Probabilistic Set Covering with Correlations

Shabbir Ahmed and Dimitri J. Papageorgiou

School of Industrial & Systems Engineering
Georgia Institute of Technology, 765 Ferst Drive, Atlanta, GA 30332

July 30, 2011

Abstract

We consider a probabilistic set covering problem where there is uncertainty regarding whether a selected set can cover an item, and the objective is to determine a minimum-cost combination of sets so that each item is covered with a pre-specified probability. To date, literature on this problem has focused on the special case in which uncertainties are independent. In this paper, we formulate conservative deterministic mixed-integer programming models for probabilistic set covering problems with correlated uncertainties. By exploiting the supermodularity properties of the probabilistic covering constraints and analyzing their polyhedral structure, we develop strong valid inequalities to strengthen the formulations. Computational results illustrate that our modeling approach outperforms formulations in which correlations are ignored and that our algorithms can significantly reduce overall computation time.

Keywords: ambiguous chance constraint, integer programming, set covering, stochastic programming, supermodularity.

1 Introduction

We consider the following probabilistic set covering (PSC) problem

$$\begin{aligned} \min_x \quad & \sum_{j \in N} c_j x_j \\ \text{s.t.} \quad & \mathbb{P}\left[\sum_{j \in N} \tilde{a}_{ij} x_j \geq 1\right] \geq 1 - \epsilon_i \quad \forall\, i \in M \\ & x \in S \subseteq \{0,1\}^n \ , \end{aligned} \tag{1}$$

where $N := \{1, \ldots, n\}$ is a collection of columns (or sets), $M := \{1, \ldots, m\}$ is a collection of rows (or items), $\tilde{a}_{ij}$ is a Bernoulli random variable indicating whether row $i \in M$ can be covered by column $j \in N$, and $\epsilon_i \in (0, 1)$ is a pre-specified allowed failure probability for row i. The set S represents some deterministic side constraints on the selected columns x. Problem (1) seeks a minimum-cost collection of columns that covers each row i with a probability of at least $1 - \epsilon_i$. Alternatively, we can also consider a max-min coverage probability problem, i.e., maximizing the minimum coverage probability subject to constraints on the selected columns:

$$\begin{aligned} \max_{w,x} \quad & w \\ \text{s.t.} \quad & \mathbb{P}\left[\sum_{j \in N} \tilde{a}_{ij} x_j \geq 1\right] \geq w \quad \forall\, i \in M \\ & x \in S \subseteq \{0,1\}^n \ . \end{aligned} \tag{2}$$

A similar formulation results from maximizing the sum of the coverage probabilities.

As an example of a probabilistic set covering problem of the above form, consider a surveillance problem in which we wish to detect (or "cover") m targets by placing sensors at n potential sensor sites. In the deterministic setting, we let the parameter a_{ij} take value 1 if target i can be detected by a sensor at location j, and 0 otherwise, for each target-sensor location pair. In our stochastic setting, we assume that there is a nontrivial probability that a sensor deployed at position j will be able to detect target i. Letting x_j take value 1 to denote that a sensor is placed at location j, and 0 otherwise, Formulation (1) becomes the problem of minimizing the cost of placing sensors at sites subject to independent chance-constraints that each target can be detected (covered) with some pre-specified probability.

The probabilistic set covering problem (1) falls in the class of chance-constrained stochastic programs, which has undergone extensive investigation. Prékopa [19] and Shapiro et al. [21] provide a thorough review of such problems. From a computational perspective, chance-constrained problems are typically challenging on two fronts: First, given a candidate solution x, verifying that x is feasible can be computationally demanding. Second, the feasible region defined by a chance constraint is generally not convex, which implies that even if checking feasibility is easy, finding a provable optimal solution may be elusive.

When both of these difficulties are present, as is the case in the most general model of uncertainty considered in this paper, there are two prevailing tactics - both approximation techniques - for identifying high quality solutions: tractable conservative approximations and sample average approximations. In the first approach, the strategy is to formulate a convex optimization problem, which can be solved efficiently (hence, is tractable), and produces a solution which has a high probability of being feasible to the original problem (and is, therefore, a "safe" or conservative approximation). Examples include the Bernstein approximation scheme of Nemirovski and Shapiro [16] and Pinter [17] and robust optimization [4]. All of these approaches fully exploit the convexity of the resulting feasible region, a property that is absent in our discrete setting.

The sample average approximation method, sometimes called the scenario approach, is an alternative strategy in which one attempts to solve an approximation problem based on an independent Monte Carlo sample of the random data [12]. The scenario approximation method of Calafiore and Campi [7] and extended by Nemirovski and Shapiro [15] requires that all of the constraints corresponding to the sample taken must be satisfied, whereas the sample average approximation of Luedtke and Ahmed [14] does not require that all sampled constraint sets be satisfied. Instead, the constraints which will be satisfied can be chosen optimally. The main advantage of the scenario approach is its generality as it does not require knowledge of the distribution of the random parameters. Since one only needs to be able to sample from this distribution, the scenario approach has no trouble handling correlated data. At the same time, no such method exploits the availability of correlation information, which we do in this paper.

As far as set covering is concerned, the *deterministic* set covering problem has been extensively studied in the OR literature due to its appearance as a fundamental building block in numerous discrete optimization models including facility location, vehicle routing, crew scheduling, and many others. There has been growing interest in stochastic variants of the set covering problem. Beraldi and Ruszczyński [6] and Saxena et al. [20] studied a variant of the probabilistic set covering problem in which there is a single joint chance constraint and randomness appears only in the right hand side, i.e., $\min\{cx : \mathbb{P}[Ax \geq \xi] \geq 1-\epsilon, x \in \{0,1\}^n\}$, where ξ is a random 0-1 vector whose components may be correlated. Haight et al. [11] formulated a variant of (1) in which all constraint coefficients $\tilde{a}_{ij}$ are assumed to be independent. Fischetti and Monaci [9] investigated what they call the Uncertain Probabilistic Set Covering Problem in which the columns $\tilde{a}_j$ are assumed to be independent. Goemans and Vondrak [10] consider stochastic set covering models with independent data in which decisions may or may not be adaptive, i.e., the selection of sets to cover items is made adaptively over time or in one fell swoop (non-adaptively). In contrast to the methods above, Beraldi and Bruni [5] use a scenario approach to perform computational experiments on a probabilistic set covering problem in which randomness appears in the constraint matrix and the right hand side. The randomness is explicitly given through a finite set of scenarios and, thus, data may be correlated.

With the exception of the scenario approach of Beraldi and Bruni [5], the existing research on probabilistic set covering with a random constraint matrix assumes all data or columns are independent. This assumption is often unrealistic. In this work we attempt to alleviate this shortcoming. Before outlining our approach, we address the issue of why it is important to explicitly consider correlations. Agrawal et al. [1] characterize the repercussions of overlooking correlations in the context of distributionally robust stochastic programming (DRSP) by introducing a quantity they term the *price of correlation*. Their DRSP model seeks to minimize a function $g(x)$ representing the maximum expected cost of a given cost function, where the maximum is taken over all distributions D in the family $\mathcal{D}$ of distributions that satisfy some limited distributional assumptions, e.g., all distributions with first moment ν. In other words, $g(x)$ is the expected cost under the worst-case distribution in the family $\mathcal{D}$. They define the price of correlation as the ratio

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

where x_I is an optimal decision for distributions in $\mathcal{D}$ that assume independent uncertainties, and x_R is an optimal decision for the DRSP model. A small price of correlation suggests that a modeler may incur only a small cost when using the product distribution, which assumes independent uncertainties. Indeed, Agrawal et al. [1] provide a class of instances for the probabilistic set covering problem whose price of correlation is $\Omega\left(\sqrt{n}\frac{\log\log n}{\log n}\right)$, which indicates the potential gravity of ignoring the dependence among the data.

This paper is concerned with *conservative deterministic* reformulations of the following probabilistically constrained system under various models of data correlation. Specifically, let us consider the following system associated with covering a single item, i.e., $m = 1$:

$$X = \left\{ (w, x) \in [0,1] \times \{0,1\}^n : \ w \le p(x) := \mathbb{P}\left[\sum_{j \in N} \tilde{a}_j x_j \ge 1 \right] \right\} . \tag{3}$$

Note that systems of the above form constitute the main substructure in the probabilistic set covering models (1) and (2) which can be written as

$$\min \left\{ \sum_{j \in N} c_j x_j : \ (w_i, x) \in X_i, w_i \ge 1 - \epsilon_i \ \forall\, i \in M, \ x \in S \subseteq \{0,1\}^n \right\}$$

and

$$\max\{w : \ (w_i, x) \in X_i, w_i \ge w \ \forall\, i \in M, \ x \in S \subseteq \{0,1\}^n\} ,$$

respectively, where X_i is a system of the form (3) corresponding to the i-th item. A deterministic reformulation of X can then be directly embedded in the above formulations to obtain deterministic reformulations of (1) and (2).

We consider various models of data correlation for the Bernoulli random vector $\tilde{a}$ in the probabilistically constrained system X defined by (3). First, we assume that the components of $\tilde{a} = (\tilde{a}_1, \ldots, \tilde{a}_n)$ are conditionally independent, and are correlated via a fully specified mixture model. Next, we assume that the full distribution of $\tilde{a}$ is not available, but that partial information is available in one of two ways: (a) only the marginal probabilities $p_j = \mathbb{P}[\tilde{a}_j = 1]$ for all j are known, or (b) the marginal and pair-wise marginal probabilities $p_{ij} = \mathbb{P}[\tilde{a}_i = 1 \wedge \tilde{a}_j = 1]$ for all i and j are known. For example, we assume that a sufficiently large sample of historical data is available so that these marginals can be computed with high precision using standard estimation techniques. Note that in case (b), correlations ρ_{ij} are known exactly since

$$\rho_{ij} = \frac{\text{Cov}(\tilde{a}_i, \tilde{a}_j)}{\sqrt{\text{Var}(\tilde{a}_i)\text{Var}(\tilde{a}_j)}} = \frac{p_{ij}}{\sqrt{p_i(1-p_i)p_j(1-p_j)}} \qquad \forall\ i, j \ \in N .$$

To handle these distributional uncertainties, we make use of ambiguous chance constraints to obtain distributionally-robust formulations of (3).

Our primary contributions are summarized below:

1. We conduct the first scenario-free investigation of the probabilistic set covering problem with correlated data in the constraint matrix.

2. We study four models of data correlation for the Bernoulli random vector $\tilde{a}$ in the probabilistically constrained system X defined by (3). With the exception of the case when the data are independent, the resulting models lead to mixed-binary nonlinear systems. By exploiting special properties of these systems, e.g., supermodularity, we are able to reformulate these models as mixed-integer *linear* models. When correlations are assumed to be known, we propose a branch-and-cut framework for solving the corresponding PSC models.

3. Two computational experiments lend empirical support to our claims. The first experiment illustrates that our distributionally-robust model not only outperforms a model in which correlations are completely ignored, but also maintains a relatively small optimality gap (i.e., it is not overly conservative as are many robust deterministic approximations). The second experiment shows that our branch-and-cut algorithms can significantly reduce overall computation time.

The outline of this paper is as follows. In Section 2, we discuss two special cases of problem (3) in which the constraint coefficients $\tilde{a}_{ij}$ are independent or conditionally independent. In Section 3, using the concept of ambiguous chance constraints, we study two distributionally-robust formulations of (3). This leads to two mixed-integer linear programming (MILP) reformulations that can be solved by an off-the-shelf solver for small instances. As these MILPs may not scale well, in Section 4, we present an alternative formulation with a small number of decision variables, but an exponential number of constraints, and show how such a model can be solved in a branch-and-cut framework. Finally, computational results are presented in Section 5.

Assumptions and Notation. Given independent Bernoulli random variables $\tilde{a}_j$, we assume throughout that $p_j = \mathbb{P}[\tilde{a}_j = 1] \in (0,1)$ and $q_j = 1-p_j$ for all $j \in N$. At times, we will express $p(x) = \mathbb{P}\left[\sum_{j\in N} \tilde{a}_j x_j \geq 1\right]$ as $\mathbb{P}\left[\cup_{j\in N} A_j(x)\right]$ or as $1 - \mathbb{P}\left[\cap_{j\in N} A_j^c(x)\right]$ where $A_j(x) = \{\tilde{a}_j x_j = 1\}$ and $A_j^c(x) = \{\tilde{a}_j x_j = 0\}$ for all $j \in N$. We define $N^j := N \setminus \{j\}$ for all $j \in N$. The convex hull of a set X is denoted by $\text{conv}(X)$.

2 Independent and Conditionally Independent Cases

In this section, we develop *deterministic* reformulations of the probabilistic covering set X defined by (3) when the random variables $\tilde{a}_j$ for $j \in N$ are either independent or conditionally independent. As mentioned before, the independent case has been investigated in [9] and [11]. We include it here for the sake of completeness.

Proposition 1 [9, 11] *Let $\tilde{a}_j$ be independent Bernoulli random variables with $p_j = \mathbb{P}[\tilde{a}_j = 1] \in (0,1)$ and $q_j = 1 - p_j$ for all $j \in N$. Then*

$$X = \left\{ (w,x) \in [0,1] \times \{0,1\}^n : \sum_{j\in N} (\log q_j) x_j \leq \log(1-w) \right\} . \tag{4}$$

Proof: Note that $\mathbb{P}[A_j^c(x)] = q_j^{x_j}$ since $x_j \in \{0,1\}$. Thus,

$$\begin{aligned} X &= \left\{ (w,x) \in [0,1] \times \{0,1\}^n : \mathbb{P}[\bigcup_{j\in N} A_j(x)] \geq w \right\} = \left\{ (w,x) \in [0,1] \times \{0,1\}^n : \mathbb{P}[\bigcap_{j\in N} A_j^c(x)] \leq 1 - w \right\} \\ &= \left\{ (w,x) \in [0,1] \times \{0,1\}^n : \prod_{j\in N} q_j^{x_j} \leq 1 - w \right\} = \left\{ (w,x) \in [0,1] \times \{0,1\}^n : \sum_{j\in N} (\log q_j) x_j \leq \log(1-w) \right\} , \end{aligned}$$

where the third equality follows from independence. $\square$

Noting that $w_i \geq 1-\epsilon_i$ is equivalent to $\log(1-w_i) \leq \log \epsilon_i$, it follows from Proposition 1 that, under independence of the covering coefficients, the probabilistic set covering problem (1) is equivalent to a deterministic binary integer program with m knapsack constraints:

$$\min \left\{ \sum_{j \in N} c_j x_j : \sum_{j \in N} (\log q_{ij}) x_j \leq \log \epsilon_i \ \forall\, i \in M,\ x \in S \subseteq \{0,1\}^n \right\} ,$$

where $q_{ij} = 1 - p_{ij}$. Similarly, noting that maximizing w is equivalent to minimizing $w' := \log(1-w)$, we can reformulate the max-min coverage probability problem (2) as the deterministic mixed-integer problem:

$$\min \left\{ w' : \sum_{j \in N} (\log q_{ij}) x_j \leq w' \ \forall\, i \in M,\ x \in S \subseteq \{0,1\}^n \right\} .$$

Next we consider Bernoulli mixture models, which are a natural extension of the above independent model in which the random covering coefficients are conditionally independent. That is, conditioned on some exogenous factors, the random covering coefficients are independent, but without conditioning, they may be dependent.

Proposition 2 *Suppose that the random vector $(\tilde{a}_1, \ldots, \tilde{a}_n)$ follows a Bernoulli mixture distribution. That is, given a finite prior distribution $\{\pi_1, \ldots, \pi_L\}$, the conditional probabilities $p_{i\ell} = \mathbb{P}[\tilde{a}_i = 1|\ell]$ and $p_{j\ell} = \mathbb{P}[\tilde{a}_j = 1|\ell]$ are independent for all $i, j \in N$ and $\ell \in \{1, \ldots, L\}$, and the posterior distribution is given by $p_j = \sum_{\ell=1}^{L} \pi_\ell p_{j\ell}$ for all j. Let $q_{j\ell} = 1 - p_{j\ell} \in (0,1)$ for all j and ℓ. Then*

$$X = \left\{ (w,x) \in [0,1] \times \{0,1\}^n : \ \exists\, y \in \mathbb{R}^L_+ \text{ s.t. } \sum_{\ell=1}^{L} \pi_\ell y_\ell \leq 1 - w,\ y_\ell \geq \prod_{j \in N} q_{j\ell}^{x_j} \ \ \forall\, \ell = 1, \ldots, L \right\} . \tag{5}$$

Proof: As in Proposition 1, we have

$$\begin{aligned} \mathbb{P}\left[\bigcup_{j \in N} A_j(x) \right] \geq w \quad &\Longleftrightarrow \quad \mathbb{P}\left[\bigcap_{j \in N} A_j^c(x) \right] \leq 1 - w \\ &\Longleftrightarrow \quad \sum_{\ell=1}^{L} \pi_\ell \mathbb{P}\left[\bigcap_{j \in N} A_j^c(x) \mid \ell \right] \leq 1 - w \Longleftrightarrow \sum_{\ell=1}^{L} \pi_\ell \prod_{j \in N} q_{j\ell}^{x_j} \leq 1 - w . \end{aligned}$$

Introducing a new variable y_ℓ to model the probability of failure $\prod_{j \in N} q_{j\ell}^{x_j}$ under scenario ℓ, we obtain the desired reformulation. □

The set X in (5) is a mixed-integer nonlinear set. The nonlinear constraint $y_\ell \geq \prod_{j \in N} q_{j\ell}^{x_j}$ in (5) can be written as

$$y_\ell \geq \prod_{j=1}^{n} q_{j\ell}^{x_j} = \exp\left\{ \sum_{j \in N} (\log q_{j\ell}) x_j \right\} =: \sigma_\ell(x) \qquad \forall\, \ell = 1, \ldots, L .$$

Since σ_ℓ is convex in x for all ℓ, the mixed-integer nonlinear system (5) can be linearized by outer approximating σ_ℓ using gradient inequalities. Existing solvers for convex mixed-integer nonlinear programs rely on this approach and they do not exploit any other special structure of the underlying nonlinearities. In particular, the function σ_ℓ can be expressed as

$$\sigma_\ell(x) = f\left(\sum_{j \in N} d_{j\ell} x_j \right)$$

where $f(\cdot) = \exp(\cdot)$ and $d_{j\ell} = \log q_{j\ell}$, i.e., as the composition of a univariate convex function and a linear function. In recent work, Ahmed and Atamtürk [2] studied mixed binary nonlinear systems of the form $w \geq f(\sum_{j\in N} d_j x_j)$, and developed inequalities that provide a much stronger linearization than those obtained by standard gradient inequalities. These inequalities can be directly applied to obtain a strong linear formulation of the covering set X given by (5).

3 Incomplete Distribution Information

In this section, we assume that the full distribution of the Bernoulli random vector $(\tilde{a}_1, \ldots, \tilde{a}_n)$ is not available, but that partial information is available in one of two ways: (a) only the marginal probabilities $p_j = \mathbb{P}[\tilde{a}_j = 1]$ for all j are known, or (b) the marginal and pair-wise marginal probabilities $p_{ij} = \mathbb{P}[\tilde{a}_i = 1 \wedge \tilde{a}_j = 1]$ for all i and j are known. These cases represent the two extremes of how much pair-wise correlation information is available since, in (a), we assume that no correlations are known, whereas in (b) we assume that pair-wise correlations are known exactly.

Let $\mathcal{D}$ be the family of all n-variate Bernoulli distributions μ with the specified marginals, and consider the *Ambiguous* Probabilistic Set Covering set

$$X^A = \left\{ (w,x) \in [0,1] \times \{0,1\}^n : \min_{\mu \in \mathcal{D}} \left\{ \mathbb{P}_\mu \left[\sum_{j \in N} \tilde{a}_j x_j \geq 1 \right] \right\} \geq w \right\}, \tag{6}$$

where $\mathbb{P}_\mu[A]$ denotes the probability of the event A under distribution μ. Below we provide distributionally-robust reformulations of X^A based on the amount of correlation information that is assumed to be present.

3.1 A Correlation-Agnostic Model

In this subsection, we assume that only the marginal probabilities $p_j = \mathbb{P}[\tilde{a}_j = 1]$ for all j are known. Let $\mathcal{D}_1$ be the family of all n-variate Bernoulli distributions μ with the specified marginals, where the subscript 1 denotes that only *first* moment information for the random vector $\tilde{a}$ is available. We refer to models that are robust with respect to first moment information as *correlation-agnostic* as they acknowledge that the data may be correlated, but make no attempt at estimating these correlations. In other words, the modeler believes that nothing can be known about these correlations based on the data available.

Proposition 3 *When $\mathcal{D} = \mathcal{D}_1$, the ambiguous covering set X^A given by (6) can be reformulated into the following deterministic 0-1 nonlinear set*

$$X_1 = \left\{ (w,x) \in [0,1] \times \{0,1\}^n : \max_{j \in N} \{p_j x_j\} \geq w \right\} . \tag{7}$$

Proof: Given $x \in \{0,1\}^n$, note that $\mathbb{P}_\mu[A_j(x)] = p_j x_j$ for all $\mu \in \mathcal{D}_1$. As $\mathbb{P}_\mu[\cup_{j\in N} A_j(x)] \geq \max_{j\in N} \mathbb{P}_\mu[A_j(x)] = \max_{j\in N}\{p_j x_j\}$ for all $\mu \in \mathcal{D}_1$, we have that, if $(w,x) \in X^1$ then $(w,x) \in X^A$. Thus, $X_1 \subseteq X^A$. We next show that $X^A \subseteq X_1$. Note that if $(w,x) \in X^A$ such that $x = 0$ then $w = 0$ and $(0,0) \in X_1$. So we only need to consider the case $x \neq 0$. Now consider $(w,x) \in [0,1] \times \{0,1\}^n$ such that $w > \max_{j\in N}\{p_j x_j\}$, i.e., $(w,x) \notin X_1$ and $x \neq 0$. Assume without loss of generality that $p_1 x_1 \geq p_j x_j$ for all $j \in N$. Consider a $\mu' \in \mathcal{D}_1$ with the following dependency structure: $\mathbb{P}_{\mu'}[A_j(x)|A_1(x)] = \mathbb{P}_{\mu'}[A_j(x)]/\mathbb{P}_{\mu'}[A_1(x)]$ and $\mathbb{P}_{\mu'}[A_j(x)|A_1^c(x)] = 0$, for all $j = 2, \ldots, n$. Then, $\mathbb{P}_{\mu'}[\cup_{j\in N} A_j(x)] = \mathbb{P}[A_1(x)]$ since $\cup_{j=2}^n A_j(x) \subseteq A_1(x)$. Moreover, $w > \mathbb{P}[A_1(x)] = \mathbb{P}_{\mu'}[\cup_{j\in N} A_j(x)] \geq \min_{\mu\in\mathcal{D}_1} \mathbb{P}_\mu[\cup_{j\in N} A_j(x)]$, and so $(w,x) \notin X^A$. Thus, $X^A = X_1$. □

Proposition 4 *The deterministic 0-1 nonlinear covering set X_1 can be reformulated into the deterministic* linear *set defined by the set of points $(w, x, z) \in [0,1] \times [0,1]^n \times [0,1]^n$ satisfying:*

$$\sum_{j \in N} p_j z_j \geq w \tag{8a}$$

$$\sum_{j \in N} z_j \leq 1 \qquad \forall\, j \in N \tag{8b}$$

$$z_j \leq x_j \qquad \forall\, j \in N \tag{8c}$$

Proof: If x and z were constrained to be binary vectors, it is easy to see that a point $(\widehat{w}, \widehat{x})$ belongs to X_1 if and only if it belongs to the set defined by (8). The set defined by (8) is totally unimodular, so integrality on x and z can be dropped in the absence of side constraints. □

Remark. If w is fixed, e.g., $w = \epsilon$ as in Problem (1), we define the set $J := \{j \in N : p_j \geq \epsilon\}$. Then, $X_1 = \{x \in \{0,1\}^n : \sum_{j \in J} x_j \geq 1\}$.

3.2 A Correlation-Certain Model

In this subsection, we assume that the marginal and pair-wise marginal probabilities $p_{ij} = \mathbb{P}[\tilde{a}_i = 1 \wedge \tilde{a}_j = 1]$ for all i and j are known and then provide deterministic reformulations of the ambiguous covering set X^A given by (6). Let $\mathcal{D}_2$ be the family of all n-variate Bernoulli distributions μ with the specified marginals, where the subscript 2 denotes that first and *second* moment information for the random vector $\tilde{a}$ is available. We refer to models that are robust with respect to second moment information as *correlation-certain* as they assume that all pair-wise correlations are known with certainty.

To obtain our reformulations, we take advantage of a result from Kuai et al. [13] who provide a tight lower bound on the probability of a union of a finite set of events when first and second moments are known. This results allows us to recast the ambiguous covering set X^A as a deterministic 0-1 *nonlinear* set. Next, we linearize the resulting nonlinear set to arrive at a deterministic large-scale MILP reformulation of X^A under the correlation-certain model. In Section 4, we take a different approach and study a compact mixed-integer formulation by exploiting supermodularity.

Proposition 5 *When $\mathcal{D} = \mathcal{D}_2$, the ambiguous covering set X^A given by (6) can be reformulated into the following deterministic 0-1 nonlinear set*

$$X_2 = \left\{ (w, x) \in [0,1] \times \{0,1\}^n : \sum_{j \in N} \Phi_j \left(\sum_{i \in N^j} p_{ij} x_i \right) x_j \geq w \right\}, \tag{9}$$

where

$$\Phi_j(t) := \max_{\ell = 1, \ldots, n-1} \left\{ \frac{2p_j}{\ell + 1} - \frac{t}{\ell(\ell+1)} \right\}.$$

Proof: We need to show that

$$\min_{\mu \in \mathcal{D}_2} \left\{ \mathbb{P}_\mu \left[\bigcup_{j \in N} A_j(x) \right] \right\} = \sum_{j \in N} \Phi_j \left(\sum_{i \in N^j} p_{ij} x_i \right) x_j \ . \tag{10}$$

It has been shown (see the proof of Theorem 1 in [13], see also [18]) that $\min_{\mu \in \mathcal{D}_2} \{\mathbb{P}_\mu [\cup_{j \in N} A_j(x)]\}$ is equal to the optimal value of the linear program

$$\begin{array}{lll} \min & \sum_{i \in N} \sum_{j \in N} y_{ij}/i & \\ \text{s.t.} & \sum_{i \in N} y_{ij} = a_j & \forall\, j \in N \\ & \sum_{i \in N} (i-1) y_{ij} = b_j & \forall\, j \in N \\ & y_{ij} \geq 0 & \forall\, i, j \in N\ , \end{array} \tag{11}$$

where $a_j = \mathbb{P}_\mu[A_j(x)] = p_j x_j$ and $b_j = \sum_{k \in N^j} \mathbb{P}_\mu[A_j(x) \cap A_k(x)] = \sum_{k \in N^j} p_{kj} x_k x_j$ for all $\mu \in \mathcal{D}_2$. Note that (11) is separable in j. We prove the following claim in Appendix A.

Claim 1 *Consider the LP*

$$\begin{array}{ll} \min & \sum_{i \in N} y_i / i \\ \text{s.t.} & \sum_{i \in N} y_i = a \\ & \sum_{i \in N} (i-1) y_i = b \\ & y_i \geq 0 \qquad\qquad \forall\, i \in N\ . \end{array} \tag{12}$$

If $n \geq 2$ and $a, b \geq 0$, then LP (12) has an optimal objective function value of

$$\max_{i=1,\ldots,n-1} \left\{ \frac{2a}{i+1} - \frac{b}{i(i+1)} \right\} \ .$$

It then follows from the separability of (11) and Claim 1 that

$$\min_{\mu \in \mathcal{D}} \left\{ \mathbb{P}_\mu \left[\bigcup_{j \in N} A_j(x) \right] \right\} = \sum_{j \in N} \max_{i=1,\ldots,n-1} \left\{ \frac{2 p_j x_j}{i+1} + \frac{\sum_{k \in N^j} p_{kj} x_k x_j}{i(i+1)} \right\} = \sum_{j \in N} \Phi_j \left(\sum_{i \in N^j} p_{ij} x_i \right) x_j \ . \ \square$$

Proposition 6 *The deterministic 0-1 nonlinear covering set X_2 given by (9) can be reformulated into the deterministic mixed-integer linear set defined by the set of points (λ, v, w, w', x) satisfying:*

$$\sum_{j \in N} w'_j \geq w \tag{13a}$$

$$w'_j \leq \sum_{\ell=1}^{n-1} \left(\frac{2p_j}{\ell+1} \right) \lambda_{j\ell} - \left(\sum_{\ell=1}^{n-1} \sum_{k \in N^j} \frac{p_{jk}}{\ell(\ell+1)} v_{jk\ell} \right) \qquad \forall j \in N \tag{13b}$$

$$0 \leq w'_j \leq p_j x_j \qquad \forall j \in N \tag{13c}$$

$$0 \leq \lambda_{j\ell} - v_{jk\ell} \leq (1 - x_k) \qquad \forall j,k \in N (k \neq j), \forall \ell = 1, \ldots, n-1 \tag{13d}$$

$$v_{jk\ell} \leq x_k \qquad \forall j,k \in N (k \neq j), \forall \ell = 1, \ldots, n-1 \tag{13e}$$

$$\sum_{\ell=1}^{n-1} \lambda_{j\ell} = 1 \qquad \forall j \in N \tag{13f}$$

$$\lambda_{j\ell} \geq 0 \qquad \forall j \in N, \forall \ell = 1, \ldots, n-1, \tag{13g}$$

$$v_{jk\ell} \in [0,1] \qquad \forall j,k \in N (k \neq j), \forall \ell = 1, \ldots, n-1 \tag{13h}$$

$$x_j \in \{0,1\} \qquad \forall j \in N\ . \tag{13i}$$

Proof: Let $c_{j\ell} = \frac{2p_j}{\ell+1}$ and $d_{j\ell}(t) = \frac{t}{\ell(\ell+1)}$. We can express $\Phi_j(t)$ as the objective function value of the following linear program:

$$\max \left\{ \sum_{\ell=1}^{n-1} \lambda_{j\ell} \left(c_{j\ell} - d_{j\ell}(t) \right) : \sum_{\ell=1}^{n-1} \lambda_{j\ell} = 1, \lambda_{j\ell} \geq 0, \forall \ell = 1, \ldots, n-1 \right\} \ .$$

Introducing an auxiliary variable w'_j to model $\Phi_j \left(\sum_{i \in N^j} p_{ij} x_i \right) x_j$, we can replace the nonconvex constraints

$$\sum_{j \in N} \Phi_j \left(\sum_{i \in N^j} p_{ij} x_i \right) x_j \geq w \qquad \forall\, j \in N$$

with linear constraints (13a) and (13c), and nonlinear constraints

$$w'_j \leq \sum_{\ell=1}^{n-1} \lambda_{j\ell} \left(c_{j\ell} - d_{j\ell} \left(\sum_{k \in N^j} p_{jk} x_k \right) \right) \qquad \forall\, j \in N\ . \tag{14}$$

To eliminate the nonlinear terms $\lambda_{j\ell}x_k$ that appear in (14), introduce auxiliary variables $v_{jk\ell}$ to model $\lambda_{j\ell}x_k$ and add the constraints (13d), (13e), and (13h). With this replacement, (14) can be written as (13b). □

While the above reformulation is convenient in that one can directly formulate this model as a MILP and attempt to solve it, it suffers from at least one major drawback – it possesses $O(n^3)$ decision variables, of which n are binary, and $O(2n^3)$ constraints. Worse yet, given $m > 1$ probabilistic constraints, these numbers get multiplied by m. To avoid these deficiencies, we next study a "compact" formulation which has only $2n + 1$ decision variables and which will be shown to yield superior performance.

4 A Compact Linearization of the Correlation-Certain Model

In this section, we investigate the polyhedral structure of the set X_2 given in (9). Ideally, we would like to obtain a linear description of the convex hull of X_2. However, this turns out to be a difficult task. Instead, we focus on a slightly less ambitious goal, but which is still beneficial, of obtaining a linear description of the convex hull of the set

$$Y_j := \left\{(w'_j, x) \in \mathbb{R} \times \{0,1\}^n : \ 0 \le w'_j \le g_j(x)x_j\right\} \ , \tag{15}$$

where $g_j(x) := \Phi_j\left(\sum_{i\in N^j} p_{ij}x_i\right)$ for each $j \in N$. Note that a linear description of $\text{conv}(Y_j)$ is helpful since we can express X_2 as

$$X_2 = \left\{(w, x) \in \mathbb{R} \times \{0,1\}^n : \exists\ w' \in \mathbb{R}^n \text{ s.t.} \sum_{j\in N} w'_j \ge w, (w'_j, x) \in Y_j \ \forall\ j \in N\right\} \ . \tag{16}$$

To obtain a linear description of $\text{conv}(Y_j)$, we study restrictions of Y_j in which $x_j = 1$ and only a subset S of the remaining binary variables x are permitted to take a non-zero value. That is, let

$$Y^S_{j1} := Y_j \cap \{(w'_j, x) \in \mathbb{R} \times \{0,1\}^n : \ x_j = 1, \ x_i = 0 \ \forall\ i \notin S\} \qquad \forall\ j \in N, \ \forall\ S \subseteq N^j \ .$$

The following proposition sheds light on the relationship between the facets of $\text{conv}(Y_j)$ and $\text{conv}(Y^S_{j1})$.

Proposition 7 *For any $j \in N$, $\text{conv}(Y_j)$ is completely described by the trivial bound inequalities $w'_j \ge 0$, $x_j \le 1$, $0 \le x_i \le 1$ for all $i \in N^j$, and a finite set of inequalities, each of which is of the form*

$$w'_j + (b - \sum_{i\in N^j} a_i)(1 - x_j) + \sum_{i\in N^j} a_i x_i \le b \ , \tag{17}$$

where $a \in \mathbb{R}^{n-1}_+$ and $b > 0$. Moreover, an inequality of the above form is a facet of $\text{conv}(Y_j)$ if and only if $w'_j + \sum_{i\in S} a_i x_i \le b$ is a facet of $\text{conv}(Y^S_{j1})$ where $S = \{i \in N^j : \ a_i > 0\}$.

Proof: See Appendix B. □

The above proposition indicates that we can express X_2 as a mixed integer linear set by modifying the inequalities that describe $\text{conv}(Y^S_{j1})$ for all $S \subseteq N^j$ for all $j \in N$. However, the above proposition does not provide an explicit form of the coefficients in the nontrivial inequalities. Next we exploit the fact that g_j is a supermodular function to identify the precise values of the coefficients in (17).

Definition 1 *A set function $h : 2^N \mapsto \mathbb{R}$ is* supermodular *on N if*

$$h(S \cup \{k\}) - h(S) \le h(T \cup \{k\}) - h(T) \qquad \forall\ S \subseteq T \subseteq N \text{ and } k \notin T.$$

A function h is submodular *if $-h$ is supermodular.*

Proposition 8 *The set function $g_j(x)$ is nonincreasing and supermodular on N^j.*

Proof: Observe that $g_j(S) = \Phi_j(\sum_{i\in S} p_{ij})$ for any $S \subseteq N^j$. Since $\Phi_j(t)$ is convex in t, we have $\Phi_j(t_1+\delta) - \Phi_j(t_1) \le \Phi_j(t_2+\delta) - \Phi_j(t_2)$, for $t_1 \le t_2$ and $\delta \ge 0$. For any $S \subseteq T \subseteq N$, let $t_1 = \sum_{i\in S} p_{ij}$ and $t_2 = \sum_{i\in T} p_{ij}$. Nonnegativity of p_{ij} implies $t_1 \le t_2$. For any $k \notin T$, let $\delta = p_{kj}$ and note that $g_j(S\cup\{k\}) - g_j(S) = \Phi_j(t_1+\delta) - \Phi_j(t_1) \le \Phi_j(t_2+\delta) - \Phi_j(t_2) = g_j(T\cup\{k\}) - g_j(T)$. $\square$

Remark. The function $g_j(x)x_j$ for a given j is neither sub- nor super-modular.

Having shown that the function g_j is supermodular, we pause to explain why this fact should not be surprising. As shown in Equation (10), given a fixed solution $\widehat{x}$, $\sum_{j\in N} g_j(\widehat{x})\widehat{x}_j$ represents the minimum probability of the union of a finite set of events $A_j(\widehat{x})$, where the minimum is taken over all distributions in the family $\mathcal{D}_2$. Thus, we can interpret $g_j(\widehat{x})$ as the contribution to this minimum probability from including column j in the set cover. As the columns "work together" to cover rows (i.e., to increase the probability of coverage for each row) of the probabilistic set covering problem, they can be considered *complements*[1]. There is a well-known connection between supermodularity and complementarity [22].

The next proposition from Atamtürk and Narayanan [3], which builds on earlier work of Edmonds [8], exploits the supermodularity of g_j and allows us to identify the precise values of the coefficients a and b in (17).

Proposition 9 [3] *For any $j \in N$ and any $S \subseteq N^j$ with $|S| = k$, $\mathrm{conv}(Y^S_{j1})$ is completely described by the trivial bound inequalities $w'_j \ge 0$, $x_j = 1$, $x_i = 0$ for all $i \notin S$, $0 \le x_i \le 1$ for all $i \in S$, and the nontrivial facets*

$$w'_j + \sum_{i\in S} a_{\pi(i)} x_{\pi(i)} \le b \qquad \forall\ \pi \in \Pi(S)\ , \tag{18}$$

where $\Pi(S)$ is the set of all permutations $\pi = (\pi(1),\ldots,\pi(k))$ of the elements of S, $b = g_j(0) = p_j$, and

$$a_{\pi(i)} = \begin{cases} g_j(0) - g_j(e^{\pi(1)}) & i = 1 \\ g_j(\sum_{t=1}^{i-1} e^{\pi(t)}) - g_j(\sum_{t=1}^{i} e^{\pi(t)}) & i = 2,\ldots,k\ , \end{cases} \tag{19}$$

where $e^i \in \mathbb{R}^n$ is the ith unit vector.

Example. Suppose $n = 3$, $p = (0.9, 0.8, 0.7)$, and $p_{ij} = p_i p_j$, for all $i, j \in N$ and note that $g_1(\{2\}) = 0.360$ and $g_1(\{3\}) = 0.315$. The nontrivial facets of $\mathrm{conv}(Y^S_{11})$ are

Inequality	S	π
$0.360x_2 + w'_1 \le 0.360$	$\{2\}$	(2)
$0.315x_3 + w'_1 \le 0.315$	$\{3\}$	(3)
$0.210x_2 + 0.315x_3 + w'_1 \le 0.525$	$\{2,3\}$	$(2,3)$
$0.360x_2 + 0.165x_3 + w'_1 \le 0.525$	$\{2,3\}$	$(3,2)$.

The facets above give rise to the nontrivial facets of $\mathrm{conv}(Y_1)$

$$\begin{array}{rcrcrcrcl} -0.900x_1 & & & & & + & w'_1 & \le & 0 \\ -0.540x_1 & + & 0.360x_2 & & & + & w'_1 & \le & 0.360 \\ -0.585x_1 & & & + & 0.315x_3 & + & w'_1 & \le & 0.315 \\ -0.375x_1 & + & 0.210x_2 & + & 0.315x_3 & + & w'_1 & \le & 0.525 \\ -0.375x_1 & + & 0.360x_2 & + & 0.165x_3 & + & w'_1 & \le & 0.525\ . \end{array}$$

The following corollary follows immediately from Propositions 7 and 9.

[1] A set of activities or decision variables are *complements* if the additional utility resulting from the availability of any additional activity is increasing with the set of other activities available [22].

Corollary 1 *The nontrivial facets of* $\text{conv}(Y_j)$ *are*

$$w'_j + \left(p_j - \sum_{i\in S} a_{\pi(i)}\right)(1-x_j) + \sum_{i\in S} a_{\pi(i)} x_{\pi(i)} \le p_j \qquad \forall\ j \in N, \forall\ S \subseteq N^j, \forall\ \pi \in \Pi(S)\ , \tag{20}$$

where the coefficients $a_{\pi(i)}$ *are defined in* (19).

We refer to constraints (20) as *lifted extended polymatroid* (LEP) cuts as they are derived from lifting cuts of the form (18), which are referred to as *extended polymatroid* cuts in [3].

Finally, we arrive at the following *compact* reformulation of the correlation-certain model:

$$X_2 = \left\{ (w', w, x) \in \mathbb{R}^n \times \mathbb{R}_+ \times \{0,1\}^n : \sum_{j\in N} w'_j \ge w, (20) \right\} . \tag{21}$$

Although this formulation has an exponential number of constraints, we call it compact since there are only $2n+1$ decision variables. Since the number of constraints (20) is extremely large, namely $n\sum_{t=0}^{n-1}\left(\binom{n-1}{t}!\right) >> n\sum_{t=0}^{n-1}\binom{n-1}{t} = n2^{n-1}$, when optimizing over the set X_2, we omit all but those corresponding to when $T = \emptyset$, i.e., we only include constraints $0 \le w'_j \le p_j x_j$ for all $j \in N$ in the initial relaxation, and add the others on an as-needed basis during the course of branch-and-cut procedure. Specifically, given a point $(\widehat{w}', \widehat{x}) \in \mathbb{R}^n \times \mathbb{R}^n$ such that $\widehat{w}'_j > g_j(\widehat{x})\widehat{x}_j$, we can separate an inequality of the form $w'_j + (b - \sum_{i\in S} a_i)(1-x_j) + \sum_{i\in S} a_i x_i \le b$ as follows. Since the right-hand-sides of the inequalities are identical, we only need to find a set S and coefficients a_i for $i \in S$ such that $\sum_{i\in S}(\widehat{x}_i - 1 + \widehat{x}_j)a_i$ is maximized. Thus we set $S = \{i \in N^j :\ \widehat{x}_i - 1 + \widehat{x}_j > 0\}$. Next we sort $\widehat{x}_i$ for $i \in S$ such that $\widehat{x}_{[1]} \ge \widehat{x}_{[2]} \ge \cdots \ge \widehat{x}_{[k]}$, breaking ties arbitrarily, and computing $a_{[i]}$ according to (19). A high-level description of this procedure is provided in Algorithm 1. Note that Algorithm 1 can be used for the large-scale MILP set (13) or for the compact MILP set (21).

Algorithm 1 `SeparateLEPCuts`$(\widehat{w}', \widehat{x})$

Require: A (possibly fractional) solution pair $(\widehat{w}', \widehat{x})$ to the current LP relaxation.
Sort the x_j variables in nonincreasing order $\widehat{x}_{[1]} \ge \widehat{x}_{[2]} \ge \cdots \ge \widehat{x}_{[n]}$.
for $j = 1, \ldots, n$ **do**
  if $\widehat{w}'_j > g_j(\widehat{x})\widehat{x}_j$ **then**
    Define the set $S = \{i \in N^j : \widehat{x}_i + \widehat{x}_j > 1\}$ and let $k = |S|$.
    Let $\pi = (\pi(1), \ldots, \pi(k))$ s.t. $\widehat{x}_{\pi(1)} \ge \cdots \ge \widehat{x}_{\pi(k)}$ and $\{\pi(1), \ldots, \pi(k)\} = S$.
    Set $a_{\pi(i)}$ according to Equation (19) $\forall\ i \in S$.
    if $\widehat{w}'_j - p_j\widehat{x}_j + \sum_{i\in S}(\widehat{x}_{\pi(i)} + \widehat{x}_j - 1)a_{\pi(i)} > 0$ **then**
      The most violated cut is $w'_j + (b - p_j)x_j + \sum_{i\in S} a_{\pi(i)} x_{\pi(i)} \le b$ where $b = \sum_{i\in S} a_i$.
    end if
  end if
end for

5 Computational Results

In this section, we present two computational experiments. In the first experiment, we show empirically that the explicit inclusion of correlation data within our framework can lead to superior results in comparison to ignoring correlations. In addition, we show that our correlation-certain model is not overly conservative, at least for the family of instances that we consider. In the second experiment, we highlight the significant reduction in computation times that our compact formulation (21) can provide over the MILP reformulation (13). All experiments used the MILP solver of CPLEX Version 12.2 and were carried out on a Linux machine with kernel 2.6.18 running a 64-bit x86 processor equipped with two 2.27 GHz Intel Xeon E5520 chips and 32GB of RAM. All code was compiled using GCC version 4.4.3.

5.1 Comparison of Models: What does correlation buy us?

In our first experiment, we compare our two correlation-robust models with the correlation-free model (4) over a set of Bernoulli mixture instances and show that our correlation-certain model performs well. Our rationale for choosing Bernoulli mixture instances is due to two of its convenient properties. First, we can analytically compute the coverage probability associated with a solution, rather than revert to time-consuming simulations which, at best, furnish bounds on the coverage probability. Second, we can solve the Bernoulli mixture model to provable optimality without simulation using the algorithms discussed in Section 2 following Proposition 2. At the same time, we recognize that Bernoulli mixture instances represent only one family of probabilistic set covering instances and, therefore, any desire to make general conclusions from our results must be tempered.

We frame this experiment in the context of three modelers tasked to solve the PSC problem. Specifically, we compare

- a *correlation-free* modeler, who believes that the correlations are negligible and, therefore, a product distribution can be used without any measurable deterioration in solution quality;
- a *correlation-agnostic* modeler, who believes that the data are correlated, but that nothing can be known about these correlations (i.e., estimating these correlations is impossible);
- a *correlation-certain* modeler, who believes that the data are correlated and that these correlations can be estimated accurately with very high precision so much so that they can be taken to be exact.

For this experiment we chose to maximize the coverage probability subject to a budget constraint:

$$\max_{x \in B(b)} \mathbb{P}\left\{ \sum_{j \in N} \tilde{a}_j x_j \geq 1 \right\} \tag{22}$$

where $B(b) = \{x \in \{0,1\} : c^T x \leq b\}$, $b \in \mathbb{R}_+$, and $c_j = p_j$ for all $j \in N$. We chose this approach so that all models would have identical feasible regions, but different objective functions. After solving each model for an optimal solution x^* (optimal with respect to its correlation assumptions), we compare the models by computing the true probability of coverage $1 - \sum_{\ell=1}^{L} \sigma_\ell(x^*)\pi_\ell$, where $\sigma_\ell(x) = \prod_{j=1}^{n} q_{j\ell}^{x_j}$ is the probability of failure associated with the solution x, given scenario ℓ, for a Bernoulli mixture distribution.

We generated Bernoulli mixture instances with $n = 15$ binary decision variables and $L = 5$ scenarios as follows: After generating conditional probabilities p_{jl} randomly from a `uniform`(0.1,0.9) distribution, we constructed prior vectors $\pi = \{\pi_l\}_{\ell=1}^{L}$ such that the average correlation magnitude $|\bar{\rho}| = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |\rho_{ij}| / \binom{n}{2}$ was nearly maximized. Provably maximizing $|\bar{\rho}|$ is a challenging problem in its own right, so we employed a heuristic. Our reasoning for making $|\bar{\rho}|$ large was to create random instances in which the individual correlations ρ_{ij} were not close to zero since, otherwise, the correlation-free model would likely provide a close approximation to the true problem. We also experimented with maximizing the minimum correlation magnitude $|\rho_{ij}|$, but the advantage of this metric was not noticeable.

Here we present the results of ten representative instances, whose correlation statistics are shown in Table 1. Four instances are shown below, while the remaining six can be found in Appendix C. Each figure compares the probability of coverage (as a function of the budget threshold b) of three different models:

- (Optimal) an optimal model in which the true probability of coverage $1 - \sum_{\ell=1}^{L} \sigma_\ell(x^*)\pi_\ell$ is maximized;
- (Certain) the correlation-certain model;
- (Free) the correlation-free model.

The probability of coverage of the correlation-agnostic model is not shown in the figures, but is discussed below. As one would expect, the optimal probability of coverage increases monotonically as b increases.

	Instance									
Statistic	1	2	3	4	5	6	7	8	9	10
$\bar{\rho}$	-0.0002	-0.0084	-0.0029	-0.0014	-0.0053	0.0256	0.0272	-0.0131	-0.0006	-0.0085
$\min_{i,j}\{\rho_{ij}\}$	-0.4532	-0.3660	-0.3223	-0.2869	-0.3433	-0.2877	-0.2856	-0.4143	-0.3766	-0.3085
$\max_{i,j}\{\rho_{ij}\}$	0.3693	0.2805	0.3292	0.2249	0.3197	0.3210	0.3872	0.3740	0.4332	0.3552
$\bar{\rho}^+$	0.1191	0.1035	0.1088	0.0786	0.1253	0.1254	0.0879	0.1387	0.1581	0.1195
$\bar{\rho}^-$	-0.1129	-0.1101	-0.1167	-0.0830	-0.1335	-0.1182	-0.0891	-0.1512	-0.1243	-0.1121
$\sum_{i,j} \mathbf{1}\{\rho_{ij} > 0\}$	51	50	53	53	52	62	69	50	46	47
$\sum_{i,j} \mathbf{1}\{\rho_{ij} < 0\}$	54	55	52	52	53	43	36	55	59	58
$\min_{i,j}\{\|\rho_{ij}\|\}$	0.0091	0.0023	0.0022	0.0005	0.0098	0.0108	0.0065	0.0000	0.0023	0.0016
$\|\bar{\rho}\|$	0.1159	0.1070	0.1127	0.0808	0.1294	0.1225	0.0883	0.1453	0.1391	0.1154

Table 1: Correlation statistics of instances. Here $\bar{\rho}$ denotes the average correlation over all pairs, $\bar{\rho}^+$ ($\bar{\rho}^-$) denotes the average correlation over all pairs with positive (negative) correlation, $\sum_{i,j} \mathbf{1}\{\rho_{ij} > 0\}$ ($\sum_{i,j} \mathbf{1}\{\rho_{ij} < 0\}$) counts the number of pairs with positive (negative) correlation, and $|\bar{\rho}|$ denotes the average correlation magnitude (not the magnitude of the average correlation).

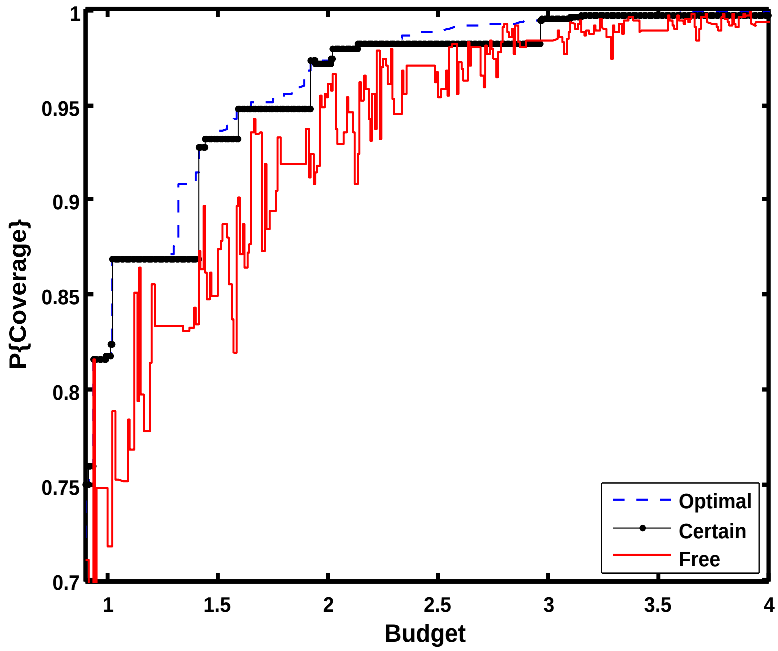


Figure 1: Instance 1

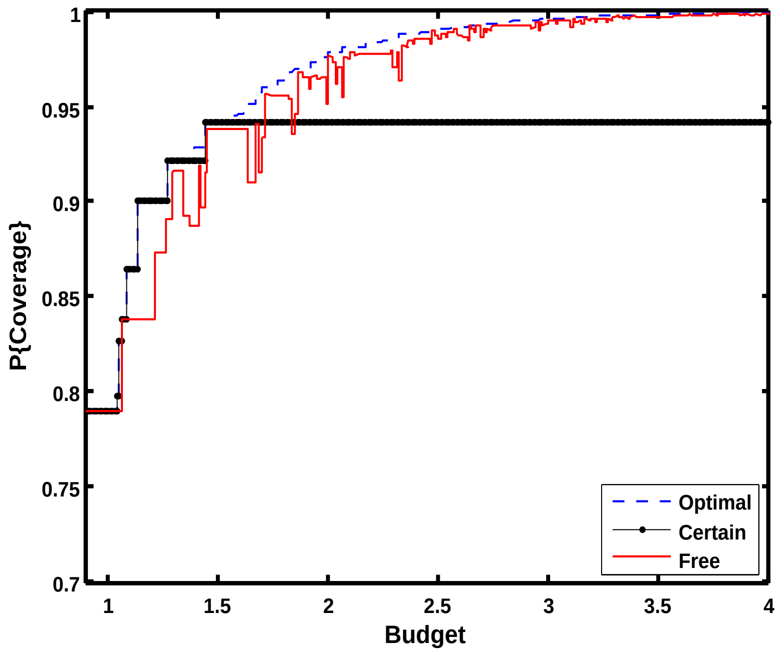


Figure 2: Instance 2

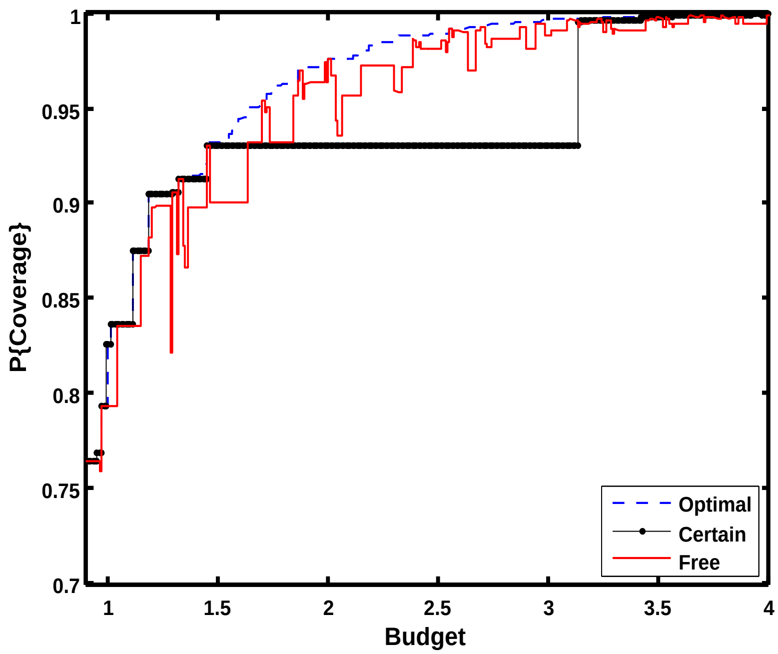


Figure 3: Instance 3

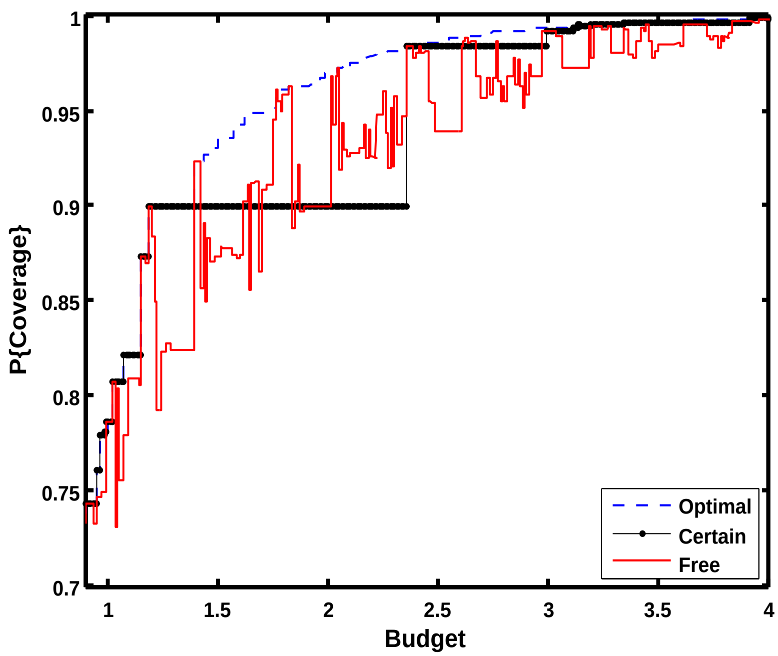


Figure 4: Instance 4

In contrast, the probability of coverage corresponding to the correlation-free and correlation-certain models does not necessarily increases monotonically since these models are optimizing different objective functions.

Instances 1 and 2 exemplify the best and worst performance of the correlation-certain model that we observed for our Bernoulli mixture instances. Figure 1 depicts an instance in which the correlation-certain model significantly outperforms the correlation-free model, while avoiding the conservatism typically found in robust models. In constrast, Figure 2 depicts an instance in which the correlation-certain model produces an optimal probability of coverage up to a budget of 1.6 before completely stagnating. Figures 3 and 4 represent average performance. What is important to remember is that the correlation-free model is making a far-fetched and altogether incorrect assumption, namely, that all correlations are 0. Hence, despite the fact that it performs well in an interval, this can be considered good fortune. For example, in Figure 4 associated with Instance 4, in the budget interval [1.3,1.7] during which the correlation-certain model is unable to find a better probability of coverage, the coverage probability associated with correlation-free model is quite unstable, dropping well below that of the correlation-certain model, while, on occasion, rising above.

Empirically, we see that for all ten instances, in the budget intervals [0,1.5] and $[3, \infty)$, the correlation-certain model consistently provided more stable and more near optimal solutions than the correlation-free model. In the budget interval [1.5,3.0], the superiority of one model over another is less clear cut. In this interval, the correlation-certain model was typically stable, but often overly conservative, whereas, the correlation-free model typically had significant oscillations, but was less suboptimal as the budget threshold b approached 3. In short, we believe these results suggest that the correlation-certain model is worthy of consideration when modeling probabilistic set covering problems with correlation uncertainty.

Finally, we answer the question posed at the outset of this subsection: What does correlation buy us? The following proposition asserts that the correlation-certain model is no worse than the correlation-agnostic model in this experiment.

Proposition 10 *For* $S \subseteq \{0,1\}^n$*,*

$$\max_{x \in S} \sum_{j \in N} \Phi_j \left(\sum_{i \in N^j} p_{ij} x_i \right) x_j \geq \max_{x \in S} \max_{j \in N} \{p_j x_j\} \ . \tag{23}$$

Proof: Let $x^* \in S$ maximize the right-hand side of (23) and let $j^* \in \arg\max_{j \in N}\{p_j x_j^*\}$. If $x_{j^*}^* = 0$, then $x^* = 0$ and the right-hand side of (23) is 0. Since 0 is feasible to the left-hand side of (23) and yields a value of 0, the inequality holds. If $x_{j^*}^* = 1$, then we can produce a new vector $x \in S$ with $x_{j^*} = 1$ and $x_k = 0$, for all $k \neq j^*$ such that $\sum_{j \in N} \Phi_j \left(\sum_{i \in N^j} p_{ij} x_i \right) x_j = \Phi_{j^*}(0) x_{j^*} = \max_{\ell = 1,\ldots,n-1}\{\frac{2p_{j^*}}{\ell+1}\} = p_{j^*}$. Since x is suboptimal to the left-hand side of (23), the inequality holds. □

In terms of our computational experiment, because the maximum probability for any individual p_j can be at most 0.9, the objective function used by the correlation-agnostic is at most 0.9, i.e., $\max_{j \in N} p_j \leq 0.9$. Meanwhile, for each instance, there is a budget threshold after which the correlation-certain model always exceeds 0.9. Moreover, in all but one of the instances presented, the correlation-certain model yields a coverage probability that approaches one as the budget increases, revealing that correlation information is helpful. This observation is not profound, but it lends support to what intuition tells us: Taking advantage of correlation information when it is available can lead to significantly better results.

5.2 Comparison of Algorithms

In our second experiment, we turn our attention away from modeling and focus on algorithmic issues associated with the correlation-certain model. Given that we have decided to use this model, we have the choice of using the large-scale MILP formulation (13), which can be fed directly into a MILP solver, or the compact formulation (21), which requires implementing a cut callback routine to separate violated lifted extended polymatroid cuts (20). Below, we compare the performance of these two approaches for PSC instances with a single ambiguous chance constraint.

The instances for this experiment were generated as follows: The n binary decision variables x_j were partitioned into consecutive blocks of k variables for $k \in \{5, 10\}$ and $n \in \{20, 30, \ldots, 70\}$. Within each block, the variables are correlated in a nested manner such that $p_{ij} = \min\{p_i, p_j\}$ implying that if the event with smaller probability occurs, then so too must the event with larger probability. Between blocks the events are independent. As before, $c_j = p_j$ for all $j \in N$.

We compare four algorithms:

- (LS) the large-scale MILP formulation (13) (denoted by LS in the tables);
- (LS+) the large-scale MILP formulation with LEP cuts added at fractional solutions through a callback;
- (C) the compact formulation (21) in which LEP cuts are added only at nodes with integral solutions;
- (C+) the compact formulation in which LEP cuts are added at nodes with both integral and fractional solutions.

The number of cuts added to each model, the number of nodes explored, and the total solution time in seconds are shown in Tables 2 and 3, with additional tables presented in Appendix D. In each table, we fix the value of ϵ and the number of variables per block as we vary the number n of decision variables in the instance. A time limit of 1800 seconds is imposed.

	# Cuts				# Nodes				Time (sec)			
n	LS	LS+	C	C+	LS	LS+	C	C+	LS	LS+	C	C+
20	0	4	22	561	10	266	217	221	15.33	2.58	**0.03**	0.19
30	0	4	24	1295	6	725	473	836	13.56	24.21	**0.06**	0.85
40	0	4	22	2466	2	1414	827	946	46.46	100.90	**0.08**	2.80
50	0	4	24	4167	5	2241	1283	1721	103.98	385.82	**0.13**	10.29
60	0	4	24	6486	14	510	1838	2358	226.42	1800.00	**0.22**	20.25
70	0	4	24	9228	272	526	2493	3266	627.79	1800.00	**0.26**	45.48

Table 2: $\epsilon = 0.05,\ k = 5$

	# Cuts				# Nodes				Time (sec)			
n	LS	LS+	C	C+	LS	LS+	C	C+	LS	LS+	C	C+
20	0	4	60	634	0	301	236	202	19.97	1.78	**0.04**	0.22
30	0	4	403	1488	34	1243	1181	461	66.15	18.19	**0.16**	0.98
40	0	4	488	2785	34	1847	1887	853	44.07	70.23	**0.28**	3.30
50	0	4	529	4623	37	2007	2593	1397	82.58	188.42	**0.44**	11.00
60	0	4	523	6936	49	1010	3392	2139	170.69	1800.00	**0.58**	21.90
70	0	4	525	9801	48	422	4336	2942	275.34	1800.00	**0.75**	43.72

Table 3: $\epsilon = 0.02,\ k = 10$

Our main finding for these instances is that, regardless of ϵ and the number of variables per block, the compact formulation in which LEP cuts are only generated at nodes with an integral solution consistently yields the fastest solution times. In general, adding LEP cuts to the large-scale MILP formulation exacerbated overall solution time and increased the number of nodes explored in the search tree. One possible explanation for this behavior is that adding cuts through a callback in CPLEX automatically disables CPLEX's dynamic search routine. As for the compact formulation, adding LEP cuts at nodes with fractional solutions increased overall solution time. Despite the fact that many more LEP cuts were generated and often reduced the number of nodes explored in the search tree, empirically it is clear that branching coupled with separating LEP cuts at nodes with integeral solutions yields better performance.

As previously stated, the large-scale MILP formulation grows large as n increases, making it more memory intensive and less tractable. Note that in these instances, we only consider a single probabilistic constraint. If there were multiple constraints, the size of the large-scale MILP formulation would grow even faster. Although not shown in the tables, LEP cuts do little to reduce the root integrality gap. In our instances, the typical root integrality gap for the compact formulation was larger than 40% and LEP cuts rarely could decrease this gap by even 1%. Finally, preliminary experimentation on the Bernoulli mixture instances of the previous subsection were consistent with our findings here.

References

[1] S. Agrawal, Y. Ding, A. Saberi, and Y. Ye. Price of correlations in stochastic optimization. Submitted to *Operations Research*, 2010.

[2] S. Ahmed and A. Atamtürk. Maximizing a class of submodular utility functions. *Mathematical Programming*, 128:149–169, 2011.

[3] A. Atamtürk and V. Narayanan. Polymatroids and mean-risk minimization in discrete optimization. *Operations Research Letters*, 36:618–622, 2008.

[4] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization.* Princeton University Press, 2009.

[5] P. Beraldi and M. E. Bruni. An exact approach for solving integer problems under probabilistic constraints with random technology matrix. *Annals of Operations Research*, 177:127–137, 2010.

[6] P. Beraldi and A. Ruszczyński. The probabilistic set covering problem. *Operations Research*, 50:956–967, 2002.

[7] G. C. Calafiore and M. C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102:25–46, 2005.

[8] J. Edmonds. Submodular functions, matroids and certain polyhedra. In: *Combinatorial Structures and their Applications*, R. Guy (Ed.), Gordon and Breach, New York, NY, 11:69–87, 1971.

[9] M. Fischetti and M. Monaci. Uncertain set covering and graph connectivity problems. Technical Report, University of Padova, 2010.

[10] M. Goemans and J. Vondrak. Stochastic covering and adaptivity. 7th Latin American Symposium on Theoretical Informatics. 3887:532–543, 2006.

[11] R. G. Haight, C. S. Revelle, and S. A. Snyder. An integer optimization approach to a probabilistic reserve site selection problem. *Operations Research*, 48:697–708, 2000.

[12] A. J. Kleywegt, A. Shapiro, and T. H. de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12:479–502, 2001.

[13] H. Kuai, F. Alajaji, and G. Takahara. A lower bound on the probability of a finite union of events. *Discrete Mathematics*, 215:147–158, 2000.

[14] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19:674–699, 2008.

[15] A. Nemirovski and A. Shapiro. Scenario approximation of chance constraints. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design Under Uncertainty*, pages 3–48. Springer, London, 2005.

[16] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17:969–996, 2006.

[17] J. Pinter. Deterministic approximations of probability inequalities. *Operations Research*, 33:219–239, 1989.

[18] A. Prékopa. Bounding the probability of the union of events by aggregation and disaggregation in linear programs. *Discrete Applied Mathematics*, 145:444–454, 2005.

[19] A. Prékopa. *Stochastic Programming.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

[20] A. Saxena, V. Goyal, and M. Lejeune. MIP reformulations of the probabilistic set covering problem. *Mathematical Programming*, 121:1–31, 2010.

[21] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory.* SIAM, Philadelphia, 2009.

[22] D. M. Topkis. *Supermodularity and complementarity.* Princeton University Press, 1998.

Appendix A

Claim 1 *Consider the LP*

$$\begin{array}{lll} \min & \sum_{i=1}^n y_i/i & \\ \text{s.t.} & \sum_{i=1}^n y_i = a & \\ & \sum_{i=1}^n (i-1)y_i = b & \\ & y_i \geq 0 & i=1,\ldots,n. \end{array} \tag{24}$$

If $n \geq 2$ and $a, b \geq 0$, then LP (24) *has an optimal objective function value of*

$$\max_{i=1,\ldots,n-1} \left\{ \frac{2a}{i+1} - \frac{b}{i(i+1)} \right\} .$$

Proof: If $n \geq 2$ and $a, b \geq 0$, then LP (24) is clearly feasible as $y_1 = a, y_2 = b, y_3 = \cdots = y_n = 0$ is a feasible solution. LP (24) is also bounded since $y_i \geq 0$ for all i. Consider the dual to LP (24):

$$\begin{array}{ll} \max & au + bv \\ \text{s.t.} & u + (i-1)v \leq \frac{1}{i} \quad i = 1, \ldots, n . \end{array} \tag{25}$$

The proof will be complete if we can show that the set E of extreme points of the dual LP (25) is given by

$$F := \left\{ \left(\frac{2}{i+1}, \frac{-1}{i(i+1)} \right) \right\}_{i=1}^{n-1} .$$

Note that each of the n constraints of the dual (25) are linearly independent. Moreover, any extreme point of (25) lies at the intersection of exactly two constraints of (25).

To prove that $F \subseteq E$, consider a point $\left(\frac{2}{i+1}, \frac{-1}{i(i+1)}\right)$ in F. This point lies at the intersection of constraints i and $i+1$ of (25). To verify that this point is feasible to (25), we check that this point satisfies each of the remaining constraints $k = 1, \ldots, n$ with $k \neq i$. Note that for the kth constraint, we need to verify that

$$\frac{2}{i+1} - \frac{k-1}{i(i+1)} \leq \frac{1}{k} \iff \frac{(i-k)(i-k+1)}{ik(i+1)} \geq 0 .$$

If $i \leq k-1$, then $i-k \leq -1$ and $i-k+1 \leq 0$. Hence, $(i-k)(i-k+1) \geq 0$ and constraint k is satisfied. If $i \geq k+1$, then $i-k \geq 1$ and $i-k+1 \geq 2$. Hence, $(i-k)(i-k+1) \geq 0$ and constraint k is satisfied. Therefore, $F \subseteq E$.

To prove that $E \subseteq F$, consider an extreme point $(\widehat{u}, \widehat{v}) \in E$ and, to arrive at a contradiction, suppose $(\widehat{u}, \widehat{v}) \notin F$. That is, $(\widehat{u}, \widehat{v})$ lies at the intersection of constraints k and $i+1$ with $k \leq i-1$ for some $i \in \{2, \ldots, n-1\}$ in (25):

$$\begin{array}{ccccc} \widehat{u} & + & (k-1)\widehat{v} & = & \frac{1}{k} \\ \widehat{u} & + & i\widehat{v} & = & \frac{1}{i+1} \end{array} .$$

Solving this system yields $\widehat{u} = \frac{k+i}{k(i+1)}$ and $\widehat{v} = \frac{-1}{k(i+1)}$. However, $(\widehat{u}, \widehat{v})$ does not satisfy constraint i since

$$\widehat{u} + (i-1)\widehat{v} = \frac{k+i}{k(i+1)} - \frac{i-1}{k(i+1)} = \frac{k+1}{k(i+1)} > \frac{i+1}{i(i+1)} = \frac{1}{i} ,$$

where the strict inequality follows from our assumption that $1 \leq k < i$, which implies that $\frac{k+1}{k} > \frac{i+1}{i}$. This contradicts our initial assumption that $(\widehat{u}, \widehat{v}) \in E$. Finally, it is easy to see that adjacent constraints in the dual give rise to F. □

Appendix B

In this appendix, we analyze the polyhedral structure of the convex hull of the set

$$Y := \{(w, x, y) \in \mathbb{R} \times \{0,1\} \times \{0,1\}^n : 0 \leq w \leq f(y)x\} , \tag{26}$$

where $f : \{0,1\}^n \mapsto \mathbb{R}_{++}$ is a nonincreasing function. Note that the set Y_j given by (15) is a special case of (26). We use the following notation throughout: $N = \{1, \ldots, n\}$, $Y_0 = Y \cap \{x : x = 0\}$, $Y_1 = Y \cap \{x : x = 1\}$ and $Y_1^S = Y_1 \cap \{y \in \{0,1\}^n : \ y_j = 0 \ \forall \ j \notin S\}$ for all $S \subseteq N$. Finally we use e^j to denote a unit vector (of appropriate dimension) with a one in the j-th position.

Proposition 11 *For any $S \subseteq N$,*

$$\dim(\text{conv}(Y_1^S)) = |S| + 1 .$$

Proof: The dimension is at most $|S|+1$. Moreover the $|S|+2$ affinely independent points $(0,0)$, $(f(0),0)$, and $\{(0, e^j)\}_{j \in S}$ are all feasible. □

Proposition 12 *For any $S \subseteq N$, the (trivial) bound inequalities $0 \leq w$ and $0 \leq y_j \leq 1$ for all $j \in S$ are valid and facet defining for $\text{conv}(Y_1^S)$.*

Proof: The inequalities are clearly valid. The $|S|+1$ affinely independent feasible points $\{(0, e^j)\}_{j \in S}$ and $(0,0)$ satisfy $0 \leq w$ at equality. For any $j \in S$, the $|S|+1$ affinely independent feasible points $\{(0, e^k)\}_{k \in S, k \neq j}$, $(f(0), 0)$ and $(0,0)$ satisfy $0 \leq y_j$ at equality. Finally, for any $j \in S$, the $|S|+1$ affinely independent feasible points $\{(0, e^j + e^k)\}_{k \in S, k \neq j}$, $(f(e^j), e^j)$, and $(0, e^j)$ satisfy $y^j \leq 1$ at equality. □

Proposition 13 *For any $S \subseteq N$,*

$$\text{conv}(Y_1^S) = \{(w, y) \in \mathbb{R}_+ \times [0,1]^{|S|} : \ w + \sum_{j \in S} a_j^i y_j \leq b^i \ \forall \ i \in M\} ,$$

with $b^i > 0$ for all $i \in M$ and $a_j^i \geq 0$ for all $i \in M$ and $j \in S$, where M is some index set of constraints.

Proof: We first establish that if $\gamma w + \sum_{j \in S} \alpha_j y_j \leq \beta$ is a nontrivial facet of $\text{conv}(Y_1^S)$ then $\gamma > 0$, $\beta > 0$, and $\alpha_j \geq 0$ for all $j \in S$. Suppose $\gamma < 0$ and consider a point $(\widehat{w}, \widehat{y}) \in \text{conv}(Y_1^S)$ with $\widehat{w} > 0$ that satisfies the given inequality at equality. Such a point exists, otherwise the given inequality is identical to the trivial facet $w \geq 0$. But then the point $(0, \widehat{y}) \in \text{conv}(Y_1^S)$ and is cut off by the given inequality. Thus $\gamma \geq 0$. Suppose $\alpha_j < 0$ for some $j \in S$ and consider a point $(\widehat{w}, \widehat{y}) \in \text{conv}(Y_1^S)$ with $\widehat{y}_j > 0$ that satisfies the given inequality at equality. Such a point exists, otherwise the given inequality is identical to the trivial facet

$y_j \geq 0$. But then the point $(\widehat{w}, \widehat{y} - \widehat{y}_j e^j) \in \text{conv}(Y_1^S)$ (since $\widehat{w} \leq f(\widehat{y}) \leq f(\widehat{y} - \widehat{y}_j e^j)$ by the nonincreasing property of f) and is cut off by the given inequality. Thus $\alpha_j \geq 0$. It also follows that $\beta \geq \sum_{j \in S} \alpha_j \geq 0$ since $(0, \sum_{j \in S} e^j) \in Y_1^S$. Now if $\gamma = 0$ then the given inequality is implied by non-negative combinations of the trivial facets $y_j \leq 1$ for all j since $\sum_{j \in S} \alpha_j y_j \leq \sum_{j \in S} \alpha_j \leq \beta$. Thus $\gamma > 0$. It follows that $\beta \geq \gamma f(0) > 0$ since $(f(0), 0) \in Y_1^S$. Thus any nontrivial facet of $\text{conv}(Y_1^S)$ is of the form $w + \sum_{j \in S} a_j y_j \leq b$. The claim then follows from Proposition 12. □

Proposition 14

$$\dim(\text{conv}(Y)) = n + 2 \ .$$

Proof: The dimension is at most $n + 2$. Moreover the $n + 3$ affinely independent points $\{(0, 0, e^j)\}_{j=1}^n$, $(0, 0, 0)$, $(0, 1, 0)$, and $(f(0), 1, 0)$, are feasible. □

Proposition 15 *The (trivial) bound inequalities $0 \leq w$, $x \leq 1$ and $0 \leq y_j \leq 1$ for all j are valid and facet defining for* $\text{conv}(Y)$.

Proof: The inequalities are clearly valid. The $n + 2$ affinely independent points $\{(0, 0, e^j)\}_{j=1}^n$, $(0, 0, 0)$ and $(0, 1, 0)$ satisfy $0 \leq w$ at equality. The $n+2$ affinely independent points $\{(0, 1, e^j)\}_{j=1}^n$, $(0, 1, 0)$ and $(f(0), 1, 0)$ satisfy $x \leq 1$ at equality. The $n + 2$ affinely independent points $\{(0, 1, e^k)\}_{k=1, k \neq j}^n$, $(f(0), 1, 0)$, $(0, 1, 0)$ and $(0, 0, 0)$ satisfy $0 \leq y_j$ at equality. Finally, the $n + 2$ affinely independent points $\{(0, 1, e^j + e^k)\}_{k=1, k \neq j}^n$, $(f(e^j), 1, e^j)$, $(0, 1, e^j)$ and $(0, 0, e^j)$ satisfy $y^j \leq 1$ at equality. □

Proposition 16 *If $\gamma w + \tau(1 - x) + \sum_j \alpha_j y_j \leq \beta$ is a nontrivial facet of* $\text{conv}(Y)$ *then $\gamma > 0$, $\beta > 0$, $\alpha_j \geq 0$ for all j and $\tau = \beta - \sum_j \alpha_j \geq 0$.*

Proof: Suppose $\gamma < 0$ and consider a point $(\widehat{w}, \widehat{x}, \widehat{y}) \in \text{conv}(Y)$ with $\widehat{w} > 0$ that satisfies the given inequality at equality. Such a point exists, otherwise the given inequality is identical to the trivial facet $w \geq 0$. But then the point $(0, \widehat{x}, \widehat{y}) \in \text{conv}(Y)$ and is cut off by the given inequality. Thus $\gamma \geq 0$. Suppose $\tau < 0$ and consider a point $(\widehat{w}, \widehat{x}, \widehat{y}) \in \text{conv}(Y)$ with $\widehat{x} < 1$ that satisfies the given inequality at equality. Such a point exists, otherwise the given inequality is identical to the trivial facet $x \leq 1$. But then the point $(\widehat{w}, 1, \widehat{y}) \in \text{conv}(Y)$ and is cut off by the given inequality. Thus $\tau \geq 0$. Similarly suppose $\alpha_j < 0$ and consider a point $(\widehat{w}, \widehat{x}, \widehat{y}) \in \text{conv}(Y)$ with $\widehat{y}_j > 0$ that satisfies the given inequality at equality. Such a point exists, otherwise the given inequality is identical to the trivial facet $y_j \geq 0$. But then the point $(\widehat{w}, \widehat{x}, \widehat{y} - \widehat{y}_j e^j) \in \text{conv}(Y)$ and is cut off by the given inequality. Thus $\alpha_j \geq 0$. Note that since $(0, 0, 1, \ldots, 1) \in Y$ it follows that $\tau + \sum_j \alpha_j \leq \beta$. Now if $\gamma = 0$ then the given inequality is implied by non-negative combinations of the trivial facets $(1 - x) \leq 1$ and $y_j \leq 1$ for all j since $\tau(1 - x) + \sum_j \alpha_j y_j \leq \tau + \sum_j \alpha_j \leq \beta$. Thus $\gamma > 0$. It follows that $\beta \geq \gamma f(0) > 0$ since $(f(0), 1, 0) \in Y$. Now $\gamma w + \sum_j \alpha_j y_j \leq \beta$ is a valid inequality for $\text{conv}(Y_1)$. Consider the lifting of this inequality to the valid inequality $\gamma w + t(1 - x) + \sum_j \alpha_j y_j \leq \beta$. The lifting coefficient $t = \min\{\beta - \sum_j \alpha_j y_j : \ y \in \{0, 1\}^n\} = \beta - \sum_j \alpha_j$. Since the given inequality is a facet we must have $\beta - \sum_j \alpha_j = t \leq \tau \leq \beta - \sum_j \alpha_j$ thus $\tau = \beta - \sum_j \alpha_j$. □

Lemma 1 *If $w + \sum_{j \in S} a_j y_j \leq b$ is a valid inequality for Y_1^S then $w + (b - \sum_{j \in S} a_j)(1 - x) + \sum_{j \in S} a_j y_j \leq b$ is valid for Y.*

Proof: Consider a point $(\widehat{w}, \widehat{x}, \widehat{y}) \in Y$. If $\widehat{x} = 0$ then $\widehat{w} = 0$ and $\sum_{j \in S} a_j y_j \leq \sum_{j \in S} a_j$ trivially. If $\widehat{x} = 1$ then construct a point y' such that $y'_j = \widehat{y}_j$ for $j \in S$ and zero otherwise. By the nonincreasing property of f, we have that $\widehat{w} \leq f(\widehat{y}) \leq f(y')$ and so $(\widehat{w}, y') \in Y_1^S$ and so satisfies $\widehat{w} + \sum_{j \in S} a_j \widehat{y}_j \leq b$.

Proposition 17 *The inequality $w + (b - \sum_{j \in N} a_j)(1 - x) + \sum_{j \in N} a_j y_j \leq b$ is a nontrivial facet of* $\text{conv}(Y)$ *if and only if $w + \sum_{j \in S} a_j y_j \leq b$ is a facet of* $\text{conv}(Y_1^S)$ *for $S = \{j \in N : \ a_j > 0\}$.*

Proof:
($\Leftarrow$) Suppose $w + \sum_{j \in S} a_j y_j \leq b$ is a facet of $\text{conv}(Y_1^S)$. By Lemma 1 the inequality $w + (b - \sum_{j \in N} a_j)(1 - x) + \sum_{j \in N} a_j y_j \leq b$ is valid for Y. Suppose $|S| = k$, thus $|N \setminus S| = n - k$. Now since $w + \sum_{j \in S} a_j y_j \leq b$ is a facet of $\text{conv}(Y_1^S)$ there are $k+1$ affinely independent feasible points $\{(w^i, y^i)\}_{i=1}^{k+1}$ that satisfy $w^i \leq f(y^i)$, $y_j^i = 0$ for all $j \notin S$, and $w^i + \sum_{j \in S} a_j y_j^i = b$ for all $i = 1, \ldots, k+1$. Now consider the points $(w, x, y) := \{\{(w^i, 1, y^i)\}_{i=1}^{k+1}, \{(0, 0, e^l + \sum_{j \in S} e^j)\}_{l \in N \setminus S}, (0, 0, \sum_{j \in S} e^j)\}$. These are $n+2$ affinely independent points, each of which satisfy $0 \leq w \leq f(y)x$ and the inequality $w + (b - \sum_{j \in S} a_j)(1-x) + \sum_{j \in S} a_j y_j \leq b$ at equality. Hence $w + (b - \sum_{j \in N} a_j)(1 - x) + \sum_{j \in N} a_j y_j \leq b$ is a facet of $\text{conv}(Y)$.

($\Rightarrow$) Note that $w + \sum_{j \in S} a_j y_j \leq b$ is valid for Y_1 and therefore valid for Y_1^S. In fact the inequality is strong in the sense that there is at least one point in $\text{conv}(Y_1^S)$ for which it is tight. Indeed, consider a point $(\widehat{w}, \widehat{x}, \widehat{y}) \in Y$ for which $w + (b - \sum_{j \in N} a_j)(1 - x) + \sum_{j \in N} a_j y_j \leq b$ is tight and $\widehat{x} = 1$ (such a point exists since the facet is nontrivial). Construct y' such that $y'_j = \widehat{y}_j$ for $j \in S$ and zero otherwise. By the nonincreasing property of f, we have that $\widehat{w} \leq f(\widehat{y}) \leq f(y')$ and so $(\widehat{w}, y') \in Y_1^S$ and $\widehat{w} + \sum_{j \in S} a_j y'_j = b$. Suppose now that $w + \sum_{j \in S} a_j y_j \leq b$ is not a facet of $\text{conv}(Y_1^S)$. Let

$$\begin{array}{rcll} w + \sum_{j \in S} \alpha_j^i y_j & \leq & \beta^i & \forall\, i \in M \\ -w & \leq & 0 & \\ -y_j & \leq & 0 & \forall\, j \in S \\ y_j & \leq & 1 & \forall\, j \in S \end{array}$$

be the complete description of $\text{conv}(Y_1^S)$. (Note that by Proposition 13 the complete description will be of this form). Taking a non-negative combination of the above inequalities, we obtain

$$\left(\sum_{i \in M} \lambda_i - \mu\right) w + \sum_{j \in S} \left(\sum_{i \in M} \lambda_i \alpha_j^i - \gamma_j + \tau_j\right) y_j \leq \sum_{i \in M} \lambda_i \beta^i + \sum_{j \in S} \tau_j.$$

Since $w + \sum_{j \in S} a_j y_j \leq b$ is not a facet of Y_1^S (we know that it is a strong valid inequality), then there exists nonnegative multipliers such that

$$\begin{array}{rcll} \sum_{i \in M} \lambda_i - \mu & = & 1 & \\ \sum_{i \in M} \lambda_i \alpha_j^i - \gamma_j + \tau_j & = & a_j & \forall\, j \in S \\ \sum_{i \in M} \lambda_i \beta^i + \sum_{j \in S} \tau_j & = & b\,. & \end{array} \tag{27}$$

Since $\gamma_j \geq 0$, we have that $\sum_{i \in M} \lambda_i \alpha_j^i + \tau_j \geq a_j$ for all $j \in S$. We consider the following two cases.

Case 1: Suppose $\sum_{i \in M} \lambda_i \alpha_j^i + \tau_j = a_j$ for all $j \in S$. By Lemma 1, the inequalities

$$w + \left(\beta^i - \sum_{j \in S} \alpha_j^i\right)(1 - x) + \sum_{j \in S} \alpha_j^i y_j \leq \beta^i \qquad \forall\, i \in M \tag{28}$$

are valid for Y. Combining the above inequalities (using multipliers λ_i) with the valid inequalities $y_j \leq 1$ for $j \in S$ (using multipliers τ_j) and $-w \leq 0$ (using multiplier μ) for $\text{conv}(Y)$ and using (27), we obtain the inequality $w + (b - \sum_{j \in N} a_j)(1 - x) + \sum_{j \in N} a_j y_j \leq b$ which then cannot be a facet.

Case 2: Suppose there exists $k \in S$ such that $\sum_{i \in M} \lambda_i \alpha_k^i + \tau_k > a_k$. Combining the inequalities

$$w + \sum_{j \in S} \alpha_j^i y_j \leq \beta^i \qquad \forall\, i \in M$$

(using multipliers λ_i) with valid inequalities $y_j \leq 1$ for $j \in S$ (using multipliers τ_j), we have that

$$w + \sum_{j \in S} \left(\sum_{i \in M} \lambda_i \alpha^i_j + \tau_j \right) y_j \leq \sum_i \lambda_i \beta^i + \sum_{j \in S} \tau_j = b$$

is valid for Y_1^S, and (by weakening the coefficients) so are the two inequalities

$$w + \sum_{j \in S \setminus \{k\}} a_j y_j + \left(\sum_{i \in M} \lambda_i \alpha^i_k + \tau_k \right) y_k \leq b$$

and

$$w + \sum_{j \in S \setminus \{k\}} a_j y_j + (0) y_k \leq b \ .$$

By Lemma 1, the following inequalities are valid for Y

$$w + \left(b - \sum_{j \in S \setminus \{k\}} a_j - \sum_{i \in M} \lambda_i \alpha^i_k - \tau_k \right) (1 - x) + \sum_{j \in S \setminus \{k\}} a_j y_j + \left(\sum_{i \in M} \lambda_i \alpha^i_k + \tau_k \right) y_k \leq b \tag{29}$$

and

$$w + \left(b - \sum_{j \in S \setminus \{k\}} a_j - 0 \right) (1 - x) + \sum_{j \in S \setminus \{k\}} a_j y_j + (0) y_k \leq b. \tag{30}$$

Multiplying (29) by $a_k / (\sum_{i \in M} \lambda_i \alpha^i_k + \tau_k)$ and (30) by $1 - a_k / (\sum_{i \in M} \lambda_i \alpha^i_k + \tau_k)$ and summing we obtain $w + (b - \sum_{j \in S} a_j)(1 - x) + \sum_{j \in S} a_j y_j \leq b$, thus it cannot be a facet. □

It follows that conv(Y) is completely described by the trivial facets $0 \leq w$, $x \leq 1$, $0 \leq y_j \leq 1$ for $j \in N$, and the nontrivial facets

$$w + (b^i - \sum_{j \in N} a^i_j)(1 - x) + \sum_{j \in N} a^i_j y_j \leq b^i \ \forall \ i \in I$$

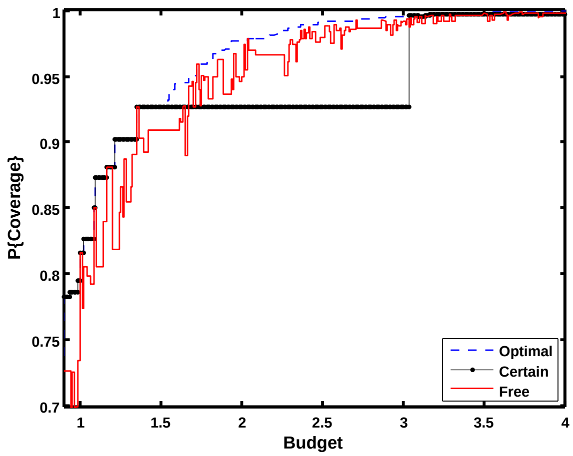
such that $b^i > 0$ for all i, $a^i_j \geq 0$ for all i and j, and for each i

$$w + \sum_{j \in S_i} a^i_j y_j \leq b^i \ ,$$

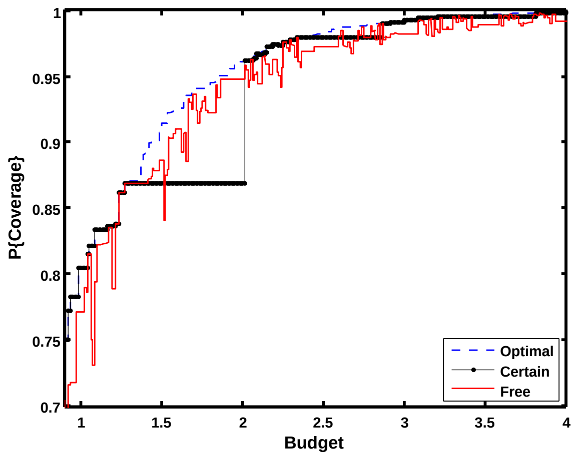
where $S_i = \{j \in N : \ a^i_j > 0\}$ is a nontrivial facet of conv($Y_1^{S_i}$). □

Appendix C

Here we present the figures associated with the remaining six instances not discussed in the main text.

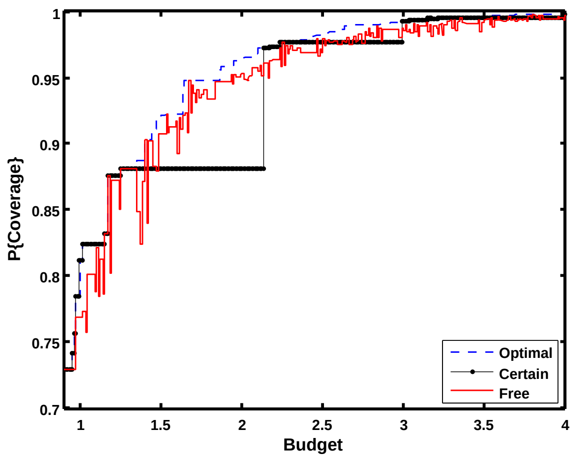


(a) Instance 5

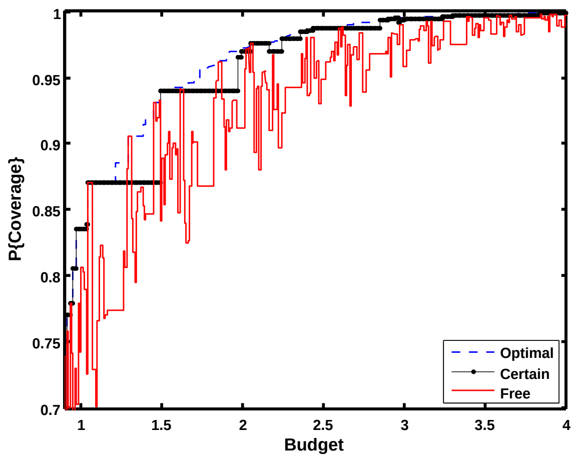


(b) Instance 6

Figure 5: Instances 5 and 6

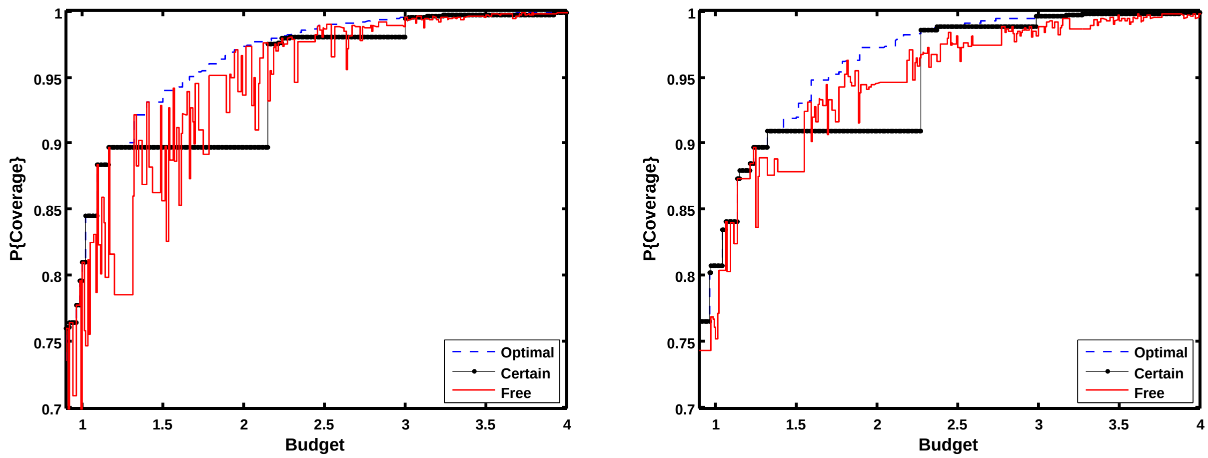


(a) Instance 7



(b) Instance 8

Figure 6: Instances 7 and 8



(a) Instance 9 (b) Instance 10

Figure 7: Instances 9 and 10

Appendix D

	# Cuts				# Nodes				Time (sec)			
n	LS	LS+	C	C+	LS	LS+	C	C+	LS	LS+	C	C+
20	0	4	58	566	36	261	235	208	17.86	1.75	**0.02**	0.17
30	0	4	62	1335	36	689	492	493	24.47	16.83	**0.06**	0.85
40	0	4	64	2573	34	1339	848	946	41.86	63.51	**0.08**	2.98
50	0	4	66	4118	37	890	1304	2295	58.61	1800.00	**0.16**	7.76
60	0	4	60	6444	37	152	1856	2300	337.83	1800.00	**0.18**	19.48
70	0	4	66	9338	631	529	2514	3191	517.65	1800.00	**0.28**	39.35

Table 4: $\epsilon = 0.05$, $k = 10$

	# Cuts				# Nodes				Time (sec)			
n	LS	LS+	C	C+	LS	LS+	C	C+	LS	LS+	C	C+
20	0	4	22	643	8	255	217	203	12.00	1.75	**0.02**	0.21
30	0	4	255	1442	5	710	741	476	13.60	15.50	**0.12**	0.96
40	0	4	305	2776	2	1383	1126	890	28.27	66.22	**0.20**	3.78
50	0	4	368	4647	5	2010	1703	1462	54.62	193.73	**0.24**	10.35
60	0	4	376	7009	14	254	2357	2217	140.03	1800.00	**0.33**	19.75
70	0	4	386	10027	32	422	3175	3079	330.16	1800.00	**0.46**	39.41

Table 5: $\epsilon = 0.02$, $k = 5$