Noname manuscript No.
(will be inserted by the editor)

On Reduced Semidefinite Programs for Second Order Moment Bounds with Applications

Karthik Natarajan · Chung-Piaw Teo

Received: date / Accepted: date

Abstract We show that the complexity of computing the second order moment bound on the expected optimal value of a mixed integer linear program with a random objective coefficient vector is closely related to the complexity of characterizing the convex hull of the points $\{\binom{1}{\boldsymbol{x}}\binom{1}{\boldsymbol{x}}' \mid \boldsymbol{x} \in \mathcal{X}\}$ where $\mathcal{X}$ is the feasible region. In fact, we can replace the completely positive programming formulation for the moment bound on $\mathcal{X}$, with an associated semidefinite program, provided we have a linear or a semidefinite representation of this convex hull. As an application of the result, we identify a new polynomial time solvable semidefinite relaxation of the distributionally robust multi-item newsvendor problem by exploiting results from the Boolean quadric polytope. For $\mathcal{X}$ described explicitly by a finite set of points, our formulation leads to a reduction in the size of the semidefinite program. We illustrate the usefulness of the reduced semidefinite programming bounds in estimating the expected range of random variables with two applications arising in random walks and best-worst choice models.

Keywords Moment bounds · Newsvendor · Random walk · Choice model

The research of the first author was partly supported by the grant IDG31300105 on 'Optimization for Complex Discrete Choice' funded by the SUTD-MIT International Design Center and the MOE Tier 2 grant number MOE2013-T2-2-168 on Distributional Robust Optimization for Consumer Choice in Transportation Systems.

Karthik Natarajan
Engineering System and Design, Singapore University of Technology and Design, Singapore 138682.
E-mail: karthik_natarajan@sutd.edu.sg

Chung-Piaw Teo
Department of Decision Sciences, NUS Business School, Singapore 117591.
E-mail: bizteocp@nus.edu.sg

1 Introduction

Given a n-dimensional real valued random vector $\tilde{\boldsymbol{c}}$ and a set $\mathcal{X} \subset \Re_n$, define $Z(\tilde{\boldsymbol{c}})$ as a maximum of a set of linear functions of the random variables:

$$Z(\tilde{\boldsymbol{c}}) = \max \left\{ \tilde{\boldsymbol{c}}'\boldsymbol{x} \;\middle|\; \boldsymbol{x} \in \mathcal{X} \right\}. \tag{1}$$

We assume that the set $\mathcal{X}$ is specified by one of the following two representations:

(a) $\mathcal{X}$ is explicitly given as a finite set of points:

$$\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}, \tag{2}$$

(b) $\mathcal{X}$ is the bounded feasible region to a mixed integer linear program:

$$\mathcal{X} = \left\{ \boldsymbol{x} \in \Re_n \;\middle|\; \boldsymbol{x} \geq \boldsymbol{0},\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ x_i \in \mathcal{Z},\ \forall i \in \mathcal{B} \subseteq [n] \right\}. \tag{3}$$

Given a finite mean $\boldsymbol{\mu}$ and second moment matrix $\boldsymbol{\Pi}$ of a random vector $\tilde{\boldsymbol{c}}$ with support contained in $\Re_n$, we focus on the following moment problem:

$$Z^* = \sup \left\{ E_\theta \left(Z(\tilde{\boldsymbol{c}}) \right) \;\middle|\; E_\theta \left(\tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}' \right) = \boldsymbol{\Pi}, E_\theta \left(\tilde{\boldsymbol{c}} \right) = \boldsymbol{\mu}, E_\theta \left(1 \right) = 1, \theta \in \mathbb{M}(\Re_n) \right\}, \tag{4}$$

where $\mathbb{M}(\Re_n)$ is the set of finite positive Borel measures supported on $\Re_n$ and θ is a probability measure for the random vector satisfying the first and second moment conditions. The study of moment bounds on functions of random variables play an important role in diverse areas such as random walks (see Feller [19], Anis and Lloyd [4], Hurst [24]), option pricing (see Boyle and Lin [11], Bertsimas and Popescu [8]) and distributionally robust optimization (see Ben-Tal et. al. [7], Delage and Ye [14], Bertsimas et. al. [10], Zymler, Kuhn and Rustem [57], Natarajan, Teo and Zheng [40], Hanasusanto et. al. [22]). The typical approach to computing moment bounds is with semidefinite programming. In this paper, we develop a new semidefinite programming formulation to bound the expected value of the maximum of linear functions of random variables with second order moment information. The structure and contributions of the paper are as follows:

(a) In Section 2, we review the current semidefinite programs and completely positive programs used to find the moment bound.

(b) In Section 3, we provide a technique to extend a rank one factorization of a smaller matrix to a larger positive semidefinite matrix. Using this technique, we develop a new semidefinite programming formulation and show that the complexity of computing the moment bound is closely related to the complexity of characterizing the convex hull of the points $\{\binom{1}{\boldsymbol{x}}\binom{1}{\boldsymbol{x}}' \mid \boldsymbol{x} \in \mathcal{X}\}$. When the set $\mathcal{X}$ is given by a finite set of points, the new semidefinite program is significantly smaller than the current semidefinite programming formulation.

(c) In Section 4, we develop moment bounds for combinatorial optimization problems with random objective by using valid inequalities from the deterministic problem. We identify a new polynomial time solvable semidefinite relaxation of the distributionally robust multi-item newsvendor problem by exploiting results from the Boolean quadric polytope that strengthens the current results.

(d) In Section 5, we apply the results to the range of random variables. We show that size of the semidefinite program to find the maximum expected range of a set of random variables with given mean, variance and covariance information can be reduced from $O(n^4)$ variables to $O(n^2)$ variables. Furthermore, we provide applications of the bound to estimating the expected range of partial sums for a problem arising in random walks and to computing best-worst choice probabilities for a problem arising in discrete choice models before concluding in Section 6.

1.1 Terminology and Notation

We use standard letters such as x to denote scalars, bold letters such as $\boldsymbol{x}$ to denote vectors, bold capital letters such as $\boldsymbol{X}$ to denote matrices, tilde notation such as $\tilde{c}$ to denote random variables. The notation $[n]$ is used to denote an index set $\{1, \ldots, n\}$. $\Re_n$ denotes the n-dimensional real space and $\Re_n^+$ denotes the nonnegative orthant. $\mathcal{Z}$ denotes the set of integers. The transpose of a column vector $\boldsymbol{x} \in \Re_n$ is denoted by $\boldsymbol{x}'$. The trace of a matrix $\boldsymbol{X}$, denoted by trace($\boldsymbol{X}$), is sum of the diagonal entries of the matrix. For two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\Re_n$, the inner product is $\boldsymbol{x}'\boldsymbol{y} = x_1y_1 + \ldots + x_ny_n$, the element-wise product of the vectors is $\boldsymbol{x} \circ \boldsymbol{y} = (x_1y_1, \ldots, x_ny_n)'$ while the inner product between two matrices of the same dimension is denoted as $\boldsymbol{X} \cdot \boldsymbol{Y} = \text{trace}(\boldsymbol{X}'\boldsymbol{Y})$. The vector $\boldsymbol{e}$ consists of all components equal to one, the vector $\boldsymbol{e}_i$ is the vector with the ith component equal to 1 and 0 otherwise while the vector $\boldsymbol{e}_{i,j}$ is the vector with the ith and jth components equal to 1 and 0 otherwise. We use $0_{n,m}$ to denote a matrix of n by m zeros and suppress the dimension often to simplify notations. Diag($\boldsymbol{x}$) denotes a diagonal matrix with the vector $\boldsymbol{x}$ along the diagonal entries and diag($\boldsymbol{X}$) denotes a vector formed with the diagonal entries of $\boldsymbol{X}$. $\Re_{m\times n}$ is the set of $m \times n$ matrices, $\mathcal{S}_n$ is the set of $n \times n$ symmetric matrices, $\mathcal{S}_n^+$ is the set of $n \times n$ symmetric positive semidefinite matrices and $\mathcal{S}_n^{++}$ is the set of $n \times n$ symmetric positive definite matrices. We use $\boldsymbol{X} \succeq 0$ to represent $\boldsymbol{X} \in \mathcal{S}_n^+$ and $\boldsymbol{X} \succ 0$ to represent $\boldsymbol{X} \in \mathcal{S}_n^{++}$. For a closed convex cone $\mathcal{K} \subset \Re_n$, the generalized completely positive cone over $\mathcal{K}$ is defined as the set of $n \times n$ symmetric matrices that is representable as a sum of rank one matrices of the type:

$$\mathcal{C}(\mathcal{K}) = \left\{ \boldsymbol{X} \in \mathcal{S}_n \, : \, \exists \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p \in \mathcal{K} \text{ such that } \boldsymbol{X} = \sum_{k \in [p]} \boldsymbol{v}_k \boldsymbol{v}_k' \right\}.$$

For $\mathcal{K} = \Re_n^+$, $\mathcal{C}(\Re_n^+)$ is the cone of completely positive matrices. The convex hull of a set $\mathcal{X}$ is denoted as $\text{conv}(\mathcal{X})$.

2 Literature Review

In this section, we review the following two approaches to find second order moment bounds: (1) Semidefinite programming based approach, and (2) Completely positive programming based approach.

2.1 Semidefinite Programming Bounds

Semidefinite programming has emerged as a computational paradigm under which bounds on functions of random vectors are computed with moment information. In a series of papers over fifty years ago, Olkin and Pratt [42] and Marshall and Olkin [33,34] developed bounds on expected value of functions of random variables using mean, variance and covariance information. In these papers, the authors describe the possibility of using an optimization technique over the cone of positive semidefinite matrices. However no computational evidence was provided due to the unavailability of semidefinite programming solvers at that time. The book of Karlin and Studden [26] provides a detailed discussion on these bounds. With the development of interior point methods and convex optimization solvers, solving moment problems with semidefinite programming is however now possible (see Nesterov [41], Bertsimas and Popescu [9], Lassere [27]). We review the approach next.

Consider the moment problem in (4) with $\mathcal{X}$ described by a finite set of points:

$$Z^* = \sup \left\{ E_\theta \left(\max_{j \in [m]} \tilde{\boldsymbol{c}}' \boldsymbol{x}_j \right) \;\middle|\; E_\theta \left(\tilde{\boldsymbol{c}} \tilde{\boldsymbol{c}}' \right) = \boldsymbol{\Pi}, E_\theta \left(\tilde{\boldsymbol{c}} \right) = \boldsymbol{\mu}, E_\theta \left(1 \right) = 1, \theta \in \mathbb{M}(\Re_n) \right\}. \tag{5}$$

Define dual variables $\boldsymbol{Y} \in \mathcal{S}_n$, $\boldsymbol{y} \in \Re_n$ and $y_0 \in \Re$ for the moment constraints in Problem (5). The dual problem is formulated as:

$$\begin{aligned} Z_D^* = \inf\ & \boldsymbol{Y} \cdot \boldsymbol{\Pi} + \boldsymbol{y}' \boldsymbol{\mu} + y_0 \\ \text{s.t. } & \boldsymbol{c}' \boldsymbol{Y} \boldsymbol{c} + \boldsymbol{y}' \boldsymbol{c} + y_0 \geq \max_{j \in [m]} \boldsymbol{c}' \boldsymbol{x}_j, \ \forall \boldsymbol{c} \in \Re_n. \end{aligned}$$

Isii [25] showed that strong duality holds under the assumption that $\boldsymbol{\Pi} - \boldsymbol{\mu}\boldsymbol{\mu}' \succ 0$. Under this assumption, $Z^* = Z_D^*$ and the dual optimal solution is attained. Disaggregating the constraints in the dual and replacing the inf with min results in the equivalent semidefinite program:

$$\begin{aligned} Z_D^* = \min\ & \boldsymbol{Y} \cdot \boldsymbol{\Pi} + \boldsymbol{y}' \boldsymbol{\mu} + y_0 \\ \text{s.t. } & \begin{pmatrix} \boldsymbol{Y} & \dfrac{\boldsymbol{y} - \boldsymbol{x}_j}{2} \\ \dfrac{\boldsymbol{y}' - \boldsymbol{x}_j'}{2} & y_0 \end{pmatrix} \succeq 0, \ \forall j \in [m]. \end{aligned} \tag{6}$$

It is easy to verify that the dual formulation in (6) is strictly feasible by setting $\boldsymbol{y} = 0$, $Y_{ij} = 0$ for $i \neq j$ and choosing $Y_{ii} > \frac{1}{2}\max_j |\boldsymbol{e}_i'\boldsymbol{x}_j|$ and $y_0 > \frac{1}{2}\sum_i \max_j |\boldsymbol{e}_i'\boldsymbol{x}_j|$. This implies that the primal optimal solution is also attained. By a similar argument, we can ensure that for the moment problems considered throughout this paper, strictly feasible solutions exist for both the primal and dual problems which implies that strong duality holds and the primal and dual optimal solutions are attained. The dual of (6) provides the primal semidefinite programming formulation:

$$\begin{aligned} Z^* = \max \ & \sum_{j \in [m]} \boldsymbol{w}_j'\boldsymbol{x}_j \\ \text{s.t.} \ & \sum_{j \in [m]} \begin{pmatrix} \boldsymbol{W}_j & \boldsymbol{w}_j \\ \boldsymbol{w}_j' & \alpha_j \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{\mu} \\ \boldsymbol{\mu}' & 1 \end{pmatrix}, \\ & \begin{pmatrix} \boldsymbol{W}_j & \boldsymbol{w}_j \\ \boldsymbol{w}_j' & \alpha_j \end{pmatrix} \succeq 0, \qquad \forall j \in [m], \end{aligned} \tag{7}$$

where the decision variables are $\boldsymbol{W}_j \in \mathcal{S}_n$, $\boldsymbol{w}_j \in \Re_n$ and $\alpha_j \in \Re$ for $j \in [m]$. Mishra, Natarajan and Teo [37] constructed an extremal distribution from the optimal solution to (7) that attains the bound Z^*. At optimality, if this extremal distribution has a unique solution for each realization of the random vector, the decision variables in (7) can be interpreted as scaled conditional moments:

$$\begin{pmatrix} \boldsymbol{W}_j & \boldsymbol{w}_j \\ \boldsymbol{w}_j' & \alpha_j \end{pmatrix} = P\left(j\text{th point is max}\right) \begin{pmatrix} E\left(\tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}' \,\middle|\, j\text{th point is max}\right) & E\left(\tilde{\boldsymbol{c}} \,\middle|\, j\text{th point is max}\right) \\ E\left(\tilde{\boldsymbol{c}}' \,\middle|\, j\text{th point is max}\right) & 1 \end{pmatrix},$$

where α_j is the probability that the jth point attains the maximum value. More generally, α_j provides a lower bound on the probability that the jth point is maximum under the extremal distribution (see the discussion in Mishra, Natarajan and Teo [37]). The first set of feasibility constraints in (7) is obtained from expressing the probability, the mean and the second moment matrix as the sum of the scaled conditional moments. The positive semidefiniteness constraint is obtained from the feasibility condition on the scaled conditional moments. The primal semidefinite program has m positive semidefinite matrices of size $(n+1) \times (n+1)$ each and $O(n^2)$ moment equality constraints. When m and n are large, solving this semidefinite program is a computational challenge.

The dual formulation in (6) was developed by Boyle and Lin [11] in a finance setting. Their motivation was to find a distribution-free upper bound on the price of an European call option with payoff defined by the maximum of a set of asset prices. The primal semidefinite program in (7) was developed by Mishra, Natarajan and Teo [37] in the area of discrete choice models to find the maximum expected utility given the mean and the covariance matrix of the utilities. Bertsimas and Popescu [9,8] and Lasserre [27] have developed a general semidefinite programming approach to solve a wide range of moment problems arising from applications in probability and finance. Delage and Ye [14], Bertsimas et. al. [10], Zymler, Kuhn and Rustem [57] and Hanasusanto et. al. [22] have applied these semidefinite programs to solve distributionally robust optimization problems with second order moment information. A closely related work in this area is the tight lower bound developed by Vandenberghe, Boyd and Comanor [54] for the probability of a random vector lying in a set defined by quadratic inequalities, given the first two moments of the distribution.

2.2 Completely Positive Programming Bounds

Natarajan, Teo and Zheng [40] recently computed second order moment bounds using the theory of completely positive matrices. We present their main result in this section. Assume that $Z(\tilde{\boldsymbol{c}})$ is the optimal objective value of a maximization problem with the feasible region defined by the intersection of the nonnegative orthant with a set of equality constraints with some of the variables possibly being binary:

$$Z(\tilde{\boldsymbol{c}}) = \max \left\{ \tilde{\boldsymbol{c}}'\boldsymbol{x} \;\middle|\; \boldsymbol{x} \in \Re_n^+, \boldsymbol{a}_k'\boldsymbol{x} = b_k,\ \forall k \in [p],\ x_i \in \{0,1\},\ \forall i \in \mathcal{B} \subseteq [n] \right\}. \tag{8}$$

This includes 0-1 integer programs and linear programs as special cases. To find the tightest upper bound on the expected optimal value, they defined $\boldsymbol{x}(\boldsymbol{c})$ as the optimal value of the decision vector $\boldsymbol{x}$ to Problem (8) for a fixed $\boldsymbol{c}$. When $\tilde{\boldsymbol{c}}$ is random, $\boldsymbol{x}(\tilde{\boldsymbol{c}})$ is also random. For the objective coefficient vectors with multiple optimal solutions, $\boldsymbol{x}(\boldsymbol{c})$ is chosen arbitrarily from one of the optimal solutions, without affecting the objective value. The set of decision variables is defined as:

$$\begin{aligned} \boldsymbol{x} &= E[\boldsymbol{x}(\tilde{\boldsymbol{c}})], \\ \boldsymbol{X} &= E[\boldsymbol{x}(\tilde{\boldsymbol{c}})\boldsymbol{x}(\tilde{\boldsymbol{c}})'], \\ \boldsymbol{Y} &= E[\boldsymbol{x}(\tilde{\boldsymbol{c}})\tilde{\boldsymbol{c}}'], \end{aligned} \tag{9}$$

where $\boldsymbol{x} \in \Re_n$, $\boldsymbol{X} \in \mathcal{S}_n$ and $\boldsymbol{Y} \in \Re_{n \times n}$. Assuming that the feasible region is bounded with the following additional minor assumption:

$$\boldsymbol{x} \in \Re_n^+ \text{ such that } \boldsymbol{a}_k'\boldsymbol{x} = b_k,\ \forall k \in [p] \Longrightarrow x_i \leq 1, \forall i \in \mathcal{B},$$

Natarajan, Teo and Zheng [40] showed that the tight bound is computable by solving an optimization problem over the generalized cone of completely positive matrices:

$$\begin{aligned} Z^* = \max\ & \text{trace}(\boldsymbol{Y}) \\ \text{s.t. } & \boldsymbol{a}_k'\boldsymbol{x} = b_k, && \forall k \in [p], \\ & \boldsymbol{a}_k'\boldsymbol{X}\boldsymbol{a}_k = b_k^2, && \forall k \in [p], \\ & X_{ii} = x_i, && \forall i \in \mathcal{B} \subseteq [n], \\ & \begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{\mu} & \boldsymbol{Y}' \\ \boldsymbol{\mu}' & 1 & \boldsymbol{x}' \\ \boldsymbol{Y} & \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in \mathcal{C}(\Re_n \times \Re_{1+n}^+). \end{aligned} \tag{10}$$

The tightness of the formulation in (10) was shown using results from Burer [12]. Problem (10) is however NP-hard to solve. The difficulty arises due to the complexity of characterizing the cone $\mathcal{C}(\Re_n \times \Re_{1+n}^+)$ and it's dual cone, both of which are intractable for general n (see Murty and Kabadi [38], Dickinson and Gibjen [17]). A simple relaxation is to use the cone of doubly nonnegative matrices as follows:

$$\begin{aligned} Z_{DNN}^* = \max\ & \text{trace}(\boldsymbol{Y}) \\ \text{s.t. } & \boldsymbol{a}_k'\boldsymbol{x} = b_k, && \forall k \in [p], \\ & \boldsymbol{a}_k'\boldsymbol{X}\boldsymbol{a}_k = b_k^2, && \forall k \in [p], \\ & X_{ii} = x_i, && \forall i \in \mathcal{B} \subseteq [n], \\ & \begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{\mu} & \boldsymbol{Y}' \\ \boldsymbol{\mu}' & 1 & \boldsymbol{x}' \\ \boldsymbol{Y} & \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \succeq 0, \\ & \boldsymbol{x}, \boldsymbol{X} \geq 0. \end{aligned} \tag{11}$$

Clearly, every completely positive matrix is doubly nonnegative, i.e., it is both positive semidefinite and nonnegative. The reverse however need not be true. Hence, $Z^* \leq Z_{DNN}^*$. While the formulation in (11) is computationally more tractable, it is in general not tight.

It is useful to observe that the tightest lower bound on the expected maximum of linear functions of random variables with a given mean and covariance reduces to Jensen's lower bound $Z(\boldsymbol{\mu}) = \max\{\boldsymbol{\mu}'\boldsymbol{x} \mid \boldsymbol{x} \in \mathcal{X}\}$ where the function is evaluated at the mean $\boldsymbol{\mu}$. The validity of the lower bound is due to the convexity of the objective function. Tightness of this bound was shown in Bertsimas et. al. [10] by constructing the following random vector:

$$\tilde{\boldsymbol{c}} = \boldsymbol{\mu} + \frac{\tilde{b}\tilde{\boldsymbol{z}}}{\sqrt{\epsilon}},$$

where $\tilde{b}$ is a Bernoulli random variable with distribution:

$$\tilde{b} = \begin{cases} 0, \text{ with probability } 1 - \epsilon, \\ 1, \text{ with probability } \epsilon, \end{cases}$$

and $\tilde{\boldsymbol{z}}$ is a multivariate normal random vector that is generated independently of $\tilde{b}$ with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Pi} - \boldsymbol{\mu}\boldsymbol{\mu}'$. Then it is easy to verify that $E\left(\tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}'\right) = \boldsymbol{\Pi}$ and $E\left(\tilde{\boldsymbol{c}}\right) = \boldsymbol{\mu}$ for all $\epsilon > 0$. Taking the limit as $\epsilon \downarrow 0$, one can check that the Jensen bound is tight in this case.

3 Main Results

3.1 Extending a Rank One Factorization

In this section, we derive a factorization of a larger positive semidefinite matrix of size $(n_1 + n_2) \times (n_1 + n_2)$ by using an explicit rank one factorization of a smaller principal submatrix of size $n_2 \times n_2$. Salce and Zanardo [51] derive completely positive factorizations of matrices of size $(1+n)\times(1+n)$ using a completely positive factorization of a principal submatrix of size $n\times n$. We provide an alternate factorization approach which is the key step to proving the tightness of the new semidefinite programs for moment bounds. To develop the factorization, we start by recollecting the definition of the Moore-Penrose pseudoinverse of a matrix.

Definition 1 [Rao and Mishra [49], Penrose [47]] Let $\boldsymbol{X}$ be a matrix of dimension $n_1 \times n_2$. The Moore-Penrose pseudoinverse of $\boldsymbol{X}$ is a matrix $\boldsymbol{X}^\dagger$ of dimension $n_2 \times n_1$ and is defined as the unique solution to the set of four equations:

(1) $\boldsymbol{X}\boldsymbol{X}^\dagger\boldsymbol{X} = \boldsymbol{X}$,

(2) $\boldsymbol{X}^\dagger\boldsymbol{X}\boldsymbol{X}^\dagger = \boldsymbol{X}^\dagger$,

(3) $\boldsymbol{X}\boldsymbol{X}^\dagger = \left(\boldsymbol{X}\boldsymbol{X}^\dagger\right)'$,

(4) $\boldsymbol{X}^\dagger\boldsymbol{X} = \left(\boldsymbol{X}^\dagger\boldsymbol{X}\right)'$.

The pseudoinverse in Definition 1 also satisfies the following properties (see [49,47]):

(5) $(\boldsymbol{X}^\dagger)' = (\boldsymbol{X}')^\dagger$,

(6) $\boldsymbol{X} = (\boldsymbol{X}^\dagger)'\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}\boldsymbol{X}'(\boldsymbol{X}^\dagger)'$,

(7) $\boldsymbol{X}^\dagger = \boldsymbol{X}^{-1}$ if $\boldsymbol{X}$ is square and nonsingular.

In addition, we need the following two lemmas. The first lemma characterizes the positive semidefiniteness of a block matrix using the Schur's complement.

Lemma 1 (Albert [2]) *Let $\boldsymbol{T}$ be a $(n_1+n_2)\times(n_1+n_2)$ positive semidefinite block matrix of the form:*

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}' \\ \boldsymbol{B} & \boldsymbol{C} \end{pmatrix} \succeq 0,$$

where $\boldsymbol{A} \in \mathcal{S}_{n_1}$ and $\boldsymbol{C} \in \mathcal{S}_{n_2}$ are symmetric matrices. Then,

$$\boldsymbol{T} \succeq 0 \textit{ if and only if } \boldsymbol{C} \succeq 0,\ \boldsymbol{C}\boldsymbol{C}^\dagger\boldsymbol{B} = \boldsymbol{B},\ \boldsymbol{A} - \boldsymbol{B}'\boldsymbol{C}^\dagger\boldsymbol{B} \succeq 0.$$

The second lemma characterizes the Moore-Penrose pseudoinverse of a matrix in terms of a rank one factorization.

Lemma 2 (Salce and Zanardo [51]) *Let $\boldsymbol{C}$ be a real matrix such that $\boldsymbol{C} = \boldsymbol{V}\boldsymbol{V}'$ for some matrix $\boldsymbol{V}$. Then $\boldsymbol{C}^\dagger = (\boldsymbol{V}')^\dagger\boldsymbol{V}^\dagger$.*

This brings us to the following key result.

Theorem 1 *Assume that the following two conditions hold:*

(a) The matrix $\boldsymbol{T}$ *is a* $(n_1+n_2)\times(n_1+n_2)$ *positive semidefinite block matrix of the form:*

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}' \\ \boldsymbol{B} & \boldsymbol{C} \end{pmatrix} \succeq 0,$$

where $\boldsymbol{A} \in \mathcal{S}_{n_1}$ *and* $\boldsymbol{C} \in \mathcal{S}_{n_2}$ *are symmetric matrices.*

(b) The matrix $\boldsymbol{C}$ *lies in* $\mathcal{C}(\mathcal{K})$ *with an explicit factorization given by* $\boldsymbol{C} = \boldsymbol{V}\boldsymbol{V}'$ *where each column vector of* $\boldsymbol{V}$ *lies in the closed convex cone* $\mathcal{K}$.

Then $\boldsymbol{T}$ *lies in the generalized completely positive cone* $\mathcal{C}(\Re_{n_1} \times \mathcal{K})$ *with a factorization:*

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{B}'(\boldsymbol{V}^{\dagger})' \\ \boldsymbol{V} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}'(\boldsymbol{V}^{\dagger})' \\ \boldsymbol{V} \end{pmatrix}' + \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{0} \end{pmatrix}',$$

where the matrix $\boldsymbol{U}$ *is defined such that* $\boldsymbol{A} - \boldsymbol{B}'\boldsymbol{C}^{\dagger}\boldsymbol{B} = \boldsymbol{U}\boldsymbol{U}' \succeq 0$.

Proof: Assumption (b) and Lemma 2 implies that the pseudoinverse of the matrix $\boldsymbol{C}$ is $\boldsymbol{C}^{\dagger} = (\boldsymbol{V}')^{\dagger}\boldsymbol{V}^{\dagger}$. Assumption (a) and Lemma 1 implies that:

$$\boldsymbol{C} \succeq 0,\ \boldsymbol{C}\boldsymbol{C}^{\dagger}\boldsymbol{B} = \boldsymbol{B},\ \boldsymbol{\Delta} = \boldsymbol{A} - \boldsymbol{B}'\boldsymbol{C}^{\dagger}\boldsymbol{B} \succeq 0. \tag{12}$$

Taken together, this implies the following set of equalities:

$$\begin{aligned} \boldsymbol{B} &= \boldsymbol{C}\boldsymbol{C}^{\dagger}\boldsymbol{B}, && \text{[From the equality condition in (12)]} \\ &= \underbrace{\boldsymbol{V}\boldsymbol{V}'(\boldsymbol{V}')^{\dagger}}\boldsymbol{V}^{\dagger}\boldsymbol{B}, && \text{[Substituting } \boldsymbol{C} = \boldsymbol{V}\boldsymbol{V}' \text{ and } \boldsymbol{C}^{\dagger} = (\boldsymbol{V}')^{\dagger}\boldsymbol{V}^{\dagger}] \\ &= \boldsymbol{V}\boldsymbol{V}^{\dagger}\boldsymbol{B}, && \text{[Since } \boldsymbol{V}\boldsymbol{V}'(\boldsymbol{V}')^{\dagger} = \boldsymbol{V} \text{ from property (6)]}. \end{aligned}$$

Furthermore, $\boldsymbol{\Delta} = \boldsymbol{A} - \boldsymbol{B}'\boldsymbol{C}^{\dagger}\boldsymbol{B}$ is positive semidefinite from (12). This implies that there exists a decomposition of $\boldsymbol{\Delta}$ given by $\boldsymbol{U}$ with $\boldsymbol{\Delta} = \boldsymbol{U}\boldsymbol{U}' \succeq 0$. Hence, we have:

$$\begin{aligned} \begin{pmatrix} \boldsymbol{B}'(\boldsymbol{V}^{\dagger})' \\ \boldsymbol{V} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}'(\boldsymbol{V}^{\dagger})' \\ \boldsymbol{V} \end{pmatrix}' + \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{0} \end{pmatrix}' &= \begin{pmatrix} \boldsymbol{B}'(\boldsymbol{V}^{\dagger})'\boldsymbol{V}^{\dagger}\boldsymbol{B} + \boldsymbol{\Delta} & \boldsymbol{B}'(\boldsymbol{V}^{\dagger})'\boldsymbol{V}' \\ \boldsymbol{V}\boldsymbol{V}^{\dagger}\boldsymbol{B} & \boldsymbol{V}\boldsymbol{V}' \end{pmatrix}, \\ &= \begin{pmatrix} \boldsymbol{B}'\boldsymbol{C}^{\dagger}\boldsymbol{B} + \boldsymbol{\Delta} & \boldsymbol{B}' \\ \boldsymbol{B} & \boldsymbol{C} \end{pmatrix}, \\ &= \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}' \\ \boldsymbol{B} & \boldsymbol{C} \end{pmatrix}. \end{aligned}$$

□

This result has been independently shown by Dickinson [16] (see Section 2.5, Page 32) who developed a characterization of the generalized completely positive cone $\mathcal{C}(\Re_{n_1} \times \mathcal{K})$ using a positive semidefinite cone and a smaller generalized completely positive cone. His result is obtained by a characterization of the generalized copositive cone and then taking the dual of the cone while our result is through a characterization of the generalized completely positive cone.

3.2 A New Semidefinite Program

In this section, we use the factorization in Theorem 1 to develop a new semidefinite programming formulation to compute the second order moment bound. Our key result is provided next.

Theorem 2 *Define:*

$$Z^* = \sup \left\{ E_\theta \left(\max_{\boldsymbol{x} \in \mathcal{X}} \tilde{\boldsymbol{c}}' \boldsymbol{x} \right) \;\middle|\; E_\theta \left(\tilde{\boldsymbol{c}} \tilde{\boldsymbol{c}}' \right) = \boldsymbol{\Pi}, E_\theta \left(\tilde{\boldsymbol{c}} \right) = \boldsymbol{\mu}, E_\theta (1) = 1, \theta \in \mathbb{M}(\Re_n) \right\}.$$

where $\mathcal{X}$ is specified in representation (2) or (3). Let $\hat{Z}^*$ be the optimal objective value to the semidefinite program:

$$\begin{aligned} \hat{Z}^* = \max\ & trace(\boldsymbol{Y}) \\ s.t.\ & \begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{\mu} & \boldsymbol{Y}' \\ \boldsymbol{\mu}' & 1 & \boldsymbol{x}' \\ \boldsymbol{Y} & \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \succeq 0, \\ & \begin{pmatrix} 1 & \boldsymbol{x}' \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in conv \left\{ \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix}' \;\middle|\; \boldsymbol{z} \in \mathcal{X} \right\}. \end{aligned} \tag{13}$$

Then, $\hat{Z}^* = Z^*$.

Proof:
Step 1: To show $Z^* \leq \hat{Z}^*$.
Consider a random vector $\tilde{\boldsymbol{c}}$ with a distribution θ that is feasible with the given first and second moment conditions. The rank one matrix of size $(2n+1) \times (2n+1)$ defined in terms of the random objective vector $\tilde{\boldsymbol{c}}$ and the random optimal decision vector $\boldsymbol{x}(\tilde{\boldsymbol{c}})$ satisfies the following condition:

$$\begin{pmatrix} \tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}' & \tilde{\boldsymbol{c}} & \tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})' \\ \tilde{\boldsymbol{c}}' & 1 & \boldsymbol{x}(\tilde{\boldsymbol{c}})' \\ \boldsymbol{x}(\tilde{\boldsymbol{c}})\tilde{\boldsymbol{c}}' & \boldsymbol{x}(\tilde{\boldsymbol{c}}) & \boldsymbol{x}(\tilde{\boldsymbol{c}})\boldsymbol{x}(\tilde{\boldsymbol{c}})' \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{c}} \\ 1 \\ \boldsymbol{x}(\tilde{\boldsymbol{c}}) \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{c}} \\ 1 \\ \boldsymbol{x}(\tilde{\boldsymbol{c}}) \end{pmatrix}' \succeq 0.$$

Taking expectations and using the decision variables as defined in (9), we obtain the positive semidefinite constraint:

$$\begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{\mu} & \boldsymbol{Y}' \\ \boldsymbol{\mu}' & 1 & \boldsymbol{x}' \\ \boldsymbol{Y} & \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} = E_\theta \left[\begin{pmatrix} \tilde{\boldsymbol{c}} \\ 1 \\ \boldsymbol{x}(\tilde{\boldsymbol{c}}) \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{c}} \\ 1 \\ \boldsymbol{x}(\tilde{\boldsymbol{c}}) \end{pmatrix}' \right] \succeq 0.$$

Furthermore, feasibility of the optimal decision vector implies that:

$$\begin{pmatrix} 1 \\ \boldsymbol{x}(\tilde{\boldsymbol{c}}) \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{x}(\tilde{\boldsymbol{c}}) \end{pmatrix}' \in \left\{ \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix}' \;\middle|\; \boldsymbol{z} \in \mathcal{X} \right\}.$$

Taking expectations, implies that the $(n+1)\times(n+1)$ matrix satisfies the following constraint:

$$\begin{pmatrix} 1 & \boldsymbol{x}' \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} = E_\theta\left[\begin{pmatrix} 1 \\ \boldsymbol{x}(\tilde{\boldsymbol{c}}) \end{pmatrix}\begin{pmatrix} 1 \\ \boldsymbol{x}(\tilde{\boldsymbol{c}}) \end{pmatrix}'\right] \in \text{conv}\left\{\begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix}\begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix}' \;\middle|\; \boldsymbol{z}\in\mathcal{X}\right\}.$$

Since $E_\theta(\tilde{\boldsymbol{c}}'\boldsymbol{x}(\boldsymbol{c})) = \text{trace}(\boldsymbol{Y})$, it follows that $Z^* \le \hat{Z}^*$.
Step 2: To show $\hat{Z}^* \le Z^*$.
Consider an optimal solution to the semidefinite program (13) given by $(\boldsymbol{x}^*, \boldsymbol{X}^*, \boldsymbol{Y}^*)$. Note that the feasible region to the semidefinite program is bounded and the optimal solution is attained. Applying Carathéodory's theorem, the convex hull constraint implies that the $(n+1)\times(n+1)$ principal submatrix has a factorization given by:

$$\begin{pmatrix} 1 & \boldsymbol{x}^{*\prime} \\ \boldsymbol{x}^* & \boldsymbol{X}^* \end{pmatrix} = \sum_{\boldsymbol{x}\in\hat{\mathcal{X}}} \alpha^*_{\boldsymbol{x}} \begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix}\begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix}',$$

where $\hat{\mathcal{X}}$ is a subset of the original set $\mathcal{X}$ consisting of $(n+1)(n+2)/2$ or fewer points with $\alpha^*_{\boldsymbol{x}} > 0$ for all $\boldsymbol{x}\in\hat{\mathcal{X}}$. Theorem 1 implies that there exists a decomposition of the $(2n+1)\times(2n+1)$ matrix that extends the convex hull decomposition of the principal submatrix of size $(n+1)\times(n+1)$ as follows:

$$\begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{\mu} & \boldsymbol{Y}^{*\prime} \\ \boldsymbol{\mu}' & 1 & \boldsymbol{x}^{*\prime} \\ \boldsymbol{Y}^* & \boldsymbol{x}^* & \boldsymbol{X}^* \end{pmatrix} = \sum_{\boldsymbol{x}\in\hat{\mathcal{X}}} \alpha^*_{\boldsymbol{x}} \begin{pmatrix} \boldsymbol{d}^*_{\boldsymbol{x}} \\ 1 \\ \boldsymbol{x} \end{pmatrix}\begin{pmatrix} \boldsymbol{d}^*_{\boldsymbol{x}} \\ 1 \\ \boldsymbol{x} \end{pmatrix}' + \begin{pmatrix} \boldsymbol{\Delta}^* & \boldsymbol{0}_{n,1} & \boldsymbol{0}_{n,n} \\ \boldsymbol{0}_{1,n} & 0 & \boldsymbol{0}_{1,n} \\ \boldsymbol{0}_{n,n} & \boldsymbol{0}_{n,1} & \boldsymbol{0}_{n,n} \end{pmatrix},$$

where $\boldsymbol{d}^*_{\boldsymbol{x}} \in \Re_n$, $\alpha^*_{\boldsymbol{x}} > 0$ and $\boldsymbol{\Delta}^* \in \mathcal{S}^+_n$. Generate a probability distribution θ^* for the random vector $\tilde{\boldsymbol{c}}$ as follows:

(a) Choose a point $\boldsymbol{x}\in\hat{\mathcal{X}}\subseteq\mathcal{X}$ with probability $\alpha^*_{\boldsymbol{x}}$.

(b) Generate a normally distributed random vector $\tilde{\boldsymbol{d}}_{\boldsymbol{x}}$ in $\Re_n$ with mean vector $\boldsymbol{d}^*_{\boldsymbol{x}}$ and second moment matrix $\boldsymbol{d}^*_{\boldsymbol{x}}\boldsymbol{d}^{*\prime}_{\boldsymbol{x}} + \boldsymbol{\Delta}^*$ (since $\boldsymbol{\Delta}^* \succeq 0$). Set $\tilde{\boldsymbol{c}} = \tilde{\boldsymbol{d}}_{\boldsymbol{x}}$.

Under the distribution θ^*, the first two moments of the random vector satisfy:

$$\begin{aligned} E_{\theta^*}\left[\begin{pmatrix} \tilde{\boldsymbol{c}} \\ 1 \end{pmatrix}\begin{pmatrix} \tilde{\boldsymbol{c}} \\ 1 \end{pmatrix}'\right] &= \sum_{\boldsymbol{x}\in\hat{\mathcal{X}}} \alpha^*_{\boldsymbol{x}} \begin{pmatrix} \boldsymbol{d}^*_{\boldsymbol{x}}\boldsymbol{d}^{*\prime}_{\boldsymbol{x}} + \boldsymbol{\Delta}^* & \boldsymbol{d}^*_{\boldsymbol{x}} \\ \boldsymbol{d}^{*\prime}_{\boldsymbol{x}} & 1 \end{pmatrix}, \\ &= \begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{\mu} \\ \boldsymbol{\mu}' & 1 \end{pmatrix}. \end{aligned}$$

Lastly, evaluating the expected maximum value of the linear functions of the random variables leads to the following set of inequalities:

$$\begin{aligned}
Z^* &\geq E_{\theta^*}\left[\max_{\boldsymbol{x}\in\hat{\mathcal{X}}} \tilde{\boldsymbol{c}}'\boldsymbol{x}\right], && \text{[Since } \theta^* \text{ is a feasible distribution]}\\
&= \sum_{\boldsymbol{x}\in\hat{\mathcal{X}}} \alpha^*_{\boldsymbol{x}} E_{\theta^*}\left[\max_{\boldsymbol{y}\in\hat{\mathcal{X}}} \tilde{\boldsymbol{d}}_{\boldsymbol{x}}'\boldsymbol{y}\right], && \text{[Using conditional expectations]}\\
&\geq \sum_{\boldsymbol{x}\in\hat{\mathcal{X}}} \alpha^*_{\boldsymbol{x}} E_{\theta^*}\left[\tilde{\boldsymbol{d}}_{\boldsymbol{x}}'\boldsymbol{x}\right], && \text{[Using the } \boldsymbol{x} \text{ solution]}\\
&= \sum_{\boldsymbol{x}\in\hat{\mathcal{X}}} \alpha^*_{\boldsymbol{x}} \boldsymbol{d}^{*\prime}_{\boldsymbol{x}}\boldsymbol{x},\\
&= \text{trace}(\boldsymbol{Y}^*),\\
&= \hat{Z}^*.
\end{aligned}$$

Taken together, this implies that $Z^* = \hat{Z}^*$. □

3.3 Reduced Formulation

When the feasible region is given as a set of points $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$ with $Z(\tilde{\boldsymbol{c}}) = \max\left\{\tilde{\boldsymbol{c}}'\boldsymbol{x}_j \;\middle|\; j \in [m]\right\}$ the semidefinite program in Theorem 2 is given as:

$$\begin{aligned}
Z^* = \max\ & \text{trace}(\boldsymbol{Y})\\
\text{s.t.}\ & \begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{\mu} & \boldsymbol{Y}' \\ \boldsymbol{\mu}' & 1 & \sum_{j\in[m]} \alpha_j \boldsymbol{x}_j' \\ \boldsymbol{Y} & \sum_{j\in[m]} \alpha_j \boldsymbol{x}_j & \sum_{j\in[m]} \alpha_j \boldsymbol{x}_j \boldsymbol{x}_j' \end{pmatrix} \succeq 0,\\
& \sum_{j\in[m]} \alpha_j = 1,\\
& \alpha_j \geq 0, \qquad \forall j \in [m].
\end{aligned} \tag{14}$$

Let $(\boldsymbol{Y}^*, \boldsymbol{\alpha}^*)$ be an optimal solution to the semidefinite program (14). Define the matrix $\boldsymbol{V}$ of size $(n+1)\times m$ as:

$$\boldsymbol{V} = \begin{pmatrix} \sqrt{\alpha_1^*} & \ldots & \sqrt{\alpha_m^*} \\ \sqrt{\alpha_1^*}\boldsymbol{x}_1 & \ldots & \sqrt{\alpha_m^*}\boldsymbol{x}_m \end{pmatrix},$$

and the matrix $\boldsymbol{\Delta}^*$ of size $n \times n$ as:

$$\boldsymbol{\Delta}^* = \boldsymbol{\Pi} - \begin{pmatrix} \boldsymbol{\mu} & \boldsymbol{Y}^{*\prime} \end{pmatrix} \begin{pmatrix} 1 & \sum_{j\in[m]} \alpha_j^* \boldsymbol{x}_j' \\ \sum_{j\in[m]} \alpha_j^* \boldsymbol{x}_j & \sum_{j\in[m]} \alpha_j^* \boldsymbol{x}_j \boldsymbol{x}_j' \end{pmatrix}^{\dagger} \begin{pmatrix} \boldsymbol{\mu} & \boldsymbol{Y}^{*\prime} \end{pmatrix}'.$$

The proof of Theorem 2 implies that the extremal distribution θ^* for the random vector $\tilde{\boldsymbol{c}}$ that attains the moment bound Z^* is constructed as:

(a) Choose a point $\boldsymbol{x}_j$ for $j \in [m]$ with probability α_j^*.

(b) Generate a normally distributed random vector $\tilde{\boldsymbol{d}}_j$ in $\Re_n$ with mean vector $\boldsymbol{d}_j^*$ set to be the jth column vector of the matrix $(\boldsymbol{\mu}\ \boldsymbol{Y}^{*\prime})(\boldsymbol{V}^{\dagger})'$ and second moment matrix $\boldsymbol{d}_j^* \boldsymbol{d}_j^{*'} + \boldsymbol{\Delta}^*$. Set $\tilde{\boldsymbol{c}} = \tilde{\boldsymbol{d}}_j$.

At optimality if the solutions are unique for all realizations, P_{θ^*} (jth point is max) $=$ α_j^* else it provides a lower bound with P_{θ^*} (jth point is max) $\geq \alpha_j^*$. This formulation has 1 positive semidefinite matrix variable of size $(2n+1) \times (2n+1)$ with m additional nonnegative variables and $O(n^2)$ moment equality constraints. On the other hand, the semidefinite program (7) involves optimization over m positive semidefinite matrices, each of dimension $(n+1) \times (n+1)$ and $O(n^2)$ moment equality constraints. When m is large, the size of the semidefinite program (14) is much smaller than the size of the semidefinite program (7).

To compare the running time to solve the semidefinite programs, we consider a simple numerical example where $\mathcal{X} = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ with $Z(\tilde{\boldsymbol{c}}) = \max\{\tilde{c}_i \mid i \in [n]\}$. In this instance, $m = n$. The numerical experiments were implemented in Matlab R2014b on an Intel Core i7-5600U CPU, 2.6 GHz laptop with 4 GB of RAM. The semidefinite programs were solved with SDPT3-4.0 [52, 53] using the YALMIP interface [30]. In the numerical experiments, for each size from $n = 2$ to $n = 23$, we assume that $\tilde{c}_1, \ldots, \tilde{c}_n$ is a set of n correlated random variables with means generated randomly in the range $[0, 1]$ and standard deviations generated randomly in the range $[0, 1]$. The correlation matrices are generated randomly with eigenvalues chosen from a uniform distribution using the MATLAB command *gallery('randcorr',n)*. For each size n, we solve five random instances of the problem and compare the average running time to solve the semidefinite programs. The results are provided in Figure 1. As is clear from the figure, the time to solve the reduced semidefinite programs in (14) is significantly smaller than the time to solve the semidefinite programs (7).

3.4 Alternate Formulation

We start by reformulating the moment bound in Theorem 2 into a nonlinear but concave trace maximization problem over a smaller set of variables. This approach has been used recently by Ahipasaoglu, Li and Natarajan [1] to develop first order algorithms to compute moment bounds when $\mathcal{X}$ is a simplex. We provide an extension of their formulation to more general sets $\mathcal{X}$. Using Schur's complement in formulation (13), we obtain an equivalent reformulation

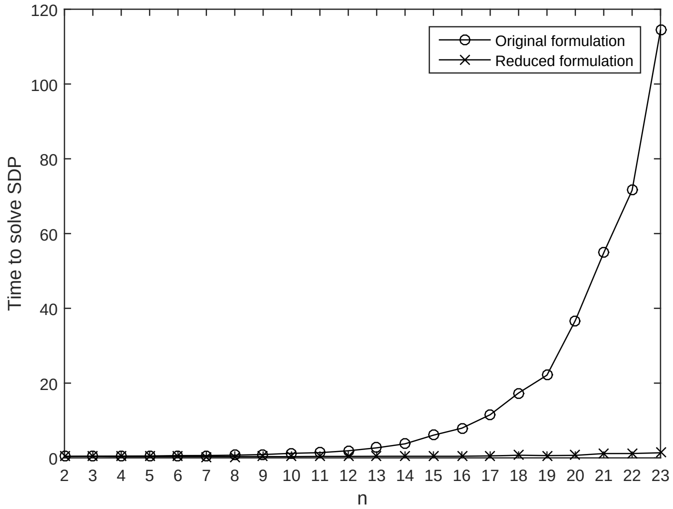

Fig. 1 Average time to solve the SDP across five instances using the original formulation (7) and reduced formulation (14) for $Z(\tilde{\boldsymbol{c}}) = \max\{\tilde{c}_i \mid i \in [n]\}$.

of the semidefinite program as follows:

$$\begin{aligned} Z^* = \max\ & \text{trace}(\boldsymbol{Y}) \\ \text{s.t.}\ & \begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{Y}' \\ \boldsymbol{Y} & \boldsymbol{X} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{x} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{x} \end{pmatrix}' \succeq 0, \\ & \begin{pmatrix} 1 & \boldsymbol{x}' \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in \text{conv}\left\{ \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix}' \;\middle|\; \boldsymbol{z} \in \mathcal{X} \right\}. \end{aligned}$$

Let $\boldsymbol{\Sigma} = \boldsymbol{\Pi} - \boldsymbol{\mu}\boldsymbol{\mu}'$ and define $\hat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{x}\boldsymbol{\mu}'$. Then, we can reformulate the semidefinite program as:

$$\begin{aligned} Z^* = \max\ & \text{trace}(\hat{\boldsymbol{Y}}) + \boldsymbol{\mu}'\boldsymbol{x} \\ \text{s.t.}\ & \begin{pmatrix} \boldsymbol{\Sigma} & \hat{\boldsymbol{Y}}' \\ \hat{\boldsymbol{Y}} & \boldsymbol{X} - \boldsymbol{x}\boldsymbol{x}' \end{pmatrix} \succeq 0, \\ & \begin{pmatrix} 1 & \boldsymbol{x}' \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in \text{conv}\left\{ \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix}' \;\middle|\; \boldsymbol{z} \in \mathcal{X} \right\}. \end{aligned}$$

The following result then helps to reformulate the problem.

Proposition 1 (Shapiro [50]) *The problem*

$$\max\left\{ \mathit{trace}(\boldsymbol{X}_{12}) \;\middle|\; \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{X}_{12} \\ \boldsymbol{X}_{12}' & \boldsymbol{S}_{22} \end{pmatrix} \succeq 0 \right\},$$

has an optimal solution $\boldsymbol{X}_{12} = \boldsymbol{S}_{11}\boldsymbol{S}_{22}^{\frac{1}{2}}\big(\boldsymbol{S}_{22}^{\frac{1}{2}}\boldsymbol{S}_{11}\boldsymbol{S}_{22}^{\frac{1}{2}}\big)^{-\frac{1}{2}}\boldsymbol{S}_{22}^{\frac{1}{2}}$ *with* $\mathit{trace}(\boldsymbol{X}_{12}) = \mathit{trace}\big(\boldsymbol{S}_{22}^{\frac{1}{2}}\boldsymbol{S}_{11}\boldsymbol{S}_{22}^{\frac{1}{2}}\big)^{\frac{1}{2}}$.

Applying this, we obtain the following result.

Corollary 1 *The formulation in Theorem 2 can be equivalently reformulated as:*

$$\begin{aligned} Z^* = \max\ & trace\left\{ \left(\boldsymbol{\Sigma}^{\frac{1}{2}} \left(\boldsymbol{X} - \boldsymbol{x}\boldsymbol{x}' \right) \boldsymbol{\Sigma}^{\frac{1}{2}} \right)^{\frac{1}{2}} \right\} + \boldsymbol{\mu}'\boldsymbol{x} \\ s.t.\ & \begin{pmatrix} 1 & \boldsymbol{x}' \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in conv \left\{ \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix}' \ \middle|\ \boldsymbol{z} \in \mathcal{X} \right\}. \end{aligned}$$

This formulation uses a smaller number of variables to transform our problem to a concave maximization problem. Since the square root function is operator concave, we can use the following trace inequality to derive upper bounds for our problem.

Lemma 3 (Peierls Inequality (Carlen [13])) *Let $\boldsymbol{A}$ be a positive semidefinite matrix, and $f(\cdot)$ be any concave function on $\mathbb{R}_+$. Let $\{\boldsymbol{u_1}, ..., \boldsymbol{u_n}\}$ be any orthonormal basis of $\Re_n$. Then*

$$\sum_{j=1}^{n} f(\boldsymbol{u_j}'\boldsymbol{A}\boldsymbol{u_j}) \geq trace[f(\boldsymbol{A})],$$

with equality iff each $\boldsymbol{u_j}$ is an eigenvector of $\boldsymbol{A}$, and $f(\cdot)$ is strictly concave.

Let $\boldsymbol{\Sigma}^{\frac{1}{2}} = \left(\boldsymbol{v_1}, \ldots, \boldsymbol{v_n} \right)$, $\boldsymbol{A} = \boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{x}\boldsymbol{x}')\boldsymbol{\Sigma}^{\frac{1}{2}}$, and using the orthonormal basis $\boldsymbol{u_j} = \boldsymbol{e_j}$, we have the following upper bound to our problem:

$$\begin{aligned} Z^* \leq \max\ & \sum_{j \in [n]} \sqrt{\boldsymbol{v_j}' \left(\boldsymbol{X} - \boldsymbol{x}\boldsymbol{x}' \right) \boldsymbol{v_j}} + \boldsymbol{\mu}'\boldsymbol{x} \\ \text{s.t.}\ & \begin{pmatrix} 1 & \boldsymbol{x}' \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in \text{conv} \left\{ \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix}' \ \middle|\ \boldsymbol{z} \in \mathcal{X} \right\}. \end{aligned}$$

If $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma}^2)$, then we have the following upper bound, in the case of 0-1 optimization problem (with $\boldsymbol{X}_{jj} = \boldsymbol{x}_j$):

$$\begin{aligned} Z^* \leq \max\ & \sum_{j \in [n]} \sigma_j \sqrt{x_j(1 - x_j)} + \boldsymbol{\mu}'\boldsymbol{x} \\ \text{s.t.}\ & \boldsymbol{x} \in \text{conv}(\mathcal{X}). \end{aligned} \tag{15}$$

This recovers the bound obtained using only the marginal moments without accounting for the fact that the cost function is uncorrelated (see Li et. al. [29] for a recent discussion on the marginal moment model).

4 Moment Bounds for Integer Programs

In this section, we apply the results to develop new moment bounds for integer programs that tighten some of the existing results in this area.

4.1 Maximum Clique Problem with Random Weights

When $\mathcal{X}$ is specified as the feasible region of a mixed integer linear program, Theorem 2 shows that the computing Z^* is related to the complexity of characterizing the set:

$$\text{conv}\left\{ \begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix}' \,\middle|\, \boldsymbol{x} \in \Re_n^+,\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ x_i \in \mathcal{Z},\ \forall i \in \mathcal{B} \subseteq [n] \right\}. \tag{16}$$

This transforms the difficulty of the stochastic problem of evaluating the maximum expected optimal value with mean, variance and covariance information to the difficulty of evaluating the convex hull of quadratic polynomials defined on the feasible region. The set (16) has been of keen interest to researchers interested in developing algorithms for quadratic mixed integer nonlinear programming problems. In general, characterizing this set is a difficult problem, since quadratic optimization problem over $\mathcal{X}$ is NP-hard.

Consider the case where the elements in $\mathcal{X}$ corresponds to the incidence vectors of a clique in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, then naturally $x_i + x_j \leq 1$ for $(i,j) \notin \mathcal{E}$, since every pair of nodes in a clique must share a common edge. In this case,

$$\text{conv}\left\{ \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix}' \,\middle|\, \boldsymbol{z} \in \mathcal{X} \right\} \subseteq \left\{ \begin{pmatrix} 1 & \boldsymbol{x}' \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \succeq 0 \,\middle|\, X_{ii} = x_i, \forall i \in \mathcal{V},\ X_{ij} = 0, \forall (i,j) \notin \mathcal{E} \right\}.$$

The convex object on the right hand side, known as Lovasz's Theta Body in the literature, is a well known semidefinite relaxation of the clique polytope in combinatorial optimization. In the simplest case when $\mathcal{G}$ is a null graph, then a clique is just a singleton, and so $\mathcal{X} = \left\{ \boldsymbol{x} : \sum_i x_i = 1,\ x_i \in \{0,1\} \right\}$. For $\mathcal{S} \subset \{1, \ldots, n\}$, let $\boldsymbol{e}_{\mathcal{S}} = \left(\mathbf{1}_{\mathcal{S}}(1), \mathbf{1}_{\mathcal{S}}(2), \ldots, \mathbf{1}_{\mathcal{S}}(n)\right)' \in \Re_n$, where $\mathbf{1}_{\mathcal{S}}(i) = 1$ if $i \in \mathcal{S}$ and 0 otherwise. It is easy to see that

$$\begin{aligned} \text{conv}\left\{ \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{z} \end{pmatrix}' \,\middle|\, \boldsymbol{z} \in \mathcal{X} \right\} &= \left\{ \sum_i x_i \begin{pmatrix} 1 \\ \boldsymbol{e}_i \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{e}_i \end{pmatrix}' \,\middle|\, \sum_{i \in \mathcal{V}} x_i = 1,\ x_i \geq 0, \forall i \in \mathcal{V} \right\}. \\ &= \left\{ \begin{pmatrix} 1 & \boldsymbol{x}' \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \,\middle|\, \begin{array}{ll} \sum_{i \in \mathcal{V}} x_i = 1, & X_{ii} = x_i, \forall i \in \mathcal{V} \\ x_i \geq 0, \forall i \in \mathcal{V}, & X_{ij} = 0, \forall i \neq j \end{array} \right\}. \end{aligned}$$

$Z(\tilde{\boldsymbol{c}})$ reduces to finding the maximum element in a set of random variables, a basic problem in the order statistic literature. Our approach produces the tightest estimation to this problem using only the first two moments of the random variables.

In the case when $\mathcal{G}$ is a bipartite graph with edge set $\mathcal{E}$, a clique in $\mathcal{G}$ has size of at most 2. Let $\mathcal{V}^o$ denote the set of singletons in $\mathcal{G}$. Now, we can use a natural decomposition argument to show that the set in (16) is given as:

$$\left\{ \sum_{i \in \mathcal{V}^o} z_i \begin{pmatrix} 1 \\ \boldsymbol{e}_i \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{e}_i \end{pmatrix}' + \sum_{(i,j) \in \mathcal{E}} z_{i,j} \begin{pmatrix} 1 \\ \boldsymbol{e}_{i,j} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{e}_{i,j} \end{pmatrix}' \,\middle|\, \sum_{i \in \mathcal{V}^o} z_i + \sum_{(i,j) \in \mathcal{E}} z_{i,j} = 1,\ z_i, z_{i,j} \geq 0 \right\}$$

$$= \left\{ \begin{pmatrix} 1 & \boldsymbol{x}' \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \;\middle|\; \begin{array}{lll} x_i = \sum_{j:(i,j)\in\mathcal{E}} X_{ij}, \forall i \notin \mathcal{V}^o, \; X_{ii} = x_i, \forall i \in \mathcal{V}, & & \boldsymbol{X} = \boldsymbol{X}' \\ X_{ij} \geq 0, \forall (i,j) \in \mathcal{E}, & X_{ij} = 0, \forall (i,j) \notin \mathcal{E}, & \sum_{i\in\mathcal{V}^o} x_i + \sum_{(i,j)\in\mathcal{E}} X_{ij} = 1 \end{array} \right\}.$$

Hence a tight bound based on the first two moments of the random weights in a maximal clique problem on bipartite graph can be obtained by solving the SDP, instead of the completely positive program.

4.2 Boolean Quadric Polytope and the Distributionally Robust Newsvendor

In this section, we focus on the moment problem for the set $\mathcal{X} = \{0,1\}^n$ with $Z(\tilde{\boldsymbol{c}}) = \sum_i (\tilde{c}_i)^+$ where $c^+ = \max(0, c)$. We derive new polynomial time computable bounds for this problem by exploiting the connection with the Boolean quadric polytope. As an application of the result, we identify a polynomial time solvable relaxation of the distributionally robust multi-item newsvendor problems under moment uncertainty that is tighter than the current state of art relaxation that uses quadratic decision rules with only marginal moment information (see Hanasusanto et. al. [22]).

The moment bound of interest is given as:

$$Z^* = \sup \left\{ E_\theta \left(\max_{\boldsymbol{x}\in\{0,1\}^n} \tilde{\boldsymbol{c}}'\boldsymbol{x} \right) \;\middle|\; E_\theta \left(\tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}' \right) = \boldsymbol{\Pi}, E_\theta \left(\tilde{\boldsymbol{c}} \right) = \boldsymbol{\mu}, E_\theta \left(1 \right) = 1, \theta \in \mathbb{M}(\Re_n) \right\}. \tag{17}$$

Hanasusanto et. al. [22] have shown that the problem of computing the bound Z^* in (17) is NP-hard. They proved this result by considering the dual formulation:

$$Z^* = \min \left\{ \boldsymbol{Y} \cdot \boldsymbol{\Pi} + \boldsymbol{y}'\boldsymbol{\mu} + y_0 \;\middle|\; \boldsymbol{c}'\boldsymbol{Y}\boldsymbol{c} + \boldsymbol{y}'\boldsymbol{c} + y_0 \geq \max_{\boldsymbol{x}\in\{0,1\}^n} \boldsymbol{c}'\boldsymbol{x}, \; \forall \boldsymbol{c} \in \Re_n \right\}. \tag{18}$$

The separation problem in this case is given as:
Given $\boldsymbol{Y} \in \mathcal{S}_n$, $\boldsymbol{y} \in \Re_n$ and $y_0 \in \Re$, check if the quadratic function $q(\boldsymbol{c}) = \boldsymbol{c}'\boldsymbol{Y}\boldsymbol{c} + \boldsymbol{y}'\boldsymbol{c} + y_0$ is greater than or equal to $\boldsymbol{c}'\boldsymbol{x}$ for all $\boldsymbol{c} \in \Re_n$ and $\boldsymbol{x} \in \{0,1\}^n$. If not, find a $\boldsymbol{c}^* \in \Re_n$ and $\boldsymbol{x}^* \in \{0,1\}^n$ such that ${\boldsymbol{c}^*}'\boldsymbol{Y}\boldsymbol{c}^* + \boldsymbol{y}'\boldsymbol{c}^* + y_0 < {\boldsymbol{c}^*}'\boldsymbol{x}^*$. The separation problem is NP-hard which follows the observation that testing if a quadratic function majorizes the L1 norm is NP-hard (see [22].) From the equivalence of separation and optimization, this implies that computing Z^* is NP-hard.

As an application, consider the multi-item newsvendor problem discussed in Hanasusanto et. al. [22]. In the multi-item newsvendor problem, a newsvendor sells n different types of products. The unit order cost for each product i is c_i while the unit selling price for each product i is v_i. Assume that any item of product i that is unsold at the end of the period is salvaged at a cost of g_i while the cost for stockout for a unit of product i is f_i. Denote the order quantity decision for product i as q_i and the demand for product i as d_i. The amount of units of product i that the newsvendor sells is $\min(q_i, d_i)$. The total cost for the newsvendor is given as:

$$
\begin{aligned}
L(\boldsymbol{q},\boldsymbol{d}) &= \sum_{i\in[n]} c_i q_i - \sum_{i\in[n]} v_i \min(q_i,d_i) - \sum_{i\in[n]} g_i\left(q_i-\min(q_i,d_i)\right) + \sum_{i\in[n]} f_i\left(d_i-\min(q_i,d_i)\right),\\
&= \boldsymbol{a}'\boldsymbol{q}+\boldsymbol{b}'\boldsymbol{d}+\sum_{i\in[n]} h_i\left(d_i-q_i\right)^+,
\end{aligned}
$$

where $\boldsymbol{a} = \boldsymbol{c} - \boldsymbol{g} \geq 0$, $\boldsymbol{b} = \boldsymbol{g} - \boldsymbol{v}$, and $\boldsymbol{h} = \boldsymbol{v} - \boldsymbol{g} + \boldsymbol{f} \geq 0$. Since the newsvendor needs to decide on the order quantities before knowing the actual demand, the demand vector $\tilde{\boldsymbol{d}}$ is assumed to be random with a distribution θ. In the distributionally robust newsvendor, the newsvendor is assumed to know the mean $\boldsymbol{\mu}$ and the second moment matrix $\boldsymbol{\Pi}$ of the random demand vector $\tilde{\boldsymbol{d}}$ but the distribution itself is unknown. The distributionally robust newsvendor chooses nonnegative order quantities accounting for the worst-case distribution of the expected costs. The distributionally robust newsvendor problem is formulated as:

$$
\min_{\boldsymbol{q}\in\Re_n^+}\sup_{\theta\in\Theta} E_\theta\left(L(\boldsymbol{q},\tilde{\boldsymbol{d}})\right) = \min_{\boldsymbol{q}\in\Re_n^+}\left\{\boldsymbol{a}'\boldsymbol{q}+\boldsymbol{b}'\boldsymbol{\mu}+\sup_{\theta\in\Theta} E_\theta \overbrace{\left(\max_{\boldsymbol{x}\in\{0,1\}^n}\sum_{i\in[n]} h_i\left(d_i-q_i\right)x_i\right)}^{\sum_{i\in[n]} h_i(d_i-q_i)^+}\right\},
$$

where $\Theta = \left\{\theta \;\middle|\; E_\theta\left(\tilde{\boldsymbol{d}}\tilde{\boldsymbol{d}}'\right)=\boldsymbol{\Pi}, E_\theta\left(\tilde{\boldsymbol{d}}\right)=\boldsymbol{\mu}, E_\theta\left(1\right)=1, \theta\in\mathbb{M}(\Re_n)\right\}$. The distributionally robust optimization problem in this case is NP-hard (see Bertsimas et. al. [10], Hanasusanto et. al. [22]). One approach to solve this problem is to enumerate all the possible 2^n combinations in the objective function and solve the exact semidefinite program using (18):

$$
\begin{aligned}
\min\;& \boldsymbol{a}'\boldsymbol{q}+\boldsymbol{b}'\boldsymbol{\mu}+\boldsymbol{Y}\cdot\boldsymbol{\Pi}+\boldsymbol{y}'\boldsymbol{\mu}+y_0\\
\text{s.t.}\;& \begin{pmatrix} \boldsymbol{Y} & (\boldsymbol{y}-\boldsymbol{h}\circ\boldsymbol{x})/2\\ (\boldsymbol{y}-\boldsymbol{h}\circ\boldsymbol{x})'/2 & y_0+\sum_{i\in[n]} h_i q_i x_i\end{pmatrix}\succeq 0,\ \forall \boldsymbol{x}\in\{0,1\}^n,\\
& \boldsymbol{q}\geq 0.
\end{aligned}
\tag{19}
$$

The semidefinite program is however exponential sized in the number of products. A simpler approach is to use the marginal moment bound which provides an upper bound on the worst-case cost. In this case, the problem reduces to:

$$
\begin{aligned}
Z^*_{\text{socp}} = \min\;& \boldsymbol{a}'\boldsymbol{q}+\boldsymbol{b}'\boldsymbol{\mu}+\sum_{i\in[n]}\frac{h_i}{2}\left(\mu_i-q_i+\sqrt{(\mu_i-q_i)^2+\sigma_i^2}\right)\\
\text{s.t.}\;& \boldsymbol{q}\geq 0,
\end{aligned}
\tag{20}
$$

which can be solved as a second order cone program. This formulation is polynomial sized in the number of products but does not account for the correlation information. We now propose an alternate relaxation for the distributionally robust newsvendor problem by using results for the Boolean quadric polytope (cf. Padberg [43]). The Boolean quadric polytope is given as:

$$\mathrm{BQP}_n = \mathrm{conv}\left\{ \begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix}' \;\middle|\; x_i \in \{0,1\},\ \forall i \in [n] \right\}. \tag{21}$$

A simple relaxation to the Boolean quadric polytope in (21) is obtained from the reformulation-linearization technique as follows:

$$\mathrm{BQP}_n \subseteq \left\{ \begin{pmatrix} 1 & \boldsymbol{x}' \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in \mathcal{S}_{n+1} \;\middle|\; \begin{array}{l} X_{ii} = x_i,\ \forall i, \quad\ X_{ij} \leq x_i,\ \forall i,j \\ X_{ij} \geq 0,\ \forall i \leq j,\ X_{ij} \geq x_i + x_j - 1,\ \forall i \leq j \end{array} \right\}.$$

Additional valid inequalities such as cut and clique inequalities have also been identified for this polytope (see Padberg [43], Deza and Laurent [15]). In our computation, we focus on the simplest relaxation where an upper bound on Z^* is obtained from using the semidefinite relaxation with the reformulation-linearization technique as follows:

$$\begin{array}{rl} Z^* \leq Z^*_{\mathrm{sdp}} = \min_{\boldsymbol{q} \in \Re^+_n} \Big\{ \boldsymbol{a}'\boldsymbol{q} + \boldsymbol{b}'\boldsymbol{\mu} + & \max\ \mathrm{trace}(\mathrm{Diag}(\boldsymbol{h})\boldsymbol{U}) - \sum_i h_i q_i x_i \Big\} \\ & \text{s.t.} \begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{\mu} & \boldsymbol{U}' \\ \boldsymbol{\mu}' & 1 & \boldsymbol{x}' \\ \boldsymbol{U} & \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \succeq 0, \\ & X_{ii} = x_i, \quad \forall i, \\ & X_{ij} \leq x_i, \quad \forall i,j, \\ & X_{ij} \geq 0, \quad \forall i \leq j, \\ & X_{ij} \geq x_i + x_j - 1, \quad \forall i \leq j. \end{array} \tag{22}$$

By taking the dual, we reformulate the problem as:

$$\begin{array}{rl} Z^*_{\mathrm{sdp}} = \min & \boldsymbol{a}'\boldsymbol{q} + \boldsymbol{b}'\boldsymbol{\mu} + \boldsymbol{Y} \cdot \boldsymbol{\Pi} + \boldsymbol{y}'\boldsymbol{\mu} + y_0 + \boldsymbol{S} \cdot (\boldsymbol{e}\boldsymbol{e}') \\ \text{s.t.} & \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{y}/2 & -Diag(\boldsymbol{h})/2 \\ \boldsymbol{y}'/2 & y_0 & (-\boldsymbol{e}'\boldsymbol{W} + (\boldsymbol{h} \circ \boldsymbol{q})')/2 - \boldsymbol{e}'\boldsymbol{S}) \\ -Diag(\boldsymbol{h})/2 & (-\boldsymbol{e}'\boldsymbol{W} + (\boldsymbol{h} \circ \boldsymbol{q})')'/2 - (\boldsymbol{e}'\boldsymbol{S})' & (\boldsymbol{W} + \boldsymbol{W}')/2 - \boldsymbol{Z} - \boldsymbol{S} \end{pmatrix} \succeq 0, \\ & \boldsymbol{W} \geq 0, \\ & \boldsymbol{Z} \geq 0, \quad \boldsymbol{Z} = \boldsymbol{Z}', \\ & \boldsymbol{S} \geq 0, \quad \boldsymbol{S} = \boldsymbol{S}', \\ & \boldsymbol{q} \geq 0. \end{array} \tag{23}$$

To test the strength of the two formulations, we consider a numerical experiment inspired by the setup in Hanasusanto et. al. [22]. For each product $i = 1, \ldots, n$, we set $v_i = 10$, $g_i = 1$, $f_i = 2.5$. The costs c_i are sampled uniformly from the box $[3, 8]^n$. The mean demand vector is sampled uniformly from $[5, 100]^n$ while the standard deviation is sampled from independent uniform distributions on $[0.1\boldsymbol{\mu}, \boldsymbol{\mu}]$. The correlations are generated by sampling a random matrix $\boldsymbol{T} \in \Re_{n \times n}$ with independent uniform random variables in $[0, 1]$ and setting the correlation matrix to $\mathrm{diag}(\boldsymbol{r})\boldsymbol{R}\mathrm{diag}(\boldsymbol{r})$ where $\boldsymbol{R} = \boldsymbol{T}'\boldsymbol{T}$

and $\boldsymbol{r}$ is a vector whose ith element is defined as $r_i = 1/\sqrt{R_{ii}}$. Note that in our data generation process, the correlations among the products is assumed to be positive by construction. For each newsvendor, we vary the number of items n from 5 to 50 in steps of 5. We solve 50 random instances for each n. As should be expected, the computational time to solve the semidefinite programs is larger than the second order cone programs. However we are able to solve on average each semidefinite programs of size up to 50 in under 150 seconds. Note that this seems to be much beyond the scale of problems that can be solved using the exact reformulation in (19) (see discussion in Hanasusanto et. al. [22]). Figure (2) plots the minimum, median and maximum gap in the objective value $|(Z^*_{\text{socp}} - Z^*_{\text{sdp}})/Z^*_{\text{sdp}}| \times 100\%$ where $Z^*_{\text{socp}} \geq Z^*_{\text{sdp}}$ over the 50 instances as we vary the number of products on a semi-log graph. The result clearly indicates that the semidefinite relaxation can improve the bound quite significantly (on an average of 4% for the larger instances).

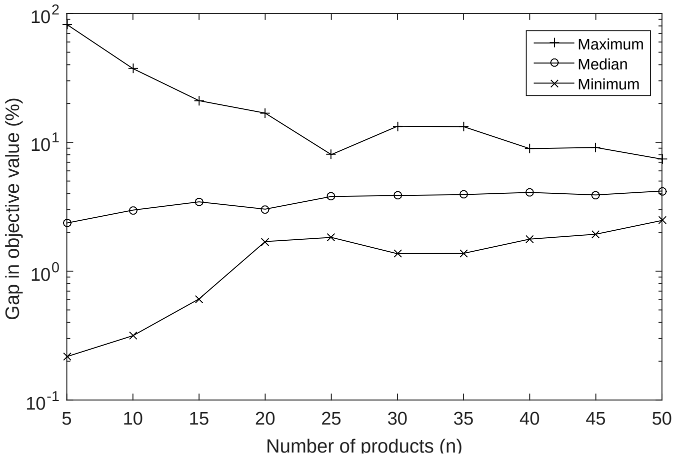

Fig. 2 Gap in the optimal objective values of the SDP and SOCP across 50 instances for different values of n.

5 Range of Dependent Random Variables

Bounds on the expected range:

$$E\left(\max_{i\in[n]} \tilde{c}_i - \min_{i\in[n]} \tilde{c}_i\right),$$

have been developed over half a decade ago with most of the early studies focussing on independent and identically distributed random variables (see Gumbel [21], Hartley and David [23] and Plackett [46]). Over the past few decades, there has been a growing interest in generalizing these results to dependent random variables. In this section, we use the semidefinite programming approach to compute bounds on the expected range of dependent random variables

Starting from the early 1950s, researchers have found closed form upper bounds on functions of order statistics of random variables using the first two moments of the random variables. Assuming identically and independently distributed random variables with mean μ and variance σ^2, Plackett [46] and Gumbel [21] evaluated an upper bound on the expected range:

$$E\left(\max_{i\in[n]}\tilde{c}_i - \min_{i\in[n]}\tilde{c}_i\right) \leq \sigma n\sqrt{\frac{2}{(2n-1)!}\left((2n-2)! - [(n-1)!]^2\right)}. \tag{24}$$

The bound in (24) was also shown to be the tightest possible, namely they constructed a distribution that for any given mean μ and variance σ^2 attains the bound. Dropping the assumptions of independence and identical distributions, Arnold and Groeneveld [3] developed a closed form bound on the expected range of order statistics using only the mean μ_i and variance σ_i^2 for each random variable:

$$E\left(\max_{i\in[n]}\tilde{c}_i - \min_{i\in[n]}\tilde{c}_i\right) \leq \sqrt{2}\sqrt{\sum_{i\in[n]}[(\mu_i-\bar{\mu})^2 + \sigma_i^2]},$$

where $\bar{\mu} = \sum_i \mu_i/n$. For the special instance of identical means and identical variances, this bound reduces to:

$$E\left(\max_{i\in[n]}\tilde{c}_i - \min_{i\in[n]}\tilde{c}_i\right) \leq \sigma\sqrt{2n}. \tag{25}$$

For identical means and covariances, Arnold and Groeneveld [3] showed that the bound is tight by constructing a distribution that attains the bound. Aven [5] and Lefevre [28] extended these bounds to dependent random variables by incorporating covariance information among the random variables. Papadatos [45] and Nagarajan [39] improved on these covariance based bounds to develop an upper bound on the expected range:

$$E\left(\max_{i\in[n]}\tilde{c}_i - \min_{i\in[n]}\tilde{c}_i\right) \leq \sqrt{2}\sqrt{\sum_{i\in[n]}\left[(\mu_i-\bar{\mu})^2 + \sigma_i^2\right] - n\text{Variance}\left[\sum_{i\in[n]}\frac{\tilde{c}_i}{n}\right]}. \tag{26}$$

Papadatos [45] showed that for exchangeable random variables with identical means $\mu_i = \mu$, identical variances $\sigma_i^2 = \sigma^2$ and identical covariances $\text{Cov}[\tilde{c}_i, \tilde{c}_j] = \rho\sigma^2$ for $i \neq j$, the bound in (26) is tight and reduces to:

$$E\left(\max_{i\in[n]}\tilde{c}_i - \min_{i\in[n]}\tilde{c}_i\right) \leq \sigma\sqrt{2(n-1)(1-\rho)}. \tag{27}$$

Note that to be a valid correlation matrix, the correlation coefficient ρ must lie in the interval $[-1/(n-1), 1]$. This guarantees that the bounds in (27) are at least as tight as the bounds in (25). To the best of our knowledge, the Papadatos and Nagarajan bound in (26) is the "sharpest" known closed form upper bound on the expected range using the mean, variance and covariance information. While this bound has a simple closed form expression, there is no guarantee that the bound is tight for arbitrary means, variances and covariances.

The tight upper bound on the expected range of a set of random variables is found by solving the following moment problem:

$$Z^*_{\text{range}} = \sup \left\{ E_\theta \left(\max_{i\in[n]} \tilde{c}_i - \min_{i\in[n]} \tilde{c}_i \right) \;\middle|\; E_\theta \left(\tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}'\right) = \boldsymbol{\Pi}, E_\theta\left(\tilde{\boldsymbol{c}}\right) = \boldsymbol{\mu}, E_\theta\left(1\right) = 1, \theta \in \mathbb{M}(\Re_n) \right\}.$$

In this case, by setting $\mathcal{X} = \{(\boldsymbol{e}_i - \boldsymbol{e}_j) \mid i, j \in [n]\}$, and using the formulation in (7), we obtain the semidefinite programming representation for the moment problem as follows:

$$\begin{aligned} Z^*_{\text{range}} = \max & \sum_{i\in[n]}\sum_{j\in[n]} (\boldsymbol{e}_i - \boldsymbol{e}_j)' \boldsymbol{w}_{ij} \\ \text{s.t.} & \sum_{i\in[n]}\sum_{j\in[n]} \begin{pmatrix} \boldsymbol{W}_{ij} & \boldsymbol{w}_{ij} \\ \boldsymbol{w}'_{ij} & \alpha_{ij} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{\mu} \\ \boldsymbol{\mu}' & 1 \end{pmatrix}, \\ & \begin{pmatrix} \boldsymbol{W}_{ij} & \boldsymbol{w}_{ij} \\ \boldsymbol{w}'_{ij} & \alpha_{ij} \end{pmatrix} \succeq 0, \qquad \forall i, j \in [n]. \end{aligned}$$

This formulation involves $m = O(n^2)$ positive semidefinite matrices, each of dimension $(n+1)\times(n+1)$. By applying the result in Section 3 (particularly formulation (14)), we obtain a reduced semidefinite program to solve the problem involving a single positive semidefinite matrix of dimension $(2n+1)\times(2n+1)$:

$$\begin{aligned} Z^*_{\text{range}} = \max\ & \text{trace}(\boldsymbol{Y}) \\ \text{s.t.}\ & \begin{pmatrix} \boldsymbol{\Pi} & \boldsymbol{\mu} & \boldsymbol{Y}' \\ \boldsymbol{\mu}' & 1 & \sum_{i\in[n]}\sum_{j\in[n]} \alpha_{ij}(\boldsymbol{e}_i' - \boldsymbol{e}_j') \\ \boldsymbol{Y} & \sum_{i\in[n]}\sum_{j\in[n]} \alpha_{ij}(\boldsymbol{e}_i - \boldsymbol{e}_j) & \sum_{i\in[n]}\sum_{j\in[n]} \alpha_{ij}(\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i' - \boldsymbol{e}_j') \end{pmatrix} \succeq 0, \\ & \sum_{i\in[n]}\sum_{j\in[n]} \alpha_{ij} = 1, \\ & \alpha_{ij} \geq 0, \qquad \forall i, j \in [n]. \end{aligned} \tag{28}$$

This leads to a reduction in the number of variables in the semidefinite program from $O(n^4)$ to $O(n^2)$.

5.1 Application to Random Walk

In this section, we use the reduced semidefinite program to find tight bounds on the expected range of partial sums from a problem arising in random walk theory. Consider a set of n random variables $\tilde{c}_1, \ldots, \tilde{c}_n$ with the partial sums defined as:

$$\tilde{s}_k = \tilde{c}_1 + \ldots + \tilde{c}_k, \quad \forall k \in [n].$$

The range of the partial sum of the random variables is defined as:

$$\tilde{r}_n = \max\left(\tilde{s}_1, \ldots, \tilde{s}_n\right) - \min\left(\tilde{s}_1, \ldots, \tilde{s}_n\right).$$

Assume that the $\tilde{c}_1, \ldots, \tilde{c}_n$ is a set of independent and identically distributed normal random variables with mean 0 and standard deviation 1. Feller [19] showed that the asymptotic scaling behavior of the expected range is given by:

$$E(\tilde{r}_n) \rightarrow 2\sqrt{\frac{2n}{\pi}} \text{ as } n \text{ approaches } \infty.$$

Anis and Lloyd [4] evaluated explicitly the exact expected range of the partial sums for a finite set of independent standard normal random variables for arbitrary n:

$$E(\tilde{r}_n) = \sqrt{\frac{2}{\pi}} \sum_{k=1}^{n-1} \frac{1}{\sqrt{k}}.$$

As n approaches ∞, this result converges to the result of Feller [19]. Hurst [24] analyzed the records of annual values of different natural phenomena such as rainfall, water levels, riverflows and found that the empirical observed range does not necessarily scale as $\sqrt{n}$ as predicted by the theory, but closer to n^α where α ranged from 0.6 to 0.8. The exponent α is known as the Hurst exponent. Hurst's empirical findings led to a flurry of research providing alternate explanations to this observation including long range dependence among the random variables, non-stationarity in the stochastic process, possibility of infinite second moments, and transient behavior of the limiting process (see Mesa and Poveda [36]).

We use the semidefinite program to compute the tight upper bound on the expected range of partial sums for uncorrelated and correlated random variables arising in random walks. We solve the following moment problem:

$$Z^*_{\text{rw}} = \sup\left\{ E_\theta\left(\max_{i\in[n]}\left(\sum_{j\in[i]} \tilde{c}_j\right) - \min_{i\in[n]}\left(\sum_{j\in[i]} \tilde{c}_j\right)\right) \;\middle|\; E_\theta\left(\tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}'\right) = \boldsymbol{\Pi}, E_\theta\left(\tilde{\boldsymbol{c}}\right) = \boldsymbol{\mu}, E_\theta\left(1\right) = 1, \theta \in \mathbb{M}(\Re_n) \right\}. \tag{29}$$

Popescu [48] and Yu et. al. [56] studied second order moment bounds of the form:

$$\sup\left\{ E_\theta\left(Z(\boldsymbol{D}\tilde{\boldsymbol{c}})\right) \;\middle|\; E_\theta\left(\tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}'\right) = \boldsymbol{\Pi}, E_\theta\left(\tilde{\boldsymbol{c}}\right) = \boldsymbol{\mu}, E_\theta\left(1\right) = 1, \theta \in \mathbb{M}(\Re_n) \right\},$$

where $\boldsymbol{D}$ is a given matrix of size $p \times n$ and $\tilde{\boldsymbol{c}}$ is a random vector of size $n \times 1$. They showed that the problem can be equivalently reformulated as:

$$\sup\left\{ E_{\hat{\theta}}\left(Z(\tilde{\boldsymbol{s}})\right) \;\middle|\; E_{\hat{\theta}}\left(\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}'\right) = \boldsymbol{D}\boldsymbol{\Pi}\boldsymbol{D}', E_{\hat{\theta}}\left(\tilde{\boldsymbol{s}}\right) = \boldsymbol{D}\boldsymbol{\mu}, E_{\hat{\theta}}\left(1\right) = 1, \hat{\theta} \in \mathbb{M}(\Re_p) \right\}.$$

where the first two moments defined through the linear transformation of the random vector $\tilde{\boldsymbol{s}} = \boldsymbol{D}\tilde{\boldsymbol{c}}$. Define the transformation matrix $\boldsymbol{D}$ of size $n \times n$ from (29) as:

$$\boldsymbol{D} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}$$

The mean $\boldsymbol{\mu}_s$ and the second moment matrix $\boldsymbol{\Pi}_s$ for $\tilde{\boldsymbol{s}}$ is computed in terms of the mean $\boldsymbol{\mu}$ and the second matrix $\boldsymbol{\Pi}$ of the random vector $\tilde{\boldsymbol{c}}$ as follows:

$$\boldsymbol{\mu}_s = \boldsymbol{D}\boldsymbol{\mu} \text{ and } \boldsymbol{\Pi}_s = \boldsymbol{D}\boldsymbol{\Pi}\boldsymbol{D}'.$$

The moment problem is then found by solving:

$$Z^*_{\text{rw}} = \sup \left\{ E_{\hat{\theta}} \left(\max_{i \in [n]} \tilde{s}_i - \min_{i \in [n]} \tilde{s}_i \right) \;\middle|\; E_{\hat{\theta}} \left(\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}' \right) = \boldsymbol{\Pi}_s, E_{\hat{\theta}} \left(\tilde{\boldsymbol{s}} \right) = \boldsymbol{\mu}_s, E_{\hat{\theta}} \left(1 \right) = 1, \hat{\theta} \in \mathbb{M}(\Re_n) \right\},$$

which can be reformulated as a semidefinite program using (28).

5.1.1 Numerical Results

In the first set of numerical experiments, we assume that $\tilde{c}_1, \dots, \tilde{c}_n$ is a set of n uncorrelated random variables with identical mean 0 and standard deviation 1. The tight Jensen's lower bound on the expected range is exactly 0 and is an optimistic estimate of the reservoir capacity. We evaluate the expected range of the partial sums for $n = 1$ to $n = 60$. For normal random variables, the expected range for the partial sums is known in closed form in this case (see Anis and Lloyd [4]). We also evaluate the tight upper bound on the expected range using semidefinite programming and the weaker Papadatos and Nagarajan upper bound. The results are provided in Figure 3. In the second set of numerical experiments, we assume that $\tilde{c}_1, \dots, \tilde{c}_n$ is a set of n correlated random variables with means generated randomly in the range $[-1, 1]$ and standard deviations generated randomly in the range $[0, 2]$. The correlation matrices are also generated randomly with eigenvalues chosen randomly from a uniform distribution using the MATLAB command *gallery('randcorr',n)*. We compare the range for the multivariate normal distribution using simulation with the tight Jensen's lower bound, the tight SDP upper bound and the weaker Papadatos and Nagarajan bound for $n = 20$. The results are provided in Figure 4 across 100 randomly generated sets of parameter values. As indicated in the figure, the SDP upper bound is the closest to the simulation estimate. Note that the SDP estimate on the expected range can also be interpreted as a pessimistic estimate on the capacity of the reservoir across all multivariate joint distributions with the given mean, variance and covariance information. For additional numerical results in random walk problems, the reader is referred to the supplemental material available online.

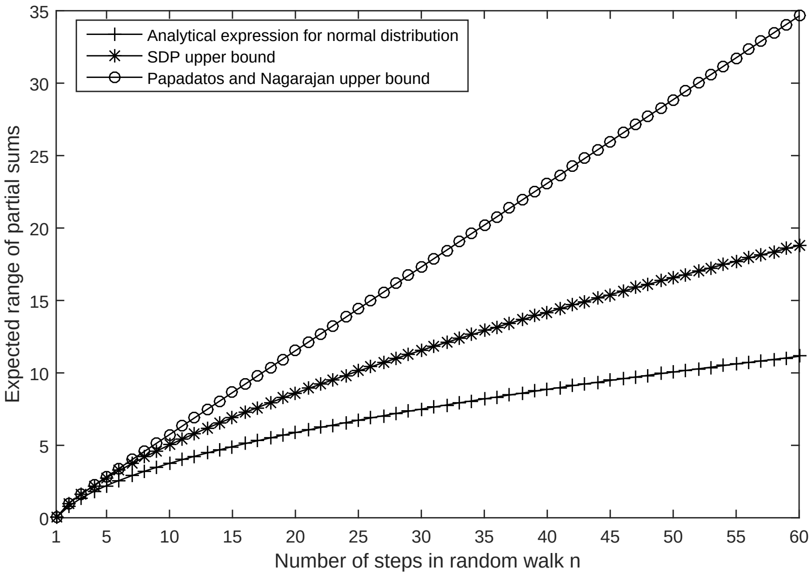

Fig. 3 Expected range of partial sums for uncorrelated random variables with $\mu = 0$ and $\sigma = 1$.

5.2 Application to Best-Worst Choice Probabilities

In this section, we use the semidefinite program to compute best, worst and best-worst choice probabilities for discrete choice models that help describe the collective choices (preferences) of a set of individuals. Discrete choice models provide a theoretical and empirical explanation of the choices that are made by individuals from among a finite set of mutually exclusive alternatives. These alternatives could represent for example an individual's choice among various brands of coffee, an individual's choice among different types of cereals or an individual's choice among the modes of transport such as car, bus, rail or taxi when they travel from home to work. These models are typically derived under an assumption of utility-maximizing behavior where each individual states or reveals his or her preference for the alternative with the maximum utility. The researcher who is analyzing the underlying choice behavior of a set of individuals however does not observe each individual's utility. Let the set of alternatives be indexed by $[n]$. From the researcher's point of view, each alternative is assumed have an additive random utility of the form:

$$\tilde{c}_i = v_i + \tilde{\epsilon}_i, \quad \forall i \in [n],$$

where v_i is the systematic term that captures the attributes of the alternative that the researcher can observe and $\tilde{\epsilon}_i$ is the random error term that captures features not accounted for in the deterministic part of the utility. For example in the choice among modes of transport, v_i captures observable attributes

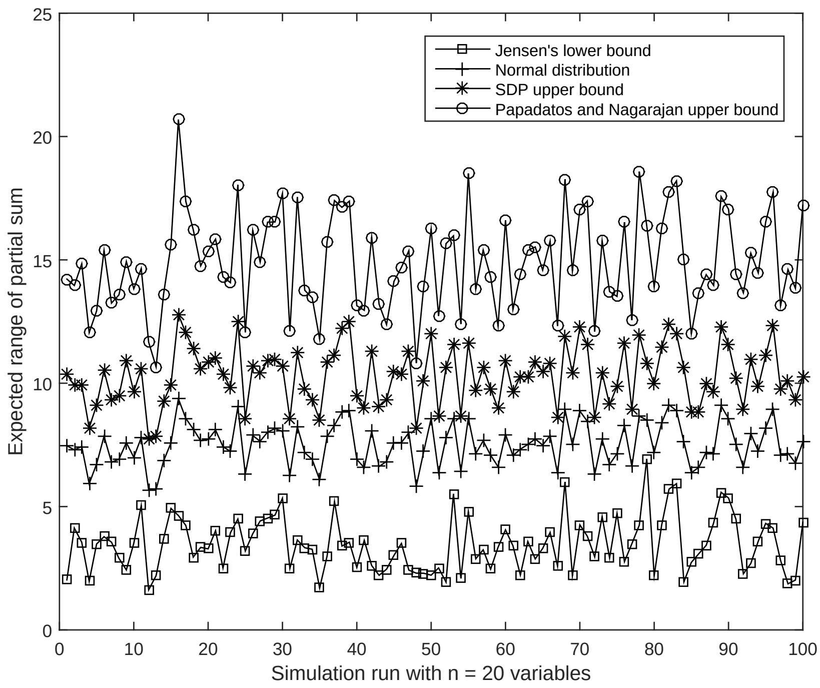

Fig. 4 Expected range of partial sums for correlated random variables.

such as the cost for a mode of transport and the time to travel for a mode of transport. From the point of the view of the individual, the realization of the $\tilde{\epsilon}_i$ values is exactly known but for the researcher, these realizations are unknown and hence treated as random variables. A fundamental question that arises in discrete choice models is to evaluate the expected maximum utility for the set of individuals and the probability that an alternative is the most preferred for this set of individuals. Suppose that the vector of random error terms is continuously distributed across the set of individuals. Assuming that each individual chooses the most preferred (or best) alternative, the expected maximum utility is given as:

$$E\left(\max_{i\in[n]}(v_i+\tilde{\epsilon}_i)\right).$$

The probability that alternative i is the best alternative is evaluated as:

$$B(i)=P\bigg(v_i+\tilde{\epsilon}_i\geq v_k+\tilde{\epsilon}_k,\ \forall k\in[n]\bigg),\quad \forall i\in[n].$$

A special instance in which a simple analytical formula for the best choice probability is available is the Multinomial Logit (MNL) model (see Luce and

Suppes [31] and McFadden [35]). In the MNL model, the error terms are assumed to be independent and identically distributed Gumbel variables with cumulative distribution function given as:

$$P(\tilde{\epsilon}_i \leq \varepsilon) = e^{-e^{-\varepsilon}}, \quad \forall i \in [n].$$

The best choice probability in MNL has a simple closed form given as:

$$B_{\text{mnl}}(i) = \frac{e^{v_i}}{\sum_{k \in [n]} e^{v_k}}, \quad \forall i \in [n].$$

This analytical formula for the best choice probabilities is the workhorse of discrete choice models in areas such as marketing (see Greene [20]) and transportation (see Ben-Akiva and Lerman [6]) where the individuals are asked to identify their most preferred alternative.

More recently, there has been a growing interest in an alternative choice design known as the "best-worst" choice design where the individual is asked to select the most preferred and the least preferred alternative (see Marley and Louviere [32]). The rationale behind the best-worst choice design is that psychologically individuals identify their extreme preferences more reliably as compared to ranking based choice designs which involves ranking all the alternatives that often leads to fatigue and lesser reliability. However even for the simplest MNL model, the worst and best-worst choice probabilities are known to have complicated closed form representations. For MNL, the worst choice probability has been explicitly evaluated as (see Van Ophem et. al. [55], Palma et. al. [44], Dijk et. al. [18]):

$$W_{\text{mnl}}(j) = \sum_{j \in \mathcal{S} \subseteq [n]} (-1)^{|\mathcal{S}|-1} \frac{e^{v_j}}{\sum_{k \in \mathcal{S}} e^{v_k}}, \quad \forall j \in [n],$$

where $|\mathcal{S}|$ is the cardinality of the set $\mathcal{S}$. This formula for the worst choice probability is derived using the inclusion-exclusion counting technique and involves the summation of 2^{n-1} terms. The formula for the joint best-worst choice probabilities for MNL is similarly given as:

$$BW_{\text{mnl}}(i,j) = \frac{e^{v_i}}{\sum_{k \in [n]} e^{v_k}} \sum_{j \in \mathcal{S} \subseteq [n] \setminus \{i\}} (-1)^{|\mathcal{S}|-1} \frac{e^{v_j}}{\sum_{k \in \mathcal{S}} e^{v_k}}, \quad \forall i, j \in [n], i \neq j.$$

Note that under the standard MNL model for best-worst choice, the expected maximum difference in the best and worst utilities is given by:

$$E\left(\max_{i \in [n]} (v_i - v_j + \tilde{\epsilon}_{ij})\right),$$

where for each i, j pair, ϵ_{ij} are (independent) random extreme value distributed random variables.

An alternative to MNL that incorporates correlation information is the Multinomial Probit (MNP) model where the error terms are assumed to be normally distributed random variables. The advantage of this model is that it captures arbitrary correlation information which logit does not. However there is no analytical formula for the best, worst and best-worst choice probabilities and simulation remains the popular approach to estimate choice probabilities for MNP models. Mishra, Natarajan and Teo [37] recently proposed using the semidefinite program in (7) to estimate the best choice probabilities assuming only knowledge on the mean and covariance information but dropping the assumption of normality. Inspired by this, we use a moment based approach to model the best-worst choice and solve:

$$Z^*_{\text{bw}} = \sup \left\{ E_\theta \left(\max_{i\in[n]} (v_i + \tilde{\epsilon}_i) - \min_{i\in[n]} (v_i + \tilde{\epsilon}_i) \right) \;\middle|\; E_{\theta_\epsilon}\left(\tilde{\boldsymbol{\epsilon}}\tilde{\boldsymbol{\epsilon}}'\right) = \boldsymbol{\Pi}_\epsilon, E_{\theta_\epsilon}\left(\tilde{\boldsymbol{\epsilon}}\right) = \boldsymbol{\mu}_\epsilon, E_{\theta_\epsilon}(1) = 1, \theta_\epsilon \in \mathbb{M}(\Re_n) \right\}.$$

Applying the results from the previous section, we can solve this as a semidefinite program:

$$\begin{aligned} Z^*_{\text{sdp}} = \max\ & \text{trace}(\boldsymbol{Y}) \\ \text{s.t.}\ & \begin{pmatrix} \boldsymbol{\Pi}_\epsilon + \boldsymbol{v}\boldsymbol{v}' + \boldsymbol{v}\boldsymbol{\mu}'_{\boldsymbol{\epsilon}} + \boldsymbol{\mu}_{\boldsymbol{\epsilon}}\boldsymbol{v}' & \boldsymbol{\mu} + \boldsymbol{v} & \boldsymbol{Y}' \\ \boldsymbol{\mu}' + \boldsymbol{v}' & 1 & \sum_{i\in[n]}\sum_{j\in[n]} \alpha_{ij}(\boldsymbol{e}_i' - \boldsymbol{e}_j') \\ \boldsymbol{Y} & \sum_{i\in[n]}\sum_{j\in[n]} \alpha_{ij}(\boldsymbol{e}_i - \boldsymbol{e}_j) & \sum_{i\in[n]}\sum_{j\in[n]} \alpha_{ij}(\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i' - \boldsymbol{e}_j') \end{pmatrix} \succeq 0, \\ & \sum_{i\in[n]}\sum_{j\in[n]} \alpha_{ij} = 1, \\ & \boldsymbol{\alpha} \geq 0, \end{aligned}$$

where $Z^*_{\text{sdp}} = Z^*_{\text{bw}}$. The best-worst choice probability for alternatives i, j under the model can be estimated as $BW_{\text{sdp}}(i,j) = \alpha^*_{ij}$ where α^*_{ij} is the optimal decision variables for the semidefinite program. The best choice probability for alternative i is given as $B_{\text{sdp}}(i) = \sum_j \alpha^*_{ij}$ while the worst choice probability for alternative j is given as $W_{\text{sdp}}(j) = \sum_i \alpha^*_{ij}$.

5.2.1 Numerical Results

In the numerical experiments, we assume that the error terms $\tilde{\epsilon}_i$ are correlated with mean $\gamma \approx 0.5772$ and standard deviation randomly chosen in the range $[0, 2\pi/\sqrt{6}]$. The error terms do not have identical distributions. The correlation matrix was randomly generated. The systematic component of the utility v_i was randomly generated in the interval $[0, 2]$ for each i independent of each other. We estimate the best, worst and best-worst choice probabilities for $n = 10$ alternatives across 100 randomly generated sets of parameter values. The probabilities for MNL were computed by ignoring the correlation information. The choice probabilities for MNP were estimated through 1000000 simulations with the normal distribution while the choice probabilities for the

SDP were evaluated by solving the semidefinite program. The choice probabilities are plotted in Figure 5. In this case the MNL model is a poor fit as should be expected since we drop correlation information in computing the choice probabilities. The MNP and SDP models are fairly similar in the scatter plot for the choice probabilities. To compare the choice probabilities across different models, we adopt the total variation distance. Consider two set of discrete probability measures $P(k)$ and $Q(k)$ defined on the same finite sample space $k \in [K]$. The total variation distance between the two probability measures is defined as:

$$d(P, Q) = \frac{1}{2} \sum_{k \in [K]} |P(k) - Q(k)| .$$

Note that the minimum total variation distance between the probability measures $P(\cdot)$ and $Q(\cdot)$ is 0 if and only if $P(k) = Q(k)$ for all k. Also, the maximum total variation distance between the probability measures $P(\cdot)$ and $Q(\cdot)$ is 1 and this happens if an only if the measures are supported on disjoint subsets of $[K]$. The minimum, average and maximum total variation distance between MNL and MNP, MNL and SDP, SDP and MNP across the 100 instances are provided in Table 1. For additional numerical results in best-worst choice problems, the reader is referred to the supplemental material available online.

Table 1 Total variation distance between MNL, MNP and SDP choice probabilities for $n = 10$ correlated random variables.

Choice Probabilities	d(MNL,MNP)			d(MNL,SDP)			d(SDP,MNP)		
	Min	Average	Max	Min	Average	Max	Min	Average	Max
Best	0.1199	0.2902	0.4305	0.0998	0.2910	0.4833	0.0187	0.0511	0.0911
Worst	0.1450	0.3305	0.5687	0.1426	0.3085	0.5764	0.0217	0.0538	0.0890
Best-Worst	0.2962	0.4809	0.6365	0.4362	0.6000	0.7509	0.1940	0.2918	0.3871

6 Conclusions

In this paper, we have developed new semidefinite programs for computing second order moment bounds on the expected value of the maximum of linear functions. Our main result shows that computing the second order moment bound is related to the complexity of characterizing the convex hull of quadratic polynomials of points in the feasible region. For the case when the feasible region is given by a finite set of points, this leads to semidefinite programs that are smaller in size than existing semidefinite programming formulations. For the case when the feasible region is given as the feasible region to a mixed integer linear program, this implies that existing valid inequalities for deterministic integer programs can be applied to obtain tighter moment bounds. This leads to smaller size semidefinite programs to find moment bounds for the maximum clique problem on a bipartite graph with random

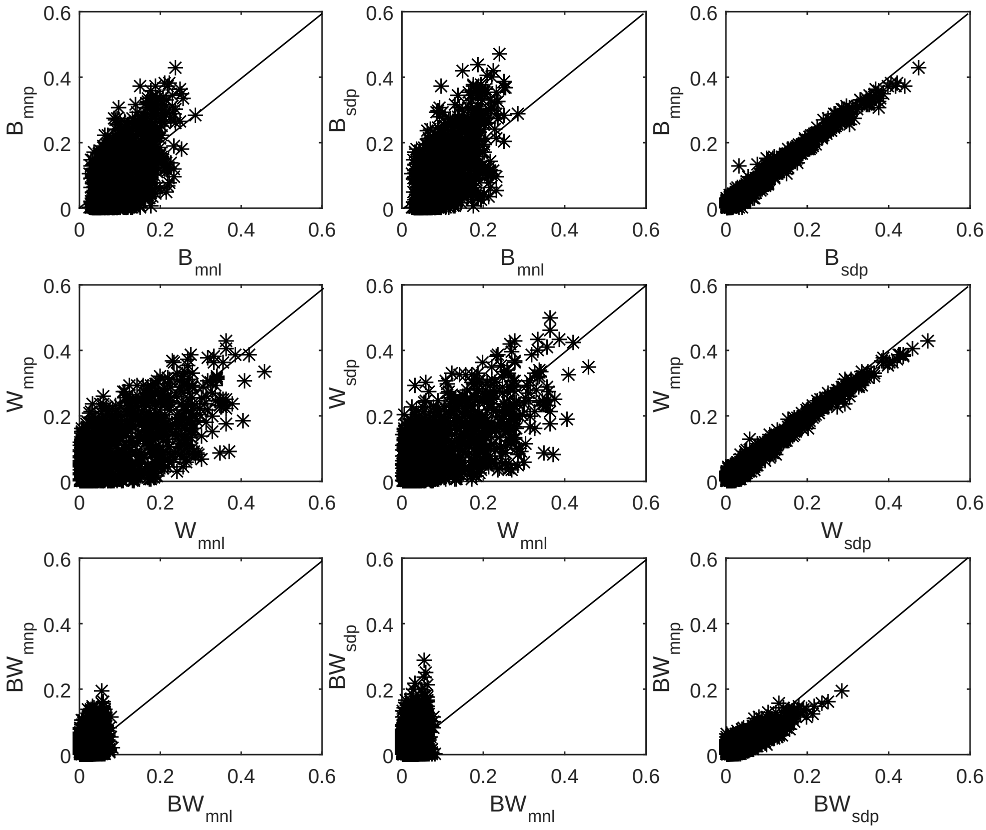

Fig. 5 Scatter plots of best, worst and best-worst choice probabilities using MNL, MNP and SDP for $n = 10$ correlated random variables.

weights. This insight also helps us identify a new polynomial time solvable relaxation of the distributionally robust newsvendor problem by exploiting results from the Boolean quadric polytope. We apply the semidefinite programs to applications in random walk problems and best-worst choice problems. The results indicate that the semidefinite programs capture correlation information, provide useful bounds and help identify important choice probabilities. Our approach uses semidefinite programming for these problems in contrast to a simulation approach.

There are some possible interesting avenues for further research. Testing the quality of cuts that have been developed in the deterministic integer quadratic programming literature and applying it to tighten the moment bounds is one such area. Theoretical and computational aspects of this approach need to be further studied in a variety of stochastic problems. Identifying instances of second order moment bounds that are computable in polynomial time is another area of important research. Lastly, incorporating additional information on the random variables such as the support to tighten the moment bounds under this approach needs to be further studied.

Acknowledgements The authors would like to thank two anonymous referees, the Associate Editor and the editor Alexander Shapiro for their very useful comments on the paper. The authors would also like to thank A. A. Marley and Zheng Zhichao for useful discussions on this topic.

References

1. Ahipasaoglu, S. D., L. Xiaobo, K. Natarajan. (2016). A convex optimization approach for computing correlated choice probabilities with many alternatives. *Under Review.*
2. Albert, A. (1969). Conditions for positive and nonnegative definiteness in terms of pseudoinverses. *SIAM Journal of Applied Mathematics*, 17(2), 434-440.
3. Arnold, B. C., R. A. Groeneveld. (1979). Bounds on expectations of linear systematic statistics based on dependent samples. *Mathematics of Operations Research*, 4(4), 441-447.
4. Anis, A. A., E. H. Lloyd. (1953). On the range of partial sums of a finite number of independent normal variates. *Biometrika*, 40(1-2), 35-42.
5. Aven, T. (1985). Upper (lower) bounds on the mean of the maximum (minimum) of a number of random variables. *Journal of Applied Probability*, 22, 723-728.
6. Ben-Akiva, M., S. R. Lerman. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge, MA: The MIT Press.
7. Ben-Tal, A., L. El. Ghaoui, A. Nemirovskii. (2009). Robust Optimization. Pricenton University Press.
8. Bertsimas, D., I. Popescu. (2002). On the relation between option and stock prices: An optimization approach. *Operations Research*, 50(2), 358-374.
9. Bertsimas, D., I. Popescu. (2005). Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal of Optimization*, 15(3), 780-804.
10. Bertsimas, D., X. V. Doan, K. Natarajan, C. P. Teo. (2010). Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3), 580-602.
11. Boyle, P., X. S. Lin. (1997). Bounds on contingent claims based on several assets. *Journal of Financial Economics*, 46, 383-400.
12. Burer, S. (2009). On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming (Series A)*, 120, 479-495.
13. Carlen, E. (2010). Trace inequalities and quantum entropy: an introductory course. Entropy and the Quantum. Contemporary Mathematics, Amer. Math. Soc., Providence, RI, 529, 73-140.
14. Delage, E., Y. Ye. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3), 596-612.
15. Deza, M. M., M. Laurent. (1997). Geometry of Cuts and Metrics. Springer, Berlin.
16. Dickinson, P. J. C. (2013). The Copositive Cone, the Completely Positive Cone and their Generalisations. PhD Thesis.
17. Dickinson, P. J. C., L. Gibjen. (2014). On the computational complexity of membership problems for the completely positive cone and its dual. *Computational Optimization and Applications*, 57(2), 403-415.
18. Dijk, B. van, D. Fok, R. Paap. (2012). A rank-ordered logit model with unobserved heterogeneity in ranking capabilities. *Journal of Applied Econometrics*, 27(5), 831-846.
19. Feller, W. (1951). The asymptotic distribution of the range of sums of independent random variables. *The Annals of Mathematical Statistics*, 22, 427-432.
20. Greene, W. H., D. A. Hensher. (2010). Modeling Ordered Choices: A Primer, Cambridge University Press.

21. Gumbel, E. J. (1954). The maximum of the mean largest value and of the range. *The Annals of Mathematical Statistics*, 25, 76-84.
22. Hanasusanto, G. A., D. Kuhn, S. W. Wallace, S. Zymler. (2015). Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming*, 152, 1-32.
23. Hartley, H. O., H. A. David. (1954). Universal bounds for mean range and extreme observations. *The Annals of Mathematical Statistics*, 25(1), 85-99.
24. Hurst, H. E. (1950). Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116, 770-799.
25. Isii, K. (1963). On the sharpness of Chebyshev-type inequalities. *The Annals of Mathematical Statistics*, 14, 185-197.
26. Karlin, S., W. J. Studden. (1966). Tchebycheff Systems: With Applications in Analysis and Statistics. Pure Appl. Math. 15, Interscience, John Wiley and Sons, New York.
27. Lasserre J. B. (2002). Bounds on measures satisfying moment conditions. *The Annals of Applied Probability*, 12(3), 1114-1137.
28. Lefevre, C. (1986). Bounds on the expectations of linear combinations of order statistics with applications to PERT networks. *Stochastic Analysis and Applications*, 4, 351-356.
29. Li, X., N. Karthik, C. P. Teo, Z. Zheng. (2014). Distributionally robust mixed integer linear programs: Persistency models with applications. *European Journal of Operational Research*, 233(3), 459-473.
30. Löfberg, J. (2004). YALMIP : A Toolbox for Modeling and Optimization in MATLAB. *In Proceedings of the CACSD Conference*, Taipei, Taiwan.
31. Luce, R. D., P. Suppes. (1965). Preference, utility and subjective probability. In R. D. Luce, R. R. Bush and E. Galanter (Eds.) *Handbook of Mathematical Psychology*, 3, 235-406, New York, NY: Wiley.
32. Marley, A. A. J., J. J. Louviere. (2005). Some probabilistic models of best, worst and best-worst choices. *Journal of Mathematical Psychology*, 49, 464-480.
33. Marshall, A. W., I. Olkin. (1960). A one-sided inequality of the Chebyshev type. *The Annals of Mathematical Statistics*, 31(2), 488-491.
34. Marshall, A. W., I. Olkin. (1960). Multivariate chebyshev inequalities. *The Annals of Mathematical Statistics*, 31(4), 1001-1014.
35. McFadden D. (1974). Conditional logit analysis of qualitative choice behavior. P. Zarembka , ed. *Frontier in Econometrics*, Academic Press, New York, 105-142.
36. Mesa, O. J., G. Poveda. (1993). The Hurst Effect: The scale of fluctuation approach. *Water Resources Research*, 29(12), 3995-4002.
37. Mishra, V. K., K. Natarajan, H. Tao, C. P. Teo. (2012). Choice modeling with semidefinite optimization when utilities are correlated. *IEEE Transactions on Automatic Control*, 57(10), 2450-2463.
38. Murty, K. G., S. N. Kabadi. (1987). Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2), 117129.
39. Nagaraja, H. N. (1981). Some finite sample results for the selection differential. *Annals of the Institute of Statistical Mathematics*, 33, 437-448.
40. Natarajan, K., C. P. Teo, Z. Zheng. (2011). Mixed zero-one linear programs under objective uncertainty: A completely positive representation. *Operations Research*, 59(3), 713-728.
41. Nesterov, Y. (2000). Squared functional systems and optimization problems. in *High Performance Optimization,*, H. Frenk et al., eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 405440.
42. Olkin, I., J. W. Pratt. (1958). A multivariate Tchebycheff inequality. *The Annals of Mathematical Statistics*, 29(1), 226-234.
43. Padberg, M. (1989). The boolean quadric polytope: Some characteristics, facets and relatives. *Mathematical Programming*, 45, 139-172.
44. Palma, A. de, K. Kilani, G. Laffond. (2013). Best and worst choices. Working Paper.
45. Papdatos, N. (2001). Expectation bounds on linear estimators from dependent samples. *Journal of Statistical Planning and inference*, 93, 17-28.
46. Plackett, R. L. (1947). Limits of the ratio of mean range to standard deviation. *Biometrika*, 34(1-2), 120-122.
47. Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, **51**, 406-413.

48. Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization and applications. *Operations Research*, 55(1), 98-112.
49. Rao, C. R., S. K. Mishra. (1988). Generalized inverse of a matrix and its applications. *Proceedings of the Cambridge Philosophical Society*, 103, 269-276.
50. Shapiro, A. (1985). Extremal problems on the set of nonnegative definite matrices. *Linear Algebra and its Applications*, 67, 7-18.
51. Salce, L., P. Zanardo. (1993). Completely positive matrices and positivity of least squares solutions. *Linear Algebra and its Applications*, 178, 201-216.
52. Toh, K. C., M. J. Todd, R. H. Tutuncu. (1999). SDPT3 - A Matlab software package for semidefinite programming. *Optimization Methods and Software*, 11, 545-581.
53. Tutuncu, R. H., K. C. Toh, M. J. Todd. (2003). Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical Programming Series B*, 95, 189-217.
54. Vandenberghe, L., S. Boyd, K. Comanor. (2007). Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49, 52-64.
55. Van Ophem, H. P. Stam, B. van Praag. (1999). Multichoice logit: Modeling incomplete preference rankings of classical concerts. *Journal of Business and Economics Statistics*, 17, 117-128.
56. Yu, Y-L., Y. Li, D. Schuurmans, C. Szepesvári. (2009). A general projection property for distribution families. *Advances in Neural Information Processing Systems*, 22.
57. Zymler, S., D. Kuhn, B. Rustem. (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2), 167-198.