OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0030-364X | EISSN 1526-5463 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

Price of Correlations in Stochastic Optimization

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

Yichuan Ding, Amin Saberi, Yinyu Ye *
y7ding@stanford.edu, saberi@stanford.edu, yyye@stanford.edu,
Department of Management Science and Engineering, Stanford university

When decisions are made in the presence of large-scale stochastic data, it is common to pay more attention to the easy-to-see statistics (e.g., mean) instead of the underlying correlations. One reason is that it is often much easier to solve a stochastic optimization problem by assuming independence across the random data. In this paper, we study the possible loss incurred by ignoring these correlations through a distributionally-robust stochastic programming model, and propose a new concept called *Price of Correlations* (POC) to quantify that loss. We show that the POC has a small upper bound for a wide class of cost functions, including uncapacitated facility location, Steiner tree and submodular functions, suggesting that the intuitive approach of assuming independent distribution may actually work well for these stochastic optimization problems. On the other hand, we demonstrate that for some cost functions, POC can be particularly large, e.g., the supermodular functions. We propose alternative ways to solve the corresponding distributionally robust models for these functions. As a byproduct, our analysis yields new results on social welfare maximization and the existence of Walrasian equilibria, which may be of independent interest.

1. Introduction

In many planning problems, the correlations among individual events contain crucial information. For example, the emergency service (medical services, fire rescue, etc) planner needs to carefully locate emergency service stations and determine the number of emergency vehicles in order to dispatch vehicles to the call points in time. If the planner assumes emergency calls rare and independent events, he simply needs to make sure that every potential call point is in the service range of at least one station; however, there might exist certain kinds of correlations 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 various risky assets as well as their individual performances, and the stochastic facility location problem, in which the supplier wants to learn more about the correlations between demands from different retailers.

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

Decision making with uncertain data is usually investigated in the context of Stochastic Programming (SP) (Ruszczynski and Shapiro (2003)). In SP, the decision maker optimizes expected value

* Research partially supported by the Boeing Company.

of an objective function that involves stochastic variables. In general, such a stochastic program can be expressed as

$$\min_{x\in C} \mathbb{E}[f(x,\xi)], \tag{1}$$

where x is the decision variable which lies in a constrained set C, and the random variable ξ cannot be observed before the decisions x is made. $f(x,\xi)$ is a *cost function* which depends on both the decision x and the random scenario ξ. Throughout this paper, we assume the cost functions are non-negative. If the underlying distribution of the random variable is unknown, then the decision maker needs to estimate it either via a parametric approach, which assumes the distribution with a certain closed form and fits its parameters by empirical data, or via a non-parametric approach, e.g., the Sample Average Approximation (SAA) method (e.g., Ahmed et al. (2002), Ruszczynski and Shapiro (2003), Swamy and Shmoys (2005), Charikar et al. (2005)), which optimizes the average objective value over a set of sampled data. However, these models are suitable only when one has access to a significant amount of reliable time-invariant statistical information. If the samples are insufficient to fit the parameters or to approximate the distribution, then SP fails to address the problem. In this case, one may instead consider optimizing the worst-case outcome, which is usually easier to characterize than estimating the joint distribution. That is,

$$\min_{x\in C} \max_{\xi} f(x,\xi), \tag{2}$$

Such a method is termed Robust Optimization (RO), following the recent literature (e.g., Ben-Tal and Nemirovski (1998), Ben-tal and Nemirovski (2000), Ben-Tal (2001)). However, the robust solution is sometimes too pessimistic compared with 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 and sometimes even not representable.

An intermediate approach that may address the limitations of SP and RO is distributionally-robust stochastic programming (DRSP). In this approach one minimizes the maximum expected cost over a collection of possible probability distributions:

$$\min_{x\in C} \max_{D\in\mathcal{D}} \mathbb{E}_D[f(x,\xi)], \tag{3}$$

where $\mathcal{D}$ is a collection of possible probability distributions of ξ, and $\mathbb{E}_D[f(x,\xi)]$ denotes the expected value of $f(x,\xi)$ over a distribution D in the collection $\mathcal{D}$. The decision maker chooses a decision x hoping to minimize the expected cost, while the nature adversarily chooses a distribution from the set $\mathcal{D}$ to maximize the expected cost of the decision.

The DRSP model was proposed by Scarf as early as the 1950s (Scarf (1958)), but it has not received much attention until recently (Popescu (2007), Delage and Ye (2008), Goh and Sim (2009)). The DRSP model accurately characterizes the challenge of making decisions with limited information about the distribution. Our model for characterizing price of correlations is based on this distributionally robust model of optimization. Because our main intention is to overcome the challenges involved in making decisions without correlation information, we will simplify the statistics of a single event. We assume that each event i of the ground set V is in a binary form that takes a value of 1 with a probability p_i. Even with such an assumption, we will show that our model covers many real-life planning problems. The marginal probability p_i reflects the chance that a particular event i occuring in the realized scenario S, and we assume that $(p_i)_{i\in V}$, as the decentralized information, are known by the centralized planner. The planning problem is then to choose a decision x that minimizes the expected cost $\mathbb{E}[f(x,S)]$ under the worst possible distribution with these marginal probabilities. This is formulated as the following in the DRSP model:

$$\min_{x\in C}\ g(x), \tag{4}$$

where $g(x)$ is the expected cost under worst-case distribution when decision x has been made, given by

$$g(x) := \max_{D \in \mathcal{D}} \quad \mathbb{E}_D[f(x,S)] \\ s.t. \ \sum_{S: i \in S} \mathbb{P}_D(S) = p_i. \ \forall i \in V. \tag{5}$$

Here, $\mathcal{D}$ is a collection of all probability distributions with a domain 2^V that satisfy the marginal probability constraint for each event i. For any distribution $D \in \mathcal{D}$, the terms $\mathbb{E}_D[f(x,S)]$ and $\mathbb{P}_D(S)$ denote expectation and probability, respectively, when random variable S has distribution D. We refer to this distributionally robust formulation of our problem as the "correlation robust model".

We believe that the correlation robust model (4)-(5) is very useful because it takes advantage of the stochasticity of the input, and at the same time efficiently utilizes the available information. On the other hand, it defines an exponential-size (exponential number of scenario S) linear program which makes the problem potentially difficult to solve. A common strategy for such linear programs is to solve the corresponding dual LP with exponential number of constraints, using the separation hyperplane approach. However, for the above model, approximating the separating hyperplane problem can be shown to be harder than the max-cut problem even for the special case, when the function f is submodular in S.

A natural question is how much risk it involves to simply ignore the correlations and minimize the expected cost of independent Bernoulli distribution (also known as product distribution) with marginals (p_i) instead of the worst case distribution. Or, in other words, how well the stochastic optimization model with independent distribution approximates the above correlation robust model. The focus of this paper is to study this 'price of correlations' incurred by the assumption of independence. For a particular problem instance $(f, V, \{p_i\})$ of stochastic optimization, let x_I be the optimal decision assuming independent distribution, and x_R is the optimal decision for the correlation robust model. Then, *price of correlations (POC)* compares the performance of x_I to x_R. That is,

$$POC = \frac{g(x_I)}{g(x_R)}$$

A small POC indicates that the decision maker can take the product distribution as an approximation of the worst case distribution without involving much risk. And a stochastic optimization problem with product distribution is often more easy to solve either by sampling or by other algorithmic techniques (e.g., Kleinberg et al. (1997), Möhring et al. (1999)). Further, 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 those cases, POC can provides a guideline to deciding how much resource should be spent on learning these correlations.

Our main result is to characterize a wide class of cost functions that have small POC. We define this class using concepts of cost-sharing from game theory. This novel application of cost-sharing schemes may be interesting in its own respect. We also provide counter-examples that demonstrate a large lower bound on POC for various other classes of functions.

Below, we summarize our key results:

- *A new approach for discrete stochastic optimization:* As a novel application of distributionally robust framework, we study this model for stochastic optimization when the random variables in question are binary random variables, i.e. form subsets of a ground set. Further, we introduce a new concept of Price of Correlations (POC) to compare the worst case distribution under given marginals to the independent Bernoulli distribution. By taking advantage of the richness of structure in these problems, we obtain many non-trivial bounds on POC, thus illustrating that the intuitive and efficient approach of assuming independent distribution may actually work well for many of these stochastic optimization problems.

• *A class of functions with nicely bounded POC:* For functions $f(x,S)$ that are non-decreasing in S and have a cross-monotone, β-budget balanced , η-summable cost-sharing scheme in S for all x, we show that POC is upper bounded by $\min\left\{2\beta, \eta\beta\left(\frac{e}{e-1}\right)\right\}$. Using this result, we prove that POC $\leq e/(e-1)$ for problems with submodular cost functions, POC ≤ 6 for uncapacitated facility location, and POC ≤ 4 for the Steiner tree cost function [1].

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

• *Polynomial-time algorithm for supermodular functions:* We analytically characterize the worst case distribution when function $f(x,S)$ is supermodular in S for all x, and consequently give a polynomial-time algorithm for the correlation robust model provided f is convex in x.

• *New results for welfare maximization problems:* As a byproduct, our result provides a $\max\{\frac{1}{2\beta}, \frac{1}{\eta\beta}(1-\frac{1}{e})\}$-approximation algorithm for the well-studied problem of social welfare maximization in combinatorial auctions, for identical utility functions that admit a (η,β)-budget balanced cross-monotonic cost-sharing scheme. Notably, this result implies $(1-1/e)$-approximation for *identical* submodular utility functions, matching the best approximation factor (Vondrak, 2008 Vondrak (2008)) for this case.

We also provide a simple counterexample for the conjecture by Bikhchandani and Mamer (1997) that markets that have buyers with identical submodular utilities admit a Walrasian price equilibria.

The rest of the paper is organized as follows. To begin, Section 2 will provide a mathematical definition of POC, and examples that illustrate a large POC for certain classes of cost functions. Here, we will introduce a secondary concept of *correlation gap*, and that an upper bound on POC can be obtained by upper bounding the correlation gap of the cost function for every decision. In Section 3, we present our main technical theorem that upper bounds the correlation gap for a wide class of cost functions, and discuss its implications on POC for various stochastic optimization problems and the welfare maximization problem. The proof of this theorem is presented in Section 4. Finally, in Section 5, we end with a direct solution of a correlation robust model for supermodular functions.

2. Price of Correlation (POC) and correlation gap

Let x_I be the optimal solution to the stochastic optimization problem (1) with independent Bernoulli distribution, and x_R be the optimal solution to the correlation robust problem (4). Then, we defined POC as the ratio of expected cost when using x_I versus using x_R, i.e.,

$$\text{POC} = \frac{g(x_I)}{g(x_R)}.$$

For a problem instance $(f, V, \{p_i\})$ and at any given decision x, we define *correlation gap* $\kappa(x)$ as the ratio between the expected cost of the worst case distribution and that of the independent distribution, i.e.,

$$\kappa(x) := \frac{\mathbb{E}_{\mathcal{D}^R(x)}[f(x,S)]}{\mathbb{E}_{\mathcal{D}^I}[f(x,S)]}, \tag{6}$$

where $\mathcal{D}^I$ is the independent Bernoulli distribution with marginals $\{p_i\}$, and $\mathcal{D}^R(x)$ is the worst-case distribution (as given by (5)) for decision x.

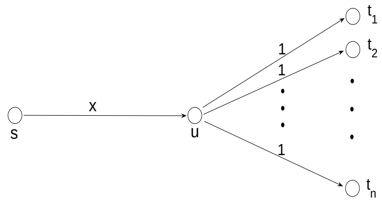

Figure 1 An example with exponential correlation gap

An upper bound on correlation gap has interesting implications on the POC of corresponding decision problem. Suppose that for particular cost function f, the correlation gap is upper bounded above by $\bar{\kappa}$ for all x, i.e. $\bar{\kappa} = \max_x \kappa(x)$. Then, POC will be upper bounded by $\bar{\kappa}$. More precisely, since

$$\begin{aligned} g(x_I) &= \mathbb{E}_{\mathcal{D}^R(x_I)}[f(x_I, S)], \quad \text{and} \\ g(x_R) &= \mathbb{E}_{\mathcal{D}^R(x_R)}[f(x_R, S)] \geq \mathbb{E}_{\mathcal{D}^I}[f(x_R, S)] \\ &\geq \mathbb{E}_{\mathcal{D}^I}[f(x_I, S)], \end{aligned} \tag{7}$$

by (7) and the definition of $\bar{\kappa}$, we have

$$\text{POC} = \frac{g(x_I)}{g(x_R)} \leq \frac{\mathbb{E}_{\mathcal{D}^R(x_I)}[f(x_I, S)]}{\mathbb{E}_{\mathcal{D}^I}[f(x_I, S)]} \leq \bar{\kappa}.$$

Hence, a uniform upper bound on correlation gap gives an upper bound on the POC for corresponding decision problem. Analyzing the correlation gap presents a relatively simpler challenge than directly analyzing POC, since the former assumes a fixed decision x. Therefore, most of our technical discussion will be concentrated around bounding the correlation gap, which will subsequently imply corresponding bounds on POC.

Unfortunately, for general cost functions, the correlation gap and hence the POC for corresponding decision problem can be large in order of n, as demonstrated by the following examples.

Example 1. (*Minimum cost flow: An instance with POC* $= \Omega(2^n)$ *for supermodular functions*) Consider a two-stage minimum cost flow problem as in Figure 1. There is a single source s, and n sinks $t_1, t_2, \ldots, t_n$. Each sink t_i has a probability $p_i = \frac{1}{2}$ to request a demand, 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 demand requests is revealed, defined as:

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

Given the first stage decision x, the cost of edges that need to be bought in the second stage to serve a set S of requests is given by: $f(x, S) = c^I(x) + c^{II}((|S| - x)^+) = c^I(x) + 2^n(|S| - x)^+$. It is easy to check that $f(x, S)$ is supermodular in S for any given x, i.e. $f(x, S \cup i) - f(x, S) \geq f(x, T \cup i) - f(x, T)$ for any $S \supseteq T$. The objective is to minimize the total expected cost $c^I(x) + \mathbb{E}[f(x, S)]$. 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 $\Pr(V) = \Pr(\emptyset) = 1/2$ and all other scenario have zero probability).

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

$$\text{POC} = g(x_I)/g(x_R) = \Omega(2^n).$$

□

EXAMPLE 2. (*Stochastic set cover: An instance with POC* $= \Omega(\sqrt{n}\frac{\log\log n}{\log n})$ *for subadditive functions*)
Consider a set cover problem with elements $V = \{1, \ldots, n\}$. Each item $j \in V$ has a marginal probability of $1/K$ to appear in the random set S. The covering sets are defined as follows. Consider a partition of V into $K = \sqrt{n}$ 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 unit cost. Then, cost of covering a set S is given by subadditive (infact, fractionally subadditive) function

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

The worst case distribution with marginal probabilities $p_i = 1/K$ is one where probabilities $\Pr(S) = 1/K$ for $S = A_i$, $i = 1, 2, \ldots, K$, and $\Pr(S) = 0$ otherwise. The expected value of $c(S)$ under this distribution is $K = \sqrt{n}$. For independent distribution, $c(S) = \max_{i=1,\ldots,K} \zeta_i$, where $\zeta_i = |S \cap A_i|$ are independent $(K, 1/K)$-binomially distributed random variables.

As K approaches ∞, since expected value remains fixed at 1, the Binomial(K, $1/K$) random variable approaches the Poisson distribution with expected value 1. Using some known results on maxima of independent poisson random variables in Kimber (1983), it can be shown that for large K, the expected value of the maximum of K i.i.d. poisson random variables is bounded by $\Theta(\log K/\log\log K)$ (refer to Appendix A for a detailed proof). This implies that $\mathbb{E}[\max_{i=1,\ldots,\sqrt{n}}\{\zeta_i\}]$ is bounded by $\Theta(\log n/\log\log n)$ for large n. So the correlation gap of cost function $c(S)$ is bounded below by $\Omega(\sqrt{n}\log\log n/\log n)$.

To obtain a lower bound on POC for a two-stage stochastic set cover (refer Swamy and Shmoys (2005) for details on this problem) instance, extend the above instance as follows. For ease of notation, let $L(n) = d\log n/\log\log n$, where d is a constant such that $\mathbb{E}[\max_i\{\zeta_i\}] \le L(n)$. Let the first stage cost of a covering set to be $w^I = (1+\epsilon)L(n)/\sqrt{n}$ for some small $\epsilon > 0$, and the second stage cost to be $w^{II} = 1$. For a given first stage cover x, let $B(x)$ be the set of elements covered by x, then total cost of covering elements in S is given by $f(x, S) = w^I|x| + c(S - B(x))$. Using above analysis for function $c(S)$, the optimal solution for independent distribution will be to buy no (or very few) sets in the first stage giving $\mathbb{E}[f(x, S)] \le L(n)$ for independent distribution, but $\Theta(\sqrt{n})$ cost for worst case distribution. On the other hand, the optimal robust solution considering worst case distribution is to cover all the elements in the first stage giving $O(L(n))$ cost in the worst case. Thus,

$$\text{POC} = \Omega(\sqrt{n}\log\log n/\log n).$$

□

These examples indicate that price of ignoring correlations and assuming independent distribution can be very high in many cases. However, below we identify a wide class of functions for which the correlation gap, and hence POC is well-bounded, indicating that independent distribution does give a close approximation to the worst-case distribution.

3. A class of functions with low POC

A key contribution of our paper is to identify a class of cost functions for which the correlation gap, and hence POC is well bounded. To our interest, many popular cost functions including submodular functions, facility location, Steiner forest, etc. belong to this class.

We derive our characterization using concepts of cost-sharing. A cost-sharing scheme is a function defining how to share the cost of a service among the serviced customers. A cost-sharing scheme is *cross-monotonic* if it satisfies the property that everyone is better off when the set of people who receive the service expands (e.g., see Nisan et al. (2007)). We consider the class of cost functions f such that for every feasible x, there exists some cost-sharing scheme for allocating the cost $f(x,S)$ among members of set S with (a) β-*budget balance* and (b) *cross-monotonicity*. Below we precisely state these properties. Since we assume that x can take any fixed value, we will abbreviate $f(x,S)$ as $f(S)$ for simplicity when clear from the context.

Define a cost-sharing scheme for function $f(S)$ as a function $\chi(i,S)$ that, for each element $i \in S$ specifies the share of i in S. Then, the scheme χ is β-budget balanced and cross-monotonic iff it satisfies the following conditions.

- β-*budget balance:* For all S,

$$f(S) \geq \sum_{i=1}^{|S|} \chi(i,S) \geq \frac{f(S)}{\beta}$$

- *Cross-monotonicity:* For all $i \in S$, $S \subseteq T$,

$$\chi(i,S) \geq \chi(i,T)$$

Let us call a cost-sharing scheme satisfying the above two properties a 'β-cost-sharing scheme'. Also, we say that a function $f(x,S)$ is non-decreasing in S if for every x and every $S \subseteq T$, $f(x,S) \leq f(x,T)$. Our main result is the following theorem, which we will prove in the next section:

THEOREM 1. *For any instance $(f,V,\{p_i\})$, for any fixed x, if the cost function $f(x,S)$ is non-decreasing in S and has a β-cost-sharing scheme for elements in S, then the correlation gap $\kappa(x)$ is upper-bounded by 2β.*

In Agrawal et al. (2010), we considered an additional property of weak η-summability for the cost-sharing schemes, derived from the concept of summable cost-sharing schemes by Roughgarden and Sundararajan (2006):

- *Weak η-summability:* A cost-sharing scheme $\chi(i,S)$ is weakly η-summable if for all S, and any pre-specified order σ_S on the elements of S,

$$\sum_{\ell=1}^{|S|} \chi(i_\ell, S_\ell) \leq \eta f(S)$$

where i_ℓ is the ℓ^{th} element and S_ℓ is the set of the first ℓ members of S according to the ordering σ_S.

Let us call an η-summable, β-budget balanced, cross-monotonic cost-sharing scheme as an '(η,β)-cost-sharing scheme'. The following result was proved in Agrawal et al. (2010):

THEOREM 2. *(Agrawal et al. (2010)) For any instance $(f,V,\{p_i\})$, for any fixed x, if the cost function $f(x,S)$ is non-decreasing in S and has an (η,β)-cost-sharing scheme for elements in S, then the correlation gap $\kappa(x)$ is upper-bounded by $\eta\beta\left(\frac{e}{e-1}\right)$.*

Combining the two results, we obtain the following corollary:

COROLLARY 1. *For any instance $(f, V, \{p_i\})$, for any fixed x, if the cost function $f(x, S)$ is non-decreasing in S and has an (η, β)-cost-sharing scheme for elements in S, then the correlation gap $\kappa(x)$ is upper-bounded by* $\min\left\{2\beta, \eta\beta\left(\frac{e}{e-1}\right)\right\}$.

As described in Section 2, this gives following corollary for upper bounding POC:

COROLLARY 2. *For instances $(f, V, \{p_i\})$ such that for all x, the cost function $f(x, S)$ is non-decreasing in S and has an (η, β)-cost-sharing scheme for elements in S, then POC is upper bounded by* $\min\left\{2\beta, \eta\beta\left(\frac{e}{e-1}\right)\right\}$.

We may point out that neither of Theorem 1 or Theorem 2 subsumes the other. As we will illustrate in the applications using example of submodular functions, in some cases there may exist a summable, budget-balanced cost-sharing scheme with particularly small η (close to 1), so that the bound of $\eta\beta e/(e-1)$ may turn out to be smaller than 2β. Otherwise, typically the bound of 2β given by Theorem 1 will be smaller and easier to use, since it does not require the additional property of summability.

Theorem 1 will be proved in the next section. Theorem 2 was proved in Agrawal et al. (2010). We have included its proof in Appendix D for completion. Before moving on to the technical proofs of these results, let us briefly discuss their implications for various stochastic optimization problems, and for a seemingly unrelated problem of welfare maximization in combinatorial auctions:

3.1. Stochastic optimization with submodular functions

A function $h: 2^V \to \mathbb{R}$ is submodular if $h(S \cup i) - h(S) \le h(T \cup i) - h(T)$ for all $S \supseteq T$, and $i \in V$. These cost functions are characterized by diminishing marginal costs, which is common for resource allocation problems where a resource can be shared by multiple users and thereby the marginal cost decreases as number of users increases. It is easy to verify that the following incremental cost-sharing scheme is 1-budget-balance and 1-weak-summable and cross-monotonic for any submodular function $h(S)$:

$$\chi(i, S) = h(S_i) - h(S_{i-1})$$

where S_i is the set of the first i members of S according to the fixed pre-specified ordering. Therefore, Corollary 2 gives an $\eta\beta\frac{e}{(e-1)} = \frac{e}{(e-1)}$ upper bound on POC for stochastic decision problems with cost function $f(x, S)$ that is submodular in S at all x. Moreover, the following example shows that the $e/(e-1)$ bound is tight for submodular functions.

EXAMPLE 3. (*Tightness*) Let $V := \{1, 2, \ldots, n\}$, define submodular function $h(S) = 1$ if $S \neq \emptyset$, and $h(\emptyset) = 0$. Let each item has a probability $p = \frac{1}{n}$. The worst case distribution that maximizes $\mathbb{E}[h(S)]$ is $Pr(\{i\}) = 1/n$ for each $i \in V$, with expected value 1. The independent distribution with the same marginals has an expected cost $1 - (1 - \frac{1}{n})^n \to 1 - 1/e$ as $n \to \infty$.

Extend this example to a stochastic decision problem with two possible decisions x_1, x_2 as follows. Define $f(x_1, S) = h(S), \forall S$, and $f(x_2, S) = 1 - \frac{1}{e} + \epsilon, \forall S$ 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 $1 - \frac{1}{e} + \epsilon$ in the worst case, giving POC $= \frac{e}{(e-1)} - \epsilon$. □

COROLLARY 3. *For stochastic problems with cost function $f(x, S)$ non-decreasing and submodular in S for all feasible x,* POC $= \frac{e}{(e-1)}$.

3.2. Stochastic Uncapacitated Facility Location (SUFL)

In 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 $x \in \{0,1\}^{|F|}$, the facilities to be build in the first stage so that the total expected cost $\mathbb{E}[f(x,S)]$ of facility location is minimized (refer to Swamy and Shmoys (2005) for further details on the problem definition).

Given a first stage decision x, the cost function $f(x,S) = w^I \cdot x + c(x,S)$, where $c(x,S)$ is the cost of deterministic UFL for set $S \subseteq V$ of customers and set F of facilities such that the facilities x already bought in first stage are available freely at no cost, while 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 Tardos (2003)). Therefore, using Corollary 2, we know that the POC for stochastic metric UFL has an upper bound of $2\beta = 6$. This observation reduces our robust facility location problem to the well-studied stochastic UFL problem under known (independent Bernoulli) distribution (Swamy and Shmoys (2005)) at the expense of a 6-approximation factor.

The next example shows POC ≥ 3 for the stochastic metric UFL problem [2].

Example 4. (*Lower bound*) Consider the following instance of two-stage metric facility location problem with n cities. Consider a partition of the n cities into $\sqrt{n}$ disjoint sets $A_1, \ldots, A_{\sqrt{n}}$ of $\sqrt{n}$ cities each. Corresponding to each set B 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 j to facility F_B such that $j \notin B$ is 3. Assume that each city has a marginal probability of $\frac{1}{\sqrt{n}}$ to appear in the demand set S. Each facility costs $w^I = 3\frac{\log n}{\sqrt{n}}$ in the first stage, and $w^{II} = 3$ in the second stage.

Then the worst case distribution is $Pr(A_i) = \frac{1}{\sqrt{n}}, i = 1, \ldots, \sqrt{n}$. The optimal solution for the worst-case distribution is to build $\sqrt{n}$ facilities in the first stage, corresponding to any $\sqrt{n}$ disjoint sets in the collection $\mathcal{B}$. These facilities will cover every city with a connection cost of 1. Thus, the expected cost for the optimal robust solution is $g(x_R) = \sqrt{n} + 3\log n$. Now 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 $3\mathbb{E}[\max_i |A_i \cap S|] + \sqrt{n}$. Using the analysis from Example 2, $\mathbb{E}[\max_i |A_i \cap S|]$ asymptotically reaches $O(\frac{\log n}{\log\log n}) = o(\log(n))$ for large n. Therefore, for any $\epsilon > 0$, for large enough n, $\mathbb{E}[\max_i |A_i \cap S|] < \epsilon \log(n)$. 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 that of the worst case, such a strategy induces an expected cost $g(x_I) \geq 3(1-\epsilon)\sqrt{n} + 3\epsilon\log(n)$, which shows $\frac{g(x_I)}{g(x_R)} \geq (3-\epsilon)$ for any $\epsilon > 0$.

□

With this example, we conclude:

Corollary 4. *For the stochastic uncapacitated facility location (metric) problem,* $3 \leq \text{POC} \leq 6$.

3.3. 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. The random set S of terminals 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 x is the edges to be bought in the first stage, and cost function $f(x,S) = w^I \cdot x + c(x,S)$, where $c(x,S)$ is the Steiner tree cost for set S given that the edges in x are already bought. Since a $\beta = 2$-budget balanced cost-sharing scheme is known for deterministic Steiner tree (Könemann et al. (2005)), we can use Corollary 2 to conclude that for this problem $POC \leq 2\beta = 4$. This observation reduces our robust 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.

The following example shows that $POC \geq 2$ for two stage stochastic Steiner tree. The construction of this example is very similar to Example 4 used to show lower bound for stochastic facility location.

Example 5. (*Lower bound of 2*) Consider the following instance of two-stage stochastic Steiner tree problem with n terminal nodes. Consider a partition of the n terminal nodes into $\sqrt{n}$ disjoint sets $A_1, \ldots, A_{\sqrt{n}}$ of $\sqrt{n}$ nodes each. Corresponding to each set B in the cartesian product $\mathcal{B} = A_1 \times \cdots \times A_{\sqrt{n}}$, there is a (non-terminal) node v_B in the graph which is connected directly via an edge to each terminal node in B. Assume that each terminal node has a marginal probability of $\frac{1}{\sqrt{n}}$ to appear in the demand set S. Each edge costs $w^I = \frac{\log n}{\sqrt{n}}$ in the first stage, and $w^{II} = 1$ in the second stage.

Then the worst case distribution is $Pr(A_i) = \frac{1}{\sqrt{n}}, i = 1, \ldots, \sqrt{n}$. The optimal solution for the worst-case distribution is to buy enough edges in the first stage so that a set of $\sqrt{n}$ non-terminal nodes $\{v_B\}$ corresponding to any $\sqrt{n}$ disjoint sets in $\mathcal{B}$ are connected to each other. By construction, any two non-terminal nodes are connected by a path of length atmost 3 to each other, therefore this requires buying atmost $3\sqrt{n}$ edges in the first stage costing atmost $3\log(n)$. Also, for any i, each node in A_i is connected directly to exactly one of these non-terminal nodes. Therefore, expected cost for this solution is $g(x_R) = 3\log(n) + \sqrt{n}$. On the other hand, for independent distribution, using arguments similar to Example 4 atmost $\epsilon\sqrt{n}$ edges will be bought in the first stage, which can make available atmost $\epsilon\sqrt{n}$ non-terminal nodes. Since no two nodes in any A_i are directly connected to each other or to any common non-terminal node, these $\epsilon\sqrt{n}$ non-terminal nodes are directly connected to atmost $\epsilon\sqrt{n}$ nodes in a set A_i. Also, each of the remaining node in A_i will require atleast two edges in order to be connected to the Steiner tree. Therefore, in the worst case, the expected cost of this decision will be at least $g(x_I) \geq 2\sqrt{n}(1-\epsilon) + \epsilon\log(n)$, which shows $\frac{g(x_I)}{g(x_R)} \geq (2-\epsilon)$ for any $\epsilon > 0$. □

Corollary 5. *For the stochastic Steiner tree problem,* $2 \leq \mathrm{POC} \leq 4$.

3.4. Welfare Maximization Problem and An Example of Non-existence of Walrasian Equilibrium

Finally, our results on correlation gap extend some existing results for *social welfare maximization* in combinatorial auctions. Consider the problem of maximizing total utility achieved by partitioning n goods among K players each with utility function $f(S)$ for subset S of goods [3]. The optimal welfare OPT is obtained by following integer program:

$$\begin{array}{ll} \max_\alpha & \sum_S \alpha_S f(S) \\ \text{s.t.} & \sum_{S: i \in S} \alpha_S = 1, \; \forall i \in V \\ & \sum_S \alpha_S = K \\ & \alpha_S \in \{0,1\}, \qquad \forall S \subseteq V \end{array} \tag{8}$$

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^* (i.e. one that maximizes expected value $\sum_S \alpha_S f(S)$ of function f) such that the marginal probability $\sum_{S: i \in S} \alpha_S$ of each element is $1/K$. Therefore:

$$\mathrm{OPT} \leq \mathbb{E}_{\alpha^*}[K f(S)]$$

Consequently, the correlation gap bounds in Theorem 1 and Theorem 2 lead to the following corollary for welfare maximization problems:

COROLLARY 6. *For welfare maximization problems 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 $\max\{\frac{1}{2\beta}, \frac{1}{\eta\beta}(1-\frac{1}{e})\}$ approximation to the optimal partition; given that function f is non-decreasing and admits an (η,β)-cost-sharing scheme.*

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

The reader may observe that even though approximating the worst case distribution directly provides a matching approximation for the corresponding welfare maximization problem, the converse is not true. In addition to having uniform probabilities $p_i = 1/K$, solutions for welfare maximization approximate the integer program (8), where as the worst case distribution requires solving the corresponding LP relaxation. The latter is a strictly harder problem unless the integrality gap is 0. A notable example is the above-mentioned case of identical submodular functions. This case was studied by Bikhchandani and Mamer (1997) in context of Walrasian equilibria who conjectured a 0 integrality gap for this problem implying the existence of Walrasian equilibria. However, in appendix C , we show a simple counter-example with non-zero integrality gap (11/12) for this problem. As a byproduct, this counter-example proves that even for identical submodular valuation functions, Walrasian equilibria may not exist.

4. Proof of Theorem 1

For a problem instance $(f,V,\{p_i\})$ and fixed x, use $\mathcal{L}(f,V,\{p_i\})$ and $\mathcal{I}(f,V,\{p_i\})$ to denote the expected cost of worst-case distribution and independent Bernoulli distribution with marginals $\{p_i\}$, respectively. In this section, we prove our main technical result that the correlation gap

$$\frac{\mathcal{L}(f,V,\{p_i\})}{\mathcal{I}(f,V,\{p_i\})} \le 2\beta$$

when f is non-decreasing and admits β-budget balanced cross-monotonic cost-sharing in S. As before, we will abbreviate $f(x,S)$ as $f(S)$ for simplicity.

The proof is structured as follows. We first focus on special instances of the problem in which all p_i's are equal to $1/K$ for some integer K, and the worst case distribution is a "K-partition-type" distribution. That is, the worst case distribution divides the elements of V into K disjoint sets $\{A_1,\dots,A_K\}$, and each A_k occurs with probability $1/K$. Observe that for such instances, the expected value under worst case distribution is $\mathcal{L}(f,V,\{p_i\}) = \frac{1}{K}\sum_k f(A_k)$. In Lemma 1, we show that for such "nice" instances the correlation gap is bounded by 2β. Then, we use a "split" operation to reduce any given instance of our problem to a "nice"' instance such that the reduction can only increase the correlation gap. This will show that the bound 2β for nice instances is an upper bound for any instance of the problem, thus concluding the proof of the theorem.

LEMMA 1. *For instances $(f,V,\{p_i\})$ such that (a) $f(S)$ is non-decreasing and admits a β-cost-sharing scheme (b) marginal probabilities p_i are all equal to $1/K$ for some integer K, and (c) the worst case distribution is a K-partition-type distribution, the correlation gap is bounded as:*

$$\frac{\mathcal{L}(f,V,\{1/K\})}{\mathcal{I}(f,V,\{1/K\})} \le (2-\frac{1}{K})\beta$$

Proof. Let the optimal K-partition corresponding to the worst case distribution is $\{A_1,\ A_2,\ \dots,A_K\}$. For any set S, denote $S_{\cap j} = S\cap A_j$, and $S_{-j} = S - A_j$, for $j = 1,\dots,K$. Let χ

is 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:

$$\mathcal{I}(f,V,\{1/K\}) = \mathbb{E}_S[f(S)] \ \geq\ \mathbb{E}_S\left[\sum_{k=1}^{K}\chi(S_{\cap k},S)\right] \tag{9}$$

where the expected value is taken over independent distribution on V with marginal probabilities $1/K$.

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

$$\begin{aligned}\mathbb{E}[\chi(S_{\cap k},S)] &\geq \mathbb{E}[\chi(S_{\cap k},S\cup A_k)]\\ &= \mathbb{E}_S[\textstyle\sum_{i\in A_k} I(i\in S_{\cap k})\chi(i,S_{-k}\cup A_k)]\\ &= \mathbb{E}_{S_{-k}}[\textstyle\sum_{i\in A_k}\mathbb{E}_{S_{\cap k}}[I(i\in S_{\cap k})\chi(i,S_{-k}\cup A_k)|S_{-k}]]\\ &= \tfrac{1}{K}\ \mathbb{E}[\textstyle\sum_{i\in A_k}\chi(i,S\cup A_k)]\\ &= \tfrac{1}{K}\ \mathbb{E}[\chi(A_k,S\cup A_k)]\end{aligned} \tag{10}$$

Here, $I(\cdot)$ denotes the indicator function. Apply the above inequality to $\gamma = \frac{1}{2-1/K}$ fraction of each term $\chi(S_{\cap k},S)$ in (9) to obtain:

$$\begin{aligned}\mathbb{E}_S[f(S)] &\geq \mathbb{E}_S\left[\textstyle\sum_{k=1}^{K}(1-\gamma)\,\chi(S_{\cap k},S) + \gamma\frac{1}{K}\chi(A_k,S\cup A_k)\right]\\ &= \mathbb{E}_S\left[\textstyle\sum_{k=1}^{K}\left(\frac{1-\gamma}{K-1}\right)(\sum_{j\neq k}\chi(S_{\cap j},S)) + \gamma\frac{1}{K}\chi(A_k,S\cup A_k)\right]\\ &= \tfrac{1}{(2K-1)}\mathbb{E}_S\left[\textstyle\sum_{k=1}^{K}(\sum_{j\neq k}\chi(S_{\cap j},S)) + \chi(A_k,S\cup A_k)\right]\\ \text{(using cross-monotonicity of } \chi) &\geq \tfrac{1}{(2K-1)}\mathbb{E}_S\left[\textstyle\sum_{k=1}^{K}(\sum_{j\neq k}\chi(S_{\cap j},S\cup A_k)) + \chi(A_k,S\cup A_k)\right]\\ \text{(using } \beta\text{-budget balance)} &\geq \tfrac{1}{(2K-1)\beta}\mathbb{E}_S\left[\textstyle\sum_{k=1}^{K} f(S\cup A_k)\right]\\ \text{(using monotonicity of } f) &\geq \tfrac{1}{(2-1/K)\beta}\left(\tfrac{1}{K}\textstyle\sum_{k=1}^{K} f(A_k)\right)\end{aligned}$$

□

Next, we reduce a general problem instance to an instance satisfying the properties required in Lemma 1. We use the following split operation.

Split: Given a problem instance $(f,V,\{p_i\})$, and integers $\{n_i \geq 1, i\in V\}$, define a new instance $(f',V',\{p'_j\})$ as follows: split each item $i\in V$ into n_i copies $C^i_1, C^i_2, \ldots, C^i_{n_i}$, and assign a marginal probability of $p'_{C^i_k} = \frac{p_i}{n_i}$ to each copy. Let V' denote the new ground set containing all the duplicates. Define the new cost function $f' : 2^{V'} \to \mathbb{R}$ as:

$$f'(S') = f(\Pi(S')),\ \text{for all } S' \subseteq V' \ , \tag{11}$$

where $\Pi(S') \subseteq V$ is the original subset of elements whose duplicates appear in S', i.e. $\Pi(S') = \{i \in V | C^i_k \in S' \text{ for some } k \in \{1,2,\ldots,n_i\}\}$.

The split operation has following properties. Their proofs will be given in Appendix B .

PROPERTY 1 *If $f(S)$ is a non-decreasing function in S that admits a β-budget balanced cross-monotonic cost sharing scheme, then so is f'.*

PROPERTY 2 *If $f(S)$ is non-decreasing in S, then splitting does not change the worst case expected value, that is:*

$$\mathcal{L}(f,V,\{p_i\}) = \mathcal{L}(f',V',\{p'_j\})$$

Property 3 *If $f(S)$ is non-decreasing in S, then splitting can only decrease the expected value over independent distribution:*

$$\mathcal{I}(f, V, \{p_i\}) \geq \mathcal{I}(f', V', \{p'_j\}).$$

The remaining proof tries to use these properties of split operation for reducing any given instance to a "nice" instance so that Lemma 1 can be invoked for proving the correlation gap bound.

Proof of Theorem 1. Suppose that the worst case distribution for instance $(f, V, \{p_i\})$ is not a partition-type distribution. Then, split any element i that appears in two different sets. Simultaneously, split the distribution by assigning probability $\alpha_{S'} = \alpha_{\Pi(S')}$ to the each set S' that contains exactly one copy of i. Repeat until the distribution becomes a partition. Since each new set in the new distribution contains exactly one copy of i, by definition of function f', this splitting does not change the expected function value. By Property 2 of Split operation, the worst case expected values for the two instances (before and after splitting) must be the same, so this partition forms a worst case distribution for the new instance. Then, we further split each element (and simultaneously the distribution) until such that the marginal probability of each new element is $1/K$ for some large enough integer K [4]. This reduces the worst case distribution to a partition $A_1, \dots, A_K$ such that each set A_k has probability $1/K$. By construction, the conditions (b) and (c) of Lemma 1 are satisfied by the reduced instance $(f', V', \{p'_i\})$. Also, by Property 1, the cost function f' obtained by splitting is non-decreasing and admits a β-cost-sharing scheme. Therefore, Lemma 1 can be invoked to bound the correlation gap by 2β for the new instance.

Now, by Property 2 and Property 3 of Split operation, the correlation gap can only become larger on splitting. Therefore, the correlation gap for the new instance bounds the correlation gap for the original instance. This completes the proof. □

5. Supermodular functions

In the end, we directly consider the correlation robust model for cost functions $f(x, S)$ which are supermodular in S. As shown in Section 2, the correlation gap for these cost functions can be exponentially high, so independent distribution does not give a good approximation to the worst case distribution. However, it is easy to characterize the worst case distribution and directly solve the correlation robust model in this case.

Lemma 2. *Given that function $f: 2^V \to \mathbb{R}$ is supermodular, the worst case distribution over S with marginals $\{p_i\}_{i \in V}$ has the following closed form*

$$\Pr(S) = \begin{cases} p_n & \text{if } S = S_n \\ p_i - p_{i+1} & \text{if } S = S_i, 1 \leq i \leq n-1 \\ 1 - p_1 & \text{if } S = \emptyset \\ 0 & o.w. \end{cases}$$

where $n = |V|$; i is the i^{th} member of V and S_i is the set of first i members of V, both with respect to a specific ordering over V such that $p_1 \geq \ldots \geq p_n$.

The lemma is simple to prove, a proof appears in appendix E . Lemma 2 implies following corollary for solving the correlation robust problem.

Corollary 7. *For cost functions $f(x,S)$ that are supermodular in S for any feasible x, the correlation robust problem with marginals $p_1 \geq p_2 \ldots \geq p_n$ is simply formulated as:*

$$\min_{x\in C} p_n f(x,S^n) + \sum_{i=1}^{n-1}(p_i - p_{i+1}) f(x,S^i) + (1-p_1) f(x,\phi)$$

Thus, if $f(x,S)$ is convex in x and C is a convex set, then it is a convex optimization problem and can be solved efficiently.

6. Discussion

We have focused our discussion on the stochastic optimization problems where each random variable takes value either 1 or 0. This typically models stochastic counterparts of discrete optimization problems. Our motivation is to simplify the individual statistics and focus on modeling the correlation information. Our bounds on POC characterize the suitability of ignoring correlations for these problems. We classify the functions as supermodular functions, sub-additive functions, functions with β-budget balance cost sharing, and submodular functions. We estimate the corresponding POC for problems associated with these functions are respectively $\Omega(2^n)$, $\Omega(\sqrt{n})$, 2β, $\frac{e}{e-1}$, which reflects a gradual change of POC with the properties of these functions. Possible progress remains in studying POC for continuous distributions case.

The reader may note that we have not compared our solutions with that obtained from other solution models for SP e.g. SAA. The reason is that we have completely different assumptions and motivations: we assume that the distribution information in unknown and we are looking for a conservative strategy to minimize the worst-case expected cost. The approximation ratio (POC) we obtained are also in terms of our DRSP model, which can not be compared with the approximation ratio for the corresponding SP problems.

Acknowledgements The authors would like to thank Ashish Goel and Mukund Sundarajan for many useful insights on the problem.

Endnotes

1. Here e is the mathematical constant $e = 2.71828\ldots$
2. The construction of this example is inspired by the example presented in Immorlica et al. (2008) for lower bounding the budget-balance factor of any cross-monotonic cost-sharing scheme for metric uncapacitated facility location.
3. A more general formulation of this problem that is often considered in the literature allows non-identical utility functions for various players.
4. Such an integer K can always be reached assuming p_is are rational.

References

Agrawal, Shipra, Yichuan Ding, Amin Saberi, Yinyu Ye. 2010. Correlation robust stochastic optimization. *SODA '10: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*.

Ahmed, Shabbir, Alexander Shapiro, Er Shapiro. 2002. The sample average approximation method for stochastic programs with integer recourse. *SIAM Journal of Optimization* **12** 479–502.

Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Mathematics of Operations Research* **23**(4) 769–805.

Ben-Tal, Aharon. 2001. Robust optimization - methodology and applications. *Mathematical Programming* **92**(3) 453–480.

Ben-tal, Aharon, Arkadi Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming* **88** 411–424.

Bertsimas, Dimitris, Melvyn Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

Bikhchandani, Sushil, John W Mamer. 1997. Competitive equilibrium in an exchange economy with indivisibilities. *Journal of Economic Theory* **74**(2) 385–413.

Calinescu, Gruia, Chandra Chekuri, Martin Pál, Jan Vondrák. 2007. Maximizing a submodular set function subject to a matroid constraint (extended abstract). *IPCO*. 182–196.

Charikar, Moses, Chandra Chekuri, Martin Pál. 2005. Sampling bounds for stochastic optimization. *APPROX-RANDOM*. 257–269.

Chen, Xin, Melvyn Sim, Peng Sun. 2007. A Robust Optimization Perspective on Stochastic Programming. *Operations Research* **55**(6) 1058–1071.

Delage, Erick, Yinyu Ye. 2008. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *to appear in Operations Research* .

Edmonds, Jack. 2003. Submodular functions, matroids, and certain polyhedra. *Combinatorial optimization - Eureka, you shrink!*. Springer-Verlag New York, Inc., 11–26.

Goh, Joel, Melvyn Sim. 2009. Distributionally robust optimization and its tractable approximations. *accepted in Operations Research* .

Gupta, Anupam, Martin Pal, R. Ravi, Amitabh Sinha. 2004. Boosted sampling: Approximation algorithms for stochastic optimization. *In Proceedings of the 36th Annual ACM Symposium on Theory of Computing*. 417–426.

Immorlica, Nicole, Mohammad Mahdian, Vahab S. Mirrokni. 2008. Limitations of cross-monotonic cost-sharing schemes. *ACM Transactions on Algorithms* **4**(2) 1–25.

Kimber, A. C. 1983. A note on poisson maxima. *Probability Theory and Related Fields* **63** 551–552.

Kleinberg, Jon, Yuval Rabani, Eva Tardos. 1997. Allocating bandwidth for bursty connections. *SIAM J. Comput* **30** 2000.

Könemann, Jochen, Stefano Leonardi, Guido Schäfer. 2005. A group-strategyproof mechanism for steiner forests. *SODA*. 612–619.

Möhring, Rolf H., Andreas S. Schulz, Marc Uetz. 1999. Approximation in stochastic scheduling: the power of lp-based priority policies. *J. ACM* **46**(6) 924–942.

Nisan, Noam, Tim Roughgarden, Eva Tardos, Vijay V. Vazirani. 2007. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA.

Pál, Martin, Éva Tardos. 2003. Group strategyproof mechanisms via primal-dual algorithms. *FOCS*. 584–593.

Popescu, Ioana. 2007. Robust mean-covariance solutions for stochastic optimization. *Operations Research* **55**(1) 98–112.

Roughgarden, Tim, Mukund Sundararajan. 2006. New trade-offs in cost-sharing mechanisms. *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*. ACM, New York, NY, USA, 79–88.

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

Scarf, Herbert E. 1958. A min-max solution of an inventory problem. *Studies in The Mathematical Theory of Inventory and Production* 201–209.

Swamy, Chaitanya, David B. Shmoys. 2005. Sampling-based approximation algorithms for multi-stage stochastic. *FOCS '05: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*. 357–366.

Vondrak, Jan. 2008. Optimal approximation for the submodular welfare problem in the value oracle model. *STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing*. 67–74.

Appendix A: Maximum of Poisson Random Variables

In this section, we show that the expected value of the maximum of a set of M independent identically distributed poisson random variables can be bounded as $O(\log M/\log\log M)$ for large M.

Let λ denote the mean, and F denote the distribution of i.i.d. poisson variables X_i. Define $G = 1 - F$. Also define continuous extension of G:

$$G_c(x) = \exp(-\lambda)\sum_{j=1}^{\infty} \lambda^{(x+j)}/\Gamma(x+j+1)$$

Note that $G(k) = G_c(k)$ for any non-negative integer k. Let $\{A_k\}_{k=1}^{\infty}$ is defined by $G_c(A_k) = 1/k$. Define continuous function $L(x) = \log(x)/\log\log(x)$. Then, in Kimber (1983), it is shown that for large k, $A_k \sim L(k)$.

We use these asymptotic results to derive a bound on expectation of $Z = \max_{i=1,\dots,M} X_i$ for large M.

$$\begin{aligned}
\mathbb{E}[Z] &= \sum_{k=0}^{\infty} \Pr(Z > k) \\
&= \sum_{k=0}^{\lceil L(M^2)\rceil} \Pr(Z > k) + \sum_{k=\lceil L(M^2)\rceil+1}^{\infty} \Pr(Z > k) \\
&\le L(M^2) + 1 + \int_{x=L(M^2)}^{\infty} \Pr(Z > x)dx \qquad (12)
\end{aligned}$$

Next, we show that the integral term on the right hand side is bounded by a constant for large M. Substituting $x = L(y)$ in the integration on the right hand side, we get

$$\begin{aligned}
&\int_{x=L(M^2)}^{\infty} \Pr(Z > x)dx \\
&= \int_{L(y)=L(M^2)}^{\infty} \Pr(Z > L(y))L'(y)dy \\
&\le \int_{y=M^2}^{\infty} \Pr(Z > L(y))\frac{1}{y}dy
\end{aligned}$$

$L'(y)$ denotes the derivative of function $L(y)$. The last step follows because $L'(y) \le \frac{1}{y}$ for large enough y (i.e. if $\log\log y \ge 1$). Further, since $\frac{\Pr(Z>L(k))}{k}$ is a decreasing function in k, it follows that:

$$\int_{y=M^2}^{\infty} \frac{\Pr(Z > L(y))}{y}dy \le \sum_{k=M^2}^{\infty} \frac{\Pr(Z > L(k))}{k}$$

Now, for large k, $L(k) \sim A_k$, and

$$\Pr(Z > A_k) \le 1 - (1 - G_c(A_k))^M = 1 - \left(1 - \frac{1}{k}\right)^M$$

Therefore, for large M,

$$\begin{aligned}
\sum_{k=M^2}^{\infty} \frac{\Pr(Z > L(k))}{k} &\le \sum_{k=M^2}^{\infty} \frac{1}{k} - \frac{1}{k}\left(1 - \frac{1}{k}\right)^M \\
&\le \sum_{k=M^2}^{\infty} \frac{2M}{k^2} \\
&\le 1
\end{aligned}$$

This proves that the integral term on the right hand side of (12) is bounded by a constant, and thus, for large M:

$$\mathbb{E}[Z] \leq L(M^2) + 2 = O(\log M / \log\log M)$$

.

Appendix B: Properties of Split Operation

Property 1. If $f(S)$ is a non-decreasing function in S that admits a β-cost sharing scheme, then so is f'.

Proof. Monotonicity holds since for any $S' \subseteq T' \subseteq V'$, $\Pi(S') \subseteq \Pi(T')$:

$$f'(S') = f(\Pi(S')) \leq f(\Pi(T')) = f'(T')$$

Given a β-cost-sharing scheme χ for f, construct cost-sharing scheme χ' for f' as follows. Consider an arbitrary but fixed ordering on elements of V. Cost-share χ' coincides with the original scheme χ for the sets without duplicates, but for a set with duplicates, it assigns the cost-share solely to the copy with smallest index (as per the ordering). That is, any $S' \subseteq V'$, and item C^i_j (j-th copy of item i) in S', allocate cost-shares as follows:

$$\chi'(C^i_k, S') = \begin{cases} \chi(i, S), & k = \min\{h: \ C^i_h \in S'\}, \\ 0, & \text{o.w.} \end{cases} \tag{13}$$

where $S = \Pi(S')$, and *min* computes the lowest index with respect to the fixed ordering on elements. It is easy to see that the property of *β-budget-balance* carries through to the new cost sharing scheme. For cross-monotonicity, consider $S' \subseteq T'$. Now, for any $i' \in S'$, if i' is not a lowest indexed copy in T', then $\chi(i', T') = 0$, so that the condition is automatically satisfied. Let i' is one of the lowest indexed copies in T', then it must have been a lowest indexed copy in S', since S' is a subset of T'. Then, by cross-monotonicity of χ:

$$\chi(i', T') = \chi(i, T) \leq \chi(i, S) = \chi(i', S')$$

where $S = \Pi(S'), T = \Pi(T')$. □

Property 2. If the cost function $f(\cdot)$ is non-decreasing in S, then the splitting procedure does not change the worst-case expected value. That is:

$$\mathcal{L}(f, V, \{p_i\}) = \mathcal{L}(f', V', \{p'_j\})$$

Proof. For any fixed x, the worst case expected cost is the optimal value of following linear program, where $\{\alpha_S\}_{S \subseteq V}$ represents a distribution over subsets of set V:

$$\begin{array}{rl} \mathcal{L}(f, V, \{p_i\}) = \max_\alpha & \sum_S \alpha_S f(x, S) \\ \text{s.t.} & \sum_{S:\ i \in S} \alpha_S = p_i, \ \forall i \in V \\ & \sum_S \alpha_S = 1 \\ & \alpha_S \geq 0, \ \forall S \subseteq V. \end{array} \tag{14}$$

Suppose item 1 is split into n_1 pieces, and each piece is assigned a probability $\frac{p_1}{n_1}$. Let $\{\alpha_S\}$ denote the optimal solution for the instance $(f, V, \{p_i\})$, then we can construct a solution for the new instance $(f', V', \{p'_j\})$ which has the same objective value by assigning non-zero probabilities to only those sets which have no duplicates.

$$\forall S' \subseteq V', \ \alpha'_{S'} = \begin{cases} \alpha_{\Pi(S')}, \text{ if } S' \text{ contains no copies of item 1} \\ \frac{p_1}{n_1}\alpha_{\Pi(S')}, \text{ if } S' \text{ contains exactly one copy of item 1} \\ 0, \text{ otherwise} \end{cases}$$

One can verify that $\{\alpha'_{S'}\}$ is a feasible distribution (i.e., feasible to the linear program (14)) for the new instance $(f', V', \{p'_j\})$, and has the same objective value as $\mathcal{L}(f, V, \{p_i\})$. Hence, $\mathcal{L}(f, V, \{p_i\}) \le \mathcal{L}(f', V', \{p'_j\})$.

For the other direction, consider an optimal solution $\{\alpha'_{S'}\}$ of the new instance. It is easy to see that there exists an optimal solution $\{\alpha'_{S'}\}$ that $\alpha'_{S'} = 0$ for all S' that contain more than one copy of item 1. To see this, assume for contradiction that some set with non-zero probability has two copies of item 1. By definition of f', removing one copy will not decrease the function value. Then, because of monotonicity of f', we can move out one copy to another set T that has no copy of item 1. Such T always exists since the probabilities of copies of item 1 must sum up to $p_1 \le 1$. So, we can assume that in the optimal solution $\alpha'_{S'} = 0$ for any set S' containing more than one copy. Thus, we can set $\alpha_S = \sum_{S':\Pi(S')=S} \alpha'_{S'}$ where S is the corresponding original set for any $S \subseteq V$. That forms a feasible solution for original instance with same objective value as $\mathcal{L}(f', V', \{p'_j\})$. We can apply the argument recursively for all the items to prove the lemma. $\square$

Next, we prove that the expected cost under independent Bernoulli distribution can only decrease by the split operation.

Property 3. If $f(\cdot)$ is non-decreasing, then after splitting

$$\mathcal{I}(f', V', \{p'_j\}) \le \mathcal{I}(f, V, \{p_i\}).$$

Proof. Let $(f', V', \{p'_j\})$ denote the new instance by splitting item 1 into n_1 pieces. Denote

$$\Lambda := \{S' \subseteq V' | S' \text{ contains at least one copy of } 1\},$$

and denote $\pi = \Pr(S' \in \Lambda)$. Consider the expected cost under independent Bernoulli distribution, by independence,

$$\begin{aligned}
\mathcal{I}(f', V', \{p'_j\}) &= \mathbb{E}_{S'}\left[f'(S')\ I(S' \in \Lambda)\right] + \mathbb{E}_{S'}\left[f'(S')\ I(S' \notin \Lambda)\right] \\
&= \pi\ \mathbb{E}_{S \subseteq V \setminus \{1\}}[f(S \cup \{1\})] + (1-\pi)\ \mathbb{E}_{S \subseteq V \setminus \{1\}}[f(S)] \\
&\le p_1\ \mathbb{E}_{S \subseteq V \setminus \{1\}}[f(S \cup \{1\})] + (1-p_1)\ \mathbb{E}_{S \subseteq V \setminus \{1\}}[f(S)] \\
&= \mathcal{I}(f, V, \{p_i\}).
\end{aligned}$$

The second last inequality holds because $\pi = 1 - (1 - \frac{p_1}{n_1})^{n_1} \le p_1$, and $f(S) \le f(S \cup \{1\})$ by monotonicity. $\square$

Appendix C: $\frac{11}{12}$ Integrality gap for SWM with identical submodular valuations

Let $V = \{1, 2, 3, 4, 5, 6\}$, $K = 3$, and construct a monotone submodular value function as

$$f(S) = \begin{cases} 0 \text{ if } S = \emptyset \\ 2 \text{ if } |S| = 1 \\ 3 \text{ if } |S \cap \{1,2,3\}| = 1 \text{ and } |S \cap \{4,5,6\}| = 1 \\ 4 \text{ if } |S \cap \{1,2,3\}| \ge 2 \text{ or } |S \cap \{4,5,6\}| \ge 2 \end{cases}$$

Then the optimal fractional solution to the LP relaxation of (8) is given by

$$\alpha_{\{1,2\}} = \alpha_{\{2,3\}} = \alpha_{\{1,3\}} = 0.5, \quad \alpha_{\{4,5\}} = \alpha_{\{5,6\}} = \alpha_{\{4,6\}} = 0.5,$$

with an optimal value 12; but the optimal integer solution will have an optimal value 11. So there is an 11/12 integrality gap.

Appendix D: Proof of Theorem 2

The proof of this theorem is along the similar lines as Theorem 1. We assume that function $f(S)$ has a cost-sharing scheme as a function $\chi(i, S, \sigma_S)$ that, for each element $i \in S$ and ordering σ_S on S, specifies the share of i in S. And, χ satisfies three properties of budget-balance, weak cross-monotonicity and weak summability stated as follows:

1. *β-budget balance:* For all S, and orderings σ_S on S:

$$f(S) \geq \sum_{i=1}^{|S|} \chi(i, S, \sigma_S) \geq \frac{f(S)}{\beta}$$

2. *Cross-monotonicity:* For all $i \in S$, $S \subseteq T$, $\sigma_S \subseteq \sigma_T$:

$$\chi(i, S, \sigma_S) \geq \chi(i, T, \sigma_T)$$

Here , $\sigma_S \subseteq \sigma_T$ means that the ordering σ_S is a restriction of ordering σ_T to subset S.

3. *Weak η-summability:* For all S, and orderings σ_S:

$$\sum_{\ell=1}^{|S|} \chi(i_\ell, S_\ell, \sigma_{S_\ell}) \leq \eta f(S)$$

where i_ℓ is the ℓ^{th} element and S_ℓ is the set of the first ℓ members of S according to ordering σ_S. And, σ_{S_ℓ} is the restriction of σ_S on S_ℓ.
Note that these conditions are weaker than those stated earlier in Section 3. We call such a cost-sharing scheme an (η, β)-cost-sharing scheme. Given, these conditions, we will first prove the result for a simpler case.

Lemma 3. *For instances $(f, V, \{p_i\})$ such that (a) $f(S)$ is non-decreasing and admits an (η, β)-cost-sharing scheme (b) marginal probabilities p_i are all equal to $1/K$ for some integer K, and (c) the worst case distribution is a K-partition-type distribution, the correlation gap is bounded as:*

$$\frac{\mathcal{L}(f, V, \{1/K\})}{\mathcal{I}(f, V, \{1/K\})} \leq \eta\beta \frac{e}{(e-1)}$$

Proof. Let the optimal K-partition corresponding to the worst case distribution is $\{A_1, A_2, \ldots, A_K\}$. Assume w.l.o.g that $f(A_1) \geq f(A_2) \geq \ldots \geq f(A_K)$. Fix an order σ on elements of V such that for all k, the elements in A_k come before A_{k-1}. For every set S, let σ_S be the restriction of ordering σ on set elements of set S. Let χ is the (η, β) cost-sharing scheme for function f, as per the assumptions of the lemma. Then by weak η-summability of χ:

$$\mathcal{I}(f, V, \{1/K\}) = \mathbb{E}_{S\subseteq V}[f(S)] \geq \tfrac{1}{\eta}\, \mathbb{E}_{S\subseteq V}\left[\textstyle\sum_{l=1}^{|S|} \chi(i_l, S_l, \sigma_{S_l})\right] \tag{15}$$

where the expected value is taken over independent distribution.

Denote $\phi(V) := \mathbb{E}_{S\subseteq V}\left[\sum_{l=1}^{|S|} \chi(i_l, S_l, \sigma_{S_l})\right]$. Let $p = 1/K$. We will show that

$$\phi(V) \geq (1-p)\phi(V \backslash A_1) + \frac{1}{\beta} f(A_1)$$

Recursively using this inequality will prove the result. To prove this inequality, denote $S_{-1} = S \cap (V \backslash A_1)$, $S_1 = S \cap A_1$, for any $S \subseteq V$. Since elements in A_1 come after the elements in $V \backslash A_1$ in ordering σ_S, note that for any $\ell \leq |S_{-1}|$, $S_\ell \subseteq S_{-1}$, and for $\ell > |S_{-1}|$, $i_\ell \in S_1$.

$$\phi(V) = \mathbb{E}_S\left[\textstyle\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_l, \sigma_{S_l})\right] \; + \; \mathbb{E}_S\left[\textstyle\sum_{l=|S_{-1}|+1}^{|S|} \chi(i_l, S_l, \sigma_{S_l})\right] \tag{16}$$

Since $S_\ell \subseteq S \cup A_1$, using cross-monotonicity of χ, the second term above can be bounded as:

$$\mathbb{E}_S[\sum_{l=|S_{-1}|+1}^{|S|} \chi(i_l, S_l, \sigma_{S_l})] \ \geq \ \mathbb{E}_S[\sum_{l=|S_{-1}|+1}^{|S|} \chi(i_l, S \cup A_1, \sigma_{S\cup A_1})] \tag{17}$$

Because S_{-1} and S_1 are mutually independent, for any fixed S_{-1}, each $i \in A_1$ will have the same conditional probability $p = 1/K$ of appearing in S_1. Therefore,

$$\begin{aligned} \mathbb{E}_S\big[\sum_{l=|S_{-1}|+1}^{|S|} \chi(i_l, S \cup A_1, \sigma_{S\cup A_1})\big] &= \mathbb{E}_{S_{-1}}\big[\mathbb{E}_{S_1}[\sum_{l=|S_{-1}|+1}^{|S|} \chi(i_l, S_{-1} \cup A_1, \sigma_{S_{-1}\cup A_1})|S_{-1}]\big] \\ &= \ p\ \mathbb{E}_{S_{-1}}[\sum_{i\in A_1} \chi(i, S_{-1} \cup A_1, \sigma_{S_{-1}\cup A_1})] \end{aligned} \tag{18}$$

Again, using independence and cross-monotonicity, analyze the first term in the right hand side of (16),

$$\begin{aligned} \mathbb{E}_S[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_l, \sigma_{S_l})] &= \mathbb{E}_{S_{-1}}[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_l, \sigma_{S_l})] \\ &\geq (1-p)\ \mathbb{E}_{S_{-1}}[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_l, \sigma_{S_l})] + p\ \mathbb{E}_{S_{-1}}[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_{-1} \cup A_1, \sigma_{S_{-1}\cup A_1})] \\ &= (1-p)\ \phi(V\backslash A_1) + p\ \mathbb{E}_{S_{-1}}[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_{-1} \cup A_1, \sigma_{S_{-1}\cup A_1})] \end{aligned} \tag{19}$$

Based on (16), (18) and (19), and the fact that the cost-sharing scheme χ is β-budget balanced, we deduce

$$\begin{aligned} \phi(V) &= (1-p)\ \phi(V\backslash A_1) + p\ \mathbb{E}_{S_{-1}}[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_{-1} \cup A_1, \sigma_{S_{-1}\cup A_1}) + \sum_{i\in A_1} \chi(i, S_{-1} \cup A_1, \sigma_{S_{-1}\cup A_1})] \\ &\geq (1-p)\ \phi(V\backslash A_1) + \tfrac{1}{\beta} p\ \mathbb{E}_{S_{-1}}[f(S_{-1} \cup A_1)] \\ &\geq (1-p)\ \phi(V\backslash A_1) + \tfrac{1}{\beta} p\ f(A_1), \end{aligned} \tag{20}$$

The last inequality follows from monotonicity of f. Expanding the above recursive inequality for A_2, ..., A_K, we get

$$\phi(V) \geq \frac{1}{\beta} p \sum_{k=1}^{K} (1-p)^{k-1} f(A_k), \tag{21}$$

Since $f(A_k)$ is decreasing in k, and $p = 1/K$ by simple arithmetic one can show

$$\begin{aligned} \phi(V) &\geq \tfrac{1}{\beta} \cdot \sum_{k=1}^{K} p f(A_k) \cdot \tfrac{(\sum_{k=1}^{K}(1-p)^{k-1})}{K} \\ &\geq \tfrac{1}{\beta} \cdot (1 - \tfrac{1}{e}) \cdot \sum_{k=1}^{K} p f(A_k) \end{aligned}$$

By definition of $\phi(V)$, this gives:

$$\mathcal{I}(f, V, \{1/K\}) \geq \frac{1}{\eta\beta}\left(1 - \frac{1}{e}\right) \mathcal{L}(f, V, \{1/K\}).$$

□

Now, we can prove Theorem 2, by reducing any general problem instance to an instance satisfying the properties required in Lemma 3 using Split operation.

Proof of Theorem 2. As in the proof of Theorem 1, we can reduce any instance $\{f, V, \{p_i\}\}$ to an instance $\{f', V', \{1/K\}\}$ by splitting so that the worst case distribution for new instance is K-partition type, and the correlation gap of the new instance is atleast as large as the original instance. Also, f' is non-decreasing, if f is. The only property we need to prove is that given the original (η, β) cost-sharing method χ for f, there exists a (η, β)-cost-sharing method χ' for f'.

Given cost-sharing scheme χ, construct χ' as follows: Cost-share χ' coincides with the original scheme χ for the sets without duplicates, but for a set with duplicates, it assigns the cost-share solely to the copy with smallest index (as per the input ordering). That is, any $S' \subseteq V'$, ordering $\sigma'_{S'}$, and item C^i_j (j-th copy of item i) in S', allocate cost-shares as follows:

$$\chi'(C^i_j, S', \sigma'_{S'}) = \begin{cases} \chi(i, S, \sigma_S), & j = \min\{h: \ C^i_h \in S'\}, \\ 0, & \text{o.w.} \end{cases} \tag{22}$$

where $S = \Pi(S')$, σ_S is the ordering of lowest index copies in $\sigma'_{S'}$, and *min* is taken with respect to the ordering $\sigma'_{S'}$.

It is easy to see that the property of *β-budget-balance* carries through to the new cost sharing scheme. *Weak η-summability* holds since

$$\sum_{\ell=1}^{|S'|} \chi'(i'_\ell, S'_\ell, \sigma_{S'\ell}) = \sum_{j=1}^{|S|} \chi(i_j, S_j, \sigma_{S_j}) \le \eta f(S) = \eta f'(S')$$

where $S = \Pi(S')$, σ_S is the ordering of lowest index copies in $\sigma'_{S'}$. Also, the scheme is cross monotone in following weaker sense. χ' is cross-monotone for any $S' \subseteq T', \sigma_{S'} \subseteq \sigma_{T'}$ such that $\sigma_{S'}, \sigma_{T'}$ respect the partial order $A_K, \ldots, A_1$ of elements, and S' is a *partial-prefix* of T', that is, for some $k \in \{1, \ldots, K\}$, $S' \subseteq A_K \cup \cdots \cup A_k$, and $T' \backslash S' \subseteq A_k \cup \cdots \cup A_1$. To prove this weaker cross-monotonicity, consider $S' \subseteq T', \sigma_{S'} \subseteq \sigma_{T'}$ such that S' is a "partial prefix" of T'. Now, for any $i' \in S'$, if i' is not a lowest indexed copy in T', then $\chi(i', T', \sigma'_{T'}) = 0$, so that the condition is automatically satisfied. Let i' is one of the lowest indexed copies in T', then it must have been a lowest indexed copy in S', since S' is a subset of T', and $\sigma_{S'} \subseteq \sigma_{T'}$. Thus,

$$\chi(i', T', \sigma'_{T'}) = \chi(i, T, \sigma_T) \le \chi(i, S, \sigma_S) = \chi(i', S', \sigma'_{S'})$$

where $S = \Pi(S'), T = \Pi(T')$, σ_S, σ_T are the orderings of lowest indexed copies in S', T' respectively. Note that the inequality in above uses cross-monotonicity of χ, which is satisfied only if in addition to $S \subseteq T$, we have that $\sigma_S \subseteq \sigma_T$. That is, if the ordering of elements of S is same in σ_S and σ_T. We show that this is true given the assumption that $\sigma_{S'}, \sigma_{T'}$ respect the partial ordering $A_K, \ldots, A_1$, and S' is a "partial prefix" of T'. To see this, observe that the splitting was performed in a manner so that atmost one copy of any element appears in each A_k. So, among the items $T' \backslash S'$, any copy of an element of S can occur only in $T' \cap \{A_{k-1}, \ldots, A_1\}$. Since $S' \subseteq A_K \cup \cdots \cup A_k$, this means that for any element $i \in S$, its copies in $T' \backslash S'$ can occur only later in the ordering than those in S'. So, elements of $T' \backslash S'$ cannot alter the order of lowest indexed copies of elements in S'. This proves that $\sigma_S \subseteq \sigma_T$.

Thus, all the conditions in Lemma 1 are satisfied by the new instance except for the cross-monotonicity. The weaker cross-monotonicity that the new instance satisfies is actually sufficient to prove Lemma 1. To see this, observe that cross monotonicity is used only in Equation 17 and 19, and at both of these places, the required prefix condition is satisfied. Thus, Lemma 1 can be invoked to bound the correlation gap for the new instance, thereby completing the proof. □

Appendix E: Proof of Lemma 2

For any fixed x, the worst case expected cost is the optimal value of following linear program, where $\{\alpha_S\}_{S \subseteq V}$ represents a distribution over subsets of set V:

$$\begin{aligned} \mathcal{L}(f, V, \{p_i\}) = \max_\alpha \ & \textstyle\sum_S \alpha_S f(x, S) \\ \text{s.t.} \ & \textstyle\sum_{S:\ i \in S} \alpha_S = p_i, \ \forall i \in V \\ & \textstyle\sum_S \alpha_S = 1 \\ & \alpha_S \ge 0, \ \forall S \subseteq V. \end{aligned} \tag{23}$$

It is easy to verify that

$$\alpha^* = \begin{cases} p_n & \text{if } S = S_n \\ (p_i - p_{i+1}) & \text{if } S = S_i, 1 \le i \le n-1 \\ 1 - p_1 & \text{if } S = \emptyset \\ 0 & o.w. \end{cases}$$

is a feasible solution to (23). Next we show that it is actually the optimal solution. The dual of linear program (23) is:

$$\begin{array}{ll} \min_{\gamma,\lambda} & \gamma + p^T \lambda \\ \text{s.t.} & f(S) - \sum_{i \in S} \lambda_i \le \gamma, \ \forall S. \end{array} \tag{24}$$

Consider the problem in λ for a given value of γ. This problem is to minimize a linear function with positive coefficients (p_i) over the supermodular polyhedron (of supermodular function $f(S) - \gamma$). Minimizing a linear function over a supermodular polyhedron can be solved by a greedy procedure Edmonds (2003), with the optimal value given by $\sum_{i=1}^{n} p_i(f(S_i) - f(S_{i-1}))$.

Then (24) can be rewritten as

$$\begin{array}{ll} \min_{\gamma} & \gamma + p_n f(S^n) + \displaystyle\sum_{i=1}^{n-1} (p_i - p_{i+1}) f(S^i) - p_1 f(\emptyset) \\ \text{s.t.} & f(\emptyset) \le \gamma. \end{array}$$

The optimal solution for above is $\gamma = f(\emptyset)$, therefore optimal value:

$$\begin{array}{l} p_n f(S^n) + \sum_{i=1}^{n-1} (p_i - p_{i+1}) f(S^i) + (1 - p_1) f(\emptyset) \\ = \sum_S \alpha^*_S f(S) \end{array}$$

This proves the lemma.