

Branch and Price for Chance Constrained Bin Packing

Zheng Zhang

Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai 200240, China,
zhang@sjtu.edu.cn

Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, zhazheng@umich.edu

Brian Denton

Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, btdenton@umich.edu

Xiaolan Xie

Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai 200240, China,
xie@sjtu.edu.cn

Center for Biomedical and Healthcare Engineering, Ecole Nationale Supérieure des Mines, Saint Etienne 42023, France,
xie@emse.fr

This article considers two versions of the stochastic bin packing problem with chance constraints. In the first version, we formulate the problem as a two-stage stochastic integer program that considers item-to-bin allocation decisions in the context of chance constraints on total item size within the bins. Next, we describe a distributionally robust formulation of the problem that assumes the item sizes are ambiguous. We present a branch-and-price approach based on a column generation reformulation that is tailored to these two model formulations. We further enhance this approach using antisymmetry branching and subproblem reformulations of the stochastic integer programming model based on conditional value at risk (CVaR) and probabilistic covers. For the distributionally robust formulation we derive a closed form expression for the chance constraints; furthermore, we show that under mild assumptions the robust model can be reformulated as a mixed integer program with significantly fewer integer variables compared to the stochastic integer program. We implement a series of numerical experiments based on real data in the context of an application to surgery scheduling in hospitals to demonstrate the performance of the methods for practical applications.

Key words: bin packing; chance constraints; branch-and-price; surgery scheduling

History: This version: November 23, 2015

1. Introduction

Bin packing has been studied in several contexts including computer scheduling (Coffman et al. (1978)), internet advertising (Adler et al. (2002)) and operating room management (Denton et al. (2010)). Stochastic versions of this problem typically assume that the item sizes are random variables, the items are allocated to bins prior to knowing the random outcome of their size, and the bins are *extensible* such that the bin size extends (at a cost) to the size necessary to contain the items. Thus, the problem becomes one of minimizing the fixed cost of bins necessary to pack the items and the expected cost of extending the bins. In this article we study a generalization of this

problem, which we refer to as the *chance constrained stochastic bin packing problem* (CCSBP) in which there are chance constraints on the sum of the item sizes. Thus, a feasible allocation of items to bins is one that minimizes the expected cost while at the same time meeting all chance constraints.

Given a finite number of scenarios for the item’s sizes, CCSBP can be formulated as a large-scale mixed-integer program (MIP); however, this problem is difficult to solve for several reasons. First, it lacks convexity because of the binary item-to-bin allocation variables, and the fact that chance constraints are generally nonconvex (Luedtke and Ahmed (2008)). Second, the linear programming (LP) relaxation of the extensive formulation is often weak because of the introduction of *big-Ms* for the chance constraints. Third, there exists symmetry for problems in which there are some identical bins (Denton et al. (2010)) and items of the same type. To overcome these challenges we propose a *branch-and-price* approach that is tailored to this problem. We show that it eliminates symmetry, and most importantly, provides opportunities to exploit properties of the stochastic knapsack problem that defines the *subproblem*. Our experimental results also suggest that it generates a tighter LP relaxation. In addition to presenting a method for the CCSBP we also present a distributionally robust optimization (DRO) model counterpart that incorporates ambiguity in item sizes. The DRO model, to the best of our knowledge, is the first MIP model of a chance constrained program that is based on a closed form expression for the worst-case probability of satisfying the chance constraint.

We use surgery scheduling in hospitals as a motivating example. In this context the bins are operating rooms (ORs) that are available for some nominal period of time during the day of surgery (e.g. 8 hours). The items represent surgeries and the surgery durations represent the item sizes. This is an important example because of the high cost of OR utilization and OR staff overtime. The chance constraints are particularly important in order to provide surgical teams with appropriate “end-of-the-day” guarantees. The CCSBP model is suited to situations in which the surgery duration distributions can be estimated using historical data. The DRO model is suited to the case in which limited information is known about surgery durations (e.g., mean and variance of surgery durations) and the true distribution is ambiguous. Aside from the practical motivation based on availability of data for estimating the probability distribution for surgery durations, the DRO model is at least an order of magnitude smaller than the CCSBP in most cases, and thus presents much less of a computational challenge.

In this article we focus on two main research questions. The first research question is: what are the advantages of using branch-and-price as a solution method to solve the chance constrained

bin packing problem? We describe a number of properties of the problem that can be exploited to achieve computational advantages in our branch-and-price implementation. We use numerical experiments based on our motivating application in the context of surgery scheduling to illustrate the performance of our approach compared to branch-and-bound for test instances based on a real-world practical application of the model. The second research question is: what is the value of the DRO model compared to the CCSBP? To investigate this we compare the DRO model to the CCSBP on the basis of computational effort required to obtain optimal solutions, and on the basis of solution quality in the context of ambiguity in item size probability distributions.

The remainder of this article is organized as follows. Section 2 reviews literature on the stochastic bin packing problem and highlights the novel contributions of this article. Section 3 defines the chance constrained stochastic bin packing problem and formulates it as a general two-stage stochastic programming model with integer recourse. Section 4 reformulates the problem using column generation, presents a branch-and-price algorithm, and analyzes theoretical properties to accelerate the computation. Section 5 describes the DRO model and its reformulation as a MIP. Section 6 presents computational results and managerial insights. Finally, Section 7 concludes the paper with a summary of the most important findings and opportunities for future work.

2. Literature Review

This section reviews the literature on stochastic bin packing, and relevant methods including column generation, branch-and-price, and distributionally robust optimization. We complete this section with a description of the main contributions of this article to the literature.

2.1. Stochastic Bin Packing

A significant portion of the literature on stochastic bin-packing problems is in the context of surgery scheduling. For example, Denton et al. (2010) considered the stochastic extensible bin-packing problem in the context of allocating surgeries (items) to operating rooms (bins). The objective included a fixed cost for each utilized operating room (OR) and a variable cost for expected overtime. They proposed a two-stage stochastic mixed-integer program (2-SMIP) to account for uncertainty in surgery durations and also a robust optimization model to account for ambiguous durations when only lower and upper bounds are available on the surgery durations. Batun et al. (2011) considered a related problem in which some surgeries could be completed in parallel, in different ORs, by the same primary surgeon assisted by surgical fellows. In both cases the L-shaped method was used to solve the problem, with additional cuts to reduce symmetry in the first stage.

The addition of chance constraints was motivated by the challenges of attributing a cost to OR overtime in practice, since much of the cost is due to a loss of good-will among nursing staff, and

is not attributable to the monetary cost of overtime alone (Deng et al. (2014)). The addition of chance constraints, although attractive for practical reasons, creates computational challenges for two reasons. First, for a given $x \in \mathcal{X}$, it is difficult to compute the probability $\mathbb{P}\{A(\xi)x \geq b\}$, as it requires multi-dimensional integration. Second, the feasible region defined by a chance constraint is generally not convex (Luedtke and Ahmed (2008)). There exist some exceptions for which the probability can be efficiently estimated, such as when ξ is normally distributed. However, surgery durations are more often associated with distributions that have a right tail, such as the lognormal distribution (Strum et al. (2003)). Shylo et al. (2012) appears to be the first to consider a chance constrained bin-packing problem. They assumed the objective was to minimize the number of ORs to open while satisfying a set of probabilistic capacity constraints. The chance constrained program was expressed as a series of MIPs based on a *normal approximation* of cumulative surgery durations. However, this approximation is not likely to be accurate when the number of surgeries is small (e.g. $n \leq 4$), which is common for many hospital inpatient surgery practices. Wang et al. (2014) considered a similar problem except that the ORs were shared with emergencies. The OR completion times were chance constrained and the risk was evaluated by an independent *Monte-Carlo simulation*. Deng et al. (2014) considered a joint OR scheduling problem similar to Batun et al. (2011) but for which the surgeon waiting times are subject to individual chance constraints and the overtime in all ORs is subject to a joint chance constraint. They proposed a 2-SMIP with integer recourse based on a *sample approximation*. Further, to speed up the computation, they used a cutting plane algorithm and lifting methods based on problem decompositions on scenarios and ORs, respectively.

Lamiri et al. (2008) used column generation to solve OR allocation problems combined with surgery date decisions. Uncertainty exists in the number of emergency surgeries that must share ORs with elective surgeries. They assumed surgery durations are deterministic and solved the subproblem as a deterministic knapsack using dynamic programming and the integer master problem using heuristics. In a related study, Wang et al. (2014) solved chance constrained OR allocation problems by column generation. The chance constraint on OR completion times was evaluated by an independent simulation on each generated column. The master problem was solved by using heuristics to obtain integer solutions.

2.2. Branch and Price

Solution methods for SMIPs or equivalent large-scale MIPs typically employ Benders decomposition (Benders (1962)), or extensions thereof, to the original model. Examples in the bin-packing context include most of the above references: Denton et al. (2010), Batun et al. (2011), Deng et al. (2014).

Unfortunately, these decompositions use a master problem whose LP relaxation is no stronger than that of the original model (Silva and Wood (2006)), leading to a poor continuous relaxation of the formulation, especially when big-M constraints exist, such as is the case for a common reformulation of chance constrained problems as we show in Section 3. In contrast, branch-and-price solves a column-oriented reformulation of the model, also by a form of decomposition, but with a normally tighter relaxation bound. A comprehensive description of branch-and-price methodology is provided in Barnhart et al. (1998), Vanderbeck (2000), and Vanderbeck and Wolsey (1996). Here, we focus on branch-and-price applied to SMIPs and literature related to our specific problem.

Silva and Wood (2006) surveyed branch-and-price for two-stage SMIPs, such as the general assignment problem, the routing and scheduling problem, the crew scheduling problem, and the integer multi-commodity flow problem. They also implemented branch-and-price for a stochastic facility location problem and reported that this method can be orders of magnitude faster than solving the original problem by branch-and-bound. Degraeve and Jans (2007) used branch-and-price to solve a capacitated lot sizing problem with setup times. They reformulated the problem as a MIP, and applied Dantzig-Wolfe decomposition. The column generation procedure was accelerated by using a combination of simplex and subgradient optimization for finding the dual prices. Similarly, Singh et al. (2009) used branch-and-price to solve a multistage stochastic capacity planning problem. They reformulated the problem as a *split-variable* model which provides a stronger LP relaxation and they implemented a branch-and-price method for their model. Because the LP relaxation usually generates an integer solution, the branch-and-price formulation can solve a large-scale MIP problem with a quarter million binary variables.

To the best of our knowledge, only Lulli and Sen (2004) incorporated chance constraints in a branch-and-price method; the authors solved a multistage stochastic batch-sizing problem for both recourse and probabilistic constrained formulations. Their computational results show that branch-and-price is faster than branch-and-bound in the model without chance constraints; however, because of the special structure of chance constraints, branch-and-price failed to outperform branch-and-bound.

Recently Song et al. (2014) provided a *probabilistic cover* based reformulation to explore the structure of chance constraints in knapsack problems. The reformulation casts the SMIP as a MIP based on probabilistic covers. They also used deterministic cover inequalities to perform approximate lifting of probabilistic cover inequalities. As we show, in Section 4, some of the properties they describe can be exploited in a branch-and-price approach for the CCSBP.

2.3. Distributionally Robust Optimization

Distributionally robust optimization formulations for stochastic programming focus on the worst-case performance over a confidence set. The confidence set represents the ambiguity and is assumed to include the true distribution. There are several possible types of ambiguous information such as moments (e.g. Scarf et al. (1958)), confidence regions of moments (e.g. Delage and Ye (2010)) and support (e.g. Denton et al. (2010)). A chance constrained version of DRO was considered by Jiang and Guan (2013) who presented a series of stochastic programs depending on different types of confidence sets.

Several papers have recently addressed DRO in scheduling problems. Kong et al. (2013) considered a stochastic appointment scheduling problem to determine arrival times for a sequence of customers. The job durations are stochastic and only the mean and covariance estimates are known. They formulated a copositive programming model to minimize the expected cost over a worst-case distribution in a set defined by moments and nonnegative support of job durations. As the problem was not necessarily polynomial-time solvable, they further proposed a semidefinite programming relaxation as a solution approach. Mak et al. (2015) considered a similar problem except that job durations were assumed to be independently distributed. With this assumption, they showed that the DRO problem can be formulated as a second-order conic program based on marginal moments, which reduced the computational complexity considerably. They also presented a closed form of the worst-case distribution; however their formulation requires the assumption that job durations could take on negative values.

Deng et al. (2014) also considered a DRO variant of the chance constrained OR scheduling problem to restrict the maximum risk of surgery delay and overtime. They built a confidence set using statistical divergence functions, which is equivalent to a chance-constrained program evaluated on an empirical probability function but with smaller risk tolerances. Once the risk tolerance is fixed, their DRO problem becomes a sample approximated chance constrained program.

2.4. Contributions to the Literature

Our approach differs from the aforementioned literature in two main aspects: First, we present a new branch-and-price approach for a SMIP with integer recourse and big-M constraints that is particularly suitable for sample approximated chance constrained programs. Lulli and Sen (2004) first applied branch-and-price to chance constrained problems, and we improve on their implementation; specifically, we adopt the probabilistic cover reformulation from Song et al. (2014) on chance constraints, which allows us to solve subproblems efficiently, and more importantly, the identification of probabilistic covers and generating columns can be done simultaneously. We show that our

approach significantly outperforms a standard branch-and-price algorithm and branch-and-bound as implemented in CPLEX 12.6. Second, we present a new DRO model for chance constrained programs in which only mean and variance of item sizes are known. A recent study by Jiang and Guan (2013) considered a series of DRO models for chance constrained programs. Our article presents, to our knowledge, the first closed form expression for chance constraints in the distributionally robust context. Based on the closed form result, we are able to build the DRO model as a MIP under mild assumptions; this differs from the widely used second-order conic program (e.g., Mak et al. (2015)) and semidefinite program (e.g., Kong et al. (2013)) in the DRO literature. Computational experiments show that practical instances of the resulting model are easily solved.

3. Problem Description and Model Formulation

We first formulate CCSBP as a two-stage stochastic mixed-integer program based on the sample average approximation (SAA) (Luedtke and Ahmed (2008)). There are n items to be allocated among m bins which are classified into $K + 1$ types of bins (we assume $n > m$ to avoid the trivial case). There are m_k bins of type k ($k \leq K$) that have chance constraints with respect to a nominal bin size, T , plus a bin extension threshold, δ_k . We assume all chance constrained bins can contain at least one item for each. We let \mathcal{I} denote the set of items and \mathcal{R} denote the set of bins. We let \mathcal{R}_k denote the set of chance constrained bins of type k . We let Ω denote the finite set of all possible scenarios based on the SAA. Finally, we let $d_i(\omega)$ denote the size of item i under scenario ω where ω indexes a finite set of realizations of the random outcome vector $\xi(\omega) = \{d_1(\omega), \dots, d_n(\omega)\}$. Item sizes are assumed to be mutually independent. Indices i and j index items, r indexes bins, k indexes bin types and ω indexes scenarios.

We consider the following decision variables:

- x_{ir} : binary variable representing whether item i is allocated to bin r ,
- $o_r(\omega)$: decision variable representing extension for bin r under scenario ω .

The extension threshold of each chance constrained bin can be exceeded with at most probability α where α is a probability tolerance. The chance constraints can be expressed as follows:

$$\mathbb{P}\left(\sum_{i \in \mathcal{I}} x_{ir} d_i(\omega) \leq T + \delta_k\right) \geq 1 - \alpha, \quad \forall k, \quad \forall r \in \mathcal{R}_k \quad (1)$$

where $\mathbb{P}(\cdot)$ is a probability distribution, and $\sum_{i \in \mathcal{I}} x_{ir} d_i(\omega)$ is the total item size for bin r under scenario ω .

Bin size T need not necessarily be the same for different bins, and the probability tolerance α is not necessarily the same for different types of bins; however, we consider identical T and α for simplicity of exposition.

Inequality (1) is, except for certain special cases, a nonlinear and nonconvex function. To linearize (1), we use the common approach of introducing auxiliary binary variables, $z_{kr}(\omega)$, with $z_{kr}(\omega) = 1$ representing that bin r satisfies the chance constraint under scenario ω , and $z_{kr}(\omega) = 0$ otherwise. The extensive form of CCSBP can now be formulated as follows:

$$(P): \quad \min \sum_{r \in \mathcal{R}} \mathbb{E}_{\omega \in \Omega} [o_r(\omega)] \quad (2a)$$

$$s.t. \quad \sum_{r \in \mathcal{R}} x_{ir} = 1, \quad \forall i \quad (2b)$$

$$o_r(\omega) \geq \sum_{i \in \mathcal{I}} x_{ir} d_i(\omega) - T, \quad \forall r, \quad \forall \omega \quad (2c)$$

$$\sum_{i \in \mathcal{I}} x_{ir} d_i(\omega) + M_{kr}(\omega) z_{kr}(\omega) \leq T + \delta_k + M_{kr}(\omega), \quad \forall k, \quad \forall r \in \mathcal{R}_k, \quad \forall \omega \quad (2d)$$

$$\sum_{\omega \in \Omega} z_{kr}(\omega) \geq (1 - \alpha) \cdot |\Omega|, \quad \forall k, \quad \forall r \in \mathcal{R}_k \quad (2e)$$

$$x_{ir}, z_{kr}(\omega) \in \{0, 1\}, o_r(\omega) \geq 0, \quad \forall i, \quad \forall r, \quad \forall k, \quad \forall \omega \quad (2f)$$

where $M_{kr}(\omega)$ is a suitably large constant and $|\cdot|$ is a cardinality function. The objective (2a) is to minimize the sum of the expected cost of bin extensions. Constraint (2b) ensures each item is allocated to exactly one bin. Constraint (2c) determines the extension of each bin. Constraints (2d) and (2e) are the linearization of chance constraint (1) where (2d) determines $z_{kr}(\omega)$ and (2e) enforces the probability tolerance. Finally, constraint (2f) restricts the decision variables. Note that our formulation assumes the number of bins is fixed at m . This is consistent with the first formulation of the extensible bin packing problem presented by Dell’Olmo et al. (1998) but slightly different from Denton et al. (2010), in which the decision to utilize a bin is a binary decision variable.

There exists symmetry among the item-to-bin allocation decisions when there are identical bins and items of the same type. The following antisymmetry constraints, which are extended from Denton et al. (2010), are added to this model to reduce the bin related symmetry:

$$\sum_{k=1}^{K+1} \sum_{r=u_k+1}^{\tilde{u}_{ik}} x_{ir} = 1, \quad \forall k, \quad \forall i \leq \max_k \{m_k\}, \quad (3a)$$

$$\sum_{r=u_k+q}^{\tilde{u}_{ik}} x_{ir} \leq \sum_{j=q-1}^{i-1} x_{j, u_k+q-1}, \quad \forall k, \quad \forall i \geq 2, \quad \forall q \leq \min(i, m_k), \quad (3b)$$

where $i = 1, \dots, n$, $k = 1, \dots, K+1$, $u_k = \sum_{v=1}^{k-1} m_v$, and $\tilde{u}_{ik} = \min(u_k + i, u_{k+1})$. Constraint set (3) ensures items are allocated to bins in lexicographic order, where (3a) requires that item i should

be allocated to one of the first i bins of a specific type, and (3b) requires that if the $(q-1)^{th}$ bin of a specific type is not allocated for any of the first $i-1$ items, then the q^{th} bin of that type should not be allocated for item i .

The following constraints eliminate symmetry across items:

$$\sum_{q=1}^r x_{jq} \leq \sum_{q=1}^r x_{iq}, \forall i, \forall j \in S_i, \forall r, \quad (4)$$

where $S_i = \{j \mid j > i, \mathbf{d}_j \stackrel{d}{=} \mathbf{d}_i\}$, meaning $\mathbf{d}_i, \mathbf{d}_j$ are identically distributed. Constraint (4) ensures that items of the same type are allocated to bins in lexicographic order.

Big-M constraints present a challenge to solving (P). The following inequalities can be used to strengthen $M_{kr}(\omega)$ in (2):

$$M_{kr}(\omega) \leq \sum_{i=r-u_{k-1}}^n d_i(\omega) - T - \delta_k, \forall k, \forall r \in \mathcal{R}_k, \forall \omega \quad (5a)$$

$$M_{kr}(\omega) \leq M_k(\omega) = \sum_{i=1}^n d_i(\omega) - \sum_{j=1}^{m-1} d_{[j]}(\omega) - T - \delta_k, \forall k, \forall r, \forall \omega \quad (5b)$$

where $d_{[j]}(\omega)$ represents the j^{th} shortest size among items in set \mathcal{I} under scenario ω . Inequality (5a) is from the symmetry breaking constraint (3a) that no item among $\{1, \dots, r - u_{k-1} - 1\}$ will be allocated in bin r where u_k is defined in (3). Inequality (5b) follows from the condition that every bin contains at least one item.

Constraints (3)-(5) can be used to strengthen model (2); however, our computational results (see Section 6) show that the strengthened model (2) is solvable via a sophisticated MIP solver such as CPLEX 12.6 only for very small problem instances. Moreover, the integer L-shaped method (Birge and Louveaux (2011)) is well known to be inefficient for formulations such as this because of weak optimality cuts resulting from big-M coefficients (Codato and Fischetti (2006)). There exist recent adaptations of Benders decomposition to reformulate big-Ms (Codato and Fischetti (2006), Luedtke (2014), Liu et al. (2014)); however, as we show, branch-and-price offers a means to avoid big-M constraints and take advantage of the special structure of the resulting subproblem.

4. Solution Methodology

In this section, we present a column generation reformulation of CCSBP and a branch-and-price algorithm to solve the integer master problem. Further, we derive some analytical properties of the problem structure that can be used to accelerate the algorithm.

4.1. Column Generation Reformulation

The column generation formulation consists of a *master problem* and a *subproblem*. The master problem assumes that there already exists a pool of packings and decides which packings to select from the pool to minimize the total expected bin extensions. The subproblem (also referred to as the *pricing problem*), on the other hand, generates new packings for the master problem until it is solved to optimality.

A packing, $p \in \mathcal{P}$, represents a group of items that will be allocated in the same bin, and it corresponds to the following parameters:

- y_{ip} : binary parameter representing whether item i is in packing p ,
- c_{kp} : binary parameter representing whether the total item size in packing p is chance-constrained for bins of type k ,
- \bar{o}_p : expected extension for packing p which is an expectation over all $o_p(\omega)$.

The decision variables in the master problem are as follows:

- λ_p : binary variable representing whether packing p is selected.

We formulate the master problem (MP) as the following mixed integer program:

$$(MP): \quad \min \sum_{p \in \mathcal{P}} \bar{o}_p \lambda_p \quad (6a)$$

$$s.t. \quad \sum_{p \in \mathcal{P}} y_{ip} \lambda_p = 1, \quad \forall i \quad (6b)$$

$$\sum_{p \in \mathcal{P}} c_{kp} \lambda_p \geq \sum_{v=1}^k m_v, \quad \forall k \quad (6c)$$

$$\sum_{p \in \mathcal{P}} \lambda_p \leq m \quad (6d)$$

$$\lambda_p \in \{0, 1\}, \quad \forall p \quad (6e)$$

The objective (6a) is to minimize the total expected bin extensions. Constraint (6b) ensures that each item should be covered in one of the selected packings. Constraint (6c) ensures that the number of chance constrained packings should be at least the number of chance constrained bins. Constraint (6d) limits the number of packings to be at most the number of bins. Finally, constraint (6e) restricts the variables to be binary.

The column generation formulation (MP) represented by (6) is equivalent to the extensive formulation (P) represented by (2) because any feasible allocation of items to bins in (2) can be regarded as packings in (6) and the optimal solution of (6) can be converted to a feasible solution to (2). Unfortunately, (MP) is, generally speaking, intractable because it is difficult to enumerate

all possible packings in \mathcal{P} and derive an integer solution. Thus, we relax (6e) to be $\lambda_p \geq 0$ and refer to the resulting relaxed master problem as (RMP). Moreover, we consider a restricted pool \mathcal{P}' ($\mathcal{P}' \subset \mathcal{P}$) with a limited number of packings, and we refer to the restricted (RMP) as (R²MP).

Given an optimal solution to (R²MP), we define the corresponding dual solution as follows:

- π_i : dual price corresponding to (6b) for the extension reduction if item i is not allocated,
- ϕ_k : dual price corresponding to (6c) for the extension reduction if bins of type k are not chance constrained,
- φ : dual price corresponding to (6d) for the extension increase if the number of bins is reduced by 1.

The above dual solution is passed to the following subproblem and additional columns are generated as needed. Based on the dual solutions, the subproblem determines which items to be selected into a new packing (i.e. $y_{ip} = 1$), whether this packing is chance constrained (i.e. $c_{kp} = 1$), and the expected extension of the new packing (i.e. $\bar{o}_p = \mathbb{E}_{\omega \in \Omega}[o_p(\omega)]$). The subproblem (SP) is formulated as the following two-stage stochastic mixed integer program:

$$(\text{SP}): \quad z_{SP} = \max \sum_{i=1}^n \pi_i y_{ip} + \sum_{k=1}^K \phi_k c_{kp} - \mathbb{E}_{\omega \in \Omega}[o_p(\omega)] \quad (7a)$$

$$s.t. \quad \sum_{i \in \mathcal{I}} y_{ip} d_i(\omega) - o_p(\omega) \leq T, \quad \forall \omega \quad (7b)$$

$$\sum_{i \in \mathcal{I}} y_{ip} d_i(\omega) + M_k^1(\omega) z_{kp}(\omega) \leq T + \delta_k + M_k^1(\omega), \quad \forall k, \quad \forall \omega \quad (7c)$$

$$\sum_{\omega \in \Omega} z_{kp}(\omega) - M_k^2 c_{kp} \geq -M_k^2 + (1 - \alpha) \cdot |\Omega|, \quad \forall k \quad (7d)$$

$$y_{ip}, c_{kp}, z_{kp}(\omega) \in \{0, 1\}, o_p(\omega) \geq 0, \quad \forall i, \quad \forall k, \quad \forall \omega, \quad (7e)$$

where $z_{kp}(\omega)$ are auxiliary binary variables representing whether packing p satisfies the chance constraint under scenario ω . $M_k^1(\omega)$ and M_k^2 are suitably large constants where $M_k^1(\omega)$ is defined by (5b) and $M_k^2 \leq (1 - \alpha) \cdot |\Omega|$. Constraint (7b) determines the extension in the new packing. Constraints (7c)-(7d) jointly ensure the chance constraints on total item size in each packing. Finally, constraint (7e) restricts decision variables.

The objective (7a) is to maximize the reduced cost with respect to the dual solutions of the current (R²MP). If $z_{SP} > \varphi$, (SP) generates a new packing that further improves (R²MP) while $z_{SP} \leq \varphi$ indicates the current solution is already optimal to (RMP). Thus, the column generation approach solves the (RMP) by iteratively solving (R²MP) and (SP). The problem is solved when (SP) cannot generate an improving packing.

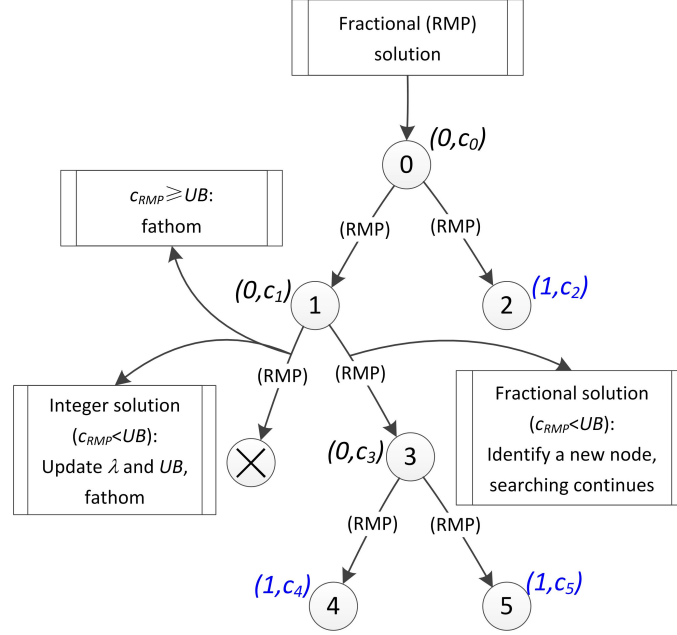


Figure 1 Outline of the branch-and-price algorithm

4.2. Branch-and-Price Algorithm

We now present a branch-and-price algorithm to find an optimal integer solution for (MP). Based on a fractional solution to (RMP), branching is carried out to identify branching nodes, and add integer cuts to (RMP). The pricing part of the algorithm solves the revised (RMP) model using column generation as discussed in Section 4.1.

Figure 1 gives an outline of the branch-and-price algorithm. Each node is characterized by a pair of elements (P_s, c_s) ; P_s is a binary parameter representing whether node s is *pending* ($P_s = 1$), or *searched* ($P_s = 0$); c_s represents the cost to (RMP) obtained by column generation on node s . We assume an initial integer solution λ^0 to (RMP) and an upper bound UB at root node 0. We also have an initial fractional solution to (RMP) with cost c_0 ($c_0 < UB$). Starting from this fractional solution, the algorithm first identifies a node from which to branch. Integer cuts are added to each branch and the revised (RMP) model for the new nodes are solved using column generation. If the cost is greater or equal to UB , the new node is fathomed. If the cost is lower than UB and the solution is integer, we update the optimal solution and UB , and no further branching is conducted on that node. If the solution is fractional, a new node is generated and added to the list of pending nodes (later we show how to generate a pending node). At each iteration, the algorithm selects the pending node with the lowest cost as the next node from which to branch. The algorithm stops when all nodes have been fathomed or their costs to (RMP) are no less than UB .

To describe the branch-and-price algorithm in detail, we define the node N (indexed by s) as follows:

$$N_s := ((i_s, j_s), c_s, P_s, PN_s, B_s)$$

where (i_s, j_s) represents a pair of items in \mathcal{I} , $i_s \neq j_s$; c_s is the cost to (RMP) up to node N_s ; P_s indicates whether N_s is pending or not; PN_s is the index of the parent node of N_s ; B_s is a binary parameter which indicates whether N_s is the left child node of PN_s ($B_s = L$), or the right child node ($B_s = R$).

We use the following pair-based branching strategy suggested by Barnhart et al. (1998). On the left branch, we allocate both items i_s and j_s to the same bin (i.e. same packing). On the right branch, we allocate i_s and j_s to separate bins. To implement this strategy, we identify infeasible columns in the (R²MP) model to avoid selecting them, and the infeasible columns are indicated as follows:

$$I_p(B, s) := \begin{cases} 1, & \forall p: y_{i_s, p} + y_{j_s, p} = 1, & \text{if } B = L, \\ 1, & \forall p: y_{i_s, p} + y_{j_s, p} = 2, & \text{if } B = R, \\ 0, & & \text{otherwise.} \end{cases}$$

We also update the (SP) model by adding integer cuts to avoid infeasible packings. The cuts are defined as follows:

$$Cut(B, s) := \begin{cases} y_{i_s, p} - y_{j_s, p} = 0, & \text{if } B = L, \\ y_{i_s, p} + y_{j_s, p} \leq 1, & \text{if } B = R. \end{cases}$$

When a fractional solution is found, we generate a new node that contains a pair of items. The pair must differ from all pairs of the predecessor nodes to avoid redundant constraints. We search for the pair in columns with fractional λ_p in the order of fewest number of items in each column.

4.3. Antisymmetry Branching

Symmetry exists in the stochastic bin packing problem due to multiple items of the same type. This is well known to create significant computational challenges (Denton et al. (2010)). We now describe a way to eliminate the symmetry across identical items in the node branching. We begin with an example with three items, i_1, j_1 and j_2 (see Figure 2) where j_1 and j_2 are identical items. We let $i_1 \oplus j_1$ denote that i_1 and j_1 are in the same bin while $i_1 \ominus j_1$ denotes they are in separate bins. Since j_1 and j_2 are identical, in Figure 2, nodes N_4 and N_5 represent the same constraint. Thus, we need to fathom either N_4 or N_5 to break the symmetry.

Our antisymmetry branching approach is implemented as follows: for each new node (say N_5 , in this example) that combines two items (in this example, say j_2 is one of them), we identify whether there exists a predecessor node of N_5 that separated two items including one (say j_1 , in

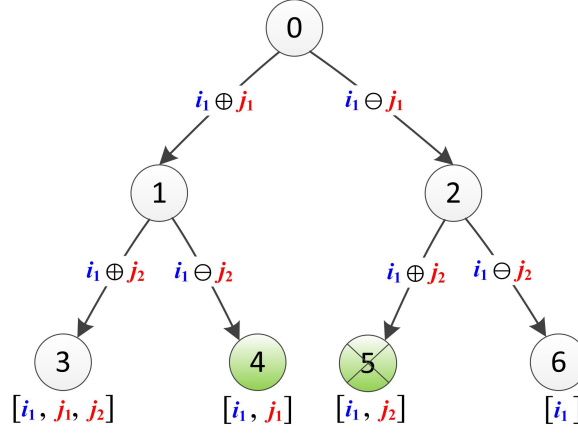


Figure 2 Illustration of antisymmetry branching

this example) identical to j_2 . If so, we fathom this new node. As a result, N_5 will be fathomed in this example. This antisymmetry branching identifies and removes redundant nodes

The complete branch-and-price algorithm is presented in Algorithm 1.

4.4. Analytical Properties of (SP)

The (SP) model is a two-stage stochastic mixed integer program and most of the computation is spent on solving (SP). However, solving this problem can be accelerated by considering the following approximations.

Lower Bound for (SP): we show that a conditional value-at-risk ($CVaR$) formulation can provide a lower bound for (SP). For a random number z , the value-at-risk (VaR) and $CVaR$ are defined as follows: $VaR(z) = \inf(\beta' \mid \mathbb{P}(z \leq \beta') \geq 1 - \alpha)$ and $CVaR(z) = \inf(\beta \mid \beta + \frac{1}{\alpha} \mathbb{E}(z - \beta)^+)$. It is known that $VaR(z) \leq CVaR(z)$ for a given probability $1 - \alpha$ (Rockafellar and Uryasev (2000)). This leads to the following theorem (all proofs to theorems, lemmas and corollaries are in the Appendix):

THEOREM 1. *The following constraints are a restriction of (7c)-(7d):*

$$\begin{cases} \beta + \frac{1}{\alpha \cdot |\Omega|} \sum_{\omega \in \Omega} \gamma(\omega) + M_k^2(\omega) c_{kp} \leq T + \delta_k + M_k^2(\omega), \quad \forall k, \\ \beta + \gamma(\omega) - \sum_{i \in \mathcal{I}} y_{ip} d_i(\omega) \geq 0, \quad \forall \omega, \\ \beta, \gamma(\omega) \geq 0, \quad \forall \omega. \end{cases} \quad (8)$$

The $CVaR$ -based approximation with (7a)-(7b),(7e) and (8) provides a lower bound on (7). We refer to this model as (LBSP). It is a two-stage stochastic program with convex recourse, which is easier to solve than (7).

Upper Bound for (SP): we show a probabilistic cover based formulation (ref. Song et al. (2014)) can provide an upper bound on (7). Define a probability function ϕ^k on a give set \mathcal{A} of items:

Algorithm 1 Branch-and-Price algorithm

Input: Root node $N_0 := ((i_0, j_0), c_0, 1, -, -)$, an integer solution λ^0 , and the *tolerance* of accuracy

Output: Optimal solution λ^*

Procedures:

- 1: Set $LB := c_0$ and $UB := \infty$. Let $\lambda^* = \lambda^0$, list of nodes $List := \{N_0\}$ and node number $l := 1$
 - 2: Select a pending node: $\hat{s} = \arg_{N_s \in List} \min\{c_s \mid P_s = 1\}$
 - 3: Add constraint $Cut(B_v, PN_v)$ to (SP) and penalize columns with $I_p(B_v, PN_v) = 1$ in (R²MP) for $v = \hat{s}$ and each of predecessor nodes. Let $\mathfrak{B} = \{R\}$ if branch L removed by antisymmetry branching, otherwise, $\mathfrak{B} = \{L, R\}$
 - 4: Branch nodes: **for** $B \in \mathfrak{B}$ **do**
 - a. Add constraint $Cut(B, \hat{s})$ to (SP), and penalize columns with $I_p(B, \hat{s}) = 1$ in (R²MP)
 - b. Solve the revised (RMP). Let $\lambda_{RMP}(\hat{s}, B)$ be the solution and $c_{RMP}(\hat{s}, B)$ be the cost
 - c. If $(UB - c_{RMP})/UB \leq tolerance$, go to step f
 - d. If $\lambda_{RMP}(\hat{s}, B)$ is fractional: set $l = l + 1$, $c_l = c_{RMP}(\hat{s}, B)$, $P_l = 1$, $PN_l = \hat{s}$, $B_l = B$ and identify a new pair of items (i_l, j_l) . Add N_l to the *List*
 - e. If $\lambda_{RMP}(\hat{s}, B)$ is integer: set $\lambda^* = \lambda_{RMP}(\hat{s}, B)$; $UB = c_{RMP}(\hat{s}, B)$
 - f. Remove the constraint $Cut(B, \hat{s})$, and remove the penalty on columns with $I_p(B, \hat{s}) = 1$
 - 5: Set $P_{\hat{s}} = 0$. Remove the constraint $Cut(B_v, PN_v)$, and remove the penalty on columns with $I_p(B_v, PN_v) = 1$ for $v = \hat{s}$ and each of predecessor nodes
 - 6: Update LB : $LB = \min(UB, \min_s\{c_s \mid P_s = 1\})$
 - 7: If $(UB - LB)/UB > tolerance$, go to step 2; otherwise stop.
-

$$\phi^k(\mathcal{A}) = \mathbb{P} \left(\sum_{i \in \mathcal{A}} y_{ip} d_i(\omega) > T + \delta_k \right).$$

Given a probability tolerance, α , a set \mathcal{C}_k is called a *probabilistic cover* if $\phi^k(\mathcal{C}_k) > \alpha$ whereas a set \mathcal{P}_k is called a *probabilistic pack* if $\phi^k(\mathcal{P}_k) \leq \alpha$. We define $\mathcal{Q}(\mathcal{P}_k)$ as a *complimentary set* of \mathcal{P}_k such that $\phi^k(\mathcal{P}_k \cup i) > \alpha$, $\forall i \in \mathcal{Q}(\mathcal{P}_k)$.

Next we have the following theorem:

THEOREM 2. *The following constraints are a relaxation of (7c)-(7d) and the relaxation bound is tight when all possible sets of \mathcal{C}_k and \mathcal{P}_k are considered:*

$$\sum_{i \in \mathcal{Q}(\mathcal{P}_k)} y_{ip} + M_{\mathcal{P}_k} \left(c_{kp} + \sum_{j \in \mathcal{P}_k} y_{jp} \right) \leq M_{\mathcal{P}_k} (|\mathcal{P}_k| + 1), \quad \forall \mathcal{P}_k, \quad (9a)$$

$$\sum_{j \in \mathcal{C}_k} y_{jp} + M_{\mathcal{C}_k} c_{kp} \leq |\mathcal{C}_k| - 1 + M_{\mathcal{C}_k}, \quad \forall \mathcal{C}_k, \quad (9b)$$

where $M_{\mathcal{C}_k}, M_{\mathcal{P}_k}$ are suitably large constants and $|\cdot|$ is the cardinality function.

The *cover-based* approximation with (7a)-(7b), (7e) and (9) provides an upper bound on (7). The $M_{\mathcal{C}_k}$ and $M_{\mathcal{P}_k}$ are bounded by the following inequalities: $M_{\mathcal{C}_k} \leq n - m + 2 - |\mathcal{C}_k|$ and $M_{\mathcal{P}_k} \leq n - m + 1 - |\mathcal{P}_k|$, respectively. We refer to this model as (UBSP). It is a MIP model which is easier to solve than (7). Although it is difficult to compute all possible sets of \mathcal{C}_k and \mathcal{P}_k , sets and their related constraints can be iteratively added into (UBSP) as needed. As the constraints are accumulative in the column generation procedure, after a few iterations, the (UBSP) becomes very efficient at generating feasible columns. As we show in Section 6, this makes the branch-and-price algorithm much more efficient.

5. Distributionally Robust Counterpart

In this section, we consider a robust counterpart to CCSBP that applies to the situation where item sizes are ambiguous. We first present a closed form expression for the worst-case probability that a bin exceeds its extension threshold, and then we formulate a MIP model based on this expression for the DRO chance constrained bin packing problem.

5.1. DRO with Moment Information

We relax the assumption that the distribution of each item size is known, and we consider the case in which only the mean μ_i , and variance σ_i^2 of the item sizes are known, which can be expressed by the following confidence set:

$$\mathcal{D} = \left\{ \mathbb{D} \in \mathcal{M}_+ : \mathbb{E}[\mathbf{d}_i] = \mu_i, \mathbb{E}[\mathbf{d}_i^2] = \mu_i^2 + \sigma_i^2, \mathbf{d}_i \in \mathbb{R}_+, \forall i \in \mathcal{I} \right\},$$

where \mathcal{M}_+ represents the set of all probability distributions.

We assume that the extension threshold of each chance constrained bin can be exceeded with at most probability α under the worst-case distribution in the confidence set, \mathcal{D} . We could employ the approach described in Jiang and Guan (2013); however, this would result in a mixed integer second-order conic program. Instead, we consider the following formulation of the worst-case chance constraint:

$$\inf_{\mathbb{D} \in \mathcal{D}} \mathbb{P} \left(\sum_{i \in \mathcal{I}} x_{ir} \mathbf{d}_i \leq T + \delta_k \right) \geq 1 - \alpha, \quad \forall k, \forall r \in \mathcal{R}_k. \quad (10)$$

The SAA formulation is no longer well defined because SAA involves sampling from a distribution that is fixed in advance, and the distribution is assumed to be ambiguous in this case. To tackle

this problem, we show that there exists a worst-case distribution that is at most a three-point distribution.

In this context we assume the objective is to minimize the total extension of bins based on the mean value of item sizes, and the DRO chance constrained bin packing problem can be expressed as follows:

$$\min_{x \in \mathcal{X}} \sum_{r \in \mathcal{R}} \left(\sum_{i \in \mathcal{I}} x_{ir} \mu_i - T \right)^+ \quad (11)$$

where \mathcal{X} is a feasible region defined by (2b), (10) and $x \in \{0, 1\}_{n \times m}$.

The worst-case expected bin extension can also be incorporated into our models as a MISOCP model as shown in the Appendix. We consider the mean-based bin extension because the chance constraints on bin extensions are our main interest; as we show later, this consideration makes it possible to reformulate the robust problem (11) as a MIP model with considerably fewer binary variables than the stochastic counterpart.

To reformulate (11), we first consider an aggregate item size X with mean μ and standard deviation σ . Its confidence set is defined as:

$$\mathcal{D}_X = \{ \mathbb{D} \in \mathcal{M}_+ : \mathbb{E}[X] = \mu, \mathbb{E}[X^2] = \mu^2 + \sigma^2, X \in \mathbb{R}_+ \}.$$

Then we consider a nominal worst-case chance constraint with respect to a given constant B :

$$\sup_{\mathbb{D} \in \mathcal{D}_X} \mathbb{P}(X \geq B) \leq \alpha. \quad (12)$$

We will give a discrete distribution expression of \mathcal{D}_X in Subsection 5.2 and a reformulation of (12) in Subsection 5.3. Finally, we reformulate the DRO problem (11) as a MIP in Subsection 5.4.

5.2. Discrete Distribution Expression of \mathcal{D}_X

We show that there exists a finite set of points, $\{x_1, \dots, x_N\}$, such that moments and chance constraints of any distribution in \mathcal{D}_X can be matched by adjusting the probabilities, p_1, \dots, p_N , associated with the points in this set.

LEMMA 1. *For any random variable, $Y \in [a, \infty)$, with finite mean and variance, there exists a one- or two-point distribution $\hat{\mathbb{D}}$ of Y with all points in $[a, \infty)$. Further, if $Y \in [a, b)$ with $b > a$, then the points of $\hat{\mathbb{D}}$ are also in $[a, b)$.*

Next we have the following theorem:

THEOREM 3. *For any distribution $\mathbb{D} \in \mathcal{D}_X$, there exists a discrete distribution*

$$\mathbb{D}' = \{(p_1, x_1), (p_2, x_2), \dots, (p_N, x_N)\}$$

such that \mathbb{D}' has the same mean and variance as \mathbb{D} and $\mathbb{P}_{X \sim \mathbb{D}}(X \geq B) = \mathbb{P}_{X \sim \mathbb{D}'}(X \geq B)$ and $N \leq 4$.

Theorem 3 establishes that all distributions in \mathcal{D}_X have a corresponding discrete distributions. This allows us to formulate an LP model for the worst-case chance probability $\sup_{\mathbb{D} \in \mathcal{D}_X} \mathbb{P}(X \geq B)$.

5.3. Closed Form Expression for the Nominal Chance Constraint

We will first give a closed form expression for $WC = \sup_{\mathbb{D} \in \mathcal{D}_X} \mathbb{P}(X \geq B)$, based on which we further linearize the nominal chance constraint (12). Note that the exact definition of worst-case chance probability is $\widehat{WC} = \sup_{\mathbb{D} \in \mathcal{D}_X} \mathbb{P}(X > B)$; however, we later show that $WC = \widehat{WC}$ in Corollary (1).

LEMMA 2. *If $\mu > B$, then $WC = 1$ and the following two-point distribution is a worst-case distribution:*

$$\mathbb{D}' = \left\{ (1-p, B), \left(p, \frac{\mu^2 + \sigma^2 - \mu B}{\mu - B} \right) \right\}, \quad p = \frac{(\mu - B)^2}{(\mu - B)^2 + \sigma^2}.$$

Using Lemma 2 and Theorem 3 we can prove the following two theorems:

THEOREM 4. *There exists a worst-case discrete distribution that is at most a three-point distribution*

THEOREM 5. *There exists a worst-case chance probability that can be expressed as follows:*

$$WC = \begin{cases} 1, & \text{if } \mu > B, \\ \mu/B, & \text{if } \mu \leq B, \mu^2 + \sigma^2 > \mu B, \\ \frac{\sigma^2}{\sigma^2 + (B - \mu)^2}, & \text{if } \mu \leq B, \mu^2 + \sigma^2 \leq \mu B. \end{cases}$$

COROLLARY 1. *WC is continuous in μ, σ and B and $WC = \widehat{WC}$.*

Note that if $\mu^2 + \sigma^2 > \mu B$, the nominal chance constraint (12) holds only if $\mu/B \leq \alpha$; thus, it is necessary that the coefficient of variation (CV) of X be such that $CV > \sqrt{\frac{1-\alpha}{\alpha}}$. As a result, we assume the following condition holds.

Condition 1. $CV \leq \sqrt{\frac{1-\alpha}{\alpha}}$.

The above condition is mild because typically the probability tolerance α is much less than 0.5 which implies $CV \leq 1.0$ which is likely to be true in most applications (see Table 2 as an example for surgery durations).

Assuming Condition 1, the nominal chance constraint (12) can be reformulated as follows:

$$\mu \leq B, \tag{13a}$$

$$\mu^2 + \sigma^2 \leq \mu B, \tag{13b}$$

$$\frac{1-\alpha}{\alpha} \sigma^2 \leq (B - \mu)^2. \tag{13c}$$

Note that constraint (13b) dominates (13a) because $\sigma^2 > 0$. Thus, (13a) can be removed from our formulation.

5.4. DRO Reformulation

Now we reformulate the distributionally robust optimization problem (11) of the chance constrained bin packing. We let $X = \sum_{i \in \mathcal{I}} x_{ir} \mathbf{d}_i$, denoting the total size of items allocated to each bin and we reformulate DRO problem (11) as a MIP as shown in the following theorem:

THEOREM 6. *The DRO problem (11) of the chance constrained bin packing problem can be reformulated as the following MIP model:*

$$(\text{ROP}): \quad \min \sum_{r \in \mathcal{R}} o_r \quad (14a)$$

$$s.t. \quad o_r \geq \sum_{i \in \mathcal{I}} x_{ir} \mu_i - T, \quad \forall r \quad (14b)$$

$$\sum_{r \in \mathcal{R}} x_{ir} = 1, \quad \forall i \quad (14c)$$

$$\sum_{i \in \mathcal{I}} x_{ir} \Gamma_{ik} + 2 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}, j > i} z_{ijr} \mu_i \mu_j \leq 0, \quad \forall k, \quad \forall r \in \mathcal{R}_k \quad (14d)$$

$$\sum_{i \in \mathcal{I}} x_{ir} \Lambda_{ik} - 2 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}, j > i} z_{ijr} \mu_i \mu_j \leq (T + \delta_k)^2, \quad \forall k, \quad \forall r \in \mathcal{R}_k \quad (14e)$$

$$z_{ijr} \geq x_{ir} + x_{jr} - 1, \quad \forall i, \quad \forall j, \quad \forall r \quad (14f)$$

$$z_{ijr} \leq x_{ir}, \quad z_{ijr} \leq x_{jr}, \quad \forall i, \quad \forall j, \quad \forall r \quad (14g)$$

$$x_{ir} \in \{0, 1\}, z_{ijr}, o_r \geq 0, \quad \forall i, \quad \forall j, \quad \forall r \quad (14h)$$

where $\Gamma_{ik} = \mu_i^2 + \sigma_i^2 - T\mu_i - \delta_k\mu_i$, $\Lambda_{ik} = \frac{1-\alpha}{\alpha}\sigma_i^2 + 2(T + \delta_k)\mu_i - \mu_i^2$.

The objective (14a) is to maximize the total extension under mean sizes which is determined by (14b). Constraint (14c) ensures every item is allocated to exactly one bin. Constraints (14d)-(14g) are chance constraints under the worst-case distribution, which are from (13b)-(13c), respectively. Constraint (14h) restricts the decision variables. Note that we can also add the antisymmetry constraints (3) and (4) to eliminate the symmetries across identical bins and items of the same type to strengthen the formulation. The column generation formulation of the DRO model (14) is relegated to the Appendix.

We summarize this section by the following main conclusions: (a) there exists a closed form expression for the chance constraint (12) based on a three-point distribution, as given in Theorem 5; (b) the worst-case chance probability is continuous in μ, σ and B ; (c) the DRO problem (11) of the chance constrained bin packing can be formulated as the mixed-integer program (14).

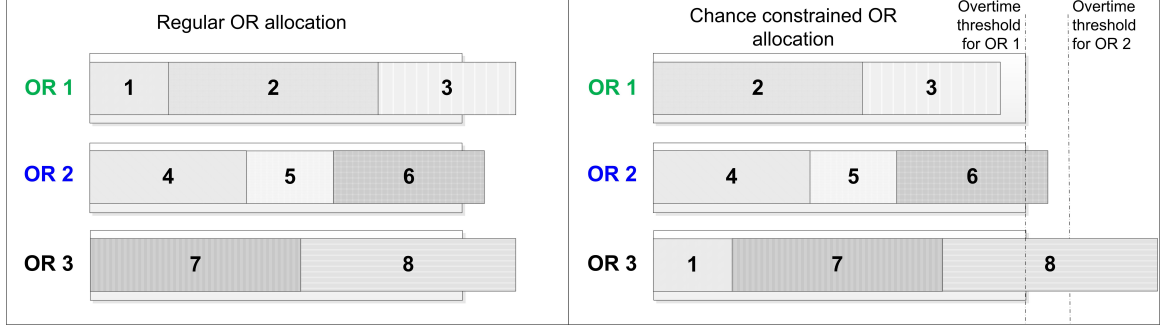


Figure 3 Comparison of regular OR allocation and chance constrained OR allocation for a small example

6. Computational Results

In this section, we test our proposed approaches on instances of CCSBP and DRO. Since there are no standard test instances for this problem yet, we construct model instances in the context of surgery scheduling based on real data and we use these instances to evaluate the performance of the branch-and-price algorithm for the CCSBP model and the DRO model. In this context the bins are ORs that are available for some nominal period of time during the day of surgery (e.g. 10 hours). The items represent surgeries and the surgery durations represent the item sizes. Figure 3 illustrates the problem with a small example that has three ORs and eight surgeries.

This is an important application due to the high cost of OR utilization and OR staff overtime (Deng et al. (2014)). The latter is associated with anxiety and poor moral among OR nurses, and ultimately a high turnover rate among nurses (see a report by Nursing Solutions (2013)). This is particularly important because there is a severe shortage of OR nurses nationwide. To address concerns about overtime, some hospital administrators are implementing a rotating schedule (as illustrated in the right side of Figure 3). Under this paradigm a portion of the ORs have low overtime so that nurses assigned to the OR have predictable shift completion times. Staff then rotate among ORs from day to day so that all staff benefit from predictable shift durations on some portion of the surgery days.

6.1. Parameter Estimation

We use data from the Inpatient Surgical Suite at Ruijin Hospital in Shanghai, China. The department consists of twenty-one ORs serving seven surgical services including general, cardiac, urology, obstetrics and gynecology, orthopaedics, neurosurgery, and Chinese physical therapy (or therapy for short). More than 16,000 procedures are performed per year. In this department, ORs are classified into *pods*, where each pod is a group of ORs associated with a certain group of surgical services. Nurses are also staffed within the same pod and planning is carried out independently for

Table 1 Set based classification of problem instances for 16 surgical days

Set No.	No. of Days	No. of Surgeries		Sum of surgery Durations per OR	
		Average (per day)	Interval (per day)	Average (per day)	Maximum (per day)
1	2	14.50	[11, 15]	5.75	6.71
2	7	22.86	[21, 25]	8.19	9.71
3	7	28.14	[26, 30]	10.61	11.43

the pods. We focus on the largest pod that consists of 9 ORs and serves for surgeries from general, urology and therapy.

Our parameter estimates are based on historical data for 16 surgical days provided by the department and our discussions with the hospital administrator. For each day, the following information is used to construct problem instances:

- Number and type of surgeries
- Actual surgery duration for each type of surgery
- Total number of surgeries and sum of surgery durations per OR on each surgical day

Since the workload of the department varies from day to day, we classify the data in Table 1 into three sets based on the number of surgeries on a single day. We classify surgeries into types and estimate their durations in Table 2 based on historic data.

A regular OR session is 10 hours in duration with overtime accruing beyond 10 hours. To provide nurses with predictable completion times, a typical plan would close 3 ORs in 10 hours, close another 3 ORs in 12 hours, and keep the rest of the ORs open until all surgeries in that pod are performed. Based on the authors' experience, this practice is similar to hospitals in many countries, including the United States. The problem instances are generated as follows: for a given problem set, we first randomly select a number of surgeries, n , from the interval, then randomly generate n surgery types based on their percentages. Finally, we use the mean and standard deviation of the n surgery types as model inputs. For the stochastic model, we further generate 500 independently and lognormally distributed scenarios based on the mean and standard deviation of each surgery type.

6.2. Computational performance of the proposed methods

We implemented the branch-and-price algorithm (Algorithm 1) based on subproblem reformulations (8) and (9) to solve column generation formulation (6) of the stochastic model. We compare its performance with a branch-and-bound algorithm as implemented by the MIP solver of CPLEX

Table 2 Duration and percentage of different type of surgeries

Surgery Type	Percentage	Mean	SD	CV
Throat	0.23	3.70	1.01	0.27
Intestines	0.18	4.32	1.69	0.39
Urology	0.13	2.69	1.19	0.44
Stomach	0.07	4.62	0.89	0.19
Laparoscopic	0.07	2.05	0.96	0.47
Therapy	0.04	0.86	0.55	0.65
Liver	0.03	4.39	1.54	0.35
Pancreas	0.02	4.11	1.11	0.27
Others	0.23	2.75	1.48	0.54

12.6 which solves the extensive formulation (2) strengthened by antisymmetry constraints (3)-(4) and big-M bounds (5). We also compare our method with a standard branch-and-price algorithm based on the original form of subproblem (7) that is solved by the MIP solver.

In our branch-and-price algorithm, the (RMP) was solved based on subproblem reformulations including (LBSP) and (UBSB). We use (LBSP) to generate columns for (R^2MP) at the beginning iterations and then switch to use (UBSP) until the (RMP) is solved to optimality. Specifically, we use (LBSP) first to generate columns that are further improved by applying a local search heuristic. At the same time, we identify probabilistic covers and add related cover constraints to (UBSB). After (LBSP) cannot improve (R^2MP), we switch to use (UBSB) for column generation. Probabilistic sets are identified when infeasible columns are generated. Since the cover constraints in (UBSB) are accumulative, our experimental results show, after a few iterations, the (UBSP) becomes very efficient at generating feasible columns. Figure 4 presents the flow chart of the algorithm that solves the (RMP).

We implemented our algorithms in Microsoft Visual Studio .NET 2012 linking with the CPLEX 12.6 callable library. Experiments were conducted on an Intel Xeon PC with processors running at 3.40 GHz and 16 GB memory under Windows 7. We limited the computation time to a maximum of 3,600 seconds. To compare the computational performance, we randomly chose 10 instances from each of the three instance sets, for a total of 30 instances. We report the solution times (in CPU seconds), the LP relaxation bound and the number of generated columns in Table 3. The relaxation bound is based on the objective ratio of integer and continuous solutions.

As can be observed from Table 3, the branch-and-price algorithms considerably outperformed the branch-and-bound algorithm as implemented by CPLEX. The branch-and-bound algorithm could not solve even for the small-scale instances of Set 1. The column generation reformulations

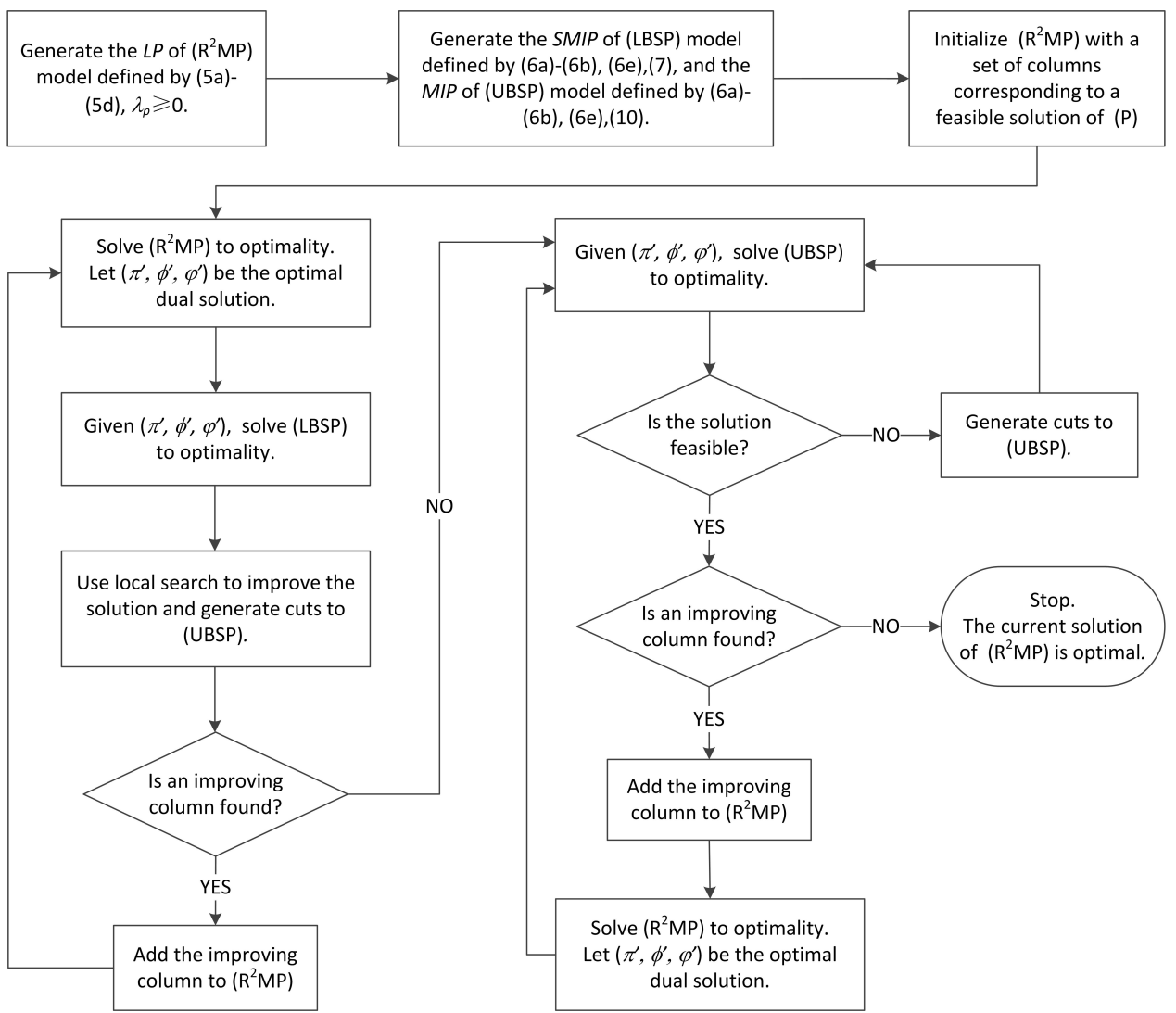


Figure 4 Flow chart for column generation approach that solves (RMP) of the stochastic model

also resulted in a much tighter LP relaxation bound than the branch-and-bound algorithm. Our branch-and-price algorithm outperformed the standard branch-and-price considerably, especially for the large-scale problem instances. Our method solved 29 out of the 30 instances to optimality. For the instance that was not solved to optimality, our method had a gap of 0.3%. The standard branch-and-price could not solve most of the large-scale instances of Set 3. This reflects the fact that the subproblem reformulation reduces the complexity of the problem considerably. Because the (UBSP) is a MIP and (LBSP) is a SMIP with convex recourse, both are easier to solve than the original formulation (SP) which is a SMIP with integer recourse.

Table 3 Performance of different methods for the stochastic model

Set No	No. of Instances	Methods	Solution time (CPU sec.) ^a		LP relaxation Bound ^b		No. of Columns ^c	
			Ave.	Max.	Ave.	Max.	Ave.	Max.
1	10	Branch & bound	581.13	(29.6%)	—	—	—	—
		Standard B&P	1.57	4.03	1.00	1.00	19.3	32
		Our B&P	1.77	3.56	1.00	1.00	18.3	30
2	10	Branch & bound	(42.9%)	(62.2%)	—	—	—	—
		Standard B&P	123.03	540.04	1.15	1.48	60.8	83
		Our B&P	54.40	153.75	1.04	1.18	69.5	117
3	10	Branch & bound	(51.0%)	(60.1%)	—	—	—	—
		Standard B&P ^d	2821.84	(21.8%)	1.17	1.28	107.9	133
		Our B&P	1040.04	(0.3%)	1.12	1.28	145.8	191

^a The optimality gap is presented in brackets for problems that could not be solved within the time limit

^b A dash in this column set indicates “the ratio is infinite”

^c A dash in this column set indicates “does not exist”

^d Results in this row are based on nine instances except one that its LP solution could not be solved within the time limit.

6.3. Value of Robust Solution

In this subsection, we present the computational results for the robust model and analyze the value of the robust solution. We used our branch-and-price algorithm to solve the column generation formulation of the robust model (14). We also compare our method with the branch-and-bound algorithm as implemented by the MIP solver which solves the extensive formulation (14) strengthened by antisymmetry constraints (3)-(4). We tested 30 instances from Sets 1-3 and we limited the computation time to a maximum of 1,500 seconds. We report the solution times (in CPU seconds), the optimality gap and the number of generated columns in Table 4.

As can be observed from Table 4, the robust model is much easier to solve than the stochastic model. All instances were solved to optimality within 500 seconds. The computation for large-scale instances is, on average, 3 times faster than the stochastic problem when both were solved by our branch-and-price algorithm. The extensive form of robust problem is also much easier to solve than the stochastic problem, and the branch-and-bound effectively solved most of the instances in less than 1,500 seconds.

Since the DRO model is much easier to solve than the stochastic model, the robust model could potentially serve as an approximation for CCSBP that obviates the long computation time. As we

Table 4 Performance of different methods for the robust model

Set No	No. of Instances	Methods	Solution time (CPU sec.)		LP relaxation Bound		No. of Columns	
			Ave.	Max.	Ave.	Max.	Ave.	Max.
1	10	Branch & bound	0.09	0.42	—	—	—	—
		Our B&P	0.05	0.17	1.00	1.00	9.0	9
2	10	Branch & bound	216.67	774.56	6.97	18.47	—	—
		Our B&P	50.39	112.78	2.67	12.04	70.5	84
3	10	Branch & bound	1201.88	(35.0%)	3.20	4.03	—	—
		Our B&P	252.83	461.30	1.64	2.49	100.9	116

show in Table 5, by adjusting the goal probability of achieving the chance constraints, the DRO has the potential to generate a similarly good solution to the stochastic problem in a much shorter computation time.

We evaluated the worst-case probability and expected overtime of the solutions to the CCSBP and DRO models with a large sample of 20,000 independently generated scenarios under lognormal (LOGN), uniform (UNIF) and gamma (GAMM) distributions. We considered different goal probabilities ranging from 0.6 - 0.9 for the DRO model while the stochastic model had a constant goal probability of 0.9. The results are reported in Table 5, where the worst-case probability $\hat{\alpha}$ ($\hat{\alpha}_{SP}$ for stochastic model, and $\hat{\alpha}_{RO}$ for DRO model) represents the lowest probability across all chance constrained ORs with respect to their overtime thresholds and the expected overtime \hat{O} (\hat{O}_{SP} for stochastic model, and \hat{O}_{RO} for DRO model) represents the total expected overtime of all ORs.

From Table 5, we can make the following observations:

- (1) The stochastic solution demonstrates moderate infeasibility with respect to the chance constraint for two reasons. The first is due to the sample approximation. When the lognormal distribution is used for both the optimization and evaluation models, it resulted in 2% and 5% probability gap for instance Sets 2 and 3, respectively. Using a larger sample size for the stochastic model would reduce this gap while at the same time increasing the computational burden; The second reason is due to the ambiguous distribution of surgery durations. When the lognormal distribution is used for optimization and the uniform distribution assumed to be the true distribution used for evaluation, it resulted in 2% extra probability gap for instance Set 3. For the stochastic model, reducing this gap could only rely on accurately predicting the distribution, which is usually hard in practice.

Table 5 Performance evaluation of stochastic and robust solutions

Set No.	No. of Ins.	Sample Type	Stochastic Solution		Robust Solution							
			$\hat{\alpha}_{SP}$	\hat{O}_{SP}	Worst-case Probability ($\hat{\alpha}_{RO}$)				Expected Overtime (\hat{O}_{RO})			
			0.9	0.9	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
1	10	LOGN	0.97	0.13	0.99	0.99	0.99	0.99	0.14	0.14	0.14	0.14
		UNIF	0.99	0.01	1.00	1.00	1.00	1.00	0.01	0.01	0.01	0.01
		GAMM	0.97	0.08	0.99	0.99	0.99	0.99	0.09	0.09	0.09	0.09
2	10	LOGN	0.88	3.38	0.99	0.96	0.93	0.89	9.65	5.52	4.21	3.64
		UNIF	0.88	3.11	1.00	0.99	0.94	0.88	9.71	5.47	4.08	3.47
		GAMM	0.88	3.32	0.99	0.96	0.93	0.89	9.67	5.51	4.18	3.61
3	10	LOGN	0.85	12.15	0.99	0.96	0.93	0.89	22.81	17.84	14.68	12.41
		UNIF	0.83	12.20	1.00	0.98	0.94	0.88	22.82	17.81	14.68	12.46
		GAMM	0.84	12.18	0.99	0.96	0.93	0.89	22.83	17.85	14.72	12.46

- (2) The robust solution is much less sensitive to incorrect assumptions about the true distribution. However, the robust solution is conservative with respect to the chance constrained ORs, and resulted in overtime that, on average, is 1-2 times higher than that of the stochastic solution.
- (3) The DRO can generate solutions similar to the CCSBP by adjusting the goal probability of the DRO model. In our test instances, setting the goal probability for the DRO model to $1 - \alpha_{RO} = 0.6$ is a very good approximation for the stochastic model with $1 - \alpha_{SP} = 0.9$ for all of the instances. For this choice of α_{RO} , the robust solution has a probability guarantee close to the desired value of 0.9 and its expected overtime is only slightly greater than that of the stochastic counterpart.

7. Conclusion

This article addresses methods for the stochastic bin packing problem with chance constraints. We proposed a two-stage stochastic mixed integer programming formulation that is similar to other previously proposed formulations. We addressed the computational challenges for this problem by reformulating a column generation form of the model that can be solved using branch-and-price. The branch-and-price approach was further improved by reformulating the subproblem of column generation based on CVaR and probabilistic covers. A series of computational results show that both the column generation formulation and the subproblem reformulations considerably reduce computation times required to solve large-scale problem instances that arise in practice.

We also presented a distributionally robust optimization (DRO) counterpart under conditions of ambiguous item sizes for which only the mean and variance of the true distribution are known. We derived a closed form expression and proved properties about the expression that permit it to be expressed as a linear set of constraints. Our branch-and-price can also be applied to solve the DRO problem and the computational results show that this problem can be much easier to solve than the stochastic problem. Moreover, we show that a robust solution with similar overtime of the stochastic solution can be achieved by adjusting the goal probability of the DRO model.

Future work will study multistage chance constrained bin packing problems with item-to-bin allocation decisions made prior to the start of packing, and dynamic reallocation decisions throughout the horizon over which the random item size outcomes are observed. The problem formulation likely requires nonanticipative constraints that depend on the scenario and the dynamic state of the bin packing system. As a result, it will be necessary to study effective formulations that accurately express the feasibility of decisions over time and for which optimal or good approximate solutions can be obtained efficiently.

Acknowledgement. We thank Siqian Shen at the University of Michigan and Jiefang Shen at the Ruijin Hospital for helpful discussions about this paper. This research was funded in part by the National Science Foundation of China under grant 71131005 (Xie), and in part by the National Science Foundation under grant CMMI-0844511 (Denton). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Adler, Micah, Phillip B Gibbons, Yossi Matias. 2002. Scheduling space-sharing for internet advertising. *Journal of Scheduling* **5**(2) 103–119.
- Barnhart, Cynthia, Ellis L Johnson, George L Nemhauser, Martin WP Savelsbergh, Pamela H Vance. 1998. Branch-and-price: Column generation for solving huge integer programs. *Operations research* **46**(3) 316–329.
- Batun, Sakine, Brian T Denton, Todd R Huschka, Andrew J Schaefer. 2011. Operating room pooling and parallel surgery processing under uncertainty. *INFORMS journal on Computing* **23**(2) 220–237.
- Benders, Jacques F. 1962. Partitioning procedures for solving mixed-variables programming problems. *Numerische mathematik* **4**(1) 238–252.
- Birge, John R, Francois Louveaux. 2011. *Introduction to stochastic programming*. Springer Science & Business Media.

-
- Codato, Gianni, Matteo Fischetti. 2006. Combinatorial benders' cuts for mixed-integer linear programming. *Operations Research* **54**(4) 756–766.
- Coffman, Edward G, Jr, Michael R Garey, David S Johnson. 1978. An application of bin-packing to multi-processor scheduling. *SIAM Journal on Computing* **7**(1) 1–17.
- Degraeve, Zeger, Raf Jans. 2007. A new dantzig-wolfe reformulation and branch-and-price algorithm for the capacitated lot-sizing problem with setup times. *Operations Research* **55**(5) 909–920.
- Delage, Erick, Yinyu Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research* **58**(3) 595–612.
- Dell'Olmo, Paolo, Hans Kellerer, Maria Grazia Speranza, Zsolt Tuza. 1998. A 13/12 approximation algorithm for bin packing with extendable bins. *Information Processing Letters* **65**(5) 229–233.
- Deng, Yan, Siqian Shen, Brian Denton. 2014. Chance-constrained surgery planning under uncertain or ambiguous surgery duration. *Available at SSRN* <http://ssrn.com/abstract=2432375> .
- Denton, Brian T, Andrew J Miller, Hari J Balasubramanian, Todd R Huschka. 2010. Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations research* **58**(4-part-1) 802–816.
- Jiang, Ruiwei, Yongpei Guan. 2013. Data-driven chance constrained stochastic program. *Available at Optimization Online* http://www.optimizationonline.org/DB_FILE/2013/09/4044.pdf .
- Kong, Qingxia, Chung-Yee Lee, Chung-Piaw Teo, Zhichao Zheng. 2013. Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research* **61**(3) 711–726.
- Lamiri, Mehdi, Xiaolan Xie, Shuguang Zhang. 2008. Column generation approach to operating theater planning with elective and emergency patients. *IIE Transactions* **40**(9) 838–852.
- Liu, Xiao, Simge Küçükyavuz, James Luedtke. 2014. Decomposition algorithms for two-stage chance-constrained programs. *Mathematical Programming* 1–25.
- Luedtke, James. 2014. A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Mathematical Programming* **146**(1-2) 219–244.
- Luedtke, James, Shabbir Ahmed. 2008. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization* **19**(2) 674–699.
- Lulli, Guglielmo, Suvrajeet Sen. 2004. A branch-and-price algorithm for multistage stochastic integer programming with application to stochastic batch-sizing problems. *Management Science* **50**(6) 786–796.
- Mak, Ho-Yin, Ying Rong, Jiawei Zhang. 2015. Appointment scheduling with limited distributional information. *Management Science* **61**(2) 316–334.
- Nursing Solutions, Inc. 2013. 2013 national healthcare and RN retention report. *Available at* <http://www.nsinursingsolutions.com/Files/assets/library/retention-institute/NationalHealthcareRNRetentionReport2013.pdf> .

-
- Rockafellar, R Tyrrell, Stanislav Uryasev. 2000. Optimization of conditional value-at-risk. *Journal of risk* **2** 21–42.
- Scarf, Herbert, KJ Arrow, S Karlin. 1958. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production* **10** 201–209.
- Shylo, Oleg V, Oleg A Prokopyev, Andrew J Schaefer. 2012. Stochastic operating room scheduling for high-volume specialties under block booking. *INFORMS Journal on Computing* **25**(4) 682–692.
- Silva, Eduardo F, R Kevin Wood. 2006. Solving a class of stochastic mixed-integer programs with branch and price. *Mathematical programming* **108**(2-3) 395–418.
- Singh, Kavinesh J, Andy B Philpott, R Kevin Wood. 2009. Dantzig-wolfe decomposition for solving multi-stage stochastic capacity-planning problems. *Operations Research* **57**(5) 1271–1286.
- Song, Yongjia, James R Luedtke, Simge Küçükyavuz. 2014. Chance-constrained binary packing problems. *INFORMS Journal on Computing* **26**(4) 735–747.
- Strum, David P, Jerrold H May, Allan R Sampson, Luis G Vargas, William E Spangler. 2003. Estimating times of surgeries with two component procedures: comparison of the lognormal and normal models. *Anesthesiology* **98**(1) 232–240.
- Vanderbeck, François. 2000. On dantzig-wolfe decomposition in integer programming and ways to perform branching in a branch-and-price algorithm. *Operations Research* **48**(1) 111–128.
- Vanderbeck, François, Laurence A Wolsey. 1996. An exact algorithm for ip column generation. *Operations research letters* **19**(4) 151–159.
- Wang, Yu, Jiafu Tang, Richard YK Fung. 2014. A column-generation-based heuristic algorithm for solving operating theater planning problem under stochastic demand and surgery cancellation risk. *International Journal of Production Economics* **158** 28–36.

Appendix

A. Proofs

Proof of Theorem 1

Proof: From the definition of $VaR(z)$, $\mathbb{P}(z \leq T) \geq 1 - \alpha$ is equivalent to $VaR(z) \leq T$. Thus, the chance constraints (7c)-(7d) can be reformulated as follows:

$$VaR\left(\sum_{i \in \mathcal{I}} y_{ip} d_i(\omega)\right) \leq T + \delta_k, \text{ if } c_{kp} = 1. \quad (15)$$

Because $VaR(z) \leq CVaR(z)$, a conservative expression of (15) is:

$$\inf_{\beta} \left(\beta + \frac{1}{\alpha} \mathbb{E} \left[\left(\sum_{i \in \mathcal{I}} y_{ip} d_i(\omega) - \beta \right)^+ \right] \leq T + \delta_k, \text{ if } c_{kp} = 1. \quad (16)$$

Since $\sum_{i \in \mathcal{I}} y_{ip} d_i(\omega) \geq 0$ and $0 < \alpha < 1$, the optimal β should be $\beta^* \geq 0$. Thus, constraint (16) is equivalent to the following constraints:

$\exists \beta \geq 0$ such that

$$\begin{cases} \beta + \frac{1}{\alpha} \mathbb{E} [\gamma(\omega)] \leq T + \delta_k, \text{ if } c_{kp} = 1, \\ \gamma(\omega) = \left(\sum_{i \in \mathcal{I}} y_{ip} d_i(\omega) - \beta \right)^+, \forall \omega. \end{cases} \quad (17)$$

Constraint is (8) a linearization of (17) and hence it is a restriction of (7c)-(7d). \square

Proof of Theorem 2

Proof: Inequality (9a) ensures that if all items in \mathcal{P}_k are selected to a chance constrained packing, it cannot select any other item from $\mathcal{Q}(\mathcal{P}_k)$. Inequality (9b) ensures that a chance constrained packing cannot select all items in \mathcal{C}_k .

From the definition of probabilistic pack, a feasible packing to chance constraints (7c)-(7d) is a probabilistic pack. So it automatically satisfies (9a). It also satisfies (9b) because a probabilistic pack cannot be a probabilistic cover. Thus, any feasible solution to (7c)-(7d) is also feasible to (9). From the definition of probabilistic cover, any packing violating chance constraints (7c)-(7d) can be regarded as a probabilistic cover. Thus, if all possible sets of \mathcal{C}_k are considered, any feasible solution to (9) also becomes feasible to (7c)-(7d). So constraint (9) with all possible sets of $\mathcal{C}_k, \mathcal{P}_k$ provides an exact relaxation of (7c)-(7d). \square

Proof of Lemma 1

Proof: Let $\hat{\mu}, \hat{\sigma}$ denote the mean and standard deviation of Y , respectively. The case $\hat{\sigma} = 0$ can be matched with the one-point distribution with $\hat{\mathbb{D}} = \hat{\mu}$. Assuming now $\hat{\sigma} > 0$, it follows that $\hat{\mu} > a$. The two-point distribution $\hat{\mathbb{D}} = \{(1-p, y_1), (p, y_2)\}$ satisfying the mean and variance constraints is as follows:

$$y_1 = \hat{\mu} - k\hat{\sigma}, \quad y_2 = \hat{\mu} + \hat{\sigma}/k, \quad k \triangleq \sqrt{\frac{p}{1-p}}. \quad (18)$$

When p increases, y_1 decreases from $\hat{\mu}$ to $-\infty$ and y_2 decreases from $+\infty$ to $\hat{\mu}$. Since $\hat{\mu} > a$, there exists $k_1 > 0$ such that

$$y_1 = a, \quad y_2 = \frac{\hat{\mu}^2 + \hat{\sigma}^2 - \hat{\mu}a}{\hat{\mu} - a}, \quad k_1 = \frac{\hat{\mu} - a}{\hat{\sigma}}. \quad (19)$$

Since $\hat{\sigma} > 0$, $y_2 \geq \hat{\mu} > a$. Consider now $Y \in [a, b)$, we have $(b - Y)(Y - a) \geq 0$. Since $\mathbb{P}(Y > a) > 0$, $\mathbb{E}[(b - Y)(Y - a)] > 0$ which implies $b\hat{\mu} + a\hat{\mu} - ab - (\hat{\mu}^2 + \hat{\sigma}^2) > 0$ leading to

$$b > \frac{\hat{\mu}^2 + \hat{\sigma}^2 - \hat{\mu}a}{\hat{\mu} - a} = y_2. \quad (20)$$

Therefore, $y_1, y_2 \in [a, b)$. □

Proof of Theorem 3

Proof: Rewrite the distribution \mathbb{D} as follows:

$$\mathbb{D} = DD \left\{ \begin{array}{l} \hat{p}_1, X_1 \\ \hat{p}_2, X_2 \end{array} \right\}$$

where $DD(\cdot)$ represents a discrete distribution and $X_1 \sim \mathbb{D}_1 = \{X \sim \mathbb{D} \mid X \in [0, B)\}$ and $X_2 \sim \mathbb{D}_2 = \{X \sim \mathbb{D} \mid X \in [B, \infty)\}$, and $\mathbb{D}_1, \mathbb{D}_2$ are conditional distributions.

From Lemma 1, X_1 and X_2 can be matched by discrete distributions $\hat{\mathbb{D}}_1$ with points in $[0, B)$ and $\hat{\mathbb{D}}_2$ with points in $[B, \infty)$, respectively. Without loss of generality, we assume $\hat{\mathbb{D}}_1 = \{(\hat{p}_{11}, x_1), (\hat{p}_{12}, x_2)\}$ and $\hat{\mathbb{D}}_2 = \{(\hat{p}_{21}, x_3), (\hat{p}_{22}, x_4)\}$. Let

$$\mathbb{D}' = \{(\hat{p}_1 \hat{p}_{11}, x_1), (\hat{p}_1 \hat{p}_{12}, x_2), (\hat{p}_2 \hat{p}_{21}, x_3), (\hat{p}_2 \hat{p}_{22}, x_4)\},$$

we have $\mu(\mathbb{D}') = \mu$ and $\sigma(\mathbb{D}') = \sigma$. Moreover, $x_1, x_2 < B$ and $x_3, x_4 \geq B$, therefore $\mathbb{P}_{X \sim \mathbb{D}'}(X \geq B) = \mathbb{P}_{X \sim \mathbb{D}}(X \geq B) = \hat{p}_2$. $N \leq 4$ because $\hat{\mathbb{D}}_1$ and $\hat{\mathbb{D}}_2$ are at most two-point distributions. □

Proof of Lemma 2

Proof: The proof is trivial because we have a \mathbb{D}' with all points in $[B, \infty)$ if $\mu > B$. Therefore

$$\mathbb{P}_{X \sim \mathbb{D}'}(X \geq B) = 1. \quad \square$$

Proof of Theorem 4

Proof: The case with $\mu > B$ is true from Lemma 2. Now consider the case with $\mu \leq B$ and assume there exists some X such that $X \geq B$ to avoid the trivial case. The worst-case chance probability WC can be formulated as the following LP model:

$$WC = \max \sum_{j=v}^N p_j \quad (21a)$$

$$s.t. \quad \sum_{j=1}^N p_j = 1 \quad (21b)$$

$$\sum_{j=1}^N p_j x_j = \mu \quad (21c)$$

$$\sum_{j=1}^N p_j x_j^2 = \mu^2 + \sigma^2 \quad (21d)$$

$$p_j \geq 0, \forall j \quad (21e)$$

where $v \triangleq \min\{i : x_i \geq B\}$. Therefore there exists a worst-case distribution on $\{x_1, \dots, x_N\}$ that is a basic solution of the above LP and hence has at most three positive probabilities, one for each constraint (21b)-(21d). \square

Proof of Corollary 1

Proof: First, $WC = 1$ for the case with $\mu > B$, thus it is continuous in μ, σ and B . Now, we consider the case with $\mu \leq B$. From Theorem (5), $WC = \min\{1, f\}$ where

$$f \triangleq \begin{cases} \mu/B, & \text{if } \mu \leq B, \mu^2 + \sigma^2 > \mu B, \\ \frac{\sigma^2}{\sigma^2 + (B-\mu)^2}, & \text{if } \mu \leq B, \mu^2 + \sigma^2 \leq \mu B. \end{cases} \quad (22)$$

Rewrite f^{-1} as follows:

$$f^{-1} \triangleq \begin{cases} g_1, & \text{if } \frac{\sigma}{\mu} \leq \frac{B-\mu}{\sigma} \\ g_2, & \text{if } \frac{\sigma}{\mu} > \frac{B-\mu}{\sigma} \end{cases} \quad (23)$$

where $g_1 \triangleq 1 + \frac{(B-\mu)^2}{\sigma^2}$, $g_2 \triangleq \frac{B}{\mu}$.

Since $\mu \leq B$, $g_1 \leq g_2 \Leftrightarrow \frac{B-\mu}{\sigma} \leq \frac{\sigma}{\mu}$ and $f^{-1} = \max(g_1, g_2) = \max\left(1 + \frac{(B-\mu)^2}{\sigma^2}, \frac{B}{\mu}\right)$. As a result,

$$WC = \min \left(\left(1 + \left(\frac{\max(B-\mu, 0)}{\sigma} \right)^2 \right)^{-1}, \frac{\mu}{B} \right) \quad (24)$$

which leads to the decreasing monotonicity of WC in B and the increasing monotonicity of WC in μ and σ .

$WC = \widehat{WC}$ because: (i) $WC \geq \widehat{WC}$ by definition (ii) $\widehat{WC}(B) \geq \lim_{h \rightarrow 0^+} WC(B+h) = WC(B)$. \square

Proof of Theorem 5

Proof: The case with $\mu > B$ is true from Lemma 2. We prove the cases with $\mu \leq B$. We begin with a two-point distribution, and then consider the three-point distribution. The two-point distribution $DD^2 = \{(1-p, x_1), (p, x_2)\}$ ($x_1 < x_2$) satisfying the mean μ and variance σ^2 is as follows:

$$x_1 = \mu - k\sigma, \quad x_2 = \mu + \sigma/k, \quad k \triangleq \sqrt{\frac{p}{1-p}}. \quad (25)$$

When p increases, k increases, and x_1 and x_2 decrease. The worst-case distribution is reached by increasing p until either $x_1 = 0$ ($k = \mu/\sigma$) or $x_2 = B$ ($k = \sigma/(B - \mu)$). The worst-case distribution of the chance probability under the two-point distribution is as follows:

$$x_1 = 0, x_2 = \frac{\mu^2 + \sigma^2}{\mu}, p = \frac{\mu^2}{\mu^2 + \sigma^2}, WC = p, \quad \text{if } \mu^2 + \sigma^2 > \mu B, \quad (26a)$$

$$x_1 = \frac{\mu B - \mu^2 - \sigma^2}{B - \mu}, x_2 = B, p = \frac{\sigma^2}{\sigma^2 + (B - \mu)^2}, WC = p, \quad \text{if } \mu^2 + \sigma^2 \leq \mu B. \quad (26b)$$

Now, we consider the three-point distribution, $DD^3 = \{(p_1, x_1), (p_2, x_2), (p_3, x_3)\}$ ($x_1 < x_2 < x_3$), satisfying the mean μ and variance σ^2 , which is as follows:

$$\begin{cases} p_1 = \frac{x_2 x_3 + \mu^2 + \sigma^2 - \mu(x_2 + x_3)}{(x_3 - x_1)(x_2 - x_1)}, \\ p_2 = \frac{-x_1 x_3 - \mu^2 - \sigma^2 + \mu(x_1 + x_3)}{(x_2 - x_1)(x_3 - x_2)}, \\ p_3 = \frac{x_1 x_2 + \mu^2 + \sigma^2 - \mu(x_1 + x_2)}{(x_3 - x_1)(x_3 - x_2)}. \end{cases} \quad (27)$$

If $x_2 < B \leq x_3$, then $WC = p_3$. Given that $x_1 < x_2 < x_3$ and $p_2 \geq 0$, we have $\partial p_1 / \partial x_2 \geq 0$, and $\partial p_3 / \partial x_2 \leq 0$. Thus, decreasing x_2 to x_1 monotonically increases WC , which makes either $p_1 = 0$ or $x_2 = x_1$, leading to a two-point distribution as shown by (26).

If $B \leq x_2 < x_3$, then $WC = p_2 + p_3$. We have $\partial WC / \partial x_2 = -\partial p_1 / \partial x_2 \leq 0$. Decreasing x_2 leads to either (1) one of the probabilities, p_1, p_2, p_3 , becomes zero or (2) $x_2^* = B$. As case (1) is dominated by the two-point distribution (26), we only need to consider the three-point distribution with $x_2^* = B$. Since $\partial WC / \partial x_1 \leq 0$, we can reduce x_1 until one of the following constraints holds at equality:

$$-x_1 x_3 - \mu^2 - \sigma^2 + \mu(x_1 + x_3) \geq 0 \quad (28a)$$

$$x_1 x_2^* + \mu^2 + \sigma^2 - \mu(x_1 + x_2^*) \geq 0 \quad (28b)$$

$$x_1 \geq 0 \quad (28c)$$

Constraint (28a) can always hold by increasing x_3 . From (28b)-(28c) and $x_2^* = B$, we have $x_1^* = 0$ if $\mu^2 + \sigma^2 > \mu B$, and $x_1^* = \frac{\mu B - \mu^2 - \sigma^2}{B - \mu}$ if $\mu^2 + \sigma^2 \leq \mu B$.

If $x_1^* = 0$, $WC = -\frac{\mu^2 + \sigma^2 - B\mu}{Bx_3} + \frac{\mu}{B}$. As a result, $WC = \mu/B$ which is reached by limiting distribution $\lim_{x_3 \rightarrow \infty} \widehat{DD^3} = \{(p_1, 0), (p_2, B), (p_3, x_3)\}$. It dominates the worst-case two-point distribution as $\frac{\mu}{B} > \frac{\mu^2}{\mu^2 + \sigma^2}$ if $\mu^2 + \sigma^2 > \mu B$.

If $x_1^* = \frac{\mu B - \mu^2 - \sigma^2}{B - \mu}$, $p_3 = 0$, leading to a two-point distribution as shown by (26b). Since we considered all possibilities in DD^2 and DD^3 , this completes the proof. \square

Proof of Theorem 6

Proof: First, (14b) is from the definition of bin extension, and (14c) from the allocation constraint. Constraints (14f)-(14h) are equivalent to $z_{ijr} = x_{ir} \cdot x_{jr}$. Constraint (14d) is from the nominal chance constraint (13b) because $(\sum_{i \in \mathcal{I}} \mu_i x_{ir})^2 = \sum_{i \in \mathcal{I}} \mu_i^2 x_{ir} + 2 \sum_{i,j \in \mathcal{I}, i < j} z_{ijr} \mu_i \mu_j$ and $\sigma^2 = \sum_{i \in \mathcal{I}} \sigma_i^2 x_{ir}$. Similarly, constraint (14e) is from the nominal chance constraint (13c). \square

B. Formulations

Column Generation Formulation of the DRO model (14)

The master problem of the DRO model is the same with (6).

The subproblem (DRO-SP) of the DRO model is formulated as the following MIP:

$$(\text{DRO-SP}): \quad z_{SP} = \max \sum_{i=1}^n \pi_i y_{ip} + \sum_{k=1}^K \phi_k c_{kp} - \bar{o}_p \quad (29a)$$

$$s.t. \quad \sum_{i \in \mathcal{I}} y_{ip} \mu_i(\omega) - \bar{o}_p \leq T \quad (29b)$$

$$\sum_{i \in \mathcal{I}} y_{ip} \Gamma_i + 2 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}, j > i} z_{ij} \mu_i \mu_j + c_{kp} M_{kp}^3 \leq M_{kp}^3, \quad \forall k \quad (29c)$$

$$\sum_{i \in \mathcal{I}} y_{ip} \Lambda_{ik} - 2 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}, j > i} z_{ij} \mu_i \mu_j + c_{kp} M_{kp}^4 \leq (T + \delta_k)^2 + M_{kp}^4, \quad \forall k \quad (29d)$$

$$z_{ij} \geq y_{ip} + y_{jp} - 1, \quad \forall i, \forall j \quad (29e)$$

$$z_{ij} \leq y_{ip}, \quad z_{ij} \leq y_{jp}, \quad \forall i, \forall j \quad (29f)$$

$$y_{ip}, c_{kp} \in \{0, 1\}, z_{ij}, \bar{o}_p \geq 0, \quad \forall i, \forall k \quad (29g)$$

where z_{ij} are auxiliary variables representing $z_{ij} = y_i \cdot y_j$, M_{kp}^3 and M_{kp}^4 are suitably large constants where $M_k^1(\omega)$ is bounded by (5b) and $M_k^2 \leq (1 - \alpha) \cdot |\Omega|$.

The objective (29a) is to maximize the reduced cost with respect to the dual solutions of the current (R²MP). Constraint (29b) determines the extension in the new packing. Constraints (29c)-(29f) jointly ensure the robust chance constraint on total item size in each packing. Finally, constraint (29g) restricts decision variables.

MISOCP model with the worst-case expected bin extension

The DRO model (14) with considering the worst-case expected bin extension can be expressed as follows:

$$\min_{x \in \hat{\mathcal{X}}} \sum_{r \in \mathcal{R}} \left[\sup_{\Sigma} \int_{\Sigma} o_r(\omega) d\mathbb{D}_r \right] \quad (30a)$$

$$s.t. \int_{\Sigma} d_i(\omega) d\mathbb{D}_r = \mu_i, \forall i, \forall r \quad (30b)$$

$$\int_{\Sigma} d_i^2(\omega) d\mathbb{D}_r = \mu_i^2 + \sigma_i^2, \forall i, \forall r \quad (30c)$$

$$\int_{\Sigma} d\mathbb{D}_r = 1, \forall i, \forall r, \quad (30d)$$

where $\hat{\mathcal{X}}$ is defined by (14c)-(14h). Formulation (30a) defines the total worst-case expected bin extension. We can use different distributions \mathbb{D}_r of item sizes for different bins because, once the allocation is fixed, the total item size in each bin become independent. Constraints (30b)-(30c) describe the mean and deviation of $d_i(\omega)$, respectively.

THEOREM 7. *The model (30) can be reformulated as the following MISOCP model:*

$$(\widehat{\text{ROP}}): \min_{x \in \hat{\mathcal{X}}} \sum_{r \in \mathcal{R}} \sum_{i \in \mathcal{I}} \mu_i p_{ir} + \sigma_i h_{ir} + \sum_{r \in \mathcal{R}} q_r \quad (31a)$$

$$s.t. \left\| \begin{matrix} p_{ir} \\ h_{ir} - s_{ir} \end{matrix} \right\| \leq h_{ir} + s_{ir}, \forall i, \forall r \quad (31b)$$

$$\left\| \begin{matrix} p_{ir} - x_{ir} \\ h_{ir} - \hat{s}_{ir} \end{matrix} \right\| \leq h_{ir} + \hat{s}_{ir}, \forall i, \forall r \quad (31c)$$

$$\sum_{i=1}^n s_{ir} - q_r \leq 0, \forall i, \forall r \quad (31d)$$

$$\sum_{i=1}^n \hat{s}_{ir} - q_r \leq T, \forall i, \forall r \quad (31e)$$

$$h_{ir}, s_{ir}, \hat{s}_{ir} \geq 0, p_{ir}, q_r \text{ free } \forall i, \forall r. \quad (31f)$$

Proof: Taking the dual of the supremum in (30), we have the following formulation:

$$\min_{x \in \mathcal{X}} \sum_{r \in \mathcal{R}} \left[\inf_{i \in \mathcal{I}} \sum \mu_i p_{ir} + \sigma_i h_{ir} + q_r \right] \quad (32a)$$

$$s.t. \sum_{i \in \mathcal{I}} h_{ir} d_i^2(\omega) + p_{ir} d_i(\omega) + q_r \geq o_r(\omega), \forall r, \forall \omega \in \Sigma. \quad (32b)$$

We rewrite (32b) as follows:

$$\sum_{i \in \mathcal{I}} h_{ir} d_i^2(\omega) + p_{ir} d_i(\omega) + q_r \geq 0, \forall r, \forall \omega \in \Sigma, \quad (33a)$$

$$\sum_{i \in \mathcal{I}} h_{ir} d_i^2(\omega) + p_{ir} d_i(\omega) + q_r \geq \sum_{i \in \mathcal{I}} x_{ir} d_i(\omega) - T, \forall r, \forall \omega \in \Sigma. \quad (33b)$$

Considering (33a), it can be reformulated as:

$$\begin{cases} h_{ir} \geq 0, h_{ir} = 0 \rightarrow p_{ir} = 0, \forall i, \forall r, \\ \sum_{\substack{i \in \mathcal{I} \\ h_{ir} \neq 0}} h_{ir} \left(d_i(\omega) - \frac{p_{ir}}{2h_{ir}} \right)^2 + q_r - \sum_{i \in \mathcal{I}} \frac{p_{ir}^2}{4h_{ir}} \geq 0, \forall r, \forall \omega \in \Sigma. \end{cases} \quad (34)$$

$\Leftrightarrow \exists s_{ir} \geq 0$ such that

$$\begin{cases} p_{ir}^2 \leq 4h_{ir}s_{ir}, \forall i, \forall r, \\ \sum_{i \in \mathcal{I}} s_{ir} - q_r \leq 0, \forall i, \forall r, \\ h_{ir} \geq 0, \forall i, \forall r. \end{cases} \quad (35)$$

Similarly, constraint (33b) can be reformulated as:

$\Leftrightarrow \exists \hat{s}_{ir} \geq 0$ such that

$$\begin{cases} (p_{ir} - x_{ir})^2 \leq 4h_{ir}\hat{s}_{ir}, \forall i, \forall r, \\ \sum_{i \in \mathcal{I}} \hat{s}_{ir} - q_r \leq 0, \forall i, \forall r, \\ h_{ir} \geq 0, \forall i, \forall r. \end{cases} \quad (36)$$

By substituting (35) with (31b),(31d), and (36) with (31c),(31e), we complete the proof. \square