Asymptotic Analysis of Sample Average Approximation for Stochastic Optimization Problems with Joint Chance Constraints via CVaR/DC Approximations

Hailin Sun[1]
Department of Mathematics, Harbin Institute of Technology, Harbin, 150001, China
(mathhlsun@gmail.com)

Huifu Xu
School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK
(h.xu@soton.ac.uk)

Yong Wang
Department of Mathematics, Harbin Institute of Technology, Harbin, 150001, China
(mathwy@hit.edu.cn)

February 23, 2012

Abstract

Conditional Value at Risk (CVaR) has been recently used to approximate a chance constraint. In this paper, we study the convergence of stationary points when sample average approximation (SAA) method is applied to a CVaR approximated joint chance constrained stochastic minimization problem. Specifically, we prove, under some moderate conditions, that optimal solutions and stationary points obtained from solving sample average approximated problems converge with probability one (w.p.1) to their true counterparts. Moreover, by exploiting the recent results on large deviation of random functions [28] and sensitivity results for generalized equations [31], we derive exponential rate of convergence of stationary points and give an estimate of sample size. The discussion is extended to the case when CVaR approximation is replaced by a DC-approximation [14]. Some preliminary numerical test results are reported.

Key words. Joint chance constraints, CVaR, DC-approximation, almost H-calmness, stationary point, exponential convergence

1 Introduction

Consider the following joint chance constrained minimization problem:

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & P(x) := \text{Prob}\{c_1(x,\xi) \le 0, \ldots, c_m(x,\xi) \le 0\} \ge 1-\alpha, \\ & x \in \mathcal{X}, \end{array} \tag{1.1}$$

where $\mathcal{X}$ is a convex compact subset of $\mathbb{R}^n$, ξ is a random vector in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with support $\Xi \subset \mathbb{R}^k$ and α is a positive number between 0 and 1; $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously

[1]The work of this author is carried out while he is visiting the second author in the School of Mathematics, University of Southampton sponsored by China Scholarship Council.

differentiable function, $c_i : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$, $i = 1, \dots, m$, is continuously differentiable w.r.t. x for every fixed ξ. Problem (1.1) has wide applications in communications and networks, product design, system control, statistics and finance, see [6] for details. For the convenience of notation, let

$$c(x,\xi) := \max\{c_1(x,\xi), \dots, c_m(x,\xi)\}. \tag{1.2}$$

Then we can rewrite (1.1) as

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & \text{Prob}\{c(x,\xi) \le 0\} \ge 1-\alpha, \\ & x \in \mathcal{X}. \end{array} \tag{1.3}$$

It is well-known that problem (1.3) is difficult to solve as the chance constraint function $P(x)$ is generally non-convex, has no closed form and is difficult to evaluate. Various approaches have been proposed in the literature to address these difficulties. For example, a number of convex conservative approximation schemes for $P(x)$ have been proposed, including the quadratic approximation [4], the CVaR approximation [23] and the Bernstein approximation [17]. These approximation schemes typically find a feasible suboptimal solution to problem (1.3). CVaR approximation is one of the most widely used approximation schemes since it is numerically tractable and enjoys nice features such as convexity and monotonicity.

Our focus here is on CVaR approximation. Recall that Value-at-Risk (VaR) of a random function $c(x,\xi)$ is defined as

$$\text{VaR}_{1-\alpha}(c(x,\xi)) := \min_{\eta\in\mathbb{R}}\{\eta : \text{Prob}\{c(x,\xi) \le \eta\} \ge 1-\alpha\}$$

and CVaR is defined as the conditional expected value of $c(x,\xi)$ exceeding VaR, that is,

$$\text{CVaR}_{1-\alpha}(c(x,\xi)) := \frac{1}{\alpha}\int_{c(x,\xi)\ge \text{VaR}_{1-\alpha}(c(x,\xi))} c(x,\xi)P(d\xi).$$

The latter can be reformulated as

$$\min_{\eta\in R}(\eta + \frac{1}{\alpha}\mathbb{E}[(c(x,\xi)-\eta)_+]),$$

see [23]. It is easy to verify that $\text{CVaR}_{1-\alpha}(c(x,\xi)) \to \text{VaR}_{1-\alpha}(c(x,\xi))$ as $\alpha \to 0$. In the case when $c(x,\xi)$ is convex in x for almost every ξ, $\text{CVaR}_{1-\alpha}(c(x,\xi))$ is a convex function, see [23, Theorem 2]. Observe that the chance constraint in problem (1.3) can be written as $\text{VaR}_{1-\alpha}(c(x,\xi)) \le 0$ while CVaR is often regarded as an approximation of VaR. Therefore we may consider the following approximation scheme for (1.1) by replacing the chance constraint with a CVaR constraint:

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & \text{CVaR}_{1-\alpha}(c(x,\xi)) \le 0, \\ & x \in \mathcal{X}. \end{array} \tag{1.4}$$

An obvious advantage of (1.4) is that the constraint function is convex as long as $c(x,\xi)$ is convex. This kind of convex approximation scheme is first proposed by Nemirovski and Shapiro [17] and has now been widely applied in stochastic programming. Using the reformulation of CVaR, i.e., [23], we may present problem (1.4) as

$$\begin{array}{ll} \min\limits_{x,\eta} & f(x) \\ \text{s.t.} & \eta + \frac{1}{\alpha}\mathbb{E}[(c(x,\xi)-\eta)_+] \le 0, \\ & x \in \mathcal{X}, \eta \in \mathbb{R}, \end{array} \tag{1.5}$$

see for instance [27, Section 1.4].

Problem (1.5) is a nonlinear stochastic optimization problem and our focus in this paper will be on the numerical method for solving the problem. The main challenge here is to handle the expected value $\mathbb{E}[(c(x,\xi)-\eta)_+]$. If we are able to obtain a closed form of the function, then problem (1.5) becomes an ordinary nonlinear programming problem (NLP). However, in practice, it is often difficult to do so not only because there is a max operation on the integrand but also it requires complete information on the distribution of ξ or multidimensional integration. A well known method to tackle this problem is sample average approximation (SAA) method which is also know wider various names as Monte Carlo method, sample path optimization (SPO) method [24] and stochastic counterpart, see [25] for a comprehensive review. The basic idea of SAA can be described as follows. Let $\xi^1,\cdots,\xi^N$ be an independent and identically distributed (i.i.d.) sampling of ξ. We consider the following sample average approximation problem for problem (1.5):

$$\begin{aligned} \min_{x,\eta} \quad & f(x) \\ \text{s.t.} \quad & \eta + \frac{1}{\alpha N}\sum_{j=1}^{N}(c(x,\xi^j)-\eta)_+ \le 0, \\ & x \in \mathcal{X}, \eta \in \mathbb{R}. \end{aligned} \tag{1.6}$$

We refer to (1.5) as the *true* problem and (1.6) as its *sample average approximation* (SAA) problem. For a fixed sample, (1.6) is a deterministic NLP and therefore any appropriate NLP code can be applied to solve the problem. Our interest here is on whether an optimal solution or a stationary point obtained from solving the SAA problem converges to their true counterpart when sample size increases, and if so at what convergence rate. The latter is practically interesting as one would like to know how large the sample size should be in order to obtain an approximate solution with a specified precision, and it is indeed technically challenging in that the constraint functions in (1.6) is often nonsmooth particularly in the joint chance constrained case. Meng et al [16] show the almost sure convergence of stationary points and exponential convergence of optimal values for the individual and mixed CVaR optimization problems in convex case. In this paper, we study the almost sure convergence and exponential convergence of stationary points for problem (1.4) in both convex and nonconvex case.

In order to carry out the convergence analysis particularly in relation to stationary points, we need to derive first order optimality conditions. We do so for both convex and nonconvex cases and resort to Hiriart-Urruty's earlier results on Karush-Kuhn-Tucker conditions for nonsmooth constrained optimization [13].

In the literature of continuous stochastic optimization, asymptotic analysis of a statistical estimator of optimal value and optimal solution is based on uniform exponential convergence of a random function which is Hölder continuous, see for instance [19, 26]. Xu [31] extends the results to a class of so-called H-calm functions which allow some extent of discontinuity and this is used to derive exponential convergence of stationary points in nonsmooth stochastic optimization. In a more recent development, the exponential convergence results are further extended [28] to a class of almost H-calm functions. Here, we present a detailed discussion about the advantage of almost H-calmness (see Remark 3.1 and Example 3.1) and strengthen [28, Theorem 3.1] by weakening a boundedness condition imposed on the random function (see Theorem 5.1). We then apply the strengthened uniform convergence result to establish exponential convergence of

Clarke stationary points (Theorem 3.2). We also do so for the case when the chance constraint is approximated by a difference of two convex functions (DC-approximation). This strengthens the existing results by Hong et al [14].

The rest of the paper is organized as follows. In section 2, we discuss optimality conditions of true problem and the SAA problem for CVaR approximated problems in both convex and nonconvex cases. In section 3, we investigate almost sure convergence of stationary points of the SAA problem as sample size increases and extend the discussion to the exponential rate of convergence. In section 4, we present similar convergence analysis to DC-approximated problem. In section 5, we report some numerical results.

2 Optimality conditions

In this section, we discuss optimality conditions and first order necessary conditions of the true problems (1.5) and its sample average approximated counterpart. This is to pave the way for the asymptotic convergence analysis of stationary points of problem (1.6) as sample size increases.

2.1 Notation and preliminaries

Throughout this paper, we use the following notation: $x^T y$ denotes the scalar product of two vectors x and y, $x \perp y$ denotes $x^T y = 0$, $\|\cdot\|$ denotes the Euclidean norm of a vector. $d(x, D) := \inf_{x' \in D} \|x - x'\|$ denotes the distance from point x to set D. For two sets D_1 and D_2, $\mathbb{D}(D_1, D_2) := \sup_{x \in D_1} d(x, D_2)$ denotes the deviation of set D_1 from set D_2. For a real valued function $h(x)$, we use $\nabla h(x)$ to denote the gradient of h at x. If $h(x)$ is vector valued, then the same notation refers to the classical Jacobian of h at x. For $x \in X$, $B(x; \rho)$ denotes a closed ρ-neighborhood of x relative to X and ρ depends on x.

Let $v : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous function. Recall that *Clarke generalized derivative* of v at point x in direction d is defined as

$$v^o(x, d) := \limsup_{y \to x, t \downarrow 0} \frac{v(y + td) - v(y)}{t}.$$

v is said to be Clarke *regular* at x if the usual one sided directional derivative, denoted by $v'(x, d)$, exists for every $d \in \mathbb{R}^n$ and $v^o(x, d) = v'(x, d)$. The *Clarke generalized gradient* (also known as Clarke subdifferential) is defined as

$$\partial v(x) := \{\zeta : \zeta^T d \leq v^o(x, d)\},$$

For a vector valued function v, it is said to be strictly differentiable at x, if v admits a strict derivative at x, an element of $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ denoted $Dv(x)$, proved for each d, the following holds:

$$\lim_{y \to x, t \downarrow 0} \frac{v(y + td) - v(y)}{t} = d^T Dv(y),$$

and provided the convergence is uniform for d in compact sets, where $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ denote a space of continuous linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Note that if v is continuously differentiable, it is strictly differentiable, see [8, Chapter 2]. Note that in this paper, we need to consider Clarke subdifferential of a random function $v(x,\xi)$ with respect to x, that is, $\partial_x v(x,\xi)$. In such a case, it is meant the Clarke subdifferential of function $v(\cdot,\xi)$ at point x for fixed ξ.

For the simplicity of notation, let

$$p(x) := \max(0, x),$$

$$g(x,\eta,\xi) := p(c(x,\xi)-\eta),\ G(x,\eta) := \mathbb{E}[g(x,\eta,\xi)],$$

$$G_N(x,\eta) := \frac{1}{N}\sum_{j=1}^{N} g(x,\eta,\xi^j),$$

$$H(x,\eta) := \eta + \frac{1}{\alpha}G(x,\eta)$$

and

$$H_N(x,\eta) := \eta + \frac{1}{\alpha}G_N(x,\eta).$$

Assumption 2.1 *There exist a point $x_0 \in \mathcal{X}$ such that $\mathbb{E}[(c(x_0,\xi))_+] < \infty$. Moreover,*

$$\mathbb{E}[\|\nabla_x c_i(x,\xi)\|] < \infty, \text{ for } i = 1,\ldots,m, x \in \mathcal{X}.$$

Proposition 2.1 *Suppose that Assumption 2.1 holds. Then*

(i) *$H(x,\eta)$ is well defined for all $x \in \mathcal{X}$ and $\eta \in \mathbb{R}$, locally Lipschitz continuous w.r.t. x, globally Lipschitz continuous w.r.t. η and*

$$\partial_{(x,\eta)} H(x,\eta) \subset \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \mathbb{E}\left[\partial_{(x,\eta)} g(x,\eta,\xi)\right], \tag{2.7}$$

where

$$\partial_{(x,\eta)} g(x,\eta,\xi) \subset \begin{pmatrix} \partial_x c(x,\xi) \\ -1 \end{pmatrix} \partial p(c(x,\xi)-\eta), \tag{2.8}$$

and $\partial_x c(x,\xi) = \mathrm{conv}\{\nabla_x c_i(x,\xi), i \in i(x)\}$, $i(x) := \{i : c(x,\xi) = c_i(x,\xi)\}$,

$$\partial p(\eta - c(x,\xi)) = \begin{cases} 1, & \text{if } c(x,\xi)-\eta > 0, \\ [0,1], & \text{if } c(x,\xi)-\eta = 0, \\ 0, & \text{if } c(x,\xi)-\eta < 0, \end{cases}$$

and the expected value of the Clarke subdifferential of the random function is in the sense of Aumann's integral [2], here and later on, "conv" denotes convex hull of a set;

(ii) *if $c(x,\xi)$ is convex or strictly differentiable w.r.t. x for all $\xi \in \Xi$, then $H(x,\eta)$ is Clarke regular w.r.t. x and η and the equality in (2.7) and (2.8) holds.*

Proof. Part (i). Verification of the well definedness and Lipschitzness is elementary given the fact that $p(c(x,\xi)-\eta)$ is a composition of the max function $p(\cdot)$ and $c(x,\xi)-\eta$. In what follows, we show inclusions (2.7) and (2.8). By the definition of Clarke subdifferential,

$$\partial_{(x,\eta)}(c(x,\xi)-\eta)) = \begin{pmatrix} \partial_x c(x,\xi) \\ -1 \end{pmatrix}.$$

Since $c_i(x,\xi)$ is continuously differentiable, by [8, Proposition 2.3.12],

$$\partial_x c(x,\xi) = \mathrm{conv}\{\nabla_x c_i(x,\xi), i \in i(x)\},$$

where $i(x) := \{i : c(x,\xi) = c_i(x,\xi)\}$. Through the chain rule ([8, Theorem 2.3.9]), we obtain

$$\partial_{(x,\eta)} g(x,\eta,\xi) \subset \begin{pmatrix} \partial_x c(x,\xi) \\ -1 \end{pmatrix} \partial p(c(x,\xi)-\eta).$$

Furthermore, it is easy to verify that the term at the right-hand side of the formula above is bounded by $\max_{i=1}^m\{\|\nabla_x c_i(x,\xi)\|\}+1$ which is integrably bounded under Assumption 2.1. Following a discussion by Artstein and Vitale [1], both $\mathbb{E}[\partial_x c(x,\xi)]$ and $\mathbb{E}\left[\partial p(c(x,\xi)-\eta)\right]$ are well defined. Further, by [31, Theorem 2.1],

$$\partial_{(x,\eta)} G(x,\eta) \subset \mathbb{E}\left[\begin{pmatrix} \partial_x c(x,\xi) \\ -1 \end{pmatrix} \partial p(c(x,\xi)-\eta)\right]$$

which implies (2.7).

Part (ii). In the case when $c(x,\xi)$ is convex or strictly differentiable w.r.t. x for all $\xi \in \Xi$, $g(x,\eta,\xi)$ is Clarke regular and equality in (2.8) holds. Subsequently we have

$$\partial_{(x,\eta)} G(x,\eta) = \mathbb{E}\left[\begin{pmatrix} \partial_x c(x,\xi) \\ -1 \end{pmatrix} \partial p(c(x,\xi)-\eta)\right]$$

and hence equality in (2.7) holds. ■

2.2 Optimality conditions of the true problem

Let us start with true problems. A widely used condition for deriving optimality conditions of a constrained convex program is Slater's constraint qualification.

Assumption 2.2 Problem (1.5) satisfies the *Slater's constraint qualification*, that is, there exists a point $(x_0,\eta_0) \in \mathcal{X} \times \mathbb{R}$ such that $H(x_0,\eta_0) < 0$.

To see how strong the assumption is, consider the case when $c(x,\xi)$ is bounded and there exists $x_0 \in \mathcal{X}$ such that $\mathbb{E}[c(x_0,\xi)] < 0$. By the definition of $H(x,\eta)$ and Assumption 2.1, $\lim_{\eta\to-\infty} H(x_0,\eta) = \mathbb{E}[c(x_0,\xi)] < 0$. This shows that we can find an η_0 sufficiently small such that $H(x_0,\eta_0) < 0$.

Let $\lambda \geq 0$ be a number and define the Lagrange function of problem (1.5):

$$\mathscr{L}(x,\eta,\lambda) := f(x) + \lambda H(x,\eta).$$

Proposition 2.2 *Assume that f and $c_i, i = 1\dots,m$, are convex w.r.t. x. Let (x^*,η^*) be an optimal solution of (1.5). Under Assumptions 2.1-2.2, there exists a number $\lambda^* \in \mathbb{R}_+$ such that*

$$\left\{\begin{array}{l} (x^*,\eta^*) \in \arg\min_{(x,\eta)\in\mathcal{X}\times R} \mathscr{L}(x,\eta,\lambda^*), \\ 0 \le -H(x^*,\eta^*)\perp\lambda^* \ge 0. \end{array}\right. \tag{2.9}$$

The set of λ^* satisfying (2.9) is nonempty, convex and bounded, and is the same for any optimal solution of the problem.

Proof. Since f and $c_i, i = 1\dots,m$ are convex functions, problem (1.5) is convex. By Assumption 2.2 and Bonnas and Shapiro [5, Proposition 2.106], Robinson's constraint qualification holds. The conclusion then follows from [5, Theorem 3.4]. ■

It is possible to characterize the optimality conditions (2.9) in terms of the derivatives of the underlying functions. In what follows, we do so for general case, by invoking to Hiriart-Urruty's KKT conditions for nonsmooth problem with equality and inequality constraints [13]. Recall that the Bouligrand tangent cone to a set $X \subset \mathbb{R}^n$ at a point $x \in X$ is defined as follows:

$$\mathcal{T}_X(x) := \{u \in \mathbb{R}^n : d(x+tu, X) = o(t), t \ge 0\}.$$

The normal cone to X at x, denoted by $\mathcal{N}_X(x)$, is the polar of the tangent cone:

$$\mathcal{N}_X(x) := \{\zeta \in \mathbb{R}^n : \zeta^T u \le 0, \forall u \in \mathcal{T}_X(x)\}$$

and $\mathcal{N}_X(x) = \emptyset$ if $x \notin X$.

Let $\Phi(x)$ be a locally Lipschitz function defined on an open subset $\mathcal{O} \subset \mathbb{R}^n$. Let $x_0 \in \mathcal{O}$ and let Q be a subset of $\mathbb{R}^n$ such that $x_0 \in \bar{Q}$ (closure of Q). We denote by $\mathscr{V}_Q(x_0)$ the filter of neighborhoods of x_0 for the topology induced on Q. The collection $(\partial_x\Phi(x)|x \in \mathcal{O}, \mathscr{V}_Q(x_0))$ is a filtered family [3, p.126]. For this family, we may consider the "lim sup" which we will denote by $\partial_x^Q\Phi(x_0)$

$$\partial_x^Q\Phi(x_0) := \bigcap_{V\in\mathscr{V}_Q(x_0)} \overline{\bigcup_{x\in V} \partial_x\Phi(x)}.$$

In other words, $\partial_x^Q\Phi(x_0)$ consists of all cluster points of sequences of matrices $M_i \in \partial_x\Phi(x_i)$ as x_i converges to x_0 in Q. Obviously, $\partial_x^Q\Phi(x_0) \subset \partial_x\Phi(x_0)$, see [13].

The following constraint qualification stems from Hiriart-Urruty [13].

Definition 2.1 Problem (1.5) is said to satisfy the *subdifferential constraint qualification* at a feasible point (x,η) if for all $\zeta \in \partial_{(x,\eta)}^{S^c} H(x,\eta)$, there exists $d \in \mathbb{R}^{n+1}$ such that $\zeta^T d < -\delta$, where S denotes the feasible set of the problem (1.5) and S^c is complementary set of S.

In the case when $\delta = 0$, the condition is regarded as an extended *Mangasarian-Fromovitz constraint qualification* (MFCQ), see discussions at pages 79-80 in Hiriart-Urruty [13]. Indeed, if $H(x,\eta)$ is differentiable and $\delta = 0$, then the subdifferential constraint qualification reduces to the classical MFCQ.

Theorem 2.1 *Let $(x^*, \eta^*) \in \mathcal{X} \times \mathbb{R}$ be a local optimal solution to the true problem (1.5). Suppose that Assumption 2.1 holds and the subdifferential constraint qualification is satisfied at (x^*, η^*). Then there exists $\lambda^* \in R_+$ such that (x^*, η^*, λ^*) satisfies the following Karush-Kuhn-Tucker (KKT for short) conditions*

$$\begin{cases} 0 \in \begin{pmatrix} \nabla f(x) \\ 0 \end{pmatrix} + \lambda \partial_{(x,\eta)} H(x, \eta) + \mathcal{N}_{\mathcal{X} \times \mathbb{R}}(x, \eta), \\ 0 \leq -H(x, \eta) \perp \lambda \geq 0, \end{cases} \tag{2.10}$$

which imply

$$\begin{cases} 0 \in \begin{pmatrix} \nabla f(x) \\ 0 \end{pmatrix} + \lambda \mathbb{E} \left[\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{1}{\alpha} \partial_{(x,\eta)} g(x, \eta, \xi) \right] + \mathcal{N}_{\mathcal{X} \times \mathbb{R}}(x, \eta), \\ 0 \leq -H(x, \eta) \perp \lambda \geq 0. \end{cases} \tag{2.11}$$

If $c(x, \xi)$ is strictly differentiable w.r.t. x, then (2.10) and (2.11) are equivalent.

Proof. The KKT conditions (2.10) follow from Proposition 2.1 and [13, Theorem 4.2]. The conditions in (2.11) follow from (2.7) and (2.10).

In the case when $c(x, \xi)$ is strictly differentiable, one can easily use Proposition 2.1 (ii) to show that KKT conditions (2.11) are equivalent to (2.10). The proof is complete. ■

Remark 2.1 In the case when $f(x)$ and $c(x, \xi)$ are convex w.r.t. x, the subdifferentiable constraint qualification in Theorem 2.1 can be weakened to Slater's constraint qualification, in which case, (2.10) and (2.11) are equivalent without strict differentiability. The claim follows from Proposition 2.2.

The KKT conditions (2.11) are weaker than those of (2.10) due to (2.7) in general, see Xu [31] for a detailed discussion on this.

A tuple $(x, \eta, \lambda) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+$ satisfying (2.10) is called a KKT pair of problem (1.5), (x, η) a *Clarke stationary point* and λ the corresponding Lagrange multiplier. Similarly, a tuple $(x, \eta, \lambda) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+$ satisfying (2.11) is called a weak KKT pair of problem (1.5), (x, η) a *weak Clarke stationary point* and λ the corresponding Lagrange multiplier.

2.3 Optimality conditions of SAA problem

We now move on to discuss the optimality conditions of SAA problem (1.6). We need the following technical results.

Proposition 2.3 *Let Assumption 2.1 holds. Let $Z =: \mathcal{X} \times \mathbb{R}$ and $\mathcal{Z}$ be a compact subset of Z. Then*

(i) w.p.1 $\frac{1}{N} \sum_{j=1}^{N} c(x, \xi^j)$ converge respectively to $\mathbb{E}[c(x, \xi)]$ uniformly over any compact subset of $\mathcal{X}$ as $N \to \infty$;

(ii) w.p.1 $\frac{1}{N} \sum_{i=1}^{N} g(x, \eta, \xi)$ converges respectively to $G(x, \eta)$ uniformly on $\mathcal{Z}$ as $N \to \infty$;

(iii) if, in addition, Assumption 2.2 holds, then the SAA problem (1.6) satisfies Slater's constraint qualification *w.p.1 for N sufficiently large, that is, there exists a point* $(x, \eta) \in \mathcal{X} \times R$ *such that*

$$H_N(x, \eta) < 0$$

w.p.1 for N sufficiently large.

Proof. Part (i) and Part (ii) follow from [25, Proposition 7]. Part (iii) is straightforward from Parts (i)-(ii) and the definition of Slater's constraint qualification. ■

From Proposition 2.3, we know that for N sufficiently large, the SAA problem (1.6) satisfies the Slater's constraint qualification w.p.1. Consequently we have the following optimality conditions for problem (1.6).

Theorem 2.2 *Assume f and $c_i, i = 1, \ldots, m$, are convex w.r.t. x and Assumptions 2.1-2.2 hold. If (x_N, η_N) is an optimal solution of the problem (1.6), then there exists a number $\lambda_N \in \mathbb{R}_+$ such that w.p.1*

$$\begin{cases} (x_N, \eta_N) \in \arg\min_{(x,\eta)\in\mathcal{X}\times\mathbb{R}} \mathscr{L}_N(x, \eta, \lambda_N), \\ 0 \le -H_N(x_N, \eta_N) \perp \lambda_N \ge 0. \end{cases} \tag{2.12}$$

Moreover, w.p.1 (x_N, η_N, λ_N) satisfies the KKT conditions

$$\begin{cases} 0 \in \begin{pmatrix} \nabla f(x) \\ 0 \end{pmatrix} + \lambda \partial_{(x,\eta)} H_N(x, \eta) + \mathcal{N}_{\mathcal{X}\times\mathbb{R}}(x, \eta), \\ 0 \le -H_N(x, \eta) \perp \lambda \ge 0, \end{cases} \tag{2.13}$$

which imply

$$\begin{cases} 0 \in \begin{pmatrix} \nabla f(x) \\ 0 \end{pmatrix} + \lambda \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{1}{\alpha N} \sum_{i=1}^{N} \partial_{(x,\eta)} g(x, \eta, \xi) \right) + \mathcal{N}_{\mathcal{X}\times\mathbb{R}}(x, \eta), \\ 0 \le -H_N(x, \eta) \perp \lambda \ge 0. \end{cases} \tag{2.14}$$

and the set of λ_N satisfying (2.12) is nonempty, convex and bounded, and is the same for any optimal solution of the problem.

Proof. By Proposition 2.3, the Slater's condition holds for problem (1.6). Similar to Proposition 2.2, (2.12) holds, which implies (2.13). Since $c(x, \xi)$ is convex, by Proposition 2.1 (ii), (2.14) and (2.13) are equivalent. The proof is complete. ■

In what follows, we derive the KKT conditions for nonconvex case. We need some additional conditions.

Definition 2.2 Problem (1.5) is said to satisfy the *strong subdifferential constraint qualification* at a feasible point (x, η) if there exist a positive number δ and $d \in \mathbb{R}^{n+1}$ such that $\zeta^T d < -\delta$ for all $\zeta \in \partial_{(x,\eta)} H(x, \eta)$.

Note that since $\partial^{S^c}_{(x,\eta)} H(x,\eta) \subset \partial_{(x,\eta)} H(x,\eta)$, the strong subdifferential constraint qualification implies the subdifferential constraint qualification. The terminology was introduced by Dentcheva and Ruszczyński in [11] for a class of semi-infinite optimization problems. Through the separation theorem in convex analysis, it is easy to observe that the constraint qualification is equivalent to $0 \notin \partial H(x,\eta)$.

Theorem 2.3 *Let $(x_N, \eta_N) \in \mathcal{X} \times \mathbb{R}$ be a local optimal solution to the sample average approximation problem (1.6). Let $\hat{Z}$ denote the subset of points in $\mathcal{X} \times \mathbb{R}$ such that*

$$\lim_{N\to\infty} d((x_N, \eta_N), \hat{Z}) \to 0$$

w.p.1. Assume that $\hat{Z}$ is bounded, Assumption 2.1 holds and problem (1.5) satisfies the strong subdifferential constraint qualification at every point $(x,\eta) \in \hat{Z}$. Then w.p.1 there exists $\lambda_N \in R_+$ such that (x_N, η_N, λ_N) satisfies the KKT conditions (2.13) which imply (2.14). Moreover, if $c(x,\xi)$ is strictly differentiable w.r.t. x for all $\xi \in \Xi$, then KKT conditions (2.13) are equivalent to that of (2.14).

Proof. Under the strong subdifferential constraint qualification at (x,η), there exist a constant $\delta > 0$ and a vector $u \neq 0$ (which depends on (x,η)) such that

$$\sup_{\zeta \in \partial_{(x,\eta)} H(x,\eta)} \zeta^T u = H^o(x,\eta;u) \leq -\delta.$$

In what follows, we show that

$$\sup_{\zeta \in \partial_{(x,\eta)} H_N(x_N,\eta_N)} \zeta^T u \leq -\delta/2 \tag{2.15}$$

w.p.1 for N sufficiently large, i.e., problem (1.6) satisfies the subdifferential constraint qualification. Let $(x,\eta) \in \hat{Z}$,

$$\partial^{\epsilon}_{(x,\eta)} H(x,\eta) := \bigcup_{(x',\eta') \in (x,\eta) + \epsilon \mathcal{B}} \partial_{(x,\eta)} H(x',\eta')$$

and

$$H^o_\epsilon(x,\eta;u) := \sup_{\zeta \in \partial^{\epsilon}_{(x,\eta)} H(x,\eta)} \zeta^T u.$$

The outer semicontinuity of $\partial_{(x,\eta)} H(\cdot,\cdot)$ allows us to find a sufficiently small ϵ (depending on (x,η)) such that

$$H^o_\epsilon(x',\eta';u) \leq -\frac{3}{4}\delta$$

for all $(x',\eta') \in B(x,\eta;\rho)$. Let $\tilde{Z}$ denote the closed ρ neighborhood of $\hat{Z}$ relative to $\mathcal{X} \times \mathbb{R}$. Applying [29, Lemma 2.1] to $\tilde{Z}$, we can find a positive number $\hat{N}$ such that

$$H^o_N(x',\eta';u) - H^o_\epsilon(x',\eta';u) \leq \frac{\delta}{2}, \forall (x',\eta') \in \tilde{Z}$$

w.p.1 for $N \geq \hat{N}$. Using this inequality, we have

$$\sup_{\zeta \in \partial_{(x,\eta)} H_N(x_N,\eta_N)} \zeta^T u = H^o_N(x_N,\eta_N;u) \leq H^o_\epsilon(x_N,\eta_N;u) + \frac{\delta}{2} \leq -\frac{\delta}{4}$$

w.p.1 as long as $(x_N, \eta_N) \in B(x, \eta; \rho)$. That means for all $\zeta \in \partial_{(x,\eta)} H_N(x_N, \eta_N)$, w.p.1 $\zeta^T u \leq -\frac{\delta}{4}$. It implies that w.p.1 problem (1.6) satisfies the strong subdifferential constraint qualification at z_N for N sufficiently large. The rest of the proof is similar to Theorem 2.1. ■

A tuple $(x, \eta, \lambda) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+$ satisfying (2.14) is called a weak KKT pair of problem (1.6) and a tuple $(x, \eta, \lambda) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+$ satisfying (2.13) is called a KKT pair of problem (1.6).

We make a blanket assumption that throughout the rest of the paper the conditions of Theorems 2.2 or 2.3 hold.

3 Convergence analysis

In this section, we discuss convergence of SAA problem (1.6) as sample size N increases. Differing from many asymptotic analysis in the literature, our focus here is on the convergence of stationary points/KKT pair of the SAA problem in that a local or global optimal solution is also a stationary point. The analysis is practically useful in that: (a) when the problem is nonconvex, it is often difficult to obtain a global or even a local optimal solution, a stationary point might provide some information on local optimality; (b) CVaR approximation problem (1.4) has potential applications in finance and engineering. Our analysis is divided into two parts: almost sure convergence and exponential convergence. The former is to examine the asymptotic consistency of stationary points obtained from solving the SAA problem and the latter is to investigate the rate of convergence through large deviation theorem. Throughout this section, we assume that the probability space $(\Omega, \mathscr{F}, P)$ is nonatomic.

3.1 Almost sure convergence

Consider the SAA problem (1.6). Assume that for each given sampling, we solve the problem and obtain a stationary point (x_N, η_N) which satisfies (2.14). We investigate the convergence of (x_N, η_N) as N increases.

Assumption 3.1 *Let $Z := \mathcal{X} \times \mathbb{R}$. There exists a compact subset $\mathcal{Z} \times \Lambda \subset Z \times \mathbb{R}_+$ and a positive number N_0 such that w.p.1 problem (1.6) has a KKT pair $(x_N, \eta_N, \lambda_N) \in \mathcal{Z} \times \Lambda$ for $N \geq N_0$.*

This assumption is standard, see for instance Ralph and Xu [21].

Recall that for a set D, the support function of D is defined as

$$\sigma(D, u) = \sup_{d \in D} d^T u.$$

Let D_1, D_2 be two convex and compact subsets of $\mathbb{R}^m$. Let $\sigma(D_1, u)$ and $\sigma(D_2, u)$ denote the support functions of D_1 and D_2 respectively. Then

$$\mathbb{D}(D_1, D_2) = \max_{\|u\| \leq 1} (\sigma(D_1, u) - \sigma(D_2, u)). \tag{3.16}$$

The above relationship is known as Hörmander's formula, see [7, Theorem II-18].

Theorem 3.1 *Let $\{(x_N, \eta_N, \lambda_N)\}$ be a sequence of KKT pairs satisfying (2.14) and (x^*, η^*, λ^*) be a cluster point. Suppose Assumption 3.1 holds. Then w.p.1 (x^*, η^*, λ^*) is a weak KKT pair of the true problem (1.5). Moreover, if $c(x, \xi)$ is convex or strictly differentiable w.r.t. x, then w.p.1 (x^*, η^*, λ^*) is a KKT pair.*

Proof. Assume without loss of generality that (x_N, η_N, λ_N) converges to (x^*, η^*, λ^*) w.p.1 as $N \to \infty$. Since $0 \leq -H(x, \eta) \perp \lambda \geq 0$ is equivalent to $\max(\lambda, -H(x, \eta)) = 0$, in view of weak KKT conditions (2.11) and (2.14), it suffices to show that w.p.1

$$\begin{aligned}\overline{\lim_{N\to\infty}} &\left\{ \begin{pmatrix} \nabla f(x_N) \\ 0 \end{pmatrix} + \left[\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{1}{\alpha N} \sum_{i=1}^{N} \partial_{(x,\eta)} g(x_N, \eta_N, \xi^i) \right] \lambda_N + \mathcal{N}_Z(x_N, \eta_N) \right\} \\ &\subset \begin{pmatrix} \nabla f(x^*) \\ 0 \end{pmatrix} + \mathbb{E} \left[\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{1}{\alpha} \partial_{(x,\eta)} g(x^*, \eta^*, \xi) \right] \lambda^* + \mathcal{N}_Z(x^*, \eta^*) \end{aligned} \tag{3.17}$$

and

$$\lim_{N\to\infty} \max(-H_N(x_N, \eta_N), \lambda_N) = \max(-H(x^*, \eta^*), \lambda^*). \tag{3.18}$$

Since $\partial_{(x,\eta)} g(x, \eta, \xi)$ is outer semicontinuous w.r.t. (x, η) for almost every ξ and integrably bounded, by [30, Theorem 4],

$$\sup_{(x,\eta)\in\mathcal{Z}} \mathbb{D} \left(\frac{1}{N} \sum_{i=1}^{N} \partial_{(x,\eta)} g(x, \eta, \xi^i), \mathbb{E} \left[\partial_{(x,\eta)} g(x, \eta, \xi) \right] \right) \to 0$$

w.p.1 as $N \to \infty$. The uniform convergence and the continuous differentiability of $f(x)$ imply (3.17).

On the other hand, under Assumption 2.1, it follows from Proposition 2.3 that $H_N(x, \eta)$ converges uniformly to $H(x, \eta)$ over $\mathcal{Z}$, which implies (3.18).

When $c(x, \xi)$ is convex or strictly differentiable, H is Clarke regular, hence the KKT systems (2.10) and (2.11) are equivalent. ■

3.2 Exponential rate of convergence

We now move on to investigate the rate of convergence to strengthen the results established in Theorem 3.1 through large deviation theorems. Results presented as such are also known as exponential convergence. While the a.s. convergence is derived fairly easily through uniform law of large numbers for random set-valued mappings, exponential convergence is far more challenging. The main technical difficulty is to estimate the rate of uniform upper semi-convergence of sample average random set-valued mappings and we propose to deal with it by exploiting a new large derivation result in [28].

Definition 3.1 *(Almost H-calmness)* Let $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real valued function and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ be a random vector defined on probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{X} \subset \mathbb{R}^n$ be a closed subset of $\mathbb{R}^n$ and $x \in \mathcal{X}$ be fixed. ϕ is said to be

(a) *almost H-calm at x from above* with modulus $\kappa_x(\xi)$ and order γ_x if for any $\epsilon > 0$, there exist an integrable function $\kappa_x : \Xi \to \mathbb{R}_+$, positive numbers γ_x, $\delta_x(\epsilon)$, K and an open set $\Xi_x(\epsilon) \subset \Xi$ such that

$$\text{Prob}(\xi \in \Xi_x(\epsilon)) \leq K\epsilon \tag{3.19}$$

and

$$\phi(x', \xi) - \phi(x, \xi) \leq \kappa_x(\xi)||x' - x||^{\gamma_x} \tag{3.20}$$

for all $\xi \in \Xi \backslash \Xi_x(\epsilon)$ and all $x' \in B(x, \delta_x) \bigcap \mathcal{X}$.

(b) *almost H-calm at x from below* with modulus $\kappa_x(\xi)$ and order γ_x if for any $\epsilon > 0$, there exist an integrable function $\kappa_x : \Xi \to \mathbb{R}_+$, positive numbers γ_x, $\delta_x(\epsilon)$, K and an open set $\Xi_x(\epsilon) \subset \Xi$ such that

$$\text{Prob}(\xi \in \Xi_x(\epsilon)) \leq K\epsilon \tag{3.21}$$

and

$$\phi(x', \xi) - \phi(x, \xi) \geq -\kappa_x(\xi)||x' - x||^{\gamma_x} \tag{3.22}$$

for all $\xi \in \Xi \backslash \Xi_x(\epsilon)$ and all $x' \in B(x, \delta_x) \bigcap \mathcal{X}$.

(c) *almost H-calm at x* with modulus $\kappa_x(\xi)$ and order γ_x if for any $\epsilon > 0$, there exist an integrable function $\kappa_x : \Xi \to \mathbb{R}_+$, positive numbers γ_x, $\delta_x(\epsilon)$, K and an open set $\Xi_x(\epsilon) \subset \Xi$ such that

$$\text{Prob}(\xi \in \Xi_x(\epsilon)) \leq K\epsilon \tag{3.23}$$

and

$$|\phi(x', \xi) - \phi(x, \xi)| \leq \kappa_x(\xi)||x' - x||^{\gamma_x} \tag{3.24}$$

for all $\xi \in \Xi \backslash \Xi_x(\epsilon)$ and all $x' \in B(x, \delta_x) \bigcap \mathcal{X}$.

Remark 3.1 The concept of almost H-calmness is recently proposed by Sun and Xu [28] to derive a uniform large deviation theorem for a class of discontinuous random functions ([28, Theorem 3.1]) where the underlying random variable satisfies a continuous distribution. It is closely related to the following calmness condition suggested by a referee:

- There exist an integrable function $\kappa_x : \Xi \to \mathbb{R}_+$, positive numbers γ_x, δ_x and a measurable subset $\Xi_x \subset \Xi$ such that

$$\text{Prob}(\xi \in \Xi_x) = 0 \tag{3.25}$$

and

$$|\phi(x', \xi) - \phi(x, \xi)| \leq \kappa_x(\xi)||x' - x||^{\gamma_x} \tag{3.26}$$

for all $\xi \in \Xi \backslash \Xi_x$ and all $x' \in B(x, \delta_x) \bigcap \mathcal{X}$.

Conditions (3.25)-(3.26) require H-calmness (3.26) to hold for *almost every* $\xi \in \Xi$ and the two conditions may be regarded as a limiting case of almost H-calmness ($\epsilon \to 0_+$). Let us call the resulting calmness as *limiting almost H-calmness*. In the case when $\Xi_x = \emptyset$, it reduces to the H-calmness of [31, Definition 2.3], which requires H-calmness condition (3.26) to hold for *every* $\xi \in \Xi$.

Let μ denote the Lebesgue measure relative to Ξ. The limiting almost calmness conditions imply $\mu(\Xi_x) = 0$. Therefore for any $\epsilon > 0$, there exists any open subset $\Xi_x^\epsilon \subset \Xi$ such that $\mu(\Xi_x^\epsilon) < \epsilon$. This means limiting almost H-calmness implies almost H-calmness. The reverse assertion may not be true.

The example below shows that an almost H-calm random function does not necessarily satisfy the limiting almost H-calmness condition and the necessity of almost H-calmness.

Example 3.1 *Consider random function*

$$\varphi(x,\xi) := \begin{cases} \frac{1}{\sqrt{|x-\xi|}}, & \text{for } x \neq \xi, \\ \infty, & \text{for } x = \xi, \end{cases}$$

where ξ is a random variable satisfying uniform distribution over $[0,1]$. For every fixed ξ, function $\varphi(\cdot,\xi)$ is calm at any point $x \in [0,1]$ except at point $x = \xi$ because it is locally continuously differentiable. However, this function does not satisfy limiting almost H-calmness in the sense that there does not exist positive numbers δ, γ and positive measurable function $\kappa(\xi)$ such that

$$|\varphi(x',\xi') - \varphi(x,\xi')| \leq \kappa(\xi')|x'-x|^\gamma \tag{3.27}$$

for all $\xi' \in \Xi\backslash\{x\}$ and $x' \in (x-\delta, x+\delta) \cap [0,1]$. On the other hand, it is easy to verify that $\varphi(x,\xi)$ is almost H-calm at any point in $[0,1]$. Indeed for every $\epsilon > 0$, Let $\delta := \epsilon/2$, $\Xi_x(\epsilon) := (x - \frac{\epsilon}{2}, x + \frac{\epsilon}{2}) \cap [0,1]$ and

$$\kappa_x(\xi) := \begin{cases} M, & \text{for } \xi \in \Xi_x^\epsilon \cap [0,1], \\ \frac{1}{2\sqrt{|x-\xi|^3}}, & \text{otherwise}, \end{cases}$$

where M is any positive constant. Then it is easy to verify that

$$|\varphi(x',\xi') - \varphi(x,\xi)| \leq \kappa_x(\xi')|x'-x| \tag{3.28}$$

for all $\xi' \in [0,1]\backslash\Xi_x(\epsilon)$ and $x' \in (x - \frac{1}{2}\epsilon, x + \frac{1}{2}\epsilon)$.

Note that the exponential convergence result based on almost H-calmness, [28, Theorem 3.1], requires $\varphi(x,\xi)$ to be bounded. However, it is easy to observe that [28, Theorem 3.1] holds when $\varphi(x,\xi)$ is integrably bounded. See Theorem 5.1 in the appendix.

In some cases, even the random function satisfies the limiting almost H-calmness condition, it is still unclear whether we are able to derive the exponential convergence through a generalization of Theorem 3.1 in [31]. To see this, consider function $\phi(x,\xi) = \max(x,\xi)$ defined on $[0,1]\times[0,1]$, where ξ is a random variable satisfying uniform distribution over $[0,1]$. This function is Lipschitz continuous w.r.t. x, therefore the Clarke generalized derivative exists. For the fixed direction $d = 1$, the derivative function can be written as:

$$\phi_x^o(x,\xi;d) = \begin{cases} 1, & \text{for } x \geq \xi, \\ 0, & \text{for } x < \xi. \end{cases}$$

It is continuous w.r.t. (x, ξ) *except at line* $x = \xi$*. In what follows, we show that it is limiting H-calm at every point in the interval* $[0, 1]$*. Let* $x \in [0, 1]$*. There exist positive constants* $\gamma_x < 1$ *and integrable function* $\kappa_x(\cdot) := 1/\sqrt{|\cdot - x|^{\gamma_x}}$ *such that*

$$|\phi_x^o(x', \xi'; 1) - \phi_x^o(x, \xi'; 1)| \leq \kappa_x(\xi')|x' - x|$$

for every $\xi' \in [0, 1]\backslash\{x\}$ *and* $x' \in [0, 1]\backslash\{x\}$*. However the moment generating function of* $\kappa_x(\xi)$ *defined as such does not seem to exist, which means we cannot apply [31, Theorem 3.1] to derive uniform exponential convergence for sample average approximation of the derivative function* $\phi_x^o(x, \xi; 1)$ *over any compact subset of* $[0, 1]$*. On the other hand, similar to the analysis for the random function in the preceding discussions of this example, we can show that this derivative function* $\phi_x^o(x, \xi; 1)$ *is indeed almost H-calm and hence we can derive the uniform exponential convergence for the sample average of the function through Theorem 5.1. We omit the details. The discussions above effectively demonstrate the advantage of almost H-calmness over H-calmness or limiting H-calmness.*

Since the subdifferential of max function take on a key role in our exponential convergence analysis and both H-calmness and limiting almost H-calmness seem too strong, we will use Theorem 5.1 which base on almost H-calmness to derive exponential convergence of stationary points obtained from the SAA problem (1.6). To this end, we need a couple of technical results.

Lemma 3.1 *Let* $X \in \mathbb{R}^n$ *be a compact set and* $x \in X$ *. Let* $K_1(x, \xi)$ *and* $K_2(x, \xi)$ *be continuously differentiable w.r.t.* (x, ξ) *and* $K(x, \xi) := \max\{K_i(x, \xi), i = 1, 2\}$*. Define*

$$\Xi_K(x) := \{\xi : K(x, \xi) = K_1(x, \xi) = K_2(x, \xi), \xi \in \Xi\}.$$

Assume that $\Xi_K(x)$ *is compact and*

$$\nabla_\xi(K_1(x, \xi) - K_2(x, \xi)) \neq 0, \forall \xi \in \Xi_K(x).$$

Then $\mu(\Xi_K(x)) = 0$*, where* μ *denotes the Lebesgue measure relative to* Ξ*. Moreover, for any* $\epsilon > 0$ *and any fixed* $x \in X$*, there exists an open set* $\Xi_K^\epsilon(x)$ *(depending on* x *and* ϵ*) such that* $\Xi_K(x) \subset \Xi_K^\epsilon(x)$ *and* $\mu(\Xi_K^\epsilon(x) \cap \Xi) \leq \epsilon$*.*

Proof. The conclusion follows in a straightforward way from [28, Lemma 4.1]. ■

Let $g_i(x, \eta, \xi) := c_i(x, \xi) - \eta$, for $i = 1, \ldots, m$, and $g_{m+1}(x, \eta, \xi) := 0$. Then $g(x, \eta, \xi) = \max_{i=1}^{m+1}\{g_i(x, \eta, \xi)\}$. For any $i, j \in \{1, \ldots, m + 1\}$, $i \neq j$, define

$$\Xi_{i,j}(x, \eta) := \{\xi \in \Xi : g(x, \eta, \xi) = g_i(x, \eta, \xi) = g_j(x, \eta, \xi)\}.$$

and

$$\Xi(x, \eta) := \bigcup_{i,j \in \{1,\cdots,m+1\}} \Xi_{i,j}(x, \eta).$$

Obviously, $\Xi(x, \eta)$ consists of the set of $\xi \in \Xi$ such that $g(\cdot, \cdot, \xi)$ is not differentiable at (x, η).

Proposition 3.1 *Let* $\mathcal{Z} \subset \mathcal{X} \times \mathbb{R}$ *be a compact set and* $(x, \eta) \in \mathcal{Z}$ *and* ξ *be a continuous random variable. Assume: (a)* $c_i(x, \xi)$ *is continuously differentiable w.r.t.* (x, ξ) *and twice continuously* *differentiable w.r.t.* x *for almost every* $\xi \in \Xi$*; (b) there exists an integrable function* $\kappa : \Xi \to \mathbb{R}$ *such that* $\nabla_x c_i(\cdot, \xi)$ *is Lipschitz continuous with modulus* $\kappa(\xi)$ *for every* $\xi \in \Xi$ *and* $\mathbb{E}[\kappa(\xi)] < \infty$*; (c)* $\Xi(x, \eta)$ *and* $\Xi_{i,j}(x, \eta)$ *are compact and*

$$\nabla_\xi (g_i(x, \eta, \xi) - g_j(x, \eta, \xi)) \neq 0, \forall \xi \in \Xi_{i,j}(x, \eta),$$

holds for all $i, j \in \{1, \ldots, m+1\}$*,* $i \neq j$*. Then*

(i) $\mathbb{E}[g^o_{(x,\eta)}(x, \eta, \xi; u)]$ *is a continuous function w.r.t.* (x, η, u)*;*

(ii) if, in addition, Ξ *is compact, then* $g^o_{(x,\eta)}(x, \eta, \xi; u)$ *is almost H-calm w.r.t.* (x, η, u) *with modulus* $\kappa(\xi)$ *and order* 1 *on* $\mathcal{Z}$*.*

Proof. Part (i). This is a well-known result. Note that condition (a) implies the twice continuous differentiability of $g_i(x, \eta, \xi)$ w.r.t. x and η and condition (b) implies locally Lipschitz continuity of $\nabla_{(x,\eta)} g_i(\cdot, \cdot, \xi)$ with modulus $\kappa(\xi)$ for every $\xi \in \Xi$. Indeed, in this case $\mathbb{E}[g(x, \eta, \xi)]$ is continuously differentiable, see for instance [20, Theorem 1].

Part (ii). Let $\epsilon > 0$ and $(\bar{x}, \bar{\eta}) \in \mathcal{Z}$ be fixed. Under condition (c), it follows by Lemma 3.1 that there exists an open subset $\Xi^\epsilon_{i,j}(\bar{x}, \bar{\eta})$ such that $\Xi \cap \Xi_{i,j}(\bar{x}, \bar{\eta}) \subset \Xi^\epsilon_{i,j}(\bar{x}, \bar{\eta})$ and $\mu(\Xi^\epsilon_{i,j}(\bar{x}, \bar{\eta})) \leq \epsilon$ for all $i, j \in \{1, \ldots, m+1\}$, $i \neq j$. Let $\Xi^\epsilon(\bar{x}, \bar{\eta}) := \bigcup_{i,j} \Xi^\epsilon_{i,j}(\bar{x}, \bar{\eta})$. Since $\Xi(\bar{x}, \bar{\eta}) = \bigcup_{i,j} \Xi_{i,j}(\bar{x}, \bar{\eta})$, we have

$$\Xi(\bar{x}, \bar{\eta}) \subset \Xi^\epsilon(\bar{x}, \bar{\eta}) \quad \text{and} \quad \mu(\Xi^\epsilon(\bar{x}, \bar{\eta})) \leq \binom{m}{2} \epsilon.$$

Let $\bar{\xi} \notin \Xi^\epsilon(\bar{x}, \bar{\eta})$. Then there exists only a single $i \in \{1, \ldots, m+1\}$ such that $g(\bar{x}, \bar{\eta}, \bar{\xi}) = g_i(\bar{x}, \bar{\eta}, \bar{\xi})$. We can find a δ-neighborhood of $(\bar{x}, \bar{\eta}, \bar{\xi})$ (depending on $\bar{x}$, $\bar{\eta}$ and $\bar{\xi}$), denoted by $B((\bar{x}, \bar{\eta}, \bar{\xi}), \delta_{\bar{x}, \bar{\eta}, \bar{\xi}})$ relative to $\mathcal{Z} \times \Xi$, such that for all $(x, \eta, \xi) \in B((\bar{x}, \bar{\eta}, \bar{\xi}), \delta_{\bar{x}, \bar{\eta}, \bar{\xi}})$, $g(x, \eta, \xi) = g_i(x, \eta, \xi)$. Let $\bar{u}$ be any direction such that $||\bar{u}|| \leq 1$, since $g_i(x, \eta, \xi)$ is continuously w.r.t. (x, η) differentiable in a neighborhood of $(\bar{x}, \bar{\eta}, \bar{\xi})$,

$$g^o_{(x,\eta)}(x, \eta, \xi; u) = \nabla_{(x,\eta)} g_i(x, \eta, \xi)^T u.$$

Under condition (b) and the compactness of Ξ, $\nabla_{(x,\eta)} g_i(\bar{x}, \bar{\eta}, \xi)$ is bounded and there exists integrable function $\bar{\kappa}$ such that

$$\begin{aligned} |g^o_{(x,\eta)}(x, \eta, \xi; u) - g^o_{(x,\eta)}(\bar{x}, \bar{\eta}, \xi; \bar{u})| &= |\nabla_{(x,\eta)} g_i(x, \eta, \xi)^T u - \nabla_{(x,\eta)} g_i(\bar{x}, \bar{\eta}, \xi)^T \bar{u}| \\ &\leq |\nabla_{(x,\eta)} g_i(x, \eta, \xi)^T u - \nabla_{(x,\eta)} g_i(x, \eta, \xi)^T \bar{u}| \\ &\quad + |\nabla_{(x,\eta)} g_i(\bar{x}, \bar{\eta}, \xi)^T \bar{u} - \nabla_{(x,\eta)} g_i(\bar{x}, \bar{\eta}, \xi)^T \bar{u}| \\ &\leq \max_{(x,\eta)} ||\nabla_{(x,\eta)} g_i(\bar{x}, \bar{\eta}, \xi)|| ||u - \bar{u}|| \\ &\quad + \kappa(\xi) ||\bar{u}|| ||(x, \eta) - (\bar{x}, \bar{\eta})|| \\ &\leq \bar{\kappa}(\xi)(||(x, \eta) - (\bar{x}, \bar{\eta})|| + ||u - \bar{u}||), \end{aligned} \tag{3.29}$$

where $\bar{\kappa}(\xi) := \max\left\{\kappa(\xi), \max_{(x,\eta)} ||\nabla_{(x,\eta)} g_i(\bar{x}, \bar{\eta}, \xi)||\right\}$. Since $\Xi \backslash \Xi^\epsilon(\bar{x}, \bar{\eta})$ is compact, we claim through the finite covering theorem that there exists a unified $\delta_{(\bar{x}, \bar{\eta})} > 0$ such that (3.29) holds for all $(x, \eta, u) \in B((\bar{x}, \bar{\eta}, \bar{u}), \delta_{((\bar{x}, \bar{\eta}))})$ and all $\xi \in \Xi \backslash \Xi^\epsilon((\bar{x}, \bar{\eta}))$. ■

Following Remark 3.1, Proposition 3.1 (ii) implies that $g^o(x, \eta, \xi; u)$ is limiting almost H-calm. Indeed condition (c) of Proposition 3.1 guarantees that set Ξ_x consists a finite number of points. This is the weakest verifiable sufficient condition that we could find to ensure $\mu(\Xi_x) = 0$ and existence of a positive constant δ_x: without this condition we are unable to show almost H-calmness of $g^o(x, \eta, \xi; u)$ or limiting almost H-calmness.

Theorem 3.2 *Let $\mathcal{Z} \times \Lambda$ be a nonempty compact subset of $Z \times \mathbb{R}_+$ and $H_N(x, \eta)$ be defined as in (1.6). Suppose, in addition to conditions of Proposition 3.1, that: (a) $c_i(x, \xi)$ is locally Lipschitz continuous w.r.t. x for every ξ with modulus $\kappa(\xi)$, where $\mathbb{E}[\kappa(\xi)] < \infty$, and (b) the support set of ξ is bounded. Let*

$$R_1^N(x, \eta, \lambda) := \mathbb{D}\left(\left(\begin{pmatrix} \nabla f(x_N) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \lambda \end{pmatrix} + \frac{\lambda}{\alpha N}\sum_{i=1}^N \partial_{(x,\eta)} g(x, \eta, \xi^i),\right.\right.$$
$$\left.\begin{pmatrix} \nabla f(x_N) \\ 0 \end{pmatrix} + \mathbb{E}\left[\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{1}{\alpha}\partial_{(x,\eta)} g(x, \eta, \xi)\right]\lambda\right)$$

and

$$R_2^N(x, \eta, \lambda) := ||\max(-H_N(x, \eta), \lambda) - \max(-H(x, \eta), \lambda)||.$$

Then with probability approaching one exponentially fast with the increase of sample size N, $\sup_{(x,\eta,\lambda)\in\mathcal{Z}\times\Lambda} R_1^N(x, \eta, \lambda)$ *and* $\sup_{(x,\eta,\lambda)\in\mathcal{Z}\times\Lambda} R_2^N(x, \eta, \lambda)$ *tend to* 0.

Proof. It is well-known that for any compact sets A, B, $\mathbb{D}(A, B) = \inf_{t>0}\{t : A \subset B + t\mathcal{B}\}$. Using this equivalence definition, one can easily derive that $\mathbb{D}(A, B) \le \mathbb{D}(A, C) + \mathbb{D}(C, B)$ and $\mathbb{D}(A + C, B + C) \le \mathbb{D}(A, B)$ for any set C. Consequently, we have

$$R_1^N(x, \eta, \lambda) \le \mathbb{D}\left(\frac{1}{\alpha N}\sum_{i=1}^N \partial_{(x,\eta)} g(x, \eta, \xi^i)\lambda, \frac{1}{\alpha}\mathbb{E}\left[\partial_{(x,\eta)} g(x, \eta, \xi)\right]\lambda\right).$$

Since the Clarke subdifferential is convex and compact set-valued, we can use the Hörmander's formula (3.16) to reformulate the right-hand side of the inequality above as:

$$\max_{\|u\|\le 1}\left[\sigma\left(\frac{1}{\alpha N}\sum_{i=1}^N \partial_{(x,\eta)} g(x, \eta, \xi^i)\lambda, u\right) - \sigma\left(\frac{1}{\alpha}\mathbb{E}\left[\partial_{(x,\eta)} g(x, \eta, \xi)\right]\lambda, u\right)\right]$$

and, by virtue of the property of support function (see [12]) and [18, Proposition 3.4], further as:

$$\frac{1}{\alpha}\max_{\|u\|\le 1}\left[\frac{1}{N}\sum_{i=1}^N g^o_{(x,\eta)}(x, \eta, \xi^i; u) - \mathbb{E}\left[g^o_{(x,\eta)}(x, \eta, \xi; u)\right]\right]\lambda. \quad (3.30)$$

Let $\Delta_N(x, \eta, u) := \frac{1}{N}\sum_{i=1}^N g^o_{(x,\eta)}(x, \eta, \xi^i; u) - \mathbb{E}[g^o_{(x,\eta)}(x, \eta, \xi; u)]$. Since Λ is a compact set, there exists a positive number M such that $\sup_{\lambda\in\Lambda}\lambda \le M$. Consequently

$$\text{Prob}\left\{\sup_{\lambda\in\Lambda,(x,\eta)\in Z,\|u\|\le 1} \Delta_N(x, \eta, u)\lambda \ge \epsilon\right\} \le \text{Prob}\left\{M \sup_{(x,\eta)\in\mathcal{Z},\|u\|\le 1} \Delta_N(x, \eta, u) \ge \epsilon\right\}. \quad (3.31)$$

By Proposition 3.1, $g^o_{(x,\eta)}(x, \eta, \xi; u)$ is almost H-calm with modulus $\kappa(\xi)$ and order 1 and $\mathbb{E}[g^o_{(x,\eta)}(x, \eta, \xi; u)]$ is a continuous function. Under condition (b), the moment generating function

$$M_g(x, \eta) := \mathbb{E}\left[e^{(g^o_{(x,\eta)}(x,\eta,\xi;u) - \mathbb{E}[g^o_{(x,\eta)}(x,\eta,\xi;u)])t}\right]$$

and

$$M_\kappa(t) := \mathbb{E}\left\{e^{[\kappa(\xi)-\mathbb{E}[\kappa(\xi)]]t}\right\}$$

are finite valued for t close to 0. By Theorem 5.1, for any $\epsilon > 0$, there exist constants $c_1(\epsilon) > 0$ and $\beta_1(\epsilon) > 0$ (independent of N) such that

$$\mathrm{Prob}\left\{\sup_{(x,\eta)\in\mathcal{Z},\|u\|\le 1} \frac{1}{N}\sum_{i=1}^N g^o_{(x,\eta)}(x,\eta,\xi^i;u) - \mathbb{E}[g^o_{(x,\eta)}(x,\eta,\xi;u)] \ge \frac{\epsilon\alpha}{M}\right\} \le c_1(\epsilon)e^{-N\beta_1(\epsilon)}.$$

Combining the inequality above with (3.30) and (3.31), we have

$$\begin{aligned}
&\mathrm{Prob}\left\{\sup_{(x,\eta,\lambda)\in\mathcal{Z}\times\Lambda} R_1^N(x,\eta,\lambda) \ge \epsilon\right\} \\
&\le \mathrm{Prob}\left\{\sup_{(x,\eta)\in\mathcal{Z},\|u\|\le 1} \frac{1}{N}\sum_{i=1}^N g(x,\eta,\xi^i,u) - \mathbb{E}[g(x,\eta,\xi,u)] \ge \frac{\epsilon\alpha}{M}\right\} \\
&\le c_1(\epsilon)e^{-\beta_1(\epsilon)N}. \qquad (3.32)
\end{aligned}$$

On the other hand, since $R_2^N(x,\eta,\lambda) \le ||H(x,\eta) - H_N(x,\eta)||$, under conditions (a) and (b), it follows by [31, Theorem 3.1] that, for any small positive number $\epsilon > 0$, there exist positive constants $c_2(\epsilon)$ and $\beta_2(\epsilon)$ (independent of N) such that

$$\begin{aligned}
\mathrm{Prob}\left\{\sup_{(x,\eta,\lambda)\in\mathcal{Z}\times\Lambda} R_2^N(x,\eta,\lambda) \ge \epsilon\right\} &\le \mathrm{Prob}\left\{\sup_{(x,\eta)\in\mathcal{Z}} \|H(x,\eta) - H_N(x,\eta)\| \ge \epsilon\right\} \\
&\le c_2(\epsilon)e^{-\beta_2(\epsilon)N}. \qquad (3.33)
\end{aligned}$$

The proof is complete. ■

Let

$$\Gamma(x,\eta,\lambda) := \begin{pmatrix} \begin{pmatrix} \nabla f(x) \\ 0 \end{pmatrix} + \mathbb{E}\left[\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{1}{\alpha}\partial_{(x,\eta)} g(x,\eta,\xi)\right]\lambda \\ \max(-H(x,\eta),\lambda) \end{pmatrix}$$

and

$$\mathcal{G}(x,\eta) := \begin{pmatrix} \mathcal{N}_{\mathcal{X}\times\mathbb{R}}(x,\eta) \\ 0 \end{pmatrix}.$$

We can rewrite (2.11) as a stochastic generalized equation

$$0 \in \Gamma(x,\eta,\lambda) + \mathcal{G}(x,\eta).$$

Likewise, we can rewrite the KKT conditions (2.14) of the SAA problem as follows:

$$0 \in \Gamma_N(x,\eta,\lambda) + \mathcal{G}(x,\eta),$$

where

$$\Gamma_N(x,\eta,\lambda) := \begin{pmatrix} \begin{pmatrix} \nabla f(x) \\ 0 \end{pmatrix} + \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{1}{\alpha N}\sum_{i=1}^N \partial_{(x,\eta)} g(x,\eta,\xi)\right)\lambda \\ \max(-H_N(x,\eta),\lambda) \end{pmatrix}.$$

It is easy to see that Theorem 3.2 implies

$$\sup_{(x,\eta,\lambda)\in\mathcal{Z}\times\Lambda} \mathbb{D}(\Gamma_N(x,\eta,\lambda),\Gamma(x,\eta,\lambda)) \to 0 \tag{3.34}$$

a.s. $N\to\infty$.

Theorem 3.3 *Assume the settings and conditions of Theorem 3.2. Under Assumption 3.1, for any $\epsilon>0$, there exist positive constants $c(\epsilon)$ and $\beta(\epsilon)$ independent of N such that*

$$\text{Prob}\{d((x_N,\eta_N),Z^*)\geq\epsilon\}\leq c(\epsilon)e^{-N\beta(\epsilon)},$$

where (x_N,η_N) denotes the KKT point satisfying (2.14) and Z^* denotes the set of weak Clarke stationary points characterized by (2.11).

Proof. The thrust of the proof is to use (3.34) and [31, Lemma 4.2]. To this end, we need to verify the outer semicontinuity of $\Gamma(x,\eta,\lambda)$. Observe that $\Gamma(x,\eta,\lambda)$ consists of two parts:

$$\begin{pmatrix}\nabla f(x)\\ 0\end{pmatrix}+\mathbb{E}\left[\begin{pmatrix}0\\1\end{pmatrix}+\frac{1}{\alpha}\partial_{(x,\eta)}g(x,\eta,\xi)\right]\lambda$$

and $\max(H(x,\eta),\lambda)$. Since f is continuously differentiable and $\partial_{(x,\eta)}g(x,\eta,\xi)$ is outer semicontinuous w.r.t. x,η for almost every ξ and integrably bounded, it follows by Aumann [2, Corollary 5.2],

$$\begin{pmatrix}\nabla f(x)\\ 0\end{pmatrix}+\mathbb{E}\left[\begin{pmatrix}0\\1\end{pmatrix}+\frac{1}{\alpha}\partial_{(x,\eta)}g(x,\eta,\xi)\right]\lambda$$

is outer semicontinuous. Moreover, since $H(x,\eta)$ is a continuous function, $\max(-H(x,\eta),\lambda)$ is continuous w.r.t. (x,η), and λ. Therefore $\Gamma(x,\eta,\lambda)$ is outer semicontinuous. The rest follows from (3.34) and [31, Lemma 4.2]. ■

It is important to note that the constants $c(\epsilon)$ and $\beta(\epsilon)$ in Theorem 3.3 may be significantly different from their counterparts in Theorem 3.2. To establish a precise relationship of these constants, we will need more information about the sensitivity of the true problem at the stationary points. One possibility is to look into the metric regularity type condition for set-valued mapping $\Gamma(x,\eta,\lambda)+\mathcal{G}(x,\eta)$. If there exists positive constants C and γ such that

$$d(x,\eta,Z^*)\leq Cd(0,\Gamma(x,\eta,\lambda)+\mathcal{G}(x,\eta))^\gamma$$

for x,η close to Z^*, then we can establish

$$d(x,\eta,Z^*)\leq C\left(||R_1^N(x,\eta,\lambda)||+||R_2^N(x,\eta,\lambda)||\right)^\gamma.$$

We refer interested readers to [22] for recent discussions on metric regularity. Under this circumstance, the constants $c(\epsilon)$ and $\beta(\epsilon)$ can be easily expressed in terms of $c(\epsilon):=c_1(\epsilon)+c_2(\epsilon)$ and $\beta(\epsilon):=\min(\beta_1(\epsilon),\beta_2(\epsilon))$ in Theorem 3.2. Moreover, following [28, Remark 3.1], under some additional conditions on the moment functions, we can obtain an estimation of sample size through (3.32) and (3.33), that is there exists a constant $\sigma>0$ such that for any $\epsilon>0$, $\text{Prob}\{d((x_N,\eta_N),Z^*)\geq\epsilon\}\leq\beta$ holds when

$$N\geq\frac{O(1)\sigma^2}{\epsilon^2}\left[n\ln\left(O(1)D\left(\frac{4\mathbb{E}[\kappa(\xi)]}{\epsilon}\right)\right)^{\frac{1}{\gamma}}+\ln\left(\frac{1}{\beta}\right)\right],$$

where $D:=\sup_{(x',\eta'),(x,\eta)\in\mathcal{Z}}\|(x',\eta')-(x,\eta)\|$ is the diameter of $\mathcal{Z}$ and $O(1)$ is a generic constant. We leave this to interested readers as it involves complex technical details.

4 DC-approximation method

Although CVaR approximation is known to be the "best" convex approximation method of chance constraints, as commented in [17], it is a convex conservative approximation, which means that there exists a gap between the CVaR approximation and the true constraint. In [14]-[15], Hong et al propose a DC-approximation method for a joint chance constraint. The numerical tests show that the DC-approximation scheme displays better results than CVaR approximation scheme. Hong et al also show almost sure convergence of optimal solution of subproblems in their algorithm called *sequential convex approximations* (SCA). Here we carry out some convergence analysis analogous to those in section 3 under the DC-approximation scheme and complement their results by showing almost sure convergence and exponential rate of convergence of stationary points.

The formulation of the DC-approximation problem is defined as follows:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & \inf_{t>0} \frac{1}{t} \left(\mathbb{E}[p(c(x,\xi)+t)] - \mathbb{E}[p(c(x,\xi))] \right) \le \alpha, \\ & x \in \mathcal{X}, \end{aligned} \tag{4.35}$$

where t is a positive number. In [14], Hong et al use ε-approximation of problem (4.35) by setting $t = \varepsilon$. The formulation of ε-approximation problem is:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & H_\varepsilon^{DC}(x) := \mathbb{E}[p(c(x,\xi)+\varepsilon)] - \mathbb{E}[p(c(x,\xi))] - \varepsilon\alpha \le 0, \\ & x \in \mathcal{X}. \end{aligned} \tag{4.36}$$

Hong et al also prove that when $\varepsilon \to 0$, the KKT point of problem (4.36) converges to that of problem (1.3).

From the formulation above, we can observe that the structure of problem (4.35) and (4.36) are similar to CVaR scheme (1.5). That means we can easily extend the results of optimality conditions and convergence analysis in the preceding sections to the DC-approximation problem. In this section, we outline those conditions and results for problem (4.36) and leave the technical details to interested readers.

The sample average approach of problem for problem (4.36) is defined as follows:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & H_N^{DC}(x) := \frac{1}{N} \sum_{j=1}^{N} \left(p(c(x,\xi^j)+\varepsilon) - p(c(x,\xi^j)) \right) - \varepsilon\alpha \le 0, \\ & x \in \mathcal{X}. \end{aligned} \tag{4.37}$$

We refer to (4.36) as the *true* problem and (4.37) as SAA problem.

4.1 First order necessary conditions

Let $\lambda \ge 0$ and define the Lagrange function of problem (4.35):

$$\mathscr{L}^{DC}(x,\lambda) := f(x) + \lambda H_\varepsilon^{DC}(x).$$

In order to derive the KKT conditions of problem (4.35), we need some assumptions:

Assumption 4.1 *$c_i(x,\xi)$, $i=1,\ldots,m$, is locally Lipschitz continuous w.r.t. x with modulus $\kappa_i(\xi)$ where $\mathbb{E}[\kappa_i(\xi)]<\infty$.*

Assumption 4.2 *$c(\cdot,\xi)$ is differentiable on $\mathcal{X}$ for a.e. ξ.*

Assumption 4.3 *Let $F(t,x) := \text{Prob}\{c(x,\xi)\le t\}$. There exists a constant $\varrho>0$ such that $F(t,x)$ is continuously differentiable on $(-\varrho,\varrho)\times\mathcal{X}$.*

Assumptions 4.1-4.3 are used and discussed in [14].

Lemma 4.1 *Suppose that Assumptions 4.1-4.3 are satisfied. Then $H_\varepsilon^{DC}(x)$ is continuously differentiable on $\mathcal{X}$ for $\varepsilon\in(-\varrho,\varrho)$ and*

$$\nabla_x H_\varepsilon^{DC}(x) = \mathbb{E}[\nabla_x c_{i(x)}(x,\xi)\cdot \mathbb{1}_{(-\varepsilon,+\infty)}(c(x,\xi))] - \mathbb{E}[\nabla_x c_{i(x)}(x,\xi)\cdot \mathbb{1}_{(0,+\infty)}(c(x,\xi))], \quad (4.38)$$

where $i(x) := \text{argmax}_{i=1,\ldots,m} c_i(x,\xi)$ and $\mathbb{1}_S(x) := \begin{cases} 0, \text{for } x\notin S, \\ 1, \text{for } x\in S. \end{cases}$

Proof. The results follow from [14, Lemma 2] and [20, Theorem 1]. ■

Problem (4.36) is said to satisfy the *differential constraint qualification* at a feasible point x if there exists $d\in\mathbb{R}^n$ such that $\nabla_x H_\varepsilon^{DC}(x)^T d < -\delta$, where $\delta>0$ is a positive constant.

Proposition 4.1 *Let $x^*\in\mathcal{X}$ be a local optimal solution to the true problem (4.36). Let Assumptions 2.1, 4.1-4.3 hold and the differential constraint qualification be satisfied at x^*. Then there exists a $\lambda^*\in R_+$ such that*

$$\begin{cases} 0\in\nabla f(x^*) + \lambda^*\nabla_x H_\varepsilon^{DC}(x^*) + \mathcal{N}_{\mathcal{X}}(x^*), \\ 0\le -H_\varepsilon^{DC}(x^*)\perp\lambda^*\ge 0. \end{cases} \quad (4.39)$$

Proof. Under Assumptions 4.2 and 4.3, we know from Lemma 4.1 that $H_\varepsilon^{DC}(x)$ is continuously differentiable. Moreover, under the differential constraint qualification, the result follows from [13, Theorem 4.2]. ■

We now move on to discuss the optimality conditions for the SAA problem (4.37).

Proposition 4.2 *Let $x_N\in\mathcal{X}$ be a local optimal solution to the sample average approximation problem (4.37). Let $\hat{X}$ denote a subset of $\mathcal{X}$ such that*

$$\lim_{N\to\infty} d(x_N,\hat{X})\to 0, w.p.1.$$

Suppose that Assumptions 2.1, 4.1-4.3 hold, $\hat{X}$ is bounded and the differential constraint qualification holds at every point $x\in\hat{X}$. Then w.p.1 problem (4.37) satisfies the differential constraint qualification for N sufficiently large, and there exists $\lambda_N\in R_+$ such that

$$\begin{cases} 0\in\nabla f(x_N) + \Phi_N(x_N)\lambda_N + \mathcal{N}_{\mathcal{X}}(x_N), \\ 0\le -H_N^{DC}(x_N)\perp\lambda_N\ge 0, \end{cases} \quad (4.40)$$

where $\Phi_N(x) := \frac{1}{N}\sum_{j=1}^N(\nabla_x c_{i(x)}(x,\xi^j)\cdot\mathbb{1}_{(-\varepsilon,+\infty)}(c(x,\xi^j)) - \nabla_x c_{i(x)}(x,\xi^j)\cdot\mathbb{1}_{(0,+\infty)}(c(x,\xi^j)))$.

Proof. By Assumptions 4.1-4.2, $c(x, \xi)$ is differentiable w.r.t. x and $\nabla_x c(x, \xi) = \nabla_x c_{i(x)}(x, \xi^i)$ w.p.1 for $x \in \mathcal{X}$. Under Assumption 4.3, we have $\text{Prob}\{c(x, \xi) = \varepsilon\} = 0$ and $\text{Prob}\{c(x, \xi) = 0\} = 0$, that is, $p(c(x, \xi) + \varepsilon)$ and $p(c(x, \xi))$ are differentiable and $\nabla_x H_N^{DC}(x) = \Phi_N(x)$ w.p.1.

By Lemma 4.1, $H_\varepsilon^{DC}(x)$ is continuously differentiable, $\nabla_x H_\varepsilon^{DC}(x)$ is continuous. Using a similar argument to that in Theorem 2.3 and [25, Proposition 7], we can show that problem (4.37) satisfies the differential constraint qualification at x_N for N sufficiently large w.p.1. The rest of the proof is similar to Proposition 4.1. ■

We call a tuple (x^*, λ^*) satisfying (4.39) a *KKT pair* of problem (4.36), x^* a *stationary point* and λ^* the corresponding Lagrange multiplier and a tuple (x_N, λ_N) satisfying (4.40) a *KKT pair* of problem (4.37). We also note that such a (x^*, λ^*) can be thought as an ε-*KKT pair* of problem (4.35).

We make a blanket assumption that throughout the rest of this section the conditions of Proposition 4.2 hold.

4.2 Convergence analysis

Assumption 4.4 *There exists a compact subset $X \times \Lambda \subset \mathcal{X} \times R_+$ and a positive number N_0 such that w.p.1 problem (4.37) has a KKT pair $(x_N, \lambda_N) \in X \times \Lambda$ for $N \geq N_0$.*

Theorem 4.1 *Let $\{(x_N, \lambda_N)\}$ be a sequence of KKT pairs of problem (4.37) and (x^*, λ^*) be a cluster point. Suppose Assumptions 4.1-4.4 hold. Then w.p.1 (x^*, λ^*) is a KKT pair of the true problem (4.36), which satisfies the KKT system (4.39).*

Proof. The claim follows straightforwardly from uniform law of large numbers [25, Proposition 7] and outer semicontinuous of the normal cone. ■

For the simplicity of notation, let $\hat{c}_i(x, \xi) := c_i(x, \xi) + \epsilon$, $i = 1, \ldots, m$, $\hat{c}_{m+1}(x, \xi) := 0$ and $\hat{c}(x, \xi) := \max_{i=1}^{m+1}\{\hat{c}_i(x, \xi)\}$; let $c_{m+1}(x, \xi) := 0$ and $\tilde{c}(x, \xi) := \max\{c(x, \xi), c_{m+1}(x, \xi)\}$. For any $i, j \in \{1, \ldots, m+1\}$, $i \neq j$. Define

$$\hat{\Xi}_{i,j}(x) := \{\xi \in \Xi : \hat{c}(x, \xi) = \hat{c}_i(x, \xi) = \hat{c}_j(x, \xi)\}$$

and

$$\tilde{\Xi}_{i,j}(x) := \{\xi \in \Xi : \tilde{c}(x, \xi) = c_i(x, \xi) = c_j(x, \xi)\}.$$

Let $\hat{\Xi}(x) := \bigcup_{i,j \in \{1,\cdots,m+1\}} \hat{\Xi}_{i,j}(x)$ and $\tilde{\Xi}(x) := \bigcup_{i,j \in \{1,\cdots,m+1\}} \tilde{\Xi}_{i,j}(x)$.

Proposition 4.3 *Let $\mathcal{X}$ be a compact set. Assume: (a) $c_i(x, \xi), i = 1, \ldots, m$, is continuously differentiable w.r.t. (x, ξ) and twice continuously differentiable w.r.t. x for almost every $\xi \in \Xi$; (b) there exists an integrable function $\kappa : \Xi \to \mathbb{R}$ such that $\nabla_x c_i(\cdot, \xi)$ is locally Lipschitz continuous with modulus $\kappa(\xi)$ for every $\xi \in \Xi$ where $\mathbb{E}[\kappa(\xi)] < \infty$; (c) $\hat{\Xi}(x)$, $\hat{\Xi}_{i,j}$, $\tilde{\Xi}(x)$ and $\tilde{\Xi}_{i,j}(x)$ are compact,*

$$\nabla_\xi(\hat{c}_i(x, \xi) - \hat{c}_j(x, \xi)) \neq 0, \forall \xi \in \hat{\Xi}_{i,j}(x),$$

and

$$\nabla_\xi(c_i(x, \xi) - c_j(x, \xi)) \neq 0, \forall \xi \in \tilde{\Xi}_{i,j}(x)$$

hold for all $i, j \in \{1, \ldots, m+1\}$, $i \neq j$. Then *(i)* $\mathbb{E}[\hat{c}_x^o(x,\xi;u)]$ *and* $\mathbb{E}[\tilde{c}_x^o(x,\xi;u)]$ *are continuous w.r.t.* (x,u)*;*

(ii) if, in addition, Ξ *is compact,* $\hat{c}_x^o(x,\xi;u)$ *and* $\tilde{c}_x^o(x,\xi;u)$ *are almost H-calm w.r.t.* (x,u) *with modulus* $\kappa(\xi)$ *and order* 1 *on* $\mathcal{X}$*.*

Proof. The conclusion follows from Proposition 3.1. ∎

Theorem 4.2 *Let* $X \times \Lambda$ *be a nonempty compact subset of* $\mathcal{X} \times \mathbb{R}_+$*. Let* $H_N^{DC}(x,\eta)$ *be defined as in (4.37) and* $\Phi_N(x)$ *be defined as in (4.40). Suppose, in addition to conditions of Proposition 4.3, that: (a) Assumptions 4.1, 4.3 and 4.4 hold; and (b) the support set of* ξ *is bounded. Then for any* $\epsilon > 0$*, there exist positive constants* $C(\epsilon)$ *and* $\beta(\epsilon)$ *independent of* N *such that*

$$\text{Prob}\{d(x_N, X^*) \geq \epsilon\} \leq C(\epsilon)e^{-N\beta(\epsilon)},$$

where x_N *denotes the Clarke stationary points characterized by (4.40) and* X^* *denotes the set of Clarke stationary points characterized by (4.39).*

Proof. Let

$$\vartheta_N(x,\lambda) := ||\nabla f(x) + \Phi_N(x)\lambda, \nabla f(x) + \nabla_x H_\varepsilon^{DC}(x)\lambda||$$

and

$$\vartheta_N^H(x,\lambda) := ||\max(-H_N^{DC}(x),\lambda), \max(-H_\varepsilon^{DC}(x),\lambda)||.$$

We show that $\sup_{(x,\lambda)\in X\times\Lambda} \vartheta_N(x,\lambda)$ and $\sup_{(x,\lambda)\in X\times\Lambda} \vartheta_N^H(x,\lambda)$ converge to 0 at exponentially rate as $N \to \infty$.

Note that condition (c) of Proposition 4.3 implies Assumption 4.2. Together with condition (a), we have

$$\nabla_x H_\varepsilon^{DC}(x) = \mathbb{E}[\nabla_x c_{i(x)}(x,\xi) \cdot \mathbb{1}_{(-\varepsilon,+\infty)}(c(x,\xi))] - \mathbb{E}[\nabla_x c_{i(x)}(x,\xi) \cdot \mathbb{1}_{(0,+\infty)}(c(x,\xi))],$$

where $i(x) := \text{argmax}_{i=1,\ldots,m}\{c_i(x,\xi)\}$, $\mathbb{E}[\nabla_x c_{i(x)}(x,\xi) \cdot \mathbb{1}_{(-\varepsilon,+\infty)}(c(x,\xi))]$ and $\mathbb{E}[\nabla_x c_{i(x)}(x,\xi) \cdot \mathbb{1}_{(0,+\infty)}(c(x,\xi))]$ are continuous. Note that for $\xi \notin \hat{\Xi}(x)$,

$$\hat{c}_x^o(x,\xi;u) = \nabla_x c_{i(x)}(x,\xi) \cdot \mathbb{1}_{(-\varepsilon,+\infty)}(c(x,\xi))^T u$$

and for $\xi \notin \tilde{\Xi}(x)$,

$$\tilde{c}_x^o(x,\xi;u) = \nabla_x c_{i(x)}(x,\xi) \cdot \mathbb{1}_{(0,+\infty)}(c(x,\xi))^T u.$$

By Proposition 4.3, $\hat{c}_x^o(x,\xi;u)$ and $\tilde{c}_x^o(x,\xi;u)$ are almost H-calm. Moreover, $\nabla_x c_{i(x)}(x,\xi) \cdot \mathbb{1}_{(-\varepsilon,+\infty)}(c(x,\xi))^T u$ and $\nabla_x c_{i(x)}(x,\xi) \cdot \mathbb{1}_{(0,+\infty)}(c(x,\xi))^T u$ are also almost H-calm w.r.t. (x,u) with modulus $\kappa(\xi)$ and order 1 on $\mathcal{X}$. Then the exponential convergence of $\sup_{(x,\lambda)\in X\times\Lambda} \vartheta_N(x,\lambda)$ and $\sup_{(x,\lambda)\in X\times\Lambda} \vartheta_N^H(x,\lambda)$ follows from the similar arguments to that in Theorem 3.2.

The rest follows by formulating the KKT conditions as generalized equations and applying a perturbation theorem [31, Lemma 4.2]. ∎

5 Numerical tests

We have carried out a number of numerical experiments on the approximation scheme for (1.1) in Matlab 7.9.0 installed in a PC with Windows XP operating system. To deal with joint chance constraint, we apply both CVaR and DC-approximation method to approximate it and then reformulate the latter as (1.5) and (4.36) respectively. In the tests, we apply SAA method to problem (1.5) and (4.36) and employ the random number generator *rand* in Matlab 7.9.0 to generate the samples and solver *fmincon* to solve the SAA problem (1.6) and (4.37). In [14], Hong et al use algorithm SCA to solve the DC-approximation scheme and then apply SAA method to the subproblem structured by algorithm SCA while we apply SAA method to DC-approximation scheme and then use the algorithm SCA to solve the SAA DC-approximation scheme. Since our focus in this paper is on convergence analysis, then our numerical test is to show the convergence behave with the sample size increase.

5.1 Algorithm

Since (4.37) is a DC problem, we will apply Algorithm SCA to solve the problem as in [14]. To this end, define problem $\mathrm{CP}_N(\tilde{x})$ as

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & \frac{1}{N}\sum_{j=1}^N \left(p(t+c(x,\xi^j)) - p(c(\tilde{x},\xi^j)) - \nabla_x c(\tilde{x},\xi^j)\mathbb{1}_{(0,+\infty)}(c(\tilde{x},\xi^j))^T(x-\tilde{x})\right) \leq t\alpha, \\ & x \in \mathcal{X}, \end{array} \quad (5.41)$$

where $\tilde{x}$ is the last iteration point. Since problem $\mathrm{CP}_N(\tilde{x})$ is a convex conservative approximation of problem (4.37) for any $\tilde{x} \in \mathcal{X}$. We can use Algorithm SCA to solve problem (4.37).

Algorithm SCA

Step 0. Given $x_0 \in \mathcal{X}$ and set $k := 0$.

Step 1. Stop if x_k satisfies the KKT conditions (4.40).

Step 2. Solve problem $\mathrm{CP}_N(x_k)$ to obtain an optimal solution x_{k+1}.

Step 3. Set $k := k+1$ and go to Step 1.

Under Properties 1-3 in [14], Hong et al show that a sequence of solutions generated by Algorithm SCA converges to a KKT point of problem (4.37).

5.2 Norm optimization problem

Let $x = (x_1, \ldots, x_d)^T$ denote a d-dimensional vector in $\mathbb{R}^d$ and $\xi \in \mathbb{R}^{d\times m}$ be a matrix of random variables. We consider the following norm optimization problem:

$$\begin{array}{ll} \min & -\sum_{j=1}^d x_1 \\ \text{s.t.} & \mathrm{Prob}\left\{\sum_{j=1}^d \xi_{ij}^2 x_j^2 \leq 100, i = 1,\ldots,m\right\} \geq 1-\alpha, \\ & x_j \geq 0, j = 1,\ldots,d. \end{array} \quad (5.42)$$

Note that (5.42) is a joint chance constraint problem, see [14, Section 5.1] for details.

Let $c_i(x,\xi) := \sum_{j=1}^{d} \xi_{ij}^2 x_j^2 - 100$, for $i = 1, \ldots, m$. For any $x \neq 0$, $c_i(x,\xi)$ is a continuous random variable and $c_i(x,\xi) = c_j(x,\xi)$ with probability 0. It is easy to verify that for any small $\varepsilon > 0$, $p(c(x,\xi) - \varepsilon)$ and $p(c(x,\xi))$ satisfy the conditions of Proposition 4.3.

We consider the case when $\xi_{ij}, i = 1, \ldots, m,\ j = 1, \ldots, d$, are independent and identically distributed random variables with stand normal distribution. Via a similar argument to that in [14], we can work out the the optimal solution of (5.42) $x^* = \{x_1^*, \ldots, x_d^*\}$, where

$$x_i^* = \frac{10}{F_{\chi_d^2}^{-1}((1-\alpha)^{1/m})}, i = 1, \ldots, d,$$

and $F_{\chi_d^2}^{-1}$ denotes the inverse distribution function of a chi-square distribution with d degrees of freedom. We use x^* as a benchmark in the CVaR and DC-approximation schemes.

We carry out some numerical experiments on this problem with the SAA scheme (1.6) and (4.37) in Matlab 7.9.0 installed in a PC with Windows XP where the SAA problem is solved by the Matlab built-in optimization solver *fmincon*.

For $d = 2$ and $m = 2$, the true optimal solution is $x^* = (4.10, 4.10)$ with optimal value $f^* = -8.20$. We set $\alpha = 0.1$ and $\epsilon = 0.05^2$ and perform comparative analysis with respect to the sample size from 100 to 2800 with increment 300. For every fixed sample size, 50 independent tests are carried out each of which solves the SAA problem and yields an approximation solution. We solve the problem (5.42) under the two approximation schemes, CVaR and DC, and use the numerical solution of CVaR problem, x_{CVaR}, as a start point of DC problem. We carry out those numerical experiments on this problem with gradient-based Monte Carlo method and Algorithm SCA, see [14], in Matlab 7.9.0 installed in a PC with windows XP where CVaR problem and every convex subproblem $\mathrm{CP}_N(x_k)$ in Algorithm SCA are solved by *fmincon*.

We use a vertical interval to indicate the range of the 50 approximate optimal values and optimal solutions based on CVaR and DC-approximation schemes. As sample size increases, we observe a trend of convergence of the range of the optimal values and solutions. Figures 1 and 2 depict the convergence of optimal values due to CVaR and DC-approximation schemes respectively. Figures 3-4 display the convergence of two components of the approximate optimal solution based on CVaR approximation while Figures 5-6 is for the DC-approximation. All of those figures show that when sample size increases from 100 to 1000, both optimal values and optimal solutions converge quickly and when sample size reaches 1300 there is not substantial changes of these quantities.

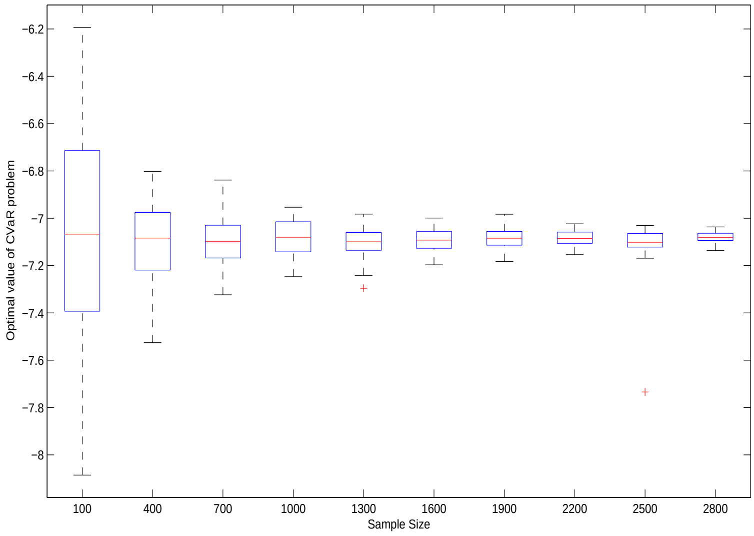

Figure 1: The convergence of the optimal values of CVaR approximation.

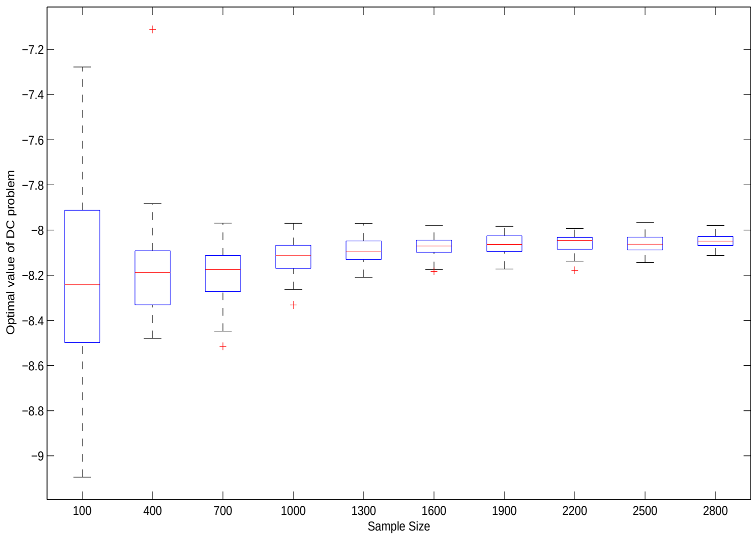

Figure 2: The convergence of the optimal values of DC-approximation.

Acknowledgements. The authors would like to thank Dr Yi Yang for helpful discussions of the algorithm of the DC-approximation method. They would also like to thank two anonymous referees for insightful comments which have substantially helped improve the quality of the paper.

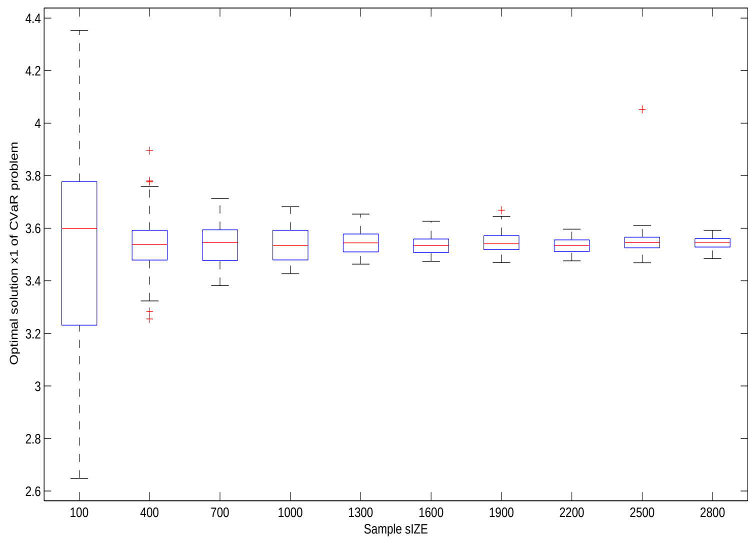

Figure 3: The convergence of the optimal solution $x(1)$ of CVaR approximation.

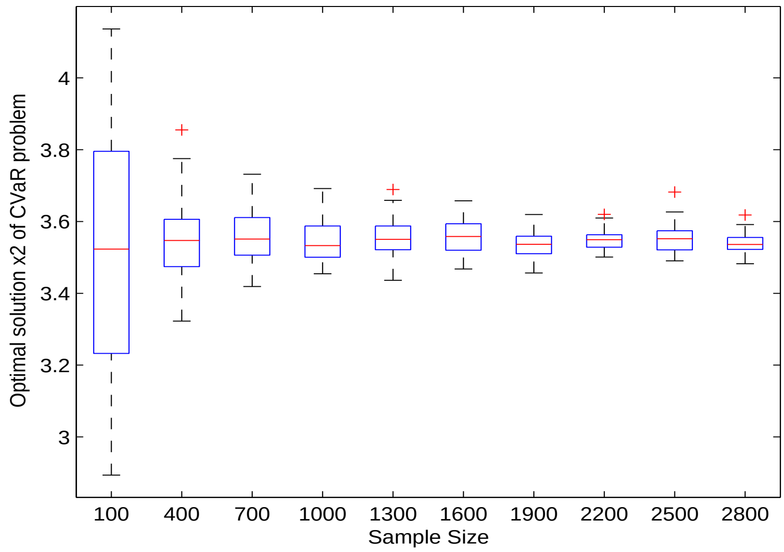

Figure 4: The convergence of the optimal solution $x(2)$ of CVaR approximation.

References

[1] Z. Artstein and R. A. Vitale, A strong law of large numbers for random compact set, *Annals of Probability*, Vol. 3, pp. 879-882, 1975.

[2] R. J. Aumann, Integrals of set-valued functions, *Journal of Mathematical Analysis and Applications*, Vol. 12, pp. 1-12, 1965.

[3] C. Berge, *Espaces topologiques, fonctions multivoques*, Dunod, Paris, 1966.

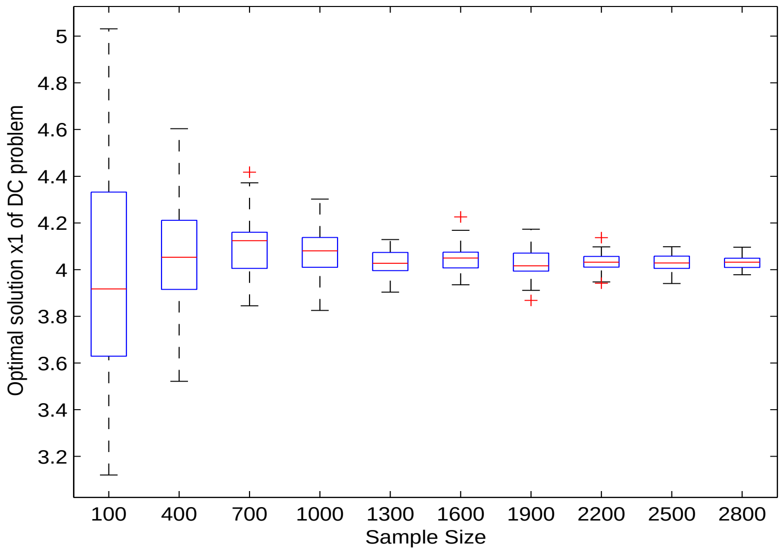

Figure 5: The convergence of the optimal solution $x(1)$ of DC-approximation.

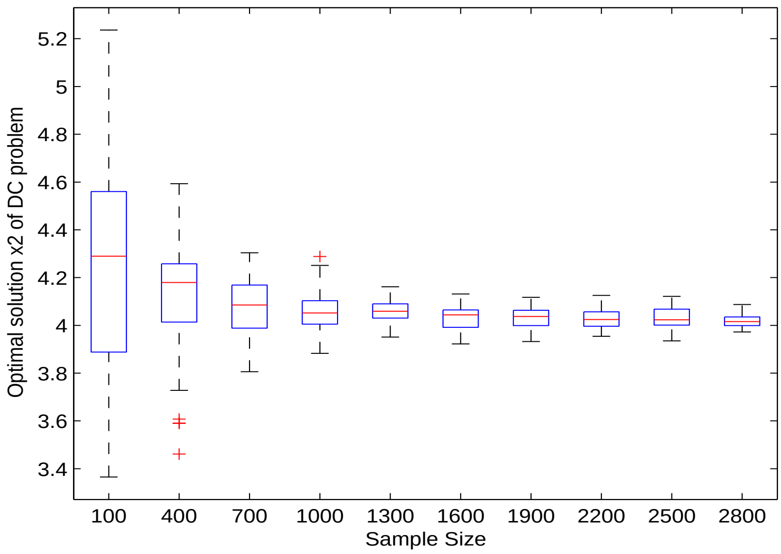

Figure 6: The convergence of the optimal solution $x(2)$ of DC-approximation.

[4] A. Ben-Tal and A. Nemirovski, Robust solutions of linear programming problems contaminated with uncertain data, *Mathematical Programming*, Vol. 88, pp. 411-424, 2000.

[5] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer series in operations research, Springer-Verlag, 2000.

[6] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[7] C. Castaing and M. Valadier, Convex analysis and measurable multifunctions, *Lecture notes in mathematics*, Springer, Berlin, 1977.

[8] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[9] A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications*, Springer-Verlag, New York, 1998.

[10] D. P. de Farias, B. Van Roy, On Constraint Sampling in the Linear Programming Approach to Approximate Dynamic Programming, *Mathematics of Operational Research*, Vol. 29, pp.462-478, 2004.

[11] D. Dentcheva and A. Ruszczyński, Composite semi-infinite optimization, *Control and Cybernetics*, Vol. 36, pp. 1-14, 2007.

[12] C. Hess, Conditional expectation and martingles of random sets, *Pattern Recognition*, Vol. 32, pp. 1543-1567, 1999.

[13] J. B. Hiriart-Urruty, Refinements of nessary optimality conditions in nondifferentiable programming I, *Applied Mathematics and Optimization*, Vol. 5, pp. 63-82, 1979.

[14] L. J. Hong and Y. Yang, and L. Zhang, Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach, *Operations Research*, Vol. 59, pp. 617-630, 2011.

[15] L. J. Hong and L. Zhang, CVaR approximation to chance constrained program: What is lost and how to find it back? Presentation.

[16] F. W. Meng, J. Sun and M. Goh, Stochastic optimization problems with CVaR risk measure and their sample average approximation, *Journal of Optimization Theory and Applications*, Vol. 146, pp. 399-418, 2010.

[17] A. Nemirovski and A. Shapiro, Convex approximations of chance constrained programs, *SIAM Journal on Optimization*, Vol. 17, pp. 969-996, 2006.

[18] N. S. Papageorgiou, On the theory of banach space valued multifunctions. I: Integration and conditional expectation, *Journal of Multivariate Analysis*, Vol. 17, pp. 185–206, 1985.

[19] G. Ch. Pflug, Stochastic optimization and statistical inference, in A. Rusczyński and A.Shapiro, editors, *Stochastic Programming*, Handbooks in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.

[20] L. Qi, A. Shapiro and C. Ling, Differentiability and semismoothness properties of integral functions and their applications, *Mathematical Programming*, Vol. 102, pp. 223 -248, 2005.

[21] D. Ralph and H. Xu, Asympototic analysis of stationary points of sample average two stage stochastic programs: A generalized equation approach. *Mathematics of Operational Research*, Vol. 36, pp. 568-592, 2011.

[22] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[23] R. T. Rockafellar and S. Uryasev, Optimization of conditional value-at-risk, *The Journal of Risk*, Vol. 2, pp. 21-41, 2000.

[24] S. M. Robinson, Analysis of sample-path optimization, *Mathematics of Operations Research*, Vol. 21, pp. 513–528, 1996.

[25] A. Shapiro, Monte Carlo sampling methods, in A. Rusczyński and A.Shapiro, editors, *Stochastic Programming*, Handbooks in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.

[26] A. Shapiro and H. Xu, Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation, *Optimization*, Vol. 57, pp. 395–418, 2008.

[27] W. Wang, Sample average approximation of risk-averse stochastic programs, A thesis presented to the academic faculty, Georgia Institute of Technology, 2007.

[28] H. Sun and H. Xu, A note on uniform exponential convergence of sample average approximation of random functions, *Journal of Mathematical Analysis and Applications*, Vol. 385, pp. 698-708, 2012.

[29] H. Sun and H. Xu, Convergence analysis of stationary points in sample average approximation of stochastic programs with second order stochastic dominance constraints, preprint, School of Mathematics, University of Southampton, Febrary 2011.

[30] P. Terán, On a uniform law of large numbers for random sets and subdifferentials of random functions, *Statistics & Probability Letters*, Vol. 78, pp. 42-49, 2008.

[31] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, *Journal of Mathematical Analysis and Applications*, Vol. 368, pp. 692-710, 2010.

Appendix

In this appendix, we strengthen [28, Theorem 3.1] by weakening a boundedness condition imposed on the random function.

Theorem 5.1 *Let $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real valued lower semicontinuous function, $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ a random vector defined on probability space $(\Omega, \mathcal{F}, P)$ and $\psi(x) := \mathbb{E}[\phi(x,\xi)]$. Let $\mathcal{X} \subset \mathbb{R}^n$ be a compact subset of $\mathbb{R}^n$. Assume: (a) for every $x \in \mathcal{X}$ the moment generating function*

$$M_x(t) := \mathbb{E}\left\{e^{t[\phi(x,\xi)-\psi(x)]}\right\}$$

is finite valued for all t in a neighborhood of zero. (b) $\psi(x)$ is continuous on $\mathcal{X}$, (c) $\phi(x,\xi)$ is bounded by an integrable function $L(\xi)$ and the moment generating function $\mathbb{E}\left[e^{(L(\xi)-\mathbb{E}[L(\xi)])t}\right]$ is finite valued for t close to 0.

(i) If $\phi(\cdot,\xi)$ is almost H-clam from above at every point $x \in \mathcal{X}$ with modulus $\kappa_x(\xi)$ and order γ_x, and the moment generating function $\mathbb{E}[e^{\kappa_x(\xi)t}]$ is finite valued for t close to 0, then for every $\epsilon > 0$, there exist positive constants $c(\epsilon)$ and $\beta(\epsilon)$, independent of N, such that

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}}(\psi_N(x) - \psi(x)) \geq \epsilon\right\} \leq c(\epsilon)e^{-N\beta(\epsilon)}. \tag{5.43}$$

(ii) If $\phi(\cdot,\xi)$ is almost H-clam from below at every point $x \in \mathcal{X}$ with modulus $\kappa_x(\xi)$ and order γ_x, and the moment generating function $\mathbb{E}[e^{\kappa_x(\xi)t}]$ is finite valued for t close to 0, then for every $\epsilon > 0$, there exist positive constants $c(\epsilon)$ and $\beta(\epsilon)$, independent of N, such that

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}}(\psi_N(x) - \psi(x)) \le -\epsilon\right\} \le c(\epsilon)e^{-N\beta(\epsilon)}. \tag{5.44}$$

(iii) If $\phi(\cdot,\xi)$ is almost H-clam at every point $x \in \mathcal{X}$ with modulus $\kappa_x(\xi)$ and order γ_x, and the moment generating function $\mathbb{E}[e^{\kappa_x(\xi)t}]$ is finite valued for t close to 0, then for every $\epsilon > 0$, there exist positive constants $c(\epsilon)$ and $\beta(\epsilon)$, independent of N, such that

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}}|\psi_N(x) - \psi(x)| \ge \epsilon\right\} \le c(\epsilon)e^{-N\beta(\epsilon)}. \tag{5.45}$$

Proof. We only prove part (iii) as part (i) and (ii) can be proved in a similar way.

For given $\epsilon > 0$ and fixed each $x \in \mathcal{X}$, it follows by [28, Lemma 3.1] that there exists $N_0 > 0$ such that for $N > N_0$,

$$\text{Prob}\{|\psi_N(x) - \psi(x)| \ge \epsilon\} \le e^{-NI_x(\epsilon)}, \tag{5.46}$$

where $I_x(\epsilon)$ is positive. Let $\nu > 0$ and $\{\bar{x}_i\}$, $i \in \{1,\dots,M\}$ be a ν-net of $\mathcal{X}$ with $M = [O(1)D/\nu]^n$, where $D := \sup_{x,x'\in\mathcal{X}}||x - x'||$, that is, for any $x \in \mathcal{X}$, there exists an index $i(x) \in \{1,\cdots,M\}$ such that $||x - \bar{x}_{i(x)}|| \le \nu$. Since $\psi(x)$ is assumed to be continuous on $\mathcal{X}$ which is a compact set, we can choose the ν-net through the finite covering theorem such that

$$|\psi(x) - \psi(\bar{x}_{i(x)})| \le \frac{\epsilon}{4} \tag{5.47}$$

for any $x \in \mathcal{X}$. On the other hand, since $\phi(x,\xi)$ is almost H-clam at x, there exist an open set $\Delta_x(\epsilon) \subset \Xi$, positive numbers $\delta_x(\epsilon)$ and γ_x such that

$$\int_{\xi\in\Delta_x(\epsilon)} dP(\xi) \le \epsilon$$

and

$$|\phi(x',\xi) - \phi(x,\xi)| \le \kappa(\xi)||x' - x||^{\gamma_x}$$

for all $\xi \notin \Delta_x(\epsilon)$ and $\|x' - x\| \le \delta_x(\epsilon)$. Let

$$\kappa(\xi) = \max_{i=1,\cdots,M} \kappa_{x_i}(\xi)$$

and

$$\gamma = \min_{i=1,\cdots,M} \gamma_{x_i},$$

and assuming without lost of generality that $\nu < 1$, we have

$$|\phi(x,\xi) - \phi(\bar{x}_{i(x)},\xi)| \le \kappa(\xi)||x - \bar{x}_{i(x)}||^{\gamma} \tag{5.48}$$

for all $\xi \notin \Delta_{\bar{x}_{i(x)}}(\epsilon)$ and $x \in \mathcal{X}$. Note that since the probability measure is absolutely continuous, we may choose $\Delta_{x_i}(\epsilon)$ such that

$$\max_{i=1,\cdots,M} \int_{\xi\in\Delta_{x_i}(\epsilon)} L(\xi)dP(\xi) \le \frac{\epsilon}{16}. \tag{5.49}$$

Let

$$\tilde{\psi}_N(x) := \frac{1}{N}\sum_{\xi^k\in\Delta_{\bar{x}_{i(x)}}(\epsilon)} \phi(x,\xi^k),\ \ \bar{\psi}_N(x) := \frac{1}{N}\sum_{\xi^k\notin\Delta_{\bar{x}_{i(x)}}(\epsilon)} \phi(x,\xi^k).$$

By (5.48), we have that

$$|\bar{\psi}_N(x)-\bar{\psi}_N(\bar{x}_{i(x)})| \le \frac{1}{N}\sum_{\xi^k\notin\Delta_{\bar{x}_{i(x)}}(\epsilon)} \kappa(\xi^k)\nu^\gamma \le \kappa^N\nu^\gamma \tag{5.50}$$

where $\kappa^N := \frac{1}{N}\sum_{k=1}^N \kappa(\xi^k)$. Since $\mathbb{E}[e^{\kappa(\xi)t}]$ is finite valued for t close to 0, by Cramér's large deviation Theorem [9], we have that for any $L' > \mathbb{E}[\kappa(\xi(\omega))]$, there exists a positive constant λ such that

$$\text{Prob}\big\{\kappa^N \ge L'\big\} \le e^{-N\lambda}$$

and hence

$$\text{Prob}\{\kappa^N\nu^\gamma \ge \frac{\epsilon}{4}\} \le e^{-N\lambda} \tag{5.51}$$

for some $\lambda > 0$ (by setting $\frac{\epsilon}{4\nu^\gamma} > \mathbb{E}[\kappa(\xi(\omega))]$). On the other hand, by condition (c) of this theorem, we have

$$\begin{aligned} |\tilde{\psi}_N(x)-\tilde{\psi}_N(\bar{x}_{i(x)})| &\le \frac{1}{N}\sum_{\xi^k\in\Delta_{\bar{x}_{i(x)}}(\epsilon)} |\phi(x,\xi^k)-\phi(\bar{x}_{i(x)},\xi^k)| = \frac{1}{N}\sum_{k=1}^N |\phi(x,\xi^k)-\phi(\bar{x}_{i(x)},\xi^k)|\eta^k \\ &\le \frac{1}{N}\sum_{k=1}^N 2L(\xi^k)\eta^k, \end{aligned} \tag{5.52}$$

where

$$\eta(\xi) := \begin{cases} 1, & \text{if } \xi\in\Delta_{\bar{x}_{i(x)}}(\epsilon), \\ 0, & \text{if } \xi\notin\Delta_{\bar{x}_{i(x)}}(\epsilon), \end{cases}$$

and

$$\eta^k := \begin{cases} 1, & \text{if } \xi^k\in\Delta_{\bar{x}_{i(x)}}(\epsilon), \\ 0, & \text{if } \xi^k\notin\Delta_{\bar{x}_{i(x)}}(\epsilon). \end{cases}$$

Applying [28, Proposition 3.1 (ii)] to $L(\xi)\eta(\xi)$, we have

$$\text{Prob}\left\{\frac{1}{N}\sum_{k=1}^N L(\xi^k)\eta^k - \mathbb{E}[L(\xi)\eta(\xi)] \ge \frac{\epsilon}{16}\right\} \le e^{-N\tilde{I}(\epsilon/16)} + e^{-N\tilde{I}(-\epsilon/16)},$$

where $\tilde{I}(z)$ is the rate function of $L(\xi)\eta(\xi)$, and by [28, Proposition 3.1 (ii)], $\tilde{I}(-\epsilon/16) > 0$ and $\tilde{I}(\epsilon/16) > 0$. On the other hand, by (5.49),

$$\mathbb{E}[2L(\xi)\eta(\xi)] = 2\int_{\xi\in\Delta_{\bar{x}_i}(\epsilon)} L(\xi)d\mu(\xi) \le \frac{\epsilon}{8}.$$

A combination of the above two inequalities yields

$$\text{Prob}\left\{\frac{1}{N}\sum_{k=1}^N 2L(\xi^k)\eta^k \ge \frac{\epsilon}{4}\right\} \le e^{-N\tilde{I}(\epsilon/16)} + e^{-N\tilde{I}(-\epsilon/16)} \tag{5.53}$$

for N sufficiently large.

Let $Z_i := |\psi_N(\bar{x}_i) - \psi(\bar{x}_i)|, i = 1, \ldots, M$. The event $\{\max_{1\le i\le M} Z_i \ge \epsilon\}$ is equal to the union of the events $\{Z_i \ge \epsilon\}, i = 1, \ldots, M$, and hence

$$\text{Prob}\left\{\max_{1\le i\le M} Z_i \ge \epsilon\right\} \le \sum_{i=1}^{M} \text{Prob}\left\{Z_i \ge \epsilon\right\}.$$

Together with (5.46), this implies that

$$\text{Prob}\left\{\max_{1\le i\le M} Z_i \ge \epsilon\right\} \le \sum_{i=1}^{M} e^{-NI_{\bar{x}_i}(\epsilon)}. \tag{5.54}$$

Combining (5.47), (5.50), (5.52) and (5.54), we obtain

$$\begin{aligned}
|\psi_N(x) - \psi(x)| &\le |\psi_N(x) - \psi_N(\bar{x}_{i(x)})| + |\psi_N(\bar{x}_{i(x)} - \psi(\bar{x}_{i(x)})| + |\psi(\bar{x}_{i(x)}) - \psi(x)| \\
&\le |\bar{\psi}_N(x) - \bar{\psi}_N(\bar{x}_{i(x)})| + |\tilde{\psi}_N(x) - \tilde{\psi}_N(\bar{x}_{i(x)})| + |\psi_N(\bar{x}_{i(x)}) - \psi(\bar{x}_{i(x)})| + \frac{\epsilon}{4} \\
&\le \kappa^N \nu^\gamma + \frac{1}{N}\sum_{k=1}^{N} 2L(\xi^k)\eta^k + |\psi_N(\bar{x}_{i(x)}) - \psi(\bar{x}_{i(x)})| + \frac{\epsilon}{4}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\text{Prob}\left\{\sup_{x\in\mathcal{X}} |\psi_N(x) - \psi(x)| \ge \epsilon\right\} \\
&\quad \le \text{Prob}\left\{\kappa^N \nu^\gamma + \frac{1}{N}\sum_{k=1}^{N} 2L(\xi^k)\eta^k + \max_{1\le i\le M} \left|\psi_N(\bar{x}_i) - \psi(\bar{x}_i)\right| \ge \frac{3\epsilon}{4}\right\}.
\end{aligned}$$

By (5.51), (5.53) and (5.54), we have

$$\begin{aligned}
\text{Prob}\left\{\sup_{x\in\mathcal{X}} |\psi_N(x) - \psi(x)| \ge \epsilon\right\} &\le e^{-N\lambda} + \text{Prob}\{\frac{1}{N}\sum_{k=1}^{N} 2L(\xi^k)\eta^k \ge \frac{\epsilon}{4}\} \\
&\quad + \text{Prob}\left\{\max_{1\le i\le M} \left|\psi_N(\bar{x}_i) - \psi(\bar{x}_i)\right| \ge \frac{\epsilon}{4}\right\} \\
&\le e^{-N\lambda} + e^{-N\tilde{I}(\epsilon/16)} + e^{-N\tilde{I}(-\epsilon/16)} + \sum_{i=1}^{M} e^{-NI_{\bar{x}_i}(\frac{\epsilon}{4})},
\end{aligned}$$

which implies (5.45) as above choice of ν-net does not depend on the sample (although it depends on ϵ), and $I_{\bar{x}_i}(\frac{\epsilon}{4})$ are positive, for $i = 1, \ldots, M$. The proof is complete. ■

Note that the exponential convergence is derived for the case when ξ satisfies a continuous distribution. In the case when ξ satisfies a discrete distribution, the concept of almost H-calmness is no longer applicable. However the uniform exponential convergence may be established in an entirely different way for a class of random function which is uniformly bounded over a considered compact set. We leave this to interested readers.