Monte Carlo Methods for Mean-Risk Optimization and Portfolio Selection

Huifu Xu and Dali Zhang
School of Mathematics
University of Southampton
Southampton, UK

(May 2009, revised March 2010)

Abstract

Stochastic programming is a well-known instrument to model many risk management problems in finance. In this paper we consider a stochastic programming model where the objective function is the variance of a random function and the constraint function is the expected value of the random function. Instead of using popular scenario tree methods, we apply the well-known sample average approximation (SAA) method to solve it. An advantage of SAA is that it can be implemented without knowing the distribution of the random data. We investigate the asymptotic properties of statistical estimators obtained from the SAA problem including examining the rate of convergence of optimal solutions of the SAA problem as sample size increases. By using the classical penalty function technique and recent results on uniform exponential convergence of sample average random functions, we show that under some mild conditions the statistical estimator of the optimal solution converges to its true counterpart at an exponential rate. We apply the proposed model and the numerical method to a portfolio management problem and present some numerical results.

Key words. Variance minimization, sample average approximation, risk management, exponential convergence.

1 Introduction

The practice of mean-risk models has been widely used in portfolio selection problems, and had a profound impact on the economic modeling of financial markets and the investment of assets since the first introduction of mean-variance models by Markowitz [16] in the financial literature. In most of real markets, returns are characterized and compared by two statistics: the expected value and the value of the risk measure. A tradeoff is often required for each portfolio selection problem between the high expected return and the low risk, where the risk measure plays an important role in making the decisions. In spite of criticism and many proposals of new risk measures such as Value-at-Risk and Conditional Value-at-Risk [4, 23], variance which calculates the spread around the expected value of a random variable is still one of the most widely used measures of risk in the portfolio selection problems.

In this paper, we study the following mean-variance minimization problem with a general

1

return function:

$$\min_x \quad \mathrm{Var}\left[f(x, \xi(\omega))\right]$$
$$\text{s.t.} \quad x \in X,$$
$$\mathbb{E}[g(x, \xi(\omega))] \leq 0, \tag{1.1}$$

where $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^m$ are locally Lipschitz continuous, $x \in X$ is a decision vector with X being a nonempty convex subset of $\mathbb{R}^n$, and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$ with support Ξ, $\mathbb{E}[\cdot]$ denotes the expected value with respect to the distribution of ξ, $\mathrm{Var}[\cdot]$ denotes the variance of a random variable, that is,

$$\mathrm{Var}\left[f(x, \xi(\omega))\right] = \mathbb{E}\left[\left[f(x, \xi(\omega)) - \mathbb{E}(f(x, \xi(\omega)))\right]^2\right] = \mathbb{E}\left[f^2(x, \xi(\omega))\right] - \left(\mathbb{E}[f(x, \xi(\omega))]\right)^2. \tag{1.2}$$

To ease the notation, we will use ξ to denote either random vector $\xi(\omega)$ or an element of $\mathbb{R}^k$ depending on the context. Throughout this paper, we assume that $\mathbb{E}[f(x, \xi)]$, $\mathbb{E}[f^2(x, \xi)]$ and $\mathbb{E}[g(x, \xi)]$ are well-defined for every $x \in X$. The stochastic programming model (1.1) is well-known for measuring risks in finance pioneered by Markowitz in 1950s [16]: it aims to minimize the variance which is often used to measure the risks subject to the constraints of the expected revenues or costs. This type of fundamental mean-variance modeling and analysis is instrumental to both practitioners and researchers in finance. For practitioners, the theory suggests that mean-variance efficient portfolios can play an important role in portfolio management applications. For researchers in finance, mean-variance analysis is central to many asset pricing theories as well as to empirical tests of those theories, see [1, 6].

The main objective of this paper is concerned with numerical methods for solving (1.1). One of the main difficulties is to deal with the expected values. In practice, it is often impossible to obtain a closed form of $\mathbb{E}[f(x, \xi)]$ and $\mathbb{E}[f^2(x, \xi)]$ either because they are computationally too expensive or the distribution function of ξ is unknown. However, it might be possible to obtain samples of ξ from past data or computer simulation. Specifically, let $\xi^1, \cdots, \xi^N$ be a sample of ξ, we consider the following sample average approximation problem for (1.1):

$$\min_x \quad \phi_N(x) := \frac{1}{N-1} \sum_{i=1}^N \left(f(x, \xi^i) - \frac{1}{N} \sum_{i=1}^N f(x, \xi^i)\right)^2$$
$$\text{s.t.} \quad x \in X,$$
$$g_N(x) := \frac{1}{N} \sum_{i=1}^N g(x, \xi^i) \leq 0. \tag{1.3}$$

We refer to (1.1) as the *true* problem and (1.3) as the *sample average approximation* (SAA) problem. SAA is a very popular method in stochastic optimization and it is known under various names such as sample path optimization (SPO) method [21], stochastic counterpart and more broadly Monte Carlo method, see [28] for a comprehensive review of the subject by Shapiro. The main benefit of SAA is that one does not have to calculate the expected values.

Note that the sample average approximation problem (1.3) is slightly different from those available in the literature: our objective function is the sample average approximation of the *variance* of a random function rather than the expected value. The difference makes it impossible to readily use the available analysis results of SAA in stochastic programming which deal

2

with expected values rather than variances. Wang and Ahmed [30] considered a stochastic programming model with a deterministic objective function and constraints of the expected value of a random function. They proposed SAA method to solve the model and demonstrated the exponential rate of convergence of the feasible set of SAA problem to its true counterpart. The stochastic programming model and SAA scheme were consequently applied to portfolio optimization problems in finance. Shapiro [27] considered a general class of stochastic programming problems which subsume (1.1) and proposed to solve them by solving a sequence of approximating problems including sample average approximation. He analyzed the asymptotic behavior of statistical estimators of the optimal values and optimal solutions obtained from solving the approximating problems. His analysis is carried out by a parametric programming approach where the approximating functionals are treated as a parameter defined on a Banach space, and the asymptotics of the optimal value and optimal solution estimators is consequently derived through an extended delta method by driving the parameter to its limit value.

Our focus here is on the asymptotic convergence of the optimal solution of SAA problem (1.3): assuming that we obtain an optimal solution to problem (1.3), 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 [27]: (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 the recently established theory of uniform exponential convergence of sample average random functions in [32] to carry out our analysis; (c) the sampling is not necessarily iid, that is, our analysis covers the case if the sampling is generated by Quasi-Monte Carlo methods. Apart from the convergence analysis of sample average approximation for mean-variance models, we also apply the numerical scheme to a classical portfolio selection problem. In most portfolio selection problems in practice, a fund manager makes his decision at the beginning of an investment period before knowing the return of each available asset. The decision is required on the amount (proportion) of capital to be invested in each of the available assets with the objective of a desired tradeoff between the variance and the average return at the end of investment. In our portfolio selection model, the investor's profit consists of two parts: the net profit from the investment in financial market and the taxation on the profit, where the taxation can be regarded as a transaction cost. In the literature of portfolio optimization, the most well-studied transaction costs are constant or linear function of the portfolio return [17]. Due to unavailability of the distribution of uncertainty or computational complexity of the mathematical models, a closed form of the optimal decision can rarely be obtained for these portfolio selection problems, see [7, 17]. In this paper we apply the SAA method to the mean-variance model and demonstrate the exponential rate of convergence of approximate optimal decision (obtained from solving sample average approximation problem) to its true counterpart as sample size increases.

The rest of this paper is organized as follows. In Section 2, we reformulate (1.3) as a minimization problem by moving the constraint to the objective through exact penalization and show the boundedness of the penalization parameters. In Section 3, we investigate the exponential rate of convergence of optimal solutions obtained from solving an exactly penalized SAA problem under general sampling. In Section 4, we apply SAA to a portfolio selection problem with taxation costs. Numerical results are presented in Section 5.

Throughout this paper, we use the following notation. $x^T y$ 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 and $\mathbb{H}(D_1, D_2)$ denotes the Hausdorff distance between the two sets, that is, $\mathbb{H}(D_1, D_2) := \max\left(\mathbb{D}(D_1, D_2), \mathbb{D}(D_1, D_2)\right)$. Finally, for a real valued function $h(x)$, we use $\nabla h(x)$ to denote the gradient of h at x which is a column vector, and when $h(x)$ is vector valued, the same notation refers to the classical Jacobian of h at x (where the gradient of each component forms a column of the Jacobian matrix).

2 Exact penalization

One of the main aims of this paper is to analyze the convergence of optimal solution x^N as sample size increases. A natural way to do this is to consider the uniform convergence of the objective function $\phi_N(x)$ and constraint function $g_N(x)$ in SAA problem (1.3). Unfortunately this approach has a main technical difficulty: the feasible set of (1.3) defined by $g_N(x) \leq 0$ varies as N increases. To get around the complication in our analysis, we reformulate (1.3) as a minimization problem by moving the constraint $g_N(x) \leq 0$ to the objective through exact penalization and subsequently leaving $x \in X$ as the only constraint which is independent of N.

2.1 Reformulation

We use the classical exact penalty function method to derive the reformulation. Let us start with the true problem (1.1). By [10, Theorem 14.3.1], there exists a positive constant $\lambda^* > 0$ such that the optimal solution to (1.1) coincides with the optimal solution of the following optimization problem:

$$\begin{aligned} \min_x \quad & \psi(x) := \mathrm{Var}\,[f(x, \xi)] + \lambda^* p(\mathbb{E}[g(x, \xi)]) \\ \text{s.t.} \quad & x \in X, \end{aligned} \tag{2.4}$$

if the second order sufficient conditions of both problems are satisfied at optimal solutions. Here $p(z) = \sum_{i=1}^m \max(0, z_i)$ for $z \in \mathbb{R}^m$ and $\lambda^* > 0$ is a constant. Likewise, the SAA problem (1.3) can be reformulated as

$$\begin{aligned} \min_x \quad & \psi_N(x) := \phi_N(x) + \lambda^N p\left(\frac{1}{N} \sum_{i=1}^N g(x, \xi^i)\right), \\ \text{s.t.} \quad & x \in X, \end{aligned} \tag{2.5}$$

if the second order sufficient conditions of (1.3) and (2.5) hold at optimal solutions, where λ^N is some positive number. To avoid technical complexity, we make a blanket assumption that the second order sufficient conditions of (1.1), (1.3), (2.4) and (2.5) hold at their optimal solutions.

The reformulation makes our asymptotic analysis easier: now we may look into the convergence of x^N to x^* by studying the uniform convergence of the objective functions $\psi_N(x)$ to $\psi(x)$ in (2.4) and (2.5) in that the feasible sets of the two problems are identical and independent of

N. This is a departure from Wang and Ahmed's analysis which depends on the sample average approximation of feasible sets ([30, Proposition 2]). The only technical issue to be resolved for the uniform convergence of $\psi_N(x)$ to $\psi(x)$ is the boundedness of the sequence of the penalization parameters $\{\lambda^N\}$. For this purpose, we need to investigate the Lagrange multiplier of the SAA problem (1.3), denoted by μ^N. Let us start by making some basic assumptions.

Assumption 2.1 *Let $f(x,\xi)$ and $g(x,\xi)$ be defined as in (1.1).*

(a) *X is a nonempty convex and compact set;*

(b) *there exists a point $x \in X$ such that $\mathbb{E}[f(x,\xi)] < \infty$, $\mathbb{E}\left[f^2(x,\xi)\right] < \infty$, $\mathbb{E}[g_i(x,\xi)] < \infty$, for $i = 1, \cdots, m$;*

(c) *f and g are Lipschitz continuous w.r.t. x and their Lipschitz modulus are bounded by $\kappa_1(\xi) > 0$, where $\mathbb{E}[\kappa_1(\xi)] < \infty$;*

(d) *$\nabla_x f(x,\xi)$ and $\nabla_x g(x,\xi)$ are Lipschitz continuous w.r.t. x and their Lipschitz modulus are bounded by $\kappa_2(\xi) > 0$, where $\mathbb{E}[\kappa_2(\xi)] < \infty$;*

(e) *$\sup_{x \in X} |f(x,\xi)| \leq \kappa_3(\xi)$, where $\mathbb{E}[\kappa_1(\xi)\kappa_3(\xi)] < \infty$.*

The assumption is standard except part (a) where we require the feasible set X to be compact. This is purely for the convenience of convergence analysis. In fact, we may relax the condition from compactness to closedness and carry out our convergence analysis on a compact subset of X which contains sequence $\{x^N\}$ w.p.1. From practical point of view, this kind of assumption is reasonable as the quantities of decision variables are usually bounded, see a similar assumption by Wang and Ahmed [30] in portfolio optimization. Under Assumption 2.1, the mean-risk model (1.1) is well defined in the sense that the underlying functions are finite valued for every $x \in X$. To see this, let us explain that under conditions (b) and (c), we have

$$|f(x',\xi)| \leq |f(x,\xi)| + \kappa_1(\xi)\|x' - x\|$$

for any $x' \in X$, which implies that $\mathbb{E}[f(x',\xi)]$ is well defined for every $x' \in X$. This comment also applies the components of $g(x,\xi)$. Moreover, from the inequality above, we have

$$\sup_{x \in X} |f(x',\xi)| \leq |f(x,\xi)| + \kappa_1(\xi)D,$$

where D denotes the diameter of set X. This means that we can choose $\kappa_3(\xi) = |f(x,\xi)| + \kappa_1(\xi)D$ in part (e) of the assumption. Let us now comment on the well-definedness of $\text{Var}\left[f(x',\xi(\omega))\right]$ for any $x' \in X$. It is easy to verify that under conditions (c) and (e) for any $x' \in X$,

$$\mathbb{E}[f^2(x',\xi)] \leq \mathbb{E}[f^2(x,\xi)] + \mathbb{E}[2\kappa_3(\xi)\kappa_1(\xi)]\|x' - x\|$$

which means that $\mathbb{E}[f^2(x',\xi)]$ is well defined for any $x' \in X$, and through (1.2), that $\text{Var}\left[f(x',\xi(\omega))\right]$ is well defined for any $x' \in X$.

For the convenience of discussion, we assume throughout this section that the sampling is iid. This does not contradict with the discussion in the next section where the sampling is assumed to be general (including both iid and non-iid). Indeed, all we need in Propositions 2.1 and 2.2 are uniform convergence of the sample average of f, g and their gradients, which can be proved through [32, Theorem 3.1] under Assumptions 3.1 and 3.2 when the sampling is non-iid.

Proposition 2.1 *Let f, g and ϕ_N be defined respectively as in (1.1) and (1.3). If the sampling is iid, then*

(i) *under Assumption 2.1 (a), (b) and (d), $\frac{1}{N}\sum_{i=1}^{N} f(x, \xi^i)$, $\frac{1}{N}\sum_{i=1}^{N} g(x, \xi^i)$ and $\frac{1}{N}\sum_{i=1}^{N} f^2(x, \xi^i)$ converge to $\mathbb{E}[f(x, \xi)]$, $\mathbb{E}[g(x, \xi)]$ and $\mathbb{E}[f^2(x, \xi)]$ respectively uniformly over X as $N \to \infty$;*

(ii) *under Assumption 2.1 (a) and (c), $\frac{1}{N}\sum_{i=1}^{N} \nabla_x f(x, \xi^i)$ and $\frac{1}{N}\sum_{i=1}^{N} \nabla_x g(x, \xi^i)$ converge to $\mathbb{E}[\nabla_x f(x, \xi)]$, $\mathbb{E}[\nabla_x g(x, \xi)]$ respectively uniformly over X as $N \to \infty$;*

(iii) *if Assumption 2.1 (a) and (d) hold and $\lambda^N \to \lambda^*$, then $\nabla \phi_N(x)$ converges to $\nabla \mathrm{Var}[f(x, \xi)]$ uniformly over X as $N \to \infty$.*

Proof. The claims can be easily proved by virtue of the classical uniform law of large numbers, see for example [26, Lemma A]. We omit the details. ∎

Let us introduce some notation. For a vector $a \in \mathbb{R}^n$, we use $a \geq 0$ to represent the componentwise nonnegativity of the vector and '$\perp$' to denote the perpendicularity of two vectors. Let $\mathcal{N}_X(x)$ denote the normal cone to X at point x, that is,

$$
\mathcal{N}_X(x) := \begin{cases} \{\eta \in \mathbb{R}^n : \eta^T(x' - x) \leq 0, \ \forall x' \in X\}, & \text{if } x \in X, \\ \emptyset, & \text{otherwise.} \end{cases}
$$

Assumption 2.2 *Consider the true problem (1.1). There is no non-zero multiplier $\mu \in \mathbb{R}_+^m$ such that*

$$
\begin{cases} 0 \in & \nabla \mathbb{E}[g(x, \xi)]\mu + \mathcal{N}_X(x), \\ 0 \leq & -\mathbb{E}[g(x, \xi)] \perp \mu \geq 0. \end{cases}
$$

Assumption 2.2 is known as *no nonzero abnormal multipliers constraint qualification* (NNAMCQ) which was proposed by Ye [34] for studying the first order optimality conditions of deterministic mathematical programs with equilibrium constraints (MPEC). The constraint qualification is a dual form of the well-known Mangasarian-Fromovitz constraint qualification (MFCQ). To see this, let us consider a simple case when $\mathcal{N}_X(x) = \{0\}$, that is, x is in the interior of X. The constraint qualification reduces to positive linear independence of the gradients of all active inequality constraints. By the Fakas lemma, there exists a vector $d \in \mathbb{R}^n$ such that $\nabla \mathbb{E}[g_i(x, \xi)]^T d < 0$ for active constraints i such that $\mathbb{E}[g_i(x, \xi)] = 0$. The latter is indeed the MFCQ, see [33, page 1696] for details.

Consider the first order necessary conditions of the SAA problem (1.3):

$$\begin{cases} 0 \in & \nabla\phi_N(x) + \nabla g_N(x)\mu + \mathcal{N}_X(x), \\ 0 \leq & -g_N(x) \perp \mu \geq 0. \end{cases} \qquad (2.6)$$

Let $\{\mu^N\}$ be a sequence of Lagrange multipliers satisfying (2.6). The following proposition states that $\{\mu^N\}$ is bounded under some moderate conditions.

Proposition 2.2 *Consider the SAA problem (1.3) with iid samples. Under Assumptions 2.1 and 2.2, the sequence of the Langrange multipliers $\{\mu^N\}$ is bounded w.p.1.*

Proof. Let x^N be a KKT point and μ^N be a corresponding vector of Lagrange multipliers. Then

$$0 \in \nabla\phi_N(x^N) + \nabla g_N(x^N)\mu^N + \mathcal{N}_X(x), \qquad (2.7)$$

and $0 \leq \mu^N \perp -g_N(x^N) \geq 0$. Assume for the sake of a contradiction that $\{\mu^N\}$ is unbounded. Then $\{\mu^N\}$ has a subsequence going to infinity. Since X is a compact, $\{x^N\}$ has a subsequence converging to some point $x^* \in X$. Assume without loss of generality that $\{x^N\} \to x^*$ and $\|\mu^N\| \to \infty$. Under Assumption 2.1, it follows from Proposition 2.1 that $\nabla\phi_N(x^N) \to \nabla\text{Var}[f(x^*,\xi)]$ and $\nabla g_N(x^N) \to \nabla\mathbb{E}[g(x^*,\xi)]$. Dividing both sides of (2.7) by $\|\mu^N\|$ and driving N to infinity, we arrive at

$$0 \in 0 + \nabla\mathbb{E}[g(x^*,\xi)]\hat{\mu} + \mathcal{N}_X(x^*), \qquad (2.8)$$

where $\hat{\mu}$ is the accumulation point of $\{\mu^N/\|\mu^N\|\}$ and hence $\hat{\mu} \geq 0$ and $\|\hat{\mu}\| = 1$. This leads to a contradiction to Assumption 2.2. The proof is complete. ∎

Proposition 2.2 gives a qualitative description of the boundedness of μ^N. In what follows, we present a quantitative estimation of the multiplier.

Proposition 2.3 *Let $\mathcal{W}^* := (X^*, \mathcal{M}^*)$ denote the set of KKT pairs of the true problem (1.1) and $\mathcal{W}_N := (X_N, \mathcal{M}_N)$ the set of KKT pairs of SAA problem (1.3). Under Assumptions 2.1 and 2.2, for every $\epsilon > 0$, there exists $N(\epsilon) > 0$ such that*

$$\mathbb{H}(\mathcal{W}_N, \mathcal{W}^*) \leq \epsilon, \qquad (2.9)$$

for $N \geq N(\epsilon)$.

Proof. We can easily reformulate the KKT conditions (2.6) as a system of generalized equations with both x and μ being treated as variables:

$$\begin{cases} 0 \in & \nabla\phi_N(x) + \nabla g_N(x)\mu + \mathcal{N}_X(x), \\ 0 = & \min\{-g_N(x), \mu\}. \end{cases} \qquad (2.10)$$

Here the operation "min" is taken componentwise. The domain of the underlying function w.r.t. variable μ is $[0, +\infty)$ but we may assume without loss of generality that $\mu^N \in [0, \bar{\mu}]$ for all N

with some sufficiently large $\bar{\mu}$ given the boundedness of $\{\mu^N\}$ proved in Proposition 2.2. Likewise the KKT conditions of (1.1) can be written as

$$\begin{cases} 0 \in & \nabla \text{Var}[f(x,\xi)] + \nabla \mathbb{E}[g(x,\xi)]\mu + \mathcal{N}_X(x), \\ 0 = & \min\{-\mathbb{E}[g(x,\xi)], \mu\}. \end{cases} \quad (2.11)$$

The set of KKT pairs $\mathcal{W}_N$ is the set of solutions of problem (2.10) and $\mathcal{W}^*$ is the set of solutions of problem (2.11). Since the vector valued function $(\nabla \phi_N(x) + \nabla g_N(x)\mu, \min(g_N(x), \mu))^T$ converges uniformly to $(\nabla \text{Var}[f(x,\xi)] + \nabla \mathbb{E}[g(x,\xi)]\mu, \min(\mathbb{E}[g(x,\xi)], \mu))^T$ over compact set $X \times [0, \bar{\mu}]$, by [31, Lemma 4.1], we immediately obtain (2.9). ∎

Remark 2.1 *From Proposition 2.2, we know that $\{\mu^N\}$ is bounded. By [10, Theorem 14.3.1], (2.5) is equivalent to (1.3) so long as we set $\lambda^N \geq \|\mu^N\|$. From here on, we assume that $\{\lambda^N\}$ is an increasing bounded sequence satisfying this.*

3 Convergence analysis

We are now ready to investigate the convergence of the statistical estimator of the optimal solution, denoted by x^N, obtained from solving (2.5) as sample size increases. We do so by looking into the uniform convergence of $\psi_N(x)$ to $\psi(x)$ which implies the convergence x^N to X^*, where X^* denotes the set of optimal solutions of the true problem (1.1), or equivalently (2.4). From computational perspective, one often need estimate the sample size N given a prescribed error bound $d(x^N, X^*)$. A popular way to address this issue is to consider the so-called exponential convergence, that is, with probability approaching one exponentially fast, $\{x^N\}$ converges to X^* based on the classical Cramér's large deviation theorem [9], see [28, 29] and the references therein.

Note that Cramér's large deviation theorem is based on independent and identically distributed (iid) sampling. In some practical instances, however, it is difficult or computationally expensive to obtain an iid sample particularly when sample size is large. Indeed, the well-known Quasi-Monte Carlo method does not require iid sampling and yet it works remarkably well. This motivates one to study SAA under non-iid sampling. Dai, Chen and Birge [8] investigated the convergence of SAA estimators under general sampling (including iid and non-iid). They used the well-known Gärther-Ellis theorem [9] to establish the exponential convergence. More recently Homem-de-Mello [11] presented a comprehensive study of this issue and derived the exponential convergence of statistical estimators of optimal solutions in stochastic programming under non-iid sampling. To broaden the scope of the convergence theorems to be established in this section, we consider general sampling.

Let us define the following moment generating functions:

$$M_f^N(t) := \mathbb{E}\left\{e^{t\left[\frac{1}{N}\sum_{i=1}^N f(x,\xi^i)\right]}\right\},$$

$$M_{f^2}^N(t) := \mathbb{E}\left\{e^{t\left[\frac{1}{N}\sum_{i=1}^N f^2(x,\xi^i)\right]}\right\},$$

and

$$M_g^N(t) := \mathbb{E}\left\{ e^{t\left[\frac{1}{N}\sum_{i=1}^N g(x,\xi^i)\right]} \right\}.$$

Assumption 3.1 *For every $x \in X$ and $t \in \mathbb{R}$, the limits*

$$M_f(t) := \lim_{N\to\infty} M_f^N(t),$$

$$M_{f^2}(t) := \lim_{N\to\infty} M_{f^2}^N(t)$$

and

$$M_g(t) := \lim_{N\to\infty} M_g^N(t)$$

exist as an extended real number and $M_f(t) < \infty$, $M_g(t) < \infty$ and $M_{f^2}(t) < \infty$ for t close to 0, where $M_h(t) = e^{ht}$ denotes the moment generating function for random variables $h = f, g, f^2$. Moreover $\nabla g_N(x)$ converges uniformly to $\mathbb{E}[\nabla g(x,\xi)]$ over X.

Note that in the case when $\xi^1, \cdots, \xi^N$ is an iid sampling, Assumption 3.1 holds as long as

$$\max(M_f(t), M_{f^2}(t), M_g(t)) < \infty$$

for t close to zero, see [9, Section 2.3]. The implication of Assumption 3.1 is that when the sampling is not necessarily iid, one may use Gärtner-Ellis' large deviation theorem, [9, 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 is well-known, see for instances [8, 9].

Lemma 3.1 *Let $f : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real valued function and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ be a random vector defined on probability space $(\Omega, \mathcal{F}, P)$. Let Assumption 3.1 hold. Then for every $x \in X$ and small positive number $\varepsilon > 0$,*

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

$$\text{Prob}\left\{ \left| \frac{1}{N}\sum_{i=1}^N f^2(x,\xi^i) - \mathbb{E}[f^2(x,\xi)] \right| \geq \varepsilon \right\} \leq e^{-NI_{f^2}(-\varepsilon)} + e^{-NI_{f^2}(\varepsilon)}$$

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

for N sufficiently large, where

$$I_h(\varepsilon) := \sup_{t\in\mathbb{R}} \left\{ \varepsilon t - \log M_h(t) \right\}$$

for $h = f, f^2, g$, and $I_h(\varepsilon)$ and $I_h(-\varepsilon)$ and $I_h(-\varepsilon)$ are positive.

To establish the uniform exponential convergence, we also need an assumption on asymptotic behavior of the sample average of the modulus of function f. Similar assumption is made in [29] for the uniform exponential convergence of sample average random functions under iid sampling.

Assumption 3.2 *Let $\kappa_1(\xi)$ and $\kappa_3(\xi)$ be defined as in Assumption 2.1. For any $L' \geq \mathbb{E}[\max\{\kappa_1(\xi), \kappa_3(\xi)\}]$, there is a positive constant τ such that*

$$\text{Prob}\left\{\min\left(\frac{1}{N}\sum_{i=1}^{N}\kappa_1(\xi^i), \frac{1}{N}\sum_{i=1}^{N}\kappa_3(\xi^i)\right) \geq L'\right\} \leq e^{-\tau N}, \tag{3.12}$$

for N sufficiently large.

Under Assumption 3.2, we may strengthen Lemma 3.1 to the uniform exponential convergence.

Theorem 3.1 *Let Assumptions 2.1, 2.2, 3.1 and 3.2 hold. Assume[1] that $\{\lambda^N\} \to \lambda^*$ as $N \to \infty$. Then for every $\varepsilon > 0$, there exist positive constants $c(\varepsilon)$ and $\beta(\varepsilon)$, independent of N, such that*

$$\text{Prob}\left\{\sup_{x \in X} |\psi_N(x) - \psi(x)| \geq \varepsilon\right\} \leq c(\varepsilon)e^{-N\beta(\varepsilon)} \tag{3.13}$$

for N sufficiently large.

Proof. By a simple calculation,

$$\phi_N(x) = \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}f^2(x, \xi^i) - \left(\frac{1}{N}\sum_{i=1}^{N}f(x, \xi^i)\right)^2\right]$$

and through (1.2)

$$\phi_N(x) - \phi(x) = \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}f^2(x, \xi^i) - \mathbb{E}[f^2(x, \xi)] - \left(\left(\frac{1}{N}\sum_{i=1}^{N}f(x, \xi^i)\right)^2 - (\mathbb{E}[f(x, \xi)])^2\right)\right]$$
$$+ \frac{1}{N-1}\phi(x). \tag{3.14}$$

On the other hand, it is easy to verify that function $p(z)$ defined in (2.4) is globally Lipschitz continuous with modulus 1 and $p(z) \leq \|z\|$ which means

$$\left|\lambda^N p\left(\frac{1}{N}\sum_{i=1}^{N}g(x, \xi^i)\right) - \lambda^* p(\mathbb{E}[g(x, \xi)])\right| \leq |\lambda^N - \lambda^*|\left\|\frac{1}{N}\sum_{i=1}^{N}g(x, \xi^i)\right\|$$
$$+ \lambda^*\left\|\frac{1}{N}\sum_{i=1}^{N}g(x, \xi^i) - \mathbb{E}[g(x, \xi)]\right\|. \tag{3.15}$$

---

[1]Under Assumptions 2.1, 2.2, 3.1 and 3.2, we can prove the uniform convergence of $\phi_N(x)$, $\nabla g_N(x)$ to $\mathbb{E}[\phi(x, \xi)]$ and $\mathbb{E}[\nabla g(x, \xi)]$ over set X. Through Proposition 2.2, this means $\{\mu^N\}$ is bounded. For simplicity of discussion, we assume $\{\lambda^N\} \to \lambda^*$ as $N \to \infty$ just for simplicity of discussion and it is justified.

Combining (3.14) and (3.15), we have

$$
\begin{aligned}
|\psi_N(x) - \psi(x)| \;\leq\; & \frac{N}{N-1}\left|\frac{1}{N}\sum_{i=1}^{N} f^2(x,\xi^i) - \mathbb{E}[f^2(x,\xi)]\right| \\
& +\frac{N}{N-1}\left|\left(\frac{1}{N}\sum_{i=1}^{N} f(x,\xi^i)\right)^2 - (\mathbb{E}[f(x,\xi)])^2\right| \\
& +\frac{1}{N-1}\phi(x) + \lambda^*\left\|\frac{1}{N}\sum_{i=1}^{N} g(x,\xi^i) - \mathbb{E}[g(x,\xi)]\right\| \\
& +|\lambda^N - \lambda^*|\left\|\frac{1}{N}\sum_{i=1}^{N} g(x,\xi^i)\right\|.
\end{aligned}
$$

Let $\epsilon > 0$ be a fixed small positive number. In what follows, we estimate

$$
\text{Prob}\left\{|\psi_N(x) - \psi(x)| \geq \epsilon\right\}.
$$

Observe first that for $N \geq 2$, we have $\frac{N}{N-1} \leq 2$. Moreover, since $\phi(x)$ is a continuous function and X is compact, there exists positive integer N such that when $N \geq \frac{5}{\epsilon}\sup_{x \in X}\phi(x) + 1$, we have $\frac{1}{N-1}\phi(x) \leq \frac{\epsilon}{5}$. Subsequently

$$
\begin{aligned}
|\psi_N(x) - \psi(x)| \;\leq\; & 2\left|\frac{1}{N}\sum_{i=1}^{N} f^2(x,\xi^i) - \mathbb{E}[f^2(x,\xi)]\right| \\
& +2\left|\left(\frac{1}{N}\sum_{i=1}^{N} f(x,\xi^i)\right)^2 - (\mathbb{E}[f(x,\xi)])^2\right| \\
& +\lambda^*\left\|\frac{1}{N}\sum_{i=1}^{N} g(x,\xi^i) - \mathbb{E}[g(x,\xi)]\right\| \\
& +|\lambda^N - \lambda^*|\left\|\frac{1}{N}\sum_{i=1}^{N} g(x,\xi^i)\right\| + \frac{\epsilon}{5}.
\end{aligned}
$$

Denote the first, second, third and fourth terms on the right hand side of the above inequality in sequel by $R_1^N(x)$, $R_2^N(x)$, $R_3^N(x)$ and $R_4^N(x)$. Under Assumption 2.1, $f^2(x,\xi)$ and $g(x,\xi)$ are Lipschitz continuous with modulus $\kappa_1(\xi)\kappa_3(\xi)$ and $\kappa_1(\xi)$. Together with the pointwise exponential convergence in Lemma 3.1, we can use [32, Theorem 3.1] to obtain the uniform exponential convergence, that is, there exist positive constants $c_l(\epsilon)$ and $\beta_l(\epsilon)$ such that

$$
\text{Prob}\left\{\sup_{x \in X} R_l^N(x) \geq \frac{\epsilon}{5}\right\} \leq c_l(\epsilon)e^{-\beta_l(\epsilon)N} \tag{3.16}
$$

for $l = 1, 3$. In what follows, we look into $R_2^N(x)$ and $R_4^N(x)$. Note that

$$
R_2^N(x) = 2\left|\frac{1}{N}\sum_{i=1}^{N} f(x,\xi^i) + \mathbb{E}[f(x,\xi)]\right|\left|\frac{1}{N}\sum_{i=1}^{N} f(x,\xi^i) - \mathbb{E}[f(x,\xi)]\right|.
$$

11

Using the conditional probability (conditional on whether $\sup_{x \in X} \left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) + \mathbb{E}[f(x, \xi)] \right|$ exceeds $2\mathbb{E}[\kappa_3(\xi)] + 1$ or not), we derive

$$\text{Prob}\left\{ \sup_{x \in X} R_2^N(x) \geq \frac{\epsilon}{5} \right\}$$

$$\leq \quad \text{Prob}\left\{ \sup_{x \in X} \left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) + \mathbb{E}[f(x, \xi)] \right| \sup_{x \in X} \left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) - \mathbb{E}[f(x, \xi)] \right| \geq \frac{\epsilon}{10} \right\}$$

$$\times \quad \text{Prob}\left\{ \sup_{x \in X} \left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) + \mathbb{E}[f(x, \xi)] \right| < 2\mathbb{E}[\kappa_3(\xi)] + 1 \right\}$$

$$+ \quad \text{Prob}\left\{ \sup_{x \in X} \left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) + \mathbb{E}[f(x, \xi)] \right| \sup_{x \in X} \left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) - \mathbb{E}[f(x, \xi)] \right| \geq \frac{\epsilon}{10} \right\}$$

$$\times \quad \text{Prob}\left\{ \sup_{x \in X} \left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) + \mathbb{E}[f(x, \xi)] \right| \geq 2\mathbb{E}[\kappa_3(\xi)] + 1 \right\}$$

$$\leq \quad \text{Prob}\left\{ \sup_{x \in X} \left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) - \mathbb{E}[f(x, \xi)] \right| \geq \frac{\epsilon}{10(2\mathbb{E}[\kappa_3(\xi)] + 1)} \right\}$$

$$+ \quad \text{Prob}\left\{ \sup_{x \in X} \left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) + \mathbb{E}[f(x, \xi)] \right| \geq 2\mathbb{E}[\kappa_3(\xi)] + 1 \right\}$$

Under Assumption 2.1 (b), f is Lipschitz continuous in x and Lemma 3.1 implies that $\frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i)$ converges to $\mathbb{E}[f(x, \xi)]$ pointwise on X at exponential rate. By [32, Theorem 3.1], $\frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i)$ converges to $\mathbb{E}[f(x, \xi)]$ uniformly on X at exponential rate, that is, there exist positive constants $c_2(\epsilon)$ and $\beta_2(\epsilon)$ such that

$$\text{Prob}\left\{ \sup_{x \in X} \left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) - \mathbb{E}[f(x, \xi)] \right| \geq \frac{\epsilon}{10(2\mathbb{E}[\kappa_3(\xi)] + 1)} \right\} \leq c_2(\epsilon) e^{-\beta_2(\epsilon)N}. \qquad (3.17)$$

On the other hand, under Assumption 2.1 (d),

$$\left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) + \mathbb{E}[f(x, \xi)] \right| \leq \frac{1}{N} \sum_{i=1}^{N} \kappa_3(\xi^i) + \mathbb{E}[\kappa_3(\xi)].$$

By Assumption 3.2,

$$\text{Prob}\left\{ \sup_{x \in X} \left| \frac{1}{N} \sum_{i=1}^{N} f(x, \xi^i) + \mathbb{E}[f(x, \xi)] \right| \geq 2\mathbb{E}[\kappa_3(\xi)] + 1 \right\} \leq \text{Prob}\left\{ \frac{1}{N} \sum_{i=1}^{N} \kappa_3(\xi^i) \geq \mathbb{E}[\kappa_3(\xi)] + 1 \right\}$$

$$\leq \quad e^{-N\tau} \qquad (3.18)$$

for N sufficiently large. Substituting (3.17) and (3.18) into the term $R_2^N(x)$ yields the following inequality

$$\text{Prob}\left\{ \sup_{x \in X} R_2^N(x) \geq \frac{\epsilon}{5} \right\} \leq c_2(\epsilon) e^{-\beta_2(\epsilon)N} + e^{-N\tau}.$$

Finally, we estimate the probability of $\sup_{x \in X} R_4^N(x) \geq \epsilon/5$, where

$$R_4^N(x) = \left| \lambda^N - \lambda^* \right| \left| \frac{1}{N} \sum_{i=1}^{N} g(x, \xi^i) \right|.$$

By Assumption 2.1 (b) and the pointwise exponential convergence of $\frac{1}{N} \sum_{i=1}^{N} g(x, \xi^i)$ to $\mathbb{E}[g(x, \xi)]$ on set X as established in Lemma 3.1, we can easily use [32, Theorem 3.1] to show the uniform exponential convergence of $\frac{1}{N} \sum_{i=1}^{N} g(x, \xi^i)$ to $\mathbb{E}[g(x, \xi)]$ on set X. Since $\sup_{x \in X} \mathbb{E}[g(x, \xi)]$ is bounded by $\mathbb{E}[g(\bar{x}, \xi)] + \|x - \bar{x}\|^\gamma \mathbb{E}[\kappa_1(\xi)]$ for any $\bar{x} \in X$, and $\lambda^N \to \lambda^*$, then for given $\varepsilon > 0$, there exist $N(\varepsilon) > 0$, $c_4(\varepsilon) > 0$ and $\beta_4(\varepsilon) > 0$ such that for any $N > N(\epsilon)$, that is,

$$\text{Prob} \left\{ \sup_{x \in X} R_4^N(x) \geq \frac{\epsilon}{5} \right\} \leq c_4(\varepsilon) e^{-\beta_4(\varepsilon)N}.$$

Combining the estimation of probabilities of $\sup_{x \in X} R_l^N(x) \geq \epsilon/5$, $l = 1, 2, 3$ and 4, we have

$$\text{Prob} \left\{ |\psi_N(x) - \psi(x)| \geq \epsilon \right\} \leq c(\epsilon) e^{-\beta(\epsilon)N},$$

where $c(\epsilon) = 2c_1(\epsilon) + c_2(\epsilon) + 1 + c_4(\epsilon)$ and $\beta(\epsilon) = \min\{\beta_1(\epsilon), \beta_2(\epsilon), \tau, c_4(\epsilon)\}$. This shows (3.13) and hence completes the proof. ∎

We now apply the uniform exponential convergence results to establish the exponential convergence of $\{x^N\}$.

Lemma 3.2 *Consider a general constrained minimization problem*

$$\begin{aligned} \min \quad & p(x) \\ \text{s.t.} \quad & x \in X \end{aligned} \tag{3.19}$$

where $p : \mathbb{R}^n \to \mathbb{R}$ and X is a subset of $\mathbb{R}^n$, and a perturbed program

$$\begin{aligned} \min \quad & \tilde{p}(x) \\ \text{s.t.} \quad & x \in X \end{aligned} \tag{3.20}$$

where $\tilde{p} : \mathbb{R}^n \to \mathbb{R}$ is a perturbation of p. Let U^* denote the set of optimal solutions of (3.19) and V^* be the set of optimal solutions of (3.20). Assume that neither U^* nor V^* is empty. Then for any $\epsilon > 0$, there exists a $\delta > 0$ (depending on ϵ) such that if $\sup_{x \in X} |\tilde{p}(x) - p(x)| \leq \delta$, $\forall x \in X$, then $\mathbb{D}(V^*, U^*) \leq \epsilon$.

The result is proved by Dai, Chen and Birge in [8, Lemma 3.2] when U^* is a singleton. It is not difficult to see from their proof that the conclusion holds when U^* contains multiple solutions.

Theorem 3.2 *Assume the setting and conditions of Theorem 3.1. Let $\{x^N\}$ be a sequence of optimal solutions to the SAA problem (1.3) and X^* be the set of optimal solutions to the true problem (1.1). For any $\epsilon > 0$, there exist constants $\beta(\epsilon) > 0$ and $c(\epsilon) > 0$ such that*

$$\text{Prob} \left\{ d\left(x^N, X^*\right) \geq \epsilon \right\} \leq c(\epsilon) e^{-\beta(\epsilon)N}. \tag{3.21}$$

Proof. Let $\epsilon > 0$ be fixed. By Lemma 3.2, there exists $\delta(\varepsilon) > 0$ such that $d(x^N, X^*) \leq \epsilon$ when $\sup_{x \in X} |\psi_N(x) - \psi(x)| \leq \delta(\varepsilon)$. For the given $\delta(\varepsilon)$, it follows from Theorem 3.1 that there exists $\beta(\delta(\varepsilon)) > 0$, $c(\delta(\varepsilon)) > 0$ (for simplicity of notation, we write them as $c(\varepsilon)$ and $\beta(\varepsilon)$) and a sufficiently large integer N_0, such that

$$\text{Prob} \left\{ \sup_{x \in X} |\psi_N(x) - \psi(x)| \geq \delta(\epsilon) \right\} \leq c(\epsilon) e^{-\beta(\epsilon)N},$$

for any $N \geq N_0$. This implies

$$\text{Prob} \left\{ d(x^N, X^*) \geq \epsilon \right\} \leq \text{Prob} \left\{ \sup_{x \in X} |\psi_N(x) - \psi(x)| \geq \delta(\epsilon) \right\} \leq c(\epsilon) e^{-\beta(\epsilon)N},$$

for any $N \geq N_0$. The proof is complete. ■

4   The portfolio selection problem with taxation costs

In this section, we apply the mean-risk model (1.1) and the sample average approximation scheme discussed in the preceding sections to a portfolio selection problem with taxation cost. The general background of this decision making problem is to address the requirement of a traditional fund manager and the tax regime imposed by a regulator over a particular investment period, see [24] for more details.

Consider a set of M assets indexed by $m \in \{1, 2, \ldots, M\}$. We denote the outcome price for per unit of asset m at the beginning and the end of the investment period by p_m^0 and p_m respectively. The unit return rate of the asset is defined as $r_m := p_m/p_m^0$. The end price p_m is often unknown at the beginning of the investment period due to uncertainties in the investment. We describe the uncertainty by a vector of random variables $\xi(\omega)$ and write p_m and r_m as a function of ξ, $p_m(\xi(\omega))$ and $r_m(\xi(\omega))$, for $m = 1, \cdots, M$. To ease the notation, we use ξ to denote either random vector $\xi(\omega)$ or an element of $\mathbb{R}^k$ depending on the context.

Let w denote the total capital available for the investment and w_m, $m = 1, \cdots, M$, to be invested in asset m. Let $x_m := w_m/w$ denote the proportion of the capital invested in asset m. Without loss of generality, we assume that the fund manager invests the rest of his capital $w_0 := w - \sum_{m=1}^{M} w_m$ in a risk-free asset (such as bond and bank account) with a deterministic return r_0. Let $x_0 := w_0/w$ and $x := (x_0, x_1, x_2, \cdots, x_M)^T$ denote the manager's decision vector. The return rate of this portfolio can be formulated as:

$$R(x, \xi) = x_0 r_0 + x_1 r_1(\xi) + x_2 r_2(\xi) + \cdots + x_M r_M(\xi).$$

By defining the vector of return rate as $r(\xi) := (r_0, r_1(\xi), \ldots, r_M(\xi))^T$, we can rewrite $R(x, \xi)$ in a concise form:

$$R(x, \xi) = x^T r(\xi). \tag{4.22}$$

From the definition, $(x_0, x_1, x_2, \ldots, x_M)$ must satisfy a set of basic constraints which specify the feasible set X of the decision vector:

$$X = \left\{ (x_0, x_1, x_2, \ldots, x_M) \Big| \sum_{m=1}^{M} x_m = 1, x_m \geq 0, \text{ for all } m = 0, 1, 2, \ldots, M \right\}. \tag{4.23}$$

14

The decision on making a choice between two portfolios $x, x' \in X$ is based on the random returns $R(x, \xi)$ and $R(x', \xi)$. Here we use a mean-variance optimization model to give a reference criterion under which portfolio vector $x \in X$ is considered 'better' or 'preferred' than others. Let us first formulate the profit that the manager might obtain from the investment. Given the total amount of capital w and the rate of return of the investment $R(x, \xi)$, the total amount of capital at the end of the investment period can be formulated as $wR(x, \xi)$. Hence the profit of the fund manager obtained from this investment period can be written as $h(x, \xi) = wR(x, \xi) - w$.

In the mean-variance optimization model to be discussed, we consider the case that tax is paid by the manager. The tax function, denoted by $T(h)$, gives the amount of capital to be taxed when the manager's profit from the investment is h. A popular tax function is defined as follows:

$$T(h) = \begin{cases} t_1 h, & \text{for } h \leq \beta; \\ t_1 \beta + t_2(h - \beta), & \text{for } h > \beta, \end{cases} \tag{4.24}$$

where t_1 and t_2 are the tax rates, β is the tax threshold from rate t_1 to rate t_2 and $T(h)$ denotes the total tax on the profit h. In the literature, $T(h)$ is known as a piecewise linear *progressive* tax function. Here the word 'progressive' is in the sense that $T(h)/h$ is an increasing function, see [18, Definition 2.4]. Observe that this tax function is nonsmooth and this might complicate the numerical solution of the problem. In what follows, we consider another strictly progressive tax function defined as follows:

$$T(h) = \begin{cases} K_2 \left(1 - e^{-\frac{h}{v_2}}\right) h, & \text{for } h \in [-w, 0), \\ K_1 \left(1 - e^{-\frac{h}{v_1}}\right) h, & \text{for } h \in [0, +\infty), \end{cases} \tag{4.25}$$

where K_1, K_2, v_1 and v_2 are positive constants with $0 < K_2 \leq K_1 < 1$, which guarantees that the tax cost is less than the overall profit h. It is easy to verify that $T(h)$ defined above is smooth (continuously differentiable). In practice, parameter K_1, taken as the ceiling rate of tax $T(h)/h$, varies for different types of taxation. The tax function has the following properties:

(a) for a sufficiently large income h, the rate of tax $T(h)/h = K_1 \left(1 - e^{-\frac{h}{v_1}}\right) \approx K_1$;

(b) if $h = 0$, i.e., there is no income from the investment, then $T(h)/h = K_1 \left(1 - e^{-\frac{h}{v_1}}\right) = 0$;

(c) if $h \in [-w, 0)$, i.e. the manager loses the capital in the investment, then the rate of tax $T(h)/h = K_2 \left(1 - e^{-\frac{h}{v_2}}\right) < 0$.

Property (c) indicates that when the profit h from the investment is negative, the taxation rate $T(h)/h$ is negative which implies certain amount of capital loss is refunded to the investor by the tax authority. This reflects practical policies taken by some governments and/or tax regimes that intend to promote investment. Note that by setting parameter v_2 sufficiently large we have from property (c) that the rate of the taxation $T(h)/h$ is close to zero. Therefore the tax policy that no compensation will be refunded for the investment can be viewed as part of the tax function (4.25).

For the simplicity of notation, we define $K(x,\xi) := \max(K_1\text{sign}(h(x,\xi)), -K_2\text{sign}(h(x,\xi)))$ and $v(x,\xi) := \max(v_1\text{sign}(h(x,\xi)), -v_2\text{sign}(h(x,\xi)))$, where

$$\text{sign}(h) = \begin{cases} 1, & h > 0, \\ 0, & h = 0, \\ -1, & h < 0. \end{cases}$$

Taking the taxation costs into account, if the manager chooses the portfolio vector $x = (x_0, x_1, \ldots, x_M)$ at the beginning of the investment, the post-tax profit obtained from this investment can be formulated as

$$H(x,\xi) = h(x,\xi) - T(h(x,\xi)) = h(x,\xi)\left((1 - K(x,\xi)) + K(x,\xi)e^{-h(x,\xi)/v(x,\xi)}\right). \quad (4.26)$$

The following lemma states the Lipschitz continuity of H and $\nabla_x H$ in x, a property that we need for the convergence analysis in Proposition 4.1.

Lemma 4.1 *Let $H(x,\xi)$ be defined as in (4.26). Then the following hold:*

(i) *H is Lipschtiz continuous w.r.t x and its Lipschtiz modulus is bounded by a positive constant $\kappa_1 > 0$;*

(ii) *$\nabla_x H(x,\xi)$ is Lipschitz continuous w.r.t x with a bounded Lipschtiz modulus.*

The proof is long but standard. We attach it in the appendix. Note that it is possible to show the continuous differentiability of function $H(x,\xi)$ although we do not need it in the following discussion.

Based on the discussions above, we are ready to develop a mean-variance optimization model for the portfolio problem: a portfolio x is chosen at the beginning of the investment period to satisfy the requirements on both expected profit and risk-aversion. Preference is then defined by considering a trade-off between a larger average return rate of the portfolio $\mathbb{E}[H(x,\xi)]$ and a smaller variance $\text{Var}[H(x,\xi)]$. Let d denote the threshold of the expected return from the portfolio. The optimal decision is to minimize the variance of the investment subject to the constraint that the expected profit is not lower than d:

$$\begin{aligned} \min_x \quad & \text{Var}[H(x,\xi)] \\ \text{s.t.} \quad & x \in X, \\ & \mathbb{E}[H(x,\xi)] \geq d. \end{aligned} \quad (4.27)$$

Our focus here is concerned with the numerical solution of problem (4.27). Obviously if the probability distribution function of $r_m(\xi)$, $m = 1, 2, \ldots, M$, is available before the investment period and we can obtain a closed form of $\mathbb{E}[H(x,\xi)]$ and $\text{Var}[H(x,\xi)]$, then (4.27) becomes a deterministic optimization problem and we can use any available nonlinear programming code to solve it. In practice, however, this is often difficult, if not impossible. Instead, it is relatively easier to acquire a sample of $r(\xi)$ from historical data. Based on this argument, we propose the

sample average approximation scheme for solving the portfolio optimization problem:

$$
\begin{aligned}
\min_{x} \quad & \frac{1}{N-1}\sum_{i=1}^{N} H^2(x,\xi^i) - \frac{N}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N} H(x,\xi^i)\right)^2 \\
\text{s.t.} \quad & x \in X, \\
& \frac{1}{N}\sum_{i=1}^{N} H(x,\xi^i) \geq d,
\end{aligned}
\tag{4.28}
$$

where $\xi^1,\cdots,\xi^N$ is a sample of ξ.

We apply Theorem 3.2 to derive the rate of convergence of optimal solution x^N (obtained from solving from problem (4.28)) to its true counterpart as sample size N increases. Observe that the tax function $T(h)$ is not strictly convex, hence problem (4.27) may have multiple optimal solutions. In a real investment problem, the random return rate r_m does not go to infinity or negative. Therefore, we may assume that for asset m, there exists a positive constant $\bar{r}_m$ such that $r_m \in [0,\bar{r}_m]$. We are now ready to state the main result of this section.

Proposition 4.1 *Let $\{x^N\}$ be a sequence of optimal solutions obtained from solving the SAA problem (4.28) and X^* the set of optimal solutions of the true problem (4.27). Assume that the sampling is iid[2]. Then*

(i) for every $\epsilon > 0$, there exists a positive integer $N_0 > 0$ such that for $N > N_0$, $d(x^N, X^*) \leq \epsilon$ w.p.1;

(ii) for any $\epsilon > 0$, there exist constants $\beta(\epsilon) > 0$ and $c(\epsilon) > 0$ such that

$$
\text{Prob}\left\{d(x^N, X^*) \geq \epsilon\right\} \leq c(\epsilon)e^{-\beta(\epsilon)N}
\tag{4.29}
$$

for N sufficiently large.

Proof. We prove the claims by virtue of Theorem 3.2. To this end, we by verify the conditions of the theorem in the context of (4.27). Let us start by verifying Assumption 2.1. From (4.23), we see that X is a convex and compact set. This verifies Assumption 2.1 (a). Assumption 2.1 (b)-(d) follow from Lemma 4.1. Assumption 2.2 is trivially satisfied because $\nabla_x \mathbb{E}[H(x,\xi)]$ is a nonzero vector for any $x \in X$.

Next we look at Assumption 3.1. Consider the moment generating function:

$$
M_H^N(t) = \mathbb{E}\left[e^{t\left[\frac{1}{N}\sum_{i=1}^{N} H(x,\xi^i)\right]}\right].
$$

By definition $|H(x,\xi)| \leq |h(x,\xi)| \leq \sum_{m=1}^{M} |r_m(\xi)| \leq \sum_{m=1}^{M} \bar{r}_m$. Consequently

$$
\begin{aligned}
M_H^N(t) \ &\leq \ \mathbb{E}\left[e^{wt\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M} |r_m(\xi^i)|\right]}\right] \\
&\leq \ \mathbb{E}\left[e^{w|t|\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M} |\bar{r}_m|\right]}\right] \\
&\leq \ e^{w|t|\sum_{m=1}^{M} \bar{r}_m},
\end{aligned}
$$

---

[2]In the case that the sampling is non-iid, we need Assumptions 3.1 and 3.2.

which implies $M_H(t) < +\infty$ for t close to 0. Since the sampling is iid, by the Cramer's large deviation theorem, $\lim_{N\to\infty} M_H^N(t)$ exits and it equals to $M_H(t)$. Likewise we can prove that $M_{H^2}(t) = \lim_{N\to\infty} M_{H^2}^N(t) < +\infty$ for t close to 0. This verifies Assumption 3.1. Finally, Assumption 3.2 follows from Cramer's large deviation theorem and Lemma 3.1. This completes the proof. ∎

To conclude this section, we comment that in some practical cases, a fund manager may be interested in pre-tax profit maximization. In this case, the optimization model becomes

$$
\begin{aligned}
\min_x \quad & \mathrm{Var}[h(x,\xi)] \\
\text{s.t.} \quad & x \in X, \\
& \mathbb{E}[h(x,\xi)] \geq d.
\end{aligned}
\tag{4.30}
$$

Applying the SAA scheme to it, we can obtain the exponential convergence of optimal solution estimators as in Proposition 4.1.

5 Numerical tests

We have carried out some numerical tests on two mean-variance optimization problems. In this section, we report the test results.

The first test problem is under the portfolio selection framework discussed in Section 4. We consider a set of stocks in iShare FTSE/Xinhua China 25 index of the New York Stock Exchange (NYSE:FXI) which consists 25 of the largest and most liquid Chinese stocks. The tests are performed over a set of 18 stocks from this index, excluding 7 stocks due to insufficient historical data. Our portfolio optimization model is similar to the mean-CVaR model in Roman, Darby-Dowman and Mitra [25, Section 5] except that we consider variance instead of CVaR in the objective as the risk measure. In the tests, we apply the sample average approximation to the problem on the basis of historical data collected from the stock market. The data set consists of daily closing prices for these 18 stocks, over a two-year period spanning from March 23rd 2007 to March 23rd 2009, download from *http://finance.google.com* with adjustment for stock splitting, including 300 historical samples of monthly return. The tests are implemented by using mathematical programming codes of GAMS installed in a PC with Windows XP operating system and the built-in solver PATHNLP.

To apply the proposed SAA framework to our numerical tests, we write the SAA of the mean-variance model (4.28) in a standard nonlinear programming form:

$$
\begin{aligned}
\min_x \quad & \frac{1}{N-1}\sum_{i=1}^{N} H^2(x,\xi^i) - \frac{N}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N} H(x,\xi^i)\right)^2 \\
\text{s.t.} \quad & x_m \geq 0,\ m = 0,1,2,\ldots,M, \\
& x_0 + x_1 + x_2 + \cdots + x_M = 1, \\
& \frac{1}{N}\sum_{i=1}^{N} H(x,\xi^i) \geq d.
\end{aligned}
\tag{5.31}
$$

Denote the set of candidate stocks by $\mathcal{M} = \{1, \cdots, M\}$, and let $x_m, m \in \mathcal{M}$ be the proportion of the total capital (normalized to $w = 1$) to be invested into stock m, where $M = 18$ in our

problem. Moreover, we use r_m, $m = 1, 2, \ldots, M$, to denote the corresponding monthly return rates, where the sample of $\{r_m\}_{m \in \mathcal{M}}$ is calculated based on a set of historical data: Let p_m^i denote the closing price in the real market of stock m on day i, we construct the data set of monthly return rates r_m^i by $r_m^i = p_m^{i+\Delta}/p_m^i$ for each $m \in \mathcal{M}$ and each time point i, where Δ is a moving-window that slides over the historical data on the time dimension i. Because what we consider is monthly return rate, the width of the moving-window Δ is set to be 30 days. In our analysis, we can take the collection of data from the market, $\{r_m^i, i = 1, 2, \ldots, S\}$, as the sample $r_m(\xi^i)$ in (5.31), where $S = 300$ is the number of samples.

First, we perform comparative static analysis on the constraint of expected return levels and the composition of the considered portfolio. One straightforward result that we may expect is that the ratio of capital invested to the risk-free asset (here, we only consider cash) decreases along with the increase requirement on the expected return level. By fixing the taxation parameters $K_1 = K_2 = 0.15$ and $v_1 = v_2 = 0.3$ in (5.31), we obtain the results on the composition of portfolios and the minimum variances for different expected return levels as follows:

Level d	Variance	Pre-tax Return	CBA	CEA	CHA	CHU	NTE	STP	CASH
0.040	0.0123	0.0539	0.1997	0.1860	0.0986	0.0000	0.1358	0.1879	0.1920
0.035	0.0076	0.0445	0.1779	0.1613	0.0484	0.0432	0.1356	0.1284	0.3051
0.030	0.0048	0.0366	0.1558	0.1397	0.0030	0.0751	0.1239	0.0876	0.4149
0.025	0.0029	0.0294	0.1249	0.1207	0.0000	0.0713	0.1092	0.0574	0.5166
0.020	0.0017	0.0227	0.0956	0.0992	0.0000	0.0600	0.0897	0.0368	0.6186
0.015	0.0009	0.0165	0.0687	0.0758	0.0000	0.0459	0.0677	0.0226	0.7193
0.010	0.0004	0.0107	0.0439	0.0511	0.0000	0.0305	0.0450	0.0125	0.8170

Table 1: The compositions of portfolios and the minimum variances w.r.t the return levels.

In Table 1 , CBA, CEA, CHA, CHU are the codes for stocks in FTSE/Xinhua China 25 index. Full name of each stock can be found at *http://finance.google.com*. Note that in our test, we consider 18 stocks in FTSE/Xinhua China 25 index. There are: ACH, CBA, CEA, CEO, CHA, CHL, CHU, CYD, GSH, HNP, LFC, NTE, PTR, SHI, SMI, STP, YZC, ZNH, and in the table we only list the stocks with non-zero capital investment. In our analysis, we normalize the total capital invested to 1 and we can regard the normalized investment on one stock as its weight in the selected portfolio. For instance, when the expected return is 0.040, the weight of CBA in the selected portfolio is 0.1997. From Table 1, we can easily see that, as the expected return level d increases the proportion of investment on CASH (risk-free assets) decreases (from 0.8170 to 0.1920). On the other hand, when the capital invested on risk-free asset (CASH) flows to the risky assets with higher return rate, the minimum variances increases, see column 2 and 10 in the table. Moreover, because stock CHA is of high return and high volatility, it is not selected in the portfolio when the investor's requirement on the expected return is not high, while its weight in the selected portfolio when the investor's expectation on the return rate increases from 0.030.

The resulting mean-variance efficient frontier is shown in Figure 1, and the changes of the weights of stocks CHA, CEA, NTE and STP are shown in Figure 2.
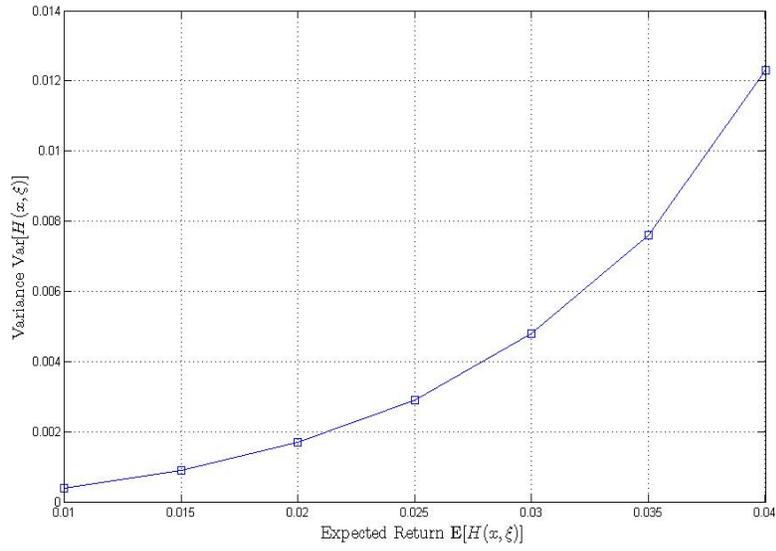
Figure 1: Efficient Frontier of the portfolio positions
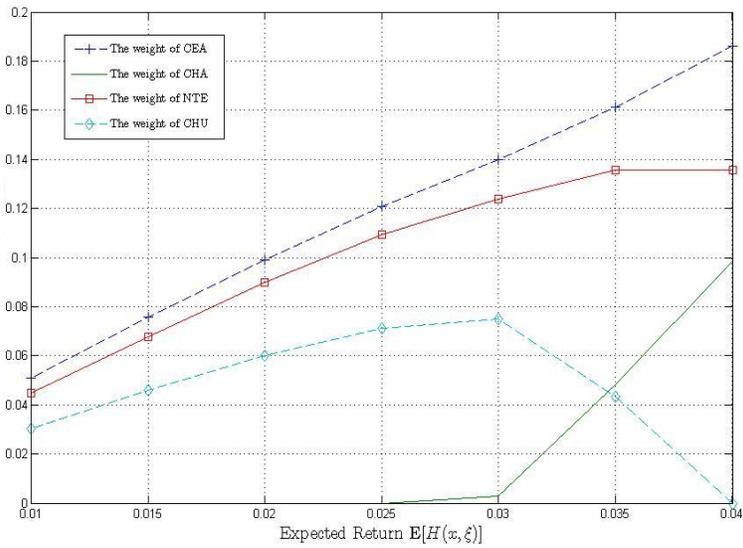
Figure 2: The weights of stocks in the portfolio positions.

Let us now look at the sensitivity of the minimum variances and the compositions of the selected portfolio with respect to the tax rate K in $H(x, \xi)$ in (5.31), where we fix the expected return rate at $d = 0.035$ and $v = 0.3$. We perform the SAA scheme to the model by varying the

tax rate K_1 and K_2 from 0 to 0.5. The results are listed in Table 2. In the table, we can see that the minimum variance increases along with the increase of the tax rate. The underlying reason of this phenomena is that to satisfy the required post-tax return level $d = 0.035$, the stocks with high return rate is preferred by the investor rather than the stocks with low variance, and hence the weight of CASH (risk-free assets) reduces along with the increase of the tax rate.

K_1, K_2	Variance	Pre-tax Return	CBA	CEA	CHA	NTE	STP	CASH
0.5	0.0076	0.0508	0.2148	0.1929	0.1214	0.1801	0.1157	0.0176
0.4	0.0062	0.0451	0.1909	0.1816	0.1154	0.1683	0.0894	0.2544
0.3	0.0054	0.0415	0.1745	0.1781	0.1120	0.1625	0.0693	0.3035
0.2	0.0049	0.0388	0.1615	0.1778	0.1088	0.1590	0.0534	0.3394
0.1	0.0045	0.0367	0.1503	0.1790	0.1051	0.1567	0.0405	0.3684
0.0	0.0042	0.0350	0.1403	0.1807	0.1009	0.1548	0.0302	0.3931

Table 2: The compositions of portfolios and the minimum variances w.r.t the tax rates.

Note that due to the limitation on the availability of the published data, we are not able to examine the rate of convergence with increasing sample size for this problem. Consequently we propose an academic example to test the convergence of the optimal values/solutions by drawing samples of increasing cardinality as a remedy.

Example 5.1 Consider a mean-variance optimization problem

$$\min_x \quad \text{Var}\,[f(x,\xi)]$$
$$\text{s.t.} \quad x \in X,$$
$$\mathbb{E}[g(x,\xi)] \leq 0,$$

where $f(x,\xi) = (x-\xi)^2 + 7\xi$ and $g(x,\xi) = 0.5(x-\xi)^2 + 0.1\xi - 20$, $X = [-10, 10]$ and ξ satisfies a normal distribution with mean 1 and standard deviation σ. It is easy to verify that the problem has a unique optimal solution $x^* = 4.5$. We carry out some numerical experiments on this problem with the SAA method (1.3) in Matlab 7.2 installed in a PC with Windows Vista where the SAA problem is solved by the Matlab built-in optimization solver $fmincon$. The numerical results are displayed in Figures 3 and 4.

The first set of tests are carried out with variance $\sigma = 0.5$. We perform comparative analysis with respect to the sample size from 200 to 9200. Figure 3 depicts how the optimal solutions obtained from solving SAA problem changes as the sample size increases. For a fixed sample size, 100 independent tests are carried out each of which solves the SAA problem and yields an approximation solution. In Figure 3, we use a vertical interval to indicate the range of the 100 approximate solutions. As sample size increases, we observe a trend of exponential convergence of the range of the approximate optimal solutions. We repeat the tests with $\sigma = 0.2$ and obtain the similar trend of convergence, see Figure 4.
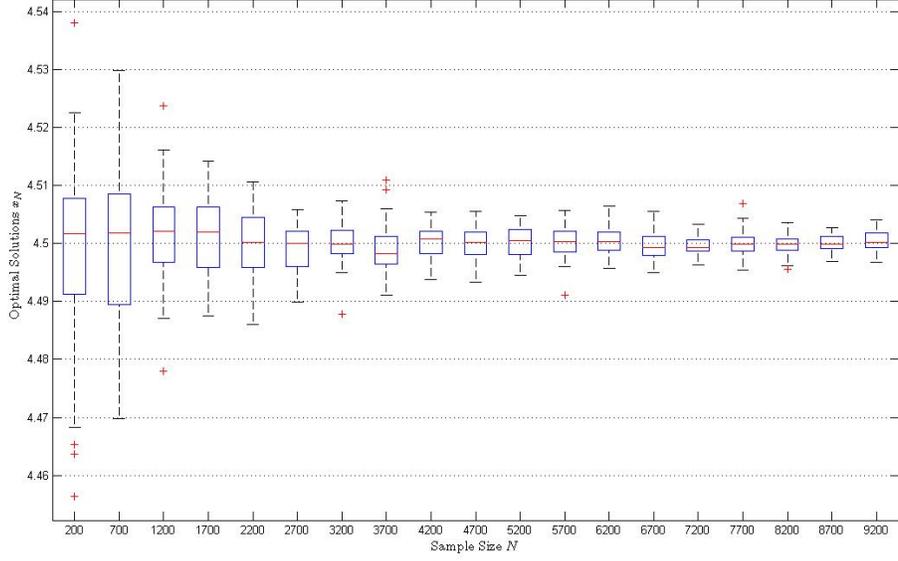
Figure 3: The convergence of the SAA problem when $\sigma = 0.5$.
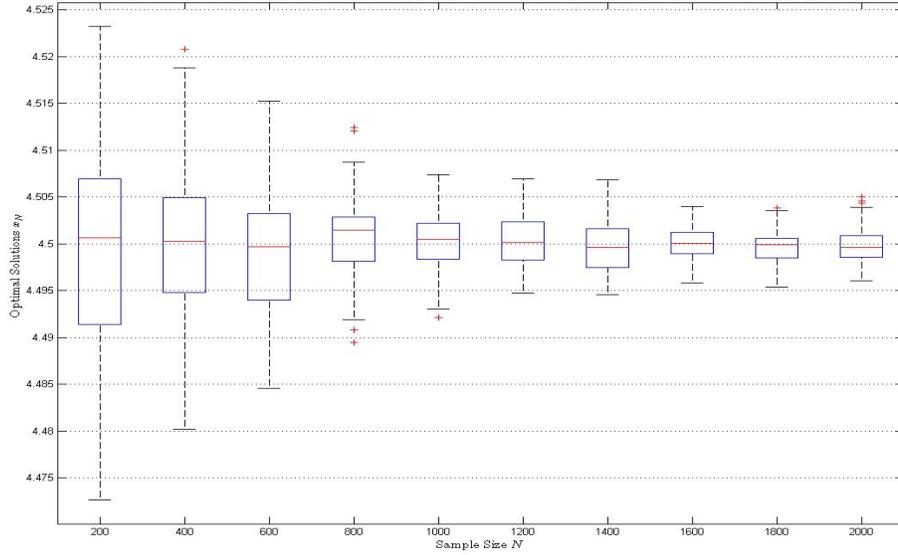
Figure 4: The convergence of the SAA problem when $\sigma = 0.2$.

References

[1] G. J. Alexander and A. M. Baptista, A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model, *Management Science*, Vol. 50, pp. 1261–1273, 2004.

[2] S. Alexander, T. F. Coleman and Y. Li, Minimizing CVaR and VaR for a portfolio of derivatives, *Journal of Banking and Finance*, Vol. 30, pp. 583–605, 2006.

[3] M. D'Antoni, Piecewise linear tax functions, progressively, and the principle of equal sacrifice, *Economics Letters*, Vol. 65, pp. 191-197, 1999.

[4] P. Artzner, F. Delbaen, J. Eber and D. Heath, Coherent measures of risks, *Mathematical Finance*, Vol. 9, pp. 203–228, 1999.

[5] Z. Artstein and R. J-B Wets, Consistency of minimizers and the SLLN for stochastic programs, *Journal of Convex Analysis*, Vol. 2, pp. 1–17, 1995.

[6] M. Britten-Jones, The sampling error in esitmates of mean-variance efficient portfolio weights, *The Journal of Finance*, Vol. 54, pp.655–671, 1999.

[7] T. R. Bielecki and S. R. Pliska, Risk sensitive asset management with transaction costs, *Finance and Stochastics*, Vol. 4, pp. 1–33, 2000.

[8] L. Dai, C. H. Chen and J. R. Birge, Convergence properties of two stage stochastic programming, *Journal of Optimization Theory and Applications*, Vol. 106, pp. 489-509, 2000.

[9] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer-Verlag, New York, 1998.

[10] R. Fletcher, *Practical Methods of Optimization*, Wiley, New York, 1987.

[11] T. Homen-De-Mello, On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling, *SIAM Journal on Optimization*, Vol. 19, pp. 524-551, 2008.

[12] Kjetil Hyland and Stein W. Wallace, Generating Scenario Trees for Multistage Decision Problems , *Management Science*, Vol. 47, pp. 295-307, 2001.

[13] A. J. King and R. T. Rockafellar, Asymptotic theory for solutions in statistical estimation and stochastic programming, *Mathematics of Operations Research*, Vol. 18, pp. 148-162, 1993.

[14] A. J. King and R. J.-B. Wets, Epi-consistency of convex stochastic programs, *Stochastics Stochastics Reports*, Vol. 34, pp. 83–92, 1991.

[15] J. Linderoth, A. Shapiro and S. Wright, The empirical behavior of sampling methods for stochastic programming, *Annals of Operations Research*, Vol. 142, pp. 215-241, 2006.

[16] H. M. Markowitz, Portfolio Selection: Efficient Diversification of Investment. John Wiley & Sons, New York, 1959.

[17] A. J. Morton and S. R. Pliska, Optimal portfolio management with fixed transaction costs, *Mathematical Finance*, Vol. 5, pp. 337–356, 2006.

[18] E. A. Ok, A note on the existence of progressive tax structures, *Soc Choice Welfare*, Vol. 14, pp. 527-543, 1997.

[19] G. Ch. Pflug, Scenario tree generation for multiperiod financial optimization by optimal discretization, *Mathematical Programming Series B*, Vol. 89, pp. 251-271, 2001.

[20] D. Ralph and H. Xu, Asympototic analysis of stationary points of sample average two stage stochastic programs: A generalized equation approach, manuscript, 2008. Submitted to *Mathematics of Operations Research.*

[21] S. M. Robinson, Analysis of sample-path optimization, *Mathematics of Operations Research*, Vol. 21, pp. 513–528, 1996.

[22] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[23] R. T. Rockafellar and S. Uryasev, Conditional value-at-risk for general loss distribution, *J. Bank. Finance*, Vol. 26, pp. 1443–1471, 2002.

[24] D. Roman, K. Darby-Dowman and G. Mitra, Porfolio construction based on stochastic dominance and target return distributions. *Mathematical Programming Series B*, Vol. 108, pp. 541–569, 2006.

[25] D. Roman, K. Darby-Dowman and G. Mitra, Mean-risk models using two risk measures: a multi-objective approach. *Quantitative Finance*, Vol. 4, pp. 443–458, 2007.

[26] R. Y. Rubinstein and A. Shapiro, *Discrete Events Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Methods*, John Wiley and Sons, New York, 1993.

[27] A. Shapiro, Asymptotic analysis of stochastic programs, *Annals of Operational Research*, Vol. 30, pp. 169-186, 1991.

[28] A. Shapiro, Monte Carlo sampling methods, in A. Rusczyński and A.Shapiro, editors, *Stochastic Programming*, Handbooks in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.

[29] A. Shapiro and H. Xu, Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation, *Optimization*, Vol. 57, pp. 395-418, 2008.

[30] W. Wang and S. Ahmed, Sample average approximation of expected value constrained stochastic programs, *Operations Research Letters*, Vol. 36, pp. 515-519, 2008.

[31] H. Xu and D. Zhang, Stochastic Nash equilibrium problems: sample average approximation and applications, Optimization online, 2009. Submitted to *Mathematical Programming.*

[32] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, preprint, School of Mathematics, University of Southampton, November 2008. Submitted to *Journal of Mathematics Analysis and Applications*, under revision.

[33] H. Xu and J.J. Ye, Necessary Optimality Conditions for two-stage Stochastic Programs with Equilibrium Constraints, *SIAM Journal on Optimization*, Vol. 20, pp. 1685–1715, 2010.

[34] J.J. Ye, Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints, *SIAM Journal on Optimization*, Vol. 10, pp. 943-962, 2000.

Appendix

Proof of Lemma 4.1. Part (i). By definition

$$
\begin{aligned}
|H(x',\xi) - H(x,\xi)| \leq \quad & |h(x',\xi) - h(x,\xi)| + |h(x',\xi)K(x',\xi) - h(x,\xi)K(x,\xi)| \\
& + \left| h(x',\xi)K(x',\xi)e^{-h(x',\xi)/v(x',\xi)} - h(x,\xi)K(x,\xi)e^{-h(x,\xi)/v(x,\xi)} \right|.
\end{aligned} \tag{5.32}
$$

Since $\|r(\xi)\| \leq \bar{r}$, by the definition of $h(x,\xi)$, we have

$$
|h(x',\xi) - h(x,\xi)| \leq \bar{r}\|x' - x\|. \tag{5.33}
$$

In what follows, we estimate the second and third terms at the right hand side of (5.32). To this end, we have to go through the cases according to the values of $h(x,\xi)$ and $h(x',\xi)$.

Case (a): $h(x,\xi) \geq 0$ and $h(x',\xi) \geq 0$. Then

$$
|h(x',\xi)K(x',\xi) - h(x,\xi)K(x,\xi)| \leq K_1 \bar{r}\|x' - x\|.
$$

Case (b): both $h(x,\xi) \leq 0$ and $h(x',\xi) \leq 0$. Then

$$
|h(x',\xi)K(x',\xi) - h(x,\xi)K(x,\xi)| \leq K_2 \bar{r}\|x' - x\|.
$$

Case (c): $h(x',\xi) \geq 0$ and $h(x,\xi) \leq 0$ ($h(x',\xi) \leq 0$ and $h(x,\xi) \geq 0$ can be discussed similarly). Let $\hat{x}$ be a point located on the line segment connecting x and x' such that $h(\hat{x},\xi) = 0$. Note that $\hat{x}$ may depend on ξ. Then $\|x' - \hat{x}\| + \|\hat{x} - x\| = \|x' - x\|$. Moreover,

$$
\begin{aligned}
|h(x',\xi)K(x',\xi) - h(x,\xi)K(x,\xi)| \quad \leq \quad & |h(x',\xi)K(x',\xi) - h(x_0,\xi)K(x_0,\xi)| \\
& + |h(x,\xi)K(x,\xi) - h(x_0,\xi)K(x_0,\xi)| \\
\leq \quad & K_1 \bar{r}\|x' - x_0\| + K_2 \bar{r}\|x - x_0\| \\
\leq \quad & \max\{K_1, K_2\}\bar{r}\|x' - x\|.
\end{aligned}
$$

The last inequality follows from Cases (a) and (b). Taking into account of all cases, we arrive at

$$
|h(x',\xi)K(x',\xi) - h(x,\xi)K(x,\xi)| \leq \max\{K_1, K_2\}\bar{r}\|x' - x\| < \bar{r}\|x' - x\|, \tag{5.34}
$$

where the second inequality is due to the fact that $0 < K_2 \leq K_1 < 1$. Analogously, we can show that

$$\left| h(x',\xi)K(x',\xi)e^{-\frac{h(x',\xi)}{v(x',\xi)}} - h(x,\xi)K(x,\xi)e^{-\frac{h(x,\xi)}{v(x,\xi)}} \right| \leq \max\{K_1,K_2\}\max\{\rho_1,\rho_2\}\bar{r}\|x'-x\|, \quad (5.35)$$

where ρ_1 and ρ_2 are the Lipschtiz modulus of functions $h(x,\xi)e^{-h(x,\xi)/v_1}$ and $h(x,\xi)e^{-h(x,\xi)/v_2}$ respectively. By combining (5.32)-(5.35), we have

$$|H(x',\xi) - H(x,\xi)| \leq \kappa_1\|x'-x\|, \quad (5.36)$$

where $\kappa_1 := \max\{\bar{r}, \max\{K_1,K_2\}\max\{\rho_1,\rho_2\}\bar{r}\}$ is deterministic and bounded.

Part (ii). Let $X(\xi) := \{x \in X : h(x,\xi) = 0\}$. From the definition of $K(x,\xi)$ and $v(x,\xi)$, $H(x,\xi)$ is continuously differentiable on $X\backslash X(\xi)$. Let $\hat{x} \in X(\xi)$ and $B(\hat{x},\delta)$ be a closed ball in $\mathbb{R}^m$ with radius δ and center $\hat{x}$. Let $S_1 = B(\hat{x},\delta)\bigcap\{x \in X : h(x,\xi) \geq 0$. Then for $x \in S_1$,

$$H(x,\xi) = h(x,\xi) - T(h(x,\xi)) = h(x,\xi)\left((1 - K_1) + K_1 e^{-h(x,\xi)/v_1}\right),$$

and

$$\nabla_x H(x,\xi) = \nabla_x h(x,\xi)\left(1 - K_1 + K_1 e^{-h(x,\xi)/v_1} - \frac{K_1 h}{v_1}e^{-h(x,\xi)/v_1}\right).$$

Letting $\delta \to 0$, we have $\nabla_x H(x,\xi) \to \nabla_x h(\hat{x},\xi)(1 - K_1 + K_1) = wr^T(\xi)$. On the other hand, letting $S_2 = B(\hat{x},\delta)\bigcap\{x \in X : h(x,\xi) \leq 0\}$, we have

$$H(x,\xi) = h(x,\xi) - T(h(x,\xi)) = h(x,\xi)\left((1 - K_2) + K_2 e^{-h(x,\xi)/v_1}\right).$$

for $x \in S_2$ and

$$\nabla_x H(x,\xi) = \nabla_x h(x,\xi)\left(1 - K_1 + K_1 e^{-h(x,\xi)/v_1} - \frac{Kh}{v}e^{-h(x,\xi)/v_1}\right).$$

Letting $\delta \to 0$, we have $\nabla_x H(x,\xi) \to \nabla_x h(\hat{x},\xi)(1 - K_2 + K_2) = wr^T(\xi)$. This shows that $\nabla_x H(\hat{x},\xi) = wr^T(\xi)$ and hence $\nabla_x H(x,\xi)$ is Lipschitz continuous w.r.t x. The proof is complete.

∎