



Sample average approximation of conditional value-at-risk based variational inequalities

Ashish Cherukuri¹

Received: 23 August 2022 / Accepted: 17 March 2023 / Published online: 6 April 2023
© The Author(s) 2023

Abstract

This paper focuses on a class of variational inequalities (VIs), where the map defining the VI is given by the component-wise conditional value-at-risk (CVaR) of a random function. We focus on solving the VI using sample average approximation, where solutions of the VI are estimated with solutions of a sample average VI that uses empirical estimates of the CVaRs. We establish two properties for this scheme. First, under continuity of the random map and the uncertainty taking values in a bounded set, we prove asymptotic consistency, establishing almost sure convergence of the solution of the sample average problem to the true solution. Second, under the additional assumption of random functions being Lipschitz, we prove exponential convergence where the probability of the distance between an approximate solution and the true solution being smaller than any constant approaches unity exponentially fast. The exponential decay bound is refined for the case where random functions have a specific separable form in the decision variable and uncertainty. We adapt these results to the case of uncertain routing games and derive explicit sample guarantees for obtaining a CVaR-based Wardrop equilibria using the sample average procedure. We illustrate our theoretical findings by approximating the CVaR-based Wardrop equilibria for a modified Sioux Falls network.

Keywords Variational inequalities · Sample average approximation · Conditional value-at-risk · Wardrop equilibrium

1 Introduction

Consider the following variational inequality problem $\text{VI}(\mathcal{X}, F)$: find $x^* \in \mathcal{X}$ such that

✉ Ashish Cherukuri
a.k.cherukuri@rug.nl

¹ ENTEG, University of Groningen, Groningen, The Netherlands

$$(x - x^*)^\top F(x^*) \geq 0, \quad \text{for all } x \in \mathcal{X}, \quad (1)$$

where $\mathcal{X} \subset \mathbb{R}^n$ is a compact set and each component $i \in \{1, 2, \dots, n\}$ of the map $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$, denoted $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$, is given by

$$F_i(x) = \text{CVaR}_\alpha^\mathbb{P}[f_i(x, u)]. \quad (2)$$

In the above equation, $f_i : \mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R}$ is referred to as the random function, the set $\mathcal{U} \subset \mathbb{R}^m$ is compact, and \mathbb{P} is the distribution of u supported over the set \mathcal{U} . We assume f_i is continuous. The map F_i gives the conditional value-at-risk (CVaR) at level $\alpha \in (0, 1)$ of the random function f_i . The CVaR computes the tail expectation of the underlying random variable [18] and can be determined by the following optimization

$$\text{CVaR}_\alpha^\mathbb{P}[f_i(x, u)] = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{\alpha} \mathbb{E}_\mathbb{P}[f_i(x, u) - t]_+ \right\}, \quad (3)$$

where $\mathbb{E}_\mathbb{P}$ is the expectation under the distribution \mathbb{P} and the operator $[\cdot]_+$ gives the positive part, i.e., $[v]_+ = \max\{0, v\}$. The parameter α characterizes the risk-averseness. When α is close to unity, the decision-maker is risk-neutral, whereas, α close to the origin implies high risk-averseness. The main purpose of the paper is to analyze the statistical properties of a sample average approximation (SAA) scheme for solving the variational inequality $\text{VI}(\mathcal{X}, F)$ given in (1). The set of solutions of this problem is denoted by $\text{SOL}(\mathcal{X}, F)$.

Variational inequality problems defined using a set of random functions is surveyed in [17]. The most widely studied VI problem in this context, termed stochastic variational inequalities (SVIs), is the one where the map defining the VI is the expectation of a random function. Risk-based VIs, where the VI map is given as the risk of a random function, naturally generalize the setup of SVI and find application in finding the Wardrop equilibria in a network routing problem where users are risk-averse. While several works explore sample average schemes for SVIs, there is no such study for risk-averse VIs. This paper aims to fill this gap.

Early investigations on statistical aspects of SAA for generalized equations and SVIs appeared in [8, 9], respectively. These works focused on asymptotic properties of the SAA schemes, that is, consistency of estimators and their asymptotic distributions. The former is concerned with showing the convergence with probability one of solutions of the SAA to solutions of the original problem as the sample size tends to infinity. The latter determines the distribution of the approximate solutions in the asymptotic limit. While these properties show the limiting behavior, they do not illustrate the guarantees in the finite-sample regime. This feature was explored in [15, 19, 27, 28] where it was shown that for generalized equilibrium problems under various set of assumptions, one can demonstrate exponential convergence of the approximate solutions. Meaning that the probability that the SAA solution is a fixed distance away from the original solution decays exponentially as the sample size tends to infinity. Technically, establishing such a property relies on conducting sensitivity analysis for the VI and then combining it with uniform large deviation bounds on random functions. All these studies share the common property that the

underlying map is the expectation of the random function, while in this paper we look at CVaR-based maps.

The works [1, 11, 21] study SAA of CVaR in the context of stochastic optimization problems, where CVaR is either being minimized or used to define the constraints. In [11, 21] asymptotic consistency and exponential convergence of Karush-Kuhn-Tucker (KKT) points of the sample average optimization problem to that of the true one was established. In [1], the SAA of CVaR is used to approximate the solution of risk-constrained optimization problem. Since CVaR is used to define a VI problem in our case, the analysis does not follow directly from these existing results. Moreover, as opposed to the general large deviation bounds provided in these works, the exponential bounds derived here are explicit without involving ambiguous constants. In another data-based approach [16], the CVaR is perceived as the expected shortfall and desirable statistical guarantees are obtained for the optimizers of its sample average.

One of the motivations for our work is to approximate the Wardrop equilibrium for a network routing problem where agents choose paths that have minimum risk. Such a setting was extensively studied in [13] where various notions of equilibrium and related computational aspects of finding them were discussed. Among other works that consider risk, [12, 14] assume the cost of each path to be the weighted sum of the mean and the variance of the uncertain cost. However, none of these works focus on CVaR-based routing. In the transportation literature, the CVaR-based equilibrium is also known as the mean excess traffic equilibrium, see e.g., [2, 29] and references therein. While these works have explored numerous algorithms for computing the equilibrium, they lack theoretical performance guarantees for sample-based solutions.

For analyzing the SAA of (1), we assume that a certain number of independent and identically distributed samples of the random variable u are available using which the expectation operator in the definition of the CVaR is replaced with its sample average. The resulting empirical CVaR gives rise to a set of functions that are sample average versions of F . Using these, we define a sample average variational inequality. Our contributions are as follows:

- (i) We establish asymptotic consistency of the sample average scheme, that is, the set of solutions of the sample average VI converge almost surely, in a set-valued sense, to the set $\text{SOL}(\mathcal{X}, F)$.
- (ii) Under the assumption that random functions are uniformly Lipschitz continuous in x , we show exponential convergence of the solution set of the sample average VI to the set $\text{SOL}(\mathcal{X}, F)$. That is, given any constant, the probability that the distance of a solution of the sample average problem from the set $\text{SOL}(\mathcal{X}, F)$ is less than that constant approaches unity exponentially with the number of samples.
- (iii) We give tighter sample guarantees with explicit expression for the coefficient in the exponential bound for a particular class of separable random functions.
- (iv) We illustrate the application of the derived approximations in computing a CVaR-based Wardrop equilibrium for a network routing problem that is defined using uncertain costs.

A preliminary version of the paper appeared as [3], where the focus was finding the Wardrop equilibrium problem for a network routing problem. As compared to it, the present article has a more general problem setup focusing not just on a Wardrop equilibrium problem, but on a general VI. In addition, the tighter sample guarantees for separable functions given in Sect. 4 are new here and the simulation example is much more elaborate.

Notation Let \mathbb{R} , $\mathbb{R}_{\geq 0}$, $\mathbb{R}_{> 0}$, and \mathbb{N} denote the set of real, nonnegative real, positive real, and natural numbers, respectively. Let $\|\cdot\|$ denote the Euclidean 2-norm. We use $[N] := \{1, \dots, N\}$ for positive integer N . For $x \in \mathbb{R}$, we let $[x]_+ = \max(x, 0)$ and $\lceil x \rceil$ be the smallest integer greater than or equal to x . The cardinality of a set \mathcal{S} is denoted by $|\mathcal{S}|$. The distance of a point $x \in \mathbb{R}^m$ to a set $\mathcal{S} \subset \mathbb{R}^m$ is denoted by $\text{dist}(x, \mathcal{S}) := \inf_{y \in \mathcal{S}} \|x - y\|$. The *deviation* of a set $\mathcal{A} \subset \mathbb{R}^m$ from \mathcal{S} is $\mathbb{D}(\mathcal{A}, \mathcal{S}) := \sup_{y \in \mathcal{A}} \text{dist}(y, \mathcal{S})$.

2 Preliminaries

Here we collect relevant mathematical background used throughout the paper.

2.1 Variational inequality

Given a map $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and a closed set $\mathcal{X} \subset \mathbb{R}^n$, the *variational inequality* (VI) problem, denoted $\text{VI}(\mathcal{X}, F)$, involves finding $x^* \in \mathcal{X}$ such that $(x - x^*)^\top F(x^*) \geq 0$ for all $x \in \mathcal{X}$. Such a point is called a *solution* of the VI problem. The set of solutions of $\text{VI}(\mathcal{X}, F)$ are denoted by $\text{SOL}(\mathcal{X}, F)$. The map F is *monotone* on the set \mathcal{X} if $(F(x) - F(x'))^\top (x - x') \geq 0$ for all $x, x' \in \mathcal{X}$. The map F is *strictly monotone* on \mathcal{X} if this inequality is strict for $x \neq x'$. Finally, F is *strongly monotone* on \mathcal{X} with modulus $\sigma > 0$ if $(F(x) - F(x'))^\top (x - x') \geq \sigma \|x - x'\|^2$ for all $x, x' \in \mathcal{X}$. If F is either strictly or strongly monotone, then $\text{SOL}(\mathcal{X}, F)$ is singleton.

2.2 Uniform convergence

A sequence of functions $\{f_N : \mathcal{X} \rightarrow \mathcal{Y}\}_{N=1}^\infty$, where \mathcal{X} and \mathcal{Y} are Euclidean spaces, is said to *converge uniformly* on a set $X \subset \mathcal{X}$ to $f : \mathcal{X} \rightarrow \mathcal{Y}$ if for any $\epsilon > 0$, there exists $N_\epsilon \in \mathbb{N}$ such that $\sup_{x \in X} \|f_N(x) - f(x)\| \leq \epsilon$, for all $N \geq N_\epsilon$. Similar definition applies for convergence in probability. That is, consider a random sequence of functions $\{f_N^\omega : \mathcal{X} \rightarrow \mathcal{Y}\}_{N=1}^\infty$ defined on a probability space (Ω, \mathcal{F}, P) . The sequence is said to *converge uniformly* to $f : \mathcal{X} \rightarrow \mathcal{Y}$ on X *almost surely* (shorthand, a.s.) if $f_N^\omega \rightarrow f$ uniformly on X for almost all $\omega \in \Omega$.

2.3 Risk measures

Next we review notions on value-at-risk and CVaR from [18]. Given a real-valued random variable Z with probability distribution \mathbb{P} , we denote the *cumulative distribution function* by $H_Z(\zeta) := \mathbb{P}(Z \leq \zeta)$. The *left-side α -quantile* of Z is defined as $H_Z^{-1}(\alpha) := \inf\{\zeta \mid H_Z(\zeta) \geq \alpha\}$. Given a probability level $\alpha \in (0, 1)$, the *value-at-risk* of Z at level α , denoted $\text{VaR}_\alpha^\mathbb{P}[Z]$, is the left-side $(1 - \alpha)$ -quantile of Z . Formally,

$$\text{VaR}_\alpha^\mathbb{P}[Z] := H_Z^{-1}(1 - \alpha) = \inf\{\zeta \mid \mathbb{P}(Z \leq \zeta) \geq 1 - \alpha\}.$$

The *CVaR*, also referred to as the average value-at-risk in [18], of Z at level α , denoted $\text{CVaR}_\alpha^\mathbb{P}[Z]$, is given as

$$\text{CVaR}_\alpha^\mathbb{P}[Z] = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{\alpha} \mathbb{E}[Z - t]_+ \right\}. \tag{4}$$

Under the continuity of the cumulative distribution function at $\text{VaR}_\alpha^\mathbb{P}[Z]$, we have that $\text{CVaR}_\alpha^\mathbb{P}[Z]$ is the expectation of Z when it takes values bigger than $\text{VaR}_\alpha^\mathbb{P}[Z]$. That is, $\text{CVaR}_\alpha^\mathbb{P}[Z] := \mathbb{E}[Z \geq \text{VaR}_\alpha^\mathbb{P}[Z]]$. The parameter α characterizes the risk-averseness. When α is close to unity, the decision-maker is risk-neutral, whereas, α close to the origin implies high risk-averseness. The minimum in (4) is attained at a point in the interval $[t^m, t^M]$, where $t^m := \inf\{\zeta \mid H_Z(\zeta) \geq 1 - \alpha\}$, and $t^M := \sup\{\zeta \mid H_Z(\zeta) \leq 1 - \alpha\}$.

3 Sample average approximation of VI(\mathcal{X}, F)

The approach in the sample average framework is to replace the expectation operator in any problem with the average over the obtained samples [18]. This is one of the main Monte Carlo methods for problems with expectations; see [5] for a detailed survey of other sample-based techniques. In our setup, for each component F_i , we will replace the expectation operator in the definition of the CVaR in (3) with the sample average. The thus formed set of functions result in a VI problem that approximates $\text{VI}(\mathcal{X}, F)$.

Note that the map F is continuous since $f_i, i \in [n]$ are so and \mathcal{X} and \mathcal{U} are compact. One can reason this fact using arguments similar to those of the proof of Lemma 4. As a consequence of the continuity of F , the set of solutions $\text{SOL}(\mathcal{X}, F)$ of the problem $\text{VI}(\mathcal{X}, F)$ is nonempty and compact [7, Corollary 2.2.5]. For convenience, we use the notation $\mathcal{S} = \text{SOL}(\mathcal{X}, F)$.

Let $\hat{\mathcal{U}}^N := \{\hat{u}^1, \hat{u}^2, \dots, \hat{u}^N\}$ be the set of $N \in \mathbb{N}$ independent and identically distributed samples of u drawn from \mathbb{P} . Then, the sample average approximation of the CVaR associated to component $i \in [n]$ is

$$\widehat{\text{CVaR}}_\alpha^N [f_i(x, u)] := \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{N\alpha} \sum_{j=1}^N [f_i(x, \hat{u}^j) - t]_+ \right\}. \tag{5}$$

The above expression is also known as the empirical estimate of the CVaR, or empirical CVaR in short. The expression is also the CVaR of the random function at level α under the empirical distribution $\frac{1}{N} \sum_{j=1}^N \delta_{\hat{u}^j}$, where $\delta_{\hat{u}^j}$ is the unit point mass at \hat{u}^j . Note that $\widehat{\text{CVaR}}_\alpha^N$ is random as it depends on the realization $\hat{\mathcal{U}}^N$ of the random variable. To emphasize this dependency, we represent with $\hat{\cdot}^N$ entities that are random. Using (5) as the approximate function, define the sample average VI problem as $\text{VI}(\mathcal{X}, \hat{F}^N)$, where

$$\hat{F}_i^N(x) := \widehat{\text{CVaR}}_\alpha^N [f_i(x, u)],$$

for all $i \in [n]$. We denote the set of solutions of $\text{VI}(\mathcal{X}, \hat{F}^N)$ by $\hat{\mathcal{S}}^N \subset \mathcal{X}$. This serves as a reminder that it approximates \mathcal{S} . The notion of approximation is made precise next. Note that $\hat{\mathcal{S}}^N$ is nonempty as \mathcal{X} is compact and \hat{F}^N is continuous.

Definition 1 (*Asymptotic consistency and exponential convergence*) The set $\hat{\mathcal{S}}^N$ is an asymptotically consistent approximation of \mathcal{S} , or in short, $\hat{\mathcal{S}}^N$ is asymptotically consistent, if any sequence of solutions $\{\hat{x}^N \in \hat{\mathcal{S}}^N\}_{N=1}^\infty$ has almost surely (a.s.) all accumulation points in \mathcal{S} . The set $\hat{\mathcal{S}}^N$ is said to converge exponentially to \mathcal{S} if for any $\epsilon > 0$, there exist positive constants c_ϵ and δ_ϵ such that for any sequence $\{\hat{x}^N \in \hat{\mathcal{S}}^N\}_{N=1}^\infty$, the following holds

$$\mathbb{P}^N \left(\text{dist}(\hat{x}^N, \mathcal{S}) \leq \epsilon \right) \geq 1 - c_\epsilon e^{-\delta_\epsilon N} \tag{6}$$

for all $N \in \mathbb{N}$. •

The asymptotic consistency of $\hat{\mathcal{S}}^N$ is equivalent to saying $\mathbb{D}(\hat{\mathcal{S}}^N, \mathcal{S}) \rightarrow 0$ a.s. as $N \rightarrow \infty$. The expression (6) gives a precise rate for this convergence. In our work, all convergence results are for $N \rightarrow \infty$ and so we drop restating this fact for convenience’s sake. In the following sections, we will establish the asymptotic consistency and the exponential convergence of $\hat{\mathcal{S}}^N$.

3.1 Asymptotic consistency of $\hat{\mathcal{S}}^N$

We begin with stating the bound on the optimizers of the problem defining the CVaR (3) and the empirical CVaR (5). This restricts our attention to compact domains for variables (x, t, u) , a property useful in showing consistency. Denote for each $i \in [n]$, functions

$$\psi_i(x, t) := t + \frac{1}{\alpha} \mathbb{E}_{\mathbb{P}} [f_i(x, u) - t]_+, \tag{7a}$$

$$\widehat{\psi}_i^N(x, t) := t + \frac{1}{N\alpha} \sum_{j=1}^N [f_i(x, \widehat{u}^j) - t]_+. \tag{7b}$$

The map $\widehat{\psi}_i^N$ is the sample average of ψ_i . Given our assumption that \mathcal{U} is compact, we have that the expected value of f_i is bounded for any $x \in \mathcal{X}$. Using this fact, one can deduce by strong law of large numbers [6] that for any fixed $(x, t) \in \mathcal{X} \times \mathbb{R}$, $\widehat{\psi}_i^N(x, t) \rightarrow \psi_i(x, t)$ a.s. We however require uniform convergence of these maps to conclude consistency, which will be established in Theorem 1 below. Observe that, by definition, $\text{CVaR}_{\alpha}^{\mathbb{P}}[f_i(x, u)] = \inf_{t \in \mathbb{R}} \psi_i(x, t)$ and $\widehat{\text{CVaR}}_{\alpha}^N[f_i(x, u)] = \inf_{t \in \mathbb{R}} \widehat{\psi}_i^N(x, t)$. The following result gives explicit bounds on the optimizers of these problems.

Lemma 1 (Bounds on optimizers of problems defining (empirical) CVaR) *For any $x \in \mathcal{X}$ and $i \in [n]$, the optimizers of the problems in (3) and (5) exist and belong to the compact set $\mathcal{T} = [\ell, L]$, where $\ell := \min\{f_i(x, u) \mid x \in \mathcal{X}, u \in \mathcal{U}, i \in [n]\}$ and $L := \max\{f_i(x, u) \mid x \in \mathcal{X}, u \in \mathcal{U}, i \in [n]\}$. Furthermore, the set of functions*

$$\phi_i(x, t, u) := t + \frac{1}{\alpha} [f_i(x, u) - t]_+, \tag{8}$$

for $i \in [n]$, satisfy for all $(x, t, u) \in \mathcal{X} \times \mathcal{T} \times \mathcal{U}$,

$$\phi_i(x, t, u) \in \left[\ell, \ell + \frac{L - \ell}{\alpha} \right]. \tag{9}$$

Proof From [18, Chapter 6], optimizers of (3) and (5) exist and they lie in the closed interval defined by the left- and the right-side $(1 - \alpha)$ -quantile of the respective random variables. Hence, they belong to \mathcal{T} . To conclude (9), notice

$$\begin{aligned} \phi_i(x, t, u) &= t + \frac{1}{\alpha} [f_i(x, u) - t]_+ \leq t + \frac{1}{\alpha} [L - t]_+ \\ &= t + \frac{1}{\alpha} (L - t) = (1 - \frac{1}{\alpha})t + \frac{1}{\alpha} L \\ &\leq (1 - \frac{1}{\alpha})\ell + \frac{1}{\alpha} L. \end{aligned}$$

Here, the first inequality follows from the bound on f_i , the first equality is because $t \in [\ell, L]$, and the second inequality is due to $\alpha < 1$. Similarly, for the lower bound, $\phi_i(x, t, u) \geq t + \frac{1}{\alpha} [\ell - t]_+ = t \geq \ell$. This completes the proof. □

We make a note here that optimizers of problems defining the CVaR in (3) and (5) exist and are bounded for more general cases, even when the support of the random variable is unbounded, see e.g., [18, Chapter 6]. Nevertheless, the above result provides an explicit bound which is used later in deriving precise exponential convergence guarantees.

As a consequence of Lemma 1, one can show uniform convergence of $\widehat{\psi}_p^N$ to ψ_p . Our next step is to analyze the sensitivity of F as one perturbs the underlying map ψ . In combination with the uniform convergence of $\widehat{\psi}_p^N$, this result leads to the uniform convergence of \widehat{F}^N to F .

Lemma 2 (Sensitivity of F with respect to ψ) *For any $\epsilon > 0$, if we have $\sup_{i \in [n], (x,t) \in \mathcal{X} \times \mathcal{T}} |\widehat{\psi}_i^N(x, t) - \psi_i(x, t)| \leq \epsilon$, where \mathcal{T} is defined in Lemma 1, then $\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| \leq \sqrt{n}\epsilon$.*

Proof The first step is to show the sensitivity of the map $\text{CVaR}_\alpha^{\mathbb{P}}[f_i(\cdot, u)]$ with respect to ψ_p . To this end, fix $i \in [n]$ and $x \in \mathcal{X}$, and let $\widehat{t}_i^N(x) \in \underset{t \in \mathbb{R}}{\text{argmin}} \widehat{\psi}_i^N(x, t)$ and $t_i(x) \in \underset{t \in \mathbb{R}}{\text{argmin}} \psi_i(x, t)$. These optimizers exist due to Lemma 1. We now have

$$\psi_i(x, t_i(x)) - \epsilon \leq \psi_i(x, \widehat{t}_i^N(x)) - \epsilon \leq \widehat{\psi}_i^N(x, \widehat{t}_i^N(x)).$$

The first inequality is due to optimality and the second inequality holds by assumption. Similarly, one can show that

$$\widehat{\psi}_i^N(x, \widehat{t}_i^N(x)) - \epsilon \leq \psi_i(x, t_i(x)).$$

The above two sets of inequalities along with the fact that $\widehat{\text{CVaR}}_\alpha^N[f_i(x, u)] = \widehat{\psi}_i^N(x, \widehat{t}_i^N(x))$ and $\text{CVaR}_\alpha^{\mathbb{P}}[f_i(x, u)] = \psi_i(x, t_i(x))$ lead to the conclusion

$$\sup_{x \in \mathcal{X}} \left| \widehat{\text{CVaR}}_\alpha^N[f_i(x, u)] - \text{CVaR}_\alpha^{\mathbb{P}}[f_i(x, u)] \right| \leq \epsilon. \tag{10}$$

Finally, the conclusion follows from the inequality $\|\widehat{F}^N(x) - F(x)\| \leq \sqrt{n} \sup_{i \in [n]} |\widehat{F}_i^N(x) - F_i(x)|$ that holds for all $x \in \mathcal{X}$. □

The final preliminary result states proximity of $\widehat{\mathcal{S}}^N$ to \mathcal{S} given that the difference between \widehat{F}^N and F is bounded. The proof is a consequence of [27, Lemma 2.1] that studies sensitivity of generalized equations.

Lemma 3 (Sensitivity of \mathcal{S} with respect to F) *For any $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that $\mathbb{D}(\widehat{\mathcal{S}}^N, \mathcal{S}) \leq \epsilon$ whenever $\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| \leq \delta(\epsilon)$.*

Next is the main result of this section, establishing the asymptotic consistency of $\widehat{\mathcal{S}}^N$.

Theorem 1 (Asymptotic consistency of $\widehat{\mathcal{S}}^N$) *We have $\mathbb{D}(\widehat{\mathcal{S}}^N, \mathcal{S}) \rightarrow 0$ almost surely.*

Proof Consider first the a.s. uniform convergence $\widehat{\psi}_i^N \rightarrow \psi_i$ over the compact set $\mathcal{X} \times \mathcal{T}$. Note that $\psi_i(x, t) = \mathbb{E}_{\mathbb{P}}[\phi_i(x, t, u)]$ where ϕ_i is given in (8) and so, $\widehat{\psi}_i^N$ is the sample average of ψ_i . For any fixed $u \in \mathcal{U}$, the map $\phi_i(\cdot, \cdot, u)$ is continuous and for any $(x, t) \in \mathcal{X} \times \mathcal{T}$, due to Lemma 1, the map $\phi_i(x, t, \cdot)$ is dominated by the integrable function (a constant in this case) $\ell + \frac{L-\ell}{\alpha}$. Hence, by the uniform law of large numbers result [18, Theorem 7.48], we conclude that $\widehat{\psi}_i^N \rightarrow \psi_i$ uniformly a.s. on $\mathcal{X} \times \mathcal{T}$. Using this fact in the sensitivity result of Lemma 2 implies that $\widehat{F}^N \rightarrow F$ uniformly a.s. on the set \mathcal{X} . Finally, we arrive at the conclusion using Lemma 3. \square

3.2 Exponential convergence of \widehat{S}^N

Here, our strategy will be to use the concentration inequality for the empirical CVaR given in [24] and derive the uniform exponential convergence of \widehat{F}^N to F . Later, we will use Lemma 3 to infer exponential convergence of \widehat{S}^N . Note that the inequality given in [24] requires compact support of the random variable and it is tight when it comes to the dependency on the risk parameter α . For unbounded support, one can use deviation inequalities from [10].

For $i \in [n]$ and $x \in \mathcal{X}$, the deviation between the CVaR and its empirical counterpart can be bounded using the results in [24, Theorem 3.1] as

$$\mathbb{P}^N \left(\left| \widehat{\text{CVaR}}_{\alpha}^N [f_i(x, u)] - \text{CVaR}_{\alpha}^{\mathbb{P}} [f_i(x, u)] \right| \geq \epsilon \right) \leq 6 \exp \left(- \frac{\alpha \epsilon^2}{11(L - \ell)^2} N \right). \tag{11}$$

In the above bound, the denominator in the exponent uses the fact that any realization of $f_i(x, u)$ given any x is supported on the compact set $[\ell, L]$. Similar to the narrative of the previous section, while the above inequality holds pointwise, what we need is uniform exponential bound for proximity of F to \widehat{F}^N . Below, we will derive such a bound under the following condition.

Assumption 2 (*Uniform Lipschitz continuity of f_i*) There exists a constant $M > 0$ such that $|f_i(x, u) - f_i(x', u)| \leq M \|x - x'\|$, for all $x, x' \in \mathcal{X}, u \in \mathcal{U}$, and $i \in [n]$. •

Under the above Lipschitz condition on the random functions, one can show the following.

Lemma 4 (Lipschitz continuity of (empirical) CVaR) *Under Assumption 2, for any $i \in [n]$, functions $x \mapsto \widehat{\text{CVaR}}_{\alpha}^N [f_i(x, u)]$ and $x \mapsto \text{CVaR}_{\alpha}^{\mathbb{P}} [f_i(x, u)]$ are Lipschitz continuous over the set \mathcal{X} with constant $\frac{M}{\alpha}$.*

Proof We will show the property for the function $x \mapsto \widehat{\text{CVaR}}_{\alpha}^N [f_i(x, u)]$. The reasoning for $x \mapsto \text{CVaR}_{\alpha}^{\mathbb{P}} [f_i(x, u)]$ follows analogously. Consider any $x, x' \in \mathcal{X}$. Recall from (7) that

$$\left| \widehat{\text{CVaR}}_\alpha^N [f_i(x, u)] - \widehat{\text{CVaR}}_\alpha^N [f_i(x', u)] \right| = \left| \inf_{t \in \mathbb{R}} \widehat{\psi}_i^N(x, t) - \inf_{t \in \mathbb{R}} \widehat{\psi}_i^N(x', t) \right|. \tag{12}$$

Assumption 2 yields the Lipschitz continuity property for the map $\widehat{\psi}_i^N$. To establish this, fix any $i \in [n]$ and $t \in \mathbb{R}$ and notice that

$$\begin{aligned} \left| \widehat{\psi}_i^N(x, t) - \widehat{\psi}_i^N(x', t) \right| &= \left| t + \frac{1}{N\alpha} \sum_{j=1}^N [f_i(x, \widehat{u}^j) - t]_+ \right. \\ &\quad \left. - \left(t + \frac{1}{N\alpha} \sum_{j=1}^N [f_i(x', \widehat{u}^j) - t]_+ \right) \right| \\ &\leq \frac{1}{N\alpha} \sum_{j=1}^N \left| [f_i(x, \widehat{u}^j) - t]_+ - [f_i(x', \widehat{u}^j) - t]_+ \right| \\ &\leq \frac{1}{N\alpha} \sum_{j=1}^N \left| f_i(x, \widehat{u}^j) - f_i(x', \widehat{u}^j) \right| \leq \frac{M}{\alpha} \|x - x'\|. \end{aligned}$$

In the above relations, the first is a consequence of the triangle inequality, the second inequality follows from the fact that the map $[\cdot]_+$ is Lipschitz continuous with constant as unity, and the last inequality uses the Lipschitz continuity property of f_i . Now let $\bar{t}, \bar{t}' \in \mathbb{R}$ be such that $\widehat{\psi}_i^N(x, \bar{t}) = \inf_{t \in \mathbb{R}} \widehat{\psi}_i^N(x, t)$ and $\widehat{\psi}_i^N(x', \bar{t}') = \inf_{t \in \mathbb{R}} \widehat{\psi}_i^N(x', t)$. Existence of such an optimizer follows from the discussion in [18, Section 6.2.4]. Next note the following sequence of inequalities that can be inferred from the optimality condition and the Lipschitz continuity property of $\widehat{\psi}_i^N$ shown above,

$$\begin{aligned} \inf_{t \in \mathbb{R}} \widehat{\psi}_i^N(x, t) &= \widehat{\psi}_i^N(x, \bar{t}) \leq \widehat{\psi}_i^N(x, \bar{t}') \leq \widehat{\psi}_i^N(x', \bar{t}') + \frac{M}{\alpha} \|x - x'\| \\ &= \inf_{t \in \mathbb{R}} \widehat{\psi}_i^N(x', t) + \frac{M}{\alpha} \|x - x'\|. \end{aligned} \tag{13}$$

One can exchange x with x' in the above reasoning and obtain

$$\inf_{t \in \mathbb{R}} \widehat{\psi}_i^N(x', t) \leq \inf_{t \in \mathbb{R}} \widehat{\psi}_i^N(x, t) + \frac{M}{\alpha} \|x - x'\|. \tag{14}$$

Inequalities (13) and (14) imply that $\left| \inf_{t \in \mathbb{R}} \widehat{\psi}_i^N(x, t) - \inf_{t \in \mathbb{R}} \widehat{\psi}_i^N(x', t) \right| \leq \frac{M}{\alpha} \|x - x'\|$. The proof concludes by using this fact in (12). □

Next, we establish the exponential convergence of \widehat{F}^N . The proof is largely inspired from the steps given in [19, Theorem 5.1] and is a standard argument in these set of results. We note that the obtained bound is very crude and in practice, the achieved performance is much better.

Proposition 1 (Uniform exponential convergence of \widehat{F}^N to F) *Under Assumption 2, for any $0 < \epsilon < \text{diam}(\mathcal{X})/2$, the following holds for all $N \in \mathbb{N}$,*

$$\mathbb{P}^N \left(\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| > \epsilon \right) \leq \gamma(\epsilon) \exp(-\beta(\epsilon)N),$$

where

$$\gamma(\epsilon) := 6n \left(\frac{12M \text{diam}(\mathcal{X})}{\epsilon \alpha} \right)^n \frac{[n/2]!}{2\pi^{n/2}}, \tag{15a}$$

$$\beta(\epsilon) := \frac{\alpha \epsilon^2}{44n(L - \ell)^2}, \tag{15b}$$

and $\text{diam}(\mathcal{X}) = \sup_{x, x' \in \mathcal{X}} \|x - x'\|$ is the diameter of \mathcal{X} .

Proof The idea of moving from the pointwise exponential bound (11) to a uniform bound is to impose the pointwise bound jointly on a finite number of points and use the Lipschitz continuity property (Lemma 4) to bound the deviation of the rest of the set from this finite set. Making precise the mathematical details, note that one can cover the set \mathcal{X} with

$$K := \left(\frac{12M \text{diam}(\mathcal{X})}{\epsilon \alpha} \right)^n \frac{[n/2]!}{2\pi^{n/2}} \tag{16}$$

number of points, labeled $\mathcal{C} := \{\tilde{x}^1, \dots, \tilde{x}^K\} \subset \mathcal{X}$, such that for any $x \in \mathcal{X}$, there exists a point $\tilde{x}^{i(x)} \in \mathcal{C}$ with

$$(M/\alpha)\|x - \tilde{x}^{i(x)}\| \leq \epsilon/4. \tag{17}$$

The number K can be computed as follows. From (17), we require $\|x - \tilde{x}^{i(x)}\| \leq \frac{\epsilon \alpha}{4M}$. Thus, from Definition 3 in Appendix, the number of points in \mathcal{C} need only be bigger than the $\frac{\epsilon \alpha}{4M}$ -covering number of \mathcal{X} . Thus, any upper bound on this covering number suffices. From Lemma 7, one such upper bound is $\left(\frac{3 \text{diam}(\mathcal{X})}{(\epsilon \alpha / 4M)} \right)^n \frac{1}{\text{vol}(B)}$, where $\text{vol}(B)$ is the volume of the unit norm ball B in \mathbb{R}^n . Since $\text{vol}(B) \geq \frac{2\pi^{n/2}}{[n/2]!}$, we get the desired value for K given in (16). Having identified the set of points \mathcal{C} , we next combine the Lipschitz bound given in Lemma 4 and the inequality (17), to get for all $i \in [n]$ and $x \in \mathcal{X}$,

$$\left| \widehat{\text{CVaR}}_\alpha^N [f_i(x, u)] - \widehat{\text{CVaR}}_\alpha^N [f_i(\tilde{x}^{i(x)}, u)] \right| \leq \epsilon/4, \tag{18a}$$

$$\left| \text{CVaR}_\alpha^{\mathbb{P}} [f_i(x, u)] - \text{CVaR}_\alpha^{\mathbb{P}} [f_i(\tilde{x}^{i(x)}, u)] \right| \leq \epsilon/4. \tag{18b}$$

The above inequalities control the deviation of the functions $\widehat{\text{CVaR}}_\alpha^N [f_i(\cdot, u)]$ and $\text{CVaR}_\alpha^{\mathbb{P}} [f_i(\cdot, u)]$ over the set \mathcal{X} from the values these functions take on the set \mathcal{C} . The next step entails bounding the deviation of the CVaR and the empirical CVaR on the set \mathcal{C} . Employing (11) and the union bound, we have

$$\begin{aligned}
 & \mathbb{P}^N \left(\sup_{i \in [n], x \in \mathcal{C}} \left| \widehat{\text{CVaR}}_\alpha^N [f_i(x, u)] - \text{CVaR}_\alpha^{\mathbb{P}} [f_i(x, u)] \right| \geq \frac{\epsilon}{2} \right) \\
 & \leq \sum_{i \in [n]} \sum_{x \in \mathcal{C}} \mathbb{P}^N \left(\left| \widehat{\text{CVaR}}_\alpha^N [f_i(x, u)] - \text{CVaR}_\alpha^{\mathbb{P}} [f_i(x, u)] \right| \geq \frac{\epsilon}{2} \right) \quad (19) \\
 & \leq 6nK \exp \left(-\frac{\alpha \epsilon^2}{44(L - \ell)^2} N \right).
 \end{aligned}$$

The next set of inequalities characterize the difference between the CVaR and the empirical CVaR over the set \mathcal{X} using the Lipschitz continuity property (18). Fix $i \in [n]$ and let $x \in \mathcal{X}$. Note that using (18),

$$\begin{aligned}
 & \left| \widehat{\text{CVaR}}_\alpha^N [f_i(x, u)] - \text{CVaR}_\alpha^{\mathbb{P}} [f_i(x, u)] \right| \\
 & \leq \left| \widehat{\text{CVaR}}_\alpha^N [f_i(x, u)] - \widehat{\text{CVaR}}_\alpha^N [f_i(\tilde{x}^{i(x)}, u)] \right| \\
 & \quad + \left| \widehat{\text{CVaR}}_\alpha^N [f_i(\tilde{x}^{i(x)}, u)] - \text{CVaR}_\alpha^{\mathbb{P}} [f_i(\tilde{x}^{i(x)}, u)] \right| \\
 & \quad + \left| \text{CVaR}_\alpha^{\mathbb{P}} [f_i(\tilde{x}^{i(x)}, u)] - \text{CVaR}_\alpha^{\mathbb{P}} [f_i(x, u)] \right| \\
 & \leq \frac{\epsilon}{2} + \left| \widehat{\text{CVaR}}_\alpha^N [f_i(\tilde{x}^{i(x)}, u)] - \text{CVaR}_\alpha^{\mathbb{P}} [f_i(\tilde{x}^{i(x)}, u)] \right|.
 \end{aligned}$$

Next, the deviation between the CVaR and its empirical counterpart is bounded using (19) and the above characterization as

$$\begin{aligned}
 & \mathbb{P}^N \left(\sup_{i \in [n], x \in \mathcal{X}} \left| \widehat{\text{CVaR}}_\alpha^N [f_i(x, u)] - \text{CVaR}_\alpha^{\mathbb{P}} [f_i(x, u)] \right| \geq \epsilon \right) \\
 & \leq \mathbb{P}^N \left(\sup_{i \in [n], x \in \mathcal{X}} \left| \widehat{\text{CVaR}}_\alpha^N [f_i(x, t)] - \text{CVaR}_\alpha^{\mathbb{P}} [f_i(x, u)] \right| \geq \frac{\epsilon}{2} \right) \quad (20) \\
 & \leq 6nK \exp \left(-\frac{\alpha \epsilon^2}{44(L - \ell)^2} N \right).
 \end{aligned}$$

The final step is to connect the above inequality to the difference between \widehat{F}^N and F . From the proof of Lemma 2, one can deduce that if $\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| > \epsilon$, then

$$\sup_{i \in [n], x \in \mathcal{X}} \left| \widehat{\text{CVaR}}_\alpha^N [f_i(x, u)] - \text{CVaR}_\alpha^{\mathbb{P}} [f_i(x, u)] \right| > \frac{\epsilon}{\sqrt{n}}.$$

Therefore, using (20) we obtain

$$\begin{aligned}
 \mathbb{P}^N \left(\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| > \epsilon \right) & \leq \mathbb{P}^N \left(\sup_{i \in [n], x \in \mathcal{X}} \left| \widehat{\text{CVaR}}_\alpha^N [f_i(x, u)] \right. \right. \\
 & \quad \left. \left. - \text{CVaR}_\alpha^{\mathbb{P}} [f_i(x, u)] \right| > \frac{\epsilon}{\sqrt{n}} \right) \\
 & \leq 6nK \exp \left(-\frac{\alpha \epsilon^2}{44n(L - \ell)^2} N \right).
 \end{aligned}$$

This concludes the proof. □

The main result is given below. The proof follows from the uniform exponential convergence of \widehat{F}^N .

Theorem 3 (Exponential convergence of $\widehat{\mathcal{S}}^N$ to \mathcal{S}) *Let Assumption 2 hold. Then, for any $0 < \epsilon < \text{diam}(\mathcal{X})/2$, and $N \in \mathbb{N}$, the following inequality holds*

$$\mathbb{P}^N(\mathbb{D}(\widehat{\mathcal{S}}^N, \mathcal{S}) \leq \epsilon) \geq 1 - \gamma(\delta(\epsilon))e^{-\beta(\delta(\epsilon))N},$$

where γ and β are given in (15) and $\delta : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ is a map such that the pair $(\epsilon, \delta(\epsilon))$ satisfies the condition of Lemma 3.

Proof Consider any $\epsilon > 0$. By Lemma 3, if $\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| \leq \delta(\epsilon)$, then $\mathbb{D}(\widehat{\mathcal{S}}^N, \mathcal{S}) \leq \epsilon$. From Proposition 1, for any $\delta(\epsilon) > 0$, there exist $\gamma(\delta(\epsilon))$ and $\beta(\delta(\epsilon))$, given in (15a) and (15b), respectively, such that

$$\mathbb{P}^N\left(\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| > \delta(\epsilon)\right) \leq \gamma(\delta(\epsilon))e^{-\beta(\delta(\epsilon))N}$$

for all N . The proof follows by using the above facts and the set of inequalities:

$$\begin{aligned} \mathbb{P}^N(\mathbb{D}(\widehat{\mathcal{S}}^N, \mathcal{S}) \leq \epsilon) &\geq \mathbb{P}^N\left(\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| \leq \delta(\epsilon)\right) \\ &= 1 - \mathbb{P}^N\left(\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| > \delta(\epsilon)\right). \end{aligned}$$

□

Remark 1 (Sample guarantees for approximating \mathcal{S} with $\widehat{\mathcal{S}}^N$) Theorem 3 implies that if one wants $\mathbb{D}(\widehat{\mathcal{S}}^N, \mathcal{S}) \leq \epsilon$ with confidence $1 - \zeta$, where $\zeta \in (0, 1)$ is a small positive number, then one would require at most

$$\begin{aligned} N(\zeta, \epsilon) &= \frac{1}{\beta(\delta(\epsilon))} \log\left(\frac{\gamma(\delta(\epsilon))}{\zeta}\right) \\ &= \frac{44n(L - \ell)^2}{\alpha\delta(\epsilon)^2} \left(\log\left(\frac{6n \lceil n/2 \rceil!}{2\pi^{n/2}\zeta}\right) + n \log\left(\frac{12M \text{diam}(\mathcal{X})}{\epsilon\alpha}\right)\right) \end{aligned}$$

number of samples of the random variable. Due to the exponential rate, a good feature of this sample guarantee is that N depends on the accuracy ζ logarithmically. That is, one can obtain high confidence bounds with fewer samples. However, the sample size grows poorly with other parameters, especially, $\delta(\epsilon)$ and the dimension n . Further, note that to obtain an accurate sample guarantee, one needs to estimate $\delta(\cdot)$ which depends on the regularity of random functions. Improving the sample complexity for specific random functions is discussed in the following section. •

4 Separable uncertain functions

Here we illustrate how specific structure of the random functions yields tighter sample guarantees. Further, we discuss the tractability of solving the sample average VI problem.

Proposition 2 (Exponential convergence for separable functions) *Assume that the random functions have the form*

$$f_i(x, u) = \tilde{f}_i(x)g_i(u) + \check{f}_i(x) \text{ for all } i \in [n], \tag{21}$$

where $\tilde{f}_i, g_i,$ and \check{f}_i are non-negative real-valued continuous functions. Then, for any $\epsilon > 0$ and $N \in \mathbb{N}$, the following holds

$$\mathbb{P}^N(\mathbb{D}(\widehat{\mathcal{S}}^N, \mathcal{S}) \leq \epsilon) \geq 1 - 6ne^{-\beta(\delta(\epsilon))N}, \tag{22}$$

where $\beta(\epsilon) := \frac{\alpha\epsilon^2}{11n(f^{\max} g^{\text{rge}})^2}$, with $f^{\max} := \sup_{i \in [n], x \in \mathcal{X}} \tilde{f}_i(x)$ and $g^{\text{rge}} := \sup_{i \in [n]} (\sup_{u \in \mathcal{U}} g_i(u) - \inf_{u \in \mathcal{U}} g_i(u))$.

Proof Since CVaR is positive-homogeneous and shift-invariant [18, Chapter 6], one gets

$$\widehat{\text{CVaR}}_\alpha^N[f_i(x, u)] = \tilde{f}_i(x)\widehat{\text{CVaR}}_\alpha^N[g_i(u)] + \check{f}_i(x), \tag{23}$$

for all $i \in [n]$. Using this fact, for any $\epsilon > 0$, we reason as

$$\begin{aligned} & \mathbb{P}^N\left(\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| > \epsilon\right) \\ & \leq \mathbb{P}^N\left(\sup_{i \in [n], x \in \mathcal{X}} \left|\widehat{\text{CVaR}}_\alpha^N[f_i(x, u)] - \text{CVaR}_\alpha^{\mathbb{P}}[f_i(x, u)]\right| > \frac{\epsilon}{\sqrt{n}}\right) \\ & \stackrel{(a)}{\leq} \mathbb{P}^N\left(\sup_{i \in [n]} \left|\widehat{\text{CVaR}}_\alpha^N[g_i(u)] - \text{CVaR}_\alpha^{\mathbb{P}}[g_i(u)]\right| > \frac{\epsilon}{\sqrt{nf^{\max}}}\right) \\ & \stackrel{(b)}{\leq} 6n \exp\left(-\frac{\alpha\epsilon^2}{11n(f^{\max} g^{\text{rge}})^2}N\right), \end{aligned} \tag{24}$$

where (a) follows from (23) (note that a similar equality as (23) holds for $\text{CVaR}_\alpha^{\mathbb{P}}$) and the fact that \tilde{f} takes non-negative values and (b) is a result of the deviation inequality (11) applied in combination with the union bound. Using (24) and proceeding along the lines of Theorem 3 we obtain (22). \square

Recounting the way we obtained exponential convergence in the previous section, the key step was in Proposition 1 where we moved from the deviation inequality for finite number of points in \mathcal{X} to the uniform exponential convergence of \widehat{F} . Such an exercise was inevitable due to the possible interdependence of x and u in the random

function. Not only that, the bound for this reason scaled poorly with many parameters, such as, the dimension n and the size of \mathcal{X} . However, when the random function takes the form (21), then one need not construct a cover for \mathcal{X} to derive uniform exponential convergence of \widehat{F} . Thus, we obtain a tighter bound (22).

Remark 2 (*Sample guarantees and tractability for separable functions*) Same as Remark 1, we deduce using Proposition 2 that for separable functions (21), the accuracy $\mathbb{D}(\widehat{\mathcal{S}}^N, \mathcal{S}) \leq \epsilon$ with confidence $1 - \zeta$ is guaranteed with

$$N(\epsilon, \zeta) = \frac{11n(f^{\max} g^{\text{rge}})^2}{\alpha\delta(\epsilon)^2} \log\left(\frac{6n}{\zeta}\right)$$

number of samples. As expected, the above sample size does not depend on the size of \mathcal{X} . Further, for the above derivation we need not assume the random function to be Lipschitz continuous.

We next comment about solving $\text{VI}(\mathcal{X}, \widehat{F}^N)$ for separable functions. For convenience, denote the concatenation of \tilde{f}_i and \check{f}_i for $i \in [n]$ with functions \tilde{f} and \check{f} , respectively. Further, let the vector $\widehat{g}^N := (\widehat{\text{CVaR}}_\alpha [g_i(u)])_{i \in [n]}$ collect the empirical CVaR of the random function of each path. Then, the aim is to solve $\text{VI}(\mathcal{X}, \tilde{f} \odot \widehat{g}^N + \check{f})$, where \odot represents component-wise product. Given samples, the approach would be to compute \widehat{g}^N and then solve the VI. Note that computing each component \widehat{g}_i^N amounts to solving a linear program:

$$\widehat{g}_i^N = \min \left\{ t + \frac{1}{N\alpha} \sum_{j=1}^N y_j \mid \begin{array}{l} y_j \geq g_p(\widehat{u}^j) - t, \quad \forall j \in [N], \\ t \in \mathbb{R}, y_j \geq 0, \quad \forall j \in [N] \end{array} \right\}.$$

The appealing part of this process is the deconstruction into two steps: computing the empirical CVaR independent of x and solving the VI without worrying about samples. Further, one can derive conditions on the underlying functions that guarantee monotonicity of $\tilde{f} \odot \widehat{g}^N + \check{f}$ that consequently lead to efficient algorithms that solve the VI, see e.g., methods given in [7]. •

Note that if F is strictly monotone, then the solution set \mathcal{S} is a singleton. In that case, asymptotic consistency implies that all sequences $\{\widehat{x}^N\}$ converge to the unique solution. If in addition F satisfies a stronger assumption, that of being strongly monotone and having a separable form, then one can estimate the map δ used in the exponential convergence bound. The next result formalizes this implication.

Lemma 5 (*Estimating δ for strongly monotone F*) Assume that the random functions are of the form $f_i(x, u) = \tilde{f}_i(x) + g_i(u)$ for all $i \in [n]$, where \tilde{f}_i and g_i are real-valued continuous functions. Suppose the concatenated function $\tilde{f} := (f_i)_{i \in [n]}$ is strongly

monotone over \mathcal{X} with modulus $\sigma > 0$. Then, $\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| \leq \epsilon$ implies

$$\mathbb{D}(\widehat{\mathcal{S}}^N, \mathcal{S}) \leq \sigma^{-1}\epsilon.$$

Proof Note that $F(x) = \tilde{f}(x) + \kappa$, where components of the vector κ are given by $\kappa_i := \text{CVaR}_\alpha^{\text{ip}}[g_i(u)]$ for all i . Similarly, we have $\widehat{F}^N(x) = \tilde{f}(x) + \widehat{\kappa}^N$ where $\widehat{\kappa}_i^N := \widehat{\text{CVaR}}_\alpha^N[g_i(u)]$. By assumption,

$$\sup_{x \in \mathcal{X}} \|\widehat{F}^N(x) - F(x)\| = \|\widehat{\kappa} - \kappa\| \leq \epsilon. \tag{25}$$

By the definition of the solution of a VI, for any $x^* \in \text{SOL}(\mathcal{X}, F)$ and $\widehat{x}^N \in \text{SOL}(\mathcal{X}, \widehat{F}^N)$, we have $(\widehat{x}^N - x^*)^\top F(x^*) \geq 0$ and $(x^* - \widehat{x}^N)^\top \widehat{F}^N(\widehat{x}^N) \geq 0$. Combining these inequalities gives us $(x^* - \widehat{x}^N)^\top (F(x^*) - \widehat{F}^N(\widehat{x}^N)) \leq 0$. Using the separable forms of the functions, we get

$$(x^* - \widehat{x}^N)^\top (\tilde{f}(x^*) - \tilde{f}(\widehat{x}^N)) \leq (x^* - \widehat{x}^N)^\top (\widehat{\kappa} - \kappa).$$

Using strong monotonicity condition $(x^* - \widehat{x}^N)^\top (\tilde{f}(x^*) - \tilde{f}(\widehat{x}^N)) \geq \sigma \|x^* - \widehat{x}^N\|^2$ and the Cauchy-Schwartz inequality in the above expression, we get

$$\sigma \|x^* - \widehat{x}^N\|^2 \leq \|x^* - \widehat{x}^N\| \|\widehat{\kappa} - \kappa\|.$$

The proof now follows from (25). □

The above result can be used to further refine the convergence rate given in Proposition 2. In the following section, we apply our results to the uncertain network routing problem.

5 Application: computing CVaR-based Wardrop equilibrium

Consider a network given by a directed graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, where \mathcal{V} and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ stand for the set of nodes and edges, respectively. The sets of origin and destination nodes¹ are the sets of sources and sinks in the network, and are denoted by $\mathcal{O} \subset \mathcal{V}$ and $\mathcal{D} \subset \mathcal{V}$, respectively. The set of origin–destination (OD) pairs is $\mathcal{W} \subseteq \mathcal{O} \times \mathcal{D}$. Let \mathcal{P}_w denote the set of available paths for the OD pair $w \in \mathcal{W}$ and let $\mathcal{P} = \cup_{w \in \mathcal{W}} \mathcal{P}_w$ be the set of all paths.² Consider the setting of nonatomic routing where numerous agents traverse the network and so each individual agent’s action has infinitesimal impact on the aggregate traffic flow. As a consequence, flow is modeled as a continuous variable. Each agent is associated with an OD pair $w \in \mathcal{W}$ and is allowed to select any path $p \in \mathcal{P}_w$. The route choices give rise to the aggregate traffic which is modeled as a flow vector $h \in \mathbb{R}_{\geq 0}^{|\mathcal{P}|}$ with h_p being the flow on a path $p \in \mathcal{P}$. The flow between each OD pair must satisfy the travel demand. We denote the demand for the OD pair $w \in \mathcal{W}$ by $d_w \in \mathbb{R}_{\geq 0}$ and the set of feasible flows by

¹ A source is a vertex with no incoming edge and a sink is a vertex with no outgoing edge.
² A path is an ordered sequence of unique vertices such that two subsequent vertices form an edge.

$$\mathcal{H} := \left\{ h \in \mathbb{R}_{\geq 0}^{|\mathcal{P}|} \mid \sum_{p \in \mathcal{P}_w} h_p = d_w \text{ for all } w \in \mathcal{W} \right\}. \tag{26}$$

Agents who choose path $p \in \mathcal{P}$ experience a non-negative uncertain cost denoted by $C_p : \mathbb{R}_{\geq 0}^{|\mathcal{P}|} \times \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$, $(h, u) \mapsto C_p(h, u)$, where $u \in \mathbb{R}^m$ models the uncertainty. Let \mathbb{P} and $\mathcal{U} \subset \mathbb{R}^m$ be the distribution and support of u , respectively. Assume that \mathcal{U} is compact. For the cost function, assume that for every $p \in \mathcal{P}$ and $u \in \mathcal{U}$, the function $h \mapsto C_p(h, u)$ is continuous. For every $p \in \mathcal{P}$ and $h \in \mathcal{H}$, the function $u \mapsto C_p(h, u)$ is measurable. In addition, for all $p \in \mathcal{P}$, assume that C_p takes finite value over $\mathcal{H} \times \mathcal{U}$. The above described elements collectively represent an uncertain routing game. To assign an appropriate objective for agents, we assume that agents are risk-averse and look for paths with least CVaR. We assume that all agents have the same risk-aversion characterized by the parameter $\alpha \in (0, 1)$. The CVaR associated to path p as a function of the flow is

$$\text{CVaR}_\alpha^\mathbb{P}[C_p(h, u)] = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{\alpha} \mathbb{E}_\mathbb{P} [C_p(h, u) - t]_+ \right\}. \tag{27}$$

The notion of equilibrium then is that of Wardrop [4], given below.

Definition 2 (*Conditional value-at-risk based Wardrop equilibrium (CWE)*) A flow vector $h^* \in \mathbb{R}_{\geq 0}^{|\mathcal{P}|}$ is called a *CVaR-based Wardrop equilibrium (CWE)* for the uncertain routing game if: (i) h^* satisfies the demand for all OD pairs and (ii) for any OD pair w , a path $p \in \mathcal{P}_w$ has nonzero flow if the CVaR of path p is minimum among all paths in \mathcal{P}_w . Formally, h^* is a CWE if $h^* \in \mathcal{H}$ and $h_p^* > 0$ for $p \in \mathcal{P}_w$ only if

$$\text{CVaR}_\alpha^\mathbb{P}[C_p(h^*, u)] \leq \text{CVaR}_\alpha^\mathbb{P}[C_q(h^*, u)], \quad \forall q \in \mathcal{P}_w. \tag{28}$$

We denote the set of CWE by $\mathcal{S}_{\text{CWE}} \subset \mathcal{H}$. •

The set \mathcal{S}_{CWE} is equivalent to the set of solutions to the variational inequality (VI) problem $\text{VI}(\mathcal{H}, G)$ (see Sect. 2 for relevant notions) [20], where

$$G_p(h) := \text{CVaR}_\alpha^\mathbb{P}[C_p(h, u)],$$

for all $p \in \mathcal{P}$. Note that the set \mathcal{H} is compact and convex. Further, the map $h \mapsto G(h)$ is continuous since C_p , $p \in \mathcal{P}$ are so and \mathcal{H} and \mathcal{U} are compact. Therefore, the set of solutions $\text{SOL}(\mathcal{H}, G)$ is nonempty and compact [7, Corollary 2.2.5]. Consequently, the set \mathcal{S}_{CWE} is nonempty and compact. Due to this connection between the CWE and the solution of the risk-based VI, we can apply the results developed in the previous section to study the sample average approximation to the CWE. We present the following main result. Given N i.i.d samples of u , the sample average approximation of the function G is defined component-wise as $\widehat{G}_p^N(h) := \widehat{\text{CVaR}}_\alpha^N[C_p(h, u)]$, where

$$\widehat{\text{CVaR}}_\alpha^N[C_p(h, u)] := \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{N\alpha} \sum_{i=1}^N [C_p(h, \hat{u}^i) - t]_+ \right\}.$$

Denote the set of solutions of the VI($\mathcal{H}, \widehat{G}^N$) by $\widehat{\mathcal{S}}_{\text{CWE}}^N$. We have the guarantee:

Theorem 4 (Convergence of $\widehat{\mathcal{S}}_{\text{CWE}}^N$ to \mathcal{S}_{CWE}) *The following hold:*

- (i) For any $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that $\mathbb{D}(\widehat{\mathcal{S}}_{\text{CWE}}^N, \mathcal{S}_{\text{CWE}}) \leq \epsilon$ whenever $\sup_{h \in \mathcal{H}} \|\widehat{G}^N(h) - G(h)\| \leq \delta(\epsilon)$.
- (ii) Almost surely $\mathbb{D}(\widehat{\mathcal{S}}_{\text{CWE}}^N, \mathcal{S}_{\text{CWE}}) \rightarrow 0$.

Furthermore, assume there exists a constant $M > 0$ such that $|C_p(h, u) - C_p(h', u)| \leq M\|h - h'\|$ for all $h, h' \in \mathcal{H}$, $u \in \mathcal{U}$, and $p \in \mathcal{P}$. Let $\ell := \min\{C_p(h, u) \mid h \in \mathcal{H}, u \in \mathcal{U}, p \in \mathcal{P}\}$ and $L := \max\{C_p(h, u) \mid h \in \mathcal{H}, u \in \mathcal{U}, p \in \mathcal{P}\}$. Then, for any $\epsilon > 0$, and $N \in \mathbb{N}$, the following inequality holds

$$\mathbb{P}^N(\mathbb{D}(\widehat{\mathcal{S}}_{\text{CWE}}^N, \mathcal{S}_{\text{CWE}}) \leq \epsilon) \geq 1 - \gamma(\delta(\epsilon))e^{-\beta(\delta(\epsilon))N},$$

where $\delta(\epsilon)$ satisfies the condition given in (i), and

$$\gamma(\epsilon) := 6|\mathcal{P}| \prod_{w \in \mathcal{W}} \left\lceil \frac{4M|\mathcal{W}|\sqrt{|\mathcal{P}_w|}}{\epsilon\alpha} \right\rceil, \quad \beta(\epsilon) := \frac{\alpha\epsilon^2}{44|\mathcal{P}|(L - \ell)^2}. \tag{29}$$

The proof follows in the same way as that of Theorem 1 and 3, using the fact that the covering number of \mathcal{H} can be bounded as given in Lemma 9 in the appendix. This sharper bound on covering number brings out a notable difference in the constants given in the exponential bound (29) as compared to those derived in Theorem 3.

6 Numerical example: Sioux falls

Here we illustrate the method of sample average approximation for the computation of the CWE through an example. We consider the Sioux Falls traffic network that consists of 24 nodes and 76 edges [23]. The (deterministic) cost associated to each edge $e \in \mathcal{E}$ of the network is the travel time and is given as an affine function of the flow on the edge, $f_e(\ell_e) := t_e \left(1 + b_e \frac{\ell_e}{c_e}\right)$, where ℓ_e is the flow on edge $e \in \mathcal{E}$, t_e is the free-flow travel time, and c_e is the capacity of the edge. The values for constants t_e and c_e are taken from the repository [23]. The constant b_e is fixed to be 100 for all edges. For simplicity, we consider three OD pairs $\mathcal{W} = \{(1, 19), (13, 8), (12, 18)\}$ and for each pair, we choose 10 paths that have the shortest free-flow travel time. The demand is given as $d_{(1,19)} = 300$, $d_{(13,8)} = 600$, and $d_{(12,18)} = 200$. The cost of each path is set to be the summation of the costs of the edges contained in it. That is, for some path $p \in \mathcal{P}$, $C_p(h) = \sum_{e \in p} f_e(\ell_e)$, where the summation is over all edges that constitute the path. Note that the flow on any edge is the sum of the flow of the paths that use that edge. That is, $\ell_e = \sum_{\{p \in \mathcal{P} \mid e \in p\}} h_p$. We assume that the cost associated to each edge that contains either of the nodes 10, 16, or 17 is uncertain. Specifically, for such an edge e , the uncertain cost is given as $J_e(\ell_e, u_e) = f_e(\ell_e) + u_e$,

where u_e has uniform distribution over the set $[0, 0.5t_e]$. For all other edges, we set u_e to be zero with probability one. All uncertainties are assumed to be mutually independent. The uncertain cost of a path $p \in \mathcal{P}$ is given as $C_p(h, u) = \sum_{e \in p} J_e(\ell_e, u_e)$. This defines completely the routing game with uncertain costs.

6.1 Affine separable costs and LCP

In the example explained above, cost C_p is linear in flow and the uncertainty is additive. Therefore, solving the sample average VI is equivalent to solving a linear complementarity problem (LCP), which in turn is a convex optimization problem with quadratic cost and affine constraints. We next derive this optimization problem. Let $Q \in \{0, 1\}^{|\mathcal{E}| \times |\mathcal{P}|}$ be the (edge, path)-incidence matrix where Q_{ep} entry is 1 if and only if edge e belongs to path p . Then, using the linear relationship between edge and path flows, the vector containing all edge flows can be written as $\ell = Qh$, where h consists of flows on paths. Stacking all uncertain edge costs J_e in a vector J and using its affine separable form, we obtain

$$J(h, u) = RQh + t + u,$$

where u and t are vector of uncertainties and free-flow travel time of all edges, respectively, and R is a diagonal matrix where the diagonal entry corresponding to edge e is $b_e t_e / c_e$. Using the above relation, the vector of costs incurred on paths takes the form

$$C(h, u) = Q^T RQh + Q^T t + Q^T u.$$

Further, since CVaR is shift-invariant, we obtain

$$G(h) = Q^T RQh + Q^T t + \text{CVaR}_\alpha^{\text{P}}[Q^T u],$$

where the last term in the above relation is the vector of element-wise CVaRs. Similarly, the sample average approximation of G is given as

$$\widehat{G}^N(h) = Q^T RQh + Q^T t + \widehat{\text{CVaR}}_\alpha^N[Q^T u].$$

Our aim is find the solution of $\text{VI}(\mathcal{H}, \widehat{G})$. To represent the set of feasible flows in a compact form, denote $B \in \{0, 1\}^{|\mathcal{W}| \times |\mathcal{P}|}$ as the (OD pair, path)-incidence matrix where B_{wp} is 1 if and only if $p \in \mathcal{P}_w$. Denoting the vector of demands as $d = (d_w)_{w \in \mathcal{W}}$, the set of feasible flows are vectors $h \geq 0$ satisfying $Bh = d$. Using this notation and the explanation given in [26, Section 2.2], finding the solution of $\text{VI}(\mathcal{H}, \widehat{G}^N)$ is equivalent to solving the LCP given as: find $x = (h; v) \in \mathbb{R}^{|\mathcal{P}| + |\mathcal{W}|}$ such that

$$x \geq 0, \quad \widehat{M}^N(x) \geq 0, \quad \text{and} \quad x^T \widehat{M}^N(x) = 0,$$

where

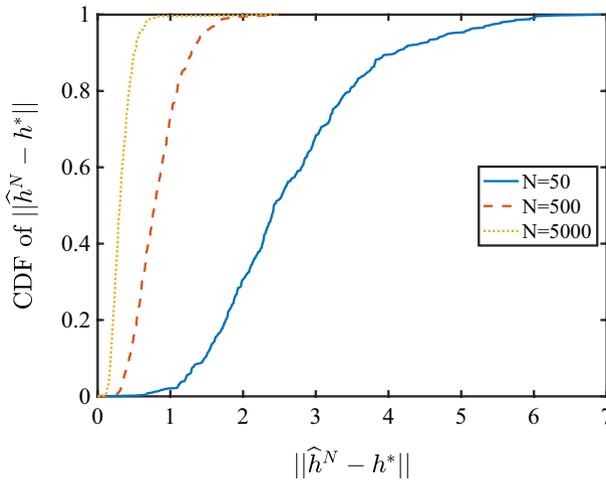


Fig. 1 Plot illustrating the convergence of the approximate solution \hat{h}^N to the CVaR-based Wardrop equilibrium h^* for Sioux falls network, see Sect. 6 for details. Each line is the cumulative distribution of $\|\hat{h}^N - h^*\|$ with a different sample size used for the approximation. The distribution is obtained using 500 runs

$$\hat{M}^N(x) := \begin{bmatrix} Q^T R Q & -B^T \\ B & 0 \end{bmatrix} x + \begin{bmatrix} Q^T t + \widehat{CVaR}_\alpha^N [Q^T u] \\ -d \end{bmatrix}.$$

Solving LCP is equivalent to finding the optimizer of the following problem

$$\begin{aligned} &\text{minimize} && x^T \hat{M}^N(x) \\ &\text{subject to} && \hat{M}^N(x) \geq 0, \\ &&& x \geq 0. \end{aligned} \tag{30}$$

Since \hat{M}^N is affine in x , the above problem is quadratic and one can show using the properties of matrices Q , R , and B , that the problem is convex. To summarize, the $VI(\mathcal{H}, G)$ can be approximated by $VI(\mathcal{H}, \hat{G}^N)$ and the latter can be solved by first finding $\widehat{CVaR}_\alpha^N [Q^T u]$ and then solving (30).

6.2 Computing CWE

For the above explained Sioux falls example, we set $\alpha = 0.05$. This defines uniquely the CWE h^* . For the sample average approximation, we consider three scenarios with different number of samples, $N \in \{50, 500, 5000\}$. We consider 500 runs for each of the scenarios. Each run collects N number of i.i.d samples of the uncertainty u , constructs the empirical $\widehat{CVaR}_\alpha^N [Q^T u]$, and computes the approximation of

the CWE \hat{h}^N by solving (30). Figure 1 illustrates our results. It plots the cumulative distribution function of the random variable $\|\hat{h}^N - h^*\|$ as estimated using the 500 runs. Note that the complete distribution moves to the left with increasing number of samples. This confirms our theoretical findings that as N increases, the approximate solution \hat{h}^N approaches the CWE almost surely.

7 Conclusions

We considered a risk-based variational inequality and studied the sample average approximation method for solving it. In particular, we derived asymptotic consistency and exponential convergence under suitable assumptions. For the case of separable random function, we derived sharper convergence bounds. Lastly we demonstrated the application of our result in the case of uncertain network routing problem where one can determine the CVaR-based Wardrop equilibrium using the sample average scheme. Future work will involve exploring tractability of the resulting sample average VI under various conditions. We also wish to investigate other efficient sampling techniques, especially when the dimension of the problem is large. Finally, we plan to investigate decentralized learning methods for finding the solution of risk-averse VIs.

Appendix

Here, we estimate the covering number of a general set \mathcal{X} and the feasible flow set \mathcal{H} related to the network routing problem. This computation helps in establishing Proposition 1 and Theorem 4. In order to present the results, we need a couple of definitions.

Definition 3 (Covering number [22, Chapter 3]) Given a set $\mathcal{X} \subset \mathbb{R}^n$ and a real value $\epsilon > 0$, a set of $m \in \mathbb{N}$ points $\{x_1, x_2, \dots, x_m\}$ is called an ϵ -cover of \mathcal{X} if $\mathcal{X} \subset \bigcup_{k=1}^m B(x_k, \epsilon)$, where $B(x, \epsilon)$ is the closed ball in Euclidean metric with center as x and radius ϵ . The minimum number of points required to form an ϵ -cover of \mathcal{X} is called the ϵ -covering number. •

From [25], we have the following result. We give the proof here for the sake of completeness.

Lemma 6 (Covering number of a set) *The ϵ -covering number of a convex set $\mathcal{X} \subset \mathbb{R}^n$ that satisfies $\epsilon B \subset \mathcal{X}$ is upper bounded by $\left(\frac{3}{\epsilon}\right)^n \frac{\text{vol}(\mathcal{X})}{\text{vol}(B)}$, where vol stands for volume and B is the unit norm ball in \mathbb{R}^n .*

Proof Note that the ϵ -covering number is bounded above by the ϵ -packing number of the set [22, Chapter 3]. The latter is defined as the maximum number of points that can be selected from \mathcal{X} such that they are mutually more than ϵ

distance apart. Let $\{x_1, x_2, \dots, x_P\} \subset \mathcal{X}$ be these points and P be the ϵ -packing number. Our next step is to derive a bound for P . Note that by definition of packing, closed balls $B(x_i, \epsilon/2) := \{y \in \mathbb{R}^n \mid \|x_i - y\| \leq \epsilon/2\}$, $i \in \{1, \dots, P\}$ are disjoint and $\cup_{i=1}^P B(x_i, \epsilon/2) \subset \mathcal{X} + (\epsilon/2)B$, where the set addition is considered to be the Minkowski sum. Taking the volume on both sides yields

$$\text{vol}(\mathcal{X} + (\epsilon/2)B) \geq \text{vol}\left(\cup_{i=1}^P B(x_i, \epsilon/2)\right) = P\text{vol}((\epsilon/2)B).$$

This implies $P \leq \frac{\text{vol}(\mathcal{X} + (\epsilon/2)B)}{\text{vol}((\epsilon/2)B)}$. Next we wish to show that

$$\mathcal{X} + (\epsilon/2)B \subset (3/2)\mathcal{X} \tag{31}$$

under our hypothesis. First note that if $x \in \mathcal{X} + (\epsilon/2)B$, then there exists $y \in \mathcal{X}$ and $z \in (\epsilon/2)B$ such that $x = y + z$. By assumption, $\epsilon B \subset \mathcal{X}$ and so $z \in (1/2)\mathcal{X}$. This implies that $x \in \mathcal{X} + (1/2)\mathcal{X}$. Thus, $\mathcal{X} + (\epsilon/2)B \subset \mathcal{X} + (1/2)\mathcal{X}$. Next using convexity one can show that $\mathcal{X} + (1/2)\mathcal{X} \subset (3/2)\mathcal{X}$. Indeed, pick any $x \in \mathcal{X} + (1/2)\mathcal{X}$, we have $y, z \in \mathcal{X}$ such that $x = y + (1/2)z$. That is, $\frac{2}{3}x = \frac{2}{3}y + \frac{1}{3}z$. Using convexity we get $\frac{2}{3}x \in \mathcal{X}$ and so $x \in \frac{3}{2}\mathcal{X}$. This establishes (31). Using this inclusion we get $\text{vol}(\mathcal{X} + (\epsilon/2)B) \leq \text{vol}((3/2)\mathcal{X})$. Finally, substituting this in the bound on P , we have

$$P \leq \frac{\text{vol}(\mathcal{X} + (\epsilon/2)B)}{\text{vol}((\epsilon/2)B)} \leq \frac{\text{vol}((3/2)\mathcal{X})}{\text{vol}((\epsilon/2)B)} = \left(\frac{3}{\epsilon}\right)^n \frac{\text{vol}(\mathcal{X})}{\text{vol}(B)}.$$

This completes the proof. □

The following is an application of the above result.

Lemma 7 (Covering number of a compact set) *The ϵ -covering number of a compact set $\mathcal{X} \subset \mathbb{R}^n$, where $\epsilon \leq \frac{\text{diam}(\mathcal{X})}{2}$, is upper bounded by $\left(\frac{3\text{diam}(\mathcal{X})}{\epsilon}\right)^n \frac{1}{\text{vol}(B)}$, where $\text{vol}(B)$ is the volume of the unit norm ball B in \mathbb{R}^n and $\text{diam}(\mathcal{X}) = \sup_{x, x' \in \mathcal{X}} \|x - x'\|$ is the diameter of \mathcal{X} .*

Proof Consider the set $\mathcal{M} := [-\text{diam}(\mathcal{X})/2, \text{diam}(\mathcal{X})/2]^n$, where $\text{diam}(\mathcal{X})$ is the diameter of the set \mathcal{X} . One can verify that the covering number of \mathcal{X} is upper bounded by that of the set \mathcal{M} . This is because the set \mathcal{X} can be entirely contained in \mathcal{M} after performing a translation operation. Note that $\text{vol}(\mathcal{M}) = \text{diam}(\mathcal{X})^n$. The result then follows from Lemma 6. □

Next, we provide a bound on the covering number of a simplex. In the consequent result, we use this bound to analyze the covering number of the feasible flow set \mathcal{H} .

Lemma 8 (Covering number for a simplex) *For the simplex $\Delta_d^n := \{x \in \mathbb{R}_{\geq 0}^n \mid \sum_{i=1}^n x_i = d\}$, the ϵ -covering number is bounded above by*

$$\binom{n + K - 1}{K - 1}, \tag{32}$$

where $K = \lceil \frac{\sqrt{nd}}{\epsilon} \rceil$.

Proof Consider the set of points

$$\mathcal{C} := \left\{ \left(\frac{i_1 d}{K}, \frac{i_2 d}{K}, \dots, \frac{i_n d}{K} \right) \mid i_s \in [K] \cup \{0\}, \forall s \in [n], \text{ and } \sum_{s=1}^n i_s = K \right\},$$

where $K = \lceil \frac{\sqrt{nd}}{\epsilon} \rceil$. Note that $\mathcal{C} \subset \Delta_d^n$. We will show that this is a valid ϵ -cover for Δ_d^n . To this end, pick any point $x \in \Delta_d^n$. We will construct a point $x^c \in \mathcal{C}$ such that $\|x - x^c\| \leq \epsilon$. Let $x^{up}, x^{dn} \in \mathbb{R}_{\geq 0}^n$ be such that each j -th component is given by

$$\begin{aligned} x_j^{up} &= \min \left\{ \frac{i * d}{K} \mid \frac{i * d}{K} \geq x_j, i \in [K] \cup \{0\} \right\}, \\ x_j^{dn} &= \max \left\{ \frac{i * d}{K} \mid \frac{i * d}{K} \leq x_j, i \in [K] \cup \{0\} \right\}. \end{aligned}$$

Note that $x^{dn} \leq x \leq x^{up}$, where \leq denotes element-wise inequality. Further, $\sum_{j=1}^n x_j^{dn} \leq \sum_{j=1}^n x_j = d \leq \sum_{j=1}^n x_j^{up}$. By construction, for any vector y satisfying $x^{dn} \leq y \leq x^{up}$, we have $\|y - x\|_\infty \leq \frac{d}{K}$. Consequently, for such a vector we have

$$\|y - x\| \leq \sqrt{n} \|y - x\|_\infty \leq \frac{\sqrt{nd}}{K} \leq \epsilon. \tag{33}$$

Thus, our aim is to find a vector y that belongs to \mathcal{C} and for which $x^{dn} \leq y \leq x^{up}$ holds. Define

$$\delta = \frac{K}{d} \sum_{j=1}^n (x_j - x_j^{dn}).$$

Note that δ is an integer as $\sum_{j=1}^n x_j = d$ and each component x_j^{dn} is a product of an integer and the quantity $\frac{d}{K}$. Now consider a vector $y^\delta \in \{0, 1\}^n$ such that $\sum_{j=1}^n y_j^\delta = \delta$.

Set $y = x^{dn} + y^\delta$. It is easy to see that by construction $x^{dn} \leq y \leq x^{up}$ and $y \in \mathcal{C}$. The former establishes $\|y - x\| \leq \epsilon$ due to the reasoning in (33). Thus, \mathcal{C} is an ϵ -cover for Δ_d^n . As a consequence, to complete the proof we need to enumerate the points in \mathcal{C} .

To this end, note that for any point $x^c = d \left(\frac{i_1}{K}, \frac{i_2}{K}, \dots, \frac{i_n}{K} \right) \in \mathcal{C}$, we have

$$\sum_{s=1}^n Kx_s^c = d \sum_{s=1}^n i_s = d \cdot K.$$

Since each i_s is a nonnegative integer, using the above inequality, the number of points in \mathcal{C} is the number of ways K identical objects can be put into n distinct bins. This number is given as (32). □

Using the above result, we next derive an upper bound on the ϵ -covering number the set \mathcal{H} .

Lemma 9 (Covering number of \mathcal{H}) *The ϵ -covering number of the set of feasible flows \mathcal{H} given in (26) is bounded above by*

$$\prod_{w \in \mathcal{W}} \binom{|\mathcal{P}_w| + K_w - 1}{K_w - 1}, \tag{34}$$

where \prod denotes the product and $K_w = \lceil \frac{|\mathcal{W}|\sqrt{\mathcal{P}_w d_w}}{\epsilon} \rceil$ for all $w \in \mathcal{W}$.

Proof First note that $\mathcal{H} = \prod_{w \in \mathcal{W}} \mathcal{H}_w$, where \prod represents the Cartesian product and

$$\mathcal{H}_w := \left\{ h^w \in \mathbb{R}_{\geq 0}^{|\mathcal{P}_w|} \mid \sum_{p \in \mathcal{P}_w} h_p^w = d_w \right\}$$

for all $w \in \mathcal{W}$. That is, \mathcal{H}_w represents the set of feasible flows for paths corresponding to the OD pair w . From Lemma 8, the number of points required to cover the set \mathcal{H}_w with balls of radius $\frac{\epsilon}{|\mathcal{W}|}$ is

$$\binom{|\mathcal{P}_w| + K_w - 1}{K_w - 1},$$

where $K_w = \lceil \frac{|\mathcal{W}|\sqrt{\mathcal{P}_w d_w}}{\epsilon} \rceil$. Consider these set of points to be represented by $\mathcal{C}_w \subset \mathcal{H}_w$. Now consider the set of points $\mathcal{C} = \{(h^w)_{w \in \mathcal{W}} \mid h^w \in \mathcal{C}_w \text{ for all } w \in \mathcal{W}\}$. The number of points in \mathcal{C} is equal to the value in (34). We show next that \mathcal{C} is an ϵ -cover for \mathcal{H} . Pick any $h = (h^w)_{w \in \mathcal{W}} \in \mathcal{H}$, we have

$$\min_{\bar{h} \in \mathcal{C}} \|h - \bar{h}\| \leq \sum_{w \in \mathcal{W}} \min_{\bar{h}^w \in \mathcal{C}_w} \|h^w - \bar{h}^w\| \leq \sum_{w \in \mathcal{W}} \frac{\epsilon}{|\mathcal{W}|} = \epsilon,$$

where the first condition follows from the triangle inequality and the second from the definition of \mathcal{C}_w . This completes the proof. □

Funding No funding was received to assist with the preparation of this manuscript.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The author declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Anderson, E., Xu, H., Zhang, D.: Varying confidence levels for CVaR risk measures and minimax limits. *Math. Program.* **180**, 327–370 (2020)
2. Chen, A., Zhou, Z.: The α -reliable mean-excess traffic equilibrium model with stochastic travel times. *Transp. Res. Part B* **44**, 493–513 (2010)
3. Cherukuri, A.: Sample average approximation of CVaR-based Wardrop equilibrium in routing under uncertain costs. In: *IEEE Conference on Decision and Control* (Nice, France, Dec. 2019), pp. 3164–3169
4. Correa, J.R., Stier-Moses, N.E.: Wardrop equilibria. *Encyclopedia of Operations Research and Management Science* (2011)
5. de Mello, T.H., Bayraksan, G.: Monte Carlo sampling-based methods for stochastic optimization. *Surv. Oper. Res. Manag. Sci.* **19**(1), 56–85 (2014)
6. Durrett, R.: *Probability: Theory and Examples*. Series in Statistical and Probabilistic Mathematics, 4th edn. Cambridge University Press (2010)
7. Facchinei, F., Pang, J.-S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, New York (2003)
8. Gürkan, G., Özge, A.Y., Robinson, S.: Sample-path solution of stochastic variational inequalities. *Math. Program.* **84**(2), 313–333 (1999)
9. King, A.J., Rockafellar, R.T.: Asymptotic theory for solutions in statistical estimation and stochastic programming. *Math. Oper. Res.* **18**(1), 148–162 (1993)
10. Kolla, R.K., Prashanth, L.A., Bhat, S.P., Jagannathan, K.: Concentration bounds for empirical conditional value-at-risk: the unbounded case. *Oper. Res. Lett.* **47**(1), 16–20 (2019)
11. Meng, F.W., Sun, J., Goh, M.: Stochastic optimization problems with CVaR risk measure and their sample average approximation. *J. Optim. Theory Appl.* **146**(2), 399–418 (2010)
12. Nikolova, E., Stier-Moses, N.E.: A mean-risk model for the traffic assignment problem with stochastic travel times. *Oper. Res.* **62**(2), 366–382 (2014)
13. Ordóñez, F., Stier-Moses, N.E.: Wardrop equilibria with risk-averse users. *Transp. Sci.* **44**(1), 63–86 (2010)
14. Prakash, A.A., Seshadri, R., Srinivasan, K.K.: A consistent reliability-based user-equilibrium problem with risk-averse users and endogenous travel time correlations: Formulation and solution algorithm. *Transp. Res. Part B* **114**, 171–198 (2018)
15. Ralph, D., Xu, H.: Convergence of stationary points of sample average two-stage stochastic programs: a generalized equation approach. *Math. Oper. Res.* **36**(3), 568–592 (2011)
16. Ramponi, F.A., Campi, M.C.: Expected shortfall: Heuristics and certificates. *Eur. J. Oper. Res.* **267**, 1003–1013 (2018)
17. Shanbhag, U.V.: *Stochastic variational inequality problems: Applications, analysis, and algorithms*. TUTORIALS in Operations Research, pp. 71–107 (2013)

18. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming. SIAM, Philadelphia (2014)
19. Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. *Optimization* **57**(3), 395–418 (2008)
20. Smith, M.J.: The existence, uniqueness and stability of traffic equilibria. *Transp. Res. Part B: Methodol.* **13**(4), 295–304 (1979)
21. Sun, H., Xu, H., Wang, Y.: Asymptotic analysis of sample average approximation for stochastic optimization problems with joint chance constraints via conditional value at risk and difference of convex functions. *J. Optim. Theory Appl.* **161**(1), 257–284 (2014)
22. Temlyakov, V.: Greedy Approximation, Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press (2011)
23. Transportation Networks for Research Core Team. Transportation networks for research. <https://github.com/bstabler/TransportationNetworks>. Accessed 12 Aug 2020
24. Wang, Y., Gao, F.: Deviation inequalities for an estimator of the conditional value-at-risk. *Oper. Res. Lett.* **38**(3), 236–239 (2010)
25. Wu, Y.: Lecture notes on informational-theoretic methods in high-dimensional statistics, Available at: <http://www.stat.yale.edu/~yw562/teaching/598/lec14.pdf> (2016)
26. Xie, Y., Shanbhag, U.V.: On robust solutions to uncertain linear complementarity problems and their variants. *SIAM J. Optim.* **26**(4), 2120–2159 (2016)
27. Xu, H.: Sample average approximation methods for a class of stochastic variational inequality problems. *Asia-Pacific J. Oper. Res.* **27**(1), 103–119 (2010)
28. Xu, H.: Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. *J. Math. Anal. Appl.* **368**, 692–710 (2010)
29. Xu, X., Chen, A., Cheng, L., Yang, C.: A link-based mean-excess traffic equilibrium model under uncertainty. *Transp. Res. Part B* **95**, 53–75 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.