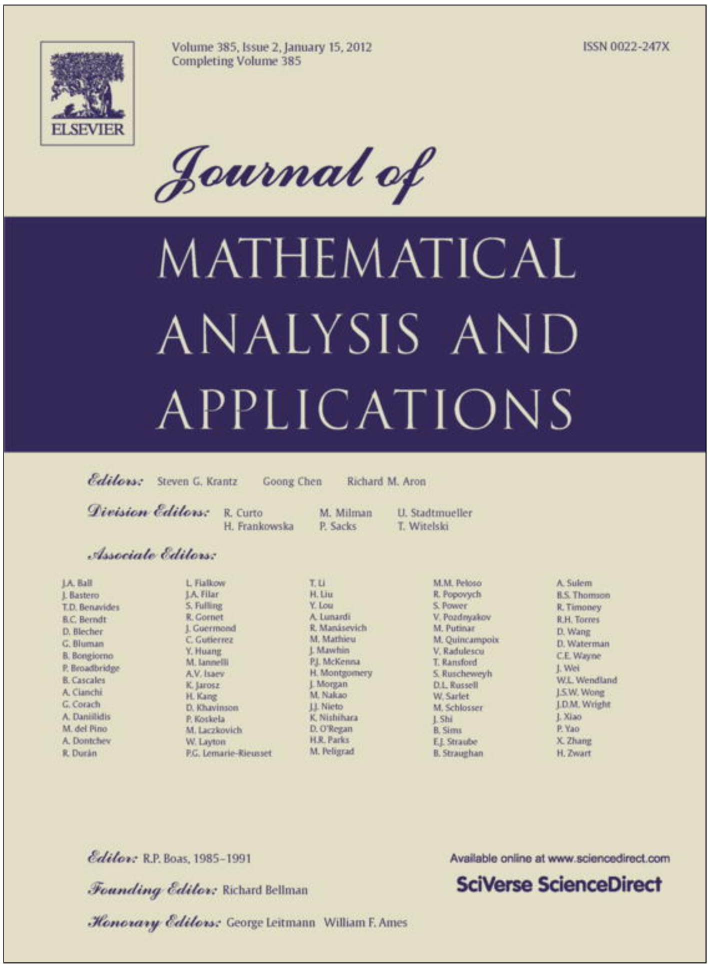
Volume 385, Issue 2, January 15, 2012
Completing Volume 385
ISSN 0022-247X
ELSEVIER
Journal of
MATHEMATICAL
ANALYSIS AND
APPLICATIONS
Editors: Steven G. Krantz Goong Chen Richard M. Aron
Division Editors: R. Curto H. Frankowska M. Milman P. Sacks U. Stadtmueller T. Witelski
Associate Editors:
Editor: R.P. Boas, 1985–1991
Founding Editor: Richard Bellman
Honorary Editors: George Leitmann William F. Ames
Available online at www.sciencedirect.com
SciVerse ScienceDirect

J. Math. Anal. Appl. 385 (2012) 698–708

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

A note on uniform exponential convergence of sample average approximation of random functions

Hailin Sun [a,*,1], Huifu Xu [b]

[a] *Department of Mathematics, Harbin Institute of Technology, Harbin 150001, China*
[b] *School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK*

ARTICLE INFO

Article history:
Received 2 January 2011
Available online 13 July 2011
Submitted by U. Stadtmueller

Keywords:
Random functions
Sample average approximation
Almost H-calm
Uniform exponential convergence

ABSTRACT

Shapiro and Xu (2008) [17] investigated uniform large deviation of a class of Hölder continuous random functions. It is shown under some standard moment conditions that with probability approaching one at exponential rate with the increase of sample size, the sample average approximation of the random function converges to its expected value uniformly over a compact set. This note extends the result to a class of discontinuous functions whose expected values are continuous and the Hölder continuity may be violated for some negligible random realizations. The extension entails the application of the exponential convergence result to a substantially larger class of practically interesting functions in stochastic optimization.

1. Introduction

Over the past decade, a number of new stochastic programming models have been proposed to deal with optimal decision making problems which involve not only uncertain data but also some specific structures such as dominance constraints, equilibrium constraints or hierarchical relationships between decision variables. Typical examples include stochastic mathematical programs with second order dominance constraints [7] and stochastic mathematical programs with equilibrium constraints (SMPEC) [10,17]. These problems are intrinsically nonsmooth and/or nonconvex and therefore one often obtains a stationary point instead of an optimal solution when a numerical method is applied to solve them. One of the most extensively studied methods for such problems is sample average approximation (SAA) method which is also known as sample path optimization, stochastic counterpart or Monte Carlo method [13,16]. The main benefit of SAA is that one may avoid computation of the expected values which are sometimes numerically intractable. The research of SAA often involves asymptotic behavior and/or asymptotic convergence of statistical estimators of optimal values and optimal solutions as sample size increases, see [11,16] for a comprehensive review.

In [17, Theorem 5.1], Shapiro and Xu studied large deviation of a class of Hölder continuous random functions. It is shown under some standard moment conditions that with probability approaching one at exponential rate with the increase of sample size, sample average approximation of a random function converges to its expected value uniformly over a compact set. The result is subsequently used to analyze the exponential rate of convergence of sample average approximation of sharp local minimizers of a two stage SMPEC. A number of generalizations have been made over the past few years primarily driven by the need to investigate the rate of convergence of stationary points rather than optimal solutions of an SAA problem particularly when the underlying functions are nonsmooth. For instance, in an early version of [12], Ralph

* Corresponding author.
E-mail addresses: mathhlsun@gmail.com (H. Sun), h.xu@soton.ac.uk (H. Xu).

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

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2011.07.007

and Xu investigated the uniform exponential rate of convergence for a class of random functions which are merely upper semicontinuous and pointwise calm from above, and applied the generalized result to analyze the convergence of stationary points of a two-stage stochastic program. The result is further consolidated in [19] to accommodate non-independent and identically distributed sampling.

In all these generalized results, a key assumption is that the Hölder continuity or calmness holds at a point for *all* realization of the random variables. Unfortunately this turns out to be rather restrictive for some practically interesting problems. For instance, the Clarke generalized derivative of a piecewise smooth function may not satisfy such a condition. This motivates us to investigate a possible relaxation of the assumption from "all" to "almost every" realization.

This note is written to give details of such relaxation and a new exponential rate of convergence result which extends [19, Theorem 3.1]. The new result is then applied to establish uniform exponential rate of convergence of the Clarke subdifferential of a piecewise smooth random function which is fundamental to many nonsmooth stochastic programming problems.

2. Existing results

Consider a real valued function $\phi(x,\xi):\mathbb{R}^n\times\Xi\to\mathbb{R}$. Let $\xi:\Omega\to\Xi\subset\mathbb{R}^r$ be a vector of random variables defined on probability space $(\Omega,\mathcal{F},P)$. Let

$$\psi(x):=\mathbb{E}\big[\phi(x,\xi)\big]$$

where $\mathbb{E}[\cdot]$ denotes the expectation with respect to the probability measure P. Let $\xi^1,\ldots,\xi^N$ be an independent and identically distributed (iid) sampling of random vector $\xi(\omega)$ and consider the corresponding sample average function

$$\phi_N(x):=\frac{1}{N}\sum_{j=1}^{N}\phi\big(x,\xi^j\big).$$

We discuss the uniform convergence of $\phi_N(x)$ to $\psi(x)$ over a compact set $\mathcal{X}\subset\mathbb{R}^n$ through large deviation theorem [5]. Let

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

denote the moment generating function of the random variable $\phi(x,\xi(\omega))-\psi(x)$.

Shapiro and Xu [17] considered the following conditions:

(C1) For every $x\in\mathcal{X}$ the moment generating function $M_x(t)$ is finite valued for all t in a neighborhood of zero.

(C2) There exist an integrable function $\kappa:\Xi\to\mathbb{R}_+$ and constant $\gamma>0$ such that

$$\big|\phi\big(x',\xi\big)-\phi(x,\xi)\big|\leqslant\kappa(\xi)\big\|x'-x\big\|^{\gamma} \tag{2.1}$$

for all $\xi\in\Xi$ and all $x',x\in\mathcal{X}$.

(C3) The moment generating function $M_\kappa(t)$ of $\kappa(\xi(\omega))$ is finite valued for all t in a neighborhood of zero.

Under these conditions, Shapiro and Xu [17] derived the following convergence result.

Theorem 2.1. *(See [17, Theorem 5.1].) Suppose that conditions* (C1)–(C3) *hold and the set $\mathcal{X}$ is compact. Then for any $\epsilon>0$ there exist positive constants $C=C(\epsilon)$ and $\beta=\beta(\epsilon)$, independent of N, such that*

$$\mathrm{Prob}\Big\{\sup_{x\in\mathcal{X}}\big|\hat{\phi}_N(x)-\psi(x)\big|\geqslant\epsilon\Big\}\leqslant C(\epsilon)e^{-N\beta(\epsilon)}. \tag{2.2}$$

Note that condition (C2) requires $\phi(\cdot,\xi)$ to be globally Hölder continuous uniformly over $\mathcal{X}$ for *all* $\xi\in\Xi$. It is possible to relax the continuity condition so that one can derive exponential convergence for some discontinuous functions.

Definition 2.1 *(H-calmness).* Let $\phi:\mathbb{R}^n\times\Xi\to\mathbb{R}$ be a real valued function and $\xi:\Omega\to\Xi\subset\mathbb{R}^k$ 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) *H-calm at x from above* with modulus $\kappa_x(\xi)$ and order γ_x if $\phi(x,\xi)$ is finite and there exist a (measurable) function $\kappa_x:\Xi\to\mathbb{R}_+$, positive numbers γ_x and δ_x such that

$$\phi\big(x',\xi\big)-\phi(x,\xi)\leqslant\kappa_x(\xi)\big\|x'-x\big\|^{\gamma_x} \tag{2.3}$$

for all $x'\in\mathcal{X}$ with $\|x'-x\|\leqslant\delta_x$ and all $\xi\in\Xi$;

(b) *H-calm at x from below* with modulus $\kappa_x(\xi)$ and order γ_x if $\phi(x,\xi)$ is finite and there exist a (measurable) function $\kappa_x:\Xi\to\mathbb{R}_+$, positive numbers γ_x and δ_x such that

$$\phi\big(x',\xi\big)-\phi(x,\xi)\geqslant-\kappa_x(\xi)\big\|x'-x\big\|^{\gamma_x} \tag{2.4}$$

for all $x'\in\mathcal{X}$ with $\|x'-x\|\leqslant\delta_x$ and all $\xi\in\Xi$;

(c) *H-calm at x* with modulus $\kappa_x(\xi)$ and order γ_x if $\phi(x,\xi)$ is finite and there exist a (measurable) function $\kappa_x : \Xi \to \mathbb{R}_+$, positive numbers γ_x and δ_x such that

$$\left|\phi(x',\xi) - \phi(x,\xi)\right| \leqslant \kappa_x(\xi)\left\|x' - x\right\|^{\gamma_x} \tag{2.5}$$

for all $x' \in \mathcal{X}$ with $\|x' - x\| \leqslant \delta_x$ and all $\xi \in \Xi$.

ϕ is said to be *H-calm from above*, *H-calm from below*, *H-calm* on set $\mathcal{X}$ if the respective properties stated above hold at every point of $\mathcal{X}$. When $\gamma = 1$, ϕ is said to be *calm from above*, *calm from below* and *calm* respectively.

The following convergence results are summarized from Lemma 3.1 in an earlier version of [12] and [19, Theorem 3.1].

Theorem 2.2. *Let $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real valued function and $\mathcal{X}$ a compact subset of $\mathbb{R}^n$. Assume:* (a) *condition* (C1) *holds*; (b) *$\psi(x)$ is Hölder continuous on $\mathcal{X}$. Then the following statements hold.*

(i) *If $\phi(\cdot,\xi)$ is H-calm from above on $\mathcal{X}$ with modulus $\kappa(\xi)$ and order γ and the moment generating function $\mathbb{E}[e^{\kappa(\xi)t}]$ 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*

$$\mathrm{Prob}\left\{\sup_{x\in\mathcal{X}}\big(\psi_N(x) - \psi(x)\big) \geqslant \epsilon\right\} \leqslant c(\epsilon)e^{-N\beta(\epsilon)}. \tag{2.6}$$

(ii) *If $\phi(\cdot,\xi)$ is H-calm from below on $\mathcal{X}$ with modulus $\kappa(\xi)$ and order γ and the moment generating function $\mathbb{E}[e^{\kappa(\xi)t}]$ 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*

$$\mathrm{Prob}\left\{\inf_{x\in\mathcal{X}}\big(\psi_N(x) - \psi(x)\big) \leqslant -\epsilon\right\} \leqslant c(\epsilon)e^{-N\beta(\epsilon)}. \tag{2.7}$$

(iii) *If $\phi(\cdot,\xi)$ is H-calm on $\mathcal{X}$ with modulus $\kappa(\xi)$ and order γ and the moment generating function $\mathbb{E}[e^{\kappa(\xi)t}]$ 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*

$$\mathrm{Prob}\left\{\sup_{x\in\mathcal{X}}\big|\psi_N(x) - \psi(x)\big| \geqslant \epsilon\right\} \leqslant c(\epsilon)e^{-N\beta(\epsilon)}. \tag{2.8}$$

The H-calmness weakens the condition on continuity of random function $\phi(x,\xi)$. However the requirements for the calmness to hold for all ξ and the expected value to be continuous seem to be too strong. Consider a simple random function

$$\phi\big(x,\xi(\omega)\big) := \begin{cases} 1, & \text{for } x \geqslant \xi,\\ 0, & \text{for } x < \xi, \end{cases} \tag{2.9}$$

where $x \in \mathbb{R}$ and ξ is a random variable with continuous distribution with support Ξ. It is easy to observe that calmness is violated in a neighborhood of the $x = \xi$ line. However some numerical experiments show that the sample average approximation of this function converges to its expected value uniformly over a compact set. This motivates us to consider a weaker condition, namely to allow the calmness to be violated for some negligible subset of Ξ. We will give a precise description of this in the next section.

3. Almost calmness and the extended exponential convergence results

In this section, we weaken the calmness condition and derive the uniform exponential convergence of random functions under the weakened conditions. Let us start with a definition corresponding to Definition 2.1.

Definition 3.1 *(Almost H-clamness).* Let $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real valued function and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ 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-clam 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, $\delta_x(\epsilon)$, C and an open set $\Delta_x(\epsilon) \subset \Xi$ such that

$$\mathrm{Prob}\big(\xi \in \Delta_x(\epsilon)\big) \leqslant C\epsilon \tag{3.10}$$

and

$$\phi\big(x',\xi\big) - \phi(x,\xi) \leqslant \kappa_x(\xi)\left\|x' - x\right\|^{\gamma_x} \tag{3.11}$$

for all $\xi \in \Xi\backslash\Delta_x(\epsilon)$ and all $x' \in B(x, \delta_x(\epsilon)) \cap \mathcal{X}$, here and later on $B(x,\delta)$ denotes the δ-neighborhood of x;

(b) *almost H-clam 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, $\delta_x(\epsilon)$ and an open set $\Delta_x(\epsilon) \subset \Xi$ such that (3.10) holds and

$$\phi(x', \xi) - \phi(x, \xi) \geqslant -\kappa_x(\xi)\|x' - x\|^{\gamma_x} \tag{3.12}$$

for all $\xi \in \Xi \backslash \Delta_x(\epsilon)$ and all $x' \in B(x, \delta_x(\epsilon)) \cap \mathcal{X}$;

(c) *almost H-clam 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, $\delta_x(\epsilon)$ and an open set $\Delta_x(\epsilon) \subset \Xi$ such that (3.10) holds and

$$\left|\phi(x', \xi) - \phi(x, \xi)\right| \leqslant \kappa_x(\xi)\|x' - x\|^{\gamma_x} \tag{3.13}$$

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

ϕ is said to be almost H-clam from above, almost H-clam from below, almost H-clam on set $\mathcal{X}$ if the respective properties stated above hold at every point x of $\mathcal{X}$. ϕ is said to be almost H-clam from above, almost H-clam from below, almost H-clam *uniformly* over set $\mathcal{X}$ if there exist $\kappa(\xi)$, $\delta(\epsilon)$, γ, all of which being independent of x, such that the respective properties stated above hold at every point x of $\mathcal{X}$. When $\gamma = 1$, ϕ is said to be almost clam from above, almost clam from below, almost clam respectively.

The key idea in the preceding definition is to allow calmness condition to be violated for some random realization with negligible probability. Consider the function defined in (2.9). Let κ and γ be any positive numbers. Then the function is almost calm at any point $x \in \mathbb{R}$ with modulus κ and order γ. Indeed for any $\epsilon > 0$, there exist $\delta = \frac{1}{2}\epsilon$ and $\Delta_x(\epsilon) = (x - \frac{1}{2}\epsilon, x + \frac{1}{2}\epsilon) \cap \Xi$, such that the (3.13) holds for all $\xi \in \Xi \backslash \Delta_x(\epsilon)$ and $x' \in (x - \frac{1}{2}\epsilon, x + \frac{1}{2}\epsilon) \cap \mathcal{X}$, where x can be any point in $\mathcal{X}$ and Ξ denotes the support of ξ. Indeed the function is uniformly almost calm over $\mathbb{R}$ as the constants κ, γ and $\delta(\epsilon)$ are independent of x.

The following result is from the Gärtner–Ellis theorem, see for instance [4,5].

Lemma 3.1 *(Pointwise exponential convergence). Let $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real valued function and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ a random vector defined on probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{X} \subset \mathbb{R}^n$ be a subset of $\mathbb{R}^n$ and $x \in \mathcal{X}$. If the moment generating function*

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

is finite for t close to 0*, then for every fixed $x \in \mathcal{X}$ and small positive number $\epsilon > 0$*

$$\text{Prob}\left\{\psi_N(x) - \psi(x) \leqslant -\epsilon\right\} \leqslant e^{-NI_x(-\epsilon)}$$

and

$$\text{Prob}\left\{\psi_N(x) - \psi(x) \geqslant \epsilon\right\} \leqslant e^{-NI_x(\epsilon)}$$

for N sufficiently large, where

$$I_x(z) := \sup_{t \in R}\left\{zt - \log M_x(t)\right\}$$

and both $I(-\epsilon)$ and $I(\epsilon)$ are positive.

Let $\xi : \Omega \to \Xi$ be a random variable and $\Delta \subset \Xi$ be an open set. Let

$$\eta(\xi(\omega)) := \begin{cases} 1, & \text{if } \xi(\omega) \in \Delta, \\ 0, & \text{if } \xi(\omega) \notin \Delta. \end{cases} \tag{3.14}$$

Then $\eta(\xi(\omega))$ is a random variable (depending on Δ) and $\mathbb{E}[\eta(\xi(\omega))] = \text{Prob}\{\xi(\omega) \in \Delta\}$. The following proposition states pointwise exponential convergence of random function $\phi(x, \xi)\eta(\xi)$.

Proposition 3.1. *Let $\phi(x, \xi)$ be defined as in Lemma* 3.1 *and $\eta : \Omega \to \{0, 1\}$ be defined by* (3.14)*. Assume that the moment generating function of $\phi(x, \xi)$ is finite valued for t close to* 0*. Then the following assertions hold.*

(i) *The moment generating function of $\phi(x, \xi)\eta(\xi)$ is finite valued for t close to* 0.
(ii) *Let $\tilde{\psi}(x) = \mathbb{E}[\phi(x, \xi)\eta(\xi)]$,*

$$\tilde{\phi}_N(x) = \frac{1}{N}\sum_{k=1}^{N} \phi(x, \xi^k)\eta^k,$$

where $\eta^k = \eta(\xi^k)$, for $k = 1, \ldots, N$. Let

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

and

$$\tilde{I}_x(z) := \sup_{t\in R}\{zt - \log \tilde{M}_x(t)\}.$$

Then for every small $\epsilon > 0$, one has

$$\text{Prob}\{\tilde{\phi}_N(x) - \tilde{\psi}(x) \leqslant -\epsilon\} \leqslant e^{-N\tilde{I}_x(-\epsilon)} \tag{3.15}$$

and

$$\text{Prob}\{\tilde{\phi}_N(x) - \tilde{\psi}(x) \geqslant \epsilon\} \leqslant e^{-N\tilde{I}_x(\epsilon)} \tag{3.16}$$

for N sufficiently large, where

$$\min(\tilde{I}_x(-\epsilon), \tilde{I}_x(\epsilon)) > 0.$$

Proof. Part (i). Using conditional expectation, we have

$$\begin{aligned}\tilde{M}_x(t) &= \mathbb{E}\big[e^{t(\phi(x,\xi)\eta(\xi)-\tilde{\psi}(x))}\big]\\ &= \mathbb{E}\big[e^{t(\phi(x,\xi)\eta(\xi)-\tilde{\psi}(x))} \mid \eta(\xi)=1\big]\text{Prob}(\eta(\xi)=1) + \mathbb{E}\big[e^{t(\phi(x,\xi)\eta(\xi)-\tilde{\psi}(x))} \mid \eta(\xi)=0\big]\text{Prob}(\eta(\xi)=0)\\ &= e^{-\tilde{\psi}(x)t}\big(\mathbb{E}\big[e^{\phi(x,\xi)t}\big]\text{Prob}(\xi\in\Delta) + \text{Prob}(\xi\notin\Delta)\big)\\ &= e^{(\psi(x)-\tilde{\psi}(x))t}\big(\mathbb{E}\big[e^{(\phi(x,\xi)-\psi(x))t}\big]\text{Prob}(\xi\in\Delta) + e^{-\psi(x)t}\,\text{Prob}(\xi\notin\Delta)\big)\\ &< \infty\end{aligned}$$

for t close to 0.

Part (ii). The conclusion follows from part (i) and Lemma 3.1. □

We are now ready to state the main results in this section.

Theorem 3.1. *Let $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real valued lower semicontinuous function and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ a random vector defined on probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{X} \subset \mathbb{R}^n$ be a compact subset of $\mathbb{R}^n$. Assume:* (a) *condition* (C1) *in Section 2 holds,* (b) *$\psi(x)$ is continuous on $\mathcal{X}$,* (c) *there exists a positive number L such that $|\phi(x,\xi)| \leqslant L$. Then the following statements hold.*

(i) *If $\phi(\cdot,\xi)$ is almost H-clam from above on $\mathcal{X}$ with modulus $\kappa(\xi)$ and order γ, and the moment generating function $\mathbb{E}[e^{\kappa(\xi)t}]$ 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*

$$\text{Prob}\Big\{\sup_{x\in\mathcal{X}}(\psi_N(x) - \psi(x)) \geqslant \epsilon\Big\} \leqslant c(\epsilon)e^{-N\beta(\epsilon)}. \tag{3.17}$$

(ii) *If $\phi(\cdot,\xi)$ is almost H-clam from below on $\mathcal{X}$ with modulus $\kappa(\xi)$ and order γ, and the moment generating function $\mathbb{E}[e^{\kappa(\xi)t}]$ 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*

$$\text{Prob}\Big\{\inf_{x\in\mathcal{X}}(\psi_N(x) - \psi(x)) \leqslant -\epsilon\Big\} \leqslant c(\epsilon)e^{-N\beta(\epsilon)}. \tag{3.18}$$

(iii) *If $\phi(\cdot,\xi)$ is almost H-clam on $\mathcal{X}$ with modulus $\kappa(\xi)$ and order γ, and the moment generating function $\mathbb{E}[e^{\kappa(\xi)t}]$ 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*

$$\text{Prob}\Big\{\sup_{x\in\mathcal{X}}|\psi_N(x) - \psi(x)| \geqslant \epsilon\Big\} \leqslant c(\epsilon)e^{-N\beta(\epsilon)}. \tag{3.19}$$

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

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

$$\text{Prob}\{\psi_N(x) - \psi(x) \geqslant \epsilon\} \leqslant e^{-NI_x(\epsilon)}, \tag{3.20}$$

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 $i(x) \in \{1,\dots,M\}$ such that $\|x - \bar{x}_{i(x)}\| \leqslant \nu$. 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

$$|\psi(x) - \psi(\bar{x}_{i(x)})| \leqslant \frac{\epsilon}{4} \tag{3.21}$$

for any $x \in \mathcal{X}$. On the other hand, since $\phi(x,\xi)$ is almost H-clam from above on $\mathcal{X}$, then for every $x \in \mathcal{X}$, there exist an open set $\Delta_x(\epsilon) \subset \Xi$ and positive number δ_x such that

$$\mathrm{Prob}\{\xi \in \Delta_x(\epsilon)\} \leqslant \frac{\epsilon}{16L} \tag{3.22}$$

and

$$\phi(x',\xi) - \phi(x,\xi) \leqslant \kappa(\xi)\|x'-x\|^\gamma$$

for all $\xi \notin \Delta_x(\epsilon)$ and $\|x'-x\| \leqslant \delta_x$. Through the finite covering theorem, this implies that our ν-net can be chosen properly so that

$$\phi(x,\xi) - \phi(\bar{x}_{i(x)},\xi) \leqslant \kappa(\xi)\|x-\bar{x}_{i(x)}\|^\gamma \tag{3.23}$$

for all $\xi \notin \Delta_{\bar{x}_{i(x)}}(\epsilon)$. Let

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

By (3.23), we have that

$$\bar{\psi}_N(x) - \bar{\psi}_N(\bar{x}_{i(x)}) \leqslant \frac{1}{N}\sum_{\xi^k \notin \Delta_{\bar{x}_{i(x)}}(\epsilon)} \kappa(\xi^k)\nu^\gamma \leqslant \kappa^N \nu^\gamma \tag{3.24}$$

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 [5], we have that for any $L' > \mathbb{E}[\kappa(\xi(\omega))]$, there exists a positive constant λ such that

$$\mathrm{Prob}\{\kappa^N \geqslant L'\} \leqslant e^{-N\lambda}$$

and hence

$$\mathrm{Prob}\left\{\kappa^N \nu^\gamma \geqslant \frac{\epsilon}{4}\right\} \leqslant e^{-N\lambda} \tag{3.25}$$

for some $\lambda > 0$ (by setting $\frac{\epsilon}{4\nu^\gamma} \geqslant \mathbb{E}[\kappa(\xi(\omega))]$). On the other hand, by using notation η defined by (3.14) and condition (c) of this theorem, we have

$$\begin{aligned}\tilde{\psi}_N(x) - \tilde{\psi}_N(\bar{x}_{i(x)}) &= \frac{1}{N}\sum_{\xi^k \in \Delta_{\bar{x}_{i(x)}}(\epsilon)} \left(\phi(x,\xi^k) - \phi(\bar{x}_{i(x)},\xi^k)\right) = \frac{1}{N}\sum_{k=1}^N \left(\phi(x,\xi^k) - \phi(\bar{x}_{i(x)},\xi^k)\right)\eta^k \\ &\leqslant \frac{1}{N}\sum_{k=1}^N 2L\eta^k, \end{aligned} \tag{3.26}$$

where

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

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

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

where $\tilde{I}(z)$ is the rate function of $L\eta(\xi)$, and by Proposition 3.1(ii), $\tilde{I}(-\epsilon/16) > 0$ and $\tilde{I}(\epsilon/16) > 0$. On the other hand, by (3.22),

$$\mathbb{E}[2L\eta(\xi)] \leqslant 2L\,\mathrm{Prob}(\xi \in \Delta_{\bar{x}_{i(x)}}(\epsilon)) \leqslant \frac{\epsilon}{8}.$$

A combination of the above two inequalities yields

$$\mathrm{Prob}\left\{\frac{1}{N}\sum_{k=1}^N 2L\eta^k \geqslant \frac{\epsilon}{4}\right\} \leqslant e^{-N\tilde{I}(\epsilon/16)} \tag{3.27}$$

for N sufficiently large.

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

$$\mathrm{Prob}\Big\{\max_{1 \leqslant i \leqslant M} Z_i \geqslant \epsilon\Big\} \leqslant \sum_{i=1}^{M} \mathrm{Prob}\{Z_i \geqslant \epsilon\}.$$

Together with (3.20), this implies that

$$\mathrm{Prob}\Big\{\max_{1 \leqslant i \leqslant M} Z_i \geqslant \epsilon\Big\} \leqslant \sum_{i=1}^{M} e^{-N I_{\bar{x}_i}(\epsilon)}. \tag{3.28}$$

Combining (3.21), (3.24) and (3.26), we obtain

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

Therefore

$$\begin{aligned}
&\mathrm{Prob}\Big\{\sup_{x \in \mathcal{X}} \big(\psi_N(x) - \psi(x)\big) \geqslant \epsilon\Big\} \\
&\quad \leqslant \mathrm{Prob}\bigg\{\kappa^N \nu^\gamma + \frac{1}{N}\sum_{k=1}^{N} 2L\eta^k + \max_{1 \leqslant i \leqslant M}\big(\psi_N(\bar{x}_i) - \psi(\bar{x}_i)\big) \geqslant \frac{3\epsilon}{4}\bigg\}.
\end{aligned}$$

By (3.25), (3.27) and (3.28), we have

$$\begin{aligned}
\mathrm{Prob}\Big\{\sup_{x \in \mathcal{X}} \big(\psi_N(x) - \psi(x)\big) \geqslant \epsilon\Big\} &\leqslant e^{-N\lambda} + \mathrm{Prob}\bigg\{\frac{1}{N}\sum_{k=1}^{N} 2L\eta^k \geqslant \frac{\epsilon}{4}\bigg\} + \mathrm{Prob}\Big\{\max_{1 \leqslant i \leqslant M}\big(\psi_N(\bar{x}_i) - \psi(\bar{x}_i)\big) \geqslant \frac{\epsilon}{4}\Big\} \\
&\leqslant e^{-N\lambda} + e^{-N\tilde{I}(\epsilon/16)} + \sum_{i=1}^{M} e^{-N I_{\bar{x}_i}(\frac{\epsilon}{4})},
\end{aligned}$$

which implies (3.17) as the 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. □

It is important to note that Theorem 3.1 requires $\psi(x)$ to be continuous. It is unclear whether or not it can be weakened to piecewise continuous. In [12], Ralph and Xu investigated sample average approximation of stochastic generalized equations where the underlying functions are set-valued and stochastic piecewise Hausdorff continuous. Under the condition that the expected value of the set-valued mapping is piecewise continuous, they derived piecewise uniform exponential convergence of sample average approximated set-valued mapping to its true counterpart, see [12, Section 4]. It will be interesting to incorporate Theorem 3.1 with those developed in [12, Section 4], we leave this for our future work.

Remark 3.1. Similarly to the discussions in [17], we may estimate the sample size. To this end, we assume that there exists a constant $\sigma > 0$ such that for all $x \in \mathcal{X}$,

$$\mathbb{E}\big[e^{(\phi(x,\xi) - \mathbb{E}[\phi(x,\xi)])t}\big] \leqslant \exp\big\{\sigma^2 t^2/2\big\}, \quad \forall t \in \mathbb{R} \tag{3.29}$$

and

$$\mathbb{E}\big[e^{(L\eta(\xi) - \mathbb{E}[L\eta(\xi)])t}\big] \leqslant \exp\big\{\sigma^2 t^2/2\big\}, \quad \forall t \in \mathbb{R}, \tag{3.30}$$

where L and $\eta(\xi)$ are defined as in Theorem 3.1. Note that equality in (3.29) and (3.30) holds if random variables $\phi(x,\xi) - \mathbb{E}[\phi(x,\xi)]$ and $L\eta(\xi) - \mathbb{E}[L\eta(\xi)]$ satisfy normal distribution with variance σ^2, see a discussion in [17]. From Theorem 3.1,

$$\mathrm{Prob}\Big\{\sup_{x \in \mathcal{X}} \big(\psi_N(x) - \psi(x)\big) \geqslant \epsilon\Big\} \leqslant e^{-N\lambda} + e^{-N\tilde{I}(\epsilon/16)} + \sum_{i=1}^{M} e^{-N I_{\bar{x}_i}(\frac{\epsilon}{4})}, \tag{3.31}$$

where $\tilde{I}$ is the rate function of $L\eta(\xi)$ and $I_{\bar{x}_i}$ is the rate function of $\phi(\bar{x}_i, \xi)$. By (3.29) and (3.30), it is easy to derive through the definition of the rate function that

$$\tilde{I}\left(\frac{\epsilon}{16}\right) \geqslant \frac{\epsilon^2}{512\sigma^2}$$

and

$$I_{\bar{x}_i}\left(\frac{\epsilon}{4}\right) \geqslant \frac{\epsilon^2}{32\sigma^2}.$$

Substituting the estimates into (3.31), we have

$$\mathrm{Prob}\Big\{\sup_{x\in\mathcal{X}}\big(\psi_N(x) - \psi(x)\big) \geqslant \epsilon\Big\} \leqslant e^{-N\lambda} + e^{-\frac{N\epsilon^2}{512\sigma^2}} + Me^{-\frac{N\epsilon^2}{32\sigma^2}}. \tag{3.32}$$

Let ϵ be sufficiently small such that $\lambda \geqslant \frac{\epsilon^2}{512\sigma^2}$ and $\beta \in (0,1)$. By (3.32), it is easy to verify that

$$\mathrm{Prob}\Big\{\sup_{x\in\mathcal{X}}\big(\psi_N(x) - \psi(x)\big) \geqslant \epsilon\Big\} \leqslant \beta \tag{3.33}$$

for

$$N \geqslant \frac{512\sigma^2}{\epsilon^2}\left[\ln(M+2) + \ln\left(\frac{1}{\beta}\right)\right].$$

In what follows, we estimate M. From the proof of Theorem 3.1, $M \leqslant (O(1)D(\frac{4\mathbb{E}[\kappa(\xi)]}{\epsilon})^{\frac{1}{\gamma}})^n$, which means that (3.33) holds for

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

4. Application

In this section, we apply the theory derived in the preceding section to study sample average approximation of Clarke subdifferentials. We focus on a particular function defined as follows:

$$g(x,\xi) := \max\big(f\big(x, \xi(\omega)\big), 0\big), \tag{4.34}$$

where $f : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}$ is Lipschitz continuous and differentiable w.r.t. x for every ξ and $\xi : \Omega \to \Xi$ is a vector of random variables defined on probability $(\Omega, \mathcal{F}, P)$ with support $\Xi \subset \mathbb{R}^q$. The reason that we consider this particular random function is that nonsmoothness in many practical stochastic optimization problems arises from max operations, see for instance [4,6,8].

Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous function. The *Clarke subdifferential* (also known as generalized gradient) of F at $x \in \mathbb{R}^n$ is defined as

$$\partial F(x) := \mathrm{conv}\Big\{\lim_{y\in D_F,\, y\to x} \nabla F(y)\Big\},$$

where D_F denotes the set of points near x at which F is Fréchet differentiable, $\nabla F(y)$ denotes the usual gradient of F and "conv" denotes the convex hull of a set. The Clarke's generalized directional derivative of F at x for a given direction $h \in \mathbb{R}^n$ is defined as:

$$F^o(x;h) = \limsup_{x'\to x,\, \tau\downarrow 0} \frac{1}{\tau}\big(F\big(x' + \tau h\big) - F\big(x'\big)\big).$$

It is well known that the Clarke generalized gradient $\partial F(x)$ is a convex compact set and as a set-valued mapping $\partial F(\cdot)$ is upper semicontinuous in the sense of Berge, that is, for any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\partial F(x + \delta\mathcal{B}) \subset \partial F(x) + \epsilon\mathcal{B}$$

where $\mathcal{B}$ denotes the unit ball in an appropriate space. Moreover, Clarke generalized directional derivative is the support of Clarke subdifferential, that is

$$F^o(x;h) = \sup_{\zeta\in\partial F(x)} \zeta^T h.$$

See Chapter 2 in [3] for details.

In what follows, we investigate the conditions under which: (a) $\mathbb{E}[g_x^o(x,\xi;h)]$ is continuous, (b) $g_x^o(x,\xi;h)$ is almost calm. Note that since f is assumed to be Lipschitz continuous w.r.t. x for each ξ, then $g(\cdot,\xi)$ is also Lipschitz continuous and hence $g_x^o(x,\xi)$ is well defined for each x.

Let $x \in \mathbb{R}^n$. Define the set

$$\Theta(x) := \{\xi:\ f(x,\xi) = 0\}.$$

Obviously $\Theta(x)$ contains the set of points ξ such that $g(\cdot,\xi)$ is not differentiable at x.

Proposition 4.1. *Consider function* (4.34). *Let $\mathcal{X} \subset \mathbb{R}^n$ be a compact subset of $\mathbb{R}^n$. Assume*:

(a) *$f(x,\xi)$ is twice continuously differentiable w.r.t. x for almost every $\xi \in \Xi$;*
(b) *$\nabla_x f(\cdot,\cdot)$ is locally Lipschitz continuous with an integrable modulus $\kappa(\xi)$;*
(c) *the Lebesgue measure of $\Theta(x)$ on Ξ is* 0;
(d) *for any $\epsilon > 0$ and any fixed $x \in \mathcal{X}$, there exists an open set $\Theta^\epsilon(x)$ (depending on both x and ϵ) such that $\Theta(x) \subset \Theta^\epsilon(x)$ and $\mu(\Theta^\epsilon(x) \cap \Xi) \leqslant \epsilon$, where μ denotes the Lebesgue measure on Ξ.*

Then

(i) *$\mathbb{E}[g_x^o(x,\xi;h)]$ is a continuous function for every $h \in \mathbb{R}^n$;*
(ii) *if Ξ is a compact set, then $g_x^o(x,\xi;h)$ is almost H-clam with modulus $\kappa(\xi)$ and order* 1 *on $\mathcal{X}$;*
(iii) *if Ξ is bounded, then the moment generation function $M_x(t) := \mathbb{E}[e^{(g_x^o(x,\xi;h)-\mathbb{E}[g_x^o(x,\xi;h)])t}]$ is finite valued for t close to* 0.

Proof. Part (i). This is a well-known result. Indeed, in this case $\mathbb{E}[g(x,\xi)]$ is continuously differentiable, see for instance [8, Theorem 1].

Part (ii). For any fixed $\epsilon > 0$ and $\bar{x} \in \mathcal{X}$, by condition (d), there exists an open subset $\Theta^\epsilon(\bar{x})$ such that $\Theta(\bar{x}) \subset \Theta^\epsilon(\bar{x})$ and $\mu(\Theta^\epsilon(\bar{x}) \cap \Xi) \leqslant \epsilon$. Let $\bar{\xi} \notin \Theta^\epsilon(\bar{x})$. Then $f(\bar{x},\bar{\xi}) \neq 0$. We only consider the case that $f(\bar{x},\bar{\xi}) > 0$ as the case when $f(\bar{x},\bar{\xi}) < 0$ can be dealt with in the same way. Under condition (b), we can find a δ-neighborhood of $(\bar{x},\bar{\xi})$ (depending on $\bar{x}$ and $\bar{\xi}$), denoted by $B((\bar{x},\bar{\xi}),\delta_{\bar{x},\bar{\xi}})$, such that for all $(x,\xi) \in B((\bar{x},\bar{\xi}),\delta_{\bar{x},\bar{\xi}}) \cap \mathcal{X} \times \Xi$, $f(x,\xi) > 0$ and

$$\left|g_x^o(x,\xi;h) - g_x^o(\bar{x},\xi;h)\right| = \left|\nabla_x f(x,\xi)^T h - \nabla_x f(\bar{x},\xi)^T h\right| \leqslant \kappa(\xi)\|x - \bar{x}\|. \tag{4.35}$$

Since $\Xi \backslash \Theta^\epsilon(\bar{x})$ is compact, we claim through the finite covering theorem that there exists a unified $\delta_{\bar{x}} > 0$ such that (4.35) holds for all $x \in B(\bar{x},\delta_{\bar{x}})$ and all $\xi \in \Xi \backslash \Theta^\epsilon(\bar{x})$. This shows that $g_x^o(x,\xi;h)$ is almost H-calm with modulus $\kappa(\xi)$ and order 1 over $\mathcal{X}$.

Part (iii) is obvious when Ξ is a bounded set. □

Remark 4.1. It might be interesting to ask how strong condition (d) is in the preceding proposition. We claim that verifiable sufficient conditions are: (a) $\Theta(x)$ is compact and (b)

$$\nabla_\xi f(x,\xi) \neq 0 \tag{4.36}$$

for $\xi \in \Theta(x)$. To see this, let $\bar{\xi} \in \Theta(x)$. Then $f(x,\bar{\xi}) = 0$ and by assumption, $\nabla_\xi f(x,\bar{\xi}) \neq 0$. Note that ξ is an r-dimensional vector in $\mathbb{R}^r$. Let us write $\xi = (\xi_1,\ldots,\xi_r)^T$. Assume without loss of generality that $\frac{\partial f(x,\bar{\xi})}{\partial \xi_1} \neq 0$. By the classical implicit function theorem, there exist a neighborhood of point $(x,\bar{\xi})$, denoted by $B((x,\bar{\xi}),\delta_{x,\bar{\xi}})$, and a unique implicit function $\xi_1(x,\xi_{-1})$, where $\xi_{-1} = (\xi_2,\ldots,\xi_r)^T$, such that $\xi_1(x,\bar{\xi}_{-1}) = \bar{\xi}_1$ and

$$f\left(x',\xi_1\left(x',\xi_{-1}\right),\xi_{-1}\right) = 0$$

for (x',ξ_{-1}) close to $(x,\bar{\xi}_{-1})$.

Let $x' = x$ be fixed. We consider the implicit function $q(\xi_{-1}) := \xi_1(x,\xi_{-1})$ defined on Ξ. The graph of the function is an $r-1$ manifold on $\mathbb{R}^r$ on Ξ. We claim that there exists a finite number of such manifolds in Ξ. Assume for a contradiction that there exists an infinite number of such manifolds, e.g., $\{\xi_1^k(x,\cdot)\}$. Let $\xi^k := (\xi_1^k(x,\xi_{-1}^k),\xi_{-1}^k)$ be a point on the k-th manifold. Since Ξ is compact, by taking a subsequence if necessary, we may assume that $\{\xi_1^k(x,\xi_{-1}^k)\} \to \xi_1^*(x,\xi_{-1}^*)$, which means that in a neighborhood of $\xi^* = (\xi_1^*,\xi_{-1}^*)$, there exists an infinite number of manifolds. This is impossible under condition (4.36) as by the implicit function theorem there exists only a unique such manifold in the neighborhood.

The discussion above shows that the Lebesgue measure of $\Theta(x)$ on Ξ is zero as the Lebesgue measure of each manifold relative to Ξ is 0. Moreover, there exists an open set $\Theta^\epsilon(x)$ such that $\Theta(x) \subset \Theta^\epsilon(x)$ and $\mu(\Theta^\epsilon(x)) \to 0$ as $\epsilon \downarrow 0$.

We now move on to investigate sample average approximation of $\mathbb{E}[\partial_x g(x,\xi)]$, where $\partial_x g(x,\xi)$ denotes the Clarke subdifferential of g w.r.t. x and the expected value of the random set-valued mapping is in the sense of Aumann [1], that is,

$$\mathbb{E}\big[\partial_x g(x,\xi)\big] := \left\{ \int_{\Xi} \zeta P(d\xi)\colon \ \zeta \text{ is a Bochner's integrable selection from } \partial_x g(x,\xi) \right\}.$$

Let $\xi^1,\ldots,\xi^N$ be iid sampling of ξ. We consider the sample average approximation

$$G_N(x) := \frac{1}{N}\sum_{i=1}^{N} g\big(x,\xi^i\big) = \frac{1}{N}\sum_{i=1}^{N} \max\big(f\big(x,\xi^i\big),0\big). \tag{4.37}$$

It is well known that if $g(x,\xi)$ is integrably bounded and Lipschitz continuous w.r.t. x, then $G_N(x)$ converges to $\mathbb{E}[g(x,\xi)]$ uniformly over any compact set as $N\to\infty$, see [14, Lemma A1]. Here we are interested in the approximation of subdifferentials. Let

$$\mathcal{A}G_N(x) := \frac{1}{N}\sum_{i=1}^{N} \partial_x g\big(x,\xi^i\big).$$

It is well known that $\partial G_N(x) \subset \mathcal{A}G_N(x)$ and equality holds when $g(\cdot,\xi^i)$ is Clarke regular at x for $i=1,\ldots,N$.

For the simplicity of notation, let $G(x) := \mathbb{E}[g(x,\xi(\omega))]$. Let $d(x,\mathcal{C}) := \inf_{x'\in\mathcal{C}} \|x-x'\|$ which is the distance from point x to $\mathcal{C}$. For two nonempty compact sets $\mathcal{C}$ and $\mathcal{D}$, $\mathbb{D}(\mathcal{C},\mathcal{D}) := \sup_{x\in\mathcal{C}} d(x,\mathcal{D})$ denotes the deviation from set $\mathcal{C}$ to set $\mathcal{D}$ (also known as *excess* of $\mathcal{C}$ over $\mathcal{D}$), and $\mathbb{H}(\mathcal{C},\mathcal{D})$ denotes the Pompeiu–Hausdorff distance between the two sets, that is, $\mathbb{H}(\mathcal{C},\mathcal{D}) := \max(\mathbb{D}(\mathcal{C},\mathcal{D}),\mathbb{D}(\mathcal{D},\mathcal{C}))$. We use $\mathcal{C}+\mathcal{D}$ to denote the Minkowski addition of the two sets, that is, $\{x+x'\colon x\in\mathcal{C},\ x'\in\mathcal{D}\}$. We are now ready to present the uniform exponential rate of convergence of $\partial G_N(x)$.

Theorem 4.1. *Let $f(x,\xi)$ and $g(x,\xi)$ be defined as in* (4.34)*. Let $\mathcal{X}$ be a compact subset of $\mathbb{R}^n$. Then the following assertions hold.*

(i) *If conditions* (a)–(b) *of Proposition* 4.1 *hold, then*

$$\lim_{N\to\infty}\sup_{x\in\mathcal{X}} \mathbb{H}\big(\mathcal{A}G_N(x),\partial G(x)\big) = 0. \tag{4.38}$$

If, in addition, condition (c) *holds, then $G(x)$ is continuously differentiable and $\partial G(x) = \nabla G(x) = \mathbb{E}[\partial_x g(x,\xi)]$. Moreover*

$$\lim_{N\to\infty}\sup_{x\in\mathcal{X}} \mathbb{H}\big(\partial G_N(x),\nabla G(x)\big) = 0. \tag{4.39}$$

(ii) *If $f(x,\xi)$ satisfies conditions* (a), (b) *and* (d) *of Proposition* 4.1 *and Ξ is bounded, then for every small positive number $\epsilon>0$, there exist $\hat{c}>0$ and $\hat{\beta}(\epsilon)>0$, independent of N, such that*

$$\mathrm{Prob}\Big\{\sup_{x\in\mathcal{X}} \mathbb{D}\big(\partial G_N(x),\partial G(x)\big) \geqslant \epsilon\Big\} \leqslant \hat{c}(\epsilon)e^{-\hat{\beta}(\epsilon)N} \tag{4.40}$$

for N sufficiently large.

Proof. Observe first that under conditions (a)–(b), both $g_x^o(x,\xi;u)$ and $\mathbb{E}[\partial_x g(x,\xi)]$ are well defined. Moreover, $g(x,\xi)$ is Clarke regular (see [3, Definition 2.3.4]). By [3, Theorem 2.7.2], $g_x^o(x;u) = \mathbb{E}[g_x^o(x,\xi;u)]$ for all $u\in\mathbb{R}^n$, which implies

$$\partial G(x) = \mathbb{E}\big[\partial_x g(x,\xi)\big]. \tag{4.41}$$

Part (i). By [18, Theorem 4],

$$\lim_{N\to\infty}\sup_{x\in\mathcal{X}} \mathbb{H}\big(\mathcal{A}G_N(x),\mathbb{E}\big[\partial_x g(x,\xi)\big]\big) = 0. \tag{4.42}$$

Combining (4.41) with (4.42), we immediately obtain (4.38).

When condition (c) holds, it follows by [8, Theorem 1] that $G(x)$ is continuously differentiable and $\partial G(x) = \nabla G(x) = \mathbb{E}[\partial_x g(x,\xi)]$. Consequently (4.39) follows from [15, Proposition 2.2]. See also [17, Proposition 4.1].

Part (ii). Let $\sigma(A,u)$ denote the support function of set A. By [9, Proposition 3.4],

$$\mathbb{E}\big[\sigma\big(\partial_x g(x,\xi),u\big)\big] = \sigma\big(\mathbb{E}\big[\partial_x g(x,\xi)\big],u\big). \tag{4.43}$$

Moreover, using the well-known Hörmander's formulae [2, Theorem II-18], we have

$$\mathbb{D}\big(\mathcal{A}G_N(x),\partial G(x)\big) = \max_{\|u\|\leqslant 1}\big(\sigma\big(\mathcal{A}G_N(x),u\big) - \sigma\big(\partial G(x),u\big)\big).$$

Since $\partial G_N(x) \subset \mathcal{A}G_N(x)$ and $\sigma(\partial_x g(x,\xi^i),u) = g_x^o(x,\xi^i;u)$, then

$$\sigma\left(\partial G_N(x),u\right) \leqslant \frac{1}{N}\sum_{i=1}^N \sigma\left(\partial_x g\left(x,\xi^i\right),u\right) = \frac{1}{N}\sum_{i=1}^N g_x^o\left(x,\xi^i;u\right).$$

Consequently, we have

$$\begin{aligned}\sup_{x\in\mathcal{X}} \mathbb{D}\left(\partial G_N(x), \mathbb{E}\left[\partial_x g(x,\xi)\right]\right) &\leqslant \sup_{x\in\mathcal{X}} \mathbb{D}\left(\mathcal{A}G_N(x), \mathbb{E}\left[\partial_x g(x,\xi)\right]\right)\\ &\leqslant \sup_{x\in\mathcal{X}} \max_{\|u\|\leqslant 1} \frac{1}{N}\sum_{i=1}^N \left[g_x^o\left(x,\xi^i;u\right) - \mathbb{E}\left[g_x^o(x,\xi;u)\right]\right]. \end{aligned} \tag{4.44}$$

By Proposition 4.1, $g_x^o(x,\xi;u)$ is almost H-clamness with modulus $\kappa(\xi)$ and order 1, $\mathbb{E}[g_x^o(x,\xi;u)]$ is a continuous function for every $u \in \mathbb{R}^n$. Moreover, since Ξ is a compact set, the moment generating functions of $g_x^o(x,\xi;u) - \mathbb{E}[g_x^o(x,\xi;u)]$ and $\kappa(\xi)$, denoted by $M_x(t)$ and $M_\kappa(t)$ respectively, are finite valued for t close to 0. Further, it is easy to verify that $|g_x^o(x,\xi;u)| \leqslant \|\nabla_x f(x,\xi)\|$. Since $\mathcal{X}$ and Ξ are compact and $\nabla_x f(\cdot,\cdot)$ is continuous by assumption, then $\|\nabla_x f(x,\xi)\|$ is bounded by a positive number L. By Theorem 3.1, for any $\epsilon > 0$, there exist positive constants $\hat{c}(\epsilon)$ and $\hat{\beta}(\epsilon)$ (independent of N) such that

$$\operatorname{Prob}\left\{\sup_{x\in\mathcal{X}} \max_{\|u\|\leqslant 1} \frac{1}{N}\sum_{i=1}^N \left[g_x^o\left(x,\xi^i;u\right) - \mathbb{E}\left[g_x^o(x,\xi;u)\right]\right] \geqslant \epsilon\right\} \leqslant \hat{c}(\epsilon)e^{-\hat{\beta}(\epsilon)N}. \tag{4.45}$$

Combining (4.41), (4.44) and (4.45), we obtain (4.40). □

The uniform exponential convergence of the subdifferentials established in Theorem 4.1 can be easily applied to derive exponential rate of convergence of stationary points in nonsmooth stochastic optimization where nonsmoothness arises from max or min functions, see [19].

Acknowledgments

The authors would like to thank an anonymous referee for insightful comments which have helped to improve the presentation of the paper. The second author would also like to thank Professors Alexander Shapiro and Daniel Ralph for helpful discussions of exponential convergence of random functions on a number of occasions from which this paper benefits.

References

[1] R.J. Aumann, Integrals of set-valued functions, J. Math. Anal. Appl. 12 (1965) 1–12.
[2] C. Castaing, M. Valadier, Convex Analysis and Measurable Multifunctions, Lecture Notes in Math., Springer, Berlin, 1977.
[3] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[4] L. Dai, C.H. Chen, J.R. Birge, Convergence properties of two stage stochastic programming, J. Optim. Theory Appl. 106 (2000) 489–509.
[5] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, Springer-Verlag, New York, 1998.
[6] D. Dentcheva, A. Ruszczyński, Semi-infinite probabilistic constraints: optimality and convexification, Optimization 53 (2004) 583–601.
[7] D. Dentcheva, A. Ruszczyński, Optimality and duality theory for stochastic optimization with nonlinear dominance constraints, Math. Program. 99 (2004) 329–350.
[8] L. Qi, A. Shapiro, C. Ling, Differentiability and semismoothness properties of integral functions and their applications, Math. Program. 102 (2005) 223–248.
[9] N.S. Papageorgiou, On the theory of Banach space valued multifunctions. I: Integration and conditional expectation, J. Multivariate Anal. 17 (1985) 185–206.
[10] M. Patriksson, L. Wynter, Stochastic mathematical programs with equilibrium constraints, Oper. Res. Lett. (1999) 159–167.
[11] G.Ch. Pflug, Stochastic optimization and statistical inference, in: A. Rusczyński, A. Shapiro (Eds.), Stochastic Programming, in: Handbooks Oper. Res. Management Sci., vol. 10, North-Holland Publishing Company, Amsterdam, 2003.
[12] D. Ralph, H. Xu, Convergence of stationary points of sample average two stage stochastic programs: A generalized equation approach, Math. Oper. Res., in press.
[13] S.M. Robinson, Analysis of sample-path optimization, Math. Oper. Res. 21 (1996) 513–528.
[14] R.Y. Rubinstein, A. Shapiro, Discrete Events Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Methods, John Wiley and Sons, New York, 1993.
[15] A. Shapiro, Asymptotic properties of statistical estimators in stochastic programming, Ann. Statist. 17 (1989) 841–858.
[16] A. Shapiro, Monte Carlo sampling methods, in: A. Rusczyński, A. Shapiro (Eds.), Stochastic Programming, in: Handbooks Oper. Res. Management Sci., vol. 10, North-Holland Publishing Company, Amsterdam, 2003.
[17] A. Shapiro, H. Xu, Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation, Optimization 57 (2008) 395–418.
[18] P. Terán, On a uniform law of large numbers for random sets and subdifferentials of random functions, Statist. Probab. Lett. 78 (2008) 42–49.
[19] H. Xuf, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, J. Math. Anal. Appl. 368 (2010) 692–710.