Asymptotics of the optimal value of SAA with AMIS on minimax stochastic programs

Wenjin Zhang[a*]

[a]College of Mathematics, Jilin University,

Changchun, 130012, P. R. China.

Abstract. The minimax stochastic programming problem is approximated in this paper using the sample average approximation with adaptive multiple importance sampling. We discuss the asymptotics and convergence of its optimal value. The core is the research and utilization of martingale difference sequences. The functional central limit theorem for martingale difference sequences is one of the main tools in studying the asymptotics. Finally, we use this result to discuss a risk averse optimization problem.

Keywords: Adaptive multiple importance sampling, minimax stochastic programming, martingale difference sequence, central limit theorem.

1 Introduction

Consider a minimax stochastic programming problem:

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\left\{f(x,y):=E[g\big(x,y,\theta\big)]\right\},\tag{1.1}$$

where $\theta:\Omega\to\Theta\subset\mathbb{R}^r$ is a random variable on $(\Omega,\mathscr{F},P)$ and $g:\mathcal{X}\times\mathcal{Y}\times\Theta\to\mathbb{R}$ with $\mathcal{X}\subset\mathbb{R}^m$ and $\mathcal{Y}\subset\mathbb{R}^n$.

The computation of the expectation is often complex and the true distribution is difficult to obtain directly. A common method for solving such problems is sample average approximation (SAA). Under the assumption that the samples are independent and identically distributed (iid), we approximate the problem

*E-mail address : zhangwenjinmails@163.com.

(1.1) by the SAA method, which is specifically expressed as follows

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\left\{\frac{1}{n}\sum_{i=1}^{n}g(x,y,\theta_i)\right\}, \tag{1.2}$$

where $\theta_1, ..., \theta_n$ are iid samples of $\theta(\omega)$. Asymptotics of the optimal values between problems (1.1) and (1.2) have been studied by Shapiro, see [14].

In recent years, the SAA method with adaptive multiple importance sampling (AMIS) has received extensive attention, see [2, 3, 4, 8, 9, 10, 13]. This method doesn't require the samples to be iid. The goal of this paper is to study the asymptotics between the optimal value obtained by this approximation method and the optimal value of the problem (1.1).

Denote $F(x,y,\theta) = g(x,y,\theta)\phi(\theta)$, where $\phi(\theta)$ represents the probability density function of $\theta(\omega)$. Thus, we can rewrite the problem (1.1) in the following form:

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\left\{f(x,y)=\int_{\Theta}F(x,y,\theta)\mathrm{d}\theta\right\}. \tag{1.3}$$

The following settings (i)-(iii) are used throughout this paper.

(i) $\{\theta_i\}_{i=1}^{\infty}$ is a sequence of random vectors on $(\Omega, \mathscr{F}, P)$, where $\theta_i : \Omega \to \mathbb{R}^r$ is $\mathscr{F}$-measurable.

(ii) $\{\mathscr{G}_i\}_{i=1}^{\infty}$ is the natural filtration sequence corresponding to the above sequence. That is, the information of $\theta_1, ..., \theta_i$ is contained in $\mathscr{G}_i$, where $\mathscr{G}_0 = \{\emptyset, \Omega\}$ and $E[\theta_1] = E[\theta_1 \mid \mathscr{G}_0]$.

(iii) For a given $\mathscr{G}_{i-1}$ with $i \in \mathbb{N}$, the conditional distribution of θ_i has a density ψ_i whose support is $\Theta_i \subset \mathbb{R}^r$.

For a given $\mathscr{G}_{i-1}$, we choose an appropriate density ψ_i and then draw a sample θ_i from it. In this way, we get $\mathscr{G}_i$. Repeating the above steps, we obtain the next sample θ_{i+1}. Clearly, θ_i can depend on the previous samples θ_1,...,θ_{i-1}. To put it succinctly, the sampling is dynamic and adaptive.

The SAA with AMIS problem associated with the problem (1.3) is as follows:

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\left\{f_n(x,y):=\frac{1}{n}\sum_{i=1}^{n}\frac{F(x,y,\theta_i)}{\psi_i(\theta_i)}\right\}. \tag{1.4}$$

It's worth noting that if $F(x,y,\theta) = g(x,y,\theta)\phi(x,y,\theta)$ with a probability density function $\phi(x,y,\theta)$, then (1.4) still works, which is pointed out in [6]. Obviously, $f(x,y) = \frac{F(x,y,\theta)}{\psi_1(\theta)} \cdot \psi_1(\theta)$. Thus, without loss of generality, let $F : \mathcal{X}\times\mathcal{Y}\times\mathbb{R}^r \to \mathbb{R}$

be a real valued function and $\mathcal{X} \times \mathcal{Y} \times \Theta$ contain its support, where $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{Y} \subset \mathbb{R}^m$ and $\Theta \subset \mathbb{R}^r$.

The rest of the paper is organized as follows. We introduce notation and preliminaries in Section 2. The main result of the paper on asymptotics of the optimal value for SAA with AMIS is shown in Section 3. In Section 4, we apply the result to the risk averse optimization problem.

2 Notation and Preliminaries

2.1 Basic Notation

Throughout this paper, we adopt the following notation:

• $(\Omega, \mathscr{F}, P)$ represents an abstract probability space.

• $E[\cdot]$ denotes the expectation with respect to the probability measure P.

• $\|\cdot\|$ stands for the Euclidean norm of a vector.

• $:=$ represents the left-hand side equal with the right-hand side by definition.

• $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution.

• $\xrightarrow{P}$ denotes convergence in probability.

• $o_p(\cdot)$ denotes a probabilistic analogue of the usual order notation $o(\cdot)$. That is, if the sequences of random variables A_n and B_n satisfy

$$\lim_{n\to\infty} \text{Prob}\left(\left|\frac{A_n}{B_n}\right| > \varepsilon\right) = 0,$$

for any $\varepsilon > 0$, then $A_n = o_p(B_n)$.

• $C(\mathcal{X}, \mathcal{Y})$ stands for the space of continuous functions $\varphi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, equipped with the sup-norm.

• $\mathcal{C}$ represents the set of convex-concave functions on $C(\mathcal{X}, \mathcal{Y})$. That is, if $\varphi \in \mathcal{C}$, then $\varphi(\cdot, y)$ is convex for any $y \in \mathcal{Y}$ and $\varphi(x, \cdot)$ is concave for $x \in \mathcal{X}$.

• $\varpi(\cdot)$ denotes a modulus of continuity. That is, $\varpi(\cdot)$ is a strictly monotonic increasing continuous function on $\mathbb{R}_+$, $\lim_{x\to 0^+} \varpi(x) = 0$ and $\limsup_{x\to 0^+} \frac{x}{\varpi(x)} < +\infty$.

• $\mathbb{D}(A, B)$ is the deviation of the set $A \subset \mathbb{R}^\iota$ from the set $B \subset \mathbb{R}^\iota$. That is, $\mathbb{D}(A, B) = \sup_{x\in A} \inf_{x'\in B} \|x - x'\|$.

2.2 Two Important Results

In this subsection, we introduce two significant results that serve as the cornerstones for our main result, Theorem 3.2. One of them is the minmax Delta theorem. It is well known that the Delta method is a useful tool in the asymptotic analysis of stochastic problems. To this end, we apply the following theorem in the next section, which is derived from [14].

Theorem 2.1. *(Minimax Delta Theorem)*

Let the sets $\mathcal{X}$ *and* $\mathcal{Y}$ *be nonempty, convex and compact. Assume that, as* $n \to \infty$*, a sequence of positive numbers* ς_n *and a random sequence* $\mathscr{Z}_n$ *in* $C(\mathcal{X}, \mathcal{Y})$ *satisfy* $\varsigma_n \to \infty$*,* $\mathscr{Z}_n \in \mathcal{C}$ *w.p.1 and* $\varsigma_n(\mathscr{Z}_n - l) \xrightarrow{\mathcal{D}} \mathscr{Z}$*, respectively, where* $\mathscr{Z} \in C(\mathcal{X}, \mathcal{Y})$ *and* $l \in \mathcal{C}$*. Denote* $\gamma := \inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \mathscr{Z}(x,y)$ *and* $\gamma_n := \inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \mathscr{Z}_n(x,y)$*. Then*

$$\varsigma_n[\gamma_n - \gamma] \xrightarrow{\mathcal{D}} \inf_{x\in\mathcal{X}^*_{\mathscr{Z}}} \sup_{y\in\mathcal{Y}^*_{\mathscr{Z}}} \mathscr{Z}(x,y),$$

and

$$\gamma_n = \inf_{x\in\mathcal{X}^*_{\mathscr{Z}}} \sup_{y\in\mathcal{Y}^*_{\mathscr{Z}}} \mathscr{Z}_n(x,y) + o_p(\varsigma_n^{-1}),$$

where $\mathcal{X}^*_{\mathscr{Z}} = \arg\min_{x\in\mathcal{X}} \left[\sup_{y\in\mathcal{Y}} \mathscr{Z}(x,y)\right]$ *and* $\mathcal{Y}^*_{\mathscr{Z}} = \arg\max_{y\in\mathcal{Y}} \left[\inf_{x\in\mathcal{X}} \mathscr{Z}(x,y)\right]$.

The other key result is a functional central limit theorem for martingale difference sequences. We briefly recall the definition of martingale difference sequences first. Assume that $\{\mathscr{F}_i\}_{i=0}^{\infty}$ is a set of sub-σ-fields of $\mathscr{F}$ with $\mathscr{F}_{i-1} \subset \mathscr{F}_i$ such that a sequence of random variables $\{X_i\}_{i=1}^{\infty}$ defined on $(\Omega, \mathscr{F}, P)$ is $\mathscr{F}_i$-measurable and $X_i \in \mathscr{F}_i$. $\{X_i, \mathscr{F}_i\}$ is called a martingale difference sequence if $E[X_i \mid \mathscr{F}_{i-1}] = 0$ for every $i \in \mathbb{N}$. Now, let us turn to this result, which is detailed in [17, *Section* 4].

Corollary 2.1. *Let* $\{X_i, \mathscr{F}_i\}_{i=1}^{\infty}$ *be a martingale difference sequence of the space* $C(\mathcal{S})$*, where* $C(\mathcal{S})$ *is the space of continuous functions on the compact set* $\mathcal{S}$ *with the sup-norm. Suppose that the following assumptions hold.*

(A1) *There exists a real nonnegative random sequence* M_i *on* $(\Omega, \mathscr{F}, P)$*, a function* $\beta : \mathcal{S} \to \mathbb{R}\backslash\{0\}$*, and a continuous distance* ρ *with* $\int_0^1 H^{\frac{1}{2}}(\mathcal{S}, \rho, r)\mathrm{d}r < \infty$ *such that for any* $s_1, s_2 \in \mathcal{S}$ *and all* $i \in \mathbb{N}$*, w.p.1*

$$|Y_i(s_1) - Y_i(s_2)| \le M_i,$$

where $Y_i(s) := |\beta(s)X_i(s)|$ *and* $\sup_{s\in\mathcal{S}} |\beta^{-1}(s)| < \infty$*.*

(A2) $\frac{1}{n}\sum_{i=1}^{n} E[M_i^2 \mid \mathscr{F}_{i-1}] \xrightarrow{P} 0.$

(A3) *For any $i \in \mathbb{N}$, there exists a constant $b > 0$ such that $E[M_i^2 \mid \mathscr{F}_{i-1}] \le b$ w.p.1.*

(A4) *For some $s_0 \in \mathcal{S}$ and all $i \in \mathbb{N}$, there exists a constant $k > 0$ such that $E[X_i^2(s_0) \mid \mathscr{F}_{i-1}] \le k$ w.p.1.*

(A5) $\frac{1}{n}\sum_{i=1}^{n} E[X_i^2(s_0) \mid \mathscr{F}_{i-1}] \xrightarrow{P} c$*, where c is a positive constant.*

(A6) *There exist a real nonnegative random sequence ς_i on $(\Omega, \mathscr{F}, P)$ such that $\sup_{n\in\mathbb{N}} \dfrac{1}{n}\sum_{i=1}^{n} E[\varsigma_i^2] < \infty$ and for any $i \in \mathbb{N}$, w.p.1*

$$|X_i(s_1) - X_i(s_2)| \le \varsigma_i \rho(s_1, s_2) \tag{2.5}$$

Then there exists a Gaussian measure μ on $C(\mathcal{S})$ such that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{\mathcal{D}} \mu \qquad \text{as} \quad n \to \infty.$$

3 Asymptotics of The Optimal Value

Assumption 3.1. *For any $i \in \mathbb{N}$, w.p.1, $\Theta \subset \Theta_i$.*

Assumption 3.2. *For any $(x, y) \in \mathcal{X} \times \mathcal{Y}$, $F(x, y, \cdot)$ is an integrable function and $f(x,y) := \int_\Theta F(x, y, \theta)\mathrm{d}\theta < \infty$.*

Lemma 3.1. *Suppose Assumptions 3.1 and 3.2 hold. Let $\Upsilon_i(x, y) := \frac{F(x,y,\theta_i)}{\psi_i(\theta_i)} - f(x, y)$ and $S_n(x, y) := \sum_{i=1}^{n} \Upsilon_i(x, y)$, where $n \in \mathbb{N}$. Then, for a given $(x, y) \in \mathcal{X} \times \mathcal{Y}$, $\{\Upsilon_i(x, y), \mathscr{G}_i\}$ is a martingale difference sequence and $\{S_n(x, y), \mathscr{G}_n\}$ is a martingale.*

This conclusion is obvious. For a pair of fixed $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and any $i \in \mathbb{N}$, w.p.1

$$\begin{aligned} E[\Upsilon_i(x, y) \mid \mathscr{G}_{i-1}] &= \int_{\Theta_i} \Upsilon_i(x, y)\psi_i(\theta_i)\mathrm{d}\theta_i \\ &= \int_{\Theta_i} \left[F(x, y, \theta_i) - f(x, y)\psi_i(\theta_i)\right]\mathrm{d}\theta_i \end{aligned}$$

$$= \quad 0.$$

By definition, Lemma 3.1 holds.

The two assumptions above are similar to those in [6]. They are viewed as basic assumptions when applying the SAA method with AMIS.

Assumption 3.3. *There exists a sequence of random measurable functions* $\alpha_i : \Theta_i \to \mathbb{R}$ *such that w.p.1*

$$\sup_{n\in\mathbb{N}} \frac{1}{n}\sum_{i=1}^{n} E[\alpha_i(\theta_i)] < \infty,$$

and w.p.1

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \big(\alpha_i(\theta_i) - E[\alpha_i(\theta_i)]\big) = 0.$$

In addition, there exists a modulus of continuity $\varpi(\cdot)$ *such that for all* $x, x' \in \mathcal{X}$, $y, y' \in \mathcal{Y}$ *and* $i \in \mathbb{N}$, *w.p.1,*

$$\left|\frac{F(x,y,\theta_i)}{\psi_i(\theta_i)} - \frac{F(x',y',\theta_i)}{\psi_i(\theta_i)}\right| \le \alpha_i(\theta_i)\varpi\big(\|x-x'\| + \|y-y'\|\big).$$

Assumption 3.4. *For every pair* $(x,y) \in \mathcal{X}\times\mathcal{Y}$, *w.p.1,*

$$\lim_{n\to\infty} \frac{1}{n} S_n(x,y) = 0.$$

Assumption 3.5. $F(x,y,\theta)$ *is a Carathéodory function, i.e.,* $F(x,y,\theta)$ *is measurable for any* $(x,y) \in \mathcal{X}\times\mathcal{Y}$ *and continuous for a.e.* $\theta \in \Theta$.

Assumption 3.6. *There is an integrable function* $L(\theta)$, *i.e.,* $\int_\Theta L(\theta)\mathrm{d}\theta < \infty$, *and an open set* $\mathcal{O} \subset \mathbb{R}^{m+n}$ *with* $\mathcal{X}\times\mathcal{Y} \subset \mathcal{O}$ *such that* $|F(x,y,\theta)| \le L(\theta)$ *for every* $(x,y) \in \mathcal{O}$ *and a.e.* $\theta \in \Theta$.

Assumption 3.7. *The sets* $\mathcal{X}$ *and* $\mathcal{Y}$ *are nonempty and compact, respectively.*

Lemma 3.2. *Suppose that Assumptions* 3.5-3.7 *hold. Then the expected value function* $f(x,y)$ *is finite valued and continuous on* $\mathcal{X}\times\mathcal{Y}$ *and the max-function* $\varphi(x) := \sup_{y\in\mathcal{Y}} f(x,y)$ *is continuous on* $\mathcal{X}$.

Proof. According to Assumption 3.6, for every pair $(x,y) \in \mathcal{X}\times\mathcal{Y}$,

$$|f(x,y)| = \left|\int_\Theta F(x,y,\theta)\mathrm{d}\theta\right| \le \int_\Theta |F(x,y,\theta)|\mathrm{d}\theta \le \int_\Theta L(\theta)\mathrm{d}\theta < \infty.$$

Therefore, $f(x,y)$ is well defined. This shows that Assumption 3.2 is satisfied. Furthermore, by Assumptions 3.5 and 3.7, applying the Lebesgue dominated convergence theorem, we obtain

$$f(x,y)=\int_\Theta \lim_{(x',y')\to(x,y)} F(x,y,\theta)\mathrm{d}\theta=\lim_{(x',y')\to(x,y)}\int_\Theta F(x,y,\theta)\mathrm{d}\theta.$$

This implies that $f(x,y)$ is continuous on $\mathcal{X}\times\mathcal{Y}$.

By Assumption 3.7, for a given $x\in\mathcal{X}$, there exists $y\in\mathcal{Y}$ such that $f(x,y)=\varphi(x)$. It is not difficult to deduce that $\varphi(x)$ is continuous on $\mathcal{X}$ by the continuity of $f(x,y)$. The proof is complete. □

The next theorem proves that the optimal value and optimal solutions of the problem (1.4) converge to the optimal value and optimal solutions of the problem (1.3), respectively. The underlying idea is to establish uniform convergence from f_n to f. To this end, we apply Theorem 3(b) in [1]. This is also exploited in [6], but our target problem is different. For ease of presentation, let ϑ and ϑ_n denote the optimal values of (1.3) and (1.4), respectively. Let $\mathcal{T}_x$ and $\mathcal{T}_{x,n}$ represent the sets of optimal solutions of (1.3) and (1.4), respectively.

Theorem 3.1. *Suppose that Assumptions* 3.1 *and* 3.3-3.7 *hold. Then* $\lim_{n\to\infty}\vartheta_n=\vartheta$ *and* $\lim_{n\to\infty}\mathbb{D}(\mathcal{T}_{x,n},\mathcal{T}_x)=0$ *w.p.1.*

Proof. In fact, $\vartheta=\inf_{x\in\mathcal{X}}\varphi(x)$. It follows from Lemma 3.2 that $\mathcal{T}_x$ is nonempty and contained in $\mathcal{X}$. Let the sample average max-function with AMIS be $\varphi_n(x):=\sup_{y\in\mathcal{Y}} f_n(x,y)$. Obviously, $\varphi_n(x)$ is continuous on $\mathcal{X}$. Then, $\vartheta_n=\inf_{x\in\mathcal{X}}\varphi_n(x)$ and $\mathcal{T}_{x,n}$ is nonempty and contained in $\mathcal{X}$.

According to Assumptions 3.1, 3.3 and 3.4, we have that $f_n(x,y)$ converges to $f(x,y)$ uniformly on $\mathcal{X}\times\mathcal{Y}$ w.p.1, see [1, *Theorem* 3(*b*)] and [6, *Theorem* 1]. By Lemma 3.2, $f(x,y)$ is finite valued and continuous on $\mathcal{X}\times\mathcal{Y}$. Consequently, applying Theorem 5.3 in [16], we complete the proof. □

Assumption 3.8. *The sets* $\mathcal{X}$ *and* $\mathcal{Y}$ *are convex, respectively.*

The dual problem of (1.3) is as follows:

$$\max_{y\in\mathcal{Y}}\min_{x\in\mathcal{X}} f(x,y). \tag{3.6}$$

Let $\mathcal{T}_y$ denote the set of optimal solutions of (3.6). If $f(x,y)\in\mathcal{C}$ and Assumptions 3.7 and 3.8 hold, then there is no duality gap between (1.3) and (3.6). That is,

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}} f(x,y)=\max_{y\in\mathcal{Y}}\min_{x\in\mathcal{X}} f(x,y), \tag{3.7}$$

and $\mathcal{T}_x \times \mathcal{T}_y$ forms the set of saddle points.

Assumption 3.9. *$F(\cdot,\cdot,\theta)$ is convex-concave on $\mathcal{X}\times\mathcal{Y}$, i.e., $F(\cdot,y,\theta)$ is convex on $\mathcal{X}$ for any $y\in\mathcal{Y}$ and $F(x,\cdot,\theta)$ is concave on $\mathcal{Y}$ for any $x\in\mathcal{X}$.*

Now, we come to state the main theorem in this paper.

Theorem 3.2. *Let Assumptions 3.1-3.2 and 3.7-3.9 hold. Suppose that the following statements hold.*
(B1) *For some $(x_0,y_0)\in\mathcal{X}\times\mathcal{Y}$, there exist constants $\tilde{c}>f^2(x_0,y_0)$ and $\tilde{k}>f^2(x_0,y_0)$ such that*

$$\frac{1}{n}\sum_{i=1}^{n} E\left[\left(\frac{F(x_0,y_0,\theta_i)}{\psi_i(\theta_i)}\right)^2 \middle| \mathscr{G}_{i-1}\right] \xrightarrow{P} \tilde{c},$$

and w.p.1

$$E\left[\left(\frac{F(x_0,y_0,\theta_i)}{\psi_i(\theta_i)}\right)^2 \middle| \mathscr{G}_{i-1}\right] \le \tilde{k}.$$

(B2) *There exists an integrable function $\alpha:\mathbb{R}^r\to\mathbb{R}_+$ whose support is contained in Θ such that for any $(x_1,y_1),(x_2,y_2)\in\mathcal{X}\times\mathcal{Y}$, w.p.1*

$$|F(x_1,y_1,\theta)-F(x_2,y_2,\theta)| \le \alpha(\theta)(\|x_1-x_2\|+\|y_1-y_2\|) \tag{3.8}$$

with $\sup_{n\in\mathbb{N}} \frac{1}{n}\sum_{i=1}^{n} E\left[\left(\frac{\alpha(\theta)}{\psi_i(\theta)}\right)^2\right] < \infty.$
(B3) *There exists a sequence of random measurable functions $\mathcal{A}_i:\Theta_i\to\mathbb{R}_+$ such that for any $i\in\mathbb{N}$, w.p.1*

$$|F(x,y,\theta)-f(x,y)\psi_i(\theta)| \le \mathcal{A}_i(\theta),$$

$$\frac{1}{n}\sum_{i=1}^{n} E\left[\left(\frac{\mathcal{A}_i(\theta_i)}{\psi_i(\theta_i)}\right)^2 \middle| \mathscr{G}_{i-1}\right] \xrightarrow{P} 0,$$

and w.p.1

$$E\left[\left(\frac{\mathcal{A}_i(\theta_i)}{\psi_i(\theta_i)}\right)^2 \middle| \mathscr{G}_{i-1}\right] \le \tilde{b}\ ,$$

where $\tilde{b}$ is a positive constant.

Then

$$\vartheta_n = \inf_{x\in\mathcal{T}_x} \sup_{y\in\mathcal{T}_y} f_n(x,y) + o_p(\tfrac{1}{\sqrt{n}}). \tag{3.9}$$

In particular, if $\mathcal{T}_x = \{\hat{x}\}$ *and* $\mathcal{T}_y = \{\hat{y}\}$ *are singletons, then*

$$\sqrt{n}(\vartheta_n - \vartheta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\hat{x},\hat{y})), \qquad \text{as} \ \ n\to\infty, \tag{3.10}$$

where $\mathcal{N}(0,\sigma^2(\hat{x},\hat{y}))$ *denotes the normal distribution with mean* 0 *and variance* $\sigma^2(\hat{x},\hat{y}) = \tilde{c} - f^2(x_0,y_0)$.

Proof. For any $(x,y)\in\mathcal{X}\times\mathcal{Y}$, $\{\Upsilon_i(x,y),\mathscr{G}_i\}$ is a martingale difference sequence by Lemma 3.1. Thus, we show that Υ_i satisfies the conditions in Corollary 2.1 at first.

It is not difficult to see that $f(x,y)$ is well defined and finite valued. Let $\tilde{\alpha} := \int_\Theta \alpha(\theta)\mathrm{d}\theta$. Obviously, $0<\tilde{\alpha}<\infty$. Integrating over (3.8), we have

$$|f(x_1,y_1) - f(x_2,y_2)| \le \tilde{\alpha}(\|x_1-x_2\| + \|y_1-y_2\|) \tag{3.11}$$

for any $(x_1,y_1),(x_2,y_2)\in\mathcal{X}\times\mathcal{Y}$. Therefore, $f(x,y)$ is Lipschitz continuous on $\mathcal{X}\times\mathcal{Y}$. Similarly, by integration, it follows from Assumptions 3.7-3.9 that $f(x,y)$ is convex-concave. Then, we have $f\in\mathcal{C}$ and $F(\cdot,\cdot,\theta)\in\mathcal{C}$ for any given $\theta\in\Theta$. By definition,

$$\Upsilon_i(x,y) = \begin{cases} \frac{F(x,y,\theta_i)}{\psi_i(\theta_i)} - f(x,y), & \theta_i\in\Theta, \\ \quad - f(x,y), & \theta_i \in \Theta_i\backslash\Theta. \end{cases}$$

Therefore, $\Upsilon_i\in\mathcal{C}$ and $f_n\in\mathcal{C}$. To simplify notation, let

$$H_n(x,y) := f_n(x,y) - f(x,y) = \frac{1}{n}\sum_{i=1}^n \Upsilon_i(x,y).$$

Then, $H_n\in\mathcal{C}$. From the previous analysis, we know that (3.7) holds. Further, $\mathcal{T}_x\times\mathcal{T}_y$ is nonempty and forms the set of saddle points.

Let $M_i = \frac{\mathcal{A}_i(\theta_i)}{\psi_i(\theta_i)}$. It follows from **(B3)** that

$$||\Upsilon_i(x_1,y_1)| - |\Upsilon_i(x_2,y_2)|| \le M_i, \tag{3.12}$$

w.p.1 for any $(x_1,y_1),(x_2,y_2)\in\mathcal{X}\times\mathcal{Y}$. Moreover, **(A1)-(A3)** are verified.

On the other hand, for $(x_0,y_0)\in\mathcal{X}\times\mathcal{Y}$ and all $i\in\mathbb{N}$, w.p.1

$$E[\Upsilon_i^2(x_0,y_0)\mid \mathscr{G}_{i-1}]$$

$$
\begin{aligned}
&= \int_{\Theta_i} \Upsilon_i^2(x_0,y_0)\psi_i(\theta_i)\mathrm{d}\theta_i \\
&= \int_{\Theta_i} \left[\frac{F(x_0,y_0,\theta_i)}{\psi_i(\theta_i)} - f(x_0,y_0)\right]^2 \psi_i(\theta_i)\mathrm{d}\theta_i \\
&= \int_{\Theta_i} \left[\frac{F^2(x_0,y_0,\theta_i)}{\psi_i(\theta_i)} - 2f(x_0,y_0)F(x_0,y_0,\theta_i) + f^2(x_0,y_0)\psi_i(\theta_i)\right]\mathrm{d}\theta_i \\
&= E\left[\left(\frac{F(x_0,y_0,\theta_i)}{\psi_i(\theta_i)}\right)^2 \middle| \mathscr{G}_{i-1}\right] - f^2(x_0,y_0).
\end{aligned}
$$

Together with **(B1)**, it is not hard to verify that $E[\Upsilon_i^2(x_0,y_0) \mid \mathscr{F}_{i-1}]$ satisfies **(A4)** and **(A5)**.

According to (3.8) and (3.11), we have

$$
\begin{aligned}
& |\Upsilon_i(x_1,y_1) - \Upsilon_i(x_2,y_2)| \\
= & \left|\frac{F(x_1,y_1,\theta_i)}{\psi_i(\theta_i)} - f(x_1,y_1) - \frac{F(x_2,y_2,\theta_i)}{\psi_i(\theta_i)} + f(x_2,y_2)\right| \\
\le & \left|\frac{F(x_1,y_1,\theta_i)}{\psi_i(\theta_i)} - \frac{F(x_2,y_2,\theta_i)}{\psi_i(\theta_i)}\right| + |f(x_1,y_1) - f(x_2,y_2)| \\
\le & \left(\frac{\alpha(\theta_i)}{\psi_i(\theta_i)} + \tilde{\alpha}\right)(\|x_1 - x_2\| + \|y_1 - y_2\|) \quad \text{w.p.1.}
\end{aligned}
$$

Let $\varsigma_i = \frac{\alpha(\theta_i)}{\psi_i(\theta_i)} + \tilde{\alpha}$. Obviously, $\varsigma_i > 0$. Hence (2.5) is satisfied. Moreover, we get

$$
\sup_{n\in\mathbb{N}} \frac{1}{n}\sum_{i=1}^{n} E\left[\varsigma_i^2\right] \le \sup_{n\in\mathbb{N}} \frac{2}{n}\sum_{i=1}^{n} E\left[\left(\frac{\alpha(\theta)}{\psi_i(\theta)}\right)^2\right] + 2\tilde{\alpha}^2 < \infty.
$$

Therefore, **(A6)** is verified.

From the above discussion, Υ_i satisfies the conditions in Corollary 2.1. Then, applying Corollary 2.1, there exists a Gaussian measure Y on $C(\mathcal{X},\mathcal{Y})$ such that for any $(x,y) \in \mathcal{X}\times\mathcal{Y}$,

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Upsilon_i(x,y) = \sqrt{n}H_n(x,y) \xrightarrow{\mathcal{D}} Y(x,y) \qquad \text{as } n \to \infty.
$$

By the properties of martingale difference sequences, we have that $Y(x,y)$ follows a normal distribution with mean 0 and variance $\sigma^2(x,y) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E[\Upsilon_i^2(x,y)]$, see Remark 3.1.

Obviously, $\sqrt{n} \to \infty$ and $\sqrt{n}(f_n - f) \xrightarrow{\mathcal{D}} Y$ as $n \to \infty$, respectively. Thus, using Theorem 2.1, we have

$$\vartheta_n = \inf_{x \in \mathcal{T}_x} \sup_{y \in \mathcal{T}_y} f_n(x, y) + o_p(\tfrac{1}{\sqrt{n}})$$

and

$$\sqrt{n}(\vartheta_n - \vartheta) \xrightarrow{\mathcal{D}} \inf_{x \in \mathcal{T}_x} \sup_{y \in \mathcal{T}_y} Y(x, y), \qquad \text{as } \ n \to \infty. \tag{3.13}$$

Hence, (3.9) is proved.

Finally, if $\mathcal{T}_x = \{\hat{x}\}$ and $\mathcal{T}_y = \{\hat{y}\}$ are singletons, we have $\inf_{x \in \mathcal{T}_x} \sup_{y \in \mathcal{T}_y} Y(\hat{x}, \hat{y}) = \mathcal{N}(0, \sigma^2(\hat{x}, \hat{y}))$. Hence (3.10) follows from (3.13). Thus, the proof is complete. □

Remark 3.1. *For any $(x, y) \in \mathcal{X} \times \mathcal{Y}$, by the definition of martingale difference sequences, we have $E[\Upsilon_i(x, y) \mid \mathscr{G}_{i-1}] = 0$. Then, we obtain*

$$\begin{aligned} E\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Upsilon_i(x, y)\right] &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E[\Upsilon_i(x, y)] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\left[E[\Upsilon_i(x, y) \mid \mathscr{G}_{i-1}]\right] \\ &= 0. \end{aligned}$$

Moreover, for any $i, j \in \mathbb{N}$ and $i > j$, we get

$$\begin{aligned} E[\Upsilon_i(x, y)\Upsilon_j(x, y)] &= E\left[E[\Upsilon_i(x, y)\Upsilon_j(x, y) \mid \mathscr{G}_j]\right] \\ &= E\left[\Upsilon_j(x, y) E[\Upsilon_i(x, y) \mid \mathscr{G}_j]\right]. \end{aligned}$$

Since $E[\Upsilon_i(x, y) \mid \mathscr{G}_j] = E[E[\Upsilon_i(x, y) \mid \mathscr{G}_{i-1}] \mid \mathscr{G}_j] = 0$, we get

$$E[\Upsilon_i(x, y)\Upsilon_j(x, y)] = 0.$$

Therefore,

$$\mathrm{Var}[\Upsilon_j(x, y)] = E[\Upsilon_j^2(x, y)], \qquad \forall j \in \mathbb{N},$$

and

$$\mathrm{Var}[\Upsilon_i(x, y) + \Upsilon_j(x, y)] = E[\Upsilon_i^2(x, y)] + E[\Upsilon_j^2(x, y)], \quad \forall i, j \in \mathbb{N} \text{ and } i \neq j.$$

Thus,

$$\mathrm{Var}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Upsilon_i(x)\right] = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}[\Upsilon_i(x)] = \frac{1}{n} \sum_{i=1}^{n} E[\Upsilon_i^2(x)].$$

By Corollary 2.1 *in [17], we have*

$$\begin{aligned}\sigma^2(x,y) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E[\Upsilon_i^2(x,y)] \\ &= \left(\tilde{c}-f^2(x_0,y_0)\right)\left(\frac{\beta(x_0,y_0)}{\beta(x,y)}\right)^2.\end{aligned}$$

where β is a nonzero real-valued function on $\mathcal{X}\times\mathcal{Y}$. According to (3.12)*, we get $\beta(x,y)=1$ for any $(x,y)\in\mathcal{X}\times\mathcal{Y}$. Hence $\sigma^2(x,y)=\tilde{c}-f^2(x_0,y_0)>0$. Notice that Y is a Gaussian measure on $C(\mathcal{X},\mathcal{Y})$. Obviously, $E[Y(x,y)]=0$ and* $\mathrm{Var}[Y(x,y)]=\sigma^2(x,y)$. *Then, $Y(x,y)\sim\mathcal{N}(0,\sigma^2(x,y))$.*

4 An Application to The Risk Averse Optimization

In recent years, the risk averse optimization has been extensively studied, see [5, 7, 11, 12, 14, 15]. A comprehensive review can be found in reference [16, *Chapter* 6]. The specific problem we focus on here is stated as follows:

$$\min_{x\in\mathcal{X}}\rho_\gamma(G(x,\theta)), \tag{4.14}$$

where $G:\mathcal{X}\times\Theta\to\mathbb{R}$ with $\mathcal{X}\subset\mathbb{R}^m$, $\theta:\Omega\to\Theta\subset\mathbb{R}^r$ is a random vector on $(\Omega,\mathscr{F},P)$, and $\rho_\gamma(\cdot)$ is the absolute semideviation risk measure with the weight constant $\gamma\in[0,1]$, that is, $\rho_\gamma(Z):=E\{Z+\gamma[Z-E(Z)]_+\}$.

Essentially, such problems are minimax problems, see [11, 16] for details. To be specific, the equivalent of (4.14) is

$$\min_{(x,t)\in\mathcal{X}\times\mathbb{R}}\ \max_{\lambda\in[0,1]}E\left\{h\big(t,\lambda,G(x,\theta)\big)\right\}, \tag{4.15}$$

where $h\big(t,\lambda,G(x,\theta)\big):=G(x,\theta)+\gamma\lambda[G(x,\theta)-t]_+ +\gamma(1-\lambda)[t-G(x,\theta)]_+$.

Let $\mathcal{X}$ be nonempty, convex and compact. Assume that $G(\cdot,\theta)$ is convex for a.e. $\theta\in\Theta$ and $E[G(x,\theta)]<\infty$ for every $x\in\mathcal{X}$. Under such assumptions, (4.15) is a convex-concave minimax problem. Furthermore, if we assume that $\mathcal{X}^\star$ is the set of optimal solutions for the problem (4.14), then the set of optimal solutions for the problem (4.15) is $\mathcal{T}=\{(x^*,t^*):x^*\in\mathcal{X}^\star,t^*=E[G(x^*,\theta)]\}$. Accordingly, the set of optimal solutions for its dual problem is $\mathcal{T}_d=[\nu^*,\nu^{**}]$, where $\nu^*=\mathrm{Prob}\big(G(x^*,\theta)<E[G(x^*,\theta)]\big)$ and $\nu^{**}=\mathrm{Prob}\big(G(x^*,\theta)\le E[G(x^*,\theta)]\big)$ with $x^*\in\mathcal{X}^\star$. Then, $\mathcal{T}\times\mathcal{T}_d$ forms the set of the saddle points. The analysis of this part is detailed in [11, 14].

If we use $\phi(\theta)$ to represent the probability density function of θ, it is not difficult to see that $F(x,t,\lambda,\theta) = h\big(t,\lambda,G(x,\theta)\big)\cdot\phi(\theta)$ satisfies Assumption 3.9, that is, $F(\cdot,\cdot,\lambda,\theta)$ is convex on $\mathcal{X}\times\mathbb{R}$ for any $\lambda\in[0,1]$ and $F(x,t,\cdot,\theta)$ is concave on $[0,1]$ for any $(x,t)\in\mathcal{X}\times\mathbb{R}$. The information above implies that Assumptions 3.2 and 3.7-3.9 hold.

Let θ_i, $\mathscr{G}_i$, ψ_i and Θ_i be defined as in the introduction, where $i\in\mathbb{N}$. Let Assumption 3.1 hold. The SAA with AMIS problem associated with problem (4.15) is as follows:

$$\min_{(x,t)\in\mathcal{X}\times\mathbb{R}}\ \max_{\lambda\in[0,1]}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{F(x,t,\lambda,\theta_i)}{\psi_i(\theta_i)}\right\}. \tag{4.16}$$

Proposition 4.1. *Let $\mathcal{X}$ be nonempty, convex and compact. Suppose that $G(\cdot,\theta)$ is convex for a.e. $\theta\in\Theta$ and $E[G(x,\theta)]<\infty$ for every $x\in\mathcal{X}$. Let Assumption* 3.1 *and the following conditions* **(B′1)**-**(B′3)** *be satisfied.*
(B′1) *For some $(x_0,t_0,\lambda_0)\in\mathcal{X}\times\mathbb{R}\times[0,1]$, there exist constants $\hat{c}>f^2(x_0,t_0,\lambda_0)$ and $\hat{k}>f^2(x_0,t_0,\lambda_0)$ such that*

$$\frac{1}{n}\sum_{i=1}^{n}E\left[\left(\frac{F(x_0,t_0,\lambda_0,\theta_i)}{\psi_i(\theta_i)}\right)^2\middle|\mathscr{G}_{i-1}\right]\xrightarrow{P}\hat{c},$$

and w.p.1

$$E\left[\left(\frac{F(x_0,t_0,\lambda_0,\theta_i)}{\psi_i(\theta_i)}\right)^2\middle|\mathscr{G}_{i-1}\right]\le\hat{k},$$

where $f(x_0,t_0,\lambda_0)=E\left\{h\big(t_0,\lambda_0,G(x_0,\theta)\big)\right\}$.
(B′2) *There exists an integrable function $\alpha:\mathbb{R}^r\to\mathbb{R}_+$ whose support is contained in Θ such that for any $(x_1,t_1,\lambda_1),(x_2,t_2,\lambda_2)\in\mathcal{X}\times\mathbb{R}\times[0,1]$, w.p.1*

$$|F(x_1,t_1,\lambda_1,\theta)-F(x_2,t_2,\lambda_2,\theta)|\le\alpha(\theta)(\|(x_1,t_1)-(x_2,t_2)\|+\|\lambda_1-\lambda_2\|),$$

with $\sup_{n\in\mathbb{N}}\frac{1}{n}\sum_{i=1}^{n}E\left[\left(\frac{\alpha(\theta)}{\psi_i(\theta)}\right)^2\right]<\infty$.
(B′3) *There exists a sequence of random measurable functions $\mathcal{A}_i:\Theta_i\to\mathbb{R}_+$ such that for any $i\in\mathbb{N}$, w.p.1*

$$|F(x,t,\lambda,\theta)-E\left\{h\big(t,\lambda,G(x,\theta)\big)\right\}\psi_i(\theta)|\le\mathcal{A}_i(\theta),$$

$$\frac{1}{n}\sum_{i=1}^{n}E\left[\left(\frac{\mathcal{A}_i(\theta_i)}{\psi_i(\theta_i)}\right)^2\middle|\mathscr{G}_{i-1}\right]\xrightarrow{P}0,$$

and w.p.1

$$E\left[\left(\frac{\mathcal{A}_i(\theta_i)}{\psi_i(\theta_i)}\right)^2 \middle| \mathscr{G}_{i-1}\right] \leq \hat{b},$$

where $\hat{b}$ is a positive constant.
Then

$$\hat{\vartheta}_n = \inf_{(x,t)\in\mathcal{T}} \sup_{\lambda\in\mathcal{T}_d} \left\{ \frac{1}{n}\sum_{i=1}^{n} \frac{F(x,t,\lambda,\theta_i)}{\psi_i(\theta_i)} \right\} + o_p(\tfrac{1}{\sqrt{n}}).$$

In particular, suppose that $\mathcal{X}^\star = \{\tilde{x}\}$ is a singleton and $\nu^* = \nu^{**}$, that is, $\text{Prob}\big(G(x^*,\theta) = E[G(x^*,\theta)]\big) = 0$. *Then*

$$\sqrt{n}(\hat{\vartheta}_n - \hat{\vartheta}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \hat{c} - f^2(x_0, t_0, \lambda_0)), \qquad \text{as} \quad n \to \infty,$$

where $\hat{\vartheta}$ and $\hat{\vartheta}_n$ denote the optimal values of (4.15) *and* (4.16), *respectively.*

Proof. It is easy to verify the proposition. By the above analysis, Assumptions 3.2 and 3.7-3.9 are satisfied. According to conditions **(B′1)**-**(B′3)**, it is straightforward to apply Theorem 3.2. Thus, the proof is complete. □

Acknowledgment

The author is very grateful to Professor Yong Li for instructions and many invaluable suggestions.

References

[1] Andrews, D.W.K.: Generic uniform convergence. Econometric Theory. 8(2), 241-257 (1992)

[2] Bugallo, M. F., Elvira, V., Martino, L., Luengo, D., Miguez, J., Djuric, P. M.: Adaptive importance sampling: The past, the present, and the future, IEEE Signal Process. Mag. 34, 60-79 (2017)

[3] Corneut, J.-M., Marin, J.-M., Mira, A., Robert, C.P.: Adaptive Multiple Importance Sampling. Scand. J. Stat. 39(4), 798-812 (2012)

[4] El-Laham, Y., Martino, L., Elvira, V., and Bugallo, M. F. Efficient adaptive multiple importance sampling. 27th European Signal Processing Conference (EUSIPCO), 1-5 (2019)

[5] Fábián, C.I., Wolf, C., Koberstein, A., Suhl, L.: Risk-averse optimization in two-stage stochastic models: computational aspects and a study. SIAM J. Optim. 25(1), 28-52 (2015)

[6] Feng, M.B., Maggiar, A., Staum, J., Wächter, A.: Uniform convergence of sample average approximation with adaptive multiple importance sampling. In 2018 Winter Simulation Conference (WSC). IEEE. 1646-1657 (2018)

[7] Lan, G., Zhang, Z.: Optimal methods for convex risk-averse distributed optimization. SIAM J. Optim. 33(3), 1518-1557 (2023)

[8] Luengo, D.; Martino, L.; Bugallo, M.; Elvira, V.; Särkkä, S. A survey of Monte Carlo methods for parameter estimation. EURASIP J. Adv. Signal Process. 2020, 1-62 (2020)

[9] Maggiar, A., Wächter, A., Dolinskaya, I.S., Staum, J.: A derivative-free trust-region algorithm for the optimization of functions smoothed via Gaussian convolution using adaptive multiple importance sampling. SIAM J. Optim. 28(2), 1478-1507 (2018)

[10] Marin, J.M., Pudlo, P., Sedki, M.: Consistency of the adaptive multiple importance sampling. arXiv preprint arXiv:1211.2548 (2012)

[11] Ogryczak, W., Ruszczyński, A.: Dual stochastic dominance and related mean-risk models. SIAM J. Optim. 13(1), 60-78 (2002)

[12] Pichler, A., Schlotter, R.: Martingale characterizations of risk-averse stochastic optimization problems. Math.Program. 181(2, Ser. B), 377-403 (2020)

[13] Retkute, R., Touloupou, P., Basáñez, M-G., Hollingsworth, T.D., Spencer, S.E.F.: Integrating geostatistical maps and infectious disease transmission models using adaptive multiple importance sampling. Ann. Appl. Stat. 15(4), 1980-1998 (2021)

[14] Shapiro, A.: Asymptotics of minimax stochastic programs. Statist. Probab. Lett. 78(2), 150-157 (2008)

[15] Shapiro, A.: Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming. European J. Oper. Res. 288(1), 1-13 (2021)

[16] Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory. Philadelphia: SIAM (2009)

[17] Zhang, W., Li, Y.: Uniform exponential convergence of SAA with AMIS and asymptotics of its optimal value.