Approximate Methods for Solving Chance-Constrained Linear Programs in Probability Measure Space

Xun Shen[1] · **Satoshi Ito[2]**

Abstract

A risk-aware decision-making problem can be formulated as a chance-constrained linear program in probability measure space. Chance-constrained linear program in probability measure space is intractable, and no numerical method exists to solve this problem. This paper presents numerical methods to solve chance-constrained linear programs in probability measure space for the first time. We propose two solvable optimization problems as approximate problems of the original problem. We prove the uniform convergence of each approximate problem. Moreover, numerical experiments have been implemented to validate the proposed methods.

1 Introduction

Let $\mathcal{X} \subset \mathbb{R}^n$ be a compact set with the infinity norm defined by $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|, x \in \mathcal{X}$. Denote $D > 0$ such that $D := \sup\{\|x - x'\|_\infty : x, x' \in \mathcal{X}\}$ for the diameter of $\mathcal{X}$. In this paper, we assume that $\mathcal{X}$ can be specified as $\mathcal{X} = \{x \in \mathbb{R}^n : g(x) \le \mathbf{0}^{n_g}\}$ where $g : \mathbb{R}^n \to \mathbb{R}^{n_g}$ is a continuously differentiable constraint function. We have the following assumption on g throughout the paper.

Communicated by Nobuo Yamashita.

✉ Xun Shen
   shenxun@eei.eng.osaka-u.ac.jp

   Satoshi Ito
   sito@ism.ac.jp

[1] Graduate School of Engineering, Osaka University, Osaka 565-0871, Japan

[2] Department of Statistical Inference and Mathematics, The Institute of Statistical Mathematics, Tokyo 190-8562, Japan

Assumption 1 Cottle Constraint Qualification (CCQ) holds at any points in $\mathcal{X}$. Namely, for any $x \in \mathcal{X}$, there is a $d \in \mathbb{R}^n$ such that

$$\nabla g(x)^\top d < \mathbf{0}^{n_g} \tag{1}$$

holds.

Let $\mathscr{B}(\mathcal{X})$ be Borel σ-algebra on metric space $\mathcal{X}$. This paper uses $\mathscr{B}(\cdot)$ to denote the Borel σ-algebra on a metric space. Notice that $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$ is a Borel space. Let μ be a Borel probability measure on $\mathscr{B}(\mathcal{X})$. Let $M(\mathcal{X})$ be the space of Borel probability measures on metric space $\mathcal{X}$. Let δ be a random vector with support $\Delta \subseteq \mathbb{R}^s$ and $\mathbb{P}\{\cdot\}$ be the probability measurable defined on Borel σ-algebra $\mathscr{B}(\Delta)$ on Δ. Let $p(\delta)$ be the probability density function associated with $\mathbb{P}\{\cdot\}$. Given a scalar function $J : \mathcal{X} \to \mathbb{R}$, and a vector-valued function $h : \mathcal{X} \times \Delta \to \mathbb{R}^m$, a chance-constrained linear program in probability measure space is formulated as:

$$
\begin{aligned}
\min_{\mu \in M(\mathcal{X})} \quad & \int_{\mathcal{X}} J(x) \mathrm{d}\mu \\
\text{s.t.} \quad & \int_{\mathcal{X}} F(x) \mathrm{d}\mu \geq 1 - \alpha,
\end{aligned}
\tag{P_α}
$$

where $\alpha \in (0, 1)$ is a given probability level and $F(x)$ is defined by

$$F(x) := \int_\Delta \mathbb{I}\{h(x, \delta)\} \mathrm{d}\mathbb{P}\{\delta\} = \int_\Delta \mathbb{I}\{h(x, \delta)\} p(\delta) \mathrm{d}\delta. \tag{2}$$

Here, $\mathbb{I}\{y\}$ presents the indicator function written as

$$\mathbb{I}\{y\} = \begin{cases} 1, & \text{if } y \leq 0, \\ 0, & \text{if } y > 0. \end{cases}$$

Note that $F(x)$ is the probability of having $h(x, \delta) \leq 0$ for given x. Throughout the paper, we assume the following conditions on $J(x)$ and $h(x, \delta)$.

Assumption 2 For functions $J(x)$ and $h(x, \delta)$, the followings are supposed to be held:

a. $J(x)$ is continuously differentiable with respect to x;
b. $h(x, \delta)$ is continuously differentiable with respect to x for any $\delta \in \Delta$;
c. For every $x \in \mathcal{X}$, $h(x, \delta)$ is continuous with respect to δ;
d. The probability density function $p(\delta)$ is continuous with respect to δ;
e. Let $\bar{h}(x, \delta) := \max_i h_i(x, \delta)$, $\mathsf{supp}\, p := \mathsf{cl}\{\delta \in \Delta : p(\delta) > 0\}$ ($\mathsf{cl}\{\cdot\}$ denotes the closure), and for each $x \in \mathcal{X}$,

$$\Delta^{\mathsf{supp}}(x) := \{\delta \in \mathsf{supp}\, p : \bar{h}(x, \delta) = 0\}.$$

For each $x \in \mathcal{X}$, the following is assumed to be true:

$$\mathbb{P}\{\Delta^{\mathsf{supp}}(x)\} = 0.$$

Besides, suppose that $h(x, \delta)$ has a continuous probability density function for every $x \in \mathcal{X}$;

f. There exists $L > 0$ such that

$$\|h(x, \delta) - h(x', \delta)\|_\infty \leq L\|x - x'\|_\infty, \quad \forall x, x' \in \mathcal{X} \text{ and } \forall \delta \in \Delta,$$

and

$$|J(x) - J(x')| \leq L\|x - x'\|_\infty, \quad \forall x, x' \in \mathcal{X}.$$

In fact, according to the content of pp. 78–79 of [18], we can obtain the continuity of $F(x)$ from Assumption 2.

Denote the feasible region of P_α as $M_\alpha(\mathcal{X}) := \{\mu \in M(\mathcal{X}) : \int_{\mathcal{X}} F(x)\mathrm{d}\mu \geq 1 - \alpha\}$. The optimal objective function value of P_α is

$$\bar{\mathcal{J}}_\alpha := \min\left\{\int_{\mathcal{X}} J(x)\mathrm{d}\mu : \mu \in M_\alpha(\mathcal{X})\right\}. \tag{3}$$

The optimal solution set of P_α is therefore written as

$$A_\alpha := \left\{\mu \in M_\alpha(\mathcal{X}) : \int_{\mathcal{X}} J(x)\mathrm{d}\mu = \bar{\mathcal{J}}_\alpha\right\}, \tag{4}$$

$\bar{\mu}_\alpha \in A_\alpha$ is called an optimal measure for P_α.

1.1 Motivation

The motivation for addressing chance-constrained linear programs in probability measure space is from seeking an optimal stochastic policy for the optimal control problem with chance constraints, which is vital for the deployment of reliable autonomous systems by control algorithms that are robust to model misspecifications and for external disturbances [2, 10, 28]. The optimal control problem with chance constraints aims at maximizing a reward function or minimizing a cost function with the constraints that the system state should locate in the safe area with a required probability. The deterministic policy has a fixed value in the decision domain at every time index. In contrast, the stochastic policy provides a probability measure on the decision domain at every time index. The deterministic policy can be regarded as a particular case of the stochastic policy by concentrating the probability measure on a fixed value in the decision domain. The existing techniques for addressing optimal control problems with chance constraints do not touch the essential parts of the problem and may require application-specific assumptions. For example, [10, 17] enforces pointwise chance constraints that ensure the independent satisfaction of each chance constraint at each time step, which leads to a more conservative solution. In general, joint chance constraints are desired, which requires all chance constraints to be satisfied jointly at all times. However, it is challenging to tackle the joint chance-constrained optimal control problem since the distribution of the state trajectory needs to be considered

fully. It is possible to address the joint chance-constrained optimal control problem by using Boole's inequality [2, 24, 36] or performing robust optimization within the bounded model parameters obtained by specifying a confident set [19]. However, these two methods are conservative. More investigations from the viewpoint of optimization theory should be addressed to enhance new breakthroughs for optimal control with chance constraints.

Obtaining open-loop stochastic optimal policies under chance constraints can be essentially written as a chance-constrained linear program in probability measure space [32]. Open-loop stochastic policies mean that the stochastic policies only depend on the initial state. Unfortunately, there is still no research on solving chance-constrained linear programs in probability measure space to our knowledge. Investigating the chance-constrained linear programs in probability measure space is vital, which can give more insights into optimal control with chance constraints.

1.2 Related Works

Optimization with finite chance constraints in finite-dimensional vector space is generally challenging due to the nonconvexity of the feasible set and intractable reformulations [9, 27]. The existing research has two major streams: (1) give assumptions that the constraint functions or the distribution of random variables has some special structure, for example, linear or convex constraint functions [23], finite sample space of random variables [21], elliptically symmetric Gaussian-similar distributions [33], or (2) extract samples [5–7, 20, 25, 26, 29, 31] or use smooth functions [13] to approximate the chance constraints. For sample-based methods, the most famous approach in the control field is scenario approach [5–8, 28]. Scenario approach generates a deterministic optimization problem as the approximation of the original one by extracting samples from the sample space of random variables. The probability of the feasibility of the approximate solution rapidly increases to one as the sample number increases. However, the convergence of the optimality of the approximate solution is not discussed. In another sample-based method, the sample-average approach [13, 20, 26, 29], both feasibility and optimality of the approximate solution are presented. However, neither scenario approach nor sample-average approach can be directly used to solve chance-constrained linear programs in probability measure space since the deduction of the convergence of either scenario approach or sample-average approach assumes that the dimension of the decision variable must be finite.

Optimization with chance/robust constraints in finite-dimensional vector space is also intensively studied, in which the number of chance constraints is infinite [1, 11, 34, 35]. In [34], the generalized differentiation of the probability function of infinite constraints is investigated. The optimality condition with an explicit formulation of subdifferentials is given. In [35], the variational tools are applied to formulate generalized differentiation of chance/robust constraints. The method of getting the explicit outer estimations of subdifferentials from data is also established. An adaptive grid refinement algorithm is developed to solve the optimization with chance/robust constraints in [1]. However, the above research on optimization with chance/robust

constraints in finite-dimensional vector space can prove convergence only when the dimension of the decision variable is finite.

Recently, chance constraints in infinite dimensions have attracted a lot of attention. In [12, 14, 15], some essential properties, such as convexity and semi-continuity, are generalized into the chance constraints in infinite dimensions. However, the results in [12] assume that the random variable should have a log-concave density to ensure the semicontinuity. In [15], the continuity of the probability function as chance constraints is proved under the assumption of continuous random distributions. The properties of chance constraints in infinite dimensions are crucial to constructing the optimality condition and implementing convergence analysis for optimization with chance constraints in infinite dimensions. In [14], chance-constrained optimization of elliptic partial differential equation systems is addressed by inner–outer approximation. It proves that the inner and outer approximation converges to the original problem and can provide approximate solutions with ensured convergence. However, the proof of the convergence requires the assumption that the state domain is convex. Besides, it concerns the specific problem in partial differential equation systems.

1.3 Overview of Proposed Method and Contributions

This paper extends the sample-based approximation method to solve chance-constrained linear programs in probability measure space. We show the relationship between chance-constrained optimization in finite-dimensional vector space and chance-constrained linear program in probability measure space. By solving a chance-constrained linear program in probability measure space, we can obtain a stochastic policy to improve the expectation of the optimal value further. We also show that the optimal objective values of the chance-constrained linear program in probability measure space and chance-constrained optimization in finite-dimensional vector space are equal if the constraints involved with random variables are required to be satisfied with probability 1. Namely, in this case, by concentrating the probability measure on an optimal solution of chance-constrained optimization in finite-dimensional vector space, we can obtain an optimal measure for the chance-constrained linear program in probability measure space. Besides, a sample approximate problem and a Gaussian mixture model approximate problem of problem P_α are proposed, by solving which the approximate solution of P_α can be obtained. The convergences of both approximate problems are investigated. Numerical examples are implemented to validate the proposed methods.

Chance-constrained linear program in probability measure space involves chance constraints in infinite dimensions. Our work differs from the [12, 15] in that our purpose is to provide numerical methods for solving chance-constrained linear programs in probability measure space. The properties of chance constraints in infinite dimensions are essential for convergence analysis.

The rest of this paper is organized as follows: Sect. 2 presents two approximate problems of P_α and gives the main results on the convergence for each approximate problem. The proofs of the main results are presented in Sect. 3. Section 4 presents

the results of two numerical examples, which show the effectiveness of our proposed methods. Section 5 concludes the whole paper.

2 Main Results

This section introduces two approximate problems of P_α. We also present the convergence for each approximate problem. The proofs are presented in Sect. 3.

2.1 Chance-Constrained Optimization in Finite Space

Chance-constrained optimization Q_α is an optimization problem with chance constraints in a finite-dimension vector space. The problem is written as

$$\min_{x \in \mathcal{X}} \ J(x)$$
$$\text{s.t.} \quad F(x) \geq 1 - \alpha, \tag{Q_α}$$

where $\alpha \in (0, 1)$ is a given probability level.

Let $\mathcal{X}_\alpha := \{x \in \mathcal{X} : F(x) \geq 1 - \alpha\}$ be the feasible domain of Q_α. Denote $\bar{J}_\alpha := \min\{J(x) : x \in \mathcal{X}_\alpha\}$ for the optimal objective value of Q_α and $X_\alpha := \{x \in \mathcal{X}_\alpha : J(x) = \bar{J}_\alpha\}$ for the optimal solution set of Q_α. We have the following assumptions over Q_α throughout the paper.

Assumption 3 There exists a globally optimal solution $\bar{x}$ of Q_α such that for any $\varepsilon > 0$ there is $x \in \mathcal{X}$ such that $0 < \|x - \bar{x}\| \leq \varepsilon$ and $F(x) > 1 - \alpha$.

The existence of chance constraints gives rise to several difficulties. First, the structural properties of $h(x, \delta)$ might not be passed to $F(x) \geq 1 - \alpha$. The feasible set $\mathcal{X}_\alpha$ can be equivalently obtained as

$$\mathcal{X}_\alpha = \bigcup_{\Delta_s \in \mathcal{F}} \bigcap_{\delta \in \Delta_s} \mathcal{X}_\delta, \tag{5}$$

where $\mathcal{X}_\delta := \{x \in \mathcal{X} : h(x, \delta) \leq 0\}$ and $\mathcal{F} := \{\Delta_s \in \mathcal{B}(\Delta) : \mathbb{P}\{\Delta_s\} \geq 1 - \alpha\}$. Even if $h_i(x, \delta), i = 1, \ldots, m$ are all linear in x for every $\delta \in \Delta$, the feasible set $\mathcal{X}_\alpha$ may not be convex due to the infinite union operations. Second, it is difficult to obtain a tractable analytical function $F(x)$ to describe the constraint or find a numerically efficient way to compute it. In most applications, $p(\delta)$ is unknown, and only samples of δ are available. We briefly review the sample-based approximation method presented in [20, 25, 26]. Let $\mathcal{D}_N = \{\delta^{(1)}, \ldots, \delta^{(N)}\}$ be a set of samples randomly extracted from Δ where $N \in \mathbb{N}$. Suppose the sample extraction is independently and identically distributed. Then, $\mathcal{D}_N$ can be regarded as a random variable from the augmented sample space Δ^N with probability measure $\mathbb{P}^N\{\cdot\}$ defined on the Borel σ-algebra $\mathcal{B}(\Delta^N)$. Giving $\mathcal{D}_N, \epsilon \in [0, \alpha)$, and $\gamma > 0$, a sample average approximate problem of Q_α, defined by $\tilde{Q}_{\epsilon,\gamma}(\mathcal{D}_N)$, is written as:

$$\min_{x \in \mathcal{X}} J(x)$$

$$\text{s.t.} \quad \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\{h(x, \delta^{(j)}) + \gamma\} \geq 1 - \epsilon. \qquad (\tilde{Q}_{\epsilon,\gamma}(\mathcal{D}_N))$$

The feasible region of $\tilde{Q}_{\epsilon,\gamma}(\mathcal{D}_N)$ is defined by

$$\tilde{\mathcal{X}}_{\epsilon,\gamma}(\mathcal{D}_N) := \left\{ x \in \mathcal{X} : \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\{h(x, \delta^{(j)}) + \gamma\} \geq 1 - \epsilon \right\}.$$

Denote $\tilde{J}_{\epsilon,\gamma}(\mathcal{D}_N) := \min\{J(x) : x \in \tilde{\mathcal{X}}_{\epsilon,\gamma}(\mathcal{D}_N)\}$ for the optimal objective function value of $\tilde{Q}_{\epsilon,\gamma}(\mathcal{D}_N)$ and $\tilde{X}_{\epsilon,\gamma}(\mathcal{D}_N) := \{x \in \tilde{\mathcal{X}}_{\epsilon,\gamma}(\mathcal{D}_N) : J(x) = \tilde{J}_{\epsilon,\gamma}(\mathcal{D}_N)\}$ for the optimal solution set of $\tilde{Q}_{\epsilon,\gamma}(\mathcal{D}_N)$. We can regard $\tilde{J}_{\epsilon,\gamma}(\mathcal{D}_N)$ as a function $\tilde{J}_{\epsilon,\gamma} : \Delta^N \to \mathbb{R}$ for given ϵ and γ. Since $\mathcal{D}_N$ is a random variable from Δ^N, $\tilde{J}_{\epsilon,\gamma}(\mathcal{D}_N)$ is consequently a random variable. The sets $\tilde{\mathcal{X}}_{\epsilon,\gamma}(\mathcal{D}_N)$ and $\tilde{X}_{\epsilon,\gamma}(\mathcal{D}_N)$ also depend on $\mathcal{D}_N$ and can be regarded as $\tilde{\mathcal{X}}_{\epsilon,\gamma} : \Delta^N \to \mathscr{B}(\mathcal{X})$ and $\tilde{X}_{\epsilon,\gamma} : \Delta^N \to \mathscr{B}(\mathcal{X})$. $\tilde{\mathcal{X}}_{\epsilon,\gamma}(\mathcal{D}_N)$ and $\tilde{X}_{\epsilon,\gamma}(\mathcal{D}_N)$ are called random sets [22]. In [20, 26], the convergence analysis on $\tilde{\mathcal{X}}_{\epsilon,\gamma}(\mathcal{D}_N)$, $\tilde{X}_{\epsilon,\gamma}(\mathcal{D}_N)$, $\tilde{J}_{\epsilon,\gamma}(\mathcal{D}_N)$ is given. We summarize Theorem 10 of [20] and Theorem 3.5 of [26] as Lemma 1.

Lemma 1 *Suppose that Assumptions* 2 *and* 3 *hold. Let* $\epsilon \in [0, \alpha)$, $\beta \in (0, \alpha - \epsilon)$ *and* $\gamma > 0$. *Then,*

$$\mathbb{P}^N\{\tilde{\mathcal{X}}_{\epsilon,\gamma}(\mathcal{D}_N) \subseteq \mathcal{X}_\alpha\} \geq 1 - \left\lceil \frac{1}{\eta} \right\rceil \left\lceil \frac{2LD}{\gamma} \right\rceil^n \exp\{-2N(\alpha - \epsilon - \beta)^2\}.$$

Besides, $\tilde{X}_{\epsilon,\gamma}(\mathcal{D}_N) \to X_\alpha$ *and* $\tilde{J}_{\epsilon,\gamma}(\mathcal{D}_N) \to \bar{J}_\alpha$ *with probability 1 when* $N \to \infty$, $\epsilon \to \alpha$, $\gamma \to 0$.

According to Lemma 1, we can obtain the solution of Q_α with probability 1 when $N \to \infty$, $\epsilon \to \alpha$, $\gamma \to 0$. A natural question arises: can we use the solution of Q_α to obtain an optimal probability measure for P_α? Let $\bar{x}_\alpha \in X_\alpha$ be an optimal solution of Q_α. Notice that we have $\{\bar{x}_\alpha\} \in \mathscr{B}(\mathcal{X})$ and thus it is possible to define a probability measure $\mu_{\bar{x}_\alpha}$ which satisfies that $\mu_{\bar{x}_\alpha}(\{\bar{x}_\alpha\}) = \mu_{\bar{x}_\alpha}(\mathcal{X}) = 1$. Then,

$$\int_{\mathcal{X}} J(x) \mathrm{d}\mu_{\bar{x}_\alpha} = \int_{\{\bar{x}_\alpha\}} J(x) \mathrm{d}\mu_{\bar{x}_\alpha} = \bar{J}_\alpha$$

and

$$\int_{\mathcal{X}} F(x) \mathrm{d}\mu_{\bar{x}_\alpha} = \int_{\{\bar{x}_\alpha\}} F(x) \mathrm{d}\mu_{\bar{x}_\alpha} = F(\bar{x}_\alpha) \geq 1 - \alpha.$$

Thus, $\mu_{\bar{x}_\alpha}$ is a feasible solution for P_α with objective value as $\bar{J}_{x,\alpha}$. However, $\mu_{\bar{x}_\alpha}$ is not sure to locate in A_α. Only when $\alpha = 0$, we have $\mu_{\bar{x}_\alpha} \in A_\alpha$. Notice that it

is not ensured that the set X_α is a Borel measurable set. However, it is possible to find a subset $X_\alpha^{\mathrm{m}} \subseteq X_\alpha$ that is Borel measurable. A particular example is to choose $X_\alpha^{\mathrm{m}} = \{\bar{x}_\alpha\}$ where $\bar{x}_\alpha \in X_\alpha$ is one element in the optimal solution set. In this paper, without loss of generality, we assume that X_α is Borel measurable for all $\alpha \in [0, 1]$. Besides, we also assume that $\mathcal{X}_0 \neq \emptyset$. Then, $\mathcal{X}_\alpha \neq \emptyset$ holds for all $\alpha \in [0, 1]$. The above content is formally summarized in Theorem 1.

Theorem 1 *Suppose that $\mathcal{X}_\alpha$ is measurable for all $\alpha \in [0, 1]$ and $\mathcal{X}_0 \neq \emptyset$. The optimal value of problem P_α satisfies $\bar{\mathcal{J}}_\alpha \leq \bar{J}_\alpha$. Besides, if $\alpha = 0$, we have*

$$\bar{\mathcal{J}}_0 = \bar{J}_0$$

and

$$A_0 = \{\mu \in M(\mathcal{X}) : \mu(\mathcal{X}) = \mu(X_0) = 1\} \tag{6}$$

with probability 1.

The proof of Theorem 1 is given in Sect. 3.1.

Remark 1 Theorem 1 implies that deterministic policy is optimal for robust optimal control where $\alpha = 0$.

2.2 Sample-Based Approximation

Let $\mathcal{X}^{\mathrm{in}}$ be the set of all interior points of $\mathcal{X}$. By using Hit-and-Run algorithm [30] and Billiard Walk algorithm [16], uniform samples can be generated from $\mathcal{X}^{\mathrm{in}}$. For a positive integer $S \in \mathbb{N}$, let $\mathcal{C}_S := \{x^{(1)}, \ldots, x^{(S)}\}$ be a set of uniform samples independently extracted from $\mathcal{X}^{\mathrm{in}}$. The set $\mathcal{C}_S$ is an element of the augmented space $(\mathcal{X}^{\mathrm{in}})^S$. Since each element $x^{(i)}, i = 1, \ldots, S$ in $\mathcal{C}_S$ is extracted independently, we define a S-fold probability $\mathbb{P}_{\mathrm{uni}}^S (= \mathbb{P}_{\mathrm{uni}} \times \cdots \times \mathbb{P}_{\mathrm{uni}}, S$ times$)$ in $(\mathcal{X}^{\mathrm{in}})^S$. Here, $\mathbb{P}_{\mathrm{uni}}$ is the probability measure of uniform distribution on $\mathcal{X}^{\mathrm{in}}$.

With $\mathcal{C}_S$ and $\mathcal{D}_N$, we can obtain a sample approximate problem of P_α defined by $\tilde{P}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$:

$$
\begin{aligned}
\min_{\mu \in U_S} \quad & \sum_{i=1}^{S} J(x^{(i)})\mu(i) \\
\text{s.t.} \quad & \sum_{i=1}^{S} \mu(i) \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\{h(x^{(i)}, \delta^{(j)})\} \geq 1 - \alpha,
\end{aligned}
\qquad (\tilde{P}_\alpha(\mathcal{C}_S, \mathcal{D}_N))
$$

where $U_S := \{\mu \in \mathbb{R}^S : \sum_{i=1}^{S} \mu(i) = 1, \mu(i) \geq 0, \ \forall i = 1, \ldots, S\}$. Define $\mathcal{F}_\alpha(\mathcal{C}_S, \mathcal{D}_N) := \{\mu \in U_S : \sum_{i=1}^{S} \mu(i) \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\{h(x^{(i)}, \delta^{(j)})\} \geq 1 - \alpha\}$ as the feasible set of $\tilde{P}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$. Denote the optimal objective function value as

$$\tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_N) := \min\left\{\sum_{i=1}^{S} J(x^{(i)})\mu(i) : \mu \in \mathcal{F}_\alpha(\mathcal{C}_S, \mathcal{D}_N)\right\}.$$

Denote the optimal solution set for $\tilde{P}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ as

$$\tilde{A}_\alpha(\mathcal{C}_S, \mathcal{D}_N) := \left\{\mu \in \mathcal{F}_\alpha(\mathcal{C}_S, \mathcal{D}_N) : \sum_{i=1}^{S} J(x^{(i)})\mu(i) = \tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_N)\right\}.$$

Let $\tilde{\mu}_\alpha \in \tilde{A}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ be an optimal measure. The optimal value $\tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ depends on $\mathcal{C}_S$ and $\mathcal{D}_S$, and thus it can be regarded as a function $\tilde{\mathcal{J}}_\alpha : \mathcal{X}^S \times \Delta^N \to \mathbb{R}$. Then, $\tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ is a random variable. Besides, $\tilde{A}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ is a random set.

The deduction of the convergences of $\tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ and $\tilde{A}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ requires another assumption on P_α. We state the assumption after a brief introduction of weak convergence.

Define a space of continuous $\mathbb{R}$-valued functions by

$$\mathscr{C}(\mathcal{X}, \mathbb{R}) := \{f : \mathcal{X} \to \mathbb{R} | f \text{ is continuous}\}. \tag{7}$$

It is able to define a metric on $\mathscr{C}(\mathcal{X}, \mathbb{R})$ by

$$\tau(f, f') := \|f - f'\|_\infty, \tag{8}$$

where $\|f\|_\infty$ is defined as

$$\|f\|_\infty := \sup_{x \in \mathcal{X}} |f(x)|.$$

The metric $\tau(\cdot, \cdot)$ turns $\mathscr{C}(\mathcal{X}, \mathbb{R})$ into a complete metric space.

The weak convergence of probability measures is defined as follows [4].

Definition 1 Let $\{\mu_k\}_{k=0}^\infty$ be a sequence in $M(\mathcal{X})$. We say that $\{\mu_k\}_{k=0}^\infty$ converges weakly to μ if

$$\lim_{k \to \infty} \left| \int_{\mathcal{X}} f(x)\mathrm{d}\mu_k - \int_{\mathcal{X}} f(x)\mathrm{d}\mu \right| = 0, \quad \text{for all} \quad f \in \mathscr{C}(\mathcal{X}, \mathbb{R}). \tag{9}$$

Since $\mathcal{X}$ is compact, $M(\mathcal{X})$ can be proved to be weakly compact by Riesz representation theorem [4]. Therefore, giving any sequence of $\{\mu_k\}_{k=0}^\infty \subset M(\mathcal{X})$, there is a subsequence which converges weakly to some $\mu \in M(\mathcal{X})$ in the sense of Definition 1. By Assumption 2, we have that $J(x)$ and $F(x)$ are continuous with respect to x. Therefore, if $\{\mu_k\}_{k=0}^\infty$ converges weakly to μ, (9) also holds for $J(x)$ or $F(x)$. We give the following assumption on Problem P_α.

Assumption 4 There exists a globally optimal solution $\mu^* \in A_\alpha$ of Problem P_α such that for any $\delta > 0$ there is $\mu \in M(\mathcal{X})$ such that $\int_{\mathcal{X}} F(x)\mathrm{d}\mu > 1 - \alpha$ and $\mathcal{W}(\mu, \mu^*) \leq \delta$, where $\mathcal{W}(\mu, \mu^*)$ is defined by

$$\mathcal{W}(\mu, \mu^*) = \left| \int_{\mathcal{X}} J(x)\mathrm{d}\mu - \int_{\mathcal{X}} J(x)\mathrm{d}\mu^* \right|. \tag{10}$$

As $S, N \to \infty$, the convergence analysis on $\tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ and $\tilde{A}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ is summarized in Theorem 2.

Theorem 2 *Consider Problem P_α with $\alpha > 0$. Suppose Assumptions 1, 2, 3, and 4 hold. As $S, N \to \infty$, we have*

$$\liminf_{S,N\to\infty} \tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_N) = \bar{\mathcal{J}}_\alpha,$$

with probability 1. Besides, as $S, N \to \infty$, we have $\tilde{A}_\alpha(\mathcal{C}_S, \mathcal{D}_N) \subset M_\alpha(\mathcal{X}) := \{\mu \in M(\mathcal{X}) : \int_{\mathcal{X}} F(x)\mathrm{d}\mu \geq 1 - \alpha\}$ with probability 1.

The proof of Theorem 2 is given in Sect. 3.2.

2.3 Gaussian Mixture Model-Based Approximation

Another option of approximation is to constrain the choice of μ in $M_\theta(\mathcal{X}) \subseteq M(\mathcal{X})$. Here, $M_\theta(\mathcal{X})$ is defined as

$$M_\theta(\mathcal{X}) := \left\{ \mu \in M(\mathcal{X}) : \mu(X) = \int_X p_\theta(x)\mathrm{d}x, \ \forall X \subseteq \mathcal{X} \right\},$$

where the probability density function $p_\theta(x)$ is written as

$$p_\theta(x) = \sum_{i=1}^{L} \omega_i \phi(x, m_i, \Sigma_i). \tag{11}$$

Here, $\omega_i \in [0, 1], \forall i = 1, \ldots, L$, $\sum_{i=1}^{L} \omega_i = 1$, and $\phi(x, m_i, \Sigma_i)$ is multivariate Gaussian distribution written by

$$\phi(x, m_i, \Sigma_i) = \frac{1}{(2\pi)^{n/2}|\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x - m_i)^\top \Sigma_i^{-1}(x - m_i) \right).$$

The notation θ denotes the parameter vector, including all the unknown parameters in $\omega_i, m_i, \Sigma_i, \forall i = 1, \ldots, L$. Denote the dimension of θ as n_θ. The feasible domain of θ is denoted by

$$\Theta := \left\{ \theta \in \mathbb{R}^{n_\theta} : \sum_{i=1}^{L} \omega_i = 1, \ \omega_i \geq 0 \right\}.$$

Then, given a data set $\mathcal{D}_N$ and the number of Gaussian distributions L, we can obtain a Gaussian mixture model-based approximate problem of P_α defined by $\hat{P}_\alpha(L, \mathcal{D}_N)$:

$$
\begin{aligned}
&\min_{\theta \in \Theta} \int_{\mathcal{X}} J(x) p_\theta(x) \mathrm{d}x \\
&\text{s.t.} \quad \int_{\mathcal{X}} \sum_{j=1}^{N} \frac{1}{N} \mathbb{I}\{h(x, \delta^{(j)})\} p_\theta(x) \mathrm{d}x \geq 1 - \alpha.
\end{aligned}
\qquad (\hat{P}_\alpha(L, \mathcal{D}_N))
$$

Denote the feasible set of $\hat{P}_\alpha(L, \mathcal{D}_N)$ as

$$
\Theta_\alpha(L, \mathcal{D}_N) := \left\{ \theta \in \Theta : \int_{\mathcal{X}} \sum_{j=1}^{N} \frac{1}{N} \mathbb{I}\{h(x, \delta^{(j)})\} p_\theta(x) \mathrm{d}x \geq 1 - \alpha \right\},
$$

and the optimal objective value as

$$
\hat{\mathcal{J}}_\alpha(L, \mathcal{D}_N) := \min \left\{ \int_{\mathcal{X}} J(x) p_\theta(x) \mathrm{d}x : \theta \in \Theta_\alpha(L, \mathcal{D}_N) \right\}.
$$

Besides, the optimal solution set is

$$
\hat{\Theta}_\alpha(L, \mathcal{D}_N) := \left\{ \theta \in \Theta_\alpha(L, \mathcal{D}_N) : \int_{\mathcal{X}} J(x) p_\theta(x) \mathrm{d}x = \hat{\mathcal{J}}_\alpha(L, \mathcal{D}_N) \right\}.
$$

The optimal objective value $\hat{\mathcal{J}}_\alpha(L, \mathcal{D}_N)$ depends on the number of used Gaussian models and the data set $\mathcal{D}_N$. Since data set $\mathcal{D}_N$ is essentially random variable with support Δ^N, $\hat{\mathcal{J}}_\alpha(L, \mathcal{D}_N)$ is also a random variable. The set $\hat{\Theta}_\alpha(L, \mathcal{D}_N)$ is a random set.

As $L, N \to \infty$, optimality and feasibility of using the optimal solution of $\hat{P}_\alpha(L, \mathcal{D}_N)$ are summarized in Theorem 3.

Theorem 3 *Consider Problem P_α with $\alpha > 0$. Suppose Assumptions 1, 2, 3, and 4 hold. As $L, N \to \infty$, we have*

$$
\liminf_{L, N \to \infty} \hat{\mathcal{J}}_\alpha(L, \mathcal{D}_N) = \bar{\mathcal{J}}_\alpha,
$$

with probability 1. Besides, let $\hat{\theta} \in \hat{\Theta}_\alpha(L, \mathcal{D}_N)$ be an optimal solution of $\hat{P}_\alpha(L, \mathcal{D}_N)$. The corresponding probability density function is $p_{\hat{\theta}}(x)$, and the obtained probability measure is

$$
\mu_{\hat{\theta}}(X) := \int_X p_{\hat{\theta}}(x) \mathrm{d}x, \quad \forall X \subseteq \mathcal{X}.
$$

We have $\mu_{\hat{\theta}} \in M_\alpha(\mathcal{X}) := \{\mu \in M(\mathcal{X}) : \int_{\mathcal{X}} F(x) \mathrm{d}\mu \geq 1 - \alpha\}$ with probability 1 as $L, N \to \infty$.

The proof of Theorem 3 is given in Sect. 3.3.

3 Proofs of Main Results

3.1 Proof of Theorem 1

Proof (Theorem 1) Define a measure by $\bar{\mu}_\alpha(\cdot)$, which satisfies that $\bar{\mu}_\alpha(X_\alpha) = 1$. Then, we have

$$\int_{\mathcal{X}_\alpha} J(x) d\bar{\mu}_\alpha = \int_{X_\alpha} J(x) d\bar{\mu}_\alpha = \bar{J}_\alpha.$$

Besides, for the constraint, we have

$$\int_{\mathcal{X}} F(x) d\bar{\mu}_\alpha(x) = \int_{X_\alpha} F(x) d\bar{\mu}_\alpha(x) \geq 1 - \alpha.$$

Then, $\bar{\mu}_\alpha(\cdot) \in M_\alpha(\mathcal{X})$ holds. Thus, we have $\bar{\mathcal{J}}_\alpha \leq \int_{\mathcal{X}} J(x) d\bar{\mu}_\alpha = \bar{J}_\alpha$.

When $\alpha = 0$, let $\mathcal{X}_0^c = \{x \in \mathcal{X} : F(x) < 1\}$ be the complement set of $\mathcal{X}_0$, namely, $\mathcal{X}_0^c \bigcup \mathcal{X}_0 = \mathcal{X}$ and $\mathcal{X}_0^c \bigcap \mathcal{X}_0 = \emptyset$. Notice that $\mathcal{X}_0^c$ is Borel measurable since $\mathcal{X}_0$ is Borel measurable. Suppose that there is $\tilde{\mu}(\cdot) \in M_0(\mathcal{X})$ such that $\tilde{\mu}(\mathcal{X}_0^c) > 0$. Then,

$$\int_{\mathcal{X}} F(x) d\tilde{\mu}(x) = \int_{\mathcal{X}_0} F(x) d\tilde{\mu}(x) + \int_{\mathcal{X}_0^c} F(x) d\tilde{\mu}(x) < \tilde{\mu}(\mathcal{X}_0) + \tilde{\mu}(\mathcal{X}_0^c) = 1, \quad (12)$$

which conflicts with that $\tilde{\mu} \in M_0(\mathcal{X})$. Therefore, we have $\mu(\mathcal{X}_0^c) = 0$ for all $\mu \in M_0(\mathcal{X})$, which implies that $\int_{\mathcal{X}} J(x) d\mu = \int_{\mathcal{X}_0} J(x) d\mu$ for all $\mu \in M_0(\mathcal{X})$.

Notice that X_0 is a Borel measurable set. Let $\mu_0^*(\cdot) \in A_0$ be an optimal probability measure for P_0 and suppose $\mu_0^*(X_0) < 1$ for deriving the contradiction. Thus, $\mu_0^*(\mathcal{X} \backslash X_0) > 0$. The corresponding objective function is

$$\begin{aligned}
\int_{\mathcal{X}} J(x) d\mu_0^* &= \int_{\mathcal{X}_0} J(x) d\mu_0^* \\
&= \int_{X_0} J(x) d\mu_0^* + \int_{\mathcal{X}_0 \backslash X_0} J(x) d\mu_0^* \\
&= \int_{X_0} \bar{J}_0 d\mu_0^* + \int_{\mathcal{X}_0 \backslash X_0} J(x) d\mu_0^* \quad (\because) \left(J(x) = \bar{J}_0, \forall x \in X_0 \right) \\
&= \bar{J}_0 \int_{X_0} d\mu_0^* + \int_{\mathcal{X}_0 \backslash X_0} J(x) d\mu_0^* \\
&= \mu_0^*(X_0) \cdot \bar{J}_0 + \int_{\mathcal{X}_0 \backslash X_0} J(x) d\mu_0^*. \quad (13)
\end{aligned}$$

Denote a measure by $\bar{\mu}_0(\cdot)$, which satisfies that $\bar{\mu}_0(X_0) = 1$. Then, we have

$$\int_{\mathcal{X}} J(x) d\bar{\mu}_0 - \int_{\mathcal{X}} J(x) d\mu_0^* = \int_{\mathcal{X}_0} J(x) d\bar{\mu}_0 - \int_{\mathcal{X}_0} J(x) d\mu_0^*$$

$$
\begin{aligned}
&= \bar{J}_0 - \mu_0^*(X_0)\bar{J}_0 - \int_{\mathcal{X}_0 \setminus X_0} J(x)\mathrm{d}\mu_0^* \\
&= (1 - \mu_0^*(X_0)) \cdot \bar{J}_0 - \int_{\mathcal{X}_0 \setminus X_0} J(x)\mathrm{d}\mu_0^* \\
&= \int_{\mathcal{X}_0 \setminus X_0} (\bar{J}_0 - J(x))\mathrm{d}\mu_0^* \\
&< \int_{\mathcal{X}_0 \setminus X_0} (J(x) - J(x))\mathrm{d}\mu_0^* = 0.
\end{aligned}
\tag{14}
$$

Thus, $\mu_0^*(\cdot)$ is not the optimal measure. Therefore, (6) holds, which leads to $\bar{\mathcal{J}}_0 = \bar{J}_0$. □

3.2 Proof of Theorem 2

Lemma 2 *Suppose that Assumption 1 holds. For any $x \in \mathcal{X}$, denote a set as*

$$
\mathcal{B}_\varepsilon(x) := \{y \in \mathcal{X} : \|x - y\| \le \varepsilon\}
$$

where $\varepsilon > 0$ is radius. For any $\varepsilon > 0$, we have

$$
\lim_{S \to \infty} \mathbb{P}_{\mathsf{uni}}^S \left\{ \mathcal{C}_S \bigcap \mathcal{B}_\varepsilon(x) \ne \emptyset \right\} = 1.
\tag{15}
$$

Lemma 2 First, we show that the interior point set $\mathcal{X}^{\mathsf{in}}$ is not empty when Assumption 1 holds. Let $\bar{x} \in \mathcal{X}$ and thus we have

$$
g(\bar{x}) \le \mathbf{0}^{n_g}.
\tag{16}
$$

By Assumption 1, CCQ holds at $\bar{x}$. Thus, there exists $d \in \mathbb{R}^n$ such that

$$
\nabla g(\bar{x})^\top d < \mathbf{0}^{n_g}.
\tag{17}
$$

Notice that (16) and (17) directly give

$$
g(\bar{x}) + \nabla g(\bar{x})^\top d < \mathbf{0}^{n_g}.
\tag{18}
$$

Since $g(\cdot)$ is continuously differentiable, there exists a small enough $\bar{\xi} > 0$ such that $g(\bar{x} + \xi d) < 0$ holds for any $\xi \in (0, \bar{\xi})$ and thus $\bar{x} + \xi d \in \mathcal{X}^{\mathsf{in}}$. It implies that $\mathcal{X}^{\mathsf{in}}$ is not empty.

We start from discussing $\mathbb{P}_{\mathsf{uni}}^S\{\mathcal{C}_S \bigcap \mathcal{B}_\varepsilon(x) \ne \emptyset\}$ for $x \in \mathcal{X}^{\mathsf{in}}$. Notice that $\mathcal{X}$ is compact and $\mathcal{C}_S$ is a set of uniform samples extracted from $\mathcal{X}^{\mathsf{in}}$. Thus, for any $x \in \mathcal{X}^{\mathsf{in}}$, the probability that a sample $x^{(i)} \in \mathcal{C}_S$, $i = 1, .., S$ locates in $\mathcal{B}_\varepsilon(x)$ is

$$
\mathbb{P}_{\mathsf{uni}}\{x^{(i)} \in \mathcal{B}_\varepsilon(x)\} > 0.
$$

Then,

$$\mathbb{P}_{\mathsf{uni}}^{S}\left\{\mathcal{C}_S \bigcap \mathcal{B}_\varepsilon(x) \neq \emptyset\right\} = 1 - \mathbb{P}_{\mathsf{uni}}^{S}\left\{\mathcal{C}_S \bigcap \mathcal{B}_\varepsilon(x) = \emptyset\right\}$$
$$\geq 1 - \left(1 - \mathbb{P}_{\mathsf{uni}}\{x^{(i)} \in \mathcal{B}_\varepsilon(x)\}\right)^{S}. \tag{19}$$

If $S \to \infty$, we have $\mathbb{P}_{\mathsf{uni}}^{S}\{\mathcal{C}_S \bigcap \mathcal{B}_\varepsilon(x) \neq \emptyset\} \geq 1$, which implies (15).

Then, we discuss $\mathbb{P}_{\mathsf{uni}}^{S}\{\mathcal{C}_S \bigcap \mathcal{B}_\varepsilon(x) \neq \emptyset\}$ for $x \in \partial \mathcal{X}$, where $\partial \mathcal{X}$ defines the boundary of $\mathcal{X}$. Let $x \in \partial \mathcal{X}$ be a boundary point. Again, by Assumption 1, x satisfies the CCQ. By replacing $\bar{x}$ in (16) and (18) by x, we have that there exists a small enough $\bar{\xi} > 0$ such that $g(x + \xi d) < 0$ holds for any $\xi \in (0, \bar{\xi})$ and thus $x + \xi d \in \mathcal{X}^{\mathsf{in}}$. Let $\varepsilon_1 \in (0, \bar{\xi})$ and we can find $x' := x + \xi d \in \mathcal{B}_{\varepsilon_1}(x) \bigcap \mathcal{X}^{\mathsf{in}}$ with a small enough ξ. Besides, the probability that a sample $x^{(i)} \in \mathcal{C}_S$, $i = 1, \ldots, S$ locates in $\mathcal{B}_{\varepsilon_1}(x')$ satisfies that $\mathbb{P}_{\mathsf{uni}}\{x^{(i)} \in \mathcal{B}_{\varepsilon_1}(x')\} > 0$. Thus, we have $\mathbb{P}_{\mathsf{uni}}\{x^{(i)} \in \mathcal{B}_{2\varepsilon_1}(x)\} > 0$. Let $\varepsilon_1 = \varepsilon/2$, and we can obtain (19) for a boundary point of $\mathcal{X}$, which completes the proof. □

With sample set $\mathcal{C}_S = \{x^{(1)}, \ldots, x^{(S)}\}$, a sample average approximate problem of P_α, defined by $\check{P}_\alpha(\mathcal{C}_S)$, is written as:

$$\min_{\mu \in U_S} \sum_{i=1}^{S} J(x^{(i)})\mu(i) \tag{$\check{P}_\alpha(\mathcal{C}_S)$}$$
$$\text{s.t.} \quad \sum_{i=1}^{S} \mu(i)F(x^{(i)}) \geq 1 - \alpha,$$

where $U_S := \{\mu \in \mathbb{R}^S : \sum_{i=1}^{S} \mu(i) = 1, \ \mu(i) \geq 0, \ \forall i = 1, \ldots, S\}$. Denote the feasible region of problem $\check{P}_\alpha(\mathcal{C}_S)$ as

$$\check{\mathcal{F}}_\alpha(\mathcal{C}_S) := \left\{\mu \in U_S : \sum_{i=1}^{S} \mu(i)F(x^{(i)}) \geq 1 - \alpha.\right\}.$$

Then, the optimal objective function value of $\check{P}_\alpha(\mathcal{C}_S)$ is defined by

$$\check{\mathcal{J}}_\alpha(\mathcal{C}_S) := \min\left\{\sum_{i=1}^{S} J(x^{(i)})\mu(i) : \mu \in \check{\mathcal{F}}_\alpha(\mathcal{C}_S)\right\}.$$

The optimal solution set for $\check{P}_\alpha(\mathcal{C}_S)$ is therefore defined by

$$\check{A}_\alpha(\mathcal{C}_S) := \left\{\mu \in \check{\mathcal{F}}_\alpha(\mathcal{C}_S) : \sum_{i=1}^{S} J(x^{(i)})\mu(i) = \check{\mathcal{J}}_\alpha(\mathcal{C}_S)\right\}.$$

A measure $\check{\mu}_\alpha \in \check{A}_\alpha(\mathcal{C}_S)$ is called an optimal measure for $\check{P}_\alpha(\mathcal{C}_S)$.

Theorem 4 *For given sample sets $\mathcal{C}_S$ and $\mathcal{D}_N$, define two functions of $\mu \in U_S$ as*

$$\check{G}_\alpha(\mu, \mathcal{C}_S) := \sum_{i=1}^{S} \mu(i) F(x^{(i)}) = \sum_{i=1}^{S} \mu(i) \int_\Delta \mathbb{I}\{h(x^{(i)}, \delta)\} p(\delta) \mathrm{d}\delta,$$

and

$$\tilde{G}_\alpha(\mu, \mathcal{C}_S, \mathcal{D}_N) := \sum_{i=1}^{S} \mu(i) \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\{h(x^{(i)}, \delta^{(j)})\}.$$

Then, $\tilde{G}_\alpha(\mu, \mathcal{C}_S, \mathcal{D}_N)$ uniformly converges to $\check{G}_\alpha(\mu, \mathcal{C}_S)$ on U_S w.p. 1, i.e.,

$$\sup_{\mu \in U} \left| \tilde{G}_\alpha(\mu, \mathcal{C}_S, \mathcal{D}_N) - \check{G}_\alpha(\mu, \mathcal{C}_S) \right| \to 0, \ w.p. \ 1 \ as \ N \to \infty.$$

Proof (Theorem 4) For any given $x^{(i)}$, $\mathbb{I}\{h(x^{(i)}, \delta)\}$ is a measurable function of δ. According to the strong Law of Large Numbers (LLN) [3], we have

$$\frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\{h(x^{(i)}, \delta^{(j)})\} - \mathbb{E}\{\mathbb{I}\{h(x^{(i)}, \delta)\}\} \to 0, \ \text{w.p. 1 as } N \to \infty,$$

where

$$\mathbb{E}\{\mathbb{I}\{h(x^{(i)}, \delta)\}\} = \int_\Delta \mathbb{I}\{h(x^{(i)}, \delta)\} p(\delta) \mathrm{d}\delta.$$

Thus, for every $\mu \in \mathcal{U}_S$, we have

$$\check{G}_\alpha(\mu, \mathcal{C}_S) - \tilde{G}_\alpha(\mu, \mathcal{C}_S, \mathcal{D}_N) = \sum_{i=1}^{S} \mu(i) \left(\frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\{h(x^{(i)}, \delta^{(j)})\} - \mathbb{E}\{\mathbb{I}\{h(x^{(i)}, \delta)\}\} \right)$$

$$\to \sum_{i=1}^{S} \mu(i) \times 0 = 0. \ \text{w.p. 1 as } N \to \infty.$$

Uniform convergence is ensured since the set U_S is compact. □

Nextly, we show that $\tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ and $\tilde{A}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ converge to $\check{\mathcal{J}}_\alpha(\mathcal{C}_S)$ and $\check{A}_\alpha(\mathcal{C}_S)$, respectively, with probability 1 as $N \to \infty$.

Theorem 5 *Consider Problem P_α with $\alpha > 0$. Assume that there exists a $x^{(i)} \in \mathcal{C}_S$ that satisfies $F(x^{(i)}) > 1 - \alpha$. As $N \to \infty$, $\tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_N) \to \check{\mathcal{J}}_\alpha(\mathcal{C}_S)$ and $\tilde{A}_\alpha(\mathcal{C}_S, \mathcal{D}_N) \to \check{A}_\alpha(\mathcal{C}_S)$ w.p. 1.*

Proof (Theorem 5) The set U_S is a compact set. The objective function $\sum_{i=1}^{S} J(x^{(i)}) \mu(i)$ is a linear function of $\mu \in U_S$. Besides, $F(x^{(i)})$ is a constant value within [0, 1] for a fixed $x^{(i)}$, which makes the constraint function $\check{G}_\alpha(\mu, \mathcal{C}_S)$ a linear function of $\mu \in U_S$. Therefore, $\check{P}_\alpha(\mathcal{C}_S)$ is a linear program. Due to the assumption that there exists $x^{(i)} \in \mathcal{C}_S$ such that $F(x^{(i)}) > 1 - \alpha$, there is $\mu \in U_S$ such that $\check{G}_\alpha(\mu, \mathcal{C}_S) > 1 - \alpha$ and thus $\check{A}_\alpha(\mathcal{C}_S)$ is nonempty. Since $\tilde{G}_\alpha(\mu, \mathcal{C}_S, \mathcal{D}_N)$ converges to $\check{G}_\alpha(\mu, \mathcal{C}_S)$ w.p. 1 by Theorem 4, there exists N_0 large enough such that $\tilde{G}_\alpha(\mu, \mathcal{C}_S, \mathcal{D}_{N_0}) \geq 1 - \alpha$ w.p. 1. Because $\tilde{G}_\alpha(\mu, \mathcal{C}_S, \mathcal{D}_{N_0})$ is a linear function of μ and U_S is compact, the feasible set of $\tilde{P}_\alpha(\mathcal{C}_S, \mathcal{D}_{N_0})$ is compact as well, and hence $\tilde{A}_\alpha(\mathcal{C}_S, \mathcal{D}_{N_0})$ is nonempty w.p. 1 for all $N \geq N_0$.

Let $\{N_k\}_{k=1}^{\infty}$ be a sequence such that $N_k \to \infty$ and $N_k \geq N_0$ holds for every $k = 1, \ldots$. Let $\tilde{\mu}_k \in \tilde{A}_\alpha(\mathcal{C}_S, \mathcal{D}_{N_0})$ such that $\tilde{G}_\alpha(\tilde{\mu}_k, \mathcal{C}_S, \mathcal{D}_{N_k}) \geq 1 - \alpha$, and $\sum_{i=1}^{S} J(x^{(i)}) \tilde{\mu}_k(i) = \tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_{N_k})$. Let $\tilde{\mu}$ be any cluster point of $\{\tilde{\mu}_k\}_{k=1}^{\infty}$. Let $\{\tilde{\mu}_t\}_{t=1}^{\infty}$ be a subsequence converging to $\tilde{\mu}$. By Theorem 4, we have

$$\check{G}_\alpha(\tilde{\mu}, \mathcal{C}_S) = \lim_{t \to \infty} \tilde{G}_\alpha(\tilde{\mu}_t, \mathcal{C}_S, \mathcal{D}_{N_t}), \text{ w.p. 1.}$$

Therefore, $\check{G}_\alpha(\tilde{\mu}, \mathcal{C}_S) \geq 1 - \alpha$ and $\tilde{\mu}$ is feasible for problem $\check{P}_\alpha(\mathcal{C}_S)$ which implies $\sum_{i=1}^{S} J(x^{(i)}) \tilde{\mu}(i) \geq \check{\mathcal{J}}_\alpha(\mathcal{C}_S)$. Note that $\tilde{\mu}_t \to \tilde{\mu}$ w.p. 1, which implies that

$$\lim_{t \to \infty} \tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_{N_t}) = \lim_{t \to \infty} \sum_{i=1}^{S} J(x^{(i)}) \tilde{\mu}_t(i) = \sum_{i=1}^{S} J(x^{(i)}) \tilde{\mu}(i) \geq \check{\mathcal{J}}_\alpha(\mathcal{C}_S), \text{ w.p. 1.}$$

Since this is true for an arbitrary point of $\{\tilde{\mu}_k\}_{k=1}^{\infty}$ in the compact set $\mathcal{U}_S$, we have

$$\lim_{k \to \infty} \tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_{N_k}) = \lim_{k \to \infty} \sum_{i=1}^{S} J(x^{(i)}) \tilde{\mu}_k(i) \geq \check{\mathcal{J}}_\alpha(\mathcal{C}_S), \text{ w.p. 1.} \qquad (20)$$

Besides, we know that there exists a globally optimal solution of $\check{P}_\alpha(\mathcal{C}_S)$, μ^*, such that for any $\varepsilon > 0$ there is $\mu \in U$ such that $0 < \|\mu - \mu^*\| \leq \varepsilon$ and $\check{G}_\alpha(\mu, \mathcal{C}_S) > 1 - \alpha$. Namely, there exists a sequence $\{\tilde{\mu}_t\}_{t=1}^{\infty} \subseteq U$ that converges to an optimal solution μ^* such that $\check{G}_\alpha(\tilde{\mu}_t, \mathcal{C}_S) > 1 - \alpha$ for all $t \in \mathbb{N}$. Notice that $\tilde{G}_\alpha(\tilde{\mu}_t, \mathcal{C}_S, \mathcal{D}_{N_k})$ converges to $\check{G}_\alpha(\tilde{\mu}_t, \mathcal{C}_S)$ w.p. 1. Then, for any fixed t, $\exists K(t)$ such that $\tilde{G}_\alpha(\tilde{\mu}_t, \mathcal{C}_S, \mathcal{D}_{N_k}) \geq 1 - \alpha$ for every $k \geq K(t)$ w.p. 1. We can assume $K(t) < K(t+1)$ for every t and define the sequence $\{\tilde{\mu}_k\}_{k=K(1)}^{\infty}$ by setting $\tilde{\mu}_k = \tilde{\mu}_t$ for all k and t with $K(t) \leq k < K(t+1)$. Then, $\tilde{G}_\alpha(\hat{\mu}_k, \mathcal{C}_S, \mathcal{D}_{N_k}) \geq 1 - \alpha$, which implies $\tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_{N_k}) \leq \sum_{i=1}^{S} J(x^{(i)}) \tilde{\mu}_k(i)$ for all $k \geq K(1)$. Thus, we have that

$$\lim_{k \to \infty} \tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_{N_k}) \leq \sum_{i=1}^{S} J(x^{(i)}) \mu^*(i) = \check{\mathcal{J}}_\alpha(\mathcal{C}_S), \text{ w.p. 1.} \qquad (21)$$

With (20) and (21), we conclude that $\tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_N) \to \check{\mathcal{J}}_\alpha(\mathcal{C}_S)$ w.p. 1 as $N \to \infty$.

The proof of $\tilde{A}_\alpha(\mathcal{C}_S, \mathcal{D}_N) \to \check{A}_\alpha(\mathcal{C}_S)$ can be referred to Theorem 5.3 of [27]. $\qquad\square$

Nextly, we show that $\check{\mathcal{J}}_\alpha(\mathcal{C}_S)$ converges to $\bar{\mathcal{J}}_\alpha$ with probability 1 as S increases.

Theorem 6 *Suppose Assumption* 2 *and* 4 *hold. As* $S \to \infty$, *with probability 1, we have*

$$\liminf_{S \to \infty} \check{\mathcal{J}}_\alpha(\mathcal{C}_S) = \bar{\mathcal{J}}_\alpha. \tag{22}$$

Proof (Theorem 6) The outline of the proof of Theorem 6 is summarized as follows:

A. Prove that the limit of lower bound of $\check{\mathcal{J}}_\alpha(\mathcal{C}_S)$ is larger than $\bar{\mathcal{J}}_\alpha$ by (23);
B. Prove that the limit of upper bound of $\check{\mathcal{J}}_\alpha(\mathcal{C}_S)$ is smaller than $\bar{\mathcal{J}}_\alpha$ by (38);

  B1. Find a sequence $\{\mu_k\}_{k=1}^\infty$ converges weakly to an optimal solution μ^* of P_α;
  B2. Show that $\int_{\mathcal{X}} F(x)\mathrm{d}\mu_k(x)$ and $\int_{\mathcal{X}} J(x)\mathrm{d}\mu_k(x)$ can be approximated by using discrete probability measure on $\mathcal{C}_S$, which refers to (34) and (35);
  B3. Show that optimal discrete probability measure on $\mathcal{C}_S$ for $\check{P}_\alpha(\mathcal{C}_S)$ has a smaller objective value than the discrete probability measure for approximating any μ_k in B2. Then, we obtain (38).

Then, we give the details of the proof.
For any discrete probability measure $\mu^S \in \check{\mathcal{F}}_\alpha(\mathcal{C}_S)$, we have

$$\int_{\mathcal{X}} F(x)\mathrm{d}\mu^S(x) = \sum_{i=1}^{S} \mu^S(\{x^{(i)}\}) F(x^{(i)}) \geq 1 - \alpha.$$

Thus, $\mu^S \in M_\alpha(x)$. Then, it holds that

$$\sum_{i=1}^{S} J(x^{(i)}) \mu^S(\{x^{(i)}\}) = \int_{\mathcal{X}} J(x)\mathrm{d}\mu^S(x) \geq \bar{\mathcal{J}}_\alpha, \quad \forall \mu^S \in \check{\mathcal{F}}_\alpha(\mathcal{C}_S).$$

Furthermore, with probability 1, we have

$$\liminf_{S \to \infty} \check{\mathcal{J}}_\alpha(\mathcal{C}_S) \geq \bar{\mathcal{J}}_\alpha. \tag{23}$$

Assumption 4 implies that there exists a sequence $\{\mu_k\}_{k=1}^\infty \subseteq M(\mathcal{X})$ that converges weakly to an optimal solution μ^* such that

$$\int_{\mathcal{X}} F(x)\mathrm{d}\mu_k(x) > 1 - \alpha \tag{24}$$

for all $k \in \mathbb{N}$. Since $\{\mu_k\}_{k=1}^\infty$ converges weakly to μ^*, we have

$$\lim_{k \to \infty} \int_{\mathcal{X}} J(x)\mathrm{d}\mu_k(x) - \int_{\mathcal{X}} J(x)\mathrm{d}\mu^*(x) = \lim_{k \to \infty} \mathcal{W}(\mu_k, \mu^*) = 0. \tag{25}$$

Notice that $\bar{\mathcal{J}}_\alpha = \int_{\mathcal{X}} J(x) \mathrm{d}\mu^*(x)$ by (3).

For any given $\varepsilon_J > 0$, $\exists K(\varepsilon_J)$, if $k \geq K(\varepsilon_J)$,

$$\int_{\mathcal{X}} J(x) \mathrm{d}\mu_k(x) - \bar{\mathcal{J}}_\alpha \leq \varepsilon_J.$$

Let $\tilde{\mathcal{C}}_{\tilde{S}}^k := \{\tilde{x}_k^{(1)}, \ldots, \tilde{x}_k^{(\tilde{S})}\}$ be a sample set obtained by sampling from $\mathcal{X}$ according to probability measure μ_k. By Law of Large Numbers (p. 457 of [27]), for any $f \in \mathscr{C}(\mathcal{X}, \mathbb{R})$, as $\tilde{S} \to \infty$, with probability 1, we have

$$\frac{1}{\tilde{S}} \sum_{i=1}^{\tilde{S}} f(\tilde{x}_k^{(i)}) \to \mathbb{E}_{x \sim \mu_k}\{f(x)\} = \int_{\mathcal{X}} f(x) \mathrm{d}\mu_k(x). \tag{26}$$

Since $J(\cdot)$ and $F(\cdot)$ are also elements in $\mathscr{C}(\mathcal{X}, \mathbb{R})$, (26) also holds by replacing $f(\cdot)$ by either $J(\cdot)$ or $F(\cdot)$. Namely, for any $\tilde{\varepsilon}_1$, there exists $\tilde{S}_{\mathsf{I}}(\tilde{\varepsilon}_J)$ such that, if $\tilde{S} \geq \tilde{S}_{\mathsf{I}}(\tilde{\varepsilon}_J)$, with probability 1, the followings hold:

$$\left| \frac{1}{\tilde{S}} \sum_{i=1}^{\tilde{S}} F(\tilde{x}_k^{(i)}) - \int_{\mathcal{X}} F(x) \mathrm{d}\mu_k(x) \right| \leq \tilde{\varepsilon}_1, \tag{27}$$

$$\left| \frac{1}{\tilde{S}} \sum_{i=1}^{\tilde{S}} J(\tilde{x}_k^{(i)}) - \int_{\mathcal{X}} J(x) \mathrm{d}\mu_k(x) \right| \leq \tilde{\varepsilon}_1. \tag{28}$$

On the other hand, according to Lemma 2, as $S \to \infty$, for any $\tilde{s} \in \{1, \ldots, \tilde{S}\}$ and $\tilde{\varepsilon}_r > 0$, with probability 1, there exists a sample $x^{(i_{\tilde{s}})} \in \mathcal{C}_S := \{x^{(1)}, \ldots, x^{(S)}\}$ such that

$$x^{(i_{\tilde{s}})} \in \mathcal{B}_{\tilde{\varepsilon}_r}(\tilde{x}_k^{\tilde{s}}). \tag{29}$$

With a little abuse of notation, let $x^{(i_{\tilde{s}})}$ be the closest sample to $\tilde{x}_k^{(\tilde{s})}$, namely, $x^{(i_{\tilde{s}})} \in \arg\min\{\|x^{(i)} - \tilde{x}_k^{(\tilde{s})}\| : x^{(i)} \in \mathcal{C}_S\}$. Define a set $I_{\tilde{S}} := \{i_1, \ldots, i_{\tilde{S}}\}$ as the set of index corresponding to $x^{(i_{\tilde{s}})}$. Without loss of generality, we assume that $x^{(i_{\tilde{s}})} \neq x^{(j_{\tilde{s}})}$ if $i_{\tilde{s}} \neq j_{\tilde{s}}$, $i_{\tilde{s}}, j_{\tilde{s}} \in I_{\tilde{S}}$. The intuitive explanation of the relationship between $\mathcal{C}_S$ and $\tilde{\mathcal{C}}_{\tilde{S}}^k$ is illustrated in Fig. 1.

Define a discrete probability measure $\mu_k^S \in \mathbb{R}^S$ such that

$$\mu_k^S(i) = \frac{1}{\tilde{S}}, \quad \forall i \in I_{\tilde{S}}, \tag{30}$$

$$\mu_k^S(i) = 0, \quad \forall i \notin I_{\tilde{S}}. \tag{31}$$

For any given positive integer $\tilde{S}$ and positive number $\tilde{\varepsilon}_2$, due to the continuity of $J(\cdot)$ and $F(\cdot)$, there exists $\bar{S}_{\mathsf{I}}(\tilde{S}, \tilde{\varepsilon}_2)$ such that, if $S > S_{\mathsf{I}}(\tilde{S}, \tilde{\varepsilon}_2)$, with probability 1, the followings hold:

$$\left| \sum_{i=1}^{S} \mu_k^S(i) F(x^{(i)}) - \frac{1}{\tilde{S}} \sum_{i=1}^{\tilde{S}} F(\tilde{x}_k^{(i)}) \right| \le \tilde{\varepsilon}_2, \tag{32}$$

$$\left| \sum_{i=1}^{S} \mu_k^S(i) J(x^{(i)}) - \frac{1}{\tilde{S}} \sum_{i=1}^{\tilde{S}} J(\tilde{x}_k^{(i)}) \right| \le \tilde{\varepsilon}_2. \tag{33}$$

By combining (27) with (32) and combining (28) with (33), then, for given $\tilde{\varepsilon}_1$, $\tilde{\varepsilon}_2$, there exist $\tilde{S}_|(\tilde{\varepsilon}_1)$ and $S_|(\tilde{S}, \tilde{\varepsilon}_2)$ such that, if $\tilde{S} > \tilde{S}_|(\tilde{\varepsilon}_1)$ and $S > S_|(\tilde{S}, \tilde{\varepsilon}_2)$, with probability 1, the following holds:

$$\left| \sum_{i=1}^{S} \mu_k^S(i) F(x^{(i)}) - \int_{\mathcal{X}} F(x) \mathrm{d}\mu_k(x) \right| \le \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2, \tag{34}$$

$$\left| \sum_{i=1}^{S} \mu_k^S(i) J(x^{(i)}) - \int_{\mathcal{X}} J(x) \mathrm{d}\mu_k(x) \right| \le \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2. \tag{35}$$

According to (24) and (34), we can find $\tilde{S}_|(\tilde{\varepsilon}_1)$ and $S_|(\tilde{S}, \tilde{\varepsilon}_2)$ such that, if $\tilde{S} > \tilde{S}_|(\tilde{\varepsilon}_1)$ and $S > S_|(\tilde{S}, \tilde{\varepsilon}_2)$, with probability 1, the following holds

$$\sum_{i=1}^{S} \mu_k^S(i) F(x^{(i)}) \ge 1 - \alpha. \tag{36}$$

Thus, μ_k^S is a feasible solution of Problem $\tilde{P}_\alpha(\mathcal{C}_S)$ and thus

$$\sum_{i=1}^{S} \mu_k^S(i) J(x^{(i)}) \ge \check{\mathcal{J}}_\alpha(\mathcal{C}_S). \tag{37}$$

Since $\int_{\mathcal{X}} J(x) \mathrm{d}\mu_k(x)$ converges to $\bar{\mathcal{J}}_\alpha$ w.p. 1 as $k \to \infty$, thus, considering (35) and (37), we have

$$\limsup_{S \to \infty} \check{\mathcal{J}}_\alpha(\mathcal{C}_S) \le \bar{\mathcal{J}}_\alpha. \tag{38}$$

With (23) and (38), we have (22).   □

The proof of Theorem 2 can be obtained immediately by using the results of Theorems 5 and 6, which is omitted here.

3.3 Proof of Theorem 3
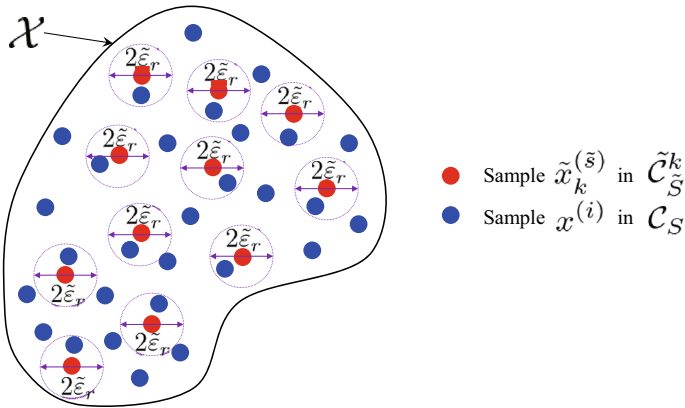
Main results of [37] are summarized as:

Fig. 1 The intuitive explanation of the relationship between $\mathcal{C}_S$ and $\tilde{\mathcal{C}}_{\tilde{S}}^k$

Lemma 3 *Let $\mathcal{X}^+$ be a compact set. Let $p : \mathbb{R}^n \to \mathbb{R}$ be a probability density function on the domain $\mathbb{R}^n$. If there exists a positive number $\rho' > 0$ such that $p \in \{p : p(x) \geq \rho', \forall x \in \mathcal{X}^+\}$, then there exists $p_\theta(x)$ defined by* (11) *such that*

$$\lim_{L \to \infty} \int_{\mathcal{X}^+} (p(x) - p_\theta(x))^2 \, dx = 0,$$

where the positive integer L is the number of Gaussian kernels in (11).

Proof (Theorem 3) For given $\mathcal{C}_S$, $\mathcal{D}_N$ and L, we have problems $\tilde{P}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ and $\hat{P}_\alpha(L, \mathcal{D}_N)$. Let $\mathcal{X}_{p,i}, i = 1, \ldots, S$ be the partitions of $\mathcal{X}$, which satisfy

(a) $x^{(i)} \in \mathcal{X}_{p,i}$;
(b) $\bigcup_{i=1}^{S} \mathcal{X}_{p,i} = \mathcal{X}$;
(c) $\mathcal{X}_{p,i} \bigcap \mathcal{X}_{p,i'} = \emptyset$ with probability 1 if $i \neq i'$.

For any $\mu^S \in U$, we can correspondingly define a Dirac measure on $\mathcal{X}$ as

$$\mu_{\mathsf{d}}^S(x) = \mu^S(x^{(i)}) \quad \text{if} \quad x \in \mathcal{X}_{p,i}.$$

Define a set of index as $I^+ = \{i : \mu^S(x^{(i)}) > 0\}$. Then, we can define a compact set

$$\mathcal{X}^+ = \bigcup_{i \in I^+} \mathcal{X}_{p,i}.$$

According to Lemma 3, there exists a sequence $\{p_\theta(x)\}_L$ such that

$$\lim_{L \to \infty} \int_{\mathcal{X}^+} \left(\mu_{\mathsf{d}}^S(x) - p_\theta(x) \right)^2 dx = 0.$$

Thus, we have

$$\lim_{L \to \infty} \int_{\mathcal{X}} J(x) \mathrm{d} p_\theta(x) = \int_{\mathcal{X}} J(x) \mathrm{d} \mu_{\mathrm{d}}^S(x)$$

and

$$\lim_{L \to \infty} \int_{\mathcal{X}} \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\{h(x, \delta^{(j)}) \le 0\} p_\theta(x) \mathrm{d}x = \int_{\mathcal{X}} \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\{h(x, \delta^{(j)}) \le 0\} \mu_{\mathrm{d}}^S(x) \mathrm{d}x.$$

For any S and N, by applying Lemma 3, we can find a sequence $\{p_\theta^*(x)\}_L$ such that

$$\lim_{L \to \infty} \int_{\mathcal{X}} J(x) \mathrm{d} p_\theta^*(x) = \tilde{\mathcal{J}}_\alpha(\mathcal{C}_S, \mathcal{D}_N) \tag{39}$$

and

$$\lim_{L \to \infty} \int_{\mathcal{X}} \sum_{j=1}^{N} \frac{1}{N} \mathbb{I}\{h(x, \delta^{(j)})\} p_\theta^*(x) \mathrm{d}x = \sum_{i=1}^{S} \mu^S(x^{(i)}) \sum_{j=1}^{N} \frac{1}{N} \mathbb{I}\{h(x^{(i)}, \delta^{(j)})\} \ge 1 - \alpha. \tag{40}$$

There exists the limit of $\hat{P}_\alpha(L, \mathcal{D}_N)$ that converges to $\tilde{P}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ as $L \to \infty$. Theorem 3 can be obtained by using Theorem 2. One point should be clarified here. In Theorem 2, the convergence holds for $S \to \infty$. In Theorem 3, $L \to \infty$ is used instead since we have (39) and (40) for any S increasing to infinite. □

4 Numerical Examples

This section provides the results of two numerical examples to validate our proposed methods. All computations were executed on Windows 10 with 32 GB RAM and an Intel(R) Core(TM) i7-1065G7 CPU running at 1.30 GHz. The algorithm and all computations were implemented in MATLAB R2021b. We check the performance of the following methods:

1. *Dirac-Delta*: solving sample average approximate problem $\tilde{Q}_{\epsilon,\gamma}(\mathcal{D}_N)$ of Q_α;
2. *Sample*: solving sample-based approximate problem $\tilde{P}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$ of P_α;
3. *GMM*: GMM-based approximate problem $\hat{P}_\alpha(L, \mathcal{D}_N)$.

We use the terminology *Dirac-Delta* for the method of solving sample average approximate problem $\tilde{Q}_{\epsilon,\gamma}(\mathcal{D}_N)$ of Q_α since it equivalently gives the measure constrained to be a Dirac-delta, namely, the measure is concentrated on one fixed solution.

4.1 One-Dimension Example

In the first numerical, we use an extremely simple example to demonstrate the concepts of Theorems 1, 2, and 3. The compact set $\mathcal{X}$ is defined by $\mathcal{X} = \{x \in \mathbb{R} : x \in [-1, 1]\}$. Moreover, the cost function $J(x)$ is
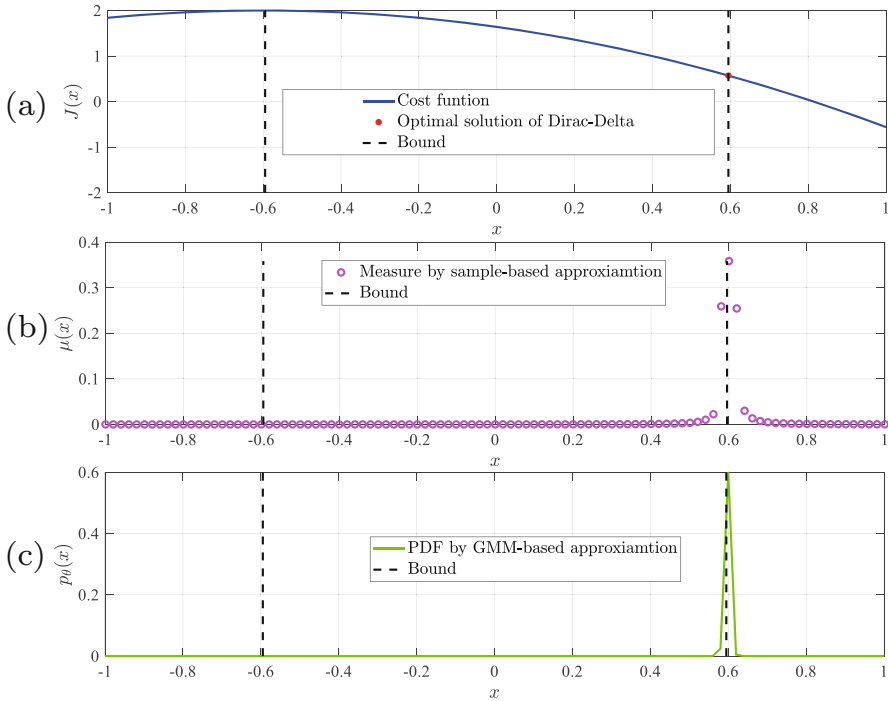
Fig. 2 Results of the numerical example 1: **a** profile of $J(x)$ and optimal solution obtained by *Dirac-Delta*; **b** optimal measure by *Sample*; **c** optimal probability density function obtained by *GMM*

$$J(x) = -(x + 0.6)^2 + 2. \tag{41}$$

The constraint function $h(x, \delta)$ is

$$h(x, \delta) = x^2 + \delta - 2 \tag{42}$$

where $\delta \sim \mathcal{N}(m_\delta, \Sigma_\delta)$, $m_\delta = 0$, and $\Sigma_\delta = 1$. The probability level α is 0.05. The optimal solution from method *Dirac-Delta* is $x_\alpha^* = 0.595$ and the optimal objective value is 0.572, which is plotted in Fig. 2a. In *Dirac-Delta*, we set $\epsilon = \alpha$, $N = 2000$, and $\gamma = 0.01$. Besides, Fig. 2b, c shows the discrete measure obtained by *Sample* and the probability density function obtained by *GMM*, respectively. For *Sample*, we choose samples $-1, -0.98, -0.96, \ldots, 0.96, 0.98, 1$ from $\mathcal{X}$ ($S = 201$) and 2000 randomly extracted samples from Δ ($N = 2000$). For *GMM*, we extracted 2000 samples from Δ randomly. Besides, we choose $L = 6$. The solutions of *Sample* and *GMM* satisfy the chance constraints. For the objective function, *Sample* achieves 0.5601 and *GMM* achieves 0.5615, which are better than the optimal objective value achieved by *Dirac-Delta*.

A more comprehensive analysis of CPU time and sample numbers is summarized in Table 1. The CPU time increases as the sample size increases for each method. Unsurprisingly, *Sample* has a very fast computation time since it only needs to solve a

Table 1 Statistics of CPU time for one-dimension example

Dirac-Delta								
Size N	1000			2000			5000		10000

Dirac-Delta			
Size N	1000	2000	5000	10000
Avg. (s)	0.0105	0.0136	0.0154	0.0169
Max. (s)	0.0281	0.0337	0.002	0.0583

Sample								
Size N	2000			5000			10000		
Size S	50	100	200	50	100	200	50	100	200
Avg. (s)	0.0113	0.0123	0.0140	0.0119	0.0145	0.0185	0.0143	0.0162	0.0229
Max. (s)	0.0263	0.0349	0.0415	0.0238	0.0341	0.0445	0.0343	0.0409	0.0530

GMM								
Size N	2000			5000			10000		
Size L	2	4	6	2	4	6	2	4	6
Avg. (s)	0.0325	0.0431	0.0991	0.0472	0.0783	0.1671	0.0773	0.1429	0.2134
Max. (s)	0.0721	0.0784	0.1562	0.0793	0.1318	0.2011	0.1267	0.1892	0.2690

linear program. In this example, since it is one dimension, the required sample number for obtaining good samples in *Sample* or approximating probability integration in *GMM* is few. It can achieve acceptable accuracy with only 50 samples. However, if the dimension of x increases, the "Curse of Dimensionality" will emerge. We will show it in the second example.

4.2 Quadrotor System Control

The second example considers a quadrotor system control problem in turbulent conditions. The control problem is expressed as follows:

$$
\begin{aligned}
\min_{\mu \in M(\mathcal{U}^T)} \quad & \mathbb{E}\{\ell^x(x) + \ell^u(u)\} \\
\text{s.t.} \quad & x_{t+1} = Ax_t + B(m)u_t + d(x_t, \varphi) + \omega_t, \ u \sim M(\mathcal{U}^T), \\
& t = 0, 1, \ldots, T-1, \\
& \Pr\{(x_t \notin \mathcal{O}, \forall t = 1, \ldots, T-1), \ (x_T \in \mathcal{X}_{\text{goal}})\} \geq 1 - \alpha,
\end{aligned}
\tag{P_{QSC}}
$$

where A, $B(m)$, $d(x_t, \varphi)$ are written by

$$
A = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ B(m) = \frac{1}{m}\begin{bmatrix} \frac{\Delta t^2}{2} & 0 \\ \Delta t & 0 \\ 0 & \frac{\Delta t^2}{2} \\ 0 & \Delta \end{bmatrix}, \ d(x_t, \varphi) = -\varphi \begin{bmatrix} \frac{\Delta t^2 |v_x| v_x}{2} \\ \Delta t |v_x| v_x \\ \frac{\Delta t^2 |v_y| v_y}{2} \\ \Delta t |v_y| v_y \end{bmatrix},
$$

and Δt is the sampling time, the state of the system is denoted as $x_t = [p_{x,t}, v_{x,t}, p_{y,t}, v_{y,t}] \in \mathbb{R}^4$, the control input of the system is $u_t = \{u_{x,t}, u_{y,t}\}$ within $\mathcal{U} := \{u_t \in \mathbb{R}^2 : -10 \le u_{x,t} \le 10, -10 \le u_{y_t} \le 10\}$, and the state and control trajectories are denoted as $x = (x_t)_{t=1}^{T}$ and $u = (u_t)_{t=1}^{T-1}$. The system starts from an initial point $x_0 = [-0.5, 0, -0.5, 0]$. The system is expected to reach the destination set $\mathcal{X}_{\text{goal}} = \{x \in \mathbb{R}^4 | \|(p_x - 10, p_y - 10)\| \le 2\}$ at time $T = 10$ while avoiding two polytopic obstacles $\mathcal{O}$ shown in Fig. 3. $\mathcal{O}$ is defined by the following constraints:

$$p_{x,t} \le 6.35, \quad p_{y,t} \ge 3.35, \quad p_{x,t} - 0.2 - p_{y,t} \ge 0,$$
$$p_{x,t} \ge 3.35, \quad p_{y,t} \le 6.35, \quad p_{x,t} + 0.2 - p_{y,t} \le 0.$$

The dynamics are parametrized by uncertain parameter vector $\delta_t = [m, \varphi]^\top$, where $m > 0$ represents the system's mass and $\varphi > 0$ is an uncertain drag coefficient. The parameter vector δ of the system is uncorrelated random variables such that $(m - 0.75)/0.5 \sim \text{Beta}(2, 2)$ and $(\varphi - 0.4)/0.2 \sim \text{Beta}(2, 5)$, where $\text{Beta}(a, b)$ denotes a Beta distribution with shape parameters (a, b). $\omega_t \in \mathbb{R}^4$ is the uncertain disturbance at time step t, which obeys multivariate normal distribution with zero means and covariance matrix

$$\Sigma = \begin{bmatrix} 0.01 & 0 & 0 & 0 \\ 0 & 0.75 & 0 & 0 \\ 0 & 0 & 0.01 & 0 \\ 0 & 0 & 0 & 0.75 \end{bmatrix}.$$

For the cost function, we adopt

$$\ell^x(x) = \frac{1}{T} \sum_{t=0}^{T-1} \left((p_{x,t+1} - p_{x,t})^2 + (p_{y,t+1} - p_{y,t})^2 \right),$$

$$\ell^u(u) = \frac{0.1}{T} \sum_{t=0}^{T-1} \left(u_{1,t}^2 + u_{2,t}^2 \right).$$

Results are shown in Fig. 3 for different methods by setting α as 15%. Figure 3 shows 5000 Monte Carlo (MC) simulations of the quadrotor system using the open-loop policy computed using *Dirac-Delta* ($\epsilon = \alpha, \gamma = 0.01, N = 2000$), *Sample* ($S = 5.1 \times 10^6, N = 2000$), and *GMM* ($L = 6, N = 2000$). When using *Dirac-Delta*, the algorithm gives a deterministic control policy that satisfies the desired success probability $1 - \alpha$. When using *Sample*, or *GMM*, the algorithm gives a stochastic control policy that satisfies the desired success probability $1 - \alpha$. The control inputs that generate trajectories passing through the riskier middle corridor between the obstacles are selected randomly for the stochastic control policies. The costs by using *Sample* and *GMM* are reduced by 8.2 and 7.9% compared to using *Dirac-Delta*. This shows that our approach can compute a better policy that solves the problem than a deterministic policy.
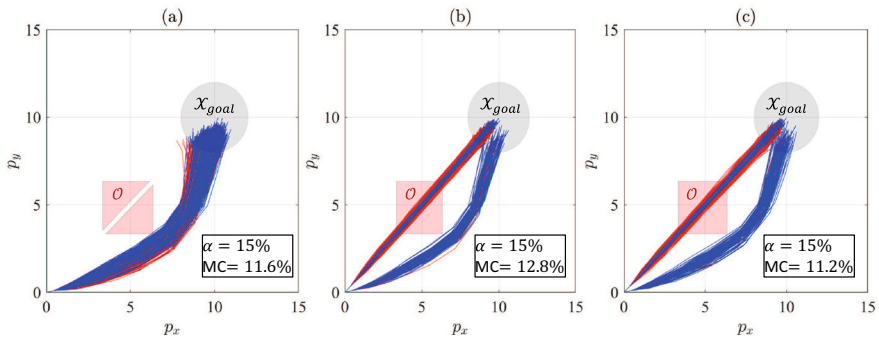
Fig. 3 Solutions from different methods for the tolerable failure probability threshold $\alpha = 15\%$. Blue trajectories from Monte Carlo (MC) simulations denote feasible trajectories that reach the goal set $\mathcal{X}_{\text{goal}}$ and avoid obstacles $\mathcal{O}$. Red trajectories violate constraints: **a** *Dirac-Delta* (MC = 11.6% represents that the violation probability is 11.6% in the MC simulations); **b** *Sample* (MC = 12.8% represents that the violation probability is 12.8% in the MC simulations); **c** *GMM* (MC = 11.2% represents that the violation probability is 11.2% in the MC simulations)
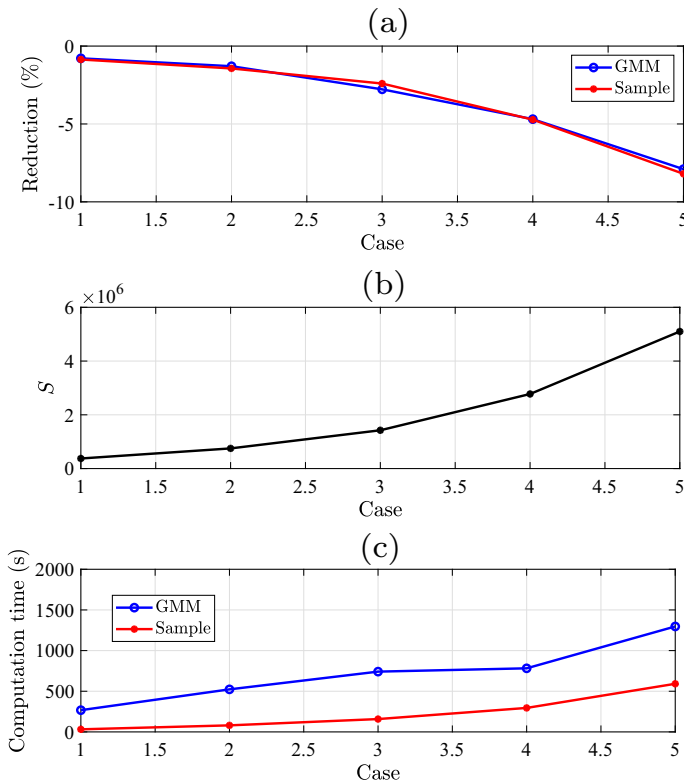
Fig. 4 The statistics of the control performance: **a** reduction of cost; **b** required samples; **c** computation time

A more comprehensive comparison between the GMM-based and sample-based approximations is plotted in Fig. 4. Five cases are considered with different sample numbers for extracting the control input. Figure 4a shows that the two algorithms similarly reduce the optimal objective function value. Figure 4b shows each case's used sample number S of decision variables. By comparing Fig. 4a, b, we can see that enough samples are required to ensure the performance of the approximations. As shown in Fig. 4c, the computation time increases dramatically as the sample number increases. In this comparison, for *GMM*, we choose $L = 6$, and the probability integration is approximated by using the same samples of *Sample*. The computation time of *GMM* is even longer than *Sample*. One way to decrease the computation time of *GMM* is to develop fast algorithms for probability integration. We leave this for future work. In this example, the dimension of the decision variable is 20. If the dimension increases, the required sample number will increase, and the computation time will consequently increase for *Sample* and *GMM*. We leave the issue of the "Curse of Dimensionality" for future work.

5 Conclusions

In conclusion, the chance-constrained linear program in probability measure space has been addressed using sample approximation or function approximation. We establish optimization problems in finite vector space as approximate problems of chance-constrained linear programs in probability measure space. By solving the approximate problems, we can obtain the approximate solution of the chance-constrained linear program in probability measure space. Numerical examples have been implemented to validate the performance of the proposed method. Future work will be focused on the following points:

– To implement sample approximation method $\tilde{P}_\alpha(\mathcal{C}_S, \mathcal{D}_N)$, samples of decision variable are required. As the dimension of the decision variable increases, the required sample number for a good approximation will also increase, bringing the issue of the "Curse of dimensionality." To overcome the issue of the "Curse of Dimensionality," it is important to develop efficient sampling algorithms to get "good but small samples" to ensure good approximation performance and mitigate the computation burden;
– For Gaussian mixture model-based approximation method $\hat{P}_\alpha(L, \mathcal{D}_N)$, the remaining issue is how to approximate the probability integration by fast algorithms when the problem is with complex cost function and constrained functions in high dimension space.

References

1. Berthold, H., Heitsch, H., Henrion, R., Schwienteck, J.: On the algorithmic solution of optimization problems subject to probabilistic/robust (probust) constraints. Math. Methods Oper. Res. **96**, 1–37 (2022)
2. Blackmore, L., Ono, M., Williams, B.C.: Chance-constrained optimal path planning with obstacles. IEEE Trans. Rob. **27**(6), 1080–1094 (2011)
3. Bertsekas, D.P., Tsitsiklis, J.N.: Introduction to Probability. Athena Scientific, Belmont (2002)
4. Billingsley, P.: Probability and Measure. John Wiley & Sons, New York (1995)
5. Calariore, G., Campi, M.C.: The scenario approach to robust control design. IEEE Trans. Autom. Control **51**(5), 742–753 (2006)
6. Campi, M.C., Garatti, S.: The exact feasibility of randomized solutions of uncertain convex programs. SIAM J. Optim. **19**(3), 1211–1230 (2008)
7. Campi, M.C., Garatti, S.: A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. J. Optim. Theory Appl. **148**(2), 257–280 (2011)
8. Campi, M.C., Garatti, S.: A general scenario theory for nonconvex optimization and decision making. IEEE Trans. Autom. Control **63**(12), 4067–4078 (2015)
9. Campi, M.C., Garatti, S., Ramponi, F.A.: Introduction to the Scenario Approach. MOS-SIAM Series on Optimization, Philadelphia (2019)
10. Castillo-Lopez, M., Ludivig, P., Sajadi-Alamdari, S.A.: A real-time approach for chance-constrained motion planning with dynamic obstacles. IEEE Robotics Autom. Lett. **5**(2), 3620–3625 (2020)
11. Chen, P., Ghattas, O.: Taylor approximation for chance constrained optimization problems governed by partial differential equations with high-dimensional random parameters. SIAM/ASA J. Uncertain. Quantif. **9**(4), 1381–1410 (2021)
12. Farshbaf-Shaker, M.H., Henrion, R., Homber, D.: Properties of chance constraints in infinite dimensions with an application to PDE constrained optimization. Set Valued Var. Anal. **26**, 821–841 (2018)
13. Geletu, A., Hoffmann, A., Kloppel, M., Li, P.: An inner-outer approximation approach to chance constrained optimization. SIAM J. Optim. **27**(3), 1834–1857 (2017)
14. Geletu, A., Hoffmann, A., Schmidt, P., Li, P.: Chance constrained optimization of elliptic PDE systems with a smoothing convex approximation. ESAIM Control Optim. Calc. Var. **26**(70) (2020)
15. Grandon, T.G., Henrion, R., Perez-Aros, P.: Dynamic probabilistic constraints under continuous random distributions. Math. Program. **196**, 1065–1096 (2022)
16. Gryazina, E., Polyak, B.: Random sampling: Billiard walk algorithm. Eur. J. Oper. Res. **238**, 497–504 (2014)
17. Hewing, L., Kabzan, J., Zeilinger, M.N.: Cautious model predictive control using Gaussian process regression. IEEE Trans. Control Syst. Technol. **28**(6), 2736–2743 (2020)
18. Kibzun, A.I., Kan, Y.S.: Stochastic Programming Problems. Wiley, West Sussex (1996)
19. Lew, T., Sharma, A., Harrison, J., Bylard, A., Pavone, M.: Safe active dynamics learning and control: asequential exploration-exploitation framework. IEEE Trans. Robotics **38**(5), 2888–2907 (2022)
20. Luedtke, J., Ahmed, S.: A sample approximation approach for optimization with probabilistic constraints. SIAM J. Optim. **19**(2), 674–699 (2008)
21. Luedtke, J., Ahmed, S., Nemhauser, G.L.: An integer programming approach for linear programs with probabilistic constraints. Math. Program. **122**, 247–272 (2010)
22. Molchanov, I.: Theory of Random Sets. Springer, London (2005)
23. Nemirovski, A., Shapiro, A.: Convex approximations of chance constrained programs. SIAM J. Optim. **17**, 969–996 (2007)
24. Ono, M., Pavone, M., Kuwata, Y., Balaram, J.: Chance-constrained dynamic programming with application to risk-aware robotic space exploration. Auton. Robots **39**(4), 555–571 (2015)
25. Pagnoncelli, B.K., Ahmed, S., Shapiro, A.: Computational study of a chance constrained portfolio selection problem. J. Optim. Theory Appl. **142**(2), 399–416 (2009)

26. Pena-Ordieres, A., Luedtke, J., Wachter, A.: Solving chance-constrained problems via a smooth sample-based nonlinear approximation. SIAM J. Optim. **30**(3), 2221–2250 (2020)
27. Shapiro, A., Dentcheva, D., Ruszczynski, A.: Lectures on Stochastic Programming: Modeling and Theory, 2nd edn. SIAM, Philadelphia (2014)
28. Shen, X., Ouyang, T., Zhang, Y., Zhang, X.: Computing probabilistic bounds on state trajectories for uncertain systems. IEEE Trans. Emerg. Top. Comput. Intell. **7**(1), 285–290 (2023)
29. Shen, X., Ouyang, T., Yang, N., Zhuang, J.: Sample-based neural approximation approach for probabilistic constrained programs. IEEE Trans. Neural Netw. Learn. Syst. **34**(2), 1058–1065 (2023)
30. Smith, R.L.: Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. Oper. Res. **32**(6), 1296–1308 (1984)
31. Sun, Y., Aw, G., Loxton, R., Teo, K.L.: Chance-constrained optimization for pension fund portfolios in the presense of default risk. Eur. J. Oper. Res. **256**(1), 205–214 (2017)
32. Thorpe, A.J., Lew, T., Oishi, M.M.K., Pavone, M.: Data-driven chance constrained control using kernel distribution embeddings. Proc. Mach. Learn. Res. **144**, 1–13 (2022)
33. van Ackooij, W., Henrion, R.: Gradient formulate for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions. SIAM J. Optim. **24**, 1864–1889 (2014)
34. van Ackooij, W., Perez-Aros, P.: Generalized differentiation of probability functions acting on an infinite system of constraints. SIAM J. Optim. **29**(3), 2179–2210 (2019)
35. van Ackooij, W., Henrion, R., Perez-Aros, P.: Generalized gradients for probabilistic/robust (probust) constraints. Optimization **69**(7–8), 1451–1479 (2020)
36. Wu, C., Teo, K.L., Wu, S.: Min-max optimal control of linear systems with uncertainty and terminal state constraints. Automatica **49**(6), 1809–1815 (2013)
37. Zeevi, A.J., Meir, R.: Density estimation through convex combinations of densities: approximation and estimation bounds. SIAM J. Optim. **10**(1), 99–109 (1997)