

Article

Computable Reformulation of Data-Driven Distributionally Robust Chance Constraints: Validated by Solution of Capacitated Lot-Sizing Problems

Hua Deng ^{1,2}  and Zhong Wan ^{3,*} ¹ School of Mathematics and Statistics, Central South University, Changsha 410083, China; huakate@163.com² School of Mathematics and Finance, Hunan University of Humanities, Science and Technology, Loudi 417000, China³ International Business School, Hunan University of Information Technology, Changsha 410151, China

* Correspondence: wanzhong@hnuit.edu.cn

Abstract

Uncertainty in optimization models often causes awkward properties in their deterministic equivalent formulations (DEFs), even for simple linear models. Chance-constrained programming is a reasonable tool for handling optimization problems with random parameters in objective functions and constraints, but it assumes that the distribution of these random parameters is known, and its DEF is often associated with the complicated computation of multiple integrals, hence impeding its extensive applications. In this paper, for optimization models with chance constraints, the historical data of random model parameters are first exploited to construct an adaptive approximate density function by incorporating piecewise linear interpolation into the well-known histogram method, so as to remove the assumption of a known distribution. Then, in view of this estimation, a novel confidence set only involving finitely many variables is constructed to depict all the potential distributions for the random parameters, and a computable reformulation of data-driven distributionally robust chance constraints is proposed. By virtue of such a confidence set, it is proven that the deterministic equivalent constraints are reformulated as several ordinary constraints in line with the principles of the distributionally robust optimization approach, without the need to solve complicated semi-definite programming problems, compute multiple integrals, or solve additional auxiliary optimization problems, as done in existing works. The proposed method is further validated by the solution of the stochastic multiperiod capacitated lot-sizing problem, and the numerical results demonstrate that: (1) The proposed method can significantly reduce the computational time needed to find a robust optimal production strategy compared with similar ones in the literature; (2) The optimal production strategy provided by our method can maintain moderate conservatism, i.e., it has the ability to achieve a better trade-off between cost-effectiveness and robustness than existing methods.



Academic Editor: Andrea Scozzari

Received: 8 December 2025

Revised: 10 January 2026

Accepted: 16 January 2026

Published: 19 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: lot-sizing problem; data-driven modeling; chance-constrained programming; scientific computing; distributionally robust optimization

MSC: 90B36; 68U01

1. Introduction

Chance-constrained optimization is a popular way to handle optimization models with random parameters, possessing significant advantages over the classical expectation-

based approaches [1]. In this paper, we consider the following chance-constrained optimization problem:

$$\min C(x) \quad (1a)$$

$$\text{s.t. } \mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha, \quad (1b)$$

$$x \in X \subseteq R^n, \quad (1c)$$

where $x \in R^n$ is the decision vector, $C : R^n \rightarrow R$ is a real-valued convex continuous objective function, $X \subset R^n$ denotes a computable bounded closed convex set, ξ is a random model parameter causing uncertainty in the coefficient matrix $A(\xi) \in R^{m \times n}$ and the capacity vector $b(\xi) \in R^m$, $\mathbb{P}\{\cdot\}$ is a probability operator of random events when the probability distributional function of ξ is $\mathbb{P}$, and $\alpha \in (0, 1)$ represents a given risk level, i.e., the tolerance of constraint violation permitted by the decision-maker.

Owing to the practical difficulties in finding the true distribution of ξ , data-driven distributionally robust chance-constrained programming (DRCCP) has been developed, which replaces the single distribution in the chance constraint (1b) with a group of potential distributions in a confidence set $\mathcal{D}$ [2–6]. Specifically, by collecting sufficient observed historical data of ξ , the so-called confidence set $\mathcal{D}$ is first constructed to depict the domain of all approximate probability distributional functions of ξ . Then, in the realm of distributionally robust optimization, Problem (1) is approximated by the following DRCCP model:

$$\min C(x) \quad (2a)$$

$$\text{s.t. } \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha, \quad (2b)$$

$$x \in X \subseteq R^n. \quad (2c)$$

Clearly, in order to guarantee the applicability of the above data-driven DRCCP approach, it is crucial to address how to choose an appropriate confidence set, especially paying attention to the computational tractability of the constraint (2b) [7]. For example, by using the first and second moments of the data [2,3], moment-based confidence sets were defined. Moment-based sets, confined by estimates of mean and covariance, often lead to tractable reformulations but may overlook crucial distributional shapes, potentially resulting in overly conservative decisions. In order to incorporate higher-order distributional information of ξ , proper divergence-based confidence sets were constructed [6,8,9]. These confidence sets typically comprised all potential probability distributions within a specified tolerance of a reference distribution (e.g., the empirical distribution), where the tolerance is quantified by a statistical divergence such as the Kullback–Leibler (KL) divergence [10,11] or Wasserstein metric [6,12]. However, the reformulation of the constraint (2b) in existing works typically involves the solution of complicated semi-definite programming (SDP) problems based on the theory of functional extrema and duality [6]. For instance, Nguyen et al. derived an equivalent SDP reformulation of the problem via an analytical formula for the Wasserstein distance between two normal distributions [13,14]. In addition, the distributionally robust chance constraints can be reformulated into tractable mixed-integer programs (MIPs) [15,16], but solving this MIP reformulation is also difficult, even with the most advanced solvers [17]. For large-scale or complex instances, the use of data-driven methods can be beneficial for improvements in solution efficiency [18].

Motivated by the need to build a more appropriate confidence set from the collected data with a complex distribution, we intend to first construct an adaptive approximate density function by incorporating piecewise linear interpolation into the well-known histogram method, so as to remove the assumption of known distributions. Then, using this estimated distribution, we define a confidence set only involving finitely many variables,

and, by the principles of distributionally robust optimization, we reformulate the original DRCCP into a computable standard optimization problem. We summarize the main contributions of this research as follows.

1. This is the first time that an adaptive probability density estimation is constructed by optimizing the bin widths in line with the distributional structure of the data, rather than using a uniform width as in the classic estimation method. In particular, for smaller fluctuations in the data, wider bin widths are chosen. For greater fluctuations in the data, smaller bin widths are used. An advantage of doing so is that it can reduce the number of used nodes when defining confidence sets and is beneficial for the dimensionality reduction of the derived robust optimization models.

2. Through variation distances, a confidence set to depict all the approximate probability density functions is constructed, which can strictly restrict the global difference between the confidence sets and the estimated discrete nominal distribution, avoiding the loss of high-order statistical information caused by moment constraints. Owing to the advantages of such a confidence set, it will be proven that the complicated chance constraint (2b) can be reformulated as an easily computable ordinary inequality constraint, which is beneficial for the development of efficient algorithms to solve the original problem (2).

3. The proposed reformulation technique is applied to solve the stochastic multiperiod capacitated lot-sizing problem (SMP-CLSP), and its advantages are validated, especially in comparison with similar ones in the literature.

The remainder of this paper is structured as follows. Section 2 introduces data-driven adaptive confidence sets with finitely many parameters, including the adaptive estimation of probability density functions and the construction of the corresponding confidence sets. Section 3 presents the reformulation of data-driven distributionally robust chance constraints. Section 4 provides numerical validation when applying the proposed computational technique to the SMP-CLSP. This research is concluded in Section 5.

2. Data-Driven Adaptive Confidence Sets with Finite Parameters

2.1. Adaptive Estimation of Probability Density Functions

Instead of fixing the bin widths when obtaining the reference distribution from the data, we first propose a numerical method to adaptively estimate the probability density function directly from the collected data by incorporating a piecewise linear interpolation technique into the well-known histogram method, where the distributional structure of the raw sample data is employed to optimize the choice of bin widths.

Let $\Xi = \{\xi^1, \xi^2, \dots, \xi^{n_s}\}$ represent a sample set of size n_s , where all the samples are independently and identically distributed (i.i.d.) realizations of a random variable ξ , governed by an unknown probability density function f . We define the sample range by the smallest and largest realizations in the sample set, denoted as $\xi^{\min} = \min_{1 \leq j \leq n_s} \xi^j$ and $\xi^{\max} = \max_{1 \leq j \leq n_s} \xi^j$, respectively. Like the classical histogram method [19], we partition $[\xi^{\min}, \xi^{\max}]$ into N subintervals $[z_i, z_{i+1})$, being referred to as N bins: $B_1, B_2, \dots, B_N$. $h_i = z_{i+1} - z_i$ is called the width of the i -th bin. For each bin B_i , let v_i denote the number of samples that fall into the bin B_i . By the Sturges' number-of-bins rule, the initial number of bins is chosen to be $N = \lceil \log_2 n_s + 1 \rceil$, where $\lceil \cdot \rceil$ represents the upper rounding-off operation. Then, by the classical histogram method, the reference density function [20] $\hat{f}_n : R \rightarrow [0, +\infty)$ is defined to be

$$\hat{f}_n(z) = \begin{cases} \frac{v_i}{n_s h_i} = \frac{1}{n_s h_i} \sum_{j=1}^{n_s} I_{[z_i, z_{i+1})}(\xi^j), & z \in B_i, i = 1, \dots, N-1; \\ \frac{v_N}{n_s h_N} = \frac{1}{n_s h_N} \sum_{j=1}^{n_s} I_{[z_N, z_{N+1})}(\xi^j), & z \in B_N, \end{cases} \quad (3)$$

where

$$I_{B_i}(z) = \begin{cases} 1, & \text{if } z \in B_i; \\ 0, & \text{Otherwise.} \end{cases}$$

Numerous studies have shown that the selection of the bin width plays a crucial role in accurately estimating the probability density [21]. In particular, for sample data with a complex distributional structure, the fixed bin width in the classical histogram method may greatly restrict the ability of the estimated density function to depict the true distribution [20]. In order to overcome this difficulty, we now present a numerical method to adaptively estimate the probability density function directly from the collected data by incorporating the piecewise linear interpolation technique into the well-known histogram method [19]. Specifically, we first define a new function $\hat{f} : [z_1, z_{N+1}] \rightarrow [0, +\infty)$ by

$$\hat{f}(z) = \begin{cases} \frac{2(\hat{f}_n(\eta_{i+1}) - \hat{f}_n(\eta_i))}{h_i + h_{i+1}}(z - \eta_i) + \hat{f}_n(\eta_i), & z_i \leq z < z_{i+1}, i = 1, 2, \dots, N-1; \\ \frac{h_N + \eta_N - z}{h_N} \hat{f}_n(\eta_N), & z_N \leq z \leq z_{N+1}, \end{cases} \quad (4)$$

where $\eta_i = \frac{z_i + z_{i+1}}{2}$, $i = 1, 2, \dots, N$. Then, in order to ensure the nonnegativity of $\hat{f}$ and achieve a better approximation, we compute

$$\begin{cases} h_i = z_{i+1} - z_i, i = 1, 2, \dots, N, \\ \eta_i = \frac{z_i + z_{i+1}}{2}, i = 1, 2, \dots, N, \\ slp_i(\hat{f}) = \frac{2(\hat{f}_n(\eta_{i+1}) - \hat{f}_n(\eta_i))}{h_i + h_{i+1}}, i = 1, 2, \dots, N-1, \\ \hat{f}[\eta_i, \eta_{i+1}, \eta_{i+2}] = \frac{4}{h_{i+1} + h_i} \left| \frac{\hat{f}_n(\eta_{i+2}) - \hat{f}_n(\eta_{i+1})}{h_{i+1} + h_{i+2}} - \frac{\hat{f}_n(\eta_{i+1}) - \hat{f}_n(\eta_i)}{h_{i+1} + h_i} \right|, \\ \quad i = 1, 2, \dots, N-2. \\ \hat{f}[\eta_{N-1}, \eta_N, z_{N+1}] = \frac{2}{h_{N-1} + h_N} \left| \frac{\hat{f}_n(\eta_N)}{h_N} - \frac{2(\hat{f}_n(\eta_N) - \hat{f}_n(\eta_{N-1}))}{h_N + h_{N-1}} \right|, \\ \Delta_{i_{max}}^2 = \max \{ \hat{f}[\eta_i, \eta_{i+1}, \eta_{i+2}], \hat{f}[\eta_{N-1}, \eta_N, z_{N+1}], i = 1, 2, \dots, N-2 \}. \end{cases}$$

It can be seen that $\hat{f}[\eta_i, \eta_{i+1}, \eta_{i+2}]$ and $\Delta_{i_{max}}^2$ reflect the information of the second-order forward difference quotient and the maximum second-order difference quotient, respectively. With this first and second-order information, we further optimize the bin widths by inserting more nodes into the bins in line with the following rules.

- (I) When $slp_i(\hat{f}) > \max \left\{ 0, \frac{2\hat{f}_n(\eta_i)}{h_i} \right\}$, a new node $z'_{i+2} = \frac{z_{i+1} + z_{i+2}}{2}$ is inserted into the $(i+1)$ -th bin. The original $(i+1)$ -th bin B_{i+1} is split into two new sub-bins: $B'_{i+1} = [z_{i+1}, z'_{i+2})$, $B'_{i+2} = [z'_{i+2}, z_{i+2})$. The midpoints of the new sub-bins B'_{i+1} and B'_{i+2} are $\eta'_{i+1} = \frac{z_{i+1} + z'_{i+2}}{2}$ and $\eta'_{i+2} = \frac{z'_{i+2} + z_{i+2}}{2}$, respectively. Apply (3) to obtain the corresponding $\hat{f}_n(\eta'_{i+1})$ and $\hat{f}_n(\eta'_{i+2})$ for the two new sub-bins. Consequently, the updated midpoints are given by $(\eta'_{i+1}, \hat{f}_n(\eta'_{i+1}))$ and $(\eta'_{i+2}, \hat{f}_n(\eta'_{i+2}))$, respectively. By rearranging the indices of all obtained bins, we insert more nodes till the inequality $slp_i(\hat{f}) \leq \frac{2\hat{f}_n(\eta_i)}{h_i}$ holds.
- (II) When $slp_i(\hat{f}) < \min \left\{ 0, \frac{-2\hat{f}_n(\eta_i)}{h_i} \right\}$, a new node $z'_{i+1} = \frac{z_i + z_{i+1}}{2}$ is inserted into the i -th bin. Split B_i into two sub-bins: $B'_i = [z_i, z'_{i+1})$, $B'_{i+1} = [z'_{i+1}, z_{i+1})$. The midpoints of the new sub-bins B'_i and B'_{i+1} are $\eta'_i = \frac{z_i + z'_{i+1}}{2}$ and $\eta'_{i+1} = \frac{z'_{i+1} + z_{i+1}}{2}$, respectively. Apply (3) to obtain the corresponding $\hat{f}_n(\eta'_i)$ and $\hat{f}_n(\eta'_{i+1})$ for the two new sub-bins. Consequently, the updated midpoints are given by $(\eta'_i, \hat{f}_n(\eta'_i))$ and $(\eta'_{i+1}, \hat{f}_n(\eta'_{i+1}))$.

$(\eta'_{i+1}, \hat{f}_n(\eta'_{i+1}))$, respectively. By rearranging the indices of all bins, we insert more nodes till $slp_i(\hat{f}) \geq \frac{-2\hat{f}_n(\eta_i)}{h_i}$ holds.

- (III) Denote by $B_{i_{max}}$, $B_{i_{max}+1}$, and $B_{i_{max}+2}$ the three adjacent bins corresponding to this maximum second-order difference quotient, respectively. Let TH be the given interpolation tolerance. If it is required that the maximum second-order difference quotient $\Delta_{i_{max}}^2 \geq TH$, then we select the bin with the maximum frequency in $B_{i_{max}}$, $B_{i_{max}+1}$, and $B_{i_{max}+2}$, denoted by B_{i_m} . Add its midpoint as the newly inserted interpolator, i.e., $z'_{i_m+1} = \frac{z_{i_m} + z_{i_m+1}}{2}$. Split B_{i_m} into two sub-bins: $B'_{i_m} = [z_{i_m}, z'_{i_m+1})$, $B'_{i_m+1} = [z'_{i_m+1}, z_{i_m+1})$. The midpoints of the new sub-bins B'_{i_m} and B'_{i_m+1} are $\eta'_{i_m} = \frac{z_{i_m} + z'_{i_m+1}}{2}$ and $\eta'_{i_m+1} = \frac{z'_{i_m+1} + z_{i_m+1}}{2}$, respectively. The updated midpoints are given by $(\eta'_{i_m}, \hat{f}_n(\eta'_{i_m}))$ and $(\eta'_{i_m+1}, \hat{f}_n(\eta'_{i_m+1}))$, respectively. Consequently, among the three bins involved in the maximum second-order difference quotient, only the bin with a larger frequency needs to be further interpolated. With this new interpolator, we calculate the largest second-order difference quotient again and repeat the above interpolation until the inequality $\Delta_{i_{max}}^2 < TH$ is satisfied.

The consecutive steps of the proposed approach to adaptively estimating the probability density based on Rules (I)–(III) are illustrated in Figure 1.

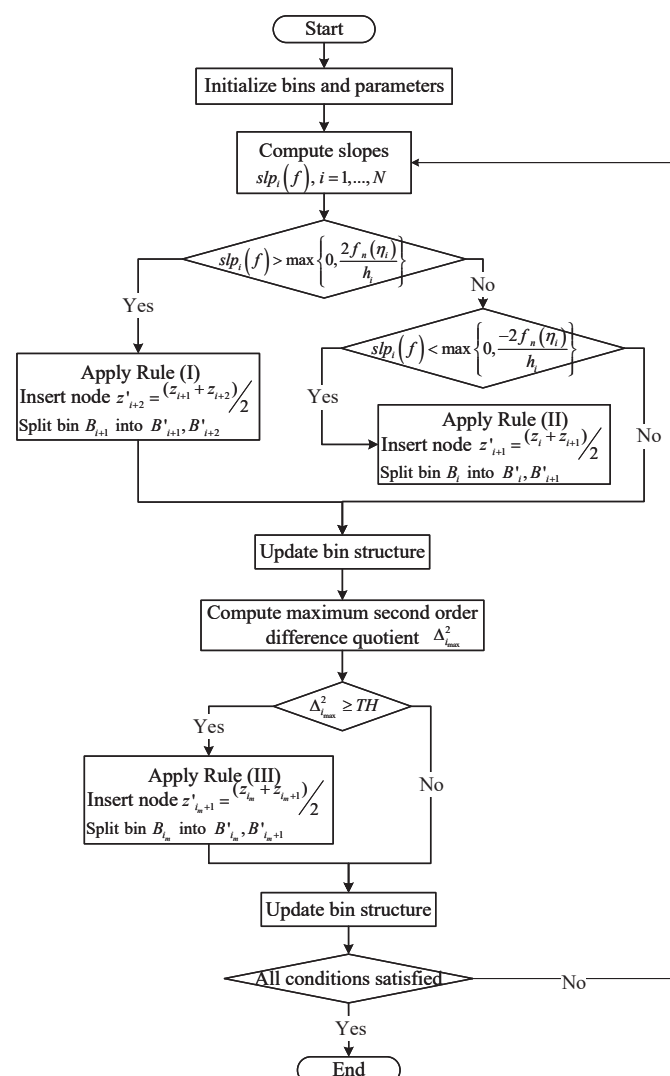


Figure 1. The structure of the proposed approach to adaptively estimating the probability density based on Rules (I)–(III).

Remark 1. As demonstrated intuitively in Figure 1, the above three rules aim to optimize the bin width by adaptively inserting new knots in each bin in line with the properties of $\hat{f}$ desired. Essentially, bins with larger slope variations will be inserted into new knots. The rate of slope changes between these bins is characterized using the second-order difference quotient.

After applying Rules (I)–(III) to optimize the bin widths, the obtained estimation of the density function can depict the distributional nature of the sample data. Without loss of generality, the final node set is still denoted by $z_{final} = \{z_1, z_2, \dots, z_{N+1}\}$, where N is the optimized number of bins. Consequently, by Rules (I), (II), and (III), $\hat{f}$ in (4) is modified to be a piecewise linear function with variable steps, denoted by $\hat{f}_n$. We can prove that $\hat{f}_n$ is an improved estimation of the true density function with adaptive bin widths, and the method is called improved density estimation with piecewise linear interpolation (IDE-PLI).

Proposition 1. Let $\hat{f}$ be defined by (4) and Rules (I), (II), and (III). Then, (1) it holds that $\hat{f}_n \geq 0$. (2) $\int_{z_1}^{z_{N+1}} \hat{f}_n(z) dz = 1$.

Proof. (1) In the case of $slp_i(\hat{f}) \geq 0$, it follows from the first equality in (4) that

$$\begin{aligned}\hat{f}(z_i) &= \frac{2(\hat{f}_n(\eta_{i+1}) - \hat{f}_n(\eta_i))}{h_i + h_{i+1}}(z_i - \eta_i) + \hat{f}_n(\eta_i) \\ &= -slp_i(\hat{f}(z), \gamma)(\eta_i - z_i) + \hat{f}_n(\eta_i) \\ &\geq -\frac{2\hat{f}_n(\eta_i)}{h_i}(\eta_i - z_i) + \hat{f}_n(\eta_i) \\ &= 0,\end{aligned}$$

where $h_i = \eta_{i+1} - \eta_i$, $h_{i+1} = \eta_{i+2} - \eta_{i+1}$, $z_i \leq z < z_{i+1}$. A similar proof can be obtained in the case of $slp_i(\hat{f}(z)) < 0$.

(2) For the bin $[z_i, z_{i+1})$, $i = 1, 2, \dots, N-1$, we have

$$\begin{aligned}\int_{z_i}^{z_{i+1}} \hat{f}(z) dz &= \int_{z_i}^{z_{i+1}} \left(\frac{2(\hat{f}_n(\eta_{i+1}) - \hat{f}_n(\eta_i))}{h_i + h_{i+1}}(z - \eta_i) \right) dz + \int_{z_i}^{z_{i+1}} \hat{f}_n(\eta_i) dz \\ &= \frac{\hat{f}_n(\eta_{i+1}) - \hat{f}_n(\eta_i)}{h_i + h_{i+1}} ((z_{i+1} - \eta_i)^2 - (z_i - \eta_i)^2) + \hat{f}_n(\eta_i)(z_{i+1} - z_i) \\ &= \frac{\hat{f}_n(\eta_{i+1}) - \hat{f}_n(\eta_i)}{h_i + h_{i+1}} h_i(z_{i+1} + z_i - 2\eta_i) + \hat{f}_n(\eta_i)h_i \\ &= \hat{f}_n(\eta_i)h_i.\end{aligned}$$

For the bin $[z_N, z_{N+1}]$, we obtain

$$\int_{z_N}^{z_{N+1}} \hat{f}(z) dz = \int_{z_N}^{z_{N+1}} \left(\frac{h_N + \eta_N - z}{h_N} \hat{f}_n(\eta_N) \right) dz = \hat{f}_n(\eta_N)h_N.$$

By the classical histogram method, we know that $\sum_{i=1}^N \hat{f}_n(\eta_i)h_i = 1$. Thus,

$$\int_{z_1}^{z_{N+1}} \hat{f}(z) dz = 1.$$

The proof is completed. $\square$

Remark 2. By Proposition 1, one knows that $\hat{f}_n$ can also be viewed as an estimation of the true density function, but its adaptive bin widths enable it to clearly depict the distributional nature of the sample data, as compared with the fixed bin width in the classical frequency histogram method.

2.2. Construction of Confidence Sets Only with Finitely Many Parameters

Using the obtained $\hat{f}_n(z)$ with $N + 1$ nodes $z_1, z_2, \dots, z_{N+1}$, we calculate the probability mass $\hat{\mathbb{P}}$ of $\hat{f}_n(z)$ on each bin B_i

$$\hat{p}_i(z_i, z_{i+1}) = \hat{\mathbb{P}}(\xi \in [z_i, z_{i+1})) = \int_{z_i}^{z_{i+1}} \hat{f}_n(z) dz, i = 1, \dots, N. \quad (5)$$

With $\hat{p}_i(z_i, z_{i+1})$, $i = 1, \dots, N$, we define a confidence set for the random variable ξ and a given divergence tolerance $\gamma > 0$ by

$$\mathcal{D} = \left\{ \mathbb{P} \left| \begin{array}{l} p_i = \mathbb{P}\{\xi = \eta_i = \frac{z_i + z_{i+1}}{2}\}, \quad i = 1, 2, \dots, N, \\ \sum_{i=1}^N p_i = 1, \\ \sum_{i=1}^N |p_i - \hat{p}_i(z_i, z_{i+1})| \leq \gamma, \\ 0 \leq p_i \leq 1, \quad i = 1, 2, \dots, N. \end{array} \right. \right\}. \quad (6)$$

Clearly, the divergence tolerance $\gamma > 0$ can control the size of the confidence set $\mathcal{D}$. Notably, different from the existing confidence sets in the literature, the set $\mathcal{D}$ in (6) only involves finitely many parameters, rather than a space of probability density functions.

Theorem 1. Let ξ be a random variable with the distribution $\mathbb{P}$ and probability density f . Partition the domain of ξ into N disjoint bins $\{B_i\}_{i=1}^N$. The sample size is denoted by n_s . Let $\hat{p}_i(z_i, z_{i+1})$ be defined by (5). Let $\mathbb{P}_{n_s}$ denote the sampling distribution induced by samples of size n_s . Then, for any confidence level $\beta \in (0, 1)$, it holds that

$$\mathbb{P}_{n_s} \left(\sum_{i=1}^N |p_i - \hat{p}_i(z_i, z_{i+1})| \leq \sqrt{\frac{N^2 \ln(2N/\beta)}{2n_s}} \right) \geq 1 - \beta. \quad (7)$$

Proof. By (5), for each bin i ,

$$\hat{p}_i(z_i, z_{i+1}) = \int_{B_i} \frac{v_i}{n_s h_i} dz = \frac{v_i}{n_s} = \frac{1}{n_s} \sum_{j=1}^{n_s} I_{B_i}(\xi^j), i = 1, \dots, N, j = 1, \dots, n_s,$$

where $I_{B_i}(\xi^j)$ is the Bernoulli variable with $0 \leq I_{B_i}(\xi^j) \leq 1$. Due to sample independence, the indicators $I_{B_i}(\xi^j)$ are independent across j , satisfying the independence and boundedness conditions required for Hoeffding's inequality [22]. For any $\epsilon > 0$,

$$\mathbb{P}_{n_s}(|\hat{p}_i(z_i, z_{i+1}) - p_i| \geq \epsilon) \leq 2 \exp(-2n_s \epsilon^2). \quad (8)$$

To control the total variation divergence across all N bins, we use the union bound to the events $\{|\hat{p}_i(z_i, z_{i+1}) - p_i| \geq \epsilon\}$, i.e.,

$$\mathbb{P}_{n_s} \left(\bigcup_{i=1}^N \{|\hat{p}_i(z_i, z_{i+1}) - p_i| \geq \epsilon\} \right) \leq \sum_{i=1}^N \mathbb{P}_{n_s}(|\hat{p}_i(z_i, z_{i+1}) - p_i| \geq \epsilon). \quad (9)$$

Substituting the Hoeffding bound from (8) into (9), we get

$$\mathbb{P}_{n_s} \left(\bigcup_{i=1}^N \{|\hat{p}_i(z_i, z_{i+1}) - p_i| \geq \epsilon\} \right) \leq 2N \exp(-2n_s \epsilon^2).$$

For a given confidence level $\beta \in (0, 1)$, the inequality $2N \exp(-2n_s \epsilon^2) \leq \beta$ yields

$$\epsilon \geq \sqrt{\frac{\ln(2N/\beta)}{2n_s}}.$$

Consequently, the desired inequality (7) holds. $\square$

Remark 3. By Theorem 1, we establish a rigorous statistical guarantee for the confidence set $\mathcal{D}$: the total deviation between the estimated probabilities $\hat{p}_i$ (from adaptive bandwidth density estimation) and the true probabilities p_i is bounded by

$$\gamma_{\beta, n_s} = \sqrt{\frac{N^2 \ln(2N/\beta)}{2n_s}}$$

with confidence level $1 - \beta$.

Notably, $\mathcal{D}$ is constructed based on direct deviations of probability mass functions, rather than the traditional moment information (e.g., mean or covariance), making it inherently suitable for scenarios where higher-order distributional properties are unknown or difficult to estimate.

By Theorem 1, we further prove the following result, which gives a distributionally robust guarantee for the chance constraint $\mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha$.

Theorem 2. Let ξ be a random variable with unknown distribution $\mathbb{P}$. Suppose that the domain of ξ is partitioned into N disjoint bins $\{B_i\}_{i=1}^N$. Let n_s be the number of independent samples used to construct the reference distribution $\hat{\mathbb{P}}$, with $\hat{p}_i$ denoting the reference probability of bin B_i . Define the confidence set $\mathcal{D}$ defined in (6), where $\gamma = \sqrt{\frac{N^2 \ln(2N/\beta)}{2n_s}}$ for some $\beta \in (0, 1)$. For the chance constraint $\mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha$, let $\hat{V}(x) = \hat{\mathbb{P}}\{A(\xi)x > b(\xi)\}$ be the reference violation probability. Then, for any $\alpha, \beta \in (0, 1)$, when

$$\hat{V}(x) \leq \alpha - \frac{\gamma}{2},$$

it holds with probability at least $1 - \beta$ that $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha$.

Moreover, the required sample size n_s to ensure $\hat{V}(x) \leq \alpha - \frac{\gamma}{2}$ satisfies

$$n_s \geq \frac{N^2 \ln(2N/\beta)}{8(\alpha - \hat{V}(x))^2}. \quad (10)$$

Proof. By Theorem 1, with probability at least $1 - \beta$, we have $\sum_{i=1}^N |p_i - \hat{p}_i| \leq \gamma$. The total variation distance [23] between $\mathbb{P}$ and $\hat{\mathbb{P}}$ is

$$\delta(\mathbb{P}, \hat{\mathbb{P}}) = \frac{1}{2} \sum_{i=1}^N |p_i - \hat{p}_i| \leq \frac{\gamma}{2}.$$

Thus, for any random event E , we know that

$$|\mathbb{P}(E) - \hat{\mathbb{P}}(E)| \leq \delta(\mathbb{P}, \hat{\mathbb{P}}) \leq \frac{\gamma}{2}.$$

In particular, taking $E = \{A(\xi)x > b(\xi)\}$, we get

$$\mathbb{P}(E) \leq \hat{\mathbb{P}}(E) + \frac{\gamma}{2} = \hat{V}(x) + \frac{\gamma}{2}.$$

Consequently,

$$\mathbb{P}\{A(\xi)x \leq b(\xi)\} = 1 - \mathbb{P}(E) \geq 1 - \hat{V}(x) - \frac{\gamma}{2}.$$

In the case that $\hat{V}(x) \leq \alpha - \frac{\gamma}{2}$, it is concluded that $\mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha$. Since this inequality holds for all $\mathbb{P} \in \mathcal{D}$, it follows that

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha.$$

Since

$$\hat{V}(x) \leq \alpha - \frac{\gamma}{2},$$

it follows from $\gamma = \sqrt{\frac{N^2 \ln(2N/\beta)}{2n_s}}$ that

$$n_s \geq \frac{N^2 \ln(2N/\beta)}{8(\alpha - \hat{V}(x))^2}.$$

This completes the proof. $\square$

Remark 4. In Theorem 2, a sample size condition (10) is provided such that the distributionally robust chance constraint holds with high probability, without requiring knowledge of the mean or covariance. Note that this condition depends on the reference violation probability $\hat{V}(x)$, which can be directly estimated from the collected data.

3. Reformulation of Data-Driven Distributionally Robust Chance Constraints

3.1. Reformulation of Distributionally Robust Chance Constraints with a Random Variable

As done in [8], it is supposed that all the model parameters in (2b) are linearly correlated with the random parameter ξ , which yields

$$A(\xi) = A^0 + A\xi, \quad b(\xi) = b^0 + b\xi.$$

Thus, for each $k \in K = \{1, 2, \dots, m\}$, it holds that

$$A(\xi)x \leq b(\xi) \iff (Ax - b)\xi \leq b^0 - A^0x \iff (A_kx - b_k)\xi \leq b_k^0 - A_k^0x,$$

where the k -th row of the matrix A is denoted by A_k .

Remark 5. When $A_kx - b_k = 0$, the inequality degenerates to $0 \leq b_k^0 - A_k^0x$, which is a deterministic constraint independent of ξ . When ξ is a continuous random parameter, the degenerate case $A_kx - b_k = 0$ can be ignored in the analysis of the chance constraint.

For any given x , define

$$c_k(x) = \left\{ z \in \mathbb{R} \left| \begin{array}{ll} z \leq \text{Top}_k^+(x) = \frac{b_k^0 - A_k^0x}{A_kx - b_k}, & \text{if } A_kx - b_k > 0, \\ z \geq \text{Top}_k^-(x) = \frac{b_k^0 - A_k^0x}{A_kx - b_k}, & \text{if } A_kx - b_k < 0. \end{array} \right. \right\}$$

Let $\mathcal{C}(x) = \bigcap_{k=1}^m c_k(x)$. Then, we can define the worst-case probability bound by

$$z_{\mathcal{D}} = \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{A(\xi)x \leq b(\xi)\} = \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi \in \mathcal{C}(x)\}. \quad (11)$$

Denote

$$\begin{aligned} S^+ &= \{k \in K : A_k x - b_k > 0\}, \quad U(x) = \min_{k \in S^+} \text{Top}_k^+(x), \\ S^- &= \{k \in K : A_k x - b_k < 0\}, \quad L(x) = \max_{k \in S^-} \text{Top}_k^-(x). \end{aligned} \quad (12)$$

We now prove that $z_{\mathcal{D}}$ in (11) has the following property.

Proposition 2. Let S^+ , S^- , $L(x)$, and $U(x)$ be defined by (12). The worst-case probability bound $z_{\mathcal{D}}$ in (11) is specified by one of the following three formulas.

(1) When $S^- = \emptyset$,

$$z_{\mathcal{D}} = \begin{cases} 0, & \text{if } U(x) < z_1, \\ \inf_{\mathbb{P} \in \mathcal{D}} \sum_{i=1}^j p_i, & \text{if } z_j \leq U(x) < z_{j+1}, j = 1, 2, \dots, N, \\ 1, & \text{if } U(x) \geq z_N. \end{cases} \quad (13)$$

(2) When $S^+ = \emptyset$,

$$z_{\mathcal{D}} = \begin{cases} 1, & \text{if } L(x) < z_1, \\ \inf_{\mathbb{P} \in \mathcal{D}} \sum_{i=j}^N p_i, & \text{if } z_j \leq L(x) < z_{j+1}, j = 1, 2, \dots, N, \\ 0, & \text{if } L(x) \geq z_{N+1}. \end{cases} \quad (14)$$

(3) When $S^+ \neq \emptyset$ and $S^- \neq \emptyset$,

$$z_{\mathcal{D}} = \begin{cases} 0, & \text{if } L(x) > U(x), \text{ or } U(x) < z_1, \text{ or } L(x) \geq z_{N+1}, \\ \inf_{\mathbb{P} \in \mathcal{D}} \sum_{i=j}^{j'} p_i, & \text{if } L(x) \in [z_j, z_{j+1}), U(x) \in [z_{j'}, z_{j'+1}) \text{ with } j, j' \in \{1, 2, \dots, N\}, \\ 1, & \text{if } L(x) \leq z_1 \text{ and } U(x) \geq z_{N+1}. \end{cases}$$

Proof. The chance constraints $\mathbb{P}\{\xi \in \mathcal{C}(x)\}$ in (11) reduce to three distinct cases.

Case 1: $S^- = \emptyset$. Then, $\mathbb{P}\{\xi \in \mathcal{C}(x)\} = \mathbb{P}\left(\xi \leq \min_{k \in K} \text{Top}_k^+(x)\right)$, matching the original formulation when all $A_k x - b_k > 0$. From (12), we know that

$$\mathbb{P}(\xi \in \mathcal{C}(x)) = \mathbb{P}(\xi \leq U(x)) = \begin{cases} 0, & \text{if } U(x) < z_1, \\ \sum_{i=1}^j p_i, & \text{if } z_j \leq U(x) < z_{j+1}, j = 1, 2, \dots, N, \\ 1, & \text{if } U(x) \geq z_N. \end{cases}$$

Thus,

$$z_{\mathcal{D}} = \begin{cases} 0, & \text{if } U(x) < z_1, \\ \inf_{\mathbb{P} \in \mathcal{D}} \sum_{i=1}^j p_i, & \text{if } z_j \leq U(x) < z_{j+1}, j = 1, 2, \dots, N, \\ 1, & \text{if } U(x) \geq z_N. \end{cases}$$

Case 2: $S^+ = \emptyset$. Then, $\mathbb{P}\{\xi \in \mathcal{C}(x)\} = \mathbb{P}\left(\xi \geq \max_{k \in K} \text{Top}_k^-(x)\right)$, matching the original formulation when all $A_k x - b_k < 0$. By (12), we get

$$\mathbb{P}(\xi \in \mathcal{C}(x)) = \mathbb{P}(\xi \geq L(x)) = \begin{cases} 1, & \text{if } L(x) < z_1, \\ \sum_{i=j}^N p_i, & \text{if } z_j \leq L(x) < z_{j+1}, j = 1, 2, \dots, N, \\ 0, & \text{if } L(x) \geq z_{N+1}. \end{cases}$$

Then,

$$z_{\mathcal{D}} = \begin{cases} 1, & \text{if } L(x) < z_1, \\ \inf_{\mathbb{P} \in \mathcal{D}} \sum_{i=j}^N p_i, & \text{if } z_j \leq L(x) < z_{j+1}, j = 1, 2, \dots, N, \\ 0, & \text{if } L(x) \geq z_{N+1}. \end{cases}$$

Case 3: $S^+ \neq \emptyset$ and $S^- \neq \emptyset$. Then,

$$\mathbb{P}\{\xi \in \mathcal{C}(x)\} = \mathbb{P}\left(\max_{k \in S^-} \text{Top}_k^-(x) \leq \xi \leq \min_{k \in S^+} \text{Top}_k^+(x)\right),$$

which is nonzero only if $\max_{k \in S^-} \text{Top}_k^-(x) \leq \min_{k \in S^+} \text{Top}_k^+(x)$. Conversely, if $\max_{k \in S^-} \text{Top}_k^-(x) > \min_{k \in S^+} \text{Top}_k^+(x)$, then $\mathbb{P}(\xi \in \mathcal{C}(x)) = 0$.

From (12), it follows that the probability $\mathbb{P}\{\xi \in \mathcal{C}(x)\}$ is expressed as follows:

$$\begin{aligned} \mathbb{P}(\xi \in \mathcal{C}(x)) &= \mathbb{P}(L(x) \leq \xi \leq U(x)) \\ &= \begin{cases} 0, & \text{if } L(x) > U(x), \text{ or } U(x) < z_1, \text{ or } L(x) \geq z_{N+1}, \\ \sum_{i=j}^{j'} p_i, & \text{if } L(x) \in [z_j, z_{j+1}), U(x) \in [z_{j'}, z_{j'+1}) \text{ with } j, j' \in \{1, 2, \dots, N\}, \\ 1, & \text{if } L(x) \leq z_1 \text{ and } U(x) \geq z_{N+1}. \end{cases} \end{aligned}$$

Then,

$$z_{\mathcal{D}} = \begin{cases} 0, & \text{if } L(x) > U(x), \text{ or } U(x) < z_1, \text{ or } L(x) \geq z_{N+1}, \\ \inf_{\mathbb{P} \in \mathcal{D}} \sum_{i=j}^{j'} p_i, & \text{if } L(x) \in [z_j, z_{j+1}), U(x) \in [z_{j'}, z_{j'+1}), \text{ with } j, j' \in \{1, 2, \dots, N\}, \\ 1, & \text{if } L(x) \leq z_1 \text{ and } U(x) \geq z_{N+1}. \end{cases}$$

With the above argument, we have completed the proof of the desired result. $\square$

By Proposition 2, we now prove that the chance constraints (2b) can be reformulated as a number of ordinary ones.

Theorem 3. Let A and b be a given matrix and a vector. Let $\hat{\mathbb{P}}$ be the estimated distribution obtained through the proposed adaptive data-driven method. Let γ be a given divergence tolerance of the ambiguous density functions in the confidence set $\mathcal{D}$ defined by (6), satisfying $2\alpha \geq \gamma > 0$. Then, the feasible set defined by the chance constraint, i.e., $\mathcal{F}_{cc} = \left\{x \in \mathbb{R}^n : \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha\right\}$, can be reformulated in the following forms.

1. The set $\mathcal{F}_{cc} = \emptyset$ if one of the following conditions holds:

- (i) For all $k \in K$, $A_k x - b_k \geq 0$, $U(x) \leq z_1$;
- (ii) For all $k \in K$, $A_k x - b_k \leq 0$, $L(x) \geq z_{N+1}$;
- (iii) There exist $k_1, k_2 \in K$ such that $A_{k_1} x - b_{k_1} \geq 0$, $A_{k_2} x - b_{k_2} \leq 0$, $L(x) \geq U(x)$;
- (iv) There exist $k_1, k_2 \in K$ such that $A_{k_1} x - b_{k_1} \geq 0$, $A_{k_2} x - b_{k_2} \leq 0$, $U(x) \leq z_1$;
- (v) There exist $k_1, k_2 \in K$ such that $A_{k_1} x - b_{k_1} \geq 0$, $A_{k_2} x - b_{k_2} \leq 0$, $L(x) \geq z_{N+1}$.

2. The set $\mathcal{F}_{cc} = \mathcal{C}_1^1 \cup \mathcal{C}_2^1 \cup \mathcal{C}_3^1$, where

$$\begin{aligned}\mathcal{C}_1^1 &= \{x \in R^n : \forall k \in K, A_k x - b_k \geq 0, U(x) \geq z_N\}, \\ \mathcal{C}_2^1 &= \{x \in R^n : \forall k \in K, A_k x - b_k \leq 0, L(x) \leq z_1\}, \\ \mathcal{C}_3^1 &= \{x \in R^n : \exists k_1, k_2 \in K, A_{k_1} x - b_{k_1} \geq 0, A_{k_2} x - b_{k_2} \leq 0, L(x) \leq z_1\} \\ &\quad \cup \{x \in R^n : \exists k_1, k_2 \in K, A_{k_1} x - b_{k_1} \geq 0, A_{k_2} x - b_{k_2} \leq 0, U(x) \geq z_{N+1}\}\end{aligned}$$

3. The set $\mathcal{F}_{cc} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3$, where

$$\begin{aligned}\mathcal{C}_1 &= \left\{x \in R^n : \sum_{i=1}^j \hat{p}_i(z_i, z_{i+1}) \geq 1 - \alpha + \frac{\gamma}{2}, A_k x - b_k \geq 0, z_j \leq U(x) < z_{j+1}, \forall k, j\right\}, \\ \mathcal{C}_2 &= \left\{x \in R^n : \sum_{i=j}^N \hat{p}_i(z_i, z_{i+1}) \geq 1 - \alpha + \frac{\gamma}{2}, A_k x - b_k \leq 0, z_j \leq L(x) \leq z_{j+1}, \forall k, j\right\}, \\ \mathcal{C}_3 &= \left\{x \in R^n : \sum_{i=j}^{j'} \hat{p}_i(z_i, z_{i+1}) \geq 1 - \alpha + \frac{\gamma}{2}, \exists k_1, k_2 \in K, A_{k_1} x - b_{k_1} \geq 0, A_{k_2} x - b_{k_2} \leq 0, \right. \\ &\quad \left. L(x) \in [z_j, z_{j+1}), U(x) \in [z_{j'}, z_{j'+1}), \text{with } j, j' \in \{1, 2, \dots, N\}\right\}.\end{aligned}$$

Proof. We prove the result in the following three cases.

Case 1: $S^- = \emptyset$, i.e., $A_k x - b_k > 0$ for all $k \in K$.

(i) When $U(x) < z_1$, from (13), it is deduced that $z_{\mathcal{D}} = 0$. Consequently, the inequality $z_{\mathcal{D}} \geq 1 - \alpha$ does not hold.

(ii) When $U(x) \geq z_N$, it follows from (13) that $z_{\mathcal{D}} = 1$, and thus $z_{\mathcal{D}} \geq 1 - \alpha$ always holds.

(iii) When $z_j \leq U(x) < z_{j+1}$, $j = 1, 2, \dots, N$, the confidence set $\mathcal{D}$ is defined by (6). By Theorem 1, for the given x and the set $\mathcal{C}(x)$, the worst-case probability bound (11) is transformed into the following programming problem:

$$[\text{Primal}] \ z_{\mathcal{D}} = \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi \in \mathcal{C}(x)\} \quad (15a)$$

$$\Leftrightarrow \min_{\mathbb{P}} \mathbb{P}\{\xi \in \mathcal{C}(x)\} \quad (15b)$$

$$\text{s.t.} \quad \sum_{i=1}^N \hat{p}_i(z_i, z_{i+1}) \left| 1 - \frac{p_i}{\hat{p}_i(z_i, z_{i+1})} \right| \leq \gamma, \quad (15c)$$

$$\sum_{i=1}^N p_i = 1, \quad (15d)$$

$$p_i \geq 0, i = 1, 2, \dots, N, \quad (15e)$$

where the constraint (15c) bounds the divergence from above by γ , and constraints (15d) and (15e) guarantee that $\mathbb{P}$ is a density function. This transformation is justified by the structural properties of the confidence set $\mathcal{D}$, which collectively constrain the feasible region of $\mathbb{P}$. Consequently, we get

$$\begin{aligned}z_{\mathcal{D}} &= \min_{\mathbb{P}} \sum_{i=1}^j p_i \\ \text{s.t.} \quad &\sum_{i=1}^N \hat{p}_i(z_i, z_{i+1}) \left| 1 - \frac{p_i}{\hat{p}_i(z_i, z_{i+1})} \right| \leq \gamma, \\ &\sum_{i=1}^N p_i = 1, \\ &p_i \geq 0, i = 1, 2, \dots, N.\end{aligned} \quad (16)$$

For the constraints in Problem (16), let λ^{eq} be the Lagrange multiplier corresponding to the equality constraint. Let λ^{ne} be the Lagrange multiplier corresponding to the inequality constraint. Then, its Lagrangian function reads

$$\begin{aligned}
 L(\mathbb{P}, \lambda^{eq}, \lambda^{ne}) &= \max_{\lambda^{eq}, \lambda^{ne} \geq 0} \min_{\mathbb{P} \geq 0} \left\{ \sum_{i=1}^j p_i + \lambda^{eq} \left(1 - \sum_{i=1}^N p_i \right) + \lambda^{ne} \left(\sum_{i=1}^N \hat{p}_i(z_i, z_{i+1}) \left| 1 - \frac{p_i}{\hat{p}_i(z_i, z_{i+1})} \right| - \gamma \right) \right\} \\
 &= \max_{\lambda^{eq}, \lambda^{ne} \geq 0} \{ \lambda^{eq} - \lambda^{ne} \gamma \} + \\
 &\quad \min_{\mathbb{P} \geq 0} \left(\sum_{i=1}^N (I_{\mathcal{C}(x)}(z) - \lambda^{eq}) p_i + \lambda^{ne} \sum_{i=1}^N \hat{p}_i(z_i, z_{i+1}) \left| 1 - \frac{p_i}{\hat{p}_i(z_i, z_{i+1})} \right| \right) \\
 &= \max_{\lambda^{eq}, \lambda^{ne} \geq 0} \{ \lambda^{eq} - \lambda^{ne} \gamma \} - \\
 &\quad \max_{\mathbb{P} \geq 0} \left(\sum_{i=1}^N (\lambda^{eq} - I_{\mathcal{C}(x)}(z)) p_i - \lambda^{ne} \sum_{i=1}^N \hat{p}_i(z_i, z_{i+1}) \left| 1 - \frac{p_i}{\hat{p}_i(z_i, z_{i+1})} \right| \right) \\
 &= \max_{\lambda^{eq}, \lambda^{ne} \geq 0} \{ \lambda^{eq} - \lambda^{ne} \gamma - \\
 &\quad \max_{\mathbb{P} \geq 0} \lambda^{ne} \sum_{i=1}^N \hat{p}_i(z_i, z_{i+1}) \left(\frac{\lambda^{eq} - I_{\mathcal{C}(x)}(z)}{\lambda^{ne}} \frac{p_i}{\hat{p}_i(z_i, z_{i+1})} - \left| 1 - \frac{p_i}{\hat{p}_i(z_i, z_{i+1})} \right| \right) \},
 \end{aligned}$$

where $I_{\mathcal{C}(x)}(z)$ is the indicator function, which takes the value of 1 when $z \in \mathcal{C}(x)$; otherwise, it takes 0. Note that the conjugate function of $\varphi(t) = |t - 1|$ is

$$\varphi^*(s) = \sup_{t \geq 0} \{ st - \varphi(t) \} = \begin{cases} -1, & \text{if } s < -1, \\ s, & \text{if } -1 \leq s \leq 1, \\ +\infty, & \text{if } s > 1. \end{cases} \quad (17)$$

Then,

$$\begin{aligned}
 z'_D &= \max_{\lambda^{eq}, \lambda^{ne} > 0} \left\{ \lambda^{eq} - \lambda^{ne} \gamma - \lambda^{ne} \sum_{i=1}^N \hat{p}_i(z_i, z_{i+1}) \varphi^* \left(\frac{\lambda^{eq} - I_{\mathcal{C}(x)}(z)}{\lambda^{ne}} \right) \right\} \\
 &= \max_{\lambda^{eq}, \lambda^{ne} > 0} \left\{ \lambda^{eq} - \lambda^{ne} \gamma - \right. \\
 &\quad \left. \lambda^{ne} \sum_{i: I_{\mathcal{C}(x)}=1} \hat{p}_i(z_i, z_{i+1}) \varphi^* \left(\frac{\lambda^{eq} - 1}{\lambda^{ne}} \right) - \lambda^{ne} \sum_{i: I_{\mathcal{C}(x)}=0} \hat{p}_i(z_i, z_{i+1}) \varphi^* \left(\frac{\lambda^{eq}}{\lambda^{ne}} \right) \right\} \\
 &= \max_{\lambda^{eq}, \lambda^{ne} > 0} \left\{ \lambda^{eq} - \lambda^{ne} \gamma - \lambda^{ne} \hat{z}_D \varphi^* \left(\frac{\lambda^{eq} - 1}{\lambda^{ne}} \right) - \lambda^{ne} (1 - \hat{z}_D) \varphi^* \left(\frac{\lambda^{eq}}{\lambda^{ne}} \right) \right\},
 \end{aligned} \quad (18)$$

where $\hat{z}_D = \sum_{i: I_{\mathcal{C}(x)}=1} \hat{p}_i(z_i, z_{i+1}) = \hat{\mathbb{P}}(\xi \in \mathcal{C}(x))$.

By the strong duality in the theory of linear programming, it holds that $z_D = z'_D$ for Problems (15) and (18). Thus, there exist constants $\lambda^{eq} > 0$ and $\lambda^{ne} > 0$ such that $z'_D \geq 1 - \alpha$, i.e.,

$$\lambda^{eq} - \lambda^{ne} \gamma - \lambda^{ne} \hat{z}_D \varphi^* \left(\frac{\lambda^{eq} - 1}{\lambda^{ne}} \right) - \lambda^{ne} (1 - \hat{z}_D) \varphi^* \left(\frac{\lambda^{eq}}{\lambda^{ne}} \right) \geq 1 - \alpha \quad (19a)$$

$$\Leftrightarrow \hat{z}_D \geq \frac{1 - \alpha - \lambda^{eq} + \lambda^{ne} \gamma + \lambda^{ne} \varphi^* \left(\frac{\lambda^{eq}}{\lambda^{ne}} \right)}{\lambda^{ne} \left(\varphi^* \left(\frac{\lambda^{eq}}{\lambda^{ne}} \right) - \varphi^* \left(\frac{\lambda^{eq} - 1}{\lambda^{ne}} \right) \right)} \quad (19b)$$

$$= 1 - \frac{\alpha + \lambda^{eq} - 1 - \lambda^{ne} \gamma - \lambda^{ne} \varphi^* \left(\frac{\lambda^{eq} - 1}{\lambda^{ne}} \right)}{\lambda^{ne} \left(\varphi^* \left(\frac{\lambda^{eq}}{\lambda^{ne}} \right) - \varphi^* \left(\frac{\lambda^{eq} - 1}{\lambda^{ne}} \right) \right)}. \quad (19c)$$

Denote $\lambda_1 = \frac{\lambda^{eq}}{\lambda^{ne}} \geq 0$ and $\lambda_0 = \frac{\lambda^{eq} - 1}{\lambda^{ne}}$. Then, $\lambda_0 = \lambda_1 - \frac{1}{\lambda^{ne}} \leq \lambda_1$.

We first prove that $\lambda_1 > 1$ is invalid. Indeed, from (17) and $\lambda^{ne} > 0$, it is inferred that when $\lambda_1 > 1$, the left side of (19a) tends towards negative infinity, violating DCC. Therefore, $0 \leq \lambda_1 \leq 1$ and $\lambda_0 \leq 1$.

Because φ^* is a conjugate function, from the properties of conjugate functions, it follows that (19c) is equivalent to

$$\hat{z}_{\mathcal{D}} \geq 1 - \sup_{\lambda_0 \leq 1, 0 \leq \lambda_1 \leq 1} \frac{\alpha + \lambda^{eq} - 1 - \lambda^{ne}\gamma - \lambda^{ne}\varphi^*\left(\frac{\lambda^{eq} - 1}{\lambda^{ne}}\right)}{\lambda^{ne}\left(\varphi^*\left(\frac{\lambda^{eq}}{\lambda^{ne}}\right) - \varphi^*\left(\frac{\lambda^{eq} - 1}{\lambda^{ne}}\right)\right)}. \quad (20)$$

We now conduct a case-by-case analysis of λ_0 as follows.

When $-1 \leq \lambda_0 \leq 1$, the right-hand side of inequality (20) is transformed into

$$1 - \sup_{\lambda_0 \leq 1, 0 \leq \lambda_1 \leq 1} \{\alpha - \lambda^{ne}\gamma\} = 1 - \alpha + \frac{\gamma}{2}. \quad (21)$$

Indeed, $\frac{1}{\lambda^{ne}} - 1 \leq \lambda_1 \leq 1 + \frac{1}{\lambda^{ne}}$ and $\lambda_1 \leq 1$. Thus, $\lambda^{ne} \geq \frac{1}{2}$.

When $\lambda_0 < -1$, the right-hand side of inequality (20) is rewritten as

$$1 - \sup_{\lambda_0 \leq 1, 0 \leq \lambda_1 \leq 1} \frac{\alpha - 1 + \lambda^{ne}(\lambda_1 - \gamma + 1)}{\lambda^{ne}(\lambda_1 + 1)} = 1 - \sup_{\lambda_0 \leq 1, 0 \leq \lambda_1 \leq 1} \left(\frac{\alpha - 1}{\lambda^{ne}(\lambda_1 + 1)} + \frac{\lambda_1 - \gamma + 1}{\lambda_1 + 1} \right). \quad (22)$$

Since $0 \leq \lambda_1 \leq 1$ and $0 \leq \alpha \leq 1$, when $\lambda^{ne} \rightarrow 0^+$, the optimization problem in (22) tends towards positive infinity—that is, there is no solution. Therefore, from (20) and (21), we get

$$\hat{z}_{\mathcal{D}} \geq 1 - \alpha + \frac{\gamma}{2}.$$

Furthermore, by (13), we have

$$\hat{z}_{\mathcal{D}} = \sum_{i=1}^j \hat{p}_i(z_i, z_{i+1}), \text{ if } S^- = \emptyset \text{ or } z_j \leq U(x) < z_{j+1}, j = 1, 2, \dots, N.$$

A similar proof can be obtained for Cases 2 and 3.

Case 2: $S^+ = \emptyset$, which means that $A_k x - b_k < 0$ for all $k \in K$.

(i) When $L(x) \geq z_{N+1}$, this implies that $z_{\mathcal{D}} = 0$.

(ii) When $L(x) < z_1$, this leads to $z_{\mathcal{D}} = 1$. As a result, the chance constraints are always satisfied.

(iii) When $z_j \leq L(x) < z_{j+1}, j = 1, 2, \dots, N$, according to (14), it yields

$$\hat{z}_{\mathcal{D}} = \sum_{i=j}^N \hat{p}_i(z_i, z_{i+1}), \text{ if } S^+ = \emptyset \text{ or } z_j \leq L(x) < z_{j+1}, j = 1, 2, \dots, N.$$

Case 3: $S^+ \neq \emptyset$ and $S^- \neq \emptyset$ —that is, $\exists k_1, k_2 \in K$, with $A_{k_1} x - b_{k_1} > 0$ and $A_{k_2} x - b_{k_2} < 0$.

(i) When $L(x) > U(x)$, or $U(x) < z_1$, or $L(x) \geq z_{N+1}$, these lead to $z_{\mathcal{D}} = 0$.

(ii) When $L(x) \leq z_1$ and $U(x) \geq z_{N+1}$, then $z_{\mathcal{D}} = 1$. Thus, the chance constraints are always satisfied.

(iii) When $L(x) \in [z_j, z_{j+1}), U(x) \in [z_{j'}, z_{j'+1})$ with $j, j' \in \{1, 2, \dots, N\}$, it yields

$$\hat{z}_{\mathcal{D}} = \sum_{i=j}^{j'} \hat{p}_i(z_i, z_{i+1}), \text{ where } S^+ = \emptyset \text{ or } S^- = \emptyset \text{ or } (L(x) \in [z_j, z_{j+1}), U(x) \in [z_{j'}, z_{j'+1}), \text{ with } j, j' \in \{1, \dots, N\}.$$

The proof has been completed. $\square$

Remark 6. By Theorem 3, constraint (2b) is reformulated by a number of ordinary inequality constraints. Notably, different from the existing results, the derived deterministic equivalent formulation (DEF) of Model (1) is a linear programming model. It does not involve solving any complicated auxiliary optimization problems [8], a semi-definite programming problem [3], or a mixed-integer second-order cone programming problem.

3.2. Extension to Multistage Chance-Constrained Programming

We further extend the results in Proposition (2) and Theorem (3) to multistage DRCCPs. Different from Model (2), a multistage DRCCP can be given by

$$\begin{aligned} \min \quad & C(x) \\ \text{s.t.} \quad & \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\mathbf{A}(\boldsymbol{\xi})x \leq \mathbf{b}(\boldsymbol{\xi})\} \geq 1 - \alpha, \\ & x \in X \subseteq R^n, \end{aligned} \quad (23)$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_T)$ is a random vector, T is the number of stages, ξ_t represents the random parameter at the t -th stage, $\mathbf{A}(\boldsymbol{\xi}) \in R^{m \times n}$ is a random coefficient matrix, and $\mathbf{b}(\boldsymbol{\xi}) \in R^m$ is a random capacity vector. Suppose that $\xi_1, \dots, \xi_T$ are assumed to be mutually independent. Thus, the constraint $\mathbf{A}(\boldsymbol{\xi})x \leq \mathbf{b}(\boldsymbol{\xi})$ is equivalent to the intersection of m scalar constraints

$$\bigcap_{s=1}^m \{A_s(\boldsymbol{\xi})x \leq b_s(\boldsymbol{\xi})\},$$

where $A_s(\boldsymbol{\xi}) \in R^n$ is the s -th row of $\mathbf{A}$, and $b_s(\boldsymbol{\xi}) \in R$ is the s -th component of $\mathbf{b}$.

Let $\alpha \in (0, 1)$ represent the total risk level, i.e., the required confidence level is $1 - \alpha$. If the distribution of each ξ_t in Model (23) is not known precisely but belongs to a confidence set $\mathcal{D}_t$, constructed from historical data, based on (6), we know that $\mathcal{D}_t$ is defined by

$$\mathcal{D}_t = \left\{ \mathbb{P}^{(t)} \left| \begin{array}{l} p_{i_t}^{(t)} = \mathbb{P}^{(t)} \left\{ \xi_t = \frac{z_{i_t}^{(t)} + z_{i_t+1}^{(t)}}{2} \right\}, \quad i_t = 1, 2, \dots, N_t, \\ \sum_{i_t=1}^{N_t} p_{i_t}^{(t)} = 1, \\ \sum_{i_t=1}^{N_t} |p_{i_t}^{(t)} - \hat{p}_{i_t}^{(t)}(z_{i_t}^{(t)}, z_{i_t+1}^{(t)})| \leq \gamma_t, \\ 0 \leq p_{i_t}^{(t)} \leq 1, \quad i = 1, 2, \dots, N_t. \end{array} \right. \right\}, t = 1, \dots, T, \quad (24)$$

where $\hat{p}_{i_t}^{(t)}(z_{i_t}^{(t)}, z_{i_t+1}^{(t)})$ is the estimated probability of ξ_t falling in $[z_{i_t}^{(t)}, z_{i_t+1}^{(t)})$, and $\gamma_t \geq 0$ controls the confidence level of $\mathcal{D}_t$. Clearly, the overall confidence set for $\boldsymbol{\xi}$ is the product of these marginal sets, i.e., $\mathcal{D} = \mathcal{D}_1 \times \dots \times \mathcal{D}_T$.

For the multistage DRCCP (23), we can prove the following result.

Theorem 4. Let $\boldsymbol{\xi} = (\xi_1, \dots, \xi_T)$ be a random vector with independent components. Let the confidence set $\mathcal{D} = \mathcal{D}_1 \times \dots \times \mathcal{D}_T$, with $\mathcal{D}_t$ being defined by (24). If there exist $\alpha_s \in [0, 1]$ ($s = 1, \dots, m$), $\sum_{s=1}^m \alpha_s \leq \alpha$, such that

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}(A_s(\boldsymbol{\xi})x \leq b_s(\boldsymbol{\xi})) \geq 1 - \alpha_s, \quad s \in \{1, \dots, m\},$$

then

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P} \left(\bigcap_{s=1}^m \{A_s(\boldsymbol{\xi})x \leq b_s(\boldsymbol{\xi})\} \right) \geq 1 - \alpha.$$

Proof. Let $E_s = \{A_s(\xi)x \leq b_s(\xi)\}$ ($s = 1, \dots, m$) denote events that can potentially be dependent on the random vector ξ . For any joint distribution $\mathbb{P} \in \mathcal{D}$, the Bonferroni inequality (union bound) gives

$$\mathbb{P}\left(\bigcup_{s=1}^m E_s^c\right) \leq \sum_{s=1}^m \mathbb{P}(E_s^c),$$

where $E_s^c = \{A_s(\xi)x > b_s(\xi)\}$ is the complement of E_s . Therefore,

$$\mathbb{P}\left(\bigcap_{s=1}^m E_s\right) = 1 - \mathbb{P}\left(\bigcup_{s=1}^m E_s^c\right) \geq 1 - \sum_{s=1}^m \mathbb{P}(E_s^c).$$

Taking the infimum over $\mathbb{P} \in \mathcal{D}$ on both sides,

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\left(\bigcap_{s=1}^m E_s\right) \geq 1 - \sup_{\mathbb{P} \in \mathcal{D}} \sum_{s=1}^m \mathbb{P}(E_s^c).$$

For each s , the individual chance constraint condition $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}(E_s) \geq 1 - \alpha_s$ implies $\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{P}(E_s^c) = \sup_{\mathbb{P} \in \mathcal{D}} [1 - \mathbb{P}(E_s)] = 1 - \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}(E_s) \leq \alpha_s$. Thus,

$$\sup_{\mathbb{P} \in \mathcal{D}} \sum_{s=1}^m \mathbb{P}(E_s^c) \leq \sum_{s=1}^m \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{P}(E_s^c) \leq \sum_{s=1}^m \alpha_s.$$

Therefore,

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\left(\bigcap_{s=1}^m E_s\right) \geq 1 - \sum_{s=1}^m \alpha_s \geq 1 - \alpha.$$

We have completed the proof. $\square$

Remark 7. In the proof of Theorem 4, the Bonferroni inequality is only employed to derive a sufficient condition by which one can easily transform a joint chance constraint into a group of individual chance constraints, so as to obtain a computationally tractable reformulation of the original model for the given confidence level of the joint chance constraint. Although this inequality seems overly conservative, Theorem 4 enables a conservative approximation of the joint chance constraint by a group of individual chance constraints with easily allocated confidence levels, such as taking $\alpha_s = \alpha/m$ for all s . Clearly, this is valuable for future research so that a more compact inequality can be used to further optimize the confidence levels of individual chance constraints.

Remark 8. Note that Theorem 4 is proven under the assumption that the random variables are independent and the confidence set is a product of marginal sets, but, by this theorem, the results in Proposition 2 and Theorem 3 can be directly applied to handle the multistage DRCCP studied in the subsequent section.

4. Numerical Tests

4.1. Stochastic Multiperiod Capacitated Lot-Sizing Problems

To assess the performance of the proposed DRCCP approach, we carry out numerical tests on our method by solving the SMP-CLSP, one of the most fundamental inventory management problems [8,24]. Mathematically, the model of this problem reads as follows:

$$\min_{x,y} \sum_{t=1}^T (\bar{c}_t y_t + (c_t + H_t)x_t) \quad (25a)$$

$$\text{s.t. } \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P} \left(\begin{array}{ccc} x_1 & & \geq \xi_1 \\ x_1 + x_2 & & \geq \xi_1 + \xi_2 \\ \vdots & & \vdots \\ x_1 + x_2 + \dots + x_T & \geq & \xi_1 + \xi_2 + \dots + \xi_T \end{array} \right) \geq 1 - \alpha, \quad (25b)$$

$$x_t \leq M_t y_t \quad t = 1, \dots, T, \quad (25c)$$

$$x_t \in \mathbb{Z}^+, \quad t = 1, \dots, T, \quad (25d)$$

$$y_t \in \{0, 1\} \quad t = 1, \dots, T, \quad (25e)$$

where the initial inventory level is set to 0, T represents the time horizon, M_t denotes an upper bound on the quantity of units manufactured during period t , c_t stands for the unit production cost in period t , H_t is the unit holding cost for period t , $\bar{c}_t$ refers to the fixed setup cost per production run, ξ_t represents the demand during period t , x_t is the number of units to be produced in period t , $\mathbb{Z}^+$ represents a set of nonnegative integers, and y_t serves as a binary variable—it takes the value of 1 when a setup is performed during period t and 0 otherwise. More precisely, $y_t = 1$ when $x_t > 0$, and $y_t = 0$ when $x_t = 0$. x_t and y_t are decision variables, where $t = 1, \dots, T$.

The constraint (25b) ensures that the cumulative production up to period t meets the cumulative demand ξ_t with a higher probability $1 - \alpha$ under the worst-case distribution from $\mathcal{D}$. Here, $\mathcal{D}$ is the product of marginal confidence sets $\mathcal{D}_t$ for each ξ_t , as defined in Equation (24), and the random variables ξ_t are assumed independent. Thus, (25b) is equivalent to

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P} \left(\bigcap_{t=1}^T \left\{ \sum_{s=1}^t x_s \geq \sum_{s=1}^t \xi_s \right\} \right) \geq 1 - \alpha.$$

By Theorem 4, if we allocate risk levels α_t such that $\sum_{t=1}^T \alpha_t \leq \alpha$ satisfying

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P} \left(\sum_{s=1}^t x_s \geq \sum_{s=1}^t \xi_s \right) \geq 1 - \alpha_t, \quad t = 1, \dots, T, \quad (26)$$

then the inequality (25b) holds. For example, we can take equal risk allocation, i.e., $\alpha_t = \alpha/T$ for all t .

4.2. Reformulation of Models and Development of Algorithms

We next reformulate Problem (25) as an ordinary optimization problem.

Let the random variable ξ_s be approximated by a confidence set D_s defined in (24), $s = 1, \dots, t$. That is to say, ξ_s is discretized by $\xi_s = \eta_{i_s}^{(s)}$ with probabilities $p_{i_s}^{(s)}$, $i_s = 1, \dots, N_s$.

Define a random variable

$$\zeta_t = \sum_{s=1}^t \xi_s.$$

Then, ζ_t takes the value $\sum_{s=1}^t \eta_{i_s}^{(s)}$ with probability $\prod_{s=1}^t p_{i_s}^{(s)}$. Let $N_\zeta = \prod_{s=1}^t N_s$ denote the number of all values taken by ζ_t . Sorting all sums $\sum_{s=1}^t \eta_{i_s}^{(s)}$ in ascending order yields N_ζ nodes $\sum_{s=1}^t \eta_{i_s^{(l)}}^{(s)}$ with probability $\prod_{s=1}^t p_{i_s^{(l)}}^{(s)}$, $l = 1, 2, \dots, N_\zeta$. Denote all nodes by $\zeta_t^{(l)}$, $l = 1, 2, \dots, N_\zeta$. Then, the cumulative distribution function (CDF) of ζ_t is specified by the following step function:

$$F_t(z) = \begin{cases} 0, & \text{if } z < \zeta_t^{(1)}; \\ \sum_{q=1}^l \prod_{s=1}^t p_{i_s^{(q)}}^{(s)}, & \text{if } \zeta_t^{(l)} \leq z < \zeta_t^{(l+1)}, \text{ for } l = 1, 2, \dots, N_\zeta; \\ 1, & \text{if } z \geq \zeta_t^{(N_\zeta)}. \end{cases}$$

Consequently, for any given x_s , $s = 1, 2, \dots, t$, the value of $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\left(\sum_{s=1}^t x_s \geq \sum_{s=1}^t \zeta_s\right)$ is computed by solving the following auxiliary problem:

$$\begin{aligned} \min \quad & F\left(\sum_{s=1}^t x_s\right) = \sum_{q=1}^l \prod_{s=1}^t p_{i_s^{(q)}}^{(s)} \\ \text{s.t.} \quad & p_{i_s^{(q)}}^{(s)} \in D_s, \quad s = 1, 2, \dots, t, \\ & \sum_{s=1}^t \eta_{i_s^{(l)}}^{(s)} \geq \sum_{s=1}^t x_s. \end{aligned} \quad (27)$$

Indeed, if $p_{i_s^{(*)}}^{(s)}$ is the optimal solution of Problem (27), then its corresponding value is the minimal value of $\mathbb{P}\left(\sum_{s=1}^t x_s \geq \sum_{s=1}^t \zeta_s\right)$ in $\mathcal{D}$.

With a given distribution F_{rb} , such as the optimal solution $p_{i_s^{(*)}}^{(s)}$ of Problem (27), we construct a master problem to solve the SMP-CLSP (25) as follows.

$$\begin{aligned} \min_{x,y} \quad & \sum_{t=1}^T (\bar{c}_t y_t + (c_t + H_t)x_t) \\ \text{s.t.} \quad & \sum_{s=1}^t x_s \geq F_t^{*-1}(1 - \alpha_t), \quad \forall t = 1, 2, \dots, T, \\ & x_t \leq M_t y_t \quad t = 1, \dots, T, \\ & x_t \in \mathbb{Z}^+, \quad t = 1, \dots, T, \\ & y_t \in \{0, 1\} \quad t = 1, \dots, T. \end{aligned} \quad (28)$$

Model (28) effectively reformulates the original problem (25) as a deterministic programming model.

With the above preparations, we now develop an algorithm to solve the original DRCCP (25), where the master problem is a mixed-integer linear programming problem and is first solved for an initial approximate solution of the auxiliary problem; then, the auxiliary problem is solved for the obtained approximate solution of the master problem. The computational procedure is specified in Algorithm 1.

However, by Theorem 3, we can further reformulate Problem (25) as another form of standard constrained optimization problem.

Indeed, for each period $t = 1, \dots, T$, using the reference distribution $\hat{\mathbb{P}}_t$ at the t -th stage, Problem (26) corresponds to Case (1) in Proposition 2 under the condition $A_k x - b_k \geq 0$, where $A_k = 0$ and $b_k = -1$. The structure of the feasible set $\mathcal{F}_{cc}$, as characterized by Theorem 3, depends on the value of $U(x) = \sum_{s=1}^t x_s$:

- (1) If $U(x) < z_1$, part 1(i) implies that $\mathcal{F}_{cc} = \emptyset$.
- (2) If $U(x) \geq z_N$, set $\mathcal{C}_1^1$ in part 2 applies, yielding $\mathcal{F}_{cc} = \{x \in R^n : U(x) \geq z_N\}$.
- (3) If $z_j \leq U(x) < z_{j+1}$ for some $j \in \{1, \dots, N\}$, set $\mathcal{C}_1$ in part 3 applies, and the feasible set is given by

$$\left\{ x \in R^n : z_j \leq U(x) < z_{j+1}, \sum_{i=1}^j \hat{p}_i(z_i, z_{i+1}) \geq 1 - \alpha_t + \frac{\gamma_t}{2T} \right\}.$$

The reformulation of the chance constraint is explicitly given by the condition for set C_1 in the third case above. Therefore, for any x satisfying $z_j \leq U(x) < z_{j+1}$, the data-driven chance constraint (26) is equivalent to

$$\mathbb{P}_t \left(\sum_{s=1}^t x_s \geq \sum_{s=1}^t \zeta_s \right) \geq 1 - \alpha_t + \frac{\gamma_t}{2T}. \quad (29)$$

Let $\hat{F}_t$ be the estimated cumulative distribution function (CDF) of $\zeta_t = \sum_{s=1}^t \zeta_s$. Then, the constraint (29) becomes

$$\hat{F}_t \left(\sum_{s=1}^t x_s \right) \geq 1 - \alpha_t + \frac{\gamma_t}{2T}.$$

Thus, the deterministic reformulation of (25) is

$$\begin{aligned} \min_{x,y} \quad & \sum_{t=1}^T (\bar{c}_t y_t + (c_t + H_t) x_t) \\ \text{s.t.} \quad & \sum_{s=1}^t x_s \geq \hat{F}_t^{-1} \left(1 - \alpha_t + \frac{\gamma_t}{2T} \right), \quad t = 1, \dots, T, \\ & x_t \leq M_t y_t, \quad t = 1, \dots, T, \\ & x_t \in \mathbb{Z}^+, \quad t = 1, \dots, T, \\ & y_t \in \{0, 1\}, \quad t = 1, \dots, T. \end{aligned} \quad (30)$$

Algorithm 1 Alternating DRCCP-LS algorithm

Input: Time horizon T ; sample sizes: n_s for $s = 1, \dots, T$; sample data: $\{\zeta_s^j\}_{j=1}^{n_s}$ for $s=1, \dots, T$; interpolation tolerance: TH ; divergence tolerance: γ_t , $t = 1, \dots, T$; risk level α with $\alpha_t = \alpha/T$, $t = 1, \dots, T$; cost parameters: $\bar{c}_t$, c_t , H_t , $t = 1, \dots, T$; capacity parameters: M_t , $t = 1, \dots, T$; tolerance ϵ

Output: Optimal production plan (x^*, y^*) and the minimum total cost TC^*

- 1: Apply Rules (I)-(III) to determine the bin parameters N_s , $\eta_{i_s}^{(s)}$, and reference distributions $\hat{p}_{i_s}^{(s)}$, $i_s = 1, \dots, N_s$, $s = 1, \dots, T$
 - 2: Initialize a production plan, denoted by $x^{(0)} \leftarrow \mathbf{1}$; the total cost $TC^{(0)} \leftarrow \infty$; the iteration counter $k \leftarrow 0$; the convergence flag converged $\leftarrow$ false
 - 3: **while** not converged **do**
 - 4: Initialize robust quantiles $Q^{(k)} \leftarrow [0, \dots, 0]$ of length T
 - 5: **for** $t = 1$ to T **do**
 - 6: Compute cumulative production $z_t \leftarrow \sum_{s=1}^t x_s^{(k-1)}$
 - 7: Solve the auxiliary problem (27) with $x_s = x^{(k-1)}$. Denote its optimal solution by $\mathbb{P}_t^{(s)*}$ for $s = 1, \dots, t$
 - 8: With $\mathbb{P}_t^{(s)*}$, $s = 1, \dots, t$, define a distribution function F_t^* , and compute $Q^{(k)}[t] \leftarrow F_t^{*-1}(1 - \alpha_t)$
 - 9: **end for**
 - 10: With $Q^{(k)}$, solve the master problem (28). The optimal solution is referred to as $(x^{(k)}, y^{(k)})$, and set $TC^{(k)} \leftarrow \sum_{t=1}^T (\bar{c}_t y_t^{(k)} + (c_t + H_t) x_t^{(k)})$
 - 11: **if** $|TC^{(k)} - TC^{(k-1)}| < \epsilon$ **then**
 - 12: converged $\leftarrow$ true
 - 13: **else**
 - 14: $k \leftarrow k + 1$
 - 15: **end if**
 - 16: **end while**
 - 17: $x^* \leftarrow x^{(k)}$, $y^* \leftarrow y^{(k)}$, $TC^* \leftarrow TC^{(k)}$
 - 18: **return** (x^*, y^*) and TC^*
-

Problem (30) is also a mixed-integer linear programming (MILP) model, which can be solved using standard optimization solvers. By virtue of this reformulation defined by Model (30), we can present a more efficient algorithm to find a robust optimal production plan. The computational procedure is specified in Algorithm 2.

Algorithm 2 Reformulated DRCCP-LS algorithm

Input: Time horizon T ; sample sizes: n_s for $s = 1, \dots, T$; sample data: $\{\tilde{\zeta}_{s,j}^i\}_{j=1}^{n_s}$ for $s = 1, \dots, T$; interpolation tolerance: TH ; divergence tolerance: γ_t , $t = 1, \dots, T$; risk parameters: $\alpha, \alpha_t = \alpha/T$, $t = 1, \dots, T$; cost parameters: $\bar{c}_t, c_t, H_t$, $t = 1, \dots, T$; capacity parameters: M_t , $t = 1, \dots, T$

Output: Optimal production plan (x^*, y^*) and the minimum total cost TC^*

- 1: Apply Rules (I)-(III) to determine the bin parameters $N_s, \eta_{i_s}^{(s)}$, and reference distributions $\hat{p}_{i_s}^{(s)}$, $s = 1, \dots, T$
 - 2: Compute the estimated cumulative distribution functions $\hat{F}_t$ for $\zeta_t = \sum_{s=1}^t \zeta_s$, $t = 1, \dots, T$, using the reference distributions $\hat{p}_{i_s}^{(s)}$
 - 3: **for** $t = 1$ to T **do**
 - 4: Compute the quantile $\hat{Q}_t = \hat{F}_t^{-1}\left(1 - \alpha_t + \frac{\gamma_t}{2T}\right)$
 - 5: **end for**
 - 6: Solve the MILP problem (30) to obtain (x^*, y^*) and TC^*
 - 7: **return** (x^*, y^*) and TC^*
-

When only the traditional histogram, instead of the approach proposed in this paper, is employed to define the confidence set, Algorithm 2 is replaced by Algorithm 1.

Remark 9. Clearly, compared with Algorithm 2, Algorithm 1 involves alternative solutions of a mixed-integer linear master problem and an auxiliary nonlinear programming problem. In contrast, owing to the used reformulation technique in Model (30), Algorithm 2 is developed to directly solve a single mixed-integer linear program, being beneficial in the use of any off-the-shelf MILP solver. In the subsequent section, we will further validate its efficiency advantages through numerical tests.

4.3. Numerical Tests

In our computational experiments, we consider a planning horizon of $T = 2$ periods, which is sufficient to capture the essential dynamics of the multiperiod DRCCP. The initial inventory is initialized to zero. Both the holding cost and setup cost are assumed to be constant throughout the experiment, with values of 0.50 and 48, respectively. The production capacity is fixed at 50 units per period. The variable production costs are dependent on the period, and they are 28 and 15 when $t = 1$ and $t = 2$, respectively.

To simulate realistic and challenging demand patterns, we construct nonstandard demand distributions. The demand in each period is a nonnegative random variable, and demands across periods are mutually independent. For $t = 1$, the demand follows a mixture of two negative binomial distributions, $0.3\text{NB}(18, 0.8) + 0.7\text{NB}(56, 0.8116)$, where the parameters r and p in $\text{NB}(r, p)$ represent the number of successes and the success probability in each trial, respectively. This mixture distribution can capture the multimodal characteristics that are representative of an uncertain market. For $t = 2$, the demand is generated from a Poisson distribution with mean $\lambda = 15$, incorporating outliers that occur with a probability of 6%.

In Figures 2 and 3, we show the application of the proposed IDE-PLI method and the traditional histogram-based approach to estimating the probability density functions for the demand distributions at $t = 1$ and $t = 2$ under different sample sizes.

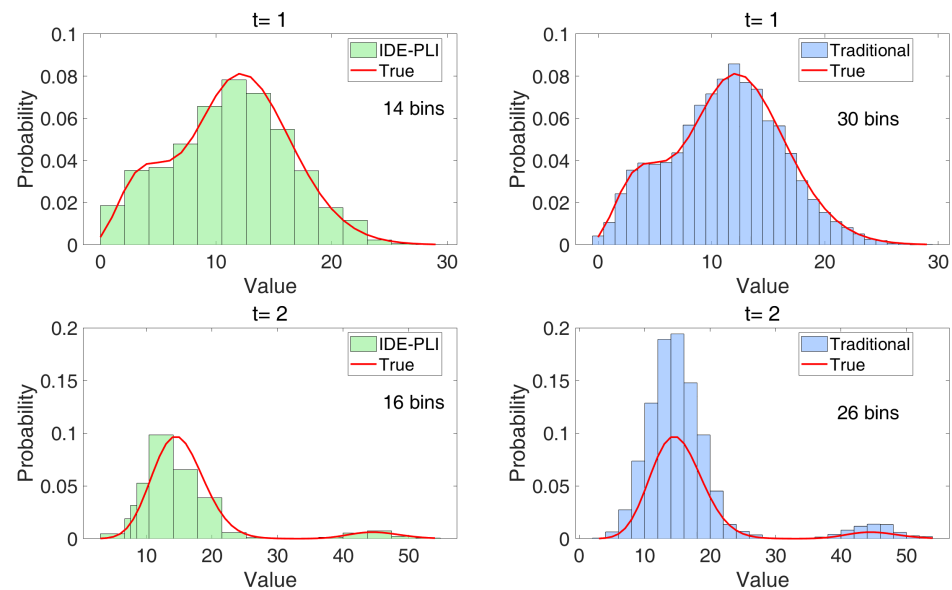


Figure 2. Distribution estimations with 5000 samples by the IDE-PLI method and the traditional histogram-based method.

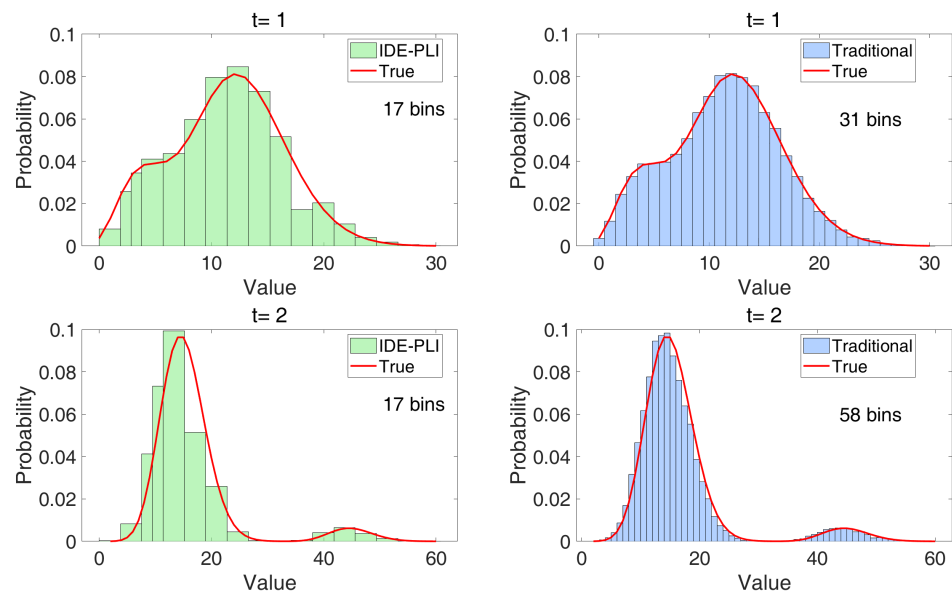


Figure 3. Distribution estimations with 20,000 samples by the IDE-PLI method and the traditional histogram-based method.

From the results in Figures 2 and 3, the following can be observed.

(1) When the sample size is 5000 and the data contain more outliers, the fixed binning strategy of the traditional histogram method results in the substantial deviation of the histogram shape from the true distribution (see the last sub-figure in Figure 2).

(2) Although the traditional histogram method can better capture the distribution characteristics when the sample size increases to 20,000 (see the last sub-figure in Figure 3), the number of bins in this method grows significantly, thereby increasing the model complexity. In contrast, the IDE-PLI method can effectively capture the underlying multimodal structure and tail behavior of the empirical data at both sample sizes without overfitting to sampling noise.

(3) The proposed IDE-PLI method always requires fewer bins compared to the traditional histogram method. This significant reduction can directly improve the computational efficiency in solving subsequent robust DEFs.

In order to assess the applicability of the proposed IDE-PLI method, we further apply this method to solve the SMP-CLSP, especially in comparison with other methods, when all of them are employed to construct distinct confidence sets. Besides Algorithm 1 (denoted by DRCCP-A) and Algorithm 2 (denoted by DRCCP-E), the other three compared methods are stated as follows.

CCC: The classical chance-constrained programming approach without dependence on constructing any confidence set, which assumes that the true demand distribution is known.

DRCCP-AT: An algorithm with the reference distribution in Algorithm 1 being replaced by the traditional histogram.

DRCCP-KL: The method in [25], where the classic empirical distribution is used as the reference distribution in the construction of confidence sets, and the confidence set is defined by the Kullback–Leibler divergence.

In Table 1, we present a comparison of the computational running times of various methods with the sample sizes of 5000 and 50,000 under different risk aversion levels.

Table 1. Running time comparison.

Sample Size	α	Time (s)				
		CCC	DRCCP-E	DRCCP-A	DRCCP-AT	DRCCP-KL
5000	0.05	0.42688	0.29209	13.371	17.849	51.000
	0.10	0.41052	0.28372	13.093	17.635	50.675
	0.15	0.38937	0.27614	13.261	17.658	51.035
50,000	0.05	0.40226	0.25176	13.819	32.574	60.924
	0.10	0.40113	0.23571	3.7973	32.282	59.093
	0.15	0.38458	0.25093	3.8887	32.151	59.934

From the numerical results in Table 1, the following conclusions can be drawn.

(1) The proposed DRCCP-E method exhibits remarkable computational efficiency and is even superior to the CCC method. The CCC model runs the fastest without considering the robustness of optimal solutions, but its assumption of a known true distribution is often impractical.

(2) As the sample size increases from 5000 to 50,000, DRCCP-A exhibits a significantly shorter running time compared with DRCCP-AT under the same setting of α . This result highlights the computational advantage of the adaptive binning strategy employed in DRCCP-A over the traditional histogram approach used in DRCCP-AT.

(3) The DRCCP-KL method consistently records the highest computational costs owing to the solution of complicated reformulated models therein. Indeed, its running time increases as the sample size grows, which underscores the numerical challenges associated with the Kullback–Leibler divergence constraint.

Considering changing risk levels (α) and changing divergence tolerances (γ), we conduct a sensitivity analysis of the minimized total cost and the optimal solutions derived from different methods, so as to validate the ability of the proposed data-driven robust approach in terms of achieving a trade-off between cost and robustness, as well as between optimality and robustness.

For this purpose, we change the value of the divergence tolerance from 0.03 to 0.09 with a step size of 0.02, and we change the value of α from 0.05 to 0.15 with a step size of 0.05. The minimized total cost and the corresponding optimal solutions across different combinations of α and γ are listed in Tables 2 and 3, respectively. In Table 3, each optimal solution is represented as $(x_1, x_2; y_1, y_2)$. All experimental results are derived from a training sample size of 50,000, ensuring statistical significance.

Table 2. Comparison of the total costs for different methods.

α	Method	Minimized Total Cost			
		$\gamma = 0.03$	$\gamma = 0.05$	$\gamma = 0.07$	$\gamma = 0.09$
$\alpha = 0.05$	CCC	1048	1048	1048	1048
	DRCCP-E	1336.5	1367.5	1398.5	1514.5
	DRCCP-A	1367.5	1472	1615	1619
	DRCCP-AT	1383	1526	1545.5	1657.5
	DRCCP-KL	1472	1418	1485	1471.5
$\alpha = 0.10$	CCC	972.5	972.5	972.5	972.5
	DRCCP-E	1201	1216.5	1232	1263
	DRCCP-A	1232	1278.5	1367.5	1441
	DRCCP-AT	1218.5	1278.5	1338.5	1412
	DRCCP-KL	1338.5	1383	1317	1303.5
$\alpha = 0.15$	CCC	912.5	912.5	912.5	912.5
	DRCCP-E	988	1019	1050	1112
	DRCCP-A	1096.5	1158.5	1216.5	1263
	DRCCP-AT	1112	1158.5	1218.5	1249.5
	DRCCP-KL	1247.5	1425	1350	1230

Table 3. Comparison of optimal solutions under different methods.

α	Method	Solution			
		$\gamma = 0.03$	$\gamma = 0.05$	$\gamma = 0.07$	$\gamma = 0.09$
$\alpha = 0.05$	CCC	(20,24;1,1)	(20,24;1,1)	(20,24;1,1)	(20,24;1,1)
	DRCCP-E	(23,37;1,1)	(23,39;1,1)	(23,41;1,1)	(27,41;1,1)
	DRCCP-A	(23,39;1,1)	(25,42;1,1)	(31,40;1,1)	(29,44;1,1)
	DRCCP-AT	(23,40;1,1)	(29,38;1,1)	(27,43;1,1)	(33,39;1,1)
	DRCCP-KL	(25,42;1,1)	(21,46;1,1)	(34,26;1,1)	(33,27;1,1)
$\alpha = 0.10$	CCC	(19,21;1,1)	(19,21;1,1)	(19,21;1,1)	(19,21;1,1)
	DRCCP-E	(21,32;1,1)	(21,33;1,1)	(21,34;1,1)	(21,36;1,1)
	DRCCP-A	(21,34;1,1)	(21,37;1,1)	(23,39;1,1)	(25,40;1,1)
	DRCCP-AT	(20,35;1,1)	(21,37;1,1)	(22,39;1,1)	(24,40;1,1)
	DRCCP-KL	(22,39;1,1)	(23,40;1,1)	(25,32;1,1)	(24,33;1,1)
$\alpha = 0.15$	CCC	(18,19;1,1)	(18,19;1,1)	(18,19;1,1)	(18,19;1,1)
	DRCCP-E	(19,22;1,1)	(19,24;1,1)	(19,26;1,1)	(19,30;1,1)
	DRCCP-A	(19,29;1,1)	(19,33;1,1)	(21,33;1,1)	(21,36;1,1)
	DRCCP-AT	(19,30;1,1)	(19,33;1,1)	(20,35;1,1)	(20,37;1,1)
	DRCCP-KL	(21,35;1,1)	(33,24;1,1)	(24,36;1,1)	(22,32;1,1)

From the results in Table 2, the following observations can be made.

(1) For the distributionally robust models (DRCCP-E, DRCCP-A, DRCCP-AT), the total cost increases as γ increases, reflecting the price of robustness against a larger confidence set. In contrast, the DRCCP-KL method displays nonmonotonic cost behavior, which can be attributed to the nonlinear nature of KL constraints, which often introduce numerical instabilities during optimization and lead to inconsistent convergence. The total cost of the CCC model remains constant since it is not influenced by γ . Among them, the performance of the DRCCP-A and DRCCP-AT methods is comparable, although the former outperforms the latter when $\gamma = 0.05$.

(2) For each of the fixed α values, the proposed DRCCP-E model consistently yields the lowest costs for the majority of the given values of γ among all compared robust methods.

This advantage in terms of the total cost underscores its effectiveness in striking a balance between economic efficiency and robustness.

Furthermore, from the results in Table 3, the following observations are derived.

(1) For all DRCCP methods except DRCCP-KL, the optimal quantity of single-period production is generally increasing with increased values of γ , and the total quantity of two-period production consistently also displays an increasing trend in accordance with theoretical expectations. Indeed, a larger value of γ can expand the size of the confidence set, enforcing robustness by a wider range of distributions but easily leading to an overconservative solution.

(2) The classical CCC method, which assumes a known distribution, always yields the smallest quantity of production among the compared methods since it does not account for distributional ambiguity. However, this advantage may disappear if the nominal distribution is misspecified.

Since any chance-constrained programming approach aims to minimize the constraint violation degree of the obtained optimal strategy when applied to unrealized scenarios of random model parameters in the future, we investigate the performance of the compared five methods in this context, i.e., the percentage of test samples that causes the demand to exceed the production capacity and the risk of constraint violation. In Table 4, we list the violation probabilities of the constraints under different risk levels and divergence tolerances, where 5000 samples are generated for the learning of the distribution and 500 samples are generated for an out-of-sample test.

Table 4. Violation probabilities of different methods under varying risk levels and divergence tolerances.

α	Method	Violation Probability			
		$\gamma = 0.03$	$\gamma = 0.05$	$\gamma = 0.07$	$\gamma = 0.09$
$\alpha = 0.05$	CCC	0.047	0.047	0.047	0.047
	DRCCP-E	0.043	0.040	0.020	0.010
	DRCCP-A	0.040	0.030	0.020	0.017
	DRCCP-AT	0.040	0.030	0.017	0.017
	DRCCP-KL	0.043	0.043	0.043	0.043
$\alpha = 0.10$	CCC	0.067	0.067	0.067	0.067
	DRCCP-E	0.047	0.047	0.047	0.043
	DRCCP-A	0.047	0.043	0.043	0.040
	DRCCP-AT	0.047	0.047	0.043	0.040
	DRCCP-KL	0.047	0.043	0.043	0.043
$\alpha = 0.15$	CCC	0.103	0.103	0.103	0.103
	DRCCP-E	0.050	0.050	0.047	0.047
	DRCCP-A	0.047	0.047	0.047	0.047
	DRCCP-AT	0.047	0.047	0.047	0.047
	DRCCP-KL	0.047	0.047	0.043	0.047

From the listed results in Table 4, we derive the following observations.

(1) Owing to the assumption of knowing the true demand distribution, the violation probability of the CCC method remains stable across different γ and is generally below the corresponding targeted level of α , which is attributed to finite-sample estimation errors in this method.

(2) All DRCCP methods exhibit consistently and significantly lower violation probabilities than those in the CCC method, further validating their characteristic superiority in view of robustness. By considering the ambiguity of the demand distribution, all DRCCP

methods can provide a more conservative production strategy, so as to hedge against the worst-case scenario.

(3) As γ increases, the violation probabilities of all DRCCP methods (DRCCP-E, DRCCP-A, DDRCCP-AT, and most cases of DRCCP-KL) decrease monotonically. This inverse relationship between γ and the violation probability demonstrates that one can effectively tune γ to achieve a desired risk protection level, albeit at the potential cost of higher production expenses.

(4) The primary practical advantage of the proposed DRCCP-E method lies in its balanced performance. Although it does not always achieve the lowest violation probability, it consistently maintains violation probabilities below the required risk level (α). From the previously conducted cost analysis, it is concluded that DRCCP-E generally achieves the minimum total cost among all compared methods. Thus, it is capable of successfully achieving a trade-off between reliability and cost, i.e., it can provide an optimal robust solution with an acceptable violation degree regarding constraints while avoiding the excessive costs associated with overconservative strategies.

(5) The violation probabilities of DRCCP-A and DRCCP-AT typically lie between those of CCC and DRCCP-E. This similar performance indicates that both of the histogram-based confidence sets (adaptive vs. traditional) offer comparable robustness. The observed minor variations between them arise from their differences in choosing the bins and allocating the probability mass.

In summary, it is concluded that all the DRCCP methods can reduce the violation probabilities compared to the CCC approach. The proposed DRCCP-E method provides the best balance between reliable risk control and a minimal production cost. The CCC method offers no protection against distributional misspecification. The divergence tolerance γ can serve as a critical model parameter employed to control the robustness levels in violating the constraints.

To conclude this section, we evaluate the performance of the compared distributionally robust optimization models when the confidence level and the sample size vary. In Figure 4, we present the numerical results obtained by taking values of $1 - \alpha$ from 0.85 to 0.95 under different sample sizes of 1000, 5000, 10,000, 20,000, and 50,000.

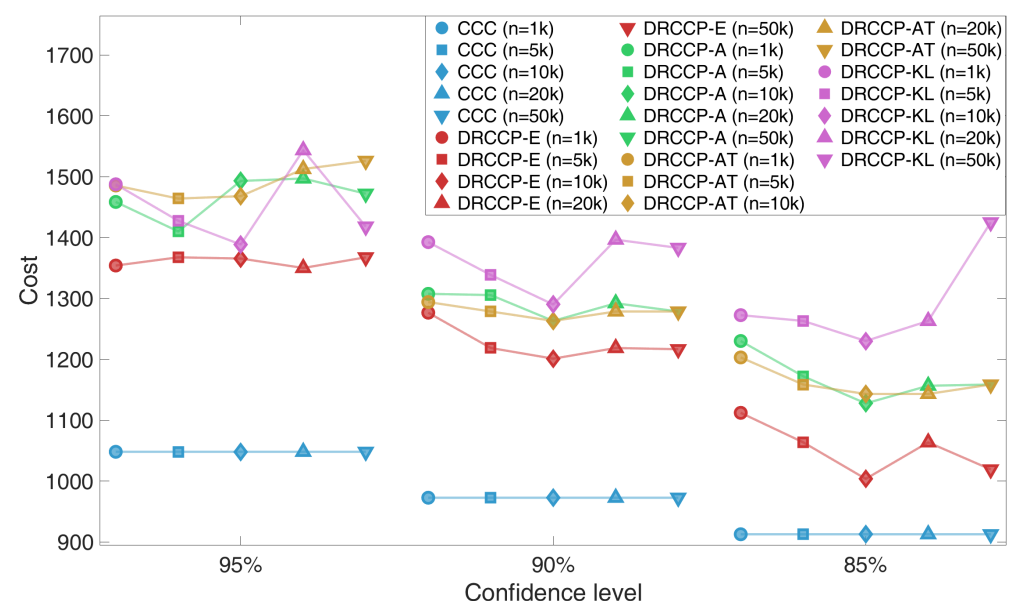


Figure 4. Cost values corresponding to different methods at various confidence levels when $\gamma = 0.05$.

Looking at Figure 4, the following conclusions can be drawn.

(1) The total cost obtained by each method consistently decreases as the confidence level decreases, except for DRCCP-KL. This fundamentally reflects the trade-off between risk and cost, as well as the inherent conservatism in robust optimization. A higher confidence level implies a low probability of stockout. Consequently, the model adopts a conservative strategy, typically involving higher production to build safety stock against demand uncertainty, thereby increasing the total cost. DRCCP-KL occasionally deviates from this trend, which can be attributed to the numerical instabilities introduced by the non-linear structure of the KL divergence constraints during optimization, as noted previously.

(2) The proposed DRCCP-E method displays remarkable stability across varying sample sizes. In contrast, DRCCP-A and DRCCP-AT show greater sensitivity to the sample size. Furthermore, as the confidence level decreases and the sample size increases, the minimal costs obtained by the DRCCP-A and DRCCP-AT methods converge. This convergence occurs because both methods are based on the same Algorithm 1, and, with a sufficiently large sample size, the distributions estimated by both the traditional histogram approach and the proposed IDE-PLI method converge to the true distribution. This result further validates the effectiveness of the proposed density estimation method.

(3) Across all settings of changing parameters, the proposed DRCCP-E method displays competitive performance, with the total cost lying between those of the conservative methods (DRCCP-AT, DRCCP-A, DRCCP-KL) and the nonrobust CCC benchmark.

In conclusion, the DRCCP-E method has the advantage of achieving a more appropriate balance between cost-effectiveness and robustness.

5. Conclusions

In this study, we propose a novel approach to constructing a data-driven adaptive confidence set for optimization problems with chance constraints, which only involves finitely many unknown parameters. By leveraging this confidence set, the original complicated chance constraint is reformulated into some tractable ordinary constraints. It is remarkable that, compared with existing works, the proposed reformulation does not involve solving complicated additional auxiliary optimization problems, a semi-definite programming problem, or a mixed-integer second-order cone programming problem.

We further expand the scope of this work to address stochastic multistage DRCCPs. A complete alternating solution method and a reformulation-based solution strategy are presented for the SMP-CLSP under demand uncertainty, along with detailed algorithmic frameworks established for each approach. In the alternating solution strategy, the auxiliary problem computes robust quantiles using nonlinear programming, while the master problem solves a deterministic mixed-integer programming problem. The resulting alternating DRCCP-LS algorithm efficiently handles distributional uncertainty while maintaining computational tractability. In the reformulation-based strategy, the DRCCP is transformed into a standard constrained optimization form, leading to the reformulated DRCCP-LS algorithm. Numerical experiments demonstrate that the proposed DRCCP-E method achieves a more appropriate balance between cost-effectiveness and robustness.

In future research, the proposed approach can be extended to handle chance constraints associated with multiple random factors, as well as developing efficient algorithms to solve the derived DEFs based on the property analysis of models. A straightforward yet computationally costly generalization method would be to construct a grid over the multidimensional sample space. Thus, investigating computational strategies such as adaptive partitioning or dimension reduction techniques will be essential. Additional directions could involve the extension of the proposed methodology to more complicated scenarios with time-varying confidence sets. Another promising direction is the application of the proposed methodology to practical operational decision-making problems in engineering

and management under uncertain environments. These could include robust supply chain management, portfolio optimization, production planning, and the robust utilization of renewable energy sources such as wind and solar power.

Author Contributions: Conceptualization, Z.W.; validation, H.D.; formal analysis, H.D. and Z.W.; investigation, H.D.; writing—original draft, H.D.; writing—review and editing, Z.W.; supervision, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Social Science Foundation of China (Grant No. 21BGL122).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: We declare that all authors have no conflicts of interest regarding the submission and publication of this paper.

References

1. Nemirovski, A.; Shapiro, A. Convex approximations of chance constrained programs. *SIAM J. Optim.* **2007**, *17*, 969–996. [\[CrossRef\]](#)
2. Calafiore, G.C.; Ghaoui, L.E. On distributionally robust chance-constrained linear programs. *J. Optim. Theory Appl.* **2006**, *130*, 1–22. [\[CrossRef\]](#)
3. Delage, E.; Ye, Y. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **2010**, *58*, 595–612. [\[CrossRef\]](#)
4. Jiang, N.; Xie, W. ALSO-X#: Better convex approximations for distributionally robust chance constrained programs. *Math. Program.* **2025**, *213*, 575–638.
5. Meng, Q.; Jin, X.; Luo, F.; Wang, Z.; Hussain, S. Distributionally robust scheduling for benefit allocation in regional integrated energy system with multiple stakeholders. *J. Mod. Power Syst. Clean Energy* **2024**, *12*, 1631–1642. [\[CrossRef\]](#)
6. Mohajerin Esfahani, P.; Kuhn, D. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Program.* **2018**, *171*, 115–166. [\[CrossRef\]](#)
7. Ning, C.; You, F. Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Comput. Chem. Eng.* **2019**, *125*, 434–448. [\[CrossRef\]](#)
8. Jiang, R.; Guan, Y. Data-driven chance constrained stochastic program. *Math. Program.* **2016**, *158*, 291–327. [\[CrossRef\]](#)
9. Zymler, S.; Kuhn, D.; Rustem, B. Distributionally robust joint chance constraints with second-order moment information. *Math. Program.* **2013**, *137*, 167–198. [\[CrossRef\]](#)
10. Shapiro, A.; Zhou, E.; Lin, Y. Bayesian distributionally robust optimization. *SIAM J. Optim.* **2023**, *33*, 1279–1304. [\[CrossRef\]](#)
11. Jiang, Y.; Ren, Z.; Li, W. Committed carbon emission operation region for integrated energy systems: Concepts and analyses. *IEEE Trans. Sustain. Energy* **2023**, *15*, 1194–1209. [\[CrossRef\]](#)
12. Kuhn, D.; Shafiee, S.; Wiesemann, W. Distributionally robust optimization. *Acta Numer.* **2025**, *34*, 579–804. [\[CrossRef\]](#)
13. Nguyen, V.A.; Kuhn, D.; Mohajerin Esfahani, P. Distributionally robust inverse covariance estimation: The Wasserstein shrinkage estimator. *Oper. Res.* **2022**, *70*, 490–515. [\[CrossRef\]](#)
14. Rahimian, H.; Mehrotra, S. Frameworks and results in distributionally robust optimization. *Open J. Math. Optim.* **2022**, *3*, 4. [\[CrossRef\]](#)
15. Chen, Z.; Kuhn, D.; Wiesemann, W. Data-driven chance constrained programs over Wasserstein balls. *Oper. Res.* **2024**, *72*, 410–424. [\[CrossRef\]](#)
16. Zhong, J.; Zhao, Y.; Li, Y.; Yan, M.; Peng, Y.; Cai, Y.; Cao, Y. Synergistic operation framework for the energy hub merging stochastic distributionally robust chance-constrained optimization and Stackelberg game. *IEEE Trans. Smart Grid* **2024**, *16*, 1037–1050. [\[CrossRef\]](#)
17. Küçükyavuz, S.; Jiang, R. Chance-constrained optimization under limited distributional information: A review of reformulations based on sampling and distributional robustness. *EURO J. Comput. Optim.* **2022**, *10*, 100030. [\[CrossRef\]](#)
18. Zhang, B.; Meng, L.L.; Lu, C.; Han, Y.Y.; Sang, H.Y. Automatic design of constructive heuristics for a reconfigurable distributed flowshop group scheduling problem. *Comput. Oper. Res.* **2024**, *161*, 106432. [\[CrossRef\]](#)
19. Pearson, K. Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. A* **1894**, *185*, 71–110. [\[CrossRef\]](#)
20. Scott, D.W. *Multivariate Density Estimation: Theory, Practice, and Visualization*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
21. Van Kerm, P. Adaptive kernel density estimation. *Stata J.* **2003**, *3*, 148–156. [\[CrossRef\]](#)
22. Hoeffding, W. Probability Inequalities for Sums of Bounded Random Variables. *J. Am. Stat. Assoc.* **1963**, *58*, 13–30. [\[CrossRef\]](#)
23. Levin, D.A.; Peres, Y. *Markov Chains and Mixing Times*; American Mathematical Society: Providence, RI, USA, 2017.

24. Beraldi, P.; Ruszczyński, A. A branch and bound method for stochastic integer problems under probabilistic constraints. *Optim. Methods Softw.* **2002**, *17*, 359–382. [[CrossRef](#)]
25. Chen, Y.; Guo, Q.; Sun, H.; Li, Z.; Wu, W.; Li, Z. A distributionally robust optimization model for unit commitment based on Kullback–Leibler divergence. *IEEE Trans. Power Syst.* **2018**, *33*, 5147–5160. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.