We consider two variants of a probabilistic set covering (PSC) problem. The first variant assumes that 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. The second variant seeks to maximize the minimum probability that a selected set can cover all items. To date, literature on this problem has focused on the special case in which uncertainties are independent. In this paper, we formulate deterministic mixed-integer programming models for distributionally robust PSC problems with correlated uncertainties. By exploiting the supermodularity of certain substructures and analyzing their polyhedral properties, we develop strong valid inequalities to strengthen the formulations. Computational results illustrate that our modeling approach can outperform formulations in which correlations are ignored and that our algorithms can significantly reduce overall computation time.

Subject classifications: distributionally robust models; integer programming; set covering; stochastic programming; supermodularity.

Area of review: Optimization.

History: Received July 2011; revisions received March 2012; accepted July 2012.

1. Introduction

We consider the following probabilistic set covering (PSC) problem

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

where \( N := \{1, \ldots, n\} \) is a collection of columns (or sets), \( M := \{1, \ldots, m\} \) is a collection of rows (or items), \( c \in \mathbb{R}^n \), \( \tilde{a}_{ij} \) 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 \( B \) 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:
Ahmed and Papageorgiou: Probabilistic Set Covering with Correlations

Article submitted to Operations Research; manuscript no. OPRE-2011-07-370.R2

\[
\max_{w,x} w \quad \text{s.t.} \quad \mathbb{P} \left[ \sum_{j \in N} \tilde{a}_{ij} x_j \geq 1 \right] \geq w \quad \forall \ i \in M
\]

(2)

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

As an example of a probabilistic set covering problem, 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 probabilistic or 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 (1995) and Shapiro et al. (2009) 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 (2006a) and Pintér (1989) and robust optimization (Ben-Tal et al. 2009). 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 (Kleywegt et al. 2002). The scenario approximation method of Calafiore and Campi (2005) and its extensions by Nemirovski and Shapiro (2006b) require that
all of the constraints corresponding to the sample taken must be satisfied, whereas the sample average approximation of Luedtke and Ahmed (2008) 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 (2002) and Saxena et al. (2010) 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., \( \min \{ cx : P[Ax \geq \xi] \geq 1 - \epsilon, x \in \{0,1\}^n \} \), where \( \xi \) is a random 0-1 vector whose components may be correlated. Haight et al. (2000) formulated a variant of (1) in which all constraint coefficients \( \tilde{a}_{ij} \) are assumed to be independent. Fischetti and Monaci (2012) investigated what they call the Uncertain Probabilistic Set Covering Problem in which the columns \( \tilde{a}_j \) are assumed to be independent. Goemans and Vondrák (2006) 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 (2010) 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 (2010), 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. (2012) characterize the repercussions of overlooking correlations in the context of distributionally robust stochastic programming (DRSP) by introducing a quantity they term the \textit{price of correlations}. 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 a family \( \mathcal{D} \) of distributions that satisfy some limited distributional
assumptions, e.g., all distributions with first moment $\nu$. In other words, $g(x)$ is the expected cost under the worst-case distribution in the family $D$. They define the price of correlations as the ratio

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

where $x_I$ is an optimal decision for distributions in $D$ that assume independent uncertainties, and $x_R$ is an optimal decision for the DRSP model. A small price of correlations suggests that a modeler may incur only a small cost when using the product distribution, which assumes independent uncertainties. Indeed, Agrawal et al. (2012) provide a class of instances for the probabilistic set covering problem with $n$ columns whose price of correlations is $\Omega\left(\sqrt{n \log \frac{\log n}{\log \log n}}\right)$, which indicates the potential gravity of ignoring the dependence among the data.

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

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

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

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

and

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

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 $p_{ij} = \mathbb{P}[\tilde{a}_i = 1 \land \tilde{a}_j = 1]$ 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 $\rho_{ij}$ are known exactly since

$$\rho_{ij} = \frac{\text{Cov}(\tilde{a}_i, \tilde{a}_j)}{\sqrt{\text{Var}(\tilde{a}_i)\text{Var}(\tilde{a}_j)}} = \frac{p_{ij} - p_i p_j}{\sqrt{p_i(1-p_i) p_j(1-p_j)}} \quad \forall \ i, j \in N.$$
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 a scenario-free investigation of the probabilistic set covering problem with correlated data in the constraint matrix. 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.

2. 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 what we call our “correlation-certain” model, in which correlations are known, can not only outperform a model in which correlations are completely ignored, but can also maintain 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 \( \tilde{a}_{ij} \), we assume throughout that \( p_j = \Pr(\tilde{a}_j = 1) \in (0,1) \) and \( q_j = 1 - p_j \) for all \( j \in N \). At times, we will express \( p(x) = \Pr\left[\sum_{j \in N} \tilde{a}_j x_j \geq 1\right] \) as \( \Pr[\cup_{j \in N} A_j(x)] \) or as \( 1 - \Pr[\cap_{j \in N} A^c_j(x)] \) where \( A_j(x) = \{ \tilde{a}_j x_j = 1 \} \) and \( A^c_j(x) = \{ \tilde{a}_j x_j = 0 \} \) for all \( j \in N \). We define \( N^\setminus j := N \setminus \{ j \} \) 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 Fischetti and Monaci (2012) and Haight et al. (2000). We include it here for the sake of completeness.
Let \( \tilde{a}_j \) be independent Bernoulli random variables with \( p_j = P[\tilde{a}_j = 1] \in (0, 1) \) and \( q_j = 1 - p_j \) for all \( j \in N \). Then

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

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

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

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:

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

where \( q_{ij} = 1 - p_{ij} \). Similarly, noting that maximizing \( w \) is equivalent to minimizing \( \log(1 - w) \), 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 \( p_{\ell i} = P[\tilde{a}_i = 1|\ell] \) and \( p_{\ell j} = P[\tilde{a}_j = 1|\ell] \) are independent for all \( i, j \in N \) and \( \ell \in \{1, \ldots, L\} \), and the posterior distribution is given by \( p_j = \sum_{\ell=1}^L \pi_{\ell} p_{\ell j} \) for all \( j \). Let \( q_{\ell j} = 1 - p_{\ell j} \in (0, 1) \) for all \( j \) and \( \ell \). Then

\[
X = \left\{ (w, x) \in [0, 1] \times \{0, 1\}^n : \exists y \in \mathbb{R}_+^L \text{ s.t. } \sum_{\ell=1}^L \pi_{\ell} y_{\ell} \leq 1 - w, \ y_\ell \geq \prod_{j \in N} q_{\ell j}^{x_{j\ell}} \forall \ell = 1, \ldots, L \right\}.
\]
Proof. As in Proposition 1, we have
\[ P \left[ \bigcup_{j \in N} A_j(x) \right] \geq w \iff P \left[ \bigcap_{j \in N} A_j^c(x) \right] \leq 1 - w \]
\[ \iff \sum_{\ell=1}^{L} \pi_{\ell} P \left[ \bigcap_{j \in N} A_j^c(x) \mid \ell \right] \leq 1 - w \iff \sum_{\ell=1}^{L} \pi_{\ell} \prod_{j \in N} q_{j}^{x_{j}} \leq 1 - w. \]
Introducing a new variable \( y_{\ell} \) to model the probability of failure \( \prod_{j \in N} q_{j}^{x_{j}} \) under scenario \( \ell \), we obtain the desired reformulation. □

The set \( X \) in (5) is a mixed-integer nonlinear set. The nonlinear constraint \( y_{\ell} \geq \prod_{j \in N} q_{j}^{x_{j}} \) in (5) can be written as
\[ y_{\ell} \geq \prod_{j=1}^{n} q_{j}^{x_{j}} = \exp \left\{ \sum_{j \in N} (\log q_{j}) x_{j} \right\} =: \sigma_{\ell}(x) \quad \forall \, \ell = 1, \ldots, L. \]
Since \( \sigma_{\ell} \) is convex in \( x \) for all \( \ell \), the mixed-integer nonlinear system (5) can be linearized by outer approximating \( \sigma_{\ell} \) 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 \( \sigma_{\ell} \) can be expressed as
\[ \sigma_{\ell}(x) = f \left( \sum_{j \in N} d_{j} x_{j} \right) \]
where \( f(\cdot) = \exp(\cdot) \) and \( d_{j} = \log q_{j} \), i.e., as the composition of a univariate convex function and a linear function. In recent work, Ahmed and Atamtürk (2011) studied mixed binary nonlinear systems of the form \( w \geq f \left( \sum_{j \in N} d_{j} x_{j} \right) \), 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 \( p_j = P[\tilde{a}_j = 1] \) for all \( j \) are known, or (b) the marginal and pair-wise marginal probabilities \( p_{ij} = P[\tilde{a}_i = 1 \land \tilde{a}_j = 1] \) 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 \( \mu \) with moments corresponding to the specified marginals, and consider the Ambiguous Probabilistic Set Covering set
\[ X^A = \left\{ (w, x) \in [0, 1] \times \{0, 1\}^n : \min_{\mu \in \mathcal{D}} \left\{ P_{\mu} \left[ \sum_{j \in N} a_j x_j \geq 1 \right] \right\} \geq w \right\}, \]
where $\mathbb{P}_\mu[A]$ denotes the probability of the event $A$ under distribution $\mu$. Ambiguous chance constraints for robust optimization are discussed in Erdogan and Iyengar (2006). 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 $p_j = \mathbb{P}[\bar{a}_j = 1]$ for all $j$ are known. Let $\mathcal{D}_1$ be the family of all $n$-variate Bernoulli distributions $\mu$ with moments corresponding to the specified marginals, where the subscript 1 denotes that only first moment information for the random vector $\bar{a}$ 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

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

**Proof.** Given $x \in \{0, 1\}^n$, note that $\mathbb{P}_\mu[A_j(x)] = p_j x_j$ for all $\mu \in \mathcal{D}_1$ and for all $j \in N$. As $\mathbb{P}_\mu[\bigcup_{j \in N} A_j(x)] \geq \max_{j \in N} \mathbb{P}_\mu[A_j(x)] = \max_{j \in N} \{p_j x_j\}$ 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 $\mathbb{P}_{\mu'}[A_j(x) | A_1(x)] = 0$, for all $j = 2, \ldots, n$. Then, $\mathbb{P}_{\mu'}[\bigcup_{j \in N} A_j(x)] = \mathbb{P}[A_1(x)]$ since $\bigcup_{j=2}^n A_j(x) \subseteq A_1(x)$. Moreover, $w > \mathbb{P}[A_1(x)] = \mathbb{P}_{\mu'}[\bigcup_{j \in N} A_j(x)] \geq \min_{\mu \in \mathcal{D}_1} \mathbb{P}_\mu[\bigcup_{j \in N} A_j(x)]$, and so $(w, x) \notin X^A$. Thus, $X^A = X_1$. \qed

**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:

$$\sum_{j \in N} p_j z_j \geq w \quad (8a)$$

$$\sum_{j \in N} z_j \leq 1 \quad (8b)$$

$$z_j \leq x_j \quad \forall j \in N.$$  

$$\sum_{j \in N} z_j \leq 1 \quad (8c)$$
Proof. If \( x \) and \( z \) were constrained to be binary vectors, it is easy to see that a point \((\hat{w}, \hat{x})\) 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 \( p_{ij} = P[\tilde{a}_i = 1 \land \tilde{a}_j = 1] \) for all \( i \) and \( j \) are known and then provide deterministic reformulations of the ambiguous covering set \( X^A \) given by (6). Let \( D_2 \) be the family of all \( n \)-variate Bernoulli distributions \( \mu \) with moments corresponding to 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. (2000) 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 result 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 \( D = D_2 \), the ambiguous covering set \( X^A \) given by (6) can be reformulated into the following deterministic 0-1 nonlinear set

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

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

\[
\min_{\mu \in D_2} \left\{ \mathbb{P}_\mu \left[ \bigcup_{j \in N} A_j(x) \right] \right\} = \sum_{j \in N} \Phi_j \left( \sum_{i \in N} p_{ij} x_i \right) x_j \quad \forall \ x \in \{0,1\}^n.
\]

It has been shown (see the proof of Theorem 1 in Kuai et al. (2000), see also Prékopa and Gao (2005)) that \( \min_{\mu \in D_2} \{ \mathbb{P}_\mu [ \bigcup_{j \in N} A_j(x) ] \} \) is equal to the optimal value of the linear program

\[
\min \sum_{i \in N} \sum_{j \in N} y_{ij} / i \\
\text{subject to:} \sum_{i \in N} y_{ij} = a_j \quad \forall \ j \in N \\\n\sum_{i \in N} (i - 1) y_{ij} = b_j \quad \forall \ j \in N \\\ny_{ij} \geq 0 \quad \forall i, j \in N.
\]
where \( a_j = P_\mu[A_j(x)] = p_j x_j \) and \( b_j = \sum_{k \in N_j} P_\mu[A_j(x) \cap A_k(x)] = \sum_{k \in N_j} p_k x_k \) for all \( \mu \in \mathcal{D}_2 \). Note that (11) is separable in \( j \). We prove the following claim in the Appendix.

**Claim 1.** Consider the LP

\[
\begin{align*}
\min \sum_{i \in N} y_i / i \\
\text{s.t.} \sum_{i \in N} y_i = a \\
\sum_{i \in N} (i - 1) y_i = b \\
y_i \geq 0 \quad \forall \ i \in N.
\end{align*}
\]

(12)

If \( n \geq 2 \) and \( a \geq b \geq 0 \), 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

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

**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 \((\lambda, v, w, w', x)\) satisfying:

\[
w'_j \leq \sum_{\ell=1}^{n-1} \left( \frac{2p_j}{\ell + 1} \right) \lambda_{j\ell} - \left( \sum_{\ell=1}^{n-1} \sum_{k \in N_j} \frac{p_k}{\ell(\ell + 1)} v_{j\ell k} \right) \quad \forall j \in N
\]

(13a)

\[
0 \leq w'_j \leq p_j x_j \quad \forall j \in N \quad (13b)
\]

\[
0 \leq \lambda_{j\ell} - v_{j\ell k} \leq (1 - x_k) \quad \forall j, k \in N(k \neq j), \forall \ell = 1, \ldots, n - 1 \quad (13c)
\]

\[
v_{j\ell k} \leq x_k \quad \forall j, k \in N(k \neq j), \forall \ell = 1, \ldots, n - 1 \quad (13d)
\]

\[
\sum_{\ell=1}^{n-1} \lambda_{j\ell} = 1 \quad \forall j \in N \quad (13e)
\]

\[
\lambda_{j\ell} \geq 0 \quad \forall j \in N, \forall \ell = 1, \ldots, n - 1, \quad (13f)
\]

\[
v_{j\ell k} \in [0, 1] \quad \forall j, k \in N(k \neq j), \forall \ell = 1, \ldots, n - 1 \quad (13g)
\]

\[
x_j \in \{0, 1\} \quad \forall j \in N. \quad (13h)
\]

**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:

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

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

\[
\sum_{j \in N} \Phi_j \left( \sum_{i \in N_j} p_{ij} x_i \right) x_j \geq w \quad \forall \ j \in N
\]

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

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

Constraints (13c) ensure that when \( x_j = 0 \), the term \( \Phi_j \left( \sum_{i \in N} p_{ij} x_i \right) x_j = 0 \) for all \( j \in N \). To eliminate the nonlinear terms \( \lambda_{j\ell} x_k \) that appear in (14), introduce auxiliary variables \( v_{j\ell k} \) to model \( \lambda_{j\ell} x_k \) and add the constraints (13d), (13e), and (13h). With this replacement, (14) can be written as (13b).

### 3.3. Comparison of Correlation-Robust Models: What does correlation buy us?

Having introduced two correlation-robust sets \( X_1 \) and \( X_2 \), we now compare their performance in the context of the following two distributionally robust models:

(Min) \[
\begin{align*}
(\text{Min}) \quad & \min_x c^T x \\
\text{s.t.} \quad & \min_{\mu \in D} \mathbb{P}_\mu[\tilde{a}^T x \geq 1] \geq 1 - \epsilon \\
& x \in S \subseteq \{0, 1\}^n, \\
\end{align*}
\]

and

(Max) \[
\begin{align*}
(\text{Max}) \quad & \max_x \min_{\mu \in D} \mathbb{P}_\mu[\tilde{a}^T x \geq 1] \\
\text{s.t.} \quad & c^T x \leq b \\
& x \in \{0, 1\}^n, \\
\end{align*}
\]

where \( b \in \mathbb{R} \) and all other parameters were defined in the introduction. Model (15) is a distributionally robust version of Model (1) with a single constraint. Model (16) is a distributionally robust version of Model (2), but simplified with \( M = 1 \) and \( S = \{ x \in \{0, 1\}^n : c^T x \leq b \} \). When Models (15) and (16) are specialized to the case with \( D = D_1 \), we refer to them as MinCA and MaxCA, respectively, where CA stands for correlation-agnostic. Similarly, when \( D = D_2 \), we refer to them as MinCC and MaxCC, respectively, where CC stands for correlation-certain. At first glance, these two models appear to be mirror images of one another. As we will see below, there are subtle differences between the claims that we can make about the solutions to these robust models.

The next proposition confirms the following intuition: When solving a distributionally robust version of Problem (1) when first and second moment information is known, exploiting correlation information by solving MinCC can only help (produce lower costs) compared to solving MinCA and ignoring correlations altogether.

**Proposition 7.** For any distribution \( \tilde{\mu} \in D_2 \), \( v^{CC} \leq v^{CA} \) where \( v^{CC} \) and \( v^{CA} \) are the objective function values of MinCC and MinCA, respectively.

**Proof.** Since \( D_2 \subseteq D_1 \), it follows that \( \mathbb{P}_{\tilde{\mu}}[\bigcup_{j \in N} A_j(x)] \geq \min_{\mu \in D_2} \mathbb{P}_\mu[\bigcup_{j \in N} A_j(x)] \geq \min_{\mu \in D_1} \mathbb{P}_\mu[\bigcup_{j \in N} A_j(x)] \) for all \( x \in S \subseteq \{0, 1\}^n \). Hence, any feasible solution to MinCA is also feasible to MinCC, but not necessarily vice versa. \( \square \)
Although Proposition 7 tells us that, given any distribution in $\mathcal{D}_2$, any optimal solution to MinCC is no worse (in terms of cost) than any optimal solution to MinCA, the same “any-any” statement cannot be said about MaxCC and MaxCA. That is, ideally, we would like to claim that every optimal solution to MaxCC is superior (in terms of true probability) to any optimal solution to MaxCA. Instead, the best we can say is that any optimal solution to MaxCC is no worse (in terms of cost) than any unit vector optimal solution to MaxCA. Note that, if a nontrivial solution to MaxCA exists, then an optimal unit vector solution must exist since, by Proposition 3, we can set $x_j^* = 1$ for some $j^* \in \text{arg}\max_{j \in \mathbb{N}} \{ p_j : c_j \leq b \}$ and $x_k = 0$ for all $k \notin \mathbb{N} \setminus \{ j^* \}$.

**Proposition 8.** For all distributions $\mu \in \mathcal{D}_2$,

$$
\mathbb{P}_\mu \left[ \bigcup_{j \in \mathbb{N}} A_j(x^{CC}) \right] \geq \mathbb{P}_\mu \left[ \bigcup_{j \in \mathbb{N}} A_j(x^{CA}) \right]
$$

where $x^{CC}$ is any optimal solution to MaxCC and $x^{CA}$ is any unit vector optimal solution to MaxCA, i.e., $x^{CA} = e_j$ for some $j \in \mathbb{N}$.

**Proof.** Let $S = \{ x \in \{0,1\}^n : c^T x \leq b \}$. Since Equation (10) implies that

$$
\mathbb{P}_\mu \left[ \bigcup_{j \in \mathbb{N}} A_j(x) \right] \geq \sum_{j \in \mathbb{N}} \Phi_j \left( \sum_{i \in \mathbb{N} \setminus \{ j \}} p_{ij} x_i \right) x_j \quad \forall x \in S, \forall \mu \in \mathcal{D}_2,
$$

we have

$$
\mathbb{P}_\mu \left[ \bigcup_{j \in \mathbb{N}} A_j(x^{CC}) \right] \geq \sum_{j \in \mathbb{N}} \Phi_j \left( \sum_{i \in \mathbb{N} \setminus \{ j \}} p_{ij} x_i^{CC} \right) x_j^{CC} \quad \forall \mu \in \mathcal{D}_2.
$$

Moreover, as $x^{CA}$ is suboptimal to MaxCC, we have

$$
\sum_{j \in \mathbb{N}} \Phi_j \left( \sum_{i \in \mathbb{N} \setminus \{ j \}} p_{ij} x_i^{CA} \right) x_j^{CA} \geq \sum_{j \in \mathbb{N}} \Phi_j \left( \sum_{i \in \mathbb{N} \setminus \{ j \}} p_{ij} x_i^{CC} \right) x_j^{CC} \quad \forall \mu \in \mathcal{D}_2.
$$

Finally, since $x^{CA}$ is any unit vector optimal solution, it follows that

$$
\sum_{j \in \mathbb{N}} \Phi_j \left( \sum_{i \in \mathbb{N} \setminus \{ j \}} p_{ij} x_i^{CA} \right) x_j^{CA} = \sum_{j \in \mathbb{N}} \Phi_j (0) x_j^{CA} = \max_{x \in S} \max_{j \in \mathbb{N}} \{ p_j x_j \} = \mathbb{P}_\mu \left[ \bigcup_{j \in \mathbb{N}} A_j(x^{CA}) \right] \quad \forall \mu \in \mathcal{D}_2. \quad \Box
$$

The reason that we cannot make a stronger claim than what is stated in Proposition 8 is that MaxCA typically has many alternative optimal solutions. The following example illustrates that there may exist an optimal solution to MaxCA with a larger true probability than MaxCC.

**Example 1.** Let $n = 3$, $p = (0.9, 0.8, 0.7)$, $p_{ij} = p_i p_j$ for all $i, j \in \mathbb{N}$, and $S = \{0,1\}^n$. The unique optimal solution of the correlation-certain model MaxCC is $x^{CC} = (1, 1, 0)$ with true probability $p(x^{CC}) = 0.98$. When $x = (1, 1, 1)$, $\sum_{j \in \mathbb{N}} \Phi_j (\sum_{i \in \mathbb{N}} p_{ij} x_i) x_j = 0.963 < 0.98 = $
\[ \sum_{j \in N} \Phi_j \left( \sum_{i \in N} P_{ij} x_{ij}^{CC} \right) x_j^{CC} \] showing that even when the data are independent, the correlation-agnostic model may not find a true optimal solution. Meanwhile, there are \( 2^{n-1} = 4 \) optimal solutions to the correlation-agnostic model MaxCA, the worst (in terms of true probability) being \( x_{CA,1} = (1, 0, 0) \) with \( p(x_{CA,1}) = 0.9 \) and the best being \( x_{CA,2} = (1, 1, 1) \) with \( p(x_{CA,2}) = 0.994 \).

Despite the theoretical guarantee, Proposition 8 can be very weak since a unit vector optimal solution to MaxCA can be arbitrarily worse (at least in terms of the ratio \( p(x^{CC})/p(x^{CA}) \)) than an optimal solution to MaxCC as seen in the following example.

**Example 2.** Suppose that there are \( n \geq 2 \) equally likely pairwise disjoint events \( A_1, \ldots, A_n \), i.e., \( P[A_j] = 1/n \) and \( P[A_i \cap A_j] = 0 \) for all \( i, j \in N \). Since events are disjoint, the objective function of MaxCC becomes

\[ \sum_{j \in N} \Phi_j \left( \sum_{i \in N} P_{ij} x_i \right) x_j = \sum_{j \in N} \Phi_j (0) x_j = \sum_{j \in N} p_j x_j. \]

Let \( S = \{ x \in \{0,1\}^n : \sum_{j \in N} c_j x_j \leq b \} \) with \( c_j = p_j \) for all \( j \in N \) and \( b = 1 \). Then MaxCC will choose \( x^{CC} = (1, 1, \ldots, 1) \) with true probability \( p(x^{CC}) = 1 \). Meanwhile, there are \( n \) optimal unit vector solutions \( x^{CA} \) to MaxCA each with a true probability of \( 1/n \). As \( n \to \infty \), the ratio \( p(x^{CC})/p(x^{CA}) \to \infty. \)

Since a unit vector optimal solution can yield poor performance, we conclude this section by trying to address the question: Given multiple optima to MaxCA, is there an intelligent way to select one that outperforms an optimal solution to MaxCC? Since \( P[\bigcup_{j \in N} A_j(x)] \geq \max_{j \in N} P[\mu(A_j(x))] = \max_{j \in N} \{ p_j x_j \} \) for all \( \mu \in D_1 \) and for all \( x \in \{0,1\}^n \), it is clear that any optimal solution in which at least two components are set to one can be no worse than a unit vector optimal solution. Thus, after the component with the largest marginal probability is set one (subject to the budget constraint), it seems natural for one to solve an auxiliary optimization problem to choose which additional components, if any, should be set to one. However, it is not clear what this auxiliary optimization problem should be. Since only first moment information is known in the correlation-agnostic model, one possibility is to solve the binary knapsack problem

\[ \max \left\{ \sum_{j \in N} p_j x_j : \sum_{j \in N} c_j x_j \leq b, x_{j^*} = 1, x \in \{0,1\}^n \right\} \tag{17} \]

for some \( j^* \in \arg \max_{j \in N} \{ p_j : c_j \leq b \} \). This seems sensible, at least at first glance, since maximizing the sum of the marginal probabilities of the remaining components subject to the remaining budget will yield solutions with higher true probability. However, this logic ignores correlations which may lead to two problems. First, as shown in Example 3 below, a solution in which the marginal probabilities are maximized does not imply that the true probability of coverage is maximized. Second, as shown in Example 4, solving the knapsack problem may produce solutions in which the true probability of coverage deteriorates as the budget increases.
Example 3. Consider a Bernoulli mixture distribution with \( n = 3 \), \( L = 2 \), prior \( \pi = (0.5, 0.5) \), conditional probability matrix \( P = [p_{jl}] = \begin{bmatrix} 0.60 & 0.35 \\ 0.10 & 0.40 \\ 0.30 & 0.25 \end{bmatrix} \), and marginal probabilities \( p = (0.475, 0.250, 0.275) \). Let \( c_j = p_j \) for all \( j \in N \) and \( b = 0.75 \). Consider two solutions to the auxiliary knapsack problem: \( x^1 = (1, 0, 1) \) and \( x^2 = (1, 1, 0) \). Although \( x^1 \) is optimal to the knapsack problem, one can check that \( 0.625 = 1 - \sum_{\ell=1}^L \sigma_{\ell}(x^2) \pi_{\ell} > 1 - \sum_{\ell=1}^L \sigma_{\ell}(x^1) \pi_{\ell} = 0.616 \), i.e., \( x^2 \) has a larger true coverage probability, which illustrates that using marginal probabilities to select additional components can be misleading. □

Example 4. Consider a Bernoulli mixture distribution with \( n = 4 \), \( L = 2 \), prior \( \pi = (0.5, 0.5) \), conditional probability matrix \( P = [p_{jl}] = \begin{bmatrix} 0.35 & 0.35 \\ 0.11 & 0.22 \\ 0.22 & 0.11 \\ 0.32 & 0.32 \end{bmatrix} \), and marginal probabilities \( p = (0.350, 0.165, 0.165, 0.320) \). Let \( c_j = p_j \) for all \( j \in N \). With a budget of 0.67, the optimal solution to the auxiliary knapsack problem is \( (1, 0, 0, 1) \) which has a true probability of coverage of 0.558. When the budget is increased to 0.68, the optimal solution to the knapsack problem becomes \( (1, 1, 1, 0) \) with a true probability of coverage of 0.549. □

In summary, using correlations when they are available for Model (15) make the correlation-certain model clearly superior to the correlation-agnostic model. On the other hand, when solving Model (16) the correlation-certain model is no longer guaranteed to be superior, but it is not at all clear how to use the correlation-agnostic model to obtain a better solution. Perhaps one can attempt to obtain a solution to both models, test the quality of each solution, and use the better of the two.

4. A Compact Linearization of the Correlation-Certain Model

While the reformulation given by (13) is convenient in that one can directly formulate the correlation-certain 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.

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

\[
Y_j := \{ (w'_j, x) \in \mathbb{R} \times \{0, 1\}^n : 0 \leq w'_j \leq g_j(x)x_j \} ,
\]
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

\[
X_2 = \left\{ (w, x) \in \mathbb{R} \times \{0,1\}^n : \exists w' \in \mathbb{R}^n \text{ s.t. } \sum_{j \in N} w'_j \geq w, (w'_j, x) \in Y_j \forall j \in N \right\}.
\] (19)

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

\[
Y_{j_1}^S := Y_j \cap \{ (w'_j, x) \in \mathbb{R} \times \{0,1\}^n : x_j = 1, x_i = 0 \forall i \notin S \} \quad \forall j \in N, \forall S \subseteq N^j.
\]

The following proposition sheds light on the relationship between the facets of \( \text{conv}(Y_j) \) and \( \text{conv}(Y_{j_1}^S) \).

**Proposition 9.** For any \( j \in N \), \( \text{conv}(Y_j) \) is completely described by the trivial bound inequalities \( w'_j \geq 0 \), \( x_j \leq 1 \), \( 0 \leq x_i \leq 1 \) for all \( i \in N^j \), and a finite set of inequalities, each of which is of the form

\[
w'_j + (b - \sum_{i \in N^j} a_i)(1 - x_j) + \sum_{i \in N^j} a_ix_i \leq b,
\] (20)

where \( a \in \mathbb{R}_n^{n-1} \) and \( b > 0 \). Moreover, an inequality of the above form is a facet of \( \text{conv}(Y_j) \) if and only if \( w'_j + \sum_{i \in S} a_i x_i \leq b \) is a facet of \( \text{conv}(Y_{j_1}^S) \) where \( S = \{ i \in N^j : a_i > 0 \} \).

**Proof.** See the e-companion to this paper. \( \square \)

The above proposition indicates that we can express \( X_2 \) as a mixed integer linear set by modifying the inequalities that describe \( \text{conv}(Y_{j_1}^S) \) 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 (20).

**Definition 1.** A set function \( h : 2^N \rightarrow \mathbb{R} \) is **supermodular** on \( N \) if

\[
h(S \cup \{k\}) - h(S) \leq h(T \cup \{k\}) - h(T) \quad \forall S \subseteq T \subseteq N \text{ and } k \notin T.
\]

A function \( h \) is **submodular** if \( -h \) is supermodular.

**Proposition 10.** 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 \( \Phi_j(t_1 + \delta) - \Phi_j(t_1) \leq \Phi_j(t_2 + \delta) - \Phi_j(t_2) \), for \( t_1 \leq t_2 \) and \( \delta \geq 0 \). 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 \( t_1 \leq t_2 \). For any \( k \notin T \), let \( \delta = p_{kj} \) and note that

\[
g_j(S \cup \{k\}) - g_j(S) = \Phi_j(t_1 + \delta) - \Phi_j(t_1) \leq \Phi_j(t_2 + \delta) - \Phi_j(t_2) = g_j(T \cup \{k\}) - g_j(T). \quad \square
\]
**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 $\hat{x}$, $\sum_{j \in N} g_j(\hat{x})x_j$ represents the minimum probability of the union of a finite set of events $A_j(\hat{x})$, where the minimum is taken over all distributions in the family $D_2$. Thus, we can interpret $g_j(\hat{x})$ 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**. Note that 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. There is a well-known connection between supermodularity and complementarity (Topkis 1998).

The next proposition from Atamtürk and Narayanan (2008), which builds on earlier work of Edmonds (1971), exploits the supermodularity of $g_j$ and allows us to identify the precise values of the coefficients $a$ and $b$ in (20).

**PROPOSITION 11.** [Atamtürk and Narayanan (2008)] For any $j \in N$ and any $S \subseteq N^j$ with $|S| = k$, $\text{conv}(Y_{j1}^S)$ is completely described by the trivial bound inequalities $w'_j \geq 0$, $x_j = 1$, $x_i = 0$ for all $i \notin S$, $0 \leq x_i \leq 1$ for all $i \in S$, and the nontrivial facets

$$w'_j + \sum_{i \in S} a_{\pi(i)}x_{\pi(i)} \leq b \quad \forall \pi \in \Pi(S),$$

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

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

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

The following corollary follows immediately from Propositions 9 and 11.

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

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

where the coefficients $a_{\pi(i)}$ are defined in (22).

We refer to constraints (23) as **lifted extended polymatroid** (LEP) cuts as they are derived from lifting cuts of the form (21), which are referred to as **extended polymatroid** cuts in Atamtürk and Narayanan (2008).
EXAMPLE 5. Suppose \( n = 3, p = (0.9, 0.8, 0.7), \) and \( p_{ij} = p_i p_j, \) for all \( i, j \in N. \) Note that \( g_1(\emptyset) = 0.900, \) \( g_1(\{2\}) = 0.540, \) \( g_1(\{3\}) = 0.585, \) and \( g_1(\{2, 3\}) = 0.375. \) In agreement with Proposition 11, the nontrivial facets of \( \text{conv}(Y_1^S) \) are

\[
0.360x_2 + w'_1 \leq 0.900 \\
0.315x_3 + w'_1 \leq 0.900 \\
0.360x_2 + 0.165x_3 + w'_1 \leq 0.900 \\
0.210x_2 + 0.315x_3 + w'_1 \leq 0.900
\]

In agreement with Corollary 1, the facets above give rise to the nontrivial facets of \( \text{conv}(Y_1) \)

\[
-0.900x_1 + w'_1 \leq 0 \\
-0.540x_1 + 0.360x_2 + w'_1 \leq 0.360 \\
-0.585x_1 + 0.315x_3 + w'_1 \leq 0.315 \\
-0.375x_1 + 0.210x_2 + 0.315x_3 + w'_1 \leq 0.525 \\
-0.375x_1 + 0.360x_2 + 0.165x_3 + w'_1 \leq 0.525
\]

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

\[
X_2 = \left\{ (w', w, x) \in \mathbb{R}^n \times \mathbb{R}^n \times \{0, 1\}^n : \sum_{j \in N} w'_j \geq w, \right\}.
\]

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 (23) is extremely large, namely \( n \sum_{t=0}^{n-1} \binom{n-1}{t} \) is \( n2^{n-1}, \) when optimizing over the set \( X_2, \) we omit all but those corresponding to when \( T = \emptyset, \) i.e., we only include constraints \( 0 \leq w'_j \leq p_j x_j \) 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 \((\hat{w}', \hat{x}) \in \mathbb{R}^n \times \mathbb{R}^n\) such that \( \hat{w}'_j > g_j(\hat{x}) \hat{x}_j, \) we can separate an inequality of the form \( w'_j + (b - \sum_{i \in S} a_i)(1 - x_j) + \sum_{i \in S} a_i x_i \leq b \) 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 \( \sum_{i \in S} (\hat{x}_i - 1 + \hat{x}_j) a_i \) is maximized. Thus we set \( S = \{ i \in N' : \hat{x}_i - 1 + \hat{x}_j > 0 \}. \) Next we sort \( \hat{x}_i \) for \( i \in S \) such that \( \hat{x}_{i[1]} \geq \hat{x}_{i[2]} \geq \cdots \geq \hat{x}_{i[k]} \), breaking ties arbitrarily, and computing \( a_{i[j]} \) according to (22). 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 (24).

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-robust models are often 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
Algorithm 1 SeparateLEPCuts($\hat{w}'$, $\hat{x}$)

 Require: A (possibly fractional) solution pair $(\hat{w}'$, $\hat{x}$) to the current LP relaxation.

 1: Sort the $x_j$ variables in nonincreasing order $\hat{x}[1] \geq \hat{x}[2] \geq \cdots \geq \hat{x}[n]$.

 2: for $j = 1, \ldots, n$ do
      3: if $\hat{w}'_j > g_j(\hat{x})$ then
          4: Define the set $S = \{i \in N^j : \hat{x}_i + \hat{x}_j > 1\}$ and let $k = |S|$.
          5: Let $\pi = (\pi(1), \ldots, \pi(k))$ s.t. $\hat{x}_{\pi(1)} \geq \cdots \geq \hat{x}_{\pi(k)}$ and $\{\pi(1), \ldots, \pi(k)\} = S$.
          6: Let $a_{\pi(i)}$ according to Equation (22) $\forall i \in S$.
          7: if $\hat{w}'_j - p_j \hat{x}_j + \sum_{i \in S} (\hat{x}_{\pi(i)} + \hat{x}_j - 1) a_{\pi(i)} > 0$ then
              8: The most violated cut is $w'_j + (b - p_j)x_j + \sum_{i \in S} a_{\pi(i)}x_{\pi(i)} \leq b$ where $b = \sum_{i \in S} a_i$.
          end if
      end if
  end for

formulation (24) 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. Empirical Comparison of Models

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 the 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 (labeled “Free” in the figures), 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 (“Agnostic”), 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 (“Certain”), 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:

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

where $B(b) = \{ x \in \{0,1\}^n : c^T x \leq b \}$, $b \in \mathbb{R}_+$, 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 $\ell$, for a Bernoulli mixture distribution.

As discussed at the end of Section 3.3, since the correlation-agnostic model typically has many alternative optimal solutions, there are several optimization problems that one could solve to find one that admits the best true coverage probability. To reiterate, these procedures are heuristics because a correlation-agnostic modeler only knows first moment information (individual marginal probabilities). We experimented with two approaches. The first solves the auxiliary knapsack problem given by Model (17), while the second solves

$$\max \left\{ \sum_{j \in N} -\log(q_j) x_j : \sum_{j \in N} c_j x_j \leq b, x_j^* = 1, x \in \{0,1\}^n \right\}$$

for some $j^* \in \arg \max_{j \in N} \{ p_j : c_j \leq b \}$, i.e., after the component with the largest marginal probability is selected (subject to the budget constraint), the correlation-agnostic modeler assumes that the data are independent in hopes of finding a solution with a larger true coverage probability. For these experiments, solving Model (25) was no worse than Model (17) over all instances for almost all budget values, so we only present the results for Model (25). To be clear, results labeled “Agnostic” in the figures refer to a correlation-agnostic modeler solving Model (25).

We generated 50 Bernoulli mixture instances with $n = 15$ binary decision variables and $L$ scenarios, for $L \in \{5,25,50\}$, (for a total of 150 instances) as follows: After generating conditional probabilities $p_{jl}$ randomly from a uniform$(0.1,0.9)$ distribution and normalizing so that $\sum_{j=1}^n p_{jl} = 1$ for all $l = 1,\ldots,L$, we constructed prior vectors $\pi = \{\pi_{\ell}\}_{\ell=1}^L$ 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 \(\rho_{ij}\) 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.

Figures 1–3 show the average and maximum absolute gaps over all 50 instances for \(L \in \{5, 25, 50\}\) and for each budget value \(b\). The absolute gap refers to the difference between the optimal coverage probability and the true coverage probability associated with the optimal solution returned by a particular model. Focusing first on the average gap, we see that the correlation-free and correlation-agnostic models behave very similarly for all but a few budget intervals in which the correlation-free model does better. Meanwhile, the correlation-certain model consistently performs the best for a budget value up to 1.3. For \(L = 5\), beyond a budget of 1.5, the correlation-certain model is overly conservative. For \(L = 25\) and \(L = 50\), in the budget intervals \([1.3, 1.9]\) and \([2.4, 2.8]\) the correlation-certain model is overly conservative, but is otherwise competitive with, if not superior to, the other models.

Perhaps most conspicuous is the large gap associated with the correlation-certain model for all scenarios in the budget interval \([1.4, 2.0]\) and the large gap associated with the correlation-agnostic model in the budget interval \([0.7, 1.0]\). For the correlation-certain model, this large gap appears to be due to the model finding the same solution as the budget increases from 1.4 to 2.0 and therefore maintaining the same true coverage probability. This type of stagnation is exemplified in Figure 5. For the correlation-agnostic model, this large gap is due to the fact that the model must use the component with the largest marginal probability and, with its remaining budget, can only choose another component with a small marginal probability. On the other hand, an optimal solution is often to choose two components with average marginal probabilities.

The maximum gap over all budget values also shows some interesting behavior. For instances with 5 scenarios, the correlation-free model obtains the largest absolute gap over all instances. The correlation-certain model also has a large maximum gap for budgets greater than 2.5, which appears to be due to the stagnation found in instances like the one depicted in Figure 5. For instances with 25 and 50 scenarios, the two correlation-robust models perform poorly in certain budget intervals.

While the aggregate information shown in Figures 1–3 is insightful, it does not tell the entire story. It is also helpful to look at the results of individual instances as shown in Figures 4–6 and also those in the e-companion to this paper. As one would expect, the optimal probability of coverage increases monotonically as \(b\) increases. In contrast, the probability of coverage corresponding to
the correlation-free, agnostic, and certain models does not necessarily increase monotonically since these models are optimizing different objective functions. Although the correlation-certain model might display a large average absolute gap in the budget interval [1.4, 1.8], Figure 4 shows that it can be superior in that budget interval and over all budgets. Figure 5 illustrates that the correlation-certain model can outperform the other models up to a particular budget after which point it becomes “stuck” unable to find an improving solution. Figure 6 reveals the issue of variability over different budget values within the same instance. In particular, in the budget interval [1.1, 1.7], during which the correlation-certain model finds an optimal solution to the true problem early on, but is then is unable to find one with a better probability of coverage, the coverage probability

Figure 1  Average and maximum absolute gap for $L = 5$ scenarios

Figure 2  Average and maximum absolute gap for $L = 25$ scenarios
associated with correlation-free and correlation-agnostic models is quite unstable, dropping well below that of the correlation-certain model, while, only in the latter part of the interval, rising above.

We conclude this section with some general remarks. For these Bernoulli mixture model instances,
Figure 5  Instance 31 for $L = 5$

Figure 6  Instance 29 for $L = 5$

ignoring correlations altogether results in surprisingly decent performance. However, 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. The variability in the gap from one budget value to another when using the correlation-free model may be an issue for a modeler looking for stable solutions. For a budget level below 1.3 and above 3, at least one of the correlation-robust models performs well. Perhaps one could attempt to solve both models. 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.

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 (24), which requires implementing a cut callback routine to separate violated lifted extended polymatroid cuts (23). 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 (24) 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 1 and 2, with additional tables presented in the e-companion to this paper. In each table, we fix the value of $\epsilon$ 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.

Our main finding for these instances is that, regardless of $\epsilon$ 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 formu-
lation 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.

Appendix. Proof of Claim 1

CLAIM 1. Consider the LP

$$\min \sum_{i \in N} \frac{y_i}{i}$$

s.t. $\sum_{i \in N} y_i = a$

$$\sum_{i \in N} (i - 1)y_i = b$$

$y_i \geq 0 \quad \forall \ i \in N$ (26)
If \( n \geq 2 \) and \( a \geq b \geq 0 \), then LP (26) 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 \( a \geq b \geq 0 \), then LP (26) is clearly feasible as \( y_1 = a - b, y_2 = b, y_3 = \cdots = y_n = 0 \) is a feasible solution. LP (26) is also bounded since \( y_i \geq 0 \) for all \( i \). Consider the dual to LP (26):

\[
\begin{align*}
\max & \quad au + bv \\
\text{s.t.} & \quad u + (i-1)v \leq \frac{1}{i} \quad i = 1,\ldots,n.
\end{align*}
\]  

(27)

The proof will be complete if we can show that the set \( E \) of extreme points of the dual LP (27) is given by

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

Note that each of the \( n \) constraints of the dual (27) are linearly independent. Moreover, any extreme point of (27) lies at the intersection of exactly two constraints of (27).

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 (27). To verify that this point is feasible to (27), we check that this point satisfies each of the remaining constraints \( k = 1,\ldots,n \) with \( k \neq i \). Note that for the \( k \)th 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 \( (\hat{u}, \hat{v}) \in E \) and, to arrive at a contradiction, suppose \( (\hat{u}, \hat{v}) \notin F \). That is, \( (\hat{u}, \hat{v}) \) lies at the intersection of constraints \( k \) and \( i+1 \) with \( k \leq i-1 \) for some \( i \in \{2,\ldots,n-1\} \) in (27):

\[
\hat{u} + (k-1)\hat{v} = \frac{1}{k}, \quad \hat{u} + \frac{1}{i} \hat{v} = \frac{1}{i+1}.
\]

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

\[
\hat{u} + (i-1)\hat{v} = \frac{k+i}{k(i+1)} - \frac{i-1}{k(i+1)} = \frac{k+1}{k(i+1)} - \frac{i+1}{i(i+1)} = \frac{1}{i+1},
\]

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 \( (\hat{u}, \hat{v}) \in E \). Finally, it is easy to see that adjacent constraints in the dual give rise to \( F \). \( \square \)

**Acknowledgments**

Shabbir Ahmed was partially supported by the National Science Foundation Grant CMMI 1129871. Dimitri Papageorgiou was partially supported by a National Science Foundation Graduate Research Fellowship. We wish to thank two anonymous referees for several instructive comments and suggestions that improved the final version of this paper.
References


Polyhedral Results Related to Proposition 9

Here we analyze the polyhedral structure of the convex hull of the set

\[ Y := \{ (w, x, y) \in \mathbb{R} \times \{0, 1\} \times \{0, 1\}^{n} : 0 \leq w \leq f(y) x \}, \tag{28} \]

where \( f : \{0, 1\}^{n} \rightarrow \mathbb{R}_{++} \) is a nonincreasing function. Note that the set \( Y_j \) given by (18) is a special case of (28). 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 \( Y^S_1 = Y_1 \cap \{y \in \{0, 1\}^n : y_j = 0 \ \forall \ j \notin S \} \) 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 13.** For any \( S \subseteq N \),

\[ \dim(\text{conv}(Y^S_1)) = |S| + 1. \]

*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. \( \square \)

**Proposition 14.** 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 \( \text{conv}(Y^S_1) \).

*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 \( \{(0, e^j + e^k)\}_{k \in S, k \neq j} \), \( (f(0), 0) \) 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. \( \square \)
Proposition 15. For any $S \subseteq N$,

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

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 $(\hat{w}, \hat{y}) \in \text{conv}(Y_1^S)$ with $\hat{w} > 0$ 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 $(0, \hat{y}) \in \text{conv}(Y_1^S)$ 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 $(\hat{w}, \hat{y}) \in \text{conv}(Y_1^S)$ with $\hat{y}_j > 0$ 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 $(\hat{w}, \hat{y} - \hat{y}_j e^j) \in \text{conv}(Y_1^S)$ (since $\hat{w} \leq f(\hat{y}) \leq f(\hat{y} - \hat{y}_j e^j)$ 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 14. □

Proposition 16.

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

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

Proposition 17. 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 $\{(0,0,e^i)\}_{i=1}^{n}$, $(0,0,0)$ and $(0,1,0)$ satisfy $0 \leq w$ at equality. The $n + 2$ affinely independent points $\{(0,1,e^i)\}_{i=1}^{n}$, $(0,1,0)$ and $(f(0),1,0)$ satisfy $x \leq 1$ at equality. The $n + 2$ affinely independent points $\{(0,1,e^k)\}_{k=1,k \neq 1}^{n}$, $(f(0),1,0)$, $(0,1,0)$ and $(0,0,0)$ satisfy $0 \leq y_j$ at equality. Finally, the $n + 2$ affinely independent points $\{(0,1,e^i + e^k)\}_{i=1,k \neq j}^{n}$, $(f(e^i),1,e^j)$, $(0,1,e^i)$ and $(0,0,e^i)$ satisfy $y^j \leq 1$ at equality. □

Proposition 18. 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 \((\hat{w}, \hat{x}, \hat{y}) \in \text{conv}(Y)\) with \(\hat{w} > 0\) 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 \((0, \hat{x}, \hat{y}) \in \text{conv}(Y)\) and is cut off by the given inequality. Thus \(\gamma \geq 0\). Suppose \(\tau < 0\) and consider a point \((\hat{w}, \hat{x}, \hat{y}) \in \text{conv}(Y)\) with \(\hat{x} < 1\) 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 \((\hat{w}, 1, \hat{y}) \in \text{conv}(Y)\) and is cut off by the given inequality. Thus \(\tau \geq 0\). Similarly suppose \(\alpha_j < 0\) and consider a point \((\hat{w}, \hat{x}, \hat{y}) \in \text{conv}(Y)\) with \(\hat{y}_j > 0\) 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 \((\hat{w}, \hat{x}, \hat{y} - \hat{y}_je^j) \in \text{conv}(Y)\) 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 \((\hat{w}, \hat{x}, \hat{y}) \in Y\). If \(\hat{x} = 0\) then \(\hat{w} = 0\) and \(\sum_{j \in S} a_j y_j \leq \sum_{j \in S} a_j\) trivially. If \(\hat{x} = 1\) then construct a point \(y'\) such that \(y'_j = \hat{y}_j\) for \(j \in S\) and zero otherwise. By the nonincreasing property of \(f\), we have that \(\hat{w} \leq f(\hat{y}) \leq f(y')\) and so \((\hat{w}, y') \in Y_1^S\) and so satisfies \(\hat{w} + \sum_{j \in S} a_j \hat{y}_j \leq b\). □

Proposition 19. 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 \(\{(w', y'_i)\}_{i=1}^{k+1}\) such that \(w' \leq f(y'_i), y'_j = 0\) for all \(j \notin S\), and \(w' + \sum_{j \in S} a_j y'_j = b\) for all \(i = 1, \ldots, k+1\). Now consider the points \((w, x, y) := \{(w', 1, y')\}_{i=1}^{k+1}, \{(0, 0, e') + \sum_{j \in S} e^j\}_{j \in N \setminus S}, (0, 0, \sum_{j \in S} e^j)\}\). 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 \(\text{conv}(Y)\).
(⇒) Note that \( w + \sum_{j \in S} a_j y_j \leq b \) is valid for \( Y \) and therefore valid for \( Y^S \). In fact the inequality is strong in the sense that there is at least one point in \( \text{conv}(Y^S) \) for which it is tight. Indeed, consider a point \((\hat{w}, \hat{x}, \hat{y}) \in Y\) 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 \( \hat{x} = 1 \) (such a point exists since the facet is nontrivial). Construct \( \gamma \) such that \( y_j = \hat{y}_j \) for \( j \in S \) and zero otherwise. By the nonincreasing property of \( f \), we have that \( \hat{w} \leq f(\hat{y}) \leq f(y') \) and so \((\hat{w}, y') \in Y^S\) and \( \hat{w} + \sum_{j \in S} a_j y'_j = b \). Suppose now that \( w + \sum_{j \in S} a_j y_j \leq b \) is not a facet of \( \text{conv}(Y^S) \). Let

\[
\begin{align*}
\sum_{j \in S} \alpha_j y_j &\leq \beta^i \quad \forall \ i \in M \\
-w &\leq 0 \\
y_j &\leq 0 \quad \forall \ j \in S \\
y_j &\leq 1 \quad \forall \ j \in S
\end{align*}
\]

be the complete description of \( \text{conv}(Y^S) \). (Note that by Proposition 15 the complete description will be of this form). Taking a non-negative combination of the above inequalities, we obtain

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

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

\[
\begin{align*}
\sum_{i \in M} \lambda_i - \mu &= 1 \\
\sum_{i \in M} \lambda_i \alpha^i_j - \gamma_j + \tau_j &= a_j \quad \forall \ j \in S \\
\sum_{i \in M} \lambda_i \beta^i + \sum_{j \in S} \tau_j &= b.
\end{align*}
\]

(29)

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

**Case 1:** Suppose \( \sum_{i \in M} \lambda_i \alpha^i_j + \tau_j = a_j \) for all \( j \in S \). By Lemma 1, the inequalities

\[
w + \left( \beta^i - \sum_{j \in S} \alpha^i_j \right) (1 - x) + \sum_{j \in S} \alpha^i_j y_j \leq \beta^i \quad \forall \ i \in M
\]

are valid for \( Y \). Combining the above inequalities (using multipliers \( \lambda_i \)) with the valid inequalities \( y_j \leq 1 \) for \( j \in S \) (using multipliers \( \tau_j \)) and \(-w \leq 0 \) (using multiplier \( \mu \)) for \( \text{conv}(Y) \) and using (29), 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 \( \sum_{i \in M} \lambda_i \alpha^i_k + \tau_k > a_k \). Combining the inequalities

\[
w + \sum_{j \in S} \alpha^i_j y_j \leq \beta^i \quad \forall \ i \in M
\]

(29)

(29)

(29)

(29)

(using multipliers \( \lambda_i \)) with valid inequalities \( y_j \leq 1 \) for \( j \in S \) (using multipliers \( \tau_j \)), 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 \( Y^S_i \), and (by weakening the coefficients) so are the two inequalities

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

and

\[
w + \sum_{j \in S \backslash \{k\}} a_j y_j + (0) y_k \leq b.
\]

By Lemma 1, the following inequalities are valid for \( Y \)

\[
w + \left( b - \sum_{j \in S \backslash \{k\}} a_j - \sum_{i \in M} \lambda_i \alpha^i_k - \tau_k \right) (1 - x) + \sum_{j \in S \backslash \{k\}} a_j y_j + \left( \sum_{i \in M} \lambda_i \alpha^i_k + \tau_k \right) y_k \leq b \tag{31}
\]

and

\[
w + \left( b - \sum_{j \in S \backslash \{k\}} a_j - 0 \right) (1 - x) + \sum_{j \in S \backslash \{k\}} a_j y_j + (0) y_k \leq b. \tag{32}
\]

Multiplying (31) by \( a_k / (\sum_{i \in M} \lambda_i \alpha^i_k + \tau_k) \) and (32) 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 \( \text{conv}(Y) \) 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 \( S_i = \{ j \in N \colon a^i_j > 0 \} \) is a nontrivial facet of \( \text{conv}(Y^S_i) \). □

**Additional Figures**

Here we present some figures associated with interesting instances not discussed in the main text.

Figure 7(a) illustrates the large variability in true probability of coverage that can occur when using the correlation-free or -agnostic models. Figure 7(b) depicts an instance in which both the correlation-free and -certain models are optimal for virtually every budget. In Figure 8, two instances are shown in which the probability of coverage behaves like a step function for all models. Although the probability of coverage corresponding to the correlation-certain model almost always increases as the budget \( b \) increases, Figure 8(a) provides an example in which the probability of coverage drops with an increase in \( b \), namely just before \( b = 2.0 \) and in the budget interval \( [3.5, 4.0] \). Figure 9 reveals two instances in which the correlation-certain model can be vastly inferior to correlation-free and -agnostic models in a particular budget interval.
Additional Tables to Compare Algorithms

Here we present two tables for the cases not presented in the main text. These tables reinforce our main finding: Regardless of $\epsilon$ 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.
Figure 9  Instances 9 and 10

(a) Instance 20 for $L = 50$

(b) Instance 48 for $L = 50$

<table>
<thead>
<tr>
<th>$n$</th>
<th># Cuts</th>
<th># Nodes</th>
<th>Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LS</td>
<td>LS+</td>
<td>C</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>4</td>
<td>58</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>4</td>
<td>62</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
<td>4</td>
<td>64</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td>4</td>
<td>66</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>70</td>
<td>0</td>
<td>4</td>
<td>66</td>
</tr>
</tbody>
</table>

Table 3  $\epsilon = 0.05$, $k = 10$

<table>
<thead>
<tr>
<th>$n$</th>
<th># Cuts</th>
<th># Nodes</th>
<th>Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LS</td>
<td>LS+</td>
<td>C</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>4</td>
<td>255</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
<td>4</td>
<td>305</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td>4</td>
<td>368</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>4</td>
<td>376</td>
</tr>
<tr>
<td>70</td>
<td>0</td>
<td>4</td>
<td>386</td>
</tr>
</tbody>
</table>

Table 4  $\epsilon = 0.02$, $k = 5$