ISSN 1392-8619 print/ISSN 1822-3613 online
ŪKIO TECHNOLOGINIS IR EKONOMINIS VYSTYMAS
TECHNOLOGICAL AND ECONOMIC DEVELOPMENT OF ECONOMY
http://www.tede.vgtu.lt
2006, Vol XII, No 4, 314–320

APPLICATION OF STATISTICAL CRITERIA TO OPTIMALITY TESTING IN STOCHASTIC PROGRAMMING

Leonidas Sakalauskas[1], Kestutis Žilinskas[2]

[1]Institute of Mathematics and Informatics, Akademijos g. 4, 08663 Vilnius, Lithuania
E-mail: sakal@ktl.mii.lt
[2]University of Šiauliai, P. Višinskio g. 19, 77156 Šiauliai, Lithuania
E-mail: kest.zil@gmail.com

Received 15 June 2006; accepted 20 November 2006

Abstract. In this paper the stochastic adaptive method has been developed to solve stochastic linear problems by a finite sequence of Monte-Carlo sampling estimators. The method is grounded on adaptive regulation of the size of Monte-Carlo samples and the statistical termination procedure, taking into consideration the statistical modeling error. Our approach distinguishes itself by treatment of the 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. The adjustment of sample size, when it is taken inversely proportional to the square of the norm of the Monte-Carlo estimate of the gradient, guarantees the convergence a. s. at a linear rate.
We examine four estimators for stochastic gradient: by the differentiation of the integral with respect to x, the finite difference approach, the Simulated Perturbation Stochastic Approximation approach, the Likelihood Ratio approach.
The numerical study and examples in practice corroborate the theoretical conclusions and show that the procedures developed make it possible to solve stochastic problems with a sufficient agreeable accuracy by means of the acceptable amount of computations.

Keywords: linear programming, stochastic programming, optimality, statistical criteria, Monte-Carlo method.

1. Introduction

Stochastic programming deals with a class of optimization models in which some of the 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. Stochastic gradient type search is often applied in constructing numerical methods for stochastic problems. The procedures are usually constructed expressing gradient as an expectation and then evaluating this expectation by means of statistical simulation by the Monte-Carlo method. Yudin (1965) was the first to suggest such an idea of realizing by means of smoothing operators. Later on, the application of smoothing and statistical simulation to stochastic optimization was considered in many papers and books (see, e.g., Rubinstein and Shapiro (1993), Ermolyev and Norkin (1995), Sakalauskas (2002), Marti (2005), etc.).

Hence, results of the stochastic problems can be described as samples of random data, for which analysis, consequently, we can use statistical methods. Since asymptotic distribution of sampling estimators can be approximated by the one- and multidimensional Gaussian laws (see, e.g., Bhattacharya and Ranga Rao, (1976)) we may apply well developed theory of normal statistics to evaluation of statistical error of sampling estimators as well as decision making about optimum finding.

The purpose of this work is to examine the convergence of estimators obtained during simulation of stochastic problems and consider applicability of this approach to testing optimality hypothesis in stochastic programming.

2. Monte-Carlo estimators for stochastic optimization

We consider a nonlinear stochastic optimization problem:

$$F(x) \equiv Ef(x,\xi) \to \min_{x \in X \subset \Re^n}, \qquad (1)$$

where the objective function is expectation of random function $f(x,\xi)$, depending on random vector ξ, defined by the distribution density function $p(x,\cdot)$, and the feasible set $x \in X \subset \Re^n$ is a bounded and convex linear set in general:

$$X = \{x | Ax = b, x \ge 0\},$$

$b \in R^m$, A is the $n \times m$-matrix, $X \ne \varnothing$ (see Ziemba & Mulvey (1998), Sakalauskas (2000), Andersson et al (2001), etc.).

The gradient search is the most often used way of constructing methods for numerical optimization. Since mathematical expectations in (1) are computed explicitly only in rare cases, it is complicated all the more to analytically compute the gradients of functions, containing this expression. The Monte-Carlo method is a universal and convenient tool of estimating these expectations and we try to apply it to estimate derivatives, too. Let us introduce a set of Monte-Carlo estimators needed for construction of stochastic optimization procedure. First, let us consider the expectation:

$$F(x) = Ef(x,\omega) \equiv \int_{R^n} f(x,y) \cdot p(x,y)dy, \qquad (2)$$

where the function f and the density function p are assumed differentiable with respect to x in the entire space $\Re^n$.

The differentiability of integrals of this kind has been studied rather well, and there exists 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 ans Shapiro (1993), Uriasyev (1994), Prekopa (1999), Ermolyev et al (2003), etc.). Let us denote the support of random vector as $S(x) = \{y | p(x,y) > 0\}, x \in R^n$. Then it is not difficult to see that the vector-column of gradient of this function could be expressed as (Sakalauskas (2002)):

$$\nabla F(x) = E\big(\nabla_x f(x,\omega) + (f(x,\omega) - f(x, E\omega)\big) \cdot \nabla_x \ln p(x,\omega)). \qquad (3)$$

We can see that it is possible to express the expectation and its gradient through a linear operator from the same probability space. Hence, operators (2) and (3) can be estimated by means of the same Monte-Carlo sample.

Thus, assume here that the Monte-Carlo samples of a certain size N are possible to provide for any $x \in R^n$:

$$Y = (y^1, y^2, ..., y^N), \qquad (4)$$

where y^i are independent random variables, identically distributed with the density $p(\cdot): \Omega \to R_+^n$, and the sampling estimators are computed:

$$\tilde{F}(x) = \frac{1}{N}\sum_{j=1}^{N} f(x, y^j), \qquad (5)$$

$$\tilde{D}^2(x) = \frac{1}{N-1}\sum_{i=1}^{N}(f(x,y^j) - \tilde{F}(x))^2. \qquad (6)$$

The estimator of a gradient:

$$\tilde{G}(x) = \frac{1}{N}\sum_{j=1}^{N} g(x, y^j) \qquad (7)$$

and the sampling covariance matrix

$$Z(x) = \frac{1}{N-n} \cdot \sum_{j=1}^{N}\left((g(x,y^j) - \tilde{G}) \cdot (g(x,y^j) - \tilde{G})'\right), \qquad (8)$$

where $g(x,\cdot)$ is the stochastic gradient, i.e., such random vector that $Eg(x,\xi) = \nabla F(x)$, will be of use further.

Note, the distribution of sampling estimators (5) and (7) can be approximated by the one- and multidi-mensional Gaussian laws (see, e.g., Bhattacharya and Ranga Rao, (1976)). In solution of problem (1) the gradient is zero (Polyak (1987)). Therefore it is convenient to test the validity of the stationary condition by means of the well-known multidimensional Hotelling T^2-statistics (see, e.g., Krishnaiah, and Lee (1980), etc.). Namely, the optimality hypothesis (i.e. of equality of the gradient to zero) might be accepted for some point x with significance $1-\mu$ according to Fisher criteria if:

$$(N^t - n) \cdot (\tilde{G}(x^t)) \cdot (Z(x^t))^{-1} \cdot (\tilde{G}(x^t))/n \le Fish(\mu, n, N^t - n). \qquad (9)$$

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

$$2 \cdot \eta_\beta \cdot \tilde{D}(x)/\sqrt{N} \le \varepsilon, \qquad (10)$$

where η_β is the β – quantile of the standard normal distribution.

3. Computer simulation of Monte-Carlo estimators

Let us consider the example:

$$F(x) \equiv Ef_0(x,\xi) \to \min_{x \in \Re^n},$$

where

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

y_i are random and normally $N(x_i, d^2)$ 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 $2 \le n \le 100$.

We examine four estimators for stochastic gradient following from the expression (3). Assume for simplicity the sample (4) being standard normal in our example. First estimator easily follows by the differentiation of the integral (2) with respect to x:

$$g(x,y) = \nabla_x f(x + d \cdot y) . \quad (11)$$

Since analytical gradient in (11) is not always available the finite difference approach is of interest. In this approach each i[th] component of the stochastic gradient is computed as:

$$g_i(x,y) = \frac{f(x + d \cdot y + \delta \cdot \varsigma_i) - f(x + d \cdot y)}{\delta}, \quad (12)$$

ς_i is the vector with zero components except i[th] one, equal to 1, δ is some small value.

Since expression (12) requires to compute the function $n + 1$ times the Simulated Perturbation Stochastic Approximation approach (Spall (1992)), which require only one additional function value computation was examined, too:

$$g(x,y) = \frac{f(x + d \cdot y + \delta \cdot \nu) - f(x + d \cdot y - \delta \cdot \nu)}{2 \cdot \delta}, \quad (13)$$

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

Let us consider Likelihood Ratio approach to obtain the next expression for stochastic gradient, which also requires only one additional function computation (Rubinstein and Shapiro (1983)). Namely, since random error in the example (10) is additive we may change variables in the integral (3) and evaluate stochastic gradient by:

$$g(x,y) = \frac{\left(f(x + d \cdot y) - f(x, E\xi)\right) \cdot y}{d} . \quad (14)$$

Let us consider simulation results. The optimal point is known for the task considered: $x_+ = 0$. Thus, 400 Monte-Carlo samples of size N = (50, 100, 200, 500, 1000) were generated and the T^2 – statistic in criterion (9) were computed for each sample using estimators (11)–(14). The hypothesis on the difference of empirical distribution of this statistic from the Fisher distribution was tested according to the criteria ω^2 and Ω^2. Values of ω^2 and Ω^2 statistics computation for estimator (14) on variable number and sample size are given in Table 1 and 2. The critical value is $\omega^2 = 0.46$ ($p = 0.05$), and that of the next one is $\Omega^2 = 2.49$ ($p = 0.05$). Values of statistics exceeding critical value are bolded.

As follows from simulation results the distribution of Hotelling statistics (9) can be approximated by Fisher distribution appropriately choosing sample size. The required Monte-Carlo sample size for the dimensionality of task necessary to approximate the Hotelling statistics in (9) by Fisher distribution is given in Table 3. Similar results are obtained for other estimators, too.

Besides, the dependencies of the frequency of optimality hypothesis (equality of gradient to zero) according to criterion (9) on the distance $r = \left| x - x^+ \right|$ to the optimal point were studied.

These dependencies for $n = 2$ and more variables are presented in Fig 1–5 (for confidence $\alpha = 0.95$). Thus, the computation results show that analytical (11) and difference approach (12) estimators provide a good coincidence of Hotelling statistics with Fisher distribution, which can be applied for optimality testing in a wide range of dimensionality of stochastic optimization problem ($2 \le n \le 100$).

Table 1. ω^2 criteria results by variable number 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 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. Required Monte-Carlo sample size by number of variables

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

However, Simulation Perturbation Stochastic Approximation (13) and Likelihood Ration (14) estimators can be applied for stochastic gradient estimation only for tasks of not very large dimensionality: $1 \le n \le 20$.

4. Study of two-stage stochastic problem

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

$$F(x) = c \cdot x + E\{Q(x,w)\} \to \min_{x \in D} \qquad (15)$$

subject to the feasible set

$$D = \left\{ x \middle| \; A \cdot x = b, \; x \in \Re_{+}^{n} \right\}, \qquad (16)$$

where

$$Q(x,\omega) = \min_y [q \cdot y \,|\, W \cdot y + T \cdot x \le h, \;\; y \in R_{+}^{m}], \qquad (17)$$

the vectors b, q, h and matrices A, W, T are of the appropriate dimensionality. Assume vectors q, h and matrices W, T random in general, and, consequently, depending on an

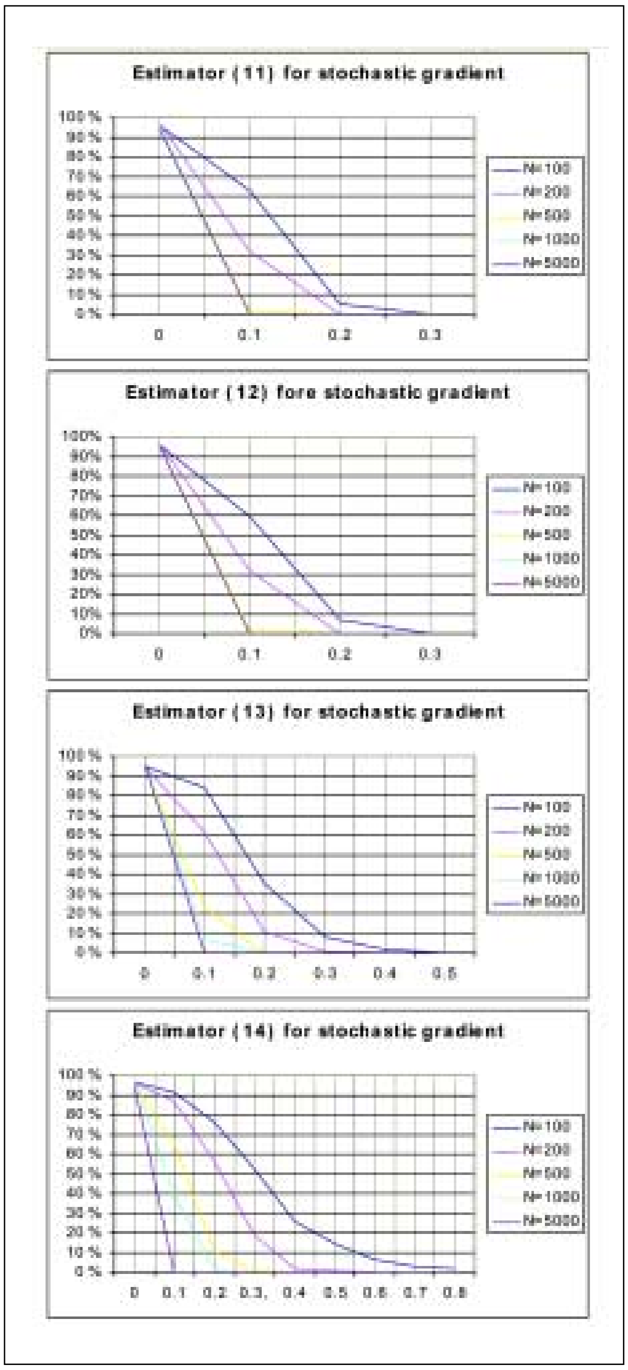

Fig 1. Frequency of optimality hypothesis ($n = 2$)

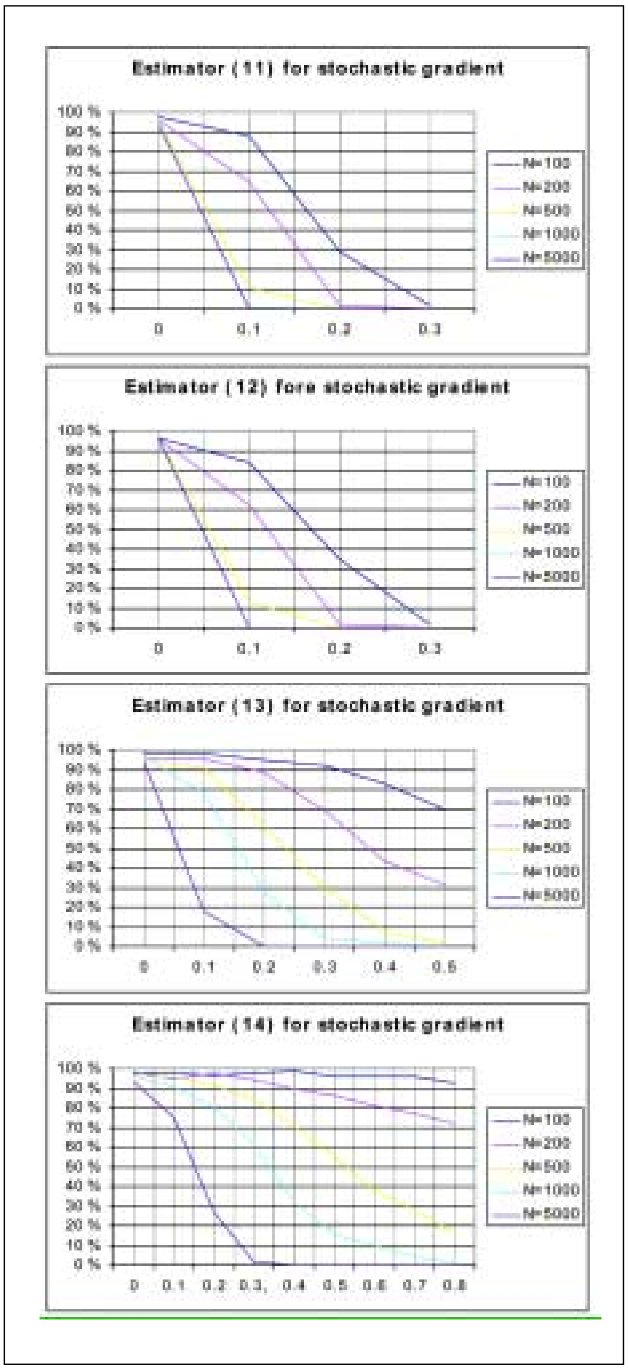

Fig 2. Frequency of optimality hypothesis ($n = 10$)

elementary event $\omega \in \Omega$ from certain probability space (Ω, Σ, P).

Let us derive the analytical gradient estimator for this problem. First, by duality of linear programming we have that:

$$F(x) = c \cdot x + E\left\{\max_u \begin{bmatrix} (h - T \cdot x) \cdot u \mid \\ u \cdot W^T + q \geq 0, \\ u \in \Re_+^n \end{bmatrix}\right\}. \quad (18)$$

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

$$\nabla_x F(x) = E(g(x, \omega)), \quad (19)$$

where $g(x, \omega) = c - T \cdot u^*$ is given by the a set of solutions of the dual problem:

$$(h - T \cdot x)^T \cdot u^* = \max_u [(h - T \cdot x)^T \cdot u \mid u \cdot W^T + q \geq 0, \quad u \in \Re^m] \quad (20)$$

(details are given in Rubinstein and Shapiro (1993), Shapiro (2000), Sakalauskas (2004), etc.)

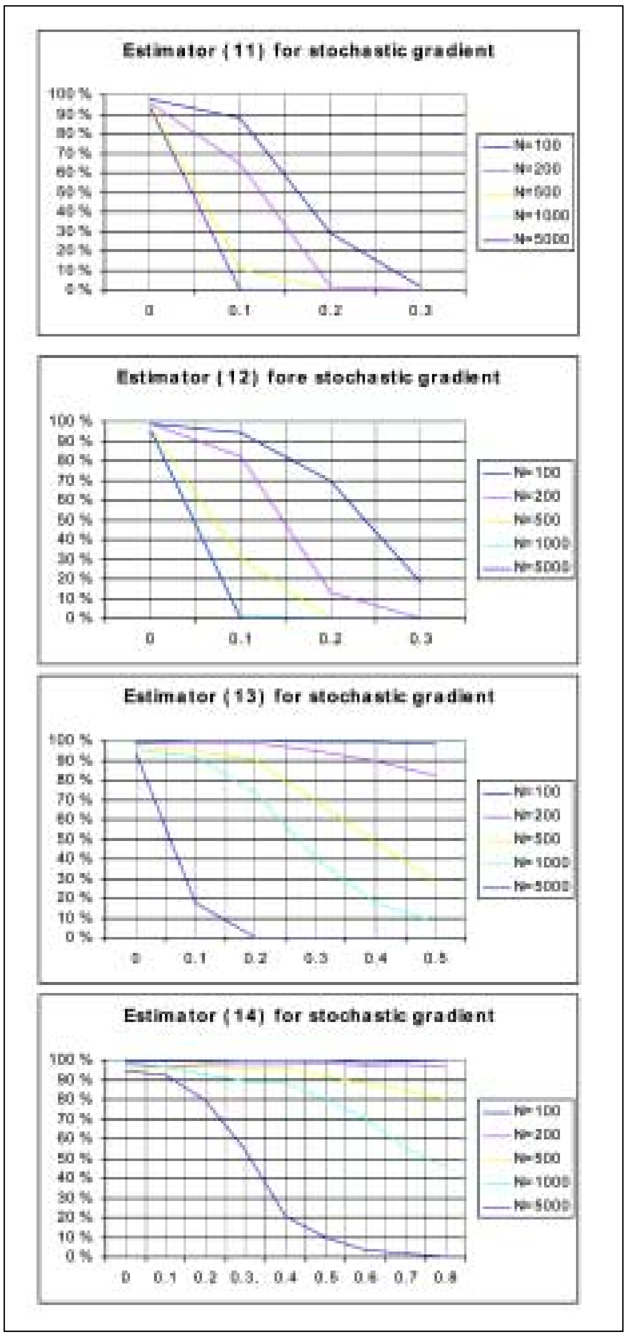

Fig 3. Frequency of optimality hypothesis (n = 20)

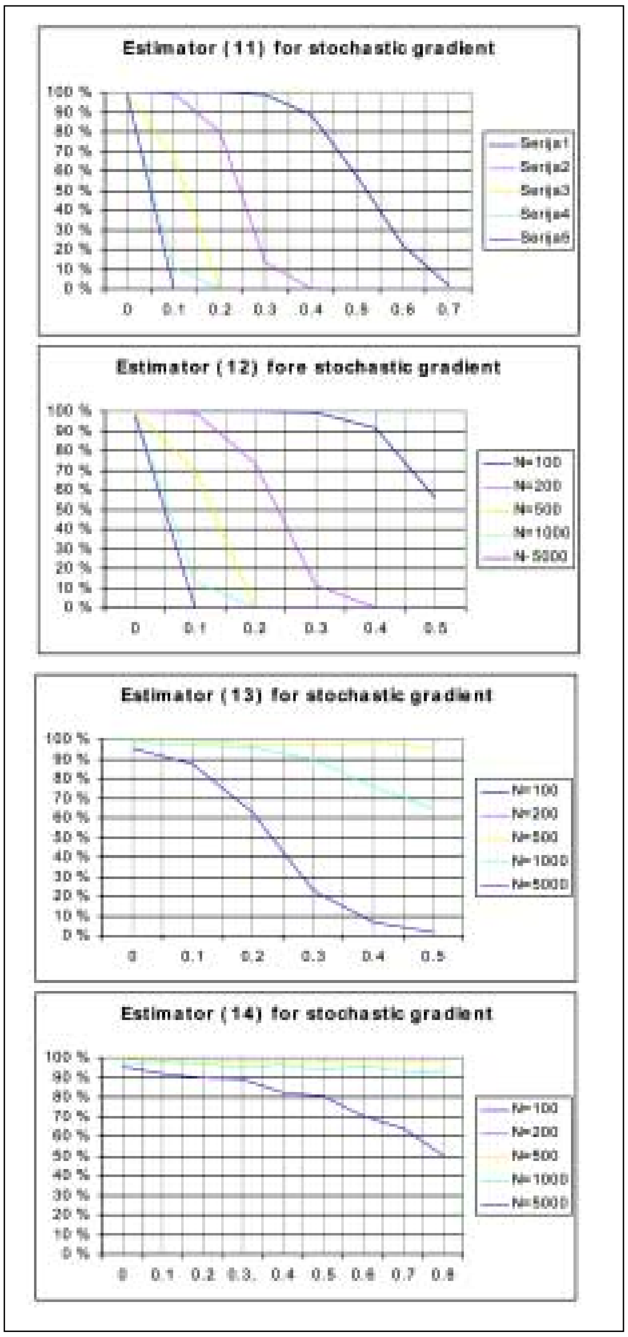

Fig 4. Frequency of optimality hypothesis (n = 50)

So the gradient search approach with projection to the ε – feasible set would be a chance to create the optimizing sequence for problem (15).

The iterative stochastic procedure of gradient search could be used (Sakalauskas (2004)):

$$x^{t+1} = x^t - \rho^t \cdot \tilde{G}(x^t), \tag{21}$$

where $\tilde{G}^t = \tilde{G}(x^t)$ is gradient projection to the ε-feasible set, ρ^t is a certain step-length multiplier. For starting point it can be obtained as the solution of the deterministic linear problem:

$$(x^0, y^0) = \arg\min_{x,y}[c \cdot x + q \cdot y \mid A \cdot x = b,$$
$$W \cdot y + T \cdot x \le h,\ y \in R_+^m, x \in R_+^n]. \tag{22}$$

Example 1. Two-stage stochastic linear optimisation problem. In this example we consider the two-stage stochastic linear optimization problem. Data of the problem are taken from a database at the address http://www.math.bme.hu/~deak/twostage/l1/20×20.2.

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.

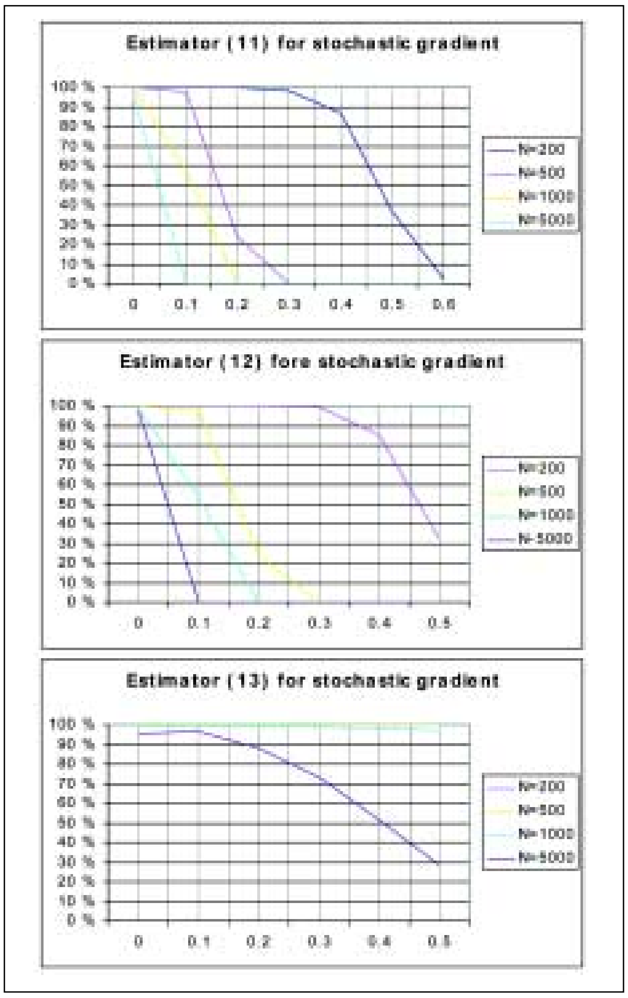

Fig 5. Frequency of optimality hypothesis (n = 100)

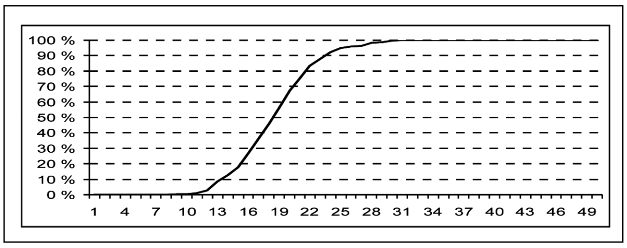

Fig 6. Frequency of termination of two-stage linear stochastic problem

The estimate of the optimal value of the objective function given in the database is 182.94234 ± 0.066. Application of the approach considered allows to improve the estimate of the optimal value up to 182.59248±0.03300.

Now, let us consider the results, obtained in solving this task 400 times by iterative approach using estimators (5)–(6) and terminating conditions (9), (10), i.e., generation of samples (4) and computation of estimators trials was broken when the estimated confidence interval of the objective function exceeds admissible value ε and the criterion (9) does not contradict to hypothesis of gradient equality to zero. Termination occurs for all times, the frequency of termination under the number of iterations is given in the Fig 6.

5. Discussion and conclusions

Thus, the computation results show that analytical (11) and difference approach (12) estimators provide a good coincidence of Hotelling statistics with Fisher distribution, which can be applied for optimality testing in a wide range of dimensionality of stochastic optimization problem ($2 \le n \le 100$). However, Simulation Perturbation Stochastic Approximation (13) and Likelihood Ration (14) estimators can be applied for stochastic gradient estimation only for tasks of not very large dimensionality: $1 \le n \le 20$.

The termination procedure proposed allows to test the optimality hypothesis and to evaluate the confidence intervals of the objective and constraint functions in a statistical way.

The numerical study and an example in practice corroborate the theoretical conclusions on the method convergence and show that the procedures developed make it possible to solve stochastic problems with a sufficient agreeable accuracy by means of the acceptable amount of computations.

References

Bhattacharya, R. N.; Ranga Rao, R. (1976). Normal Approximation and Asymptotic Expansions. John Wiley, New York, London, Toronto. 274 p.

http://www.math.bme.hu/~deak/twostage/l1/20×20.2 [revised 12 01 2006].

Yudin, D. B. (1965). Qualitative methods for analysis of complex systems (Количественные методы анализа сложных систем). Izv. AN SSSR, ser. *Technicheskaya Kibernetika*, No 1, p. 3–13.

Krishnaiah, P. R.; Lee, J. C. (1980). Handbook of Statistics, Vol 1 (Analysis of Variance). North-Holland, N.Y. 1002 p.

Marti, K. (2005). Stochastic Optimization Methods. Springer: N.Y. 316 p.

Polyak, B. T. (1987). Introduction to optimization. Translations Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division, New York. 677 p.

Rubinstein, R, and Shapiro, A. (1993). Discrete events systems: sensitivity analysis and stochastic optimization by the score function method. Wiley & Sons, N. Y, p. 353.

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

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

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

Shapiro, A. (2000). Stochastic Programming by Monte-Carlo simulation methods. Stochastic Programming E-Print Series, p. 1–29.

Spall J. C. (1992). Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, Vol 37, p. 332–341.

STOCHASTINIO PROGRAMAVIMO STATISTINIŲ KRITERIJŲ TAIKYMAS OPTIMALUMUI TESTUOTI

L. Sakalauskas, K. Žilinskas

Santrauka

Išnagrinėtas stochastinis taikomasis metodas stochastiniams tiesiniams uždaviniams spręsti naudojant baigtines Monte Karlo imtis. Šis metodas remiasi Monte Karlo imties reguliavimo taisykle ir statistine stabdymo procedūra, naudojančia statistinę modeliavimo paklaidą. Metodas skiriasi nuo kitų sprendinio tikslumo statistiniu tyrimu, optimalumo hipotezės tikrinimu, remiantis statistiniais kriterijais, ir tikslo funkcijos bei ribojimų funkcijų pasikliautinųjų intervalų įvertinimu. Imties ilgis nustatomas atvirkščiai proporcingai gradiento Monte Karlo įverčio normos kvadratui, ir tai garantuoja konvergavimą tiesiniu greičiu.

Nagrinėjami keturi stochastinio gradiento įverčiai: analitiškai diferencijuojant integralą x atžvilgiu, skirtuminiu, modeliuojamojo pokyčio ir tikėtinumo santykio metodais. Skaitinis tyrimas ir pavyzdžiai praktiškai patvirtina teorines prielaidas ir parodo, kad sukurtos procedūros leidžia spręsti stochastinius uždavinius gana tiksliai naudojant priimtiną skaičiavimų apimtį.

Reikšminiai žodžiai: tiesinis programavimas, stochastinis programavimas, optimizavimas, statistiniai kriterijai, Monte Karlo metodas.

Leonidas SAKALAUSKAS. Doctor Habil, Professor. Department of Information Technologies, Vilnius Gediminas Technical University.

Doctor (1974), Kaunas University of Technology. Doctor Habil (2000), Institute of Mathematics and Informatics. Professor (2005). Research visits to International Centre of Theoretical Physics (ICTP) (Italy, 1996, 1998).

L.Sakalauskas is a member of the New-York Academy of Sciences (1997), vice-president of the Lithuanian Operational Research society (2001), Elected Member of the International Statistical Institute (2002), member of International Association of Official Statistic (2001), member of European Working Groups on Continuous Optimization, Financial Modelling and Multicriterial Decisions. Author of more than 100 scientific articles.

Research interests: continuous optimization, stochastic approximation, data mining, Monte-Carlo method, optimal design.

Kęstutis ŽILINSKAS. PhD student of Department of Operational Research of the Institute of Mathematics and Informatics.

First degree in mathematics and physics, Šiauliai Pedagogical Institute (1981). Master of Science (1994). First degree in informatics, Šiauliai University (1999). Author of 5 scientific articles.

Research interests: linear and stochastic programming, Monte-Carlo method, optimal design.