**A Semidefinite Programming Approach
to Optimal Moment Bounds
for Convex Classes of Distributions**

by

I. Popescu

2004/30/TM

(Revised Version of 2002/54/TM)

A Semidefinite Programming Approach to Optimal Moment Bounds for Convex Classes of Distributions

Ioana Popescu [*]

March 2002, revised March 2004

Abstract

We provide an optimization framework for computing optimal upper and lower bounds on functional expectations of distributions with special properties, given moment constraints. Bertsimas and Popescu [3, 4] have already shown how to obtain optimal moment inequalities for arbitrary distributions via semidefinite programming. These bounds are not sharp if the underlying distributions possess additional structural properties, including symmetry, unimodality, convexity or smoothness. For convex distribution classes that are in some sense generated by an appropriate parametric family, we use conic duality to show how optimal moment bounds can be efficiently computed as semidefinite programs. In particular, we obtain generalizations of Chebyshev's inequality for symmetric and unimodal distributions, and provide numerical calculations to compare these bounds given higher order moments. We also extend these results for multivariate distributions.

[*]INSEAD, Technology Management Area, Boulevard de Constance, Fontainebleau 77300, FRANCE, ioana.popescu@insead.edu, http://faculty.insead.edu/popescu/ioana

1   Introduction

A generalized moment bound is a problem of the following type: Given "moment" information, in the form $E[f_i(X)] = q_i, i = 1, \ldots, n$, about a random variable X, what are the "best possible" upper and/or lower bounds on the expectation of a related quantity, $\phi(X)$, that can be derived from the available information? We can formulate the problem of finding such optimal upper (and similarly lower) bounds as an optimization program:

$$(P) \quad \max_{X \in \mathcal{P}} \quad E[\phi(X)]$$
$$\text{s.t.} \quad E[f_i(X)] = q_i, \ i = 0, \ldots, n.$$

where the optimization is taken over all possible distributions of the random variable X in the class $\mathcal{P}$. This setup is made rigorous in the next section.

This formulation is powerful because of the variety of interpretations that can be given to the random variable X and the underlying class $\mathcal{P}$, as well as the generality of the objective and constraint functions ϕ and f_i. These are not assumed to be continuous or bounded, to allow for "moments" such as $P(X \geq a)$ and distributions with unbounded support. The problem (P) provides a general framework for studying a multitude of moment problems, with applications. For example, moment inequalities are used to provide robust estimates for financial quantities, such as option and stock prices [22, 14, 4], wealth balance in option hedging [46] or value at risk [13]. In the decision sciences literature, Smith [36] explores several areas of application of moment bounds, including dynamic programming, decision analysis with incomplete information (see also [43, 21]) and Bayesian statistics.

The question of feasibility of Problem (P) given standard moment constraints $E[X^i] = q_i, \ i = 1, \ldots, n$, is the classical moment problem. It has been investigated by probabilists since the nineteenth century, most notably by Chebyshev [7], Markov [24], Stieltjes [38], Akhiezer and Krein [1], Karlin and Studden [16]. For collected works on moment problems, see also Shohat and Tamarkin [35], Tong [40], Landau [20] and references therein. Given the first and second moment of a univariate random variable, Chebyshev's inequality [7, 24] gives a bound on the distribution function. A generalization of this result is due to Bertsimas and Popescu [3] who compute optimal bounds on arbitrary distributions given any finite number of generalized moments using semidefinite programming.

Moment bounds are used to provide robust, worst case estimates of unknown random quantities. These estimates, however, can be overly pessimistic. The reason is that moment bounds are achieved by discrete distributions, which are not always realistic for practical applications. For example, in finance, Yamada and Primbs [46] observe that their upper and lower moment bounds can be far apart, hence not providing much valuable information. This is partly due to ignoring additional structural properties of the underlying stock price distributions. In an inventory control application, Scarf [33] and Gallego [12] derive worst

case order quantities given mean-variance demand information. The resulting bound is very conservative, as it corresponds to an unrealistic two point distribution of demand.

In contrast to worst case moment based estimates, an alternate approach taken in the literature is to fit a (functional) parametric distribution to the moment data. For example, unknown financial quantities are usually modeled as normal or lognormal distributions. In a decision sciences context, Soll and Klayman [37] provide measures of overconfidence by estimating mean absolute dispersion and other distributional properties given sample quantiles of a distribution which is known to be continuous and unimodal. To make the analysis tractable, they fit a beta distribution to the data. This approach, akin to the method of moments estimators, is ubiquitous in a variety of settings for statistical estimation. Lanckriet, El Ghaoui, Bhattacharya and Jordan [19] compares general distribution-free moment bounds with traditional approaches based on normal densities. Their numerical data-classification tests on medical disease data show a significant gap between the estimates provided by the two methods.

Our results are useful in settings where one needs to provide measurements of random quantities when incomplete distributional information is available. Instead of assuming a particular distribution type (normal, lognormal, beta), or solving a pure moment problem, we propose an intermediary approach by incorporating structural distributional properties into the moment problem.

The contribution of this paper is a general approach for deriving moment bounds that are tight for some special convex classes of distributions. Our main result (Theorem 4) states that for piecewise polynomial objective and constraint functions ϕ and f_i, the optimal bound in Problem (P) can be efficiently[1] computed as a semidefinite program (SDP), when the underlying distributions form a convex class that can be "generated" by an appropriate parametric family (see Assumptions [**A, B**]). Such classes include symmetric and/or unimodal distributions, distributions with convex and/or concave densities, and slope constraints. The tighter bounds for these classes extend the results of Bertsimas and Popescu [3] for arbitrary distributions.

For example, suppose we want to find the best upper bound on the probability $P(X - M \geq 2\sigma)$ that a realization of a random variable X, with mean M and variance σ^2, falls at least 2 standard deviations above the mean. Furthermore, suppose X is known to be symmetric and unimodal. The one-sided Chebyshev inequality provides a worst case estimate of 0.20. For a normal random variable the true value is 0.025, compared to which Chebyshev's bound is disappointingly weak. However, by incorporating symmetry and unimodality conditions, but without assuming normality, the tight bound can be reduced to 0.05, a 75% error improvement over Chebyshev (see Proposition 10 Section 7). Our numerical results show even greater improvements for higher order moments.

---

[1]Throughout the paper we refer to an efficient, or polynomial time algorithm when it takes polynomial time in the problem data and $\log(1/\epsilon)$, and it computes a bound within ϵ of the optimal bound for all $\epsilon > 0$.

Univariate moment bounds for unimodal and higher order convexity classes have been previously investigated in the remarkable monograph of Karlin and Studden [16] (Chapter 12), who also provide closed form solutions for the mean-variance case. Their proof relies on a clever integration by parts argument which is valid only if the random variables have unbounded support. A similar approach is taken by Mulholland and Rogers [25], who also characterize the corresponding extremal distributions, thereby extending ad-hoc results of Mallows [23]. None of these papers, however, provides any computational approach. The main contribution of our results is to provide a simple and efficient method for computing the optimal solution for special classes of moment problems via semidefinite programming. Moreover, our duality-based approach allows to characterize more general classes of distributions, including multivariate extensions.

From a methodological standpoint, mathematical programming tools, such as conic duality and semidefinite optimization, provide a powerful framework for solving efficiently what otherwise appears to be a difficult problem in probability. Our results exploit and bring a new dimension – that of special convex distribution classes – to the intriguing connection between moment problems and semidefinite optimization.

Section 2 formalizes the problem and reviews the main known results. Sections 3 develops the conic duality framework and our main result for univariate convex distribution classes generated by a one-parameter family. Section 4 derives optimal moment bounds for symmetric and unimodal distributions. Section 5 investigates the moment problem for distributions with convex, respectively concave densities, including bounds on the slope. Several results for combined classes are also outlined. A general multivariate result, and applications for unimodal and symmetric classes are obtained in Section 6. As an illustration of our approach, in Section 7, we derive SDP formulations for sharp moment bounds on tail probabilities for some special classes of distributions. In particular, we obtain analytical analogues and extensions of Chebyshev's inequality. Numerical results illustrate the comparative performance of these bounds.

2 Moment Bounds for Arbitrary Distributions via SDP

In this section, we briefly review the general moment bounds and corresponding SDP approach from Bertsimas and Popescu [3]. We first set up the problem by introducing some definitions and notation.

2.1 The Problem and Notation

The goal is to find the best upper bound on $E[\phi(X)]$, given expectations q_i of functionals $f_i(X)$ of the random vector X, defined on a closed set $\Omega \subseteq \mathbf{R}^m$, endowed with a Borel sigma algebra $\mathcal{B} = \mathcal{B}_\Omega$, typically omitted for notational convenience. The distribution[2]

---

[2]Throughout the paper, we use the terms measure and distribution interchangeably.

$\mu : \mathcal{B} \to [0,1]$ of X, defined by $\mu(S) = P(X \in S)$, is restricted to a particular convex class $\mathcal{P}$. For a vector of moment functions $f = (f_0, f_1, \ldots, f_n) : \Omega \to \mathbf{R}^n$ we denote the set of *feasible moment sequences* by $Q_{\mathcal{P}}(f) = \{q = (q_0, q_1, \ldots, q_n) \mid q = E_\mu[f(X)] \text{ for some } \mu \in \mathcal{P}\}$. We explicitly include the probability mass constraint by setting $q_0 = 1$, $f_0 \equiv \mathbf{1}_\Omega$, where $\mathbf{1}_S(x)$ is the indicator function of a set S. The central problem of this research is:

$$(P) \quad \max_{\mu \overset{\mathcal{P}}{\sim} (f,q)} E_\mu[\phi(X)] \overset{def}{=} \max_{\mu \in \mathcal{P}} \quad E_\mu[\phi(X)]$$
$$\text{s.t.} \quad E_\mu[f(X)] = q . \tag{1}$$

In this paper, by solving Problem (P) we mean calculating its optimal value. Our use of "max" and "min" operators does not automatically imply that the corresponding optimal value is attained. We focus on Problem (P) as an upper bound problem, but all results hold true for the lower bound problem as well, via a simple change of sign transformation. Moreover, the mathematical programming setup extends to problems involving inequality constraints on the moments.

Consider the set $\mathcal{X}^+ = \mathcal{X}^+(\Omega)$ of all measures on $(\Omega, \mathcal{B})$ such that the functions ϕ and f_i are μ-measurable for all $\mu \in \mathcal{X}^+$. Let $\mathcal{M}^+ = \mathcal{M}^+(\Omega)$ denote the corresponding subset of probability measures (i.e. $\mu(\Omega) = 1$) and let $\mathcal{X} = \mathcal{X}(\Omega)$ denote the span of $\mathcal{M}^+$, which is a linear space of signed measures. Let $\mathcal{X}^*$ denote the span of $\phi, f_0, \ldots, f_n$. The vector spaces $\mathcal{X}$ and $\mathcal{X}^*$ are paired by a bilinear form (scalar product) given by the integral operator $\langle h \mid \mu \rangle := \int_\Omega h d\mu = \int_\Omega h(\omega) d\mu(\omega)$. For probability measures, this is the expectation operator: $E_\mu[h(X)] = \int_\Omega h d\mu$.

The convex hull of a set $\mathcal{T}$ is denoted $\mathbf{cx}(\mathcal{T})$, and $\mathbf{co}(\mathcal{T})$ denotes the cone generated by a set $\mathcal{T}$. Since we are working on infinite dimensional spaces, we extend these concepts to allow for a continuum of convex combinations. We define the convex set of generalized mixtures of distributions $\tau \in \mathcal{T}$ by:

$$\mathbf{mix}(\mathcal{T}) = \{\mu \in \mathcal{M}^+ : \mu(A) = \int_\mathcal{T} \tau(A) d\nu(\tau) \ \forall A \in \mathcal{B} \mid \nu \in \mathcal{M}^+(\mathcal{T})\}, \tag{2}$$

where $\mathcal{M}^+(\mathcal{T})$ is the set of all *mixing distributions* ν over $\mathcal{T}$, with the sigma algebra on $\mathcal{T}$ defined so that the functions $\tau \to \tau(A)$ are measurable in $\tau \in \mathcal{T}$, for every $A \in \mathcal{B}$. In functional analysis, μ is known as the *resultant* or *barycenter* of ν.

Throughout the paper, we denote closure with a bar. We implicitly work with the standard topology of the reals and the weak topology of measures (e.g. see Billingsley [5] or Parthasarathy [28]), unless stated otherwise.

Finally, we use standard notation $\langle a, b \rangle = \{\alpha a + (1 - \alpha)b \mid \alpha \in [0,1]\}$ for the closed segment between vectors a and b, and $(x)^+ = \max(0, x)$ for the positive part of x.

2.2 Review of General Moment Bounds

Consider the moment problem (P) over $\mathcal{P} = \mathcal{M}^{+}$. We denote this problem by (P_0). The dual of this problem can be written as a linear semiinfinite program:

$$
(D_0) \quad \begin{aligned} &\min_y \quad y'q \\ &\text{s.t.} \quad y'f(x) - \phi(x) \geq 0 \quad \forall\, x \in \Omega \end{aligned} \tag{3}
$$

If the support set Ω is finite, this problem is a simple linear program.

Isii [15] shows that solving Problem (P_0) is equivalent to solving its dual, under standard regularity assumptions (e.g. the vector q is in the interior of $Q_{\mathcal{M}^+}(f)$, in the norm topology). For duality results on moment problems see also [16, 1, 36, 34].

Theorem 1 (Linear Duality) *Strong duality holds between Problems (P_0) and (D_0) under standard Slater conditions.*

A characterization of the set of extremal[3] distributions is due to Rogosinsky [32] Thm. 1 (see also Shapiro [34] Lemma 3.1):

Proposition 1 (Extremal Distributions) *The extremal distributions for Problem (P_0) are discrete with support on at most $n + 1$ points, where n is the number of moment constraints.*

Bertsimas and Popescu [3, 4] provide an efficient method for solving a very general class of moment bounds via semidefinite programming (SDP). Semidefinite optimization problems are linear programs with linear matrix inequality (LMI) constraints, i.e. semidefinite positivity constraints on matrices of variables. SDPs preserve the strong duality properties of linear programs, and are efficiently solvable via interior point methods (see [45] and references therein). The following *univariate* result from [3, 4] constitutes a key ingredient in this paper:

Theorem 2 (Moment Bounds via SDP) *The univariate moment Problem (P_0) reduces to solving a semidefinite program, provided that the dual feasible set $\Delta_0 = \{y|\ y'f(x) - \phi(x) \geq 0\,,\ \forall x \in \Omega \subseteq \mathbf{R}\}$ is semi-algebraic.*

A set is semi-algebraic whenever it can be characterized by inequality constraints among polynomials. The proof relies on the fact that a polynomial is non-negative (on an interval) if and only if it is a sum of squares, which can be expressed as an LMI condition (see Nesterov [26] or Bertsimas and Popescu [3, 4]):

---

[3]μ is an *extreme point* of a set $\mathcal{P}$ if it cannot be represented as the average of two distinct elements of $\mathcal{P}$.

Proposition 2 *The set $\{y \mid h_y(x) = \sum_{r=0}^{d} y_r x^r \geq 0 , \forall\, x \in I\}$, where $I \subseteq \mathbf{R}$ is a possibly infinite interval, can be reduced to an LMI of polynomial size in d.*

This result shows that Theorem 2 holds for functions ϕ and f that are piecewise polynomial (henceforth denoted pp). The pp functions form a fairly general class, as most functions of practical interest can be approximated by piecewise polynomials. Examples of pp functions are interval indicators $h(x) = \mathbf{1}_{[a,b]}(x)$ and option-like functions $h(x) = (x - a)^+$.

3 Moment Bounds for General Convex Classes

We extend the results of the previous section to the case when the measures underlying the Problem (P) are restricted to a convex subset $\mathcal{P} \subseteq \mathcal{M}^+$.

3.1 Conic Duality, Generating Sets and Polar Representations

The first step in our developments is to observe that Problem (P) over the convex set $\mathcal{P}$ of probability measures is equivalent to its relaxation over the corresponding cone of measures $\mathcal{C} = \mathbf{co}(\mathcal{P})$. Since we included the probability mass constraint ($f_0 \equiv 1$, $q_0 = 1$) among the moment constraints, we can relax Problem (P) to the following equivalent problem:

$$(P')\quad Z^* = \max_{\mu \in \mathcal{C}} \quad \int_{\Omega} \phi\, d\mu$$
$$\text{s.t.} \quad \int_{\Omega} f\, d\mu = q\,. \tag{4}$$

Given a cone of measures $\mathcal{C} \subseteq \mathcal{X}$, we remind, within our context, the notion of polarity. For a general treatment, see Rockafellar [31].

Definition 1 *The polar of a cone $\mathcal{C} \subseteq \mathcal{X}$ is defined as $\mathcal{C}^* = \{h \in \mathcal{X}^* \mid \int h\, d\mu \geq 0, \ \forall\, \mu \in \mathcal{C}\}$.*

For any set $\mathcal{C}$, its polar $\mathcal{C}^*$ is a pointed convex cone. Shapiro [34] (Prop. 3.4, also 3.1 and 3.3; see also Duffin [10]) provides various specific sufficient conditions for the following conic duality result to hold, as well as for the existence of an optimal solution:

Theorem 3 (Conic Duality) *Under certain Slater conditions (e.g., the moment vector q is interior[4] to the set of feasible moments: $q \in \ int[Q_{\mathcal{C}}(f)]$), then the optimal value of the Primal Problem (P') equals that of the following Dual Problem:*

$$(D)\quad Z^* = \min_y \quad y'q$$
$$\text{s.t.} \quad y'f - \phi \in \mathcal{C}^* \tag{5}$$

---

[4]The strong (norm) topology is needed here.

6

where $\mathcal{C}^*$ is the polar cone of $\mathcal{C}$.

Since $\mathcal{C} = \mathbf{co}(\mathcal{P})$, its polar can be written as $\mathcal{C}^* = \{h \in \mathcal{X}^* \mid \int h d\mu \geq 0,\ \forall\, \mu \in \mathcal{P}\}$. Duality allows to reduce the primal Problem (P) over infinite dimensional variables, to the dual Problem (D) in $n+1$ variables, but with an infinite number of constraints, indexed by the set of probability measures $\mathcal{P}$. This type of indexing is typically difficult to work with. However, the constraint set of the dual problem can be significantly reduced if the set $\mathcal{P}$ is in some sense generate by a convenient class $\mathcal{T}$. We consider several alternate concepts of "generating class":

Assumption [A] *There exists a subset $\mathcal{T}$ of the convex measure set $\mathcal{P}$ underlying Problem (P) such that one the following conditions holds:*

[A1] $\mathcal{P} = \mathbf{cx}(\mathcal{T})$;

[A2] $\mathcal{P} = \mathbf{mix}(\mathcal{T})$;

[A3] $\mathcal{P} = \bar{\mathbf{cx}}(\mathcal{T})$, *and f, ϕ are continuous and bounded;*

[A4] $\mathcal{P} = \bar{\mathbf{cx}}(\mathcal{T})$, *and $\mathcal{T}$ is (weakly) closed.*

The following lemma is a key milestone for our main result, presented in the next section:

Lemma 1 *Suppose that the set of measures $\mathcal{P}$ is generated by a set $\mathcal{T}$ in the sense of Assumption [A]. Then the polar of $\mathcal{C} = \mathbf{co}(\mathcal{P})$ equals $\mathcal{C}^* = \{h \in \mathcal{X}^* \mid \int h d\tau \geq 0,\ \forall\, \tau \in \mathcal{T}\}$.*

This result together with Theorem 3 implies that the dual (D) of Problem (P) is equivalent to the following relaxation:

$$(D_{\mathcal{T}}) \quad Z^* = \min_y \quad y'q$$
$$\text{s.t.} \quad \int y'f - \phi d\tau \geq 0,\ \forall\, \tau \in \mathcal{T}\,.$$

For example, consider $\mathcal{P} = \mathcal{M}^+$ and let $\mathcal{T}_\delta$ denote the set of Dirac measures $\delta_x, x \in \Omega$. Since

$$\mathcal{M}^+ = \bar{\mathbf{cx}}(\mathcal{T}_\delta) = \mathbf{mix}(\mathcal{T}_\delta), \tag{6}$$

by Lemma 1 ([**A2**] or [**A4**]), the corresponding polar cone is $\mathcal{C}^* = \{h \in \mathcal{X}^* \mid h(x) \geq 0,\ \forall x \in \Omega\}$. In this case, the conic duality result of Theorem 3 reduces to the linear duality result of Theorem 1. By Lemma 1 ([**A1**]), the same polar representation holds for the set $\mathcal{D}^+$ of discrete probability measures on Ω, in which case $\mathcal{D}^+ = \mathbf{cx}(\mathcal{T}_\delta)$.

The applications presented in this paper rely mostly on Assumptions [**A2**] and [**A4**]. A closed convex hull ([**A4**]) representation is typically easier and more intuitive to obtain than an integral representation ([**A2**]), hence more useful in practical applications (see

e.g. Section 4 and 5). Assumption [**A3**] allows a weakening of [**A4**] when the underlying functions are continuous and bounded, whereas Assumption [**A1**] is useful in the case of discrete distributions.

Proof of Lemma 1: Under Assumption [**A1**], the lemma follows by simple application of convexity to the linearity of the expectation operator.

Under [**A2**], for any $\mu \in \mathcal{P}$ we can write $\mu = \int_{\mathcal{T}} \tau d\nu(\tau)$ for a certain mixing measure ν. If $h \in \mathcal{X}^*$ then $\int_{\Omega} h d\tau$ is ν-measurable and $\int_{\Omega} h d\mu = \int_{\mathcal{T}} \left(\int_{\Omega} h d\tau \right) d\nu(\tau)$ (see e.g. Dudley [9] p.269). The desired result follows.

Under [**A3**], for any $\mu \in \bar{\mathbf{cx}}(\mathcal{T})$, consider a sequence of measures $\mu_n \in \mathbf{cx}(\mathcal{T})$ converging weakly to μ. Since $\mathcal{X}^*$ consists of continuous bounded functions, it follows that $0 \leq \int h d\mu_n \to \int h d\mu$ for any $h \in \mathcal{D}^*$, where $\mathcal{D} = \mathbf{co}(\mathbf{cx}((\mathcal{T}))$. Therefore $\int h d\mu \geq 0$, which proves $\mathcal{D}^* \subseteq \mathcal{C}^*$. The opposite inclusion is trivial.

Under [**A4**], the lemma is a consequence of the second part of the following result, proved in Appendix A. $\square$

Proposition 3 *Let $\mathcal{T}$ be a set of Borel probability measures on $\Omega \subseteq \mathbf{R}^m$. We have that $\mathbf{cx}(\mathcal{T}) \subseteq \mathbf{mix}(\mathcal{T}) \subseteq \bar{\mathbf{cx}}(\mathcal{T})$. Moreover, if Ω and $\mathcal{T}$ are closed, then $\bar{\mathbf{cx}}(\mathcal{T}) = \mathbf{mix}(\mathcal{T})$.*

In a general topological context, the existence of generalized mixture (a.k.a. integral) representations and its relationship with closed convex hull representations is the focus of Choquet theory, for which a standard reference is Phelps [29]. The main results are the theorems of Krein-Milman and Choquet, providing topological conditions under which a set admits a closed convex hull representation, and respectively a general mixture representation in terms of its extreme points. Extensions and applications in the context of measure sets can be found in Rogosinsky [32], Winkler [44], Weizsäcker and Winkler [42] and Karr [17]. None of these results, however, implies Proposition 3 in the non-compact case. Moreover, we intentionally do not require the generating set $\mathcal{T}$ to coincide with the extreme points of $\mathcal{P}$, since these can be difficult to characterize, and to parametrize as required further by Assumption [**B**].

3.2   The Main Result for Convex Classes of Univariate Distributions

Our main result requires, beside Assumption [**A**], a parametrization of the generating set:

Assumption [B] *The generating set $\mathcal{T} \subseteq \mathcal{P}$ consists of Borel measures μ_t on a real interval Ω, such that $\mu(S, t) = \mu_t(S) : \mathcal{B} \times I_T \to [0, 1]$ is a Borel measurable function of t and I_T is a real interval.*

Assumption [**B**] effectively says that $\mathcal{T}$ is defined by a Markov kernel. Extensions dealing with multivariate distributions ($\Omega \subseteq \mathbf{R}^m$) and multi-parameter generating classes

$(I_T \subseteq \mathbf{R}^m)$ are investigated in Section 6. Our main result extends Theorem 2 by providing an SDP solution for univariate moment problems over convex classes of distributions:

Theorem 4 (Moment Bounds for Convex Distribution Classes via SDP) *Suppose that the Slater condition holds for Problem (P) satisfying Assumptions [**A**] and [**B**]. Then the following results are true:*

(a) *The optimal bound in Problem (P) can be efficiently computed as a semidefinite program whenever the set $\{y \mid \int (y'f - \phi)d\mu_t \geq 0, \forall t \in I_T \subseteq \mathbf{R}\}$ is semi-algebraic.*

(b) *If the bound in Problem (P) is achievable, then there exists an optimal measure which is a convex combination of $n+1$ probability measures from the generating set $\mathcal{T}$.*

Proof: (a) For any function $h \in \mathcal{X}^*$, define the linear transform $H_\mathcal{T}(t) = \int_\Omega h d\mu_t$. Based on Theorem 3 and Lemma 1, the dual of Problem (P) over the convex class $\mathcal{P}$ is the following semi-infinite linear program:

$$(D_\mathcal{T}) \quad Z^* = \min_y \quad y'q$$
$$\text{s.t.} \quad y'F_\mathcal{T}(t) - \Phi_\mathcal{T}(t) \geq 0 , \ \forall t \in I_T. \tag{7}$$

By Theorem 2, this reduces to an SDP whenever the dual feasible set is semi-algebraic.

(b) Assumption [**B**] implies that $\Phi_\mathcal{T}$ and $F_\mathcal{T}$ are measurable functions of t. The problem:

$$(P_\mathcal{T}) \quad \max_{\nu \in \mathcal{M}^+(I_T)} \quad \int \Phi_\mathcal{T} d\nu$$
$$\text{s.t.} \quad \int F_\mathcal{T} d\nu = q . \tag{8}$$

is a special case of Problem (P_0) over $\Omega = I_T$ with $(f, \phi) := (F_\mathcal{T}, \Phi_\mathcal{T})$. The mass constraint is satisfied since $F_\mathcal{T}^0(t) = \int f_0 d\mu_t = 1$.

One can easily see that the Slater condition is verified for Problem $(P_\mathcal{T})$ whenever it is verified for (P). So by Theorem 1, Problem (P) has the same optimal value as Problem $(P_\mathcal{T})$. Moreover, if the bound in Problem (P) is achievable, then so is the bound in Problem $(P_\mathcal{T})$, since for an optimal measure μ for (P), the corresponding mixing measure $\nu \in \mathcal{M}^+(I_T)$ is optimal for $(P_\mathcal{T})$.

By Proposition 1, the extremal measures for Problem $(P_\mathcal{T})$ are of the type $\nu^* = \sum_{i=0}^n w_i \delta_{t_i}$, $t_i \in I_\mathcal{T}$ with $\sum w_i = 1$, $w_i \geq 0$. The feasibility and optimality conditions for ν^* are:

$$q = \sum_{i=0}^n w_i \int f d\mu_{t_i} = \int f d \left(\sum_{i=0}^n w_i \mu_{t_i} \right)$$
$$Z^* = \sum_{i=0}^n w_i \int \phi d\mu_{t_i} = \int \phi d \left(\sum_{i=0}^n w_i \mu_{t_i} \right).$$

9

This shows that $\mu^* = \sum_{i=0}^{n} w_i \mu_{t_i} \in \mathbf{cx}(\mathcal{T}) \subseteq \mathcal{P}$ achieves the optimum in Problem (P), since it satisfies the moment conditions, and achieves the optimal value Z^*. $\qquad\square$

Our main result requires the feasible set of the Dual Problem $(D_{\mathcal{T}})$ to be semi-algebraic. When f and ϕ are piecewise polynomial (pp), this condition is satisfied for example if $\mathcal{T} = \{\mu_t\}_{t \in I_T}$ is a parametric class of continuous measures whose densities $\pi_t(x) = \pi(x, t)$ are pp in x and t. In this case, the following functions are pp in t:

$$
\begin{aligned}
F_{\mathcal{T}}(t) &= \int f d\mu_t &= \int f(x) \pi(x, t) dx, \\
\Phi_{\mathcal{T}}(t) &= \int \phi d\mu_t &= \int \phi(x) \pi(x, t) dx.
\end{aligned}
$$

The semi-algebraic condition also allows for generating densities that are pp in x, but fractional in the parameter t. In this case, the functions $\Phi_{\mathcal{T}}$ and $F_{\mathcal{T}}$ are piecewise polynomial fractions (ppf), that is fractions of polynomials on an interval partition, and the dual constraints can be reduced to semi-algebraic conditions (since checking $p(x)/q(x) \geq 0$ is equivalent to $p(x)q(x) \geq 0$). Examples are Cauchy, Pareto and uniform densities.

Another case of potential interest is that of exponentially damped moments $f_i(x) = f_i^0(x)e^{-\lambda(x)}$ and $\phi(x) = \phi^0(x)e^{-\lambda(x)}$, for arbitrary functions $\lambda(\cdot)$ and where $f^0(\cdot)$, $\phi^0(\cdot)$ are ppf. In this case, checking $y'f(x) - \phi(x) \geq 0$ amounts to $y'f^0(x) - \phi^0(x) \geq 0$, a semi-algebraic condition.

4   Bounds for Symmetric and Unimodal Distributions

In this section we apply Theorem 4 to derive optimal moment bounds for distributions that are: (1) symmetric, (2) unimodal with given mode, (3) unimodal with bounds on the mode, (4) unimodal and symmetric (including bounds on the mean). Multivariate analogues of these properties are investigated in Section 6.

4.1   Symmetric Distributions

Let $\Omega = I \subseteq \mathbf{R}$, be either a compact interval, or $I = \mathbf{R}$. A measure μ on $(I, \mathcal{B})$ is M-symmetric if $\mu[M - x, M] = \mu[M, M + x]$ for all $x \in I_M$, where $I_M = \{x \geq 0 \,|\, M - x \in I$ and $M + x \in I\}$.

Since any convex combinations of M-symmetric probability measures is M-symmetric, the set of M-symmetric probability measures $\mathcal{P}_M^{\mathbf{S}}$ is convex. Symmetry is preserved under weak limits, so $\mathcal{P}_M^{\mathbf{S}}$ is closed.

Lemma 2 *The closed convex set of M-symmetric probability measures can be generated by pairs of symmetric Diracs as $\mathcal{P}_M^{\mathbf{S}} = \bar{\mathbf{cx}}(\mathcal{T}_M^{\mathbf{S}})$, where $\mathcal{T}_M^{\mathbf{S}} = \{\mu_t = \frac{1}{2}\delta_{M+t} + \frac{1}{2}\delta_{M-t} \,|\, t \in I_M\}$.*

The proof trivially follows from (6) applied on the half-interval $\Omega = M + I_M$. Since $\mathcal{T}_M^{\mathtt{S}}$ is closed, by Lemma 1 ([**A4**]), the polar cone of M-symmetric measures is:

$$\mathcal{C}_M^{\mathtt{S}\,*} = \{h \in \mathcal{X}^* \mid H_M^{\mathtt{S}}(t) = h(M+t) + h(M-t) \geq 0, \quad \forall\, t \in I_M\}.$$

Therefore, the dual of the moment Problem for M-symmetric distributions is:

$$\begin{aligned} \min_y \quad & y'q \\ \text{s.t.} \quad & y'(f(M-t) + f(M+t)) - (\phi(M-t) + \phi(M+t)) \geq 0 \quad, \forall\, t \in I_M \end{aligned} \tag{9}$$

The constraints amount to checking polynomial positivity whenever f, ϕ are pp, so Theorem 4 implies the following result:

Proposition 4 *If ϕ and f are* pp *and the Slater condition holds, then:*

(a) *Problem (P) over the convex set $\mathcal{P}_M^{\mathtt{S}}$ of M-symmetric probability measures on $(I, \mathcal{B})$ can be efficiently solved as a semidefinite program.*

(b) *If the corresponding bound is achievable, there exists an optimal measure which is a convex combinations of $n+1$ pairs of M-symmetric Diracs.*

When the mean M of the distribution is unspecified, our approach does not directly provide sharper bounds based on symmetry information. This is because the class of symmetric probability measures is non-convex, and its convex hull is all of $\mathcal{M}^+$.

4.2   Unimodal Distributions

Let $\Omega = I \subseteq \mathbf{R}$ be a possibly infinite interval. A measure μ is m-unimodal on $I \ni m$ if the corresponding distribution function is convex to the left of m and concave to the right of m. For a continuous measure, this amounts to its density function being increasing[5] to the left of m and decreasing to the right of m. Any m-unimodal measure can be decomposed into a continuous m-unimodal measure and a Dirac at m. The class of continuous m-unimodal probability measures $\mathcal{P}_m^{\mathtt{u}}$ is convex, and its closure $\bar{\mathcal{P}}_m^{\mathtt{u}}$ is the class of m-unimodal measures.

Examples of continuous unimodal densities include the normal (possibly truncated), Betas, monotone densities and uniform densities. Measures with uniform densities on closed m-bounded segments I are also called m-*rectangular* measures, and denoted δ_I. The following characterization of unimodal measures can be found in Dharmadhikari and Joag-Dev [8] (Theorem 1.2). The mixture representation is attributed to Khintchine [18].

Lemma 3 *The set of (continuous) m-unimodal measures can be generated using m-rectangulars as* $\mathcal{P}_m^{\mathtt{u}} = \mathbf{mix}(\mathcal{T}_m^{\mathtt{u}})$, *respectively* $\bar{\mathcal{P}}_m^{\mathtt{u}} = \bar{\mathbf{cx}}(\mathcal{T}_m^{\mathtt{u}})$, *where* $\mathcal{T}_m^{\mathtt{u}} = \{\delta_{\langle t,m \rangle} \mid t \in I, \, t \neq m\}$.

---

[5]Throughout the paper, we will use the terms increasing/decreasing in their weak sense.

For the closed convex hull representation, Assumption [**A4**] requires closure of the generating set, so a Dirac at m should be added to $\mathcal{T}_m^{\mathtt{u}}$. By Lemmas 1 and 3, the dual of the moment problem for m-unimodal distributions is:

$$
\begin{aligned}
\min_y \quad & y'q \\
\text{s.t.} \quad & y' \int_{\langle m,t \rangle} f(x)dx - \int_{\langle m,t \rangle} \phi(x)dx \geq 0 \quad , \ \forall \, t \in I \,, \ t \neq m \\
& y' f(m) - \phi(m) \geq 0.
\end{aligned}
\tag{10}
$$

The last constraint should be relaxed for the case of *continuous m-unimodal measures*.

The integral transform $\int_{\langle m,t \rangle} h(x)dx$ is $\mathtt{pp}$ in t for any function $h \in \mathcal{X}^*$ that is $\mathtt{pp}$ in x, so Theorem 4 implies the following result:

Proposition 5 *Suppose that ϕ and f are $\mathtt{pp}$ and the Slater condition holds. Then the following results are true:*

(a) *Problem (P) over the convex class of (continuous) m-unimodal probability measures, can be efficiently solved as a semidefinite program.*

(b) *If the corresponding bound is achievable, then there exists an optimal measure which is a convex combinations of $n+1$ m-rectangulars, possibly including a Dirac at m.*

4.3 Constraints on the mode and $[m_1, m_2]$-modal distributions.

Suppose that the underlying distribution in Problem (P) is unimodal, but the mode is only known to belong to a certain interval: $m \in [m_1, m_2]$. Define $\mathcal{P}_{m_1,m_2}^{\mathtt{u}} = \mathbf{cx}(\bigcup_{m \in [m_1,m_2]} \mathcal{P}_m^{\mathtt{u}})$, to be the convex hull of the (non-convex) class of unimodal probability measures with mode between m_1 and m_2. We refer to the elements of the class $\mathcal{P}_{m_1,m_2}^{\mathtt{u}}$ as $[m_1, m_2]$-*modal probability measures*. By Lemma 3, an obvious generating family for $\mathcal{P}_{m_1,m_2}^{\mathtt{u}}$ is the class $\bigcup_{m \in [m_1,m_2]} \mathcal{T}_m^{\mathtt{u}}$ of rectangulars $\delta_{\langle s,t \rangle}$ for $s \in I$, $t \in [m_1, m_2]$, and Diracs δ_t for $t \in [m_1, m_2]$. However, this class is double indexed by s and t, hence does not satisfy Assumption [**B**].

Any measure in $\mathcal{P}_{m_1,m_2}^{\mathtt{u}}$ can be decomposed into an m_1 unimodal to the left of m_1, an m_2 unimodal to the right of m_2 and an arbitrary distribution on $[m_1, m_2]$. Lemma 3 implies the following representation:

Lemma 4 *The closed convex set of $[m_1, m_2]$-modal measures can be generated by the closed set of probability measures $\mathcal{T}_{m_1,m_2}^{\mathtt{u}} = \{\mu_t\}_{t \in I}$, where:*

$$
\mu_t = \begin{cases}
\delta_{[t,m_1]} & , \ \text{if} \ t < m_1 \\
\delta_{[m_2,t]} & , \ \text{if} \ t > m_2 \ . \\
\delta_t & , \ \text{if} \ t \in [m_1, m_2]
\end{cases}
$$

12

By Lemma 1, the dual of Problem (P) over $\mathcal{P}^{\mathtt{u}}_{m_1,m_2}$ can be written as:

$$
\begin{aligned}
(D^{\mathtt{u}}_{m_1,m_2}) \quad \min_y \quad & y'q \\
\text{s.t.} \quad & \int_t^{m_1} y'f(x) - \phi(x)dx \geq 0 \quad , \; \forall\, t \leq m_1, \; x \in I \\
& \int_{m_2}^t y'f(x) - \phi(x)dx \geq 0 \quad , \; \forall\, t \geq m_2, \; x \in I \\
& y'f(t) - \phi(t) \geq 0 \qquad\qquad\quad , \; \forall\, t \in [m_1, m_2].
\end{aligned}
\tag{11}
$$

By Theorem 4, we have the following result:

Proposition 6 *If f and ϕ are* pp *and the Slater condition holds, then:*

(a) *Problem (P) over the closed convex set of $[m_1, m_2]$-unimodals can be efficiently solved as an SDP.*

(b) *If the corresponding bound is achievable, then there exists an optimal measure which is a convex combinations of $n + 1$ probability measures from the class $\mathcal{T}^{\mathtt{u}}_{m_1,m_2}$.*

If the mode of the underlying distribution is completely unspecified, then we cannot improve the moment bounds by adding unimodality conditions in a duality framework. This is because the convex hull of all unimodal probability measures is all of $\mathcal{M}^+$.

4.4 Symmetric Unimodal Distributions

We now combine the results for unimodal and symmetric distributions. The set of continuous M-symmetric unimodal probability distributions is convex, denoted $\mathcal{P}^{\mathtt{su}}_M = \mathcal{P}^{\mathtt{s}}_M \bigcap \mathcal{P}^{\mathtt{u}}_M$, and its closure $\bar{\mathcal{P}}^{\mathtt{su}}_M = \mathcal{P}^{\mathtt{s}}_M \bigcap \bar{\mathcal{P}}^{\mathtt{u}}_M$, is the set of all M-symmetric unimodals. Recall the notation $I_M = \{t \geq 0 \,|\, M - t \in I \text{ and } M + t \in I\}$.

Lemma 5 *The convex set of M-symmetric unimodal measures can be generated using M-centered rectangulars as $\bar{\mathcal{P}}^{\mathtt{su}}_M = \bar{\mathbf{cx}}(\mathcal{T}^{\mathtt{su}}_M)$, respectively $\mathcal{P}^{\mathtt{su}}_M = \mathbf{mix}(\mathcal{T}^{\mathtt{su}}_M)$ for continuous measures, where $\mathcal{T}^{\mathtt{su}}_M = \{\delta_{[M-t,M+t]} \,|\, t \in I_M, \; t \neq 0\}$.*

The result is a direct consequence of Lemma 3. Again, for $\bar{\mathcal{P}}^{\mathtt{su}}_M$, a Dirac at M should be added to the generating set to insure closure ([**A4**]). Based on this characterization, Lemma 1 yields the following dual formulation (after basic simplifications):

$$
\begin{aligned}
\min_y \quad & y'q \\
\text{s.t.} \quad & y' \int_{M-t}^{M+t} f(x)dx - \int_{M-t}^{M+t} \phi(x)dx \geq 0, \; \forall\, t \in I_M, t \neq 0 \\
& y'f(M) - \phi(M) \geq 0
\end{aligned}
\tag{12}
$$

The last constraint should be dropped for continuous M-symmetric unimodal measures.

For any $\mathtt{pp}$ function $h \in \mathcal{X}^*$ the corresponding transform $H_M^{\mathtt{su}}(t) = \int_{M-t}^{M+t} h(x)dx$ is $\mathtt{pp}$ in t, so Theorem 4 implies the following result:

Proposition 7 *Suppose that ϕ and f are $\mathtt{pp}$ and the Slater condition holds. Then the following results are true:*

(a) *Problem (P) over the class of (continuous) M-symmetric unimodal probability measures can be efficiently solved as a semidefinite program.*

(b) *If the bound is achievable, there exists an optimal measure which is a convex combination of $n+1$ M-symmetric rectangulars, possibly including a Dirac at M.*

Unspecified mean. Given bounds on the mean, the convex hull of the class of symmetric unimodal distributions with mean $M \in [M_1, M_2]$ is the class of $[M_1, M_2]$-unimodals, with mean in $[M_1, M_2]$. For this class, we derived bounds in Section 4.3.

Without any prior bounds on the mean, our approach does not directly improve the bounds for arbitrary distributions. This is because the convex hull of symmetric unimodal probability distributions is all of $\mathcal{M}^+$. In this case, we propose an alternative two step approach:

1. Find optimal lower and upper bounds M_1, M_2 on the mean given the moment constraints, by solving the corresponding moment problem for arbitrary distributions.

2. Solve the moment problem for $[M_1, M_2]$-unimodals, with mean in $[M_1, M_2]$.

5 Bounds for Distributions with Convex/Concave Densities.

The last section showed how monotonicity properties of the distributions in Problem (P) translate into optimizing over certain combinations of $n+1$ uniforms. The resulting bounds are tighter than those for arbitrary distributions, which are achieved by discrete distributions (this is also illustrated numerically in Section 7). However, approximations by uniforms can be rough if the distributions satisfy additional convexity or smoothness properties. In this section we improve the moment bounds by incorporating conditions on the slope of the underlying densities.

5.1 Monotone convex densities.

Denote $\mathcal{P}_{m,a}^{\mathtt{dx}}$ the set of continuous probability measures on $\Omega = [m, m+a]$, which admit **d**ecreasing and conve**x** densities. Similarly, consider the set $\mathcal{P}_{m,a}^{\mathtt{ix}}$ of continuous probability measures that admit **i**ncreasing and conve**x** densities on an interval $[m-a, m]$. In this section only, we abuse notation and allow $a \geq 0$ to be infinite.

Both classes are convex, but not closed, since limiting distributions (such as δ_m) do not necessarily have a density. The respective closures $\bar{\mathcal{P}}_{m,a}^{\mathtt{dx}}$ ($\bar{\mathcal{P}}_{m,a}^{\mathtt{ix}}$) consist of distributions that can be decomposed into a continuous part, with a decreasing (increasing) and convex density, and possibly a Dirac at m. These can be generated using *right (left) m-triangular* densities:

Definition 2 *A probability density function is said to be:*

(a) *right m-triangular, if it is given by* $\pi_{m,t}^{+}(x) = \dfrac{2}{t^2}(m + t - x)^{+}$, *for* $x \geq m, t > 0$.

(b) *left m-triangular, if it is given by* $\pi_{m,t}^{-}(x) = \dfrac{2}{t^2}(x - m + t)^{+}$, *for* $x \leq m, t > 0$.
   The corresponding triangular measures are denoted $\gamma_{m,t}^{+}$, *respectively* $\gamma_{m,t}^{-}$.

   For $t \to 0$, *the degenerate triangular measure* $\gamma_{m,0}^{\pm}$ *is a Dirac at* m.

Lemma 6 (a) *We have that* $\bar{\mathcal{P}}_{m,a}^{\mathtt{dx}} = \bar{\mathbf{cx}}(\mathcal{T}_{m,a}^{\mathtt{dx}})$ *and* $\mathcal{P}_{m,a}^{\mathtt{dx}} = \mathbf{mix}(\mathcal{T}_{m,a}^{\mathtt{dx}})$, *where* $\mathcal{T}_{m,a}^{\mathtt{dx}} = \{\gamma_{m,t}^{+} \mid t \in [0, a)\} \cup \{\delta_{[m,m+a]}\}$.

(b) *We have that* $\bar{\mathcal{P}}_{m,a}^{\mathtt{ix}} = \bar{\mathbf{cx}}(\mathcal{T}_{m,a}^{\mathtt{ix}})$ *and* $\mathcal{P}_{m,a}^{\mathtt{ix}} = \mathbf{mix}(\mathcal{T}_{m,a}^{\mathtt{ix}})$, *where* $\mathcal{T}_{m,a}^{\mathtt{ix}} = \{\gamma_{m,t}^{-} \mid t \in [0, a)\} \cup \{\delta_{[m-a,m]}\}$.
   If a is infinite, the rectangular should be omitted from the generating sets.

The proof is provided in Appendix B. By Lemma 1, this characterization leads to the following equivalent formulation of the dual problem over the class $\bar{\mathcal{P}}_{m,a}^{\mathtt{dx}}$:

$$\begin{aligned}
\min \quad & y'q \\
\text{s.t.} \quad & y' \int_{m}^{m+a}(m + t - x)^{+}f(x)dx - \int_{m}^{m+a}(m + t - x)^{+}\phi(x)dx \geq 0, \ \forall\, t \in (0, a] \\
& y'f(m) - \phi(m) \geq 0.
\end{aligned}$$

The last constraint should be dropped if we are only interested in continuous measures ($\mathcal{P}_{m,a}^{\mathtt{dx}}$). The dual constraints are $\mathtt{ppf}$ in t, whenever f and ϕ are $\mathtt{pp}$, hence the problem can be stated as an SDP.

We obtain analogous results for the convex class $\mathcal{P}_{m,a}^{\mathtt{ix}}$. By combining the two, we can obtain SDP formulations for optimal moment bounds for measures with U-shaped densities with given mode (point of minimum) m. Summarizing, by Theorem 4 we have:

Proposition 8 *If ϕ and f are $\mathtt{pp}$ and the Slater condition holds, then:*

(a) *Problem (P) over the classes $\mathcal{P}_{m,a}^{\mathtt{dx}}, \mathcal{P}_{m,a}^{\mathtt{ix}}$ and their respective closures can be efficiently solved as a semidefinite program.*

(b) *If the corresponding bound is achievable, there exists an optimal solution which is a convex combination of $n+1$ right/left triangulars, possibly including a rectangular on the whole domain and a Dirac at m.*

Constraints on the maximal slope. We can improve these bounds by further limiting the slope of the densities underlying the moment problems. This basically amounts to putting bounds on the values t that index the generating set.

Consider for example the moment problem (P) for measures that admit convex increasing densities on $(-\infty, m]$ with maximal slope bounded in a finite interval $[\alpha, \beta]$. This class is closed and convex (since β is finite, the Dirac is not a possible limiting distribution, hence these measures are continuous). By Proposition 8, this problem admits an optimal solution whose density is of the type:

$$\pi^*(x) = \sum_{i=0}^{n} w_i \pi_{m,t_i}^+(x) = \sum_{i=0}^{n} w_i \frac{2}{t_i^2} (x - m + t_i)^+, \ x \le m, \ t_i > 0 \tag{13}$$

with $w_i \ge 0$. Since π^* must correspond to a probability distribution, the mass constraint implies that $\sum w_i = 1$. The maximum slope of the corresponding density is:

$$\sum_{i=0}^{n} \frac{2}{t_i^2} w_i \in (\min_i \frac{2}{t_i^2}, \max_i \frac{2}{t_i^2}).$$

In order for the maximal slope to be between α and β, we should restrict the generating class to $\alpha \le 2/t^2 \le \beta$, that is:

$$\mathcal{T}_{m,\alpha,\beta}^{\mathtt{dx}} = \{\gamma_{m,t}^+ \mid \sqrt{2/\beta} \le t \le \sqrt{2/\alpha}\}.$$

We can similarly treat the case of convex decreasing densities, by incorporating bounds on the maximal absolute slope.

Approximations by triangulars can be regarded as a second stage in a series of approximations by means of increasing numbers of independent uniforms. As pointed out by Bell [2], triangular distributions can provide fairly good approximations to normal distributions. Buslenko et al. [6] present a method of construction of so-called "random normal deviates" using triangular distributions. However, normal densities are concave around the mean and convex on the tails. Therefore, if one wants to provide more suitable approximations for such distributions, it is desirable to model concave densities in the moment problem.

5.2 Monotone concave densities.

Let $m, a \in \mathbf{R}, a \ge 0$ and consider the class $\mathcal{P}_{m,a}^{\mathtt{dv}}$ of probability measures on a finite interval $[m, m + a]$ that admit **d**ecreasing and conca**v**e densities. Similarly, consider the class $\mathcal{P}_{m,a}^{\mathtt{iv}}$ of probability measures that admit **i**ncreasing and conca**v**e densities on $[m - a, m]$. Both

16

classes are closed and convex, and can be generated using *m-trapezoidal densities*.

Definition 3 *A probability density function π is said to be:*

(a) *right m-trapezoidal on $[m, m+a]$, if for some $0 \le t < a$ it is given by*

$$\pi_{m,t,a}^+(x) = \frac{2}{a+t} \min\left(1, \frac{m+a-x}{a-t}\right), \ m \le x \le m+a.$$

(b) *left m-trapezoidal on $[m-a, m]$, if for some $0 \le t < a$ it is given by*

$$\pi_{m,t,a}^-(x) = \frac{2}{a+t} \min\left(1, \frac{x-m+a}{a-t}\right), \ m-a \le x \le m.$$

The corresponding trapezoidal probability measures are denoted $\zeta_{m,t,a}^+$, respectively $\zeta_{m,t,a}^-$.

If $t = 0$, the degenerate trapezoidal is a triangular. As $t \to a$ above, the degenerate trapezoidal measures denoted $\zeta_{m,a,a}^\pm$ are rectangulars on the whole domain.

Lemma 7 (a) *The closed convex class $\mathcal{P}_{m,a}^{\mathtt{dv}}$ can be generated as $\mathcal{P}_{m,a}^{\mathtt{dv}} = \bar{\mathbf{cx}}(\mathcal{T}_{m,a}^{\mathtt{dv}})$, where $\mathcal{T}_{m,a}^{\mathtt{dv}} = \{\zeta_{m,t,a}^+ \mid t \in [0, a]\}$.*

(b) *The closed convex class $\mathcal{P}_{m,a}^{\mathtt{iv}}$ can be generated as $\mathcal{P}_{m,a}^{\mathtt{iv}} = \bar{\mathbf{cx}}(\mathcal{T}_{m,a}^{\mathtt{iv}})$, where $\mathcal{T}_{m,a}^{\mathtt{iv}} = \{\zeta_{m,t,a}^- \mid t \in [0, a]\}$.*

The proof is provided in Appendix B. Since the generating classes are closed, by Lemma 1 ([**A4**]), the dual of Problem (P) over the class $\mathcal{P}_{m,a}^{\mathtt{dv}}$ is:

$$\begin{aligned} \min \ & y'q \\ \text{s.t.} \ & \int_m^{m+a} \min(a-t, m+a-x)(y'f(x) - \phi(x))dx \ge 0, \ \forall \, t \in [0, a]. \end{aligned}$$

This can be expressed as an SDP whenever f and ϕ are $\mathtt{pp}$.

We obtain analogous results for the convex class $\mathcal{P}_{m,b}^{\mathtt{iv}}$, generated by left m-trapezoidal probability measures. By combining the two, we obtain optimal moment bounds for (symmetric) measures with m-unimodal and concave densities on $[m-b, m+a]$. These are generated by (symmetric pairs of) right and left m-trapezoidal distributions.

Proposition 9 *If ϕ and f are $\mathtt{pp}$ and the Slater condition holds, then:*

(a) *Problem (P) over the closed convex class of m-unimodal probability measures with concave densities can be efficiently solved as a semidefinite program.*

(b) *If the corresponding bound is achievable, then there exists an optimal measure which is a convex combination of $n+1$ m-trapezoidals, possibly degenerate.*

Constraints on the maximal slope. Consider for example the moment problem $(P_{m,a}^{\mathtt{iv}})$ for concave increasing densities on $[m-a, m]$. By Proposition 9, this problem admits an optimal solution whose density has the following structure:

$$\pi^*(x) = \sum_{i=0}^n w_i \pi_{m,t_i,a}^+(x) = \sum_{i=0}^n w_i \frac{2}{a^2 - t_i^2} \min(a - t_i, x - m + a), \; x \leq m, \; 0 \leq t_i < a$$

with $w_i \geq 0$ and $\sum w_i = 1$. Therefore, in order to restrict the maximal slope between α and β, we should define the generating class as:

$$\mathcal{T}_{m,\alpha,\beta}^{\mathtt{dv}} = \{\zeta_{m,t,a}^+ \mid \sqrt{a^2 - 2/\beta} \leq t \leq \sqrt{a^2 - 2/\alpha}\,\}.$$

We can similarly treat the case of concave decreasing densities, by incorporating bounds on the maximal absolute slope.

Finally, by putting all the pieces together, we obtain bounds on (symmetric) m-unimodal distributions that are concave on $[m-b, m+a]$ and convex on the tails $(-\infty, m-b]$, $[m+a, \infty)$. Furthermore, slope constraints can be added at the inflexion points. Similarly, we obtain bounds for (symmetric) m-unimodal distributions with convex tails outside an interval $[m-b, m+a]$, and other such combinations of convex properties. These results can be extended to incorporate higher order convexity and smoothness information, by using generating families of piecewise polynomials densities.

6 Multivariate Extensions

Our results so far concern univariate distributions generated by single-parameter classes. In this section we generalize these results to multiple dimensions, and apply them for distributions satisfying multivariate unimodality and symmetry properties.

All the results of Sections 2 and 3.1 hold over $\mathbf{R}^m$, with the exception of Theorem 2. Bertsimas and Popescu [3] extend this result in the context of a moment problem over $\mathbf{R}^m$. They prove that the dual problem of optimizing a linear objective over a general semi-algebraic set (i.e. defined by multivariate polynomial inequalities) has an equivalent SDP formulation of exponential size in m. They also obtain an improving sequence of polynomial size SDP relaxations. Based on Putinar [30], polynomial positivity in the dual formulation are relaxed to sum of square conditions (the two are not equivalent in the multivariate case), which in turn are expressed as SDPs. These results lead to a natural multivariate generalization of our main result (Theorem 4). The proof follows the same principles, and is omitted here for brevity.

A multivariate version [**B***] of Assumption [**B**] is required, where $\Omega \subseteq \mathbf{R}^m$ is closed and the generating set $\mathcal{T}$ of probability measures μ_t is parametrized by a **vector** $t \in I_T \subseteq \mathbf{R}^m$, and I_T is polyhedral.

Theorem 5 *Suppose that the Slater condition holds for Problem (P) satisfying Assumptions [**A**] and [**B***]. Then the following results are true:*

(a) *Problem (P) can be approximated by a converging sequence of SDP relaxations whenever the set $\{y \mid \int (y'f - \phi)d\mu_t \geq 0, \forall\, t \in I_T \subseteq \mathbf{R}^m\}$ is semi-algebraic.*

(b) *There exists a probability measure achieving the optimal bound for this problem which is a convex combination of $n + 1$ measures from the generating set $\mathcal{T}$.*

We use this theorem to extend the results of Section 4 for various multivariate generalizations of unimodality and symmetry. A very good reference on multivariate unimodality is Dharmadhikari and Joag-Dev [8], from which we adopt the concepts and generating classes described below. Throughout this section, generating classes refer to both integral and closed convex hull representations.

Definition 4 *A set $S \in \mathbf{R}^m$ is said to be star-shaped about $s \in S$ if for every $x \in S$, the segment $\langle x, s \rangle$ is completely contained in S.*

Definition 5 *A set $S \in \mathbf{R}^m$ is centrally symmetric if $x \in S$ implies $-x \in S$. A distribution μ on $\mathbf{R}^m$ is centrally symmetric if $\mu(A) = \mu(-A)$ for all Borel sets A in $\mathbf{R}^m$.*

Star unimodality. A probability measure on $\mathbf{R}^m$ is star unimodal about 0 if it belongs to the closed convex hull of the set of all uniform distributions on sets $S \subseteq \mathbf{R}^m$ which are star shaped about 0. Intuitively, the corresponding density function (if it exists) is decreasing along any ray away from the origin. The following parametric generating class satisfies the assumptions of Theorem 5:

$$\mathcal{T}_0^{\mathtt{s-u}} = \{\mu_t = \delta_{\langle 0, t \rangle} \mid t \in \mathbf{R}^m\},$$

where $\delta_{\langle 0, t \rangle}$ is the rectangular probability measure on the segment $\langle 0, t \rangle$. We include the Dirac at 0 as a degenerate case $\delta_0 = \delta_{\langle 0, 0 \rangle}$.

The corresponding dual feasible set is given by:

$$\{y \mid \int_{\langle 0, t \rangle} (y'f(x) - \phi(x))dx = \int_0^1 (y'f_t(a)da - \phi_t(a))da \geq 0 \, , \ \forall t \in \mathbf{R}^m\},$$

where we denoted $h_t(a) = h(ta) \, , \ t \in \mathbf{R}^r \, , \ a \in [0, 1]$.

Centrally symmetric star unimodality. The class of centrally symmetric star unimodal probability measures about 0 is the closed convex hull of uniform distributions on centrally symmetric star-shaped sets about 0. This class is generated by the parametric family:

$$\mathcal{T}_0^{\texttt{cs-u}} = \{\mu_t = \delta_{\langle -t,t\rangle} \,|\, t \in \mathbf{R}^m\},$$

where $\delta_{\langle -t,t\rangle}$ is the rectangular probability measure on the segment $\langle -t, t\rangle$.

The dual feasible set is given by

$$\{y \,|\, \int_{\langle -t,t\rangle} (y'f(x) - \phi(x))dx = \int_{-1}^{1} (y'f_t(a) - \phi_t(a))da \ge 0 \,,\; \forall t \in \mathbf{R}^m\}.$$

Block unimodality. A probability measure on $\mathbf{R}^m$ is block unimodal about 0 if it belongs to the closed convex hull of the set of all uniform distributions on sets $R \subseteq \mathbf{R}^m$ which are rectangles containing 0 and having edges parallel to the coordinate axes. This is a subset of the class of star unimodals, that can be generated by the parametric family:

$$\mathcal{T}_0^{\texttt{b-u}} = \{\mu_t = \rho_{\langle 0,t\rangle} \,|\, t \in \mathbf{R}^m\},$$

where $\rho_{\langle 0,t\rangle}$ is the uniform distribution on the rectangle with edges parallel with the axes and with opposite vertices 0 and t. Equivalently, $\rho_{\langle 0,t\rangle}$ is the distribution of $(U_1 t_1, \ldots, U_n t_n)$ where U_i are independent and uniform on $(0, 1)$.

The dual feasible set is given by

$$\{y \,|\, \iint_0^t (y'f(x) - \phi(x))dx = \int_0^{t_1}\cdots\int_0^{t_n} (y'f(x) - \phi(x))dx \ge 0 \,,\; \forall t \in \mathbf{R}^m\}.$$

Centrally symmetric block unimodality. The class of centrally symmetric block unimodals about 0 can be generated by the following set:

$$\mathcal{T}_0^{\texttt{cb-u}} = \{\mu_t = \rho_{\langle -t,t\rangle} \,|\, t \in \mathbf{R}^m\},$$

where $\rho_{\langle -t,t\rangle}$ is the uniform distribution on the rectangle with edges parallel with the axes and opposite vertices $-t$ and t. The dual feasible set is given by

$$\{y \,|\, \iint_{-t}^t (y'f(x) - \phi(x))dx = \int_{-t_1}^{t_1}\cdots\int_{-t_n}^{t_n} (y'f(x) - \phi(x))dx \ge 0 \,,\; \forall t \in \mathbf{R}^m\}.$$

By Theorem 5, we can obtain a converging sequence of SDP relaxations for the moment

problems for star and block unimodal distributions and their centrally symmetric analogues, provided that the dual feasible set in the above problems is semi-algebraic. In particular, this is the case if f and ϕ are polynomial, or piecewise polynomial on multidimensional rectangles. The latter follows by a natural multivariate extension of Proposition 2.

Other unimodal classes, such as *linear unimodal* or *convex unimodal* measures are not convex, therefore cannot be treated with our approach. On the other hand, the **spherically symmetric unimodal** probability measures form a closed convex set that can be parametrized using a single real parameter. This coincides with the classes of spherically symmetric star-shaped, block and convex unimodals, respectively, and can be generated using uniforms on disks $D(t) \subseteq \mathbf{R}^m$ of radius $t \in \mathbf{R}_+$ centered at 0. The dual feasible set is given by:

$$\{y \mid \iint_{D(t)} (y'f(x) - \phi(x))dx \geq 0 \, , \, \forall \, t \in \mathbf{R}_+\},$$

By changing to polar coordinates, the dual feasible set can be characterized by inequalities in terms of univariate polynomials (in t), whenever f and ϕ are polynomial on rectangles. Therefore Theorem 4 provides an equivalent SDP formulation for this problem.

7 Generalizations of Chebyshev's Inequality

In this section, we obtain optimal upper bounds on the survival distribution (sdf) $P(X \geq a)$ of a random variable X satisfying standard moment constraints, for the case when X is symmetric, respectively symmetric and unimodal. We generalize some known extensions of Chebyshev's mean-variance bound for higher order moments and show how these bounds can be computed by solving an SDP. Our numerical results compare the performance of these bounds relative to those from Bertsimas and Popescu [3], and relative to the normal distribution.

7.1 Chebyshev Bounds for Symmetric and Unimodal Distributions

Given the mean and variance of the random variable X, the best bound on the upper tail $P(X \geq a)$ is given by the one-sided Chebyshev inequality (see [41] or [3]), which we present here for completeness. Bertsimas and Popescu [3] generalize this result to the case of an arbitrary number of moments. The two sided version of the mean-variance bound for unimodal (not necessarily symmetric) random variables is known as the Gauss inequality (see Karlin and Studden [16]). One can easily show that this is the same as the optimal one-sided bound for symmetric and unimodal random variables. We generalize this bound for higher order moments. The proof of the following result is given in Appendix C.

Proposition 10 *Given the mean M and variance σ^2 of a random variable X, the following bounds on $P(X \geq a)$ are optimal:*

	$\sup P(X \geq a)$	
X	$a > M$	$a \leq M$
arbitrary	$\dfrac{\sigma^2}{\sigma^2 + (M-a)^2}$	1
symmetric	$\dfrac{1}{2} \min\left\{ 1, \dfrac{\sigma^2}{(M-a)^2} \right\}$	1
symmetric unimodal	$\dfrac{1}{2} \min\left\{ 1, \dfrac{4}{9}\dfrac{\sigma^2}{(M-a)^2} \right\}$	1

Given any number of moments $M = (M_0 = 1, M_1, \ldots M_{2n})$ of a random variable X, the optimal upper bounds on $P(X \geq a)$ can be computed as semidefinite programs as follows:

(a) *For arbitrary random variables:*

$$
\begin{aligned}
\min \quad & \sum_{i=0}^{n} y_i M_i \\
\text{s.t.} \quad & 0 = \ diag_{2l-1}(U), && l = 1, \ldots, 2n, \\
& y_0 - 1 + \sum_{i=1}^{n} y_i a^i = diag_0(U), \\
& \sum_{i=l}^{n} \binom{i}{l} a^{i-l} y_i = \ diag_{2l}(U), && l = 1, \ldots, 2n, \\
& 0 = \ diag_{2l-1}(V), && l = 1, \ldots, 2n, \\
& (-1)^l \sum_{i=l}^{n} \binom{i}{l} a^{i-l} y_i = \ diag_{2l}(V), && l = 0, \ldots, 2n, \\
& U_{(n \times n)}, V_{(n \times n)} \succeq 0.
\end{aligned}
\tag{14}
$$

where $\ diag_d(X) = \displaystyle\sum_{i,j: \ i+j=d} x_{ij}$ *to denotes the anti-diagonal sums of the matrix X.*

(b.1) *For symmetric random variables with mean $M_1 = 0$ and $a \leq 0$:*

$$
\begin{aligned}
\min \quad & \sum_{i=0}^{n} y_i M_{2i} \\
\text{s.t.} \quad & 0 = \ diag_{2l-1}(U) = \ diag_{2l-1}(V), && l = 1, \ldots, n, \\
& y_l = \ diag_{2l}(U), && l = 1, \ldots, n, \\
& y_0 - \frac{1}{2} = \ diag_0(U), \\
& (y_0 - 1)\binom{n}{l} + \sum_{i=1}^{l} y_i a^{2i} = \ diag_{2l}(V) && l = 1, \ldots, n, \\
& U_{(n \times n)}, V_{(n \times n)} \succeq 0
\end{aligned}
\tag{15}
$$

(b.2) *For symmetric random variables with mean $M_1 = 0$ and $a > 0$:*

$$
\begin{aligned}
\min \quad & \sum_{i=0}^{n} y_i M_{2i} \\
\text{s.t.} \quad & 0 = \mathrm{diag}_{2l-1}(U) = \mathrm{diag}_{2l-1}(V), & l = 1, \ldots, n, \\
& y_l = \mathrm{diag}_{2l}(U), & l = 0, \ldots, n, \\
& \sum_{i=l}^{n} \binom{i}{l} y_i a^{2(i-l)} = \mathrm{diag}_{2l}(V) & l = 1, \ldots, n, \\
& (y_0 - \tfrac{1}{2}) + \sum_{i=1}^{l} y_i a^{2i} = \mathrm{diag}_0(V) \\
& U_{(n\times n)}, V_{(n\times n)} \succeq 0
\end{aligned}
\tag{16}
$$

(c.1) *For unimodal symmetric random variables with mean $M_1 = 0$ and $a \leq 0$:*

$$
\begin{aligned}
\min \quad & \sum_{i=0}^{n} y_i M_{2i} \\
\text{s.t.} \quad & 0 = \mathrm{diag}_{2l-1}(U), & l = 1, \ldots, n, \\
& (y_0 - 1)\binom{n}{l} + \sum_{i=1}^{l} \frac{y_i}{2i+1}\binom{n-l}{l-i} a^{2i} = \mathrm{diag}_{2l}(U), & l = 0, \ldots, n, \\
& 0 = \mathrm{diag}_{2l-1}(V), & l = 1, \ldots, 2n, \\
& \sum_{i=\lfloor \frac{l-1}{2} \rfloor}^{n} \frac{y_i}{2i+1}\binom{2i+1}{l}(-a)^{2i+1-l} = \mathrm{diag}_{2l}(V) & l = 2, \ldots, 2n, \\
& y_0 - \frac{1}{2} + \sum_{i=1}^{n} y_i a^{2i} = \mathrm{diag}_2(V) \\
& -(y_0 - 1)a - \sum_{i=0}^{n} \frac{y_i}{2i+1} a^{2i+1} = \mathrm{diag}_0(V) \\
& U_{(n\times n)}, V_{((2n+1)\times(2n+1))} \succeq 0
\end{aligned}
\tag{17}
$$

(c.2) *For unimodal symmetric random variables with mean $M_1 = 0$ and $a > 0$:*

$$
\begin{aligned}
\min \quad & \sum_{i=0}^{n} y_i M_{2i} \\
\text{s.t.} \quad & 0 = \mathrm{diag}_{2l-1}(U), & l = 1, \ldots, n, \\
& \frac{y_l}{2l+1} = \mathrm{diag}_{2l}(U), & l = 0, \ldots, n, \\
& 0 = \mathrm{diag}_{2l-1}(V), & l = 1, \ldots, 2n, \\
& 0 = \mathrm{diag}_{4l}(V), & l = 1, \ldots, n, \\
& \frac{y_l}{2l+1} = \mathrm{diag}_{4l+2}(V), & l = 1, \ldots, n, \\
& y_0 - \frac{1}{2} = \mathrm{diag}_2(V), \\
& \frac{a}{2} = \mathrm{diag}_0(V), \\
& U_{(n\times n)}, V_{((2n+1)\times(2n+1))} \succeq 0
\end{aligned}
\tag{18}
$$

23

7.2 Numerical results

This numerical study compares the sharp bounds for arbitrary, symmetric, respectively symmetric unimodal distributions, given higher order moments.

The standard normal survival distribution function (sdf) $N(a) = 1 - \Phi(a)$ for $a \geq 0$ is benchmarked against moment bounds on the survival distribution $P(X \geq a)$ for arbitrary, symmetric, and symmetric unimodal distributions, with the same moments[6] up to order k, where $k = 2, \ldots, 10$. Figure 1 plots the corresponding upper bounds $A_k(a), S_k(a), SU_k(a)$ and the normal sdf $N(a)$ as functions of $a \geq 0$; the three plots correspond to $k = 2, 4$ and respectively 10 moments. While all bounds perform well in the tails (beyond 3-4 std), the bound for symmetric and unimodal distributions exhibits a good performance also around the mean.
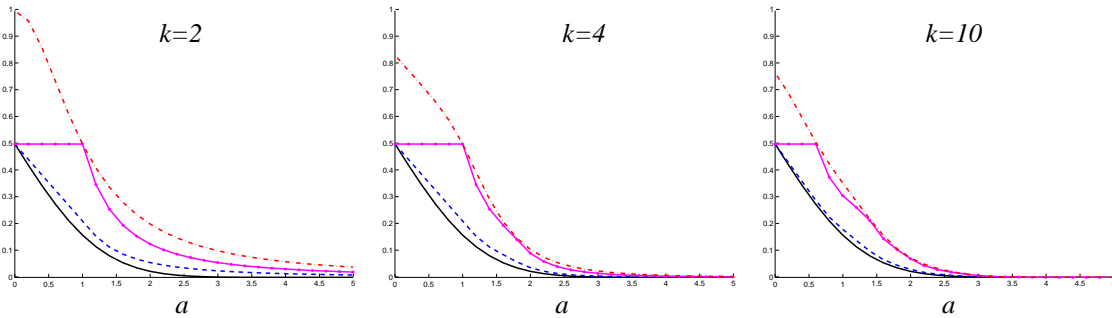

Figure 1: From left to right: bounds given $k = 2, 4, 10$ moments respectively. Each plot compares $A_k(a), S_k(a), SU_k(a)$ and $N(a)$ for $a \in [0, 5]$.
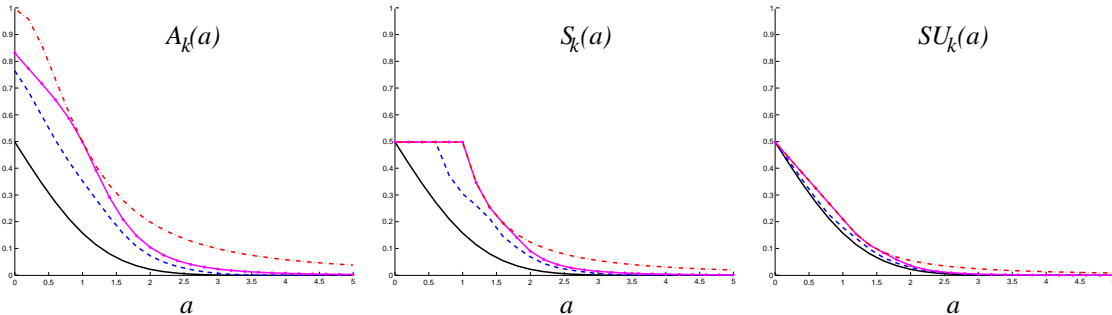

Figure 2: From left to right: $A_k(a)$, $S_k(a)$ and $SU_k(a)$, $a \in [0, 5]$. Each plot compares sharp bounds given $k = 2, 4, 10$ moments vs. the normal sdf.

In Table 1 we compare the improvement of the bound for symmetric unimodal distributions over the bound for arbitrary distributions, relative to the normal sdf benchmark, calculated as $\Delta_k(a) = (A_k(a) - SU_k(a))/(A_k(a) - N(a))$. Remarkably, $\Delta_k(a)$ is consistently

---

[6]The first ten moments of the standard normal distribution are $M = (0, 1, 0, 3, 0, 15, 0, 105, 0, 945)$.

a	$\Delta_2(a)$	$\Delta_4(a)$	$\Delta_6(a)$	$\Delta_8(a)$	$\Delta_{10}(a)$
0	100	100	100	100	100
0.2	96.01	93.92	96.27	95.46	96.69
0.4	92.29	89.33	93.13	92.33	94.24
0.6	88.61	86.32	91.16	90.95	93.77
0.8	85.62	84.82	91.82	91.79	93.27
1.0	84.59	84.59	90.91	90.20	88.98
1.2	86.70	85.95	87.01	86.39	85.99
1.4	87.32	84.59	82.90	87.19	87.16
1.6	85.85	80.26	79.80	89.04	88.37
1.8	83.64	79.88	79.75	88.81	86.58
2.0	81.49	83.03	81.48	86.79	85.24
2.2	79.66	85.28	81.35	83.02	86.15
2.4	78.24	86.20	79.17	79.03	85.10
2.6	77.29	86.26	80.60	78.92	82.56
2.8	76.74	85.56	82.71	78.30	79.25
3.0	76.29	84.44	82.42	75.81	75.00
3.2	76.22	84.27	82.54	80.56	78.13
3.4	76.17	83.92	82.22	81.82	76.47
3.6	76.33	84.35	83.87	84.62	75.00
3.8	76.35	84.04	82.61	87.50	75.00
4.0	76.36	84.42	82.35	83.33	76.67
4.2	76.49	84.38	84.62	75.00	100
4.4	76.58	84.91	80.00	100	100
4.6	76.72	84.09	85.71	100	100
4.8	76.92	86.49	83.33	100	100
5.0	76.88	84.38	75.00	100	100

Table 1: Percentage improvement of the bound for symmetric unimodal distributions over the bound for arbitrary distributions (relative to the normal), given k =2,4,6,8, and 10 moments.

(i.e. for all a and k) above 75%. It is higher around the mean, for any number of moments, and it is also higher in the tails for moments of higher order.

It is interesting to understand how much the various bounds improve as higher order moments are given. Each plot in Figure 2 compares the bounds given 2, 4 and 10 moments (against the normal sdf benchmark) for a category of distributions: arbitrary, symmetric, respectively unimodal and symmetric. The value of adding higher order moment information appears to be higher under less distributional assumptions. In fact, for unimodal and symmetric distributions, the mean-variance bound ($k = 2$) is already fairly strong. This

a	$\Delta_2^{\mathtt{su}}(a)$	$\Delta_4^{\mathtt{su}}(a)$	$\Delta_6^{\mathtt{su}}(a)$	$\Delta_8^{\mathtt{su}}(a)$
0.2	0	0	0	0
0.4	0	41.20	0	29.92
0.6	0	44.61	0	33.94
0.8	0	50.86	0	43.80
1.0	0	62.41	0	33.02
1.2	0	58.94	0	0.46
1.4	0	29.34	13.72	0
1.6	0	2.76	44.16	0
1.8	0.0469	0	60.66	0.83
2.0	0.3089	0	57.08	0
2.2	0.5732	0	37.14	14.77
2.4	0.7156	0	10.99	38.27
2.6	78.62	0	0	23.08
2.8	84.05	43.90	0	4.35
3.0	85.04	54.29	6.25	0
3.2	86.67	60.71	36.36	0
3.4	87.83	65.22	50.00	0
3.6	89.35	72.22	60.00	0
3.8	90.20	73.33	75.00	0
4.0	91.37	75.00	66.67	0
4.2	92.06	80.00	50.00	100
4.4	93.04	75.00	100	NaN
4.6	93.33	85.71	100	NaN
4.8	94.79	80.00	100	NaN
5.0	94.38	80.00	100	NaN

Table 2: Percentage improvement in the bound for symmetric unimodal distributions relative to the normal sdf due to higher moment information. (NaN stand for division by zero.)

also shows that approximating a unimodal and symmetric distribution by a normal with the same mean and variance yields a reasonably good approximation.

Another interesting observation is that fourth moment information has a most significant impact in the tails (beyond $2 - 2.5$ standard deviations), but less so around the mean. This can also be observed from Table 2, which illustrates the relative improvement in the bound for symmetric unimodal distributions from using higher order moments $\Delta_k^{\tt su}(a) = (SU_k(a) - SU_{k+2}(a))/(SU_k(a) - N(a))$. The relative improvement from using higher order moments appears to be much stronger in the tails than around the mean. Nevertheless, no monotonicity relationship can be observed.

The numerical experiments have been conducted in MATLAB5.3 under WindowsNT, using the SOS Toolbox [27]- an optimization package over semi-algebraic sets, based on the SeDuMi [39] solver for semidefinite programming.

8   Conclusions.

In this paper we showed how optimal moment bounds over convex classes of distributions generated by one-parameter families can be efficiently computed using semidefinite programming. Polar representations and conic duality provide a general and tractable framework for solving Problem (P) for convex classes of distributions, satisfying special properties such as symmetry, unimodality, convexity and smoothness. We also extended these results to obtain approximate SDP solutions for multi-parameter classes and multivariate distributions. Finally, we applied these results to obtain generalizations of Chebyshev's inequality given higher order moments via SDP. Numerical computation shows that accounting for structural properties such as symmetry and unimodality can improve the quality of the moment bounds by at least 75%.

Acknowledgements

References

[1] N.I. Akhiezer and M. Krein. *Some Questions in the Theory of Moments*, volume 2 of *Translations of Math. Monographs*. American Math. Soc., Prov, R.I., 1962.

[2] S. Bell. Approximating the normal distribution with the triangular. Technical Report 494, Sandia Corporation Report, 1962.

[3] D. Bertsimas and I. Popescu. Optimal inequalities in probability: A convex programming approach. Working paper, INSEAD, http://faculty.insead.edu/popescu/ioana, 1998.

[4] D. Bertsimas and I. Popescu. On the relation between option and stock prices: An optimization approach. *Operations Research*, 50(2):358–374, 2002.

[5] P. Billingsley. *Weak Convergence of Measures: Applications in Probability.* SIAM, CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia, PA, 1971.

[6] N.P. Buslenko, D.I. Golenko, Yu.A. Shreider, I.M. Sobol, and V.G. Sragovich. *The Monte Carlo Method.* Oxford: Pergamon, 1966.

[7] P. Chebyshev. Sur les valeurs limites des intégrales. *Jour. Math. Pur. Appl.*, 19(2):157–160, 1874.

[8] S. Dharmadhikari and K. Joag-Dev. *Unimodality, convexity and applications.* Probability and Mathematical Statistics. Academic Press, San Diego, CA, 1988.

[9] R.M. Dudley. *Real Analysis and Probability.* Chapman and Hall Mathematics Series, New York, 1989.

[10] R.J. Duffin. Infinite programs. In H. W. Kuhn and A. W. Tucker, editors, *Linear inequalities and related systems*, pages 157–170. Princeton University Press, Princeton, NJ, 1956.

[11] G.A. Edgar. personal communication, March 2004.

[12] G. Gallego. A minmax distribution free procedure for the (q,r) inventory model. *Oper. Res. Letters*, 11:55–60, 1992.

[13] L. El Ghaoui, F. Oustry, and M. Oks. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, July-August 2003.

[14] B. Grundy. Option prices and the underlying asset's return distribution. *Journal of Finance*, 46(3):1045–1070, July 1991.

[15] K. Isii. On the sharpness of chebyshev-type inequalities. *Ann. Inst. Stat. Math.*, 14:185–197, 1963.

[16] S. Karlin and W.J. Studden. *Tchebycheff Systems: with Applications in Analysis and Statistics.* Pure and Applied Mathematics, A Series of Texts and Monographs. Interscience Publishers, John Wiley and Sons, 1966.

[17] A.F. Karr. Extreme points of certain sets of probability measures with applications. *Math. of O.R.*, 8(1):74–85, 1983.

[18] A.Y. Khintchine. On unimodal distributions. *Izv. Naucho-Isled. Inst. Mat. Mech. Tomsk. Gos. Univ.*, (2):1–7, 1938.

[19] G. Lanckriet, L. El Ghaoui, C. Bhattacharya, and M. Jordan. Minimax probability machine. working paper, University of California, Bekeley, 2001.

[20] H.J. Landau. Classical background on the moment problem. In ed. H.J. Landau, editor, *Moments in Mathematics*, number 37 in Proc. Sympos. Appl. Math., Series III, pages 1–15, Providence, RI, 1987. American Mathematical Society.

[21] M. LiCalzi. Upper and lower bounds for expected utility. *Economic Theory*, 16:489–502, 2000.

[22] A. Lo. Semiparametric upper bounds for option prices and expected payoffs. *Journal of Financial Economics*, 19:373–388, 1987.

[23] C.L. Mallows. Generatizations of tchebysheff's inequalities. *J. Roy. Stat. Soc.*, 18(2):139–168, 1956.

[24] A. Markov. *On Certain Applications of Algebraic Continued Fractions*. PhD thesis, St. Petersburg, 1884. in Russian.

[25] H.P. Mulholland and C.A. Rogers. Representation theorems for distribution functions. *Proc. London Math. Soc.*, 3(8):177–223, 1958.

[26] Y. Nesterov. Structure of non-negative polynomial and optimization problems. Preprint DP 9749, Louvan-la-Neuve, Belgium, 1997.

[27] P. Parrilo. Sos toolbox (algorithm for optimization over semi-algebraic sets). ETH, Zurich, Switzerland, http://www.cds.caltech.edu/sostools/, 2002.

[28] K.R. Parthasarathy. *Probability Measures on Metric Spaces*. Academic Press, New York etc., 1967.

[29] R.R. Phelps. *Lectures on Choquet's Theorem*. van Nostrand Mathematical Studies, New York etc., 1966.

[30] M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.

[31] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[32] W.W. Rogosinsky. Moments of non-negative mass. *Proc. Roy. Soc. London. Ser. A.*, 245:1–27, 1958.

[33] H. Scarf. A min-max solution of an inventory problem. In *Studies in the mathematical theory of inventory and production*, pages 201–209, Stanford University Press, Stanford, CA, 1958. K.J. Arrow, S. Karlin and H. Scarf, eds.

[34] A. Shapiro. On duality theory of conic linear problems. In M.A. Goberna and M.A. Lopez, editors, *Semi-infinite programming: recent advances*, pages 135–165. Kluwer Academic Publishers, 2001.

[35] J.A. Shohat and J.D. Tamarkin. The problem of moments. In *Mathematical Surveys*. American Mathematical Society, New York, 1943.

[36] J. Smith. Generalized chebyshev inequalities: Theory and applications in decision analysis. *Operations Research*, 43(5):807–825, 1995.

[37] J. Soll and J. Klayman. Measuring overconfidence. INSEAD working paper, 2001.

[38] T.J. Stieltjes. Recherches sur les fractions continues. *Annales de la Faculté de Sciences de Toulouse*, 9(1):5–47, 1895.

[39] J. Sturm. Sedumi (semidefinite programming algorithm), version 1.05. Department of Econometrics, Tilburg University, The Netherlands, http://fewcal.kub.nl/sturm/software/sedumi.html, 1998.

[40] Y.L. Tong. *Probability Inequalities in Multivariate Distributions*. New York, Academic Press, 1980.

[41] J.V. Uspensky. *Introduction to Mathematical Probability*. McGraw-Hill, New York, 1937.

[42] H. Weizsäcker and G. Winkler. Integral representation in the set of solutions of a generalized moment problem. *Math. Ann.*, 246:23–32, 1979.

[43] Y. Willassen. Best upper and lower tchebycheff bounds on expected utility. *Review of Economic Studies*, (57):513–520, 1990.

[44] G. Winkler. On the integral representation in convex noncompact sets of tight measures. *Math. Z.*, 158:71–77, 1978.

[45] H. Wolkowicz, R. Saigal, and L. Vandengerghe editors. *Handbook of Semidefinite Programming, Theory, Algorithms, and Applications*. Kluwer Academic Publishers, 2000.

[46] Y. Yamada and J. Primbs. Value-at-risk estimation for dynamic hedging. California Institute of Technology working paper, May 2000.

9 Appendix A. Proof of Proposition 3

The first part of the proposition is a known result. We provide a proof for completeness:

Clearly $\mathbf{cx}(\mathcal{T}) \subseteq \mathbf{mix}(\mathcal{T})$. To prove $\mathbf{mix}(\mathcal{T}) \subseteq \bar{\mathbf{cx}}(\mathcal{T})$, let $\mu = \int \tau d\nu(\tau) \in \mathbf{mix}(\mathcal{T})$. The set of measures with finite support is weakly dense in $\mathcal{M}^+(\mathcal{T})$ (a separable metric space[7]), so there exists $\nu_n \to \nu$ such that ν_n has support on (at most) n points. Let $\mu_n = \int \tau d\nu_n(\tau) \in \mathbf{cx}(\mathcal{T})$. For any bounded continuous h, $\int h d\tau$ is bounded and continuous in $\tau \in \mathcal{T}$ (in the inherited weak topology), implying $\int h d\mu_n = \int (\int h d\tau) d\nu_n \to \int (\int h d\tau) d\nu = \int h d\mu$, so $\mu \in \bar{\mathbf{cx}}(\mathcal{T})$. This concludes the proof of the first part.

The second part of Proposition 3 is not new in the compact case. It follows from the following result (Phelps [29] Proposition 1.2):

Proposition 11 *Suppose that X is a compact subset of a locally convex[8] topological space. Then $\bar{\mathbf{cx}}(X) = \mathbf{mix}(X)$.*

Since the space of probability measures can be appropriately metrized, weak compactness is sufficient. By Prohorov's Theorem (see [5]), a sufficient condition for weak compactness is that $\mathcal{T}$ be weakly closed and *uniformly tight*, i.e. for any $\epsilon > 0$ there exists a compact K_ϵ with $\tau(K_\epsilon) > 1 - \epsilon$ for every $\tau \in \mathcal{T}$. Uniform tightness is satisfied if Ω is compact.

For non-compact intervals $\Omega \subseteq \mathbf{R}$, the following proof has been suggested by Edgar [11]. Consider for example $\Omega = [0, \infty)$ (the other cases work similarly), which is not compact, but can be compactified in $\bar{\mathbf{R}}$ by adding the point $\{\infty\}$. This yields $\bar{\Omega} = [0, \infty]$, which is homeomorphic to $[0, 1]$. The set of probability measures on Ω can be identified with the following subset of probability measures on $\bar{\Omega}$: $\mathcal{M}^+(\Omega) = \{\mu \in \mathcal{M}^+(\bar{\Omega}) \mid \mu(\{\infty\}) = 0\}$.

Let $\mathcal{S}$ denote the closure of $\mathcal{T}$ in $\mathcal{M}^+(\bar{\Omega})$. Since $\mathcal{T}$ is closed in $\mathcal{M}^+(\Omega)$, it follows that $\mathcal{T} = \{\mu \in \mathcal{S} \mid \mu(\{\infty\}) = 0\}$. By Proposition 11 we have $\bar{\mathbf{cx}}^*(\mathcal{S}) = \mathbf{mix}(\mathcal{S})$, where the star refers to closure in $\mathcal{M}^+(\bar{\Omega})$.

We want to show that $\bar{\mathbf{cx}}(\mathcal{T}) \subseteq \mathbf{mix}(\mathcal{T})$, where the closure is in $\mathcal{M}^+(\Omega)$. Let $\mu_0 \in \bar{\mathbf{cx}}(\mathcal{T})$. Then $\mu_0 \in \bar{\mathbf{cx}}^*(\mathcal{S})$ and $\mu_0(\{\infty\}) = 0$. Let $\nu_0 \in \mathcal{M}^+(\mathcal{S})$ be a mixing measure for μ_0. It remains to show that ν_0 is concentrated on $\mathcal{T}$. For this we rely on the following result (Phelps [29] Chapter 12 p.100):

Proposition 12 *If X is a compact convex subset of a locally convex space and if ν is a probability measure on X with resultant x, then $f(x) = \int f d\nu$ for each affine function f of first Baire class.*

Affine functions of first Baire class are those affine functions that are the pointwise limit of a sequence of continuous (not necessarily affine) functions on X. The map $\mu \to \mu(\{\infty\})$ is affine and of first Baire class (although not continuous). To see this, use continuous

---

[7]It is possible to metrize the space of probability measures as a separable complete metric space such that convergence in the metric is equivalent to weak convergence (see [28] Theorem 6.2 Ch.II)

[8]A space is locally convex if it admits a convex base. Any metric space is locally convex.

increasing functions $f_n(x)$ that are zero for $x < n$, and one for $x > n+1$. Then each $\int f_n d\mu$ is continuous, and $\mu(\{\infty\})$ is the limit of that sequence.

Therefore, by Proposition 12, $0 = \mu_0(\{\infty\}) = \int_{\mathcal{S}} \mu(\{\infty\}) d\nu_0(\mu)$. This integral of a non-negative function is zero, so the set $\{\mu \in \mathcal{S} \mid \mu(\{\infty\}) = 0\} = \mathcal{T}$ has ν_0-measure 1. This shows that ν_0 is supported on $\mathcal{T}$, so $\mu_0 \in \mathbf{mix}(\mathcal{T})$, as claimed.

The above argument easily extends for closed sets $\Omega \subseteq \mathbf{R}^m$, using the one-point compactification of $\mathbf{R}^m$ (i.e. add one extra point ∞, with neighborhoods consisting of the complements of the compact sets of $\mathbf{R}^m$). This result becomes relevant in Section 6, when we deal with multivariate extensions.

Appendix B. Proofs of Lemmas 6 and 7.

Proof of Lemma 6: We only prove part (a); the second part is analogous. Any measure in $\mathcal{S} = \mathbf{cx}(\mathcal{T}_{m,a}^{\mathtt{dx}})$ admits a continuous piecewise linear decreasing and convex density. Conversely, any continuous piecewise linear decreasing and convex density function on $[m, m+a]$, can be written as $p(x) = s_0 + \sum_{i=1}^{l} s_i(b_i - x)^+$, with $s_i \geq 0$ and $m \leq b_l \leq \cdots \leq b_1 \leq m + a$, where the left-slopes at breakpoints b_k equal $-\sum_{i=1}^{k} s_i$. Denote $t_i = b_i - m$ and $w_i = s_i t_i^2 / 2$, so $p(x) = s_0 + \sum_{i=1}^{l} w_i \frac{2}{t_i^2}(m + t_i - x)^+$. Since $p(x)$ is a density, the weights sum up to 1. Hence the measure with density $p(x)$ is a convex combination of triangulars and a rectangular (with weight s_0), so it belongs in $\mathcal{S}$. Hence $\mathcal{S}$ is the set of continuous probability measures that admit piecewise linear decreasing and convex densities.

We now show that $\mathcal{S} \subseteq \bar{\mathcal{P}}_{m,a}^{\mathtt{dx}} \subseteq \bar{\mathcal{S}}$, hence $\bar{\mathcal{P}}_{m,a}^{\mathtt{ix}} = \bar{\mathcal{S}}$ as desired. Let $\mu \in \bar{\mathcal{P}}_{m,a}^{\mathtt{ix}}$, so $\mu = \alpha\mu_0 + (1-\alpha)\delta_m$, where μ_0 is a continuous measure that admits an increasing convex density π_0. There exists an increasing sequence of non-negative piecewise linear decreasing concave measurable functions π_n converging pointwise to π_0. By monotone convergence, $\int_A \pi_n(x) dx \to \int_A \pi_0(x) dx$ for any measurable set A. Let $\int \pi_n(x) = \frac{1}{c_n} \leq 1$, so $c_n \to 1$ and $c_n \pi_n$ is the density function of a measure in $\mathcal{S}$. Therefore the sequence of measures $\mu_n = \alpha c_n \int \pi_n(x) dx + (1-\alpha)\gamma_{m,\frac{1}{n}}^+ \in \mathcal{S}$ converges weakly to μ_0.

Finally, by Proposition 3, we have $\bar{\mathcal{P}}_{m,a}^{\mathtt{dx}} = \mathbf{mix}(\mathcal{T}_{m,a}^{\mathtt{dx}} \cup \{\delta_m\})$, where the Dirac at m is added to insure closure of the generating set. Without it, we obtain the corresponding set of continuous measures $\mathcal{P}_{m,a}^{\mathtt{dx}} = \mathbf{mix}(\mathcal{T}_{m,a}^{\mathtt{dx}})$.

$\square$

Proof of Lemma 7: We only prove the first part, the second part is analogous. The proof follows the same lines as the previous lemma. Any measure in $\mathcal{S} = \mathbf{cx}(\mathcal{T}_{m,a}^{\mathtt{dv}})$ admits a continuous piecewise linear decreasing and concave density. Any continuous piecewise linear decreasing and concave density function $v(x) \geq 0$ on $[m, m + a]$ can be written as $v(x) = M - \sum_{i=0}^{l} w_i(x - m - t_i)^+$, where $w_i \geq 0$, $0 \leq t_i \leq a$ and $M = v(m + a)$ is the

32

maximal value of v. It follows that

$$v(x) = \sum_{i=0}^{l} w_i(a - t_i - (x - m - t_i)^+) = \sum_{i=0}^{l} w_i \min(a - t_i, a + m - x),$$

which is a convex combination of right m-trapezoidal densities with weights $\rho_i = w_i \frac{a^2 - t_i^2}{2} \geq 0$. Since v is a density, the weights must sum to 1. This shows that $\mathcal{S}$ is exactly the set of distributions that admit a piecewise linear decreasing and concave density. We now show that $\mathcal{S} \subseteq \mathcal{P}_{m,a}^{\mathsf{dv}} \subseteq \bar{\mathcal{S}}$. Closure of $\mathcal{P}_{m,a}^{\mathsf{dv}}$ implies the desired equality.

Let π_0 be an increasing convex density for $\mu \in \mathcal{P}_{m,a}^{\mathsf{dv}}$. Let π_n be an increasing sequence of non-negative piecewise linear increasing convex measurable functions converging pointwise to π_0. By monotone convergence, $\int_A \pi_n(x)dx \rightarrow \int_A \pi_0(x)dx$ for any measurable set A. Let $\int \pi_n(x) = \frac{1}{c_n} \leq 1$, so $c_n \rightarrow 1$ and $c_n\pi_n$ is the density function of a measure in $\mathcal{S}$. So $c_n \int_A \pi_n(x)dx \rightarrow \int_A \pi_0(x)dx$ for any measurable set A, and the sequence of measures $\mu_n = c_n \int \pi_n(x)dx \in \mathcal{S}$ converges weakly to μ. $\qquad\square$

Appendix C. Proof of Proposition 10.

The result for arbitrary distributions is from Bertsimas and Popescu [3]. In order to simplify notation, in the following we assume that the mean $M_1 = 0$, and hence all odd order moments are also null.

Symmetric random variables. According to the formulation (9) in Section 4.1, the dual of this problem can be written as:

$$\begin{aligned}
\min_{y} \quad & y'M \\
\text{s.t.} \quad & \sum_{i=0}^{n} y_i((-t)^{2i} + t^{2i}) \geq \mathbf{1}_{(t \leq -a)} + \mathbf{1}_{(t \geq a)} \quad, \ \forall\, t \geq 0.
\end{aligned} \tag{19}$$

We must distinguish two cases:

Case 1: $a \leq 0$. By the change of variables $t^2 = z$, we can rewrite the dual as follows:

$$\begin{aligned}
\min_{y} \quad & y'M \\
\text{s.t.} \quad & \sum_{i=0}^{n} y_i z^i \geq \frac{1}{2} \quad, \ \text{for } \ z \geq 0 \\
& \sum_{i=0}^{n} y_i z^i \geq 1 \quad, \ \text{for } \ z \leq a^2.
\end{aligned}$$

The corresponding SDP formulation is a simple application of Proposition 2 (see the explicit

33

formulation in [3]). Suppose now that $n = 1$, and $\sigma^2 = M_2$. The optimal Chebyshev bound for symmetric distributions is the solution of the following program:

$$
\begin{aligned}
\min_{y} \quad & y_0 + y_1 \sigma^2 \\
\text{s.t.} \quad & y_0 + y_1 z \geq \tfrac{1}{2} \quad , \text{ for } z \geq 0 \\
& y_0 + y_1 z \geq 1 \quad , \text{ for } z \leq a^2,
\end{aligned}
$$

The constraints imply that $y_0 \geq 1$ and $y_1 \geq 0$. It follows that the optimal bound is 1.

Case 2: $a > 0$. By the same change of variables $t^2 = z$, the dual can be written as:

$$
\begin{aligned}
\min_{y} \quad & y'M \\
\text{s.t.} \quad & \sum_{i=0}^{n} y_i z^i \geq 0 \quad , \text{ for } z \geq 0 \\
& \sum_{i=0}^{n} y_i z^i \geq \frac{1}{2} \quad , \text{ for } z \geq a^2.
\end{aligned}
$$

The corresponding SDP formulation is a simple application of Proposition 2 as explicitly stated in [3]. Suppose now that $n = 1$, and $\sigma^2 = M_2$. The optimal Chebyshev bound for symmetric distributions is the solution of the following program:

$$
\begin{aligned}
\min_{y} \quad & y_0 + y_1 \sigma^2 \\
\text{s.t.} \quad & y_0 + y_1 z \geq 0 \quad , \text{ for } z \geq 0 \\
& y_0 + y_1 z \geq \tfrac{1}{2} \quad , \text{ for } z \geq a^2,
\end{aligned}
$$

The first constraint set is equivalent to $y_0, y_1 \geq 0$. In this case, the second constraint set is equivalent to $y_0 + y_1 a^2 \geq \frac{1}{2}$. So the problem can be reformulated as

$$
\min_{y_1 \leq \frac{1}{2} a^2} \frac{1}{2} + y_1 (\sigma^2 - a^2) =
\begin{cases}
\dfrac{1}{2} & , \text{ if } \sigma^2 \geq a^2 \\[2mm]
\dfrac{\sigma^2}{2a^2} & , \text{ if } \sigma^2 < a^2
\end{cases}
= \frac{1}{2} \min\left(1, \frac{\sigma^2}{a^2}\right).
$$

This concludes the proof of this part.

Unimodal and symmetric random variables. According to formulation (12) in Section 4.3, the problem can be written as:

$$
\begin{aligned}
\min_{y} \quad & y'M \\
\text{s.t.} \quad & \sum_{i=0}^{n} \frac{2y_i}{2i+1} t^{2i+1} \geq \int_{-t}^{t} \mathbf{1}_{(x \geq a)} dx \quad , \forall\, t \geq 0.
\end{aligned}
\tag{20}
$$

34

We consider two separate cases:

Case 1: $a \leq 0$ **.** The equivalent dual formulation is

$$
\min_{y} \quad y'M
$$
$$
\text{s.t.} \quad \sum_{i=0}^{n} \frac{2y_i}{2i+1} t^{2i+1} \geq \begin{cases} 2t & , \ \forall \, 0 \leq t \leq -a \\ t - a & , \ \forall \, t \geq -a. \end{cases} \tag{21}
$$

The degree of the first constraint set can be reduced after simplification by $2t$ and the change of variable $z = t^2$, to $\sum_{i=0}^{n} \frac{y_i}{2i+1} z^i \geq 1$, $\forall \, 0 \leq z \leq a^2$. The corresponding SDP formulation is a direct consequence of Proposition 2 (see [3] for the explicit formulation).

For $n = 1$, and $\sigma^2 = M_2$, the desired bound is the solution to the following program:

$$
\min_{y} \quad y_0 + y_1 \sigma^2
$$
$$
\text{s.t.} \quad y_0 + \frac{y_1}{3} z \geq 1 \qquad\qquad , \ \forall \, 0 \leq z \leq a^2 \tag{22}
$$
$$
y_0 t + \frac{y_1}{3} t^3 \geq \frac{t - a}{2} \qquad , \ \forall \, t \geq -a
$$

The second constraint implies that $y_1 \geq 0$, whereas the first implies $y_0 \geq 1$. The optimum is achieved when equality holds in both, and the bound equals 1.

Case 2: $a > 0$ **.** The equivalent dual formulation is:

$$
\min_{y} \quad y'M
$$
$$
\text{s.t.} \quad \sum_{i=0}^{n} \frac{2y_i}{2i+1} t^{2i+1} \geq \begin{cases} 0 & , \ \forall \, t \geq 0 \\ t - a & , \ \forall \, t \geq 0 \end{cases} . \tag{23}
$$

We degree of the first constraint can be reduced after dividing by $2t$, and by the change of variable $z = t^2$, obtaining: $\sum_{i=0}^{n} \frac{y_i}{2i+1} z^i \geq 0$, $\forall \, z \geq 0$. The corresponding SDP formulation is a direct consequence of Proposition 2.

In particular, for $n = 1$ and $\sigma^2 = M_2$, we have the following program:

$$
\min_{y} \quad y_1 \sigma^2 + y_0
$$
$$
\text{s.t.} \quad \frac{y_1}{3} z + y_0 \geq 0 \qquad\qquad , \ \forall \, z \geq 0 \tag{24}
$$
$$
\frac{y_1}{3} t^3 + (y_0 - \frac{1}{2}) t + \frac{a}{2} \geq 0 \qquad , \ \forall \, t \geq 0
$$

The first condition set implies $y_0 \geq 0$. The second condition implies $y_1 \geq 0$, in which case the first constraint set is always satisfied. For $y_1 = 0$, or $y_0 \geq \frac{1}{2}$, the bound cannot exceed

$\frac{1}{2}$, obtained for $y_0 = \frac{1}{2}$, $y_1 = 0$.

Suppose now that $y_1 > 0$ and $0 \le y_0 < \frac{1}{2}$. Denote $h(t) = \frac{y_1}{3}t^3 + (y_0 - \frac{1}{2})t + \frac{a}{2}$. Notice that the inflexion point is at 0 and $h(0) = a/2 \ge 0$, whereas $h(a) \ge 0$. The local optima are $t_\pm = \pm\sqrt{(\frac{1}{2} - y_0)/y_1}$. Therefore, we can have $h(t) \ge 0$ for all $t \ge a$ if and only if $h(t_+) \ge 0$, or else $t_+ \le a$. The latter case leads to a contradiction. The condition $h(t_+) \ge 0$ is equivalent to $y_1 \ge \frac{2}{9}\frac{(1-2y_0)^3}{a^2}$. The minimum is achieved for $y_0 = 0$ or $y_0 = \frac{1}{2}$, and the value of the bound is $\min\left(\frac{1}{2}, \frac{2}{9}\frac{\sigma^2}{a^2}\right)$. This concludes the proof. $\square$