COMPUTATIONAL SCIENCE AND TECHNIQUES
Volume 2, Number 2, 2014, 289-312

Publisher: Klaipėda University
http://journals.ku.lt/index.php/CST
Online ISSN: 2029-9966
DOI: 10.15181/csat.v2i2.891

TWO-STAGE STOCHASTIC LINEAR PROGRAMMING BY A SERIES OF MONTE-CARLO ESTIMATORS

Kęstutis Žilinskas
University of Šiauliai
P. Višinskio st 25, Šiauliai, Lithuania
E-mail: kest.zil@gmail.com

In this paper a stochastic adaptive method has been developed to solve stochastic linear problems by a finite sequence of Monte-Carlo sampling estimators. The method is based on the adaptive regulation of the size of Monte-Carlo samples and a statistical termination procedure taking into consideration statistical modelling accuracy. Our approach distinguishes itself by the treatment of accuracy of the solution in a statistical manner, testing the hypothesis of optimality according to statistical criteria, and estimating confidence intervals of the objective and constraint functions. To avoid "jamming" or "zigzagging" solving a constraint problem we implement the ε–feasible direction approach. The proposed adjustment of a sample size, when it is taken inversely proportional to the square of the norm of the Monte-Carlo estimate of the gradient, guarantees convergence a. s. at a linear rate. The numerical study and examples in practice corroborate theoretical conclusions and show that the developed procedures make it possible to solve stochastic problems with sufficient accuracy by the means of an acceptable size of computations.

Introduction

In tasks of resource and finance planning, job scheduling management, various problem with nondeterministic parameters and various kind of uncertainty are often being faced. This uncertainty often is described by statistical probabilistic methods. These tasks are solving by stochastic linear and nonlinear methods. Two-stage or multistage stochastic linear problems are extension of the classic linear programming, when parameters of the problem may be random variables.

Stochastic linear programming was developed on basis of the linear programming at second half of 20th century by demand to solve technical, economic and financial problems. Linear programming problems couldn't evaluate uncertainty of planning parameters. Usage of random parameters in linear programming models leads to complicated nonlinear optimization problems, which usually couldn't be solved by direct nonlinear programming methods. Stochastic methods for solving stochastic problems must be developed and applied. These stochastic methods generalize deterministic linear and nonlinear programming methods.

Main problems for stochastic programming are complicated computation of precise values of the objective function and verification of the optimality of the solution. Solving of the stochastic linear problems under admissible accuracy is actual and imperfect investigated theoretical and practical problem.

Two stage stochastic linear programming can be applied at various fields: electric energy production, manpower management, portfolio management, logistics, analysis of the biological systems, etc.

Let us describe example of the power plant investment planning. An energetic concern must invest in a system of power plants to meet its current and future demand for electrical power. These plants were to be built for the first year only, and were expected to operate over the next 15 years. The budget for construction of power plants was $10 billion, which was to be allocated for four different types of plants: gas turbine, coal, nuclear power, and hydroelectric. Power plants were priced according to their electric capacity, measured in gigawatts. The objective was to find the power plant allocation which had minimized the sum of the investment cost and the expected value of the operating cost over 15 years. The operating cost was stochastic due to uncertainty in future demand and each year demand grows with some rate.

Stochastic programming deals with a class of optimization models in which some data may be subject to significant uncertainty. Such models are appropriate when data evolve over time and decisions have to be made prior to observing the entire data streams. Although widespread applicability of stochastic programming models has attracted considerable attention of researchers, stochastic linear models remain one of more challenging optimisation problems.

Methods based on approximation and decomposition are often applied to solve stochastic programming tasks (see, e.g., Ermolyev and Wets, 1988, Prekopa, 1995, Marti, 2005, etc.), however, they can lead to very large-scale problems, and, thus, require very large computational resources. Therefore the study of stochastic programming algorithms has led to alternative ways of approximating problems, some of which obey certain asymptotic properties. This reliance on approximations has prompted to study asymptotic convergence of solutions of approximate problems to a solution of original (see, Rubinstein and Shapiro, 1993, Sen, 2001, Marti, 2005, etc.), and consider adaptive methods for approximations (Shapiro and Homem-de-Mello, 1998, Higle and Sen, 1999). In this paper we have developed an adaptive approach for solving stochastic linear problems by the Monte-Carlo method based on asymptotic properties of Monte-Carlo sampling estimators. This approach is grounded on the treatment of a statistical simulation error in a statistical manner and the rule for iterative regulation of the size of Monte-Carlo samples (Sakalauskas, 2002, Sakalauskas, 2004).

Let us consider a two-stage stochastic optimization problem with a complete recourse:

$$F(x) = c \cdot x + E\left\{Q(x,\xi)\right\} \to \min_{x \in D \subset \Re_+^n} \tag{1}$$

subject to a feasible set

$$D = \left\{ x \middle| \; A \cdot x = b, \; x \in \Re_+^n \right\} \tag{2}$$

where

$$Q(x,\xi) = \min_y [q \cdot y \,|\, W \cdot y + T \cdot x \le h, y \in R_+^m] \tag{3}$$

vectors b, q, h and full rank matrices A, W, T are of appropriate dimensionality. Let the feasible set D be nonempty and bounded. Assume vectors q, h and matrices W, T be random in general, and, consequently, depending on an elementary event $\xi \in \Omega$ from a certain probability space (Ω,Σ,P). Thus, under uncertainty the modelled system operates in an environment of uncontrollable parameters, which are modelled using random variables. Hence, the performance of such a system can also be viewed as a random variable. Let the measure P be absolutely continuous and defined by a probability density function $p(x,\cdot): \Re^n \times \Omega \to \Re_+$, depending on the decision variable x in general. Moreover, assume that the solution of the second stage problem (3) and the values of the function Q almost surely (a. s.) exist and are bounded.

1. Stochastic differentiation and Monte-Carlo estimators

In this section we discuss some basic ideas applied in stochastic gradient search in two-stage programming with recourse. The procedures of gradient evaluation are often constructed by expressing a gradient as an expectation and then evaluating this expectation by the means of statistical simulation (see, e.g., Rubinstein and Shapiro, 1993, Marti, 1996, Shapiro, 2000, Sakalauskas, 2002).

We treat the problem (1) as an optimization problem of an expected value function subject to linear deterministic constraints. Thus, in general we deal with a nonlinear stochastic optimization problem:

$$F(x) \equiv Ef(x,\xi) \to \min_{x \in D \subset \Re_+^n} \tag{4}$$

where the objective function is expectation of the random function $f(x,\xi) = c \cdot x + Q(x,\xi)$, depending on random vector ξ, defined by the distribution density function $p(x,\cdot)$, and the feasible set $x \in D \subset \Re_+^n$ is a nonempty, bounded and convex linear set in general.

First, let us consider the expectation

$$F(x) = Ef(x,\xi) \equiv \int_{\Omega} f(x,y) \cdot p(x,y)dy \tag{5}$$

when function f and density function p are assumed differentiable with respect to x for any $x \in D \subset \Re_+^n$ and any $y \in S(x)$, $S(x) = \{y | p(x,y) > 0\}$ is a support of random vector.

Differentiability of integrals of such kind has been studied rather deeply, and a technique for stochastic differentiation to express such an objective function and its gradient both together as expectations in the same probability space (see, Rubinstein, 1983, Uriasyev, 1994, Shapiro and Homem-de-Mello, 1998, Prekopa, 1999, Ermolyev et al, 2003, etc.) exists. It is not difficult to see that the vector-column of the gradient of a function (5) could be expressed as (Sakalauskas, 2002)

$$\nabla_x F(x) = E\left(\nabla_x f(x,\xi) + (f(x,\xi) - f(x,E\xi))\cdot\nabla_x \ln p(x,\xi)\right) \equiv Eg(x,\xi) \tag{6}$$

where the generalized gradient

$$\nabla_x \ln p(x,y) = \begin{cases} \dfrac{1}{p(x,y)}\dfrac{dp(x,y)}{dx}, & y \in S(x), \\ 0, & y \notin S(x) \end{cases}$$

is defined and it is assumed that uniformly $\dfrac{1}{p(x,y)}\dfrac{dp(x,y)}{dx} < \infty,\ y \in S(x), x \in D$. We see that it is possible to express expectation and its gradient through a linear operator from the same probability space. Hence, operators (5) and (6) can be estimated by the means of the same Monte-Carlo sample. Solving the problem (4) suppose it is possible to get finite sequences of realizations (trials) of ξ at any point $x \in D \subset \Re_+^n$. Hence, here we assume that Monte-Carlo samples of a certain size N are provided for any:

$$Y = (y^1, y^2, ..., y^N), \tag{7}$$

where y^i are independent random variables identically distributed at density $p(x,\cdot)$. Sampling estimator of the objective function is very simple and can be computed as:

$$\tilde{F}(x) = \frac{1}{N}\sum_{j=1}^{N} f(x, y^j). \tag{8}$$

Sampling variance can also be computed what is useful to evaluate accuracy of an estimator (8):

$$\tilde{D}^2(x) = \frac{1}{N-1}\sum_{i=1}^{N}(f(x,y^j) - \tilde{F}(x))^2. \tag{9}$$

The gradient is evaluated using the same random sample:

$$\tilde{G}(x) = \frac{1}{N}\sum_{j=1}^{N} g(x, y^j). \tag{10}$$

We use the sampling covariance matrix

$$Z(x) = \frac{1}{N-n}\sum_{j=1}^{N}(g(x,y^j) - \tilde{G})\cdot(g(x,y^j) - \tilde{G})' \tag{11}$$

later on for normalising a gradient estimator (10).

We examine several estimators for stochastic gradient following from expression (6). Assume randomness be exogenous and non affected by a decision variable, i.e. density $p(\cdot)$ does not depend on the variable *x*. In this case we obtain from (6):

$$g(x,y) = \nabla_x f(x,y). \tag{12}$$

Let us consider the analytical approach (AA) to estimate (12) two-stage stochastic programming with recourse (1). Indeed, by the duality of linear programming we have that

$$F(x) = c \cdot x + E\left\{\max_u \left[(h - T \cdot x) \cdot u \mid u \cdot W^T + q \geq 0,\ u \in \Re_+^n\right]\right\}. \tag{13}$$

Further it can be derived that under the assumption on the existence of a solution to the second stage problem in (3) and continuity of the measure *P*, the objective function (4) is smoothly differentiable and its gradient is expressed as

$$\nabla_x F(x) = E\left(g^i\left(x,\xi\right)\right), \tag{14}$$

where

$$g^1(x,\xi) = c - T \cdot u^* \tag{15}$$

is given by a set of solutions of a dual problem

$$(h - T \cdot x)^T \cdot u^* = \max_u [(h - T \cdot x)^T \cdot u \mid u \cdot W^T + q \geq 0, \quad u \in \Re^m]$$

,

(details are given in Rubinstein and Shapiro, 1993, Shapiro, 2000, etc.).

Let us compare the estimate (14), (15) with several ones. Since the analytical gradient is not always available, the finite difference (FD) approach is of interest. In this approach each ith component of the stochastic gradient $g^2(x,\xi)$ is computed as:

$$g_i^2(x,y) = \frac{f(x + \delta \cdot \varsigma_i, y) - f(x,y)}{\delta}, \tag{16}$$

ς_i is the vector with zero components except ith, equal to 1, δ is some small value.

Since the expression (16) requires computation of the function *n+1* times, the Simulated Perturbation Stochastic Approximation (SPSA) approach (Spall, 1992), which requires only one additional function value computation is also examined:

$$g^3(x,y)=\frac{f(x+\delta\cdot\nu,y)-f(x-\delta\cdot\nu,y)}{2\cdot\delta}, \tag{17}$$

where ν is the random vector obtaining values 1 or -1 with the probability of p=0.5 (see Spall, 1992), δ is some small value.

Let us additionally consider the Likelihood Ratio (LR) approach to obtain the expression of the stochastic gradient, which also requires only one additional function computation (Rubinstein, 1983). Say a random error in the objective function is an additive noise. Then we may change the variables in the integral (5) and evaluate the stochastic gradient using (6):

$$g^4(x,y)=\left(f(x+y)-f(x)\right)\cdot\nabla_y \ln p(y). \tag{18}$$

Numerical comparison of the estimators (15), (16), (17) and (18) is given in Section 5.

2. Stochastic procedure for optimisation

The gradient search approach with projection to a feasible set would be a chance to create optimizing sequence; however, the problems of "jamming" or "zigzagging" are typical in this case. To avoid them we implement the ε–feasible direction approach.

Let us define the set of *feasible directions* as follows:

$$V(x)=\left\{g\in\Re^n \,\middle|\, Ag=0,\ \forall_{1\le i\le n}\left(g_j\le 0,\ if\ x_j=0\right)\right\},\ x\in D. \tag{19}$$

Denote, g_U is projection of the vector g onto the set U. Since the objective function is differentiable, the solution $x\in D$ is optimal if (Bertsekas, 1982):

$$\nabla F(x)_V=0. \tag{20}$$

Assume a certain multiplier $\hat{\rho}>0$ is given. Define the function $\rho_x: V(x)\to\Re_+$ by

$$\rho_x(g)=\min\left\{\hat{\rho},\ \min_{\substack{g_j>0,\\ 1\le j\le n}}\left(\frac{x_j}{g_j}\right)\right\},\ \exists_{1\le j\le n}\left(g_j>0\right), \tag{21}$$

$\rho_x(g)=\hat{\rho}$, if $\forall_{1\le j\le n}(g_j\le 0)$. Thus, $\left(x+\rho\cdot g\right)\in D$, when $\rho=\rho_x(g)$, for any $g\in V(x), x\in D$, Now, let a certain small value $\hat{\varepsilon}>0$ be given. Then we introduce the function $\varepsilon_x: V(x)\to\Re_+$

$$\varepsilon_x(g)=\hat{\varepsilon}\cdot\max_{\substack{1\le j\le n\\ g_j>0}}\left\{\min\left\{x_j,\hat{\rho}\cdot g_j\right\}\right\},\ \exists_{1\le j\le n}\left(g_j>0\right),$$

$\varepsilon_x(g) = 0$, if $\forall_{1 \le j \le n}(g_j \le 0)$, and define the ε *-feasible set*

$$V_\varepsilon(x) = \left\{ g \in \Re^n \,\middle|\, Ag = 0, \forall_{1 \le i \le n} \left(g_j \le 0, \; if \; \left(0 \le x_j \le \varepsilon_x(g) \right) \right) \right\} \tag{22}$$

Now we start developing the procedure of stochastic optimization. Let some initial point $x^0 \in D$ be given, a random sample (7) of a certain initial size N^0 be generated at this point, and Monte-Carlo estimates (8), (9), (10), (11) be computed.

For instance, the starting point can be obtained as a solution of a deterministic linear problem:

$$(x^0, y^0) = \arg\min_{x,y} [c \cdot x + q \cdot y \mid A \cdot x = b, W \cdot y + T \cdot x \le h, \;\; y \in R_+^m, x \in R_+^n]. \tag{23}$$

The iterative stochastic procedure of gradient search could be used further:

$$x^{t+1} = x^t - \rho^t \cdot \tilde{G}(x^t), \tag{24}$$

where $\rho^t = \rho_{x^t}(\tilde{G}^t)$ is a step-length multiplier defined by (21), and $\tilde{G}_\varepsilon^t = \tilde{G}(x^t)_{V_\varepsilon\left(x^t\right)}$ is projection of a gradient estimator to the ε -feasible set.

Let us consider the choice of the Monte-Carlo sample size in more detail. Note, that there is no great necessity starting optimisation to compute estimators with high accuracy because then it suffices only to approximately evaluate the direction leading to the optimum. Therefore, one can obtain not so large samples at the beginning of optimum search and, later on, increase the size of samples so as to get the estimate of the objective function with desired accuracy just at the time of decision making on finding a solution to an optimisation problem. We can pursue this purpose by choosing a sample size at next iteration inversely proportional to the square of a gradient estimator from the current iteration:

$$N^{t+1} \ge \frac{\hat{\rho} \cdot C}{\rho^t \cdot \left| \tilde{G}_\varepsilon^t \right|^2}, \tag{25}$$

where $C > 0$ is a certain constant. On the other hand, this rule enables us to ensure the condition of proportionality of stochastic gradient variance to the square of the gradient norm, which is sufficient for convergence. Thus, under certain wide conditions of the existence of expectations of estimators this rule guarantees convergence of a. s. to an optimal solution, i.e., starting from any initial approximation $x^0 \in D$ and $N^0 > 1$, formulae (24), (25) define sequence $\left\{ x^t, N^t \right\}_0^\infty$ so that $x^t \in D$, and the values $\overline{\rho} > 0$, $\varepsilon_0 > 0$, $\overline{C} > 0$ exist so that

$$\lim_{t\to\infty}\left|\nabla F(x^t)_{V_\varepsilon\left(x^t\right)}\right|^2 = 0 \ (\mathrm{mod}\ (\mathrm{P})), \tag{26}$$

for $0 < \hat{\rho} \le \overline{\rho}$, $0 < \varepsilon < 1$, $C \ge \overline{C}$. Proof is given in (Sakalauskas, 2004).

Let us discuss the choice of parameters of the method. Step length ρ in (24) can be determined experimentally. The choice of the constant C or that of the best metrics for computing the stochastic gradient norm in (24) requires a separate study. For instance, the choice $C = n \cdot Fish(\gamma, n, N^t - n) \approx \chi^2_\gamma(n)$, where $Fish(\gamma, n, N^t - n)$ is the γ -quantile of the Fisher distribution with $(n, N^t - n)$ degrees of freedom, and the estimation of the gradient norm in metric induced by the sampling covariance matrix (11), ensure that a random error of the stochastic gradient does not exceed the gradient norm approximately with the probability of $1-\gamma$. Thus, we propose the following version of (25) for regulating the sample size in practice:

$$N^{t+1} = \min\left\{\max\left\{\left[\frac{n \cdot Fish(\gamma, n, N^t - n)}{\rho^t \cdot (\tilde{G}(x^t) \cdot (Z(x^t))^{-1} \cdot (\tilde{G}(x^t)}\right] + n,\ N_{\min}\right\},\ N_{\max}\right\}, \tag{27}$$

where minimal N_{min} (usually ~20-100) and maximal N_{max} (usually ~ 10000-20000) values are introduced to avoid great fluctuations of the sample size in iterations. Note that N_{max} may also be chosen from the conditions of a permissible confidence interval of the estimates of the objective function.

3. Statistical testing of the optimality hypothesis

A possible decision on finding an optimal solution should be examined at each step of the optimization process. Since we know only Monte-Carlo estimates of the objective function and that of its gradient, we can test only the statistical optimality hypothesis. As far as a stochastic error of these estimates in essence depends on the size of Monte-Carlo samples, a possible optimal decision could be made, if, first, there is no reason to reject the hypothesis of equality to zero of a gradient, and, second, the sample size is sufficient to estimate the objective function with desired accuracy.

Note that the distribution of sampling averages (8) and (10) can be approximated by the one- and multidimensional Gaussian laws (see, e.g., (Bhattacharya and Ranga Rao, 1976). Therefore, it is convenient to test the validity of a stationarity condition (20) the means of the well-known multidimensional Hotelling T^2-statistics (Krishnaiah and Lee, (1980)). Hence, the optimality hypothesis could be accepted for some point x^t with significance $1-\mu$, if the following condition is met:

$$(N^t - n) \cdot (\tilde{G}(x^t)) \cdot (Z(x^t))^{-1} \cdot (\tilde{G}(x^t)) / n \le Fish(\mu, n, N^t - n). \tag{28}$$

Next, again we can use asymptotic normality and decide that the objective function is estimated with permissible accuracy ε, if its confidence bound does not exceed this value:

$$2 \cdot \eta_\beta \cdot \tilde{D}(x^t) / \sqrt{N^t} \le \varepsilon , \tag{29}$$

where η_β is the β -quantile of a standard normal distribution. Thus, the procedure (24) is iterated adjusting the sample size according to (25) and testing conditions (28) and (29) at each iteration. If the latter conditions are met at some iteration, then there are no reasons to reject the hypothesis on finding the optimum. Therefore, there is the basis to terminate optimization and make a decision on the optimum finding with permissible accuracy. If at least one condition out of (28), (29) is not met, then the next sample is generated and optimization is continued. As it follows from the previous section, optimization should be terminated having generated the finite number of Monte-Carlo samples.

4. Computer simulation of stochastic gradient estimators

In this section a computer simulation study on the gradient estimators considered in Section 3 is presented using testing examples given in Appendix.

Let us consider Example 1 (see Appendix). Due to symmetry this function has the minimum at the point $x_+ = 0$. Thus, 400 Monte-Carlo samples of the size N = (1-10, 20, 40, 60, 80, 100) were generated at this point and T^2-statistic in criterion (20) was computed for each sample using the estimators (10), (11) and various stochastic gradients given by (15)-(18). The hypothesis on the difference of empirical distribution of this statistics from Fisher distribution was tested according to the criteria ω^2 and Ω^2. The values of ω^2 and Ω^2 statistics computation for the estimator (15) on the variable number and sample size are given in Table 1 and Table 2. The critical value ω^2=0.46 (p=0.05), and that of the next one is Ω^2=2.49 (p=0.05). The values of statistics exceeding the critical value are bolded. Thus the minimal size of the Monte Carlo sample necessary to approximate the distribution of the Hotelling statistics by the Fisher distribution depends on the dimensionality of the task n. Thus the requiring Monte-Carlo sample size depending on dimensionality is given in Table 3. Similar results are obtained for other estimators, too.

Table 1. ω^2 criteria results by the number of variables and sample size

N / n	50	100	200	500	1000
2	0.30	0.24	0.10	0.08	0.04
3	0.37	0.12	0.09	0.06	0.04
4	0.19	0.19	0.13	0.08	0.04
5	**0.75**	0.13	0.12	0.08	0.06
6	**1.53**	0.34	0.10	0.10	0.08
7	**1.56**	0.39	0.13	0.08	0.09
8	**1.81**	0.42	0.27	0.18	0.10
9	**4.18**	**0.46**	0.26	0.20	0.12
10	**8.12**	**0.56**	**0.53**	0.25	0.17

Table 2. Ω^2 criteria results by the number of variables and sample size

N / n	50	100	200	500	1000
2	**2.57**	1.14	0.66	0.65	0.42
3	**2.78**	0.82	0.65	0.60	0.27
4	**3.75**	1.17	0.79	0.53	0.31
5	**4.34**	1.46	0.85	0.64	0.36
6	**8.31**	2.34	0.79	0.79	0.76
7	**8.14**	**2.72**	1.04	0.52	0.45
8	**10.22**	**2.55**	1.87	0.89	0.52
9	**20.86**	**2.59**	1.57	1.41	0.78
10	**40.57**	**3.69**	**3.51**	1.56	0.98

Table 3. Requiring Monte-Carlo sample size by the number of variables

Number of variables	Monte-Carlo sample size
10	500
20	1000
40	2200
60	3300
80	4500
100	6000

Similar experiments were performed with Example 2 and Example 3. The optimum point in these examples was established by stochastic optimization. The values of ω^2 and Ω^2 statistics computation for Example 3 on the sample size are given in Table 4.

Table 4. ω^2 and Ω^2 criteria results by the sample size

Monte-Carlo sample size	ω^2 criteria (Critical value 0.46)	Ω^2 criteria (Critical value 2.56)
500	**1,95**	**9,53**
700	**0,64**	**3,39**
800	**0,61**	**3,33**
900	**0,49**	**2,57**
1000	0,30	1,89
1500	0,12	0,80
2000	0,06	0,56

As follows from the results of this simulation, the distribution of the Hotelling statistics can be approximated by the Fisher distribution appropriately choosing the sample size.

Further dependencies of the frequency of the optimality hypothesis (gradient equality to zero) according to the criterion (28) on the distance $r = \left|x - x^+\right|$ to the optimal point and the Monte Carlo sample size N for various gradient estimators were studied. The purpose of this study is to answer how the used estimators are good to reject the hypothesis of optimality at the point of

solution which differs from the optimal one. These dependencies for n=2 (Example 1) and more variables are presented in **Fig. 1-5** (for confidence α =0.95). Thus, the computation results show that the AA (15) and the FD approach (16) provide estimators for reliable checking of the optimality hypothesis in a wide range of dimensionality of the stochastic optimization problem ($2 \le n \le 100$). However, the SPSA (17) and the LR (18) estimators can be applied for stochastic gradient estimation only for the tasks of not very large dimensionality: $1 \le n \le 20$. Similar results were obtained investigating Example 2 and Example 3.

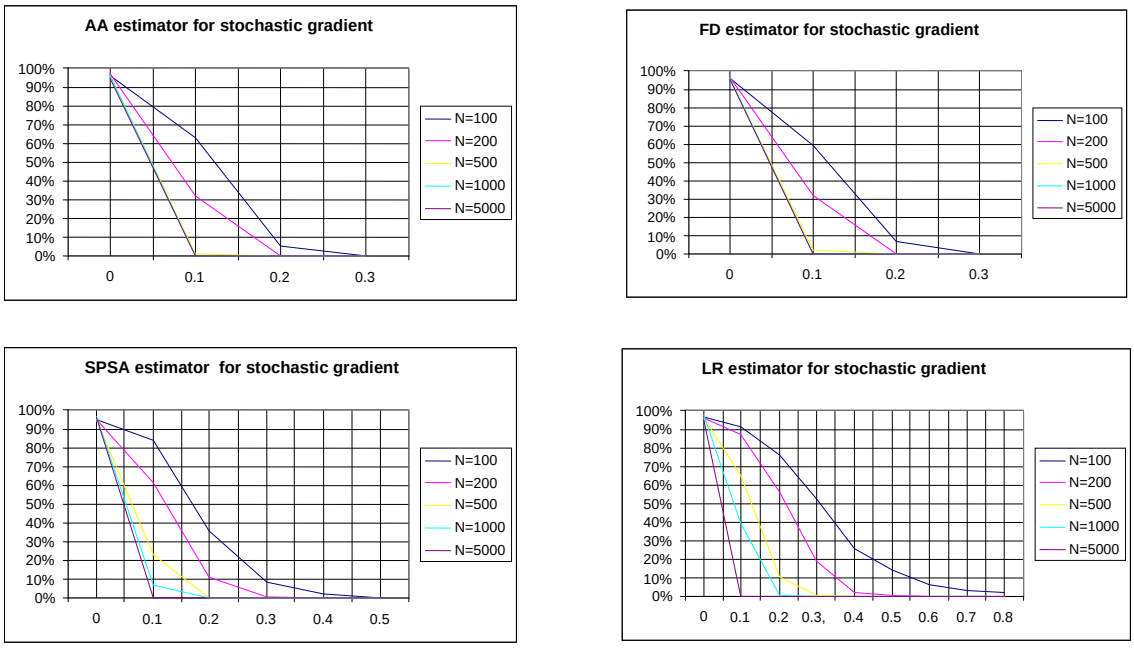

Fig 1. **Frequency of the optimality hypothesis (n=2).**

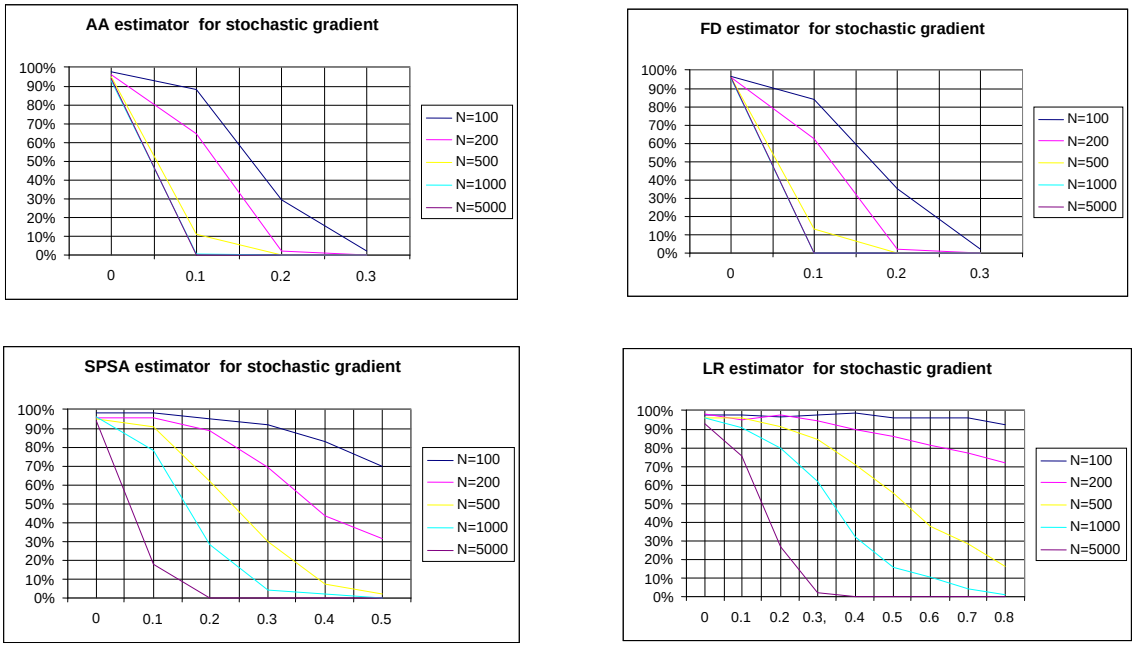

Fig 2. **Frequency of the optimality hypothesis (n= 10)**

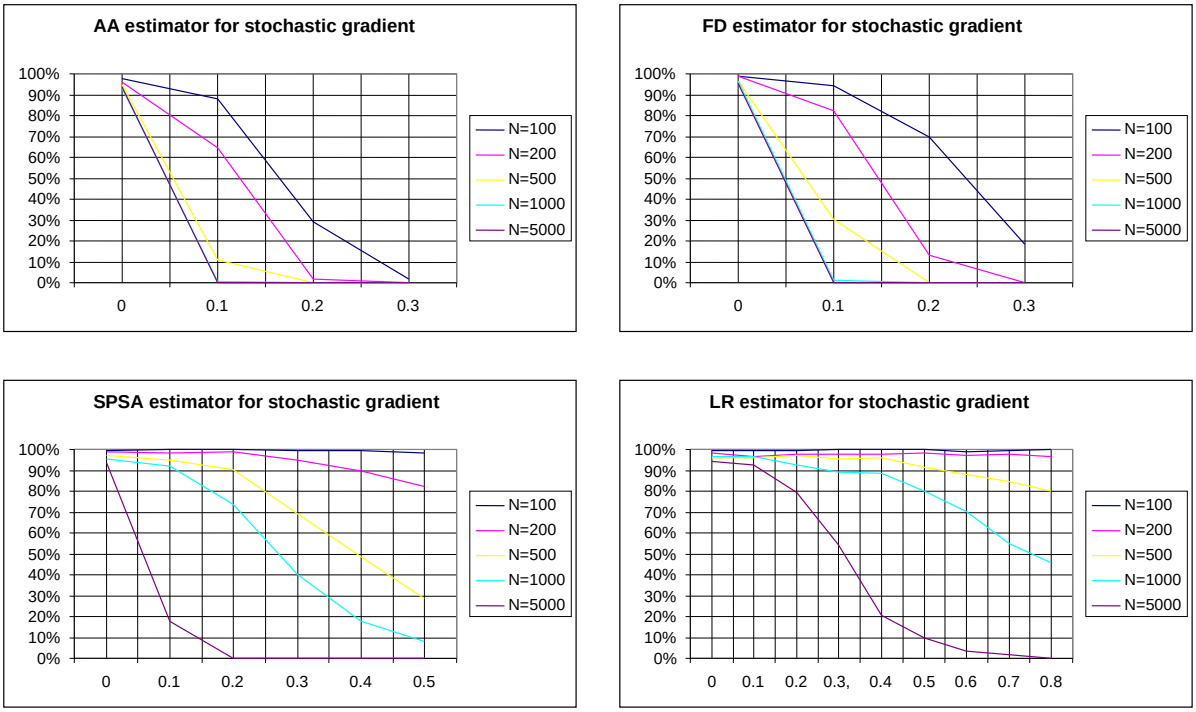

Fig 3. Frequency of the optimality hypothesis (n=20)

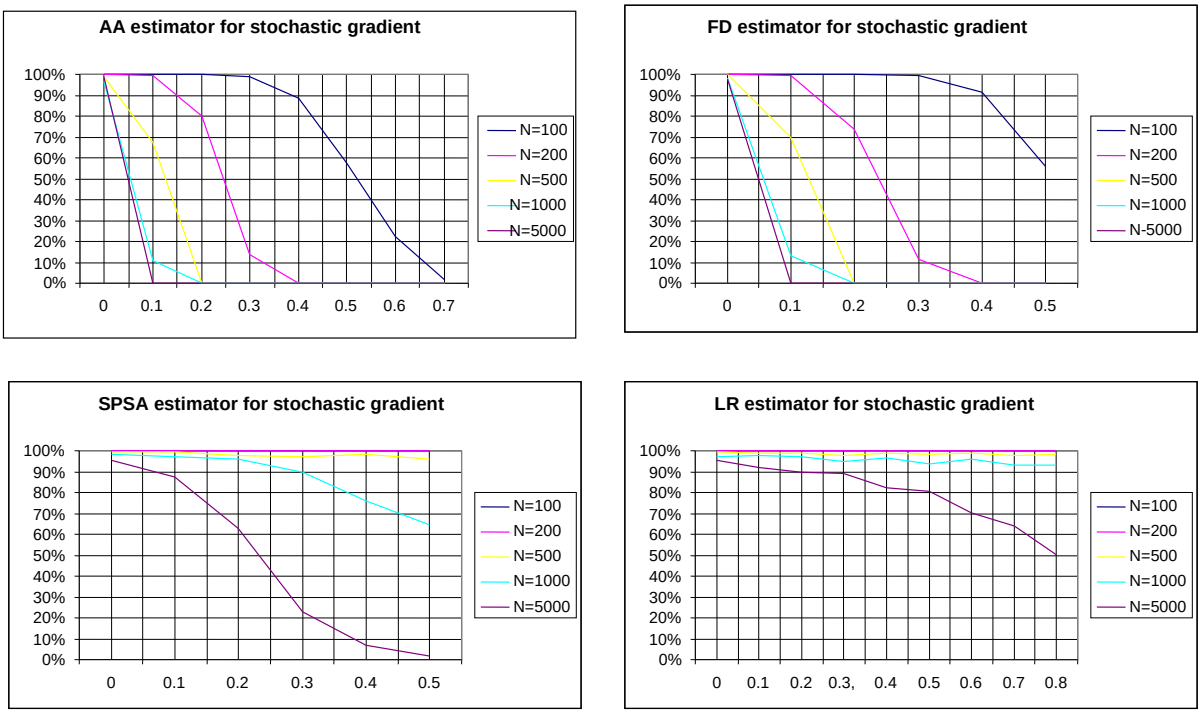

Fig 4. **Frequency of the optimality hypothesis (n=50)**

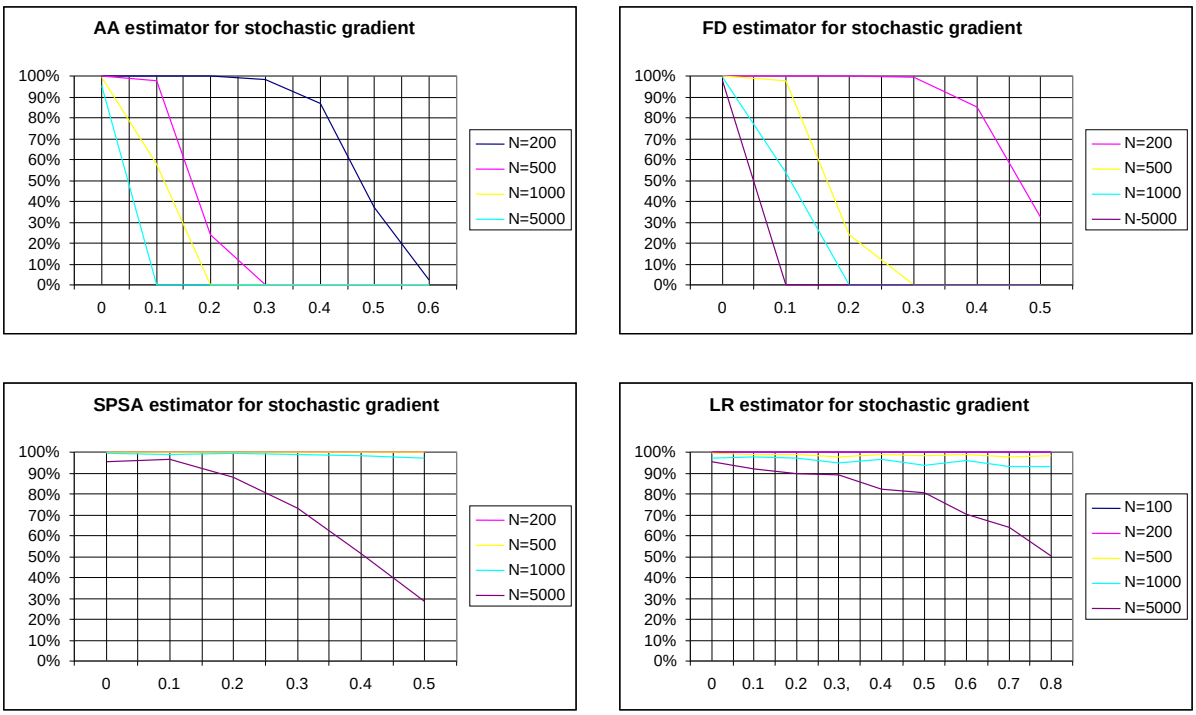

Fig 5. **Frequency of the optimality hypothesis (n=100)**

5. Computer study of the convergence of the stochastic optimization method

In this section we study the convergence of the approach developed by computer simulation using the testing examples given in Appendix.

Let us consider the results solving a manpower problem (Example 2). Optimal solutions of solving this task of varying variation η are given in Table 5 (manpower costs are given in USD, an admissible confidence interval of the objective function is 0,1 (100 USD).

Table 5. Manpower amount on the levels and costs (in dependence of the variation η).

	base level of regular staff			cost of manpower (conf. interval 0,1)
η	X_1	X_2	X_3	F
0	9222	5533	1106	94, 899
1	9222	5533	1106	94,899
10	9376	5616	1106	96,832
30	9452	5672	1036	96,614

It is proved (Sakalauskas, 2000, 2002) that the considered approach ensures a linear rate of convergence. It follows from the linearity of this rate that the total amount of Monte-Carlo trials $\sum_{j=1}^{t} N^j$ performed to get the optimal solution is approximately proportional to the amount of trials N^t necessary to solve the problem with admissible accuracy.

Moreover, the ratio $\sum_{j=1}^{t} \frac{N^j}{N^t}$ is determined mostly by a positive definiteness of Hessian of the objective function and almost does not depend on the admissible accuracy ε (see details in Sakalauskas, 2002). Table 6 illustrates this fact where this ratio is presented for various ε and η. Thus, a conclusion follows that if we have a certain resource to compute one value of the objective function with admissible accuracy, then optimization requires only several times more computations. This enables us to construct reasonable from a computational viewpoint stochastic methods for stochastic programming with admissible accuracy.

Table 6. Ratio $\sum_{j=1}^{t} \frac{N^j}{N^t}$ under admissible interval ε and variation η.

ε	0,05	0,1	0,2
η =10	16.1	10.7	10.9
η =30	21.4	21.3	20.2

Let us study the solution of the examples taken from the Internet database of test problems (http://www.math.bme.hu/~deak/twostage/). The solution taken from the database and the one obtained by the developed approach are presented in Appendix. As one can see, the developed adaptive method enables us neither to find the given solutions nor to improve them, sometimes rather essentially (for instance, in Example 4).

Now let us consider more detailed results obtained in solving Example 3, 400 times by the method (15). The initial data were as follows: $\gamma = \beta = 0.95$, $\mu = 0.99$, $\varepsilon = 0.1; 0.2; 0.5; 1.0$, N^0=N_{min}=100, maximal number of iterations $t_{\max} = 100$, generation of trials was broken when the estimated confidence interval of the objective function exceeded the admissible value ε.

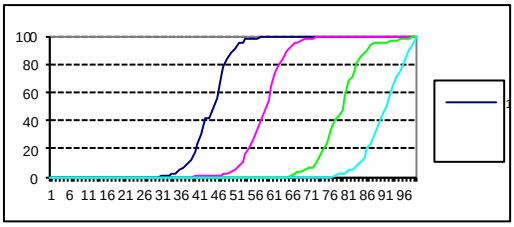

Figure 6. **Frequency of stopping under admissible interval ε.**

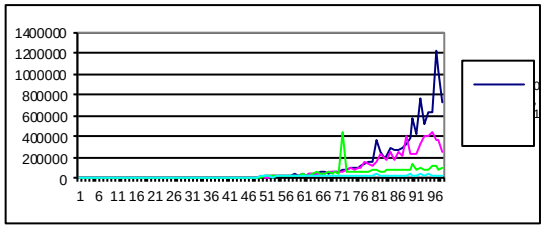

Figure 7. **Change of the sample size under admissible interval ε.**

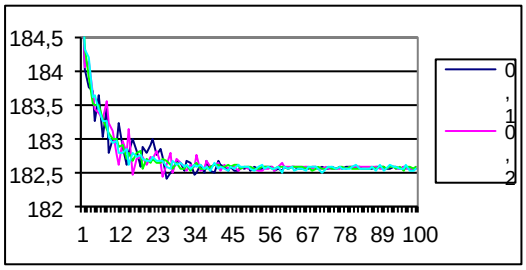

Figure 8. **Change of the objective function under admissible interval ε.**

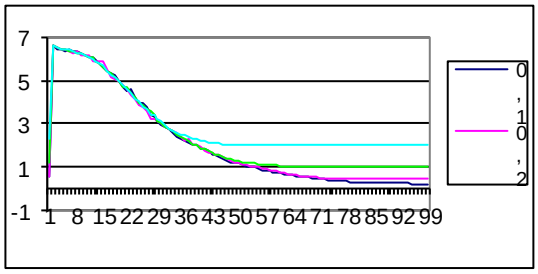

Figure 9. **Change of confidence interval under admissible interval ε.**

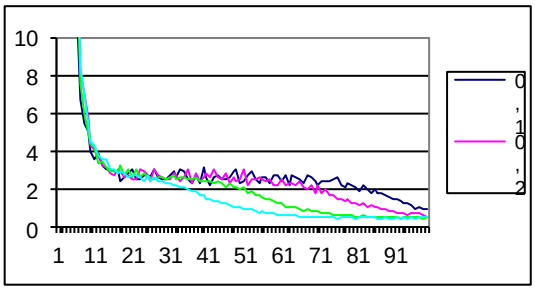

Figure 10. **Change of Hotelling statistics under admissible interval ε.**

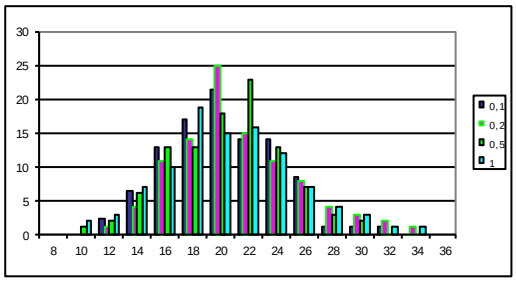

Figure 11. **Histogram of ratio $\sum_{j=1}^{t} \frac{N^j}{N^t}$ under admissible interval ε.**

Table 7. Value of the objective function F and the ratio $\sum_{j=1}^{t} \frac{N^j}{N^t}$ under admissible interval ε (Example 3).

Estimated confidence interval, ε	Value of the objective function, F	Ratio, $\sum_{j=1}^{t} \frac{N^j}{N^t}$
0.1	182.6101	20.14
0.2	182.6248	19.73
0.5	182.7186	19.46
1	182.9475	19.43

Termination conditions were satisfied at least one time for all paths of optimization. Thus, a conclusion on the optimum finding with admissible accuracy could be made for all paths (sampling frequency of termination after t iterations with confidence intervals is presented in Fig. 6). Average dependencies of the sample size, objective function, confidence interval, the Hotelling statistics on the iteration number t are given (Fig's. 7-10) to illustrate the convergence and behavior of the optimization process. Also, one path of realization of the optimization process illustrates a stochastic character of this process in these figures. In Fig. 11 the histogram of the ratio $\sum_{j=1}^{t} \frac{N^j}{N^t}$ is depicted.

Discussion and conclusions

Thus, the stochastic iterative method has been developed to solve stochastic linear programming (SLP) problems by a finite sequence of Monte-Carlo sampling estimators. Since in iterative stochastic optimization only the first order methods are working, we have confined ourselves by gradient-descent type methods showing that a typical deterministic approach of constrained optimization might be generalized in a stochastic case. The proposed method was studied by numerical ways using the examples taken from literature and standard database of two stage programming tests.

The approach presented in this paper is grounded on the stopping procedure and the rule for adaptive regulation of the size of Monte-Carlo samples, taking into account statistical modelling accuracy. Several stochastic gradient estimators were compared by computer simulation studying the workability of the estimators for testing the optimality hypothesis by statistical criteria. It was demonstrated that a minimal size of the Monte Carlo sample necessary to approximate the distribution of the Hotelling statistics, computed using gradient estimators, by the Fisher distribution depends on an approximation approach and dimensionality of the task n. The computation results show that an analytical and difference approach provide estimators for reliable checking of the optimality hypothesis in a wide range of dimensionality of the stochastic optimization problem ($2 \le n \le 100$). However, the SPSA and the LR estimators can be applied for stochastic gradient estimation only for tasks of not very large dimensionality $1 \le n \le 20$. The proposed termination procedure allows us to test the optimality hypothesis and to evaluate reliably confidence intervals of objective and constraint functions in a statistical way.

The regulation of a sample size in case this size is taken inversely proportional to the square of the norm of the gradient of the Monte-Carlo estimator allows us to solve SLP problems rationally from a computational viewpoint and guarantees convergence a. s. The linear rate of the proposed convergence was studied by a numerical way. In our approach optimization can usually require only several times more computations as compared with the computation of one function value. The numerical study corroborates theoretical conclusions on the convergence method and shows that the developed procedures make it possible to solve stochastic problems with sufficiently agreeable accuracy by the means of an acceptable amount of computations. All test examples taken from standard database were solved successfully and some solutions given on this base were improved.

References

Bertsekas D. P. (1982). Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York, London.

Bhattacharya, R.N., and Ranga Rao, R. (1976). Normal Approximation and Asymptotic Expansions. *John Wiley*, New York, London, Toronto.

Ermolyev, Ju.M. (1976). Methods of Stochastic Programming. *Nauka*, Moscow. 240 pp. (in Russian).

Ermolyev, Yu., and Wets R. (1988). Numerical Techniques for Stochastic Optimization. *Springer-Verlag*, Berlin.

Higle, J. L., and S. Sen (1999). Statistical approximations for stochastic linear programming problems. *Annuals of Operations Research* 85, 173-192.

Krishnaiah, P.R., and Lee, J.C. (1980). Handbook of Statistics, vol. 1 (Analysis of Variance). *North-Holland, N.Y.*

Kushner, H.J., and Yin, G. G. (2003). Stochastic Approximation and Recursive Algorithms and Applications. *Springer*: N.Y., Heidelberg, Berlin.

Marti. K. (1988). Descent stochastic quasi-gradient methods. In Ermolyev, Yu., and Wets R. Numerical Techniques for Stochastic Optimization. *Springer-Verlag*, Berlin, pp. 393-400.

Marti, K. (1989). Optimal semi-stochastic approximation procedures. ii. Z. Angew. Math. Mech. 69, No.4, pp. 67-69.

Marti, K. (1996). Differentiation formulas for probability functions: The transformation method. Mathematical Programming, 75(2), pp. 201-220.

Marti, K. (2005). Stochastic Optimization Methods. *Springer*: N.Y.

Mikhalevitch, V.S., Gupal, A.M., and Norkin, V.I. (1987). Methods of Non-convex Optimization. *Nauka*: Moscow (in Russian).

Polyak, B.T. (1987). Introduction to optimization. *Translations Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division*, New York.

Prekopa, A. (1995). Stochastic Programming, *Kluwer*.

Rubinstein, R, and Shapiro, A. (1993). Discrete events systems: sensitivity analysis and stochastic optimization by the score function method. *Wiley & Sons, N.Y.*

Sakalauskas, L. (2002). Nonlinear stochastic programming by Monte-Carlo estimators. *European Journal on Operational Research*, 137, 558-573.

Sakalauskas, L. (2004). Application of the Monte-Carlo method to nonlinear stochastic optimization with linear constraints. *Informatica*, **15**(2), 271-282.

Sen, S. (2001). Stochastic Programming: Computational Issues and Challenges. *Encyklopedia of OR/MS*, Gass S., and Harris C. (eds.), *Kluwer Academic Publishers*, Dordrecht, pp. 784-789.

Shapiro, A., and Homem-de-Mello, T. (1998). A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming*, 81, pp. 301-325.

Shapiro, A.(2000). Stochastic Programming by Monte-Carlo simulation methods. *Stochastic Programming E-Print Series.*

http://www.math.bme.hu/~deak/twostage/

K. Žilinskas is Associate Professor, Phd, Dept of Informatics. Šiauliai University. First degree in mathematics and physics (Šiauliai Pedagogical Institute, 1981), Master of Science (1994), Phd (2007). Member of European Working Group on continuous optimization (2007), member of Council of the Lithuanian Operational Research Society (2007). Author of 10 scientific articles. Research interests: optimisation, stochastic programming, Monte Carlo method, statistical methods.

DVIEJŲ ETAPŲ STOCHASTINIS TIESINIS PROGRAMAVIMAS NAUDOJANT MONTE-KARLO IMČIŲ ĮVERČIUS

Kęstutis Žilinskas

Santrauka

Straipsnyje pateikiamas iteracinis stochastinio tiesinio programavimo uždavinių sprendimo metodas naudojant baigtinių Monte Karlo imčių serijų įverčius. Metodas pagrįstas Monte Karlo imčių ilgio reguliavimo taisykle, kai imties ilgis atvirkščiai proporcingas tikslo funkcijos gradiento Monte Karlo įverčio normos kvadratui, bei statistinėmis algoritmo stabdymo sąlygomis, kurios atsižvelgia į statistinio modeliavimo tikslumą. Uždavinio sprendinio optimalumas bei tikslumas yra vertinami statistiniais būdais, patikrinant statistinę hipotezę apie sprendinio optimalumą bei skaičiuojant tikslo funkcijos pasikliautinąjį intervalą. Iteracinio procese taikomas ε-leistinųjų krypčių metodas. Skaitiniai eksperimentai patvirtina teorines prielaidas bei parodo, kad pasiūlytas metodas leidžia išspręsti stochastinius uždavinius reikiamu tikslumu priimtinu skaičiavimų kiekiu.

Pagrindiniai žodžiai: stochastinis programavimas, Monte Karlo metodas, stochastinis gradientas, statistiniai kriterijai, ε-leistinoji kryptis.

Appendix

For numerical investigation of the developed approach the following test examples were studied.

Example 1. Let us consider the function (Sakalauskas, 2002)

$$F(x) \equiv Ef_0(x+\xi)$$

being an expectation of

$$f_0(y) = \sum_{i=1}^{n}(a_i y_i^2 + b_i \cdot (1-\cos(c_i \cdot y_i)))$$

where y_i are random and normally $N(0, d^2)$ distributed, d=0.5, a_i are uniformly distributed in [2, 5], b_i - in [1, 2] and c_i - in [-0.5, 0.5] and $2 \le n \le 100$.

Example 2. Manpower-planning problem

We consider the manpower-planning problem (Ermolyev and Wets, 1988) where the employer must decide upon the base level of the regular staff at various skill levels. The available recourse actions are regular staff overtime or outside temporary help in order to meet unknown demand for service at a minimal cost. The problem is as follows: choose $x = (x_1, x_2, x_3)$ to minimize

$$F(x,z) = \sum_{j=1}^{3} c_j \cdot x_j + \sum_{t=1}^{12} E \min\left(\sum_{j=1}^{3} (q_j \cdot y_{j,t} + r_j \cdot z_{j,t}) \right)$$

subject to

$$x_j \geq 0,\ y_{j,t} \geq 0,\ z_{j,t} \geq 0,$$

$$\sum_{j=1}^{3} (y_{j.t} + z_{j,t}) \geq w_t - \alpha_t \cdot \sum_{j=1}^{3} x_j,\ t = 1,2,\ldots,12,$$

$$y_{j,t} \leq 0.2 \cdot \alpha_t x_j,\ j = 1,2,3,\ t = 1,2,\ldots,12,$$

$$\gamma_{j-1}(x_j + y_{j-1,t} + z_{j-1,t}) - (x_j + y_{j-1,t} + z_{j-1,t}) \geq 0,$$

$$j = 1,2,3,\ t = 1,2,\ldots,12,$$

where

x_j	base level of the regular staff at the skill level j = 1, 2, 3,
$y_{j,t}$	amount of overtime help,
$z_{j,t}$	amount of temporary help,
c_j	cost of the regular staff at the skill level j = 1, 2, 3,
q_j	cost of overtime,
r_j	cost of temporary help,
w_t	demand for services at the period *t*,
α_t	anticipated absentee rate for the regular staff at the time *t*,
γ_{j-1}	ratio of the amount of the skill level *j* per amount of *j-1* required,

the demands w^t are independently normal: $N(\mu, \sigma^2)$, where $\mu_t = \eta \cdot \sigma_t^2$. The initial data and other details can be found in (Ermolyev and Wets, 1988). This problem has 3 variables at the first stage and 72 linear inequalities with 75 variables at the second stage.

Example 3.

Two-stage stochastic linear optimisation problem. The dimensions of the task are as follows: the first stage has 10 rows and 20 variables; the second stage has 20 rows and 30 variables. The data of the problem are taken from the database at the address http://www.math.bme.hu/~deak/twostage/ l1/20x20.1 (accessed on 2006-01-20).

The estimate of the optimal value of the objective function given in the database is 182.94234 ± 0.066 and the optimal point is:

```
0.000000    0.000000    0.410936    0.107915    0.020078
0.000000    0.000000    0.280964    0.038423    0.085044
0.000000    0.161090    0.000000    0.758401    0.551627
0.266730    0.309641    0.000000    0.125513    0.357334
```

The application of the considered approach allows us to improve the estimate of the optimal value up to 182.59248 ± 0.033 and the optimal point is:

```
0.000000    0.000000    0.438473    0.105014    0.027074
0.000000    0.000000    0.297970    0.029859    0.104539
0.000000    0.170788    0.000000    0.737453    0.580798
0.279175    0.325201    0.000000    0.127925    0.352842
```

Example 4.

Two-stage stochastic linear optimisation problem. The dimensions of the task are as follows: the first stage has 10 rows and 20 variables; the second stage has 20 rows and 30 variables. The data of the problem are taken from the database at the address http://www.math.bme.hu/~deak/twostage/l1/20x20.2.

The estimate value of the objective function given in the database is 266.68373 ± 0.187 and the solution point is:

```
0.000000    2.491040    0.000000    1.140885    0.290554
0.000000    0.033435    0.023088    0.406078    0.000000
0.000000    0.625346    0.000000    0.000000    0.000000
0.084560    0.119499    0.281248    0.192027    0.152020
```

The application of the considered approach allows us to improve the estimate of the function value up to 266.22764 ± 0.066 and the solution point is:

```
0.000000    2.596710    0.000000    1.195545    0.281074
0.000000    0.032765    0.020928    0.406148    0.000000
0.000000    0.638956    0.000029    0.000000    0.000000
0.083820    0.118879    0.282258    0.191927    0.150970
```

Example 5.

Two-stage stochastic linear optimisation problem. The dimensions of the task are as follows: the first stage has 30 rows and 60 variables; the second stage has 60 rows and 90 variables. The data of the problem are taken from the database at the address http://www.math.bme.hu/~deak/twostage/l1/60x60.4.

The estimate value of the objective function given in the database is 300.84160 ± 0.039 and the solution point is:

```
0.1969    0.0000    0.0260    0.2892    0.1560    0.0000    0.0000
0.0172    0.0000    0.0793
0.0000    0.0000    0.0000    0.0499    0.2118    0.0881    0.0000
0.0425    0.0967    0.0501
0.6735    0.0000    0.0619    0.0203    0.0000    0.0000    0.1780
0.0000    0.0414    0.3646
0.3090    0.3593    0.0000    0.0000    0.0000    0.0899    0.2509
0.1090    0.0375    0.0000
0.0656    0.0450    0.0501    0.0000    0.0000    0.1950    0.0633
0.0000    0.0000    0.0633
0.6750    0.0026    0.0000    0.0000    0.2335    0.3067    0.0000
0.8165    0.0641    0.0000
```

Application of the considered approach allows us to improve the estimate of the function value up to 300.66896 ± 0.033 and the solution point is:

```
0.1940    0.0000    0.0224    0.2781    0.1651    0.0000    0.0000
0.0187    0.0000    0.0804
0.0000    0.0000    0.0000    0.0594    0.2019    0.0837    0.0000
0.0395    0.0984    0.0536
0.6311    0.0000    0.0636    0.0342    0.0000    0.0000    0.1745
0.0000    0.0318    0.3610
0.3057    0.3738    0.0000    0.0000    0.0000    0.0812    0.2623
0.0915    0.0385    0.0000
0.0651    0.0367    0.0592    0.0000    0.0000    0.1919    0.0666
0.0000    0.0000    0.0680
0.6700    0.0051    0.0000    0.0000    0.2367    0.3040    0.0000
0.7703    0.0563    0.0000
```

Example 6.

Two-stage stochastic linear optimisation problem. The dimensions of the task are as follows: the first stage has 40 rows and 80 variables; the second stage has 80 rows and 120 variables. The data of the problem are taken from the database at the address http://www.math.bme.hu/~deak/twostage/l1/80x80.3.

The estimate value of the objective function given in the database is 586.32985 ± 0.327 and the solution point is:

```
0.0000    0.0000    0.2811    0.9275    0.0000    0.0000    0.0000
0.0000    0.1368    0.2596
0.2590    0.0000    0.1567    0.0727    0.0311    0.3451    0.0000
0.0917    0.0262    0.0276
0.3560    0.0000    0.0853    0.0000    0.6380    0.0000    0.6734
0.0000    0.1170    0.0723
0.0000    0.1558    0.1571    0.0000    0.1542    0.0000    0.0653
0.1334    0.0000    0.0000
0.0000    0.0000    0.0548    0.0000    0.1130    0.0000    0.0153
0.0735    0.1191    0.0000
0.0401    0.0251    0.0519    0.0000    0.0319    0.0000    0.0229
0.0000    0.0406    0.1074
0.0101    0.0000    0.0000    0.0000    0.3822    0.0000    0.0624
0.0000    0.0368    0.0186
0.1950    0.1213    0.0000    0.0000    0.0000    0.0144    0.3296
0.0000    0.1363    0.0950
```

The application of the considered approach allows us to improve the estimate of the function value up to 475.01266 ± 0.99999 and the solution point is:

```
0.1451    0.0000    0.0000    0.0704    0.1418    0.0000    0.0020
0.2071    0.0000    0.2293
0.0000    0.1992    0.2146    0.0387    0.0175    0.0000    0.0000
0.1091    0.1159    0.1331
0.0000    0.0000    0.0126    0.0529    0.8415    0.0000    0.2138
0.0671    0.1740    0.1710
0.0000    0.2721    0.1630    0.0000    0.0589    0.0000    0.1285
0.1268    0.0000    0.0000
0.2703    0.0000    0.0351    0.0000    0.6207    0.2879    0.1270
0.1280    0.1745    0.0375
0.0567    0.0751    0.1603    0.0000    0.0000    0.0000    0.0244
0.4942    0.0539    0.1429
0.0639    0.0000    0.0000    0.1672    0.0913    0.0000    0.0000
0.0619    0.0000    0.0149
0.1967    0.1349    0.0000    0.0000    0.0000    0.0110    0.2659
0.0000    0.0938    0.0000
```