Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Price of Correlations in Stochastic Optimization

Shipra Agrawal, Yichuan Ding, Amin Saberi, Yinyu Ye,

Operations Research
Vol. 60, No. 1, January–February 2012, pp. 150–162
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.1110.1011

Price of Correlations in Stochastic Optimization

Shipra Agrawal
Department of Computer Science, Stanford University, Stanford, California 94305,
shipra@cs.stanford.edu

Yichuan Ding, Amin Saberi, Yinyu Ye
Department of Management Science and Engineering, Stanford University, Stanford, California 94305
{y7ding@stanford.edu, saberi@stanford.edu, yyye@stanford.edu}

When decisions are made in the presence of high-dimensional stochastic data, handling joint distribution of correlated random variables can present a formidable task, both in terms of sampling and estimation as well as algorithmic complexity. A common heuristic is to estimate only marginal distributions and substitute joint distribution by independent (product) distribution. In this paper, we study possible loss incurred on ignoring correlations through a distributionally robust stochastic programming model, and we quantify that loss as *price of correlations* (POC). Using techniques of cost sharing from game theory, we identify a wide class of problems for which POC has a small upper bound. To our interest, this class will include many stochastic convex programs, uncapacitated facility location, Steiner tree, and submodular functions, suggesting that the intuitive approach of assuming independent distribution acceptably approximates the robust model for these stochastic optimization problems. Additionally, we demonstrate hardness of bounding POC via examples of subadditive and supermodular functions that have large POC. We find that our results are also useful for solving many deterministic optimization problems like welfare maximization, k-dimensional matching, and transportation problems, under certain conditions.

Subject classifications: stochastic optimization; submodularity; cost sharing; correlation gap; joint distributions.
Area of review: Optimization.
History: Received February 2010; revisions received September 2010, February 2011; accepted June 2011.

1. Introduction

In many planning problems, it is crucial to consider correlations[1] among individual events. For example, an emergency service (medical services, fire rescue, etc.) planner needs to carefully locate emergency service stations and determine the number of emergency vehicles that need to be maintained in order to dispatch vehicles to the call points in time. If the planner assumes that emergency calls are rare and independent events, he simply needs to make sure that every potential call point is in service range of at least one station; however, there might exist a certain kind of dependence between those rare events, so that the planner cannot ignore the chance of simultaneous occurrences of those emergency events. The underlying correlations, possibly caused by some common trigger factors (e.g., weather, festivals), are often difficult to predict or analyze, which makes the planning problem complicated. Other examples include the portfolio selection problem, in which the risk-averse investor has to take into account the correlations among multiple risky assets as well as their individual performances, and the stochastic facility location problem, in which the supplier needs to consider the correlations between demands from different retailers.

As these examples illustrate, information about correlations can be crucial for operational planning, especially for large system planning. However, estimating the joint distribution in presence of correlations is usually difficult, and much harder than, for example, estimating marginal distributions. Reasons for this include the huge sample size required to characterize joint distribution accurately, and the practical difficulty of retrieving centralized information, e.g., the retailers may only be able to provide statistics about the demand for their own products. The question that arises is: how should one make decisions in presence of a large number of uncertain parameters when the correlations are not known?

Decision making under uncertainty is usually investigated in the context of stochastic programming (SP) (e.g., see Ruszczynski and Shapiro 2003 and references therein). In SP, the decision maker optimizes expected value of an objective function that involves random parameters. In general, a stochastic program is expressed as

$$\text{(SP)} \quad \underset{x\in C}{\text{minimize}}\ \mathbb{E}[h(\mathbf{x}, \boldsymbol{\xi})], \tag{1}$$

where $\mathbf{x}$ is the decision variable constrained to lie in set C, and the random variable $\boldsymbol{\xi} \in \Omega$ cannot be observed before the decision $\mathbf{x}$ is made. The *cost function* $h(\mathbf{x}, \boldsymbol{\xi})$ depends on both the decision $\mathbf{x} \in C$ and the random vector $\boldsymbol{\xi} \in \Omega$. If underlying distribution of random variables is unknown, then the decision maker needs to estimate it either via a parametric approach, which assumes that the distribution has a certain closed form and fits its parameters by empirical data, or via a nonparametric approach, e.g., the sample average approximation (SAA) method (e.g., Ahmed and Shapiro 2002, Ruszczynski and Shapiro 2003, Swamy and Shmoys 2005, Charikar et al. 2005), which optimizes the average objective value over a set of samples. However, these models are suitable only when one has access to a significant amount of reliable time-invariant statistical information. If the available samples are insufficient to fit the parameters of the distribution or to accurately estimate the expected value of the cost function, then SP fails to address the problem.

One alternate approach is to instead optimize the worst-case outcome, which is usually easier to characterize than estimating the joint distribution. That is,

$$\text{(RO)} \quad \underset{x\in C}{\text{minimize}}\ \underset{\xi\in\Omega}{\text{maximize}}\ h(\mathbf{x}, \boldsymbol{\xi}). \tag{2}$$

Such a method is termed robust optimization (RO) following the recent literature (e.g., Ben-Tal and Nemirovski 1998, 2000; Ben-Tal 2001). However, such a robust solution is often too pessimistic compared to SP (e.g., see Ben-Tal and Nemirovski 2000, Bertsimas and Sim 2004, Chen et al. 2007) because the worst-case scenario can be very unlikely. In particular, this model does not utilize any available or easy-to-estimate information about the distribution, such as marginal distributions of random variables.

An intermediate approach that may address the limitations of SP and RO is distributionally robust stochastic programming (DRSP). In this approach, one minimizes the expected cost over the worst joint distribution among all probability distributions consistent with the available information. That is,

$$\text{(DRSP)} \quad \underset{x\in C}{\text{minimize}}\ \underset{p\in\mathcal{P}}{\text{maximize}}\ \mathbb{E}_p[h(\mathbf{x}, \boldsymbol{\xi})], \tag{3}$$

where $\mathcal{P}$ is a collection of possible probability distributions on Ω, and for any $\mathbf{x}\in C$, $\mathbb{E}_p[h(\mathbf{x}, \boldsymbol{\xi})]$ denotes the expected value of $h(\mathbf{x}, \boldsymbol{\xi})$ over a distribution p on $\boldsymbol{\xi}$. The DRSP model can be interpreted as a two-person game. The decision maker chooses a decision $\mathbf{x}$ hoping to minimize the expected cost, whereas the nature adversarially chooses a distribution p from the collection $\mathcal{P}$ to maximize the expected cost of the decision.

Our model for characterizing price of correlations is based on this distributionally robust model of optimization. We consider the case when only the marginal distributions of involved random variables are known and no information about any correlations between them is available. Formally, we are given an n-dimensional random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ taking values in $\Omega = \Omega_1 \times \cdots \times \Omega_n$. For each i, we are given p_i, a probability measure over Ω_i. Then we consider the set of possible distributions $\mathcal{P}$ as the collection of all multivariate probability measures p over Ω such that each component ξ_i has a fixed marginal distribution p_i. That is,

$$\mathcal{P} = \left\{ p\colon \int_\Omega I(\xi_i = \theta_i)\, dp(\boldsymbol{\xi}) = p_i(\theta_i), \forall \theta_i \in \Omega_i, i = 1, \ldots, n \right\}, \tag{4}$$

where $I(\cdot)$ denotes the indicator function. In practice, for large domains the marginal distributions could be available explicitly in closed parametric form or as a black box sampling oracle.

The DRSP model, also known as *minimax stochastic programs*, was proposed by Scarf as early as 1958 (Scarf 1958) and Žáčková (1966). Since its introduction, it has gained extensive interest in context of making robust decisions when distribution information is limited (e.g., see Dupacová 1987, 2001; Shapiro and Kleywegt 2002 and references therein). The inner maximization problem has also been studied as moment problem (e.g., in Rogosinski 1958 and in Landau 1987). The applicability of various existing results depends on the assumed form of set $\mathcal{P}$ and the properties of objective function h. In Lagoa and Barmish (2002) and in Shapiro (2006), the authors consider a set containing unimodal distributions that satisfy some given support constraints. Under some conditions on $h(\mathbf{x}, \boldsymbol{\xi})$, they characterize the worst distribution as being the uniform distribution. The most popular type of distributional set $\mathcal{P}$ imposes linear constraints on moments of the distribution, as considered in Scarf (1958), in Dupacová (1987), in Prékopa (1995), in Bertsimas et al. (2000), and in Bertsimas and Popescu (2005). Scarf (1958) exploited the fact that for the newsvendor problem the worst distribution of demand with given mean and variance could be chosen to be one with all of its weight on two points. This idea was reused in Yue et al. (2006) and in Popescu (2007), although for more general forms of objective function. More recently, Delage and Ye (2010) showed that if the distributions have a fixed mean, bounded (by the positive semidefinite partial order) covariance matrix, and are supported on a closed convex set, then DRSP can be reformulated as a semidefinite program, solvable in polynomial time. Goh and Sim (2010) could develop tractable approximations to the DRSP problem by restricting to linear decision rules. Global optimization methods have been suggested to compute the worst-case distribution in Ermoliev et al. (1985) and in Gaivoronski (1991). Shapiro and Ahmed (2004) use duality to reduce any minimax program with given moment constraints to a minimization-type stochastic program and suggest using sample average approximation (SAA) method to solve this stochastic program. They do not explicitly derive polynomial-time bounds on the sample size required by the SAA method, which may depend on the objective function and constraints of the formulated stochastic program.

The DRSP models closest to ours are those considered in Klein Haneveld (1986) and Bertsimas et al. (2005). Klein Haneveld (1986) considers the problem of finding the worst-case distribution for a PERT-type project-planning problem under given marginal distributions. Due to special structure of this application, the dual problem can be reduced to a finite-dimensional convex program. Bertsimas et al. (2005) study the worst-case expectation of optimal value of generic combinatorial optimization problems with random objective coefficients, under limited marginal distribution information. They show that tight upper and lower bounds on this worst-case expected value can be computed in polynomial time, under certain conditions. They also analyze the asymptotic behavior of this expected value under knowledge of complete marginal distributions.

In general, computing the worst-case joint distribution or the corresponding expected value is difficult. The DRSP problem cannot be solved to optimality in polynomial time, especially when the support of the distributions is restricted. Bertsimas and Popescu (2005) show that the problem is NP-hard if moments of third or higher order are given, or if moments of second order are given and the domain of each random variable is restricted to nonnegative real numbers. For binary random variables, i.e. if the domain is restricted to $\{0, 1\}$, the problem of finding worst case distribution with given marginals is equivalent to finding worst-case distribution with given mean or first order moments. We show that this problem is NP-hard even when restricted to objective functions that are monotone and submodular in the random variable. In fact, we prove a stronger result that this problem is hard to approximate within any reasonable factor even with specific assumptions on the objective function.

Theorem 1. *Given a function of n binary random variables, the problem of computing its expected value under worst-case distribution with given mean is NP-hard, even when restricted to functions that are monotone and submodular in the random variables. The problem cannot be approximated within a factor better than $O(1/\sqrt{n})$ in polynomial time for some monotone and subadditive functions.*

The proof of NP-hardness is based on the observation that even though the problem of finding worst-case distribution with given mean can be formulated as a linear program (with exponential number of variables), the problem of computing a separating hyperplane for this linear program is at least as hard as MAX-CUT problem. The proof of this theorem is not central to the focus of this paper, and appears in Appendix A.

1.1. Price of Correlations

Considering the difficulty of computing the worst-case distribution, a natural question is how much risk it involves to simply ignore the correlations and, given marginal distributions $\{p_i\}$, minimize the expected cost under the independent or product distribution

$$\hat{p}(\boldsymbol{\xi}) = \prod_i p_i(\xi_i)$$

instead of the worst-case distribution. Or, in other words, how well does the stochastic optimization model with independent distribution approximate the robust DRSP model. The focus of this paper is to study the "price of correlations" incurred by this assumption of independence. Given a problem instance $(h, \Omega, \{p_i\})$, let $\mathbf{x}_I$ be the optimal decision assuming independent or product distribution and $\mathbf{x}_R$ be the optimal decision for the DRSP model. Then, *price of correlations* (POC) compares the performance of $\mathbf{x}_I$ to $\mathbf{x}_R$.

Definition 1. Given a problem instance $(h, \Omega, \{p_i\})$, where $H\colon C \times \Omega \to \mathbb{R}_+$. For any $\mathbf{x} \in C$, define

$$g(\mathbf{x}) = \sup_{p \in \mathcal{P}} \mathbb{E}_p[h(\mathbf{x}, \boldsymbol{\xi})]. \tag{5}$$

Let $\mathbf{x}_I = \arg\min_{x\in C} \mathbb{E}_{\hat{p}}[h(\mathbf{x}, \boldsymbol{\xi})]$, $\mathbf{x}_R = \arg\min_{x\in C} g(\mathbf{x})$. Then POC is defined as

$$\mathrm{POC} = \frac{g(\mathbf{x}_I)}{g(\mathbf{x}_R)}. \tag{6}$$

We redefine POC $= 1$, if $g(\mathbf{x}_I) = g(\mathbf{x}_R) = 0$, or if $g(\mathbf{x}_R) = \infty$. 1 defined.

Here $\mathbb{R}_+$ denotes the set of nonnegative real numbers. Note that POC $\geqslant 1$, and POC $= 1$ corresponds to the case where a stochastic program with product distribution yields the same result as the DRSP or minimax approach. A small upper bound on POC would indicate that the optimal solution obtained assuming independence is almost as robust as the most robust solution, and a stochastic optimization problem with independent (product) distribution is often relatively easier to solve, either by sampling or by other algorithmic techniques (e.g., see Kleinberg et al. 1997, Möhring et al. 1999). Thus, a small upper bound on POC would yield a simple approximation technique for the DRSP problem proven earlier to be difficult to solve.

However, our motivation for studying POC is more than getting an approximation algorithm for the DRSP problem. In many real data collection scenarios, practical constraints can make it very difficult (or costly) to learn the complete information about correlations in data. In lack of sufficient data, a widely adopted strategy in practice is to use independent distribution as a simple substitute for joint distribution. DRSP provides an alternate optimization-based approach to this problem promoting worst-case distribution under given marginals as a substitute to the joint distribution. However, a practitioner may doubt that the worst-case distribution is very pessimistic for the problem at hand, especially if she expects the random variables not to be very correlated. We believe that POC provides a conceptual understanding of the value of correlations in a decision problem involving uncertainties. It quantifies the gap between the two approaches of assuming no correlations and assuming worst-case correlations, providing a guiding principle for the decision maker. A small POC indicates that the solution obtained assuming independence is reasonably robust against correlations. On the other hand, if POC is very high, and from experience a practitioner expects the involved random variables to be not very correlated, she may decide that for this problem the DRSP approach is indeed very pessimistic, and it is essential to invest in learning the joint distribution. In this case, even getting poor estimates of correlations may be a better approach than ignoring the correlations or assuming the worst case. As we shall demonstrate later in the text, bounding POC is also useful as an algorithmic concept to obtain efficient approximation algorithms for many *deterministic optimization* problems.

This paper focuses on providing bounds on POC using structure of objective functions involved. As one would expect, POC can be arbitrarily large in general. In fact, as we demonstrate in this paper, except for the special case of submodular functions, almost all currently popular classifications of functions like subadditive functions, supermodular functions, convex functions, concave functions, contain some example with arbitrarily large POC, and hence are unsuitable for our characterization. Our main result is to characterize a new wide and useful class of cost functions that have small POC. We define this class using conceptsof cost sharing from game theory. This novel application of cost-sharing schemes may be interesting in its own respect.

1.2. Our Results

Below, we summarize our key results:

• *Upper bound on* POC *for submodular functions*: For functions $h(\mathbf{x}, \boldsymbol{\xi})$ that are monotone and submodular in $\boldsymbol{\xi}$ for all $\mathbf{x}$, we show an upper bound of $e/(e-1)$ on POC.[2] For binary random variables, this result can also be derived using a result by Calinescu et al. (2007). Our more general result, proven using different techniques, makes no assumption on the domains of random variables, and holds for submodular functions of general discrete variables and even continuous submodular functions.

• *Upper bounds on* POC *via cost-sharing:* For functions $h(\mathbf{x}, \boldsymbol{\xi})$ that are monotone in $\boldsymbol{\xi}$ and have a cross-monotonic β-budget balanced cost-sharing scheme in $\boldsymbol{\xi}$ for all $\mathbf{x}$, we show that POC is upper bounded by 2β. Using this result, we can obtain constant upper bounds on POC for many applications involving *nonsubmodular functions*, like POC $\leqslant 6$ for two-stage stochastic (metric) uncapacitated facility location, and POC $\leqslant 4$ for two-stage stochastic Steiner tree problem.

• *Lower bounds on* POC (*Hardness results*): We provide examples that prove POC can be as large as $\Omega(2^n)$ for functions $h(\mathbf{x}, \boldsymbol{\xi})$ that are monotone and supermodular in $\boldsymbol{\xi}$, and $\Omega(\sqrt{n}\log\log n/\log n)$ for monotone (fractionally) subadditive functions.[3] Further, we show examples with POC $\geqslant 3$ for stochastic uncapacitated facility location, POC $\geqslant 2$ for stochastic Steiner tree, and POC $\geqslant e/(e-1)$ for submodular functions, thus demonstrating the tightness of our upper bounds for these functions.

• *New results for deterministic optimization problems*: As a by-product, our result also provides new approximation algorithms for many deterministic optimization problems. For the *d-dimensional maximum matching problem*, transportation problems, and the *welfare maximization problem*, we obtain a $1-1/e$ approximation when the utility function (or weight matrix) in question satisfies monotonicity and submodularity, and a $1/2\beta$ approximation when it admits a β-budget balanced cross-monotonic cost-sharing scheme. For the special case of submodular welfare maximization problem, the result matches an earlier result proven using some other techniques in Vondrák (2008).

The rest of the paper is organized as follows. In §2, we give formal statements of our main results on upper bounding POC and provide rigorous proofs. In §3, we derive lower bounds on POC for many classes of cost functions. Finally, in §4, we discuss applications to stochastic and deterministic optimization problems. An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

2. Upper Bounds on POC

We assume that there is a complete ordering, denoted by $\leqslant$, on each Ω_i, $i = 1, \ldots, n$. We also use $\leqslant$ to denote the induced product order on Ω. That is, for any $\boldsymbol{\xi}, \boldsymbol{\theta} \in \Omega$, $\boldsymbol{\xi} \leqslant \boldsymbol{\theta}$ iff $\xi_i \leqslant \theta_i$, for all i. Also, for any $\boldsymbol{\xi}, \boldsymbol{\theta} \in \Omega$, let $\sup\{\boldsymbol{\xi}, \boldsymbol{\theta}\}$ and $\inf\{\boldsymbol{\xi}, \boldsymbol{\theta}\}$ denote supremum and infimum, respectively, of the two vectors taken with respect to the partial order $\leqslant$ on Ω. That is, $(\sup\{\boldsymbol{\xi}, \boldsymbol{\theta}\})_i = \max\{\xi_i, \theta_i\}$, $(\inf\{\boldsymbol{\xi}, \boldsymbol{\theta}\})_i = \min\{\xi_i, \theta_i\}$, for all i.

Our first result is to show that POC has a small upper bound for *monotone submodular functions*. Submodular functions are defined as follows:

Definition 2. A function $f(\boldsymbol{\xi})$: $\Omega \to \mathbb{R}$ is *submodular* iff

$$f(\sup\{\boldsymbol{\xi}, \boldsymbol{\theta}\}) + f(\inf\{\boldsymbol{\xi}, \boldsymbol{\theta}\}) \leqslant f(\boldsymbol{\xi}) + f(\boldsymbol{\theta}), \quad \forall\, \boldsymbol{\xi}, \boldsymbol{\theta} \in \Omega. \tag{7}$$

Property (7) also appears as the Monge property for matrices in the literature. An n-dimensional matrix M with k columns in each dimension is a Monge matrix iff the function f: $\{1, \ldots, k\}^n \to \mathbb{R}$ defined to take values in the corresponding cells of the matrix M is submodular. For functions of binary variables f: $2^V \to \mathbb{R}$, where $V = \{1, \ldots, n\}$, submodularity condition is equivalent to

$$f(S \cup i) - f(S) \geqslant f(T \cup i) - f(T), \quad \forall\, S \subseteq T \subseteq V, i \notin T.$$

Intuitively, submodularity of a utility function corresponds to decreasing marginal utilities—the marginal utility of an item decreases as the set of other items increases. The notion is very similar to the property of gross substitutes, which is, however, a stronger property in the sense that a utility function that satisfies the gross substitutes property is always submodular (see Gul and Stacchetti 1999, Kelso and Crawford 1982).

We will prove the following theorem:

THEOREM 2. *For any instance $(h, \Omega, \{p_i\})$, if for every feasible* $\mathbf{x}$ *the cost function $f(\boldsymbol{\xi}) = h(\mathbf{x}, \boldsymbol{\xi})$ is nonnegative, monotone, and submodular in $\boldsymbol{\xi}$ and has a finite expected value on product distribution $\hat{p}(\boldsymbol{\xi}) = \prod_i p_i(\xi_i)$, then* POC $\leqslant e/(e-1)$.

For the special case of $\Omega = \{0, 1\}^n$, the above result for submodular functions can also be derived from a result in Calinescu et al. (2007). However, our more general result makes no assumption on Ω_is and holds even for the continuous case of $\Omega = \mathbb{R}^n$ and the discrete case of $\Omega = \mathbb{N}^n$. To our understanding, the technique in Calinescu et al. (2007) cannot be easily applied to prove these general cases. For the continuous case, we have the following corollary:

COROLLARY 1. *For instances $(h, \Omega, \{p_i\})$ such that for all feasible* $\mathbf{x}$, *$f(\boldsymbol{\xi}) = h(\mathbf{x}, \boldsymbol{\xi})$ is nondecreasing, continuous twice differentiable, and $\nabla^2_{ij} f(\boldsymbol{\xi}) \leqslant 0$, for all $i \neq j$, then* POC $\leqslant e/(e-1)$.

An example of a convex function that satisfies the above property is L_q norm $f(\boldsymbol{\xi}) = \|\boldsymbol{\xi}\|_q$, when $q \geqslant 1$ and $\boldsymbol{\xi} \in \mathbb{R}^n_+$. Unfortunately, the condition of submodularity can be quite restrictive in practice. Many popular applications such as stochastic facility location and stochastic Steiner tree involve cost functions that are subadditive but not submodular in the random variables. Also, as we demonstrate in §3, even for the class of monotone subadditive or fractionally subadditive functions, POC can be arbitrarily large for large n. Therefore, it is apparent that we need a different characterization of functions for bounding POC. We derive our characterization using the concept of *cost-sharing schemes* from game theory. A cost-sharing scheme refers to a scheme for dividing the cost $f(\boldsymbol{\xi})$ among the n components.

DEFINITION 3. Given the total cost $f(\boldsymbol{\xi})$ of servicing $\boldsymbol{\xi} \in \Omega$, a cost allocation is a function $\Psi\colon [n] \to \mathbb{R}_+$, which for every $i = 1, \ldots, n$ specifies the share $\Psi(i)$ of i in the total cost $f(\boldsymbol{\xi})$. A *cost-sharing scheme* $\chi\colon [n] \times \boldsymbol{\xi} \to \mathbb{R}_+$ is a collection of cost allocations for all $\boldsymbol{\xi} \in \Omega$.

We will use a more compact notation $\chi_i(\boldsymbol{\xi})$ to denote the cost share $\chi(i, \boldsymbol{\xi})$ in the remaining text.

Ideally, we want the cost-sharing scheme (and corresponding cost allocations) to be *budget balanced*, that is, $\sum_{i=1}^n \chi_i(\boldsymbol{\xi}) = f(\boldsymbol{\xi})$. However, it is not always possible to achieve budget balance in combination with other properties. Therefore, a relaxed notion of the β-budget balanced cost-sharing scheme is often considered.

DEFINITION 4. A cost-sharing scheme χ is *β-budget balanced* if

$$\frac{f(\boldsymbol{\xi})}{\beta} \leqslant \sum_{i=1}^n \chi_i(\boldsymbol{\xi}) \leqslant f(\boldsymbol{\xi}), \ \ \forall\, \boldsymbol{\xi} \in \Omega.$$

For our characterization of functions with small POC, we are interested in cost-sharing schemes with additional properties of *cross monotonicity*. Cross monotonicity (or population monotonicity) was studied by Moulin (1999) and Moulin and Shenker (2001) in order to design group-strategy-proof mechanisms and has recently received considerable attention in the computer science literature (see, for example, Pál and Tardo 2003, Mahdian and Pál 2003, Könemann et al. 2005, Immorlica et al. 2008, Nisan et al. 2007, and references therein). This property captures the notion that an agent should not be penalized as the demands of other agents grow.

DEFINITION 5. A cost-sharing scheme χ is *cross-monotonic* if for all i, $\xi_i \in \Omega_i$, $\boldsymbol{\xi}_{-i}$, $\boldsymbol{\theta}_{-i} \in \Pi_{j \neq i} \Omega_j$,

$$\chi_i(\xi_i, \boldsymbol{\xi}_{-i}) \geqslant \chi_i(\xi_i, \boldsymbol{\theta}_{-i}), \quad \text{if } \boldsymbol{\xi}_{-i} \leqslant \boldsymbol{\theta}_{-i}.$$

Our main result is the following theorem, which uses the concept of a β-budget-balanced cross-monotonic cost-sharing scheme.

DEFINITION 6. We call a cost-sharing scheme χ a *β-cost-sharing scheme* if it is cross-monotonic and β-budget balanced.

THEOREM 3. *For any instance $(h, \Omega, \{p_i\})$, if for every feasible* $\mathbf{x}$ *the cost function $f(\boldsymbol{\xi}) = h(\mathbf{x}, \boldsymbol{\xi})$ is nonnegative, monotone, has a β-cost-sharing scheme in $\boldsymbol{\xi}$ for some $\beta < \infty$, and has a finite expected value on product distribution $\hat{p}(\boldsymbol{\xi}) = \prod_i p_i(\xi_i)$, then* POC $\leqslant 2\beta$.

Note that the above theorem requires that a β-cost-sharing scheme exists for every feasible $\mathbf{x}$, where β does not depend on $\mathbf{x}$.

The above result connecting cost-sharing schemes to POC is particularly interesting because it allows us to use the already existing nontrivial work on designing β-cost-sharing schemes for complex cost functions such as the facility location cost function, Steiner forest cost function, etc. (see Pál and Tardo 2003, Könemann et al. 2005, Leonardi and Schaefer 2004). We will derive the following corollaries.

COROLLARY 2. *For the two-stage stochastic uncapacitated facility location* (*metric*) *problem,* POC $\leqslant 6$.

COROLLARY 3. *For the two-stage stochastic Steiner tree problem,* POC $\leqslant 4$.

In the following subsections, we provide a rigorous proof of the upper bounds on POC given by Theorems 2 and 3. The proof will proceed as follows. First, we show that an upper bound on a simpler quantity "correlation gap" implies the same upper bound on POC. Then, we bound the correlation gap (by $e/(e-1)$ and 2β, respectively) to complete the proof.

In the main text, we provide the proofs for binary domains and finite domains only. The proofs for countably infinite domains and uncountable case of $\Omega = \mathbb{R}^n$ appear in Appendices D and E, respectively.

2.1. Bounding POC via Correlation Gap

We bound POC by bounding a simpler quantity of the *correlation gap*. The correlation gap of a function f will compare its expected value on independent distribution to its expected value on the worst-case distribution.

Definition 7. Given the function $f\colon \prod_i \Omega_i \to \mathbb{R}_+$, and probability measures p_i on Ω_i for each i, we define *correlation gap* as the ratio:

$$\kappa = \sup_{p\in\mathcal{P}} \frac{\mathbb{E}_p[f(\boldsymbol{\xi})]}{\mathbb{E}_{\hat{p}}[f(\boldsymbol{\xi})]},$$

where $\hat{p}$ denotes the product distribution, $\hat{p}(\boldsymbol{\xi}) = \prod_i p_i(\xi_i), \forall\, \boldsymbol{\xi}$.

It is easy to show that a uniform bound on the correlation gap for all $\mathbf{x}$ will bound POC. Let $\kappa(\mathbf{x})$ denote the correlation gap of function $h(\mathbf{x}, \boldsymbol{\xi})$ at $\mathbf{x}$, i.e.,

$$\kappa(\mathbf{x}) = \sup_{p\in\mathcal{P}} \frac{\mathbb{E}_p[h(\mathbf{x}, \boldsymbol{\xi})]}{\mathbb{E}_{\hat{p}}[h(\mathbf{x}, \boldsymbol{\xi})]},$$

Let for all feasible $\mathbf{x}$,

$$\kappa(\mathbf{x}) \leqslant \bar{\kappa}.$$

Then,

$$g(\mathbf{x}_I) = \sup_{p\in\mathcal{P}} \; \mathbb{E}_p[h(\mathbf{x}_I, \boldsymbol{\xi})],$$

$$g(\mathbf{x}_R) = \sup_{p\in\mathcal{P}} \; \mathbb{E}_p[h(\mathbf{x}_R, \boldsymbol{\xi})] \geqslant \mathbb{E}_{\hat{p}}[h(\mathbf{x}_R, \boldsymbol{\xi})] \geqslant \mathbb{E}_{\hat{p}}[h(\mathbf{x}_I, \boldsymbol{\xi})],$$

and hence,

$$\text{POC} = \frac{g(\mathbf{x}_I)}{g(\mathbf{x}_R)} \leqslant \sup_{p\in\mathcal{P}} \frac{\mathbb{E}_p[h(\mathbf{x}_I, \boldsymbol{\xi})]}{\mathbb{E}_{\hat{p}}[h(\mathbf{x}_I, \boldsymbol{\xi})]} = \kappa(\mathbf{x}_I) \leqslant \bar{\kappa}.$$

Therefore, we can concentrate on deriving a uniform upper bound on correlation gap $\kappa(\mathbf{x})$ for all fixed $\mathbf{x}$. In particular, observe that the proof of Theorems 2 and 3 will follow directly from the following theorem.

Theorem 4. *Consider any instance $(f, \Omega, \{p_i\})$, where $f\colon \Omega \to \mathbb{R}_+$, $\Omega = \prod_i \Omega_i$, and each p_i is a probability measure on Ω_i. Then, the correlation gap of instance $(f, \Omega, \{p_i\})$ is bounded by*:

- *$e/(e-1)$, if $f(\boldsymbol{\xi})$ is monotone and submodular.*
- *2β, if $f(\boldsymbol{\xi})$ is monotone and has a β-cost-sharing scheme.*

We also make a technical assumption that f has finite expected value on the product distribution.

In the next two subsections, we will prove the above theorem for the case of binary domains and finite domains, respectively. The proofs for countably infinite domains and for the uncountable case of $\Omega = \mathbb{R}^n$ appear in Appendices D and E, respectively.

2.2. Proof for Binary Random Variables

For submodular functions on binary variables, a bound of $e/(e-1)$ on the correlation gap was proven earlier in Agrawal et al. (2010) and can also be derived from a result by Calinescu et al. (2007).

Theorem (Calinescu et al. 2007, Agrawal et al. 2010): *For any instance $(f, \Omega, \{p_i\})$, where $\Omega = \{0,1\}^n$ and f is a nonnegative, monotone, and submodular function, correlation gap is bounded by $e/(e-1)$.*

In this section, we present the proof of the correlation gap bound of 2β for functions on binary random variables that admit a β-cost-sharing scheme.

Lemma 1. *For any instance $(f, \Omega, \{p_i\})$, where $\Omega = \{0,1\}^n$ and f is a nonnegative monotone function with a β-cost-sharing scheme, the correlation gap is bounded by 2β.*

Proof. Because we are considering the binary random vector $\boldsymbol{\xi}$, we can equivalently represent it by the corresponding random subset S of set $V = \{1, \dots, n\}$. That is, $\Omega = 2^V$, the set of all subsets of set V. Also, for each i, p_i is the probability of i to appear in random set S.

First, we consider a simplified problem. We assume that (a) all p_is are equal to $1/K$ for some finite integer $K > 0$, and (b) the worst-case distribution is a "K-partition-type" distribution. That is, the worst-case distribution has support on K disjoint sets $\{A_1, \dots, A_K\}$ that form a partition of V, and each A_k occurs with probability $1/K$. Let us call such instances $(f, 2^V, \{1/K\})$ "nice" instances. Here, we show that the correlation gap is bounded by 2β for all "nice" instances. In Lemma 2, we show that it is sufficient to consider only the "nice" instances and thus complete the proof.

For any set $S \subseteq V$, denote $S_{\cap k} = S \cap A_k$ and $S_{-k} = S \backslash A_k$ for $k = 1, \dots, K$. Let χ be the β-cost-sharing scheme for function f, as per the assumptions of the lemma. Also, for any subset T of S, denote $\chi(T, S) := \sum_{i\in T} \chi_i(S)$. Then, by the budget balance property of χ, expected value under the independent distribution

$$\mathbb{E}_S[f(S)] \geqslant \mathbb{E}_S\left[\sum_{k=1}^{K} \chi(S_{\cap k}, S)\right]. \tag{8}$$

Note that under independent distribution, the marginal probability that an element $i \in A_k$ appears in random set $S_{\cap k}$ is $1/K$. Using this observation along with the cross monotonicity of cost-sharing scheme χ and properties of independent distribution, we can derive that for any k,

$$\begin{aligned}\mathbb{E}_S[\chi(S_{\cap k}, S)] &\geqslant \mathbb{E}_S[\chi(S_{\cap k}, S \cup A_k)]\\ &= \mathbb{E}_S\left[\sum_{i\in A_k} I(i \in S_{\cap k})\chi(i, S_{-k} \cup A_k)\right]\\ &= \mathbb{E}_{S_{-k}}\left[\sum_{i\in A_k} \mathbb{E}_{S_{\cap k}}[I(i \in S_{\cap k})\chi(i, S_{-k} \cup A_k) \mid S_{-k}]\right]\end{aligned}$$

$$= \frac{1}{K}\ \mathbb{E}\bigg[\sum_{i\in A_k}\chi(i, S\cup A_k)\bigg]$$
$$= \frac{1}{K}\ \mathbb{E}[\chi(A_k, S\cup A_k)]. \tag{9}$$

Here, $I(\cdot)$ denotes the indicator function. Apply the above inequality to the $\gamma = 1/((2-1)/K)$ fraction of each term $\chi(S_{\cap k}, S)$ in (8) to obtain

$$\mathbb{E}_S[f(S)]$$
$$\geqslant \mathbb{E}_S\bigg[\sum_{k=1}^{K}(1-\gamma)\chi(S_{\cap k}, S) + \gamma\frac{1}{K}\chi(A_k, S\cup A_k)\bigg]$$
$$= \mathbb{E}_S\bigg[\sum_{k=1}^{K}\bigg(\frac{1-\gamma}{K-1}\bigg)\bigg(\sum_{j\neq k}\chi(S_{\cap j}, S)\bigg) + \gamma\frac{1}{K}\chi(A_k, S\cup A_k)\bigg]$$
$$= \frac{1}{(2K-1)}\mathbb{E}_S\bigg[\sum_{k=1}^{K}\bigg(\sum_{j\neq k}\chi(S_{\cap j}, S)\bigg) + \chi(A_k, S\cup A_k)\bigg]$$

(using cross monotonicity of χ)

$$\geqslant \frac{1}{(2K-1)}\mathbb{E}_S\bigg[\sum_{k=1}^{K}\bigg(\sum_{j\neq k}\chi(S_{\cap j}, S\cup A_k)\bigg) + \chi(A_k, S\cup A_k)\bigg]$$

(using β-budget balance)

$$\geqslant \frac{1}{(2K-1)\beta}\mathbb{E}_S\bigg[\sum_{k=1}^{K}f(S\cup A_k)\bigg]$$

(using monotonicity of f)

$$\geqslant \frac{1}{(2-1/K)\beta}\bigg(\frac{1}{K}\sum_{k=1}^{K}f(A_k)\bigg).$$

Under the assumption of the "nice" instance, the expected value on worst-case distribution is given by $(1/K)\sum_{k=1}^{K}f(A_k)$. Therefore, the correlation gap is bounded by 2β for nice instances. The next lemma shows that it is sufficient to consider only the nice instances, and completes the proof. ☐

LEMMA 2. *For every instance $(f, 2^V, \{p_i\})$ such that $f(S)$ is nondecreasing and has a β-cost-sharing scheme (submodular), there exists a nice instance $(f', 2^{V'}, \{1/K\})$ for some integer $K > 0$ such that f' is nondecreasing and has a β-cost-sharing scheme (submodular) and the correlation gap of instance $(f', 2^{V'}, \{1/K\})$ is at least as large as $(f, 2^V, \{p_i\})$.*

We make a technical assumption that all p_is are rational and nonzero.

PROOF. We use the following split operation.

Split: Given a problem instance $(f, 2^V, \{p_i\})$ and integers $\{m_i \geqslant 1, i\in V\}$, the split operation defines a new instance $(f', 2^{V'}, \{p'_r\})$ as follows: split each item $i\in V$ into m_i copies $C_1^i, C_2^i, \ldots, C_{m_i}^i$, and assign a marginal probability of $p'_{C_j^i} = p_i/m_i$ to each copy. Let V' denote the new ground set of size $\sum_i m_i$ that contains all the duplicates. Define the new cost function $f'\colon 2^{V'}\to\mathbb{R}$ as:

$$f'(S') = f(\Pi(S')), \quad \text{for all } S'\subseteq V', \tag{10}$$

where $\Pi(S')\subseteq V$ is the original subset of elements whose duplicates appear in S', i.e., $\Pi(S') = \{i\in V \mid C_j^i\in S' \text{ for some } j\in\{1, 2, \ldots, m_i\}\}$.

We claim that splitting an instance

- does not change the expected value under the worst-case distribution.
- can only decrease the expected value under independent distribution.
- preserves the monotonicity and β-cost-sharing/submodularity of the function.

The proofs of these claims appear in Appendix B.1.

Thus, the new instance generated on splitting has a correlation gap at least as large as the original instance. The remaining proof tries to use split operation to reduce any given instance to a "nice" instance. Let p be the worst-case distribution for instance $(f, 2^V, \{p_i\})$. The set of distributions $\mathcal{P}$ is defined by a rational linear program with $|V|$ constraints, so $\mathcal{P}$ is compact, and p exists. Suppose that p is not a partition-type distribution. Then, split any element i that appears in two different sets in the support of the worst-case distribution. Simultaneously, split the distribution by assigning probability $p'(S') = p(\Pi(S'))$ to each set S' that contains exactly one copy of i, and probability 0 to all other sets. Because each set in the support of new distribution contains exactly one copy of every element i, by definition of function f', the expected function value of f' on p' is the same as that of f on p. By properties of split operation, the worst-case expected values for the two instances (before and after splitting) must be the same, so the distribution p' forms a worst-case distribution for the new instance. Repeat the splitting until the distribution becomes a partition-type distribution. Then, we further split each element (and simultaneously the distribution) until the marginal probability of each new element is $1/K$ for some large enough integer K. Note that such a finite K always exists, assuming that p_is are rational and nonzero. ☐

REMARK 1. In the above proof, we assumed for simplicity that all p_is are rational and nonzero. In Appendix B.2, we show that the results hold even if the p_is are not rational. The assumption of p_i being nonzero is without loss of generality, because we could simply remove an item in V from the problem if its probability to appear in a random set is 0.

2.3. Proof for Finite Domains

LEMMA 3. *Consider any instance $(f, \Omega, \{p_i\})$, $f\colon \Omega\to\mathbb{R}_+$, where $\Omega = \prod_i\Omega_i$ and for all i, $|\Omega_i| < \infty$. Then, the correlation gap of instance $(f, \Omega, \{p_i\})$ is bounded by*

- *$e/(e-1)$, if $f(\boldsymbol{\xi})$ is monotone and submodular.*
- *2β, if $f(\boldsymbol{\xi})$ is monotone and has a β-cost-sharing scheme.*

PROOF. We prove this lemma by first reducing the problem to a problem with binary random variables only. Then, the results in the previous subsection on bounding the correlation gap for instances with binary random variables will complete the proof.

Without loss of generality, let $\Omega_i = \{0, \dots, K_i\}$, where $K_i < \infty$. Given instance $(f, \Omega, \{p_i\})$, we create a new instance $(f', \Omega', \{p'_{ij}\})$ with *binary variables only* as follows. For every variable ξ_i in the original instance, we create K_i new binary variables $\{\xi'_{ij}\}_{j=1}^{K_i}$ in the new instance, and set the marginal probability of ξ'_{ij} to take value 1 as: $p'_{ij} = p_i(j)$. Also, given $f\colon \prod_i \Omega_i \to \mathbb{R}_+$, we define new function $f'\colon \prod_{i=1}^n \{0,1\}^{K_i} \to \mathbb{R}$ as,

$$f'(\boldsymbol{\xi}') := f(\Theta(\boldsymbol{\xi}')),$$

where for $i = 1, \dots, n$, the value of $\Theta(\boldsymbol{\xi}')_i$ is given by the largest j for which ξ'_{ij} is nonzero. That is, for all i,

$$\Theta(\boldsymbol{\xi}')_i = \max\{j\colon \xi'_{ij} = 1\},$$

assuming that $\max\{j\colon \xi'_{ij} = 1\}$ returns 0 if none of ξ'_{ij}, $j = 1, \dots, K_i$ is 1.

Next, we compare the problem instance $(f, \prod_i \Omega_i, \{p_i\})$ to the reduced 0-1 instance $(f', \prod_{i=1}^n \{0,1\}^{K_i}, \{p'_{ij}\})$. We show that this reduction has the properties of preserving monotonicity, submodularity, β-cost-sharing, and the expected value over worst-case distribution. Also, the expected value over independent distribution can only decrease. The proofs of these claims use very similar ideas, as in the proof of Lemma 2. See Appendix C for details.

Thus, we get a new instance $(f', \Omega', \{p'_r\})$ where $\Omega' = \{0,1\}^{n'}$, f' is monotone, and has a β-cost-sharing scheme (or is submodular). Also, the correlation gap of the new instance bounds the correlation gap of the original instance. Because the correlation gap of the new 0-1 instance is bounded as required by the results in the previous subsection, this completes the proof of the lemma. □

The reader may refer to Appendices D and E for the proofs for countably infinite domains and the uncountable case of $\Omega = \mathbb{R}^n$, respectively.

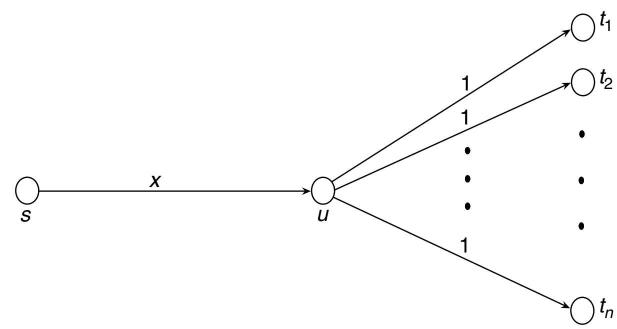
3. Lower Bounds

In this section, we demonstrate the difficulty of bounding POC by showing examples that have provably large POC. We construct examples of functions that are monotone supermodular and monotone subadditive, respectively, in the random variables, but could still have arbitrarily large POC if n is large. This illustrates the importance of characterization using techniques like cost sharing in order to get upper bounds, as in Theorem 3, that do not depend on n.

We also show tightness of our upper bound for submodular functions via a counterexample. In the next section, we will prove the tightness of our bounds for specific applications like facility location and Steiner forest by showing close lower bounds.

Lemma 4. *There exists an instance $(h, 2^V, \{p_i\})$ with function $h(x, S)$ that is nondecreasing and supermodular in S, and* POC $\geqslant \Omega(2^n)$. *Here, $n = |V|$.*

Figure 1. An example with exponential correlation gap.



Proof. Consider a two-stage minimum cost flow problem as in Figure 1. There is a single source s, and n sinks $t_1, t_2, \dots, t_n$. Each sink t_i has a probability $p_i = \frac{1}{2}$ to demand a flow, and then a unit flow has to be sent from s to t_i. Each edge (u, t_i) has a fixed capacity 1, but the capacity of edge (s, u) needs to be purchased. The cost of capacity x on edge (s, u) is $c^I(x)$ in the first stage, and $c^{II}(x)$ in the second stage after the set of demands is revealed, defined as

$$c^I(x) = \begin{cases} x, & x \leqslant n-1, \\ n+1, & x = n, \end{cases} \qquad c^{II}(x) = 2^n x.$$

Given a first-stage decision x, the total cost of edges that need to be bought in order to serve a set S of requests is given by: $h(x, S) = c^I(x) + c^{II}((|S| - x)^+) = c^I(x) + 2^n(|S| - x)^+$. It is easy to check that $h(x, S)$ is nondecreasing and supermodular in S for any given x, i.e., $h(x, S \cup i) - h(x, S) \geqslant h(x, T \cup i) - h(x, T)$ for any $S \supseteq T$. The objective is to minimize the total expected cost $\mathbb{E}[h(x, S)]$. If the decision maker assumes independent demands from the sinks, then $x_I = n - 1$ minimizes the expected cost, and the expected cost is n; however, for the worst-case distribution the expected cost of this decision will be $g(x_I) = 2^{n-1} + n - 1$ (when $\Pr(\{1, \dots, n\}) = \Pr(\varnothing) = 1/2$ and all other scenarios have zero probability).

Hence, the correlation gap at x_I is exponentially high. A risk-averse strategy is to instead use the robust solution $x_R = n$, which leads to a cost $g(x_R) = n + 1$. Thus, POC $= g(x_I)/g(x_R) = \Omega(2^n)$. □

We may remark here that although independent distribution does not appear to be a good substitute of worst-case distribution for supermodular functions, the worst-case joint distribution actually has a nice closed form in this case, and the DRSP model is directly solvable in polynomial time. Refer to Agrawal et al. (2010) for details.

Lemma 5. *There exists an instance $(h, 2^V, \{p_i\})$ with function $h(\mathbf{x}, S)$ that is nondecreasing and (fractionally) subadditive*[4] *in S, and* POC $\geqslant \Omega(\sqrt{n}\log\log(n)/\log(n))$. *Here $n = |V|$.*

Proof. Consider a set cover problem with elements $V = \{1, \ldots, n\}$. Each item $i \in V$ has a marginal probability of $1/K$ to appear in the random set S where $K = \sqrt{n}$. The covering sets are defined as follows. Consider a partition of V into K sets $A_1, \ldots, A_K$, each containing K elements. The covering sets are all the sets in the Cartesian product $A_1 \times \cdots \times A_K$. Each set has a unit cost. Then, the cost of covering a set S is given by the subadditive (in fact, fractionally subadditive) function

$$c(S) = \max_{k=1,\ldots,K} |S \cap A_k| \quad \forall S \subseteq V.$$

The worst-case distribution with marginal probabilities $p_i = 1/K$ is one where probabilities $\Pr(S) = 1/K$ for $S = A_k$, $k = 1, 2, \ldots, K$, and $\Pr(S) = 0$ otherwise. The expected value of $c(S)$ under this distribution is $K = \sqrt{n}$. For independent distribution, $c(S) = \max_{k=1,\ldots,K} \zeta_k$, where $\zeta_k = |S \cap A_k|$ are independent $(K, 1/K)$-binomially distributed random variables. Using some known statistical results in Kimber (1983), it can be shown that as $K = \sqrt{n}$ approaches ∞, $\mathbb{E}[c(S)]$ approaches $\Theta(\log n/\log\log n)$. See Appendix F for details.

Therefore, the correlation gap of cost function $c(S)$ is bounded below by $\Omega(\sqrt{n}\log\log n/\log n)$. To get the corresponding lower bound on POC, consider a two-stage stochastic set cover problem where sets cost little more than $\Omega(\log n/\sqrt{n}\log\log n)$ in the first stage, and 1 in the second stage. Then, on assuming independent distribution, the optimal decision is to buy no or very few sets in the first stage. However, for the worst-case distribution, the expected second-stage cost of this decision will be $\sqrt{n}$. On the other hand, a robust solution considering the worst-case distribution is to cover all the elements in the first stage costing $O(\log(n)/\log\log(n))$, and nothing in the second stage. Thus, $\mathrm{POC} \geqslant \Omega(\sqrt{n}\log\log n/\log n)$. □

Lemma 6. *There exists an instance $(h, 2^V, \{p_i\})$ with function $h(x, S)$ that is nondecreasing and submodular in S, and* $\mathrm{POC} = e/(e-1)$. *Thus, the upper bound given by Theorem 2 is tight.*

Proof. Let $V := \{1, 2, \ldots, n\}$, define the submodular function $f(S) = 1$ if $S \neq \varnothing$, and $h(\varnothing) = 0$. Let each item $i \in V$ have marginal probability $p_i = 1/n$ to appear in random set S. The worst-case distribution that maximizes $\mathbb{E}[f(S)]$ is the one with $\Pr(\{i\}) = 1/n$ for all $i \in V$, with expected value 1. On the independent distribution with the same marginals, f has an expected cost $1 - (1 - 1/n)^n \to 1 - 1/e$ as $n \to \infty$. Thus, the correlation gap is $e/(e-1)$.

To obtain the corresponding lower bound on POC, we can extend this example to a stochastic decision problem with two possible decisions x_1, x_2 as follows. Define $h(x_1, S) = f(S), \forall S$, and $h(x_2, S) = 1 - 1/e + \epsilon, \forall S$ for some arbitrarily small $\epsilon > 0$. Then, on assuming independent distribution, x_1 seems to be the optimal decision; however, it will have expected cost 1 on the worst-case distribution. On the other hand, decision x_2 would cost $1 - 1/e + \epsilon$ in the worst case, giving $\mathrm{POC} = e/(e-1) - \epsilon$. □

4. Applications

In this section, we discuss applications of our results to a variety of stochastic and deterministic optimization problems. The first two applications—stochastic facility location and stochastic Steiner tree—apply Theorem 3 to functions on binary random variables ($\Omega = \{0, 1\}^n$) with β-cost-sharing schemes. Next, for stochastic bottleneck matching, we apply Theorem 2 to continuous submodular functions ($\Omega = \mathbb{R}^n$). Finally, we discuss welfare maximization and d-dimensional matching as applications of our results to deterministic optimization problems. Welfare maximization reduces to a problem of bounding correlation gap for functions on binary random variables, whereas d-dimensional matching involves a function on *general discrete domain* ($\Omega = \{1, \ldots, T\}^n$).

4.1. Stochastic Uncapacitated Facility Location (SUFL)

In the two-stage stochastic facility location problem, any facility $j \in F$ can be bought at a low cost w_j^I in the first stage, and higher-cost $w_j^{II} > w_j^I$ in the second stage, that is, after the random set $S \subseteq V$ of cities to be served is revealed. The decision maker's problem is to decide $\mathbf{x} \in \{0, 1\}^{|F|}$, the facilities to be built in the first stage so that the total expected cost $\mathbb{E}[h(\mathbf{x}, S)]$ of facility location is minimized (refer to Swamy and Shmoys 2005 for further details on the problem definition).

Proposition 1. *For the two-stage stochastic metric uncapacitated facility location* (SUFL) *problem,* $\mathrm{POC} \leqslant 6$. *Also, there exists an instance of this problem with* $\mathrm{POC} \geqslant 3$.

Proof. Given a first-stage decision $\mathbf{x}$, the cost function $h(\mathbf{x}, S) = w^I \cdot \mathbf{x} + c(\mathbf{x}, S)$, where $c(\mathbf{x}, S)$ is the cost of deterministic UFL for set $S \subseteq V$ of customers and set F of facilities such that the facilities $\mathbf{x}$ already bought in the first stage are available freely at no cost, whereas any other facility j costs w_j^{II}. For deterministic metric UFL there exists a cross-monotonic, 3-budget balanced, cost-sharing scheme (Pál and Tardo 2003). Therefore, using Theorem 3, we know that the POC for stochastic metric UFL has an upper bound of $2\beta = 6$. The next example shows an instance of this problem with $\mathrm{POC} \geqslant 3$.

Consider the following two-stage stochastic metric facility location instance with n cities. There is a partition of the n cities into $\sqrt{n}$ disjoint sets $A_1, \ldots, A_{\sqrt{n}}$ containing $\sqrt{n}$ cities each. Corresponding to each set B of cities in the Cartesian product $\mathcal{B} = A_1 \times \cdots \times A_{\sqrt{n}}$, there is a facility F_B with connection cost 1 to each city in B. The remaining connection costs are defined by extending the metric, that is, the cost of connecting any city i to facility F_B such that $i \notin B$ is 3. Assume that each city has a marginal probability of $1/\sqrt{n}$ to appear in the random demand set S. Each facility costs $w^I = 3\log n/\sqrt{n}$ in the first stage and $w^{II} = 3$ in the second stage.

First, consider the independent distribution case. Regardless of how many facilities are opened in the first stage, the expected cost in the second stage will be no more than $3\mathbb{E}[\max_k |A_k \cap S|] + \sqrt{n}$. $\mathbb{E}[\max_k |A_k \cap S|]$ asymptotically reaches $O(\log n/(\log\log n)) = o(\log(n))$ for large n. Therefore, for any $\epsilon > 0$, for large enough n, $\mathbb{E}[\max_k |A_k \cap S|] < \epsilon \log(n)$. As a result, if the decision maker assumes independent distribution, she will never buy more than $\sqrt{n}\epsilon$ facilities in the first stage, which would cost her $3\epsilon\log(n)$. However, if the distribution turns out to be of form $\Pr(A_k) = 1/\sqrt{n}$, $k = 1, \ldots, \sqrt{n}$, then such a strategy induces an expected cost $g(\mathbf{x}_I) \geqslant 3(1-\epsilon)\sqrt{n} + 3\epsilon\log(n)$. A robust solution is to instead build $\sqrt{n}$ facilities in the first stage, corresponding to a collection of $\sqrt{n}$ disjoint sets in the collection $\mathscr{B}$. These facilities will cover every city with a connection cost of 1. Thus, the worst-case expected cost for robust solution $g(\mathbf{x}_R) \leqslant 3\log n + \sqrt{n}$. This shows that $g(\mathbf{x}_I) \geqslant (3-\epsilon)g(\mathbf{x}_R)$ for any $\epsilon > 0$. □

The above proposition reduces the distributionally robust facility location problem to the well-studied (e.g., see Swamy and Shmoys 2005) stochastic UFL problem under known (independent Bernoulli) distribution at the expense of a 6-approximation factor.

4.2. Stochastic Steiner Tree (SST)

In the two-stage stochastic Steiner tree problem, we are given a graph $G = (V, E)$. An edge $e \in E$ can be bought at cost w_e^I in the first stage. A random set $S \subseteq V$ of terminal nodes to be connected are revealed in the second stage. More edges may be bought at a higher cost w_e^{II}, $e \in E$ in the second stage after observing the actual set of terminals. Here, decision variable $\mathbf{x}$ is the edges to be bought in the first stage, and cost function $h(\mathbf{x}, S) = w^I \cdot \mathbf{x} + c(\mathbf{x}, S)$, where $c(\mathbf{x}, S)$ is the deterministic Steiner tree cost for connecting nodes in set S, given that the edges in $\mathbf{x}$ are already bought.

Proposition 2. *For the two-stage stochastic Steiner tree* (SST) *problem,* POC $\leqslant 4$. *Also, there exists an instance of this problem with* POC $\geqslant 2$.

Proof. Because a 2-budget balanced cross-monotonic cost-sharing scheme is known for the deterministic Steiner tree (see Könemann et al. 2005), we can use Theorem 3 to conclude that for this problem POC $\leqslant 2\beta = 4$. The following example shows an instance of the two-stage stochastic Steiner tree with POC $\geqslant 2$. The construction of this example is very similar to the example used in the previous subsection to show a lower bound on POC for stochastic facility location.

Consider the following instance of the two-stage stochastic Steiner tree problem with n terminal nodes. There is a partition of the n terminal nodes into $\sqrt{n}$ disjoint sets $A_1, \ldots, A_{\sqrt{n}}$ containing $\sqrt{n}$ nodes each. Corresponding to each set B in the Cartesian product $\mathscr{B} = A_1 \times \cdots \times A_{\sqrt{n}}$, there is a (nonterminal) node v_B in the graph that is connected directly via an edge to each terminal node in B. Assume that each terminal node has a marginal probability of $1/\sqrt{n}$ to appear in the demand set S. Each edge $e \in E$ costs $w_e^I = \log n/\sqrt{n}$ in the first stage, and $w_e^{II} = 1$ in the second stage.

Then, in the optimal decision made using independent distribution, at most $\epsilon\sqrt{n}$ edges will be bought in the first stage, which can provide at most $\epsilon\sqrt{n}$ nonterminal nodes. Because no two nodes in any A_k are directly connected to each other or to any common nonterminal node, these $\epsilon\sqrt{n}$ nonterminal nodes are directly connected to at most $\epsilon\sqrt{n}$ nodes in a set A_k. Each of the remaining nodes in A_k will require at least two edges in order to be connected to the Steiner tree. Therefore, if the distribution is of form $\Pr(A_k) = 1/\sqrt{n}$, $k = 1, \ldots, \sqrt{n}$, then the expected cost of this decision will be $g(\mathbf{x}_I) \geqslant 2\sqrt{n}(1-\epsilon) + \epsilon\log(n)$. A robust solution is to instead buy enough edges in the first stage so that a set of $\sqrt{n}$ nonterminal nodes $\{v_B\}$ corresponding to a collection of $\sqrt{n}$ disjoint sets in $\mathscr{B}$ are connected to each other. By construction, any two nonterminal nodes are connected by a path of length of at most 3 to each other; therefore, this requires buying at most $3\sqrt{n}$ edges in the first stage, costing at most $3\log(n)$. Also, for any k, each node in A_k is connected directly to one of these nonterminal nodes. Therefore, the worst-case expected cost for this solution is $g(\mathbf{x}_R) \leqslant 3\log(n) + \sqrt{n}$. This shows $g(\mathbf{x}_I) \geqslant (2-\epsilon)g(\mathbf{x}_R)$ for any $\epsilon > 0$. □

The above proposition reduces the distributionally robust stochastic Steiner tree problem to the well-studied (for example, see Gupta et al. 2004) SST problem under known (independent Bernoulli) distribution at the expense of a 4-approximation factor.

4.3. Stochastic Bottleneck Matching

Consider a graph $G = (V, E)$, and let $\mathscr{M}$ denote the set of all perfect matchings in graph G. Associate a cost $\xi_{ij} \in \mathbb{R}_+$ with every edge $(i, j) \in E$. The bottleneck matching problem (Derigs 1980) is to find a perfect matching of minimum cost, where the cost of a matching is determined by the most expensive edge in the matching. Formally,

$$\underset{\sigma\in\mathscr{M}}{\text{minimize}}\ \underset{(i,j)\in\sigma}{\text{maximize}}\ \xi_{ij}.$$

In the stochastic version of this problem, the edge costs $\{\xi_{ij}\}$ are random variables, and the objective is to find the matching σ that minimizes expected cost $\mathbb{E}[h(\sigma, \boldsymbol{\xi})]$, where $h(\sigma, \boldsymbol{\xi}) = \max_{(i,j)\in\sigma} \xi_{ij}$. It is easy to verify that for every fixed matching σ, the objective function $h(\sigma, \boldsymbol{\xi})$ is nondecreasing and submodular in $\boldsymbol{\xi}$. Therefore, applying Theorem 2 for submodular functions of *continuous* random variables ($\Omega = \mathbb{R}^n$), we obtain the following result.

Proposition 3. *For stochastic bottleneck matching,* POC $\leqslant e/(e-1)$.

Thus, if correlations are unknown, random variables $\{\xi_{ij}\}$ can be assumed to be independent to get $e/(e-1)$ approximation for the corresponding (DRSP) model.

4.4. d-Dimensional Maximum Matching

A d-dimensional matching is a generalization of bipartite matching (a.k.a. 2-dimensional matching) to d-uniform hypergraphs. A d-uniform hypergraph consists of d disjoint sets of vertices $V_1, \ldots, V_d$. The set of hyperedges is given by $E = V_1 \times V_2 \times \cdots \times V_d$, and each edge is associated with a nonnegative weight given by a d-dimensional weight matrix W. For an edge $(i_1, \ldots, i_d) \in E$, we denote its weight by $W[i_1, \ldots, i_d]$. A set of edges $M \subseteq E$ forms a d-dimensional matching if every vertex appears in at most one edge in M. The d-dimensional maximum matching problem is to find a matching M of maximum weight. Assuming w.l.o.g. that $|V_1| = \cdots = |V_d| = T$, the problem is formulated as

$$\begin{aligned}
\max_x \quad & \sum_{1\leqslant i_1,\ldots,i_d \leqslant T} W[i_1,\ldots,i_d]x[i_1,\ldots,i_d] \\
\text{s.t.} \quad & \sum_{\{i_1,\ldots,i_d \mid i_j = t\}} x[i_1,\ldots,i_d] = 1 \\
& \qquad\qquad \text{for } 1 \leqslant j \leqslant d, 1 \leqslant t \leqslant T \qquad (11) \\
& x[i_1,\ldots,i_d] \in \{0,1\} \quad \text{for all } i_1,\ldots,i_d,
\end{aligned}$$

where $W[i_1, \ldots, i_d]$ denotes the weight of hyperedge $(i_1, \ldots, i_d)$, and $x[i_1, \ldots, i_d]$ denotes the decision whether to include the hyperedge $(i_1, \ldots, i_d)$ in the matching. Every node should be included in exactly one hyperedge in a matching.

d-dimensional maximum matching is a notoriously hard problem. To date the best approximation known for general d is $2/d$ (Hurkens and Schrijver 1989, Berman 2000). Also, it is known that there is no polynomial-time algorithm that achieves an approximation factor of $\Omega(\ln(d)/d)$ unless $P = NP$ (Hazan et al. 2006).

Our result on the correlation gap will provide a $1 - 1/e$ approximation for this problem when the weight matrix W is monotone and satisfies “Monge property.” Monge property has been studied extensively earlier for the d-dimensional *minimum* matching and transportation problem (e.g., in Burkard et al. 1996, Bein et al. 1995), in which case it characterizes precisely the instances of the minimization problem that are solvable by greedy algorithm. However, in the case of maximum matching, it is easy to show that the approximation factor for the greedy algorithm can be as bad as $O(1/n)$, even under Monge property. To our knowledge, ours is the first result that shows that the maximum matching problem has a constant approximation under monotonicity and the Monge property.

Proposition 4. *A uniformly random matching created by randomly permuting the values of each coordinate i_j gives a $1/\bar{\kappa}$-approximate solution for the d-dimensional maximum matching problem, if the correlation gap of the function $W(i_1, \ldots, i_d)$ is at most $\bar{\kappa}$. $\bar{\kappa} \leqslant e/(e-1)$ if matrix W is monotone in $i_1, \ldots, i_d$ and satisfies the Monge property.*

Proof. Observe that on relaxing the integrality constraints of (11) and scaling the variables by T, we get exactly the problem of finding the worst-case joint distribution when the marginal distribution on each variable $i_j \in \{1, \ldots, T\}$ is the uniform distribution. Therefore, a solution generated by sampling from the product of uniform distributions will give $1/\bar{\kappa}$ approximation to this problem. When W is a monotone Monge matrix, the function $W(i_1, \ldots, i_d)$ is monotone and submodular in $(i_1, \ldots, i_d)$, therefore $\bar{\kappa} \leqslant e/(e-1)$. □

4.5. Welfare Maximization

Consider the problem of maximizing the total utility achieved by partitioning a set V of n goods among K players, all with identical utility functions $f(S)$ for a subset S of goods.[5] The optimal welfare OPT is obtained by the following integer program:

$$\begin{aligned}
\max_\alpha \quad & \sum_S \alpha_S f(S) \\
\text{s.t.} \quad & \sum_{S:\, i\in S} \alpha_S = 1, \quad \forall i \in V \\
& \sum_S \alpha_S = K \\
& \alpha_S \in \{0,1\}, \quad \forall S \subseteq V.
\end{aligned} \qquad (12)$$

Observe that on relaxing the integrality constraints on α and scaling it by $1/K$, the above problem reduces to that of finding the worst-case distribution α^* such that the marginal probability $\sum_{S:\, i\in S} \alpha_S$ of each element $i \in V$ is $1/K$. Therefore,

$$\text{OPT} \leqslant \mathbb{E}_{\alpha^*}[Kf(S)].$$

Consequently, our correlation gap bounds lead to the following corollary for welfare maximization problems:

Proposition 5. *For welfare maximization problem (12) with n goods and K players with identical utility functions f, the randomized algorithm that assigns goods independently to each of the K players with probability $1/K$ gives $1/(2\beta)$ (or, $1 - 1/e$) approximation if function f is nondecreasing and admits a β-cost-sharing scheme (or, is submodular).*

For submodular functions, the above result matches the $1 - 1/e$ approximation factor that was proven earlier in the literature (Calinescu et al. 2007, Vondrák 2008), for the case of identical monotone submodular functions. Also, it extends the result to problems with *nonsubmodular* functions not previously studied in the literature.

5. Conclusions

In this paper, we investigated how well the simple approach of ignoring correlations and assuming independence approximates the distributionally robust stochastic programming model. We introduced a new concept of price of correlations (POC) to measure the approximation ratio achieved. We believe that the concept of POC is especially attractive because it characterizes the cases when the seemingly pessimistic worst-case joint distribution is close to the more natural independent distribution, in the sense that the former can be substituted for by the latter. By proving the upper and lower bounds on POC for a wide range of problems, our research provides important insights on when correlations can be ignored in practice. We also show that many deterministic optimization problems that involve matching or partitioning constraints can be formulated as the problem of computing the worst-case distribution with given marginals. Hence, our results provide approximation algorithms for those as well. Finally, our methodology of bounding POC using cost-sharing schemes is a novel application of these algorithmic game theory techniques, and deserves further study.

Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Endnotes

1. Here, "correlation" refers to any departure of two or more random variables from probabilistic independence.
2. Here e is the mathematical constant $e = 2.71828\ldots$
3. For any two functions $f(n)$, $g(n)$, $f(n) = \Omega(g(n))$ if there exists some constant c_1 such that $f(n) \geqslant c_1 g(n)$; and $f(n) = O(g(n))$, if there exists some constant c_2 such that $f(n) \leqslant c_2 g(n)$.
4. A function $f\colon 2^V \to \mathbb{R}$ is subadditive iff $f(S \cup T) \leqslant f(S) + f(T), \forall S, T \subseteq V$. f is fractionally subadditive iff $f(S) \leqslant \sum_j a_j f(T_j)$ for all $S \subseteq V$ and collection $\{T_j\}$ of subsets of S that forms a fractional cover of S, i.e., $\forall\, i \in S$, $\sum_{j:\, i \in T_j} a_j \geqslant 1$.
5. A more general formulation of this problem that is often considered in the literature allows nonidentical utility functions for different players.

Acknowledgments

The authors thank Ashish Goel and Mukund Sundararajan for useful insights on the problem. The research was partially supported by the Boeing Company.

References

Agrawal, S., Y. Ding, A. Saberi, Y. Ye. 2010. Correlation robust stochastic optimization. *SODA '10*: *Proc. 16th Annual ACM-SIAM Sympos. Discrete Algorithms*, SIAM, Philadelphia.

Ahmed, S., A. Shapiro. 2002. The sample average approximation method for stochastic programs with integer recourse. *SIAM J. Optim.* **12**(2) 479–502.

Bein, W. W., P. Brucker, J. K. Park, P. K. Pathak. 1995. A Monge property for the d-dimensional transportation problem. *Discrete Appl. Math.* **58**(2) 97–109.

Ben-Tal, A. 2001. Robust optimization—Methodology and applications. *Math. Programming* **92**(3) 453–480.

Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Math. Oper. Res.* **23**(4) 769–805.

Ben-Tal, A., A. Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming* **88** 411–424.

Berman, P. 2000. A $d/2$ approximation for maximum weight independent set in d-claw free graphs. *Nordic J. Comput.* **7** 178–184.

Bertsimas, D., I. Popescu. 2005. Optimal inequalities in probability theory: A convex optimization approach. *SIAM J. Optim.* **15**(3) 780–804.

Bertsimas, D., M. Sim. 2004. The price of robustness. *Oper. Res.* **52**(1) 35–53.

Bertsimas, D., K. Natarajan, C.-P. Teo. 2005. Probabilistic combinatorial optimization: Moments, semidefinite programming, and asymptotic bounds. *SIAM J. Optim.* **15** 185–209.

Bertsimas, D., I. Popescu, J. Sethuraman. 2000. *Moment Problems and Semidefinite Programming*. Kluwer Academic Publishers, Norwell, MA.

Burkard, R. E., B. Klinz, R. Rudolf. 1996. Perspectives of Monge properties in optimization. *Discrete Appl. Math.* **70**(2) 95–161.

Calinescu, G., C. Chekuri, M. Pál, J. Vondrák. 2007. Maximizing a submodular set function subject to a matroid constraint (extended abstract), IPCO, Springer-Verlag, Berlin, 182–196.

Charikar, M., C. Chekuri, M. Pál. 2005. Sampling bounds for stochastic optimization. *APPROX-RANDOM, Lecture Notes in Computer Science* 3624, Springer, 257–269.

Chen, X., M. Sim, P. Sun. 2007. A robust optimization perspective on stochastic programming. *Oper. Res.* **55**(6) 1058–1071.

Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **58** 595–612.

Derigs, U. 1980. On two methods for solving the bottleneck matching problem. *Lecture Notes in Control and Inform. Sci.* **23** 176–184.

Dupacová, J. 1987. The minimax approach to stochastic programming and an illustrative application. *Stochastics* **20**(1) 73–88.

Dupacová, J. 2001. Stochastic programming: Minimax approach. *Encyclopedia of Optimization*, Vol. 5. Kluwer Academic Publishers, Boston, 327–330.

Ermoliev, Y., A. Gaivoronski, C. Nedeva. 1985. Stochastic optimization problems with incomplete information on distribution functions. *SIAM J. Control Optim.* **23** 697–716.

Gaivoronski, A. A. 1991. A numerical method for solving stochastic programming problems with moment constraints on a distribution function. *Ann. Oper. Res.* **31**(1) 347–369.

Goh, J., M. Sim. 2010. Distributionally robust optimization and its tractable approximations. *Oper. Res.* **58**(4) 902–917.

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

Gul, F., E. Stacchetti. 1999. Walrasian equilibrium with gross substitutes. *J. Econom. Theory* **87**(1) 95–124.

Gupta, A., M. Pál, R. Ravi, A. Sinha. 2004. Boosted sampling: Approximation algorithms for stochastic optimization. *Proc. 36th Annual ACM Sympos. Theory Comput.*, ACM, New York, 417–426.

Hazan, E., S. Safra, O. Schwartz. 2006. On the complexity of approximating k-set packing. *Comput. Complexity* **15**(1) 20–39.

Hurkens, C. A. J., A. Schrijver. 1989. On the size of systems of sets every t of which have an SDR, with an application to the worst-case ratio of heuristics for packing problems. *SIAM J. Discrete Math.* **2**(1) 68–72.

Immorlica, N., M. Mahdian, V. S. Mirrokni. 2008. Limitations of cross-monotonic costsharing schemes. *ACM Trans. Algorithms* **4**(2) 1–25.

Kelso, A. S., Jr., V. P. Crawford. 1982. Job matching, coalition formation, and gross substitutes. *Econometrica: J. Econometric Soc.* **50**(6) 1483–1504.

Kimber, A. C. 1983. A note on Poisson maxima. *Probab. Theory Related Fields* **63** 551–552.

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

Kleinberg, J., Y. Rabani, É. Tardos. 1997. Allocating bandwidth for bursty connections. *Proc. 29th Annual ACM Sympos. Theory Computing. STOC '97*, ACM, New York, 664–673.

Könemann, J., S. Leonardi, G. Schäfer. 2005. A group-strategyproof mechanism for Steiner forests. *Proc. 16th Annual ACM-SIAM Sympos. Discrete Algorithms, SODA '05*, SIAM, Philadelphia, 612–619.

Lagoa, C. M., B. R. Barmish. 2002. Distributionally robust Monte Carlo simulation: A tutorial survey. *Proc. IFAC World Congress*, 1–12.

Landau, H. J. 1987. *Moments in Mathematics: Lecture Notes Prepared for the AMS Short Course*. American Mathematical Society, Providence, RI.

Leonardi, S., G. Schaefer. 2004. Cross-monotonic cost-sharing methods for connected facility location games. *EC '04: Proc. 5th ACM Conf. Electronic Commerce*, ACM, New York, 242–243.

Mahdian, M., M. Pál. 2003. Universal facility location. *Proc. ESA 03*, Springer, Berlin, 409–421.

Möhring, R. H., A. S. Schulz, M. Uetz. 1999. Approximation in stochastic scheduling: The power of LP-based priority policies. *J. ACM* **46**(6) 924–942.

Moulin, H. 1999. Incremental cost sharing: Characterization by coalition strategy-proofness. *Social Choice Welfare* **16** 279–320.

Moulin, H., S. Shenker. 2001. Strategyproof sharing of submodular costs: budget balance versus efficiency. *Econom. Theory* **18**(3) 511–533.

Nisan, N., T. Roughgarden, E. Tardos, V. V. Vazirani. 2007. *Algorithmic Game Theory*. Cambridge University Press, New York.

Pál, M., É. Tardo. 2003. Group strategyproof mechanisms via primal-dual algorithms. *Proc. 44th Annual IEEE Sympos. Foundations Comput. Sci. FOCS '03*, IEEE Computer Society, Washington, DC, 584–593.

Popescu, I. 2007. Robust mean-covariance solutions for stochastic optimization. *Oper. Res.* **55**(1) 98–112.

Prékopa, A. 1995. *Stochastic Programming*. Kluwer Academic Publishers, Norwell, MA.

Rogosinski, W. W. 1958. Moments of non-negative mass. *Proc. Royal Soc. London. Series A, Math. Physical Sci.* **245**(1240) 1–27.

Ruszczynski, A., A. Shapiro, eds. 2003. *Stochastic Programming*. Handbooks in Operations Research and Management Science, Vol. 10. Elsevier, Amsterdam.

Scarf, H. E. 1958. A min-max solution of an inventory problem. K. J. Arrow, S. Karlin, H. E. Scarf, eds. *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, Stanford, CA, 201–209.

Shapiro, A. 2006. Worst-case distribution analysis of stochastic programs. *Math. Programming* **107**(1) 91–96.

Shapiro, A., S. Ahmed. 2004. On a class of minimax stochastic programs. *SIAM J. Optim.* **14**(4) 1237–1252.

Shapiro, A., A. Kleywegt. 2002. Minimax analysis of stochastic problems. *Optim. Methods Software* **17**(3) 523–542.

Swamy, C., D. B. Shmoys. 2005. Sampling-based approximation algorithms for multistage stochastic optimization. *Proc. 46th Annual IEEE Sympos. Foundations Comput. Sci.*, IEEE Computer Society, Washington, DC, 357–366.

Vondrák, J. 2008. Optimal approximation for the submodular welfare problem in the value oracle model. *STOC '08: Proc. 40th Annual ACM Sympos. Theory Comput.*, ACM, New York, 67–74.

Yue, J., B. Chen, M.-C. Wang. 2006. Expected value of distribution information for the newsvendor problem. *Oper. Res.* **54**(6) 1128–1136.

Žáčková, J. 1966. On minimax solutions of stochastic linear programming problems. *Časopis Pro Pěstování Matematiky* **91**(4) 423–430.