OPERATIONS RESEARCH
Vol. 60, No. 1, January–February 2012, pp. 138–149
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.1110.1005

On the Complexity of Nonoverlapping Multivariate Marginal Bounds for Probabilistic Combinatorial Optimization Problems

Xuan Vinh Doan
DIMAP and ORMS Group, Warwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom, xuan@doanwbs.ca.uk

Karthik Natarajan
Department of Management Sciences, College of Business, City University of Hong Kong, Hong Kong, knataraj@cityu.edu.hk

Given a combinatorial optimization problem with an arbitrary partition of the set of random objective coefficients, we evaluate the tightest-possible bound on the expected optimal value for joint distributions consistent with the given multivariate marginals of the subsets in the partition. For univariate marginals, this bound was first proposed by Meilijson and Nadas [Meilijson, I., A. Nadas. 1979. Convex majorization with an application to the length of critical path. *J. Appl. Probab.* **16**(3) 671–677]. We generalize the bound to nonoverlapping multivariate marginals using multiple-choice integer programming. New instances of polynomial-time computable bounds are identified for discrete distributions. For the problem of selecting up to M items out of a set of N items of maximum total weight, the multivariate marginal bound is shown to be computable in polynomial time, when the size of each subset in the partition is $O(\log N)$. For an activity-on-arc PERT network, the partition is naturally defined by subsets of incoming arcs into nodes. The multivariate marginal bound on expected project duration is shown to be computable in time polynomial in the maximum number of scenarios for any subset and the size of the network. As an application, a polynomial-time solvable two-stage stochastic program for project crashing is identified. An important feature of the bound developed in this paper is that it is exactly achievable by a joint distribution, unlike many of the existing bounds.

Subject classifications: probability bounds; integer programming; PERT.
Area of review: Optimization.
History: Received October 2010; revisions received March 2011, May 2011; accepted June 2011.

1. Introduction

The analysis of combinatorial optimization problems with random objective coefficients is an important but challenging problem. As motivation, we consider a timing analysis application arising in the design and analysis of digital integrated circuits. The timing characteristic of a circuit is significantly affected by variations in fabrication process parameters and variations in operating environmental factors such as temperature and supply voltages. A detailed description on the different sources of variations that influence circuit behavior can be found in the book of Sapatnekar (2004). In conjunction with shrinking device sizes, there has been an increasing use of statistical methods in the design and analysis of digital circuits (Agarwal et al. 2003, Visweswariah et al. 2006, Blaauw et al. 2008). This is known as *statistical static timing analysis* (SSTA). Although early research in this area dates back to the 1960s–1970s (Kirkpatrick and Clark 1966, Nadas 1979), the past decade has seen a widespread adoption of SSTA by the electronic design automation (EDA) community. SSTA has now been integrated into several EDA software tools such as the IBM EinsStat, the Altos Variety, and the Cadence Encounter Timing System. The basic model in SSTA is motivated from the Project Evaluation and Review Technique (PERT) in project management. A PERT network is a directed acyclic graph representation of a project that consists of several activities with partially specified precedence relationships among the activities. The project completion time is given by the length of the longest path on this graph between a fixed start and sink node. The goal is to estimate the probability distribution and moments of the project completion time, given probability distribution information on the random activity durations. In SSTA, the network is a timing graph with nodes representing input and output pins of gates and arcs representing input-output delays of gates. Using the probability distributions of the individual input-output delays of gates, SSTA aims to find the distribution and expected latest arrival time at the sink node of the timing graph. However, in contrast with deterministic analysis, it is well known that the probabilistic analysis of PERT networks is much more difficult (see Hagstrom 1988). The timing graphs of industrial application-specific integrated circuits are also very large (often with millions of gates) with gate delays that are possibly dependent on each other. Developing tractable methods that can perform accurate probabilistic analysis of these networks remains a significant challenge. Our goal in this paper is to propose a new approach that provides efficiently computable bounds on such networks while capturing partial dependence information among the random variables. The approach is inspired by a bound proposed by Meilijson and Nadas (1979) and Nadas (1979), which we review next.

Consider a generic linear combinatorial optimization problem in maximization form with objective coefficients $\mathbf{c} = (c_1, c_2, \ldots, c_N)$:

$$Z(\mathbf{c}) = \max \quad \mathbf{c}'\boldsymbol{x} \qquad \text{s.t.} \quad \boldsymbol{x} \in \mathscr{X} \subseteq \{0, 1\}^N. \tag{1}$$

Suppose $\tilde{\mathbf{c}} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_N)$ is random. Then, the optimal value $Z(\tilde{\mathbf{c}})$ is random and is the object of our interest. To avoid trivialities, we assume that the sets $\{\mathbf{x} \in \mathscr{X} \mid x_i = 0\}$ and $\{\mathbf{x} \in \mathscr{X} \mid x_i = 1\}$ are nonempty for each index $i = 1, \ldots, N$. Meilijson and Nadas (1979) proposed an upper bound on the expected optimal objective value in (1) using only the univariate marginal distributions of $\tilde{\mathbf{c}}$. Their problem was motivated in the context of finding a worst-case upper bound on the expected project duration that is valid over all joint distributions of the activity durations with the given marginals. The upper bound was obtained through the solution of the following convex minimization problem over the decision variables $\mathbf{d} = (d_1, d_2, \ldots, d_N) \in \mathbb{R}^N$:

$$\mathbb{E}[Z(\tilde{\mathbf{c}})] \leqslant \inf_{\mathbf{d}} \left(Z(\mathbf{d}) + \sum_{i=1}^{N} \mathbb{E}[\tilde{c}_i - d_i]^+ \right), \tag{2}$$

where $[y]^+ = \max(y, 0)$. Furthermore, they showed that the bound in (2) was tight by constructing a joint distribution for $\tilde{\mathbf{c}}$ with the correct marginals that attained the upper bound exactly. The bound can thus be interpreted as being *robust against dependence*. For PERT networks, Klein Haneveld (1986) interpreted the formulation on the right-hand side of (2) as finding reference values $\mathbf{d}$ for the durations of the activities, such that the project completion time based on $\mathbf{d}$ is balanced with the sum of the expected delays of the activity durations beyond $\mathbf{d}$. For $Z(\mathbf{c}) = \max_i c_i$, this bound reduces to the maximally dependent bound of Lai and Robbins (1976):

$$\mathbb{E}\left[\max_{i=1,\ldots,N} \tilde{c}_i \right] \leqslant \inf_{d} \left(d + \sum_{i=1}^{N} \mathbb{E}[\tilde{c}_i - d]^+ \right). \tag{3}$$

The result extends to the *increasing convex order* bound, which provides a tight upper bound on $\mathbb{E}[Z(\tilde{\mathbf{c}}) - T]^+$ for a given T. For $Z(\mathbf{c}) = \sum_i c_i$, this reduces to the comonotonic upper bound discussed in Rüschendorf (1983):

$$\mathbb{E}\left[\sum_{i=1}^{N} \tilde{c}_i - T \right]^+ \leqslant \inf_{\sum_i d_i = T} \left(\sum_{i=1}^{N} \mathbb{E}[\tilde{c}_i - d]^+ \right). \tag{4}$$

McNeil et al. (2005) have discussed the relevance of these bounds to the actuarial sciences and portfolio risk management community. Weiss (1986) evaluated the bound for combinatorial optimization problems such as the shortest-path, maximum flow, and reliability problem. Extensions to incompletely specified univariate marginal distributions with moment information have been proposed in Klein Haneveld (1986), Birge and Maddox (1995), and Bertsimas et al. (2004, 2006). Meilijson (1991), and Natarajan et al. (2009) have extended the univariate marginal bound to integer programs using a binary reformulation.

In this paper, we generalize the result for probabilistic combinatorial optimization problems by assuming that information on nonoverlapping multivariate marginals are available. A popular tool to construct multivariate distributions from univariate distributions is the *copula* that helps distinguish the dependencies from the marginals. Formally, an N-dimensional copula is defined as a distribution function on the unit hypercube $[0, 1]^N$ with standard uniform marginal distributions (see McNeil et al. 2005). Sklar (1959) showed that for all multivariate distributions F with marginal distributions $F_1, F_2, \ldots, F_N$, there exists a copula $\mathrm{C}\colon [0, 1]^N \to [0, 1]$ such that

$$F(c_1, c_2, \ldots, c_n) = \mathrm{C}\big(F_1(c_1), F_2(c_2), \ldots, F_N(c_N)\big) \quad \text{for all } (c_1, c_2, \ldots, c_n) \in [-\infty, \infty]^N.$$

For a multivariate distribution with continuous marginals, the copula is uniquely defined as

$$\mathrm{C}(u_1, u_2, \ldots, u_n) = F\big(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_N^{-1}(u_N)\big) \quad \text{for all } (u_1, u_2, \ldots, u_n) \in [0, 1]^N.$$

The copula can be used for constructing multivariate discrete distributions too. However, the copula might no longer be unique. As compared to univariate marginals, analysis under multivariate marginals is far more challenging. The concept of a copula is known to be inadequate in this setting (see Scarsini 1989). Genest et al. (1995) showed that the only copula consistent with all nonoverlapping multivariate marginals is the independence copula. Li et al. (1996b) proposed a linkage function to characterize distributions with nonoverlapping multivariate marginals by emphasizing the separate roles of the dependence structure between the marginals, and the dependence structure within each of the marginals. Difficulties in constructing distributions with prescribed multivariate marginals has also resulted in fewer known bounds. The reader is referred to Li et al. (1996a), Rüschendorf (2004), and Embrechts and Pucetti

(2006) for some of the known bounds. However, none of these bounds are directly applicable to the combinatorial optimization problem.

Our interest in multivariate marginal bounds are motivated by realistic assumptions on dependence information among random parameters in applications from different areas such as risk management and PERT networks discussed earlier:

(a) Consider an N-dimensional vector of nonnegative random losses that can be partitioned into subvectors representing losses for policies within specific risk categories. The goal is to compute the worst-case expected aggregate loss of a financial position, given subvector loss distributions but allowing for arbitrary dependencies between subvectors. These subvectors could represent losses from companies in industry sectors such as health care, energy, and the Internet, or from countries in different geographical locations. Our focus is on aggregate loss defined by the sum of the M highest losses from the set of N losses. For $M = 1$, this reduces to maximum loss, whereas for $M = N$, this reduces to sum of the losses.

(b) Consider the estimation of the expected project duration in an activity-on-arc PERT network with random activity durations. A simplifying assumption often made in the analysis of PERT networks is statistical independence among the activity durations. Ball et al. (1995) and Möhring and Radermacher (1989) review methods that compute, bound, or approximate the expected project duration with independent activity durations. Ringer (1971) and van Dorp and Duffey (1999), however, argue that in construction projects the activity durations are often correlated due to dependence on factors such as weather, manpower skills, site conditions, and supervision quality. Fulkerson (1962) proposed a lower bound by using dependence information among activity durations incoming into each node. For his result to be a valid lower bound, an explicit assumption of independence among the activity durations entering different nodes needed to be made. Kleindorfer (1971) and Shogan (1983) developed both upper and lower bounds using dependencies among durations of activities entering a node and independence among activities entering different nodes. On the other hand, we are interested in developing a worst-case upper bound on the expected completion time that uses dependency information for activities entering a node, but does not assume independence among activities entering different nodes. For the SSTA application discussed earlier, a common assumption is that the input-output delays of gates that are nearby are highly correlated due to spatial proximity (see Sapatnekar 2004, Visweswariah et al. 2006). Our worst-case estimate on the expected timing of this circuit would then be consistent with the correlation information of nearby gates and allow for factors that might make the delays of gates that are far away arbitrarily correlated.

1.1. Problem Description

The formal description of the problem is provided next. Consider a partition of the index set $\mathcal{N} = \{1, 2, \ldots, N\}$ into subsets $\mathcal{N}_1, \ldots, \mathcal{N}_R$ such that

$$\mathcal{N} = \bigcup_{r=1}^{R} \mathcal{N}_r \quad \text{and} \quad \mathcal{N}_r \cap \mathcal{N}_s = \varnothing \quad \text{for all } r \neq s.$$

Given a vector $\boldsymbol{c} \in \mathbb{R}^N$, let $\boldsymbol{c}_r \in \mathbb{R}^{N_r}$ denote the subvector formed with the elements in the rth subset $\mathcal{N}_r$ where $N_r = |\mathcal{N}_r|$ is the size of the subset. The probability measures P_r for the subvectors $\tilde{\mathbf{c}}_r$ are assumed to be known. Let $\mathcal{P}(P_1, \ldots, P_R)$ denote the set of joint probability measures for the random vector $\tilde{\mathbf{c}}$ consistent with the prescribed probability measures for the subvectors $\tilde{\mathbf{c}}_r$. No assumption on the dependencies between random variables in distinct subsets are made. The independence measure among the subvectors thus forms one feasible distribution. For $R > 1$, the joint distribution is incompletely specified. For $R = N$, only the univariate marginals are specified. Our goal is to compute the supremum of the expected optimal objective value in (1) consistent with the nonoverlapping multivariate marginals:

$$Z^* = \sup_{P \in \mathcal{P}(P_1, \ldots, P_R)} \int Z(\mathbf{c})\, dP(\mathbf{c}). \tag{5}$$

For ease of exposition, we restrict our attention in the paper to discrete multivariate marginals with bounded support. It is useful to note that Theorems 1 and 3 and Propositions 1 and 2 are also applicable to continuous multivariate marginals with finite second moments.

Assumption. *The discrete probability distribution for the subvectors $\tilde{\mathbf{c}}_r$ are defined by the scenarios $\mathbf{c}_{rk}$ for $k = 1, \ldots, K_r$ with probabilities p_{rk} satisfying $\sum_k p_{rk} = 1$:*

$$P_r(\tilde{\mathbf{c}}_r = \mathbf{c}_{rk}) = p_{rk} \quad \textit{for all } k = 1, \ldots, K_r,\ r = 1, \ldots, R.$$

We obtain the following key results:

(a) In §2, we generalize the Meilijson and Nadas (1979) bound to nonoverlapping multivariate marginals. Using an expanded set of decision variables, the computation of the tight bound Z^* is shown to be related to solving a multiple-choice integer program. This leads to a polynomial-time computable bound for the problem selecting up to M items out a set of N items of maximum total weight when the size of each subset in the partition is $O(\log N)$. This extends the polynomial complexity result of Meilijson and Nadas (1979), where the size of each subset in the partition is $O(1)$.

(b) In §3, we identify a weaker upper bound based on a reduced-integer program. A condition is identified under which the bound is tight. This leads to polynomial-time computable bounds for worst-case expected project duration in PERT networks with the partition defined by subsets of incoming arcs into nodes. A two-stage stochastic program in project crashing is identified for which a polynomial-time algorithm is provided.

2. A Multivariate Marginal Formulation

Let $\mathscr{X}_r$ denote the projection of $\mathscr{X}$ onto the space of the decision variables in the rth subset:

$$\mathscr{X}_r = \text{proj}_r(\mathscr{X}) = \{\mathbf{x}_r \mid \mathbf{x} \in \mathscr{X}\} \subseteq \{0,1\}^{N_r}.$$

For example, the projection onto the space of a single variable is the set $\{0,1\}$. Our first theorem provides the generalization of the Meilijson and Nadas bound in (2) using an expanded set of decision variables.

THEOREM 1. *Let $\mathbf{d}_r = (d_r(\mathbf{x}_r))_{\boldsymbol{x}_r \in \mathscr{X}_r}$ be a decision vector for $r = 1, \dots, R$ with $d_r(\mathbf{0}) = 0$. Define*

$$\hat{Z}_u^* = \min_{\mathbf{d}_1,\dots,\mathbf{d}_R} \left(\max_{\mathbf{x}\in\mathscr{X}} \sum_{r=1}^R d_r(\boldsymbol{x}_r) + \sum_{r=1}^R \mathbb{E}_{P_r}\left[\max_{\boldsymbol{x}_r\in\mathscr{X}_r}(\tilde{\mathbf{c}}_r'\boldsymbol{x}_r - d_r(\boldsymbol{x}_r))\right]\right). \tag{6}$$

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

PROOF.

Step 1. **Prove that $Z^* \leqslant \hat{Z}_u^*$.** For any feasible solution $\boldsymbol{x} \in \mathscr{X}$ and a collection of vectors $\mathbf{d}_1, \dots, \mathbf{d}_R$ with $d_r(\mathbf{0}) = 0$,

$$\mathbf{c}'\boldsymbol{x} = \sum_{r=1}^R \mathbf{c}_r'\boldsymbol{x}_r = \sum_{r=1}^R d_r(\boldsymbol{x}_r) + \sum_{r=1}^R (\mathbf{c}_r'\boldsymbol{x}_r - d_r(\boldsymbol{x}_r)).$$

Upper bounding $\sum_r d_r(\boldsymbol{x}_r)$ by $\max_{\boldsymbol{x}\in\mathscr{X}} \sum_r d_r(\boldsymbol{x}_r)$ and $\mathbf{c}_r'\boldsymbol{x}_r - d_r(\boldsymbol{x}_r)$ by $\max_{\boldsymbol{x}_r\in\mathscr{X}_r}(\mathbf{c}_r'\boldsymbol{x}_r - d_r(\boldsymbol{x}_r))$, we obtain:

$$\mathbf{c}'\boldsymbol{x} \leqslant \max_{\boldsymbol{x}\in\mathscr{X}} \sum_{r=1}^R d_r(\boldsymbol{x}_r) + \sum_{r=1}^R \max_{\boldsymbol{x}_r\in\mathscr{X}_r}(\mathbf{c}_r'\boldsymbol{x}_r - d_r(\boldsymbol{x}_r)).$$

Because the right-hand side is independent of any particular feasible solution, the following inequality holds:

$$Z(\mathbf{c}) \leqslant \max_{\boldsymbol{x}\in\mathscr{X}} \sum_{r=1}^R d_r(\boldsymbol{x}_r) + \sum_{r=1}^R \max_{\boldsymbol{x}_r\in\mathscr{X}_r}(\mathbf{c}_r'\boldsymbol{x}_r - d_r(\boldsymbol{x}_r)).$$

Taking expectations with respect to probability measures $P \in \mathscr{P}(P_1, \dots, P_R)$ and minimum with respect to all the $\mathbf{d}_r$ variables, we get

$$\mathbb{E}_P[Z(\tilde{\mathbf{c}})] \leqslant \min_{\mathbf{d}_1,\dots,\mathbf{d}_R} \left(\max_{\boldsymbol{x}\in\mathscr{X}} \sum_{r=1}^R d_r(\boldsymbol{x}_r) + \sum_{r=1}^R \mathbb{E}_{P_r}\left[\max_{\boldsymbol{x}_r\in\mathscr{X}_r}(\tilde{\mathbf{c}}_r'\boldsymbol{x}_r - d_r(\boldsymbol{x}_r))\right]\right) \quad \text{for all } P \in \mathscr{P}(P_1,\dots,P_R).$$

Hence, $Z^* \leqslant \hat{Z}_u^*$.

Step 2. **Prove that $Z^* \geqslant \hat{Z}_u^*$.** We provide an explicit construction of a distribution $\bar{P} \in \mathscr{P}(P_1, \dots, P_R)$ such that $\mathbb{E}_{\bar{P}}[Z(\tilde{\mathbf{c}})] \geqslant \hat{Z}_u^*$. The upper bound $\hat{Z}_u^*$ can be computed as the optimal objective value to a linear program with decision variables $(d_r(\boldsymbol{x}_r), t, y_{rk})_{r,k,\boldsymbol{x}_r}$:

$$\begin{aligned} \hat{Z}_u^* = \min \quad & t + \sum_{r=1}^R \sum_{k=1}^{K_r} y_{rk} \\ \text{s.t.} \quad & t \geqslant \sum_{r=1}^R d_r(\boldsymbol{x}_r), \quad \forall \boldsymbol{x} \in \mathscr{X}, \\ & y_{rk} \geqslant p_{rk}(\mathbf{c}_{rk}'\boldsymbol{x}_r - d_r(\boldsymbol{x}_r)), \quad \forall \boldsymbol{x}_r \in \mathscr{X}_r, \\ & \qquad k = 1, \dots, K_r,\ r = 1, \dots, R. \end{aligned} \tag{7}$$

Using strong duality for linear programming, $\hat{Z}_u^*$ is also the optimal objective value to the dual linear program with decision variables $(\lambda(\boldsymbol{x}), \gamma_{rk}(\boldsymbol{x}_r))_{r,k,\boldsymbol{x},\boldsymbol{x}_r}$:

$$\begin{aligned} \hat{Z}_u^* = \max \quad & \sum_{r=1}^R \sum_{k=1}^{K_r} \sum_{\boldsymbol{x}_r\in\mathscr{X}_r} p_{rk}\mathbf{c}_{rk}'\boldsymbol{x}_r\gamma_{rk}(\boldsymbol{x}_r) \\ \text{s.t.} \quad & \sum_{\boldsymbol{x}\in\mathscr{X}} \lambda(\boldsymbol{x}) = 1, \\ & \sum_{\boldsymbol{x}_r\in\mathscr{X}_r} \gamma_{rk}(\boldsymbol{x}_r) = 1, \quad \forall k = 1, \dots, K_r, \\ & \qquad r = 1, \dots, R, \\ & \sum_{\boldsymbol{v}\in\mathscr{X}:\, \boldsymbol{v}_r = \boldsymbol{x}_r} \lambda(\boldsymbol{v}) - \sum_{k=1}^{K_r} p_{rk}\gamma_{rk}(\boldsymbol{x}_r) = 0, \\ & \qquad \forall \boldsymbol{x}_r \in \mathscr{X}_r,\ r = 1, \dots, R, \\ & \lambda(\boldsymbol{x}) \geqslant 0, \quad \forall \boldsymbol{x} \in \mathscr{X}, \\ & \gamma_{rk}(\boldsymbol{x}_r) \geqslant 0, \quad \forall \boldsymbol{x}_r \in \mathscr{X}_r,\ k = 1, \dots, K_r, \\ & \qquad r = 1, \dots, R. \end{aligned} \tag{8}$$

Consider a set of optimal solutions $(d_r^*(\boldsymbol{x}_r), t^*, y_{rk}^*)_{r,k,\boldsymbol{x}_r}$ and $(\lambda^*(\boldsymbol{x}), \gamma_{rk}^*(\boldsymbol{x}_r))_{r,k,\boldsymbol{x},\boldsymbol{x}_r}$ to the primal and dual linear programs, respectively. Construct a mixture distribution $\bar{P}$ as follows:

(a) Pick a random feasible solution $\boldsymbol{x} \in \mathscr{X}$ with probability $\lambda^*(\boldsymbol{x})$.

(b) For each r, the random subvector $\tilde{\mathbf{c}}_r$ is given by the scenarios $\mathbf{c}_{rk}$ with probabilities $q_{rk}^*(\boldsymbol{x}_r)$ defined as

$$q_{rk}^*(\boldsymbol{x}_r) = \bar{P}_{r,\boldsymbol{x}_r}(\tilde{\mathbf{c}}_r = \mathbf{c}_{rk}) = \frac{p_{rk}\gamma_{rk}^*(\boldsymbol{x}_r)}{\sum_{l=1}^{K_r} p_{rl}\gamma_{rl}^*(\boldsymbol{x}_r)} \quad \text{for } k = 1, \dots, K_r.$$

Clearly, $\sum_{k=1}^{K_r} q_{rk}^*(\boldsymbol{x}_r) = 1$ with $q_{rk}^*(\boldsymbol{x}_r) \geqslant 0$. For $\bar{P}$, the marginal probabilities can be evaluated as

$$\begin{aligned} \bar{P}_r(\tilde{\mathbf{c}}_r = \mathbf{c}_{rk}) &= \sum_{\boldsymbol{x}\in\mathscr{X}} \lambda^*(\boldsymbol{x}) q_{rk}^*(\boldsymbol{x}_r) \\ &= \sum_{\boldsymbol{x}_r\in\mathscr{X}_r} \sum_{\boldsymbol{v}\in\mathscr{X}:\, \boldsymbol{v}_r = \boldsymbol{x}_r} \lambda^*(\boldsymbol{v}) \left(\frac{p_{rk}\gamma_{rk}^*(\boldsymbol{x}_r)}{\sum_{l=1}^{K_r} p_{rl}\gamma_{rl}^*(\boldsymbol{x}_r)} \right) \\ &= \sum_{\boldsymbol{x}_r\in\mathscr{X}_r} p_{rk}\gamma_{rk}^*(\boldsymbol{x}_r) \\ &= p_{rk}. \end{aligned}$$

Hence, $\bar{P} \in \mathcal{P}(P_1, \ldots, P_R)$. The expected optimal value under the distribution $\bar{P}$ satisfies

$$\begin{aligned}\mathbb{E}_{\bar{P}}[Z(\tilde{\mathbf{c}})] &\geqslant \sum_{\boldsymbol{x}\in\mathcal{X}} \lambda^*(\boldsymbol{x}) \sum_{r=1}^{R} \mathbb{E}_{\bar{P}_{r,\boldsymbol{x}_r}}[\tilde{\mathbf{c}}_r' \boldsymbol{x}_r] \\ &= \sum_{\boldsymbol{x}\in\mathcal{X}} \lambda^*(\boldsymbol{x}) \sum_{r=1}^{R}\sum_{k=1}^{K_r} q_{rk}^*(\boldsymbol{x}_r)\mathbf{c}_{rk}'\boldsymbol{x}_r \\ &= \sum_{r=1}^{R}\sum_{\boldsymbol{x}_r\in\mathcal{X}_r}\sum_{\boldsymbol{v}\in\mathcal{X}:\boldsymbol{v}_r=\boldsymbol{x}_r} \lambda^*(\boldsymbol{v})\left(\frac{\sum_{k=1}^{K_r} p_{rk}\gamma_{rk}^*(\boldsymbol{x}_r)\mathbf{c}_{rk}'\boldsymbol{x}_r}{\sum_{l=1}^{K_r} p_{rl}\gamma_{rl}^*(\boldsymbol{x}_r)}\right) \\ &= \sum_{r=1}^{R}\sum_{k=1}^{K_r}\sum_{\boldsymbol{x}_r\in\mathcal{X}_r} p_{rk}\gamma_{rk}^*(\boldsymbol{x}_r)\mathbf{c}_{rk}'\boldsymbol{x}_r \\ &= \hat{Z}_u^*.\end{aligned}$$

The first inequality is obtained by evaluating the objective function value at the feasible solution $\boldsymbol{x} \in \mathcal{X}$ chosen at step (a) of the distribution instead of the corresponding optimal solution. The remaining equalities follow from dual feasibility and strong duality. Hence, $Z^* \geqslant \mathbb{E}_{\bar{P}}[Z(\tilde{\mathbf{c}})] \geqslant \hat{Z}_u^*$. From steps 1 and 2, $Z^* = \hat{Z}_u^*$. $\square$

For the special case of the maximum of random variables and the sum of random variables, we use the result in Theorem 1 to extend the univariate marginal bounds in (3) and (4) to multivariate marginal bounds.

Proposition 1. (i) *For* $Z(\mathbf{c}) = \max_i c_i$, *the following inequality holds*:

$$\mathbb{E}_P\left[\max_{i\in\mathcal{N}} \tilde{c}_i\right] \leqslant \min_d \left(d + \sum_{r=1}^{R} \mathbb{E}_{P_r}\left[\max_{i\in\mathcal{N}_r} \tilde{c}_i - d\right]^+\right) \quad \textit{for all } P \in \mathcal{P}(P_1, \ldots, P_R),$$

and the bound is tight.

(ii) *For* $Z(\mathbf{c}) = \sum_i c_i$ *and* $T \in \mathbb{R}$, *the following inequality holds*:

$$\mathbb{E}_P\left[\sum_{i\in\mathcal{N}} \tilde{c}_i - T\right]^+ \leqslant \min_{\sum_r d_r = T}\left(\sum_{r=1}^{R} \mathbb{E}_{P_r}\left[\sum_{i\in\mathcal{N}_r} \tilde{c}_i - d_r\right]^+\right) \quad \textit{for all } P \in \mathcal{P}(P_1, \ldots, P_R),$$

and the bound is tight.

Proof. (i) Let $\mathcal{X} = \{\mathbf{e}_1^{(N)}, \mathbf{e}_2^{(N)}, \ldots, \mathbf{e}_N^{(N)}\}$ where $\mathbf{e}_i^{(N)}$ is the unit vector in $\mathbb{R}^N$ with 1 in the ith position and 0 otherwise. Then, $\max\{\mathbf{c}'\mathbf{x} \mid \mathbf{x} \in \mathcal{X}\} = \max_{i\in\mathcal{N}} c_i$. If $R > 1$, the projection of the feasible region is $\mathcal{X}_r = \{\mathbf{e}_1^{(N_r)}, \mathbf{e}_2^{(N_r)}, \ldots, \mathbf{e}_{N_r}^{(N_r)}, \mathbf{0}\}$. Using Theorem 1 and noting that $\mathbf{0} \in \mathcal{X}_r$ with $d_r(\mathbf{0}) = 0$, the tight upper bound is given as

$$\min_{d_1,\ldots,d_N}\left(\max_{i\in\mathcal{N}} d_i + \sum_{r=1}^{R}\mathbb{E}\left[\max_{i\in\mathcal{N}_r}(\tilde{c}_i - d_i)\right]^+\right).$$

It is easy to check that there exists an optimal solution such that all the d_i values are equal. Let $d_1 = \max_{i\in\mathcal{N}} d_i$. For any $i \neq 1$, by increasing d_i up to d_1, the first term $\max_{i\in\mathcal{N}} d_i$ remains unaffected whereas the second term does not increase, but possibly decreases. Hence, there exists an optimal solution with all the d_i values equal. This leads to the single-variable optimization problem:

$$\min_d\left(d + \sum_{r=1}^{R}\mathbb{E}\left[\max_{i\in\mathcal{N}_r}\tilde{c}_i - d\right]^+\right).$$

(ii) Let $\mathcal{X} = \{\mathbf{e}^{(N+1)}, \mathbf{0}\}$, where $\mathbf{e}^{(N+1)}$ is the vector in $\mathbb{R}^{N+1}$ with all ones. Then we have

$$\max\left\{\sum_{i\in\mathcal{N}} c_i x_i - T x_{N+1} \,\middle|\, \mathbf{x} \in \mathcal{X}\right\} = \left[\sum_{i\in\mathcal{N}} c_i - T\right]^+.$$

Consider the modified problem in $N+1$ dimensions with $R+1$ partitions, of which the last partition is $\mathcal{N}_{R+1} = \{N+1\}$ corresponding to the variable x_{N+1}. The projection of the feasible region is $\mathcal{X}_r = \{\mathbf{e}^{(N_r)}, \mathbf{0}\}$ for all $r = 1, \ldots, R+1$. Using Theorem 1, the tight upper bound is given as

$$\min_{d_1,\ldots,d_R,d_{R+1}}\left(\left[\sum_{r=1}^{R} d_r + d_{R+1}\right]^+ + \sum_{r=1}^{R}\mathbb{E}\left[\sum_{i\in\mathcal{N}_r}\tilde{c}_i - d_r\right]^+ + [-T - d_{R+1}]^+\right).$$

We have $x^+ + y^+ \geqslant [x+y]^+$ for all $x, y \in \mathbb{R}$ and the equality can happen when $x = 0$ or $y = 0$. Thus, we can claim that by setting $d_{R+1} = -T$, the tight upper bound is still obtained by solving:

$$\min_{d_1,\ldots,d_R}\left(\left[\sum_{r=1}^{R} d_r - T\right]^+ + \sum_{r=1}^{R}\mathbb{E}\left[\sum_{i\in\mathcal{N}_r}\tilde{c}_i - d_r\right]^+\right).$$

The term $[\sum_r d_r - T]^+$ is nondecreasing in d_r. If the term $\sum_r d_r - T = \epsilon > 0$, we can decrease at least one of the d_rs by ϵ such that the first term decreases by ϵ whereas one of the expectation terms would increase by at most ϵ. Using a similar argument for a negative ϵ, we can verify that there exists an optimal solution that satisfies $\sum_r d_r = T$. Thus, the tight upper bound is hence the optimal value of the optimization problem:

$$\begin{aligned}\min_{d_1,\ldots,d_R}\quad & \sum_{r=1}^{R}\mathbb{E}\left[\sum_{i\in\mathcal{N}_r}\tilde{c}_i - d_r\right]^+ \\ \text{s.t.}\quad & \sum_{r=1}^{R} d_r = T. \quad \square\end{aligned}$$

From Theorem 1, given a set of vectors $\mathbf{d}_1, \ldots, \mathbf{d}_R$, evaluating the upper bound reduces to

1. Computing the optimal value to the deterministic maximization problem $\max_{\mathbf{x}\in\mathcal{X}} \sum_r d_r(\boldsymbol{x}_r)$ and
2. Computing expectations of the random terms $\max_{\mathbf{x}_r\in\mathcal{X}_r}(\tilde{\mathbf{c}}_r'\boldsymbol{x}_r - d_r(\boldsymbol{x}_r))$ for $r = 1, \ldots, R$.

The feasible region $\mathscr{X}$ for combinatorial optimization problems can be represented using binary variables and linear inequalities as

$$\mathscr{X} = \left\{ \boldsymbol{x} \in \{0,1\}^N \;\middle|\; \sum_{r=1}^{R} \mathbf{A}_r \boldsymbol{x}_r \leqslant \mathbf{b} \right\}.$$

The next proposition uses this representation of the feasible region to reformulate the deterministic term in the upper bound as a multiple-choice integer program. A multiple-choice integer program is a linear binary optimization problem in which the variables are partitioned and precisely one variable from each subset in the partition is selected (see Bean 1984).

Proposition 2. *The tight upper bound Z^* in Theorem 1 is computable as*

$$Z^* = \min_{\mathbf{d}_1,\ldots,\mathbf{d}_R} \left(\hat{Z}(\mathbf{d}_1,\ldots,\mathbf{d}_R) + \sum_{r=1}^{R} \mathbb{E}_{P_r}\left[\max_{\boldsymbol{x}_r \in \mathscr{X}_r} (\tilde{\mathbf{c}}_r' \boldsymbol{x}_r - d_r(\boldsymbol{x}_r)) \right] \right),$$

where $\hat{Z}(\mathbf{d}_1,\ldots,\mathbf{d}_R)$ is the optimal objective value to a multiple-choice integer program over the decision vectors $\mathbf{z}_r = (z_r(\mathbf{x}_r))_{\boldsymbol{x}_r \in \mathscr{X}_r}$ for $r = 1,\ldots,R$:

$$\begin{aligned} &\hat{Z}(\mathbf{d}_1,\ldots,\mathbf{d}_R) \\ &= \max_{\mathbf{z}_1,\ldots,\mathbf{z}_R} \quad \sum_{r=1}^{R} \sum_{\boldsymbol{x}_r \in \mathscr{X}_r} d_r(\boldsymbol{x}_r) z_r(\boldsymbol{x}_r) \\ &\qquad \text{s.t.} \quad \sum_{r=1}^{R} \sum_{\boldsymbol{x}_r \in \mathscr{X}_r} \mathbf{A}_r \boldsymbol{x}_r z_r(\boldsymbol{x}_r) \leqslant \mathbf{b}, \\ &\qquad\qquad \sum_{\boldsymbol{x}_r \in \mathscr{X}_r} z_r(\boldsymbol{x}_r) = 1, \quad \forall\, r = 1,\ldots,R, \\ &\qquad\qquad z_r(\boldsymbol{x}_r) \in \{0,1\}, \quad \forall\, \boldsymbol{x}_r \in \mathscr{X}_r,\ r = 1,\ldots,R. \end{aligned} \tag{9}$$

Proof. Let $\mathscr{Z}$ denote the feasible region to the multiple-choice integer program (9). Thus,

$$\hat{Z}(\mathbf{d}_1,\ldots,\mathbf{d}_R) = \max_{(\mathbf{z}_1,\ldots,\mathbf{z}_R) \in \mathscr{Z}} \sum_{r=1}^{R} \sum_{\boldsymbol{x}_r \in \mathscr{X}_r} d_r(\boldsymbol{x}_r) z_r(\boldsymbol{x}_r).$$

Step 1. **Prove that** $\hat{Z}(\mathbf{d}_1,\ldots,\mathbf{d}_R) \leqslant \max_{\boldsymbol{x} \in \mathscr{X}} \sum_r d_r(\boldsymbol{x}_r)$**.** Given vectors $\mathbf{d}_1,\ldots,\mathbf{d}_R$, consider an optimal solution $(\mathbf{z}_1^*,\ldots,\mathbf{z}_R^*)$ to formulation (9). Set $\mathbf{x}_r^* = \sum_{\boldsymbol{x}_r \in \mathscr{X}_r} \mathbf{x}_r z_r^*(\mathbf{x}_r)$. Then, we check for the feasibility of $\boldsymbol{x}^*$:

$$\sum_{r=1}^{R} \mathbf{A}_r \boldsymbol{x}_r^* = \sum_{r=1}^{R} \sum_{\boldsymbol{x}_r \in \mathscr{X}_r} \mathbf{A}_r \boldsymbol{x}_r z_r^*(\boldsymbol{x}_r) \leqslant \mathbf{b}.$$

Furthermore, $\boldsymbol{x}_r^* \in \mathscr{X}_r \subseteq \{0,1\}^{N_r}$ implies that $\boldsymbol{x}^* \in \mathscr{X} \subseteq \{0,1\}^N$. The objective value satisfies

$$\begin{aligned} \hat{Z}(\mathbf{d}_1,\ldots,\mathbf{d}_R) &= \sum_{r=1}^{R} \sum_{\boldsymbol{x}_r \in \mathscr{X}_r} d_r(\boldsymbol{x}_r) z_r^*(\boldsymbol{x}_r) \\ &= \sum_{r=1}^{R} d_r(\boldsymbol{x}_r^*) \\ &\leqslant \max_{\boldsymbol{x} \in \mathscr{X}} \sum_{r=1}^{R} d_r(\boldsymbol{x}_r). \end{aligned}$$

Step 2. **Prove that** $\hat{Z}(\mathbf{d}_1,\ldots,\mathbf{d}_R) \geqslant \max_{\boldsymbol{x} \in \mathscr{X}} \sum_r d_r(\boldsymbol{x}_r)$**.** Given vectors $\mathbf{d}_1,\ldots,\mathbf{d}_R$, consider an optimal solution $\boldsymbol{x}^*$ to $\max_{\boldsymbol{x} \in \mathscr{X}} \sum_r d_r(\boldsymbol{x}_r)$. For each r, set $z_r^*(\boldsymbol{x}_r^*) = 1$. Set $z_r^*(\boldsymbol{x}_r) = 0$ for all $\mathbf{x}_r \in \mathscr{X}_r, \mathbf{x}_r \neq \boldsymbol{x}_r^*$. Clearly, $\sum_{\boldsymbol{x}_r \in \mathscr{X}_r} z_r^*(\boldsymbol{x}_r) = 1$. Thus, $(\mathbf{z}_1^*,\ldots,\mathbf{z}_R^*) \in \mathscr{Z}$ because

$$\begin{aligned} \sum_{r=1}^{R} \sum_{\boldsymbol{x}_r \in \mathscr{X}_r} \mathbf{A}_r \boldsymbol{x}_r z_r^*(\boldsymbol{x}_r) &= \sum_{r=1}^{R} \mathbf{A}_r \boldsymbol{x}_r^* \\ &\leqslant \mathbf{b}. \end{aligned}$$

The objective value satisfies

$$\begin{aligned} \max_{\boldsymbol{x} \in \mathscr{X}} \sum_{r=1}^{R} d_r(\boldsymbol{x}_r) &= \sum_{r=1}^{R} d_r(\boldsymbol{x}_r^*) \\ &= \sum_{r=1}^{R} \sum_{\boldsymbol{x}_r \in \mathscr{X}_r} d_r(\boldsymbol{x}_r) z_r^*(\boldsymbol{x}_r) \\ &\leqslant \hat{Z}(\mathbf{d}_1,\ldots,\mathbf{d}_R). \end{aligned}$$

From steps 1 and 2, $\hat{Z}(\mathbf{d}_1,\ldots,\mathbf{d}_R) = \max_{\boldsymbol{x} \in \mathscr{X}} \sum_r d_r(\boldsymbol{x}_r)$, proving the desired result. □

2.1. Application to Subset Selection

Consider the problem of selecting up to M items out of a total of N items of maximum total weight:

$$\begin{aligned} Z(\mathbf{c}) = \max \quad & \mathbf{c}' \boldsymbol{x} \\ \text{s.t.} \quad & \mathbf{e}^{(N)'} \boldsymbol{x} \leqslant M, \\ & \boldsymbol{x} \in \{0,1\}^N. \end{aligned} \tag{10}$$

In a risk management context, $\mathbf{c}$ denotes a nonnegative loss vector and $Z(\mathbf{c})$ defines the sum of the M highest losses in this set. Our next theorem provides an instance of a polynomial computable bound on the expected value of $Z(\tilde{\mathbf{c}})$ in (10) using multivariate marginal distribution information of $\tilde{\mathbf{c}}$.

Theorem 2. *Given a scenario representation for the random weights, the tight multivariate marginal upper bound Z^* on the expected optimal value in* (10) *is computable in time polynomial in the maximum number of scenarios in any subset and N when the size of each subset in the partition is $O(\log N)$.*

Proof. Consider a partition of the index set $\mathscr{N} = \{1,\ldots,N\}$. Assume that the size of each subset $\mathscr{N}_r$ in the partition is $N_r = O(\log N)$. The projection of the feasible region of (10) on the space of decision variables in the rth subset is

$$\mathscr{X}_r = \left\{ \boldsymbol{x}_r \in \{0,1\}^{N_r} \;\middle|\; \mathbf{e}^{(N_r)'} \boldsymbol{x}_r \leqslant M \right\}.$$

Using Proposition 2, Z^* can be computed as

$$\begin{aligned} Z^* = \min \quad & t + \sum_{r=1}^{R}\sum_{k=1}^{K_r} y_{rk} \\ \text{s.t.} \quad & t \geqslant \hat{Z}(\mathbf{d}_1, \ldots, \mathbf{d}_R), \\ & y_{rk} \geqslant p_{rk}\mathbf{c}'_{rk}\boldsymbol{x}_r - p_{rk}d_r(\boldsymbol{x}_r) \quad \forall \boldsymbol{x}_r \in \mathscr{X}_r, \\ & \qquad k = 1, \ldots, K_r,\ r = 1, \ldots, R. \end{aligned} \tag{11}$$

Let $|\mathscr{X}_r|$ denote the size of the set $\mathscr{X}_r$ and $K_{\max} = \max_r K_r$ denote the maximum number of scenarios in any subset. Let $\text{poly}(N) = N^{O(1)}$ denote a polynomial in N. The total number of decision variables in formulation (11) is polynomial in the size of the input because

number of decision variables in (11)

$$\begin{aligned} &= \underbrace{1}_{t} + \underbrace{\sum_{r=1}^{R}|\mathscr{X}_r|}_{d_r(\boldsymbol{x}_r)} + \underbrace{\sum_{r=1}^{R}K_r}_{y_{rk}} \\ &= 1 + \sum_{r=1}^{R} 2^{O(\log N)} + \sum_{r=1}^{R} K_r \\ &= 1 + \text{poly}(N) + O(RK_{\max}). \end{aligned}$$

To analyze the complexity of the optimization problem (11), consider the separation version of the problem (see Grötschel et al. 1988). Since testing for the feasibility of the second set of inequalities in (11) is easy, we restrict our attention to the following:

Given t and $\mathbf{d}$, check if

$$t \geqslant \hat{Z}(\mathbf{d}_1, \ldots, \mathbf{d}_R).$$

If not, find a violated inequality.

From Proposition 2, $\hat{Z}(\mathbf{d}_1, \ldots, \mathbf{d}_R)$ is the optimal objective value to the multiple choice knapsack problem:

$$\begin{aligned} &\hat{Z}(\mathbf{d}_1, \ldots, \mathbf{d}_R) \\ &\quad = \max \sum_{r=1}^{R}\sum_{\boldsymbol{x}_r\in\mathscr{X}_r} d_r(\boldsymbol{x}_r) z_r(\boldsymbol{x}_r) \\ &\qquad \text{s.t.} \sum_{r=1}^{R}\sum_{\boldsymbol{x}_r\in\mathscr{X}_r} \mathbf{e}^{(N_r)'}\boldsymbol{x}_r z_r(\boldsymbol{x}_r) \leqslant M, \\ &\qquad\quad \sum_{\boldsymbol{x}_r\in\mathscr{X}_r} z_r(\boldsymbol{x}_r) = 1, \quad \forall r = 1, \ldots, R, \\ &\qquad\quad z_r(\boldsymbol{x}_r) \in \{0, 1\}, \quad \forall \boldsymbol{x}_r \in \mathscr{X}_r, \\ &\qquad\qquad\qquad r = 1, \ldots, R. \end{aligned} \tag{12}$$

Dudzinski and Walukiewicz (1987) provide a pseudo-polynomial time dynamic programming algorithm for solving multiple choice knapsack problems. We outline the algorithm in our context. Let $Z_S(B)$ be the optimal solution to the multiple choice knapsack problem (12) when restricted to the first S subsets with a knapsack capacity restricted to B:

$$\begin{aligned} Z_S(B) = \max \quad & \sum_{r=1}^{S}\sum_{\boldsymbol{x}_r\in\mathscr{X}_r} d_r(\boldsymbol{x}_r) z_r(\boldsymbol{x}_r) \\ \text{s.t.} \quad & \sum_{r=1}^{S}\sum_{\boldsymbol{x}_r\in\mathscr{X}_r} \mathbf{e}^{(N_r)'}\boldsymbol{x}_r z_r(\boldsymbol{x}_r) \leqslant B, \\ & \sum_{\boldsymbol{x}_r\in\mathscr{X}_r} z_r(\boldsymbol{x}_r) = 1, \quad \forall r = 1, \ldots, S, \\ & z_r(\boldsymbol{x}_r) \in \{0, 1\}, \quad \forall \boldsymbol{x}_r \in \mathscr{X}_r,\ r = 1, \ldots, S. \end{aligned} \tag{13}$$

Define $Z_0(B) = 0$ for $B = 0, 1, \ldots, M$. To compute $Z_S(B)$ for $S = 1, \ldots, R$, the dynamic programming recursion is set up as

$$\begin{aligned} Z_S(B) = \max \quad & Z_{S-1}(B - \mathbf{e}^{(N_S)'}\boldsymbol{x}_S) + d_S(\boldsymbol{x}_S) \\ \text{s.t.} \quad & \boldsymbol{x}_S \in \mathscr{X}_S, \\ & \mathbf{e}^{(N_S)'}\boldsymbol{x}_S \leqslant B, \end{aligned} \tag{14}$$

where $Z_S(B)$ is $-\infty$ if the set is infeasible. The optimal objective value $\hat{Z}(\mathbf{d}_1, \ldots, \mathbf{d}_R)$ is equal to $Z_R(M)$. The overall time complexity of implementing this method is $O(M\sum_r |\mathscr{X}_r|)$. Since $|\mathscr{X}_r|$ is polynomially bounded in the N, $\hat{Z}(\mathbf{d}_1, \ldots, \mathbf{d}_R)$ can be computed in time polynomial in the size of the input. From the equivalence on separation and optimization (Grötschel et al. 1988), the result follows. ☐

3. A Reduced Formulation

The total number of decision variables in Theorem 1 could be very large compared to the number of decision variables in the original deterministic optimization problem. In this section, we provide a weaker upper bound wherein the number of decision variables in the formulation does not grow exponentially in size of the problem. For univariate marginals, this bound reduces to the Meilijson and Nadas (1979) result. An advantage of the bound is that the computation relates directly to the complexity of solving the original combinatorial optimization problem. This property has been exploited by Meilijson and Nadas (1979), Weiss (1986), Klein Haneveld (1986), and Bertsimas et al. (2004, 2006) in proposing algorithms for univariate marginals.

Theorem 3. (i) *Let $\mathbf{d} = (d_1, \ldots, d_N) \in \mathbb{R}^N$ be a decision vector. An upper bound on Z^* is computable as*

$$Z^* \leqslant Z_u^* = \min_{\mathbf{d}} \left(Z(\mathbf{d}) + \sum_{r=1}^{R} \mathbb{E}_{P_r}\left[\max_{\boldsymbol{x}_r\in\mathscr{X}_r} (\tilde{\mathbf{c}}'_r\boldsymbol{x}_r - \mathbf{d}'_r\boldsymbol{x}_r) \right] \right). \tag{15}$$

(ii) *Assume that for each $r = 1, \ldots, R$, the set of subvectors $\boldsymbol{x}_r \in \mathscr{X}_r/\{\mathbf{0}\}$ are linearly independent. Then, $Z^* = Z_u^*$.*

Proof. (i) Consider an optimal solution $\mathbf{d}^*$ to formulation (15). For the upper bound defined in Theorem 1, set $d_r^*(\mathbf{x}_r) = \mathbf{d}_r^{*'}\mathbf{x}_r$. Then the tight upper bound satisfies

$$\begin{aligned} Z^* &\leqslant \max_{\mathbf{x}\in\mathscr{X}} \sum_{r=1}^R d_r^*(\boldsymbol{x}_r) + \sum_{r=1}^R \mathbb{E}_{P_r}\Big[\max_{\boldsymbol{x}_r\in\mathscr{X}_r}(\tilde{\mathbf{c}}_r'\boldsymbol{x}_r - d_r^*(\boldsymbol{x}_r))\Big] \\ &= \max_{\boldsymbol{x}\in\mathscr{X}} \mathbf{d}^{*'}\boldsymbol{x} + \sum_{r=1}^R \mathbb{E}_{P_r}\Big[\max_{\boldsymbol{x}_r\in\mathscr{X}_r}(\tilde{\mathbf{c}}_r'\boldsymbol{x}_r - \mathbf{d}_r^{*'}\mathbf{x}_r)\Big] \\ &= Z(\mathbf{d}^*) + \sum_{r=1}^R \mathbb{E}_{P_r}\Big[\max_{\boldsymbol{x}_r\in\mathscr{X}_r}(\tilde{\mathbf{c}}_r'\boldsymbol{x}_r - \mathbf{d}_r^{*'}\mathbf{x}_r)\Big] \\ &= Z_u^*. \end{aligned}$$

The first inequality is valid because the set of $(d_r^*(\boldsymbol{x}_r))_{\boldsymbol{x}_r\in\mathscr{X}_r}$ forms a feasible solution in Theorem 1. Thus, $Z^* \leqslant Z_u^*$.

(ii) The bound Z_u^* is the optimal objective value to the linear program with decision variables $(\mathbf{d}_r, t, y_{rk})_{r,k}$:

$$\begin{aligned} Z_u^* = \min \quad & t + \sum_{r=1}^R\sum_{k=1}^{K_r} y_{rk} \\ \text{s.t.} \quad & t \geqslant \sum_{r=1}^R \mathbf{d}_r'\boldsymbol{x}_r, \quad \forall\, \boldsymbol{x}\in\mathscr{X}, \qquad (16) \\ & y_{rk} \geqslant p_{rk}(\mathbf{c}_{rk}'\boldsymbol{x}_r - \mathbf{d}_r'\boldsymbol{x}_r), \quad \forall\, \boldsymbol{x}_r\in\mathscr{X}_r, \\ & \qquad k=1,\dots,K_r,\ r=1,\dots,R. \end{aligned}$$

Using strong duality for linear programming, Z_u^* is the optimal objective value to the dual linear program with decision variables $(\lambda(\boldsymbol{x}), \gamma_{rk}(\boldsymbol{x}_r))_{r,k,\boldsymbol{x},\boldsymbol{x}_r}$:

$$\begin{aligned} Z_u^* = \max \quad & \sum_{r=1}^R\sum_{k=1}^{K_r}\sum_{\boldsymbol{x}_r\in\mathscr{X}_r} p_{rk}\mathbf{c}_{rk}'\boldsymbol{x}_r\gamma_{rk}(\boldsymbol{x}_r) \\ \text{s.t.} \quad & \sum_{\boldsymbol{x}\in\mathscr{X}} \lambda(\boldsymbol{x}) = 1, \\ & \sum_{\boldsymbol{x}_r\in\mathscr{X}_r} \gamma_{rk}(\boldsymbol{x}_r) = 1, \quad \forall\, k=1,\dots,K_r,\ r=1,\dots,R, \\ & \sum_{\boldsymbol{x}\in\mathscr{X}} \mathbf{x}_r\lambda(\mathbf{x}) - \sum_{k=1}^{K_r}\sum_{\boldsymbol{x}_r\in\mathscr{X}_r} p_{rk}\boldsymbol{x}_r\gamma_{rk}(\boldsymbol{x}_r) = 0, \quad \forall\, r=1,\dots,R, \\ & \lambda(\boldsymbol{x}) \geqslant 0, \quad \forall\, \boldsymbol{x}\in\mathscr{X}, \\ & \gamma_{rk}(\boldsymbol{x}_r) \geqslant 0, \quad \forall\, \boldsymbol{x}_r\in\mathscr{X}_r,\ k=1,\dots,K_r,\ r=1,\dots,R. \end{aligned}$$

Consider a set of optimal solutions $(\mathbf{d}_r^*, t^*, y_{rk}^*)_{r,k}$ and $(\lambda^*(\boldsymbol{x}), \gamma_{rk}^*(\boldsymbol{x}_r))_{r,k,\boldsymbol{x},\boldsymbol{x}_r}$ to the primal and dual linear programs, respectively. From the dual feasibility conditions, we have

$$\sum_{\boldsymbol{x}_r\in\mathscr{X}_r}\sum_{\boldsymbol{v}\in\mathscr{X}:\,\boldsymbol{v}_r=\boldsymbol{x}_r} \lambda^*(\boldsymbol{v}) = 1, \quad \forall\, r=1,\dots,R,$$

and

$$\sum_{\boldsymbol{x}_r\in\mathscr{X}_r} \boldsymbol{x}_r\Bigg(\sum_{\boldsymbol{v}\in\mathscr{X}:\,\boldsymbol{v}_r=\boldsymbol{x}_r}\lambda^*(\boldsymbol{v})\Bigg) = \sum_{\boldsymbol{x}_r\in\mathscr{X}_r}\boldsymbol{x}_r\Bigg(\sum_{k=1}^{K_r} p_{rk}\gamma_{rk}^*(\boldsymbol{x}_r)\Bigg), \quad \forall\, r=1,\dots,R.$$

Given that the subvectors $\boldsymbol{x}_r \in \mathscr{X}_r/\{\mathbf{0}\}$ are linearly independent, this implies

$$\sum_{\boldsymbol{v}\in\mathscr{X}:\,\boldsymbol{v}_r=\boldsymbol{x}_r}\lambda^*(\boldsymbol{v}) = \sum_{k=1}^{K_r} p_{rk}\gamma_{rk}^*(\boldsymbol{x}_r), \quad \forall\, \mathbf{x}_r\in\mathscr{X}_r,\ r=1,\dots,R.$$

With this equality, it is easy to verify that the distribution $\bar{P}$ constructed in Theorem 1 provides the tight upper bound as before. Because the construction is identical to Theorem 1, we omit it for brevity. Hence, $Z^* = Z_u^*$. □

The next example shows that the bound in Theorem 3 might not be the tightest bound in general. Consider an instance of the subset selection problem with $N = 4$, $M = 2$, and $R = 2$. The subsets are $\mathscr{N}_1 = \{1,2\}$ and $\mathscr{N}_2 = \{3,4\}$. The feasible set $\mathscr{X}$ contains a total of 11 feasible solutions and $\mathscr{X}_1 = \mathscr{X}_2 = \{0,1\}^2$. Define a total of 8 decision variables $d_r(\boldsymbol{x}_r)$ for $\boldsymbol{x}_r \in \{0,1\}^2$, $r = 1,2$, with $d_1(0,0) = d_2(0,0) = 0$. The tight bound Z^* in Theorem 1 can be calculated as

$$\begin{aligned} Z^* = \min \quad & t + \sum_{r=1}^R\sum_{k=1}^{K_r} y_{rk} \\ \text{s.t.} \quad & t \geqslant 0, \\ & t \geqslant \max\big(d_1(0,1), d_1(1,0), d_1(1,1)\big), \\ & t \geqslant \max\big(d_2(0,1), d_2(1,0), d_2(1,1)\big), \\ & t \geqslant \max\big(d_1(0,1)+d_2(0,1), d_1(0,1)+d_2(1,0)\big), \\ & t \geqslant \max\big(d_1(1,0)+d_2(0,1), d_1(1,0)+d_2(1,0)\big), \\ & y_{rk} \geqslant p_{rk}\mathbf{c}_{rk}'\boldsymbol{x}_r - p_{rk}d_r(\boldsymbol{x}_r), \\ & \qquad \forall\, \boldsymbol{x}_r\in\{0,1\}^2,\ k=1,\dots,K_r,\ r=1,2. \end{aligned}$$

This can be reformulated as linear optimization problem and solved by a linear programming solver. To compute the upper bound Z_u^* in Theorem 3, we define (d_1, d_2, d_3, d_4) as decision variables:

$$\begin{aligned} Z_u^* = \min \quad & t + \sum_{r=1}^R\sum_{k=1}^{K_r} y_{rk} \\ \text{s.t.} \quad & t \geqslant 0, \\ & t \geqslant \max\big(d_1, d_2, d_3, d_4\big), \\ & t \geqslant \max\big(d_1+d_2, d_1+d_3, d_1+d_4\big), \\ & t \geqslant \max\big(d_2+d_3, d_2+d_4, d_3+d_4\big), \\ & y_{rk} \geqslant p_{rk}\mathbf{c}_{rk}'\boldsymbol{x}_r - p_{rk}\mathbf{d}_r'\boldsymbol{x}_r, \\ & \qquad \forall\, \boldsymbol{x}_r\in\{0,1\}^2,\ k=1,\dots,K_r,\ r=1,2. \end{aligned}$$

Let P_1 be a discrete uniform distribution with $K_1 = 3$ scenarios: $(6, 8)$, $(5, 1)$, and $(7, 9)$, each occurring with probability $1/3$. Likewise, let P_2 be a discrete uniform distribution with $K_2 = 3$ scenarios: $(1, 9)$, $(9, 3)$, and $(9, 6)$, each occurring with probability $1/3$. By solving the linear programs, we verify that the tight bound $Z^* = 50/3$ is strictly smaller than the bound $Z_u^* = 17$.

An example of combinatorial optimization problem that satisfies the linear independence condition in Theorem 3(ii) is the PERT problem with the partition formed by subsets of incoming arcs into nodes. In the next section, we identify instances of the PERT problem where the multivariate marginal bound is computable in polynomial time.

3.1. Application to PERT Networks

Let $\mathscr{V} = \{1, \ldots, M\}$ denote the set of nodes in a PERT network where nodes 1 and M represent the start and end of the project and $\mathscr{E}$ denotes the set of arcs. A total of N arcs are present in the network. Each arc is associated with a length or the time needed to complete the activity. Let (i, j) denote an arc originating from node i and terminating at node j with arc length c_{ij}. The project duration is determined by the length of the longest path from the start node to end node in this network. It can be computed as the optimal objective value to the combinatorial optimization problem:

$$
\begin{aligned}
Z(\mathbf{c}) = \max \quad & \sum_{(i,j)\in\mathscr{E}} c_{ij} x_{ij} \\
\text{s.t.} \quad & \sum_{i:(j,i)\in\mathscr{E}} x_{ji} - \sum_{i:(i,j)\in\mathscr{E}} x_{ij} \\
& = \begin{cases} 1, & \text{if } j = 1, \\ -1, & \text{if } j = M, \\ 0, & \text{if } j = 2, \ldots, M-1, \end{cases} \\
& x_{ij} \in \{0, 1\}, \quad \forall (i,j) \in \mathscr{E}.
\end{aligned}
\tag{17}
$$

For project networks with deterministic activity durations, $Z(\mathbf{c})$ can be computed in polynomial time by solving the linear programming relaxation of formulation (17). This is due to the unimodular structure of the constraint matrix for network flow problems. Whereas the deterministic optimization problem is easy, the computation of the expected optimal value is still a challenging problem. Hagstrom (1988) formally established the complexity of the stochastic version of PERT networks under the assumption of independent activity durations. Specifically, she addressed the complexity of the following two problems:

(1) **MEAN:** Given a PERT network with discrete, independent activity durations, compute the expected project duration.

(2) **CDF:** Given a PERT network with discrete, independent activity durations and a due date T, compute the probability that the project will be completed in time less than or equal to T.

Her main results are summarized in the next theorem.

Theorem 4 (Hagstrom 1988). (i) **CDF** *and the two-state version of* **MEAN** *are* $\#\mathscr{P}$*-complete.*

(ii) *Furthermore,* **MEAN** *and* **CDF** *cannot be computed in time polynomial in the number of points in the range of the project duration unless* $\mathscr{P} = \mathscr{NP}$.

The classes $\#\mathscr{P}$ and $\#\mathscr{P}$-complete are the counting versions of the $\mathscr{NP}$ and $\mathscr{NP}$-complete recognition problems. This computational complexity class was first introduced by Valiant (1979). Any counting problem is at least as hard as the corresponding recognition problem. However, there exist recognition problems that are solvable in polynomial time for which the counting versions are $\#\mathscr{P}$-complete. Valiant (1979) and Provan and Ball (1983) provide instances of such problems. The PERT problem falls under this category. Because the existence of a polynomial-time algorithm for solving a $\#\mathscr{P}$-complete problem would imply $\mathscr{P} = \mathscr{NP}$, it seems highly unlikely that these problems can be solved in polynomial time. Although the complexity of **MEAN** with activity durations taking more than two values is still open, part (ii) of Theorem 4 indicates that problems with longer encoding might be difficult to solve.

A special class of graphs for which efficient algorithms have been derived are *series-parallel graphs*. These graphs can be constructed by a sequence of series and parallel compositions starting from single-arc graphs. The inverse operations, series and parallel reductions, recursively reduce series-parallel graphs to single arcs. Given independent activity durations, the computation for these graphs can therefore be reduced to a sequence of convolutions (series reductions) and products (parallel reductions) of distribution functions. However, even in this case, there is an inherent complexity to the PERT problem that depends on the way distributions are generated along the reduction sequence. Potentially, the number of points in the range of the project duration could be exponential in the size of the network. The complexity results for these graphs are summarized in the next theorem.

Theorem 5 (Ball et al. 1995, Möhring 2001). (i) *The two-state version of* **MEAN** *and* **CDF** *for a series-parallel graph is* $\mathscr{NP}$*-hard in the weak sense.*

(ii) *For a series-parallel graph with arbitrary discrete distributions,* **MEAN** *and* **CDF** *can be computed in time polynomial in the size of the network and the maximum number of distinct values that the project duration takes along a series-parallel reduction sequence.*

(iii) *For a series-parallel graph with activity durations restricted to the set* $\{0, 1, 2, \ldots, q\}$, **MEAN** *and* **CDF** *can be computed in time polynomial in the size of the network and* q.

Several authors, including Kleindorfer (1971), Spelde (1976), Dodin (1985) and Möhring (2001), have proposed bounds for general graphs by transforming them to series-parallel graphs and using bounds for series-parallel graphs.

In the PERT context, a natural partition is formed by the set of incoming arcs into each node. Given a joint (possibly dependent) discrete distribution for the set of activities coming into each node and assuming independence among activity durations at different nodes, Fulkerson (1962) and Shogan (1983) developed bounds on the expected project completion time. Meilijson and Nadas (1979) developed a polynomial-time computable upper bound on the expected project completion time, assuming marginal distribution information allowing for any possible dependence among activity durations. Our next theorem identifies instances of the PERT problem when the multivariate marginal bound is computable in polynomial time.

THEOREM 6. (i) *Given a scenario representation for the set of activity durations entering each node, the tight multivariate marginal upper bound Z^* on the expected project duration is computable in time polynomial in the maximum number of scenarios in any subset and the size of the network.*

(ii) *Given a discrete distribution for each activity duration with independence among the activity durations that enter a node, the tight multivariate marginal upper bound Z^* on the expected project duration is computable in time polynomial in the maximum number of supporting points of the activity duration distribution and the size of the network.*

PROOF. (i) Consider the subvector of random activity durations $\tilde{\mathbf{c}}_j = (\tilde{c}_{ij})_{i:\,(i,j)\in\mathscr{E}}$ for arcs entering a node j. Let $\mathbf{c}_{jk}$ denote a realization of this subvector occurring with probability $P_j(\tilde{\mathbf{c}}_j = \mathbf{c}_{jk}) = p_{jk}$ for $k = 1,\ldots,K_j$ where $\sum_k p_{jk} = 1$. Without loss of generality, we assume that for each node $j = 2,\ldots,M-1$, there exists at least one directed path from the start to the end node that passes through j and one directed path that does not pass through j. Else, we can decompose the problem into two independent problems of finding the longest path from node start 1 to node j and the longest path from node j to end node M. The projection of the feasible region of (17) onto the space of incoming arcs into node j is then the extreme points of the standard simplex. This follows by observing that any directed path in the acyclic PERT network from the start to the end node consists of at most one arc entering node j. Hence, Theorem 3(ii) is applicable to this problem. The tight multivariate marginal upper bound is

$$Z^* = \min_{\mathbf{d}}\left(Z(\mathbf{d}) + \sum_{j=2}^{M-1}\mathbb{E}\Big[\max_{i:\;(i,j)\in\mathscr{E}}(\tilde{c}_{ij} - d_{ij})\Big]^+ + \mathbb{E}\Big[\max_{i:\;(i,M)\in\mathscr{E}}(\tilde{c}_{iM} - d_{iM})\Big]\right). \tag{18}$$

By using the dual formulation for the linear programming representation of $Z(\mathbf{d})$, Z^* is computed as

$$\begin{aligned} Z^* = \min\quad & w_M - w_1 + \textstyle\sum_{j=2}^{M}\sum_{k=1}^{K_j} y_{jk}\\ \text{s.t.}\quad & w_j - w_i \geqslant d_{ij}, \quad \forall\,(i,j)\in\mathscr{E},\\ & y_{jk} \geqslant p_{jk}c_{ijk} - p_{jk}d_{ij}, \quad \forall\, i\colon (i,j)\in\mathscr{E},\\ &\qquad k = 1,\ldots,K_j,\; j = 2,\ldots,M,\\ & y_{jk} \geqslant 0, \quad \forall\, k = 1,\ldots,K_j,\; j = 2,\ldots,M-1. \end{aligned} \tag{19}$$

The linear program in (19) is polynomial sized in the size of the network and the maximum number of scenarios in any subset, proving the desired result.

(ii) To prove the polynomial complexity of computing the bound, we reformulate (18) as follows:

$$\begin{aligned} Z^* = \min\quad & t\\ \text{s.t.}\quad & t \geqslant Z(\mathbf{d}) + \sum_{j=2}^{M-1}\mathbb{E}\Big[\max_{i:\;(i,j)\in\mathscr{E}}(\tilde{c}_{ij} - d_{ij})\Big]^+\\ &\qquad + \mathbb{E}\Big[\max_{i:\;(i,M)\in\mathscr{E}}(\tilde{c}_{iM} - d_{iM})\Big]. \end{aligned} \tag{20}$$

The separation version of this problem is as follows: Given t and $\mathbf{d}$, check if

$$t \geqslant Z(\mathbf{d}) + \sum_{j=2}^{M-1}\mathbb{E}\Big[\max_{i:(i,j)\in\mathscr{E}}(\tilde{c}_{ij} - d_{ij})\Big]^+ + \mathbb{E}\Big[\max_{i:(i,M)\in\mathscr{E}}(\tilde{c}_{iM} - d_{iM})\Big].$$

If not, find a violated inequality.

For a fixed $\mathbf{d}$, $Z(\mathbf{d})$ can be evaluated in polynomial time in the size of the network. We focus on the computation of the expected value of the maximum of independent random variables; namely, a parallel graph. Assuming that the maximum number of supporting points among the activity distributions is $K_{\max} = \max_i K_i$, the random variable $\tilde{c}_{\max} = \max_{i=1,\ldots,n}\tilde{c}_i$ has at most $nK_{\max}$ supporting points. From Theorem 5(ii), **MEAN** and **CDF** problem for this parallel graph is solvable in time polynomial in n and $K_{\max}$. The maximum number of incoming arcs into a node is $M-1$. Thus, testing for feasibility in Problem (20) is possible in polynomial time in the maximum number of supporting points $K_{\max}$ and the size of the network.

To find a violated inequality, consider the case when the solution is infeasible. We compute subgradients of $Z(\mathbf{d})$ and functions of the form $\mathbb{E}[\max_{i=1,\ldots,n}(\tilde{c}_i - d_i)]$ with respect to $\mathbf{d}$. For $Z(\mathbf{d})$, the optimal solution to Problem (17) provides the subgradient. The piecewise-linear function $f(\mathbf{d};\mathbf{c}) = \max_{i=1,\ldots,n}(\tilde{c}_i - d_i)$ has $g(\mathbf{d};\mathbf{c}) = -\mathbf{e}^{(n)}_{i^{\boldsymbol{d}}_{\min}(\boldsymbol{c})}$, where $i^{\boldsymbol{d}}_{\min}(\mathbf{c}) = \min\{j\colon c_j - d_j = f(\mathbf{d};\mathbf{c})\}$, as one of its subgradients. Therefore, $\mathbb{E}[g(\mathbf{d};\tilde{\mathbf{c}})]$ is a subgradient of $\mathbb{E}[f(\mathbf{d};\tilde{\mathbf{c}})]$. In order to calculate $\mathbb{E}[g(\mathbf{d};\tilde{\mathbf{c}})]$, we need to calculate

$P_i(\mathbf{d}) = P(\tilde{c}_i - d_i = f(\mathbf{d}; \tilde{\mathbf{c}}), i = i^{\mathbf{d}}_{\min}(\tilde{\mathbf{c}}))$ for all $i = 1, \ldots, n$. We have

$$P_i(\mathbf{d}) = P\Big(\tilde{c}_i - d_i > \max_{j=1,\ldots,i-1}(\tilde{c}_j - d_j),\ \tilde{c}_i - d_i \geqslant \max_{j=i+1,\ldots,n}(\tilde{c}_j - d_j)\Big),\ i = 1, \ldots, n.$$

The random variable $\tilde{c}^{i}_{\max}(\mathbf{d}) = \max_{j=1,\ldots,i-1}(\tilde{c}_j - d_j)$ has at most $(i-1)K_{\max}$ supporting points and its distribution again can be specified in polynomial time (instances of **CDF** problem). We have similar results for the random variable $\tilde{c}^{-i}_{\max}(\mathbf{d}) = \max_{j=i+1,\ldots,n}(\tilde{c}_j - d_j)$. The random variables $\tilde{c}_i$, $\tilde{c}^{i}_{\max}(\mathbf{d})$, and $\tilde{c}^{-i}_{\max}(\mathbf{d})$ are independent. The total supporting points of the joint distribution is bounded from above by $(i-1)(n-i-1)K^3_{\max}$. Thus, the probability $P_i(\mathbf{d})$ can be calculated in polynomial time, which implies that the subgradient $\mathbb{E}[g(\mathbf{d}; \tilde{\mathbf{c}})]$ is computable in polynomial time. The violated inequality is then constructed in polynomial time in the maximum number of supporting points $K_{\max}$ and the size of the network. Combining these two results, we obtain the polynomial-time solvability of the separation problem. The equivalence of separation and optimization (Grötschel et al. 1988) proves the desired result. □

As an application of the worst-case bound on the expected project completion time, we consider a two-stage stochastic program for project crashing. Formally, the duration of each activity $(i,j) \in \mathscr{E}$ is described as $\tilde{c}_{ij} - t_{ij}$, where $\tilde{c}_{ij}$ is the bounded discrete random variable and the decision variable t_{ij} is bounded between 0 and u_{ij}. The crashing cost is assumed to be linear and the cost per unit change in t_{ij} is $f_{ij} \geqslant 0$ for all $(i,j) \in \mathscr{E}$. The goal is to determine the crashing duration for each activity that minimizes the sum of the crashing costs and the expected project completion time. The standard two-stage stochastic optimization formulation for this problem is provided by Wollmer (1985):

$$\begin{aligned} \min\ & \alpha \sum_{(i,j)\in\mathscr{E}} f_{ij} t_{ij} + \mathbb{E}[Z(\tilde{\mathbf{c}} - \mathbf{t})] \\ \text{s.t.}\ & 0 \leqslant t_{ij} \leqslant u_{ij}, \quad \forall (i,j) \in \mathscr{E}, \end{aligned} \tag{21}$$

where $\alpha \geqslant 0$ is a parameter that provides a trade-off between crashing costs and expected project completion time. One can conceive more complicated variants of the model to account for budget constraints of shared resources. For simplicity, we focus on the basic model and develop a distributional robust optimization counterpart of this problem. Assume as in part (i) of Theorem 6 that a scenario representation is provided for the set of activity durations entering a node. However, the entire joint distribution for all activity durations in the project is unknown. The stochastic program in (21) is replaced by a distributional robust optimization problem:

$$\begin{aligned} \min & \Big(\alpha \sum_{(i,j)\in\mathscr{E}} f_{ij} t_{ij} + \sup_{P\in\mathscr{P}(P_1,\ldots,P_R)} \mathbb{E}_P[Z(\tilde{\mathbf{c}} - \mathbf{t})]\Big) \\ \text{s.t.}\ & 0 \leqslant t_{ij} \leqslant u_{ij}, \quad \forall (i,j) \in \mathscr{E}, \end{aligned} \tag{22}$$

where $\mathscr{P}$ is the set of all probability measures P with the given marginals. Because the worst-case value is used in computing the second-stage expected project duration, the first-stage crashing decisions are robust to dependence. A straightforward application of the dual formulation in (19) to the inner supremum in (22) provides a linear optimization formulation for the distributional robust problem:

$$\begin{aligned} \min\ & \alpha \sum_{(i,j)\in\mathscr{E}} f_{ij} t_{ij} + w_M - w_1 + \sum_{j=2}^{M}\sum_{k=1}^{K_j} y_{jk} \\ \text{s.t.}\ & 0 \leqslant t_{ij} \leqslant u_{ij}, \quad \forall (i,j) \in \mathscr{E}, \\ & w_j - w_i \geqslant d_{ij}, \quad \forall (i,j) \in \mathscr{E}, \\ & y_{jk} \geqslant p_{jk} c_{ijk} - p_{jk} t_{ij} - p_{jk} d_{ij}, \quad \forall i\colon (i,j) \in \mathscr{E}, \\ & \qquad k = 1, \ldots, K_j,\ j = 2, \ldots, M, \\ & y_{jk} \geqslant 0, \quad \forall k = 1, \ldots, K_j,\ j = 2, \ldots, M-1. \end{aligned} \tag{23}$$

Because this linear program is polynomial sized, the distributional robust project-crashing problem under multivariate marginals can be solved in polynomial time.

4. Conclusion

In this paper, the Meilijson and Nadas (1979) bound for probabilistic combinatorial optimization problems is extended from univariate marginals to nonoverlapping multivariate marginals. The bound is robust against dependence and valid across all joint distributions with the given marginals. Furthermore, this bound is tight, in that there exists a multivariate distribution that attains the bound. Our result thus provides a way to improve on the univariate marginal bound when additional distributional information is available. Importantly, we identify new instances in the subset selection and PERT network problem, where the bound on the expected value is computable in polynomial time. One interesting question that remains is whether these bounds can be tightened when information on overlapping multivariate marginals are available. Furthermore, is it possible to develop polynomial-time computable tight bounds in this case?

Acknowledgments

The authors thank the associate editor and two anonymous reviewers for their valuable comments and suggestions on improving the manuscript. The research was done when the first author was at the Operations Research Center at Massachusetts Institute of Technology and the Department of Combinatorics and Optimization at the University of Waterloo.

References

Agarwal, A., V. Zolotov, D. T. Blaauw. 2003. Statistical timing analysis using bounds and selective enumeration. *IEEE Trans. Comput.-Aided Design Integrated Circuits Systems* **22**(9) 1243–1260.

Ball, M. O., C. J. Colbourn, J. S. Provan. 1995. Network reliability. M. O. Ball, T. L. Magnanti, C. L. Monma, G. L. Nemhauser, eds. *Network Models*. Handbooks in Operations Research and Management Science, Vol. 7. Elsevier Science, Amsterdam, 673–762.

Bean, J. C. 1984. A Lagrangian algorithm for the multiple choice integer program. *Oper. Res.* **32**(5) 1185–1193.

Bertsimas, D., K. Natarajan, C. P. Teo. 2004. Probabilistic combinatorial optimization: Moments, semidefinite programming and asymptotic bounds. *SIAM J. Optim.* **15**(1) 185–209.

Bertsimas, D., K. Natarajan, C. P. Teo. 2006. Persistence in discete optimization under data uncertainty. *Math. Programming* **108**(2–3) 251–274.

Birge, J. R., M. J. Maddox. 1995. Bounds on expected project tardiness. *Oper. Res.* **43**(5) 838–850.

Blaauw, D. T., K. Chopra, A. Srivastava, L. Scheffer. 2008. Statistical timing analysis: From basic principles to state of the art. *IEEE Trans. Comput.-Aided Design Integrated Circuits Systems* **27**(4) 589–607.

Dodin, B. 1985. Bounding the project completion time distribution in PERT networks. *Oper. Res.* **33**(4) 862–881.

Dudzinski, K., S. Walukiewicz. 1987. Exact methods for the knapsack problem and its generalizations. *Eur. J. Oper. Res.* **28**(1) 3–21.

Embrechts, P., G. Pucetti. 2006. Bounds for functions of multivariate risks. *J. Multivariate Anal.* **97**(2) 526–547.

Fulkerson, D. R. 1962. Expected critical path length in PERT networks. *Oper. Res.* **10**(6) 808–817.

Genest, C., J. J. Quesada Molina, J. A. Rodríguez Lallena. 1995. De l'impossibilité de construire des lois à marges multidimensionnelles données à partir de copules. *Comptes rendus de l'Académie des sciences de Paris* **320**(6) 723–726.

Grötschel, M., L. Lovász, A. Schrijver. 1988. *Geometric Algorithms and Combinatorial Optimization*, Vol. 2. Springer-Verlag, Berlin.

Hagstrom, J. N. 1988. Computational complexity of PERT problems. *Networks* **18**(2) 139–147.

Kirkpatrick, T., N. Clark. 1966. PERT as an aid to logic design. *IBM J. Res. Development* **10**(2) 135–141.

Kleindorfer, G. B. 1971. Bounding distributions for a stochastic acyclic network. *Oper. Res.* **19**(7) 1586–1601.

Klein Hanevel, W. K. 1986. Robustness against dependence in PERT: An application of duality and distributions with known marginals. *Math. Programming Stud.* **27** 153–182.

Lai, T. L., H. Robbins. 1976. Maximally dependent random variables. *Proc. Natl. Acad. Sci. USA* **73**(2) 286–288.

Li, H., M. Scarsini, M. Shaked. 1996a. Bounds for the distribution of a multivariate sum. L. Rüschendorf, B. Schweizer, M. D. Taylor, eds. *Distributions with Fixed Marginals and Related Topics. IMS Lecture Notes*, Vol. 28. Institute of Mathematical Statistics, Hayward, CA, 198–212.

Li, H., M. Scarsini, M. Shaked. 1996b. Linkages: A tool for the construction of multivariate distributions with given nonoverlapping multivariate marginals. *J. Multivariate Anal.* **56**(1) 20–41.

McNeil, A. J., R. Frey, P. Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. *Princeton Series in Finance*. Princeton University Press, Princeton, NJ.

Meilijson, I. 1991. Sharp bounds on the largest of some linear combinations of random variables with given marginal distributions. *Probab. Engrg. Informational Sci.* **5**(1) 1–14.

Meilijson, I., A. Nadas. 1979. Convex majorization with an application to the length of critical path. *J. Appl. Probab.* **16**(3) 671–677.

Möhring, R. H. 2001. Scheduling under uncertainty: Bounding the makespan distribution. H. Alt, ed. *Computational Discrete Mathematics: Advanced Lectures*, *Lecture Notes in Computer Science*, Vol. 2122. Springer, Berlin, 79–97.

Möhring, R. H., F. J. Radermacher. 1989. The order-theoretic approach to scheduling: The stochastic case. R. Slowiński, J. Wegarlz, eds. *Advances in Project Scheduling. IMS Lecture Notes*. Elsevier, Amsterdam, 479–531.

Nadas, A. 1979. Probabilistic PERT. *IBM J. Res. Development* **23**(3) 339–347.

Natarajan, K., M. Song, C.-P. Teo. 2009. Persistency model and its applications in choice modeling. *Management Sci.* **55**(3) 453–469.

Provan, J. S., M. O. Ball. 1983. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM J. Comput.* **12**(4) 777–788.

Ringer, L. J. 1971. A statistical theory for PERT in which completion times of activities are inter-dependent. *Management Sci.* **17**(11) 717–723.

Rüschendorf, L. 1983. Solution of statistical optimization problem by rearrangement methods. *Metrika* **30** 55–61.

Rüschendorf, L. 2004. Comparison of multivariate risks and positive dependence. *J. Appl. Probab.* **41**(2) 391–406.

Sapatnekar, S. 2004. *Timing.* Kluwer Academic Publishers, Boston.

Scarsini, M. 1989. Copulae of probability measures on product spaces. *J. Multivariate Anal.* **31**(2) 201–219.

Shogan, A. W. 1983. Bounding distributions for a stochastic PERT network. *Networks* **12**(4) 777–788.

Sklar, A. 1959. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* **8** 229–231.

Spelde, H. G. 1976. Stochastische Netzpläne und ihre Anwendungim Baubetrieb. Ph.D. thesis, Rheinisch-Westfälische Technische Hochschule, Aachen, Germany.

Valiant, L. G. 1979. The complexity of enumeration and reliability problems. *SIAM J. Comput.* **8**(3) 410–421.

van Dorp, J. R., M. R. Duffey. 1999. Statistical dependence in risk analysis for project networks using Monte Carlo methods. *Internat. J. Production Econom.* **58**(1) 17–29.

Visweswariah, C., K. Ravindran, K. Kalafala, S. G. Walker, S. Narayan, D. K. Beece, J. Piaget, N. Venkateswaran, J. G. Hemmett. 2006. First-order incremental block-based statistical timing analysis. *IEEE Trans. Comput.-Aided Design Integrated Circuits Systems* **25**(10) 2170–2179.

Weiss, G. 1986. Stochastic bounds on distributions of optimal value functions with applications to PERT, network flows and reliability. *Oper. Res.* **34**(4) 595–605.

Wollmer, R. D. 1985. Critical path planning under uncertainty. *Math. Programming Stud.* **25** 164–171.