# Optimal Inequalities in Probability Theory: A Convex Optimization Approach

Dimitris Bertsimas*and Ioana Popescu,†

Sloan WP #4025
June 1998

*Boeing Professor of Operations Research, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Mass. 02139. Research supported by NSF grant DMI-9610486.

†Department of Mathematics and Operations Research Center, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

# Optimal Inequalities in Probability Theory: A Convex Optimization Approach

Dimitris Bertsimas * and Ioana Popescu †

May, 1998

**Abstract**

We derive optimal inequalities for $P(X \in S)$, for a multivariate random variable $X$ that has a given collection of moments, and $S$ is an arbitrary set. Using convex optimization methods, we find explicit tight bounds that generalize and improve upon the classical Markov, and Chebyshev inequalities, when moments up to second order are known, and the set $S$ is convex. We find extremal distributions that achieve the new bounds, and examine implications of the tight bounds to the law of large numbers, and the central limit theorem. We also characterize sharply the complexity of finding optimal bounds. We show that, while we can find optimal bounds in polynomial time, when moments up to second order are known and the domain of $X$ is $R^n$, it is NP-hard to obtain such bounds when moments of third or higher order are given, or if moments of second order are given and the domain of $X$ is $R^n_+$.

*Boeing Professor of Operations Research, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Mass. 02139. Research supported by NSF grant DMI-9610486.

†Department of Mathematics and Operations Research Center, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

# 1 Introduction.

The problem of deriving bounds on the probability that a certain random variable belongs in a set, given information on some of the moments of this random variable, has a rich history, which is very much connected with the development of probability theory in the twentieth century. The inequalities due to Markov, Chebyshev and Chernoff are some of the classical and widely used results of modern probability theory. Natural questions, however, that arise are:

1. *Are such bounds "best possible," i.e., do there exist distributions that match them? A concrete and simple question in the univariate case: Is the Chebyshev inequality "best possible"?*

2. *Can such bounds be generalized in multivariate settings?*

3. *Can we develop a general theory based on optimization methods to address problems in probability theory?*

In order to answer these questions we first define the notion of a *feasible moment sequence.*

**Definition 1** *A sequence $\bar{\sigma} : (\sigma_{k_1 \cdots k_n})_{k_1+\cdots+k_n \leq k}$ is a feasible $(n, k, \Omega)$-moment vector (or sequence), if there is a random variable $X = (X_1, \ldots, X_n)$ with domain $\Omega \subseteq R^n$, whose moments are given by $\bar{\sigma}$, that is $\sigma_{k_1 \cdots k_n} = E[X_1^{k_1} \cdots X_n^{k_n}]$, $\forall\, k_1 + \cdots + k_n \leq k$. We say that any such random variable $X$ has a $\bar{\sigma}$-feasible distribution and denote this as $X \sim \bar{\sigma}$.*

We denote by $\mathcal{M} = \mathcal{M}(n, k, \Omega)$ the set of feasible $(n, k, \Omega)$-moment vectors. For the univariate case ($n = 1$), the problem of deciding if $\bar{\sigma} = (M_1, M_2, \ldots, M_k)$ is a feasible $(1, k, \Omega)$-moment vector is the classical moment problem, which has been completely characterized by necessary and sufficient conditions (see Karlin and Shapley [5], Akhiezer [1], Siu, Sengupta & Lind [14] and Kemperman [6]). For univariate, nonnegative random variables ($\Omega = R_+$), these conditions can be expressed by the semidefiniteness of the following matrices:

$$R_{2n} = \begin{pmatrix} 1 & M_1 & \ldots & M_n \\ M_1 & M_2 & \ldots & M_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ M_n & M_{n+1} & \ldots & M_{2n} \end{pmatrix} \succeq 0,$$

$$R_{2n+1} = \begin{pmatrix} M_1 & M_2 & \dots & M_{n+1} \\ M_2 & M_3 & \dots & M_{n+2} \\ \vdots & \vdots & \ddots & \vdots \\ M_{n+1} & M_{n+2} & \dots & M_{2n+1} \end{pmatrix} \succeq 0.$$

For univariate random variables with $\Omega = R$, the necessary and sufficient condition for a vector $\bar{\sigma} = (M_1, M_2, \dots, M_k)$ to be a feasible $(1, k, R)$-moment sequence is that $R_{2\lfloor \frac{k}{2} \rfloor} \succeq 0$. In the multivariate case however, the sufficiency part of the moment problem has not been resolved.

Suppose that $\bar{\sigma}$ is a feasible moment sequence and $X$ has a $\bar{\sigma}$-feasible distribution. We now define the central problem that the paper addresses:

## The $(\mathbf{n}, \mathbf{k}, \boldsymbol{\Omega})$-Bound Problem.

Given a sequence $\bar{\sigma}$ of up to $k$th order moments

$$\sigma_{k_1 k_2 \cdots k_n} = E[X_1^{k_1} X_2^{k_2} \cdots X_n^{k_n}], \qquad k_1 + k_2 + \cdots + k_n \leq k,$$

of a multivariate random variable $X = (X_1, X_2, \dots, X_n)$ on $\Omega \subseteq R^n$, find the "best possible" or "tight" upper and lower bounds on $P(X \in S)$, for arbitrary events $S \subseteq \Omega$.

The term "best possible" or "tight" upper (and by analogy lower) bound above is defined as follows.

**Definition 2** *We say that $\alpha$ is a tight upper bound on $P(X \in S)$ if :*

**(a)** *it is an upper bound, i.e., $P(X \in S) \leq \alpha$ for all random variables $X \sim \bar{\sigma}$;*

**(b)** *it cannot be improved, i.e., for any $\epsilon > 0$ there is a random variable $X_\epsilon \sim \bar{\sigma}$ for which $P(X_\epsilon \in S) > \alpha - \epsilon$.*

We will denote such a tight upper bound by $\sup_{X \sim \bar{\sigma}} P(X \in S)$. Note that a bound can be tight without necessarily being achievable, i.e., there is a random variable $\tilde{X} \sim \bar{\sigma}$ for which $P(\tilde{X} \in S) = \alpha$, but only asymptotically achievable.

The well known inequalities due to Markov, Chebyshev and Chernoff, which are widely used if we know the first moment, the first two moments, and all moments (i.e., the generating function) of a random variable, respectively, are feasible but not necessarily optimal solutions to the $(n, k, \Omega)$-bound problem, i.e., they are not necessarily tight bounds. In the univariate case, the idea that optimization methods and duality theory can be used to address these type of questions is due to Isii [4]. Thirty years later, Smith [16] has generalized this work to the multivariate case, and proposed interesting applications in decision analysis, dynamic programming, statistics and finance. He also introduces a computational procedure for the $(n, k, R^n)$-bound problem, although he does not refer to it in this way. Unfortunately, the procedure is far from an actual algorithm, as there is no proof of convergence, and no investigation (theoretical or experimental) of its efficiency. It is fair to say that understanding of the complexity of the $(n, k, \Omega)$-bound problem is still lacking.

Another line of research loosely connected to the present paper, is the work of Pitowski [11], [12] who makes use of duality results to prove general theorems in probability (weak and strong laws of large numbers, approximate central limit, the Linial-Nissan theorem etc.). The author uses different linear programming formulations to define and study geometric and complexity properties of correlation polytopes, which arise naturally in probability and logic.

Our goal in this paper is to examine the existence of an algorithm that on input $\langle n, k, \bar{\sigma}, \epsilon \rangle$ computes a value $\alpha \in [\alpha^* - \epsilon, \alpha^* + \epsilon]$, where $\alpha^* = \sup_{X \sim \bar{\sigma}} P(X \in S)$, and runs in time polynomial in $n, k, \log \sigma_{max}$ and $\log \frac{1}{\epsilon}$, where $\sigma_{max} = \max(\sigma_{k_1 \cdots k_n})$. We assume the availability of an oracle to test membership in $S \subset \Omega$, and we allow our algorithm to make a polynomial number of oracle queries of the type "is $x$ in $S$ ?".

The contributions of the present paper are as follows:

1. We completely characterize the complexity of the $(n, k, \Omega)$-bound problem. We show that the $(n, 1, \Omega)$, $(n, 2, R^n)$-bound problems can be solved in polynomial time, where as the $(n, 2, R^n_+)$-bound problem is NP-hard, as well as all $(n, k, R^n)$-bound problems for $k \geq 3$. Our derivation of the tight bounds uses convex optimization methods, and Lagrangean and Gauge duality.

2. If the set $S$ in the definition of the $(n, k, R^n)$-bound problem for $k = 1, 2$ is convex, we find the best possible bound for $P(X \in S)$ explicitly as a solution of $n$ (for $k = 1$),

and a single (for $k = 2$) convex optimization problems. These bounds represent natural extensions and improvements of the Markov ($k = 1$) and Chebyshev ($k = 2$) inequalities in multivariate settings. They retain the simplicity and attractiveness of the univariate case, as they only use the mean and covariance matrix of a multivariate random variable. We also provide explicit constructions of distributions that achieve the bounds.

3. We examine applications of the derived bounds to the law of large numbers by showing a necessary and sufficient condition for the law of large numbers to hold for correlated random variables. We also show, as an application of our constructions, that the central limit theorem fails to hold if the random variables involved are uncorrelated instead of independent.

4. We investigate in detail the univariate case, i.e., the $(1, k, \Omega)$-bound problem for $\Omega = R, R_+$. We derive optimal bounds for tail probability events in closed form. If we specialize our multivariate bound for the univariate case, it coincides with the Markov inequality (and thus the Markov inequality is best possible), but, surprisingly, we obtain an improvement of the Chebyshev inequality that retains the simplicity of the bound. We also derive closed form tight bounds for $k = 3$, and generalizations for arbitrary $k$.

The structure of the paper is as follows: In Section 2, we formulate the $(n, k, \Omega)$-bound problem as an optimization problem and present duality results that are used throughout the paper. In Section 3, we solve for the case when the set $S$ is convex (a) the $(n, 1, R^n_+)$-bound problem, as $n$ convex optimization problems, and (b) the $(n, 2, R^n)$-bound problem as a single convex optimization problem. We construct extremal distributions that achieve these bounds either exactly or asymptotically. We also provide a polynomial time algorithm to solve the $(n, 1, \Omega), (n, 2, R^n)$-bound problems for the case when the set $S$ is the union of disjoint convex sets. In Section 4, we consider several applications of the bounds derived in the previous section: we prove necessary and sufficient conditions for the Law of Large Numbers to hold for correlated random variables, we discuss the validity of the Central Limit Theorem, and we present a multivariate generalization of Markov's and Chebyshev's inequality. In Section 5, we restrict our attention to the univariate case, and we derive closed form tight bounds on tail probabilities. We compare these bounds with known inequalities such as the Markov, and the Chebyshev bounds and investigate their tightness. Finally, we derive closed form tail probability bounds when higher order moments are known. In Section 6, we prove that the $(n, 2, R_+^n)$-bound problem and the $(n, k, R^n)$-bound problem for $k \geq 3$ are NP-hard. The last section contains some concluding remarks.

## 2 Primal and Dual Formulations of The $(\mathbf{n}, \mathbf{k}, \boldsymbol{\Omega})$-Bound Problem.

In this section, we formulate the $(n, k, \Omega)$-upper bound problem as an optimization problem, where $\Omega$ is the domain of the random variables we consider. We examine the corresponding dual problem and present weak and strong duality results that permit us to develop algorithms for the problem. The same approach and results apply to the $(n, k, \Omega)$-lower bound problem.

The $(n, k, \Omega)$-upper bound problem can be formulated as the following optimization problem $(P)$:

$$
\begin{array}{lll}
(P) \quad Z_P = \text{maximize} & \displaystyle\int_S f(\bar{z})d\bar{z} & \\
\text{subject to} & \displaystyle\int_\Omega z_1^{k_1} \cdots z_n^{k_n} f(\bar{z})d\bar{z} = \sigma_{k_1 \ldots k_n}, & \forall\, k_1 + \cdots + k_n \leq k, \\
& f(\bar{z}) = f(z_1, \ldots, z_n) \geq 0, & \forall\, \bar{z} = (z_1, \ldots, z_n) \in \Omega.
\end{array}
$$

Notice that if Problem $(P)$ is feasible, then $\bar{\sigma}$ is a feasible moment sequence, and any feasible distribution $f(\bar{z})$ is a $\bar{\sigma}$-feasible distribution. The feasibility problem is exactly the classical multidimensional moment problem.

In the spirit of linear programming duality theory, we associate a dual variable $u_{k_1 \ldots k_n}$ with each equality constraint of the primal. We can identify the vector of dual variables with a $k$-degree, $n$-variate dual polynomial:

$$g(x_1, \ldots, x_n) = \sum_{k_1 + \cdots + k_n \leq k} u_{k_1 \ldots k_n} x_1^{k_1} \cdots x_n^{k_n}.$$

The dual objective translates to finding the smallest value of:

$$\sum u_{k_1 \ldots k_n} \sigma_{k_1 \ldots k_n} = \sum u_{k_1 \ldots k_n} E[X_1^{k_1} \cdots X_n^{k_n}] = E[g(X)],$$

where the expected value is taken over any $\bar{\sigma}$-feasible distribution. In this framework, the Dual Problem $(D)$ corresponding to Problem $(P)$ can be written as:

$$
\begin{aligned}
(D) \quad Z_D = \text{minimize} \quad & E[g(X)] \\
\text{subject to} \quad & g(x) \ k\text{-degree, } n\text{-variate polynomial,} \\
& g(x) \geq \chi_S(x), \ \forall \, x \in \Omega,
\end{aligned}
$$

where $\chi_S(x)$ is the indicator function of the set $S$, defined by:

$$
\chi_S(x) = \begin{cases} 1, & \text{if } x \in S, \\ 0, & \text{otherwise.} \end{cases}
$$

Notice that in general the optimum may not be achievable. Whenever the primal optimum is achieved, we call the corresponding distribution **an extremal distribution.** We next establish weak duality.

**Theorem 1 (Weak duality)** $Z_P \leq Z_D$.

**Proof:** Let $f(\bar{z})$ be a primal optimal solution and let $g(\bar{z})$ be any dual feasible solution. Then:

$$
Z_P = \int_S f(\bar{z})d\bar{z} = \int_\Omega \chi_S(\bar{z}) f(\bar{z}) d\bar{z} \leq \int_\Omega g(\bar{z}) f(\bar{z}) d\bar{z} = E[g(X)],
$$

and hence $Z_P \leq \inf_{g(\cdot) \leq \chi_S(\cdot)} E[g(X)] = Z_D$. ∎

Theorem 1 indicates that by solving the Dual Problem $(D)$ we obtain an upper bound on the primal objective and hence on the probability we are trying to bound. Under some mild restrictions on the moment vector $\bar{\sigma}$, the dual bound turns out to be tight. This strong duality result follows from a more general theorem first proved in one dimension by Isii [4], and in arbitrary dimensions by Smith [16]. The following theorem holds for arbitrary distributions and it is a consequence of their work:

**Theorem 2 (Strong Duality and Complementary Slackness)**
*If the moment vector $\bar{\sigma}$ is an interior point of the set $\mathcal{M}$ of feasible moment vectors, then the following results hold:*

1. *Strong Duality:* $Z_P = Z_D$.

2. *Complementary Slackness: If the dual is bounded, there exists a dual optimal solution* $g_{opt}(\cdot)$ *and a discrete extremal distribution concentrated on points* $x$*, where* $g_{opt}(x) = \chi_S(x)$*, that achieves the bound.*

It can also be shown that if the dual is unbounded, then the primal is infeasible, i.e., the multidimensional moment problem is infeasible. Moreover, if $\bar{\sigma}$ is a boundary point of $\mathcal{M}$, then it can be shown that the $\bar{\sigma}$-feasible distributions are concentrated on a subset $\Omega_0$ of $\Omega$, and strong duality holds provided we relax the dual to $\Omega_0$ (see Smith [16] , p. 824). In the univariate case, Isii [4] proves that if $\bar{\sigma}$ is a boundary point of $\mathcal{M}$ , then exactly one $\bar{\sigma}$-feasible distribution exists.

If strong duality holds, then by optimizing over Problem $(D)$ we obtain a **tight** bound on $P(X \in S)$. On the other hand, solving Problem $(D)$ is equivalent to solving the corresponding separation problem, under certain technical conditions (see Grötschel, Lovász and Schrijver [3]). In the next section, we show that the separation problem is polynomially solvable for the cases $(n, 1, \Omega)$ and $(n, 2, R^n)$, and in Section 6, we show that it is NP-hard for the cases $(n, 2, R^n_+)$ and $(n, k, R^n)$ for $k \geq 3$.

# 3 Efficient Algorithms for The $(\mathbf{n}, \mathbf{1}, \mathbf{\Omega})$, $(\mathbf{n}, \mathbf{2}, \mathbf{R^n})$-Bound Problems.

In this section, we address the $(n, 1, \Omega)$, and $(n, 2, R^n)$-bound problems. We present tight bounds as solutions to $n$ convex optimization problems for the $(n, 1, R^n_+)$-bound problems, and as a solution to a single convex optimization problem for the $(n, 2, R^n)$-bound problem for the case when the event $S$ is a convex set. We present a polynomial time algorithm for more general sets.

## 3.1 The $(\mathbf{n}, \mathbf{1}, \mathbf{R^n_+})$-Bound Problem for Convex Sets.

In this case, we are given a vector $M$ that represents the vector of means of a random variable $X$ defined in $R^n_+$, and we would like to find tight bounds on $P(X \in S)$ for a convex set $S$.

**Theorem 3** *The tight $(n, 1, R^n_+)$-upper bound for an arbitrary convex event $S$ is given by:*

$$\sup_{X \sim M} P(X \in S) = \min\left(1, \max_{i=1,\ldots,n} \frac{M_i}{\inf\limits_{x \in S_i} x_i}\right), \tag{1}$$

*where $S_i = S \cap (\cap_{j \neq i}\{x \in R^n_+ |\ M_i x_j - M_j x_i \leq 0\})$.*

**Proof:** Problem $(D)$ can be written as follows for this case:

$$\begin{aligned} Z_D = \text{minimize} \quad & a'M + b \\ \text{subject to} \quad & a'x + b \geq 1, \quad \forall x \in S, \\ & a'x + b \geq 0, \quad \forall x \in R^n_+. \end{aligned}$$

If the optimal solution $(a_0, b_0)$ satisfies $\min\limits_{x \in S} a_0'x + b_0 = \alpha > 1$, then the solution $\left(\frac{a_0}{\alpha}, \frac{b_0}{\alpha}\right)$ has value $Z_D/\alpha < Z_D$. Therefore, $\inf\limits_{x \in S} a'x + b = 1$. By a similar argument we have that $b_0 \leq 1$. Moreover, since $a'x + b \geq 0,\ \forall x \in R^n_+$, $a \geq 0$, and $b \geq 0$. We thus obtain:

$$\begin{aligned} Z_D = \text{minimize} \quad & a'M + b \\ \text{subject to} \quad & \inf_{x \in S} a'x = 1 - b. \\ & a \geq 0,\ 0 \leq b \leq 1. \end{aligned}$$

Without loss of generality we let $a = \lambda v$, where $\lambda$ is a nonnegative scalar, and $v$ is a nonnegative vector with $||v|| = 1$. Thus, we obtain:

$$\begin{aligned} Z_D = \text{minimize} \quad & (1 - b)\frac{v'M}{\inf\limits_{x \in S} v'x} + b \\ \text{subject to} \quad & v \geq 0,\ ||v|| = 1,\ 0 \leq b \leq 1. \end{aligned}$$

Thus,

$$\begin{aligned} Z_D &= \min\left(1, \min_{||v||=1, v \geq 0} \frac{v'M}{\inf\limits_{x \in S} v'x}\right) \\ &= \min\left(1, \min_{||v||=1, v \geq 0} \sup_{x \in S} \frac{v'M}{v'x}\right) \\ &= \min\left(1, \sup_{x \in S} \min_{||v||=1, v \geq 0} \frac{v'M}{v'x}\right) \end{aligned} \tag{2}$$

$$= \min\left(1, \sup_{x\in S} \min_{i=1,\ldots,n} \frac{M_i}{x_i}\right) \tag{3}$$

$$= \min\left(1, \max_{i=1,\ldots,n} \frac{M_i}{\inf\limits_{x\in S_i} x_i}\right), \tag{4}$$

where $S_i = S \cap (\cap_{j\neq i}\{x \in R^n_+ |\ M_i x_j - M_j x_i \leq 0\})$ is a convex set. Note that in Eq. (2) we exchanged the order of min and sup (see Rockafellar [13], p. 382). In Eq. (3), we used $\min\limits_{||v||=1, v\geq 0} \frac{v'M}{v'x}$ is attained at $v = e_j$, where

$$\frac{M_j}{x_j} = \min_{i=1,\ldots,n} \frac{M_i}{x_i}.$$

In order to understand Eq. (4), we let $\phi(x) = \min\limits_{i=1,\ldots,n} \frac{M_i}{x_i}$. Note that $\phi(x) = \frac{M_i}{x_i}$, when $x \in \{x \in R^n_+ |\ M_i x_j - M_j x_i \leq 0\}$. Then, we have

$$\sup_{x\in S} \phi(x) = \max_{i=1,\ldots,n} \sup_{x\in S_i} \phi(x) = \max_{i=1,\ldots,n} \sup_{x\in S_i} \frac{M_i}{x_i} = \max_{i=1,\ldots,n} \frac{M_i}{\inf\limits_{x\in S_i} x_i}.$$

■

## 3.2 Extremal Distributions for The $(\mathrm{n}, 1, \mathrm{R}^{\mathrm{n}}_+)$-Bound Problem.

In this section, we construct a distribution that achieves Bound (1). We will say that the Bound (1) is achievable, when there exists an $x^* \in S$ such that

$$\min\left(1, \max_{i=1,\ldots,n} \frac{M_i}{\inf\limits_{x\in S_i} x_i}\right) = \frac{M_i}{x_i^*} < 1.$$

In particular, the bound is achievable when the set $S$ is closed and $M \notin S$.

**Theorem 4** **(a)** *If $M \in S$ or if the Bound (1) is achievable, then there is an extremal distribution that exactly achieves it.*

**(b)** *Otherwise, there is a sequence of distributions defined on $R^n_+$ with mean $M$, that asymptotically achieve it.*

**Proof:** **(a)** If $M \in S$, then the extremal distribution is simply $P(X = M) = 1$. Now suppose that $M \notin S$ and the Bound (1) is achievable. We assume without loss of generality that the bound equals $\frac{M_1}{x_1^*} < 1$, and it is achieved at $x^* \in S$. Therefore, $\frac{M_1}{x_1^*} = \min_{i=1,\ldots,n} \frac{M_i}{x_i^*}$. We consider the following random variable $X$ defined on $R_+^n$:

$$X = \begin{cases} x^*, & \text{with probability} \quad p = \frac{M_1}{x_1^*}, \\ v = \frac{x_1^* M - M_1 x^*}{x_1^* - M_1}, & \text{with probability} \quad 1 - p = 1 - \frac{M_1}{x_1^*}. \end{cases}$$

Note that $E[X] = M$, and $v_i = \frac{M_i x_1^* - M_1 x_i^*}{x_1^* - M_1} \geq 0$ for all $i = 1, \ldots, n$, since $\frac{M_1}{x_1^*} = \min_{i=1,\ldots,n} \frac{M_i}{x_i^*}$. Moreover, $v \notin S$, or else by the convexity of $S$, we have that $M = px^* + (1-p)v \in S$, a contradiction. Therefore,

$$P(X \in S) = P(X = x^*) = \frac{M_1}{x_1^*}.$$

**(b)** If $M \notin S$ and the Bound (1) is not achievable, then we construct a sequence of non-negative distributions with mean $M$ that approach it. Suppose without loss of generality that $\max_{i=1,\ldots,n} \frac{M_i}{\inf_{x \in S_i} x_i}$ equals $\frac{M_1}{x_1^*}$, for $x^* \in \overline{S}_1$ (the closure of $S_1$), so Bound (1) is equal to $\min\left(1, \frac{M_1}{x_1^*}\right)$. Consider a sequence $x^k \in S_1$, $x^k \to x^*$, so that $\lim_{k\to\infty} \min_{i=1,\ldots,n} \frac{M_i}{x_i^k} = \frac{M_1}{x_1^*}$, and a sequence $p_k$, $0 < p_k < \min\left(1, \frac{M_1}{x_1^k}\right)$ so that $p_k \to \min\left(1, \frac{M_1}{x_1^*}\right)$. Consider the sequence of distributions:

$$X_k = \begin{cases} x^k, & \text{with probability} \quad p_k, \\ v^k = \frac{M - p_k x^k}{1 - p_k}, & \text{with probability} \quad 1 - p_k. \end{cases}$$

Clearly, the random variables $X_k$ are nonnegative with mean $E[X_k] = M$. Also $v^k \notin S$ or else $M \in S$, so $P(X_k \in S) = P(X_k = x^k) = p_k \to \min\left(1, \frac{M_1}{x_1^*}\right)$. This shows that the sequence of nonnegative, distributions $X_k$ with mean $M$ asymptotically achieve the Bound (1). ■

### 3.3 The $(\mathbf{n}, 2, \mathbf{R}^n)$-Bound Problem for Convex Sets.

We first rewrite the $(n, 2, R^n)$-bound problem in a more convenient form. Rather than assuming that $E[X]$ and $E[XX']$ are known, we assume equivalently that the vector $M =$

$E[X]$ and the covariance matrix $\Gamma = E[(X-M)(X-M)']$ are known. Given a set $S \subseteq R^n$, we find tight upper bounds, denoted by $\sup_{X\sim(M,\Gamma)} P(X \in S)$, on the probability $P(X \in S)$ for all multivariate random variables $X$ defined on $R^n$ with mean $M = E[X]$ and covariance matrix $\Gamma = E[(X-M)(X-M)']$.

First, notice that a necessary and sufficient condition for the existence of such a random variable $X$, is that the covariance matrix $\Gamma$ is symmetric and positive semidefinite. Indeed, given $X$, for an arbitrary vector $a$ we have:

$$0 \leq E[(a'(X-M))^2] = a'E[(X-M)(X-M)']a = a'\Gamma a,$$

so $\Gamma$ must be positive semidefinite. Conversely, given a symmetric semidefinite matrix $\Gamma$ and a mean vector $M$, we can define a multivariate normal distribution with mean $M$ and covariance $\Gamma$. Moreover, notice that $\Gamma$ is positive definite if and only if the components of $X - M$ are linearly independent. Indeed, the only way that $0 = a'\Gamma a = E[(a'(X-M))^2]$ for a nonzero vector $a$ is that $a'(X-M) = 0$.

We assume that $\Gamma$ has full rank and is positive definite. This does not reduce the generality of the problem, it just eliminates redundant constraints, and thereby insures that Theorem 2 holds. Indeed, the tightness of the bound is guaranteed by Theorem 2 whenever the moment vector is interior to $\mathcal{M}$. If the moment vector is on the boundary, it means that the covariance matrix of $X$ is not of full rank, implying that the components of $X$ are linearly dependent. By eliminating the dependent components, we reduce without loss of generality the problem to one of smaller dimension for which strong duality holds. Hence, the primal and the dual problems $(P)$ and $(D)$ satisfy $Z_P = Z_D$. Our main result in this section is as follows.

**Theorem 5** *The tight $(n, 2, R^n)$-upper bound for an arbitrary convex event $S$ is given by:*

$$\sup_{X\sim(M,\Gamma)} P(X \in S) = \frac{1}{1+d^2}, \tag{5}$$

*where* $d^2 = \inf_{x\in S}(x-M)'\Gamma^{-1}(x-M)$, *is the squared distance from $M$ to the set $S$, under the norm induced by the matrix $\Gamma^{-1}$.*

The proof of the theorem consists of two parts: First, we formulate a restricted dual problem, and prove the restriction to be exact whenever the set $S$ is convex. Second, we calculate the optimal value of the restricted problem and show that it is equal to the expression given in Eq. (5). Before we proceed to formulate the restricted problem, we need the following preliminary result, which holds regardless of the convexity assumption on the set $S$:

**Lemma 1** *There exists an optimal dual solution for the $(n, 2, R^n)$-bound problem of the form $g(x) = \| A'(x - x_0) \|^2$, for some square matrix $A$ and vector $x_0$.*

**Proof:** Let $g(x) = x'Hx + c'x + d$ be an optimal solution to Problem (D). Then, $H$ must be positive semidefinite, since $g(x) \geq 0 \; \forall x \in R^n$, and we can assume without loss of generality that $H$ is symmetric. This is equivalent to the existence of a square matrix $A$ such that $H = AA'$. Notice that whenever $x'Hx = 0$, or equivalently $A'x = 0$, we must have $c'x = 0$ by the nonnegativity of $g(x)$. This means that $c$ is spanned by the columns of $A$, so we can write $c = 2Ab$, and $g(x) = x'AA'x + 2b'A'x + d = \| A'x + b \|^2 + d - \| b \|^2$. Since we seek to minimize $E[g(X)]$, we should make the constant term as small as possible, yet keeping $g(x)$ nonnegative. Thus $\| b \|^2 - d = \min \| A'x + b \|^2 = \| A'x_0 + b \|^2$, where $x_0$ satisfies $AA'x_0 + Ab = 0$, from the first order conditions. It follows that

$$g(x) = \| A'x + b \|^2 - \| A'x_0 + b \|^2 = \| A'(x - x_0) \|^2 .$$

■

Lemma 1 shows that the Dual Problem $(D)$ is equivalent to:

$$\begin{aligned} Z_D = \text{minimize} \quad & E[\, \| A'(X - b) \|^2] \\ \text{subject to} \quad & \inf_{x \in S} \| A'(x - b) \|^2 = 1. \end{aligned} \tag{6}$$

The reason we wrote equality in Eq. (6) above is that if $A, b$ are optimal solutions, and $\inf_{x \in S} \| A'(x - b) \|^2 = \alpha^2 > 1$, then by letting $A' = A/\alpha$, we can decrease the objective value further, thus contradicting the optimality of $(A, b)$.

We formulate the following restricted dual problem:

$$\begin{aligned} (RD) \;\; Z_{RD} = \text{minimize} \quad & E[(a'(X - b))^2] \\ \text{subject to} \quad & \inf_{x \in S} a'(x - b) = 1. \end{aligned}$$

Clearly $Z_D \leq Z_{RD}$, since for any feasible solution $(a, b)$ to $(RD)$ we have a corresponding feasible solution of $(D)$ with the same objective value, namely: $(A = (a, 0, \ldots, 0)\ ,\ b)$. We next show that if $S$ is a convex set, this restriction is actually exact, thereby reducing the dual problem to one which is easier to solve.

**Lemma 2** *If $S$ is a convex set, then $Z_D = Z_{RD}$.*

**Proof:** We only need to show $Z_D \geq Z_{RD}$.

Let $(A, b)$ be an optimal solution to Problem (6), and let $\inf_{x \in S} \| A'(x-b) \|^2 = \| A'(x_0 - b) \|^2 = 1$, for some minimizer $x_0 \in S$. If the optimum value is not attained, we can consider a sequence in $S$ that achieves it. By the Cauchy-Schwartz inequality we have:

$$\| A'(x-b) \|^2 = \| A'(x-b) \|^2 \cdot \| A'(x_0 - b) \|^2 \geq ((x_0 - b)' AA'(x-b))^2.$$

Let $a = AA'(x_0 - b)$, so $((x_0 - b)'AA'(x-b))^2 = (a'(x-b))^2 \leq \| A'(x-b) \|^2$. We next show that $(a, b)$ is feasible for $(RD)$. Indeed, taking expectations, we obtain that $Z_{RD} \leq E[(a'(X-b))^2] \leq E[\ \| A'(X-b) \|^2] = Z_D$.

We now prove that $(a, b)$ is feasible for $(RD)$, as desired. Notice that $a'(x_0 - b) = 1$; it remains to show that $a'(x-b) \geq 1$, for all other $x \in S$. We have that

$$\inf_{x \in S} \| A'(x-b) \|^2 = \| A'(x_0 - b) \|^2 = 1.$$

We rewrite this as $\inf_{v \in S_{A,b}} \| v \|^2 = \| v_0 \|^2 = 1$, where

$$S_{A,b} = \{\ A'(x-b) \mid x \in S\ \},\ v = A'(x-b),\ v_0 = A'(x_0 - b) \in S_{A,b}.$$

Clearly $S_{A,b}$ is a convex set, since it is obtained from the convex set $S$ by a linear transformation. It is well known (see Kinderlehrer and Stampacchia [7]) that for every convex function $F: R^n \to R$, and convex set $K$, $z_0$ is an optimal solution to the problem $\inf_{z \in K} F(z)$ if and only if

$$\nabla F(z_0)'(z - z_0) \geq 0, \qquad \forall z \in K. \tag{7}$$

Applying this result for $F(z) = \frac{1}{2} z'z$, $K = S_{A,b}$, and $z_0 = v_0$, we obtain that $v_0'(v - v_0) \geq 0$, that is $v_0'v \geq v_0'v_0 = 1$, for all $v \in S_{A,b}$. But notice that $v_0'v = (x_0 - b)'AA'(x-b)$. This shows that $a'(x-b) = (x_0 - b)'AA'(x-b) \geq 1$ for all $x \in S$, so $(a, b)$ is feasible for $(RD)$. ■

**Proof of Theorem 3:**

The previous two lemmas show that Problem (D) is equivalent to the following restricted problem:

$$\begin{aligned} Z_D = \text{minimize} \quad & E[(a'(X-M-c))^2] = \min\ a'\Gamma a + (a'c)^2 \\ \text{subject to} \quad & \inf_{x\in S} a'(x-M-c) = 1, \end{aligned}$$

where we substituted $b = c + M$ in the Formulation $(RD)$. Substituting $a'c = \inf_{x\in S} a'(x - M) - 1$ back into the objective, the problem can be further rewritten as:

$$Z_D = \min_a\ a'\Gamma a + (1 - a'(x_a - M))^2,$$

where $x_a$ is an optimizer of $\inf\limits_{x\in S} a'(x-M)$ (again if the optimum is not attained we can consider a sequence in $S$ converging to $x_a$).

From the Cauchy-Schwartz inequality we have

$$(a'(x-M))^2 \le ||\Gamma^{\frac{1}{2}}a||^2||\Gamma^{-\frac{1}{2}}(x-M)||^2.$$

Therefore,

$$\inf_{x\in S} a'(x-M) \le \inf_{x\in S} |a'(x-M)| \le ||a'\Gamma^{\frac{1}{2}}|| \inf_{x\in S} ||\Gamma^{-\frac{1}{2}}(x-M)||.$$

Let $d = \inf_{x\in S} ||\Gamma^{-\frac{1}{2}}(x-M)||$. Thus,

$$Z_D = \min_a \left(a'\Gamma a + [1 - \inf_{x\in S}(a'(x-M))]^2\right) \ge \begin{cases} \min\limits_a \left(a'\Gamma a + [1-(a'\Gamma a)^{\frac{1}{2}}d]^2\right), & \text{if } a'\Gamma a \le \dfrac{1}{d^2}, \\ \min\limits_a\ a'\Gamma a, & \text{if } a'\Gamma a \ge \dfrac{1}{d^2}. \end{cases}$$

If $a'\Gamma a \ge \frac{1}{d^2}$, then $Z_D \ge \frac{1}{d^2}$. Otherwise, let $\alpha = (a'\Gamma a)^{\frac{1}{2}}$. Then,

$$\min_a \left(a'\Gamma a + [1 - \inf_{x\in S}(a'(x-M))]^2\right) \ge \min_\alpha \left(\alpha^2 + (1-\alpha d)^2\right).$$

Optimizing over the right hand side we obtain that $\alpha^* = d/(1+d^2) < \frac{1}{d}$, and the optimal value is $\dfrac{1}{1+d^2}$. Thus, in this case,

$$\min_a \left( a'\Gamma a + [1 - \min_{x\in S}(a'(x-M))]^2 \right) \geq \frac{1}{1+d^2}.$$

Since $\dfrac{1}{d^2} \geq \dfrac{1}{1+d^2}$, we have in all cases:

$$Z_D \geq \frac{1}{1+d^2}.$$

To prove equality, let $x^*$ be an optimizer of $\inf_{x\in S} ||\Gamma^{-\frac{1}{2}}(x-M)||$ (again if the optimum is not attained, we consider a sequence $x^k \in S$ converging to $x^*$). Applying (7) with $F(z) = (z-M)'\Gamma^{-1}(z-M)$, $z_0 = x^*$, and $K = S$, and since $S$ is convex, we have that for all $x \in S$:

$$(x^* - M)'\Gamma^{-1}(x - x^*) \geq 0,$$

and therefore,

$$a_0'(x-M) \geq a_0'(x^*-M),$$

with $a_0 = \theta\Gamma^{-1}(x^*-M)$, and $\theta = \dfrac{1}{1+d^2}$. Hence,

$$\inf_{x\in S} a_0'(x-M) = a_0'(x^*-M) = \theta d^2,$$

and therefore,

$$\left( a_0'\Gamma a_0 + [1 - \inf_{x\in S}(a_0'(x-M))]^2 \right) = \frac{1}{1+d^2}.$$

Therefore, $Z_D = \dfrac{1}{1+d^2}$. ■

### 3.4 Extremal Distributions for The $(\mathrm{n}, 2, \mathbf{R}^{\mathrm{n}})$-Bound Problem.

In this section, we construct an extremal distribution of a random variable $X \sim (M, \Gamma)$, so that $P(X \in S) = 1/(1+d^2)$ with $d^2 = \inf_{x\in S}(x-M)'\Gamma^{-1}(x-M)$. We will say that the bound $d$ is achievable, when there exists an $x^* \in S$ such that $d^2 = (x^* - M)'\Gamma^{-1}(x^* - M)$. In particular, $d$ is achievable if the set $S$ is closed.

**Theorem 6** **(a)** *If $M \notin S$ and if $d^2 = \inf_{x \in S}(x - M)'\Gamma^{-1}(x - M)$ is achievable, then there is an extremal distribution that exactly achieves the Bound (5).*

**(b)** *Otherwise, if $M \in S$ or if $d^2$ is not achievable, then there is a sequence of $(M, \Gamma)$-feasible distributions that asymptotically approach the Bound (5).*

**Proof:**

**(a)** Suppose that the bound $d^2$ is achievable and $M \notin S$. We show how to construct a random variable $X \sim (M, \Gamma)$ that achieves the bound: $P(X \in S) = \dfrac{1}{1+d^2}$. Note that

$$d^2 = \inf_{x \in S} ||\Gamma^{-\frac{1}{2}}(x - M)||^2 = \inf_{y \in T} ||y||^2,$$

where $T = \{y |\ y = \Gamma^{-\frac{1}{2}}(x - M),\ x \in S\}$. Since we assumed that the bound is achievable, there exists a vector $v_0 \in T$, such that $d^2 = ||v_0||^2$. Since $M \notin S$, it follows $0 \notin T$, and therefore, $v_0 \neq 0$.

We first construct a discrete random variable $Y \sim (0, I)$, that has the property that $P(Y \in T) \geq \frac{1}{1+d^2}$. By letting $X = \Gamma^{\frac{1}{2}}Y + M$, we obtain a discrete distribution $X \sim (M, \Gamma)$ that satisfies:

$$P(X \in S) = P(Y \in T) \geq \frac{1}{1+d^2}.$$

The distribution of $Y$ is as follows:

$$Y = \begin{cases} v_0, & \text{with probability} \quad p_0 = \dfrac{1}{1+d^2}, \\ v_i, & \text{with probability} \quad p_i,\ i = 1, \ldots, n. \end{cases}$$

We next show how the vectors $v_i$, and the probabilities $p_i$, $i = 1, \ldots, n$ are selected. Let

$$V_0 = I - \frac{1}{1+d^2}(v_0 \cdot v_0').$$

The matrix $V_0$ is positive definite. Indeed, using the Cauchy-Schwartz inequality, we obtain:

$$v'\, V_0\, v = \|\, v\,\|^2 - \frac{1}{1+d^2}\,(v'v_0)^2 \geq \frac{(v'v_0)^2}{\|\, v_0\,\|^2} - \frac{1}{1+d^2}\,(v'v_0)^2 = (v'v_0)^2 \left(\frac{1}{d^2} - \frac{1}{1+d^2}\right) \geq 0,$$

and equality holds in both inequalities above if and only if $v'v_0 = 0$, and (since $v_0 \neq 0$) $v'v = 0$, that is $v = 0$. Since $V_0$ is positive definite we can decompose it as $V_0 = Q \cdot Q'$, where $Q$ is a nonsingular matrix. Notice that, by possibly multiplying it by an orthonormal rotation matrix, we can choose $Q$ in such a way that $Q^{-1}v_0 \leq 0$.

We select the vector of probabilities $p = (p_1, \ldots, p_n)$ as follows:

$$\sqrt{p} = (\sqrt{p_1}, \ldots, \sqrt{p_n}) = -\frac{1}{1+d^2} \cdot Q^{-1}v_0 \geq 0.$$

Note that

$$e'p = \|\, \sqrt{p}\,\|^2 = \frac{1}{(1+d^2)^2} \cdot v_0'(QQ')^{-1}v_0 = \frac{1}{(1+d^2)^2} \cdot v_0'\,(\, I + v_0 v_0'\,)\, v_0 = \frac{d^2}{1+d^2},$$

since $v_0'v_0 = \|\, v_0\,\|^2 = d^2$ and $(QQ')^{-1} = V_0^{-1} = I + v_0 v_0'$. Therefore,

$$\sum_{i=0}^{n} p_i = p_0 + e'p = \frac{1}{1+d^2} + \frac{d^2}{1+d^2} = 1.$$

Let $V$ denote the square $n \times n$ matrix with rows $v_i'$. We select the matrix $V$ as follows:

$$V = I_{\sqrt{p}}^{-1} Q',$$

where $I_{\sqrt{p}}$ is a diagonal matrix, whose $i$th diagonal entry is $\sqrt{p_i}$, $i = 1, \ldots, n$. Note that

$$V'p = Q\sqrt{p} = Q\left(-\frac{1}{1+d^2}\right) Q^{-1}v_0 = -\frac{1}{1+d^2} v_0,$$

and therefore, $E[Y] = \sum_{i=0}^{n} v_i p_i = 0$. Moreover,

$$V' I_p V = QQ' = V_0 = I - \frac{1}{1+d^2}\,(v_0 \cdot v_0').$$

Hence,

$$E[YY'] = \sum_{i=0}^{n} p_i(v_i \cdot v_i') = V' I_p V + p_0 v_0 \cdot v_0' = I.$$

Finally, since the bound is achievable, the vector $v_0 \in T$. Therefore,

$$P(X \in S) = P(Y \in T) \geq P(Y = v_0) = p_0 = \frac{1}{1+d^2}.$$

From Eq. (5), we know that $P(X \in S) \leq \frac{1}{1+d^2}$, and thus the random variable $X$ satisfies the bound with equality.

**(b)** If $M \in S$, then the upper bound in Eq. (5) equals 1. Let $X_\epsilon = M + \frac{1}{\sqrt{\epsilon}} B_\epsilon \cdot Z$, where $B_\epsilon$ is a Bernoulli random variable with success probability $\epsilon$, and $Z \sim N(0, \Gamma)$ is a multivariate normal random variable independent of $B_\epsilon$. One can easily check that $X_\epsilon \sim (M, \Gamma)$ and $P(X_\epsilon = M) \geq 1-\epsilon$. Therefore, for any event $S$ than contains $M$, we have $P(X_\epsilon \in S) \geq 1-\epsilon$.

If the bound $d^2$ is not achievable, then we can construct a sequence $X_k = \Gamma^{\frac{1}{2}} Y_k + M$ of $(M, \Gamma)$-feasible random variables that approach the bound in Eq. (5) in the following way: Let $(v_0^k) \to v_0$ with $v_0^k \in T$ , and $d_k^2 = \| v_0^k \|^2$, so $d_k \to d$. We define for each $k \geq 1$, the random varaible $Y_k$ in the same way as we constructed $Y$ in part (a), so $Y_k \sim (0, I)$ and $P(Y_k \in T) \geq P(Y_k = v_0^k) = \frac{1}{1+d_k^2} \longrightarrow \frac{1}{1+d^2}$. This shows that the sequence of $(0, I)$-feasible random variables $Y_k$, and thus the sequence of $(M, \Gamma)$-feasible random variables $X_k = \Gamma^{\frac{1}{2}} Y_k + M$, asymptotically approach the bound (5). ■

### 3.5 A Polynomial Time Algorithm for Unions of Convex Sets.

In this section, we present polynomial time algorithms that compute tight $(n, 1, \Omega)$ and $(n, 2, R^n)$-bounds for any event $S$ that can be decomposed as a disjoint union of a polynomial (in $n$) number of convex sets. We further assume that the set $\Omega$ can be decomposed as a disjoint union of a polynomial (in $n$) number of convex sets. Our overall strategy is to formulate the problem as an optimization problem, consider its dual and exhibit an algorithm that solves the corresponding separation problem in polynomial time.

**The Tight $(\mathbf{n}, \mathbf{1}, \boldsymbol{\Omega})$-Bound.**

We are given the mean-vector $M = (M_1, \ldots, M_n)$ of an $n$-dimensional random variable $X$ with domain $\Omega$ that can be decomposed in a polynomial (in $n$) number of convex sets, and we want to derive tight bounds on $P(X \in S)$. Problem $(D)$ can be written as follows:

$$\begin{aligned} Z_D = \text{minimize} \quad & u'M + u_0 \\ \text{subject to} \quad & g(x) = u'x + u_0 \geq \chi_S(x), \ \forall x \in \Omega. \end{aligned} \tag{8}$$

The separation problem associated with Problem (8) is defined as follows: Given a vector $a$ and a scalar $b$ we want to check whether $g(x) = a'x + b \geq \chi_S(x)$, $\forall x \in \Omega$, and if not, we want to exhibit a violated inequality. The following algorithm achieves this goal.

**Algorithm A:**

1. Solve the problem $\inf_{x \in \Omega} g(x)$ (note that the problem involves a polynomial number of convex optimization problems; in particular if $\Omega$ is polyhedral, this is a linear optimization problem). Let $z_0$ be the optimal solution value and let $x_0 \in \Omega$ be an optimal solution.

2. If $z_0 < 0$, then we have $g(x_0) = z_0 < 0$: this constitutes a violated inequality;

3. Otherwise, we solve $\inf_{x \in S} g(x)$ (again, the problem involves a polynomial number of convex optimization problems, while if $S$ is polyhedral, this is a linear optimization problem). Let $z_1$ be the optimal solution value and let $x_1 \in S$ be an optimal solution.

   (a) If $z_1 < 1$, then for $x_1 \in S$ we have $g(x_1) = z_1 < 1$: this constitutes a violated inequality.

   (b) If $z_1 \geq 1$, then $a, b$ are feasible.

The above algorithm solves the separation problem in polynomial time, since we can solve any convex optimization problem in polynomial time (see Nesterov and Nemirowskii [10], Nemhauser and Wolsey [9]). Therefore, the $(n, 1, \Omega)$-upper bound problem is polynomially solvable.

**The Tight $(\mathbf{n}, \mathbf{2}, \mathbf{R^n})$-Bound.**

We are given first and second order moment information $(M, \Gamma)$ on the $n$-dimensional random variable $X$, and we would like to compute $\sup_{X \sim (M,\Gamma)} P(X \in S)$. Recall that the corresponding dual problem can be written as:

$$\begin{aligned} Z_D = \text{minimize} \quad & E[g(X)] \\ \text{subject to} \quad & g(x) = x'Hx + c'x + d \geq \chi_S(x), \qquad \forall\, x \in R^n. \end{aligned} \tag{9}$$

The separation problem corresponding to Problem (9) can be stated as follows: Given a matrix $H$, a vector $c$ and a scalar $d$, we need to check whether $g(x) = x'Hx + c'x + d \geq \chi_S(x)$, $\forall x \in R^n$, and if not, find a violated inequality. Notice that we can assume without loss of generality that the matrix $H$ is symmetric.

The following algorithm solves the separation problem in polynomial time.

**Algorithm B:**

1. If $H$ is not positive semidefinite, then we find a vector $x_0$ so that $g(x_0) < 0$. We decompose $H = Q'\Lambda Q$, where $\Lambda =$ diag$(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of eigenvalues of $H$. Let $\lambda_i < 0$ be a negative eigenvalue of $H$. Let $y$ be vactor with $y_j = 0$, for all $j \neq i$, and $y_i$ large enough so that $\lambda_i y_i^2 + (Qc)_i y_i + d < 0$. Let $x_0 = Q'y$. Then,

$$\begin{aligned} g(x_0) &= x_0'Hx_0 + c'x_0 + d \\ &= y'QQ'\Lambda QQ'y + c'Q'y + d \\ &= y'\Lambda y + c'Q'y + d \\ &= \sum_{j=1}^{n} \lambda_j y_j^2 + \sum_{j=1}^{n} (Qc)_j y_j + d \\ &= \lambda_i y_i^2 + (Qc)_i y_i + d < 0. \end{aligned}$$

   This produces a violated inequality.

2. Otherwise, if $H$ is positive semidefinite, then:

   (a) We test if $g(x) \geq 0$, $\forall x \in R^n$ by solving the convex optimization problem:

$$\inf_{x \in R^n} g(x).$$

Let $z_0$ be the optimal value. If $z_0 < 0$, we find $x_0$ such that $g(x_0) < 0$, which represents a violated inequality. Otherwise,

(b) We test if $g(x) \geq 1, \ \forall x \in S$ by solving a polynomial collection of convex optimization problems

$$\inf_{x \in S} g(x).$$

Let $z_1$ be the optimal value. If $z_1 \geq 1$, then $g(x) \geq 1$, $\forall x \in S$, and thus $(H, c, d)$ is feasible. If not, we exhibit an $x_1$ such that $g(x_1) < 1$, and thus we identify a violated inequality.

Since we can solve the separation problem in polynomial time, we can also solve (within $\epsilon$) the $(n, 2, R^n)$-bound problem in polynomial time (in the problem data and $\log \frac{1}{\epsilon}$).

# 4 Applications.

In this section, we provide several applications of the bounds we derived in the previous section.

## 4.1 On The Law of Large Numbers for Correlated Random Variables.

Consider a sequence of random variables $X^{(n)} = (X_1, \ldots, X_n)$. If $X^{(n)} \sim (\mu \cdot e, \Gamma^{(n)})$, i.e., all members of the sequence have the same mean, and $\mathrm{Var}(X_i) < \infty$, $i = 1, \ldots n$, under what conditions does the law of large numbers hold, i.e., as $n \to \infty$

$$P\left(\frac{\sum_{i=1}^{n} X_i}{n} = \mu\right) \longrightarrow 1?$$

In preparation to answering this question we first derive simple tight closed form bounds for $P(X^{(n)} \in S)$ for particular sets $S$.

**Proposition 1** *For any vector $\alpha$ and constant $\tau$, we have:*

$$\sup_{X \sim (M, \Gamma)} P\left(\alpha' X \geq \tau\right) = \begin{cases} \dfrac{\alpha' \Gamma \alpha}{\alpha' \Gamma \alpha + (\tau - \alpha' M)^2} & , \ \textit{if } \tau \geq \alpha' M, \\ 1 & , \ \textit{otherwise.} \end{cases} \tag{10}$$

$$\inf_{X\sim(M,\Gamma)} P\left(\alpha' X > \tau\right) = \begin{cases} \dfrac{(\tau-\alpha' M)^2}{\alpha'\Gamma\alpha + (\tau-\alpha' M)^2} & , \textit{ if } \tau \le \alpha' M, \\ 0 & , \textit{ otherwise.} \end{cases} \tag{11}$$

**Proof:** From Eq. (5) we have that

$$\sup_{X\sim(M,\Gamma)} P\left(\alpha' X \ge \tau\right) = \frac{1}{1+d^2},$$

where

$$\begin{aligned} d^2 = \text{minimize} &\quad (x-M)'\Gamma^{-1}(x-M) \\ \text{subject to} &\quad \alpha' x \ge \tau. \end{aligned}$$

Applying the Kuhn-Tucker conditions, we easily obtain that $d^2 = \lambda^2 \alpha'\Gamma\alpha$, and $\lambda = \frac{\tau-\alpha' M}{\alpha'\Gamma\alpha}$ if $\tau - \alpha' M \ge 0$, and $\lambda = 0$, otherwise. The Bound (10) then follows.
For the infimum, we observe that

$$\inf_{X\sim(M,\Gamma)} P\left(\alpha' X > \tau\right) = 1 - \sup_{X\sim(M,\Gamma)} P\left(\alpha' X \le \tau\right).$$

Since $\{x \mid \alpha' x \le \tau\}$ is a convex set, Eq. (11) follows similarly by applying Eq. (5). ∎

**Theorem 7 (The Law of Large Numbers for correlated random variables)** *A sequence of correlated random variables $X^{(n)} = (X_1, \ldots, X_n)$ with $X^{(n)} \sim (\mu \cdot e, \Gamma^{(n)})$ satisfies the law of large numbers, i.e., $P\left(\frac{\sum_{i=1}^n X_i}{n} = \mu\right) \longrightarrow 1$, as $n \to \infty$ if and only if*

$$\lim_{n\to\infty} \operatorname{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sum\limits_{i,j=1}^{n} \Gamma_{i,j}^{(n)}}{n^2} = 0.$$

**Proof:** Applying Proposition 1 with $\alpha = \frac{1}{n}e$, we obtain that for any $n \geq 1$:

$$\sup_{X^{(n)} \sim (\mu \cdot e, \Gamma^{(n)})} P\left(\frac{\sum_{i=1}^n X_i}{n} > \tau\right) = \begin{cases} \dfrac{1}{1 + (\tau - \mu)^2 \cdot \dfrac{n^2}{\sum\limits_{i,j=1}^{n} \Gamma_{i,j}^{(n)}}} & , \text{ if } \tau > \mu, \\ 1 & , \text{ if } \tau \leq \mu . \end{cases}$$

Therefore, if $\sum\limits_{i,j=1}^{n} \Gamma_{i,j}^{(n)}/n^2$ converges to 0 as $n \to \infty$, then :

$$\sup_{X^{(n)} \sim (\mu \cdot e, \Gamma^{(n)})} P\left(\frac{\sum_{i=1}^n X_i}{n} > \tau\right) \longrightarrow \begin{cases} 0 & , \text{ if } \tau > \mu, \\ 1 & , \text{ if } \tau \leq \mu . \end{cases}$$

This shows that for any such infinite sequence of random variables, the Law of Large Numbers holds.

Conversely, if $\sum\limits_{i,j=1}^{n} \Gamma_{i,j}^{(n)}/n^2$ does not converge to 0, then there is a subsequence $\sum\limits_{i,j=1}^{n_k} \Gamma_{i,j}^{(n_k)}/n_k^2$ that converges to a constant $\theta$, or it diverges to infinity. We found in Theorem 6 a sequence of extremal distributions that satisfies

$$P\left(\frac{\sum_{i=1}^{n_k} X_i}{n_k} > \tau\right) = \begin{cases} \dfrac{1}{1 + \dfrac{(\tau - \mu)^2}{\theta}} & , \text{ if } \tau > \mu, \\ 1 & , \text{ if } \tau \leq \mu . \end{cases}$$

Such a subsequence clearly violates the Law of Large Numbers. ■

**Remark:** The law of large numbers for independent, identically distributed random variables assumes that $E[|X_i|] < \infty$. This implies that $\mathrm{Var}(X_i) < \infty$, and thus we have $\lim\limits_{n \to \infty} \mathrm{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = 0$. Therefore, in this case the usual law of large numbers follows from Theorem 7.

## 4.2 Fat Tails and The Central Limit Theorem for Uncorrelated Random Variables.

Consider a sequence of random variables $X^{(n)} = (X_1, \ldots, X_n)$. If the random variables $X_i$ are independent and identically distributed, then the central limit theorem holds. Suppose, we relax the independence condition by only assuming instead that $X^{(n)} \sim (\mu \cdot e, \sigma^2 I)$, i.e., $X_i$ are identically distributed and uncorrelated but not necessarily independent. Is it true that the central limit theorem holds in this case?

Applying Proposition 1 with $\alpha = e$, $\tau = t\sqrt{e'\Gamma^{(n)}e} + e'M^{(n)}$ we obtain that for any $n \geq 1$:

$$\sup_{X\sim(\mu\cdot e,\sigma^2 I)} P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \geq t\right) = \begin{cases} \dfrac{1}{1+t^2} & \text{, if } t > 0, \\ 1 & \text{, if } t \leq 0\,. \end{cases}$$

$$\inf_{X\sim(\mu\cdot e,\sigma^2 I)} P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \geq t\right) = \begin{cases} \dfrac{t^2}{1+t^2} & \text{, if } t \leq 0, \\ 0 & \text{, if } t > 0\,. \end{cases}$$

Moreover, from Theorem 6 there exist extremal distributions that achieve these bounds. Such distributions clearly violate the central limit theorem, as they induce much "fatter tails" for $\sum_{i=1}^n X_i$ than the one (normal distribution) predicted by the central limit theorem.

### 4.3 The Multivariate Markov Inequality.

Given a vector $M = (M_1, \ldots, M_n)'$, we derive in this section tight bounds on the following upper tail of a multivariate nonnegative random variable $X = (X_1, \ldots, X_n)'$ with mean $M = E[X]$:

$$P(X > M_{e+\delta}) = P(X_i > (1+\delta_i)M_i\,,\ \forall i = 1, \ldots, n).$$

where $\delta = (\delta_1, \ldots, \delta_n)'$, and we denote by $M_\delta = (\delta_1 M_1, \ldots, \delta_n M_n)'$.

The bounds we derive constitute multivariate generalizations of Markov's inequality.

**Theorem 8** *The tight multivariate Markov bound is*

$$\sup_{X\sim M} P(X > M_{e+\delta}) = \min_{i=1,\ldots,n} \frac{1}{1+\delta_i}. \tag{12}$$

**Proof:** Applying the bound (1) for $S = \{x|\ x_i \geq (1+\delta_i)M_i\,,\ \forall i = 1, \ldots, n\}$, we obtain Eq. (12). ∎

The bound (12) is a natural generalization of the Markov inequality. In particular, for a nonnegative univariate random variable, in the case that $S = [(1+\delta)M, \infty)$, then the bound (12) is exactly Markov inequality:

$$\sup_{X \sim M} P(X \geq (1+\delta)M) = \frac{1}{1+\delta}.$$

### 4.4 The Multivariate Chebyshev Inequality.

Given a vector $M = (M_1, \ldots, M_n)'$, and an $n \times n$ positive definite, full rank matrix $\Gamma$, we derive in this section tight bounds on the following upper, lower, and two-sided tail probabilities of a multivariate random variable $X = (X_1, \ldots, X_n)'$ with mean $M = E[X]$ and covariance matrix $\Gamma \doteq E[(X-M)(X-M)']$:

$$\begin{aligned} P(X > M_{e+\delta}) &= P(X_i > (1+\delta_i)M_i\ ,\ \forall i = 1, \ldots, n), \\ P(X < M_{e-\delta}) &= P(X_i < (1-\delta_i)M_i\ ,\ \forall i = 1, \ldots, n), \\ P(X > M_{e+\delta} \text{ or } X < M_{e-\delta}) &= P(|X_i - M_i| > \delta_i M_i\ ,\ \forall i = 1, \ldots, n), \end{aligned}$$

where $\delta = (\delta_1, \ldots, \delta_n)'$, and we denote by $M_\delta = (\delta_1 M_1, \ldots, \delta_n M_n)'$.

The bounds we derive constitute multivariate generalizations of Chebyshev's inequality. Surprisingly, they improve upon the Chebyshev's inequality for scalar random variables. In order to obtain non-trivial bounds we require that not all $\delta_i M_i \leq 0$, which expresses the fact that the tail event does not include the mean vector.

### The One-Sided Chebyshev Inequality.

In this section, we find a tight bound for $P(X > M_{e+\delta})$. The bounds immediately extends to $P(X < M_{e-\delta})$.

**Theorem 9** **(a)** *The tight multivariate one-sided Chebyshev bound is*

$$\sup_{X \sim (M,\Gamma)} P(X > M_{e+\delta}) = \frac{1}{1+d^2}, \tag{13}$$

*where $d^2$ is given by:*

$$d^2 = \text{minimize} \quad x'\Gamma^{-1}x \tag{14}$$

$$\text{subject to} \quad x \geq M_\delta,$$

*or alternatively $d^2$ is given by the Gauge dual problem of (14):*

$$\begin{aligned} \frac{1}{d^2} = \text{minimize} \quad & x'\Gamma x \\ \text{subject to} \quad & x' M_\delta = 1 \\ & x \geq 0. \end{aligned} \tag{15}$$

**(b)** *If $\Gamma^{-1} M_\delta \geq 0$, then the tight bound is expressible in closed form:*

$$\sup_{X \sim (M,\Gamma)} P(X > M_{e+\delta}) = \frac{1}{1 + M_\delta' \Gamma^{-1} M_\delta}. \tag{16}$$

**Proof:** **(a)** Applying the Bound (5) for $S = \{x |\ x_i > (1+\delta_i) M_i\ ,\ \forall i = 1, \ldots, n\}$, and changing variables we obtain Eq. (13). The alternative expression (15) for $d^2$ follows from elementary Gauge duality theory (see Freund [2]).

**(b)** The Kuhn-Tucker conditions for Problem (14) are as follows:

$$2\Gamma^{-1} x - \lambda = 0, \qquad \lambda \geq 0, \qquad x \geq M_\delta, \qquad \lambda'(x - M_\delta) = 0.$$

The choice $x = M_\delta$, $\lambda = 2\Gamma^{-1} M_\delta \geq 0$ (by assumption) satisfies the Kuhn-Tucker conditions, which are sufficient (this is a convex quadratic optimization problem). Thus, $d^2 = M_\delta' \Gamma^{-1} M_\delta$, and hence, Eq. (16) follows. ■

### The Two-Sided Chebyshev Inequality.

In this section, we find a tight bound for $P(X > M_{e+\delta}$ or $X < M_{e-\delta})$.

**Theorem 10** **(a)** *The tight multivariate two-sided Chebyshev bound is*

$$\sup_{X \sim (M,\Gamma)} P(X > M_{e+\delta} \text{ or } X < M_{e-\delta}) = \min(1, t^2), \tag{17}$$

*where*

$$t^2 = \text{minimize} \quad x'\Gamma x \tag{18}$$

$$\text{subject to} \quad x'M_\delta = 1$$
$$x \geq 0.$$

**(b)** *If* $\Gamma^{-1}M_\delta \geq 0$*, then the tight bound is expressible in closed form:*

$$\sup_{X\sim(M,\Gamma)} P(X > M_{e+\delta} \text{ or } X < M_{e-\delta}) = \min\left(1, \frac{1}{M_\delta'\Gamma^{-1}M_\delta}\right). \tag{19}$$

**Proof:** Problem $(D)$ in this particular case becomes:

$$\begin{aligned} Z_D = \text{minimize} \quad & E[g(X)] \\ \text{subject to} \quad & g(x) \text{ 2-degree } n\text{-variate polynomial} \\ & g(x) \geq \begin{cases} 1, & \text{if } x > M_{e+\delta} \text{ or } x < M_{e-\delta}, \\ 0, & \text{otherwise} . \end{cases} \end{aligned}$$

Similar to Lemma 2, we show in an analogous way that either the dual optimum is 1, or else there exists an optimal solution of the form $g(x) = (a'(x - M))^2$, for some vector $a$. Therefore, the dual problem is equivalent to:

$$\begin{aligned} Z_D = \text{minimize} \quad & E[g(X)] \\ \text{subject to} \quad & g(x) = 1, \quad \forall x \in R^n \\ & \quad \text{or} \\ & g(x) = (a'(x - M))^2 \geq \begin{cases} 1, & \text{if } x > M_{e+\delta} \text{ or } x < M_{e-\delta}, \\ 0, & \text{otherwise} . \end{cases} \end{aligned} \tag{20}$$

Suppose that $g(x) = (a'(x-M))^2$ is optimal for Problem (20). Then, $g(M_{e+\delta}) = g(M_{e-\delta}) = 1$, that is $(a'M_\delta)^2 = 1$. The feasibility constraints are $g(x + M_{e+\delta}) \geq 1$, $\forall x \geq 0$ and $g(-x + M_{e-\delta}) \geq 1$, $\forall x \geq 0$, or equivalently $(a'(x + M_\delta))^2 \geq (a'M_\delta)^2$, $\forall x \geq 0$, which is further equivalent to $a \geq 0$ or $a \leq 0$. Therefore, the dual problem can be reformulated as:

$$\begin{aligned} Z_D = \text{minimize} \quad & (1, E[(a'(X - M))^2]) = \min(1, a'\Gamma a) \\ \text{subject to} \quad & a'M_\delta = 1 \\ & a \geq 0, \end{aligned}$$

from which Eq. (17) follows.

**(b)** If $\Gamma^{-1}M_\delta \geq 0$, then $a_0 = \frac{\Gamma^{-1}M_\delta}{M_\delta \Gamma^{-1} M_\delta}$ is feasible and $a_0'\Gamma a_0 = \frac{1}{M_\delta\Gamma^{-1}M_\delta}$. By the Cauchy-Schwartz inequality, for an arbitrary $a$:

$$1 = (a'M_\delta)^2 \leq (a'\Gamma a)(M_\delta'\Gamma^{-1}M_\delta),$$

or equivalently $a'\Gamma a \geq \frac{1}{M_\delta\Gamma^{-1}M_\delta} = a_0'\Gamma a_0$, which means $a_0$ is optimal and the closed form bound is indeed $\dfrac{1}{M_\delta\Gamma^{-1}M_\delta}$. ■

In the univariate case $M_\delta = \delta M$ and $\Gamma = \sigma^2$. Therefore, $\Gamma^{-1}M_\delta = \dfrac{\delta M}{\sigma^2} \geq 0$, and the closed form bound applies, i.e.,

$$P(X > (1+\delta)M) \leq \frac{C_M^2}{\delta^2 + C_M^2}, \tag{21}$$

where $C_M^2 = \dfrac{\sigma^2}{M^2}$ is the coefficient of variation of the random variable $X$. The usual Chebyshev inequality is given by $P(X > (1+\delta)M) \leq \frac{C_M^2}{\delta^2}$. Surprisingly, Inequality (21) is always stronger. Moreover, as we showed in Theorem 6 there exist extremal distributions that satisfy it with equality.

# 5 Closed Form Bounds for Univariate Tail Probabilities.

In this section, we restrict our attention to univariate random variables. Given the first $k$ moments $M_1, \ldots, M_k$ of a real random variable $X$ with domain $\Omega$, we are interested in deriving tight bounds on the tail probabilities:

$$P(X > (1+\delta)M_1), \qquad P(X < (1-\delta)M_1), \qquad P(|X - M_1| > \delta M_1),$$

for positive deviations $\delta M_1 > 0$.

We define the squared coefficient of variation: $C_M^2 = \dfrac{M_2 - M_1^2}{M_1^2}$, and the third order coefficient of variation $D_M^2 = \dfrac{M_1M_3 - M_2^2}{M_1^4}$.

**Theorem 11** *The following bounds in Table 1 are tight for $k = 1, 2, 3$. The bounds for arbitrary $k$ are not necessarily tight. When $k$ is odd, $\Omega = R_+$, while for even $k$, $\Omega = R$.*

| $(k,\ \Omega)$ | $P(X > (1+\delta)M_1)$ | $P(X < (1-\delta)M_1)$ | $P(\lvert X - M_1\rvert > \delta M_1)$ |
|---|---|---|---|
| $(1, R_+)$ | $\frac{1}{1+\delta}$ | $1$ | $1$ |
| $(2, R)$ | $\frac{C_M^2}{C_M^2+\delta^2}$ | $\frac{C_M^2}{C_M^2+\delta^2}$ | $\min\left(1, \frac{C_M^2}{\delta^2}\right)$ |
| $(3, R_+)$ | $f_1(C_M^2, D_M^2, \delta)$ | $f_2(C_M^2, D_M^2, \delta)$ | $f_3(C_M^2, D_M^2, \delta)$ |
| $(k, \Omega)$ | $\min\limits_{\gamma\in\Omega} \frac{\sum_i C(k,i) M_i \gamma^{k-i}}{((1+\delta)M_1+\gamma)^k}$ | $\min\limits_{\gamma\in\Omega} \frac{\sum_i C(k,i) M_i \gamma^{k-i}}{((1-\delta)M_1+\gamma)^k}$ | $\min\limits_{\gamma\in\Omega} \frac{\sum_i C(k,i) M_i \gamma^{k-i}}{((1-\delta)M_1+\gamma)^k}$ |

Table 1: Bounds for tail probabilities. We use the notation $C(k,i) = \binom{k}{i}$.

The following definitions are used:

$$f_1(C_M^2, D_M^2, \delta) = \begin{cases} \min\left(\frac{C_M^2}{C_M^2+\delta^2}, \frac{1}{1+\delta}\cdot\frac{D_M^2}{D_M^2+(C_M^2-\delta)^2}\right), & \text{if } \delta \geq C_M^2, \\ \frac{1}{1+\delta}\cdot\frac{D_M^2+(1+\delta)(C_M^2-\delta)}{D_M^2+(1+C_M^2)(C_M^2-\delta)}, & \text{if } \delta < C_M^2, \end{cases}$$

$$f_2(C_M^2, D_M^2, \delta) = 1 - \frac{(C_M^2+1)^3}{(D_M^2+(C_M^2+1)(C_M^2+\delta))(D_M^2+(C_M^2+\delta)^2)},$$

$$f_3(C_M^2, D_M^2, \delta) = \min\left(1, 1+3^3\frac{d_M^2+C_M^4-\delta^2}{4+3(1+3\delta^2)+2(1+3\delta^2)^{\frac{3}{2}}}\right).$$

**Proof:**

The inequality for $k=1$ and $\Omega = R_+^n$ follows from Eq. (12) (Markov's inequality). It is also tight as indicated from the following distribution:

$$X = \begin{cases} 0, & \text{with probability} \quad \dfrac{\delta}{1+\delta}, \\ (1+\delta)M_1, & \text{with probability} \quad \dfrac{1}{1+\delta}. \end{cases}$$

The one-sided tail inequalities for $k = 2$ and $\Omega = R^n$ follow from Eq. (21). They are also tight as indicated by Theorem 6. The two-sided tail inequality for for $k = 2$ and $\Omega = R^n$ follows from Eq. (17). It is tight as indicated from the following distribution:

$$X = \begin{cases} (1+\delta)M_1, & \text{with probability} \quad \dfrac{C_M^2}{2\delta^2}, \\ (1-\delta)M_1, & \text{with probability} \quad \dfrac{C_M^2}{2\delta^2}, \\ M_1, & \text{with probability} \quad 1 - \dfrac{C_M^2}{\delta^2}. \end{cases}$$

### The $(1, 3, \mathbf{R}_+)$-Bound.

Let $j = (1+\delta)M_1$. The necessary and sufficient condition for $(M_1, M_2, M_3)$ to be a valid sequence is $C_M^2 = M_2 - M_1^2 \geq 0$, and $D_M^2 = M_1M_3 - M_2^2 \geq 0$. The dual feasible solution $g(x)$ needs to satisfy $g(x) \geq 0$ for all $x \geq 0$, and $g(x) \geq 1$, for all $x \geq j$. At optimality $g(j) = 1$, otherwise we can decrease the objective function further. Therefore, there are three possible types of dual optimal functions $g(\cdot)$:

**(a)** (See Figure 1) $g(x) = \left(\dfrac{x-\gamma}{j-\gamma}\right)^3$, $\gamma \leq 0$.

**(b)** (See Figure 2) $g(x) = \dfrac{(x-\gamma_1)(x-\gamma_2)^2}{(j-\gamma_1)(j-\gamma_2)^2}$, $\gamma_1 < 0, \gamma_1 < \gamma_2 < j$.

**(c)** (See Figure 3) $g(x) = a\,\dfrac{(x-\gamma)^2(x-j)}{\gamma^2 j} + 1$, $a \leq 1, \gamma \geq j$.

The tight bound using only $M_1$ and $M_2$ is:

$$P(X > j) \leq Z_0 = \min\left(\frac{1}{1+\delta}, \frac{C_M^2}{C_M^2+\delta^2}\right) = \begin{cases} \dfrac{1}{1+\delta}, & \text{for} \quad C_M^2 > \delta, \\ \dfrac{C_M^2}{C_M^2+\delta^2}, & \text{for} \quad C_M^2 \leq \delta. \end{cases} \tag{22}$$

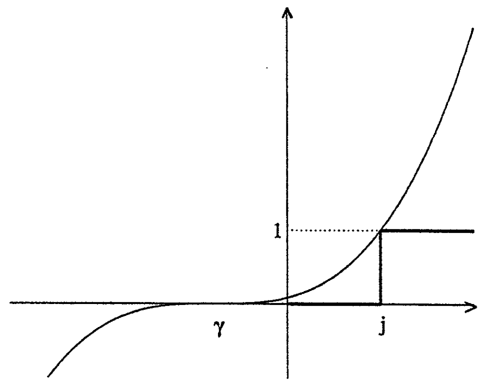


Figure 1: The function $g(x) = \left(\frac{x-\gamma}{j-\gamma}\right)^3$, $\gamma \le 0$.

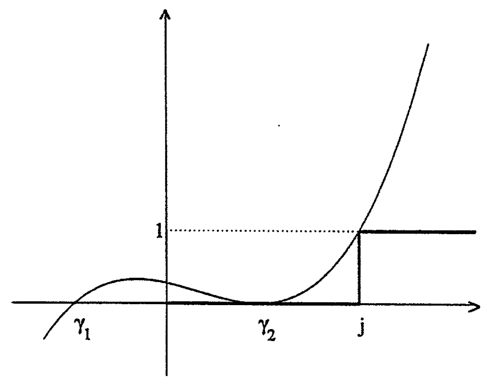


Figure 2: The function $g(x) = \frac{(x-\gamma_1)(x-\gamma_2)^2}{(j-\gamma_1)(j-\gamma_2)^2}$, $\gamma_1 < 0, \gamma_1 < \gamma_2 < j$.

We next examine the bounds obtained by optimizing undetermined parameters in cases (a), (b), and (c).

**Case (a).** The best possible bound in this case is:

$$Z_a = \min_{\gamma \le 0} \frac{E[(X-\gamma)^3]}{(j-\gamma)^3} = \min_{\gamma \le 0} \frac{M_3 - 3\gamma M_2 + 3\gamma^2 M_1 - \gamma^3}{(j-\gamma)^3}.$$

We differentiate with respect to $\gamma$ and we obtain that the critical point satisfies:

$$E[(X-\gamma)^3] = (j-\gamma)E[(X-\gamma)^2],$$

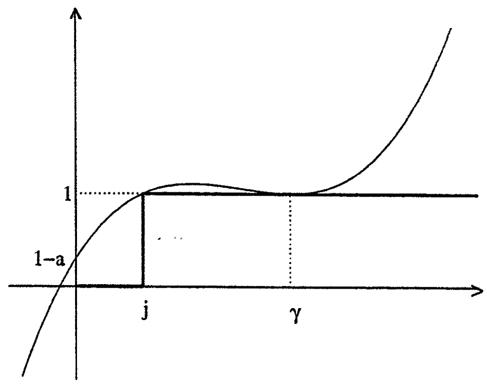


Figure 3: The function $g(x) = a\,\dfrac{(x-\gamma)^2(x-j)}{\gamma^2 j} + 1,\ a \le 1, \gamma \ge j.$

which leads to :

$$\gamma^2(M_1 - j) - 2\gamma(M_2 - M_1 j) + M_3 - M_2 j = 0. \tag{23}$$

There are two possibilities to consider:

(i) If $M_3 \ge jM_2$, then Eq. (23) has a feasible solution $\gamma^* < 0$. The dual objective function thus becomes:

$$Z_a' = \frac{E[(X-\gamma^*)^2]}{(j-\gamma^*)^2} \ge \min_{\gamma} \frac{E[(X-\gamma)^2]}{(j-\gamma)^2} = \frac{C_M^2}{C_M^2+\delta^2} \ge Z_0,$$

and thus this bound is dominated by $Z_0$.

(ii) If $M_3 < jM_2$, then $M_2 \le jM_1$, otherwise $M_1M_3 - M_2^2 < 0$, and thus $(M_1, M_2, M_3)$ is not a valid moment sequence. Therefore, there does not exist a solution of Eq. (23) with $\gamma^* < 0$. Thus, the optimal solution is for $\gamma^* = 0$, and the dual objective function becomes:

$$Z_a = \frac{M_3}{j^3} = \frac{D_M^2 + (C_M^2+1)^2}{(1+\delta)^3}. \tag{24}$$

**Case (b).** The best possible bound in this case is:

$$Z_b = \min_{\gamma_1,\gamma_2} \frac{E[(X-\gamma_1)(X-\gamma_2)^2]}{(j-\gamma_1)(j-\gamma_2)^2} = \min_{\gamma_1,\gamma_2} \frac{1}{(j-\gamma_1)} \cdot \frac{E[(X-j)(X-\gamma_2)^2]}{(j-\gamma_2)^2} + E\left[\left(\frac{X-\gamma_2}{j-\gamma_2}\right)^2\right].$$

In order for an optimal solution to produce a non-dominated bound, it must be that $E[(X-\gamma_2)^2(X-j)] < 0$, or else the dual objective is at least: $E\left[\left(\frac{X-\gamma_2}{j-\gamma_2}\right)^2\right] \geq \frac{C_M^2}{C_M^2+\delta^2}$. Therefore, in such an optimal solution we should set $(j-\gamma_1)$ as small as possible, so $\gamma_1 = 0$. The dual objective becomes:

$$Z_b = \min_{0\leq\gamma_2<j} \frac{E[X(X-\gamma_2)^2]}{j(j-\gamma_2)^2} .$$

When we differentiate the objective function with respect to $\gamma_2$, we obtain that the critical point must satisfy

$$E[X(X-\gamma_2)^2] = (j-\gamma_2)E[X(X-\gamma_2)], \tag{25}$$

which leads to:

$$\gamma_2^* = \frac{M_3 - jM_2}{M_2 - jM_1} = M_1\left[1 + C_M^2 + \frac{D_M^2}{C_M^2 - \delta}\right].$$

There are two possibilities to consider:

(i) If $\delta \leq C_M^2$, then $\gamma_2^* \geq j$, and the optimum is obtained by setting $\gamma_2 = 0$, which produces the dominated bound:

$$Z_b' = \frac{M_3}{j^3} = \frac{D_M^2 + (C_M^2+1)^2}{(1+\delta)^3} \geq \frac{1}{1+\delta} = Z_0 .$$

(ii) If $\delta > C_M^2$, then $\gamma_2^* < j$. Then, substituting $\gamma_2^*$, we obtain the bound:

$$Z_b = \frac{1}{1+\delta} \cdot \frac{D_M^2}{D_M^2 + (C_M^2 - \delta)^2}. \tag{26}$$

**Case (c).** In this case $g(x) = a\frac{(x-\gamma)^2(x-j)}{\gamma^2 j} + 1,\ a \leq 1, \gamma > j$ . First notice that $a$ must be 1 in an optimal solution, and the bound becomes:

$$Z_c = \min_{\gamma\geq j} \frac{E[(X-\gamma)^2(X-j)]}{\gamma^2 j} + 1 = \min_{\gamma\geq j} \frac{M_3 - M_2(2\gamma+j) + M_1(\gamma^2 + 2\gamma j)}{\gamma^2 j} .$$

Again, by differentiating with respect to $\gamma$, we obtain the same critical point: $\gamma^* = \dfrac{M_3 - jM_2}{M_2 - jM_1}$, which satisfies

$$E[(X-\gamma)^2(X-j)] = \gamma E[(X-\gamma)(X-j)].$$

There are two possibilities:

(i) If $C_M^2 \geq \delta$, then $\gamma^* \geq j$, and we obtain the bound

$$Z_c = \frac{1}{1+\delta} \cdot \frac{D_M^2 + (1+\delta)(C_M^2 - \delta)}{D_M^2 + (1+C_M^2)(C_M^2 - \delta)}. \tag{27}$$

(ii) If $C_M^2 < \delta$, then $\gamma^* < j$, and the optimum is obtained by setting $\gamma = j$, which produces the dominated bound

$$Z_c' = \frac{M_3 - 3jM_2 + 3j^2M_1}{j^3} = \frac{1}{(1+\delta)^3}[D_M^2 + (C_M^2 - \delta)(C_M^2 - 2\delta - 1)] + \frac{1}{1+\delta} > \frac{1}{1+\delta} \geq Z_0.$$

Combining all previous case, we obtain that

$$Z_D = \begin{cases} \min(Z_0, Z_a, Z_b) & \text{, if } C_M^2 < \delta, \\ Z_c & \text{, if } C_M^2 \geq \delta \end{cases}$$

Moreover, one can easily check that:

$$Z_b = \frac{1}{1+\delta} \cdot \frac{D_M^2}{D_M^2 + (C_M^2 - \delta)^2} \leq \frac{D_M^2 + (C_M^2 + 1)^2}{(1+\delta)^3} = Z_a,$$

and the theorem follows. The formulae for the left tail and the two-sided inequality follow by a similar construction.

**On the $(1, 2m, R), (1, 2m+1, R_+)$-Bound.**

We next proceed to find bounds for general $k$, and consider $\sup_{X \sim (M_1,\ldots,M_k)} P(X > (1+\delta)M_1)$. Problem $(D)$ in this case becomes:

$$\begin{aligned} Z_D = \text{minimize} \quad & u_0 + u_1 M_1 + \cdots + u_k M_k \\ \text{subject to} \quad & g(x) = u_0 + u_1 x + \cdots + u_k x^k \geq \begin{cases} 0, & \text{if } x < j, \\ 1, & \text{if } x \geq j, \end{cases} \end{aligned} \tag{28}$$

where $j = (1+\delta)M_1$ for the right tail and $j = (1-\delta)M_1$ for the left tail.

We assume that $k$ is even (if $k$ is odd and $\Omega = R$, then the tight bound is 1). Notice that for a $k$-degree polynomial $g(\cdot)$ to be feasible, all its roots must have even multiplicity, otherwise the nonnegativity constraint will be violated. Moreover, any optimal solution must satisfy $g(j) = 1$. We can obtain a valid (but not necessarily tight) bound by optimizing Problem (28) over a smaller range of functions $g(\cdot)$ that satisfy the above necessary conditions, and have a single root of multiplicity $k$:

$$g(x) = \left(\frac{x+\gamma}{j+\gamma}\right)^k,$$

where $\gamma$ is a parameter we can optimize over. This leads to the following bounds:

$$\begin{aligned} P(X > (1+\delta)M_1) &\leq \min_{\gamma \in \Omega} \frac{\sum_i C(k,i) M_i \gamma^{k-i}}{((1+\delta)M_1 + \gamma)^k}, \\ P(X < (1-\delta)M_1) &\leq \min_{\gamma \in \Omega} \frac{\sum_i C(k,i) M_i \gamma^{k-i}}{((1-\delta)M_1 + \gamma)^k}, \end{aligned}$$

where $C(k,i) = \binom{k}{i}$. When $k$ is odd, and $\Omega = R^n_+$, the same construction works, but we need to restrict $\gamma$ to be positive, i.e., $\gamma \in \Omega$. ■

**Remark:** When $X \sim (M_1, \ldots, M_k)$, we can obtain weaker closed-form bounds by selecting specific values for $\gamma$. For example, for the right tail, if we let $\gamma = -M_1$ we obtain the bound:

$$P(X > (1+\delta)M_1) \leq \frac{1}{(-\delta)^k} \sum_{i=1}^{k} \frac{C(k,i) M_i}{(-M_1)^i}.$$

If we select $\gamma = \dfrac{M_2 - M_1^2}{\delta M_1} - M_1$, which is the value that yields the tight bound (13) applied for $n = 1, k = 2$, we obtain the bound:

$$P(X > (1+\delta)M_1) \leq \left(\frac{1}{C_M^2 + \delta^2}\right)^k \sum_{i=1}^{k} C(k,i) \frac{M_i}{M_1^i} \delta^i (C_M^2 - \delta)^{k-i}.$$

The same procedure applies for deriving corresponding closed-form left tail bounds.

# 6 The Complexity of The $(\mathbf{n}, 2, \mathbf{R}_+^\mathbf{n})$, $(\mathbf{n}, \mathbf{k}, \mathbf{R}^\mathbf{n})$-Bound Problems.

In this section, we show that the separation problem associated with Problem $(D)$ for the cases $(n, 2, R_+^n)$, $(n, k, R^n)$-bound problems are NP-hard for $k \geq 3$. By the equivalence of optimization and separation (see Grötschel, Lovász and Schrijver [3]), solving Problem $(D)$ is NP-hard as well. Finally, because of Theorem 2, solving the $(n, 2, R_+^n)$, $(n, k, R^n)$-bound problems with $k \geq 3$ is NP-hard.

## 6.1 The Complexity of The $(\mathbf{n}, 2, \mathbf{R}_+^\mathbf{n})$-Bound Problem.

The separation problem can be formulated as follows in this case:

**Problem 2SEP**: Given a multivariate polynomial $g(x) = x'Hx + c'x + d$, and a set $S \subseteq R_+^n$, does there exist $x \in S$ such that $g(x) < 0$ ?

If we consider the special case $c = 0$, $d = 0$, and $S = R_+^n$, Problem 2SEP reduces to the question whether a given matrix $H$ is co-positive, which is NP-hard (see Murty and Kabadi [8]).

## 6.2 The Complexity of The $(\mathbf{n}, \mathbf{k}, \mathbf{R}^\mathbf{n})$-Bound Problem for $k \geq 3$.

For $k \geq 3$, the separation problem can be formulated as follows:

**Problem 3SEP**: Given a multivariate polynomial $g(\cdot)$ of degree $k \geq 3$, and a set $S \subseteq R^n$, does there exist $x \in S$ such that $g(x) < 0$ ?

We show that problem $3SEP$ is NP-hard by performing a reduction from $3SAT$ (see Sipser [15]).

**Theorem 12** *Problem $3SAT$ polynomially reduces to $3SEP$.*

**Proof:** For an arbitrary $3SAT$ instance $\phi$ (a 3CNF boolean formula in $n$ variables), we consider the following arithmetization $g_\phi(\cdot)$ of $\phi$: we replace each boolean variable $x_i$ by the monomial $1 - x_i$, its negation $\bar{x}_i$ by $x_i$, and we convert $\bigwedge$'s into additions and $\bigvee$'s to multiplications. For example, the arithmetization of the formula $\phi = (\bar{x}_1 \vee x_2 \vee \bar{x}_3) \wedge (x_1 \vee x_3 \vee \bar{x}_4)$, is: $g_\phi(x) = x_1(1-x_2)x_3 + (1-x_1)(1-x_3)x_4$.

As motivation for the proof, note that $g_\phi(\cdot)$ is a 3-degree polynomial in $n$ variables, evaluating to zero at any satisfying assignment of $\phi$. Also note that $g_\phi(x)$ is a nonnegative integer for any boolean assignment $x \in \{0,1\}^n$. Thus if $\phi$ is unsatisfiable, then $g_\phi(x) \geq 1$ for any boolean assignment $x \in \{0,1\}^n$.

Starting with an instance $\phi$ of 3SAT with $n$ variables and $m$ clauses, we construct an instance $(g(\cdot), S)$ of $3SEP$ as follows:

$$g(x) = 2g_\phi(x) + (24m)^2 \sum_{i=1}^{n} x_i(1-x_i) - 1, \qquad S = [0,1]^n.$$

Note that the construction can be done in polynomial time. We next show that formula $\phi$ is satisfiable if and only if there exists $x \in S$ such that $g(x) < 0$.

Clearly if $\phi$ is satisfiable, there exists a satisfying assignment corresponding to a vector $x_0 \in \{0,1\}^n$. Clearly $g_\phi(x_0) = 0$, and thus $g(x_0) = -1 < 0$.

Conversely, suppose $\phi$ is not satisfiable. We will show that for all $x \in S = [0,1]^n$, $g(x) \geq 0$. Let $\epsilon = \dfrac{1}{24m}$. For any $x \in S = [0,1]^n$, there are two possibilities:

**(a)** There exists a boolean vector $y \in \{0,1\}^n$ such that $|x_i - y_i| < \epsilon$, $\forall i$.

If we expand the term in $g_\phi(\cdot)$ corresponding to each of the $m$ clauses of $\phi$ as a polynomial, we obtain a sum of at most one monomial of degree three, three monomials of degree two, three monomials of degree one, and one monomial of degree zero. Let $S_k$ be the set $k$-tuples corresponding to the monomials of degree $k$, $k = 1,2,3$. Then, $|S_1| \leq 3m$, $|S_2| \leq 3m$, $|S_3| \leq m$. Matching corresponding monomials for $x$ and $y$, canceling constants, and applying the triangle inequality, we obtain:

$$|g_\phi(x) - g_\phi(y)| \leq \sum_{(i,j,k)\in S_3} |x_i x_j x_k - y_i y_j y_k| + \sum_{(i,j)\in S_2} |x_i x_j - y_i y_j| + \sum_{i\in S_1} |x_i - y_i|.$$

Since $|x_i - y_i| < \epsilon$, $\forall i$, we obtain:

$$\begin{aligned} |x_i x_j x_k - y_i y_j y_k| &\leq |x_i x_j x_k - y_i x_j x_k| + |y_i x_j x_k - y_i y_j x_k| + |y_i y_j x_k - y_i y_j y_k| \\ &= |x_i - y_i| x_j x_k + y_i |x_j - y_j| x_k + y_i y_j |x_k - y_k| \leq 3\epsilon, \end{aligned}$$

since $x, y \in [0,1]^n$. Similarly, $|x_i x_j - y_i y_j| \leq 2\epsilon$. Therefore,

$$|g_\phi(x) - g_\phi(y)| \leq 3\epsilon|S_3| + 2\epsilon|S_2| + \epsilon|S_1| \leq 12m\epsilon = \frac{1}{2}.$$

Thus, $g_\phi(x) > g_\phi(y) - \frac{1}{2}$. Since $\phi$ is not satisfiable, we have $g_\phi(y) \geq 1$, for any boolean vector $y \in \{0,1\}^n$. Thus, $g_\phi(x) \geq 1 - \frac{1}{2} = \frac{1}{2}$, and hence $g(x) \geq 2g_\phi(x) - 1 \geq 0$.

**(b)** There exists at least one $i$ for which $\epsilon \leq x_i \leq 1 - \epsilon$ . This implies $x_i(1 - x_i) \geq \epsilon^2$, and, since $g_\phi(x) \geq 0$, $\forall\, x \in S$, it follows that $g(x) \geq (24m)^2\epsilon^2 - 1 = 0$.

Therefore, if $\phi$ is not satisfiable, then all $x \in S = [0,1]^n$, satisfy $g(x) \geq 0$, and the theorem follows. ■

# 7 Concluding Remarks.

In this paper, we completely characterized the complexity of the $(n, k, \Omega)$-bound problem, by providing polynomial time algorithms for the $(n, 1, \Omega)$, $(n, 2, R^n)$-bound problems, and by showing that the $(n, 2, R^n_+)$, $(n, k, R_n)$-bound problems for $k \geq 3$ are NP-hard. We found explicit tight bounds that generalize and improve upon the classical Markov, and Chebyshev inequalities. Finally, we found extremal distributions that achieve the new bounds, and examined implications of the tight bounds to the law of large numbers, and the central limit theorem. We find it interesting that classical results of probability can be attacked using mathematical optimization methods. We expect that using ideas from optimization to classical probability problems is a fertile area of research that can lead to new insights in probability theory.

## References


[1] N. I. Akhiezer. *The Classical Moment Problem.* Hafner, New York, NY, 1965.

[2] R. Freund. Dual gauge programs, with applications to quadratic programming and the minimum-norm problem. *Mathematical Programming*, 38:47–67, 1987.

[3] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization.* Algorithms and Combinatorics; 2. Springer, Berlin-Heidelberg, 1988.

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

[5] S. Karlin and L.S. Shapley. Geometry of moment spaces. *Memoirs Amer. Math. Soc.*, 12, 1953.

[6] J. H. B. Kempermann. On the role of duality in the theory of moments. In *Lecture Notes in Econom. and Math. Systems*, volume 215, pages 63–92, Berlin-New York, 1983. Springer.

[7] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and Applications.* Academic Press, New York, NY, 1980.

[8] K. G. Murty and S. N. Kabadi. Some np-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:117–129, 1987.

[9] G.L. Nemhauser and L.A. Wolsey. *Integer and Combinatorial Optimization.* Wiley, New York, NY, 1988.

[10] Y. Nesterov and A. Nemirovskii. Interior point polynomial algorithms for convex programming. *Studies in Applied Mathematics*, 13, 1994.

[11] I. Pitowski. Correlation polytopes: Their geometry and complexity. *Mathematical Programming*, 50:395–414, 1991.

[12] I. Pitowski. Correlation polytopes ii: The geometry of limit laws in probability. preprint, 1996.

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

[14] S. Sengupta, W. Siu, and N.C. Lind. A linear programming formulation of the problem of moments. *Z. Angew. Math. Mecg.*, 10:533–537, 1979.

[15] M. Sipser. *Introduction to the Theory of Computation.* PWS, Boston, MA, 1996.

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