Available online at www.sciencedirect.com
ScienceDirect
Statistics & Probability Letters 78 (2008) 150–157
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Asymptotics of minimax stochastic programs ☆

Alexander Shapiro

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA

Received 10 November 2006; received in revised form 3 April 2007; accepted 23 May 2007
Available online 11 June 2007

Abstract

We discuss in this paper asymptotics of the sample average approximation (SAA) of the optimal value of a minimax stochastic programming problem. The main tool of our analysis is a specific version of the infinite dimensional delta method. As an example, we discuss asymptotics of SAA of risk averse stochastic programs involving the absolute semideviation risk measure. Finally, we briefly discuss exponential rates of convergence of the optimal value of SAA problems.

Keywords: Sample average approximation; Infinite dimensional delta method; Functional central limit theorem; Minimax stochastic programming; Absolute semideviation risk measure; Exponential rate of convergence

1. Introduction

Consider the following minimax stochastic problem:

$$\min_{x\in X}\max_{y\in Y}\{f(x,y):=\mathbb{E}[F(x,y,\xi)]\}, \tag{1.1}$$

where $X\subset\mathbb{R}^m$, $Y\subset\mathbb{R}^n$, ξ is a random vector having probability distribution P supported on a set $\Xi\subset\mathbb{R}^d$ and $F: X\times Y\times\Xi\to\mathbb{R}$. (Unless stated otherwise we assume throughout the paper that the expectations are taken with respect to the probability distribution P.) Suppose that we have an iid sample $\xi^1,\dots,\xi^N$ of N realizations of the random vector ξ. Then we can approximate problem (1.1) by the so-called sample average approximation (SAA) minimax problem:

$$\min_{x\in X}\max_{y\in Y}\left\{\hat{f}_N(x,y):=\frac{1}{N}\sum_{j=1}^{N}F(x,y,\xi^j)\right\}. \tag{1.2}$$

We discuss in this paper asymptotic properties of the optimal value $\hat{v}_N$ of the SAA problem (1.2) viewed as an estimator of the optimal value v^* of the "true" problem (1.1).

Of course, if the set Y is a singleton, say $Y=\{y^*\}$, then problem (1.1) becomes a stochastic programming problem of minimization of the expectation function $f(x,y^*)$ over $x\in X$. Statistical properties of SAAs of

☆ Supported in part by the National Science Foundation awards DMS-0510324 and DMI-0619977.
E-mail address: ashapiro@isye.gatech.edu

0167-7152/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2007.05.012

such (minimization) problems were studied extensively (see, e.g., Shapiro, 2003, and references therein). This study is motivated by the following example. Consider the following risk averse stochastic problem:

$$\min_{x \in X} \rho_\lambda[G(x,\xi)], \tag{1.3}$$

where $G : \mathbb{R}^m \times \Xi \to \mathbb{R}$ and $\rho_\lambda[Z] := \mathbb{E}[Z] + \lambda \mathbb{E}\{[Z - \mathbb{E}(Z)]_+\}$ is the so-called absolute semideviation risk measure with $\lambda \in [0,1]$ being the weight constant and $[a]_+ := \max\{0, a\}$. Such risk averse stochastic problems were investigated in a number of recent publications (see, e.g., Ogryczak and Ruszczyński, 2002; Ahmed, 2006, and references therein). It is possible to show (cf. Ogryczak and Ruszczyński, 2002) that

$$\rho_\lambda[Z] = \inf_{t \in \mathbb{R}} \sup_{\alpha \in [0,1]} \mathbb{E}\{Z + \lambda\alpha[Z - t]_+ + \lambda(1-\alpha)[t - Z]_+\}, \tag{1.4}$$

and hence to represent problem (1.3) as a minimax stochastic problem of form (1.1).

Of course, there is a large number of publications on all types of minimax estimation, decision rules, etc. in the statistics literature. It will be beyond the scope of this paper to give even a brief survey of this topic.

We assume that the sets X and Y are nonempty, closed and convex, and for every $\xi \in \Xi$ the function $F(\cdot,\cdot,\xi)$ is *convex–concave*, i.e., for all y the function $F(\cdot, y, \xi)$ is convex, and for all x the function $F(x,\cdot,\xi)$ is concave. The main result of this paper is that, under mild additional conditions,

$$\hat{v}_N = \min_{x \in X_f^*} \max_{y \in Y_f^*} \hat{f}_N(x,y) + o_p(N^{-1/2}), \tag{1.5}$$

where $X_f^* \times Y_f^*$ is the set of saddle points of the minimax problem (1.1). The asymptotic result (1.5) can be viewed as an extension of a theory presented in Shapiro (1991) where a similar result was established for minimization-type stochastic problems. The main tool of our analysis will be the infinite dimensional delta method, which was initiated in the works of Gill (1989), Grübel (1988), King (1989) and Shapiro (1991) (for a recent survey see Römisch, 2005).

The remainder of this paper is organized as follows. In the next section we derive a form of a delta theorem giving asymptotics of the optimal value of minimax stochastic programming problems. In Section 3 we employ this theorem to describe asymptotics of the optimal value of minimax SAA problems and, in particular, asymptotics of SAA risk averse stochastic programs involving the absolute semideviation risk measure. Finally, in Section 4 we briefly discuss exponential rates of convergence of the optimal value of SAA problems.

2. Minimax delta theorem

We assume in this section that the sets X and Y are nonempty convex and *compact*. Consider the space $C(X,Y)$ of continuous functions $\phi : X \times Y \to \mathbb{R}$ equipped with the corresponding sup-norm, and set $\mathfrak{C} \subset C(X,Y)$ formed by convex–concave functions, i.e., $\phi \in \mathfrak{C}$ iff $\phi(\cdot, y)$ is convex for every $y \in Y$ and $\phi(x,\cdot)$ is concave for every $x \in X$. It is not difficult to see that the set $\mathfrak{C}$ is a closed (in the norm topology of $C(X,Y)$) and convex cone. Consider the optimal value function $V : C(X,Y) \to \mathbb{R}$ defined as

$$V(\phi) := \inf_{x \in X} \sup_{y \in Y} \phi(x,y). \tag{2.1}$$

Note that the function $x \mapsto \sup_{y \in Y} \phi(x,y)$ is continuous, and hence the minimax problem in the right-hand side of (3.1) has a finite optimal value and attains its minimal optimal value at a nonempty compact subset X_ϕ^* of X. Similarly, the function $y \mapsto \inf_{x \in X} \phi(x,y)$ attains its maximal optimal value at a nonempty compact subset Y_ϕ^* of Y. If $\phi \in \mathfrak{C}$, then by the classical minimax theorem (due to von Neumann) we have that $X_\phi^* \times Y_\phi^*$ forms the set of saddle points of $\phi(x,y)$ and

$$\inf_{x \in X} \sup_{y \in Y} \phi(x,y) = \sup_{y \in Y} \inf_{x \in X} \phi(x,y). \tag{2.2}$$

Note also that for any $\phi_1, \phi_2 \in C(X, Y)$,

$$|V(\phi_1) - V(\phi_2)| \leqslant \sup_{x\in X}\left|\sup_{y\in Y}\phi_1(x,y) - \sup_{y\in Y}\phi_2(x,y)\right| \leqslant \sup_{(x,y)\in X\times Y}|\phi_1(x,y) - \phi_2(x,y)|, \tag{2.3}$$

i.e., the function $V(\cdot)$ is Lipschitz continuous (with Lipschitz constant one).

We denote by $T_{\mathfrak{C}}(\phi)$ the tangent cone to the set $\mathfrak{C}$ at a point $\phi \in \mathfrak{C}$. Recall that $\gamma \in T_{\mathfrak{C}}(\phi)$ iff the distance $\operatorname{dist}(\phi + t\gamma, \mathfrak{C}) = o(t)$ for $t \geqslant 0$. It is said that $V(\cdot)$ is Hadamard directionally differentiable at $\phi \in \mathfrak{C}$ tangentially to the set $\mathfrak{C}$ if the following limit exists for any $\gamma \in T_{\mathfrak{C}}(\phi)$:

$$V'_\phi(\gamma) := \lim_{\substack{t\downarrow 0\\ \mathfrak{C}\ni\gamma'\to\gamma}} \frac{V(\phi + t\gamma') - V(\phi)}{t}. \tag{2.4}$$

We will need the following which is an adaptation of a result going back to Gol'shtein (1972).

Proposition 2.1. *The optimal value function $V(\cdot)$ is Hadamard directionally differentiable at any $\phi \in \mathfrak{C}$ tangentially to the set $\mathfrak{C}$ and*

$$V'_\phi(\gamma) = \inf_{x\in X^*_\phi}\sup_{y\in Y^*_\phi}\gamma(x,y) \tag{2.5}$$

for any $\gamma \in T_{\mathfrak{C}}(\phi)$.

Proof. Consider a sequence $\gamma_k \in C(X, Y)$ converging (in the norm topology) to $\gamma \in T_{\mathfrak{C}}(\phi)$ and such that $\zeta_k := \phi + t_k\gamma_k \in \mathfrak{C}$ for some sequence $t_k \downarrow 0$. For a point $x^* \in X^*_\phi$ we have

$$V(\phi) = \sup_{y\in Y}\phi(x^*, y) \quad \text{and} \quad V(\zeta_k) \leqslant \sup_{y\in Y}\zeta_k(x^*, y).$$

Since Y is compact and $\zeta_k(x^*, \cdot)$ is continuous, we have that the set $\arg\max_{y\in Y}\zeta_k(x^*, y)$ is nonempty. Let $y_k \in \arg\max_{y\in Y}\zeta_k(x^*, y)$. We have that $\arg\max_{y\in Y}\phi(x^*, y) = Y^*_\phi$ and, since ζ_k tends to ϕ, we have that y_k tends in distance to Y^*_ϕ (i.e., the distance from y_k to Y^*_ϕ tends to zero as $k \to \infty$). By passing to a subsequence if necessary we can assume that y_k converges to a point $y^* \in Y$ as $k \to \infty$. It follows that $y^* \in Y^*_\phi$, and of course we have that $\sup_{y\in Y}\phi(x^*, y) \geqslant \phi(x^*, y_k)$. We obtain

$$V(\zeta_k) - V(\phi) \leqslant \zeta_k(x^*, y_k) - \phi(x^*, y_k) = t_k\gamma_k(x^*, y_k) = t_k\gamma(x^*, y^*) + o(t_k).$$

We obtain that for any $x^* \in X^*_\phi$ there exists $y^* \in Y^*_\phi$ such that

$$\limsup_{k\to\infty}\frac{V(\zeta_k) - V(\phi)}{t_k} \leqslant \gamma(x^*, y^*).$$

It follows that

$$\limsup_{k\to\infty}\frac{V(\zeta_k) - V(\phi)}{t_k} \leqslant \inf_{x\in X^*_\phi}\sup_{y\in Y^*_\phi}\gamma(x,y). \tag{2.6}$$

In order to prove the converse inequality we proceed as follows. Consider a sequence $x_k \in \arg\min_{x\in X}\theta_k(x)$, where $\theta_k(x) := \sup_{y\in Y}\zeta_k(x, y)$. We have that $\theta_k : X \to \mathbb{R}$ are continuous functions converging uniformly in $x \in X$ to the function $\vartheta(x) := \sup_{y\in Y}\phi(x, y)$. Consequently, x_k converges in distance to the set $\arg\min_{x\in X}\vartheta(x)$, which is equal to X^*_ϕ. By passing to a subsequence if necessary we can assume that x_k converges to a point $x^* \in X^*_\phi$. For any $y \in Y^*_\phi$ we have $V(\phi) \leqslant \phi(x_k, y)$. Since $\zeta_k \in \mathfrak{C}$, i.e., $\zeta_k(x, y)$ is convex–concave, it has a nonempty set of saddle points $X^*_{\zeta_k} \times Y^*_{\zeta_k}$. We have that $x_k \in X^*_{\zeta_k}$, and hence $V(\zeta_k) \geqslant \zeta_k(x_k, y)$ for any $y \in Y$. It follows that for any $y \in Y^*_\phi$ the following holds:

$$V(\zeta_k) - V(\phi) \geqslant \zeta_k(x_k, y) - \phi(x_k, y) = t_k\gamma_k(x^*, y) + o(t_k),$$

and hence

$$\liminf_{k\to\infty} \frac{V(\zeta_k) - V(\phi)}{t_k} \geqslant \gamma(x^*, y).$$

Since y was an arbitrary element of Y^*_ϕ, we obtain that

$$\liminf_{k\to\infty} \frac{V(\zeta_k) - V(\phi)}{t_k} \geqslant \sup_{y\in Y^*_\phi} \gamma(x^*, y),$$

and hence

$$\liminf_{k\to\infty} \frac{V(\zeta_k) - V(\phi)}{t_k} \geqslant \inf_{x\in X^*_\phi} \sup_{y\in Y^*_\phi} \gamma(x, y). \tag{2.7}$$

The assertion of the theorem follows from (2.6) and (2.7). □

We use the following version of (infinite dimensional) delta theorem (Shapiro, 1991, Theorem 2.1).

Proposition 2.2. *Let $\mathscr{X}$ be a separable Banach space, equipped with its Borel sigma algebra, $\mathscr{C}$ be a closed convex subset of $\mathscr{X}$, $g : \mathscr{X} \to \mathbb{R}$ be a measurable and Hadamard directionally differentiable at a point $\mu \in \mathscr{C}$ tangentially to the set $\mathscr{C}$ function, $\tau_N \to \infty$ be a sequence of positive numbers and $\{\psi_N\}$ be a sequence of random elements of $\mathscr{X}$ such that $\psi_N \in \mathscr{C}$ w.p.1 and $\tau_N(\psi_N - \mu)$ converges in distribution to a random element Z, written as $\tau_N(\psi_N - \mu) \xrightarrow{\mathscr{D}} Z$. Then,*

$$g(\psi_N) = g(\mu) + g'_\mu(\psi_N - \mu) + o_p(\tau_N^{-1}) \tag{2.8}$$

and

$$\tau_N(g(\psi_N) - g(\mu)) \xrightarrow{\mathscr{D}} g'_\mu(Z). \tag{2.9}$$

Propositions 2.1 and 2.2 imply the following result.

Theorem 2.1. *Consider the space $\mathscr{X} := C(X, Y)$ and the set $\mathfrak{C} \subset \mathscr{X}$ of convex–concave functions. Let τ_N be a sequence of positive numbers tending to infinity, $\psi \in \mathfrak{C}$ and $\{\hat{\psi}_N\}$ be a sequence of random elements of $\mathscr{X}$ such that $\hat{\psi}_N \in \mathfrak{C}$ w.p.1 and $\tau_N(\hat{\psi}_N - \psi)$ converges in distribution to a random element Ψ of $\mathscr{X}$. Denote $\vartheta^* := \inf_{x\in X}\sup_{y\in Y} \psi(x, y)$ and $\hat{\vartheta}_N := \inf_{x\in X}\sup_{y\in Y} \hat{\psi}_N(x, y)$. Then,*

$$\hat{\vartheta}_N = \inf_{x\in X^*_\psi} \sup_{y\in Y^*_\psi} \hat{\psi}_N(x, y) + o_p(\tau_N^{-1}) \tag{2.10}$$

and

$$\tau_N(\hat{\vartheta}_N - \vartheta^*) \xrightarrow{\mathscr{D}} \inf_{x\in X^*_\psi} \sup_{y\in Y^*_\psi} \Psi(x, y). \tag{2.11}$$

Proof. Consider the optimal value function $V : \mathscr{X} \to \mathbb{R}$. This function is continuous and hence measurable. Clearly, we have that $\vartheta^* = V(\psi)$ and $\hat{\vartheta}_N = V(\hat{\psi}_N)$. By Proposition 2.1, the optimal value function $V(\cdot)$ is Hadamard directionally differentiable at ψ tangentially to the set $\mathfrak{C}$ and formula (2.5) holds. Together with formula (2.9) of the above delta theorem this implies (2.11). Formula (2.10) follows by (2.8) and by noting that $\psi(x^*, y^*) = \vartheta^*$ for any $(x^*, y^*) \in X^*_\psi \times Y^*_\psi$. □

3. Applications to minimax stochastic problems

Consider the minimax stochastic problem (1.1) and its SAA (1.2) based on an iid sample. Let us make the following assumptions:

(A1) The sets X and Y are nonempty, convex and closed, and for every $\xi \in \Xi$ the function $F(\cdot, \cdot, \xi)$ is convex–concave on $X \times Y$.

(A2) Problem (1.1) and its dual

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

have nonempty and bounded sets of optimal solutions $X_f^* \subset X$ and $Y_f^* \subset Y$, respectively.

(A3) For every $(x,y)\in X\times Y$, the function $F(x,y,\cdot)$ is measurable.

(A4) For some point $(\bar{x},\bar{y})\in X\times Y$, the expectation $\mathbb{E}[F(\bar{x},\bar{y},\xi)^2]$ is finite.

(A5) There exists a measurable function $c:\Xi\to\mathbb{R}_+$ such that $\mathbb{E}[c(\xi)^2]$ is finite and the inequality

$$|F(x',y',\xi)-F(x,y,\xi)|\leqslant c(\xi)(\|x'-x\|+\|y'-y\|) \tag{3.2}$$

holds for all $(x,y),(x',y')\in X\times Y$ and a.e. $\xi\in\Xi$.

It follows that the expected value function $f(x,y)$ is well defined, finite valued and Lipschitz continuous with Lipschitz constant $\mathbb{E}[c(\xi)]$. By assumption (A1) we have that the function $f(x,y)$ is convex–concave, and by (A2) that the optimal values of problems (1.1) and (3.1) are equal to each other and $X_f^*\times Y_f^*$ forms the set of saddle points of these problems. Moreover, optimal solutions of the respective SAA problems converge (in distance) w.p.1 to X_f^* and Y_f^*, respectively (e.g., Shapiro, 2003, Theorem 4, p. 360). Therefore, without loss of generality we can assume that the sets X and Y are compact.

The above assumptions (A3)–(A5) are sufficient for the central limit theorem to hold in the Banach space $C(X,Y)$. That is, the sequence $N^{1/2}(\hat{f}_N-f)$, of random elements of $C(X,Y)$, converges in distribution to a (Gaussian) random element Ψ (see, e.g., Araujo and Giné, 1980, Chapter 7, Corollary 7.17). For any fixed point $(x,y)\in X\times Y$, $\Psi(x,y)$ is a real valued random variable having normal distribution with zero mean and variance $\sigma^2(x,y)$ equal to the variance of $F(x,y,\xi)$, i.e.,

$$\sigma^2(x,y)=\mathbb{E}[F(x,y,\xi)^2]-f(x,y)^2. \tag{3.3}$$

As a consequence of Theorem 2.1 we have the following result.

Theorem 3.1. *Consider the minimax stochastic problem* (1.1) *and the SAA problem* (1.2) *based on an iid sample. Suppose that assumptions* (A1)–(A5) *hold and let* v^* *and* $\hat{v}_N$ *be the optimal values of* (1.1) *and* (1.2), *respectively. Then,*

$$\hat{v}_N=\min_{x\in X_f^*}\max_{y\in Y_f^*}\hat{f}_N(x,y)+o_p(N^{-1/2}). \tag{3.4}$$

Moreover, if the sets $X_f^*=\{x^*\}$ *and* $Y_f^*=\{y^*\}$ *are singletons, then* $N^{1/2}(\hat{v}_N-v^*)$ *converges in distribution to normal with zero mean and variance* $\sigma^2=\mathbb{V}\mathrm{ar}[F(x^*,y^*,\xi)]$.

Let us finally discuss application of the above results to problem (1.3). We assume that the set X is nonempty convex and closed, the expectation $\mathbb{E}[G(x,\xi)]$ is well defined and finite valued for every $x\in X$, for a.e. $\xi\in\Xi$ the function $G(\cdot,\xi)$ is convex and problem (1.3) has a nonempty and bounded set X^* of optimal solutions.

For a random variable Z having finite first order moment we have that

$$\sup_{\alpha\in[0,1]}\mathbb{E}\{Z+\lambda\alpha[Z-t]_+ +\lambda(1-\alpha)[t-Z]_+\}=\mathbb{E}[Z]+\lambda\max\{\mathbb{E}([Z-t]_+),\mathbb{E}([t-Z]_+)\} \tag{3.5}$$

and $\mathbb{E}([Z-t]_+)=\mathbb{E}([t-Z]_+)$ if $t=\mathbb{E}[Z]$. Consequently, the minimum, over $t\in\mathbb{R}$, of the right-hand side of (3.5) is attained at $t^*=\mathbb{E}[Z]$, and hence formula (1.4) follows. Because of (1.4) we can write problem (1.3) in the following equivalent form:

$$\min_{(x,t)\in X\times\mathbb{R}}\max_{\alpha\in[0,1]}\mathbb{E}[H_\lambda(G(x,\xi)-t,\alpha)+t], \tag{3.6}$$

where

$$H_\lambda(z,\alpha):=z+\lambda\alpha[z]_+ +\lambda(1-\alpha)[-z]_+.$$

The function $H_\lambda(z,\alpha)$ is convex in z and linear (and hence concave) in α. Moreover, for any $\alpha \in [0,1]$ and $\lambda \in [0,1]$, this function is monotonically nondecreasing in z. It follows that the function $H_\lambda(G(x,\xi)-t,\alpha)+t$ is convex in (x,t) and concave (linear) in α.

The SAA of problem (1.3) is obtained by replacing the "true" distribution P with its sample approximation $P_N := N^{-1}\sum_{j=1}^N \Delta(\xi^j)$, where $\Delta(\xi)$ denotes measure of mass one at point $\xi \in \Xi$. That is, the SAA problem associated with problem (1.3) is

$$\min_{x\in X} \frac{1}{N}\sum_{j=1}^N \{G(x,\xi^j) + \lambda[G(x,\xi^j) - \hat{g}_N(x)]_+\}, \tag{3.7}$$

where $\hat{g}_N(x) := N^{-1}\sum_{j=1}^N G(x,\xi^j)$. Alternatively this SAA problem can be written in the minimax form:

$$\min_{(x,t)\in X\times\mathbb{R}} \max_{\alpha\in[0,1]} \left\{ \frac{1}{N}\sum_{j=1}^N H_\lambda(G(x,\xi^j)-t,\alpha)+t \right\}. \tag{3.8}$$

The sets of optimal solutions of the minimax problem (3.6) and its dual can be described as follows. We have that if (x^*,t^*) is an optimal solution of (3.6), then $t^* = \mathbb{E}[G(x^*,\xi)]$ and the set of optimal solutions of the dual of the problem (3.6) is interval $A^* = [\alpha', \alpha^*]$, where

$$\alpha' := \text{Prob}\{G(x^*,\xi) < \mathbb{E}[G(x^*,\xi)]\} \quad \text{and} \quad \alpha^* := \text{Prob}\{G(x^*,\xi) \leqslant \mathbb{E}[G(x^*,\xi)]\}. \tag{3.9}$$

It follows that this interval A^* is the same for any optimal solution $x^* \in X^*$. Theorem 3.1 can be applied now in a straightforward way. That is, under additional assumptions (A3)–(A5) adapted to the considered problem, we have that

$$\hat{v}_N = \min_{\substack{x\in X^* \\ t=\mathbb{E}[G(x,\xi)]}} \max_{\alpha\in A^*} \frac{1}{N}\sum_{j=1}^N \{G(x,\xi^j) + \lambda\alpha[G(x,\xi^j)-t]_+ + \lambda(1-\alpha)[t-G(x,\xi^j)]_+\} + o_p(N^{-1/2}). \tag{3.10}$$

In particular, if $X^* = \{x^*\}$ is a singleton and $\text{Prob}\{G(x^*,\xi) = \mathbb{E}[G(x^*,\xi)]\} = 0$, then $A^* = \{\alpha^*\}$ is a singleton. In that case $N^{1/2}(\hat{v}_N - v^*)$ converges in distribution to normal with zero mean and variance σ^2 given by

$$\sigma^2 = \mathbb{V}\text{ar}\{G(x^*,\xi) + \lambda\alpha^*[G(x^*,\xi) - \mathbb{E}[G(x^*,\xi)]]_+ + \lambda(1-\alpha^*)[\mathbb{E}[G(x^*,\xi)] - G(x^*,\xi)]_+\}. \tag{3.11}$$

4. Exponential rate of convergence

It is also possible to show that, under mild regularity conditions, the optimal value $\hat{v}_N$ of the SAA problem converges to its true counterpart v^* at an *exponential* rate. For this type of results there is no need to develop a new theory. Let us briefly outline the derivations.

Let $\varepsilon > 0$ be a given constant. We have that if

$$\sup_{(x,y)\in X\times Y} |\hat{f}_N(x,y) - f(x,y)| \leqslant \varepsilon, \tag{4.1}$$

then $|\hat{v}_N - v^*| \leqslant \varepsilon$ (this follows from (2.3)). It is possible to show that, under certain regularity conditions, probability of the above event tends to one at an exponential rate. That is, let us make the following assumptions. For $(x,y) \in X \times Y$ we denote by

$$M_{(x,y)}(t) := \mathbb{E}[\exp\{t(F(x,y,\xi) - f(x,y))\}]$$

the moment generating function of the random variable $F(x,y,\xi)$ corrected for its mean.

(B1) The sets X and Y are nonempty and compact, and the expectation function $f(x,y)$ is well defined and finite valued.

(B2) There exists a constant $L > 0$ such that the inequality

$$|F(x',y',\xi) - F(x,y,\xi)| \leqslant L(\|x'-x\| + \|y'-y\|) \tag{4.2}$$

holds for all $(x,y),(x',y') \in X \times Y$ and a.e. ξ.

(B3) For every $(x,y) \in X \times Y$ the moment generating function $M_{(x,y)}(t)$ is finite valued for all t in a neighborhood of zero.

It is not difficult to show that if assumptions (B1)–(B3) hold, then there exist positive constants $C = C(\varepsilon)$ and $\beta = \beta(\varepsilon)$, independent of N, such that

$$\operatorname{Prob}\left\{ \sup_{(x,y)\in X\times Y} |\hat{f}_N(x,y) - f(x,y)| > \varepsilon \right\} \leqslant C\exp\{-N\beta\}. \tag{4.3}$$

Moreover, suppose that assumption (B3) is strengthened to:

(B4) There is a constant $\theta > 0$ such that the inequality

$$M_{(x,y)}(t) \leqslant \exp\{\theta^2 t^2/2\} \quad \forall t \in \mathbb{R}, \tag{4.4}$$

holds for all $(x,y) \in X \times Y$.

Then the constants C and β in the (uniform) exponential bound (4.3) can be estimated in the form

$$\operatorname{Prob}\left\{ \sup_{(x,y)\in X\times Y} |\hat{f}_N(x,y) - f(x,y)| > \varepsilon \right\} \leqslant \left[\frac{\kappa DL}{\varepsilon}\right]^{m+n} \exp\left\{-\frac{N\varepsilon^2}{32\theta^2}\right\}, \tag{4.5}$$

where $D := \sup_{(x,y),(x',y')\in X\times Y}(\|x-x'\| + \|y-y'\|)$ is the diameter of the set $X \times Y$ and $\kappa > 0$ is a constant depending on the chosen norms (for an elementary derivations of these estimates see, e.g., Shapiro, 2006).

Theorem 4.1. *Suppose that assumptions* (B1)–(B3) *hold. Then there exist positive constants* $C = C(\varepsilon)$ *and* $\beta = \beta(\varepsilon)$, *independent of* N, *such that*

$$\operatorname{Prob}\{|\hat{v}_N - v^*| > \varepsilon\} \leqslant C\exp\{-N\beta\}. \tag{4.6}$$

Moreover, if assumption (B4) *holds, then for* $\alpha \in (0,1)$ *and sample size* N *satisfying*

$$N \geqslant \frac{32\theta^2}{\varepsilon^2}\left[(m+n)\log\left(\frac{\kappa DL}{\varepsilon}\right) + \log(\alpha^{-1})\right], \tag{4.7}$$

we have that

$$\operatorname{Prob}\{|\hat{v}_N - v^*| \leqslant \varepsilon\} \geqslant 1 - \alpha. \tag{4.8}$$

Proof. Estimate (4.6) follows from (4.3). Estimate (4.8) is obtained from (4.5) by setting the right-hand side of (4.5) less than α and solving the obtained inequality. □

The above assumption (B2) can be relaxed to assume that the Lipschitz constant L is a function of ξ (as in assumption (A5)) such that its moment generating function is finite valued. Assumptions (B3) and (B4) postulate that distributions of random variables $F(x,y,\xi)$, $(x,y) \in X \times Y$, have sufficiently light tails. It is also interesting to note that the above exponential bounds do not involve the assumption that the function $F(\cdot,\cdot,\xi)$ is convex–concave.

References

Ahmed, S., 2006. Convexity and decomposition of mean-risk stochastic programs. Math. Programming 106, 433–446.

Araujo, A., Giné, E., 1980. The Central Limit Theorem for Real and Banach Valued Random Variables. Wiley, New York.

Gill, R.D., 1989. Non-and-semiparametric maximum likelihood estimators and the von Mises method (part I). Scand. J. Statist. 16, 97–124.

Gol'shtein, E.G., 1972. Theory of Convex Programming. Translations of Mathematical Monographs, vol. 36. American Mathematical Society, Providence, RI.

Grübel, R., 1988. The length of the short. Ann. Statist. 16, 619–628.

King, A.J., 1989. Generalized delta theorems for multivalued mappings and measurable selections. Math. Oper. Res. 14, 720–736.

Ogryczak, W., Ruszczyński, A., 2002. Dual stochastic dominance and related mean-risk models. SIAM J. Optim. 13, 60–78.
Römisch, W., 2005. Delta method, infinite dimensional. In: Kotz, S., Read, C.B., Balakrishnan, N., Vidakovic, B. (Eds.), Encyclopedia of Statistical Sciences, second ed. Wiley, New York.
Shapiro, A., 1991. Asymptotic analysis of stochastic programs. Ann. Oper. Res. 30, 169–186.
Shapiro, A., 2003. Monte Carlo sampling methods. In: Ruszczyński, A., Shapiro, A. (Eds.), Stochastic Programming. Handbook in Operations Research and Management Science, vol. 10. North-Holland, Amsterdam.
Shapiro, A., 2006. On complexity of multistage stochastic programs. Oper. Res. Lett. 34, 1–8.