

ASYMPTOTICS OF THE OPTIMAL VALUE OF SAA WITH AMIS ON MINIMAX STOCHASTIC PROGRAMS

Wenjin Zhang^{1,†}

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 apply this result to discuss a risk averse optimization problem.

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

MSC(2010) 90C15, 90C47.

1. Introduction

Studying the optimal values and optimal solutions of optimization problems is often of great significance in real life, see [8, 9, 19]. In this paper, we consider a minimax stochastic programming problem:

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

where $\theta : \Omega \rightarrow \Theta \subset \mathbb{R}^r$ is a random variable on (Ω, \mathcal{F}, P) and $g : \mathcal{X} \times \mathcal{Y} \times \Theta \rightarrow \mathbb{R}$ 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 (1.1) by the SAA method, which is specifically expressed as follows:

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

where $\theta_1, \dots, \theta_n$ are iid samples of $\theta(\omega)$. Asymptotics of the optimal values between problems (1.1) and (1.2) have been studied by Shapiro, see [16].

In recent years, the SAA method with adaptive multiple importance sampling (AMIS) has received extensive attention, see [2–4, 10–12, 15]. This method doesn't

[†]The corresponding author.

¹College of Mathematics, Jilin University, Changchun 130012, China
Email: zhangwenjinmails@163.com (W. Zhang)

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).

To simplify notation, we use θ to represent either a random or a deterministic vector, depending on the context. Denote $F(x, y, \theta) = g(x, y, \theta)\phi(\theta)$, where $\phi(\theta)$ represents the probability density function of θ . Thus, we can rewrite the problem (1.1) in the following form:

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

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

- (i) $\{\theta_i\}_{i=1}^{\infty}$ is a sequence of random vectors on (Ω, \mathcal{F}, P) , where $\theta_i : \Omega \rightarrow \mathbb{R}^r$ is \mathcal{F} -measurable.
- (ii) $\{\mathcal{G}_i\}_{i=1}^{\infty}$ is the natural filtration sequence corresponding to the above sequence. That is, the information of $\theta_1, \dots, \theta_i$ is contained in \mathcal{G}_i , where $\mathcal{G}_0 = \{\emptyset, \Omega\}$ and $E[\theta_1] = E[\theta_1 | \mathcal{G}_0]$.
- (iii) For a given \mathcal{G}_{i-1} ($i \in \mathbb{N}$), the conditional distribution of θ_i has a density ψ_i whose support is $\Theta_i \subset \mathbb{R}^r$.

For a given \mathcal{G}_{i-1} , we choose an appropriate density ψ_i and then draw a sample θ_i from it. In this way, we get \mathcal{G}_i . Repeating the above steps, we obtain the next sample θ_{i+1} . Clearly, θ_i can depend on the previous samples $\theta_1, \dots, \theta_{i-1}$. To put it succinctly, the sampling is dynamic and adaptive.

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

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

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, $f(x, y) = \int_{\Theta} \frac{F(x, y, \theta)}{\psi_1(\theta)} \psi_1(\theta) d\theta$. Thus, without loss of generality, let $F : \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^r \rightarrow \mathbb{R}$ be a real valued function and $\mathcal{X} \times \mathcal{Y} \times \Theta$ contain its support, where $\mathcal{X} \subset \mathbb{R}^m$, $\mathcal{Y} \subset \mathbb{R}^n$ 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:

- (Ω, \mathcal{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

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

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

- $C(\mathcal{X}, \mathcal{Y})$ stands for the space of continuous functions $\varphi : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$, 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}_+ , $\lim_{x \rightarrow 0^+} \varpi(x) = 0$ and $\limsup_{x \rightarrow 0^+} \frac{x}{\varpi(x)} < +\infty$.

• $\mathbb{D}(A, B)$ is the deviation of the set $A \subset \mathbb{R}^t$ from the set $B \subset \mathbb{R}^t$. 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. We apply the following theorem in the next section, which is derived from [16].

Theorem 2.1 (Minimax Delta Theorem). *Let the sets \mathcal{X} and \mathcal{Y} be nonempty, convex and compact. Assume that, as $n \rightarrow \infty$, a sequence of positive numbers $\{\varsigma_n\}$ and a random sequence $\{\mathcal{Z}_n\}$ in $C(\mathcal{X}, \mathcal{Y})$ satisfy $\varsigma_n \rightarrow \infty$, $\mathcal{Z}_n \in \mathcal{C}$ w.p.1 and $\varsigma_n(\mathcal{Z}_n - l) \xrightarrow{\mathcal{D}} \mathcal{Z}$, respectively, where $\mathcal{Z} \in C(\mathcal{X}, \mathcal{Y})$ and $l \in \mathcal{C}$. Denote $\gamma := \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \mathcal{Z}(x, y)$ and $\gamma_n := \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \mathcal{Z}_n(x, y)$. Then*

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

and

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

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

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 $\{\mathcal{F}_i\}_{i=0}^{\infty}$ is a set of sub- σ -fields of \mathcal{F} with $\mathcal{F}_{i-1} \subset \mathcal{F}_i$ and $\{X_i\}_{i=1}^{\infty}$ is a sequence of random variables defined on (Ω, \mathcal{F}, P) , where X_i is \mathcal{F}_i -measurable. $\{X_i, \mathcal{F}_i\}$ is called a martingale difference sequence if $E[X_i | \mathcal{F}_{i-1}] = 0$ for every $i \in \mathbb{N}$. Now, let us turn to this result, which is detailed in [20, Section 4].

Theorem 2.2. Let $\{X_i, \mathcal{F}_i\}_{i=1}^\infty$ 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 (Ω, \mathcal{F}, P) and a function $\beta : \mathcal{S} \rightarrow \mathbb{R} \setminus \{0\}$ such that for any $s_1, s_2 \in \mathcal{S}$ and all $i \in \mathbb{N}$, w.p.1

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

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

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

(A3) For any $i \in \mathbb{N}$, there exists a constant $b > 0$ such that $E[M_i^2 \mid \mathcal{F}_{i-1}] \leq b$ 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 $|X_i(s_0)| \leq k$ w.p.1.

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

(A6) There exists a real nonnegative random sequence $\{\varsigma_i\}$ on (Ω, \mathcal{F}, P) and a continuous distance ρ with $\int_0^1 H^{\frac{1}{2}}(\mathcal{S}, \rho, r) dr < \infty$ such that $\sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^n E[\varsigma_i^2] < \infty$ and for any $i \in \mathbb{N}$, w.p.1

$$|X_i(s_1) - X_i(s_2)| \leq \varsigma_i \rho(s_1, s_2). \quad (2.1)$$

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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \xrightarrow{\mathcal{D}} \mu.$$

3. Asymptotics of the optimal value

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

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

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

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

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

$$= 0.$$

By the definitions of martingale difference and martingale, 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 $\{\alpha_i : \Theta_i \rightarrow \mathbb{R}\}$ such that*

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

and w.p.1

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

In addition, there exists a modulus of continuity $\varpi(\cdot)$ such that w.p.1

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

for all $(x, y), (x', y') \in \mathcal{X} \times \mathcal{Y}$ and $i \in \mathbb{N}$.

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

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

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

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

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

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

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

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. \square

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 $\lim_{n \rightarrow \infty} \vartheta_n = \vartheta$ and $\lim_{n \rightarrow \infty} \mathbb{D}(\mathcal{T}_{x,n}, \mathcal{T}_x) = 0$ 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 and 3.3-3.4, 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 [18], we complete the proof. \square

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

The dual problem of (1.3) is as follows:

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} f(x, y). \quad (3.1)$$

Let \mathcal{T}_y denote the set of optimal solutions of (3.1). If $f \in \mathcal{C}$ and Assumptions 3.7-3.8 hold, then there is no duality gap between (1.3) and (3.1). That is,

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y) = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} f(x, y), \quad (3.2)$$

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 exists a constant $\tilde{c} > f^2(x_0, y_0)$ such that*

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

and

$$\sup_{i \in \mathbb{N}} \sup_{\theta_i \in \Theta_i} \left| \frac{F(x_0, y_0, \theta_i)}{\psi_i(\theta_i)} \right| < \infty.$$

(B2) *There exists an integrable function $\alpha : \mathbb{R}^r \rightarrow \mathbb{R}_+$ 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*

$$|F(x_1, y_1, \theta) - F(x_2, y_2, \theta)| \leq \alpha(\theta)(\|x_1 - x_2\| + \|y_1 - y_2\|) \quad (3.3)$$

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

(B3) *There exists a sequence of random measurable functions $\{\mathcal{A}_i : \Theta_i \rightarrow \mathbb{R}_+\}$ and a function $\mathcal{V} : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_+$ such that for any $i \in \mathbb{N}$, w.p.1*

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

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

and w.p.1

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

where \tilde{b} is a positive constant and $\sup_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \mathcal{V}(x, y) < \infty$.

Then

$$\vartheta_n = \inf_{x \in \mathcal{T}_x} \sup_{y \in \mathcal{T}_y} f_n(x, y) + o_p\left(\frac{1}{\sqrt{n}}\right). \quad (3.4)$$

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

$$\sqrt{n}(\vartheta_n - \vartheta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\hat{x}, \hat{y})), \quad (3.5)$$

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

Proof. According to Lemma 3.1, $\{\Upsilon_i(x, y), \mathcal{G}_i\}$ is a martingale difference sequence for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$. First, we show that Υ_i fulfills the conditions in Theorem 2.2.

It is not difficult to see that $f(x, y)$ is well defined and finite valued. Let $\tilde{\alpha} := \int_{\Theta} \alpha(\theta) d\theta$. Obviously, $0 < \tilde{\alpha} < \infty$. Integrating over (3.3), we have

$$|f(x_1, y_1) - f(x_2, y_2)| \leq \tilde{\alpha}(\|x_1 - x_2\| + \|y_1 - y_2\|) \quad (3.6)$$

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. Thus, we have $f \in \mathcal{C}$ and $F(\cdot, \cdot, \theta) \in \mathcal{C}$ for any given $\theta \in \Theta$. Then, $f_n \in \mathcal{C}$. From the previous analysis, we know that (3.2) holds. Further, $\mathcal{T}_x \times \mathcal{T}_y$ is nonempty and forms the set of saddle points.

Let $M_i = \frac{\mathcal{A}_i(\theta_i)}{\psi_i(\theta_i)}$ and $\beta(x, y) = \mathcal{V}^{-1}(x, y)$. It follows from **(B3)** that

$$|\beta(x_1, y_1)\Upsilon_i(x_1, y_1)| - |\beta(x_2, y_2)\Upsilon_i(x_2, y_2)| \leq M_i$$

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

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

Together with **(B1)**, it is not hard to verify that

$$\frac{1}{n} \sum_{i=1}^n E[\Upsilon_i^2(x_0, y_0) \mid \mathcal{F}_{i-1}] \xrightarrow{P} c,$$

where $c = \tilde{c} - f^2(x_0, y_0)$. Furthermore, w.p.1

$$|\Upsilon_i(x_0, y_0)| \leq \left| \frac{F(x_0, y_0, \theta_i)}{\psi_i(\theta_i)} \right| + |f(x_0, y_0)|.$$

Then, applying **(B1)** again, it is clear that **(A4)** and **(A5)** are satisfied.

According to (3.3) and (3.6), we have

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

Let $\varsigma_i = \frac{\alpha(\theta_i)}{\psi_i(\theta_i)} + \tilde{\alpha}$. Obviously, $\varsigma_i > 0$. Hence (2.1) is satisfied. Moreover, we get

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

This implies that **(A6)** is verified.

From the above discussion, Υ_i satisfies the conditions in Theorem 2.2. Then, applying Theorem 2.2, there exists a Gaussian measure Y on $C(\mathcal{X}, \mathcal{Y})$ such that for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \Upsilon_i(x, y) = \sqrt{n}(f_n(x, y) - f(x, y)) \xrightarrow{\mathcal{D}} Y(x, y) \quad \text{as } n \rightarrow \infty.$$

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

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

and

$$\sqrt{n}(\vartheta_n - \vartheta) \xrightarrow{\mathcal{D}} \inf_{x \in \mathcal{T}_x} \sup_{y \in \mathcal{T}_y} Y(x, y). \quad (3.7)$$

By the properties of martingale difference sequences, we have that $Y(x, y)$ follows a normal distribution with mean 0 and variance $\sigma^2(x, y) = \frac{c\mathcal{V}^2(x, y)}{\mathcal{V}^2(x_0, y_0)}$, see Remark 3.1 for details. Hence, (3.4) is proved.

Finally, if $\mathcal{T}_x = \{\hat{x}\}$ and $\mathcal{T}_y = \{\hat{y}\}$ are singletons, then (3.5) follows from (3.7). Thus, the proof is complete. \square

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

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

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

$$E[\Upsilon_i(x, y)\Upsilon_j(x, y)] = E[E[\Upsilon_i(x, y)\Upsilon_j(x, y) \mid \mathcal{G}_j]] = E[\Upsilon_j(x, y)E[\Upsilon_i(x, y) \mid \mathcal{G}_j]].$$

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

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

Therefore,

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

and

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

Thus, $\text{Var}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \Upsilon_i(x, y)\right] = \frac{1}{n} \sum_{i=1}^n \text{Var}[\Upsilon_i(x, y)] = \frac{1}{n} \sum_{i=1}^n E[\Upsilon_i^2(x, y)]$. It follows from the proof of Theorem 4.1 in [20], we have

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

where β is a nonzero real-valued function on $\mathcal{X} \times \mathcal{Y}$. Since $\beta(x, y) = \mathcal{V}^{-1}(x, y)$ in Theorem 3.2, we get $\sigma^2(x, y) = [\tilde{c} - f^2(x_0, y_0)] \frac{\mathcal{V}^2(x, y)}{\mathcal{V}^2(x_0, y_0)} > 0$ for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Notice that Y is a Gaussian measure on $C(\mathcal{X}, \mathcal{Y})$. Obviously, $E[Y(x, y)] = 0$ and $\text{Var}[Y(x, y)] = \sigma^2(x, y)$. 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, 13, 14, 16, 17]. A comprehensive review can be found in reference [18, Chapter 6]. The specific problem we focus on here is stated as follows:

$$\min_{x \in \mathcal{X}} \rho_\gamma(G(x, \theta)), \quad (4.1)$$

where $G : \mathcal{X} \times \Theta \rightarrow \mathbb{R}$ with $\mathcal{X} \subset \mathbb{R}^m$, $\theta : \Omega \rightarrow \Theta \subset \mathbb{R}^r$ is a random vector on (Ω, \mathcal{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 [13, 18] for details. To be specific, the equivalent form of (4.1) is

$$\min_{(x, t) \in \mathcal{X} \times \mathbb{R}} \max_{\lambda \in [0, 1]} E\{h(t, \lambda, G(x, \theta))\}, \quad (4.2)$$

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

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, (4.2) is a convex-concave minimax problem. Furthermore, if we assume that \mathcal{X}^* is the set of optimal solutions for the problem (4.1), then the set of optimal solutions for the problem (4.2) is $\mathcal{T} = \{(x^*, t^*) : x^* \in \mathcal{X}^*, t^* = E[G(x^*, \theta)]\}$. Accordingly, the set of optimal solutions for its dual problem is $\mathcal{T}_d = [\nu^*, \nu^{**}]$, where $\nu^* = \text{Prob}\{G(x^*, \theta) < E[G(x^*, \theta)]\}$ and $\nu^{**} = \text{Prob}\{G(x^*, \theta) \leq E[G(x^*, \theta)]\}$ with $x^* \in \mathcal{X}^*$. The analysis of this part is detailed in [13, 16].

If we use $\phi(\theta)$ to represent the probability density function of θ , it is not difficult to see that $H(x, t, \lambda, \theta) := h(t, \lambda, G(x, \theta)) \cdot \phi(\theta)$ 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 $(x, t) \in \mathcal{X} \times \mathbb{R}$. According to the previous analysis, t^* is related to x^* . Then, we have

$$\begin{aligned} \min_{(x, t) \in \mathcal{X} \times \mathbb{R}} \max_{\lambda \in [0, 1]} E\{h(t, \lambda, G(x, \theta))\} &= \min_{(x, t) \in \mathcal{X} \times \mathbb{R}} \max_{\lambda \in [0, 1]} \int_{\Theta} H(x, t, \lambda, \theta) d\theta \\ &= \min_{x \in \mathcal{X}} \max_{\lambda \in [0, 1]} \int_{\Theta} H(x, E[G(x, \theta)], \lambda, \theta) d\theta. \end{aligned} \quad (4.3)$$

Denote $\hat{F}(x, \lambda, \theta) := H(x, E[G(x, \theta)], \lambda, \theta)$ and $\hat{f}(x, \lambda) = \int_{\Theta} \hat{F}(x, \lambda, \theta) d\theta$. The information above implies that Assumptions 3.2 and 3.7-3.9 hold. Let $\theta_i, \mathcal{G}_i, \psi_i$ and Θ_i be defined as in the introduction, where $i \in \mathbb{N}$. Let Assumption 3.1 hold. Then, the SAA with AMIS problem associated with problem (4.3) is as follows:

$$\min_{x \in \mathcal{X}} \max_{\lambda \in [0, 1]} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\hat{F}(x, \lambda, \theta_i)}{\psi_i(\theta_i)} \right\}. \quad (4.4)$$

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 $(x_0, \lambda_0) \in \mathcal{X} \times [0, 1]$, there exist constants $\hat{c} > \hat{f}^2(x_0, \lambda_0)$ such that*

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

and

$$\sup_{i \in \mathbb{N}} \sup_{\theta_i \in \Theta_i} \left| \frac{\hat{F}(x_0, \lambda_0, \theta_i)}{\psi_i(\theta_i)} \right| < \infty.$$

(B'2) There exists an integrable function $\alpha : \mathbb{R}^r \rightarrow \mathbb{R}_+$ whose support is contained in Θ such that for any $(x_1, \lambda_1), (x_2, \lambda_2) \in \mathcal{X} \times [0, 1]$, w.p.1

$$|\hat{F}(x_1, \lambda_1, \theta) - \hat{F}(x_2, \lambda_2, \theta)| \leq \alpha(\theta)(\|x_1 - x_2\| + \|\lambda_1 - \lambda_2\|),$$

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

(B'3) There exists a sequence of random measurable functions $\{\mathcal{A}_i : \Theta_i \rightarrow \mathbb{R}_+\}$ and a function $\mathcal{V} : \mathcal{X} \times [0, 1] \rightarrow \mathbb{R}_+$ such that for any $i \in \mathbb{N}$, w.p.1

$$\begin{aligned} |\hat{F}(x, \lambda, \theta) - \hat{f}(x, \lambda)\psi_i(\theta)| &\leq \mathcal{V}(x, \lambda)\mathcal{A}_i(\theta), \\ \frac{1}{n} \sum_{i=1}^n E \left[\left(\frac{\mathcal{A}_i(\theta_i)}{\psi_i(\theta_i)} \right)^2 \middle| \mathcal{G}_{i-1} \right] &\xrightarrow{P} 0, \end{aligned}$$

and w.p.1

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

where \hat{b} is a positive constant and $\sup_{(x, \lambda) \in \mathcal{X} \times [0, 1]} \mathcal{V}(x, \lambda) < \infty$.

Then

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

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

$$\sqrt{n}(\hat{\vartheta}_n - \hat{\vartheta}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(x^*, \nu^*)),$$

where $\hat{\vartheta}$ and $\hat{\vartheta}_n$ denote the optimal values of (4.2) and (4.4), respectively, and $\sigma^2(x^*, \nu^*) = [\hat{c} - \hat{f}^2(x_0, y_0)] \frac{\mathcal{V}^2(x^*, \nu^*)}{\mathcal{V}^2(x_0, y_0)}$.

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. \square

Acknowledgements

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

References

- [1] D. W. K. Andrews, *Generic uniform convergence*, Econometric Theory, 1992, 8(2), 241–257.
- [2] M. F. Bugallo, V. Elvira, L. Martino, D. Luengo, J. Miguez and P. M. Djuric, *Adaptive importance sampling: The past, the present, and the future*, IEEE Signal Process. Mag., 2017, 34(4), 60–79.
- [3] J.-M. Corneut, J.-M. Marin, A. Mira and C. P. Robert, *Adaptive multiple importance sampling*, Scand. J. Stat., 2012, 39(4), 798–812.
- [4] Y. El-Laham, L. Martino, V. Elvira and M. F. Bugallo, *Efficient adaptive multiple importance sampling*, 2019 27th European Signal Processing Conference (EUSIPCO), 2019. DOI: 10.23919/EUSIPCO.2019.8902642.
- [5] C. I. Fábían, C. Wolf, A. Koberstein and L. Suhl, *Risk-averse optimization in two-stage stochastic models: Computational aspects and a study*, SIAM J. Optim., 2015, 25(1), 28–52.
- [6] M. B. Feng, A. Maggiar, J. Staum and A. Wächter, *Uniform convergence of sample average approximation with adaptive multiple importance sampling*, 2018 Winter Simulation Conference (WSC), 2018. DOI: 10.1109/WSC.2018.8632370.
- [7] G. Lan and Z. Zhang, *Optimal methods for convex risk-averse distributed optimization*, SIAM J. Optim., 2023, 33(3), 1518–1557.
- [8] T. Latunde, J. O. Richard, O. O. Esan and D. D. Dare, *Optimal value and post optimal solution in a transportation problem*, J. Nonl. Mod. Anal., 2021, 3(3), 335–348.
- [9] A. Lodi, E. Malaguti, G. Nannicini and D. Thomopulos, *Nonlinear chance-constrained problems with applications to hydro scheduling*, Math. Program., 2022, 191(1), 405–444.
- [10] D. Luengo, L. Martino, M. Bugallo, V. Elvira and S. Särkkä, *A survey of Monte Carlo methods for parameter estimation*, EURASIP J. Adv. Signal Process., 2020. DOI: 10.1186/s13634-020-00675-6.
- [11] A. Maggiar, A. Wächter, I. S. Dolinskaya and J. Staum, *A derivative-free trust-region algorithm for the optimization of functions smoothed via Gaussian convolution using adaptive multiple importance sampling*, SIAM J. Optim., 2018, 28(2), 1478–1507.
- [12] J.-M. Marin, P. Pudlo and M. Sedki, *Consistency of the adaptive multiple importance sampling*, arXiv preprint, 2012. arXiv:1211.2548.
- [13] W. Ogryczak and A. Ruszczyński, *Dual stochastic dominance and related mean-risk models*, SIAM J. Optim., 2002, 13(1), 60–78.
- [14] A. Pichler and R. Schlotter, *Martingale characterizations of risk-averse stochastic optimization problems*, Math. Program., 2020, 181(2, Ser.B), 377–403.
- [15] R. Retkute, P. Touloupou, M.-G. Basáñez, T. D. Hollingsworth and S. E. F. Spencer, *Integrating geostatistical maps and infectious disease transmission models using adaptive multiple importance sampling*, Ann. Appl. Stat., 2021, 15(4), 1980–1998.

- [16] A. Shapiro, *Asymptotics of minimax stochastic programs*, Statist. Probab. Lett., 2008, 78(2), 150–157.
- [17] A. Shapiro, *Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming*, European J. Oper. Res., 2021, 288(1), 1–13.
- [18] A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.
- [19] B. Singh and B. Knueven, *Lagrangian relaxation based heuristics for a chance constrained optimization model of a hybrid solar-battery storage system*, J. Global Optim., 2021, 80(4), 965–989.
- [20] W. Zhang and Y. Li, *Uniform exponential convergence of SAA with AMIS and asymptotics of its optimal value*, Submitted for Publication.