Available online at www.sciencedirect.com
SCIENCE DIRECT®
Operations Research Letters 34 (2006) 1–8
Operations Research Letters
www.elsevier.com/locate/orl

On complexity of multistage stochastic programs

Alexander Shapiro*

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

Received 5 January 2005; accepted 16 February 2005
Available online 31 May 2005

Abstract

In this paper we derive estimates of the sample sizes required to solve a multistage stochastic programming problem with a given accuracy by the (conditional sampling) sample average approximation method. The presented analysis is self-contained and is based on a relatively elementary, one-dimensional, Cramér's Large Deviations Theorem.

Keywords: Stochastic programming; Monte Carlo sampling; Sample average method; Large deviations exponential bounds; Complexity

1. Introduction

Consider the following stochastic programming problem:

$$\operatorname{Min}_{x\in\mathscr{X}} \{f(x) := \mathbb{E}[F(x,\xi)]\}, \tag{1}$$

where ξ is a random vector supported on a set $\Xi \subset \mathbb{R}^d$, the expectation in (1) is taken with respect to a (known) probability distribution of ξ, $\mathscr{X}$ is a nonempty subset of $\mathbb{R}^n$ and $F : \mathscr{X} \times \Xi \to \mathbb{R}$. In the case of two-stage stochastic programming, the function $F(x,\xi)$ is given as the optimal value of a corresponding second-stage problem. In that case the assumption that $F(x,\xi)$ is real valued for all $x \in \mathscr{X}$ and $\xi \in \Xi$ can only hold if the corresponding recourse is relatively complete.

Only in very specific situations the expected value function $f(x)$ can be written in a closed form. Therefore, it should be calculated by a numerical integration. Already for the number of random variables $d \geqslant 5$, it is typically impossible to evaluate the corresponding multidimensional integral (expectation) with a high accuracy. This makes stochastic programming problems of the form (1) really difficult. A way of estimating the expected value function is suggested by the Monte Carlo method. That is, a random sample $\xi^1, \ldots, \xi^N$ of N realizations of ξ is generated, and the expected value function $f(x)$ is approximated by the sample average function $\hat{f}_N(x) := N^{-1}\sum_{i=1}^N F(x,\xi^i)$. This is the basic idea of the so-called sample average approximation (SAA) method.

It is possible to show, under mild regularity conditions, that for $\varepsilon > 0$ and $\alpha \in (0,1)$ the sample size

$$N \geqslant \frac{O(1)\sigma^2}{\varepsilon^2}\left[n\log\left(\frac{DL}{\varepsilon}\right) + \log\left(\frac{O(1)}{\alpha}\right)\right] \tag{2}$$

* Tel.: +1 404 894 6544; fax: +1 404 894 2301.
E-mail address: ashapiro@isye.gatech.edu.

0167-6377/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.orl.2005.02.003

guarantees that any $\varepsilon/2$-optimal solution of the SAA problem is an ε-optimal solution of the true problem with probability at least $1-\alpha$ (see [3,7,8]). Here $O(1)$ is a generic constant, D is the diameter of the set $\mathscr{X}$ (assumed to be finite), L is a Lipschitz constant of $f(x)$ and σ^2 is a certain constant measuring variability of the objective function $F(x,\xi)$. Recall that for $\varepsilon>0$ it is said that $\bar{x}$ is an ε-optimal solution of problem (1) if $\bar{x}\in\mathscr{X}$ and $f(\bar{x})\leqslant\inf_{x\in\mathscr{X}} f(x)+\varepsilon$. In a sense estimate (2) of the sample size gives a bound on the complexity of solving, with a specified probability, the (true) problem (1) by using the sample average approximation. Note that the estimated sample size grows linearly in the dimension n of the first-stage problem and is proportional to the squared ratio of the variability coefficient σ to the desired accuracy ε. (The following Example 1 shows that this estimate cannot be significantly improved.) This indicates that one may expect to solve the true problem (1) with a manageable sample size to a reasonable accuracy by using the SAA method. And, indeed, this was verified in various numerical experiments (cf., [4,5,9]).

Example 1. Consider problem (1) with $F(x,\xi):=\|x\|^{2k}-2k\langle\xi,x\rangle$, where k is a positive integer, $\mathscr{X}:=\{x\in\mathbb{R}^n:\|x\|\leqslant 1\}$, $\langle x,y\rangle$ denotes the standard scalar product of two vectors $x,y\in\mathbb{R}^n$ and $\|x\|=\sqrt{\langle x,x\rangle}$. Suppose, further, that the random vector ξ has the normal distribution $N(0,\sigma^2 I_n)$, where σ^2 is a positive constant, i.e., components ξ_i of ξ are independent and $\xi_i\sim N(0,\sigma^2)$, $i=1,\ldots,n$. It follows that $f(x)=\|x\|^{2k}$, and hence for $\varepsilon\in[0,1]$ the set of ε-optimal solutions of the true problem (1) is $\{x:\|x\|^{2k}\leqslant\varepsilon\}$. Now let $\xi^1,\ldots,\xi^N$ be an iid random sample of ξ and $\bar{\xi}_N:=(\xi^1+\cdots+\xi^N)/N$. The corresponding sample average function is $\hat{f}_N(x)=\|x\|^{2k}-2k\langle\bar{\xi}_N,x\rangle$, and the optimal solution $\hat{x}_N$ of the SAA problem is $\hat{x}_N=\|\bar{\xi}_N\|^{-\gamma}\bar{\xi}_N$, where $\gamma:=(2k-2)/(2k-1)$ if $\|\bar{\xi}_N\|\leqslant 1$, and $\gamma=1$ if $\|\bar{\xi}_N\|>1$. It follows that, for $\varepsilon\in(0,1)$, the optimal solution of the corresponding SAA problem is an ε-optimal solution of the true problem iff $\|\bar{\xi}_N\|^{v}\leqslant\varepsilon$, where $v:=2k/(2k-1)$. We have that $\bar{\xi}_N\sim N(0,\sigma^2N^{-1}I_n)$, and hence $N\|\bar{\xi}_N\|^2/\sigma^2$ has the chi-square distribution with n degrees of freedom. Consequently, the probability that $\|\bar{\xi}_N\|^{v}>\varepsilon$ is equal to the probability $\mathbb{P}(\chi^2_n>N\varepsilon^{2/v}/\sigma^2)$. Moreover, $\mathbb{E}[\chi^2_n]=n$ and the probability $\mathbb{P}(\chi^2_n>n)$ increases and tends to $1/2$ as n increases, e.g., $\mathbb{P}(\chi^2_1>1)=0.3173$, $\mathbb{P}(\chi^2_2>2)=0.3679$, $\mathbb{P}(\chi^2_3>3)=0.3916$, etc. Consequently, for $\alpha\in(0,0.3)$ and $\varepsilon\in(0,1)$, for example, the sample size N should satisfy

$$N>\frac{n\sigma^2}{\varepsilon^{2/v}} \tag{3}$$

in order to have the following property: "with probability $1-\alpha$ an (exact) optimal solution of the SAA problem is an ε-optimal solution of the true problem". Compared with (2), the lower bound (3) also grows linearly in n and is proportional to $\sigma^2/\varepsilon^{2/v}$. It remains to note that the constant v decreases to one as k increases.

The aim of this paper is to extend this analysis of the SAA method to the multistage stochastic programming (MSP) setting. A discussion of complexity of the MSP can be found in [8]. It was already argued there that the complexity of the SAA method, when applied to the MSP, grows fast with increase of the number of stages and seemingly simple MSP problems can be computationally unmanageable. We estimate sample sizes, required to solve the true problem with a given accuracy, by using tools of the large deviations (LD) theory (see, e.g., [2] for a thorough discussion of the LD theory). In that respect our analysis is self-contained and rather elementary since we only employ the upper bound of the (one-dimensional) Cramér's LD Theorem. That is, if $X_1,\ldots,X_N$ is a sequence of iid realizations of a random variable X and $\bar{X}_N:=N^{-1}\sum_{i=1}^N X_i$ is the corresponding average, then

$$\mathbb{P}(\bar{X}_N\geqslant a)\leqslant \mathrm{e}^{-NI(a)}. \tag{4}$$

Here $\mathbb{P}(A)$ denotes probability of event A,

$$I(z):=\sup_{t\in\mathbb{R}}\{tz-\log M(t)\}$$

is the so-called rate function, and $M(t):=\mathbb{E}[\mathrm{e}^{tX}]$ is the moment generating function of the random variable X.

Let us make the following simple observation which will be used in our derivations. Let X and Y be two random variables and $\varepsilon\in\mathbb{R}$. We have that if $X\leqslant\varepsilon_1$ and $Y\leqslant\varepsilon_2$, where $\varepsilon_1+\varepsilon_2=\varepsilon$, then $X+Y\leqslant\varepsilon$, and hence

$$\{\omega: X(\omega)+Y(\omega)>\varepsilon\}\subset\{\omega: X(\omega)>\varepsilon_1\}\cup\{\omega: Y(\omega)>\varepsilon_2\}.$$

This implies the following inequality for the corresponding probabilities:

$$\mathbb{P}(X+Y>\varepsilon)\leqslant \mathbb{P}(X>\varepsilon_1)+\mathbb{P}(Y>\varepsilon_2). \tag{5}$$

2. Sample average approximations of multistage stochastic programs

Consider the following T-stage stochastic programming problem:

$$\min_{x_1\in\mathscr{X}_1} F_1(x_1)+\mathbb{E}\left[\inf_{x_2\in\mathscr{X}_2(x_1,\xi_2)} F_2(x_2,\xi_2) +\mathbb{E}\left[\cdots+\mathbb{E}\left[\inf_{x_T\in\mathscr{X}_T(x_{T-1},\xi_T)} F_T(x_T,\xi_T)\right]\right]\right] \tag{6}$$

driven by the random data process $\xi_2,\ldots,\xi_T$. Here $x_t\in\mathbb{R}^{n_t}$, $t=1,\ldots,T$, are decision variables, $F_t:\mathbb{R}^{n_t}\times\mathbb{R}^{d_t}\to\mathbb{R}$ are continuous functions and $\mathscr{X}_t:\mathbb{R}^{n_{t-1}}\times\mathbb{R}^{d_t}\rightrightarrows\mathbb{R}^{n_t}$, $t=2,\ldots,T$, are measurable multifunctions, the function $F_1:\mathbb{R}^{n_1}\to\mathbb{R}$ and the set $\mathscr{X}_1\subset\mathbb{R}^{n_1}$ are deterministic. We assume that the set $\mathscr{X}_1$ is nonempty. For example, in the linear case $F_t(x_t,\xi_t):=\langle c_t,x_t\rangle$, $\mathscr{X}_1:=\{x_1: A_1x_1=b_1,\ x_1\geqslant 0\}$,

$$\mathscr{X}_t(x_{t-1},\xi_t):=\{x_t: B_tx_{t-1}+A_tx_t=b_t,\ x_t\geqslant 0\},\quad t=2,\ldots,T,$$

$\xi_1:=(c_1,A_1,b_1)$ is known at the first stage (and hence is nonrandom), and $\xi_t:=(c_t,B_t,A_t,b_t)\in\mathbb{R}^{d_t}$, $t=2,\ldots,T$, are data vectors, some (all) elements of which can be random. In the sequel we use ξ_t to denote the random data vector and its particular realization. Which one of these two meanings will be used in a particular situation will be clear from the context.

If we denote by $Q_2(x_1,\xi_2)$ the optimal value of the $(T-1)$-stage problem

$$\min_{x_2\in\mathscr{X}_2(x_1,\xi_2)} F_2(x_2,\xi_2) +\mathbb{E}\left[\cdots+\mathbb{E}\left[\min_{x_T\in\mathscr{X}_T(x_{T-1},\xi_T)} F_T(x_T,\xi_T)\right]\right], \tag{7}$$

then we can write the T-stage problem (6) in the following form of a two-stage programming problem:

$$\min_{x_1\in\mathscr{X}_1} F_1(x_1)+\mathbb{E}[Q_2(x_1,\xi_2)]. \tag{8}$$

Note, however, that if $T>2$, then problem (7) in itself is a stochastic programming problem. Consequently, if the number of scenarios involved in (7) is very large, or infinite, then the optimal value $Q_2(x_1,\xi_2)$ can be calculated only approximately, say by sampling.

For the sake of simplicity we make the following derivations for the three-stage problem, i.e., we assume that $T=3$ (it will be clear how the obtained results can be extended to an analysis of $T>3$). In that case $Q_2(x_1,\xi_2)$ is given by the optimal value of the problem

$$\min_{x_2\in\mathscr{X}_2(x_1,\xi_2)} F_2(x_2,\xi_2)+\mathbb{E}[Q_3(x_2,\xi_3)|\xi_2], \tag{9}$$

where the expectation is taken with respect to the conditional distribution of ξ_3 given ξ_2 and

$$Q_3(x_2,\xi_3):=\inf_{x_3\in\mathscr{X}_3(x_2,\xi_3)} F_3(x_3,\xi_3).$$

We make the following assumption:

- For every $x_1\in\mathscr{X}_1$ the expectation $\mathbb{E}[Q_2(x_1,\xi_2)]$ is well defined and finite valued.

Of course, finite valuedness of $\mathbb{E}[Q_2(x_1,\xi_2)]$ can only hold if $Q_2(x_1,\xi_2)$ is finite valued for a.e. ξ_2, which in turn implies that $\mathscr{X}_2(x_1,\xi_2)$ is nonempty for a.e. ξ_2, etc. That is, the above assumption implies that the recourse is *relatively complete*.

Now let ξ_2^i, $i=1,\ldots,N_1$, be a random sample of independent realizations of the random vector ξ_2. We can approximate problem (8) by the following SAA problem:

$$\min_{x_1\in\mathscr{X}_1}\left\{\hat{f}_{N_1}(x_1):=F_1(x_1)+\frac{1}{N_1}\sum_{i=1}^{N_1}Q_2(x_1,\xi_2^i)\right\}. \tag{10}$$

Since $Q_2(x_1,\xi_2^i)$ are not given explicitly, we need to estimate these values by *conditional sampling* (note that in order for the SAA method to produce *consistent* estimators, conditional sampling is required, see [6]). That is, we generate random sample ξ_3^{ij}, $j=1,\ldots,N_2$, of N_2 independent realizations according to the conditional distribution of ξ_3 given ξ_2^i, $i=1,\ldots,N_1$.

Consequently, we approximate $Q_2(x_1, \xi_2^i)$ by

$$\hat{Q}_{2,N_2}(x_1, \xi_2^i) := \inf_{x_2 \in \mathscr{X}_2(x_1, \xi_2^i)} \left\{ F_2(x_2, \xi_2^i) + \frac{1}{N_2} \sum_{j=1}^{N_2} Q_3(x_2, \xi_3^{ij}) \right\}. \tag{11}$$

Finally, we approximate the true (expected value) problem (8) by the following so-called SAA problem:

$$\operatorname*{Min}_{x_1 \in \mathscr{X}_1} \left\{ \tilde{f}_{N_1,N_2}(x_1) := F_1(x_1) + \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2,N_2}(x_1, \xi_2^i) \right\}. \tag{12}$$

The above SAA problem is obtained by approximating the objective function

$$f(x_1) := F_1(x_1) + \mathbb{E}[Q_2(x_1, \xi_2)]$$

of problem (8) with $\tilde{f}_{N_1,N_2}(x_1)$.

3. Sample size estimates

In order to proceed with our analysis, we need to estimate the probability

$$\mathbb{P}\left\{ \sup_{x_1 \in \mathscr{X}_1} |f(x_1) - \tilde{f}_{N_1,N_2}(x_1)| > \varepsilon \right\} \tag{13}$$

for an arbitrary constant $\varepsilon > 0$. To this end we use the following result about LD bounds for the uniform convergence of sample average approximations.

Consider a function $h : \mathscr{X} \times \Xi \to \mathbb{R}$ and the corresponding expected value function $\phi(x) := \mathbb{E}[h(x, \xi)]$, where the expectation is taken with respect to the probability distribution P of the random vector $\xi = \xi(\omega)$, $\mathscr{X}$ is a nonempty closed subset of $\mathbb{R}^n$ and $\Xi \subset \mathbb{R}^d$ is the support of the probability distribution P. Assume that for every $x \in \mathscr{X}$ the expectation $\phi(x)$ is well defined, i.e., $h(x, \cdot)$ is measurable and P-integrable. Let $\xi^1, \ldots, \xi^N$ be an iid sample of the random vector $\xi(\omega)$, and $\hat{\phi}_N(x) := (1/N)\sum_{i=1}^N h(x, \xi^i)$ be the corresponding sample average function.

Theorem 1. *Suppose that the set $\mathscr{X}$ has finite diameter D, and the following conditions hold*: (i) *there exists a constant $\sigma > 0$ such that*

$$M_x(t) \leqslant \exp\{\sigma^2 t^2/2\}, \quad \forall t \in \mathbb{R}, \ \forall x \in \mathscr{X}, \tag{14}$$

where $M_x(t)$ is the moment generating function of the random variable $h(x, \xi) - \phi(x)$,

(ii) *there exists a constant $L > 0$ such that*

$$|h(x', \xi) - h(x, \xi)| \leqslant L\|x' - x\|, \quad \forall \xi \in \Xi, \ \forall x', x \in \mathscr{X}. \tag{15}$$

Then for any $\varepsilon > 0$,

$$\mathbb{P}\left\{ \sup_{x \in \mathscr{X}} |\hat{\phi}_N(x) - \phi(x)| \geqslant \varepsilon \right\} \leqslant O(1)\left(\frac{DL}{\varepsilon}\right)^n \exp\left\{ -\frac{N\varepsilon^2}{16\sigma^2} \right\}. \tag{16}$$

To make the paper self-contained we give a proof of this theorem in the appendix.

We can apply the LD bound (16) to obtain estimates of the probability (13). We have that

$$\begin{aligned} &\sup_{x_1 \in \mathscr{X}_1} |f(x_1) - \tilde{f}_{N_1,N_2}(x_1)| \\ &\quad \leqslant \sup_{x_1 \in \mathscr{X}_1} |f(x_1) - \hat{f}_{N_1}(x_1)| \\ &\qquad + \sup_{x_1 \in \mathscr{X}_1} |\hat{f}_{N_1}(x_1) - \tilde{f}_{N_1,N_2}(x_1)|, \end{aligned}$$

and hence, by (5),

$$\begin{aligned} &\mathbb{P}\left\{ \sup_{x_1 \in \mathscr{X}_1} |f(x_1) - \tilde{f}_{N_1,N_2}(x_1)| > \varepsilon \right\} \\ &\quad \leqslant \mathbb{P}\left\{ \sup_{x_1 \in \mathscr{X}_1} |f(x_1) - \hat{f}_{N_1}(x_1)| > \varepsilon/2 \right\} \\ &\qquad + \mathbb{P}\left\{ \sup_{x_1 \in \mathscr{X}_1} |\hat{f}_{N_1}(x_1) - \tilde{f}_{N_1,N_2}(x_1)| > \varepsilon/2 \right\}. \end{aligned} \tag{17}$$

Note that

$$f(x_1) - \hat{f}_{N_1}(x_1) = \mathbb{E}[Q_2(x_1, \xi_2)] - \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1, \xi_2^i)$$

and

$$\hat{f}_{N_1}(x_1) - \tilde{f}_{N_1,N_2}(x_1) = \frac{1}{N_1} \sum_{i=1}^{N_1} [Q_2(x_1, \xi_2^i) - \hat{Q}_{2,N_2}(x_1, \xi_2^i)].$$

Let us assume, for the sake of simplicity, the *between stages independence* of the random process. That is, suppose that the following condition holds:

(A1) The random vectors ξ_2 and ξ_3 are independent.

Of course, under this condition the conditional expectation in formula (9) does not depend on ξ_2. Also in that case the conditional sample ξ_3^{ij} has the (marginal) distribution of ξ_3 and is independent of ξ_2^i, and can be generated in two ways. Namely, we can either generate the same random sample ξ_3^{ij} for each $i=1,\ldots,N_1$, or these samples can be generated independently of each other.

Let us further make the following assumptions.

(A2) The set $\mathscr{X}_1$ has finite diameter D_1.

(A3) There is a constant $L_1>0$ such that

$$|Q_2(x_1', \xi_2) - Q_2(x_1, \xi_2)| \leqslant L_1\|x_1' - x_1\| \quad (18)$$

for all $x_1', x_1 \in \mathscr{X}_1$ and a.e. ξ_2.

(A4) There exists a constant $\sigma_1>0$ such that for any $x_1 \in \mathscr{X}_1$ it holds that

$$M_{1,x_1}(t) \leqslant \exp\{\sigma_1^2 t^2/2\}, \quad \forall t \in \mathbb{R}, \quad (19)$$

where $M_{1,x_1}(t)$ is the moment generating function of $Q_2(x_1, \xi_2) - \mathbb{E}[Q_2(x_1, \xi_2)]$.

(A5) There is a positive constant D_2 such that for every $x_1 \in \mathscr{X}_1$ and a.e. ξ_2 the set $\mathscr{X}_2(x_1, \xi_2)$ has a finite diameter less than or equal to D_2.

(A6) There is a constant $L_2>0$ such that

$$|F_2(x_2', \xi_2) - F_2(x_2, \xi_2) + Q_3(x_2', \xi_3) - Q_3(x_2, \xi_3)| \leqslant L_2\|x_2' - x_2\| \quad (20)$$

for all $x_2', x_2 \in \mathscr{X}_2(x_1, \xi_2)$, $x_1 \in \mathscr{X}_1$ and a.e. ξ_2 and ξ_3.

(A7) There exists a constant $\sigma_2>0$ such that for any $x_2 \in \mathscr{X}_2(x_1, \xi_2)$ and all $x_1 \in \mathscr{X}_1$ and a.e. ξ_2 it holds that

$$M_{2,x_2}(t) \leqslant \exp\{\sigma_2^2 t^2/2\}, \quad \forall t \in \mathbb{R}, \quad (21)$$

where $M_{2,x_2}(t)$ is the moment generating function of $Q_3(x_2, \xi_3) - \mathbb{E}[Q_3(x_2, \xi_3)]$.

By Theorem 1, under assumptions (A2)–(A4), we have that

$$\mathbb{P}\left\{\sup_{x_1\in\mathscr{X}_1} |f(x_1) - \hat{f}_{N_1}(x_1)| > \varepsilon/2\right\} \leqslant O(1)\left(\frac{D_1L_1}{\varepsilon}\right)^{n_1} \exp\left\{-\frac{O(1)N_1\varepsilon^2}{\sigma_1^2}\right\}. \quad (22)$$

For ξ_2 and the random sample $\xi_3^1,\ldots,\xi_3^{N_2}$ of N_2 independent replications of ξ_3, consider the function

$$\hat{\psi}_{N_2}(x_2, \xi_2) := F_2(x_2, \xi_2) + \frac{1}{N_2}\sum_{j=1}^{N_2} Q_3(x_2, \xi_3^j)$$

and its expected value

$$\psi(x_2, \xi_2) = F_2(x_2, \xi_2) + \mathbb{E}[Q_3(x_2, \xi_3)].$$

By Theorem 1, under assumptions (A5)–(A7), we have that for any $x_1 \in \mathscr{X}_1$,

$$\mathbb{P}\left\{\sup_{x_2\in\mathscr{X}_2(x_1,\xi_2)} |\hat{\psi}_{N_2}(x_2, \xi_2) - \psi(x_2, \xi_2)| > \varepsilon/2\right\} \leqslant C_\varepsilon(N_2), \quad (23)$$

where

$$C_\varepsilon(N_2) = O(1)\left(\frac{D_2L_2}{\varepsilon}\right)^{n_2} \exp\left\{-\frac{O(1)N_2\varepsilon^2}{\sigma_2^2}\right\}.$$

It follows that

$$\mathbb{P}\left\{\left|\inf_{x_2\in\mathscr{X}_2(x_1,\xi_2)} \hat{\psi}_{N_2}(x_2, \xi_2) - \inf_{x_2\in\mathscr{X}_2(x_1,\xi_2)} \psi(x_2, \xi_2)\right| > \varepsilon/2\right\} \leqslant C_\varepsilon(N_2). \quad (24)$$

Note that

$$\inf_{x_2\in\mathscr{X}_2(x_1,\xi_2)} \psi(x_2, \xi_2) = Q_2(x_1, \xi_2),$$

and for $\xi_2 = \xi_2^i$,

$$\inf_{x_2\in\mathscr{X}_2(x_1,\xi_2)} \hat{\psi}_{N_2}(x_2, \xi_2) = \hat{Q}_{2,N_2}(x_1, \xi_2^i).$$

It follows from (24) that (for both strategies of using the same or independent samples for each ξ_2^i) the following inequality holds:

$$\mathbb{P}\{|\hat{f}_{N_1}(x_1) - \tilde{f}_{N_1,N_2}(x_1)| > \varepsilon/2\} \leqslant C_\varepsilon(N_2). \quad (25)$$

Suppose further that:

• There is $L_3 > 0$ such that $\hat{f}_{N_1}(\cdot) - \tilde{f}_{N_1,N_2}(\cdot)$ is Lipschitz continuous on $\mathscr{X}_1$ with constant L_3.

Then by constructing a ν-net in $\mathscr{X}_1$ and using (25) it can be shown (in a way similar to the proof of Theorem 1 in the Appendix) that

$$\mathbb{P}\left\{\sup_{x_1\in\mathscr{X}_1} |\hat{f}_{N_1}(x_1) - \tilde{f}_{N_1,N_2}(x_1)| > \varepsilon/2\right\} \leqslant O(1)\left(\frac{D_1L_3}{\varepsilon}\right)^{n_1}\left(\frac{D_2L_2}{\varepsilon}\right)^{n_2} \times \exp\left\{-\frac{O(1)N_2\varepsilon^2}{\sigma_2^2}\right\}. \tag{26}$$

Combining (17) with estimates (22) and (26) gives an upper bound for the probability (13). Let us also observe that if $\hat{x}_1$ is an $\varepsilon/2$-optimal solution of the SAA problem (12) and $\sup_{x_1\in\mathscr{X}_1}|f(x_1)-\tilde{f}_{N_1,N_2}(x_1)|\leqslant\varepsilon/2$, then $\hat{x}_1$ is an ε-optimal solution of the true problem (8). Therefore, we obtain the following result.

Theorem 2. *Under the specified assumptions and for $\varepsilon > 0$ and $\alpha \in (0, 1)$, and the sample sizes N_1 and N_2 satisfying*

$$O(1)\left[\left(\frac{D_1L_1}{\varepsilon}\right)^{n_1}\exp\left\{-\frac{O(1)N_1\varepsilon^2}{\sigma_1^2}\right\} + \left(\frac{D_1L_3}{\varepsilon}\right)^{n_1}\left(\frac{D_2L_2}{\varepsilon}\right)^{n_2} \times \exp\left\{-\frac{O(1)N_2\varepsilon^2}{\sigma_2^2}\right\}\right] \leqslant \alpha, \tag{27}$$

we have that any $\varepsilon/2$-optimal solution of the SAA problem (12) *is an ε-optimal solution of the true problem* (8) *with probability at least $1-\alpha$.*

In particular, suppose that $N_1 = N_2$. Then for $L := \max\{L_1, L_2, L_3\}$, $D := \max\{D_1, D_2\}$ and $\sigma := \max\{\sigma_1, \sigma_2\}$ we can use the following estimate of the required sample size $N_1 = N_2$:

$$O(1)\left(\frac{DL}{\varepsilon}\right)^{n_1+n_2}\exp\left\{-\frac{O(1)N_1\varepsilon^2}{\sigma^2}\right\} \leqslant \alpha, \tag{28}$$

which is equivalent to

$$N_1 \geqslant \frac{O(1)\sigma^2}{\varepsilon^2}\left[(n_1+n_2)\log\left(\frac{DL}{\varepsilon}\right) + \log\left(\frac{O(1)}{\alpha}\right)\right]. \tag{29}$$

4. Discussion

Estimate (29), for three-stage programs, looks similar to estimate (2) for two-stage programs. Note, however, that if we use the SAA method with conditional sampling and respective sample sizes N_1 and N_2, then the total number of scenarios is $N = N_1N_2$. Therefore, our analysis seems to indicate that for three-stage problems we need random samples with the total number of scenarios of the order of the *square* of the corresponding sample size for two-stage problems. This analysis can be extended to T-stage problems with the conclusion that the total number of scenarios needed to solve the true problem with a reasonable accuracy grows *exponentially* with increase of the number of stages T. Some numerical experiments seem to confirm this conclusion (cf. [1]). Of course, it should be mentioned that the above analysis does *not* prove in a rigorous mathematical sense that complexity of multistage programming grows exponentially with increase of the number of stages. It only indicates that the SAA method, which showed a considerable promise for solving two-stage problems, could be practically inapplicable for solving multistage problems with a large (say greater than 5) number of stages.

Our analysis was performed under several simplifying assumptions. In particular, we considered a three-stage setting and assumed the between-stages independence condition. An extension of the analysis from 3 to a higher number of stages is straightforward. Removing the between-stages independence assumption may create technical difficulties and requires a further investigation. The Lipschitz constant L_1 in (18) is assumed to be independent of the random vector ξ_2. If the distribution of ξ_2 has a bounded (compact) support Ξ_2 and $Q_2(\cdot, \xi_2)$ has a Lipschitz constant $L(\xi_2)$, which is a continuous function of ξ_2, then one can take L_1 to be the maximum of $L(\xi_2)$ over $\xi_2 \in \Xi_2$. A similar remark applies to the Lipschitz constant in (20). Condition (14), for the moment generating function,

means that the tails of the distribution of the random variable $h(x,\xi)$ are sufficiently light. This holds, in particular, if the distribution of ξ has a bounded support and $h(x,\cdot)$ is continuous. If $h(x,\xi)$ has a normal distribution, then (14) holds with σ^2 being the variance of $h(x,\xi)$. Similar remarks can be said for assumptions (A4) and (A7).

5. Appendix

Proof of Theorem 1. By the LD bound (4) we have that for any $x \in \mathscr{X}$ and $\varepsilon > 0$ it holds that

$$\mathbb{P}\{\hat{\phi}_N(x) - \phi(x) \geqslant \varepsilon\} \leqslant \exp\{-NI_x(\varepsilon)\}, \tag{30}$$

where

$$I_x(z) := \sup_{t\in\mathbb{R}}\{zt - \log M_x(t)\} \tag{31}$$

is the rate function of the random variable $h(x,\xi) - \phi(x)$. Similarly,

$$\mathbb{P}\{\hat{\phi}_N(x) - \phi(x) \leqslant -\varepsilon\} \leqslant \exp\{-NI_x(-\varepsilon)\},$$

and hence

$$\begin{aligned}&\mathbb{P}\{|\hat{\phi}_N(x) - \phi(x)| \geqslant \varepsilon\}\\ &\quad\leqslant \exp\{-NI_x(\varepsilon)\} + \exp\{-NI_x(-\varepsilon)\}.\end{aligned} \tag{32}$$

For a constant $v > 0$, let $\bar{x}_1, \ldots, \bar{x}_M \in \mathscr{X}$ be such that for every $x \in \mathscr{X}$ there exists $\bar{x}_\ell$, $\ell \in \{1,\ldots,M\}$, such that $\|x - \bar{x}_\ell\| \leqslant v$. Such a set $\{\bar{x}_1,\ldots,\bar{x}_M\}$ is called a v-net in $\mathscr{X}$. We can choose this net in such a way that $M \leqslant O(1)(D/v)^n$, where $D := \sup_{x',x\in\mathscr{X}}\|x' - x\|$ is the diameter of $\mathscr{X}$ and $O(1)$ is a generic constant. By (15) we have that

$$|\phi(x') - \phi(x)| \leqslant L\|x' - x\| \tag{33}$$

and

$$|\hat{\phi}_N(x') - \hat{\phi}_N(x)| \leqslant L\|x' - x\| \tag{34}$$

for any $x, x' \in \mathscr{X}$. It follows by (32) that

$$\begin{aligned}&\mathbb{P}\left(\max_{1\leqslant\ell\leqslant M}|\hat{\phi}_N(\bar{x}_\ell) - \phi(\bar{x}_\ell)| \geqslant \varepsilon\right)\\ &\quad\leqslant \mathbb{P}\left(\bigcup_{\ell=1}^{M}\{|\hat{\phi}_N(\bar{x}_\ell) - \phi(\bar{x}_\ell)| \geqslant \varepsilon\}\right)\\ &\quad\leqslant \sum_{\ell=1}^{M}\mathbb{P}(|\hat{\phi}_N(\bar{x}_\ell) - \phi(\bar{x}_\ell)| \geqslant \varepsilon)\\ &\quad\leqslant 2\sum_{\ell=1}^{M}\exp\{-N[I_{\bar{x}_\ell}(\varepsilon) \wedge I_{\bar{x}_\ell}(-\varepsilon)]\}.\end{aligned} \tag{35}$$

For an $x \in \mathscr{X}$ consider $\ell(x) \in \arg\min_{1\leqslant\ell\leqslant M}\|x - \bar{x}_\ell\|$. By construction of the v-net we have that $\|x - \bar{x}_{\ell(x)}\| \leqslant v$ for every $x \in \mathscr{X}$. Then

$$\begin{aligned}|\hat{\phi}_N(x) - \phi(x)| &\leqslant |\hat{\phi}_N(x) - \hat{\phi}_N(\bar{x}_{\ell(x)})| + |\hat{\phi}_N(\bar{x}_{\ell(x)})\\ &\quad - \phi(\bar{x}_{\ell(x)})| + |\phi(\bar{x}_{\ell(x)}) - \phi(x)|\\ &\leqslant Lv + |\hat{\phi}_N(\bar{x}_{\ell(x)}) - \phi(\bar{x}_{\ell(x)})| + Lv.\end{aligned}$$

Let us now take a v-net with such a v that $Lv = \varepsilon/4$, i.e., $v := [\varepsilon/(4L)]$. Then

$$\begin{aligned}&\mathbb{P}\left\{\sup_{x\in\mathscr{X}}|\hat{\phi}_N(x) - \phi(x)| \geqslant \varepsilon\right\}\\ &\quad\leqslant \mathbb{P}\left\{\max_{1\leqslant\ell\leqslant M}|\hat{\phi}_N(\bar{x}_\ell) - \phi(\bar{x}_\ell)| \geqslant \varepsilon/2\right\},\end{aligned}$$

which together with (35) implies that

$$\begin{aligned}&\mathbb{P}\left\{\sup_{x\in\mathscr{X}}|\hat{\phi}_N(x) - \phi(x)| \geqslant \varepsilon\right\}\\ &\quad\leqslant 2\sum_{\ell=1}^{M}\exp\{-N[I_{\bar{x}_\ell}(\varepsilon/2) \wedge I_{\bar{x}_\ell}(-\varepsilon/2)]\}.\end{aligned} \tag{36}$$

Moreover, because of condition (i) we have that $\log M_x(t) \leqslant \sigma^2t^2/2$, and hence

$$I_x(z) \geqslant \frac{z^2}{2\sigma^2}, \ \forall z \in \mathbb{R}. \tag{37}$$

It follows from (36) and (37) that

$$\begin{aligned}&\mathbb{P}\left\{\sup_{x\in\mathscr{X}}|\hat{\phi}_N(x) - \phi(x)| \geqslant \varepsilon\right\}\\ &\quad\leqslant 2M\exp\left\{\frac{-N\varepsilon^2}{16\sigma^2}\right\}.\end{aligned} \tag{38}$$

Finally, since $M \leqslant O(1)(D/v)^n = O(1)(DL/\varepsilon)^n$, we obtain that (38) implies (16), and hence the proof is complete. □

References

[1] J. Blomvall, A. Shapiro, Solving multistage asset investment problems by Monte Carlo based optimization, E-print available at: http://www.optimization-online.org, 2004.

[2] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, Springer, New York, NY, 1998.

[3] A.J. Kleywegt, A. Shapiro, T. Homem-De-Mello, The sample average approximation method for stochastic discrete optimization, SIAM J. Optim. 12 (2001) 479–502.

[4] J. Linderoth, A. Shapiro, S. Wright, The empirical behavior of sampling methods for stochastic programming, Annals Oper. Res. to appear.

[5] W.K. Mak, D.P. Morton, R.K. Wood, Monte Carlo bounding techniques for determining solution quality in stochastic programs, Oper. Res. Lett. 24 (1999) 47–56.

[6] A. Shapiro, Inference of statistical bounds for multistage stochastic programming problems, Math. Methods Oper. Res. 58 (2003) 57–68.

[7] A. Shapiro, Monte Carlo sampling methods. in: A. Rusczyński, A. Shapiro (Eds.), Stochastic Programming, Handbooks in Operations Research and Management Science, Vol. 10, North-Holland, Amsterdam, 2003.

[8] A. Shapiro, A. Nemirovski, On complexity of stochastic programming problems, E-print available at: http://www.optimization-online.org, 2004.

[9] B. Verweij, S. Ahmed, A.J. Kleywegt, G. Nemhauser, A. Shapiro, The sample average approximation method applied to stochastic routing problems: a computational study, Comput. Optim. Appl. 24 (2003) 289–333.