Theory of Stochastic Processes
Vol. 11 (27), no. 3–4, 2005, pp. 97–103

UDC 519.21

P. S. KNOPOV AND E. J. KASITSKAYA

ON LARGE DEVIATIONS IN ESTIMATION PROBLEM WITH DEPENDENT OBSERVATIONS

The paper is devoted to the stochastic optimization problem with a stationary ergodic random sequence satisfying the hypermixing condition. It is assumed that we have the finite number of observed elements in the sequence, and instead of solving the former problem we investigate the empirical function, find its points of minimum, and study their asymptotic properties. More precisely we consider the probabilities of large deviations of minimizers and the minimal value of the empirical criterion function from the corresponding characteristics of the main problem. The conditions under which the probabilities of the large deviations decrease exponentially are found.

We investigate the stochastic optimization problem: minimize

$$Ef(x) = Ef(x,\xi_0), \quad x \in X, \tag{1}$$

where $\{\xi_i, i \in Z\}$ is a stationary in a strict sense ergodic random sequence defined on a probability space (Ω, F, P) and with values in some measurable space $(Y, \Im)$, X is a compact subset of some Banach space S with the norm $\|\cdot\|$, $f : X * Y \to R$ is some known function continuous in the first argument and measurable in the second one.

Instead of (1) we will minimize the empirical function

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} f(x,\xi_k), \quad x \in X, \tag{2}$$

with $\{\xi_k, k = 1, \dots, n\}$ observed elements of the sequence $\{\xi_i\}$.

If we have

$$E\left\{\max(|f(x,\xi_0)|, x \in X)\right\} < \infty.$$

then there exists a solution x^* to the problem (1), and we suppose that it is unique.

It is known that there exists a minimum point $x_n(\omega)$ of the function (2). Under some sufficiently simple conditions (see [1]) $x_n(\omega)$ tends to x^* with probability 1 as $n \to \infty$.

The aim of the paper is to estimate the large deviations of x_n and $F_n(x_n)$.

Let us recollect some facts from functional analysis. For any $y \in Y$ the function $f(\circ, y)$ belongs to the space $C(X)$ of continuous real functions on X. We assume that for all $y \in Y$ we have $f(\circ, y) - Ef(\circ) \in K$, where K is some convex compact set from $C(X)$. Therefore for any n $F_n(\circ) - Ef(\circ)$ is a random element defined on the probability space (Ω, F, P) and with values in K.

Definition 1. [2]. Let $(V, \|\circ\|)$ be a normed linear space, $B(x, \rho)$ - a closed ball in V with the radius ρ and the center x, $f : V \to [-\infty, +\infty]$ -some function, and $f(x_f) = \min\{f(x), x \in V\}$. *A condition function* ψ for f at x_f is a monotone increasing function

2000 *AMS Mathematics Subject Classification.* Primary 62M10, 62M40.

Key words and phrases. Large deviations, stochastic programming problem, empirical estimates, hypermixing.

$\psi : [0, +\infty) \to [0, +\infty]$ with $\psi(0) = 0$ such that for some $\rho > 0$ and for all $x \in B(x_f, \rho)$ we have

$$f(x) \geq f(x_f) + \psi\left(\|x - x_f\|\right).$$

Assume that $V_0 \subset V$, and denote by δ_{V_0} the indicator function of V_0:

$$\delta_{V_0}(x) = 0,\ x \in V_0; \quad \delta_{V_0}(x) = +\infty,\ x \notin V_0.$$

Theorem 1. [2]. *Let $(V, \|\circ\|)$ be a normed linear space, $V_0 \subset V$ closed, and $f_0, g_0 : V \to R$ be continuous functions on V. Suppose that*

$$\varepsilon = \sup\left\{\left|f_0(x) - g_0(x)\right|, x \in V_0\right\}.$$

Define the functions $f, g : V \to (-\infty, +\infty]$ in the following way: $f = f_0 + \delta_{V_0}$, $g = g_0 + \delta_{V_0}$. Then

$$|\inf\{f(x), x \in V\} - \inf\{g(x), x \in V\}| \leq \varepsilon.$$

Next, let x_f be a minimum point of f:

$$f(x_f) = \inf\{f(x), x \in V\}.$$

Assume that ψ is a condition function for f at x_f with some coefficient $\rho > 0$. If ε is sufficiently small so that for all x when $\psi\left(\|x - x_f\|\right) \leq 2\varepsilon$ we have $\|x - x_f\| \leq \rho$, then for any $x_g \in \arg\min\{g(x), x \in B(x_f, \rho)\}$ the following inequality is fulfilled:

$$\psi\left(\|x_f - x_g\|\right) \leq 2\varepsilon.$$

When ψ is convex and strictly increasing on $[0, \rho]$, the preceding inequality can also be expressed in the following way: if ε is small enough so that $\psi^{-1}(2\varepsilon) \leq \rho$, then for any $x_g \in \arg\min\{g(x), x \in B(x_f, \rho)\}$ one has

$$\|x_f - x_g\| \leq \psi^{-1}(2\varepsilon).$$

Theorem 2. [3, p.53]. *Let $\{\mu_\varepsilon : \varepsilon > 0\}$ be a family of probability measures on G, where G is a closed convex subset of a separable Banach space S. Assume that*

$$\Lambda(\lambda) \equiv \lim_{\varepsilon \to 0} \varepsilon \Lambda_{\mu_\varepsilon}(\lambda/\varepsilon)$$

exists for every $\lambda \in S^*$, where S^* is the dual space of S, and

$$\Lambda_\mu(\lambda) = \ln\left(\int_E \exp\left[\langle\lambda, x\rangle\right] \mu(dx)\right)$$

for an arbitrary probability measure μ on S , where $\langle\lambda, x\rangle$ is the corresponding duality relation. Denote

$$\Lambda^*(q) = \sup\left\{\langle\lambda, q\rangle - \Lambda(\lambda), \lambda \in S^*\right\},\ q \in G.$$

Then the function Λ^* is nonnegative, lower semicontinuous and convex, and for any compact set $A \subset G$

$$\limsup_{\varepsilon \to 0} \varepsilon \ln\left(\mu_\varepsilon(A)\right) \leq -\inf\{\Lambda^*(q), q \in A\}$$

holds.

Definition 2. [3]. Let Σ be a separable Banach space, $\{\xi_i, i \in Z\}$ –a stationary in a strict sense random sequence defined on a probability space (Ω, F, P) with values in Σ. Let B_{mk} denote the σ -algebra over Ω generated by the random elements $\{\xi_i, m \le i \le k\}$. For given $l \in N$ the real random variables $\eta_1, \dots, \eta_p$, $p \ge 2$ are called l *-measurably separated* if $-\infty \le m_1 \le k_1 < m_2 \le k_2 < \dots < m_p \le k_p \le +\infty$;

$$m_j - k_{j-1} \ge l, \quad j = 2, \dots, p$$

and for each $j \in \{1, \dots, p\}$ the random variable η_j is $B_{m_j k_j}$ -measurable.

Definition 3. [3]. A random sequence $\{\xi_i\}$ from Definition 2 is called a sequence with *hypermixing* if there exist a number $l_0 \in N \bigcup \{0\}$ and non-increasing functions

$$\alpha, \beta : \{l > l_0\} \to [1, +\infty)$$

and $\gamma : \{l > l_0\} \to [0, 1]$ which satisfy

$$\lim_{l\to\infty} \alpha(l) = 1, \quad \limsup_{l\to\infty} l(\beta(l) - 1) < \infty, \quad \lim_{l\to\infty} \gamma(l) = 0,$$

and for which

$$\|\eta_1 \dots \eta_p\|_{L^1(P)} \le \prod_{j=1}^{p} \|\eta_j\|_{L^{\alpha(l)}(P)} \tag{H-1}$$

whenever $p \ge 2$, $l > l_0$ and $\eta_1, \dots, \eta_p$ are l -measurably separated functions. Here

$$\|\eta\|_{L^r(P)} = \left(\int_\Omega |\eta(\omega)|^r \, dP \right)^{1/r},$$

and

$$\left| \int_\Omega \left(\xi(\omega) - \int_\Omega \xi(\omega) dP \right) \eta(\omega) dP \right| \le \gamma(l) \, \|\xi\|_{L^{\beta(l)}(P)} \, \|\eta\|_{L^{\beta(l)}(P)} \tag{H-2}$$

for all $l > l_0$ and $\xi, \eta \in L^1(P)$ l -measurably separated.

It is known (see [4]), that $C(X)^* = M(X)$ is the set of bounded signed measures on X, and

$$\langle g, Q \rangle = \int_X g(x) Q(dx)$$

for any $g \in C(X)$, $Q \in M(X)$.

Theorem 3. *Suppose that $\{\xi_i, i \in Z\}$ is a stationary in a strict sense ergodic random sequence satisfying the hypothesis (H-1) of the hypermixing condition, defined on a probability space (Ω, F, P) with values in a compact convex set $K \subset C(X)$.*

Then for any measure $Q \in M(X)$ there exists

$$\Lambda(Q) = \lim_{n\to\infty} \frac{1}{n} \ln \left(\int_\Omega \exp \left\{ \sum_{i=1}^{n} \int_X \xi_i(\omega)(x) Q(dx) \right\} dP \right)$$

and for any closed $A \subset K$

$$\limsup_{n\to\infty} \frac{1}{n} \ln \left(P \left\{ \frac{1}{n} \sum_{i=1}^{n} \xi_i \in A \right\} \right) \le -\inf\{\Lambda^*(g), g \in A\},$$

where $\Lambda^*(g) = \sup\left\{\int_X g(x)Q(dx) - \Lambda(Q),\, Q \in M(X)\right\}$ *is the nonnegative, lower semi-continuous and convex function.*

Proof. Let us consider any $Q \in M(X)$. Assume that l_0 is the number from the hypothesis (H-1). Fix $l > l_0$ and $m, n \in N$, where $l < m < n$. Then

$$n = N_n m + r_n,\ N_n \in N,\ r_n \in N \bigcup \{0\},\ r_n < m.$$

We will use the following notation:

$$\|g\| = \max\{|g(x)|\,,\, x \in X\},\ g \in C(X),$$

$$f_n = \ln\left(\int_\Omega \exp\left\{\sum_{i=1}^n \int_X \xi_i(\omega)(x)Q(dx)\right\} dP\right),\ c = \max\{\|g\|\,,\quad g \in K\}, \tag{3}$$

$$v(Q, X) = \sup\left\{\sum_{i=1}^k |Q(E_i)|\,,\, E_i \bigcap E_j = \emptyset, \quad i \neq j, \quad E_i \in B(X),\, k \in N\right\} < \infty,$$

where $B(X)$ is the Borel σ–algebra on X, $Q \in M(X)$, where the last formula is taken from [Dunford and Schwartz (1957)]. For all ω we have

$$\sum_{i=1}^n \int_X \xi_i(\omega)(x)Q(dx) = \sum_{j=0}^{N_n-1} \sum_{i=jm+1}^{(j+1)m-l} \int_X \xi_i(\omega)(x)Q(dx)$$

$$+ \sum_{j=0}^{N_n-1} \sum_{i=(j+1)m-l+1}^{(j+1)m} \int_X \xi_i(\omega)(x)Q(dx) + \sum_{i=N_n m+1}^{n} \int_X \xi_i(\omega)(x)Q(dx). \tag{4}$$

Further, in view of (3) for each i, ω

$$\left|\int_X \xi_i(\omega)(x)Q(dx)\right| \le cv(Q, X). \tag{5}$$

Due to (5) for any ω we have

$$\sum_{j=0}^{N_n-1} \sum_{i=(j+1)m-l+1}^{(j+1)m} \int_X \xi_i(\omega)(x)Q(dx) \le cv(Q, X)lN_n, \tag{6}$$

$$\sum_{i=N_n m+1}^{n} \int_X \xi_i(\omega)(x)Q(dx) \le cv(Q, X)r_n. \tag{7}$$

For each ω denote

$$V_1 = \sum_{j=0}^{N_n-1} \sum_{i=jm+1}^{(j+1)m-l} \int_X \xi_i(\omega)(x)Q(dx),$$

$$V_2 = \sum_{j=0}^{N_n-1} \sum_{i=(j+1)m-l+1}^{(j+1)m} \int_X \xi_i(\omega)(x)Q(dx),$$

$$V_3 = \sum_{i=N_n m+1}^{n} \int_X \xi_i(\omega)(x)Q(dx).$$

The inequalities (6) and (7) imply that

$$\exp\{V_1 + V_2 + V_3\} \le$$

$$\leq \exp\{V_1\}\exp\{cv(Q,X)lN_n\}\exp\{cv(Q,X)r_n\}, \quad \omega \in \Omega. \tag{8}$$

It follows from (8) that

$$\int_\Omega \exp\{V_1+V_2+V_3\}\,dP \leq \exp\{cv(Q,X)lN_n\}\exp\{cv(Q,X)r_n\}\int_\Omega \exp\{V_1\}\,dP.$$

Due to the properties of $\{\xi_i\}$ we obtain

$$\int_\Omega \prod_{j=0}^{N_n-1}\exp\left\{\sum_{i=jm+1}^{(j+1)m-l}\int_X \xi_i(\omega)(x)Q(dx)\right\}dP \leq$$

$$\leq \prod_{j=0}^{N_n-1}\left(\int_\Omega\left(\exp\left\{\sum_{i=jm+1}^{(j+1)m-l}\int_X \xi_i(\omega)(x)Q(dx)\right\}\right)^{\alpha(l)}dP\right)^{1/\alpha(l)}, \tag{9}$$

$$\int_\Omega \exp\left\{\alpha(l)\sum_{i=jm+1}^{(j+1)m-l}\int_X \xi_i(\omega)(x)Q(dx)\right\}dP =$$

$$= \int_\Omega \exp\left\{\alpha(l)\sum_{i=1}^{m-l}\int_X \xi_i(\omega)(x)Q(dx)\right\}dP,\ j=1,\dots,N_n-1. \tag{10}$$

In view of (9) and (10) we get

$$\int_\Omega \exp\{V_1\}\,dP \leq \left(\int_\Omega \exp\left\{\alpha(l)\sum_{i=1}^{m-l}\int_X \xi_i(\omega)(x)Q(dx)\right\}dP\right)^{N_n/\alpha(l)}.$$

¿From (4) we get

$$f_n = \ln\left(\int_\Omega \exp\{V_1+V_2+V_3\}\,dP\right) \leq cv(Q,X)lN_n + cv(Q,X)r_n +$$

$$+\ln\left[\left(\int_\Omega \exp\left\{\alpha(l)\sum_{i=1}^{m-l}\int_X \xi_i(\omega)(x)Q(dx)\right\}dP\right)^{N_n/\alpha(l)}\right]$$

$$= cv(Q,X)lN_n + cv(Q,X)r_n +$$

$$+\frac{N_n}{\alpha(l)}\ln\left(\int_\Omega \exp\left\{(\alpha(l)-1)\sum_{i=1}^{m-l}\int_X \xi_i(\omega)(x)Q(dx)+\sum_{i=1}^{m-l}\int_X \xi_i(\omega)(x)Q(dx)\right\}dP\right)\leq$$

$$\leq cv(Q,X)lN_n + cv(Q,X)r_n + \frac{N_n}{\alpha(l)}\left(\alpha(l)-1\right)(m-l)\,cv(Q,X)+$$

$$+\frac{N_n}{\alpha(l)}\ln\left(\int_\Omega \exp\left\{\sum_{i=1}^{m-l}\int_X \xi_i(\omega)(x)Q(dx)\right\}dP\right)\leq$$

$$\leq cv(Q,X)lN_n + cv(Q,X)r_n + \left(\alpha(l)-1\right)(m-l)\,cv(Q,X)N_n +$$

$$+\frac{N_n}{\alpha(l)}\ln\left(\int_\Omega \exp\left\{\sum_{i=1}^{m}\int_X \xi_i(\omega)(x)Q(dx)-\sum_{i=m-l+1}^{m}\int_X \xi_i(\omega)(x)Q(dx)\right\}dP\right)\leq$$

$$\leq cv(Q,X)lN_n + cv(Q,X)r_n + (\alpha(l)-1)\,cv(Q,X)mN_n + \frac{N_n}{\alpha(l)}cv(Q,X)l +$$

$$+\frac{N_n}{\alpha(l)}\ln\left(\int_\Omega \exp\left\{\sum_{i=1}^m \int_X \xi_i(\omega)(x)Q(dx)\right\}dP\right) \leq$$

$$\leq 2cv(Q,X)lN_n + cv(Q,X)r_n + (\alpha(l)-1)\,cv(Q,X)mN_n + \frac{N_n}{\alpha(l)}f_m\ . \quad (11)$$

The inequality (11) implies that

$$\frac{f_n}{n} \leq \frac{2N_n cv(Q,X)l}{N_n m} + \frac{cv(Q,X)r_n}{n} + (\alpha(l)-1)\,cv(Q,X) + \frac{N_n f_m}{\alpha(l)\,(N_n m + r_n)} =$$

$$= \frac{2cv(Q,X)l}{m} + \frac{cv(Q,X)r_n}{n} + (\alpha(l)-1)\,cv(Q,X) + \frac{f_m}{\alpha(l)\,(m + r_n/N_n)}.$$

Therefore

$$\limsup_{n\to\infty}\frac{f_n}{n} \leq \frac{2cv(Q,X)l}{m} + (\alpha(l)-1)\,cv(Q,X) + \frac{1}{\alpha(l)}\frac{f_m}{m}.$$

Passing to the lim inf as $m\to\infty$, we obtain

$$\limsup_{n\to\infty}\frac{f_n}{n} \leq (\alpha(l)-1)\,cv(Q,X) + \frac{1}{\alpha(l)}\liminf_{m\to\infty}\frac{f_m}{m},$$

and letting $l\to\infty$, we have

$$\limsup_{n\to\infty}\frac{f_n}{n} \leq \liminf_{m\to\infty}\frac{f_m}{m}.$$

Consequently, there exists

$$\lim_{n\to\infty}\frac{f_n}{n} = \Lambda(Q).$$

Now we can see that the theorem follows from Theorem 2. Indeed, for

$$G = K,\ S = C(X),\ S^* = M(X),\ \langle Q, g\rangle = \int_X g(x)Q(dx),\ \varepsilon = \frac{1}{n},$$

and for $\mu_\varepsilon = \mu_{1/n}$ the probability measure on K, defined by the distribution of a random element $\frac{1}{n}\sum_{i=1}^n \xi_i$, we have

$$\lim_{\varepsilon\to 0}\varepsilon\Lambda_{\mu_\varepsilon}(Q/\varepsilon) = \lim_{n\to\infty}\frac{1}{n}\ln\left(\int_K \exp\left\{\int_X g(x)nQ(dx)\right\}\mu_{1/n}(dg)\right) =$$

$$(12)\qquad = \lim_{n\to\infty}\frac{1}{n}\ln\left(\int_\Omega \exp\left\{\int_X \frac{1}{n}\sum_{i=1}^n \xi_i(\omega)(x)nQ(dx)\right\}dP\right) = \lim_{n\to\infty}\frac{f_n}{n} = \Lambda(Q).$$

The proof is complete.

Now let us consider the problems (1) and (2). Suppose that the given sequence $\{\xi_i,\ i\in Z\}$ satisfies the hypothesis (H-1) of the hypermixing condition. Then the sequence

$$\zeta_i = f(\circ,\xi_i) - Ef(\circ),\quad i\in Z,$$

satisfies (H-1) too.

Denote

$$A_\varepsilon = \{z\in K : \|z\| \geq \varepsilon\},$$

$$I(z) = \Lambda^*(z) = \sup\left\{ \int_X z(x)Q(dx) - \Lambda(Q),\, Q \in M(X) \right\}.$$

Theorem 4. *Under the hypothesis (H-1) of the hypermixing condition we have*

$$\limsup_{n\to\infty} \frac{1}{n} \ln P\left\{ |\min\{Ef(x),\, x \in X\} - \min\{F_n(x),\, x \in X\}| \geq \varepsilon \right\}$$

$$\leq -\inf\{I(z),\, z \in A_\varepsilon\}. \tag{13}$$

Assume that there exists a condition function ψ for Ef at x^* with some constant ρ. Let x_n be a point of the minimum of the function (2) on the set $B(x^*, \rho)$. If ε is sufficiently small so that the condition

$$\psi(\|x - x^*\|) \leq 2\varepsilon \;\Rightarrow\; \|x - x^*\| \leq \rho,$$

is fulfilled, then

$$\limsup_{n\to\infty} \frac{1}{n} \ln P\{\psi(\|x_n - x^*\|) \geq 2\varepsilon\} \leq -\inf\{I(z),\, z \in A_\varepsilon\}. \tag{14}$$

Moreover, if ψ is convex and strictly increasing on $[0, \rho]$, then

$$\limsup_{n\to\infty} \frac{1}{n} \ln P\left\{ \|x_n - x^*\| \geq \psi^{-1}(2\varepsilon) \right\} \leq -\inf\{I(z),\, z \in A_\varepsilon\}. \tag{15}$$

Proof. Theorem 1 implies that for each ω

$$|\min\{Ef(x),\, x \in X\} - \min\{F_n(x),\quad x \in X\}| \leq \|F_n - Ef\|. \tag{16}$$

Then, the conditions of Theorem 3 are fulfilled for the sequence $\{\varsigma_i\}$. Therefore for any $\varepsilon > 0$

$$\limsup_{n\to\infty} \frac{1}{n} \ln P\{\|F_n - Ef\| \geq \varepsilon\} \leq -\inf\{I(z),\, z \in A_\varepsilon\}. \tag{17}$$

The inequality (13) follows from (16) and (17). For the proof of the second part of the theorem we also use Theorem 1. Under the conditions of the theorem we have for all ω

$$\psi(\|x^* - x_n\|) \leq 2\|F_n - Ef\|, \tag{18}$$

or

$$\|x_n - x^*\| \leq \psi^{-1}(2\|F_n - Ef\|). \tag{19}$$

Taking into account (17), the inequalities (18) and (19) imply (14) and (15) respectively. The theorem is proved.

Bibliography

1. Knopov, P.S. and Kasitskaya, E.J., *Empirical estimates in stochastic optimization and identification*, Kluwer Academic Publishers,, Dordrecht, 2002.
2. Kaniovski, Yu.M., King, A.J. and Wets, R.J-B., *"Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems"*, Ann.Oper.Res. 56, 189 - 208, 1995.
3. Deuschel, J.-D. and Stroock, D.W., *"Large Deviations"*, Academic Press, Boston, 1989.
4. Dunford, N. and Schwartz, J., *"Linear Operators, Part I: General Theory"*, Interscience Publishers, New York, 1958.

V.M.Glushkov Institute of Cybernetics National Academy of Science of Ukraine, 40, Acad. V.M.Glushkov Ave., 03187, Kiev, Ukraine

E-mail: knopov1@yahoo.com