Asia-Pacific Journal of Operational Research
Vol. 27, No. 1 (2010) 103–119

DOI: 10.1142/S0217595910002569

SAMPLE AVERAGE APPROXIMATION METHODS FOR A CLASS OF STOCHASTIC VARIATIONAL INEQUALITY PROBLEMS

HUIFU XU

School of Mathematics, University of Southampton
Southampton, SO17 1BJ, UK
h.xu@soton.ac.uk

In this paper we apply the well known sample average approximation (SAA) method to solve a class of stochastic variational inequality problems (SVIPs). We investigate the existence and convergence of a solution to the sample average approximated SVIP. Under some moderate conditions, we show that the sample average approximated SVIP has a solution with probability one and with probability approaching one exponentially fast with the increase of sample size, the solution converges to its true counterpart. Finally, we apply the existence and convergence results to SAA method for solving a class of stochastic nonlinear complementarity problems and stochastic programs with stochastic constraints.

Keywords: Stochastic variational inequality; stochastic complementarity problem; sample average approximation; exponential convergence.

1. Introduction

In this paper, we consider the following stochastic variational inequality problem (SVIP): find $x \in \mathcal{K} \subset \mathbb{R}^n$ such that

$$(y - x)^T \mathbb{E}\left[F(x, \xi(\omega))\right] \geq 0, \quad \forall\, y \in \mathcal{K}, \tag{1.1}$$

where $F : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ is a continuous function, $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$, $\mathbb{E}$ denotes the mathematical expectation with respect to the distribution of ξ, and $\mathcal{K}$ is a nonempty closed convex set. We make a blanket assumption that for every $x \in \mathcal{K}$, $\mathbb{E}\left[F(x, \xi(\omega))\right]$ is well defined. To ease notation, we will use ξ to denote either the random vector $\xi(\omega)$ or an element of $\mathbb{R}^k$, depending on the context.

SVIP model (1.1) is a natural extension of deterministic variational inequality models. Over the past few decades, deterministic variational inequality has been extensively studied for its extensive application in engineering, economics, game theory and networks, see the book on the topic by Facchinei and Pang (2003). While many practical problems only involve deterministic data, there are some important instances where problem data contain some uncertainties and consequently SVIP

models are proposed to reflect the uncertainties. For instances, Watling (2006) uses an SVIP model to investigate user equilibrium problems in transportation where uncertainty arises from road conditions and travel times. Similar models can also be used to describe transmission of electricity in a network where generators and retailers are located at spatially separated nodes of the network and uncertainty arises from nodal demands, see Jiang and Xu (2008, Example 6.4). Application of SVIP models can also be found in signal transmission in a wireless network (Ngo and Krishnamurthy, 2007), inventory or pricing competition among several firms that provide substitutable goods or services (Chen *et al.*, 2004; Mahajan and Ryzin, 2001) and stochastic dynamic games (Basar and Olsder, 1999; Filar and Vrieze, 1997).

In this paper, we are concerned with numerical solution of SVIP. The key issue is to deal with the expected value of $F(x, \xi)$. If we are able to obtain a closed form of $\mathbb{E}[F(x, \xi)]$, then SVIP (1.1) becomes a deterministic VIP and existing numerical methods for the latter (Facchinei and Pang, 2003) can be applied directly to it. However, in practice, obtaining a closed form of $\mathbb{E}[f(x, \xi)]$ or computing the value of it numerically is often difficult either due to the unavailability of the distribution of ξ or because it involves multiple integration. Instead, it is often possible to obtain a sample of the random vector ξ either from past data or from computer simulation. Consequently one may consider an approximate solution to (1.1) based on sampling.

Let $\xi^1, \ldots, \xi^N$ be a sampling of $\xi(\omega)$. A well known approach based on the sampling is so-called Sample Average Approximation (SAA) method, that is, using sample average value of $F(x, \xi)$ to approximate its expected value because the classical law of large number for random functions ensures that $\hat{F}^N(x)$ converges with probability 1 to $\mathbb{E}[F(x, \xi)]$ (Rubinstein and Shapiro, 1993, Section 2.6) when the sampling is independent and identically distributed (iid for short). Specifically, we can write down the SAA of our SVIP (1.1) as follows: find $x \in \mathcal{K}$ such that

$$(y - x)^T \hat{F}^N(x) \geq 0, \quad \forall\, y \in \mathcal{K}, \tag{1.2}$$

where

$$\hat{F}^N(x) := \frac{1}{N} \sum_{j=1}^{N} F(x, \xi^j).$$

We call problem (1.2) the *SAA problem* and (1.1) the *true problem*. Shapiro (2003, Section 7) first discussed the above described sample average approximation approach for SVIP under the framework of stochastic generalized equations. He carried out comprehensive analysis including the existence and convergence of solutions to (1.2) as N increases. The topic of stochastic generalized equations and its asymptotic approximation can be traced down to King and Rockafellar's early work (King and Rockafellar, 1992, 1993) which is partly motivated by the asymptotic analysis of statistical estimators in stochastic programming. Of course, King and Rockafellar's focus is slightly different from Shapiro's in that the latter looks

into the asymptotic behavior of the approximate solutions using the sensitivity analysis of parametric generalized equations where the sample averaged functions are treated as a parameter (in a functional space). On a similar but slightly broader track, Gürkan *et al.* (1996, 1999) proposed simulation based sample-path optimization (SPO) approach for solving SVIP (1.1). SPO is proposed by Robinson (1996) and essentially coincides with SAA although SPO could be slightly broader.

More recently, Jiang and Xu (2008) proposed a stochastic approximation (SA) method for solving SVIP (1.1). The method is an iterative scheme where at each iterate a correction is made and the correction is obtained by sampling or other stochastic approximation. SA is well known in stochastic optimization but the application of SA to SVIP is new. As far as sampling is concerned, the main difference between SA and SAA is that SA usually requires a small sample size at each iterate (typically with $N = 1$) and hence converges slowly, while SAA requires a larger sample size and a new SAA problem has to be solved when sampling is updated.

In this paper, we follow Shapiro's approach to revisit the issues concerning the existence and convergence of solutions of the SAA problem (1.2). We complement Shapiro's results on the issues essentially on two-fold: (a) we use the concept of subinvertibility of a set-valued mapping due to King and Rockafellar (1992) rather than Robinson's strong regularity as in Shapiro (2003) to derive weaker conditions for the existence of a solution to SAA problem (1.2); we establish the exponential rate of convergence, instead of almost sure convergence as in Shapiro (2003), of solution of SAA problem (1.2) to its true counterpart as sample size increases under general sampling rather than iid sampling. We also apply the results to analyze the convergence of SAA method for a class of stochastic nonlinear complementarity problems and stochastic programs with stochastic constraints.

A few words about notation. Throughout this paper, $x^T y$ denotes the scalar products of two vectors x and y, $\|\cdot\|$ denotes the Euclidean norm of a vector and a compact set of vectors. $d(x, D) := \inf_{x' \in D} \|x - x'\|$ denotes the distance from point x to set D. For two sets D_1 and D_2, $\mathbb{D}(D_1, D_2) := \sup_{x \in D_1} d(x, D_2)$ denotes the deviation from set D_1 to set D_2 (which is also known as excess of D_1 over D_2), and $\mathbb{H}(D_1, D_2)$ denotes the Hausdorff distance between the two sets, that is, $\mathbb{H}(D_1, D_2) := \max(\mathbb{D}(D_1, D_2), \mathbb{D}(D_1, D_2))$.

2. Convergence Analysis of SAA Problem

In this section, we discuss the convergence of (1.2) to (1.1) as N increases. Specifically we investigate two issues: (a) the existence of a solution to the SAA problem (1.2), (b) the convergence of solution of (1.2) to its true counterpart as $N \to \infty$. Observe that both (1.1) and (1.2) can be written as generalized equations:

$$0 \in \hat{F}(x) + \mathcal{N}_{\mathcal{K}}(x) \tag{2.1}$$

and

$$0 \in \hat{F}^N(x) + \mathcal{N}_{\mathcal{K}}(x) \tag{2.2}$$

where $\hat{F}(x) = \mathbb{E}[F(x,\xi)]$, $\hat{F}^N(x) = \frac{1}{N}\sum_{i=1}^N F(x,\xi^i)$, and $\mathcal{N}_{\mathcal{K}}(x)$ denotes the normal cone (the usual one of convex analysis) to $\mathcal{K}$ at point x. Consequently, we may investigate the convergence of solutions of (2.2) to the set of solutions of (2.1).

2.1. *Existence of a solution*

It is not unusual in asymptotic analysis of SAA method for stochastic programming that existence of a solution to the sample average approximate problem is ignored by simply assuming the problem has a solution. This is not because the issue of existence is not important but because it is often difficult to derive sensible results. Moreover, when the feasible set is compact, then the existence often becomes trivial. In this paper, we investigate this issue as it is discussed in the literature. Let us start with a review of Shapiro's established results on this issue. The following concept of strong regularity is due to Robinson (1980) and considered by Shapiro (2003, Section 7) for the purpose of stochastic generalized equations.

Definition 2.1. Let $\hat{F}(x)$ be defined as in (2.1) and $\hat{F}(x)$ be continuously differentiable. Let x^* be a solution of the true problem (2.1). x^* is said to be *strongly regular*, if there exist neighborhoods U_1 and U_2 of $0 \in \mathbb{R}^n$ and x^* respectively such that for $\delta \in U_1$, the following generalized equation

$$\delta \in \hat{F}(x^*) + \nabla \hat{F}(x^*)(x - x^*) + \mathcal{N}_{\mathcal{K}}(x)$$

has a unique solution in U_2, denoted by $x(\delta)$, and $x(\cdot)$ is Lipschitz continuous on U_1.

Under the strong regularity condition, Shapiro established the following existence and convergence result for the solutions of (2.2).

Proposition 2.1. (Shapiro, 2003, Proposition 21) *Let x^* be a strongly regular solution to the true problem (2.1), and suppose that both $\hat{F}(x)$ and $\hat{F}^N(x)$ are continuously differentiable in a neighborhood of x^* and w.p.1 $\hat{F}^N(x)$ converges to $\hat{F}(x)$ uniformly in a neighborhood of x^*. Then w.p.1 for N large enough the SAA problem (2.2) possesses a unique solution x^N in a neighborhood of x^* and $x^N \to x^*$ w.p.1 as $N \to \infty$.*

In what follows we revisit the existence issue from a slightly different angle: instead of using Robinson's strong regularity, we apply King and Rockafellar's sensitivity analysis results on generalized equations to derive the existence of solution to (2.2). Two reasons underly our discussion: one is that the function $\hat{F}(x)$ is continuous but not necessarily continuously differentiable in some practical cases; the other is that if we just want existence of a solution to (2.2) without further requiring the uniqueness and Lipschitz continuity of the solution, then we might establish the existence result under a slightly weaker condition than those imposed in Proposition 2.1. The following concept of subinvertibility of is due to King and Rockafellar (1992).

Definition 2.2. Let $\Phi(x) := \mathbb{E}[F(x,\xi)] + \mathcal{N}_{\mathcal{K}}(x)$ and $x^* \in \mathbb{R}^n$. Φ is said to be *subinvertible* at $(x^*, 0)$, if one has $0 \in \Phi(x^*)$ and there exist a compact convex neighborhood U of x^* in $\mathbb{R}^n$, a positive constant $\epsilon > 0$, and a nonempty convex-valued mapping $G : \epsilon B \to U \subset \mathbb{R}^n$ such that the graph of $G, gphG$, is closed, the point x^* belongs $G(0)$, and $G(y)$ is contained in $\Phi^{-1}(y)$ for all $y \in \epsilon B$.

An instance of Φ being subinvertible at $(x^*, 0)$ is that there exists a continuous selection $x(y)$ of Φ^{-1} on a compact neighborhood of 0 such that $x(0) = x^*$.

King and Rockafellar (1992) presented intensive discussions about the sufficient conditions for subinvertibility of a set-valued mapping in terms of contingent derivative of the set-valued mapping. In this context, their results indicate that: (a) if $\Phi(x)$ is maximal monotone (that is, monotone and its graph is not contained properly by any other monotone set-valued mapping), then single valuedness of the contingent derivative of Φ^{-1} at 0 is enough to imply subinvertibility; (b) when $\mathcal{N}_{\mathcal{K}}(\cdot)$ is polyhedral, B-differentiability of $\hat{F}(x)$ and Robinson's strong regularity (in terms of B-derivative) imply the subinvertibility of Φ, see King and Rockafellar (1992, Sections 5–6) for details. The main interest in employing the concept of subinvertibility here is to extend Proposition 2.1 so that it covers the stochastic generalized equations (2.1) where $\hat{F}(x)$ is not necessarily continuously differentiable.

Proposition 2.2. *Let* $\Phi(x) := \mathbb{E}[F(x,\xi)] + \mathcal{N}_{\mathcal{K}}(x)$ *and* x^* *be a solution of* $0 \in \Phi(x)$. *Assume*: (*a*) $F(x,\xi)$ *is Lipschitz continuous w.r.t.* x *and its Lipschitz modulus is bounded by* $\kappa(\xi) > 0$, *where* $\mathbb{E}[\kappa(\xi)] < \infty$; (*b*) $\Phi(x)$ *is subinvertible at* $(x^*, 0)$ *in* $\mathbb{R}^n$. *Then there exist a compact convex neighborhood* $U \subset \mathbb{R}^n$ *of* x^* *such that w.p.1* (1.2) *has at least one solution in the neighborhood* U *for* N *sufficiently large.*

Proof. Under the subinvertibility condition (b), it follows from King and Rockafellar's (1992, Proposition 3.1) that there exist a compact neighborhood U of x^* and a small positive number ϵ such that when

$$\sup_{x \in U} \left\| \mathbb{E}[F(x,\xi)] - \frac{1}{N} \sum_{i=1}^{N} F(x,\xi^i) \right\| \leq \epsilon \tag{2.3}$$

the SAA problem (2.2) has a solution x^N in U. On the other hand, for the given ϵ and under condition (a), the classical law of large numbers (see, e.g., Rubinstein and Shapiro, 1993, Lemma A) ensures that (2.3) holds w.p.1 for N sufficiently large. The conclusion follows. □

2.2. *Exponential rate of convergence*

We now move on to discuss the convergence of SAA problem (1.2) to the true problem (1.1) as sample increases, that is, the convergence of x^N (a solution to (1.2)) to its true counterpart as $N \to \infty$. Shapiro (2003, Section 7) established almost sure convergence of x^N, that is, w.p.1, $x^N \to x^*$ as $N \to \infty$, where x^* is a solution to the true problem (1.1). See Proposition 2.1 and Shapiro (2003,

Theorem 22) for details. Our focus here is on the rate of convergence, that is, how fast x^N converges to x^*. From computational perspective, this is an important issue because it concerns the efficiency of the SAA method. We use the classical large deviation theorem (Dembo and Zeitouni, 1998) and some recently established sensitivity results for generalized equations (Xu, 2008) to establish the exponential convergence of x^N to x^* in hope to complement the existing results by Shapiro and others on this topic.

For $i = 1, \ldots, n$, let $F_i(x, \xi)$ denote the ith component of $F(x, \xi)$, and

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

denote the moment generating function of the random variable $F_i(x, \xi) - \mathbb{E}[F_i(x, \xi)]$, let

$$M_{F_i}^N(t) := \mathbb{E}\{e^{t(\frac{1}{N}\sum_{j=1}^N F_i(x,\xi^j)-\mathbb{E}[F_i(x,\xi)])}\}.$$

We make the following assumption.

Assumption 2.1. Let $\mathcal{X}$ be a compact subset of $\mathcal{K}$. For $i = 1, \ldots, n$, the limits $M_{F_i}(t) := \lim_{N\to\infty} M_{F_i}^N(t)$, exist for every $x \in \mathcal{X}$ and $t \in \mathbb{R}$.

In the case when $\xi^1, \ldots, \xi^N$ is an iid sampling, Assumption 2.1 holds so long as $M_{F_i}(t) < 0$ for t close to zero, see Dembo and Zeitouni (1998, Section 2.3). In practice, sampling may not necessarily be iid, yet it satisfies Assumption 2.1 particularly when the sampling is generated by quasi-Monte Carlo method, see detailed discussions about this issue by Homem-de-Mello (2008). Under these circumstances, one may use Gärtner-Ellis' large deviation theorem (Dembo and Zeitouni, 1998, Theorem 2.3.6) instead of Cramér's large deviation theorem to establish the exponential convergence of the sample averages. The following pointwise exponential convergence indeed follows from Assumption 2.1.

Proposition 2.3. *Let Assumption 2.1 hold and $\mathcal{X}$ be a compact subset of $\mathcal{K}$. Then for $i = 1, \ldots, n$, every $x \in \mathcal{X}$ and small positive number $\epsilon > 0$,*

$$\text{Prob}\left\{\left|\frac{1}{N}\sum_{j=1}^N F_i(x,\xi^i) - \mathbb{E}[F_i(x,\xi)]\right| \geq \epsilon\right\} \leq e^{-NI_x^i(-\epsilon)} + e^{-NI_x^i(\epsilon)},$$

for N sufficiently large, where

$$I_x^i(z) := \sup_{t\in\mathbb{R}}\{zt - \log M_x^i(t)\}$$

and both $I_x^i(\epsilon)$ and $I_x^i(-\epsilon)$ are positive.

The above pointwise convergence is inadequate for us to derive the exponential of x^N, we state it to pave the way for the uniform exponential convergence of the components of $\hat{F}^N(x)$ over a compact set. We need the following additional assumptions for the latter.

Assumption 2.2. Let $\mathcal{X}$ be a compact subset of $\mathcal{K}$. For $i = 1, \ldots, n$,

(a) for every $x \in \mathcal{X}$, the moment generating function $M_x^i(t)$ of $F_i(x, \xi) - \hat{F}_i(x)$ is finite valued for all t in a neighborhood of zero;
(b) there exist a (measurable) function $\kappa_i : \Xi \to \mathbb{R}_+$ and constant $\gamma_i > 0$ such that

$$|F_i(x', \xi) - F_i(x, \xi)| \leq \kappa_i(\xi)\|x' - x\|^{\gamma_i} \tag{2.4}$$

for all $\xi \in \Xi$ and all $x', x \in \mathcal{X}$;
(c) the moment generating function $M_{\kappa_i}(t)$ of $\kappa_i(\xi)$ is finite valued for all t in a neighborhood of zero.

Under Assumption 2.2 and Proposition 2.3, we are able to derive the uniform exponential convergence by virtue of a recently established result (Xu, 2008, Theorem 3.1). We omit details of proof.

Proposition 2.4. *Let Assumptions 2.1 and 2.2 hold and $\mathcal{X}$ be a compact subset of $\mathcal{K}$. Then for any $\epsilon > 0$, there exist positive constants $c(\epsilon)$ and $\beta(\epsilon)$, independent of N, such that*

$$\text{Prob}\left\{\sup_{x \in \mathcal{X}} \|\hat{F}^N(x) - \hat{F}(x)\| \geq \epsilon\right\} \leq c(\epsilon)e^{-N\beta(\epsilon)}. \tag{2.5}$$

We need to translate the uniform exponential convergence of F^N to $\hat{F}$ into the exponential convergence of x^N to the solution set of true problem (2.1). To this end, we need some sensitivity analysis of generalized equations discussed in Xu (2008).

Consider the following generalized equation

$$0 \in G(x) + \mathcal{N}_{\mathcal{C}}(x), \tag{2.6}$$

where $G : \mathcal{C} \to \mathbb{R}^n$ is a vector valued function, $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^n$. Let $\tilde{G}(x)$ be a perturbation of $G(x)$ and we consider the perturbed equation

$$0 \in \tilde{G}(x) + \mathcal{N}_{\mathcal{C}}(x). \tag{2.7}$$

The following lemma states that when $\sup_{x \in \mathcal{C}} \|\tilde{G}(x) - G(x)\|$ is sufficiently small, the solution set of (2.7) is close to the solution set of (2.6) provided that both sets are nonempty.

Lemma 2.1. *Let S^* denote the set of solutions to (2.6) and $\tilde{S}$ the set of solutions to (2.7). Assume that both S^* and $\tilde{S}$ are nonempty. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that if $\sup_{x \in \mathcal{C}} \|\tilde{G}(x) - G(x)\| < \delta$, then $\mathbb{D}(\tilde{S}, S^*) < \epsilon$.*

Proof. The result is a corollary of Xu (2008, Lemma 4.2). Here we include a proof for completeness. Let

$$R(\epsilon) := \inf_{x : d(x, S^*) \geq \epsilon} d(0, G(x) + \mathcal{N}_{\mathcal{C}}(x)).$$

Then $R(\epsilon) > 0$. Let $\delta = R(\epsilon)/2$ and $\sup_{x\in\mathcal{C}} \|\tilde{G}(x) - G(x)\| < \delta$. For any point $x \in \mathcal{C}$ with $d(x, S^*) > \epsilon$,

$$\begin{aligned} d(0, \tilde{G}(x) + \mathcal{N}_{\mathcal{C}}(x)) &\geq d(0, G(x) + \mathcal{N}_{\mathcal{C}}(x)) - \mathbb{D}(\tilde{G}(x) + \mathcal{N}_{\mathcal{C}}(x), G(x) + \mathcal{N}_{\mathcal{C}}(x)) \\ &\geq d(0, G(x) + \mathcal{N}_{\mathcal{C}}(x)) - \|\tilde{G}(x) - G(x))\| \\ &> d(0, G(x) + \mathcal{N}_{\mathcal{C}}(x)) - \delta \geq 2\delta - \delta > 0. \end{aligned}$$

(Here we use the following properties of the excess function $\mathbb{D}$: for sets A, B, C, $\mathbb{D}(A, B) \leq \mathbb{D}(A, C) + \mathbb{D}(C, B)$ and $\mathbb{D}(A + C, B + C) \leq \mathbb{D}(A, B)$. This is evident from the fact that $\mathbb{D}(A, B) = \inf_{t>0}\{t : A \subset B + t\mathcal{B}\}$ where $\mathcal{B}$ denotes the unit ball. Note also that when A reduces to a singleton $\{a\}$, $\mathbb{D}(A, B) = d(a, B)$). This shows that $x \notin \tilde{S}$. Hence for any $y^* \in \tilde{S}$, $d(y^*, S^*) \leq \epsilon$, which implies $\mathbb{D}(\tilde{S}, S^*) < \epsilon$. □

Theorem 2.1. *Let x^N be a solution to the SAA problem (1.2) and X^* the set of solutions to the true problem (2.1). Assume: (a) w.p.1 the sequence $\{x^N\}$ is located in a compact subset $\mathcal{X}$ of $\mathcal{K}$, (b) Assumptions 2.1 and 2.2 hold. Then for every $\epsilon > 0$, there exist positive constants $\hat{c}(\epsilon)$ and $\hat{\beta}(\epsilon)$, independent of N, such that*

$$\text{Prob}\{d(x^N, X^*) \geq \epsilon\} \leq \hat{c}(\epsilon)e^{-N\hat{\beta}(\epsilon)} \tag{2.8}$$

for N sufficiently large.

Proof. Let $\epsilon > 0$ be any small positive number. By Lemma 2.1, there exists a $\delta(\epsilon) > 0$ such that if $\sup_{x\in\mathcal{X}} \|\hat{F}^N(x) - \hat{F}(x)\| < \delta(\epsilon)$ then $d(x^N, X^*) < \epsilon$. On the other hand, under Assumptions 2.1 and 2.2, we have from Proposition 2.4 that for the $\delta(\epsilon) > 0$ there exist positive constants $c(\delta(\epsilon))$ and $\beta(\delta(\epsilon))$ (for the simplicity of notation we write them as $\hat{c}(\epsilon)$ and $\hat{\beta}(\epsilon)$) independent of N, such that

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}} \|\hat{F}^N(x) - \hat{F}(x)\| \geq \delta_{(\epsilon)}\right\} \leq \hat{c}(\epsilon)e^{-N\hat{\beta}(\epsilon)}$$

for N sufficiently large. Consequently we have

$$\text{Prob}(d(x^N, X^*) \geq \epsilon) \leq \text{Prob}\left\{\sup_{x\in\mathcal{X}} \|\hat{F}^N(x) - \hat{F}(x)\| \geq \delta_{(\epsilon)}\right\} \leq \hat{c}(\epsilon)e^{-N\hat{\beta}(\epsilon)}.$$

The proof is complete. □

2.3. *Metric regularity*

In Theorem 2.1, the dependence of δ on ϵ is implicit in the sense that we don't have an explicit quantitative relationship between the two constants. This makes the convergence results weak because if for a given ϵ, δ is very small, then we will need a large sample size for x^N to converge to a solution of (2.1). The issue has nothing to do with the SAA method but it has to do with the sensitivity of the true

problem at a solution point. For this purpose, we resort to the concept of metric regularity.

Definition 2.3. Let $\mathcal{X}$ be a closed set of $\mathbb{R}^m$, $\Phi : \mathcal{X} \to 2^{\mathbb{R}^m}$ be a closed set valued mapping. For $\bar{x} \in \mathcal{X}$ and $\bar{y} \in \Phi(\bar{x})$, Φ is said to be *metrically regular* at $\bar{x}$ for $\bar{y}$ if there exists a constant $\alpha > 0$ such that

$$d(x, \Phi^{-1}(y)) \leq \alpha d(y, \Phi(x)) \text{ for all } (x, y) \text{ close to} (\bar{x}, \bar{y}).$$

Here the inverse mapping Φ^{-1} is defined as $\Phi^{-1}(y) = \{x \in \mathcal{X} : y \in \Phi(x)\}$ and the minimal constant $\alpha < \infty$ which makes the above inequality hold is called *regularity modulus*.

Metric regularity is a generalization of derivative nonsingularity of a function to that of a set-valued mapping (Robinson, 1976). The property is equivalent to nonsingularity of the coderivative of Φ at $\bar{x}$ for $\bar{y}$ and to Aubin's property of Φ^{-1} (Pseudo Lipschitz continuity). For a comprehensive discussion of the history and recent development of the notion, see Dontchev *et al.* (2004), Rockafellar and Wets (1998) and references therein.

Using the notion of metric regularity, we can analyze the sensitivity of generalized equations.

Lemma 2.2. *Let $\Phi, \Psi : \mathcal{X} \to 2^{\mathbb{R}^m}$ be two set valued mappings. Let $\bar{x} \in \mathcal{X}$ and $0 \in \Phi(\bar{x})$. Let $0 \in \Psi(x)$ with x being close to $\bar{x}$. Suppose that Φ is metrically regular at $\bar{x}$ for 0. Then*

$$d(x, \Phi^{-1}(0)) \leq \alpha \mathbb{D}(\Psi(x), \Phi(x)), \tag{2.9}$$

where α is the regularity modulus of Φ at $\bar{x}$ for 0.

Proof. Since Φ is metrically regular at $\bar{x}$ for 0, there exists a constant $\alpha > 0$ such that $d(x, \Phi^{-1}(0)) \leq \alpha d(0, \Phi(x))$. Since $0 \in \Psi(x)$, then $d(0, \Phi(x)) \leq \mathbb{D}(\Psi(x), \Phi(x))$. (2.9) follows. The proof is complete. □

The above result can be explained as follows. Suppose we solve a generalized equation $0 \in \Phi(x)$ by solving an approximate equation $0 \in \Psi(x)$ where Ψ is an approximation of Φ, and obtain a solution x for the approximate equation. Suppose also that x is close to a true solution $\bar{x} \in \Phi^{-1}(0)$ and Φ is metrically regular at $\bar{x}$, then the distance between x and $\Phi^{-1}(0)$ is bounded by the distance of the set-valued mapping from $\Psi(x)$ to $\Phi(x)$. This type of error bound is numerically useful because we are guaranteed local stability against the data perturbation of the underlying functions.

Theorem 2.2. *Assume the setting and conditions of Theorem 2.1. Let $x^N \to x^* \in X^*$ w.p.1 and $\Phi(x)$ be metric regular at x^* for 0 with regularity modulus α. Then for any small $\epsilon > 0$, there exist positive constants $\hat{c}(\epsilon) := c(\epsilon/\alpha)$ and $\hat{\beta}(\epsilon) =$*

$\beta(\epsilon/\alpha)$, independent of N, such that (2.8) holds, where $c(\epsilon)$ and $\beta(\epsilon)$ are given in Proposition 2.4.

Proof. Under the metric regularity, it follows from Lemma 2.2 that for x^N close x^*,

$$d(x^N, X^*) \leq \alpha \|\hat{F}^N(x^N) - \hat{F}(x^N)\| \leq \alpha \sup_{x\in\mathcal{X}} \|\hat{F}^N(x) - \hat{F}(x)\|. \tag{2.10}$$

Let $B(x^*)$ denote a compact neighborhood of x^* relative to $\mathcal{K}$ such that for $x^N \in B(x^*)$, (2.10) holds. Then by applying Proposition 2.4 on $B(x^*)$, there exist positive constants $c(\epsilon)$ and $\beta(\epsilon)$, independent of N, such that

$$\text{Prob}(d(x^N, X^*) \geq \epsilon) \leq \text{Prob}\left\{\sup_{x\in\mathcal{B}(x^*)} \|\hat{F}^N(x) - \hat{F}(x)\| \geq \epsilon/\alpha\right\} \leq c(\epsilon/\alpha)e^{-N\beta(\epsilon/\alpha)}$$

for N sufficiently large. The conclusion follows. □

In the case when $F(x,\xi)$ is strongly monotone w.r.t. x on $\mathcal{K}$ with modulus $\sigma(\xi) > 0$, that is,

$$(F(x,\xi) - F(y,\xi))^T(x-y) \geq \sigma(\xi)\|x-y\|^2$$

for all $x, y \in \mathcal{K}$, where $\mathbb{E}[\sigma(\xi)] \in (0,\infty)$, the subinvertibility holds trivially. Consequently we have the following existence and convergence results.

Corollary 2.1. *Let Assumptions 2.1 and 2.2 hold and $F(\cdot,\xi)$ be strongly monotone on $\mathcal{K}$ for each ξ. Then the sample average VIP (1.2) has a unique solution x^N and w.p.1 approaching 1 exponentially fast with the increase of sample size N, sequence $\{x^N\}$ converges to the unique solution x^* of SVIP (1.1).*

3. Stochastic Complementarity Problems

When $\mathcal{K} = \mathbb{R}^n_+$, SVIP (1.1) reduces to a stochastic nonlinear complementarity problem (SNCP):

$$\hat{F}(x) \geq 0, \quad x \geq 0, \quad \hat{F}(x)^T x = 0. \tag{3.1}$$

Likewise, the SAA problem (1.2) reduces to an NCP

$$\hat{F}^N(x) \geq 0, \quad x \geq 0, \quad (\hat{F}^N(x))^T x = 0, \tag{3.2}$$

where $\hat{F}(x) = \mathbb{E}[F(x,\xi)]$ and $\hat{F}^N(x) = \frac{1}{N}\sum_{j=1}^N F(x,\xi^j)$. It is well known that a nonlinear complementarity problem can be reformulated as a system of nonsmooth equations using some elementary functions (also called NCP functions) such as the min-function $\min(a,b)$ and Fischer-Burmeister function $\phi : \mathbb{R}^2 \to \mathbb{R}$, which is defined as

$$\phi(a,b) = \sqrt{a^2+b^2} - a - b.$$

Using the FB-function, the SNCP (3.1) can be reformulated as a system of stochastic equations

$$H(x) \equiv \begin{pmatrix} \phi(x_1, \mathbb{E}[F_1(x,\xi)]) \\ \vdots \\ \phi(x_n, \mathbb{E}[F_n(x,\xi)]) \end{pmatrix} = 0 \tag{3.3}$$

in the sense that the solution set of (3.1) coincides with that of (3.3). One of the main benefits in using the Fischer-Burmeister function is that $H(x)$ is semismooth everywhere and continuously differentiable of any order at any point except the origin, and it is globally Lipschitz continuous. Similarly, the SAA problem (3.2) can be reformulated as

$$\hat{H}^N(x) \equiv \begin{pmatrix} \phi(x_1, \hat{F}_1^N(x)]) \\ \vdots \\ \phi(x_n, \hat{F}_n^N(x)]) \end{pmatrix} = 0 \tag{3.4}$$

where $\hat{F}_i^N(x) = \frac{1}{N}\sum_{j=1}^N F_i(x,\xi^j)$, for $i = 1,\ldots,n$.

Theorem 3.1. *Let x^* be a solution to the true SNCP problem (3.1). Let Assumptions 2.1 and 2.2 hold. Then the following statements apply to (3.1) and (3.2).*

(i) *If $H(x) = y$ has a unique solution in a compact neighborhood U of x^* relative to $\mathbb{R}^n_+$ for y close to 0, and $F(x,\xi)$ is Lipschitz continuous w.r.t. x and its Lipschitz modulus is bounded by $\kappa(\xi) > 0$, where $\mathbb{E}[\kappa(\xi)] < \infty$, then there exist a compact convex neighborhood $U \subset \mathbb{R}^n$ of x^* such that with probability approaching one exponentially fast against the increase of N, (3.4) has a solution x^N in the neighborhood U for N sufficiently large.*

(b) *If, in addition, H^{-1} is Lipschitz continuous near x^*, then for every $\epsilon > 0$, there exist positive constants $\hat{c}(\epsilon)$ and $\hat{\beta}(\epsilon)$, independent of N, such that*

$$\text{Prob}\{\|x^N - x^*\| \geq \epsilon\} \leq \hat{c}(\epsilon)e^{-N\hat{\beta}(\epsilon)} \tag{3.5}$$

for N sufficiently large.

Proof. Part (i). The existence and uniqueness of an implicit function, denoted by $H^{-1}(y)$, for $y \in U$ implies the invertibility (and hence subinvertibility) of H in the neighborhood. Therefore when

$$\sup_{x\in U} \|H^N(x) - H(x)\| \leq \epsilon \tag{3.6}$$

for some sufficiently small positive number ϵ, the equation (3.4) has a solution x^N in the neighborhood U. On the other hand, under Assumptions 2.1 and 2.2, it follows from Proposition 2.4 that $\sup_{x\in U} \|\hat{F}^N(x) - F(x)\| \to 0$ at an exponential rate. Since the FB-function is Lipschitz continuous , this implies $\sup_{x\in U} \|\hat{H}^N(x) - H(x)\| \to 0$ at an exponential rate as $N \to \infty$. The conclusion follows.

Part (ii). There exists $\alpha > 0$ such that

$$\|x^N - x^*\| \leq \alpha \sup_{x \in U} \|H^N(x) - H(x)\|.$$

The rest follows from part (i). □

Note that SNCP model (3.1) is different from the stochastic complementarity models recently considered in Chen and Fukushima (2005). The latter finds a deterministic solution to NCPs parameterized by all possible realizations of a random variate. This results in a deterministic overdetermined system of NCPs which usually do not have a solution. Chen and Fukushima (2005) used NCP functions to reformulate NCPs into systems of nonsmooth equations and consider least-squared minimization of the residual of the reformulated equations. Consequently it can be proved that solutions for such a reformulated problem exist under suitable conditions, see a recent survey by Fukushima and Lin (2009) on this model and other related models.

4. Application to Stochastic Programs with Stochastic Constraints

We now consider the following stochastic programs with stochastic constraints

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \mathbb{E}[f_0(x, \xi)] \\ \text{s.t.} \ & \mathbb{E}[f_i(x, \xi)] \leq 0, \quad i = 1, \ldots, s, \\ & \mathbb{E}[f_i(x, \xi)] = 0, \quad i = s+1, \ldots, m, \\ & x \in X \subset \mathbb{R}^n \end{aligned} \tag{4.1}$$

where $f_i : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$, $i = 0, \ldots, m$, is continuously differentiable. $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$, and X is a closed subset of $\mathbb{R}^n$. The stochastic program model covers many interesting problems in stochastic programming.

Let us consider sample average approximation of (4.1) as follows:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \hat{f}_0^{\,N}(x) \\ \text{s.t.} \ & \hat{f}_i^{\,N}(x) \leq 0, \quad i = 1, \ldots, s, \\ & \hat{f}_i^{\,N}(x) = 0, \quad i = s+1, \ldots, m, \\ & x \in X, \end{aligned} \tag{4.2}$$

where

$$\hat{f}_i^{\,N}(x) := \frac{1}{N} \sum_{j=1}^{N} f_i(x, \xi^j), \quad i = 0, \ldots, m$$

and $\xi^1, \ldots, \xi^N$ is a sampling of random vector ξ.

Shapiro (1991) proposed a general approximating scheme which includes sample average approximation for solving (4.2). He investigated the asymptotic behavior of the optimal values and optimal solutions obtained from solving the approximating problems and he did so by using a parametric programming approach where the approximating functionals are treated as parameters defined on a Banach space, and then deriving the asymptotics of the optimal value and optimal solution estimators by an extended delta method.

Our focus here is on the asymptotic behavior of the stationary points of the SAA problem (4.2): assuming that SAA obtains a stationary point to problem (4.2), denoted by x^N, we investigate the convergence of x^N to its true counterpart, denoted by x^*, as sample size increases. There are three differences between our analysis and Shapiro's asymptotic analysis in Shapiro (1991): (a) we intend to estimate the rate of convergence of x^N to x^* rather than the asymptotic distribution of $x^N - x^*$; (b) instead of using sensitivity analysis of parametric programming, we use our established convergence results for SVIP (1.1); (c) the sampling is not necessarily iid, that is, our analysis covers the case if the sampling is generated by Quasi-Monte Carlo methods.

Let us start our analysis by describing the Karush-Kuhn-Tucker conditions of the true problem and the SAA problem. The following result is well known in deterministic optimization and it is also a corollary of Wets (1989, Proposition 5.1) when $X = \mathbb{R}^n$.

Proposition 4.1. *Consider the stochastic program* (4.1) *and suppose that for* $i = 1, \ldots, m$, *the function* $\mathbb{E}[f_i(x, \xi)]$, *are continuously differentiable and* $X = \mathbb{R}^n$. *If* x^* *is a local minimizer at which the Magsarian-Fromovitz constraint qualifications hold, that is,* $\mu = 0$ *is the only solution of the system*

$$0 = \mu_i \mathbb{E}[f_i(x^*, \xi)], \quad i = 1, \ldots, s, \tag{4.3}$$

$$0 \leq \mu_i, -\mathbb{E}[f_i(x^*, \xi)]; \quad i = 1, \ldots, s, \tag{4.4}$$

$$0 = \sum_{i=1}^{m} \mu_i \nabla \mathbb{E}[f_i(x^*, \xi)], \tag{4.5}$$

then there exists a multiplier μ^* *such that*

$$0 \in \nabla \mathbb{E}[f_0(x^*, \xi)] + \sum_{i=1}^{m} \mu_i^* \nabla \mathbb{E}[f_i(x^*, \xi)] + \mathcal{N}_X(x), \tag{4.6}$$

$$0 = \mu_i^* \mathbb{E}[f_i(x^*, \xi)], \quad i = 1, \ldots, s, \tag{4.7}$$

$$0 \leq \mu_i^*, -\mathbb{E}[f_i(x^*, \xi)], \quad i = 1, \ldots, s, \tag{4.8}$$

$$0 = \mathbb{E}[f_i(x^*, \xi)], \quad i = s + 1, \ldots, m. \tag{4.9}$$

Let

$$G(x,\mu) := \begin{pmatrix} \nabla \mathbb{E}[f_0(x,\xi)] + \sum_{i=1}^{m} \mu_i \nabla \mathbb{E}[f_i(x,\xi)] \\ \min(\mu_1, \mathbb{E}[f_1(x,\xi)]) \\ \vdots \\ \min(\mu_s, \mathbb{E}[f_s(x,\xi)]) \\ \mathbb{E}[f_{s+1}(x,\xi)] \\ \vdots \\ \mathbb{E}[f_m(x,\xi)] \end{pmatrix}$$

and $\Gamma(x,\mu) := \mathcal{N}_X(x) \times \{0\}$, where $0 \in \mathbb{R}^m$. Then we can rewrite the KKT conditions (4.6)–(4.9) as

$$0 \in G(x,\mu) + \Gamma(x,\mu). \tag{4.10}$$

Similarly, we can write down the KKT conditions of the SAA problem (4.2) as follows:

$$0 \in \hat{G}^N(x,\mu) + \Gamma(x,\mu), \tag{4.11}$$

where

$$\hat{G}^N(x,\mu) := \begin{pmatrix} \frac{1}{N}\sum_{j=1}^{N} \nabla_x f_0(x,\xi^j) + \sum_{i=1}^{m} \mu_i \frac{1}{N}\sum_{j=1}^{N} \nabla_x f_i(x,\xi^j) \\ \min(\mu_1, \frac{1}{N}\sum_{j=1}^{N} f_1(x,\xi^j)) \\ \vdots \\ \min(\mu_s, \frac{1}{N}\sum_{j=1}^{N} f_s(x,\xi^j)) \\ \frac{1}{N}\sum_{j=1}^{N} f_{s+1}(x,\xi^j) \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N} f_m(x,\xi^j) \end{pmatrix}.$$

Let $H : \mathbb{R}^n \to \mathbb{R}^n$ be a locally Lipschitz continuous function. Recall that the Clarke generalized Jacobian (Clarke, 1983) of H at $x \in \mathbb{R}^n$ is defined as

$$\partial H(x) := \operatorname{conv}\left\{ \lim_{\substack{y \in D_H \\ y \to x}} \Delta H(y) \right\},$$

where D_H denotes the set of points at which H is Frechét differentiable, $\nabla H(y)$ denotes the usual Jacobian of H, "conv" denotes the convex hull of a set. $\partial H(x)$ is said to be nonsingular if every matrix in $\partial H(x)$ is nonsingular.

Theorem 4.1. *Let $x^* \in X^*$ be a KKT point of the true problem (4.1) and μ^* be the corresponding vector of Lagrange multipliers. Assume: (a) for $i = 0, \ldots, m, \nabla_x f_i(x, \xi)$ is Lipschitz w.r.t. x and its Lipschitz modulus is bounded by an integrable function $\kappa_i(\xi)$; (b) Assumptions 2.1 and 2.2 hold for $F_i(x, \xi) = f_i(x, \xi), i = 1, \ldots, m$ and $F_i(x, \xi) = \nabla_x f_i(x, \xi)$ for $i = 0, \ldots, m$; (c) the Clarke generalized Jacobian of $G(x, \mu)$ is nonsingular at (x^*, μ^*). Then the following statements apply to (4.10) and (4.11).*

(i) *There exists a compact convex neighborhood U of (x^*, μ^*) such that with probability approaching one exponentially fast against the increase of sample size N, (4.11) has a solution (x^N, μ^N) in the neighborhood U for N sufficiently large.*

(ii) *For every $\epsilon > 0$, there exist positive constants $\hat{c}(\epsilon)$ and $\hat{\beta}(\epsilon)$, independent of N, such that*

$$\text{Prob}\{\|x^N - x^*\| + \|\mu^N - \mu^*\| \geq \epsilon\} \leq \hat{c}(\epsilon) e^{-N\hat{\beta}(\epsilon)} \tag{4.12}$$

for N sufficiently large.

Proof. Part (i). Let $\Phi(x, \mu) = G(x, \mu) + \Gamma(x, \mu)$. Under condition (c), $\Phi(x, \mu)$ is subinvertible at $((x^*, \mu^*), 0)$, where $0 \in \mathbb{R}^{m+n}$. Therefore when

$$\sup_{x \in U} \|\hat{G}^N(x, \mu) - G(x, \mu)\| \leq \epsilon \tag{4.13}$$

for some sufficiently small positive number ϵ, the equation (4.11) has a solution (x^N, μ^N) in the neighborhood U. On the other hand, under Assumptions 2.1 and 2.2, it follows from Proposition 2.4 that

$$\sup_{x \in U} \left\| \frac{1}{N} \sum_{j=1}^{N} f_i(x, \xi^j) - \mathbb{E}[f_i(x, \xi)] \right\|$$

and

$$\sup_{x \in U} \left\| \frac{1}{N} \sum_{j=1}^{N} \nabla_x f_i(x, \xi^j) - \mathbb{E}[\nabla_x f_i(x, \xi)] \right\|$$

converge to 0 at an exponential rate, for $i = 0, \ldots, m$. Observe that under condition (a), $\mathbb{E}[\nabla_x f_i(x, \xi)] = \nabla \mathbb{E}[f_i(x, \xi)]$ and the min-function is Lipschitz continuous. Then

$$\sup_{x \in U} \|\hat{G}^N(x, \mu) - G(x, \mu)\| \to 0$$

at an exponential rate as $N \to \infty$. The conclusion follows.

Part (ii). Under condition (c), the set-valued mapping $G(x,\mu)+\Gamma(x,\mu)$ is metric regular at (x^*,μ^*). By Lemma 2.2, there exists $\alpha > 0$ such that

$$\|(x^N,\mu^N)-(x^*,\mu^*)\| \le \alpha \sup_{x\in U} \|\hat{G}^N(x,\mu)-G(x,\mu)\|.$$

The rest follows from part (i). □

Acknowledgments

The author would like to thank two referees for helpful comments and Professor Wenyu Sun for handling the review of this paper.

References

Basar, T and G Olsder (1999). *Dynamic Noncooperative Game Theory*. SIAM.

Chen, FY, H Yan and L Yao (2004). A newsvendor pricing game. *IEEE Transactions on Systems, Man & Cybernetics: Part A*, 34, 450–456.

Chen, X and M Fukushima (2005). Expected residual minimization method for stochastic linear complementarity problems. *Mathematics of Operational Research*, 30, 1022–1038.

Clarke, FH (1983). *Optimization and Nonsmooth Analysis*. New York: Wiley.

Dai, L, CH Chen and JR Birge (2000). Convergence properties of two stage stochastic programming. *Journal of Optimization Theory and Applications*, 106, 489–509.

Dembo, A and O Zeitouni (1998). *Large Deviations Techniques and Applications*. New York: Springer-Verlag.

Dontchev, AL, AS Lewis and RT Rockafellar (2004). The radius of metric regularity. *Transactions of the American Mathematics Society*, 355, 493–517.

Facchinei, F and J-S Pang (2003). *Finite-Dimensional Variational Inequalities and Complementarity Problems*. New York: Springer.

Filar, J and K Vrieze (1997). *Competitive Markov Decision Processes*. Springer-Verlag.

Fukushima, M and GH Lin, (2009). Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: A survey. Technical Report 2009-008, Department of Applied Mathematics and Physics, Kyoto University, April.

Gürkan, G and AY Özge and SM Robinson (1996). Sample path solution of stochastic variational inequalities, with applications to option pricing, In JM Charnes and DM Morrice and DT Brunner and JJ Swai (eds.), *Proceedings of the 1996 Winter Simulation Conference*, pp. 337–344.

Gürkan, G, AY Özge and SM Robinson (1999). Sample-path solution of stochastic variational inequalities. *Mathematical Programming*, 84, 313-333.

Homen-De-Mello, T (2008). On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling. *SIAM Journal on Optimization*, 19, 524–551.

Jiang, H and H Xu (2008). Stochastic Approximation Approaches to the Stochastic Variational Inequality Problem. *IEEE Transactions on Automatic Control*, 53, 1462–1475.

King, AJ and RT Rockafellar (1992). Sensitivity analysis for nonsmooth generalized equations, 55, 193–212.

King, AJ and RT Rockafellar (1993). Asymptotic theory for solutions in statistical estimation and stochastic programming. *Mathematics of Operations Research*, 18, 148–162.

Mahajan, S and G van Ryzin, Inventory competition under dynamic consumer choice. *Operations Research*, 49, 646–657.

Ngo, MH and VK Krishnamurthy (2007). Game theoretic cross-layer transmission policies in multipacket reception wireless networks. *IEEE Transactions on Signal Processing*, 55, 1911–1926.

Robinson, SM (1976). Regularity and stability for convex multivalued functions. *Mathematics of Operations Research*, 1, 131–143.

Robinson, SM (1980). Strongly regular genarlized equations. *Mathematics of Operations Research*, 5, 43–62.

Robinson, SM (1996). Analysis of sample-path optimization. *Mathematics of Operations Research*, 21, 513–528.

Rockafellar, RT and RJ-B Wets (1998). *Variational Analysis*. Berlin: Springer-Verlag.

Rubinstein, RY and A Shapiro (1983). *Discrete Events Systems*: *Sensitivity Analysis and Stochastic Optimization by the Score Function Methods*. New York: John Wiley and Sons.

Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operational Research*, 30, 169–186.

Shapiro, A. (2003). Monte Carlo sampling methods. In Rusczyński and Shapiro, (eds.) *Stochastic Programming*, Handbooks in OR & MS, Vol. 10. Amsterdam: North-Holland Publishing Company.

Shapiro, A and H Xu (2008). Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. *Optimization*, 57, 395–418.

Watling, D (2006). User equilibrium traffic network assignment with stochastic travel times and late arrival penalty. *European Journal of Operational Research*, 175, 1539–1556.

Wets, R (1989). *Stochastic Programming*. In Handbooks in OR & MS, Vol. 1, G. L. Nemhauser *et al.*, eds. pp. 573–629.

Xu, H (2008). Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. To appear in *Journal of Mathematical Analysis and Applications*.

Huifu Xu received his B.Sc. degree in Mathematics (with computational mathematics specialization) from Nanjing University (China) in 1986 and Ph.D. degree in Mathematics from Ballarat University (Australia) in 1999. He is a Senior Lecturer in Operations Research in the School of Mathematics, University of Southampton. His research interest covers a number of areas in continuous optimization and operational research including numerical methods for solving nonsmooth/nonconvex optimization problems and nonsmooth equations, modelling optimal bidding behaviour of generators in deregulated electricity markets, stochastic mathematical programs with equilibrium constraints (SMPECs), nonsmooth stochastic optimization. He has published papers in *Journal of Optimization Theory and Applications, Optimization*, *Matematical Methods of Operations Research*, *Mathematical Programming*, *SIAM Journal on Optimization*, *Mathematics of Operations Research,* etc.