SIAM J. OPTIM.
Vol. 35, No. 2, pp. 739–766

A NONPARAMETRIC ROBUST OPTIMIZATION APPROACH FOR CHANCE-CONSTRAINED KNAPSACK PROBLEM*

LIANG XU†, CHAO ZHANG‡, ZHOU XU§, AND DANIEL ZHUOYU LONG¶

Abstract. A chance-constrained knapsack problem (CCKP) is a knapsack problem restricted by a chance constraint, which ensures that the total capacity constraint under uncertain volume can be violated only up to a given probability threshold. CCKP is challenging to solve due to its combinatorial nature and the involvement of its chance constraint. Existing solution methods for CCKP with tractability guarantees mainly focus on two approaches: (1) a full-information approach (stochastic programming) that assumes the uncertain volume follows certain distributions, such as normal or empirical distribution; (2) a partial-information approach (robust optimization) that adopts specific statistics of the unknown distribution, such as the mean and variance. The existing full-information approach lacks robustness under limited samples due to its strong assumption; the existing partial-information approach can be further improved, as the uncertainty set or distributional ambiguity set can be ameliorated. With these concerns in mind, we propose a nonparametric robust approach for CCKP by involving a novel nonparametric statistic to form a new distributional ambiguity set. Furthermore, we develop an upper bound on the violation probability of the chance constraint under the distributional ambiguity set to approximate CCKP by a deterministic robust counterpart. In terms of solution methodology, we decompose the deterministic robust counterpart into cardinality-constrained knapsack problems, which can be efficiently solved by the proposed dynamic programming algorithm. Computational results show that our proposed solution methods produce better solutions to CCKP compared with existing approaches.

Key words. robust optimization, knapsack problem, chance constraint, nonparametric

MSC codes. 90C15, 90C27, 90C59

DOI. 10.1137/23M1620867

1. Introduction.

The knapsack problem is a fundamental and widely studied combinatorial optimization problem [45]. It can be used to model a wide range of industrial problems, including facility location, machine scheduling, and resource allocation problems [17, 24]. Most importantly, the knapsack problem often serves as a subproblem in large-scale discrete optimization problems. For instance, the assignment problem and the capacitated P-median problem involve the knapsack problem as a subproblem when implementing column generation algorithms [44, 36, 19]. With this, the knapsack problem has several variants, including the binary knapsack problem, the cardinality-constrained knapsack problem, and the quadratic knapsack problem. The binary knapsack problem is a classic and fundamental variant, known to be NP-hard, yet solvable in pseudopolynomial time via dynamic programming algorithms. Many efficient algorithms have been extensively studied in the literature

*Received by the editors November 30, 2023; accepted for publication (in revised form) December 23, 2024; published electronically April 15, 2025.
https://doi.org/10.1137/23M1620867

Funding: The work is supported by the National Natural Science Foundation of China (NSFC) (grants 72401121, 71971177, and 72342012).

†School of Business Administration, Southwestern University of Finance and Economics, Chengdu, China; Big Data Laboratory on Financial Security and Behavior, SWUEF (Laboratory of Philosophy and Social Sciences, Ministry of Education), Chengdu, China (arecxuliang1112@swufe.edu.cn).

‡Corresponding author. School of Medical Informatics, Chongqing Medical University, Chongqing, China (103396@cqmu.edu.cn).

§Department of Logistics Management and Maritime Studies, Hong Kong Polytechnic University, Hong Kong, China (zhou.xu@polyu.edu.hk).

¶Department of System Engineering and Engineering Management, Chinese University of Hong Kong, Hong Kong, China (zylong@se.cuhk.edu.hk).

to solve other variants [20, 41, 12, 1, 39, 34]. For a detailed review of knapsack problems, the reader may refer to [37] and [29]. In this study, we focus on the binary knapsack problem.

The knapsack problem often has uncertain volumes in practical applications [30, 38], motivating decision-makers to apply a chance constraint that restricts the probability of constraint violation below an allowed risk tolerance [23, 32, 48, 27]. In this article, we investigate the chance-constrained knapsack problem (CCKP) defined as

$$\text{(CCKP)} \max_{x\in\{0,1\}^n} \left\{ \sum_{i=1}^n c_i x_i : \text{Prob}\left(\sum_{i=1}^n \tilde{a}_i x_i > b \right) \le \alpha \right\}, \tag{1.1}$$

where c_i is the deterministic objective coefficient, and $\tilde{a}_i$ is the uncertain volume. The probability measure $\text{Prob}(\cdot)$ represents the probability of an event, and the parameter α describes the risk tolerance of violation probability that often takes a small value (e.g., 0.05). As such, the probability of constraint violation $\text{Prob}(\sum_{i=1}^n \tilde{a}_i x_i > b)$ is required to be less than the predetermined α. The CCKP is inherently challenging to solve, primarily due to the fact that the chance constraint is typically nonconvex and the decision variables are discrete. Nonetheless, two prominent approaches have been identified to recapture the tractability of chance-constrained models: stochastic optimization and (distributionally) robust optimization.

Stochastic optimization typically requires full information about the probability distribution. One research stream assumes knowing the explicit form of the probability distribution [26, 30]. Although this approach is implausible, it may offer computational tractability, which is appealing to decision-makers. For example, when random variables follow a joint normal distribution [24], the CCKP can be reformulated as a deterministic mixed integer second-order conic program, enabling the use of off-the-shelf optimization solvers. However, in many cases, only historical data is available, and the probability distribution is approximated by its empirical form [46]. Under these circumstances, the probability of constraint violation can be simulated using the Monte Carlo method [31] or approximated by conservative measures such as conditional value-at-risk (CVaR) [42] and entropic value-at-risk (EVaR) [2]. The optimal solution of chance-constrained models can be sensitive to the constructed empirical distribution, resulting in poor out-of-sample performances. As demonstrated by [49], the solution obtained under a limited sample size can deviate significantly from true reality, even for the relatively simple newsvendor problem with only one uncertain coefficient. Therefore, the stochastic optimization may lack robustness when the sample size is limited.

Robust optimization basically relies on partial information about the probability distribution, such as the support for a random variable. It employs a feasible set of possible realizations, referred to as the uncertainty set, based on the adopted partial information. Notable pioneering works include the ellipsoidal uncertainty set [5, 3] and the budget uncertainty set [9]. Computationally tractable reformulations are proposed to solve their robust counterparts [8, 7]. [33] proposes an uncertainty set based on a finite number of scenarios to optimize the worst-case performance. [15] introduces an uncertainty set using new deviation measures for bounded random variables, which can capture distributional asymmetry to improve the objective function. [40] incorporates asymmetric distribution information into robust optimization. In these robust optimization approaches, constructing uncertainty sets focuses on parametric information about the probability distribution. To the best of our knowledge, only the robust counterparts of ellipsoidal and budget uncertainty sets are computationally tractable for CCKPs.

Distributionally robust optimization further involves a plausible set of probability distributions, known as the ambiguity set. Common ambiguity sets include probability distributions restricted by moment information, such as the first and second moments, and semivariance [21, 50, 51]. Focusing on research that addresses CCKPs, we note that [16] proposes distributionally robust CVaR and [35] proposes distributionally robust EVaR. [50] reformulates the distributionally chance-constrained binary program as 0-1 second-order cone (SOC) programs, exploiting the submodularity of the 0-1 SOC constraints under special and general covariance matrices. By utilizing submodularity and lifting, they derive extended polymatroid inequalities to strengthen the 0-1 SOC formulations. It is noted that (distributionally) robust optimization yields conservative solutions that offer opportunities for improvement.

To address the challenges of robustness and conservativeness, this article proposes a nonparametric robust approach. Previous studies have demonstrated the effectiveness of incorporating nonparametric information, such as monotonicity and convexity, in the newsvendor problem [49]. It is essential to identify valid nonparametric information for the knapsack problem. To this end, we propose a novel supportwise statistic, the maximum average tail probability distribution (MATPD), which exploits the nonparametric information of underlying distributions. The proposed MATPD induces an ambiguity set of candidate distributions. Based on this ambiguity set, we then present a distributionally robust CCKP (DR-CCKP) formulation, which explicitly derives an upper bound on the probability of constraint violation and yields an approximation for the CCKP. Furthermore, we propose tractable reformulations via subproblems of cardinality-constrained knapsack problems. In addition to utilizing off-the-shelf optimization solvers, we provide an exact solution technique based on dynamic programming for these subproblems. The numerical experiment demonstrates that our nonparametric robust approach outperforms distributionally robust optimization in terms of conservativeness and stochastic optimization in terms of robustness.

Our major contributions are summarized as follows.

1. We propose a novel supportwise MATPD to construct the ambiguity set for the DR-CCKP. The proposed MATPD leverages the nonparametric information of distributions.
2. We derive an explicit upper bound on the probability of constraint violation in the DR-CCKP. Specifically, the derived upper bound is tight when uncertain volumes follow uniform distributions.
3. We equivalently reformulate the DR-CCKP into a set of deterministic cardinality-constrained knapsack problems, which can be efficiently solved using off-the-shelf solvers or dynamic programming techniques.

Organization. The rest of this article is organized as follows. Section 2 presents our new distributionally robust optimization approaches. Section 3 introduces the explicit upper bound for the probability of constraint violation. Section 4 provides solution methods. Section 5 applies our robust optimization approaches in a data-driven setting and extends our robust optimization approach for the multidimentional CCKP. Section 6 conducts numerical experiments comparing our framework with existing approaches. Section 7 provides the conclusions and potential research directions.

Notation. We denote random variables with the tilde sign (e.g., $\tilde{a}_i$). Lowercase letters represent scalars, such as x_i. Uppercase letters represent sets, such as N.

2. Distributionally robust optimization approach for the CCKP. Let $N := \{1, 2, \ldots, n\}$ denote the set of subscripts for the uncertain volume $\tilde{a}_i$. We assume that these coefficients are independent and distributed within bounded intervals. Despite its ostensible deviation from many practical situations, the independence assumption is common in myriad studies engaged in chance constraints [35, 32, 10]. For each uncertain volume $\tilde{a}_i$, we define its nominal value as $\bar{a}_i$ and its deviation from the nominal value as $\hat{a}_i$. Specifically, when the support set of $\tilde{a}_i$ is bounded by the interval $[s_i, t_i]$, the nominal value has $\bar{a}_i = (t_i + s_i)/2$ and the deviation has $\hat{a}_i = (t_i - s_i)/2$. Many have scaled the uncertain volume by $(\tilde{a}_i - \bar{a}_i)/\hat{a}_i$ [4, 8]. Crucially, this representation can derive an explicitly formed but computationally intractable formulation in our approach. For that reason, we introduce the scaled uncertain volume $\tilde{z}_i = (\tilde{a}_i - \bar{a}_i)/\hat{a}$, where $\hat{a}$ denotes the maximum value of deviations such that $\hat{a} = \max\{\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n\}$. With this tailored representation, our numerical studies indicate that the robust optimization approach based on $\hat{a}$ still has superior performance to many existing approaches on robustness and conservativeness. As such, we can represent $\tilde{z}_i$ as an independent random variable distributed within the interval $[-1, 1]$. We consider continuous random variables $\tilde{z}_i$ with probability density functions (PDFs) $f_i(z)$ and cumulative distribution functions (CDFs) $F_i(z)$. However, a fundamental challenge in the CCKP is that the full knowledge of the probability distribution may not be accessible. Instead of estimating the empirical probability distribution of $\tilde{z}_i$, we employ a set of plausible probability distributions, termed the ambiguity set $\mathcal{F}_i$. The DR-CCKP is defined as

$$\text{(DR-CCKP)} \max_{x \in \{0,1\}^n} \left\{ \sum_{i=1}^n c_i x_i : \sup_{F \in \mathcal{F}} \text{Prob}\left(\sum_{i=1}^n (\bar{a}_i + \hat{a}\tilde{z}_i) x_i > b \right) \le \alpha \right\}, \tag{2.1}$$

where $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \ldots \mathcal{F}_n$ represents the ambiguity set of the joint distribution because $\tilde{z}_i$ is independent for all $i \in N$. Given the binary decision vector $x \in \{0, 1\}^n$, let $I(x) := \{i \in N : \ x_i = 1\}$ denote the set of subscripts corresponding to the selected uncertain volumes. For notation convenience, we equivalently denote $I(x) := \{i_1, i_2, \ldots, i_{m(x)}\}$ where $m(x)$ is the cardinality of the set $I(x)$. Before constructing the ambiguity set $\mathcal{F}$, we propose a novel upper bound for the violation probability by introducing a nonnegative and integrable functions $g_i(z)$ for $i \in N$, which first-order stochastically dominates $f_i(z)$, in the following Theorem 2.1.

THEOREM 2.1. *If there exists a nonnegative and integrable function $g_i(z)$ that first-order stochasticallly dominates the PDF $f_i(z)$ such that*

$$\int_y^1 g_i(z) dz \ge \int_y^1 f_i(z) dz \, \forall \mathrm{y} \in [-1, 1], i \in N, \tag{2.2}$$

then the violation probability satisfies

$$\text{Prob}\left(\sum_{i=1}^n (\bar{a}_i + \hat{a}\tilde{z}_i) x_i > b \right) \le \int \cdots \int_{z \in S(x)} \left[\prod_{j=1}^{m(x)} g_{i_j}(z_{i_j}) \right] dz_{i_1} \ldots dz_{i_{m(x)}}, \tag{2.3}$$

where $S(x) := \{z = (z_{i_1}, z_{i_2}, \ldots, z_{i_{m(x)}}) \in [-1, 1]^{m(x)} : \sum_{j=1}^{m(x)} \hat{a} z_{i_j} > b - \sum_{j=1}^{m(x)} \bar{a}_{i_j}\}$ and $x \in \{0, 1\}^n$.

Proof. We use mathematical induction to prove this theorem.

First, we prove that the statement holds when $m(x) = 1$. We assume that $b \in [\bar{a}_{i_1} - \hat{a}, \bar{a}_{i_1} + \hat{a}]$ such that $-1 \le (b - \bar{a}_{i_1})/\hat{a} \le 1$. Otherwise, the violation probability is zero when b is larger and one when b is smaller. The violation probability satisfies

$$\text{(2.4)}\quad \text{Prob}\left(\sum_{i=1}^{n}(\bar{a}_i+\hat{a}\tilde{z}_i)x_i>b\right)=\int_{(b-\bar{a}_{i_1})/\hat{a}}^{1} f_{i_1}(z_{i_1})dz_{i_1}$$
$$\le \int_{(b-\bar{a}_{i_1})/\hat{a}}^{1} g_{i_1}(z_{i_1})dz_{i_1}=\int_{z_{i_1}\in S(x)} g_{i_1}(z_{i_1})dz_{i_1},$$

where $S(x)=\{z_{i_1}\in[-1,1],\hat{a}z_{i_1}>b-\bar{a}_{i_1}\}$. The inequality in (2.4) derives from the definition of first-order stochastic dominance in (2.2).

Second, we assume that the statement holds when $m(x)=k-1\ge 2$. Then, we prove that the statement holds when $m(x)=k$. To exclude trivial cases, we similarly assume that $b\in[\sum_{j=1}^{m(x)}\bar{a}_{i_j}-m(x)\hat{a},\sum_{j=1}^{m(x)}\bar{a}_{i_j}+m(x)\hat{a}]$. Denoting $\beta(x)=b-\sum_{j=1}^{m(x)}\bar{a}_{i_j}$, it holds that

$$\text{(2.5)}\qquad \text{Prob}\left(\sum_{i=1}^{n}(\bar{a}_i+\hat{a}\tilde{z}_i)x_i>b\right)=\int\cdots\int_{z\in S(x)}\left[\prod_{j=1}^{m(x)} f_{i_j}(z_{i_j})\right]dz_{i_1}\dots dz_{i_{m(x)}},$$

where $S(x):=\{(z_{i_1},z_{i_2},\dots,z_{i_k})\in[-1,1]^k:\sum_{j=1}^{k}\hat{a}z_{i_j}\ge\beta(x)\}$. Then, we show that

(2.6)

$$\int\cdots\int_{(z_{i_1},z_{i_2},\dots,z_{i_k})\in S(x)} f_{i_1}(z_{i_1})f_{i_2}(z_{i_2})\dots f_{i_k}(z_{i_k})dz_{i_1}dz_{i_2}\dots dz_{i_k}$$
$$=\int_{z'_{i_k}}^{1} f_{i_k}(z_{i_k})dz_{i_k}\text{Prob}\left(\sum_{j=1}^{k-1}\hat{a}\tilde{z}_{i_j}>\beta(x)-\hat{a}z_{i_k}\right)$$
$$\le\int_{z'_{i_k}}^{1} f_{i_k}(z_{i_k})dz_{i_k}\int\cdots\int_{(z_{i_1},z_{i_2},\dots,z_{i_{k-1}})\in S_{i_k}(z_{i_k})}\left[\prod_{j=1}^{k-1} g_{i_j}(z_{i_j})\right]dz_{i_1}\dots dz_{i_{k-1}}$$
$$\le\int_{z'_{i_k}}^{1} g_{i_k}(z_{i_k})dz_{i_k}\int\cdots\int_{(z_{i_1},z_{i_2},\dots,z_{i_{k-1}})\in S_{i_k}(z_{i_k})}\left[\prod_{j=1}^{k-1} g_{i_j}(z_{i_j})\right]dz_{i_1}\dots dz_{i_{k-1}}$$
$$=\int\cdots\int_{z\in S(x)}\left[\prod_{j=1}^{k} g_{i_j}(z_{i_j})\right]dz_{i_1}\dots dz_{i_k},$$

where $S_{i_k}(z_{i_k})=\{(z_{i_1},z_{i_2},\dots,z_{i_{k-1}})\in[-1,1]^{k-1},\sum_{j=1}^{k-1}\hat{a}z_{i_j}\ge\beta(x)-\hat{a}z_{i_k}\}$ and $z'_{i_k}=\max\{-1,\frac{\beta(x)-(k-1)\hat{a}}{\hat{a}}\}$. The first inequality in (2.6) holds by the induction assumption. The second inequality in (2.6) holds by the following arguments, in which

$$\text{(2.7)}\qquad F(z_{i_k})=\int\cdots\int_{(z_{i_1},z_{i_2},\dots,z_{i_{k-1}})\in S_{i_k}(z_{i_k})}\left[\prod_{j=1}^{k-1} g_{i_j}(z_{i_j})\right]dz_{i_1}\dots dz_{i_{k-1}}.$$

For any given $z'_{i_k}\in[-1,1]$, according to the second mean value theorem for definite integrals [18], there exists $\zeta\in(z'_{i_k},1)$ such that

$$
\begin{aligned}
(2.8)\quad & \int_{z'_{i_k}}^{1} [f_{i_k}(z_{i_k}) - g_{i_k}(z_{i_k})]F(z_{i_k})dz_{i_k} \\
&= F(z'_{i_k}) \int_{z'_{i_k}}^{\zeta} [f_{i_k}(z_{i_k}) - g_{i_k}(z_{i_k})]dz_{i_k} + F(1) \int_{\zeta}^{1} [f_{i_k}(z_{i_k}) - g_{i_k}(z_{i_k})]dz_{i_k} \\
&\leq F(z'_{i_k}) \int_{z'_{i_k}}^{\zeta} [f_{i_k}(z_{i_k}) - g_{i_k}(z_{i_k})]dz_{i_k} + F(z'_{i_k}) \int_{\zeta}^{1} [f_{i_k}(z_{i_k}) - g_{i_k}(z_{i_k})]dz_{i_k} \\
&= F(z'_{i_k}) \int_{z'_{i_k}}^{1} [f_{i_k}(z_{i_k}) - g_{i_k}(z_{i_k}))]dz_{i_k} \leq 0.
\end{aligned}
$$

The first inequality in (2.8) holds by $\int_{\zeta}^{1} [f_{i_k}(z_{i_k}) - g_{i_k}(z_{i_k})]dz_{i_k} \leq 0$ and $F(z'_{i_k}) \leq F(1)$. The second inequality in (2.8) holds by $F(z'_{i_k}) \geq 0$ and $\int_{z'_{i_k}}^{1} [f_{i_k}(z_{i_k}) - g_{i_k}(z_{i_k})]dz_{i_k} \leq 0$. Consequently, the statement holds for all $m(x) \in N$, and the proof is complete. □

Theorem 2.1 establishes a sufficient condition in (2.2) for deriving an upper bound on the violation probability. Notably, if we find a nonnegative and integrable function $g_i(z)$ that can bound the tail probability distribution $\int_y^1 f_i(z)dz \ \forall\ y \in [-1, 1]$, then we obtain an upper bound of the violation probability, which is expressed in (2.3) as $\int \cdots \int_{z\in S(x)} \left[\prod_{j=1}^{m(x)} g_{i_j}(z_{i_j})\right] dz_{i_1} \ldots dz_{i_{m(x)}}$. For the sake of computational tractability toward this newly proposed upper bound, it is convenient to select a constant value for $g_i(z)$ that satisfies the sufficient condition. Intuitively, a larger selected constant results in a higher upper bound, thereby yielding a more conservative approximation of the violation probability. Motivated by this observation, we seek to identify a minimum constant that satisfies the sufficient condition. Subsequently, we introduce the concept of the MATPD in Definition 2.2 and demonstrate its minimality in Theorem 2.3.

Definition 2.2 (maximum average tail probability distribution). *For a random variable $\tilde{z}_i$ with the integrable PDF $f_i(z)$ over its support set* $[-1, 1]$*, the MATPD is defined as*

$$
\text{(MATPD)}\ \frac{1}{\mu_i} := \max_{y\in[-1,1]} h_i(y), \tag{2.9}
$$

where

$$
h_i(y) = \begin{cases} \frac{\int_y^1 f_i(z)dz}{1-y}, & y \in [-1, 1), \\ f_i(1), & y = 1. \end{cases} \tag{2.10}
$$

Definition 2.2 stipulates that the MATPD of random variable $\tilde{z}_i$ is defined as $1/\mu_i$, which corresponds to the maximum value of $h_i(y)$ within its support set $[-1, 1]$. Notably, the tail probability distribution $\int_y^1 f_i(z)dz$ is positive for any $y \in [-1, 1]$, and the average tail probability distribution $h_i(y)$ is scaled over the interval $[y, 1]$. When the variable y equals -1, the average tail probability distribution $h_i(y)$ is $1/2$. As the variable y approaches one, the function $h_i(y)$ converges to $f_i(1)$. Thus, our MATPD is properly defined on the support set $[-1, 1]$. Next, we present several useful properties about the MATPD.

Theorem 2.3. *The MATPD of random variable $\tilde{z}_i$ satisfies the following properties:*

(i) *Dominance:* $\int_y^1 1/\mu_i dz \geq \int_y^1 f_i(z)dz\ \forall y \in [-1,1]$.

(ii) *Minimum:* $1/\mu_i$ *is the minimum constant* c *that satisfies* $\int_y^1 c dz \geq \int_y^1 f_i(z)dz$ $\forall y \in [-1,1]$.

(iii) *Uniqueness:* $1/\mu_i$ *equals* $1/2$ *if* $f_i(z)$ *is a nonincreasing PDF.*

Proof. (i) According to Definition 2.2, it holds that $1/\mu_i \geq \int_y^1 f_i(z)dz/(1-y)$ for any $y \in [-1,1]$. By simple reformulations, it holds that $\int_y^1 1/\mu_i dz \geq \int_y^1 f_i(z)dz$ or, equivalently, $F_i(y) \geq 1 - 1/\mu_i + y/\mu_i$, for any $y \in [-1,1]$. Thus, (i) is proven.

(ii) It follows directly from the definition of MATPD. Thus, (ii) is proven.

(iii) First, we prove that $1/\mu_i \geq 1/2$. According to Definition 2.2, it holds that $1/\mu_i \geq \int_y^1 f_i(z)dz/(1-y)$ for any $y \in [-1,1]$. Then, by setting $y=-1$, it holds that $1/\mu_i \geq \int_{-1}^1 f_i(z)dz/2 = 1/2$.

Second, we prove that $1/\mu_i \leq 1/2$. (1) We prove that $1/\mu_i \leq 1/2$ when optimal $y=1$. Since $f_i(z)$ is a nonincreasing PDF, it holds that $f_i(1) \leq 1/2$. Otherwise, $\int_{-1}^1 f_i(z)dz > \int_{-1}^1 1/2 dz = 1$ contradicts the definition of PDFs. Therefore, $1/\mu_i \leq f_i(1) \leq 1/2$. (2) We prove that $1/\mu_i \leq 1/2$ when optimal $y \in [-1,1)$. By the definition of the MATPD, it holds that $1/\mu_i \leq 1/2$ is equivalent to $F_i(y) - 1/2 - y/2 \geq 0$. Denote $T_i(y) = F_i(y) - 1/2 - y/2$. Then, we only need to verify that $T_i(y) \geq 0$ for all $y \in [-1,1)$. To exclude trivial cases, we here consider that the PDF $f_i(z)$ is first-order differentiable and strictly decreasing. The function $T_i(y)$ is concave in $y \in [-1,1]$, due to $\partial^2 T_i(y)/\partial y^2 = df_i(y)/dy < 0$. Besides, the function $T_i(y)$ equals zero on the end points such that $T_i(-1) = T_i(1) = 0$. Consequently, $T_i(y) \geq \max(T_i(-1), T_i(1)) = 0$ for all $y \in [-1,1)$. To conclude, (iii) is proven and the proof is completed. ∎

Theorem 2.3 provides several prominent properties of the proposed MATPD. The first property ensures that the MATPD dominates the PDF $f_i(z)$, thereby satisfying the sufficient condition presented in (2.2), achieved by setting a constant function $g_i(z) = 1/\mu_i$ for $z \in [-1,1]$. This property motivates us to adopt the MATPD as a constant proxy of $f_i(z)$, enabling us to derive an upper bound of the violation probability. The second property states that the MATPD is the minimum constant that dominates the PDF $f_i(z)$. This property guarantees that our MATPD is the best among all constant proxies of $f_i(z)$, resulting in the tightest upper bound of the violation probability. The third property reveals that the MATPD equals $1/2$ for any nonincreasing PDF. This property is particularly useful when dealing with random variables that exhibit nonincreasing PDFs.

Based on the proposed MATPD, we now formally present the ambiguity set $\mathcal{F}_i$ for the uncertain volume $\tilde{z}_i$. Given the CDF $F_i(y)$ for the uncertain volume $\tilde{z}_i$, we know that $F_i(y)$ is bounded such that $F_i(y) \geq 1 - \int_y^1 1/\mu_i$ or, equivalently, $F_i(y) \geq 1 - 1/\mu_i + y/\mu_i$ (see the proof of (i) in Theorem 2.3). Then, the ambiguity set $\mathcal{F}_i$ induced by the MATPD $1/\mu_i$ is defined as

$$\mathcal{F}_i := \left\{ F_i : F_i(y) \geq 1 - \frac{1}{\mu_i} + \frac{y}{\mu_i}, y \in [-1,1] \right\}. \tag{2.11}$$

Intuitively, the ambiguity set $\mathcal{F}_i$ contains more distributions as the MATPD $1/\mu_i$ decreases. The motivation for using the ambiguity set $\mathcal{F}_i$ may be attributed to the following reasons. It is simple to approximate the violation probability, as the general function $g_i(y)$ is replaced by the MATPD, the minimum constant, to facilitate computations. Second, it can derive a tight upper bound of the violation probability, when uncertain volumes follow uniform distributions. This conclusion directly follows from Remarks 3.4 and 3.5. In our analysis later, we show that the proposed ambiguity set $\mathcal{F}_i$ derives an explicit upper bound of the violation probability in section 3 and possesses computational tractability for DR-CCKP in section 4.

Afterward, we first analyze truncated normal distributions to demonstrate the benefits of choosing the MATPD. For example, we consider three truncated normal distributions with zero expectations and standard deviations of 1, 2, and 3, respectively. We compare our proposed ambiguity set $\mathcal{F}_i$ with the ambiguity set $\mathcal{F}_i' = \{\tilde{z}_i \in [-1,1], E(\tilde{z}_i) = 0\}$ proposed by [35]. Besides the support, both $\mathcal{F}_i$ and $\mathcal{F}_i'$ incorporate one more statistic into the ambiguity set (i.e., $\mathcal{F}_i$ with the MATPD and $\mathcal{F}_i'$ with the mean). Under these distributions, the MATPD equals 0.50, 0.51, and 0.52, respectively. In this example, we can tailor our ambiguity set $\mathcal{F}_i$ for each distribution but construct an identical ambiguity set $\mathcal{F}_i'$ for all distributions. Consequently, our MATPD can build a more accurate ambiguity set based on distribution information, thereby reducing the model's conservativeness. In our numerical studies later, we show that our DR-CCKP yields higher objective values compared with existing (distributionally) robust optimization approaches.

Then, we analyze the consistency of MATPDs under different sample sizes. For instance, we consider a truncated exponential distribution with the rate parameter being one. Then, we generate 50, 500, and 5000 samples to estimate the rate parameter and the MATPD. In this example, we find that the largest gap is 11.5% between the estimated rate parameters and 1.25% between the estimated MATPD. This observation motivates us to adopt MATPDs under limited sample sizes. In our numerical studies later, we show that our DR-CCKP yields more stable objective values compared with existing stochastic optimization approaches.

3. Approximating chance constraint under distributional ambiguity set. The CCKP is generally challenging to solve due to the combinatorial nature of decision variables and the nonconvexity of chance constraints. Particularly, the chance constraint in (1.1) represents a value-at-risk (VaR) constraint, which is formulated as $\mathrm{VaR}_{1-\alpha}\big(\sum_{i=1}^n \tilde{a}_i x_i\big) < b$ [43]. To solve CCKPs, several celebrated tractable approximations have been identified to restore the convexity of VaR constraints. For instance, CVaR and EVaR have been proposed by [2] and [42], respectively. More sophisticated tractable approximations have been designed for chance-constrained programs with continuous decision variables [25, 27].

In this section, we approximate the chance constraint under the ambiguity set $\mathcal{F}$ constructed on the supportwise statistic MATPD. Specifically, we set out to derive the violation probability $\sup_{F\in\mathcal{F}} \mathrm{Prob}\big(\sum_{i=1}^n (\hat{a}\tilde{z}_i)x_i > \beta(x)\big)$. For any given $x \in \{0,1\}^n$, we focus on $b \in [\sum_{j=1}^{m(x)} \bar{a}_{i_j} - m(x)\hat{a}, \sum_{j=1}^{m(x)} \bar{a}_{i_j} + m(x)\hat{a}]$. Otherwise, the violation probability equals one when b is smaller and zero when b is larger. After eliminating these trivial cases, We establish Theorem 3.1 as follows.

Theorem 3.1. *For any given $x \in \{0,1\}^n$ and MATPDs $1/\mu_i$ of $\tilde{z}_i$ for all $i \in N$, the violation probability of the chance constraint under the ambiguity set $\mathcal{F}$ satisfies*

$$\sup_{F\in\mathcal{F}} \mathrm{Prob}\left(\sum_{i=1}^n (\bar{a}_i + \hat{a}\tilde{z}_i)x_i > b\right) \le u(x,\beta(x)), \tag{3.1}$$

where the upper bound $u(x,\beta(x))$ is

$$u(x,\beta(x)) := \sum_{k=0}^{m(x)} (-1)^k C_{m(x)}^k \frac{\max\{[m(x)-2k]\hat{a} - \beta(x), 0\}^{m(x)}}{m(x)! \prod_{i=1}^n (\hat{a}\mu_i)^{x_i}}, \tag{3.2}$$

and $\beta(x) := b - \sum_{j=1}^{m(x)} \bar{a}_{i_j}$. *Here,* $C_{m(x)}^k$ *denotes the number of combinations of* $m(x)$ *taken* k *at a time.*

Proof. First, the violation probability under the ambiguity set $\mathcal{F}$ satisfies

$$\begin{aligned}\sup_{F\in\mathcal{F}} \text{Prob}\left(\sum_{i=1}^{n}(\bar{a}_i + \hat{a}\tilde{z}_i)x_i > b\right) &= \sup_{F\in\mathcal{F}} \int \dots \int_{z\in S(x)} \left[\prod_{j=1}^{m(x)} f_{i_j}(z_{i_j})\right] dz_{i_1}\dots dz_{i_{m(x)}} \\ &\le \int \dots \int_{z\in S(x)} \left[\prod_{l=1}^{m(x)} \frac{1}{\mu_{i_l}}\right] dz_{i_1}\dots dz_{i_{m(x)}} \\ &= \left[\prod_{l=1}^{m(x)} \frac{1}{\mu_{i_l}}\right] \int \dots \int_{z\in S(x)} dz_{i_1}\dots dz_{i_{m(x)}},\end{aligned}$$

where $S(x) := \{(z_{i_1}, z_{i_2}, \dots, z_{i_{m(x)}}) \in [-1,1]^{m(x)} : \sum_{l=1}^{m(x)} \hat{a}z_{i_l} > \beta(x)\}$ is measurable, and the inequality derives from Theorem 2.1 and Definition 2.2.

Second, the upper bound of the violation probability relies on solving the definite integral $\int \dots \int_{z\in S(x)} dz_{i_1}\dots dz_{i_{m(x)}}$, which represents the volume of set $S(x)$. Let the set function $|\cdot|$ denote the volume of measurable sets. Define sets $S'(x)$ and $S_j(x)$ as follows:

$$S'(x) := \left\{(z_{i_1}, z_{i_2}, \dots, z_{i_{m(x)}}) \in [-\infty, 1]^{m(x)} : \sum_{l=1}^{m(x)} \hat{a}z_{i_l} > \beta(x)\right\}, \tag{3.3}$$

and

$$S_j(x) := S'(x) \bigcap \{(z_{i_1}, z_{i_2}, \dots, z_{i_{m(x)}}) : z_{i_j} < -1\}\ \forall j \in I(x). \tag{3.4}$$

It holds that $|S(x)| = |S'(x)| - |\bigcup_{j=1}^{m(x)} S_j(x)|$, due to the fact that $S(x) = S'(x) \setminus (\bigcup_{j=1}^{m(x)} S_j(x))$. Let P_0 denote $|S'(x)|$. As shown in Lemma 3.2, P_0 has an analytic expression.

Then, by applying the principle of inclusion-exclusion, it holds that

$$\left|\bigcup_{j=1}^{m(x)} S_j(x)\right| = \sum_{k=1}^{m(x)} (-1)^{k-1} P_k,$$

where $P_k = \sum_{\{j_1,j_2,..,j_k\}\in\Omega(k)} |\bigcap_{j\in\{j_1,j_2,..,j_k\}} S_j(x)|$, and the set collection $\Omega(k)$ denotes the sigma-algebra on $I(x)$ wherein each set element has cardinality k, $\forall k \in \{1, 2, \dots, m(x)\}$. For example, if $I(x) = \{1,2,3\}$, then $\Omega(1) = \{\{1\},\{2\},\{3\}\}$, $\Omega(2) = \{\{1,2\},\{2,3\},\{1,3\}\}$, and $\Omega(3) = \{\{1,2,3\}\}$. As shown in Lemma 3.3, P_k, $\forall k \in \{1, 2, \dots, m(x)\}$, also has an analytic expression.

As a result, it obtains that $|S(x)| = P_0 + \sum_{k=1}^{m(x)} (-1)^{k-1} P_k$. After substituting P_0 and P_k obtained in Lemmas 3.2 and 3.3, the proof is completed. ∎

Theorem 3.1 provides an upper bound $u(x, \beta(x))$ for the violation probability under the ambiguity set $\mathcal{F}$. Intuitively, the tightness of the upper bound depends on the MATPD. One concern is that the MATPD may be loose when the underlying distribution is concentrated. However, we demonstrate that the MATPD slightly increases when the PDF becomes more concentrated. To illustrate this observation, we compare the MATPD (i.e., $1/\mu_i$) with the peak of density (i.e., $UB_i := \max_{y\in[-1,1]} f_i(y)$).

Let y^* be the optimal solution to the optimization problem $\max_{y\in[-1,1]} h_i(y)$. By the first mean value theorem, there exists a $\xi \in (y^*, 1)$ such that $1/\mu_i = f_i(\xi) \le UB_i$, because $\int_{y^*}^1 f_i(y)dz = f_i(\xi)$. Although a highly concentrated $f_i(y)$ can lead to a large UB_i, the obtained $1/\mu_i$ can be significantly smaller than the UB_i. For example, consider two truncated normal distributions in $[-1,1]$ with zero expectation, one with a standard deviation of $\sigma = 0.1$ and the other with $\sigma = 1$. The UB_i for the former is 4.99, and that for the latter is 0.58. However, the MATPD for the former is 1.03, and that for the latter is 1.92. Despite the significant difference in the distributions, the MATPD for the concentrated one is not far from that for the dispersed one. Thus, the proposed upper bound $u(x, \beta(x))$ may not be sensitive to the form of underlying distributions. Numerical studies in section 6 indicate that our approach performs well across various underlying distributions.

LEMMA 3.2. *For any* $x \in \{0,1\}^n$, *it holds that* $P_0 := \frac{(m(x)\hat{a}-\beta(x))^{m(x)}}{[m(x)]!\prod_{i\in I(x)}(\hat{a})}$ *where* $\beta(x) = b - \sum_{j=1}^{m(x)} \bar{a}_{i_j}$.

Proof. We use mathematical induction to prove this lemma.

First, we prove that the statement holds when $m(x) = 1$. It holds that

$$P_0 = \int_{z_{i_1}\in S'(x)} dz_{i_1} = \int_{\beta(x)/\hat{a}}^1 dz_{i_1} = \frac{\hat{a} - \beta(x)}{\hat{a}}.$$

Second, we assume that that the statement holds when $m(x) = k-1 \ge 2$. Then, we prove that the statement holds when $m(x) = k$. It holds that

$$\begin{aligned} P_0 &= \int \cdots \int_{(z_{i_1}, z_{i_2}, \ldots, z_{i_{m(x)}})\in S'(x)} dz_{i_1} dz_{i_2} \ldots dz_{i_k}, \\ &= \int_{\frac{\beta(x)-\sum_{l=1}^{k-1}\hat{a}}{\hat{a}}}^1 dz_{i_k} \int \cdots \int_{(z_{i_1}, z_{i_2}, \ldots, z_{i_{k-1}})\in S(z_{i_k})} dz_{i_1} dz_{i_2} \ldots dz_{i_{k-1}}, \\ &= \frac{(m(x)\hat{a} - \beta(x))^{m(x)}}{[m(x)]!\prod_{i\in I(x)}(\hat{a})}, \end{aligned}$$

where $S(z_{i_k}) = \{(z_1, z_2, \ldots, z_{k-1}) : z_{i_l} \in [-\infty, 1]\, \forall l \in \{1, 2, \ldots, k-1\}, \sum_{l=1}^{k-1} \hat{a} z_{i_l} > \beta(x) - \hat{a} z_{i_k}\}$. The proof is completed. ∎

LEMMA 3.3. *For any* $x \in \{0,1\}^n$, *it holds that*

$$P_k = (-1)^k C_{m(x)}^k \frac{\max\left\{m(x)\hat{a} - \beta(x) - 2k\hat{a}, 0\right\}^{m(x)}}{[m(x)]!\prod_{i\in I(x)}(\hat{a})},$$

where $\beta(x) = b - \sum_{j=1}^{m(x)} \bar{a}_{i_j}$.

Proof. We use mathematical induction to prove this lemma.

First, we prove that the statement holds when $m(x) = 1$. For simplicity of notation, suppose that $I(x) \setminus J = \{i_1, i_2, \ldots, i_{m(x)-1}\}$ and $J = \{i_{m(x)}\}$. It holds that

$$\begin{aligned} &\int \cdots \int_{(z_{i_1}, z_{i_2}, \ldots, z_{i_{m(x)}})\in \bigcap_{j\in J} S_j} dz_{i_1} dz_{i_2} \ldots dz_{i_{m(x)}} \\ &\quad = \int_{\frac{\beta(x)-(m(x)-1)\hat{a}}{\hat{a}}}^{-1} dz_{i_{m(x)}} \int_{(z_{i_1}, z_{i_2}, \ldots, z_{i_{m(x)-1}})\in S(z_{i_{m(x)}})} dz_{i_1} \ldots d_{z_{i_{m(x)-1}}}, \end{aligned}$$

$$= \int_{\frac{\beta(x)-(m(x)-1)\hat{a}}{\hat{a}}}^{-1} dz_{i_{m(x)}} \frac{\max\{(m(x)-1)\hat{a} - \beta(x) + \hat{a}z_{i_{m(x)}}, 0\}^{m(x)-1}}{[m(x)-1]! \prod_{i=1}^{m(x)-1}(\hat{a})},$$
$$= \frac{\max\{m(x)\hat{a} - \beta(x) - 2\sum_{i\in J}\hat{a}, 0\}^{m(x)}}{[m(x)]! \prod_{i\in I(x)}(\hat{a})},$$

where $S(z_{i_{m(x)}}) = \{(z_{i_1}, z_{i_2}, \ldots, z_{i_{m(x)-1}}) : z_{i_l} \in [-\infty, 1] \forall l \in \{1, 2, \ldots, m(x)-1\}, \sum_{l=1}^{m(x)-1} \hat{a}z_{i_l} \geq \beta(x) - \hat{a}z_{i_{m(x)}}\}$.

Second, we assume that that the statement holds when $m(x) = k-1 \geq 2$. Then, we prove that the statement holds when $m(x) = k$. For brevity of notation, we assume that $J = \{i_{m(x)-m+1}, i_{m(x)-m+2}, \ldots, i_{m(x)}\}$ and $I(x) \setminus J = \{i_1, i_2, \ldots, i_{m(x)-m}\}$. It holds that

$$\int \ldots \int_{(z_{i_1}, z_{i_2}, \ldots, z_{i_{m(x)}}) \in \bigcap_{k\in J} S_k} dz_{i_1} dz_{i_2}, \ldots, dz_{i_{m(x)}}$$
$$= \int_{\frac{\beta(x)-(m(x)-1)\hat{a}}{\hat{a}}}^{-1} dz_{i_{m(x)}} \int_{(z_{i_1}, z_{i_2}, \ldots, z_{i_{m(x)-1}}) \in S(z_{i_{m(x)}})} dz_{i_1} \ldots d_{z_{i_{m(x)-1}}},$$
$$= \frac{\max\{m(x)\hat{a} - \beta(x) - 2m\hat{a}, 0\}^{m(x)}}{[m(x)]! \prod_{i\in I(x)}(\hat{a})},$$

where $S(z_{i_{m(x)}}) = \{(z_{i_1}, z_{i_2}, \ldots, z_{i_{m(x)-1}}) : z_{i_l} \in [-\infty, 1], \forall l \in \{1, 2, \ldots, m(x)-m\}, z_{i_l} \in [-\infty, -1], \forall l \in \{m(x)-m+1, m(x)-m+2, \ldots, m(x)\}, \sum_{l=1}^{m(x)-1} \hat{a}z_{i_l} \geq \beta - \hat{a}z_{i_{m(x)}}\}$. Together with the fact that $|\Omega(k)| = C_{m(x)}^k$, the proof is completed. ∎

Remark 3.4. If uncertain volumes $\tilde{a}_i$ for $i \in N$ follow uniform distributions with a homogeneous deviation $\hat{a}$, the approximation of chance constraint is tight such that $\text{Prob}\big(\sum_{i=1}^n (\bar{a}_i + \hat{a}\tilde{z}_i)x_i > b\big) = u(x, \beta(x))$.

To verify Remark 3.4, we note that the violation probability $\text{Prob}(\sum_{i=1}^n \tilde{a}_i x_i > b)$ in (1.1) equals $\text{Prob}(\sum_{i=1}^n (\bar{a}_i + \hat{a}\tilde{z}_i)x_i > b)$ due to the homogeneous deviation $\hat{a}$. Then, both the PDF and the MATPD of random variable $\tilde{z}_i$ have the same value of $1/2$ when they follow uniform distributions (as shown in (iii) of Theorem 2.3). This implies that the true violation probability $\text{Prob}(\sum_{i=1}^n (\bar{a}_i + \hat{a}\tilde{z}_i)x_i > b)$ equals its approximation $\sup_{F\in\mathcal{F}} \text{Prob}(\sum_{i=1}^n (\bar{a}_i + \hat{a}\tilde{z}_i)x_i > b)$ under the ambiguity set $\mathcal{F}$. Consequently, this remark holds true.

Afterward, we address the issue of computing $u(x, \beta(x))$. One concern is that computing $u(x, \beta(x))$ may result in unaffordable memory consumption, as the numerator and denominator of $u(x, \beta(x))$ could be very large numbers. For example, the factorial calculation $m(x)!$ may require excessive memory consumption. To tackle this computational issue, we scale the numerator and denominator by multiplying $1/(m(x)\hat{a})^{m(x)}$ simultaneously. As such, $u(x, \beta(x))$ is equivalently reformulated as $A^{m(x)}/[B \prod_{i\in I(x)}(\mu_i)]$, where $A = \max\{[m(x)-2k]/m(x) - \beta(x)/[m(x)\hat{a}], 0\}$ and $B = m(x)!/m(x)^{m(x)}$ take values between zero and one. Thus, the proposed $u(x, \beta(x))$ is capable of handling large-scale DR-CCKPs.

Remark 3.5. For any given $x \in \{0,1\}^n$ and MATPDs $1/\mu_i$ of $\tilde{z}_i\ \forall\ i \in N$, the violation probability of the chance constraint under the ambiguity set $\mathcal{F}$ satisfies

$$\sup_{F\in\mathcal{F}} \text{Prob}\left(\sum_{i=1}^{n}(\hat{a}_i\tilde{z}_i)x_i > \beta(x)\right) \le \left[1+\sum_{J\subseteq I(x)}(-1)^{|J|}[\rho(J)]^{m(x)}\right]\frac{(\sum_{i\in I(x)}\hat{a}_i-\beta(x))^{m(x)}}{[m(x)]!\prod_{i\in I(x)}(\hat{a}_i\mu_i)},$$

where $\rho(J):=\frac{\max\{\sum_{i\in I(x)}\hat{a}_i-\beta(x)-2\sum_{i\in J}\hat{a}_i,0\}}{\sum_{i\in I(x)}\hat{a}_i-\beta(x)}$ and $\beta(x):=b-\sum_{j=1}^{m(x)}\bar{a}_{i_j}$.

Proof. According to Theorem 2.1 and Definition 2.2, it is noted that

$$\text{Prob}\left(\sum_{i=1}^{n}(\hat{a}_i\tilde{z}_i)x_i > \beta(x)\right) \le \int\cdots\int_{z\in S(x)}\left[\prod_{l=1}^{m(x)}\frac{1}{\mu_{i_l}}\right]dz_{i_1}\dots dz_{i_{m(x)}},$$

where $S(x):=\{(z_{i_1},z_{i_2},\dots,z_{i_{m(x)}})\in[-1,1]^{m(x)}:\sum_{l=1}^{m(x)}\hat{a}_{i_l}z_{i_l}>\beta(x)\}$.

To obtain an upper bound for $\int\cdots\int_{z\in S(x)}\left[\prod_{l=1}^{m(x)}\frac{1}{\mu_{i_l}}\right]dz_{i_1}\dots dz_{i_{m(x)}}$, define

$$V':=\int\cdots\int_{(z_{i_1},z_{i_2},\dots,z_{i_{m(x)}})\in S'(x)}\left[\prod_{l=1}^{m(x)}\frac{1}{\mu_{i_l}}\right]dz_{i_1}\dots dz_{i_{m(x)}},$$

and

$$V_J:=\int\cdots\int_{(z_{i_1},z_{i_2},\dots,z_{i_{m(x)}})\in\bigcap_{j\in J}S_j(x)}\left[\prod_{l=1}^{m(x)}\frac{1}{\mu_{i_l}}\right]dz_{i_1}\dots dz_{i_{m(x)}},$$

where J denotes any subset of $I(x)$.

Denote $S'(x)=\{(z_{i_1},z_{i_2},\dots,z_{i_{m(x)}})\in[-\infty,1]^{m(x)}:\sum_{l=1}^{m(x)}\hat{a}_{i_l}z_{i_l}>\beta(x)\}$ and $S_j(x)=\{(z_{i_1},z_{i_2},\dots,z_{i_{m(x)}}):z_{i_l}\in[-\infty,1],\forall l\in I(x)\setminus j, z_{i_j}\in[-\infty-1],\sum_{l=1}^{m(x)}\hat{a}_{i_l}z_{i_l}>\beta\}$. By a similar argument in the proof for Theorem 3.1, it holds that

$$\int\cdots\int_{(z_{i_1},z_{i_2},\dots,z_{i_{m(x)}})\in S(x)}\left[\prod_{l=1}^{m(x)}\frac{1}{\mu_{i_l}}\right]dz_{i_1}\dots dz_{i_{m(x)}}=V'+\sum_{J\subseteq I(x)}(-1)^{|J|}V_J.$$

After substituting V' and V_J obtained in Lemmas 3.6 and 3.7, the proof is completed. ∎

Remark 3.5 offers an alternative upper bound for the violation probability under the ambiguity set $\mathcal{F}$. Specifically, it seeks to derive the violation probability $\sup_{F\in\mathcal{F}}\text{Prob}\big(\sum_{i=1}^{n}(\hat{a}_i\tilde{z}_i)x_i>\beta(x)\big)$, which directly incorporates heterogeneous deviations $\hat{a}_i$ for $i\in N$. This approximation differs from the one in Theorem 3.1, where $\hat{a}$ is set to $\max\{\hat{a}_1,\hat{a}_2,\dots,\hat{a}_N\}$. On the flip side, it has an exponential complexity, as its computation enumerates all possible combinations of $J\subseteq I(x)$. For this reason, we concentrate on the polynomial-complexity approximation presented in Theorem 3.1. Following a similar argument for Remark 3.4, one can verify that the upper bound is also tight for uniform distributions even when $\hat{a}_i$ is heterogeneous.

LEMMA 3.6. *It holds that* $V'=\frac{(\sum_{i\in I(x)}\hat{a}_i-\beta)^{m(x)}}{[m(x)]!\prod_{i\in I(x)}(\hat{a}_i\mu_i)}$ *for any* $x\in\{0,1\}^n$, *where* $\beta(x)=b-\sum_{j=1}^{m(x)}\bar{a}_{i_j}$.

Proof. This lemma is proven using induction. If $m(x)=1$, the analysis follows a similar approach as the proof of Lemma 3.2. Then, supposing the theorem holds for $m(x)=k-1$ and $k-1\geq 1$, we prove that the theorem holds for $m(x)=k$. We have

$$\begin{aligned}
V' &= \int\cdots\int_{(z_{i_1},z_{i_2},\ldots,z_{i_{m(x)}})\in S'(x)} \frac{1}{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}} dz_{i_1}dz_{i_2}\ldots dz_{i_k}\\
&= \frac{1}{\prod_{l=1}^{k}(\mu_{i_l})}\int\cdots\int_{(z_{i_1},z_{i_2},\ldots,z_{i_{m(x)}})\in S'(x)} dz_{i_1}dz_{i_2}\ldots dz_{i_k}\\
&= \frac{1}{\prod_{l=1}^{k}(\mu_{i_l})}\int_{\frac{\beta(x)-\sum_{l=1}^{k-1}\hat a_{i_l}}{\hat a_{i_k}}}^{1} dz_{i_k}\frac{(\sum_{l=1}^{k-1}\hat a_{i_l}-\beta(x)+\hat a_{i_k}z_{i_k})^{k-1}}{(k-1)!\prod_{l=1}^{k-1}(\hat a_{i_l})}\\
&= \frac{(\sum_{i\in I(x)}\hat a_i-\beta(x))^{m(x)}}{[m(x)]!\Pi_{i\in I(x)}(\mu_i\hat a_i)}.
\end{aligned}$$

Here, $S(z_{i_k})=\{(z_{i_1},z_{i_2},\ldots,z_{i_{k-1}}): z_{i_l}\in[-\infty,1],\forall l\in\{1,2,\ldots,k-1\},\sum_{l=1}^{k-1}\hat a_{i_l}z_{i_l}>\beta(x)-\hat a_{i_k}z_{i_k}\}$. The second equality holds by the induction assumption. The proof is completed. ∎

LEMMA 3.7. *It holds that* $V_J=[(-1)^{|J|}[\rho(J)]^{m(x)}]\frac{(\sum_{i\in I(x)}\hat a_i-\beta(x))^{m(x)}}{[m(x)]!\prod_{i\in I(x)}(\hat a_i\mu_i)}$ *for any* $x\in\{0,1\}^n$, *where* $\rho(J)=\frac{\max\{\sum_{i\in I(x)}\hat a_i-\beta(x)-2\sum_{i\in J}\hat a_i,0\}}{\sum_{i\in I(x)}\hat a_i-\beta(x)}$ *and* $\beta(x)=b-\sum_{j=1}^{m(x)}\bar a_{i_j}$.

Proof. It is noted that $\bigcap_{j\in J}S_j=\{(z_{i_1},z_{i_2},\ldots,z_{i_{m(x)}}): z_{i_l}\in[-\infty,1],\forall l\in I(x)\setminus J, z_{i_l}\in[-\infty,-1],\forall l\in J,\sum_{l=1}^{m(x)}\hat a_{i_l}z_{i_l}>\beta(x)\}$. If $\sum_{l\in I(x)}\hat a_l-\beta-2\sum_{l\in J}\hat a_l<\beta(x)$, then it can be concluded that $\bigcap_{j\in J}S_j=\emptyset$, which implies that $V_J=0$; otherwise, we have $\bigcap_{j\in J}S_j\neq\emptyset$.

We prove this by induction. When $|J|=1$, for simplicity of notation, suppose $I(x)\setminus J=\{i_1,i_2,\ldots,i_{m(x)-1}\}$ and $J=\{i_{m(x)}\}$. It holds that

$$\begin{aligned}
&\int\cdots\int_{(z_{i_1},z_{i_2},\ldots,z_{i_{m(x)}})\in\bigcap_{j\in J}S_j}\frac{1}{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_{m(x)}}}dz_{i_1}dz_{i_2}\ldots dz_{i_{m(x)}}\\
&= \frac{1}{\prod_{l=1}^{m(x)}(\mu_{i_l})}\int_{\frac{\beta(x)-\sum_{l=1}^{m(x)-1}\hat a_{i_l}}{\hat a_{i_{m(x)}}}}^{-1}dz_{i_{m(x)}}\int_{(z_{i_1},z_{i_2},\ldots,z_{i_{m(x)-1}})\in S(z_{i_{m(x)}})}dz_{i_1}\ldots d_{z_{i_{m(x)-1}}}\\
&= \frac{1}{\prod_{l=1}^{m(x)}(\mu_{i_l})}\int_{\frac{\beta(x)-\sum_{l=1}^{m(x)-1}\hat a_{i_l}}{\hat a_{i_{m(x)}}}}^{-1}dz_{i_{m(x)}}\frac{(\sum_{l=1}^{m(x)-1}\hat a_{i_l}-\beta(x)+\hat a_{i_{m(x)}}z_{i_{m(x)}})^{m(x)-1}}{[m(x)-1]!\Pi_{l=1}^{m(x)-1}\hat a_{i_l}}\\
&= \frac{(\sum_{l\in I(x)}\hat a_l-2\sum_{l\in J}\hat a_l-\beta(x))^{m(x)}}{[m(x)]!\Pi_{l\in I(x)}(\mu_l\hat a_l)}\\
&= \rho(J)^{m(x)}\frac{(\sum_{i\in I(x)}\hat a_i-\beta(x))^{m(x)}}{[m(x)]!\Pi_{i\in I(x)}(\mu_i\hat a_i)},
\end{aligned}$$

where $S(z_{i_{m(x)}})=\{(z_{i_1},z_{i_2},\ldots,z_{i_{m(x)-1}}): z_{i_l}\in[-\infty,1],\forall l\in\{1,2,\ldots,m(x)-1\},\sum_{l=1}^{m(x)-1}\hat a_{i_l}z_{i_l}\geq\beta(x)-\hat a_{i_{m(x)}}z_{i_{m(x)}}\}$.

When the lemma holds for $|J|=m-1$ and $m-1\geq 1$, we prove as follows that it holds for $|J|=m$. Without loss of generality, we assume that $J=\{i_{m(x)-m+1}, i_{m(x)-m+2},\ldots,i_{m(x)}\}$ and $I(x)\setminus J=\{i_1,i_2,\ldots,i_{m(x)-m}\}$. It holds that

$$
\begin{aligned}
&\int \cdots \int_{(z_{i_1}, z_{i_2}, \ldots, z_{i_{m(x)}}) \in \bigcap_{k \in J} S_k} \frac{1}{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_{m(x)}}} dz_{i_1} dz_{i_2}, \ldots, dz_{i_{m(x)}} \\
&= \frac{1}{\prod_{l=1}^{m(x)}(\mu_{i_l})} \int_{\frac{\beta(x) - \sum_{l=1}^{m(x)-m} \hat{a}_{i_l} + \sum_{l=m(x)-m+1}^{m(x)-1} \hat{a}_{i_l}}{\hat{a}_{i_{m(x)}}}}^{-1} dz_{i_{m(x)}} \\
&\quad \times \int_{(z_{i_1}, z_{i_2}, \ldots, z_{i_{m(x)-1}}) \in S(z_{i_{m(x)}})} dz_{i_1} \ldots d_{z_{i_{m(x)-1}}} \\
&= \frac{1}{\prod_{l=1}^{m(x)}(\mu_{i_l})} \int_{\frac{\beta(x) - \sum_{l=1}^{m(x)-m} \hat{a}_{i_l} + \sum_{j=m(x)-m+1}^{m(x)-1} \hat{a}_{i_l}}{\hat{a}_{i_{m(x)}}}}^{-1} dz_{i_{m(x)}} \\
&\quad \times \frac{(\sum_{l=1}^{m(x)-1} \hat{a}_{i_l} - \beta(x) - 2\sum_{l=m(x)-m+1}^{m(x)-1} \hat{a}_{i_l} + \hat{a}_{i_{m(x)}} z_{i_{m(x)}})^{m(x)-1}}{[m(x)-1]! \Pi_{l=1}^{m(x)-1} \hat{a}_{i_l}} \\
&= \frac{(\sum_{l \in I(x)} \hat{a}_l - 2\sum_{l \in J} \hat{a}_l - \beta(x))^{m(x)}}{[m(x)]! \Pi_{l \in I(x)}(\mu_l \hat{a}_l)} \\
&= \rho(J)^{m(x)} \frac{(\sum_{i \in I(x)} \hat{a}_i - \beta(x))^{m(x)}}{[m(x)]! \Pi_{i \in I(x)}(\mu_i \hat{a}_i)},
\end{aligned}
$$

where $S(z_{i_{m(x)}}) = \{(z_{i_1}, z_{i_2}, \ldots, z_{i_{m(x)-1}}) : z_{i_l} \in [-\infty, 1], \forall l \in \{1, 2, \ldots, m(x) - m\}, z_{i_l} \in [-\infty, -1], \forall l \in \{m(x) - m + 1, m(x) - m + 2, \ldots, m(x)\}, \sum_{l=1}^{m(x)-1} \hat{a}_{i_l} z_{i_l} \geq \beta(x) - \hat{a}_{i_{m(x)}} z_{i_{m(x)}}\}$. The proof is completed. ∎

Finally, with the upper bound $u(x, \beta(x))$ for the violation probability under the ambiguity set $\mathcal{F}$, we can approximately transform the DR-CCKP into its deterministic robust counterpart, denoted by RCP-CCKP, as follows:

$$
\text{(3.5a)} \qquad \text{(RCP-CCKP)} \quad Z^* := \max_{x \in \{0,1\}^n} \sum_{i=1}^{n} c_i x_i
$$

$$
\text{(3.5b)} \qquad \text{s.t.}\, u(x, \beta(x)) \leq \alpha,
$$

where $\beta(x) = b - \sum_{i=1}^{n} \bar{a}_i x_i$. Straightforwardly, RCP-CCKP is feasible to DR-CCKP, due to the upper bound of the violation probability.

4. Solution methods for the robust counterpart of CCKP. This section establishes solution methods for the RCP-CCKP. In section 4.1, we decompose the RCP-CCKP into a set of binary programs. In section 4.2, we design an exact solution method for binary programs by using a dynamic programming approach.

4.1. Decomposition into binary programs. The RCP-CCKP cannot be directly handled by the off-the-shelf solvers, as $u(x, \beta(x))$ is highly nonlinear in decision variable x. The following Theorem 4.1 decomposes the RCP-CCKP into a set of binary programs.

THEOREM 4.1. *Solving the RCP-CCKP is equivalent to solving* $\mid W_1 \mid \times \mid W_2 \mid$ *binary programs*

$$
\text{(4.1)} \qquad \textit{(BP-CCKP)} \quad Z^* := \max_{(w_1, w_2) \in W_1 \times W_2} \Omega(w_1, w_2),
$$

where $W_1 = \{\sum_{i=1}^{n} x_i : x \in \{0,1\}^n\}$, $W_2 = \{\sum_{i=1}^{n} \ln(\hat{a}\mu_i) x_i : x \in \{0,1\}^n\}$, *and* $\Omega(w_1, w_2)$ *is defined as*

$$\Omega(w_1, w_2) := \max_{x \in \{0,1\}^n} \sum_{i=1}^{n} c_i x_i \tag{4.2a}$$

$$s.t. \sum_{i=1}^{n} \bar{a}_i x_i + \beta_\alpha(w_1, w_2) \le b, \tag{4.2b}$$

$$\sum_{i=1}^{n} x_i = w_1, \tag{4.2c}$$

$$\sum_{i=1}^{n} \ln(\hat{a}\mu_i) x_i = w_2, \tag{4.2d}$$

and $\beta_\alpha(w_1, w_2)$ *is given by*

$$\beta_\alpha(w_1, w_2) := \min\left\{\beta : \sum_{k=0}^{w_1} (-1)^k C_{w_1}^k \frac{\max\{[w_1 - 2k]\hat{a} - \beta, 0\}^{w_1}}{w_1! e^{w_2}} \le \alpha\right\}. \tag{4.2e}$$

Proof. First, we prove that the optimal solution x^* to the RCP-CCKP is a feasible solution to $\Omega(w_1, w_2)$ for w_1^* and w_2^*, where $w_1^* = \sum_{i=1}^n x_i^*$ and $w_2^* = \sum_{i=1}^n \ln(\hat{a}\mu_i) x_i^*$. Note that x^* satisfies that $u(x^*, \beta(x^*)) \le \alpha$. By the definition of $\beta_\alpha(w_1^*, w_2^*)$, we have $\beta_\alpha(w_1^*, w_2^*) \le \beta(x^*) = b - \sum_{i=1}^n \bar{a}_i x_i^*$. Since $\sum_{i=1}^n \bar{a}_i x_i^* + \beta(x^*) = b$ and $\beta_\alpha(w_1^*, w_2^*) \le \beta(x^*)$, we have $\sum_{i=1}^n \bar{a}_i x_i^* + \beta_\alpha(w_1^*, w_2^*) \le b$, which implies that x^* is a feasible solution to $\Omega(w_1^*, w_2^*)$.

Second, we prove that the optimal solution $\tilde{x}$ to the BP-CCKP is a feasible solution to the RCP-CCKP. Let $\tilde{w}_1 = \sum_{i=1}^n \tilde{x}$ and $\tilde{w}_2 = \sum_{i=1}^n \ln(\hat{a}\mu_i)\tilde{x}_i$. Since $\beta_\alpha(\tilde{w}_1, \tilde{w}_2)$ satisfies that $\sum_{i=1}^n \bar{a}_i \tilde{x}_i + \beta_\alpha(\tilde{w}_1, \tilde{w}_2) \le b$, noticing that $\beta(\tilde{x}) = b - \sum_{i=1}^n \bar{a}_i \tilde{x}_i$, we have $\beta_\alpha(\tilde{w}_1, \tilde{w}_2) \le \beta(\tilde{x})$. Considering that $\sum_{k=0}^{\tilde{w}_1} (-1)^k C_{\tilde{w}_1}^k \frac{\max\{[\tilde{w}_1 - 2k]\hat{a} - \beta, 0\}^{\tilde{w}_1}}{\tilde{w}_1! e^{\tilde{w}_2}}$ is nonincreasing in β, and that $\sum_{k=0}^{\tilde{w}_1} (-1)^k C_{\tilde{w}_1}^k \frac{\max\{[\tilde{w}_1 - 2k]\hat{a} - \beta_\alpha(\tilde{w}_1, \tilde{w}_2), 0\}^{\tilde{w}_1}}{\tilde{w}_1! e^{\tilde{w}_2}} \le \alpha$, we can obtain that $\sum_{k=0}^{\tilde{w}_1} (-1)^k C_{\tilde{w}_1}^k \frac{\max\{[\tilde{w}_1 - 2k]\hat{a} - \beta(\tilde{x}), 0\}^{\tilde{w}_1}}{\tilde{w}_1! e^{\tilde{w}_2}} \le \alpha$, which implies that $u(\tilde{x}, \beta(\tilde{x})) \le \alpha$. Thus, we have $\tilde{x}$ is a feasible solution to the RCP-CCKP, which completes the proof. ∎

Now, we elaborate on computing $\beta_\alpha(w_1, w_2)$ with the binary search approach. First, expression $\sum_{k=0}^{w_1} (-1)^k C_{w_1}^k \frac{\max\{[w_1 - 2k]\hat{a} - \beta, 0\}^{w_1}}{w_1! e^{w_2}}$ is continuous and nonincreasing in β, which takes values in the bounded interval $[0, n\hat{a}]$. Hence, the binary search approach can efficiently optimize $\beta_\alpha(w_1, w_2)$ in (4.2e). Second, we can only approximate this expression with a certain level of accuracy that depends on the computing precision. With this, evaluating this expression may encounter numerical instability, where number values exceed the numerical range allowed by the computing precision. For example, if w_2 exceeds 710, e^{w_2} will surpass the numerical range allowed by the double precision. To tackle this issue, one can simultaneously scale down the numerator and denominator of the expression by multiplying by an appropriate factor. For example, if the factor is $1/(w_1\hat{a})^{w_1}$, the expression is reformulated as $\sum_{k=0}^{w_1} (-1)^k C_{w_1}^k \frac{\max\{1 - (2k\hat{a} + \beta)/(w_1\hat{a}), 0\}^{w_1}}{[\prod_{i=0}^{w_1 - 1}(w_1 - i)](e^{w_2 - w_1})[(e/\hat{a})^{w_1}]}$. Thus, when terms $w_2 - w_1$ and $\hat{a}$ take intermediate values, computing transitional terms such as $e^{w_2 - w_1}$ and $(e/\hat{a})^{w_1}$ is more numerically stable, ensuring guaranteed computing precision.

Then, we analyze the cardinalities of sets W_1 and W_2. The cardinality of W_1 is n, due to the combinatorial nature of x. However, the cardinality of W_2 could be exponentially large, as the coefficient $\ln(\hat{a}\mu_i)$ could be an irrational number. To restrict the problem scale, we introduce a rounding procedure to approximate W_2.

Given a rounding precision $\epsilon \in (0,1]$, we define $r_i' = \epsilon\lceil \ln(\hat{a}\mu_i)/\epsilon\rceil \ \forall\ i \in N$, where $\lceil\cdot\rceil$ is the ceiling function. By replacing $\ln(\hat{a}\mu_i)$ with r_i', we denote $W_2' = \{\sum_{i=1}^n r_i' x_i : x \in \{0,1\}^n\}$.

Instead, we analyze the cardinality of the set W_2'. Let $\mu_{\min}$ (resp., $\mu_{\max}$) denote the minimum (resp., maximum) value of μ_i across $i \in N$. Since r_i' is bounded such that $\epsilon\lceil\ln(\hat{a}\mu_{\min})/\epsilon\rceil \le r_i' \le \epsilon\lceil\ln(\hat{a}\mu_{\max}/\epsilon)\rceil$, the cardinality of W_2' is no greater than $n\max\{\lceil\ln(\hat{a}\mu_{\max})/\epsilon\rceil - \lceil\ln(\hat{a}\mu_{\min})/\epsilon\rceil, 1\}$, which linearly relates to the number of random variable (i.e., n). Specifically, when $\mu_{\max} = \mu_{\min}$, which means all random variables share the same value of MATPDs, then the cardinality of W_2' equals n. Consequently, random variables with less dispersed distributions potentially relate to a lesser MATPD gap (i.e., $\mu_{\max} - \mu_{\min}$), leading to a smaller cardinality of the set W_2'.

It is possible to question whether $|W_1| \times |W_2'|$ is a large number. However, we can demonstrate that, in a medium case where all random variables follow truncated normal distributions, the cardinality of $|W_1| \times |W_2'|$ is at most 2500. For instance, consider a medium-sized problem with $n = 50$, $\epsilon = 0.05$, $\hat{a} = 10$ (i.e., the largest $\hat{a}_i$ is 10), $\mu_{max} = 0.519$ (i.e., the smallest variance of $\tilde{z}_i$ is 3), and $\mu_{min} = 0.502$ (i.e., the largest variance of $\tilde{z}_i$ is 1). Under these conditions, we find that $|W_2'|$ is at most 50, which equals n. As a result, the cardinality of $|W_1| \times |W_2'|$ is indeed 2500.

Finally, we examine the error term of violation probability when replacing W_2 with W_2'. For any given $w_2 \in W_2$, let w_2' denote its rounded value. Then, we have $w_2' - w_2 = \sum_i r_i' x_i - \sum_i \ln(\hat{a}\mu_i)x_i \le w_1\epsilon$ for any $w_1 \in W_1$. After replacing w_2 with w_2' in (4.2d), the error term of violation probability is $e^{w_2' - w_2} \le e^{w_1\epsilon}$. This implies that the violation probability is actually scaled up by a factor of $e^{w_1\epsilon}$, relative to its original level α. As such, the tolerance α in $\Omega(w_1, w_2)$ should be replaced by $\alpha/e^{w_1\epsilon}$, ensuring that the chance constraint still holds at the tolerance α. For example, when $\alpha = 0.05$, $n = 100$, and $\epsilon = 0.01$, to guarantee that the violation probability is not greater than 0.05, the tolerance α in $\Omega(w_1, w_2)$ should be reset accordingly: 0.018 when $w_1 = 100$, 0.03 when $w_1 = 50$, and 0.045 when $w_1 = 10$.

4.2. Exact solution method for each binary program based on dynamic programming. As is well established, the classical knapsack problem admits an exact algorithm based on dynamic programming, which exhibits a pseudopolynomial time complexity of $O(nb)$ [28]. In this section, we develop a dynamic programming approach for the binary program $\Omega(w_1, w_2)$, as presented in Theorem 4.1. For description convenience, we assume that $\ln(\hat{a}\mu_i)\ \forall\ i \in N$ has been rounded to integers, denoted by r_i'. We define (i,j,k,w) as the state variable, where $i \in \{0,1,\dots,n\}$, $j \in \{0,1,\dots,w_1\}$, $k \in \{0,1,\dots,w_2\}$, and $w \in \{0,1,\dots,\lfloor b - \beta_\alpha(w_1,w_2)\rfloor\}$. Now we present the following optimization problem:

$$g(i,j,k,w) = \max_{x_\ell \in \{0,1\}^n} \sum_{\ell=1}^{i} c_i x_i \tag{4.3a}$$

$$\text{s.t.} \sum_{\ell=1}^{i} x_\ell = j, \tag{4.3b}$$

$$\sum_{\ell=1}^{i} r_\ell' x_\ell = k, \tag{4.3c}$$

$$\sum_{\ell=1}^{i} \bar{a}_\ell x_\ell = w. \tag{4.3d}$$

The optimal objective of $\Omega(w_1, w_2)$ can be obtained by $\max\{g(i, j, k, w) : w \leq b - \beta_\alpha(w_1, w_2)\}$. The dynamic programming is stated as follows: The initial condition is $g(0, 0, 0, 0) = 0$ and $g(0, j, k, w) = -\infty$ for any $j \in \{1, 2, \ldots, w_1\}$, $k \in \{1, 2, \ldots, w_2\}$, and $w \in \{1, 2, \ldots, \lfloor b - \beta_\alpha(w_1, w_2)\rfloor\}$; the recursion equation is $g(i, j, k, w) = \max\{g(i-1, j, k, w), g(i-1, j-1, k-r'_i, w-\bar{a}_i)\}$, where $g(i, j, k, w)$ is set to $-\infty$ for any negative attribute j, k, or w.

The dynamic programming approach presented herein can be viewed as a generalization of the cardinality-constrained knapsack problem with the time complexity being $O(n^2 \mid W'_2 \mid)$ (see [11] for more details). Notably, the largest cardinality of $\mid W'_2 \mid$ is bounded by $n \max\{\lceil \ln(\hat{a}\mu_{\max})/\epsilon \rceil - \lceil \ln(\hat{a}\mu_{\min})/\epsilon \rceil, 1\}$, which implies that the time complexity of the proposed dynamic programming is $O(n^3 \max\{\lceil \ln(\hat{a}\mu_{\max})/\epsilon \rceil - \lceil \ln(\hat{a}\mu_{\min})/\epsilon \rceil, 1\})$.

5. Applications and extensions. We have developed distributionally robust approaches for CCKPs with known PDFs. In section 5.1, we apply our proposed robust approach in a data-driven setting when only historical data is available. In section 5.2, we extend the CCKP to the multidimensional CCKP (MD-CCKP) by incorporating multiple chance constraints.

5.1. Applications in a data-driven setting. Let $M := \{1, 2, \ldots, m\}$ denote the set of subscripts for samples $(y_{i,j})_{j \in M}$, which are drawn independently from the uncertain volume $\tilde{a}_i$ for any $i \in N$. We define $y_{i,\min}$ and $y_{i,\max}$ as the minimum and maximum samples, respectively. To avoid underestimating the support interval, we assume that the uncertain volume $\tilde{a}_i$ takes values between $y'_{i,\min} = y_{i,\min} - \delta$ and $y'_{i,\max} = y_{i,\max} + \delta$, where δ takes a positive and small value (e.g., $\delta = 0.01$). Then, the corresponding estimators can be obtained as follows: $\bar{a}_i = (y'_{i,\max} + y'_{i,\min})/2$ and $\hat{a}_i = (y'_{i,\max} - y'_{i,\min})/2$. Subsequently, we define the scaled samples $d_{i,j} = (y_{i,j} - \bar{a}_i)/\hat{a}$ for any $i \in N$ and $j \in M$. As one can observe, the scaled samples $(d_{i,j})_{j \in M}$ are drawn independently from the scaled uncertain volume $\tilde{z}_i$ and take values within the interval $(-1, 1)$.

Following the statistics theory, the empirical CDF of $\tilde{z}_i$ is given by

$$F_i^m(y) := \frac{1}{m} \sum_{j=1}^{m} \mathbb{I}_{(-\infty, y]}(d_{i,j}), \tag{5.1}$$

where the indicator function $\mathbb{I}_{(-\infty,y]}(d_{i,j})$ equals one if $d_{i,j} \in (-\infty, y]$ and zero otherwise. We consider the estimator of the MATPD defined in (2.9). When $h_i(y)$ equals $\int_y^1 f_i(z)dz/(1-y)$, it can be estimated by its empirical form $[1 - F_i^m(y)]/(1-y)$. When $h_i(y)$ equals $f_i(1)$, it can be estimated by $[F_i^m(1) - F_i^m(1-\gamma)]/\gamma$ for a positive and small γ [46]. We can always select a γ that is less than $\delta/\hat{a}_i$, ensuring that $F_i^m(1) - F_i^m(1-\delta) = 0$ and $h_i(1) = 0$. With this, the estimator of the MATPD with m samples is obtained as follows:

$$\frac{1}{\mu_i} = \max_{y \in \{d_{i,1}, d_{i,2}, \ldots, d_{i,m}\}} \frac{1 - F_i^m(y)}{1-y}. \tag{5.2}$$

To examine the stability of various estimating approaches, we further conduct a comparative analysis of three statistics (i.e., the MATPD, the first moment, and the empirical distribution) under limited sample sizes, employing the deviation ratio as the stability indicator.

First, we consider the deviation ratio for the first moment of $\tilde{z}_i$. The sample mean is $y_{i,m} = \sum_{j=1}^{m} d_{i,j}/m$ when the sample size is m and $y_{i,m+1} = \sum_{j=1}^{m+1} d_{i,j}/(m+1)$ when the sample size is $m+1$. The deviation ratio for the first moment equals

$$\frac{y_{i,m+1} - y_{i,m}}{y_{i,m}} = \frac{d_{i,m+1}}{(m+1)y_{i,m}} - \frac{1}{m+1}. \tag{5.3}$$

Intuitively, the deviation ratio tends to increase significantly as $y_{i,m}$ approaches zero, a phenomenon commonly encountered in symmetric distributions (e.g., scaled normal distributions), so that the sample mean can exhibit substantial changes in response to even slight increases in the small sample size. In this sense, the first moment is unstable under limited sample conditions.

Second, we consider the deviation ratio for the empirical distribution. The empirical CDF is $F_i^m(y) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{I}_{(-\infty,y]}(d_{i,j})$ when the sample size is m and $F_i^{m+1}(y) = \frac{1}{1+m}\sum_{j=1}^{m+1} \mathbb{I}_{(-\infty,y]}(d_{i,j})$ when the sample size is $m+1$. We can verify that $F_i^m(y)$ and $F_i^{m+1}(y)$ satisfy that $F_i^{m+1}(y) = \frac{mF_i^m(y)}{m+1}$ for $y \in [-\hat{a}, d_{i,m+1})$ and $F_i^{m+1}(y) = \frac{mF_i^m(y)+1}{m+1}$ for $y \in [d_{i,m+1}, \hat{a}]$. Based on these equations, the deviation ratio for the empirical distribution is

$$\frac{F_i^{m+1}(y) - F_i^m(y)}{F_i^m(y)} = \begin{cases} \dfrac{-1}{m+1} + \dfrac{1}{(m+1)F_i^m(y)}, & z \in [-\hat{a}, d_{i,m+1}), \\ \dfrac{-1}{m+1}, & z \in [d_{i,m+1}, \hat{a}]. \end{cases} \tag{5.4}$$

As can be observed, the deviation ratio tends to be large when $F_i^m(y)$ approaches zero, so that the change of empirical CDFs can be substantial. This suggests that the empirical distribution is also unstable under limited sample conditions.

Third, we consider the deviation ratio for the MATPD. Denote d_{i,s^*} as the optimal solution to $\max_{y\in[-1,1)} \frac{1-F_i^m(y)}{1-y}$, and denote d_{i,t^*} as the optimal solution to $\max_{y\in[-1,1)} \frac{1-F_i^{m+1}(y)}{1-y}$. For notation brevity, we set $\frac{1}{\mu_m} = \frac{1-F_i^m(d_{i,s^*})}{1-d_{i,s^*}}$ and $\frac{1}{\mu_{m+1}} = \frac{1-F_i^{m+1}(d_{i,t^*})}{1-d_{i,t^*}}$. Recall that $d_{i,j} \in (-1,1)$ for any $i \in N$ and $j \in M \cup \{m+1\}$, implying that $0 < 1 - d_{i,j} < 2$. Then, we consider the following four cases:

(1) $d_{i,s^*} < d_{i,m+1}$ and $d_{i,t^*} < d_{i,m+1}$. For this case, it must hold that $d_{is^*} \geq d_{it^*}$ and

$$\begin{aligned} \frac{\mu_m}{(1-d_{i,s^*})(m+1)} - \frac{1}{m+1} \leq \left(\frac{1}{\mu_{m+1}} - \frac{1}{\mu_m}\right) \Big/ \frac{1}{\mu_m} \\ \leq \frac{\mu_m}{(1-d_{i,t^*})(m+1)} - \frac{1}{m+1}. \end{aligned}$$

(2) $d_{i,s^*} \geq d_{i,m+1}$ and $d_{i,t^*} \geq d_{i,m+1}$. For this case, it must hold that

$$\left(\frac{1}{\mu_{m+1}} - \frac{1}{\mu_m}\right) \Big/ \frac{1}{\mu_m} = -\frac{1}{m+1}.$$

(3) $d_{i,t^*} \leq d_{i,m+1} \leq d_{i,s^*}$. For this case, it must hold that

$$-\frac{1}{m+1} \leq \left(\frac{1}{\mu_{m+1}} - \frac{1}{\mu_m}\right) \Big/ \frac{1}{\mu_m} \leq \frac{\mu_m}{(1-d_{i,t^*})(m+1)} - \frac{1}{m+1}.$$

(4) $d_{i,s^*} \leq d_{i,m+1} \leq d_{i,t^*}$. For this case, its condition does not hold.

Compared to the first moment and the empirical distribution, the deviation ratio of the MATPD remains bounded, regardless of the sample size. This observation can be attributed to the fact that the sample is scaled by $d_{i,j} = (y_{i,j} - \bar{a}_i)/\hat{a}$ for any $i \in N$ and $j \in M \cup \{m+1\}$, ensuring that all boundaries in cases (1)–(3) take finite values. For example, in the lower bound of case (1), $1/(1-d_{i,s^*})$ takes a finite value because $d_{i,s^*} \neq 1$. This analysis leads to the conclusion that the MATPD exhibits greater stability than the first moment and empirical distribution under limited sample conditions.

5.2. Extensions to the multidimensional CCKP. The MD-CCKP extends the CCKP by involving multiple chance constraints. According to the CCKP formulation (1.1), the MD-CCKP is formulated as follows:

$$\text{(MD-CCKP)} \max_{x\in\{0,1\}^n} \sum_{j=1}^{n} c_j x_j \tag{5.5a}$$

$$\text{s.t. Prob}\left(\sum_{j=1}^{n} \tilde{a}_{i,j} x_j > b_i\right) \leq \alpha_i, i \in K. \tag{5.5b}$$

Let $K := \{1, 2, \dots, k\}$ denote the set of subscripts for knapsacks. Following the independence assumption for the CCKP, we assume that all random variables in each constraint are independent. For each knapsack $i \in K$, the uncertain volume $\tilde{a}_{i,j}$ has the nominal value as $\bar{a}_{i,j}$ and the largest deviation from its nominal value as $\hat{a}_{i,j}$. For notation brevity, let $\hat{a}_i$ denote the maximum value $\max\{\hat{a}_{i,1}, \hat{a}_{i,2}, \dots, \hat{a}_{i,n}\}$. By introducing the scaled uncertain volume $\tilde{z}_{i,j} = (\tilde{a}_{i,j} - \bar{a}_{i,j})/\hat{a}_i$, we can represent $\tilde{z}_{i,j}$ as an independent random variable distributed within the interval $[-1, 1]$.

Similar to the DR-CCKP, we construct the ambiguity set $\mathcal{F}^i$ for each chance constraint $i \in K$. These ambiguity sets are also induced by the MATPD, as described in (2.11). Then, the distributionally robust MD-CCKP (DR-MD-CCKP) can be formulated as

$$\text{(DR-MD-CCKP)} \max_{x\in\{0,1\}^n} \sum_{j=1}^{n} c_j x_j \tag{5.6a}$$

$$\text{s.t.} \sup_{F^i \in \mathcal{F}^i} \text{Prob}\left(\sum_{j=1}^{n} \tilde{a}_{i,j} x_j > b_i\right) \leq \alpha_i, i \in K. \tag{5.6b}$$

According to Theorem 3.1, we can derive the upper bound $u_i(x, \beta_i(x))$ under the ambiguity set $\mathcal{F}^i$ similarly for each chance constraint $i \in K$. Then, we can approximately transform the DR-MD-CCKP into its deterministic robust counterpart, denoted by RCP-MD-CCKP, as follows:

$$\text{(RCP-MD-CCKP)} \max_{x\in\{0,1\}^n} \sum_{j=1}^{n} c_j x_j \tag{5.7a}$$

$$\text{s.t. } u_i(x, \beta_i(x)) \leq \alpha_i, i \in K, \tag{5.7b}$$

where the upper bound $u_i(x, \beta_i(x))$ is

$$u_i(x, \beta_i(x)) := \sum_{k=0}^{m(x)} (-1)^k C_{m(x)}^k \frac{\max\{[m(x) - 2k]\hat{a}_i - \beta_i(x), 0\}^{m(x)}}{m(x)! \prod_{j=1}^{n} (\hat{a}_i \mu_{i,j})^{x_j}},$$

and $\beta_i(x) := b_i - \sum_{j=1}^{n} \bar{a}_{i,j} x_j$. Here, $C_{m(x)}^k$ denotes the number of combinations of $m(x)$ taken k at a time.

Following Theorem 4.1, the RCP-MD-CCKP can be decomposed into a multitude of binary programs, analogous to the RCP-CCKP. However, the number of binary programs in RCP-MD-CCKP is substantially higher, due to the presence of multiple chance constraints. To overcome this computational complexity, we propose an alternative upper bound to replace the original one $u_i(x, \beta_i(x))$, motivated by finding new bounds on $\sum_{j=1}^{n} \ln(\hat{a}_i \mu_{i,j}) x_j$.

THEOREM 5.1. *Solving the RCP-MD-CCKP is feasible to solve* $\mid W_0 \mid$ *binary programs*

$$\text{(BP-MD-CCKP)} \quad W^* := \max_{w_0 \in W_0} \Gamma(w_0), \tag{5.8}$$

where $W_0 = \{\sum_{j=1}^{n} x_j : x \in \{0,1\}^n\}$ *and* $\Gamma(w_0)$ *is defined as*

$$\Gamma(w_0) := \max_{x \in \{0,1\}^n} \sum_{j=1}^{n} c_j x_j \tag{5.9a}$$

$$\text{s.t.} \sum_{j=1}^{n} \bar{a}_{i,j} x_j + h_i(w_0) \le b_i \;\; \forall i \in K, \tag{5.9b}$$

$$\sum_{j=1}^{n} x_j = w_0, \tag{5.9c}$$

where $h_i(w_0) := \min\{\beta_i : \sum_{k=0}^{w_0} (-1)^k C_{w_0}^k \frac{\max\{[w_0-2k]\hat{a}_i - \beta_i, 0\}^{w_0}}{w_0! e^{B_{i,k}}} \le \alpha_i\}$ *and* $B_{i,k}$ *is*

$$B_{i,k} := \begin{cases} \min_{\{z_j \in [0,1], \sum_{j=1}^{n} z_j = w_0\}} \sum_{j=1}^{n} \ln(\hat{a}_i \mu_{i,j}) z_j & \text{if } k \text{ is even,} \\ \max_{\{z_j \in [0,1], \sum_{j=1}^{n} z_j = w_0\}} \sum_{j=1}^{n} \ln(\hat{a}_i \mu_{i,j}) z_j & \text{if } k \text{ is odd.} \end{cases}$$

Proof. Originally, to solve the RCP-MD-CCKP, the $h_i(w_0)$ in (5.9b) should be replaced by $h_i'(w_0)$, defined as follows:

$$h_i'(w_0) = \min \left\{ \beta_i : \sum_{k=0}^{w_0} (-1)^k C_{w_0}^k \frac{\max\{[w_0 - 2k]\hat{a}_i - \beta_i, 0\}^{w_0}}{w_0! e^{\sum_{j=1}^{n} \ln(\hat{a}_i \mu_{i,j}) x_j}} \le \alpha_i \right\}.$$

According to the definition of $B_{i,k}$, it holds that $\sum_{j=1}^{n} \ln(\hat{a}_i \mu_{i,j}) x_j \ge B_{i,k}$ when k is even and $\sum_{j=1}^{n} \ln(\hat{a}_i \mu_{i,j}) x_j \le B_{i,k}$ when k is odd. As a result, it holds that $h_i'(w_0) \le h_i(w_0)$, implying that the optimal solution of (5.9) is a feasible solution to (5.7). The proof is completed. ∎

6. Numerical studies. In this section, we numerically evaluate the performance of our proposed nonparametric robust optimization approaches for CCKPs. In section 6.1, we describe the experimental setups of the CCKP instances. Then, we elaborate on comparing with alternative robust and stochastic optimization approaches. In section 6.2, we present computational results for CCKPs with a single chance constraint.

6.1. Computational setups. To evaluate the performance of our proposed robust optimization approach, we consider two types of distributions: symmetric and asymmetric distributions. On the one hand, the symmetric distribution comprises uniform distributions with various intervals of support. For brevity, we denote Uniform-E (resp., Uniform-U) as uniform distributions with equal (resp., unequal) interval lengths. In this numerical study, Uniform-E has a homogeneous interval length of 30, while Uniform-U has interval lengths varying from 20 to 30. On the other hand, the asymmetric distribution includes right-skewed distributions with light and heavy tails. Specifically, we use exponential distributions to represent the light-tailed distributions and Pareto distributions to describe the heavy-tailed distributions. Compared to the exponential distribution, the Pareto distribution has a higher probability of extreme values, resulting in a heavier tail on its right side. In this numerical study, both distributions are truncated to bounded intervals with lengths ranging from 20 to 30. The rate parameter of the exponential distributions is set to the centroid of the intervals, while the scale parameter of the Pareto distributions is set to the left endpoint of the intervals.

We compare our RCP-CCKP with robust optimization approaches incorporating partial distributional information (i.e., interval, mean, and variance). Specifically, these robust optimization approaches include classic Ellipsoidal and Budget uncertainty sets proposed by [4] and [8], respectively, and distributionally robust CVaR and EVaR (abbreviated as DCVaR and DEVaR, respectively) proposed by [14] and [35], respectively. Additionally, we apply two ambiguous methods under the mean-variance ambiguity set proposed by [50] and under the ∞-Wasserstein ambiguity set provided by [27] (abbreviated as DCBP and DALSO-X+, respectively). Both ambiguous methods offer tractable approximations for the distributionally robust chance-constrained program. To estimate parameters for each robust optimization approach, we adopt 500 samples for each random volume.

We also compare our RCP-CCKP with stochastic optimization approaches encompassing full distributional information (i.e., empirical distributions). In particular, these full-information models include a tractable approximation called ALSO-X+, which outperforms the CVaR approximation as demonstrated by [27]. Additionally, except for the EVaR approximation, we also appraise a full-information optimization approach by assuming the Gaussian distribution of random volumes, which is proposed by [24]. As stochastic optimization approaches with limited samples generally lack sufficient protection against distributional ambiguity, we particularly focus on their limited sample performances. Thus, we adopt 50 samples for each random volume.

All computations are performed on a workstation with an Intel Core i5-12400 CPU at 2.50 GHz and 16 GB of memory. We implement the optimization approaches and the branch-and-cut algorithm using the commercial solver Gurobi 11.0 via Python 3.9.12. The Gurobi default settings are used to optimize all the binary programs, and we set the number of threads to four. Throughout the numerical studies, the rounding precision is set to 0.1 to maintain one decimal place.

6.2. Computational results with a single chance constraint. We investigate the computational results for CCKPs with a single chance constraint. The number of items (i.e., n) and the capacity of the knapsack (i.e., b) are set to 50 and 400, respectively. The objective coefficients (i.e., c_i) are deterministic integers randomly generated from uniform distribution $U[1, 10]$. The risk tolerance of violation probability (i.e., α) is set to 0.05 and 0.1 to check the sensitivity of our results.

6.2.1. Comparison with robust optimization approaches. This section compares our RCP-CCKP model with existing robust optimization approaches. The Budget and DEVaR models are reformulated to solve a polynomial number of deterministic knapsack problems and optimized using the branch-and-cut algorithm. The uncertainty sets in both models depend on the range of random volumes. Then, the Ellipsoidal, DCBP, and DCVaR models are representable as SOC binary programs, which are often time-consuming to solve due to the binary restrictions on decision variables. When implementing the branch-and-cut algorithm, we utilize the extended polymatroid inequalities, which are derived from the submodularity of the SOC binary constraints, to accelerate the computing procedures (see section 4 in [50]). Furthermore, the DALSO-X+ model resembles a bilevel optimization problem, where the upper-level problem is to find the best objective value, and the lower-level problem is to minimize the expectation of constraint violations (see Algorithm 4 in [27]). In our numerical settings, these alternative models incorporate various distributional information: the Ellipsoidal and Budget models adopt the range and mean information, the DCBP and DCVaR models rely on the mean and variance information, and the DALSO-X+ model sets the Wasserstein radius as the standard deviation. To solve these robust models, we estimate the empirical uncertainty sets from 500 in-sample data points, along with the empirical MATPDs using (5.2).

Tables 1 and 2 report our computational performances on three indicators for various models. The first column (i.e., Obj.) reports the optimal objective value (i.e., Z^*). The percentage in brackets reflects our RCP-CCKP model's relative improvement. The second column (i.e., P%) reports the out-of-sample probability of constraint violation, given the optimal solution (i.e., x^*) calculated on the empirical uncertainty set with in-sample data. When the out-of-sample probability exceeds the risk tolerance of violation probability (i.e., α), we disregard the optimal objective due to its unreliability. The third column reports the CPU time (in seconds). The time limit for computing each instance is set to 1200s, where all instances are solved to global optimality.

Table 1 summarizes the computational performances on symmetric distributions and provides several observations. First, we find that our RCP-CCKP model demonstrates the highest objective value, while the Budget model reports the lowest one. When the risk tolerance equals 0.05, our RCP-CCKP model improves upon the Budget model by 34.7% and 21.3% under Uniform-E and Uniform-U distributions, respectively. The performance of our RCP-CCKP model is even more appealing, which can improve the objective value of the DALSO-X+ model by around 10%–26%. Second, we find that our RCP-CCKP model has better out-of-sample probabilities of constraint violation, which are closer to the risk tolerance. For instance, under the Uniform-E distribution, the out-of-sample probabilities of constraint violation are

Table 1
Comparison with robust optimization approaches on symmetric distributions.

α	Model	Uniform-E			Uniform-U		
		Obj.	P (%)	Time (s)	Obj.	P (%)	Time (s)
0.05	RCP-CCKP	**163** (0.0%)	**0.4**	13.7	**165** (0.0%)	0.0	12.8
0.05	Ellipsoidal	143 (14.0%)	0.0	2.3	157 (5.1%)	0.0	0.5
0.05	Budget	121 (34.7%)	0.0	0.1	133 (24.1%)	0.0	0.1
0.05	DCBP	136 (19.9%)	0.0	16.9	151 (9.3%)	0.0	2.3
0.05	DCVaR	143 (14.0%)	0.0	2.7	156 (5.8%)	0.0	0.7
0.05	DALSO-X+	129 (26.4%)	0.0	8.1	143 (15.4%)	0.0	2.7
0.05	DEVaR	121 (34.7%)	0.0	0.1	136 (21.3%)	0.0	0.6
0.1	RCP-CCKP	**168** (0.0%)	**0.6**	14.9	**168** (0.0%)	0.0	14.2
0.1	Ellipsoidal	150 (12.0%)	0.0	1.8	163 (3.1%)	0.0	0.5
0.1	Budget	121 (38.8%)	0.0	0.1	133 (26.3%)	0.0	0.1
0.1	DCBP	157 (7.0%)	0.0	1.1	168 (0.0%)	0.0	0.4
0.1	DCVaR	163 (3.1%)	0.2	0.4	168 (0.0%)	0.0	1.0
0.1	DALSO-X+	136 (10.3%)	0.0	9.2	143 (14.0%)	0.0	2.5
0.1	DEVaR	121 (38.8%)	0.0	0.2	136 (23.5%)	0.0	0.5

TABLE 2
Comparison with robust optimization approaches on asymmetric distributions.

α	Model	Light-tailed			Heavy-tailed		
		Obj.	P (%)	Time (s)	Obj.	P (%)	Time (s)
0.05	RCP-CCKP	**180** (0.0%)	0.0	6.2	175 (0.0%)	0.0	0.5
0.05	Ellipsoidal	157 (14.6%)	0.0	0.9	157 (11.5%)	0.0	0.6
0.05	Budget	134 (34.3%)	0.0	0.0	135 (29.6%)	0.0	0.1
0.05	DCBP	157 (14.6%)	0.0	1.1	157 (11.5%)	0.0	0.3
0.05	DCVaR	162 (11.1%)	0.0	0.3	161 (8.7%)	0.0	0.3
0.05	DALSO-X+	169 (6.5%)	0.0	4.7	**184** (-4.9%)	0.0	4.7
0.05	DEVaR	136 (32.4%)	0.0	0.5	137 (27.7%)	0.0	0.5
0.1	RCP-CCKP	**182** (0.0%)	0.0	6.4	176 (0.0%)	0.0	0.5
0.1	Ellipsoidal	163 (11.7%)	0.0	0.4	163 (8.0%)	0.0	0.3
0.1	Budget	134 (35.8%)	0.0	0.1	135 (30.4%)	0.0	0.1
0.1	DCBP	171 (6.4%)	0.0	0.7	169 (4.1%)	0.0	0.5
0.1	DCVaR	175 (4.0 %)	0.0	0.2	175 (0.6%)	0.0	0.2
0.1	DALSO-X+	174 (4.6%)	0.0	2.2	**188** (-6.4%)	0.0	5.9
0.1	DEVaR	143 (27.3%)	0.0	0.4	143 (23.1%)	0.0	0.4

0.4% and 0.6% when the risk tolerance equals 0.05 and 0.1, respectively. This observation explains why the RCP-CCKP achieves better objective values. Third, the CPU time of the RCP-CCKP is comparable to that of alternative models, due to the fact that the number of binary programs (i.e., $|W_1| \times |W_2'|$) is not as daunting as it appears. To conclude, our RCP-CCKP outperforms alternative robust optimization approaches with improved conservativeness and tolerable computation times.

Table 2 summarizes the computational performances on asymmetric distributions. We see that the performance of RCP-CCKP and DALSO-X+ models belongs to the first tier. Indeed, the DALSO-X+ model can provide better objective values than our RCP-CCKP model under the heavy-tailed distribution, implying a stronger bound than our $u(x, \beta(x))$ under this setting. Since ALSO-X+ is a better approximation than CVaR, which has been known to be the best for more than a decade, its worst-case version (i.e., DALSO-X+) can provide noninferior objective values as expected. Besides, the DCBP and DCVaR models perform similarly to our RCP-CCKP model. Compared with other alternative models, the DCBP and DCVaR models incorporate more parametric information (i.e., the second moment), which results in comparable robustness to the nonparametric information (i.e., the MATPD). Moreover, we observe that our RCP-CCKP model yields greater improvement under light-tailed distributions than under heavy-tailed ones. For example, when the risk tolerance is set to 0.05, our RCP-CCKP model outperforms the DCBP model by 14.6% and 11.5% under light-tailed and heavy-tailed distributions, respectively. This is intuitively attributed to the fact that heavy-tailed distributions typically require a higher value of MATPD, which compromises the optimality of the solution.

6.2.2. Comparison with stochastic optimization approaches. In this section, we compare our RCP-CCKP with existing stochastic optimization approaches. We employ the sample average approximation (SAA) approach to approximate the hinge-loss function (i.e., $\mathbb{E}[\cdot]_+$) exploited by CVaR and ALSO-X+ models and the moment generating function employed by the EVaR model. Then, we implement ALSO-X+ using Algorithm 4 in [27], whereas we solve CVaR and EVaR using Gurobi directly. Additionally, we formulate the Gaussian model as an SOC binary program, assuming a Gaussian distribution of random volumes. Similarly to the Ellipsoidal and DCVaR models, we utilize valid inequalities for the submodular SOC binary

TABLE 3
Comparison with stochastic optimization approaches on symmetric distributions.

α	Model	Uniform-E			Uniform-U		
		Obj.	P (%)	Time (s)	Obj.	P (%)	Time (s)
0.05	RCP-CCKP	151 (0.0%)	0.0	24.6	156 (0.0%)	0.0	22.1
0.05	Gaussian	**181** (-16.6%)	**5.0**	0.3	**187** (-16.6%)	**4.0**	0.3
0.05	CVaR	-	8.8	0.2	-	11.6	0.1
0.05	ALSO-X+	-	17.4	0.5	-	18.6	0.5
0.05	EVaR	121 (24.8%)	0.0	1.0	136 (14.7%)	0.0	0.5
0.1	RCP-CCKP	154 (0.0%)	0.0	23.6	**157** (0.0%)	0.0	24.3
0.1	Gaussian	-	10.6	0.2	-	12.0	0.2
0.1	CVaR	**182** (-15.4%)	**8.8**	0.2	-	11.6	0.1
0.1	ALSO-X+	-	17.2	0.7	-	23.4	0.3
0.1	EVaR	128 (20.3%)	0.0	1.1	136 (15.4%)	0.0	0.5

constraints, and implement the branch-and-cut algorithm. To solve these stochastic models, we estimate the empirical distribution of random volumes from 50 in-sample data points, as well as the empirical MATPD using (5.2). We disregard the optimal objective (denoted by $-$) when the out-of-sample probability exceeds the risk tolerance.

Table 3 summarizes the computational performance of stochastic optimization approaches on symmetric distributions. Compared with Gaussian, CVaR, and ALSO-X+ models, although our RCP-CCKP model does not always yield the highest objective value, its probability of constraint violation is always less than the risk tolerance. The Gaussian model is unreliable when the underlying distribution deviates from the Gaussian assumption, regardless of sample size. Our observation is also consistent with the SAA approach, which suggests that CVaR and ALSO-X+ models may be treacherous under limited sample sizes. We conduct additional experiments with the same computational setups, finding that CVaR and ALSO-X+ models converge when the sample size increases to more than 100. Additionally, compared with EVaR, the performance of our RCP-CCKP model is even more appealing, as it can improve the solution quality by around 15%–25%. This improvement stems from the better approximation of the violation probability, as the Chernoff inequality adopted by the EVaR model lacks tightness for most distributions.

In terms of computational time, all instances are solved within seconds. Although our RCP-CCKP model is more time-consuming, its computational complexity may be significantly lower than that of CVaR and ALSO-X+ models. Following the same computational setups with 200 random volumes, we conduct additional experiments and find that CVaR, ALSO-X+, and RCP-CCKP models take 123.3s, 380.1s, and 30.8s, respectively. In conclusion, our RCP-CCKP model can provide necessary protection against data uncertainty when limited sample sizes are available.

Table 4 reveals that our RCP-CCKP model has a more stable out-of-sample performance than the Gaussian, CVaR, and ALSO-X+ models, which may backfire under the light-tailed distributions. This observation echoes the analysis of various statistics in section 5.1, which states that the MATPD (i.e., the RCP-CCKP model) is more stable than the first moment (i.e., the Gaussian model) and the empirical distribution (i.e., CVaR and ALSO-X+ models) under the limited sample conditions. As mentioned earlier, the Gaussian model is only reliable when the underlying assumption holds, and empirical (i.e., CVaR and ALSO-X+) models converge when the sample size increases (to more than 100 in our experiments). This suggests that our RCP-CCKP model is generally preferable, when limited sample sizes are available.

TABLE 4
Comparison with stochastic optimization approaches on asymmetric distributions.

α	Model	Light-tailed			Heavy-tailed		
		Obj.	P (%)	Time (s)	Obj.	P (%)	Time (s)
0.05	RCP-CCKP	**174** (0.0%)	0.0	15.3	179 (0.0%)	0.0	15.8
0.05	Gaussian	-	7.2	0.2	**251** (-28.7%)	**3.8**	0.2
0.05	CVaR	-	10.2	0.1	242 (-26.0%)	3.4	0.1
0.05	ALSO-X+	-	10.2	0.3	-	5.2	0.2
0.05	EVaR	143 (21.7%)	0.0	0.5	143 (25.2%)	0.0	0.5
0.1	RCP-CCKP	**175** (0.0%)	0.0	17.4	187 (0.0%)	0.0	6.3
0.1	Gaussian	-	13.6	0.2	**257** (-27.2%)	**8.4**	0.4
0.1	CVaR	-	10.2	0.1	247 (-24.3%)	2.0	0.1
0.1	ALSO-X+	-	11.2	0.4	-	15.4	0.2
0.1	EVaR	148 (18.2%)	0.0	0.6	149 (25.5%)	0.0	0.5

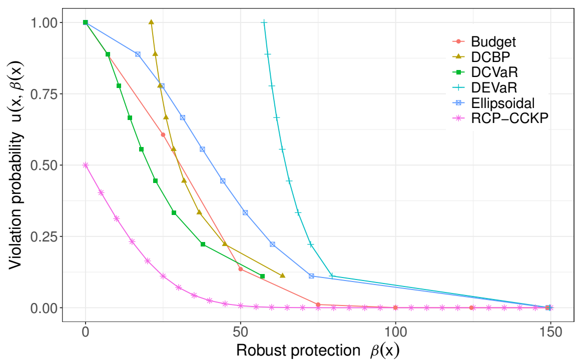

FIG. 1. *Approximated violation probabilities under various robust optimization approaches.*

6.2.3. Comparison on the approximations of chance constraint violation. This section evaluates the approximation of violation probability embedded in robust optimization approaches. Theorem 3.1 provides a novel upper bound $u(x,\beta(x))$ to the violation probability $\text{Prob}\big(\sum_{i=1}^n(\hat{a}\tilde{z}_i)x_i > \beta(x)\big)$. We refer to $\beta(x)$ as the robust protection, as it aims to protect the chance constraint against potential violations. The upper bound $u(x,\beta(x))$ is contingent upon the underlying ambiguity (or uncertainty) set $\mathcal{F}$. For example, the Ellipsoidal model is characterized by $u(x,\beta(x)) = \exp(-\omega^2/2)$, where ω is a given nonnegative parameter determined by $\beta(x) = \omega\sqrt{\sum_{i=1}^n(\hat{a}_i x_i)^2}$ (see [4]). Interested readers are referred to [8, 14, 35, 50] for other comparable models.

Figure 1 illustrates the approximated violation probabilities by varying the robust protection under various robust optimization approaches. The violation probabilities are simulated based on 50 items and Uniform-E distribution with interval $[0,10]$. As shown in Figure 1, the approximated violation probability in our RCP-CCKP model is approximately 0.5 when the robust protection is zero and decreases exponentially thereafter. Given that the approximation in the RCP-CCKP model is tight for the chance constraint with the Uniform-E distribution by Remark 3.4, the remaining models evaluated are all conservative approximations. Consequently, these conservative approximations exhibit a notable gap compared with our RCP-CCKP model.

7. Conclusions. For the knapsack problem under uncertainty, apart from the chance constraint, other distributionally robust optimization approaches are introduced to cope with the uncertainty. For example, [6] addressed the problem of evaluating the expected optimal objective value of a binary program under uncertainty. For other distributionally robust optimization studies, readers may refer to [47], [22], and [13] as examples. Compared with other distributionally robust optimization methods, the chance constraint has important applications, particularly with the "stop-loss" constraint in the risk management of portfolio selection or project selection.

In future studies, we will further enhance the performance of the proposed method by improving the upper bounds of the violation probability. Following this direction, we will investigate how to utilize additional structural information, such as monotonicity and convexity. Specifically, we can divide the support for random variables into mutually exclusive partitions. With monotonicity information, we can divide the support into partitions where the probability density function (PDF) is either nonincreasing or nondecreasing. With convexity information, each monotone partition can be further refined into concave or convex subpartitions. We can approximate the PDF by using a piecewise linear proxy, such that the provided proxy is stochastically dominated by the densities. However, the explicit form for the upper bound on the violation probability can hardly be expected. Instead, based on the distribution ambiguity set with support and moment information provided in [35], we can incorporate the piecewise linear proxy into the distribution ambiguity set to narrow down the distribution ambiguity set and improve the performance under EVaR. This also presents research opportunities to extend our work to continuous optimization problems.

Acknowledgments. The authors are grateful to the editor-in-chief, the associated editor, and two anonymous reviewers for their valuable comments.

REFERENCES

[1] A. Abdi and R. Fukasawa, *On the mixing set with a knapsack constraint*, Math. Program., 157 (2016), pp. 191–217, https://doi.org/10.1007/s10107-016-0979-5.

[2] A. Ahmadi-Javid, *Entropic value-at-risk: A new coherent risk measure*, J. Optim. Theory Appl., 155 (2012), pp. 1105–1123, https://doi.org/10.1007/s10957-011-9968-2.

[3] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski, *Adjustable robust solutions of uncertain linear programs*, Math. Program., 99 (2004), pp. 351–376, https://doi.org/10.1007/s10107-003-0454-y.

[4] A. Ben-Tal and A. Nemirovski, *Robust convex optimization*, Math. Oper. Res., 23 (1998), pp. 769–805, https://doi.org/10.1287/moor.23.4.769.

[5] A. Ben-Tal and A. Nemirovski, *Robust solutions of linear programming problems contaminated with uncertain data*, Math. Program., 88 (2000), pp. 411–424, https://doi.org/10.1007/PL00011380.

[6] D. Bertsimas, K. Natarajan, and C.-P. Teo, *Probabilistic combinatorial optimization: Moments, semidefinite programming, and asymptotic bounds*, SIAM J. Optim., 15 (2004), pp. 185–209, https://doi.org/10.1137/S1052623403430610.

[7] D. Bertsimas, D. Pachamanova, and M. Sim, *Robust linear optimization under general norms*, Oper. Res. Lett., 32 (2004), pp. 510–516, https://doi.org/10.1016/j.orl.2003.12.007.

[8] D. Bertsimas and M. Sim, *Robust discrete optimization and network flows*, Math. Program., 98 (2003), pp. 49–71, https://doi.org/10.1007/s10107-003-0396-4.

[9] D. Bertsimas and M. Sim, *The price of robustness*, Oper. Res., 52 (2004), pp. 35–53, https://doi.org/10.1287/opre.1030.0065.

[10] G. C. Calafiore and L. El Ghaoui, *On distributionally robust chance-constrained linear programs*, J. Optim. Theory Appl., 130 (2006), pp. 1–22, https://doi.org/10.1007/s10957-006-9084-x.

[11] A. Caprara, H. Kellerer, U. Pferschy, and D. Pisinger, *Approximation algorithms for knapsack problems with cardinality constraints*, European J. Oper. Res., 123 (2000), pp. 333–345, https://doi.org/10.1016/S0377-2217(99)00261-1.

[12] A. Caprara, D. Pisinger, and P. Toth, *Exact solution of the quadratic knapsack problem*, INFORMS J. Comput., 11 (1999), pp. 125–137, https://doi.org/10.1287/ijoc.11.2.125.
[13] W. Chen and M. Sim, *Goal-driven optimization*, Oper. Res., 57 (2009), pp. 342–357, https://doi.org/10.1287/opre.1080.0570.
[14] W. Chen, M. Sim, J. Sun, and C.-P. Teo, *From CVaR to uncertainty set: Implications in joint chance-constrained optimization*, Oper. Res., 58 (2010), pp. 470–485, https://doi.org/10.1287/opre.1090.0712.
[15] X. Chen, M. Sim, and P. Sun, *A robust optimization perspective on stochastic programming*, Oper. Res., 55 (2007), pp. 1058–1071, https://doi.org/10.1287/opre.1070.0441.
[16] J. Cheng, E. Delage, and A. Lisser, *Distributionally robust stochastic knapsack problem*, SIAM J. Optim., 24 (2014), pp. 1485–1506, https://doi.org/10.1137/130915315.
[17] I. R. De Farias Jr and G. L. Nemhauser, *A polyhedral study of the cardinality constrained knapsack problem*, Math. Program., 96 (2003), pp. 439–467, https://doi.org/10.1007/s10107-003-0420-8.
[18] A. Dixon, *The second mean value theorem in the integral calculus*, Math. Proc. Cambridge Philos. Soc., 25 (1929), pp. 282–284.
[19] T. Fahle, U. Junker, S. E. Karisch, N. Kohl, M. Sellmann, and B. Vaaben, *Constraint programming based column generation for crew assignment*, J. Heuristics, 8 (2002), pp. 59–81, https://doi.org/10.1023/A:1013613701606.
[20] F. D. Fomeni and A. N. Letchford, *A dynamic programming heuristic for the quadratic knapsack problem*, INFORMS J. Comput., 26 (2014), pp. 173–182, https://doi.org/10.1287/ijoc.2013.0555.
[21] S. Ghosal and W. Wiesemann, *The distributionally robust chance-constrained vehicle routing problem*, Oper. Res., 68 (2020), pp. 716–732, https://doi.org/10.1287/opre.2019.1924.
[22] J. Goh and M. Sim, *Distributionally robust optimization and its tractable approximations*, Oper. Res., 58 (2010), pp. 902–917, https://doi.org/10.1287/opre.1090.0795.
[23] V. Goyal and R. Ravi, *A PTAS for the chance-constrained knapsack problem with random item sizes*, Oper. Res. Lett., 38 (2010), pp. 161–164, https://doi.org/10.1016/j.orl.2010.01.003.
[24] J. Han, K. Lee, C. Lee, K.-S. Choi, and S. Park, *Robust optimization approach for a chance-constrained binary knapsack problem*, Math. Program., 157 (2016), pp. 277–296, https://doi.org/10.1007/s10107-015-0931-0.
[25] L. J. Hong, Y. Yang, and L. Zhang, *Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach*, Oper. Res., 59 (2011), pp. 617–630, https://doi.org/10.1287/opre.1100.0910.
[26] H. Iida, *A note on the max-min* 0-1 *knapsack problem*, J. Comb. Optim., 3 (1999), pp. 89–94, https://doi.org/10.1023/A:1009821323279.
[27] N. Jiang and W. Xie, *ALSO-X and ALSO-X+: Better convex approximations for chance constrained programs*, Oper. Res., 70 (2022), pp. 3581–3600, https://doi.org/10.1287/opre.2021.2225.
[28] H. Kellerer, U. Pferschy, and D. Pisinger, *Introduction to np-completeness of knapsack problems*, in Knapsack Problems, Springer, New York, 2004, pp. 483–493.
[29] H. Kellerer, U. Pferschy, and D. Pisinger, *Multiple knapsack problems*, in Knapsack Problems, Springer, New York, 2004, pp. 285–316.
[30] A. J. Kleywegt and J. D. Papastavrou, *The dynamic and stochastic knapsack problem with random sized items*, Oper. Res., 49 (2001), pp. 26–41, https://doi.org/10.1287/opre.49.1.26.11185.
[31] A. J. Kleywegt, A. Shapiro, and T. Homem-de-Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM J. Optim., 12 (2002), pp. 479–502, https://doi.org/10.1137/S1052623499363220.
[32] O. Klopfenstein and D. Nace, *A robust approach to the chance-constrained knapsack problem*, Oper. Res. Lett., 36 (2008), pp. 628–632, https://doi.org/10.1016/j.orl.2008.03.006.
[33] P. Kouvelis and G. Yu, *Robust Discrete Optimization and Its Applications*, Nonconvex Optim. Appl. 14, Springer, New York, 2013.
[34] A. Lodi and M. Monaci, *Integer linear programming models for* 2*-staged two-dimensional knapsack problems*, Math. Program., 94 (2003), pp. 257–278, https://doi.org/10.1007/s10107-002-0319-9.
[35] D. Z. Long and J. Qi, *Distributionally robust discrete optimization with entropic value-at-risk*, Oper. Res. Lett., 42 (2014), pp. 532–538, https://doi.org/10.1016/j.orl.2014.09.004.
[36] L. A. Lorena and E. L. Senne, *A column generation approach to capacitated* p*-median problems*, Comput. Oper. Res., 31 (2004), pp. 863–876, https://doi.org/10.1016/S0305-0548(03)00039-X.

[37] S. Martello, *Knapsack Problems: Algorithms and Computer Implementations*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, New York, 1990.
[38] M. Monaci and U. Pferschy, *On the robust knapsack problem*, SIAM J. Optim., 23 (2013), pp. 1956–1982, https://doi.org/10.1137/120880355.
[39] J. J. Moré and S. A. Vavasis, *On the solution of concave knapsack problems*, Math. Program., 49 (1990), pp. 397–411, https://doi.org/10.1007/BF01588800.
[40] K. Natarajan, D. Pachamanova, and M. Sim, *Incorporating asymmetric distributional information in robust value-at-risk optimization*, Manage. Sci., 54 (2008), pp. 573–585, https://doi.org/10.1287/mnsc.1070.0769.
[41] W. D. Pisinger, A. B. Rasmussen, and R. Sandvik, *Solution of large quadratic knapsack problems through aggressive reduction*, INFORMS J. Comput., 19 (2007), pp. 280–290, https://doi.org/10.1287/ijoc.1050.0172.
[42] R. T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, J. Risk, 2 (2000), pp. 21–42, https://doi.org/10.21314/JOR.2000.038.
[43] S. Sarykalin, G. Serraino, and S. Uryasev, *Value-at-risk vs. Conditional Value-at-risk in Risk Management and Optimization*, Tutorials in Operations Research, INFORMS, 2008, pp. 270–294, https://doi.org/10.1287/educ.1080.0052.
[44] M. Savelsbergh, *A branch-and-price algorithm for the generalized assignment problem*, Oper. Res., 45 (1997), pp. 831–841, https://doi.org/10.1287/opre.45.6.831.
[45] S. Walukiewicz, *Integer Programming*, Math. Appl. 46, Springer, New York, 2013.
[46] M. Waterman and D. Whiteman, *Estimation of probability densities by empirical density functions*, Int. J. Math. Educat. Sci. Technol., 9 (1978), pp. 127–137, https://doi.org/10.1080/0020739780090201.
[47] W. Wiesemann, D. Kuhn, and M. Sim, *Distributionally robust convex optimization*, Oper. Res., 62 (2014), pp. 1358–1376, https://doi.org/10.1287/opre.2014.1314.
[48] W. Xie and S. Ahmed, *On deterministic reformulations of distributionally robust joint chance constrained optimization problems*, SIAM J. Optim., 28 (2018), pp. 1151–1182, https://doi.org/10.1137/16M1094725.
[49] L. Xu, Y. Zheng, and L. Jiang, *A robust data-driven approach for the newsvendor problem with nonparametric information*, Manuf. Service Oper. Manage., 24 (2022), pp. 504–523, https://doi.org/10.1287/msom.2020.0961.
[50] Y. Zhang, R. Jiang, and S. Shen, *Ambiguous chance-constrained binary programs under mean-covariance information*, SIAM J. Optim., 28 (2018), pp. 2922–2944, https://doi.org/10.1137/17M1158707.
[51] Z. Zhang, C. Zhang, Q.-C. He, and P. Wang, *Robust integrated planning for LEO satellite network design and service operations*, Oper. Res. Lett., 51 (2023), pp. 575–582, https://doi.org/10.1016/j.orl.2023.09.005.