Invited Review

Distributionally Robust Mixed Integer Linear Programs: Persistency Models with Applications

Xiaobo Li[*] Karthik Natarajan[†] Chung-Piaw Teo[‡] Zhichao Zheng[§]

Abstract

In this paper, we review recent advances in the distributional analysis of mixed integer linear programs with random objective coefficients. Suppose that the probability distribution of the objective coefficients is incompletely specified and characterized through partial moment information. Conic programming methods have been recently used to find distributionally robust bounds for the expected optimal value of mixed integer linear programs over the set of all distributions with the given moment information. These methods also provide additional information on the probability that a binary variable attains a value of 1 in the optimal solution for 0-1 integer linear programs. This probability is defined as the persistency of a binary variable. In this paper, we provide an overview of the complexity results for these models, conic programming formulations that are readily implementable with standard solvers and important applications of persistency models. The main message that we hope to convey through this review is that tools of conic programming provide important insights in the probabilistic analysis of discrete optimization problems. These tools lead to distributionally robust bounds with applications in activity networks, vertex packing, discrete choice models, random walks and sequencing problems, and newsvendor problems.

[*]Engineering Systems and Design, Singapore University of Technology and Design, Singapore 138682. Email: li xiaobo@sutd.edu.sg

[†]Engineering Systems and Design, Singapore University of Technology and Design, Singapore 138682. Email: natarajan_karthik@sutd.edu.sg

[‡]Department of Decision Sciences, NUS Business School, Singapore 117591. Email: bizteocp@nus.edu.sg

[§]Lee Kong Chian School of Business, Singapore Management University, Singapore 178899. Email: danielzheng@smu.edu.sg

1 Introduction

Consider a mixed 0-1 linear program (LP) in maximization form,

$$Z(\boldsymbol{c}) = \max\left\{\boldsymbol{c}^T\boldsymbol{x} \;:\; \boldsymbol{x} \in \boldsymbol{\mathcal{X}}\right\}, \tag{1}$$

where the feasible region is described as

$$\boldsymbol{\mathcal{X}} = \left\{\boldsymbol{x} \in \Re_n^+ : \boldsymbol{a}_j^T\boldsymbol{x} = b_j,\ \forall j = 1, \ldots, m;\ x_i \in \{0,1\},\ \forall i \in \mathcal{B} \subseteq \{1, \ldots, n\}\right\}. \tag{2}$$

The set of decision variables $\{1, \ldots, n\}$ includes 0-1 decision variables indexed by the set $\mathcal{B}$ and nonnegative decision variables indexed by $\{1, \ldots, n\} \backslash \mathcal{B}$. The class of mixed 0-1 linear programs has been used extensively in business, engineering, and economic applications, to model diverse types of problems arising from production planning, logistics deployment, scheduling of jobs and machines, among others. However, in practice, the input parameters (i.e. $\boldsymbol{c}$, $\boldsymbol{a}_j$, b_j) are often not known with certainty, and need to be estimated in the modeling process. Since the optimal solution is sometimes very sensitive to the input parameters, one needs to be careful in modeling the uncertainty in these problems. In this review, we focus ourselves on the uncertainty inside the objective coefficient vector, $\boldsymbol{c}$.

1.1 Probabilistic analysis of mixed 0-1 linear programs

Formally, the problem of interest is:

> Given the mixed 0-1 linear program in (1) and a probability measure θ for the random objective coefficient vector $\boldsymbol{c}$, compute the expected optimal value, i.e.,
>
> $$(\textbf{MEAN}) \qquad E_\theta\Big(Z(\boldsymbol{c})\Big) = \int Z(\boldsymbol{c})d\theta(\boldsymbol{c}).$$

Clearly, computing **MEAN** is at least as hard as solving the deterministic mixed 0-1 linear program. **MEAN** is computable in time polynomial in the size of the instance[1] when the deterministic problem is solvable in polynomial time and θ is a discrete distribution with a polynomial number of support points. However, for general distributions, the computation of **MEAN** is significantly more challenging than solving the deterministic problem. An example is the problem of finding the longest path on a directed acyclic graph. The deterministic version of this problem is to find a longest path between a source node s and a sink node t in a directed acyclic graph $G(\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of vertices, $\mathcal{E}$ is the set of edges

[1] Here, the "instance" refers to the problem input string that encodes all the necessary parameters of the optimization problem, $Z(\boldsymbol{c})$, and all the support points of $\boldsymbol{c}$.

with associated arc lengths c_{ij} for each arc $(i,j) \in \mathcal{E}$. The longest path problem (LPP) is formulated as the following 0-1 integer linear program,

$$Z_{LPP}(\boldsymbol{c}) = \max \left\{ \sum_{(i,j)\in\mathcal{E}} c_{ij}x_{ij} \ : \ \sum_{j:(i,j)\in\mathcal{E}} x_{ij} - \sum_{j:(j,i)\in\mathcal{E}} x_{ji} = b_i, \ \forall i \in \mathcal{V}; \ x_{ij} \in \{0,1\}, \ \forall (i,j) \in \mathcal{E} \right\},$$

where b_i is defined to be 1 for $i = s$, -1 for $i = t$ and 0 otherwise. For a fixed $\boldsymbol{c}$, $Z_{LPP}(\boldsymbol{c})$ is computable in polynomial time. The linear programming relaxation solves the integer program in this case. The complexity of computing $E(Z_{LPP}(\boldsymbol{c}))$ for independent discrete distributions was resolved by Hagstrom [28].

Theorem 1 *(Hagstrom [28]) For a directed acyclic graph with arc lengths that are independently distributed and restricted to taking two possible values each, computing the expected value of the longest path is $\#\mathcal{P}$-complete. Furthermore it cannot be computed in time polynomial in the number of points in the range of the longest path unless $\mathcal{P} = \mathcal{NP}$.*

Roughly speaking, the difficulty stems from the observation that an exponential number of support points for the multivariate independent discrete distribution can be supported by an exponential number of optimal solutions. The difficulty of this problem has led to the development of methods such as Monte Carlo simulations [73, 15], PERT approximations [50], exact methods in special instances such as series-parallel graphs [56] and upper and lower bounds on the expected longest path [25, 22, 38].

Explicit formulas for the expected optimal value have been developed in the asymptotic analysis of random combinatorial optimization problems. This area has its roots in the pioneering work of Beardwood et al. [5] who characterized the asymptotic behavior of a traveling salesperson tour length for points randomly generated on the Euclidean plane. The book by Steele [70] provides an introduction to the asymptotic analysis of combinatorial optimization problems under the *Euclidean model*. More closely related to the theme of this paper is the probabilistic analysis of combinatorial optimization problems under the *mean field model*. In the mean field model, the nodes of a graph are assumed to be fixed while the arc lengths are independently chosen from a probability distribution. The arc lengths need not satisfy the triangle inequality. The origins of this model lies in an early paper of Karp [35] who analyzed the asymmetric traveling salesperson problem and its linear assignment relaxation for distances drawn independently from an uniform distribution in $[0,1]$. The linear assignment problem (LAP) is formulated as

$$Z_{LAP}(\boldsymbol{c}) = \min \left\{ \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij} \ : \ \begin{array}{l} \sum_{i=1}^{n} x_{ij} = 1, \ \forall j = 1,\ldots,n; \ \sum_{j=1}^{n} x_{ij} = 1, \ \forall i = 1,\ldots,n; \\ x_{ij} \in \{0,1\}, \ \forall i,j = 1,\ldots,n \end{array} \right\}.$$

Explicit asymptotic expressions for the expected value of the assignment problem under random costs have been developed for uniform and exponential distributions by several authors (see Krokhmal and Pardalos [39] for a review on this topic). Aldous [2] in 2001 rigorously proved the following result which was initially conjectured by Mézard and Parisi [54] in 1985 using ideas in statistical physics.

Theorem 2 *(Aldous [2]) Let the random variables c_{ij} for $i, j = 1, \ldots, n$ be independent with uniform distribution on [0,1] or exponentially distributed with parameter 1. Then*

$$\lim_{n\to\infty} E\Big(Z_{LAP}(\boldsymbol{c})\Big) = \zeta(2) = \frac{\pi^2}{6} \approx 1.645.$$

Asymptotic expressions for the expected value of other combinatorial optimization problems have since been developed under the mean-field model (see Aldous [3]). However, explicit formulas for finite size instances of combinatorial optimization problems are much more difficult to obtain. One such formula that was conjectured by Parisi in 1998 [64] and recently proved by two sets of authors is provided next.

Theorem 3 *(Linusson and Wästlund [46], Nair, Prabhakar, and Sharma [59]) Let the random variables c_{ij} for $i, j = 1, \ldots, n$ be independent and exponentially distributed with parameter 1. Then*

$$E\Big(Z_{LAP}(\boldsymbol{c})\Big) = \sum_{i=1}^{n} \frac{1}{i^2}.$$

While this result is elegant and surprising, the techniques in its proof is typically difficult to use for the analysis of general discrete optimization problems with non-identical, or non-independently distributed uncertainties. In this paper, we review an alternate approach for the probabilistic analysis of discrete optimization problems that relaxes the assumption of independence. Instead of fixing the joint distribution of the random parameters, we allow for the joint distribution to be incompletely specified by the partial moment information. Surprisingly, the probabilistic analysis of the problem becomes tractable for a wide class of mixed integer linear programs, under appropriate assumptions of the input distributions. The moment information can be viewed as imposing boundary conditions on the way the scenarios of the extremal distribution are generated so as to guard against extremely unrealistic distributions.

1.2 Organization of the paper

In Section 2, the central problem of finding the tight bound on the expected optimal value of a mixed 0-1 linear program is introduced. The formal characterization of the set of probability distributions using the theory of moments is provided in this section. The notion of persistency - the probability that a binary variable attains a value of 1 in the optimal solution is also discussed in this section. In Section 3, conic programming methods to compute the bounds and estimate persistency are reviewed. Polynomial time computable bounds and bounds that are $\mathcal{NP}$-hard to compute are identified in this section. In Section 4, we review applications of this approach in activity networks, vertex packing discrete choice models, random walk and sequencing problems, and newsvendor problems. We conclude in Section 5.

1.3 Notation

Throughout the paper, standard letters such as x denote scalars, bold letters such as $\boldsymbol{x}$ denote vectors, bold capital letters such as $\boldsymbol{X}$ denote matrices and calligraphic fonts such as $\boldsymbol{\mathcal{X}}$ denotes sets. The notation c^+ represents $\max(0, c)$. The transpose of a column vector $\boldsymbol{c}$ is denoted as $\boldsymbol{c}^T$. For two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of dimension n, $\boldsymbol{x}^T\boldsymbol{y}$ denotes $x_1y_1 + \ldots + x_ny_n$, and $\boldsymbol{x} \circ \boldsymbol{y}$ denotes the column vector $(x_1y_1, \ldots, x_ny_n)^T$. The set $\mathcal{S}_n$ denotes the set of $n \times n$ symmetric matrices equipped with the standard inner product $\boldsymbol{A} \cdot \boldsymbol{B} = \sum_{i=1}^n \sum_{j=1}^n A_{ij}B_{ij}$. The vector $\boldsymbol{e}$ has all components equal to one, and the vector $\boldsymbol{e_i}$ has 1 in its ith component and 0 otherwise. The cone of $n \times n$ nonnegative matrices is defined as $\mathcal{N}_n = \{\boldsymbol{A} \in \mathcal{S}_n : \boldsymbol{A} \geq 0\}$. The cone of $n \times n$ positive semidefinite matrices is defined as $\mathcal{S}_n^+ = \left\{\boldsymbol{A} \in \mathcal{S}_n : \boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v} \geq 0, \ \forall \boldsymbol{v} \in \Re_n\right\}$. The cone of $n \times n$ copositive matrices is defined as $\mathcal{CO}_n = \left\{\boldsymbol{A} \in \mathcal{S}_n : \boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v} \geq 0, \ \forall \boldsymbol{v} \in \Re_n^+\right\}$. The cone of $n \times n$ completely positive matrices is defined as $\mathcal{CP}_n = \left\{\boldsymbol{A} \in \mathcal{S}_n : \exists \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k \in \Re_n^+ \text{ such that } \boldsymbol{A} = \sum_{i=1}^k \boldsymbol{v}_i\boldsymbol{v}_i^T\right\}$. For a cone $\mathcal{K}$, the dual cone is defined as $\mathcal{K}^* = \{\boldsymbol{A} : \boldsymbol{A} \cdot \boldsymbol{B} \geq 0, \ \forall \boldsymbol{B} \in \mathcal{K}\}$. The nonnegative and positive semidefinite cones are self-dual while the copositive and completely positive cones are duals of each other. Namely, $(\mathcal{N}_n)^* = \mathcal{N}_n$, $(\mathcal{S}_n^+)^* = \mathcal{S}_n^+$ and $(\mathcal{CO}_n)^* = \mathcal{CP}_n$. The closure of a cone $\mathcal{K}$ is denoted as $\overline{\mathcal{K}}$ and the interior is denoted as $\text{Int}(\mathcal{K})$. The convex hull of a set $\mathcal{K}$ is given as $\text{conv}(\mathcal{K})$.

2 Bounds on the expected optimal value and persistency

Suppose the exact probability distribution θ of the parameters is unknown. Rather, θ is only known to lie in the set of probability distributions Θ. Formally, the central problem of interest is:

> Given the mixed 0-1 linear program in maximization form in (1) and a nonempty set of probability measures Θ for the random objective coefficient vector $\boldsymbol{c}$, compute the tightest upper bound on the expected optimal value, i.e.,

$$\textbf{(MEAN-UB)} \qquad Z = \sup_{\theta \in \Theta} E_\theta\Big(Z(\boldsymbol{c})\Big) = \sup_{\theta \in \Theta} \int Z(\boldsymbol{c}) d\theta(\boldsymbol{c}).$$

We discuss a few important aspects of this problem next:

(a) The most commonly used model in the probabilistic analysis of optimization problems is the independent distribution model. However, modeling data as independent random variables is sometimes unrealistic. For example, in supply chain networks one often needs to deal with correlated demands, and in activity networks one needs to deal with correlated activity durations due to resource dependencies. By dropping the explicit assumption of independence in the description of Θ, it is possible to capture the effect of dependencies.

(b) The upper bound of interest is valid across all distributions in the set Θ and is as tight as possible. Tightness implies that either there exists a feasible distribution that attains the upper bound exactly or there exists a sequence of distributions that attains the bound and is feasible in a limiting sense. Hence, this bound is termed as a *distributionally robust bound.* For the longest path problem on directed acyclic graphs with random arc lengths as arising in activity networks, the upper bound corresponds to a worst-case expected project completion time. For the maximum flow problem with random arc capacities, the upper bound corresponds to the worst-case expected maximum flow that is supported by the network. While **MEAN-UB** deals with upper bounds for maximization problems, by simply replacing $\boldsymbol{c}$ with $-\boldsymbol{c}$ transforms the bounds to lower bounds on minimization problems.

(d) Under reasonable assumptions on the set of distributions Θ, the tight bound turns out to be efficiently computable with convex conic programming. In instances where the tight bound is not efficiently computable, conic programming relaxations provide weaker upper bounds. An example of a bound on the expected optimal value of a discrete optimization problem through convex quadratic programming is the following elegant result of Lyons, Pemantle and Peres [49] and Lovász [47].

Theorem 4 *(Lyons, Pemantle, and Peres [49], Lovász [47]) Let Θ be the set of distributions for a nonnegative random vector $\boldsymbol{c}$ whose joint survival function $S(t_1, \ldots, t_n) = P(c_1 \geq t_1, \ldots, c_n \geq t_n)$ is log-concave.*

(a) Suppose $\boldsymbol{\mathcal{X}}$ is the feasible region for a shortest s-t path problem, minimum s-t cut problem or minimum linear assignment problem. Then

$$\inf_{\theta \in \Theta} E_\theta \left(\min_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}} \boldsymbol{c}^T \boldsymbol{x} \right) \geq \min_{\boldsymbol{x} \in conv(\boldsymbol{\mathcal{X}})} \left(\sum_{i=1}^n E(c_i) x_i^2 \right).$$

(b) The lower bound in (a) is tight for the set $\boldsymbol{\mathcal{X}} = \left\{ \boldsymbol{x} \in \Re_n^+ : \boldsymbol{e}^T \boldsymbol{x} = 1, x_i \in \{0,1\} \ \forall i = 1, \ldots, n \right\}$. The bound is attained by independent exponential random variables, i.e.,

$$\min_{\theta \in \Theta} E_\theta \left(\min_{i=1,\ldots,n} c_i \right) = \left(\sum_{i=1}^n \frac{1}{E(c_i)} \right)^{-1}.$$

The bound in Theorem 4(a) is the optimal value to a separable convex quadratic minimization problem over 0-1 polytopes and efficiently computable. This result implies that for independent exponentially distributed random edge lengths in an undirected graph, the expected length of the shortest path between any two nodes is bounded from below by the resistance between the nodes, where the resistance of an edge is defined as the expectation of its length. While the lower bound in Theorem 4(a) is not tight in general (see Example 5, Page 372 in [47] for a counterexample), Theorem 4(b) identifies a particular instance where the bound is tight. The models discussed in Section 3 provide tight bounds for general mixed integer linear programs using conic optimization. A formal description of the set of distributions Θ based on the theory of moments is provided next.

2.1 Moment representation of the set of distributions Θ

A simple and popular characterization of the set of distributions Θ is based on only the first two moments - the mean and covariance matrix. Let $\mathbf{\Omega}$ be a given subset of $\Re_n$. Define $\mathbb{M}(\mathbf{\Omega})_+$ to be the space of finite positive Borel measures defined on the domain $\mathbf{\Omega}$. The set of probability measures with first moment vector $\boldsymbol{\mu}$ and second moment matrix $\mathbf{\Pi}$ is defined as

$$\Theta = \left\{\theta \in \mathbb{M}(\mathbf{\Omega})_+ \ : \ 1 = \int_{\mathbf{\Omega}} d\theta(\boldsymbol{c}), \ \boldsymbol{\mu} = \int_{\mathbf{\Omega}} \boldsymbol{c} d\theta(\boldsymbol{c}), \ \mathbf{\Pi} = \int_{\mathbf{\Omega}} \boldsymbol{c}\boldsymbol{c}^T d\theta(\boldsymbol{c}) \right\}. \quad (3)$$

Under this description, the distributional robust bound is found by solving the *generalized moment problem*,

$$\begin{array}{rrll} \textbf{(MEAN-UB)} \quad Z & = & \displaystyle\sup_{\theta \in \mathbb{M}(\mathbf{\Omega})_+} & \displaystyle\int_{\mathbf{\Omega}} Z(\boldsymbol{c}) d\theta(\boldsymbol{c}) \\ & & \text{s.t.} & \displaystyle\int_{\mathbf{\Omega}} d\theta(\boldsymbol{c}) = 1 \\ & & & \displaystyle\int_{\mathbf{\Omega}} \boldsymbol{c} d\theta(\boldsymbol{c}) = \boldsymbol{\mu} \\ & & & \displaystyle\int_{\mathbf{\Omega}} \boldsymbol{c}\boldsymbol{c}^T d\theta(\boldsymbol{c}) = \mathbf{\Pi}. \end{array} \quad (4)$$

For an in-depth discussion of the moment problem, the reader is referred to the classic book of Karlin and Studden [34]. A more recent algorithmic exposition on the moment problem using conic programming is found in the book of Lasserre [43]. We review the results from this theory of moments that are particularly relevant to this paper.

The moment feasibility problem characterizes necessary and sufficient conditions that the moments must satisfy so that the set of distributions Θ as defined in (3) is nonempty. Towards this, define the moment cone of order 2 supported on $\mathbf{\Omega} \subseteq \Re_n$ as[2]

$$\mathcal{M}_2(\mathbf{\Omega}) = \left\{ \lambda \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \mathbf{\Pi} \end{pmatrix} \in \mathcal{S}_{n+1} \; : \; \begin{array}{l} \lambda \geq 0; \; 1 = \int_{\mathbf{\Omega}} d\theta(\boldsymbol{c}), \; \boldsymbol{\mu} = \int_{\mathbf{\Omega}} \boldsymbol{c} d\theta(\boldsymbol{c}), \\ \mathbf{\Pi} = \int_{\mathbf{\Omega}} \boldsymbol{c}\boldsymbol{c}^T d\theta(\boldsymbol{c}), \text{ for some } \theta \in \mathbb{M}(\mathbf{\Omega})_+ \end{array} \right\}.$$

From the theory of moments, the dual of this moment cone is the cone of all non-homogeneous quadratic polynomials that is nonnegative over $\mathbf{\Omega}$ and defined by

$$\mathcal{P}_2(\mathbf{\Omega}) \;=\; \mathcal{M}_2(\mathbf{\Omega})^* \;=\; \left\{ \begin{pmatrix} w_0 & \boldsymbol{w}^T/2 \\ \boldsymbol{w}/2 & \boldsymbol{W} \end{pmatrix} \in \mathcal{S}_{n+1} \; : \; w_0 + \boldsymbol{w}^T \boldsymbol{c} + \boldsymbol{c}^T \boldsymbol{W} \boldsymbol{c} \geq 0, \; \forall \; \boldsymbol{c} \in \mathbf{\Omega} \right\}.$$

The dual of the cone of nonnegative polynomials is the closure of the moment cone, namely $\mathcal{P}_2(\mathbf{\Omega})^* = \overline{\mathcal{M}_2(\mathbf{\Omega})}$. The dual of the moment problem in (4) is hence formulated as:

$$\begin{aligned} Z_D \;=\; \inf_{w_o, \boldsymbol{w}, \boldsymbol{W}} \quad & w_0 + \boldsymbol{w}^T \boldsymbol{\mu} + \boldsymbol{W} \cdot \mathbf{\Pi} \\ \text{s.t.} \quad & w_0 + \boldsymbol{w}^T \boldsymbol{c} + \boldsymbol{c}^T \boldsymbol{W} \boldsymbol{c} \geq Z(\boldsymbol{c}) \quad \forall \boldsymbol{c} \in \mathbf{\Omega}, \end{aligned} \tag{5}$$

where $w_0 \in \Re$, $\boldsymbol{w} \in \Re_n$ and $\boldsymbol{W} \in \mathcal{S}_n$. Problems (4) and (5) are related through conic duality.

Theorem 5 *(Karlin and Studden [34], Lasserre [43])*

(a) Weak duality: The optimal primal and dual objective satisfy $Z \leq Z_D$.

(b) Strong duality: If the moments lie in the interior of the moment cone, i.e.,

$$\begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \mathbf{\Pi} \end{pmatrix} \in Int\left(\mathcal{M}_2(\mathbf{\Omega})\right),$$

then the optimal primal and dual objectives are equal and satisfy $Z = Z_D$.

Alternate conditions to guarantee strong duality for the moment problem have also been identified in the literature (see Zuluaga and Pena [76] and Shapiro [68] for some examples). The primary method to compute Z and Z_D uses convex conic programming techniques such as linear and semidefinite programming. The connection between the cone of moments, the cone of nonnegative quadratic polynomials and the semidefinite cone is provided in the next theorem for the domain $\mathbf{\Omega} = \Re_n$ and $\mathbf{\Omega} = \Re_n^+$.

Theorem 6 *(Karlin and Studden [34], Kemperman [36])*

[2]The definition of the moment cone based on symmetric matrices in $\mathcal{S}_{n+1}$ is a slight modification of the definition from the literature (cf. [34, 43]) that uses vector notation in $\Re_{(n+1)(n+2)/2}$.

(a) For $\mathbf{\Omega} = \Re_n$, the following cones are equivalent,

$$\begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \mathbf{\Pi} \end{pmatrix} \in \overline{\mathcal{M}_2(\Re_n)} \Longleftrightarrow \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \mathbf{\Pi} \end{pmatrix} \in \mathcal{S}^+_{n+1},$$

$$\begin{pmatrix} w_0 & \boldsymbol{w}^T/2 \\ \boldsymbol{w}/2 & \boldsymbol{W} \end{pmatrix} \in \mathcal{P}_2(\Re_n) \Longleftrightarrow \begin{pmatrix} w_0 & \boldsymbol{w}^T/2 \\ \boldsymbol{w}/2 & \boldsymbol{W} \end{pmatrix} \in \mathcal{S}^+_{n+1}.$$

(b) For $\mathbf{\Omega} = \Re^+_n$, the following cones are equivalent,

$$\begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \mathbf{\Pi} \end{pmatrix} \in \overline{\mathcal{M}_2(\Re^+_n)} \Longleftrightarrow \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \mathbf{\Pi} \end{pmatrix} \in \mathcal{CP}_{n+1},$$

$$\begin{pmatrix} w_0 & \boldsymbol{w}^T/2 \\ \boldsymbol{w}/2 & \boldsymbol{W} \end{pmatrix} \in \mathcal{P}_2(\Re^+_n) \Longleftrightarrow \begin{pmatrix} w_0 & \boldsymbol{w}^T/2 \\ \boldsymbol{w}/2 & \boldsymbol{W} \end{pmatrix} \in \mathcal{CO}_{n+1}.$$

Theorem 6(a) is from Karlin and Studden [34]. Essentially, for $\mathbf{\Omega} = \Re_n$, testing feasibility in the cones $\overline{\mathcal{M}_2(\mathbf{\Omega})}$ and $\mathcal{P}_2(\mathbf{\Omega})$ are easy since these are equivalent to testing the positive semidefiniteness of a symmetric matrix. Theorem 6(b) is from Kemperman [36]. For $\mathbf{\Omega} = \Re^+_n$, testing feasibility in $\overline{\mathcal{M}_2(\mathbf{\Omega})}$ is equivalent to verifying if a matrix is completely positive, and testing feasibility in $\mathcal{P}_2(\mathbf{\Omega})$ is equivalent to verifying if a matrix is copositive. For a detailed introduction to completely positive matrices, the reader is referred to the book of Berman and Shaked-Monderer [6]. Three recent surveys on completely positive and copositive matrices with emphasis on optimization are found in Bomze [13], Dür [24] and Hiriart-Urruty and Seeger [31]. Unlike the positive semidefinite cone, testing feasibility in the completely positive and copositive cones are known to be difficult. For instance, Murty and Kabadi [58] showed that the problem of verifying if a given matrix is copositive is co-$\mathcal{NP}$-complete. A popular relaxation to the completely positive cone is the doubly nonnegative cone which is defined as the intersection of the positive semidefinite and the nonnegative cone. The following well-known relationship holds among these cones,

$$\mathcal{CP}_n \subseteq \mathcal{S}^+_n \cap \mathcal{N}_n \subset \mathcal{S}^+_n + \mathcal{N}_n \subseteq \mathcal{CO}_n.$$

Equality holds for the leftmost and rightmost inclusion only for $n \leq 4$ from a result of Diananda [19]. To obtain better approximations to the completely positive and copositive cones, hierarchies of convex cones based on positive semidefinite and nonnegative cones have been developed by several researchers (see Parillo [63], Lasserre [41], Bomze and de Klerk [14], Laurent [45], Zuluaga and Pena [76] for details on these hierarchies). While these hierarchies of cones converge to the completely positive and copositive cones asymptotically, the size of the formulations grow so rapidly that the higher order approximations are intractable from a computational viewpoint. The conic representations in Theorem 6 are extendable to arbitrary sets $\mathbf{\Omega} \subseteq \Re_n$ by using a generalized notion of complete positivity and copositivity over the domain $\mathbf{\Omega}$ (see Kemperman [36], Sturm and Zhang [71]).

2.2 Persistency in combinatorial optimization

A related parameter in the analysis of optimization problems under uncertainty is the distribution of the optimal decision vector. Define the mapping $\boldsymbol{\mathcal{X}}_{opt}(\boldsymbol{c})$ as the set of all optimal solutions for a given $\boldsymbol{c}$,

$$\boldsymbol{\mathcal{X}}_{opt}(\boldsymbol{c}) = \left\{ \boldsymbol{x} \in \boldsymbol{\mathcal{X}} \ : \ \boldsymbol{c}^T\boldsymbol{x} \geq \boldsymbol{c}^T\boldsymbol{y}, \ \forall \boldsymbol{y} \in \boldsymbol{\mathcal{X}} \right\}. \tag{6}$$

The number of solutions in $\boldsymbol{\mathcal{X}}_{opt}(\boldsymbol{c})$ could be one or many depending on the uniqueness of the optimal solution. For continuous distributions such as the multivariate normal distribution with a positive definite covariance matrix, the probability measure of the support over which there are multiple optimal solutions is zero. This however need not be the case for discrete distributions. For a random vector $\boldsymbol{c}$, any optimal solution $\boldsymbol{x}(\boldsymbol{c}) \in \boldsymbol{\mathcal{X}}_{opt}(\boldsymbol{c})$ is a random vector. Bertsimas et al. [10] proposed a definition of persistency for combinatorial optimization problems based on the components of the random vector $\boldsymbol{x}(\boldsymbol{c})$.

Definition 1 *(Bertsimas, Natarajan, and Teo [10]) The persistency of the binary variable x_i is defined as the probability that x_i takes the value of 1 in some optimal solution, i.e.,*

$$\begin{aligned} \textit{Persistency of binary variable } x_i &= P\left(x_i(\boldsymbol{c}) = 1, \ \textit{for some } \boldsymbol{x}(\boldsymbol{c}) \in \boldsymbol{\mathcal{X}}_{opt}(\boldsymbol{c}) \right) \\ &= P\left(Z(\boldsymbol{c}) = \max_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}} : x_i = 1} \boldsymbol{c}^T\boldsymbol{x} \right). \end{aligned}$$

For activity networks, “persistency” is equivalent to the concept of the “criticality index” of an activity. A criticality index of an activity measures the probability that an activity is on the longest path in an activity network. A higher criticality index roughly indicates higher importance of that activity in the successful completion of the project on time. Criticality indices have been previously estimated using simulations [73, 15] and approximations [23]. The conic optimization methods discussed in the next section provide an alternate method to estimate the persistency of a binary variable.

It is important to point out that an alternate definition of “persistency” for deterministic 0-1 optimization problems has been previously proposed. Adams et al. [1] and Hammer [29] define an optimal solution to the continuous relaxation of a mixed 0-1 linear program to be persistent if the set of 0-1 variables realizing binary values in the continuous relaxation retain those same binary values in at least one integer optimum. A mixed 0-1 linear program possesses the persistency property if every optimal solution to the continuous relaxation is a persistent solution. Weighted vertex packing is the first example of a combinatorial optimization problem that is shown to possess the persistency property (see Nemhauser and Trotter [62]). The persistency property has also been identified in unconstrained quadratic 0-1 optimization by Hammer et al. [29], unconstrained polynomial 0-1 optimization by Lu and Williams [48], and integer programs with at most two variables per inequality by Hochbaum et al. [32]. The motivation of identifying the persistency property in this line of research is to reduce the computational time to solve deterministic mixed 0-1 linear programs. This is done through a preprocessing step that solves the continuous relaxation first and fixes the persistent variables to their respective binary values. The reduction in the search space then provides computational benefits in the second step where either the optimal solution or near-optimal solution is found through exact methods or heuristics. Definition 1 of persistency has a similar motivation of helping identify variables that often take a value of 1 in the optimal solution. The main distinction is that while Definition 1 is for the stochastic 0-1 optimization problem, the earlier definition was for the deterministic 0-1 optimization problem.

3 Conic optimization based methods

There is a vast literature on moment bounds in areas such as inventory control [67], queueing systems [75], finance [16, 11], decision theory [69], activity networks [12] and probability and statistics [52, 42, 72]. In this section, we review conic optimization based methods to compute **MEAN-UB** and the corresponding complexity of obtaining the bounds. The following generic "primal" proof technique is adopted in developing the conic programs to compute **MEAN-UB**:

(a) Define the "appropriate" decision variables using moments of the objective coefficient vector $\boldsymbol{c}$ and the optimal solution vector $\boldsymbol{x}(\boldsymbol{c})$.

(b) Identify necessary constraints that these variables must satisfy for all distributions in Θ. Express the objective function in terms of the decision variables and find a upper bound on the expected value of $Z(\boldsymbol{c})$ using conic optimization.

(c) Show that the constraints are sufficient by constructing a distribution (or a sequence of distributions) in Θ that attains the upper bound (in a limiting sense). We outline steps (a) and (b) in this section to provide a flavor of the proof technique. The tightness argument of step (c) is found in the specific references. It is also possible to derive these results through an alternative dual approach (see [34, 52, 11]).

To develop the conic programs to compute **MEAN-UB**, we make use of three different moment representations for the set of distributions Θ:

(a) **Cross moments:** Given cross moment information that includes the means and covariances, we show that the distributionally robust bound is $\mathcal{NP}$-hard in general. We provide two different conic programming approaches in Section 3.1. The first approach uses a complete enumeration of the extreme points to construct the SDP formulation, whereas the second approach uses the constraint formulation to derive a completely positive conic program for this problem. Both approaches are exact as there are extremal distributions that match the bound.

(b) **Marginal moments:** Given univariate marginal moment information that includes the means and variances, we show that the distributionally robust bound is computationally tractable if the deterministic 0-1 linear program is solvable in polynomial time. In this moment representation, no assumption on the dependency among the random variables is made. We describe the convex formulation for these models in Section 3.2.

(c) **Nonoverlapping multivariate marginal moments:** We conclude by discussing a hybrid approach in Section 3.3, that uses nonoverlapping marginal multivariate information to compute the distributionally robust bound. In this hybrid approach, the random objective coefficients are partitioned into several subsets, and cross moment information for each of the subsets is assumed to be known. However, the dependence structure across different subsets is unknown. A natural application of this approach is in activity networks.

Table 1 provides a summary of the key results from this section.

Description of Θ	Cross moments Mean and covariance	Marginal moments Mean and variance	Nonoverlapping multivariate moments Partitioned mean and covariance
Complexity	$\mathcal{NP}$-hard (Theorem 7)	$\mathcal{P}$	$\mathcal{P}$ fmor activity networks
Formulation	Exponential sized SDP (Theorem 8) CPP (Theorem 9)	SOCP (Theorem 10)	SDP (Theorem 15)

Table 1: **MEAN-UB** for polynomial ($\mathcal{P}$) time solvable 0-1 linear programs with $\mathbf{\Omega} = \Re_n$

Notes: Polynomial complexity ($\mathcal{P}$) in the table refers to an algorithm that takes polynomial time in the size of the instance and $\log(1/\epsilon)$, and computes a bound within ϵ of the optimal bound for all $\epsilon > 0$. SDP, SOCP, and CPP refer to semidefinite program, second-order cone program, and completely positive program, respectively.

3.1 Cross moments (mean and covariance)

Computing the bound on the expected optimal value with mean and covariance information is unfortunately $\mathcal{NP}$-hard even for the class of polynomial time solvable mixed integer linear programs. The complexity result is formally described in the next theorem.

Theorem 7 *(Bertsimas and Popescu [11], Bertsimas, Doan, Natarajan, and Teo [8])*

(i) For $\mathbf{\Omega} = \Re_n^+$, computing **MEAN-UB** *with mean and covariance information is $\mathcal{NP}$-hard even when $Z(\boldsymbol{c}) = \boldsymbol{c}^T\boldsymbol{x}$ is just a linear function on $\boldsymbol{c}$.*

(ii) For $\mathbf{\Omega} = \Re_n$, computing **MEAN-UB** *with mean and covariance information is $\mathcal{NP}$-hard even for linear programs.*

The key step that is used to prove the hardness results in Theorem 7 is to show that the separation version of the dual problem is $\mathcal{NP}$-hard. Then from the equivalence of separation and optimization (see Grötschel et al. [26]), Theorem 7 follows. The separation version of the dual problem is:

> *Given a function $Z(\cdot)$, a set $\mathbf{\Omega}$, a scalar $w_0 \in \Re$, a vector $\boldsymbol{w} \in \Re_n$ and a matrix $\boldsymbol{W} \in \mathcal{S}_n^+$, verify if $w_0 + \boldsymbol{w}^T\boldsymbol{c} + \boldsymbol{c}^T\boldsymbol{W}\boldsymbol{c} \geq Z(\boldsymbol{c})$, for all $\boldsymbol{c} \in \mathbf{\Omega}$. Otherwise, find a violated inequality.*

The separation problem is difficult for $\mathbf{\Omega} = \Re_n^+$ since it is equivalent to testing if a matrix is copositive (see Murty and Kabadi [58]). For $\mathbf{\Omega} = \Re_n$, the separation problem is easy when $Z(\boldsymbol{c})$ is given by the maximum of a polynomial number of linear functions. However, Bertsimas et al. [8] showed that the separation problem is $\mathcal{NP}$-hard when $Z(\boldsymbol{c})$ is the optimal objective value to a linear program where $Z(\boldsymbol{c})$ is given by the maximum of a exponential number of linear functions. The result is proved by a reduction from the two norm maximization problem over a polytope, which was shown to be $\mathcal{NP}$-hard by Mangasarian and Shiau [51]. We now discuss conic programming formulations and relaxations for these $\mathcal{NP}$-hard problems.

3.1.1 Vertex based formulation

In this section, we discuss an explicit conic program to compute **MEAN-UB** given a vertex representation of the convex hull of the feasible region for the mixed integer linear program. Let

$$\text{conv}(\boldsymbol{\mathcal{X}}) = \left\{ \sum_{k=1}^{K} \lambda_k \boldsymbol{x}^{(k)} \ : \ \sum_{k=1}^{K} \lambda_k = 1;\ \lambda_k \geq 0,\ \boldsymbol{x}^{(k)} \in \boldsymbol{\mathcal{X}},\ \forall\ k = 1, \ldots, K \right\}, \tag{7}$$

where $\boldsymbol{x}^{(k)}$'s represent the vertices of the convex hull. Then $Z(\boldsymbol{c})$ can be evaluated using the vertex based formulation, i.e.,

$$Z(\boldsymbol{c}) = \max \left\{ \sum_{k=1}^{K} \lambda_k \boldsymbol{c}^T \boldsymbol{x}^{(k)} \ : \ \sum_{k=1}^{K} \lambda_k = 1;\ \lambda_k \geq 0,\ \forall k = 1, \ldots, K \right\}. \tag{8}$$

Note that in general the number of vertices K is exponential in the size of the problem.

Theorem 8 *(Bersimas and Popescu [11], Bertsimas, Doan, Natarajan, and Teo [8], Zuluaga and Pena [76], Mishra, Natarajan, Tao, and Teo [55]) For the vertex based formulation in (8) with mean and covariance information,* **MEAN-UB** *is computed by solving the conic optimization problem:*

$$\begin{aligned} Z \;=\; \max_{\lambda_k, \boldsymbol{w}_k, \boldsymbol{W}_k, k=1,\ldots,K} \quad & \sum_{k=1}^{K} \boldsymbol{w}_k^T \boldsymbol{x}^{(k)} \\ \text{s.t.} \quad & \sum_{k=1}^{K} \begin{pmatrix} \lambda_k & \boldsymbol{w}_k^T \\ \boldsymbol{w}_k & \boldsymbol{W}_k \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \boldsymbol{\Pi} \end{pmatrix} \\ & \begin{pmatrix} \lambda_k & \boldsymbol{w}_k^T \\ \boldsymbol{w}_k & \boldsymbol{W}_k \end{pmatrix} \in \overline{\mathcal{M}_2(\boldsymbol{\Omega})} \qquad \forall k = 1, \ldots, K. \end{aligned} \tag{9}$$

Step (a): The decision variables in formulation (9) are defined as the scaled conditional moments,

$$\begin{aligned} \lambda_k &= P\left(\boldsymbol{x}(\boldsymbol{c}) = \boldsymbol{x}^{(k)}\right), \\ \boldsymbol{w}_k &= E\left(\boldsymbol{c} | \boldsymbol{x}(\boldsymbol{c}) = \boldsymbol{x}^{(k)}\right) P\left(\boldsymbol{x}(\boldsymbol{c}) = \boldsymbol{x}^{(k)}\right), \\ \boldsymbol{W}_k &= E\left(\boldsymbol{c}\boldsymbol{c}^T | \boldsymbol{x}(\boldsymbol{c}) = \boldsymbol{x}^{(k)}\right) P\left(\boldsymbol{x}(\boldsymbol{c}) = \boldsymbol{x}^{(k)}\right). \end{aligned}$$

Step (b): The objective function is expressed as the weighted sum of conditional moments,

$$E\Big(Z(\boldsymbol{c})\Big) = E\left(\boldsymbol{c}^T \boldsymbol{x}(\boldsymbol{c})\right) = \sum_{k=1}^{K} E\Big(\boldsymbol{c}^T \boldsymbol{x}^{(k)} | \boldsymbol{x}(\boldsymbol{c}) = \boldsymbol{x}^{(k)}\Big) P\Big(\boldsymbol{x}(\boldsymbol{c}) = \boldsymbol{x}^{(k)}\Big) = \sum_{k=1}^{K} \boldsymbol{w}_k^T \boldsymbol{x}^{(k)}.$$

The equality constraint is obtained by defining the moment matrix as the sum of the conditional moments,

$$\begin{aligned} \sum_{k=1}^{K} \begin{pmatrix} \lambda_k & \boldsymbol{w}_k^T \\ \boldsymbol{w}_k & \boldsymbol{W}_k \end{pmatrix} &= \sum_{k=1}^{K} P\Big(\boldsymbol{x}(\boldsymbol{c}) = \boldsymbol{x}^{(k)}\Big) \begin{pmatrix} 1 & E\left(\boldsymbol{c}^T | \boldsymbol{x}(\boldsymbol{c}) = \boldsymbol{x}^{(k)}\right) \\ E\left(\boldsymbol{c} | \boldsymbol{x}(\boldsymbol{c}) = \boldsymbol{x}^{(k)}\right) & E\left(\boldsymbol{c}\boldsymbol{c}^T | \boldsymbol{x}(\boldsymbol{c}) = \boldsymbol{x}^{(k)}\right) \end{pmatrix} \\ &= \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \boldsymbol{\Pi} \end{pmatrix}. \end{aligned}$$

The conic constraints in formulation (9) are obtained from moment feasibility on the domain $\boldsymbol{\Omega}$.

A natural implication of this result is that for $\boldsymbol{\Omega} = \Re_n$ with the number of vertices K polynomially bounded in n, **MEAN-UB** is computable in polynomial time by solving the semidefinite program in Theorem 8. This problem was first studied by Boyle and Lin [16] for an option pricing problem with $Z(\boldsymbol{c}) = (\max_i c_i - k)^+$ where k is the strike price and the option payoff is determined by the maximum of asset prices. Delage and Ye [20] extended this result by relaxing the assumption on the exact knowledge of the two moments and incorporating additional support information. In their model, the set of distributions is defined with a compact convex support $\boldsymbol{\Omega}$, mean $\boldsymbol{\mu}$ and second moment matrix bounded from above in the positive semidefinite order by $\boldsymbol{\Pi}$, i.e.,

$$\Theta = \left\{ \theta \in \mathbb{M}(\boldsymbol{\Omega})_+ \ : \ 1 = \int_{\boldsymbol{\Omega}} d\theta(\boldsymbol{c}), \ \boldsymbol{\mu} = \int_{\boldsymbol{\Omega}} \boldsymbol{c} d\theta(\boldsymbol{c}), \ \boldsymbol{\Pi} \succeq \int_{\boldsymbol{\Omega}} \boldsymbol{c}\boldsymbol{c}^T d\theta(\boldsymbol{c}) \right\}. \tag{10}$$

Delage and Ye [20] showed that with the number of vertices K polynomially bounded in n, **MEAN-UB** is computable in polynomial time for the set of distributions defined in (10) under reasonable assumptions on the convex set $\boldsymbol{\Omega}$.

3.1.2 Constraint based formulation

Theorem 8 is useful when the number of vertices K of the feasible region is not too large. However, K is often exponential in the size of the problem. The next theorem provides a completely positive program to compute **MEAN-UB** for mixed 0-1 linear programs using a constraint based representation. In this part, we work on the original constraint based formulation of $Z(\boldsymbol{c})$ as defined in (1)-(2). The constraint based formulation is derived using an interesting result of Burer [17] who showed that any mixed 0-1 linear program with a mixture of binary and continuous variables can be formulated as a completely positive program. Natarajan et al. [61] extended this result to mixed 0-1 linear programs under objective uncertainty by formulating a completely positive cross moment model (CPCMM).

Theorem 9 *(Natarajan, Teo, and Zheng [61]) For the class of mixed 0-1 linear programs with mean and covariance information,* **MEAN-UB** *is computed by solving the following completely positive program,*

$$\begin{array}{rcll} Z & = & \displaystyle\max_{\boldsymbol{x},\boldsymbol{X},\boldsymbol{Y}} & \displaystyle\sum_{i=1}^{n} Y_{ii} \\ & & s.t. & \boldsymbol{a}_j^T \boldsymbol{x} = b_j \qquad\qquad\qquad\qquad \forall j = 1,\ldots,m \\ & & & \boldsymbol{a}_j^T \boldsymbol{X} \boldsymbol{a}_j = b_j^2 \qquad\qquad\qquad \forall j = 1,\ldots,m \\ & & & X_{ii} = x_i \qquad\qquad\qquad\qquad \forall i \in \mathcal{B} \subseteq \{1,\ldots,n\} \\ & & & \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{x}^T \\ \boldsymbol{\mu} & \boldsymbol{\Pi} & \boldsymbol{Y}^T \\ \boldsymbol{x} & \boldsymbol{Y} & \boldsymbol{X} \end{pmatrix} \in \overline{\mathcal{M}_2(\boldsymbol{\Omega} \times \Re_n^+)}. \end{array} \tag{11}$$

Step (a): The decision variables in this formulation are defined as

$$\begin{array}{rcl} \boldsymbol{x} & = & E(\boldsymbol{x}(\boldsymbol{c})), \\ \boldsymbol{Y} & = & E(\boldsymbol{x}(\boldsymbol{c})\boldsymbol{c}^T), \\ \boldsymbol{X} & = & E(\boldsymbol{x}(\boldsymbol{c})\boldsymbol{x}(\boldsymbol{c})^T). \end{array}$$

Step (b): The objective function is expressed as

$$E\Big(Z(\boldsymbol{c})\Big) = \sum_{i=1}^{n} E\left(c_i x_i(\boldsymbol{c})\right) = \sum_{i=1}^{n} Y_{ii}.$$

The first two constraints in formulation (11) are obtained by taking the expectations,

$$\begin{array}{rclcl} b_j & = & \boldsymbol{a}_j^T \boldsymbol{x} & = & E\left(\boldsymbol{a}_j^T \boldsymbol{x}(\boldsymbol{c})\right), \\ b_j^2 & = & \boldsymbol{a}_j^T \boldsymbol{X} \boldsymbol{a}_j & = & E\left(\left(\boldsymbol{a}_j^T \boldsymbol{x}(\boldsymbol{c})\right)^2\right). \end{array}$$

The third constraint is obtained from taking the expectation of the equality constraint $x_i(\boldsymbol{c})^2 = x_i(\boldsymbol{c})$ for the binary variables $x_i(\boldsymbol{c}) \in \{0, 1\}$,

$$X_{ii} = E(x_i(\boldsymbol{c})^2) = E(x_i(\boldsymbol{c})) = x_i.$$

The validity of the conic constraint follows from $\boldsymbol{c} \in \boldsymbol{\Omega}$ and $\boldsymbol{x}(\boldsymbol{c}) \geq \boldsymbol{0}$, since

$$\begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{x}^T \\ \boldsymbol{\mu} & \boldsymbol{\Pi} & \boldsymbol{Y}^T \\ \boldsymbol{x} & \boldsymbol{Y} & \boldsymbol{X} \end{pmatrix} = \begin{pmatrix} 1 & E(\boldsymbol{c}^T) & E(\boldsymbol{x}(\boldsymbol{c})^T) \\ E(\boldsymbol{c}) & E(\boldsymbol{c}\boldsymbol{c}^T) & E(\boldsymbol{c}\boldsymbol{x}(\boldsymbol{c})^T) \\ E(\boldsymbol{x}(\boldsymbol{c})) & E(\boldsymbol{x}(\boldsymbol{c})\boldsymbol{c}^T) & E(\boldsymbol{x}(\boldsymbol{c})\boldsymbol{x}(\boldsymbol{c})^T) \end{pmatrix} \in \overline{\mathcal{M}_2(\boldsymbol{\Omega} \times \Re_n^+)}.$$

For $\boldsymbol{\Omega} = \Re_n^+$, the conic constraint is equivalent to complete positivity of the matrix variable. While the number of constraints and variables in Formulation (11) are polynomial in the size of the problem, the difficulty lies in the completely positive cone constraint which is intractable. The advantage of this formulation is that it directly uses the constraint based representation instead of the vertex based representation. A simple polynomial time computable upper bound is obtained by using the doubly nonnegative relaxation of the completely positive cone,

$$\begin{array}{rcll} Z & \leq & \displaystyle\max_{\boldsymbol{x},\boldsymbol{X},\boldsymbol{Y}} & \displaystyle\sum_{i=1}^{n} Y_{ii} \\ & & \text{s.t.} & \boldsymbol{a}_j^T \boldsymbol{x} = b_j \qquad\qquad \forall j = 1, \ldots, m \\ & & & \boldsymbol{a}_j^T \boldsymbol{X} \boldsymbol{a}_j = b_j^2 \qquad \forall j = 1, \ldots, m \\ & & & X_{ii} = x_i \qquad\qquad \forall i \in \mathcal{B} \subseteq \{1, \ldots, n\} \\ & & & \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{x}^T \\ \boldsymbol{\mu} & \boldsymbol{\Pi} & \boldsymbol{Y}^T \\ \boldsymbol{x} & \boldsymbol{Y} & \boldsymbol{X} \end{pmatrix} \in \mathcal{S}_{2n}^+ \cap \mathcal{N}_{2n}. \end{array} \tag{12}$$

By using higher order approximations to the completely positive cone, it is possible to get better approximations to **MEAN-UB**.

3.1.3 Extension to higher order moments

The constraint based formulation in Theorem 9 uses the first two moments of the random vector $\boldsymbol{c}$ and the random optimal solution vector $\boldsymbol{x}(\boldsymbol{c})$ as decision variables,

$$\Big(1, E\left(\boldsymbol{c}\right), E\left(\boldsymbol{x}(\boldsymbol{c})\right), E\left(\boldsymbol{c}\boldsymbol{c}^T\right), E\left(\boldsymbol{c}\boldsymbol{x}(\boldsymbol{c})^T\right), E\left(\boldsymbol{x}(\boldsymbol{c})\boldsymbol{x}(\boldsymbol{c})^T\right)\Big).$$

By allowing for higher order moments, Lasserre [44] has generalized the approach to the class of parametric polynomial optimization problems which includes mixed 0-1 linear programs as a special case. Note that binary variables can be modeled in polynomial optimization problems with constraints of the form $x_i^2 = x_i$. The parametric optimization problem studied in Lasserre [44] is of the form,

$$Z(\boldsymbol{\xi}) = \max\Big\{ f(\boldsymbol{\xi}, \boldsymbol{x}) \ : \ f_j(\boldsymbol{\xi}, \boldsymbol{x}) \geq 0, \ \forall j = 1, \ldots, m \Big\}, \tag{13}$$

where $\boldsymbol{\xi}$ is a random parameter vector that lies in a compact set $\boldsymbol{\Omega}$ with a probability measure θ, and $\boldsymbol{x}$ is the decision vector. Define the set

$$\boldsymbol{K} = \left\{ (\boldsymbol{\xi}, \boldsymbol{x}) \ : \ \boldsymbol{\xi} \in \boldsymbol{\Omega}, \ f_j(\boldsymbol{\xi}, \boldsymbol{x}) \geq 0, \ \forall j = 1, \ldots, m \right\}.$$

Let φ denote the joint probability measure on the random vector $(\boldsymbol{\xi}, \boldsymbol{x}(\boldsymbol{\xi}))$, where $\boldsymbol{x}(\boldsymbol{\xi})$ is an optimal solution for a fixed $\boldsymbol{\xi}$. Lasserre [44] defined the infinite dimensional linear program over the measure φ as

$$\begin{aligned} \sup_{\varphi \in \mathbb{M}(\boldsymbol{K})_+} \quad & \int_{\boldsymbol{K}} f d\varphi \\ \text{s.t.} \quad & \text{proj}_{\boldsymbol{\Omega}} \varphi = \theta, \end{aligned} \tag{14}$$

where $\text{proj}_{\boldsymbol{\Omega}} \varphi$ denotes the projection of φ on the set $\boldsymbol{\Omega}$. This formulation is referred to as a "joint + marginal" formulation since φ is a joint probability measure on the parameters and optimal solutions while θ is the given probability measure on the parameters. Under appropriate compactness conditions on the feasible region (see Lasserre [44]), the optimal objective value to (14) is exactly $E_\theta(Z(\boldsymbol{\xi}))$. To solve the infinite dimensional linear program for polynomial functions $f(\cdot)$ and $f_j(\cdot)$, Lasserre proposed a hierarchy of semidefinite relaxations that is based on the theory of moments. The optimal objective value to the sequence of semidefinite relaxations converges in the limit to $E_\theta(Z(\boldsymbol{\xi}))$. The attractiveness of this technique is that it is general purpose since it can handle uncertainty in the objective and constraints and is applicable to the class of polynomial optimization problems. However the size of the semidefinite relaxation grows rapidly which makes solving the higher order semidefinite relaxations numerically challenging. In the remaining part of this section, we review sets of distributions Θ where the distributionally robust bound can be found in polynomial time using conic optimization.

3.2 Marginal moments (mean and variance)

Suppose that the support space Ω_i for each random variable c_i along with the mean $E(c_i) = \mu_i$ and the second moment $E(c_i^2) = \Pi_i$ is known. The variance of c_i is denoted by σ_i^2. However, the dependence structure among the different random variables is unknown. Let $\theta_i = \text{proj}_i\theta$ denote the projection of the multivariate measure θ to the ith random variable c_i. The marginal moment representation of the set of distributions is

$$\Theta = \left\{ \theta \in \mathbb{M}(\Omega_1 \times \ldots \times \Omega_n)_+ \ : \ \begin{array}{c} 1 = \displaystyle\int_{\Omega_i} d\theta_i(c_i), \ \mu_i = \int_{\Omega_i} c_i d\theta_i(c_i), \\ \Pi_i = \int_{\Omega_i} c_i^2 d\theta_i(c_i), \ \forall i = 1, \ldots, n \end{array} \right\}.$$

The upper bound on the expected optimal value with mean and variance information is formulated as

$$\begin{aligned} Z \quad = \quad \sup_{\theta \in \mathbb{M}(\Omega_1 \times \ldots \times \Omega_n)_+} \quad & \int_{\Omega_1 \times \ldots \times \Omega_n} Z(\boldsymbol{c}) d\theta(\boldsymbol{c}) \\ \text{s.t.} \quad & \int_{\Omega_i} 1 d\theta_i(c_i) = 1 \qquad \forall i = 1, \ldots, n \\ & \int_{\Omega_i} c_i d\theta_i(c_i) = \mu_i \qquad \forall i = 1, \ldots, n \\ & \int_{\Omega_i} c_i^2 d\theta_i(c_i) = \Pi_i \qquad \forall i = 1, \ldots, n \end{aligned} \tag{15}$$

Moment feasibility in this instance is equivalent to the feasibility of univariate moment sequences. This condition is obviously necessary. Sufficiency follows by constructing a feasible joint measure using the independent distribution. Testing moment feasibility is thus easy for the marginal moment model for both $\Omega_i = \Re$ and $\Re^+$. The next theorem provides a conic programming formulation for the class of 0-1 linear programs, i.e.,

$$\boldsymbol{\mathcal{X}} = \left\{ \boldsymbol{x} \in \Re_n^+ : \boldsymbol{a}_j^T \boldsymbol{x} = b_j, \ \forall j = 1, \ldots, m; \ x_i \in \{0, 1\}, \ \forall i \in \{1, \ldots, n\} \right\}.$$

Theorem 10 *(Bertsimas, Natarajan, and Teo [10], Natarajan, Song, and Teo [60])*

(i) For the class of 0-1 linear programs with mean and variance information, **MEAN-UB** *in (15) is computed by solving the following conic optimization problem,*

$$\begin{aligned} Z \quad = \quad \max_{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}} \quad & \sum_{i=1}^{n} y_i \\ s.t. \quad & \boldsymbol{x} \in conv(\boldsymbol{\mathcal{X}}) \\ & \begin{pmatrix} x_i & y_i \\ y_i & z_i \end{pmatrix}, \begin{pmatrix} 1 - x_i & \mu_i - y_i \\ \mu_i - y_i & \Pi_i - z_i \end{pmatrix} \in \overline{\mathcal{M}_2(\Omega_i)} \quad \forall i = 1, \ldots, n. \end{aligned} \tag{16}$$

For $\mathbf{\Omega} = \Re_n$, formulation (16) reduces to the second-order cone program,

$$Z = \max_{\boldsymbol{x} \in conv(\boldsymbol{\mathcal{X}})} \sum_{i=1}^{n} \left(\mu_i x_i + \sigma_i \sqrt{x_i(1-x_i)} \right). \tag{17}$$

(ii) Z is computable in polynomial time for a 0-1 linear program with a compact convex hull.

Step (a): The decision variables in formulation (16) are defined as

$$\begin{aligned} x_i &= P(x_i(\boldsymbol{c}) = 1), \\ y_i &= E\left(c_i x_i(\boldsymbol{c})\right) &&= E\left(c_i | x_i(\boldsymbol{c}) = 1\right) P(x_i(\boldsymbol{c}) = 1), \\ z_i &= E\left(c_i^2 x_i(\boldsymbol{c})\right) &&= E\left(c_i^2 | x_i(\boldsymbol{c}) = 1\right) P(x_i(\boldsymbol{c}) = 1). \end{aligned}$$

Step (b): The objective function is expressed in terms of the decision variables as

$$E\Big(Z(\boldsymbol{c})\Big) = \sum_{i=1}^{n} E\Big(c_i | x_i(\boldsymbol{c}) = 1\Big) P(x_i(\boldsymbol{c}) = 1) = \sum_{i=1}^{n} y_i.$$

Since the vector $(x_1(\boldsymbol{c}), \ldots, x_n(\boldsymbol{c}))$ lies in $\text{conv}(\boldsymbol{\mathcal{X}})$ for all realizations of $\boldsymbol{c}$, taking expectation, we get

$$\boldsymbol{x} = E(\boldsymbol{x}(\boldsymbol{c})) \in \text{conv}(\boldsymbol{\mathcal{X}}).$$

The conic constraints are from the moment feasibility conditions,

$$\begin{pmatrix} x_i & y_i \\ y_i & z_i \end{pmatrix} = P(x_i(\boldsymbol{c}) = 1) \begin{pmatrix} 1 & E(c_i | x_i(\boldsymbol{c}) = 1) \\ E(c_i | x_i(\boldsymbol{c}) = 1) & E(c_i^2 | x_i(\boldsymbol{c}) = 1) \end{pmatrix} \in \overline{\mathcal{M}_2(\Omega_i)},$$

$$\begin{pmatrix} 1 - x_i & \mu_i - y_i \\ \mu_i - y_i & \Pi_i - z_i \end{pmatrix} = P(x_i(\boldsymbol{c}) = 0) \begin{pmatrix} 1 & E(c_i | x_i(\boldsymbol{c}) = 0) \\ E(c_i | x_i(\boldsymbol{c}) = 0) & E(c_i^2 | x_i(\boldsymbol{c}) = 0) \end{pmatrix} \in \overline{\mathcal{M}_2(\Omega_i)}.$$

The proof of the tightness can be found in [10].

A dual representation of this conic program in Theorem 10 is discussed in Klein Haneveld [37], Birge and Maddox [12] and Bertsimas et al. [9]. The key implication of Theorem 10 is that **MEAN-UB** can be found in polynomial time for supports such as $\mathbf{\Omega} = \Re_n$ and $\mathbf{\Omega} = \Re_n^+$ for the class of 0-1 linear programs with a compact convex hull representation. This provides tight bounds for combinatorial optimization problems such as the shortest path, linear assignment, and spanning tree problems.

3.2.1 Extension to integer programs

The marginal moment model has been extended to general integer programs by Natarajan et al. [60] with a binary reformulation. Assume that the deterministic integer program with nonnegative integer variables is formulated as

$$Z(\boldsymbol{c}) = \max \left\{ \boldsymbol{c}^T \boldsymbol{x} \ : \ \sum_{i=1}^{n} a_{ji} x_i = b_j, \ \forall j = 1, \ldots, m; \ x_i \in \mathcal{X}_i, \ \forall i = 1, \ldots, n \right\},$$

where the set $\mathcal{X}_i$ consists of nonnegative integer values from α_i to β_i:

$$\mathcal{X}_i = \{\alpha_i, \alpha_i + 1, \ldots, \beta_i - 1, \beta_i\} \subseteq \mathcal{Z}^+.$$

Defining binary variables y_{ik} for $k \in \mathcal{X}_i$, $i = 1, \ldots, n$, the binary expansion of the feasible region is given as

$$\mathcal{Y} = \left\{ \boldsymbol{y} \ : \ \begin{array}{l} \sum_{i=1}^{n} \sum_{k \in \mathcal{X}_i} a_{ji} k y_{ik} = b_j, \ \forall j = 1, \ldots, m; \ \sum_{k \in \mathcal{X}_i} y_{ik} = 1, \ \forall i = 1, \ldots, n; \\ y_{ik} \in \{0, 1\}, \ \forall k \in \mathcal{X}_i, \ \forall i = 1, \ldots, n \end{array} \right\}.$$

There is an unique one to one correspondence between the extreme points of the original feasible region and the binary reformulation $\mathcal{Y}$, namely $x_i = k$ if and only if $y_{ik} = 1$. Based on this, Natarajan et al. [60] provided a second-order cone program for integer programs with mean and variance information for $\boldsymbol{\Omega} = \Re_n$.

Theorem 11 *(Natarajan, Song, and Teo [60]) For the class of integer programs with mean and variance information and* $\boldsymbol{\Omega} = \Re_n$*,* **MEAN-UB** *in (15) is computed by solving the following second-order cone program,*

$$Z = \max \ \left\{ \sum_{i=1}^{n} \left(\mu_i \sum_{k \in \mathcal{X}_i} k y_{ik} + \sigma_i \sqrt{\sum_{k \in \mathcal{X}_i} k^2 y_{ik} - \left(\sum_{k \in \mathcal{X}_i} k y_{ik} \right)^2} \right) : \boldsymbol{y} \in conv(\mathcal{Y}) \right\}. \tag{18}$$

3.2.2 Extension to marginal distributions

Finding a bound on a function of multiple random variables given only the probability measures of the individual random variables has its origins in the Monge-Kantorovich [57, 33] formulation for mass transportation problems. The reader is referred to the book of Rachev and Ruschendorf [66] for a historical account of this problem. The upper bound on the expected optimal value with given marginal distributions is formulated as

$$\begin{aligned} Z \quad = \quad & \sup_{\theta \in \mathbb{M}(\Omega_1 \times \ldots \times \Omega_n)_+} \int_{\Omega_1 \times \ldots \times \Omega_n} Z(c_1, \ldots, c_n) d\theta(\boldsymbol{c}) \\ & \text{s.t.} \quad \text{proj}_{\Omega_i} \theta = \theta_i \qquad\qquad \forall i = 1, \ldots, n. \end{aligned} \tag{19}$$

Meilijson and Nadas [53] solved this problem in a combinatorial optimization setting by estimating an upper bound on the expected longest path in a directed acyclic graph given marginal distributions of the arc lengths. Their motivation was to find the worst case expected project completion time in an activity network across all joint distributions of activity durations that are consistent with the marginal distributions. This bound is thus robust against dependence. Weiss [74] generalized this bound to combinatorial optimization problems such as the maximum flow, shortest route and reliability problems. The formulation in [53, 74] is derived from a dual convex minimization formulation.

Theorem 12 *(Meilijson and Nadas [53]) For 0-1 linear programs with given marginal distributions,* **MEAN-UB** *in (19) is computed by solving the following convex minimization problem,*

$$Z = \inf_{\boldsymbol{d}} \left(Z(\boldsymbol{d}) + \sum_{i=1}^{n} E_{\theta_i}(c_i - d_i)^+ \right). \tag{20}$$

Natarajan et al. [60] provided a primal approach to compute this bound for continuous marginal distributions.

Theorem 13 *(Natarajan, Song, and Teo [60]) For 0-1 linear programs with continuous marginal distributions $c_i \sim F_i(\cdot)$,* **MEAN-UB** *in (19) is computed by solving the following concave maximization problem,*

$$Z = \sup_{\boldsymbol{x} \in conv(\boldsymbol{\mathcal{X}})} \sum_{i=1}^{n} \int_{1-x_i}^{1} F_i^{-1}(t)dt. \tag{21}$$

In the next example, we compare the probabilistic bounds for combinatorial optimization problems given marginal distributions with and without the assumption of independence.

Example: Probabilistic analysis of combinatorial optimization problems

Bertsimas et al. [9] applied the marginal distribution model to find the expected value of combinatorial optimization problems when the assumption of independence among the random costs is dropped. For the linear assignment problem with random costs identically distributed $c_{ij} \sim F(\cdot)$, the tight lower bound on the expected optimal value is found by solving the following convex minimization problem,

$$Z_{LAP} = \min \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} \int_{0}^{x_{ij}} F^{-1}(t)dt \ : \ \begin{array}{l} \sum_{i=1}^{n} x_{ij} = 1, \ \forall i = 1, \ldots, n; \\ \sum_{j=1}^{n} x_{ij} = 1, \ \forall j = 1, \ldots, n; \ x_{ij} \geq 0, \ \forall i, j = 1, \ldots, n \end{array} \right\}.$$

Theorem 14 *(Bertsimas, Natarajan, and Teo [9]) Let the random variables c_{ij} for $i, j = 1, \ldots, n$ be identically distributed with density function $f(\cdot)$ and distribution function $F(\cdot)$. Then*

$$Z_{LAP} = n^2 \int_{0}^{F^{-1}(1/n)} cf(c)dc.$$

A comparison of the bound in Theorem 14 with results for the independence model is provided next:

(a) For the uniform distribution in $[0, 1]$, the lower bound in Theorem 14 is $Z_{LAP} = 1/2$. This lower bound is tight for all n. Namely there exists a joint distribution with uniform marginals that attains this bound for each n. In contrast under the assumption of independence, only the explicit asymptotic limit $\lim_{n\to\infty} E(Z_{LAP}(\boldsymbol{c})) = \pi^2/6$ is known.

(b) For the exponential distribution with parameter 1, the lower bound is $Z_{LAP} = n+n(n-1)\ln(1-1/n)$, while under independence $E(Z_{LAP}(\boldsymbol{c})) = \sum_{i=1}^{n} 1/i^2$.

As highlighted by these examples, the extremal distributions under the marginal distribution model provide new and non-trivial limits on the asymptotic behavior of optimization problems. It is interesting to compare the proof technique for the marginal distribution model with that of the independence model. The proof of the former model is based on convex optimization, while for the later model, the proof is based on sophisticated probabilistic techniques (see [2]).

3.3 Nonoverlapping multivariate marginal moments

Doan and Natarajan [21] recently developed the bound **MEAN-UB** for a set of distributions Θ that lies between the two extremes of the cross moment and marginal moment information. In this distribution model, the random objective coefficients are assumed to be partitioned into subsets with information on the moments of the random parameters in each subset. The dependence structure between any random parameters for different subsets is assumed to be unknown. To describe the formulation, we use the example of activity networks. In the directed acyclic graph representation of an activity network, a natural partition is formed by the set of arcs (activities) entering each node. Define $\boldsymbol{c}_j = (c_{ij})_{i:(i,j)\in\mathcal{E}}$ to be the sub-vector of random arc lengths for the arcs entering node $j \in \mathcal{V}$, where n is the total number of nodes in the graph. Denote the dimension of $\boldsymbol{c}_j$ as n_j. Suppose that the support $\boldsymbol{\Omega}_j$ for each random sub-vector $\boldsymbol{c}_j$ along with the mean $E(\boldsymbol{c}_j) = \boldsymbol{\mu}_j$ and the second moment moment matrix $E(\boldsymbol{c}_j\boldsymbol{c}_j^T) = \boldsymbol{\Pi}_j$ is known. For example, in projects where different teams are responsible for the set of activities entering different nodes, it is reasonable to assume that each team is knowledgeable about the joint distribution of the activities for which they are responsible. The project manager is interested in evaluating the worst-case expected project completion time that is compatible with these factors. The correlation among the arc lengths c_{ij} and c_{kl} entering two different nodes j and l is unknown under this model. Let θ_j denote the projection of the measure θ for the random sub-vector $\boldsymbol{c}_j$. The upper bound on the expected longest path with nonoverlapping mean, variance and covariance information is formulated as

$$
\begin{aligned}
Z_{LPP} \quad = \sup_{\theta \in \mathbb{M}(\mathbf{\Omega}_1 \times \ldots \times \mathbf{\Omega}_n)_+} \quad & \int_{\mathbf{\Omega}_1 \times \ldots \times \mathbf{\Omega}_n} Z_{LPP}(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n) d\theta(\boldsymbol{c}) \\
\text{s.t.} \quad & \int_{\mathbf{\Omega}_j} 1 d\theta_j(\boldsymbol{c}_j) = 1 && \forall j \in \mathcal{V} \\
& \int_{\mathbf{\Omega}_j} \boldsymbol{c}_j d\theta_j(\boldsymbol{c}_j) = \boldsymbol{\mu}_j && \forall j \in \mathcal{V} \\
& \int_{\mathbf{\Omega}_j} \boldsymbol{c}_j \boldsymbol{c}_j^T d\theta_j(\boldsymbol{c}_j) = \mathbf{\Pi}_j && \forall j \in \mathcal{V}.
\end{aligned}
\tag{22}
$$

Theorem 15 *For the longest path problem in a directed acyclic graph with nonoverlapping multivariate marginal moment information at each node,* **MEAN-UB** *in (22) is computed by solving the following conic optimization problem:*

$$
\begin{aligned}
Z_{LPP} = \max_{x_{ij}, \boldsymbol{w}_{ij}, \boldsymbol{W}_{ij}, (i,j) \in \mathcal{E}} \quad & \sum_{(i,j) \in \mathcal{E}} \boldsymbol{e}_{ij}^T \boldsymbol{w}_{ij} \\
s.t. \quad & \sum_{j:(i,j) \in \mathcal{E}} x_{ij} - \sum_{j:(j,i) \in \mathcal{E}} x_{ji} = \begin{cases} 1, & \text{if } i = s \\ -1, & \text{if } i = t \\ 0, & \text{if } i \in \mathcal{V} \end{cases} \\
& \begin{pmatrix} x_{ij} & \boldsymbol{w}_{ij}^T \\ \boldsymbol{w}_{ij} & \boldsymbol{W}_{ij} \end{pmatrix} \in \overline{\mathcal{M}_2(\mathbf{\Omega}_j)} && \forall (i,j) \in \mathcal{E} \\
& \begin{pmatrix} 1 & \boldsymbol{\mu}_j^T \\ \boldsymbol{\mu}_j & \mathbf{\Pi}_j \end{pmatrix} - \sum_{i:(i,j) \in \mathcal{E}} \begin{pmatrix} x_{ij} & \boldsymbol{w}_{ij}^T \\ \boldsymbol{w}_{ij} & \boldsymbol{W}_{ij} \end{pmatrix} \in \overline{\mathcal{M}_2(\mathbf{\Omega}_j)} && \forall j \in \mathcal{V},
\end{aligned}
\tag{23}
$$

where $\boldsymbol{e}_{ij}$ *is a vector of dimension* n_j *with* 1 *in its ith component and 0 otherwise. For* $\mathbf{\Omega} = \Re_n$, **MEAN-UB** *is computable in polynomial time as a semidefinite program.*

Step (a): The decision variables in this model are defined as

$$
\begin{aligned}
x_{ij} &= P(x_{ij}(\tilde{\boldsymbol{c}}) = 1), && && \forall (i,j) \in \mathcal{E}, \\
\boldsymbol{w}_{ij} &= E(\boldsymbol{c_j} x_{ij}(\tilde{\boldsymbol{c}})) &&= E(\boldsymbol{c_j} | x_{ij}(\tilde{\boldsymbol{c}}) = 1) P(x_{ij}(\tilde{\boldsymbol{c}}) = 1), && \forall (i,j) \in \mathcal{E}, \\
\boldsymbol{W}_{ij} &= E(\boldsymbol{c_j} \boldsymbol{c_j}^T x_{ij}(\tilde{\boldsymbol{c}})) &&= E(\boldsymbol{c_j} \boldsymbol{c_j}^T | x_{ij}(\tilde{\boldsymbol{c}}) = 1) P(x_{ij}(\tilde{\boldsymbol{c}}) = 1), && \forall (i,j) \in \mathcal{E}.
\end{aligned}
$$

Step (b): The objective function is expressed as

$$
E\Big(Z_{LPP}(\boldsymbol{c})\Big) = \sum_{(i,j) \in \mathcal{E}} E\Big(c_{ij} | x_{ij}(\boldsymbol{c}) = 1\Big) P(x_{ij}(\tilde{\boldsymbol{c}}) = 1) = \sum_{(i,j) \in \mathcal{E}} \boldsymbol{e}_{ij}^T \boldsymbol{w}_{ij}.
$$

The first constraint is obtained by taking the expectation of the vector $\boldsymbol{x}(\boldsymbol{c}) \in \text{conv}(\boldsymbol{\mathcal{X}})$ for all realizations of $\boldsymbol{c}$,

$$
\boldsymbol{x} = E(\boldsymbol{x}(\boldsymbol{c})) \in \text{conv}(\boldsymbol{\mathcal{X}}).
$$

The conic constraint is obtained from moment feasibility on the domain $\mathbf{\Omega}$. The last conic constraint is obtained from the equality,

$$
\begin{aligned}
&\begin{pmatrix} 1 & \boldsymbol{\mu}_j^T \\ \boldsymbol{\mu}_j & \mathbf{\Pi}_j \end{pmatrix} - \sum_{i:(i,j)\in\mathcal{E}} \begin{pmatrix} x_{ij} & \boldsymbol{w}_{ij}^T \\ \boldsymbol{w}_{ij} & \boldsymbol{W}_{ij} \end{pmatrix} \\
&= P\left(x_{ij}(\tilde{\boldsymbol{c}}) = 0, \forall i \in \mathcal{V}\right) \begin{pmatrix} 1 & E(\boldsymbol{c}_j^T | x_{ij}(\tilde{\boldsymbol{c}}) = 0, \forall i \in \mathcal{V}) \\ E(\boldsymbol{c}_j^T | x_{ij}(\tilde{\boldsymbol{c}}) = 0, \forall i \in \mathcal{V}) & E(\boldsymbol{c}_j \boldsymbol{c}_j^T | x_{ij}(\tilde{\boldsymbol{c}}) = 0, \forall i \in \mathcal{V}) \end{pmatrix} \\
&\in \overline{\mathcal{M}_2(\mathbf{\Omega}_j)}.
\end{aligned}
$$

The proof of tightness can be found in Doan and Natarajan [21].

4 Applications

The conic programming method provides a flexible and simple way to analyze mixed integer linear programs with random objective. Since there is a huge number of problems under the umbrella of mixed integer LP, we review only a few applications of the approach.

4.1 Activity networks

The activity network example in Figure 1 is inspired from van Slyke [73] and discussed in more details in Bertsimas et al. [10]. This example serves as a benchmark for comparison of the conic programs with alternative methods for estimating the project completion time and the criticality indices of activities under random activity durations.

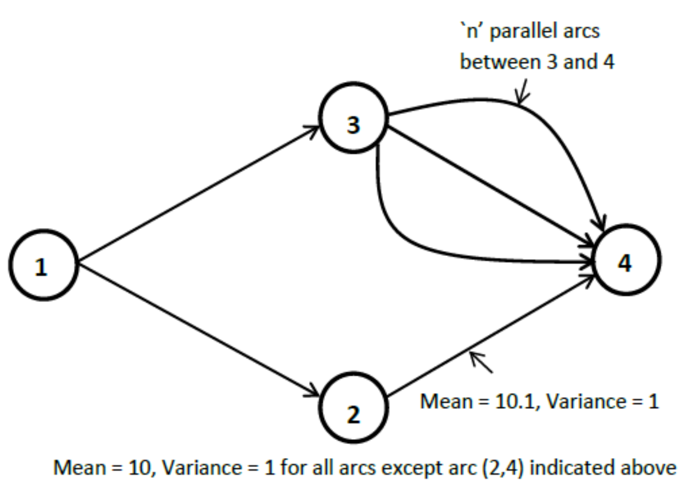

Figure 1: Simple activity network from Van Slyke [73]

The deterministic critical path method (CPM) uses the expected value of the activity durations to compute the critical path. The deterministic critical path approach identifies activities $(1,2)$ and $(2,4)$ as being critical with expected project duration of 20.1 irrespective of the number of parallel arcs between nodes 3 and 4. Simulation is a popular approach to analyze activity networks under uncertainty. By simulating durations from a joint probability distribution, it is possible to analyze the project by solving a longest path problem for each sample. However, this comes at the cost of computational expense for large projects comprising of several thousand activities. For this project, we use the multivariate normal distribution to simulate activity durations. Table 2 provides a comparison of CPM and the simulation method with the SDP and SOCP methods discussed in Section 3.

	CPM		Simulation		Cross moment		Marginal moment		Nonoverlap moment	
n	$Z_{LPP}(\boldsymbol{\mu})$	x_{13}	$E(Z_{LPP}(\boldsymbol{c}))$	x_{13}	Z_{LPP}	x_{13}	Z_{LPP}	x_{13}	Z_{LPP}	x_{13}
1	20.1	0.000	20.848	0.481	21.051	0.475	22.050	0.487	21.757	0.485
2	20.1	0.000	21.117	0.597	21.504	0.594	22.431	0.563	22.122	0.559
5	20.1	0.000	21.490	0.715	22.349	0.733	23.147	0.656	22.842	0.664
10	20.1	0.000	21.768	0.785	23.271	0.820	23.960	0.728	23.672	0.746
15	20.1	0.000	21.917	0.817	23.973	0.862	24.595	0.770	24.322	0.792

Table 2: Project statistics for different numbers of parallel arcs (n) between nodes 3 and 4

Notes: $Z_{LPP}(\boldsymbol{\mu})$ *is the project duration using the means.* $E(Z_{LPP}(\boldsymbol{c}))$ *is the expected project duration for the normal distribution and* $Z_{LPP} = \sup_{\theta \in \Theta} E(Z_{LPP}(\boldsymbol{c}))$ *is the worst case expected project duration. The value* x_{13} *is the criticality index of activity* $(1,3)$ *for the different models.*

The simulation results in Table 2 are obtained with independent activity durations. It is clear from the table that the deterministic critical path method severely underestimates the expected project duration. Furthermore it fails to identify activity $(1,3)$ as being the most important especially when the number of parallel arcs between nodes 3 and 4 increases. The criticality index of activity $(1,3)$ is clearly larger than the criticality index of activity $(1,2)$ for $n > 1$ based on the simulation results. This is due to the simple observation that it is very likely that any one of the upper paths will be critical in comparison to the lower path due to the presence of parallel independent arcs. Using a deterministic approach would imply that the project manager focuses on the wrong activity $(1,2)$. For the cross moment and nonoverlapping multivariate marginal moments, SDP is used while for univariate marginal moments, SOCP is used to compute the worst case expected project completion time. All three methods help identify the importance of activity $(1,3)$ under distributional uncertainty. As should be expected, the worst case expected project duration for the cross moment model is lesser than that of the nonoverlapping marginal moment model which in turn is lesser than marginal moment model.

In Table 3, the effect of partial correlations is tested on the project performance. Three multivariate normal distributions are simulated. The correlations between all activities are set to zero except for the correlation between activity $(1, 2)$ and $(1, 3)$, which takes values of -0.9, 0 and 0.9. From Table 3, it is clear that actual criticality index is sensitive to the correlation structure as expected. In fact, as the degree of dependence among activity durations increases, the variation in the criticality indices potentially become significant. Although SOCP based marginal moment model still identifies activity $(1, 3)$ as the most critical activity, it does not capture explicit dependence information. In contrast, the more computationally intensive SDP models help provide closer fits to the exact simulation values.

	Simulation		Cross moment		Nonoverlap moment	
$\rho_{(1,2),(1,3)}$	$E(Z_{LPP}(\boldsymbol{c}))$	x_{13}	Z_{LPP}	x_{13}	Z_{LPP}	x_{13}
-0.9	21.6401	0.6787	22.5514	0.6990	23.0975	0.6432
0	21.4908	0.7151	22.3496	0.7334	22.8428	0.6645
0.9	21.3058	0.7966	22.0933	0.8024	22.3966	0.7243

Table 3: Project statistics for $n = 5$ with different correlations $\rho_{(1,2),(1,3)}$ between activity $(1, 2)$ and $(1, 3)$

4.2 Vertex packing

In this example, we compare the persistency obtained from the SOCP in Formulation (17) with known persistency results for deterministic combinatorial optimization problems. The deterministic weighted vertex packing problem is: Given an undirected graph $G(\mathcal{V}, \mathcal{E})$ with weights c_i for each vertex $i \in \mathcal{V}$, find a subset of vertices $\mathcal{S} \subseteq \mathcal{V}$ such that $(i, j) \notin E$ for all $i, j \in \mathcal{S}$ with maximum total sum of the weights of nodes in the set $\mathcal{S}$. The integer programming formulation for the weighted vertex packing problem is

$$Z_{WVP}(\boldsymbol{c}) = \max \left\{ \sum_{i \in \mathcal{V}} c_i x_i \ : \ x_i + x_j \leq 1, \ \forall (i, j) \in \mathcal{E}; \ x_i \in \{0, 1\}, \ \forall i \in \mathcal{V} \right\}, \tag{24}$$

with its linear programming relaxation given as

$$\bar{Z}_{WVP}(\boldsymbol{c}) = \max \left\{ \sum_{i \in \mathcal{V}} c_i x_i \ : \ x_i + x_j \leq 1, \ \forall (i, j) \in \mathcal{E}; \ x_i \geq 0, \ \forall i \in \mathcal{V} \right\}. \tag{25}$$

Let $\boldsymbol{x}_{WVP}(\boldsymbol{c})$ and $\bar{\boldsymbol{x}}_{WVP}(\boldsymbol{c})$ denote the optimal solutions to (24) and (25) for objective vector $\boldsymbol{c}$. The vertex packing problem has shown to be persistent in the deterministic discrete optimization context (see Nemhauser and Trotter [62]) - namely for every optimal solution to the linear programming relaxation

$\bar{\boldsymbol{x}}_{WVP}(\boldsymbol{c})$, the set of variables realizing binary values retains the same binary values in at least one optimal solution $\boldsymbol{x}_{WVP}(\boldsymbol{c})$. If we allow for uncertainty in the weights, e.g., given $E(c_i) = \mu_i$ and $Var(c_i) = \sigma_i^2$, $\forall i \in \mathcal{V}$, then under the marginal moment model, the upper bound on the expected optimal value is computed by solving

$$Z_{WVP} = \max \left\{ \sum_{i \in \mathcal{V}} \left(\mu_i x_i + \sigma_i \sqrt{x_i(1-x_i)} \right) \ : \boldsymbol{x} \in \text{conv}(\boldsymbol{\mathcal{X}}_{WVP}) \right\}, \tag{26}$$

where $\boldsymbol{\mathcal{X}}_{WVP}$ is the feasible region in formulation (24). The persistency of each binary variable x_i is obtained by using the optimal solution to the second-order cone program. However, for the vertex packing problem, the convex hull of feasible region is not easily characterizable. In this case, it is appealing to use linear programming relaxation to approximate $\text{conv}(\boldsymbol{\mathcal{X}}_{WVP})$. This results in a weaker upper bound,

$$\bar{Z}_{WVP} = \max \left\{ \sum_{i \in \mathcal{V}} \mu_i x_i + \sigma_i \sqrt{x_i(1-x_i)} \ : x_i + x_j \leq 1 \ \forall (i,j) \in \mathcal{E}, \ x_i \geq 0 \ \forall i \in \mathcal{V} \right\}. \tag{27}$$

Clearly $Z_{WVP} \leq \bar{Z}_{WVP}$. We use the simple graph in Figure 2 to compare the persistency values.

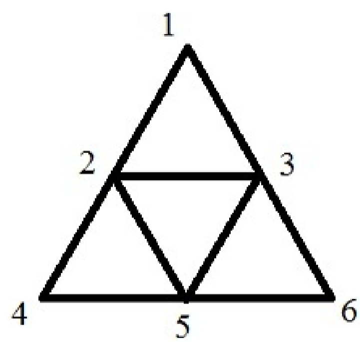

Figure 2: Vertex packing example

The set of all feasible solutions for this problem is

$$\boldsymbol{\mathcal{X}} = \left\{ \begin{array}{lllll} (1,0,0,0,0,0), & (0,1,0,0,0,0), & (0,0,1,0,0,0), & (0,0,0,1,0,0), & (0,0,0,0,1,0), \\ (0,0,0,0,0,1), & (1,0,0,1,0,0), & (1,0,0,0,0,1), & (0,0,0,1,0,1), & (1,0,0,0,1,0), \\ (0,1,0,0,0,1), & (0,0,1,1,0,0), & (1,0,0,1,0,1), & (0,0,0,0,0,0) & \end{array} \right\}.$$

Assume $\sigma_i = \sigma$, $\forall i \in \mathcal{V}$. Let $\boldsymbol{x}_{WVP}(\boldsymbol{\mu}, \sigma)$ and $\bar{\boldsymbol{x}}_{WVP}(\boldsymbol{\mu}, \sigma)$ denote the optimal solutions to (26) and (27), respectively. In Table 4, two sets of mean parameters are considered. The first column corresponds to mean $\boldsymbol{\mu} = (2,1,1,1,1,1)$ for which the deterministic problem has the unique optimal solution $(1,0,0,1,0,1)$. The second column corresponds to mean $\boldsymbol{\mu} = (3,1,1,3,6,3)$ for which the deterministic problem has two optimal solutions $(1,0,0,1,0,1)$ and $(1,0,0,0,1,0)$. In the first example as the standard deviation $\sigma \to 0$,

$\boldsymbol{x}_{WVP}(\boldsymbol{\mu},\sigma) \to \boldsymbol{x}_{WVP}(\boldsymbol{\mu})$. However in the second example this is not true due to the presence of multiple optimal solutions. Specifically, the nonlinear part of the objective function $\sum_{i\in\mathcal{V}} \sigma_i \sqrt{x_i(1-x_i)}$ pulls the optimal solution to the middle of the true optimal solutions. The more tractable linear programming formulation has a similar behavior for small values of σ. However for larger values of σ, the two solutions $\boldsymbol{x}_{WVP}(\boldsymbol{\mu},\sigma)$ and $\bar{\boldsymbol{x}}_{WVP}(\boldsymbol{\mu},\sigma)$ can be much further apart.

	$\boldsymbol{\mu} = (2,1,1,1,1,1)$	$\boldsymbol{\mu} = (3,1,1,3,6,3)$
$\boldsymbol{x}_{WVP}(\boldsymbol{\mu},1)$	$(0.7582, 0.1209, 0.1209, 0.6139, 0.2652, 0.6139)$	$(0.9484, 0.0258, 0.0258, 0.4914, 0.4828, 0.4914)$
$\bar{\boldsymbol{x}}_{WVP}(\boldsymbol{\mu},1)$	$(0.5822, 0.4178, 0.4178, 0.5822, 0.4178, 0.5822)$	$(0.6581, 0.3419, 0.3419, 0.5000, 0.5000, 0.5000)$
$\boldsymbol{x}_{WVP}(\boldsymbol{\mu},0.1)$	$(0.9949, 0.0026, 0.0026, 0.9780, 0.0194, 0.9780)$	$(0.9994, 0.0003, 0.0003, 0.4999, 0.4998, 0.4999)$
$\bar{\boldsymbol{x}}_{WVP}(\boldsymbol{\mu},0.1)$	$(0.9287, 0.0713, 0.0713, 0.9287, 0.0713, 0.9287)$	$(0.9789, 0.0211, 0.0211, 0.5000, 0.5000, 0.5000)$
$\boldsymbol{x}_{WVP}(\boldsymbol{\mu},0.01)$	$(0.9999, 0.0000, 0.0000, 0.9998, 0.0002, 0.9998)$	$(1.0000, 0.0000, 0.0000, 0.4999, 0.5001, 0.4999)$
$\bar{\boldsymbol{x}}_{WVP}(\boldsymbol{\mu},0.01)$	$(0.9991, 0.0009, 0.0009, 0.9991, 0.0009, 0.9991)$	$(0.9998, 0.0002, 0.0002, 0.4999, 0.5001, 0.4999)$
$\lim_{\sigma\to 0} \boldsymbol{x}_{WVP}(\boldsymbol{\mu},\sigma)$	$(1,0,0,1,0,1)$	$(1,0,0,0.5,0.5,0.5)$
$\lim_{\sigma\to 0} \bar{\boldsymbol{x}}_{WVP}(\boldsymbol{\mu},\sigma)$	$(1,0,0,1,0,1)$	$(1,0,0,0.5,0.5,0.5)$

Table 4: Persistency $\boldsymbol{x}_{WVP}(\boldsymbol{\mu},\sigma)$ and $\bar{\boldsymbol{x}}_{WVP}(\boldsymbol{\mu},\sigma)$ for different values of the mean $\boldsymbol{\mu}$ and standard deviation σ

4.3 Discrete choice models

The results described earlier indicate that random discrete optimization problem with a polynomial number of extreme points can be analyzed using compact convex programs. One important example is the class of discrete choice models. These models predict the probability that customers choose an item from a finite set of alternatives. Consider a set of alternatives $\mathcal{N} = \{1,\ldots,n\}$. Assume that the utility that an individual customer assigns to alternative $k \in \mathcal{N}$ is given by

$$c_k = v_k + \epsilon_k,$$

where v_k is the deterministic component that relates to the known attributes of the alternative, and ϵ_k is the random error associated with the model due to uncontrolled factors. The random utility maximization problem faced by the customer is then formulated as

$$Z(\boldsymbol{c}) = \max\left\{\sum_{k\in\mathcal{N}} c_k x_k : \sum_{k\in\mathcal{N}} x_k = 1;\ x_k \in \{0,1\},\ \forall k \in \mathcal{N}\right\}.$$

Let P_j denote the probability that alternative j is selected by the customer. This choice probability is the persistency value,

$$P_k = P(x_k(\boldsymbol{c}) = 1) = P\left(c_k \geq c_j,\ \forall j \in \mathcal{N}\right).$$

The classical logit model for choice prediction starts with the assumption that the error terms ϵ_k's are modeled by independent extreme value distributions,

$$F(\epsilon_k \leq t) = e^{-e^{-t}},$$

for which the following elegant closed form solution for the choice probabilities can be obtained,

$$P_k = \frac{e^{v_k}}{\sum\limits_{j \in \mathcal{N}} e^{v_j}}.$$

However, this approach has some drawbacks. For example, the formula implies the Independence of Irrelevant Alternatives (IIA) property wherein the relative ratio of the choice probabilities for two alternatives is independent of the remaining alternatives. This property is not always observed in practice where the entire choice set helps in determining the relative probabilities. The probit model, another classical choice prediction model, using correlated normal distributions, can overcome this shortcoming, but at the added cost of finding choice probabilities through extensive simulation. In this case, no simple closed-form solution exists.

An alternative approach to find probabilities in discrete choice models is through conic optimization methods we reviewed in the previous section. Suppose that the n dimensional vector of random errors $\boldsymbol{\epsilon}$ is characterized by marginal moments, i.e., the mean vector $\mathbf{0}$ and the second moment matrix $\mathbf{\Pi} \succ 0$. The tight upper bound on the expected random utility is found by solving the following moments problem,

$$Z = \sup \left\{ \int_{\Re_n} \max_{k \in \mathcal{N}} \left(v_k + \epsilon_k\right) d\theta(\boldsymbol{\epsilon}) : \int_{\Re_n} d\theta(\boldsymbol{\epsilon}) = 1, \int_{\Re_n} \boldsymbol{\epsilon} d\theta(\boldsymbol{\epsilon}) = \mathbf{0}, \int_{\Re_n} \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T d\theta(\boldsymbol{\epsilon}) = \mathbf{\Pi} \right\}. \tag{28}$$

The constraints in (28) ensure that θ is a joint probability distribution consistent with the mean and covariance matrix of the random variables. The equivalent primal semidefinite program using Theorem 8 is

$$\begin{aligned} Z = \max_{\lambda_k, \boldsymbol{w}_k, \boldsymbol{W}_k, k \in \mathcal{N}} \quad & \sum_{k \in \mathcal{N}} \boldsymbol{e}_k^T \boldsymbol{w}_k \\ \text{s.t.} \quad & \sum_{k \in \mathcal{N}} \begin{pmatrix} \lambda_k & \boldsymbol{w}_k^T \\ \boldsymbol{w}_k & \boldsymbol{W}_k \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{v}^T \\ \boldsymbol{v} & \boldsymbol{v}\boldsymbol{v}^T + \mathbf{\Pi} \end{pmatrix} \\ & \begin{pmatrix} \lambda_k & \boldsymbol{w}_k^T \\ \boldsymbol{w}_k & \boldsymbol{W}_k \end{pmatrix} \succeq 0 \qquad \forall k \in \mathcal{N}, \end{aligned} \tag{29}$$

where $\boldsymbol{e}_k$ is a vector of dimension n with 1 in its kth entry and 0 otherwise. In this formulation λ_k is the choice probability that alternative k is selected by the customer under the extremal distribution. Alternatively, we can describe $\boldsymbol{\epsilon}$ by marginal distributions, i.e., ϵ_j's distribution function is known as $F_j(\cdot)$.

From Theorem 13, the upper bound on the expected random utility is found by maximizing the concave function over the simplex, i.e.,

$$\begin{aligned} \sup_{\boldsymbol{x}} \quad & \sum_{k\in\mathcal{N}} \left(v_k x_k + \int_{1-x_k}^{1} F_k^{-1}(t)dt \right) \\ \text{s.t.} \quad & \sum_{k\in\mathcal{N}} x_k = 1 \\ & x_k \geq 0 \qquad\qquad \forall k \in \mathcal{N}. \end{aligned} \tag{30}$$

In this formulation x_k is the choice probability that alternative k is selected by the customer under the extremal distribution. Natarajan et al. [60] and Mishra et al. [55] provide detailed numerical comparisons of these conic optimization based choice models with classical discrete choice models such as logit and multinomial probit models.

4.4 Random walks and sequencing

In this section, we discuss the sequencing problem with random costs, which is a notoriously difficult problem. To illustrate the viability of the moment based approach, we compare the persistency values obtained from the SDP models for a simple random walk with the values that are exactly known. This example demonstrates that the distributionally robust assumption in the models are sufficient to obtain near exact approximations to the actual persistency values.

Let c_i, $i = 1, \ldots, n$ be a sequence of independent and identically distributed random variables. For each positive integer k, define the partial sums,

$$S_k = c_1 + \ldots + c_k, \ \forall k = 1, \ldots, n,$$

and $S_0 = 0$. The sequence $S_0, S_1, \ldots, S_n$ is a random walk. Let $M_n = \max\{S_0, \ldots, S_n\}$ denote the maximum partial sum in the first n steps, and $K_n = \min\{k : S_k = M_n\}$ denote the first time step at which the maximum partial sum is obtained in the first n steps. Note that both M_k and S_k are random variables. A basic problem in the random walk theory is to estimate the probability distribution of K_n, i.e.,

$$P\Big(K_n = k\Big) = P\Big(S_k > S_0, \ldots, S_k > S_{k-1}, S_k \leq S_{k+1}, \ldots, S_k \leq S_n\Big).$$

Define $p_0 = q_0 = 1$, $p_n = P(S_1 > 0, \ldots, S_n > 0)$, $q_n = P(S_1 \leq 0, \ldots, S_n \leq 0)$, and $a_n = P(S_n > 0)$, for all $n \geq 1$. The random walk theory provides an explicit form for this probability as

$$P\Big(K_n = k\Big) = p_k q_{n-k}, \ \forall n \geq k \geq 0.$$

For random variables c_i with a symmetric continuous distribution function, the limiting distribution is given by the arcsine law,

$$\lim_{n\to\infty} P\left(\frac{K_n}{n} < x\right) = \frac{2}{\pi}\text{arcsin}\sqrt{x},\ \forall x \in (0,1).$$

The corresponding limiting density function is:

$$\frac{1}{\pi\sqrt{x(1-x)}}.$$

Note that in contrast to common intuition, the arcsine law shows that the two end points ($k = 0$ or $k = n$) have the highest probability of attaining the maximum, while the minimum takes place around $k \approx n/2$. The problem is similar to a discrete choice problem, where the utility of alternative k is given by the summand $S_k = \sum_{j=1}^{k} c_j$. Figure 3 shows the choice prediction of the random walk model, based on simulation and the cross moment model, for $n = 20$, where c_i has mean $\mu_i = 0$, standard deviation $\sigma_i = 1$. The simulation results are based on a multivariate normal distribution. Interestingly, the figure clearly shows that the cross moment model approximately returns the arcsine law behaviour of the choice probabilities.

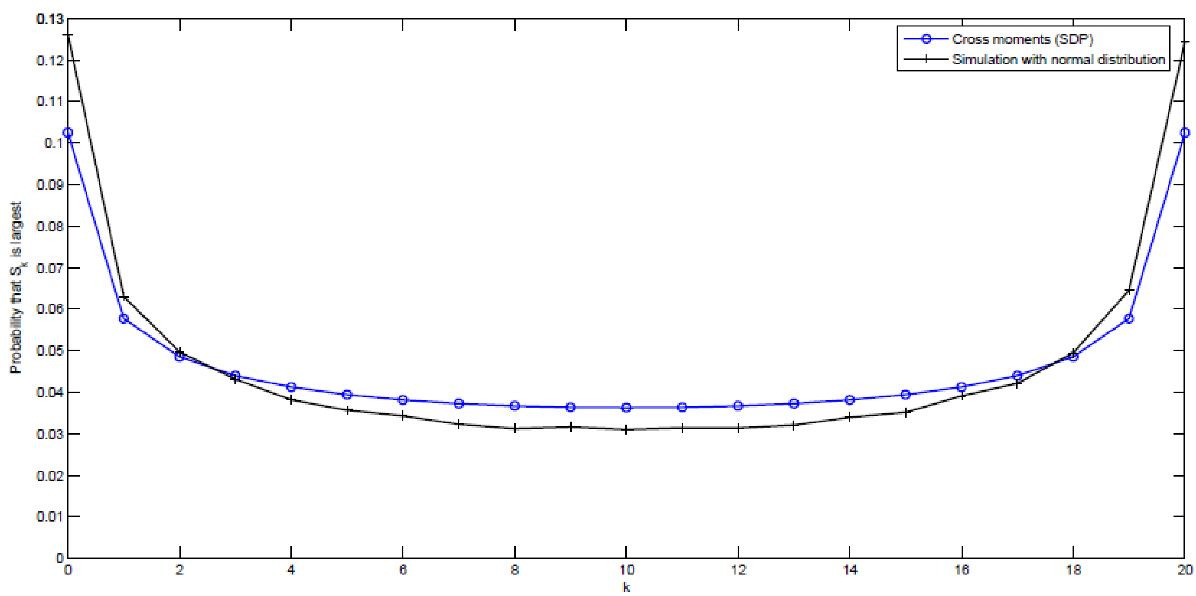

Figure 3: Comparison of probabilities under simulation and cross moment model

These techniques can be used to investigate complex sequencing problems. Consider an oil field exploration problem, where the objective is to determine the optimal sequence to explore a set of n oil fields. The valuation of oil field i is random and denoted by v_i. For ease of exposition, assume that the cost to explore each oil field is μ_i where $\mu_i = E(v_i)$. With a little abuse of notation, denote the maximum loss given a sequence π of explorations as

$$Z(\pi) = \max\left(0, \mu_{\pi(1)} - v_{\pi(1)}, (\mu_{\pi(1)} - v_{\pi(1)}) + (\mu_{\pi(2)} - v_{\pi(2)}), \ldots, \sum_{i=1}^{n}(\mu_{\pi(i)} - v_{\pi(i)})\right),$$

where $\pi(i) = j$ indicates that the oil field j is the ith to explore under the sequence π. The goal is to find the optimal sequence that minimizes the expected maximum loss given the joint distribution of the valuations denoted as θ, i.e.,

$$\min_{\pi} E_{\theta}\left(Z(\pi)\right).$$

An alternative approach is to use a distributional robust model where the exact joint distribution of the valuations is unknown, and θ is known only to lie in a set of distributions Θ. The distributional robust sequencing problem is formulated as

$$\min_{\pi} \sup_{\theta \in \Theta} E_{\theta}(Z(\pi)).$$

Suppose the valuation v_i has mean μ_i and variance σ_i^2, and valuations of different fields are uncorrelated to each other. The inner problem is similar to a discrete choice problem, where the utility function of the $(k+1)$th alternative is described as

$$\sum_{i=1}^{k}(\mu_{\pi(i)} - v_{\pi(i)}),$$

with mean 0 and variance $\sigma_{\pi(1)}^2 + \ldots + \sigma_{\pi(i)}^2$. A simple approximation is to use the marginal moment model which provides a distributional robust approximation to this problem,

$$\begin{aligned}
\min_{\pi} \quad \max_{\boldsymbol{y}} \quad & \sum_{i=1}^{n}\left(\sqrt{\sigma_{\pi(1)}^2 + \ldots + \sigma_{\pi(i)}^2}\sqrt{y_i(1-y_i)}\right) \\
\text{s.t.} \quad & \sum_{i=1}^{n} y_i = 1 \\
& y_i \geq 0 \qquad\qquad i = 1, \ldots, n.
\end{aligned} \tag{31}$$

It follows simply that the optimal sequence π in Formulation (31) is obtained by the $s$mallest variance first rule. i.e., explore the oil fields starting from the smallest to the largest variance. However, this sequencing rule is not optimal in general, when the distributions are explicitly given, or even when the cross moments are known. It remains an open problem to find the optimal sequencing rule in these cases. For other applications of the persistency model in appointment scheduling, we refer the readers to Kong et al. [40].

4.5 Newsvendor planning

Consider a newsvendor planning problem where the seller needs to determine the order quantity q of an item that maximizes her expected profit under random demand $\boldsymbol{D}$. The unit cost of the item is c and the selling price is p, with $p > c$. As the items are perishable, any unsold item will be deemed wasted. It is well known that the optimal ordering quantity q^* satisfies the critical fractile rule: $P(\boldsymbol{D} \leq q^*) = (p-c)/p$. Scarf [67] analyzed a maximin newsvendor problem who finds the optimal ordering quantity only based on the mean μ and variance σ^2 of the demand. In the maximin approach, the newsvendor chooses the order quantity that maximizes the minimum expected profit over all possible demand distributions Θ with the given mean and variance, i.e.,

$$\max_{q\geq 0} \quad \inf_{\theta\in\Theta} \quad pE_\theta(\min(q,\boldsymbol{D})) - cq. \tag{32}$$

To see that this problem fits into our framework, note that $\min(q, \boldsymbol{D})$ can be written as

$$\min\left\{qx + \boldsymbol{D}y : x + y = 1, x, y \geq 0\right\}.$$

For given mean and variance only, we have the following distributionally robust formulation for the newsvendor problem:

$$\begin{array}{llll} \max\limits_{q\geq 0} & \min\limits_{\lambda_1,\lambda_2,w_1,w_2,W_1,W_2} & pw_1 + pq\lambda_2 - cq & \\ & \text{s.t.} & \begin{pmatrix} \lambda_1 & w_1 \\ w_1 & W_1 \end{pmatrix} + \begin{pmatrix} \lambda_2 & w_2 \\ w_2 & W_2 \end{pmatrix} = \begin{pmatrix} 1 & \mu \\ \mu & \sigma^2+\mu^2 \end{pmatrix} & \\ & & \begin{pmatrix} \lambda_1 & w_1 \\ w_1 & W_1 \end{pmatrix} \succeq 0, \ \begin{pmatrix} \lambda_2 & w_2 \\ w_2 & W_2 \end{pmatrix} \succeq 0, \ w_1, w_2 \geq 0. & \end{array} \tag{33}$$

The inner semidefinite program in (33) decomposes the demand into two events: Event 1 when $D < q$, and Event 2 when $D \geq q$. By taking the dual of the inner SDP, the single item minimax newsvendor problem is formulated as a SDP which can be solved analytically as in Scarf [67]. For other results in this area, the reader is referred to the works of Popescu [65], Chen et. al. [18], Goh and Sim [27], Delage and Ye [20], Doan et. al. [21] and Zymler et. al. [77].

We end this section by providing the extension of the maximin approach to the multi-dimensional newsvendor problem. Let $\boldsymbol{q} \in \Re_m^+$ be the vector of resource order quantities, and $\boldsymbol{c}$ be the unit cost vector of the resources. The matrix $\boldsymbol{A}$ is the technology matrix whose (i,j) component represents the amount of resource i required to produce one unit of product j. Let $\boldsymbol{p} \in \Re_n^+$ be the unit revenue vector for the set of n products. The product demand is random and denoted by $\boldsymbol{D} \in \Re_n^+$. Harrison and Van Mieghem [30] formulated the multi-dimensional newsvendor problem as

$$\max_{\boldsymbol{q}\geq \boldsymbol{0}} \quad E(\Psi(\boldsymbol{q},\boldsymbol{D})) - \boldsymbol{c}^T\boldsymbol{q}, \tag{34}$$

where the recourse problem is

$$
\begin{aligned}
\Psi(\boldsymbol{q}, \boldsymbol{D}) = \quad \max_{\boldsymbol{y}} \quad & \boldsymbol{p}^T \boldsymbol{y} \\
\text{s.t.} \quad & \boldsymbol{y} \leq \boldsymbol{D} \\
& \boldsymbol{A}\boldsymbol{y} \leq \boldsymbol{q} \\
& \boldsymbol{y} \geq 0.
\end{aligned} \tag{35}
$$

The maximin multi-dimensional newsvendor problem is to find the resource vector that maximizes the minimum expected profit over all nonnegative demand distributions θ with the given mean vector $\boldsymbol{\mu}$ and second moment matrix $\boldsymbol{\Pi}$, i.e.,

$$
\max_{\boldsymbol{q} \geq \boldsymbol{0}} \quad \inf_{\theta \in \Theta} \quad E(\Psi(\boldsymbol{q}, \boldsymbol{D})) - \boldsymbol{c}^T \boldsymbol{q}. \tag{36}
$$

We show that the maximin multi-dimensional newsvendor problem can be reformulated as a copositive program. To see this, consider the more general distributionally robust stochastic linear program,

$$
Z = \min_{\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\, \boldsymbol{x} \geq \boldsymbol{0}} \left(\boldsymbol{c}^T \boldsymbol{x} + \sup_{\theta \in \Theta} \mathbb{E}_P \left[\mathcal{Q}(\tilde{\boldsymbol{h}}, \boldsymbol{x}) \right] \right), \tag{37}
$$

where the recourse problem is defined as

$$
\begin{aligned}
\mathcal{Q}(\tilde{\boldsymbol{h}}, \boldsymbol{x}) = \min_{\boldsymbol{w}} \quad & \boldsymbol{q}^T \boldsymbol{w} \\
\text{s.t.} \quad & \boldsymbol{W}\boldsymbol{w} \geq \tilde{\boldsymbol{h}} - \boldsymbol{T}\boldsymbol{x}.
\end{aligned}
$$

From the strong duality of linear programming problem,

$$
\begin{aligned}
\mathcal{Q}(\tilde{\boldsymbol{h}}, \boldsymbol{x}) = \max_{\boldsymbol{p}} \quad & (\tilde{\boldsymbol{h}} - \boldsymbol{T}\boldsymbol{x})^T \boldsymbol{p} \\
\text{s.t.} \quad & \boldsymbol{W}^T \boldsymbol{p} = \boldsymbol{q} \\
& \boldsymbol{p} \geq \boldsymbol{0}.
\end{aligned}
$$

By directly applying Theorem 11 on the inner problem of (37) and taking the dual of the completely positive program, the distributionally robust stochastic program can be formulated as the following copositive program,

$$
\begin{aligned}
Z = \min_{\boldsymbol{x}, \boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{Y}, \boldsymbol{y}, y_0} \quad & \boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{q}^T \boldsymbol{w}_1 + (\boldsymbol{q} \circ \boldsymbol{q})^T \boldsymbol{w}_2 + \boldsymbol{Q} \cdot \boldsymbol{Y} + \boldsymbol{\mu}^T \boldsymbol{y} + y_0 \\
\text{s.t.} \quad & \begin{pmatrix} y_0 & \boldsymbol{y}^T/2 & (\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{w}_1)^T/2 \\ \boldsymbol{y}/2 & \boldsymbol{Y} & -\boldsymbol{I}/2 \\ (\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{w}_1)/2 & -\boldsymbol{I}/2 & \boldsymbol{W}\mathrm{diag}(\boldsymbol{w}_2)\boldsymbol{W}^T \end{pmatrix} \in \mathcal{CO}_{2n+1} \\
& \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0},
\end{aligned} \tag{38}
$$

where $\boldsymbol{I}$ represents the identity matrix of appropriate dimension, and $\mathrm{diag}(\boldsymbol{w}_2)$ denotes the diagonal matrix formed with the diagonal entries being the entries of the vector $\boldsymbol{w}$. This formulation provides the generalization of the Scarf formulation to the multi-dimensional newsvendor problem.

5 Conclusion

In this paper, we review some of the recent advances on the distributional robust analysis of mixed integer linear programs using conic programming techniques. Evaluating bounds on the expected optimal value is a classical problem that has been well studied over several decades. However, it is only more recently that conic programming methods have been used in the probabilistic analysis of discrete optimization problems. The strong connection between the theory of moments and conic programming provides an important analytical tool for this class of problems. Besides bounds, these methods also aid in estimating parameters of interest such as the persistency of a binary variable. This review paper also discusses the complexity results for this class of problems, and a few of the important applications in areas such as activity networks, discrete choice models, random walks, and newsvendor problems.

Acknowledgements

The authors would like to thank the Editor Immanuel Bomze and two anonymous reviewers for their valuable comments and suggestions on improving the manuscript.

References

[1] W. P. Adams, J. B. Lassiter, H. D. Sherali. 1998. Persistency in 0-1 polynomial programming. *Mathematics of Operations Research*, 23, No. 2, 359-389.

[2] D. Aldous. 2001. The $\zeta(2)$ limit in the random assignment problem. *Random Structures and Algorithms*, 18, No. 4, 381-418.

[3] D. Aldous, J. M. Steele. 2003. The objective method: Probabilistic combinatorial optimization and local weak convergence. *Probability on Discrete Structures*, Edited by Harry Kesten, 1-72, Springer.

[4] Ben-Tal, A., L . El Ghaoui, A. Nemirovki. 2009. Robust Optimization. Priceton Series in Applied Mathematics.

[5] J. Beardwood, J. H. Halton, J. M. Hammersley. 1959. The shortest path through many points. *Proceedings of the Cambridge Philosophical Society*, 55, No. 4, 299-327.

[6] A. Berman, N. Shaked-Monderer. 2003. Completely Positive Matrices. Singapore: World Scientific.

[7] D. Bertsimas, D. B. Brown, C. Caramanis. 2011. Theory and Applications of Robust Optimization. *SIAM Review*, 53, 464-501.

[8] D. Bertsimas, X. V. Doan, K. Natarajan, C. P. Teo. 2010. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35, No. 3, 580-602.

[9] D. Bertsimas, K. Natarajan, C. P. Teo. 2004. Probabilistic combinatorial optimization: Moments, semidefinite programming and asymptotic bounds. *SIAM Journal of Optimization*, 15, No. 1, 185-209.

[10] D. Bertsimas, K. Natarajan, C. P. Teo. 2006. Persistence in discrete optimization under data uncertainty. *Mathematical Programming*, Series B, Issue on Opimization under Uncertainty, 108, No. 2-3, 251-274.

[11] D. Bertsimas, I. Popescu. 2002. On the relation between option and stock prices: A convex optimization approach. *Operations Research*, 50, No. 2, 358-374.

[12] J. R. Birge, M. J. Maddox. 1995. Bounds on expected project tardiness. *Operations Research*, 43, No. 5, 838-850.

[13] I. M. Bomze. 2012. Copositive optimization: Recent developments and applications. Invited Review. *European Journal of Operational Research*, 126, No. 3, 509-520.

[14] I. M. Bomze, E. de Klerk. 2002. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *Journal of Global Optimization*, 24, No. 2, 163-185.

[15] R. A. Bowman. 1995. Efficient estimation of arc criticalities in stochastic activity networks. *Management Science*, 41, No. 1, 58-67.

[16] P. P. Boyle, X. S. Lin. 1997. Bounds on contingent claims based on several assets. *Journal of Financial Economics*, 46, No. 3, 383-400.

[17] S. Burer. 2009. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming*, 120, No. 2, 479-495.

[18] W. Chen, M. Sim, J. Sun and C. P. Teo. 2010. From CVaR to Uncertainty Set: Implications in Joint Chance Constrained Optimization. *Operations Research*, 58, 470-485.

[19] P. H. Diananda. 1962. On nonnegative forms in real variables some or all of which are nonnegative. *Mathematical Proceedings of the Cambridge Philosophical Society*, 58, 17-25.

[20] E. Delage, Y, Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58, No. 3, 596-612.

[21] X. V. Doan, K. Natarajan. 2011. On the complexity of non-overlapping multivariate marginal bounds for probabilistic combinatorial optimization problems. *Operations Research*, 60, No. 1, 138-149, 2012.

[22] B. Dodin. 1985. Bounding the project completion time distribution in PERT networks. *Operations Research*, 33, No. 4, 862-881.

[23] B. Dodin, S. E. Elmaghraby. 1985. Approximating the criticality indices of activites in PERT networks. *Management Science*, 31, No. 2, 207-223.

[24] M. Dür. 2010. Copositive programming: A survey. In M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels, editors. *Recent Advances in Optimization and its Applications in Engineering*, 3-20. Springer Berlin Heidelberg.

[25] D. R. Fulkerson. 1962. Expected critical path length in PERT networks. *Operations Research* 10, No. 6, 808-817.

[26] M. Grötschel, L. Lovász, A. Schrijver. 1988. Geometric Algorithms and Combinatorial Optimization. Springer, Berlin, Heidelberg.

[27] J. Goh, M. Sim. 2010. Distributionally Robust Optimization and its Tractable Approximations. *Operations Research*, 58, No. 4, 902-917.

[28] J. N. Hagstrom. 1988. Computational complexity of PERT problems. *Networks*, 18, No. 2, 139-147.

[29] P. L. Hammer, P. Hansen, B. Simeone. 1984. Roof duality, complementation and persistency in quadratic 0-1 optimization. *Mathematical Programming*, 28, No. 2, 121-155.

[30] J. M. Harrison, J. A. Van Mieghem. Multi-resource investment strategies: Operational hedging under demand uncertainty. *European Journal of Operational Research*, 113, 17-29.

[31] J.-B. Hiriart-Urruty, A. Seeger. 2010. A variational approach to copositive matrices. *SIAM Review*, 52, No. 4, 593-629.

[32] D. S. Hochbaum, N. Megiddo, J. Naor, A. Tamir. 1993. Tight bounds and 2-approximation algorithms for integer programs with two variables per inequality. *Mathematical Programming*, 62, No. 1-3, 69-83.

[33] L. Kantorovich. 1958. On the translocation of masses. *Management Science*, 5, No. 1, 1-4.

[34] S. Karlin, W. J. Studden. 1966. *Tchebycheff Systems: With Applications in Analysis and Statistics* (Pure and applied mathematics), Interscience Publishers.

[35] R. M. Karp. 1979. A patching algorithm for the nonsymmetric traveling-salesman problem. *SIAM Journal on Computing*, 8, No. 4, 561-573.

[36] J. H. B. Kemperman, M. Skibinsky. 1992. Covariance spaces for measures on polyhedral sets. *Stochastic Inequalities*, IMS Lecture Notes - Monograph Series, 22, 182-195.

[37] W. K. Klein Haneveld. 1986. Robustness against dependence in PERT: An application of duality and distributions with known marginals. *Mathematical Programming Studies*, 27, 153-182.

[38] G. B. Kleindorfer. 1971. Bounding distributions for a stochastic acyclic network. *Operations Research*, 19, No. 7, 1586-1601.

[39] P. Krokhmal, P. Pardalos. 2009. Random assignment problems. *European Journal of Operational Research*, 194, No. 1, 1-17.

[40] Q. Kong, C. Y Lee, CP Teo and Z. Zheng. 2013. Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research*, 61, No. 3, 711-726.

[41] J. B. Lasserre. 2001. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11, No. 3, 796-817.

[42] J. B. Lasserre. 2002. Bounds on measures satisfying moment conditions. *Annals of Applied Probability*, 12, No. 3, 1114-1137.

[43] J. B. Lasserre. 2009. Moments, Positive Polynomials and Their Applications. Imperial College Press.

[44] J. B. Lasserre. 2010. A "joint+marginal" approach to parametric polynomial optimization. *SIAM Journal of Optimization*, 20, 1995-2022.

[45] M. Laurent. 2003. A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. *Mathematics of Operations Research*, 28, 470-496.

[46] S. Linusson, J. Wästlund. 2004. A proof of Parisi's conjecture on the random assignment problem. *Probability Theory and Related Fields*, 128, No. 3, 419-440.

[47] L. Lovász. 2001. Energy of convex sets, shortest paths and resistance. *Journal of Combinatorial Theory, Series A*, 94, 363-382.

[48] S. H. Lu, A. C. Williams. 1987. Roof duality for polynomial 0-1 optimization. *Mathematical Programming*, 37, No. 3, 357-360.

[49] R. Lyons, R. Pemantle, Y. Peres. 1999. Resistance bounds for first-passage percolation and maximum flow. *Journal of Combinatorial Theory, Series A*, 86, No. 1, 158-168.

[50] D. G. Malcolm, J. H. Roseboom, C. E. Clark, W. Fazar. 1959. Application of a technique for research and development program evaluation. *Operations Research*, 7, No. 5, 646-669.

[51] O. J. Mangasarian, T. H. Shiau. 1986. A variable-complexity norm maximization problem. *SIAM Journal on Algebraic and Discrete Methods*, 7, No. 3, 455-461.

[52] A. W. Marshall, I. Olkin. 1960. Multivariate Chebyshev Inequalities. *Annals of Mathematical Statistics*, 31, No. 4, 1001-1014.

[53] I. Meilijson, A. Nadas. 1979. Convex majorization with an application to the length of critical path. *Journal of Applied Probability*, 16, No. 3, 671-677.

[54] M. Mézard, G. Parisi. 1985. Replicas and optimization. *Journal de Physique Lettres*, 46, 771-778.

[55] V. K. Mishra, K. Natarajan, H. Tao, C. P. Teo. 2012. Choice prediction with semidefinite optimization when utilities are correlated. *IEEE Transactions on Automatic Control*, 57, No. 10, 2450-2463.

[56] R. H. M. Möhring. 2001. Scheduling under uncertainty: Bounding the makespan distribution. In H. Alt, editor, Computational Discrete Mathematics: Advanced Lectures, volume 2122 of Lecture Notes in Computer Science, 7997, Springer.

[57] G. Monge. 1781. Mémoire sur la théorie des déblais et des remblais. Histoire de lAcadémie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la méme année, 666-704.

[58] K. G. Murty, S. N. Kabadi. 1987. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39, No. 2, 117-129.

[59] C. Nair, B. Prabhakar, M. Sharma. 2005. Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures. *Random Structures & Algorithms*, 27, No. 4, 413-444.

[60] K. Natarajan, M. Song, C. P. Teo. 2009. Persistency model and its applications in choice modeling. *Management Science*, 55, No. 3, 453-469.

[61] K. Natarajan, C. P. Teo, Z. Zheng. 2011. Mixed zero-one linear programs under ojective uncertainty: A completely positive representation. *Operations Research*, 59, No. 3, 713-728.

[62] G. L. Nemhauser, L. E. Trotter. 1975. Vertex packings: structural properties and algorithms. *Mathematical Programming*, 8, No. 1, 232-248.

[63] P. A. Parillo. 2000. Structured Semidefinite Programs and Semi-algebraic Geometry Methods in Robustness and Optimization, Ph.D. thesis, California Institute of Technology.

[64] G. Parisi. 1998. A conjecture on random bipartite matching. Physics e-Print archive.

[65] I. Popescu. 2007. Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55, No. 1, 98-112.

[66] S. Z. Rachev, L. Rüschendorf. 1998. *Mass Transportation Problems* (Probability and Its Applications), Springer.

[67] H. Scarf. 1958. A min-max solution of an inventory problem. *Studies in The Mathematical Theory of Inventory and Production*, Stanford University Press, California, 201-209.

[68] A. Shapiro. 2001. On duality theory of conic linear problems. Miguel A. Goberna and Marco A. Lṕpez, Eds., *Semi-Infinite Programming: Recent Advances*, Kluwer Academic Publishers, 135-165.

[69] J. E. Smith. 1995. Generalized Chebychev inequalities: Theory and applications in decision analysis. *Operations Research*, 43, No. 5, 807-825.

[70] J. M. Steele. 1997. *Probability Theory and Combinatorial Optimization.* CBMS-NSF Regional Conference Series in Applied Mathematics, 69.

[71] J. F. Sturm, S. Zhang. 2003. On cones of nonnegative quadratic functions. *Mathematics of Operations Research*, 28, No. 2, 246-267.

[72] L. Vandenberghe, S. Boyd, K. Comanor. 2007. Generalized Chebyshev bounds via Semidefinite Programming. *SIAM Review*, 49, No. 1, 52-64.

[73] R. M. Van Slyke. 1963. Monte Carlo methods and the PERT problem. *Operations Research*, 11, No. 5, 839-860.

[74] G. Weiss. 1986. Stochastic bounds on distributions of optimal value function with applications to PERT, network flows and reliability. *Operations Research*, 34, No. 4, 595-605.

[75] W. Whitt. 1984. On approximations for queues, I. Extremal distributions. *AT&T Bell Labs Technical Journal*, 63 115138.

[76] L. Zuluaga, J. F. Pena. 2005. A conic programming approach to generalized tchebycheff inequalities. *Mathematics of Operations Research*, 30, No. 2, 369-388.

[77] S. Zymler, D. Kuhn, B. Rustem. 2013. Distributionally robust joint chance constraints with second-order moment informmation. *Mathematical Programming*, 137, No. 1-2, 167-198.