

On the Complexity of Non-Overlapping Multivariate Marginal Bounds for Probabilistic Combinatorial Optimization Problems

Xuan Vinh Doan *

Karthik Natarajan †

First submitted: October 2010. Revised: March 7, 2011

Abstract

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 (Journal of Applied Probability, 1979). We generalize the bound to non-overlapping multivariate marginals using multiple choice integer programming. For discrete distributions, new instances of polynomial time computable multivariate marginal bounds are identified. For the problem of selecting up to M items out a set of N items of maximum total weight, the 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 worst-case 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. An instance of a polynomial time solvable two stage stochastic program arising from project crashing is identified. An important feature of the bound is that it is exactly achievable by a joint distribution, unlike many of the existing bounds.

*Department of Combinatorics and Optimization, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada, vanxuan@uwaterloo.ca. The research was partly done when the author was in the Operations Research Center at Massachusetts Institute of Technology.

†Department of Management Sciences, College of Business, City University of Hong Kong. Email: knataraj@cityu.edu.hk.

1 Introduction

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

$$\begin{aligned} Z(\mathbf{c}) = \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X} \subseteq \{0, 1\}^N. \end{aligned} \tag{1}$$

Suppose $\tilde{\mathbf{c}} = (\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_N)$ 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 $\{\mathbf{x} \in \mathcal{X} \mid x_i = 0\}$ and $\{\mathbf{x} \in \mathcal{X} \mid x_i = 1\}$ are nonempty for each index $i = 1, \dots, N$. Meilijson and Nadas [20] 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 bounding expected project duration in PERT networks. A PERT network is a directed acyclic graph that represents a project of several activities with partially specified precedence relationships among the activities. In an activity-on-arc PERT network, an arc represents an activity and a node represents the completion of all the activities entering it. Let $\mathcal{V} = \{1, \dots, M\}$ denote the set of nodes in the network where nodes 1 and M represent the start and end of the project and \mathcal{E} denote 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:

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

The challenge in analyzing PERT networks stems from the random activity duration. Meilijson and Nadas [20] developed 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.

For the generic combinatorial optimization problem (1), their upper bound was obtained through the solution of a convex minimization problem over the decision variables $\mathbf{d} = (d_1, d_2, \dots, d_N) \in \mathbb{R}^N$:

$$\mathbb{E}[Z(\tilde{\mathbf{c}})] \leq \inf_{\mathbf{d}} \left(Z(\mathbf{d}) + \sum_{i=1}^N \mathbb{E}[\tilde{c}_i - d_i]^+ \right), \tag{3}$$

where $[y]^+ = \max(y, 0)$. Furthermore, they showed that the bound in (3) was tight by constructing a joint distribution for $\tilde{\mathbf{c}}$ with the correct marginals, that attained the upper bound exactly. The bound can hence be interpreted as being *robust against dependence*. For PERT networks, Klein Haneveld [13] interpreted the formulation on the right hand side of (3) 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 [15]:

$$\mathbb{E} \left[\max_{i=1, \dots, N} \tilde{c}_i \right] \leq \inf_d \left(d + \sum_{i=1}^N \mathbb{E} [\tilde{c}_i - d]^+ \right). \quad (4)$$

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 Rüschemdorf [26]:

$$\mathbb{E} \left[\sum_{i=1}^N \tilde{c}_i - T \right]^+ \leq \inf_{\sum_i d_i = T} \left(\sum_{i=1}^N \mathbb{E} [\tilde{c}_i - d]^+ \right). \quad (5)$$

McNeil et al. [18] have discussed the relevance of these bounds to the actuarial sciences and portfolio risk management community. Weiss [34] 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 [13], Birge and Maddox [5] and Bertsimas et al. [3], [4]. Meilijson [19] and Natarajan et al. [23] 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 non-overlapping multivariate marginals are available. A popular tool to construct multivariate distributions from univariate distributions is the *copula* which helps distinguish the dependencies from the marginals. Formally, a 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. [18]). Sklar [30] showed that for all multivariate distributions F with marginal distributions F_1, F_2, \dots, F_N , there exists a copula $C : [0, 1]^N \rightarrow [0, 1]$ such that:

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

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

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

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 [28]). Genest et al. [10] showed that the only copula consistent with all non-overlapping multivariate marginals is the independence copula. Li et al. [17] proposed a linkage function to characterize distributions with non-overlapping 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. [16], Rüschendorf [27] and Embrechts and Pucetti [8] 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 two applications in risk management and PERT networks:

- (a) Consider a N dimensional vector of nonnegative random losses that can be partitioned into sub-vectors representing losses for policies within specific risk categories. The goal is to compute the worst-case expected aggregate loss of a financial position, given sub-vector loss distributions but allowing for arbitrary dependencies between sub-vectors. These sub-vectors could represent losses from companies in industry sectors such as healthcare, 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 while 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. [1] and Möhring and Radermacher [22] review methods that compute, bound or approximate the expected project duration with independent activity durations. Ringer [25] and van Dorp and Duffey [33] 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 [9] 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 [14] and Shogan [29]

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. In the construction project example with each project team responsible for the set of activities entering a particular node, it is reasonable to assume that the team is knowledgeable about the joint distribution of the activities that they are responsible for. The project manager is then interested in evaluating the worst-case expected project completion time that is compatible with these estimates, but allowing for factors that might make activity durations handled by different teams correlated.

1.1 Problem Description

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

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

Given a vector $\mathbf{c} \in \mathbb{R}^N$, let $\mathbf{c}_r \in \mathbb{R}^{N_r}$ denote the sub-vector formed with the elements in the r th subset \mathcal{N}_r where $N_r = |\mathcal{N}_r|$ is the size of the subset. The probability measures P_r for the sub-vectors $\tilde{\mathbf{c}}_r$ are assumed to be known. Let $\mathcal{P}(P_1, \dots, P_R)$ denote the set of joint probability measures for the random vector $\tilde{\mathbf{c}}$ consistent with the prescribed probability measures for the sub-vectors $\tilde{\mathbf{c}}_r$. No assumption on the dependencies between random variables in distinct subsets are made. The independence measure among the sub-vectors 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 non-overlapping multivariate marginals:

$$Z^* = \sup_{P \in \mathcal{P}(P_1, \dots, P_R)} \int Z(\mathbf{c}) dP(\mathbf{c}). \quad (6)$$

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 sub-vectors $\tilde{\mathbf{c}}_r$ are defined by the scenarios

\mathbf{c}_{rk} for $k = 1, \dots, K_r$ with probabilities p_{rk} satisfying $\sum_k p_{rk} = 1$:

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

The key results that we obtain in the paper are:

- (a) In Section 2, we generalize the Meilijson and Nadas [20] bound to non-overlapping 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 [20], where the size of each subset in the partition is $O(1)$.
- (b) In Section 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 \mathcal{X}_r denote the projection of \mathcal{X} onto the space of the decision variables in the r th subset:

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

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 (3) using an expanded set of decision variables.

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

$$\hat{Z}_u^* = \min_{\mathbf{d}_1, \dots, \mathbf{d}_R} \left(\max_{\mathbf{x} \in \mathcal{X}} \sum_{r=1}^R d_r(\mathbf{x}_r) + \sum_{r=1}^R \mathbb{E}_{P_r} \left[\max_{\mathbf{x}_r \in \mathcal{X}_r} (\tilde{\mathbf{c}}_r' \mathbf{x}_r - d_r(\mathbf{x}_r)) \right] \right). \quad (7)$$

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

Proof.

Step 1: Prove that $Z^* \leq \hat{Z}_u^*$.

For any feasible solution $\mathbf{x} \in \mathcal{X}$ and a collection of vectors $\mathbf{d}_1, \dots, \mathbf{d}_R$ with $d_r(\mathbf{0}) = 0$,

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

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

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

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

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

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

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

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

Step 2: Prove that $Z^* \geq \hat{Z}_u^*$.

We provide an explicit construction of a distribution $\bar{P} \in \mathcal{P}(P_1, \dots, P_R)$ such that $\mathbb{E}_{\bar{P}} [Z(\tilde{\mathbf{c}})] \geq \hat{Z}_u^*$.

The upper bound \hat{Z}_u^* can be computed as the optimal objective value to a linear program with decision variables $(d_r(\mathbf{x}_r), t, y_{rk})_{r,k,\mathbf{x}_r}$:

$$\begin{aligned} \hat{Z}_u^* = \min \quad & t + \sum_{r=1}^R \sum_{k=1}^{K_r} y_{rk} \\ \text{s.t.} \quad & t \geq \sum_{r=1}^R d_r(\mathbf{x}_r), \quad \forall \mathbf{x} \in \mathcal{X}, \\ & y_{rk} \geq p_{rk} (\mathbf{c}'_{rk} \mathbf{x}_r - d_r(\mathbf{x}_r)), \quad \forall \mathbf{x}_r \in \mathcal{X}_r, k = 1, \dots, K_r, r = 1, \dots, R. \end{aligned} \tag{8}$$

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

$$\begin{aligned}
\hat{Z}_u^* = \max & \sum_{r=1}^R \sum_{k=1}^{K_r} \sum_{\mathbf{x}_r \in \mathcal{X}_r} p_{rk} \mathbf{c}'_{rk} \mathbf{x}_r \gamma_{rk}(\mathbf{x}_r) \\
\text{s.t.} & \sum_{\mathbf{x} \in \mathcal{X}} \lambda(\mathbf{x}) = 1, \\
& \sum_{\mathbf{x}_r \in \mathcal{X}_r} \gamma_{rk}(\mathbf{x}_r) = 1, \quad \forall k = 1, \dots, K_r, \quad r = 1, \dots, R, \\
& \sum_{\mathbf{v} \in \mathcal{X}: \mathbf{v}_r = \mathbf{x}_r} \lambda(\mathbf{v}) - \sum_{k=1}^{K_r} p_{rk} \gamma_{rk}(\mathbf{x}_r) = 0, \quad \forall \mathbf{x}_r \in \mathcal{X}_r, \quad r = 1, \dots, R, \\
& \lambda(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in \mathcal{X}, \\
& \gamma_{rk}(\mathbf{x}_r) \geq 0, \quad \forall \mathbf{x}_r \in \mathcal{X}_r, \quad k = 1, \dots, K_r, \quad r = 1, \dots, R.
\end{aligned} \tag{9}$$

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

- (a) Pick a random feasible solution $\mathbf{x} \in \mathcal{X}$ with probability $\lambda^*(\mathbf{x})$.
- (b) For each r , the random sub-vector $\tilde{\mathbf{c}}_r$ is given by the scenarios \mathbf{c}_{rk} with probabilities $q_{rk}^*(\mathbf{x}_r)$ defined as:

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

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

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

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

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

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

From steps 1 and 2, $Z^* = \hat{Z}_u^*$. □

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 (4) and (5) to multivariate marginal bounds.

Proposition 1

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

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

and the bound is tight.

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

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

and the bound is tight.

Proof.

- (i) Let $\mathcal{X} = \{\mathbf{e}_1^{(N)}, \mathbf{e}_2^{(N)}, \dots, \mathbf{e}_N^{(N)}\}$ where $\mathbf{e}_i^{(N)}$ is the unit vector in \mathbb{R}^N with 1 in the i th 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 $\mathcal{X}_r = \{\mathbf{e}_1^{(N_r)}, \mathbf{e}_2^{(N_r)}, \dots, \mathbf{e}_{N_r}^{(N_r)}, \mathbf{0}\}$. 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:

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

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 while 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:

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

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 $r = 1, \dots, R + 1$. Using Theorem 1, the tight upper bound is given as:

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

We have: $x^+ + y^+ \geq [x + y]^+$ 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:

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

The term $[\sum_r d_r - T]^+$ is non-decreasing in d_r . If the term $\sum_r d_r - T = \epsilon > 0$, we can decrease at least one of the d_r 's by ϵ such that the first term decreases by ϵ while 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 which satisfies $\sum_r d_r = T$. Thus the tight upper bound is hence the optimal value of the optimization problem:

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

□

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

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

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

$$\mathcal{X} = \left\{ \mathbf{x} \in \{0, 1\}^N \mid \sum_{r=1}^R \mathbf{A}_r \mathbf{x}_r \leq \mathbf{b} \right\}.$$

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 [2]).

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

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

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

$$\begin{aligned} \hat{Z}(\mathbf{d}_1, \dots, \mathbf{d}_R) = \quad & \max_{\mathbf{z}_1, \dots, \mathbf{z}_R} \sum_{r=1}^R \sum_{\mathbf{x}_r \in \mathcal{X}_r} d_r(\mathbf{x}_r) z_r(\mathbf{x}_r) \\ \text{s.t.} \quad & \sum_{r=1}^R \sum_{\mathbf{x}_r \in \mathcal{X}_r} \mathbf{A}_r \mathbf{x}_r z_r(\mathbf{x}_r) \leq \mathbf{b}, \\ & \sum_{\mathbf{x}_r \in \mathcal{X}_r} z_r(\mathbf{x}_r) = 1, \quad \forall r = 1, \dots, R, \\ & z_r(\mathbf{x}_r) \in \{0, 1\}, \quad \forall \mathbf{x}_r \in \mathcal{X}_r, r = 1, \dots, R. \end{aligned} \tag{10}$$

Proof.

Let \mathcal{Z} denote the feasible region to the multiple choice integer program (10). Thus,

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

Step 1: Prove that $\hat{Z}(\mathbf{d}_1, \dots, \mathbf{d}_R) \leq \max_{\mathbf{x} \in \mathcal{X}} \sum_r d_r(\mathbf{x}_r)$.

Given vectors $\mathbf{d}_1, \dots, \mathbf{d}_R$, consider an optimal solution $(\mathbf{z}_1^*, \dots, \mathbf{z}_R^*)$ to formulation (10). Set $\mathbf{x}_r^* = \sum_{\mathbf{x}_r \in \mathcal{X}_r} \mathbf{x}_r z_r^*(\mathbf{x}_r)$. Then, we check for the feasibility of \mathbf{x}^* :

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

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

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

Step 2: Prove that $\hat{Z}(\mathbf{d}_1, \dots, \mathbf{d}_R) \geq \max_{\mathbf{x} \in \mathcal{X}} \sum_r d_r(\mathbf{x}_r)$.

Given vectors $\mathbf{d}_1, \dots, \mathbf{d}_R$, consider an optimal solution \mathbf{x}^* to $\max_{\mathbf{x} \in \mathcal{X}} \sum_r d_r(\mathbf{x}_r)$. For each r , set $z_r^*(\mathbf{x}_r^*) =$

1. Set $z_r^*(\mathbf{x}_r) = 0$ for all $\mathbf{x}_r \in \mathcal{X}_r, \mathbf{x}_r \neq \mathbf{x}_r^*$. Clearly, $\sum_{\mathbf{x}_r \in \mathcal{X}_r} z_r^*(\mathbf{x}_r) = 1$. Thus, $(z_1^*, \dots, z_R^*) \in \mathcal{Z}$ since:

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

The objective value satisfies:

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

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

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:

$$\begin{aligned} Z(\mathbf{c}) = \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{e}^{(N)'}\mathbf{x} \leq M, \\ & \mathbf{x} \in \{0, 1\}^N. \end{aligned} \tag{11}$$

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 (11) 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 (11) 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 $\mathcal{N} = \{1, \dots, N\}$. Assume that the size of each subset \mathcal{N}_r in the partition is $N_r = O(\log N)$. The projection of the feasible region of (11) on the space of decision variables in the r th subset is:

$$\mathcal{X}_r = \left\{ \mathbf{x}_r \in \{0, 1\}^{N_r} \mid \mathbf{e}^{(N_r)'}\mathbf{x}_r \leq M \right\}.$$

Using Proposition 2, Z^* can be computed as:

$$\begin{aligned} Z^* = \min \quad & t + \sum_{r=1}^R \sum_{k=1}^{K_r} y_{rk} \\ \text{s.t.} \quad & t \geq \hat{Z}(\mathbf{d}_1, \dots, \mathbf{d}_R), \\ & y_{rk} \geq p_{rk}\mathbf{c}'_k\mathbf{x}_r - p_{rk}d_r(\mathbf{x}_r) \quad \forall \mathbf{x}_r \in \mathcal{X}_r, k = 1, \dots, K_r, r = 1, \dots, R, \end{aligned} \tag{12}$$

Let $|\mathcal{X}_r|$ denote the size of the set \mathcal{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 (12) is polynomial in the size of the input since:

$$\begin{aligned} \text{Number of decision variables in (12)} &= \underbrace{1}_t + \underbrace{\sum_{r=1}^R |\mathcal{X}_r|}_{d_r(\mathbf{x}_r)} + \underbrace{\sum_{r=1}^R K_r}_{y_{rk}} \\ &= 1 + \sum_{r=1}^R 2^{O(\log N)} + \sum_{r=1}^R K_r \\ &= 1 + \text{poly}(N) + O(RK_{\max}), \end{aligned}$$

To analyze the complexity of the optimization problem (12), consider the separation version of the problem (see Grötschel, Lovász and Schrijver [11]). Since testing for the feasibility of the second set of inequalities in (12) is easy, we restrict our attention to:

Given t and \mathbf{d} , check if

$$t \geq \hat{Z}(\mathbf{d}_1, \dots, \mathbf{d}_R).$$

If not, find a violated inequality.

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

$$\begin{aligned} \hat{Z}(\mathbf{d}_1, \dots, \mathbf{d}_R) = \max & \sum_{r=1}^R \sum_{\mathbf{x}_r \in \mathcal{X}_r} d_r(\mathbf{x}_r) z_r(\mathbf{x}_r) \\ \text{s.t.} & \sum_{r=1}^R \sum_{\mathbf{x}_r \in \mathcal{X}_r} \mathbf{e}^{(N_r)'} \mathbf{x}_r z_r(\mathbf{x}_r) \leq M, \\ & \sum_{\mathbf{x}_r \in \mathcal{X}_r} z_r(\mathbf{x}_r) = 1, & \forall r = 1, \dots, R, \\ & z_r(\mathbf{x}_r) \in \{0, 1\}, & \forall \mathbf{x}_r \in \mathcal{X}_r, r = 1, \dots, R. \end{aligned} \quad (13)$$

Dudzinski and Walukiewicz [7] 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 (13) when restricted to the first S subsets with a knapsack capacity restricted to B :

$$\begin{aligned} Z_S(B) = \max & \sum_{r=1}^S \sum_{\mathbf{x}_r \in \mathcal{X}_r} d_r(\mathbf{x}_r) z_r(\mathbf{x}_r) \\ \text{s.t.} & \sum_{r=1}^S \sum_{\mathbf{x}_r \in \mathcal{X}_r} \mathbf{e}^{(N_r)'} \mathbf{x}_r z_r(\mathbf{x}_r) \leq B, \\ & \sum_{\mathbf{x}_r \in \mathcal{X}_r} z_r(\mathbf{x}_r) = 1, & \forall r = 1, \dots, S, \\ & z_r(\mathbf{x}_r) \in \{0, 1\}, & \forall \mathbf{x}_r \in \mathcal{X}_r, r = 1, \dots, S. \end{aligned} \quad (14)$$

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

$$\begin{aligned} Z_S(B) = \max & Z_{S-1} \left(B - \mathbf{e}^{(N_S)'} \mathbf{x}_S \right) + d_S(\mathbf{x}_S) \\ \text{s.t.} & \mathbf{x}_S \in \mathcal{X}_S, \\ & \mathbf{e}^{(N_S)'} \mathbf{x}_S \leq B, \end{aligned} \quad (15)$$

where $Z_S(B)$ is $-\infty$ if the set is infeasible. The optimal objective value $\hat{Z}(\mathbf{d}_1, \dots, \mathbf{d}_R)$ is equal to $Z_R(M)$. The overall time complexity of implementing this method is $O(M \sum_r |\mathcal{X}_r|)$. Since $|\mathcal{X}_r|$ is polynomially bounded in the N , $\hat{Z}(\mathbf{d}_1, \dots, \mathbf{d}_R)$ can be computed in time polynomial in the size of the input. From the equivalence on separation and optimization [11], the result follows. \square

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 [20] 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 [20], Weiss [34], Klein Haneveld [13] and Bertsimas et al. [3, 4] in proposing algorithms for univariate marginals.

Theorem 3

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

$$Z^* \leq Z_u^* = \min_{\mathbf{d}} \left(Z(\mathbf{d}) + \sum_{r=1}^R \mathbb{E}_{P_r} \left[\max_{\mathbf{x}_r \in \mathcal{X}_r} (\tilde{\mathbf{c}}_r' \mathbf{x}_r - \mathbf{d}_r' \mathbf{x}_r) \right] \right). \quad (16)$$

(ii) Assume that for each $r = 1, \dots, R$, the set of sub-vectors $\mathbf{x}_r \in \mathcal{X}_r / \{\mathbf{0}\}$ are linearly independent.

Then, $Z^* = Z_u^*$.

Proof.

(i) Consider an optimal solution \mathbf{d}^* to formulation (16). 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:

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

The first inequality is valid since the set of $(\mathbf{d}_r^*(\mathbf{x}_r))_{\mathbf{x}_r \in \mathcal{X}_r}$ forms a feasible solution in Theorem 1. Thus $Z^* \leq Z_u^*$.

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

$$\begin{aligned} Z_u^* = \min \quad & t + \sum_{r=1}^R \sum_{k=1}^{K_r} y_{rk} \\ \text{s.t.} \quad & t \geq \sum_{r=1}^R \mathbf{d}'_r \mathbf{x}_r, \quad \forall \mathbf{x} \in \mathcal{X}, \\ & y_{rk} \geq p_{rk} (\mathbf{c}'_{rk} \mathbf{x}_r - \mathbf{d}'_r \mathbf{x}_r), \quad \forall \mathbf{x}_r \in \mathcal{X}_r, k = 1, \dots, K_r, r = 1, \dots, R. \end{aligned} \tag{17}$$

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

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

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

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

and

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

Given that the sub-vectors $\mathbf{x}_r \in \mathcal{X}_r / \{\mathbf{0}\}$ are linearly independent, this implies:

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

With this equality, it is easy to verify that the distribution \bar{P} constructed in Theorem 1 provides the tight upper bound as before. Since 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 $\mathcal{N}_1 = \{1, 2\}$ and $\mathcal{N}_2 = \{3, 4\}$. The feasible set \mathcal{X} contains a total of 11 feasible solutions and $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}^2$. Define a total of 8 decision variables $d_r(\mathbf{x}_r)$ for $\mathbf{x}_r \in \{0, 1\}^2$, $r = 1, 2$, with $d_1(0, 0) = d_2(0, 0) = 0$. The tight bound Z^* in Theorem 1 can be calculated as:

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

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:

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

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

For project networks with deterministic activity durations, $Z(\mathbf{c})$ can be computed in polynomial time by solving the linear programming relaxation of formulation (2):

$$\begin{aligned}
Z(\mathbf{c}) = \max \quad & \sum_{(i,j) \in \mathcal{E}} c_{ij} x_{ij} \\
\text{s.t.} \quad & \sum_{i:(j,i) \in \mathcal{E}} x_{ji} - \sum_{i:(i,j) \in \mathcal{E}} x_{ij} = \begin{cases} 1, & \text{if } j = 1, \\ -1, & \text{if } j = M, \\ 0, & \text{if } j = 2, \dots, M-1, \end{cases} \\
& x_{ij} \geq 0, \quad \forall (i,j) \in \mathcal{E}.
\end{aligned} \tag{18}$$

This is due to the unimodular structure of the constraint matrix for network flow problems. While the deterministic optimization problem is easy, the computation of the expected optimal value is still a challenging problem. Hagstrom [12] 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 [12])

- (i) **CDF** and the two-state version of **MEAN** are $\#\mathcal{P}$ -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 $\mathcal{P} = \mathcal{NP}$.

The classes $\#\mathcal{P}$ and $\#\mathcal{P}$ -complete are the counting versions of the \mathcal{NP} and \mathcal{NP} -complete recognition problems. This computational complexity class was first introduced by Valiant [32]. 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 $\#\mathcal{P}$ -complete. Valiant [32] and Provan and Ball [24] provide instances of such problems. The PERT problem falls under this category. Since the existence of a polynomial-time algorithm for solving a $\#\mathcal{P}$ -complete problem would imply $\mathcal{P} = \mathcal{NP}$, it seems highly unlikely that these problems can be solved in polynomial time. While 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 theorem next.

Theorem 5 (Ball et al. [1], Möhring [21])

- (i) *The two-state version of **MEAN** and **CDF** for a series-parallel graph is \mathcal{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 $\{0, 1, 2, \dots, q\}$, **MEAN** and **CDF** can be computed in time polynomial in the size of the network and q .*

Several authors including Kleindorfer [14], Spelde [31], Dodin [6] and Möhring [21] 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 [9] and Shogan [29] developed bounds on the expected project completion time. Meilijson and Nadas [20] 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 sub-vector of random activity durations $\tilde{\mathbf{c}}_j = (\tilde{c}_{ij})_{i:(i,j) \in \mathcal{E}}$ for arcs entering a node j . Let \mathbf{c}_{jk} denote a realization of this sub-vector occurring with probability $P_j(\tilde{\mathbf{c}}_j = \mathbf{c}_{jk}) = p_{jk}$ for $k = 1, \dots, K_j$ where $\sum_k p_{jk} = 1$. Without loss of generality, we assume that for each node $j = 2, \dots, M - 1$, 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 (2) 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:

$$Z^* = \min_{\mathbf{d}} \left(Z(\mathbf{d}) + \sum_{j=2}^{M-1} \mathbb{E} \left[\max_{i:(i,j) \in \mathcal{E}} (\tilde{c}_{ij} - d_{ij}) \right]^+ + \mathbb{E} \left[\max_{i:(i,M) \in \mathcal{E}} (\tilde{c}_{iM} - d_{iM}) \right] \right). \quad (19)$$

By using the dual formulation for the linear programming representation of $Z(\mathbf{d})$, Z^* in (18) is computed as:

$$\begin{aligned}
Z^* = \min \quad & w_M - w_1 + \sum_{j=2}^M \sum_{k=1}^{K_j} y_{jk} \\
\text{s.t.} \quad & w_j - w_i \geq d_{ij}, \quad \forall (i, j) \in \mathcal{E}, \\
& y_{jk} \geq p_{jk} c_{ijk} - p_{jk} d_{ij}, \quad \forall i : (i, j) \in \mathcal{E}, k = 1, \dots, K_j, j = 2, \dots, M, \\
& y_{jk} \geq 0, \quad \forall k = 1, \dots, K_j, j = 2, \dots, M - 1.
\end{aligned} \tag{20}$$

The linear program in (20) 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 (19) as follows:

$$\begin{aligned}
Z^* = \min \quad & t \\
\text{s.t.} \quad & t \geq Z(\mathbf{d}) + \sum_{j=2}^{M-1} \mathbb{E} \left[\max_{i:(i,j) \in \mathcal{E}} (\tilde{c}_{ij} - d_{ij}) \right]^+ + \mathbb{E} \left[\max_{i:(i,M) \in \mathcal{E}} (\tilde{c}_{iM} - d_{iM}) \right].
\end{aligned} \tag{21}$$

The separation version of this problem is:

Given t and \mathbf{d} , check if

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

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 $\tilde{c}_{\max} = \max_{i=1, \dots, n} \tilde{c}_i$ 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 (21) 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 $\mathbb{E} [\max_{i=1, \dots, n} (\tilde{c}_i - d_i)]$ with respect to \mathbf{d} . For $Z(\mathbf{d})$, the optimal solution to Problem (2) provides the subgradient. The piecewise linear function

$f(\mathbf{d}; \mathbf{c}) = \max_{i=1, \dots, n} (\tilde{c}_i - d_i)$ has $g(\mathbf{d}; \mathbf{c}) = -\mathbf{e}_{i_{\min}^{\mathbf{d}}(\mathbf{c})}^{(n)}$, where $i_{\min}^{\mathbf{d}}(\mathbf{c}) = \min\{j : c_j - d_j = f(\mathbf{d}; \mathbf{c})\}$, 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 $P_i(\mathbf{d}) = P(\tilde{c}_i - d_i = f(\mathbf{d}; \tilde{\mathbf{c}}), i = i_{\min}^{\mathbf{d}}(\tilde{\mathbf{c}}))$ for all $i = 1, \dots, n$. We have:

$$P_i(\mathbf{d}) = P\left(\tilde{c}_i - d_i > \max_{j=1, \dots, i-1} (\tilde{c}_j - d_j), \tilde{c}_i - d_i \geq \max_{j=i+1, \dots, n} (\tilde{c}_j - d_j)\right), i = 1, \dots, n.$$

The random variable $\tilde{c}_{\max}^i(\mathbf{d}) = \max_{j=1, \dots, i-1} (\tilde{c}_j - d_j)$ 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 $\tilde{c}_{\max}^{-i}(\mathbf{d}) = \max_{j=i+1, \dots, n} (\tilde{c}_j - d_j)$. The random variables \tilde{c}_i , $\tilde{c}_{\max}^i(\mathbf{d})$, and $\tilde{c}_{\max}^{-i}(\mathbf{d})$ are independent. The total supporting points of the joint distribution is bounded from above by $(i-1)(n-i-1)K_{\max}^3$. Thus the probability $P_i(\mathbf{d})$ can be calculated in polynomial time, which implies 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 [11] 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 $(i, j) \in \mathcal{E}$ 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 $f_{ij} \geq 0$ for all $(i, j) \in \mathcal{E}$. 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 [35]:

$$\begin{aligned} \min \quad & \alpha \sum_{(i,j) \in \mathcal{E}} f_{ij} t_{ij} + \mathbb{E}[Z(\tilde{\mathbf{c}} - \mathbf{t})] \\ \text{s.t.} \quad & 0 \leq t_{ij} \leq u_{ij}, \quad \forall (i, j) \in \mathcal{E}, \end{aligned} \tag{22}$$

where $\alpha \geq 0$ is a parameter that provides a tradeoff 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 (22) is replaced by a distributional robust optimization problem:

$$\begin{aligned} \min \quad & \left(\alpha \sum_{(i,j) \in \mathcal{E}} f_{ij} t_{ij} + \sup_{P \in \mathcal{P}(P_1, \dots, P_R)} \mathbb{E}_P [Z(\tilde{\mathbf{c}} - \mathbf{t})] \right) \\ \text{s.t.} \quad & 0 \leq t_{ij} \leq u_{ij}, \quad \forall (i, j) \in \mathcal{E}, \end{aligned} \quad (23)$$

where \mathcal{P} is the set of all probability measures P with the given marginals. Since 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 (20) to the inner supremum in (23) provides a linear optimization formulation for the distributional robust problem:

$$\begin{aligned} \min \quad & \alpha \sum_{(i,j) \in \mathcal{E}} f_{ij} t_{ij} + w_M - w_1 + \sum_{j=2}^M \sum_{k=1}^{K_j} y_{jk} \\ \text{s.t.} \quad & 0 \leq t_{ij} \leq u_{ij}, \quad \forall (i, j) \in \mathcal{E}, \\ & w_j - w_i \geq d_{ij}, \quad \forall (i, j) \in \mathcal{E}, \\ & y_{jk} \geq p_{jk} c_{ijk} - p_{jk} t_{ij} - p_{jk} d_{ij}, \quad \forall i : (i, j) \in \mathcal{E}, k = 1, \dots, K_j, j = 2, \dots, M, \\ & y_{jk} \geq 0, \quad \forall k = 1, \dots, K_j, j = 2, \dots, M - 1. \end{aligned} \quad (24)$$

Since 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 [20] bound for probabilistic combinatorial optimization problems is extended from univariate marginals to non-overlapping 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?

Acknowledgment

We would like to thank the associate editor and two anonymous reviewers for their valuable comments and suggestions on improving the manuscript.

References

- [1] M. O. Ball, C. J. Colbourn, and J. S. Provan. Network reliability. In *Handbook in Operations Research and Management Science: Network Models*, volume 7, pages 673–762. Elsevier Science, 1995.
- [2] J. C. Bean. A Lagrangian algorithm for the multiple choice integer program. *Operations Research*, 32(5):1185–1193, 1984.
- [3] D. Bertsimas, K. Natarajan, and C. P. Teo. Probabilistic combinatorial optimization: moments, semidefinite programming and asymptotic bounds. *SIAM Journal of Optimization*, 15(1):185–209, 2004.
- [4] D. Bertsimas, K. Natarajan, and C. P. Teo. Persistence in discrete optimization under data uncertainty. *Mathematical Programming*, 108(2-3):251–274, 2006.
- [5] J. R. Birge and M. J. Maddox. Bounds on expected project tardiness. *Operations Research*, 43(5):838–850, 1995.
- [6] B. Dodin. Bounding the project completion time distribution in PERT networks. *Operations Research*, 33(4):862–881, 1985.
- [7] K. Dudzinski and S. Walukiewicz. Exact methods for the knapsack problem and its generalizations. *European Journal of Operational Research*, 28:3–21, 1987.
- [8] P. Embrechts and G. Pucetti. Bounds for functions of multivariate risks. *Journal of Multivariate Analysis*, 97(2):526–547, 2006.
- [9] D. R. Fulkerson. Expected critical path length in PERT networks. *Operations Research*, 10(6):808–817, 1962.
- [10] C. Genest, J. J. Quesada Molina, and J. A. Rodríguez Lallena. 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, 1995.
- [11] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*. Springer, 1988.
- [12] J. N. Hagstrom. Computational complexity of PERT problems. *Networks*, 18(2):139–147, 1988.
- [13] W. K. Klein Haneveld. Robustness against dependence in PERT: An application of duality and distributions with known marginals. *Mathematical Programming Studies*, 27:153–182, 1986.
- [14] G. B. Kleindorfer. Bounding distributions for a stochastic acyclic network. *Operations Research*, 19(7):1586–1601, 1971.
- [15] T. L. Lai and H. Robbins. Maximally dependent random variables. *Proceedings of the National Academy of Sciences*, 73(2):286–288, 1976.
- [16] H. Li, M. Scarsini, and M. Shaked. Bounds for the distribution of a multivariate sum. In L. Rüschendorf, B. Schweizer, and M. D. Taylor, editors, *Distributions with Fixed Marginals and Related Topics*, volume 28 of *IMS Lecture Notes*, pages 198–212. Institute of Mathematical Statistics, 1996.
- [17] H. Li, M. Scarsini, and M. Shaked. Linkages: A tool for the construction of multivariate distributions with given nonoverlapping multivariate marginals. *Journal of Multivariate Analysis*, 56:20–41, 1996.
- [18] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton Series in Finance. Princeton University Press, 2005.

- [19] I. Meilijson. Sharp bounds on the largest of some linear combinations of random variables with given marginal distributions. *Probability in the Engineering and Informational Sciences*, 5:1–14, 1991.
- [20] I. Meilijson and A. Nadas. Convex majorization with an application to the length of critical path. *Journal of Applied Probability*, 16:671–677, 1979.
- [21] R. H. Möhring. Scheduling under uncertainty: Bounding the makespan distribution. In H. Alt, editor, *Computational Discrete Mathematics: Advanced Lectures*, volume 2122 of *Lecture Notes in Computer Science*, pages 79–97. Springer, 2001.
- [22] R. H. Möhring and F. J. Radermacher. The order-theoretic approach to scheduling: The stochastic case. In R. Slowiński and J. Wegarz, editors, *Advances in Project Scheduling*, IMS Lecture Notes, pages 479–531. Elsevier, Amsterdam, 1989.
- [23] K. Natarajan, M. Song, and C. P. Teo. Persistency model and its applications in choice modeling. *Management Science*, 55(3):453–469, 2009.
- [24] J. S. Provan and M. O. Ball. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM Journal of Computing*, 12:777–788, 1983.
- [25] L. J. Ringer. A statistical theory for pert in which completion times of activities are inter-dependent. *Management Science*, 17:717–723, 1971.
- [26] L. Rüschendorf. Solution of statistical optimization problem by rearrangement methods. *Metrika*, 30:55–61, 1983.
- [27] L. Rüschendorf. Comparison of multivariate risks and positive dependence. *Journal of Applied Probability*, 41(2):391–406, 2004.
- [28] M. Scarsini. Copulae of probability measures on product spaces. *Journal of Multivariate Analysis*, 31(2):201–219, 1989.
- [29] A. W. Shogan. Bounding distributions for a stochastic PERT network. *Networks*, 12:777–788, 1983.
- [30] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231, 1959.
- [31] H. G. Spelde. Bounds for the distribution function of network variables. In *Proceedings of the 1st Symposium of Operations Research*, pages 113–123. University Heidelberg, 1977.
- [32] L. G. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal of Computing*, 8:410–421, 1979.
- [33] J. R. van Dorp and M.R. Duffey. Statistical dependence in risk analysis for project networks using monte carlo methods. *International Journal of Production Economics*, 58:17–29, 1999.
- [34] G. Weiss. Stochastic bounds on distributions of optimal value functions with applications to PERT, network flows and reliability. *Operations Research*, 34(4):595–605, 1986.
- [35] R. D. Wollmer. Critical path planning under uncertainty. *Mathematical Programming Studies*, 25:164–171, 1985.