically increasing function of K. Since $W(K)$ is a continuous function with $W(1) = -1$, there exists a $1 \leq K_0 < K$ such that $W(K_0) = \frac{1}{\sqrt{L}}$.

From Lemma A.1, for any $v \geq a > 0$, by abusing the notation, let $u = u(v)$ and $t = t(v)$, the function $g_v(x) = c(x^2 - u^2)^2 + d(x+u)^2$ with $d = \frac{2v}{(u+v)^3}$ and $c = \frac{1}{(u+v)^3(u-v)}$ is feasible for problem (A.20). Notice that $d = 2cv(u-v)$ and $u^2 - uv = \frac{(a+v)^2}{2}$, the corresponding dual objective value is:

$$\begin{aligned} P(v) &= \frac{M_4 + M_2(2v(u-v) - 2u^2) + u^4 + 2v(u-v)u^2}{(u+v)^3(u-v)} \\ &= 1 - \frac{-M_4 + M_2(3v^2 + 2av + a^2) + a^2v^2 + 2av^3}{(u+v)^3(u-v)}. \end{aligned}$$

Let

$$f(v) = -M_4 + M_2(3v^2 + 2av + a^2) + a^2v^2 + 2av^3$$

and

$$h(v) = (u+v)^3(u-v),$$

then we have

$$f'(v) = 2(3v + a)(M_2 + av),$$

and

$$\begin{aligned} h'(v) &= (u+v)^3(u'-1) + 3(u+v)^2(u-v)(u'+1) \\ &= (u+v)^2(2(2u-v)u' + 2u - 4v) \\ &= (u+v)^2\left(4\sqrt{\frac{v^2}{4}+\frac{(a+v)^2}{2}}\left(\frac{1}{2}+\frac{1}{4}\frac{2a+3v}{\sqrt{\frac{v^2}{4}+\frac{(a+v)^2}{2}}}\right)+v+2\sqrt{\frac{v^2}{4}+\frac{(a+v)^2}{2}}-4v\right) \\ &= 2(u+v)^2\left(a+2t(v)\right). \end{aligned}$$

Therefore,

$$\begin{aligned} P'(a) &= \frac{h'(a)f(a) - h(a)f'(a)}{h^2(a)} \\ &= \frac{72a^3(-M_4 + 6a^2M_2 + 3a^4) - 27a^4 \times 8a(M_2 + a^2)}{h^2(a)} \\ &= \frac{72a^3(3a^2M_2 - M_4)}{h^2(a)} \\ &= \frac{72a^5M_2(3 - KL)}{h^2(a)} \\ &\leq 0 \end{aligned}$$

Also for all v,

$$(u+v)(a+2t) - (3v+a)(a+u) = (\frac{3}{2}v + t)(a+2t) - (3v+a)(a + \frac{v}{2} + t) = 2t^2 - a^2 - 2va - \frac{3}{2}v^2 = 0.$$

Let

$$\begin{cases} v_0 = V(K_0)\sqrt{M_2}; \\ u_0 = U(K_0)\sqrt{M_2}. \end{cases}$$

Then because

$$a = \frac{1}{\sqrt{L}}\sqrt{M_2} = W(K_0)\sqrt{M_2},$$

we have

$$u_0(u_0 - v_0) = M_2\sqrt{K_0 - 1}\frac{\sqrt{K_0+3}+\sqrt{K_0-1}}{2} = \frac{(a+v_0)^2}{2}.$$

Since $u_0 \geq v_0$, we have that

$$u_0 = \frac{v_0}{2} + \sqrt{\frac{v_0^2}{4} + \frac{(a+v_0)^2}{2}} = u(v_0).$$

Notice that $a+v_0 = S(K_0)\sqrt{M_2}$ and $a+2t(v_0) = a+2u_0-v_0 = (S(K_0)+2\sqrt{K_0-1})\sqrt{M_2}$, we have

$$\begin{aligned}
& (f(v_0)+M_4-K_0M_2^2)h'(v_0)-h(v_0)f'(v_0) \\
= \;& 2(u_0+v_0)^2M_2^{5/2}\Big[(S(K_0)+2\sqrt{K_0-1})\left(-K_0+2V(K_0)^2+S(K_0)^2+V(K_0)^2(S(K_0)^2-V(K_0)^2)\right) \\
& -(U(K_0)^2-V(K_0)^2)(2V(K_0)+S(K_0))V(K_0)(U(K_0)+W(K_0))\Big] \\
= \;& 2(u_0+v_0)^2M_2^{5/2}\Big[(S(K_0)+2\sqrt{K_0-1})\sqrt{K_0^2+2K_0-3}\sqrt{K_0+3}V(K_0) \\
& -\sqrt{K_0^2+2K_0-3}V(K_0)(S(K_0)+\sqrt{K_0-1})(S(K_0)+\sqrt{K_0+3}-\sqrt{K_0-1})\Big] \\
= \;& 2(u_0+v_0)^2M_2^{5/2}V(K_0)\sqrt{K_0^2+2K_0-3}\Big[(S(K_0)+2\sqrt{K_0-1})\sqrt{K_0+3} \\
& -(S(K_0)+\sqrt{K_0-1})(S(K_0)+\sqrt{K_0+3}-\sqrt{K_0-1})\Big] \\
= \;& 0.
\end{aligned}$$

Therefore,

$$P'(v_0) = \frac{h'(v_0)f(v_0)-h(v_0)f'(v_0)}{h^2(v_0)} = (K_0-K)\frac{h'(v_0)M_2^2}{h^2(v_0)} < 0.$$

Define $v_1 = (\sqrt{3L^2+2L}-L-1)a$, and the corresponding $u_1 = u(v_1) = La$. Let $\sqrt{3L^2+2L} = r$, we have that

$$\begin{aligned}
& f(v_1)+M_4-(L+\frac{1}{L}-1)M_2^2 \\
= \;& a^4\left[-(L^3-L^2+L)+L\left(12L^2+12L+3-6(L+1)r+2r-2(L+1)+1\right)\right. \\
& \left.+\left(4L^2+4L+1-2(L+1)r\right)(2r-2L-1)\right] \\
= \;& a^4\left((6L^2+10L+4)r-(9L^3+21L^2+13L+1)\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
& (f(v_1)+M_4-(L+\frac{1}{L}-1)M_2^2)h'(v_1)-h(v_1)f'(v_1) \\
= \;& 2(u_1+v_1)^2a^5\left[\left((6L^2+10L+4)r-(9L^3+21L^2+13L+1)\right)(3L+2-r)\right. \\
& \left.-(r-1)(2L+1-r)(1+3r-3L-3)(L+r-L-1)\right] \\
= \;& 2(u_1+v_1)^2a^5.\left[(27L^3+63L^2+45L+9)r-(45L^4+123L^3+113L^2+37L+2)\right. \\
& \left.-(27L^3+63L^2+45L+9)r+(45L^4+123L^3+113L^2+37L+2)\right] \\
= \;& 0,
\end{aligned}$$

which implies that

$$P'(v_1) = \frac{h'(v_1)f(v_1)-h(v_1)f'(v_1)}{h^2(v_1)} = (L^3+L-L^2-KL^2)a^4\frac{h'(v_1)}{h^2(v_1)} \geq 0.$$

Since $L \geq 2$, $K_0 \geq 1$ and $K_0 L \geq 3$, we have that

$$v_1 \geq \sqrt{L}V(\max(1, 3/L)a \geq \sqrt{L}V(K_0)a = v_0.$$

Since $L \geq 1$, we have $v_1 \geq a$. Because $P'(v)$ is a continuous function, there exists a $\hat{v} \in [\max(a, v_0), v_1]$ such that $P'(\hat{v}) = 0$. Let $\hat{u} = u(\hat{v})$ and $\hat{t} = t(\hat{v})$. Because $u(v)$ is monotonically increasing function of v, we have that

$$a\hat{u} \leq au_1 = M_2 = u_0 v_0 \leq \hat{u}\hat{v}.$$

Since $f'(\hat{v})h(\hat{v}) = h'(\hat{v})f(\hat{v})$, it follows that

$$\frac{f(\hat{v})}{M_2 + a\hat{v}} = 2(3\hat{v} + a)\frac{f(\hat{v})}{f'(\hat{v})} = 2(3\hat{v} + a)\frac{h(\hat{v})}{h'(\hat{v})} = \frac{(\hat{u}^2 - \hat{v}^2)(3\hat{v} + a)}{a + 2\hat{t}}.$$

Since

$$(\hat{u} + \hat{v})(a + 2\hat{t}) = (3\hat{v} + a)(a + \hat{u}),$$

we have

$$\frac{f(\hat{v})}{M_2 + a\hat{v}} = \frac{(\hat{u}^2 - \hat{v}^2)(3\hat{v} + a)}{a + 2\hat{t}} = \frac{(\hat{u}^2 - \hat{v}^2)(\hat{u} + \hat{v})}{a + \hat{u}}.$$

Therefore the corresponding dual objective is

$$\hat{P} = P(\hat{v}) = 1 - \frac{M_2 + a\hat{v}}{(a + \hat{u})(\hat{u} + \hat{v})}.$$

By the definition of t and u, it's straightforward to prove that

$$(3v^2 + 2au - a^2)(a + u) = (\frac{3}{2}v^3 + 5av^2 + \frac{5}{2}a^2 v) + (a^2 + 2av + 3v^2)t = 2(u + v)^3(u - v).$$

Therefore,

$$\begin{aligned}
M_4 &= -f(\hat{v}) + (2a\hat{v}^3 + a^2\hat{v}^2 + M_2(3\hat{v}^2 + 2a\hat{v} + a^2) \\
&= 2a\hat{v}^3 + a^2\hat{v}^2 + M_2(3\hat{v}^2 + 2a\hat{v} + a^2) - (M_2 + a\hat{v})\frac{(\hat{u} + \hat{v})^2(\hat{u} - \hat{v})}{a + \hat{u}} \\
&= 2a\hat{v}^3 + a^2\hat{v}^2 + M_2(3\hat{v}^2 + 2a\hat{v} + a^2) - (M_2 + a\hat{v})\frac{3v^2 + 2au - a^2}{2} \\
&= M_2(\hat{v}^2 + a^2 + \frac{(\hat{v} + a)^2}{2} - a\hat{u} + a\hat{v}) + a\hat{v}(\hat{u}^2 - \hat{u}\hat{v} - a\hat{u}) \\
&= M_2(\hat{v}^2 + a^2 + \hat{u}^2 - \hat{u}\hat{v} - a\hat{u} + a\hat{v}) + a\hat{v}(\hat{u}^2 - \hat{u}\hat{v} - a\hat{u}) \\
&= M_2(\hat{v}^2 + a^2) - a^2\hat{v}^2 + (M_2 + a\hat{v})(\hat{u} - a)(\hat{u} - \hat{v}) \\
&= M_2(\hat{v}^2 + a^2) - a^2\hat{v}^2 + (1 - \hat{P})(\hat{u}^2 - a^2)(\hat{u}^2 - \hat{v}^2) \\
&= (M_2 - (1 - \hat{P})\hat{u}^2)(\hat{v}^2 + a^2) - \hat{P}a^2\hat{v}^2 + (1 - \hat{P})\hat{u}^4.
\end{aligned}$$

Now we define distribution X as

$$X = \begin{cases} \hat{v}, & \text{with probability } q = \frac{M_2 - \hat{P}a^2 - (1-\hat{P})\hat{u}^2}{\hat{v}^2 - a^2}; \\ -\hat{u}, & \text{with probability } p = 1 - \hat{P}; \\ a, & \text{with probability } r = \frac{\hat{P}\hat{v}^2 + (1-\hat{P})\hat{u}^2 - M_2}{\hat{v}^2 - a^2}. \end{cases}$$

Because

$$a\hat{u} \leq M_2 \leq \hat{u}\hat{v},$$

it follows that

$$\frac{\hat{u}^2 - M_2}{\hat{u}^2 - a^2} \leq \hat{P} \leq \frac{\hat{u}^2 - M_2}{\hat{u}^2 - \hat{v}^2}.$$

Therefore, $q, r \geq 0$. Since

$$f(\hat{v}) = \frac{(\hat{u}^2 - \hat{v}^2)(\hat{u} + \hat{v})(M_2 + av)}{a + \hat{u}} \geq 0,$$

and $u > v$, we have that $\hat{P} \leq 1$, therefore $p \geq 0$ and the distribution X is well defined.

Furthermore,

$$\begin{aligned}
\mathsf{E}[X] &= q\hat{v} - p\hat{u} + ra = \frac{M_2 - a\hat{P}\hat{v} - (1 - \hat{P})\hat{u}^2}{\hat{v} + a} - (1 - \hat{P})\hat{u} = 0, \\
\mathsf{E}[X^2] &= q\hat{v}^2 + p\hat{u}^2 + r\hat{a}^2 = M_2, \\
\mathsf{E}[X^4] &= q\hat{v}^4 + p\hat{u}^4 + ra^4 = (M_2 - (1 - \hat{P})\hat{u}^2)(\hat{v}^2 + a^2) - \hat{P}a^2\hat{v}^2 + (1 - \hat{P})\hat{u}^4 = M_4.
\end{aligned}$$

Therefore, X is feasible to problem (A.19).

Finally, since $\hat{v} \geq a$, the primal objective value is $\text{Prob}\{X \geq a\} = q + r = \hat{P}$, which matches the dual objective value for dual feasible solution corresponds to $\hat{v}$. Therefore, we have that

$$Z_P^2 = Z_D^2 = \hat{P}.$$

By Lemma A.1, any $v \geq a$ corresponds to a dual feasible solution with objective value $P(v)$, therefore $P(v) \geq \hat{P}$, and

$$\hat{P} = P(\hat{v}) = \min\{P(v) \mid v \geq a\}.$$

□