Data-driven Chance Constrained Stochastic Program

Ruiwei Jiang and Yongpei Guan
Department of Industrial and Systems Engineering
University of Florida, Gainesville, FL 32611, USA
Email: rwjiang@ufl.edu; guan@ise.ufl.edu

July 5, 2012

Abstract

Chance constrained programming is an effective and convenient approach to control risk in decision making under uncertainty. However, due to unknown probability distributions of random parameters, the solution obtained from a chance constrained optimization problem can be biased. In practice, instead of knowing the true distribution of a random parameter, only a series of historical data, which can be considered as samples taken from the true (while ambiguous) distribution, can be observed and stored. In this paper, we develop exact approaches to solve stochastic programs with Data-driven Chance Constraints (DCCs). For a given historical data set, we construct two types of confidence sets for the ambiguous distribution through nonparametric statistical estimation of its moments and density functions. We then formulate DCCs from the perspective of robust feasibility, by allowing the ambiguous distribution to run adversely within its confidence set. By deriving equivalent reformulations, we show that stochastic programs with DCCs under both moment-based and density-based confidence sets can be solved effectively. In addition, we derive the relationship between the conservatism of DCCs and the sample size of historical data, which shows quantitatively what we call the value of data.

Key words: stochastic optimization; chance constraints; semi-infinite programming; S-Lemma

1 Introduction

1.1 Motivation and Literature Review

To assist decision making in uncertain environments, significant research progress has been made in stochastic optimization formulations and their solution approaches. One effective and convenient way of handling uncertainty arising in constraint parameters employs chance constraints. In a chance constrained optimization problem, decision makers are interested in satisfying a constraint, which is subject to uncertainty, by at least a pre-specified probability at the smallest cost,

$$\min \quad \psi(x) \tag{1a}$$

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

$$x \in X, \tag{1c}$$

where $\psi : \mathbb{R}^n \to \mathbb{R}$ often represents a convex cost function, X represents a computable bounded convex set (e.g., a polytope) in $\mathbb{R}^n$, $\xi \in \mathbb{R}^K$ represents a K-dimensional random vector defined on a probability space $(\Omega, \mathcal{F}, \mu)$, and the set function $\mathbb{P}\{\cdot\}$ represents the probability distribution on $\mathbb{R}^K$ induced by ξ, i.e., $\mathbb{P}\{C\} = \mu\{\xi^{-1}(C)\}$, $\forall C \in \mathcal{B}(\mathbb{R}^K)$. In addition, the linear inequality system $A(\xi)x \leq b(\xi)$ represents the constraints to be satisfied, where $A(\xi) \in \mathbb{R}^{m\times n}$ and $b(\xi) \in \mathbb{R}^m$ denote the technology matrix and right hand side subject to uncertainty, and $x \in \mathbb{R}^n$ denotes the decision variable. Constraint (1b) is called a *single* chance constraint when $m = 1$ (i.e., the matrix $A(\xi)$ reduces to a row vector), and otherwise it is called a *joint* chance constraint. In addition, α value represents the risk level (or tolerance of constraint violation) allowed by the decision makers, and usually α is chosen to be small, e.g., 0.10 or 0.05.

Chance constraints emerge naturally as a modeling tool in various decision making circumstances. For example, decision makers in the finance industry may attempt to ensure that the return of their portfolio meets a target value with high probability. The study of chance constrained optimization problems has a long history dating back to Charnes et al. [8], Miller and Wagner [19], and Prékopa [27]. Unfortunately, constraint (1b) remains challenging to handle because of two key difficulties: (i) chance constraints are non-convex in general, and (ii) the probability associated with the chance constraints can be hard to compute since it requires a multi-dimensional integral. To address the first difficulty and recapture convexity, previous research identifies special cases under which chance constraints are nonlinear but convex (see, e.g., Charnes and Cooper [7], Prékopa [28], and Calafiore and El Ghaoui [6]), and provides conservative convex approximations (see, e.g., Pintér [25], Nemirovski and Shapiro [22], Rockafellar and Uryasev [29], and Chen et al. [9]). To address the second difficulty, previous research proposes scenario approximation approaches (see, e.g., Calafiore and Campi [4, 5], Nemirovski and Shapiro [21], and Luedtke and Ahmed [17]), which are computationally tractable and can guarantee to obtain a solution satisfying a chance constraint with high probability. In addition, Integer Programming (IP) techniques are successfully applied in exactly solving chance-constrained problems (see, e.g., Luedtke et al. [18], Küçükyavuz [15], and Luedtke [16]). An alternative of the chance constraint approach is the Robust Optimization (RO) approach (see, e.g., Soyster [34], Ben-Tal and Nemirovski [2], Bertsimas and Sim [3], and Ben-Tal et al. [1]), which requires the constraint $A(\xi)x \leq b(\xi)$ to be satisfied for each ξ in a pre-defined uncertainty set $U \subseteq \mathbb{R}^K$, i.e.,

$$\inf_{\xi \in U} \{A(\xi)x - b(\xi)\} \leq 0, \tag{2}$$

where the operator $\inf_{\xi \in U}\{\cdot\}$ is considered constraint-wise without loss of generality. One important merit of the RO approach is that, by a priori adjusting the uncertainty set U, one can ensure that the constraint $\mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha_0$ for a pre-specified risk level α_0 is satisfied under any probability distribution $\mathbb{P}$ (under certain mild conditions). Hence, the RO approach can be viewed as a conservative approximation of chance constraints.

A basic, and perhaps the most challenging, question on chance constraints is the accessibility of the probability distribution $\mathbb{P}$. Most literature on chance constraints listed above assumes the decision makers have perfect knowledge of $\mathbb{P}$. In practice, however, it might be unrealistic to make such an assumption. Normally, decision makers have only a series of historical data points, which can be considered as samples taken from the true (while ambiguous) distribution. Based on the given data set, there are two potential issues for the traditional chance constrained model (1): (i) it might be challenging to assume a specific probability distribution and to generate a large number of scenarios accordingly in the scenario approximations, and (ii) the solution might be sensitive to the ambiguous probability distribution and thus questionable in practice. To address these drawbacks, Distributionally Robust (or ambiguous) Chance Constrained (DRCC) models are proposed (see, e.g., Erdoğan and Iyengar [13], Calafiore and El Ghaoui [6], Nemirovski and Shapiro [22], Vandenberghe et al. [36], Zymler et al. [39], and Xu et al. [37]). In DRCC models, $\mathbb{P}$ is only assumed to belong to a pre-defined confidence set $\mathcal{D}$ (rather than being known with certainty), and the chance constraints are required to be satisfied under each probability distribution in $\mathcal{D}$:

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha. \tag{3}$$

Most current literature proposes approximation approaches to solve DRCC, and to the best of our knowledge, only Vandenberghe et al. [36] and Zymler et al. [39] consider exact solution approaches for the joint chance constraint version of DRCC. In this paper, we call the constraints in the form of

(3) Data-driven Chance Constraints (DCCs) because we attempt to construct the confidence set $\mathcal{D}$ based *only* on the historical data sampled from the true probability distribution and the statistical inferences obtained from the data. Intuitively, since real data is involved in its estimation, $\mathcal{D}$ gets tighter around the true probability distribution $\mathbb{P}$ with more data samples, and accordingly the DCC becomes less conservative. In this paper, following previous research, we propose *exact* solution approaches to handle joint DCCs under different types of confidence sets by deriving *equivalent* reformulations. Furthermore, we show the relationship between the conservatism of DCCs and the sample size of historical data, which depicts quantitatively the value of data.

1.2 Model Settings and Confidence Sets

In uncertain environments, people utilize historical data in various ways to help describe random parameters through statistical inference. For example, decision makers in the finance industry commonly describe the uncertainty in Rate of Return (RoR) of the investments by the first and second moments statistically inferred by the historical data. Hence, the confidence set $\mathcal{D}$ can be naturally defined as all probability distributions whose first and second moments agree with the inference, i.e.,

$$\mathcal{D}_1 = \{\mathbb{P} \in \mathcal{M}_+ : \quad \mathbb{E}[\xi] = \mu, \mathbb{E}[\xi\xi^\top] = \Sigma\}, \tag{4}$$

where $\mathcal{M}_+$ represents the set of all probability distributions, and μ and $\Sigma \succ \mu\mu^\top$ represent the inferred first and second moments, respectively. A more detailed description of RoR uncertainty can be obtained from Delage and Ye [11], where the uncertainty of moments is considered and the nonparametric bounds on the mean and covariance matrix are developed. In this case, the confidence set can be defined as follows:

$$\mathcal{D}_2 = \{\mathbb{P} \in \mathcal{M}_+ : \quad (\mathbb{E}[\xi] - \mu)^\top \Lambda^{-1} (\mathbb{E}[\xi] - \mu) \le \gamma_1, \ \mathbb{E}[(\xi - \mu)(\xi - \mu)^\top] \preceq \gamma_2 \Lambda\}, \tag{5}$$

where $\Lambda \succ 0$ represents the inferred covariance matrix, and $\gamma_1 > 0$ and $\gamma_2 > 1$ are two parameters obtained from the process of inference. One merit of moment-based confidence sets is that we can estimate them with a small amount of data, since usually the first and second moments of ξ can be effectively estimated by the sample mean and covariance.

Besides the moments, decision makers can resort to the density function of the random vector ξ. For example, power system operators often describe random wind power available in a time unit at a wind farm by estimating its density function. A natural extension of such a "point" estimation is a "confidence set" estimation built around the estimate, i.e., the operators might believe that although their estimate could suffer from some errors, the true density function is not too "far away" from it. One convenient and commonly used way of modeling the distance between density functions is by *Kullback-Leibler* (KL) divergence, which is defined as

$$D_{\text{KL}}(f||f_0) = \int_{\mathbb{R}^K} f(\xi) \log \frac{f(\xi)}{f_0(\xi)} d\xi,$$

where f and f_0 denote the true density function and its estimate, respectively. For general statistical divergence measures between density functions, interested readers are referred to Pardo [23] and Ben-Tal et al. [1], as well as the references therein. Based on KL divergence, the operators can build a confidence set as follows:

$$\mathcal{D}_3 = \{\mathbb{P} \in \mathcal{M}_+ : \quad D_{\text{KL}}(f||f_0) \leq d, \ f = d\mathbb{P}/d\xi\}, \tag{6}$$

where the divergence tolerance d can be chosen by the operators to represent their risk-aversion level, or can be obtained from statistical inference. To the best of our knowledge, in this paper we propose the first study on using $\mathcal{D}_3$ as a confidence set for joint DCCs, which can perform better than using $\mathcal{D}_1$ and $\mathcal{D}_2$ since we can depict the profile of the ambiguous distribution $\mathbb{P}$ more accurately by its density function than by its first two moments alone. Therefore, DCCs based on $\mathcal{D}_3$ can be less conservative than those based on $\mathcal{D}_1$ and $\mathcal{D}_2$.

The remainder of this paper is organized as follows. At the end of this section, we introduce notation and uncertainty settings to be used throughout this paper. We discuss the moment-based confidence sets ($\mathcal{D}_1$ and $\mathcal{D}_2$) in Section 2, and the density-based confidence set ($\mathcal{D}_3$) in Section 3. In both sections, we describe the construction of confidence sets and show how to equivalently reformulate DCCs under different confidence sets. We also discuss solution approaches for data-driven chance constrained programs (DCCPs) by using their equivalent reformulations. In addition, we discover the relationship between the conservatism of DCCs and the sample size of historical data under the density-based confidence set, which shows quantitatively the value of data. Finally, we summarize this paper in Section 5.

Notation and Uncertainty Settings. We specify the technology matrix $A(\xi)$ and right hand side $b(\xi)$ by assuming that $A(\xi)$ and $b(\xi)$ are affinely dependent on ξ (a K-dimensional vector in the form $\xi = (\xi_1, \ldots, \xi_K)^\top$), i.e.,

$$A(\xi) = A_0 + \sum_{k=1}^{K} A_k \xi_k, \quad b(\xi) = b_0 + \sum_{k=1}^{K} b_k \xi_k, \tag{7}$$

where A_0 and b_0 represent the deterministic part of $A(\xi)$ and $b(\xi)$, and each matrix A_k and vector b_k consists of the coefficients subject to random variable ξ_k. This uncertainty setting has been adopted in the literature of stochastic programming and robust optimization, and the readers are referred to Chen et al. [9] and Chen and Zhang [10] for more examples. Under this uncertainty setting, we can reformulate constraint $A(\xi)x \le b(\xi)$ as follows:

$$A(\xi)x \le b(\xi) \quad \Leftrightarrow \quad A_0 x + \sum_{k=1}^{K} (A_k x)\xi_k \le b_0 + \sum_{k=1}^{K} b_k \xi_k \quad \Leftrightarrow \quad \bar{A}(x)\xi \le \bar{b}(x),$$

where vector $\bar{b}(x) = b_0 - A_0 x$, and $\bar{A}(x)$ is an $m \times K$ matrix defined as

$$\bar{A}(x) = \begin{bmatrix} A_1 x - b_1, & A_2 x - b_2, & \ldots, & A_K x - b_K \end{bmatrix}.$$

Therefore, DCC (3) can be generally reformulated as

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi \in C\} \ge 1 - \alpha, \tag{8}$$

where $C = \{\xi \in \mathbb{R}^K : \ \bar{A}\xi \le \bar{b}\}$ is a polyhedron whose parameters $\bar{A}$ and $\bar{b}$ depend upon x. In the remainder of this paper, we use DCC (3) and its general reformulation (8) interchangeably for notation brevity.

2 DCC with Moment-based Confidence Set

In this section, we consider constraint (8) with $\mathcal{D} = \mathcal{D}_1$ and $\mathcal{D} = \mathcal{D}_2$. First, we discuss the construction of $\mathcal{D}_1$ and $\mathcal{D}_2$ by using a series of independent random samples $\{\xi^i\}_{i=1}^N$ obeying the true distribution $\mathbb{P}$. Second, we show how to reformulate the left hand side of (8), i.e., the subproblem for the worst-case probability bound $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi \in C\}$, as a Semidefinite Programming (SDP) problem with polyhedron C replaced by its interior, $int(C) = \{\xi \in \mathbb{R}^K : \bar{A}\xi < \bar{b}\}$. The derivation of the reformulation utilizes the S-Lemma (see, e.g, Yakubovich [38] and Pólik and Terlaky [26]) and provides a general technique for handling DCCs with moment informations. Third, we extend the application of this technique to the case with $\mathcal{D}_2$ and also obtain an SDP reformulation. Fourth, we show the continuity of $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi \in C\}$ over the polyhedron C, i.e.,

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi \in C\} = \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi \in int(C)\},$$

which accomplishes the reformulation of (8). Finally, we discuss the solution approaches to solve DCCPs. With the above five steps combined, we build a solution framework for optimization problems with DCCs. Before continuing, we highlight the contributions of our work as compared to those of Vandenberghe et al. [36] and Zymler et al. [39]. Vandenberghe et al. [36] equivalently reformulate a strict version of the joint DCCs (i.e., $\mathbb{P}\{A(\xi)x < b(\xi)\}$) and obtain a different SDP reformulation similar to our first result, and Zymler et al. [39] develop an equivalent reformulation for the single DCCs and a worst-case CVaR-based approximation for the joint DCCs. First, as compared to Vandenberghe et al. [36], we show that the reformulation also works for the soft version of the joint DCCs (i.e., $\mathbb{P}\{A(\xi)x \leq b(\xi)\}$), which makes the analysis complete. Second, as compared to Zymler et al. [39], we develop an *equivalent* reformulation for the joint DCCs. Third, as compared to both Vandenberghe et al. [36] and Zymler et al. [39], we develop a more general approach to obtain the result, which can easily be extended to obtain reformulations under other forms of moment information (e.g., $\mathcal{D}_2$). The extension might not be obvious from their approaches.

2.1 Construction of Confidence Sets

In this subsection, we discuss the construction of confidence sets $\mathcal{D}_1$ and $\mathcal{D}_2$. Given a series of independent samples $\{\xi^i\}_{i=1}^N$ drawn from the true distribution $\mathbb{P}$ of the random vector ξ, we want to estimate the first and second moments of ξ. First, the point estimate of the first two moments can be obtained by the sample moments

$$\mu = \frac{1}{N}\sum_{i=1}^{N}\xi^i, \qquad \Sigma = \frac{1}{N}\sum_{i=1}^{N}\xi^i(\xi^i)^\top,$$

where μ and Σ are unbiased estimators of $\mathbb{E}[\xi]$ and $\mathbb{E}[\xi\xi^\top]$, respectively. Hence, $\mathcal{D}_1$ can be constructed by letting $\mathbb{E}[\xi]$ and $\mathbb{E}[\xi\xi^\top]$ equal their estimates, i.e.,

$$\mathcal{D}_1 = \{\mathbb{P} \in \mathcal{M}_+ : \ \ \mathbb{E}[\xi] = \mu, \ \ \mathbb{E}[\xi\xi^\top] = \Sigma\}.$$

Second, we apply the result proposed in Delage and Ye [11] to construct a nonparametric confidence set for the mean and covariance matrix of ξ as follows:

$$\mathcal{D}_2 = \Big\{\mathbb{P} \in \mathcal{M}_+ : \ (\mathbb{E}[\xi] - \mu)^\top \Lambda^{-1}(\mathbb{E}[\xi] - \mu) \leq \gamma_1, \ \ \mathbb{E}[(\xi - \mu)(\xi - \mu)^\top] \preceq \gamma_2\Lambda\Big\},$$

where $\mu = \sum_{i=1}^N \xi^i$, $\Lambda = (1/N)\sum_{i=1}^N(\xi^i - \mu)(\xi^i - \mu)^\top$, and the parameters $\gamma_1 > 0$ and $\gamma_2 > 1$ can be obtained from the process of inference. For the sake of brevity, we do not restate their results in this paper and the interested readers are referred to, e.g, Theorem 2 in Delage and Ye [11].

2.2 Reformulation of Worst-case Probability Bound

In this subsection, we investigate the reformulation of the worst-case probability bound

$$z_{\mathcal{D}_1} = \inf_{\mathbb{E}[\xi]=\mu, \mathbb{E}[\xi\xi^\top]=\Sigma} \mathbb{P}\{\bar{A}\xi < \bar{b}\}. \tag{9}$$

The implication for $z_{\mathcal{D}_1}$ can be easily extended to the case of $\mathcal{D}_2$. Before going into the reformulation details, we first review the well-known S-Lemma, which will be applied throughout the remaining part of this section.

Theorem 1 *(S-Lemma, Yakubovich [38]) Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be quadratic functions and suppose that there is an $\bar{x} \in \mathbb{R}^n$ such that $g(\bar{x}) < 0$. Then the following two statements are equivalent.*

(i) There is no $x \in \mathbb{R}^n$ such that

$$\begin{cases} f(x) < 0 \\ g(x) \leq 0. \end{cases}$$

(ii) There is a nonnegative number $y \geq 0$ such that

$$f(x) + yg(x) \geq 0, \quad \forall x \in \mathbb{R}^n.$$

Second, We summarize our main result in this subsection by the following theorem.

Theorem 2 *Given matrix $\bar{A}$, vector $\bar{b}$, first moment μ, and second moment Σ, the worst-case probability bound $z_{\mathcal{D}_1} = \inf_{\mathbb{P}\in\mathcal{D}_1} \mathbb{P}\{\bar{A}\xi < \bar{b}\}$ with the confidence set $\mathcal{D}_1 = \{\xi \in \mathbb{R}^K : \mathbb{E}[\xi] = \mu, \mathbb{E}[\xi\xi^\top] = \Sigma\}$ equals the optimal objective value of the following SDP:*

$$\begin{aligned} z_{\mathcal{D}_1} = \max_{H,p,y,q} \quad & \Sigma \cdot H + \mu^\top p + q && \text{(10a)} \\ \text{s.t.} \quad & \begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1-q \end{bmatrix} \succeq 0, && \text{(10b)} \\ & \begin{bmatrix} -H & -\frac{1}{2}(p + y_i a_i) \\ -\frac{1}{2}(p+y_i a_i)^\top & y_i \bar{b}_i - q \end{bmatrix} \succeq 0, \quad \forall i = 1, \ldots, m, && \text{(10c)} \\ & y \geq 0, \quad H \in \mathbb{S}^{K\times K}, && \text{(10d)} \end{aligned}$$

where $\mathbb{S}^{K\times K}$ denotes the set of $K \times K$ symmetric matrices, $a_1^\top, \ldots, a_m^\top$ denote the m row vectors consisting of the matrix $\bar{A}$, and the operator "$\cdot$" in the objective function (10a) *represents the Frobenius inner product.*

Proof: First, we specify $z_{\mathcal{D}_1}$ by the following optimization problem:

$$[\text{Primal}] \quad z_{\mathcal{D}_1} = \inf_{\mathbb{E}[\xi]=\mu, \mathbb{E}[\xi\xi^\top]=\Sigma} \mathbb{P}\{\bar{A}\xi < \bar{b}\}$$

$$= \min_{\mathbb{P}} \int_{\mathbb{R}^K} I_{[\bar{A}\xi<\bar{b}]}(\xi) d\mathbb{P} \tag{11a}$$

$$\text{s.t.} \int_{\mathbb{R}^K} \xi d\mathbb{P} = \mu, \tag{11b}$$

$$\int_{\mathbb{R}^K} \xi\xi^\top d\mathbb{P} = \Sigma, \tag{11c}$$

$$\int_{\mathbb{R}^K} d\mathbb{P} = 1, \tag{11d}$$

where $I_{[\bar{A}\xi<\bar{b}]}(\cdot) : \mathbb{R}^K \to \{0, 1\}$ represents the indicator function of $int(C)$, constraints (11b) and (11c) describe the first and second moments of ξ respectively, and constraint (11d) guarantees we are considering probability distributions on $\mathbb{R}^K$. We apply the duality theory for conic linear programming problems and dualize problem (11) as

$$[\text{Dual}] \quad z_{\mathcal{D}_1} = \max_{H,p,q} \quad \Sigma \cdot H + \mu^\top p + q \tag{12a}$$

$$\text{s.t.} \quad \xi^\top H\xi + p^\top \xi + q \leq I_{[\bar{A}\xi<\bar{b}]}(\xi), \quad \forall \xi \in \mathbb{R}^K, \tag{12b}$$

where vector $p \in \mathbb{R}^K$, symmetric matrix $H \in \mathbb{R}^{K\times K}$ (i.e., $H \in \mathbb{S}^{K\times K}$), and scalar q represent the dual variables for constraints (11b), (11c), and (11d), respectively. Note that strong duality holds for problems [Primal] and [Dual] based on established conic linear programming theory (cf. Isii [14], Smith [33], and Shapiro [31]).

Second, we reformulate constraint (12b). Letting $a_1^\top, \ldots, a_m^\top$ denote the m row vectors consisting of the matrix $\bar{A}$, we first observe that

$$I_{[\bar{A}\xi<\bar{b}]}(\xi) = \prod_{i=1}^{m} I_{[a_i^\top \xi<\bar{b}_i]}(\xi), \quad \forall \xi \in \mathbb{R}^K,$$

and hence, constraint (12b) is equivalent to

$$\xi^\top H\xi + p^\top \xi + q \leq I_{[a_i^\top \xi<\bar{b}_i]}(\xi), \quad \forall \xi \in \mathbb{R}^K, \forall i = 1, \ldots, m, \tag{13}$$

which is further equivalent to

$$\begin{cases} \xi^\top H\xi + p^\top \xi + q \leq 1, \quad \forall \xi \in \mathbb{R}^K, & (14a) \\ a_i^\top \xi \geq \bar{b}_i \quad \Rightarrow \quad \xi^\top H\xi + p^\top \xi + q \leq 0, \quad \forall \xi \in \mathbb{R}^K, \ \forall i = 1, \ldots, m. & (14b) \end{cases}$$

Next we discuss the reformulation of constraint (14a) and implication (14b). First, we observe that constraint (14a) equivalently requires that the quadratic function $\xi^\top H\xi + p^\top \xi + q - 1$ is nonpositive everywhere in $\mathbb{R}^K$, and hence, is equivalent to

$$\begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1-q \end{bmatrix} \succeq 0. \tag{15}$$

Second, the implication (14b) is equivalent to the statement that the quadratic system

$$\begin{cases} \xi^\top H\xi + p^\top \xi + q > 0 \\ a_i^\top \xi - \bar{b}_i \geq 0 \end{cases} \tag{16}$$

has no solution in $\mathbb{R}^K$ for each $i = 1, \ldots, m$. We claim that this is equivalent to the following (infinite) quadratic system having a nonnegative solution y_i:

$$-(\xi^\top H\xi + p^\top \xi + q) - y_i(a_i^\top \xi - \bar{b}_i) \geq 0, \quad \forall \xi \in \mathbb{R}^K. \tag{17}$$

To see this, we discuss the following cases:

Case 1. If $a_i \neq 0$, we observe that there exists a $\bar{\xi} \in \mathbb{R}^K$ such that $a_i^\top \bar{\xi} - \bar{b}_i > 0$ since we can choose $\bar{\xi}$ to make $a_i^\top \bar{\xi}$ arbitrarily large. In this case, the assumptions of S-Lemma are satisfied, and hence, the equivalence is guaranteed.

Case 2. If $a_i = 0$ and $\bar{b}_i \leq 0$: First, suppose that (16) has no solution. Since $a_i^\top \xi - \bar{b}_i = -\bar{b}_i \geq 0$ is satisfied, it follows that $\xi^\top H\xi + p^\top \xi + q > 0$ has no solution, i.e., $\xi^\top H\xi + p^\top \xi + q \leq 0$ for all $\xi \in \mathbb{R}^K$, which implies that (17) has a solution $y_i = 0$. Second, suppose that (17) has a solution $y_i \geq 0$. Since $a_i = 0$ and $\bar{b}_i \leq 0$, it follows that $-(\xi^\top H\xi + p^\top \xi + q) \geq -y_i\bar{b}_i \geq 0$, which implies that (16) has no solution. Hence, the equivalence is guaranteed in this case.

Case 3. If $a_i = 0$ and $\bar{b}_i > 0$: Since $a_i^\top \xi - \bar{b}_i = -\bar{b}_i < 0$, (16) has no solution, and we only need to show that (17) has a nonnegative solution. But in view of (14a), we know that $y_i = 1/\bar{b}_i > 0$ is a solution for (17).

We have proved the equivalence between (16) and (17). Another way of stating (17) is that there exists some $y_i \geq 0$, such that the quadratic function $-\xi^\top H\xi - (p + y_i a_i)^\top \xi + y_i\bar{b}_i - q$ is nonnegative everywhere in $\mathbb{R}^K$, which is equivalent to

$$\begin{bmatrix} -H & -\frac{1}{2}(p + y_i a_i) \\ -\frac{1}{2}(p + y_i a_i)^\top & y_i\bar{b}_i - q \end{bmatrix} \succeq 0, \quad y_i \geq 0, \quad \forall i = 1, \ldots, m. \tag{18}$$

Therefore, we have equivalently reformulated constraint (12b) as constraints (10b)-(10d), which completes the proof. ■

Remark 1 *We summarize the equivalence implications as follows. We mark each equivalence relationship by a "⇔," and highlight where the S-Lemma is used.*

$$
(12b) \;\Leftrightarrow\; (13) \;\Leftrightarrow\; \begin{cases} (14a) \;\Leftrightarrow\; (10b) \\ (14b) \;\Leftrightarrow\; (16) \;\Leftarrow \textit{S-Lemma} \Rightarrow\; (17) \;\Leftrightarrow\; \begin{cases} (10c) \\ (10d) \end{cases} \end{cases}
$$

It is convenient to see that all the implications discussed above can be extended to the case with confidence set $\mathcal{D}_2$. We present the following parallel conclusion and provide the detailed proof in Appendix B.

Corollary 1 *Given matrix $\bar{A}$, vector $\bar{b}$, mean μ, covariance matrix Λ, and constants $\gamma_1 > 0$, $\gamma_2 > 1$, the worst-case probability bound $z_{\mathcal{D}_2} = \inf_{\mathbb{P}\in\mathcal{D}_2} \mathbb{P}\{\bar{A}\xi < \bar{b}\}$ with the confidence set $\mathcal{D}_2 = \{\mathbb{P} \in \mathcal{M}_+ :\ (\mathbb{E}[\xi]-\mu)^\top \Lambda^{-1}(\mathbb{E}[\xi]-\mu) \le \gamma_1,\ \mathbb{E}[(\xi-\mu)(\xi-\mu)^\top] \preceq \gamma_2\Lambda\}$ equals the optimal objective value of the following SDP:*

$$
\begin{aligned}
z_{\mathcal{D}_2} = \quad & \max_{G,H,p,y,r,q} && (\mu\mu^\top - \gamma_2\Lambda)\cdot G - \Lambda\cdot H + 2\mu^\top p - \gamma_1 q + r && \text{(19a)}\\
& \text{s.t.} && \begin{bmatrix} G & -(p + G\mu + \frac{1}{2}y_i a_i) \\ -(p + G\mu + \frac{1}{2}y_i a_i)^\top & y_i \bar{b}_i - r \end{bmatrix} \succeq 0, && \text{(19b)}\\
& && \begin{bmatrix} G & -(p + G\mu) \\ -(p + G\mu)^\top & 1 - r \end{bmatrix} \succeq 0, \quad \forall i = 1,\ldots,m, && \text{(19c)}\\
& && G \succeq 0, \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \succeq 0, y \ge 0, \quad G \in \mathbb{S}^{K\times K},\ H \in \mathbb{S}^{K\times K}, && \text{(19d)}
\end{aligned}
$$

where $\mathbb{S}^{K\times K}$ denotes the set of $K\times K$ symmetric matrices, and $a_1^\top,\ldots,a_m^\top$ denote the m row vectors consisting of the matrix $\bar{A}$.

A variant of $\mathcal{D}_2$ is to estimate the first two moments of each component ξ_k separately. That is, if we do not consider the correlation of each pair ξ_k and ξ_ℓ, we can build confidence intervals for $\mathbb{E}[\xi_k]$ and $\mathbb{E}[\xi_k^2]$ for $k = 1,\ldots,K$, which results in

$$
\mathcal{D}_2' = \left\{\mathbb{P} \in \mathcal{M}_+ :\ \mu_k^{\mathrm{L}} \le \mathbb{E}[\xi_k] \le \mu_k^{\mathrm{U}},\ \sigma_k^{\mathrm{L}} \le \mathbb{E}[\xi_k^2] \le \sigma_k^{\mathrm{U}},\ \forall k = 1,\ldots,K\right\},
$$

where $[\mu_k^{\mathrm{L}}, \mu_k^{\mathrm{U}}]$ and $[\sigma_k^{\mathrm{L}}, \sigma_k^{\mathrm{U}}]$ are confidence intervals of $\mathbb{E}[\xi_k]$ and $\mathbb{E}[\xi_k^2]$, respectively. It is easier to construct $\mathcal{D}_2'$ than $\mathcal{D}_2$, since we only consider the marginal distribution of the random vector ξ.

Furthermore, we can use a smaller size data set to construct $\mathcal{D}_2'$ than to construct $\mathcal{D}_2$ since we do not have to estimate the correlations. In addition, we can extend Theorem 2 to this case as well. The resulting reformulation for $\mathcal{D}_2'$ is a Second-Order Cone Program (SOCP), which can be solved more efficiently than SDP. Likewise, we provide the detailed proof in Appendix C for brevity.

Corollary 2 *Given matrix $\bar{A}$, vector $\bar{b}$, $[\mu_k^L, \mu_k^U]$ (the confidence interval of $\mathbb{E}[\xi_k]$), and $[\sigma_k^L, \sigma_k^U]$ (the confidence interval of $\mathbb{E}[\xi_k^2]$), the worst-case probability bound $z_{\mathcal{D}_2'} = \inf_{\mathbb{P}\in\mathcal{D}_2'} \mathbb{P}\{\bar{A}\xi < \bar{b}\}$ with the confidence set $\mathcal{D}_2' = \{\mathbb{P} \in \mathcal{M}_+ : \ \mu_k^L \leq \mathbb{E}[\xi_k] \leq \mu_k^U, \ \sigma_k^L \leq \mathbb{E}[\xi_k^2] \leq \sigma_k^U, \ \forall k = 1, \ldots, K\}$ equals the optimal objective value of the following SOCP:*

$$
\begin{aligned}
z_{\mathcal{D}_2'} = \quad & \max_{h,p,q,t,s} \quad \sum_{k=1}^{K} \big(\mu_k^L p_k^L - \mu_k^U p_k^U + \sigma_k^L h_k^L - \sigma_k^U h_k^U\big) + q && \text{(20a)}\\
& \quad s.t. \quad q + \sum_{k=1}^{K} t_k \leq 1, && \text{(20b)}\\
& \qquad\qquad \left\| \begin{bmatrix} p_k^L - p_k^U \\ t_k + h_k^L - h_k^U \end{bmatrix} \right\| \leq t_k - h_k^L + h_k^U, \quad \forall k = 1, \ldots, K, && \text{(20c)}\\
& \qquad\qquad q + \sum_{k=1}^{K} s_{ik} \leq y_i \bar{b}_i, \quad \forall i = 1, \ldots, m, \ \forall k = 1, \ldots, K, && \text{(20d)}\\
& \qquad\qquad \left\| \begin{bmatrix} p_k^L - p_k^U + y_i (a_i)_k \\ s_{ik} + h_k^L - h_k^U \end{bmatrix} \right\| \leq s_{ik} - h_k^L + h_k^U, \quad \forall i = 1, \ldots, m, \ \forall k = 1, \ldots, K, && \text{(20e)}\\
& \qquad\qquad t_k \geq 0, \ s_{ik} \geq 0, \quad \forall i = 1, \ldots, m, \ \forall k = 1, \ldots, K, && \text{(20f)}\\
& \qquad\qquad p^L, p^U, h^L, h^U \geq 0, \quad y \geq 0, && \text{(20g)}
\end{aligned}
$$

where $a_1^\top, \ldots, a_m^\top$ denote the m row vectors consisting of the matrix $\bar{A}$.

2.3 Continuity of the Probability Bound

In this subsection, we show the continuity of the worst-case probability bound $\inf_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\{\xi \in C\}$ over the polyhedron C. We specify the conclusion with $\mathcal{D} = \mathcal{D}_1$ by the following proposition, and the parallel claims for $\mathcal{D}_2$ and $\mathcal{D}_2'$ follow immediately.

Proposition 1 *Denote $z(C) = \inf_{\mathbb{P}\in\mathcal{D}_1} \mathbb{P}\{\xi \in C\}$, then $z(C) = z\big(int(C)\big)$, i.e.,*

$$
\inf_{\mathbb{P}\in\mathcal{D}_1} \mathbb{P}\{\xi \in C\} = \inf_{\mathbb{P}\in\mathcal{D}_1} \mathbb{P}\{\xi \in int(C)\}, \tag{21}
$$

where $C = \{\xi \in \mathbb{R}^K : \ a_i^\top \xi \leq \bar{b}_i\}$.

Proof: It is clear that $z\big(int(C)\big) \leq z(C)$ because $int(C) \subseteq C$. If we enlarge the polyhedron C by twisting the right hand side $\bar{b}$, i.e., define

$$C_\delta = \{\xi \in \mathbb{R}^K : a_i^\top \xi < \bar{b}_i + \delta, \ \ \forall i = 1, \ldots, m\} \text{ for } \delta > 0,$$

then $C \subseteq C_\delta$, and hence $z\big(int(C)\big) \leq z(C) \leq z(C_\delta)$. Next we prove the claim by showing that $z(C_\delta) \downarrow z(int(C))$ as $\delta \downarrow 0$.

Picking $\epsilon > 0$, we have to find a $\delta > 0$ such that $z(C_\delta) \leq z\big(int(C)\big) + \epsilon$. According to Theorem 2, we can compute $z\big(int(C)\big)$ by solving an SDP problem (10). Similarly, we can compute $z(C_\delta)$ by solving (10) with $\bar{b}_i$ replaced by $\bar{b}_i + \delta$ in constraint (10c). To avoid notational confusion, in the following we highlight the problem associated with $z(C_\delta)$ and call it (10-C_δ).

Now we consider a feasible and near-optimal solution (H^*, p^*, y^*, q^*) of problem (10-C_δ), such that

$$\Sigma \cdot H^* + \mu^\top p^* + q^* \geq z(C_\delta) - \epsilon/2. \tag{22}$$

First, we prove that such a near-optimal solution exists. Let set Z collect all possible objective values to problem (10-C_δ), i.e.,

$$Z = \Big\{\Sigma \cdot H + \mu^\top p + q : \ \ \exists y, \text{ such that } (H, p, y, q) \text{ satisfies constraints (10b)-(10d)}\Big\}.$$

Since i) $z(C_\delta) = \sup\{Z\}$ based on (10a), ii) $z(C_\delta)$ is bounded from above by 1 based on (9), and iii) Z is nonempty due to the fact that (10b)-(10d) has at least one feasible solution $H = 0$ and $y = p = q = 0$, there exists a feasible solution (H^*, p^*, y^*, q^*), such that inequality (22) is satisfied.

Second, starting from (H^*, p^*, y^*, q^*), we now find a feasible solution to problem (10) by perturbing q^* to

$$\bar{q} = q^* - \delta \sum_{j=1}^m y_j^*,$$

where the value of δ can be specified later. To prove its feasibility, we have to check each constraint of problem (10). First, it is clear that constraint (10d) is satisfied. Second, constraint (10b) is also satisfied since we have

$$\begin{bmatrix} -H^* & -\frac{1}{2}p^* \\ -\frac{1}{2}p^{*\top} & 1 - \bar{q} \end{bmatrix} = \begin{bmatrix} -H^* & -\frac{1}{2}p^* \\ -\frac{1}{2}p^{*\top} & 1 - q^* \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \delta \sum_{j=1}^m y_j^* \end{bmatrix} \succeq 0,$$

where $\delta > 0$ and each $y_j^* \geq 0$. Third, constraint (10c) is satisfied in the same manner, since for each $i = 1, \dots, m$ we have

$$\begin{bmatrix} -H^* & -\frac{1}{2}(p^* + y_i^* a_i) \\ -\frac{1}{2}(p^* + y_i^* a_i)^\top & y_i^* \bar{b}_i - \bar{q} \end{bmatrix} = \begin{bmatrix} -H^* & -\frac{1}{2}(p^* + y_i^* a_i) \\ -\frac{1}{2}(p^* + y_i^* a_i)^\top & y_i^*(\bar{b}_i + \delta) - q^* \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \delta \sum_{j \neq i} y_j^* \end{bmatrix} \succeq 0.$$

Now we connect $z(C_\delta)$ with $z\big(int(C)\big)$. By (22), we have

$$\begin{aligned} z(C_\delta) &\leq \Sigma \cdot H^* + \mu^\top p^* + q^* + \epsilon/2 \; = \; \Sigma \cdot H^* + \mu^\top p^* + \bar{q} + \delta \sum_{j=1}^m y_j^* + \epsilon/2 \\ &\leq z\big(int(C)\big) + \delta \sum_{j=1}^m y_j^* + \epsilon/2, \end{aligned}$$

where the second inequality uses the fact that $(H^*, p^*, y^*, \bar{q})$ is feasible to problem (10). Hence, to achieve that $z(C_\delta) \leq z\big(int(C)\big) + \epsilon$, we only need to choose an appropriate δ, such that $\delta \sum_{j=1}^m y_j^* \leq \epsilon/2$. It follows that if $y_j^* = 0$ for each $j = 1, \dots, m$, we can choose δ arbitrarily, and if $y_j^* > 0$ for some $j = 1, \dots, m$, we can choose $\delta = \epsilon \big/ (2 \sum_{j=1}^m y_j^*)$, which completes the proof. ■

Remark 2 *Proposition 1 is trivial if the probability distribution $\mathbb{P}$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^K$, since $Leb(a_i^\top \xi = \bar{b}_i) = 0, \forall i = 1, \dots, m$. However, the above proof works for general distributions.*

2.4 Solution Approach for DCCP

In this subsection, we discuss the solution approach for the following DCCP with a moment-based confidence set,

$$\begin{aligned} \text{[DCCP]} \quad \min_x \quad & \psi(x) \\ \text{s.t.} \quad & \inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha, \\ & x \in X. \end{aligned} \tag{23}$$

We restrict the discussion to the case $\mathcal{D}_1$, and the parallel solution approaches for the cases $\mathcal{D}_2$ and $\mathcal{D}_2'$ can be similarly obtained. Recall that, by Theorem 2, we can replace the left hand side of constraint (23) by the optimal objective value of an SDP problem (10). By using this result, we can reformulate [DCCP] and claim the following proposition:

Proposition 2 [DCCP] *can be reformulated as*

$$
\begin{aligned}
[\text{DCCP}^{\text{R}}] \quad \min_{x,a,\bar{b},H,p,y,q} \quad & \psi(x) && \text{(24a)}\\
\text{s.t.} \quad & \Sigma \cdot H + \mu^\top p + q \geq 1 - \alpha, && \text{(24b)}\\
& \begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1-q \end{bmatrix} \succeq 0, && \text{(24c)}\\
& \begin{bmatrix} -H & -\frac{1}{2}(p + y_i a_i) \\ -\frac{1}{2}(p + y_i a_i)^\top & y_i \bar{b}_i - q \end{bmatrix} \succeq 0, \quad \forall i = 1, \ldots, m, && \text{(24d)}\\
& a_{ik} = A_{ki}^\top x - b_{ki}, \ \forall i = 1, \ldots, m, \ \forall k = 1, \ldots, K, && \text{(24e)}\\
& \bar{b}_i = b_{0i} - A_{0i}^\top x, \ \forall i = 1, \ldots, m, && \text{(24f)}\\
& x \in X, \quad y \geq 0, \quad H \in \mathbb{S}^{K \times K}, && \text{(24g)}
\end{aligned}
$$

where a_{ik} denotes the k^{th} component of vector a_i and b_{ki} represents the i^{th} component of vector b_k.

Proof: Since the SDP problem (10) is a maximization problem, the reformulated constraint (23)

$$
\max_{\substack{H,p,y,q:\\ \text{(10b)-(10d) are satisfied}}} \quad \Sigma \cdot H + \mu^\top p + q \geq 1 - \alpha
$$

is equivalent to constraints (24b)-(24d), (24g). The conclusion then follows immediately by substituting the equivalent reformulation into [DCCP] and recalling the definition of vectors $a_1, \ldots, a_m$ and $\bar{b}$. ■

[DCCP$^{\text{R}}$] is not a convex program because it contains bilinear terms $y_i a_i$ and $y_i \bar{b}_i$ in constraints (24d). Hence, we cannot solve [DCCP$^{\text{R}}$] directly by using the powerful convex optimization solvers. However, we observe that [DCCP$^{\text{R}}$] is a convex program if variable y is fixed. In this paper, inspired by Chen et al. [9] and Zymler et al. [39], we propose an algorithm based on iteratively solving two convex optimization problems (hereafter denoted as iterative convex optimization) to solve [DCCP$^{\text{R}}$]. First, we denote

$$
z_{\mathcal{D}_1}(x, y) = \max_{\substack{a,\bar{b},H,p,q:\\ \text{(24c)-(24g) are satisfied}}} \quad \Sigma \cdot H + \mu^\top p + q,
$$

and it can be shown, by the same argument in the proof of Proposition 2, that [DCCP$^{\text{R}}$] is equivalent to

$$
\begin{aligned}
\min_{x \in X, \ y \geq 0} \quad & \psi(x) && \text{(25a)}\\
\text{s.t.} \quad & z_{\mathcal{D}_1}(x, y) \geq 1 - \alpha. && \text{(25b)}
\end{aligned}
$$

Again, we observe that problem (25) is convex if variable y is fixed. Given a feasible solution (x_0, y_0) to problem (25), we can enlarge the feasible region of problem (25) by solving the following SDP problem with variable x fixed at x_0:

$$\max_{y \geq 0} \; z_{\mathcal{D}_1}(x, y). \tag{26}$$

Indeed, suppose that y_1 is an optimal solution to problem (26), we have $z_{\mathcal{D}_1}(x_0, y_1) \geq z_{\mathcal{D}_1}(x_0, y_0) \geq 1 - \alpha$, and hence (x_0, y_1) is feasible to problem (25). Also, if we resolve problem (25) with variable y fixed at y_1 and obtain an optimal solution x_1, then we have $\psi(x_1) \leq \psi(x_0)$. This observation indicates that we can obtain a series of feasible solutions to [DCCP$^{\text{R}}$] with nonincreasing objective values by iteratively solving convex problems (25) and (26), where variables y and x are fixed respectively. We display the iterative convex optimization algorithm as follows:

Step 0 Input: a feasible solution (x_0, y_0) to [DCCP$^{\text{R}}$] and a convergence tolerance ε. Initialize the counter $i \leftarrow 0$.

Step 1 Solve problem (26) with x fixed at x_i, and record the optimal solution y^*. Set $y_{i+1} \leftarrow y^*$.

Step 2 Solve problem (25) with y fixed at y_{i+1}, and record the optimal solution x^*. Set $x_{i+1} \leftarrow x^*$.

Step 3 If $\big(\psi(x_i) - \psi(x_{i+1})\big)/|\psi(x_{i+1})| < \varepsilon$, then terminate and output the solution (x_{i+1}, y_{i+1}). Otherwise, set $i \leftarrow i + 1$ and go back to Step 1.

It is clear that the sequence $\{\psi(x_i) : \; i \geq 0\}$ converges to a finite limit since the set X is bounded by assumption. That is, we can obtain an nonincreasing sequence of objective values by using the iterative convex optimization algorithm, and the sequence converges to a finite limit.

3 DCC with Density-based Confidence Set

In this section, we consider constraint (8) when density information is taken into account, i.e., with $\mathcal{D} = \mathcal{D}_3$. One advantage of using $\mathcal{D}_3$ as a confidence set over $\mathcal{D}_1$ and $\mathcal{D}_2$ is that we can depict the profile of the ambiguous distribution $\mathbb{P}$ more accurately by its density function than by its first two moments alone. Indeed, as pointed out in the study of classical moment problems (see, e.g., Shohat and Tamarkin [32]), the first two moments are generally unable to determine a unique probability distribution. In this section, we will first discuss the construction of $\mathcal{D}_3$ by using a histogram. Then, we show that constraint (8) under $\mathcal{D}_3$ is equivalent to a traditional chance constraint (1b) with a deterministic probability distribution, which can be further solved by scenario approximation approaches. In addition, by deriving how the sample size of the data reflects the level of conservatism of DCCs, we depict quantitatively the value of data.

3.1 Construction of Confidence Set

In this subsection, we discuss the construction of confidence set $\mathcal{D}_3$. In practice, histograms are useful in depicting density function profiles. Suppose that we have a data set $\{\xi^i\}_{i=1}^N$. To draw a histogram, we first construct a nonempty partition $\{B_j: \ j = 1, \ldots, \bar{B}\}$ of the sample space, where $\Omega = \bigcup_{j=1}^{\bar{B}} B_j$ and each B_j is called a "bin." Second, we count the frequency $N_j = \sum_{i=1}^N I_{[B_j]}(\xi^i)$ for each bin B_j, where $I_{[B_j]}(\xi^i)$ equals one if $\xi^i \in B_j$, and zero otherwise. Finally, we can estimate the probability of landing in each bin, i.e., $\mathbb{P}\{B_j\}$, by its empirical relative frequency N_j/N. With the help of a histogram, we can estimate the KL divergence between the true probability distribution and its histogram estimate,

$$D_{\text{KL}}(f||f_0) = \int_{\mathbb{R}^K} f(\xi) \log \frac{f(\xi)}{f_0(\xi)} d\xi,$$

where f and f_0 denote the density function (resp. probability mass function for discrete distribution) of $\mathbb{P}$ and its estimate respectively, and the integral is with respect to Lebesgue measure on $\mathbb{R}^K$ (resp. with respect to the counting measure for discrete distribution). For the well-defineness of $D_{\text{KL}}(f||f_0)$, we define $x \log(x/0) = +\infty$ for $x > 0$ and $0 \log(0/0) = 0$. In a histogram, f corresponds to the probability mass $\mathbb{P}\{B_j\}$ of each bin B_j, and f_0 corresponds to the histogram estimate, i.e., N_j/N. Pardo [23] (see, e.g., Remark 3.1 in [23]) showed that $2ND_{\text{KL}}(f||f_0)$ converges in measure to a chi-square distributed random variable with $\bar{B} - 1$ degrees of freedom as N goes to infinity. This observation inspires us to construct $\mathcal{D}_3$ as

$$\mathcal{D}_3 = \{\mathbb{P} \in \mathcal{M}_+ : \quad D_{\text{KL}}(f||f_0) \leq d\},$$

where the divergence tolerance d can be approximated by $\chi^2_{\bar{B}-1,1-\beta}/(2N)$ with large N, where $\chi^2_{\bar{B}-1,1-\beta}$ represents the $100(1-\beta)\%$ (e.g., $\beta = 0.05$) percentile of the $\chi^2_{\bar{B}-1}$ distribution. For continuous distributions, we cannot directly use the histogram estimate for f since it is not absolutely continuous with regard to Lebesgue measure. In this paper, we propose to replace the histogram estimate by its counterpart in estimating continuous density functions, called the Kernel Density Estimator (KDE). Originally proposed in Rosenblatt [30] and Parzen [24], KDE is defined as

$$f_N(\xi) = \frac{1}{Nh_N^K}\sum_{i=1}^{N} H\Big(\frac{\xi-\xi^i}{h_N}\Big),$$

where h_N is a positive constant, and $H(\cdot)$ is a smooth function satisfying $H(\cdot) \geq 0$, $\int H(\xi)d\xi = 1$, $\int \xi H(\xi)d\xi = 0$, and $\int \xi^2 H(\xi)d\xi > 0$. One example for $H(\cdot)$ is the standard normal density function. It was shown in Devroye and Györfi [12] that KDE estimation converges in ℓ_1 sense, i.e., with probability one,

$$\int_{\mathbb{R}^K} |f_N(\xi) - f(\xi)|d\xi \ \rightarrow\ 0 \ \text{ as } N \rightarrow \infty.$$

Besides the KL divergence we discuss in this paper, there are various other divergence measures in statistics (e.g., Pearson, Cressie and Read, etc), and the interested readers are referred to Pardo [23] for more examples. Indeed, every other divergence measure can replace $D_{\text{KL}}(f||f_0)$ in the definition of $\mathcal{D}_3$ and make a new confidence set, as long as the divergence measure can be conveniently estimated from data and the corresponding DCCs can be efficiently handled. In this paper, we select KL divergence because it is commonly applied in both machine learning (see, e.g., Thollard et al. [35]) and information theory (see, e.g, Moreno et al. [20]). In a recent paper, Ben-Tal et al. [1] successfully apply general divergence measures to construct uncertainty sets in the robust optimization framework. See Ben-Tal et al. [1] and Pardo [23] for general divergence measures.

3.2 Reformulation of DCC

In this subsection, we address the reformulation of (8) using confidence set $\mathcal{D}_3$ with the KL divergence measure considered. Before giving the main result, we review the definition of conjugate duality. Given a function $g : \mathbb{R} \rightarrow \mathbb{R}$, the conjugate $g^* : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ is defined as

$$g^*(t) = \sup_{x\in\mathbb{R}}\{tx - g(x)\}.$$

Theorem 3 *Given matrix $\bar{A}$, vector $\bar{b}$, density estimate f_0, and the KL divergence tolerance d, with $\mathbb{P}_0$ representing the probability distribution defined by f_0 and f representing the density function of the ambiguous probability distribution $\mathbb{P}$, the chance constraint $\inf_{D_{KL}(f||f_0)\leq d} \mathbb{P}\{\bar{A}\xi \leq \bar{b}\} \geq 1-\alpha$ can be reformulated as*

$$\mathbb{P}_0\{\bar{A}\xi \leq \bar{b}\} \geq 1-\alpha', \quad \textit{where} \ \ 1-\alpha' = \inf_{x\in(0,1)}\Big\{\frac{e^{-d}x^{1-\alpha}-1}{x-1}\Big\}, \tag{27}$$

and $\alpha' \uparrow \alpha$ *as* $d \downarrow 0$.

Proof: We first redefine $D_{\text{KL}}(f||f_0)$ by using an auxiliary function $g(x) = x\log(x) - x + 1$ if $x > 0$ and $g(x) = +\infty$ otherwise. Also, for well-definedness, we let $0g(x/0) = +\infty$ for $x > 0$ and $0g(0/0) = 0$. Note here that $g^*(t) = e^t - 1$. It follows that

$$\begin{aligned} D_{\text{KL}}(f||f_0) &= \int_{\mathbb{R}^K} f(\xi)\log\frac{f(\xi)}{f_0(\xi)}d\xi \\ &= \int_{\mathbb{R}^K}\Big(f(\xi)\log\frac{f(\xi)}{f_0(\xi)} + f_0(\xi) - f(\xi)\Big)d\xi \\ &= \int_{\mathbb{R}^K} f_0(\xi)g\Big(\frac{f(\xi)}{f_0(\xi)}\Big)d\xi, \end{aligned}$$

where the second equality holds since both $f(\xi)$ and $f_0(\xi)$ are density functions.

Denoting set $C = \{\xi \in \mathbb{R}^K : \bar{A}\xi \le \bar{b}\}$, we rewrite the left hand side of the chance constraint as

$$z_{\mathcal{D}_3} = \min_f \quad \int_{\mathbb{R}^K} I_C(\xi)f(\xi)d\xi \tag{28a}$$

$$\text{s.t.} \quad \int_{\mathbb{R}^K} f_0(\xi)g\Big(\frac{f(\xi)}{f_0(\xi)}\Big)d\xi \le d, \tag{28b}$$

$$\int_{\mathbb{R}^K} f(\xi)d\xi = 1, \tag{28c}$$

$$f(\xi) \ge 0, \quad \forall \xi \in \mathbb{R}^K, \tag{28d}$$

where constraint (28b) bounds the KL divergence $D_{\text{KL}}(f||f_0)$ from above by d, and constraints (28c) and (28d) guarantee that f is a density function. Since problem (28) is, once again, a semi-infinite problem, we resort to duality. The Lagrangian dual of problem (28) can be written as

$$\begin{aligned} L &= \sup_{z\ge 0, z_0\in\mathbb{R}} \inf_{f(\xi)\ge 0}\left\{\int_{\mathbb{R}^K}\Big(I_C(\xi)f(\xi) - z_0 f(\xi) + z f_0(\xi) g\Big(\frac{f(\xi)}{f_0(\xi)}\Big)\Big)d\xi + z_0 - zd\right\} \\ &= \sup_{z\ge 0, z_0\in\mathbb{R}}\left\{z_0 - zd + \inf_{f(\xi)\ge 0}\Big\{\int_{\mathbb{R}^K}\Big[(I_C(\xi) - z_0)f(\xi) + z f_0(\xi)g\Big(\frac{f(\xi)}{f_0(\xi)}\Big)\Big]d\xi\Big\}\right\}, \end{aligned} \tag{29}$$

where z and z_0 represent the dual variables of constraints (28b) and (28c), respectively. Strong duality between problems (28) and (29) yields $z_{D_3} = L$.

Second we solve the max-min problem (29) to optimality. Suppose that we let $z = 0$, then we have $L = \sup_{z_0\in\mathbb{R}}\Big\{z_0 + \inf_{f(\xi)\ge 0}\int_{\mathbb{R}^K}(I_C(\xi) - z_0)f(\xi)d\xi\Big\}$. It follows that

Case 1. If $C \neq \mathbb{R}^K$, then at optimality $z_0^* = 0$ with $L = 0$, which is a lower bound for L since $L = z_{\mathcal{D}_3} \in [0, 1]$.

Case 2. If $C = \mathbb{R}^K$, then at optimality $z_0^* = 1$ with $L = 1$, which is trivial and can be immediately identified before solving problem (29).

Hence, without loss of generality $z = 0$ cannot be an optimal solution to problem (29), and we focus on the case $z > 0$ from now on. When $z > 0$, we have

$$\begin{aligned}
L &= \sup_{z>0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z \sup_{f(\xi)\geq 0} \left\{ \int_{\mathbb{R}^K} \Big[\Big(\frac{z_0 - I_C(\xi)}{z} \Big) f(\xi) - f_0(\xi) g\Big(\frac{f(\xi)}{f_0(\xi)} \Big) \Big] d\xi \right\} \right\} \\
&= \sup_{z>0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z \sup_{f(\xi)\geq 0} \left\{ \int_{[f_0(\xi)>0]} \Big[\Big(\frac{z_0 - I_C(\xi)}{z} \Big) \Big(\frac{f(\xi)}{f_0(\xi)} \Big) - g\Big(\frac{f(\xi)}{f_0(\xi)} \Big) \Big] f_0(\xi) d\xi \right\} \right\} \quad (30) \\
&= \sup_{z>0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z \int_{[f_0(\xi)>0]} \sup_{f(\xi)/f_0(\xi)\geq 0} \left\{ \Big(\frac{z_0 - I_C(\xi)}{z} \Big) \Big(\frac{f(\xi)}{f_0(\xi)} \Big) - g\Big(\frac{f(\xi)}{f_0(\xi)} \Big) \right\} f_0(\xi) d\xi \right\} \\
& \qquad (31) \\
&= \sup_{z>0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z \int_{[f_0(\xi)>0]} g^* \Big(\frac{z_0 - I_C(\xi)}{z} \Big) f_0(\xi) d\xi \right\} \quad (32) \\
&= \sup_{z>0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z \int_{[f_0(\xi)>0]} \Big[e^{\frac{z_0 - I_C(\xi)}{z}} - 1 \Big] f_0(\xi) d\xi \right\} \\
&= \sup_{z>0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z\mathbb{P}_0(C) \big[e^{(z_0 - 1)/z} - 1 \big] - z\big(1 - \mathbb{P}_0(C)\big) \big[e^{z_0/z} - 1 \big] \right\}, \quad (33)
\end{aligned}$$

where equality (30) follows from the observation that whenever $f_0(\xi) = 0$, $f(\xi)$ has to be zero to achieve optimality, equality (31) follows from the Dominated Convergence Theorem, equality (32) follows from the definition of function g and its conjugate g^*, and equality (33) follows from conditional probability, conditioning on the value of $I_C(\xi)$.

The remaining task is to optimize the outer problem. In the remainder of the proof, we denote $L = \sup_{z>0, z_0 \in \mathbb{R}} \theta(z, z_0)$. For fixed $z > 0$, since an exponential function is convex, it is clear that $\theta(z, z_0)$ is a concave function over z_0. Hence, by fixing $\partial\theta(z,z_0)/\partial z_0 \big|_{z_0 = z_0^*} = 1 - \mathbb{P}_0(C) e^{(z_0^* - 1)/z} - \big(1 - \mathbb{P}_0(C)\big) e^{z_0^*/z} = 0$, we have $e^{z_0^*/z} = \big[\mathbb{P}_0(C) e^{-1/z} + \big(1 - \mathbb{P}_0(C)\big) \big]^{-1}$, and accordingly

$$\theta(z, z_0^*) = -z \Big[d + \log\big(\mathbb{P}_0(C) e^{-1/z} + 1 - \mathbb{P}_0(C) \big) \Big].$$

Now we optimize $\theta(z, z_0^*)$ over z. Since

$$\begin{aligned}
\frac{\mathrm{d}\theta(z, z_0^*)}{\mathrm{d}z} &= -\Big(d + \log\big(\mathbb{P}_0(C) e^{-1/z} + 1 - \mathbb{P}_0(C) \big) \Big) - \frac{1}{z} \Big(\frac{\mathbb{P}_0(C) e^{-1/z}}{\mathbb{P}_0(C) e^{-1/z} + 1 - \mathbb{P}_0(C)} \Big), \text{and} \\
\frac{\mathrm{d}\theta^2(z, z_0^*)}{\mathrm{d}z^2} &= -\frac{\mathbb{P}_0(C)\big(1 - \mathbb{P}_0(C)\big) e^{-1/z}}{z^3 (\mathbb{P}_0(C) e^{-1/z} + 1 - \mathbb{P}_0(C))^2} \leq 0,
\end{aligned}$$

we know that $\theta(z, z_0^*)$ is concave over z. To compute z^*, we observe that

$$\lim_{z\to 0^+} \frac{\mathrm{d}\theta(z,z_0^*)}{\mathrm{d}z} = \begin{cases} -d, & \text{if } \mathbb{P}_0(C)=1 \\ -d-\log\big(1-\mathbb{P}_0(C)\big), & \text{if } \mathbb{P}_0(C)\in[0,1) \end{cases}, \text{ and } \lim_{z\to+\infty} \frac{\mathrm{d}\theta(z,z_0^*)}{\mathrm{d}z} = -d < 0.$$

Hence, we distinguish the following cases to ensure the chance constraint $L = z_{\mathcal{D}_3} \geq 1-\alpha$ satisfied:

Case 1. If $\mathbb{P}_0(C) = 1$, then $L = \sup_{z>0} \theta(z, z_0^*) = \lim_{z\to 0^+} \theta(z, z_0^*) = 1$, and hence $L \geq 1-\alpha$ trivially follows.

Case 2. If $\mathbb{P}_0(C) \in [0,1)$, then we have to force $-d - \log\big(1-\mathbb{P}_0(C)\big) > 0$, since otherwise $\theta(z, z_0^*)$ is nonincreasing on $(0, +\infty)$ and hence $L = \sup_{z>0} \theta(z, z_0^*) = \lim_{z\to 0^+} \theta(z, z_0^*) = 0$, violating the fact that $L \geq 1-\alpha > 0$. Thus, under the condition we have $\lim_{z\to 0^+} \mathrm{d}\theta(z, z_0^*)/\mathrm{d}z > 0$ and $\lim_{z\to+\infty} \mathrm{d}\theta(z, z_0^*)/\mathrm{d}z < 0$, which implies that the optimal $z^* > 0$, i.e., there exists a $z > 0$, such that $L = \theta(z, z_0^*)$. It follows that

$$\begin{aligned} & \inf_{D_{\mathrm{KL}}(f\|f_0)\leq d} \mathbb{P}\{\bar{A}\xi \leq \bar{b}\} \geq 1-\alpha \\ \Leftrightarrow\quad & \exists z > 0: \quad -z\Big[d + \log\big(\mathbb{P}_0(C)e^{-1/z} + 1 - \mathbb{P}_0(C)\big)\Big] \geq 1-\alpha, \\ \Leftrightarrow\quad & \mathbb{P}_0(C) \geq \inf_{z>0}\left\{\frac{e^{-d}\big(e^{-1/z}\big)^{1-\alpha}-1}{e^{-1/z}-1}\right\}, \text{ i.e., } \mathbb{P}_0(C) \geq \inf_{x\in(0,1)}\left\{\frac{e^{-d}x^{1-\alpha}-1}{x-1}\right\}. \end{aligned} \tag{34}$$

Note here based on (34) and in view that $x^{1-\alpha} < (1-\alpha)x + \alpha$ for each $x \in (0,1)$, we have

$$\mathbb{P}_0(C) \;\geq\; \inf_{x\in(0,1)}\left\{\frac{e^{-d}\big((1-\alpha)x+\alpha\big)-1}{x-1}\right\} \;=\; 1 - e^{-d}\alpha \;>\; 1 - e^{-d},$$

which ensures $-d - \log\big(1-\mathbb{P}_0(C)\big) > 0$. Thus, the initial condition $-d - \log\big(1-\mathbb{P}_0(C)\big) > 0$ is satisfied by the reformulation (34).

Therefore, the above two cases show that

$$\inf_{D_{\mathrm{KL}}(f\|f_0)\leq d} \mathbb{P}\{\bar{A}\xi \leq \bar{b}\} \geq 1-\alpha \;\Leftrightarrow\; \mathbb{P}_0(C) = 1 \text{ , or } \mathbb{P}_0(C) \geq \inf_{x\in(0,1)}\left\{\frac{e^{-d}x^{1-\alpha}-1}{x-1}\right\}.$$

Since $\lim_{x\to 0^+} \dfrac{e^{-d}x^{1-\alpha}-1}{x-1} = 1$, the case $\mathbb{P}_0(C) = 1$ is redundant. Hence, the equivalent reformulation of the DCC is proved.

It remains to show that $\alpha' \uparrow \alpha$ as $d \downarrow 0$. It is clear that $e^{-d} \uparrow 1$ as $d \downarrow 0$, and so we have

$$\alpha' = 1 - \inf_{x\in(0,1)}\left\{\frac{e^{-d}x^{1-\alpha}-1}{x-1}\right\} \;\Big\uparrow\; 1 - \inf_{x\in(0,1)}\left\{\frac{x^{1-\alpha}-1}{x-1}\right\} = 1 - (1-\alpha) = \alpha,$$

which completes the proof. ■

Henceforth, we call the chance constraint (27) a "reformulated DCC." Theorem 3 shows that a reformulated DCC is equivalent to a traditional chance constraint with the ambiguous probability distribution $\mathbb{P}$ replaced by its estimate $\mathbb{P}_0$, and the risk level α decreased to α'. As compared to the traditional chance constraints, reformulated DCCs provide us the following theoretical merits:

1. In terms of modeling, unlike relying on an ambiguous probability distribution $\mathbb{P}$, we resort to the data in a reformulated DCC, which can be represented as the historical data in practice. We can waive the "perfect information" assumption. Meanwhile, we can make a more accurate $\mathbb{P}_0$ estimate with more data on hand, and accordingly derive a less conservative reformulated DCC.

2. In terms of algorithm development, the estimate $\mathbb{P}_0$ is more accessible and controllable than the ambiguous $\mathbb{P}$. First, the samples taken from $\mathbb{P}_0$, which is definite and deterministic, is more trustable than from a guess of the ambiguous $\mathbb{P}$. Second, we can facilitate the sampling procedure by choosing density functions that are easier to sample from (e.g., normal distributions) since we have the freedom to choose function $H(\cdot)$ in KDE. This observation motivates us to solve optimization problems with reformulated DCCs by using the scenario approximation approach.

3. In terms of robustness, it is easy to observe that reducing α can increase the robustness of a traditional chance constraint. Furthermore, our result quantifies how much α needs to be reduced, and hence accurately depicts the relationship between the risk level α and the robustness.

Before closing this subsection, we show that the perturbed risk level α' is easy to obtain through numerical computation.

Proposition 3 *The perturbed risk level*

$$\alpha' = 1 - \inf_{x\in(0,1)} \left\{ \frac{e^{-d}x^{1-\alpha} - 1}{x - 1} \right\} \tag{35}$$

can be computed by using bisection line search after $\lceil \log_2(\frac{1}{\epsilon}) \rceil$ *steps to achieve* ϵ *accuracy.*

Proof: We compute α' by searching the optimal solution to the minimization problem embedded in equation (35). Granted this, we can compute α' directly. First, by denoting $1-\alpha' = \inf_{x\in(0,1)} h(x)$, we have

$$h'(x) = \frac{1 - e^{-d}\alpha x^{1-\alpha} - e^{-d}(1-\alpha)x^{-\alpha}}{(x-1)^2}, \quad \forall x \in (0,1).$$

It is clear that $(x-1)^2$ decreases as x increases. Meanwhile, since $x < 1$ and $x^{-\alpha-1} > x^{-\alpha}$, we have

$$(1 - e^{-d}\alpha x^{1-\alpha} - e^{-d}(1-\alpha)x^{-\alpha})'_x = e^{-d}\alpha(1-\alpha)(x^{-\alpha-1} - x^{-\alpha}) > 0.$$

Therefore, $h'(x)$ increase as x increases in $(0,1)$, and hence the function $h(x)$ is convex over x in $(0,1)$. Because $\lim_{x\to 0^+} h'(x) = -\infty$ and $\lim_{x\to 1^-} h'(x) = +\infty$, the infimum of $h(x)$ is attained in the interval $(0,1)$, and we can compute the optimal x^* by forcing

$$\frac{1 - e^{-d}\alpha(x^*)^{1-\alpha} - e^{-d}(1-\alpha)(x^*)^{-\alpha}}{(x^*-1)^2} = 0,$$

i.e., $(x^*)^\alpha = e^{-d}\alpha x^* + e^{-d}(1-\alpha)$. The intersection of functions x^α and $e^{-d}\alpha x + e^{-d}(1-\alpha)$ can be easily computed by a bisection line search. Finally, to achieve ϵ accuracy, i.e., $|\hat{x} - x^*| \leq \epsilon$, of the incumbent probing value $\hat{x}$, we only have to conduct S steps of bisection, such that $2^{-S} \leq \epsilon$. It follows that $S \geq \lceil \log_2(\frac{1}{\epsilon}) \rceil$. ■

3.3 The Value of Data

In this subsection, we discuss how the sample size can help decrease the conservatism of DCCs. Intuitively, as the sample size N increases we can depict the profile of the ambiguous probability distribution $\mathbb{P}$ more accurately with a smaller KL divergence tolerance d. Accordingly, from Theorem 3, a DCC becomes less conservative with a larger α' value in its reformulation and α' converges to α as N tends to be very large. To quantify how the increase of sample size can make a DCC less conservative, in this paper we introduce the notion *value of data* and define it as

$$\text{VoD}_\alpha = \frac{\mathrm{d}\alpha'}{\mathrm{d}N},$$

where risk level α is given. Intuitively, VoD_α represents the increase of α' value if we enlarge the data set. In practice it often incurs costs to collect data, and VoD_α gives decision makers a clear picture on how a group of new data can help reduce the conservatism. To evaluate VoD_α, we have to connect α' value with the sample size N. To that end, we first depict the relationship between the α' value and the KL divergence tolerance d by the following proposition.

Proposition 4 *Given risk levels α and α' such that $\alpha > \alpha'$, the DCC $\inf_{D_{KL}(f\|f_0)\leq d} \mathbb{P}\{\bar{A}\xi \leq \bar{b}\} \geq 1-\alpha$ is equivalent to the traditional chance constraint $\mathbb{P}_0\{A(\xi)x \leq b(\xi)\} \geq 1-\alpha'$ with the KL divergence* *tolerance*

$$d = \alpha \log\Big(\frac{\alpha}{\alpha'}\Big) + (1-\alpha)\log\Big(\frac{1-\alpha}{1-\alpha'}\Big). \tag{36}$$

Proof: From Theorem 3, we know that

$$1-\alpha' = \inf_{x\in(0,1)} \Big\{\frac{e^{-d}x^{1-\alpha}-1}{x-1}\Big\}. \tag{37}$$

Also, from Proposition 3, we know that the optimal value of the embedded optimization problem in equality (37) can be attained by some $\bar{x} \in (0,1)$, which is the stationary point of the objective function. It follows that

$$\begin{cases} \frac{e^{-d}\bar{x}^{1-\alpha}-1}{\bar{x}-1} = 1-\alpha' \\ \bar{x}^{\alpha} = e^{-d}\alpha\bar{x} + e^{-d}(1-\alpha). \end{cases} \tag{38}$$

Solving this nonlinear equation system by substituting the second equation into the first, we have

$$\begin{aligned} & e^{-d}\bar{x} - (1-\alpha')\bar{x}\bar{x}^{\alpha} = \alpha'\bar{x}^{\alpha} \\ \Rightarrow\ & e^{-d}\bar{x} - (1-\alpha')\bar{x}\Big(e^{-d}\alpha\bar{x} + e^{-d}(1-\alpha)\Big) = \alpha'\Big(e^{-d}\alpha\bar{x} + e^{-d}(1-\alpha)\Big) \\ \Rightarrow\ & (\bar{x}-1)\Big(\alpha(1-\alpha')\bar{x} - \alpha'(1-\alpha)\Big) = 0. \end{aligned}$$

Ruling out the solution $\bar{x} = 1$, we have $\bar{x} = \frac{\alpha'(1-\alpha)}{\alpha(1-\alpha')} \in (0,1)$. Finally, we substitute the solution of $\bar{x}$ back into the second equation in (38) and obtain

$$e^{-d} = \bar{x}^{\alpha}/(\alpha\bar{x}+1-\alpha) = \big(\alpha'/\alpha\big)^{\alpha}\big((1-\alpha')/(1-\alpha)\big)^{1-\alpha}. \tag{39}$$

Noting that $\bar{x}^{\alpha} < \alpha\bar{x} + 1 - \alpha$ since $\bar{x} \in (0,1)$ and $\alpha \in (0,1)$, we remark that $e^{-d} < 1$, and hence $d > 0$. The final conclusion immediately follows by taking the natural logarithm on both sides of equation (39). ■

We are now ready to evaluate VoD_{α}. As discussed previously, the confidence set $\mathcal{D}_3$ can be constructed from the data samples and the corresponding histogram. Given N data samples, we can partition the sample space Ω into $\bar{B}$ bins in the histogram and construct $\mathcal{D}_3$ by setting $d = \chi^2_{\bar{B}-1,1-\beta}/(2N)$, where $\chi^2_{\bar{B}-1,1-\beta}$ represents the $100(1-\beta)\%$ percentile of the $\chi^2_{\bar{B}-1}$ distribution. Hence, the prior proposition gives us the following corollary.

Corollary 3 *Given N data samples, confidence level β, and risk levels α and α' such that $\alpha > \alpha'$, the DCC $\inf_{D_{KL}(f||f_0)\leq d} \mathbb{P}\{\bar{A}\xi \leq \bar{b}\} \geq 1-\alpha$ is equivalent to the traditional chance constraint*

$\mathbb{P}_0\{A(\xi)x \le b(\xi)\} \ge 1-\alpha'$ *with the sample size* N *satisfying*

$$N = \chi^2_{\bar{B}-1,1-\beta}/(2d), \tag{40}$$

where d *is defined in* (36). *Besides, the value of data satisfies*

$$VoD_\alpha = \frac{\alpha'(1-\alpha')}{(\alpha-\alpha')}\Big(\frac{\chi^2_{\bar{B}-1,1-\beta}}{2N^2}\Big).$$

Proof: First, equality (40) follows immediately from the construction of $\mathcal{D}_3$. Second, from equation (36) we have

$$\frac{\mathrm{d}d}{\mathrm{d}\alpha'} = -\frac{\alpha}{\alpha'} + \frac{1-\alpha}{1-\alpha'} = \frac{\alpha'-\alpha}{\alpha'(1-\alpha')}.$$

Also, from equation (40) we have $\mathrm{d}N/\mathrm{d}d = -\chi^2_{\bar{B}-1,1-\beta}/(2d^2)$. It follows that

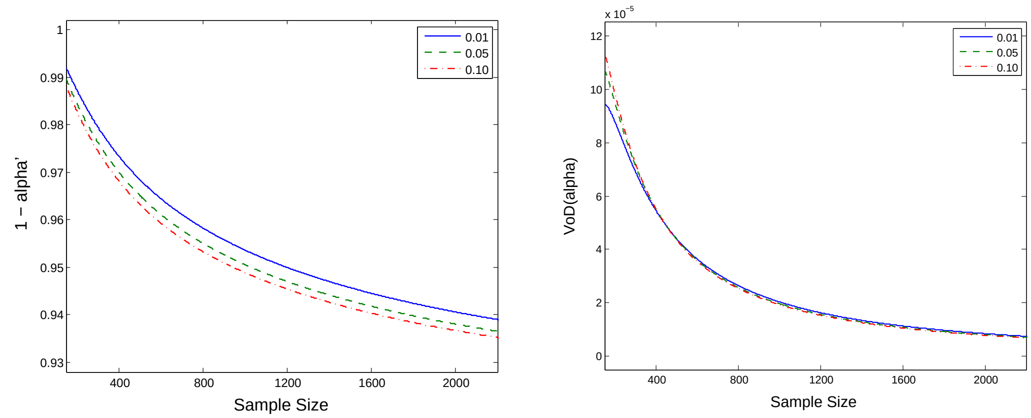
$$\frac{\mathrm{d}N}{\mathrm{d}\alpha'} = \frac{\mathrm{d}N}{\mathrm{d}d}\frac{\mathrm{d}d}{\mathrm{d}\alpha'} = \Big(\frac{\chi^2_{\bar{B}-1,1-\beta}}{2d^2}\Big)\frac{(\alpha-\alpha')}{\alpha'(1-\alpha')}.$$

It is easy to observe that $\mathrm{d}N/\mathrm{d}\alpha'$ is a monotone function of α' and $\mathrm{d}N/\mathrm{d}\alpha' \neq 0$. Hence, we have

$$\begin{aligned}\mathrm{VoD}_\alpha &= \frac{\mathrm{d}\alpha'}{\mathrm{d}N} = 1\Big/\Big(\frac{\mathrm{d}N}{\mathrm{d}\alpha'}\Big)\\ &= \Big(\frac{2d^2}{\chi^2_{\bar{B}-1,1-\beta}}\Big)\frac{\alpha'(1-\alpha')}{(\alpha-\alpha')},\end{aligned}$$

and the conclusion immediately follows by noting $d = \chi^2_{\bar{B}-1,1-\beta}/(2N)$. ■

For illustration, we depict the relationships between $1-\alpha'$ and N and between VoD_α and N in Figure 1 under confidence levels with $\alpha = 0.90$ and bin size $\bar{B} = 30$. We observe that with sample size N increasing, both $1-\alpha'$ and VoD_α decay quickly. However, we need a large sample size (e.g., $N > 2000$) to guarantee α' converging to α and VoD_α converging to zero. That is, we have to draw a large set of historical data to guarantee an almost exact description of the unknown probability distribution $\mathbb{P}$, and accordingly the risk level of the reformulated DCC, α', can be chosen to be near its deterministic counterpart α. This observation makes sense because we need a large sample size to construct the histogram in the first place, and it reflects the difficulty of estimating continuous density functions, especially when the dimension of random vectors becomes very large. In practice, we suggest using density-based confidence sets in an industry that has rich access to data and relies heavily on data to make decisions. As compared to density-based confidence sets, moment-based confidence sets are more conservative and only need small data sets for construction, and hence, are suitable to be used in an industry that has limited access to data.



(a) $1-\alpha'$ v.s. N (b) VoD_α v.s. N

Figure 1: Evolution of values $1-\alpha'$ and VoD_α against sample size under risk level $\alpha = 0.90$ and confidence levels $\beta = 0.01, 0.05, 0.10$

4 Numerical Experiments

In this section, we conduct a simple numerical experiment to illustrate the application of DCCPs. We model DCCs with both moment-based and density-based confidence sets in a portfolio optimization problem. In this experiment, a generic DCCP for the portfolio optimization problem can be formulated as

$$\begin{aligned} \text{[DCPO]} \quad & \max_{x \geq 0} \ \sum_{i=1}^{n} \mathbb{E}[\xi_i] x_i \\ \text{s.t.} \quad & \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P} \left\{ \begin{array}{l} \sum_{i=1}^{n} \xi_i x_i \geq T_0, \\ \sum_{i \in N_j} \xi_i x_i \geq T_j, \forall j = 1, \ldots, J \end{array} \right\} \geq 1 - \alpha, \quad \sum_{i=1}^{n} x_i = 1, \end{aligned}$$

where n represents the total number of investments, ξ_i represents the Rate of Return (RoR) of investment i, x_i represents the share of investment i, $N_1, \ldots, N_J$ represent different portfolio segments with $\bigcup_{j=1}^{J} N_j = \{1, \ldots, n\}$ (e.g., N_1 consists of stocks, N_2 consists of bonds, and so on), and $T_0, \ldots, T_J$ represent the investment targets of different portfolio segments. To specify which confidence set we use, we denote [DCPO-M] when using the moment-based confidence set (i.e., $\mathcal{D} = \mathcal{D}_1$), and denote [DCPO-D] when using the density-based confidence set (i.e., $\mathcal{D} = \mathcal{D}_3$). Inspired by Delage and Ye [11], we evaluate [DCPO] in this experiment by using a historical data set of 30 assets from years 2008 to 2011, obtained from the Yahoo! Finance website[1]. In each

[1] The 30 assets are AAR Corp., AT&T, Avery Denison Corp., Boeing Corp., Bristol-Myers-Squibb, Cisco Systems, Dell Computer Corp., Dow Chemical, Duke Energy Company, Du Pont, Eli Lilly and Co., Exelon Corp., FMC Corp., General Electric, Hewlett Packard, Hitachi, Honeywell, IBM Corp., Ingersoll Rand, Intel Corp., Lockheed

experiment, we randomly choose four assets, randomly assign them into $J = 2$ portfolio segments, and build a dynamic portfolio with these assets. The assets in the portfolio are updated every thirty days, through years 2008 to 2011, by adopting optimal investment decisions obtained from [DCPO]. During any day of the experiment, we collect the most recent 2000 days of RoR data to construct both confidence sets $\mathcal{D}_1$ and $\mathcal{D}_3$. In this experiment, we assume that the estimate of each mean RoR, i.e., $\mathbb{E}[\xi_i]$ for each $i = 1, \ldots, n$, is accurate based on the most recent 2000 days data. We employ the iterative convex optimization algorithm to solve [DCPO-M], and the scenario approximation approach to solve [DCPO-D].

In this experiment, we evaluate the performance of the investment decisions obtained from [DCPO] during each trading day against the real data in the following thirty days. That is, after making the investment decision during each trading day, we will hold the assets for thirty days and see how it performs in the real market. To set a benchmark, we compare [DCPO-M] and [DCPO-D] to a myopic model, which maximizes the average return over the last 2000 days. In this experiment, we run 100 replications in total by choosing four assets in each replication, and summarize the results obtained from all replications in Table 1 and Figure 2.

	Avg.	St. dev	10th Perc.	90th Perc.
DCPO-D	1.118	0.194	0.969	1.415
DCPO-M	0.995	0.410	0.519	1.272
Myopic	0.991	0.507	0.499	1.394

Table 1: Comparison of average end wealth and risk in 100 replications through years 2008-2011

From Table 1 and Figure 2, we observe that both [DCPO-D] and [DCPO-M] outperform the myopic approach in both end wealth and risk control for the years 2008-2011. In particular, [DCPO-D] largely outperforms the other two approaches, with at least 12% more in end wealth and at least 52% less in standard deviation. It indicates that the density-based DCCPs can make robust and profitable portfolio selection. [DCPO-M] slightly outperforms the myopic approach and has a nontrivial standard deviation. There might be two contributing factors to this phenomenon. First, the iterative convex optimization algorithm does not guarantee to obtain a global optimal solution to [DCPO-M], and thus the dynamic portfolio is not necessarily an optimal one. Second, the moment-based confidence set $\mathcal{D}_1$ might be insufficient to describe the uncertainty in the real

Martin, Merck and Co., Microsoft, Motorola, Northern Telecom, Oracle, Pinnacle West, Texas Instruments, United Technologies and a 0%-interest-rate deposit.

market. As pointed out in Delage and Ye [11], the uncertainty of the moments should also be taken into account by considering, e.g., $\mathcal{D}_2$ and $\mathcal{D}'_2$ in this paper.

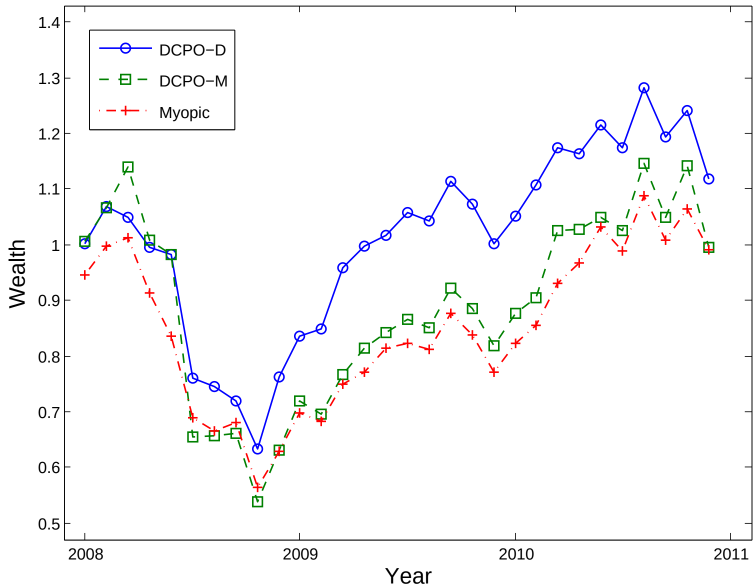


Figure 2: Comparison of wealth evolution in 100 replications through years 2008-2011

5 Conclusion

In this paper, we developed exact approaches for DCCPs. Starting from the historical data, we described how to construct moment-based and density-based confidence sets for the ambiguous probability distributions, how to equivalently reformulate DCCs, and how to effectively solve DC-CPs. In general, in this study, we proposed a framework to provide robust decisions based on the available data set information. Besides guaranteeing the robustness, our framework ensures that the proposed approach is less conservative when more data information is on hand. Possible future research directions include the study of DCCs under different confidence sets and their solution approaches. It is also interesting to apply DCCPs in a dynamic (multi-stage) decision process.

Acknowledgments

This research was partly supported by the U.S. National Science Foundation under CAREER Award CMMI0942156.

References

[1] A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *CentER Working Paper Series No. 2011-061*, 2011.

[2] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3):411–424, 2000.

[3] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.

[4] G.C. Calafiore and M.C. Campi. Uncertain convex programs: Randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[5] G.C. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[6] G.C. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[7] A. Charnes and W.W. Cooper. Deterministic equivalents for optimizing and satisficing under chance constraints. *Operations Research*, 11(1):18–39, 1963.

[8] A. Charnes, W.W. Cooper, and G.H. Symonds. Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Management Science*, 4(3):235–263, 1958.

[9] W. Chen, M. Sim, J. Sun, and C.P. Teo. From CVaR to uncertainty set: Implications in joint chance constrained optimization. *Operations Research*, 58(2):470–485, 2010.

[10] X. Chen and Y. Zhang. Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research*, 57(6):1469–1482, 2009.

[11] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[12] L. Devroye and L. Györfi. *Nonparametric Density Estimation: The ℓ_1 View.* John Wiley & Sons Inc, 1985.

[13] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1):37–61, 2006.

[14] K. Isii. The extrema of probability determined by generalized moments (I) bounded random variables. *Annals of the Institute of Statistical Mathematics*, 12(2):119–134, 1960.

[15] S. Küçükyavuz. On mixing sets arising in chance-constrained programming. *Mathematical Programming*, 132(1–2):31–56, 2010.

[16] J. Luedtke. A branch-and-cut decomposition algorithm for solving general chance-constrained

mathematical programs. *Preprint available at www.optimization-online.org*, 2011.

[17] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.

[18] J. Luedtke, S. Ahmed, and G.L. Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, 122(2):247–272, 2010.

[19] B.L. Miller and H.M. Wagner. Chance constrained programming with joint constraints. *Operations Research*, 13(6):930–945, 1965.

[20] P.J. Moreno, P. Ho, and N. Vasconcelos. A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications. *Advances in Neural Information Processing Systems*, 16:1385–1393, 2003.

[21] A. Nemirovski and A. Shapiro. Scenario approximations of chance constraints. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, pages 3–47. Springer, 2006.

[22] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2007.

[23] L. Pardo. *Statistical Inference Based on Divergence Measures*, volume 185. CRC Press, 2006.

[24] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[25] J. Pintér. Deterministic approximations of probability inequalities. *Mathematical Methods of Operations Research*, 33(4):219–239, 1989.

[26] I. Pólik and T. Terlaky. A survey of the S-Lemma. *SIAM Review*, 49(3):371–418, 2007.

[27] A. Prékopa. On probabilistic constrained programming. In *Proceedings of the Princeton Symposium on Mathematical Programming*, pages 113–138. Citeseer, 1970.

[28] A. Prékopa. *Stochastic Programming*, volume 324. Springer, 1995.

[29] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[30] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.

[31] A. Shapiro. On duality theory of conic linear problems. In *Semi-Infinite Programming*, pages 135–165. Kluwer Academic Publishers, 2000.

[32] J.A. Shohat and J.D. Tamarkin. *The Problem of Moments*, volume 1. American Mathematical Society, 1943.

[33] J. Smith. Generalized Chebychev inequalities: Theory and applications in decision analysis. *Operations Research*, 43(5):807–825, 1995.

[34] A.L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157, 1973.

[35] F. Thollard, P. Dupont, and C. Higuera. Probabilistic DFA inference using Kullback-Leibler divergence and minimality. In *Proceedings of the Seventeenth International Conference on Machine Learning*, pages 975–982. Morgan Kaufmann Publishers Inc., 2000.

[36] L. Vandenberghe, S. Boyd, and K. Comanor. Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52–64, 2007.

[37] H. Xu, C. Caramanis, and S. Mannor. Optimization under probabilistic envelope constraints. *Operations Research*, 2012. To appear.

[38] V. Yakubovich. S-procedure in nonlinear control theory. *Vestnik Leningrad University*, 4:73–93, 1977.

[39] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 2012. To appear.

Appendix A An Observation from Theorem 2

We first draw an observation from the proof of Theorem 2 which will be used throughout the proofs of Corollaries 1 and 2. The observation describes the equivalence between constraint (12b) and constraints (10c) and (10d), whose proof is also illustrated by Remark 1.

Observation 1 *Suppose we are given a matrix $\bar{A} \in \mathbb{R}^{m\times n}$ consisting of m row vectors $\{a_i^\top\}_{i=1}^m$ and a vector $\bar{b} = [\bar{b}_1, \ldots, \bar{b}_m]^\top \in \mathbb{R}^m$. Then for a matrix $H \in \mathbb{S}^{K\times K}$, a vector $p \in \mathbb{R}^K$, and a scalar q, the following are equivalent:*

(i) $\xi^\top H\xi + p^\top \xi + q \leq I_{[\bar{A}\xi<\bar{b}]}(\xi), \quad \forall \xi \in \mathbb{R}^K.$

(ii) There exist $\{y_i\}_{i=1}^m \geq 0$*, such that*

$$\begin{cases} \begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1-q \end{bmatrix} \succeq 0, \\ \begin{bmatrix} -H & -\frac{1}{2}(p+y_i a_i) \\ -\frac{1}{2}(p+y_i a_i)^\top & y_i\bar{b}_i - q \end{bmatrix} \succeq 0, \quad \forall i = 1, \ldots, m. \end{cases}$$

Appendix B Proof of Corollary 1

Proof: First, we specify $z_{\mathcal{D}_2}$ by the following optimization problem:

$$\begin{aligned}
[\text{Primal}_2] \quad z_{\mathcal{D}_2} \quad = \quad \min_{\mathbb{P}} \quad & \int_{\mathbb{R}^K} I_{[\bar{A}\xi<\bar{b}]}(\xi) d\mathbb{P} && (41a)\\
\text{s.t.} \quad & \int_{\mathbb{R}^K} \begin{bmatrix} \Lambda & \xi-\mu \\ (\xi-\mu)^\top & \gamma_1 \end{bmatrix} d\mathbb{P} \succeq 0, && (41b)\\
& \int_{\mathbb{R}^K} (\xi-\mu)(\xi-\mu)^\top d\mathbb{P} \preceq \gamma_2 \Lambda, && (41c)\\
& \int_{\mathbb{R}^K} d\mathbb{P} = 1, && (41d)
\end{aligned}$$

where $I_{[\bar{A}\xi<\bar{b}]}(\cdot) : \mathbb{R}^K \to \{0,1\}$ represents the indicator function of $int(C)$, constraints (41b) and (41c) describe the confidence sets of $\mathbb{E}[\xi]$ and $\mathbb{E}[(\xi-\mu)(\xi-\mu)^\top]$ respectively, and constraint (41d) guarantees we are considering probability distributions on $\mathbb{R}^K$. We apply the duality theory for conic linear programming problems and dualize problem (41) as

$$\begin{aligned}
[\text{Dual}_2] \quad z_{\mathcal{D}_2} = \quad \max_{G,H,p,q,r} \quad & (\mu\mu^\top - \gamma_2\Lambda)\cdot G - \Lambda \cdot H + 2\mu^\top p - \gamma_1 q + r && (42a)\\
\text{s.t.} \quad & \xi^\top(-G)\xi + 2(p+G\mu)^\top \xi + r \le I_{[\bar{A}\xi<\bar{b}]}(\xi), \quad \forall \xi \in \mathbb{R}^K, && (42b)\\
& G \succeq 0, \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \succeq 0, && (42c)
\end{aligned}$$

where matrix $\begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \in \mathbb{S}^{(K+1)\times(K+1)}$, matrix $G \in \mathbb{S}^{K\times K}$, and scalar r represent the dual variables for constraints (41b), (41c), and (41d), respectively. Again, note here that strong duality holds for problems [Primal$_2$] and [Dual$_2$] based on established conic linear programming theory. The conclusion follows immediately by applying Observation 1 in Appendix A, in which matrix H is replaced by $-G$, vector p is replaced by $(p+G\mu)$, and scalar q is replaced by r. ■

Appendix C Proof of Corollary 2

Proof: First, we specify $z_{\mathcal{D}'_2}$ by the following optimization problem:

$$\begin{aligned}
[\text{Primal}'_2] \quad z_{\mathcal{D}'_2} \quad = \quad \min_{\mathbb{P}} \quad & \int_{\mathbb{R}^K} I_{[\bar{A}\xi<\bar{b}]}(\xi) d\mathbb{P} && (43a)\\
\text{s.t.} \quad & \mu^{\text{L}} \le \int_{\mathbb{R}^K} \xi d\mathbb{P} \le \mu^{\text{U}}, && (43b)\\
& \sigma_k^{\text{L}} \le \int_{\mathbb{R}^K} \xi_k^2 d\mathbb{P} \le \sigma_k^{\text{U}}, \quad \forall k = 1,\ldots,K, && (43c)\\
& \int_{\mathbb{R}^K} d\mathbb{P} = 1, && (43d)
\end{aligned}$$

where $I_{[\bar{A}\xi<\bar{b}]}(\cdot):\mathbb{R}^K\to\{0,1\}$ represents the indicator function of $int(C)$, constraints (43b) and (43c) describe the confidence intervals of $\mathbb{E}[\xi]$ and $\mathbb{E}[\xi_k^2]$ for each $k=1,\ldots,K$ respectively, and constraint (43d) guarantees we are considering probability distributions on $\mathbb{R}^K$. We apply the duality theory for conic linear programming problems and dualize problem (43) as

$$
\begin{aligned}
[\text{Dual}_2'] \quad z_{\mathcal{D}_2'} = \quad & \max_{p,h,q} \quad p^{\text{L}}\mu^{\text{L}} - p^{\text{U}}\mu^{\text{U}} + \sum_{k=1}^{K}(\sigma_k^{\text{L}}h_k^{\text{L}} - \sigma_k^{\text{U}}h_k^{\text{U}}) + q && \text{(44a)}\\
& \text{s.t.} \quad \sum_{k=1}^{K}(h_k^{\text{L}} - h_k^{\text{U}})\xi_k^2 + (p^{\text{L}} - p^{\text{U}})^\top\xi + q \le I_{[\bar{A}\xi<\bar{b}]}(\xi), \quad \forall \xi\in\mathbb{R}^K, && \text{(44b)}\\
& \qquad\;\; p^{\text{L}}, p^{\text{U}}, h^{\text{L}}, h^{\text{U}} \ge 0, && \text{(44c)}
\end{aligned}
$$

where vectors p^{L} and p^{U}, vectors h^{L} and h^{U}, and scalar q represent the dual variables for constraints (43b), (43c), and (43d), respectively. Again, note here that strong duality holds for problems [Primal$_2'$] and [Dual$_2'$] based on established conic linear programming theory. To apply Observation 1 in Appendix A, we first simplify the notation by defining

$$p = p^{\text{L}} - p^{\text{U}} \;\text{ and }\; H = \text{diag}(h_1,\ldots,h_K), \;\text{where } h_k = h_k^{\text{L}} - h_k^{\text{U}},\; \forall k=1,\ldots,K. \tag{45}$$

Now by Observation 1, the constraint (44b) is equivalent to the following constraints

$$
\begin{aligned}
& \begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1-q \end{bmatrix} \succeq 0, && \text{(46)}\\
& \begin{bmatrix} -H & -\frac{1}{2}(p+y_ia_i) \\ -\frac{1}{2}(p+y_ia_i)^\top & y_i\bar{b}_i - q \end{bmatrix} \succeq 0, \quad \forall i=1,\ldots,m, && \text{(47)}\\
& y \ge 0.
\end{aligned}
$$

We show that (46) is second-order cone representable, and the similar reformulation of (47) follows immediately. To this end, we have

$$
\begin{aligned}
& \begin{bmatrix} -h_1 & & & \\ & \ddots & & -\frac{1}{2}p \\ & & -h_K & \\ & -\frac{1}{2}p^\top & & 1-q \end{bmatrix} \succeq 0 \\
\Leftrightarrow \quad & -\sum_{k=1}^{K} h_k\xi_k^2 - \sum_{k=1}^{K} p_k\xi_k\xi_{K+1} + (1-q)\xi_{K+1}^2 \ge 0, \quad \forall \xi_1,\ldots,\xi_K,\xi_{K+1}\in\mathbb{R} \\
\Leftrightarrow \quad & \begin{cases} h_k = 0 \;\;\Rightarrow\;\; p_k = 0, \quad \forall k=1,\ldots,K, \\ -\displaystyle\sum_{\substack{k=1\\ h_k\neq 0}}^{K} h_k\Big(\xi_k + \big(\frac{p_k}{2h_k}\big)\xi_{K+1}\Big)^2 + \Big(1 - q + \displaystyle\sum_{\substack{k=1\\ h_k\neq 0}}^{K}\frac{p_k^2}{4h_k}\Big)\xi_{K+1}^2 \ge 0, \quad \forall \xi_1,\ldots,\xi_K,\xi_{K+1}\in\mathbb{R} \end{cases}
\end{aligned}
$$

$$\Leftrightarrow \quad \begin{cases} h_k = 0 \quad \Rightarrow \quad p_k = 0, \quad \forall k = 1, \ldots, K, \\ h_k \leq 0, \\ 1 - q + \sum\limits_{\substack{k=1 \\ h_k \neq 0}}^{K} \dfrac{p_k^2}{4h_k} \geq 0 \end{cases}$$

$$\Leftrightarrow \; \exists t_k \geq 0, \quad \forall k = 1, \ldots, K, \text{ such that}$$
$$\begin{cases} p_k^2 \leq -4h_k t_k, \quad \forall k = 1, \ldots, K, \\ q + \sum\limits_{k=1}^{K} t_k \leq 1 \end{cases}$$

$$\Leftrightarrow \; \exists t_k \geq 0, \quad \forall k = 1, \ldots, K, \text{ such that}$$
$$\begin{cases} \left\| \begin{bmatrix} p_k \\ t_k + h_k \end{bmatrix} \right\| \leq t_k - h_k, \quad \forall k = 1, \ldots, K, \\ q + \sum\limits_{k=1}^{K} t_k \leq 1. \end{cases} \tag{48}$$

By substituting the simplification (45) back into equation (48), we obtain the reformulation of constraint (46) as follows

$$\begin{cases} \left\| \begin{bmatrix} p_k^{\mathrm{L}} - p_k^{\mathrm{U}} \\ t_k + h_k^{\mathrm{L}} - h_k^{\mathrm{U}} \end{bmatrix} \right\| \leq t_k - h_k^{\mathrm{L}} + h_k^{\mathrm{U}}, \quad \forall k = 1, \ldots, K, \\ q + \sum\limits_{k=1}^{K} t_k \leq 1 \\ t_k \geq 0, \quad \forall k = 1, \ldots, K. \end{cases}$$

The formulation of constraint (47) can be similarly obtained and is omitted, which completes the proof. ■