Contributions to the Asymptotic Minimax Theorem for the Two-Armed Bandit Problem

A.V. Kolnogorov

Abstract—The asymptotic minimax theorem for Bernoully two-armed bandit problem states that the minimax risk has the order $N^{1/2}$ as $N \to \infty$, where N is the control horizon, and provides lower and upper estimates. It can be easily extended to normal two-armed bandit. For normal two-armed bandit, we generalize the asymptotic minimax theorem as follows: the minimax risk is approximately equal to $0.637N^{1/2}$ as $N \to \infty$.

Keywords—two-armed bandit problem, control in a random environment, minimax and bayesian approaches, an asymptotic minimax theorem, parallel processing.

I. Introduction

We consider the two-armed bandit problem (see, e.g. [1], [2]) which is also well-known as the problem of expedient behavior in a random environment (see, e.g. [3], [4]) and the problem of adaptive control (see, e.g. [5], [6]) in the following setting. Let ξ_n, $n = 1, \dots, N$ be a controlled random process which values are treated as incomes, depend only on currently chosen actions y_n ($y_n \in \{1, 2\}$) and are normally distributed with probability densities

$$f(x|m_\ell) = (2\pi)^{-1/2} \exp\left\{-(x - m_\ell)^2/2\right\},$$

if $y_n = \ell$ ($\ell = 1, 2$). So, this is the so-called normal (or Gaussian) two-armed bandit. It can be described by a vector parameter $\theta = (m_1, m_2)$. The goal is to maximize (in some sense) the total expected income. Control strategy σ at the point of time n assigns a random choice of the action y_n depending on the current history of the process, i.e. replies $x^{n-1} = x_1, \dots, x_{n-1}$ to applied actions $y^{n-1} = y_1, \dots, y_{n-1}$:

$$\Pr(y_n = \ell | y^{n-1}, x^{n-1}) = \sigma_\ell(y^{n-1}, x^{n-1}),$$

$\ell = 1, 2$. The set of strategies is denoted by Σ.

If parameter θ is known then the optimal strategy should always apply the action corresponding to the larger value of m_1, m_2. The total expected income would thus be equal to $N(m_1 \vee m_2)$. If parameter is unknown then the loss function

$$L_N(\sigma, \theta) = N(m_1 \vee m_2) - E_{\sigma,\theta}\left(\sum_{n=1}^{N} \xi_n\right)$$

describes expected losses of total income with respect to its maximal possible value due to incomplete information. Here $E_{\sigma,\theta}$ denotes the mathematical expectation calculated over the measure generated by a strategy σ and a parameter θ. The set of parameters is assumed to be the following

$$\Theta = \{\theta : |m_1 - m_2| \le 2C\},$$

where $0 < C < \infty$. Restriction $C < \infty$ ensures the boundedness of the loss function on Θ.

According to the minimax approach the maximal value of the loss function on the set of parameters Θ should be minimized over the set of strategies Σ. The value

$$R_N^M(\Theta) = \inf_{\Sigma} \sup_{\Theta} L_N(\sigma, \theta) \tag{1}$$

is called the minimax risk and corresponding strategy (if it exists) is called the minimax strategy. The minimax approach to the problem was proposed by H. Robbins in [7]. This article caused a significant interest to considered problem. The classic object of the most of arisen articles was the so-called Bernoulli two-armed bandit which can be described by distribution

$$\Pr(\xi_n = 1 | y_n = \ell) = p_\ell, \quad \Pr(\xi_n = 0 | y_n = \ell) = q_\ell,$$

$p_\ell + q_\ell = 1$, $\ell = 1, 2$. Such bandit is described by a parameter $\theta = (p_1, p_2)$ with the set of values $\Theta = \{\theta : 0 \le p_\ell \le 1; \ell = 1, 2\}$. It was shown in [8] that explicit determination of the minimax strategy and minimax risk is practically impossible already for $N > 4$. However, the following asymptotic minimax theorem was proved by W. Vogel in [9]:

Theorem 1: The following estimates hold as $N \to \infty$ for Bernoulli two-armed bandit:

$$0.612 \le (DN)^{-1/2} R_N^M(\Theta) \le 0.752 \tag{2}$$

with $D = 0.25$ being the maximal variance of one-step income. Presented here the lower estimate was obtained in [10]. The upper estimate was obtained in [9] for the following strategy.

Thresholding strategy. *Use actions turn-by-turn until the absolute difference between total incomes for their applications exceeds the value of the threshold $\alpha(DN)^{1/2}$ or the control time expires. If the threshold has been achieved and the control time has not expired then at the rest of the control horizon use only the action corresponding to the larger value of total initial income.* The optimal value of α is $\alpha \approx 0.584$ and the maximal value of expected losses corresponds to $|m_1 - m_2| \approx 3.78(D/N)^{1/2}$.

Remark 1: The estimates (2) can be easily extended to considered normal two-armed bandit with a glance that $D = 1$ in this case.

There are some different approaches to robust control in the two-armed and multi-armed bandit problems, see, e.g.

A.V. Kolnogorov is with Applied Mathematics and Information Sciences Department of Yaroslav-the-Wise Novgorod State University, Velikiy Novgorod, 173003, Russia, e-mail: Alexander.Kolnogorov@novsu.ru

[6], [11], [12], [13]. In these articles stochastic approximation method and mirror descent algorithm are used for the control. Instead of minimax risk, the authors often consider the equivalent attitude called the guaranteed rate of convergency. The order of the minimax risk for these algorithms is $N^{1/2}$ or close to $N^{1/2}$.

The goal of the present paper is to improve the estimates (2) for the normal two-armed bandit. We propose the following estimate:

$$\lim_{N\to\infty}(DN)^{-1/2}R_N^M(\Theta)\approx 0.637. \tag{3}$$

The structure of the paper is the following. In section II we present the approach based on the main theorem of the theory of games. This approach allows to determine explicitly the value of the minimax risk as Bayesian one calculated over the worst-case prior distribution. In section III we present the estimate (3) for the case of close expectations $|m_1-m_2|\le 2cN^{-1/2}$. In section IV we consider the control at the initial stage which allows to generalize the estimate (3) to distributions which expectations are not obligatory close, i.e. $|m_1-m_2|\le 2C$. Section V contains conclusion.

II. Main Theorem of the Theory of Games Based on Approach

Another well-known approach to the problem is a Bayesian one. Denote by Λ a prior distribution of the parameter on the set Θ. The value

$$R_N^B(\Lambda)=\inf_{\Sigma}\int\limits_{\Theta}L_N(\sigma,\theta)\Lambda(d\theta) \tag{4}$$

is called Bayesian risk and corresponding strategy is called Bayesian strategy. Bayesian approach is very popular one because it allows to write recursive equations for determination of both Bayesian strategy and Bayesian risk by a dynamic programming technique. Both minimax and Bayesian approaches are integrated by the main theorem of the theory of games. According to this theorem the minimax risk (1) is equal to the Bayesian risk (4) calculated over the worst prior distribution corresponding to the maximum of the Bayesian risk. And the minimax strategy is equal to corresponding Bayesian strategy as well.

Determination of the minimax strategy and minimax risk as Bayesian ones corresponding to the worst-case prior distribution was considered in [14], [15], [16], [17]. To present these results, it is convenient to modify parameterization. Let $m_1=m+v$, $m_2=m-v$, then $\theta=(m+v,m-v)$ and $\Theta=\{\theta:|v|\le C\}$. It was proved in [14], [15] that asymptotically the worst prior distribution density can be chosen the following

$$\nu_a(m,v)=\kappa_a(m)\rho(v), \tag{5}$$

where $\kappa_a(m)$ is the uniform density on the interval $|m|\le a$, $\rho(v)$ is a symmetric density (i.e. $\rho(-v)=\rho(v)$) on the interval $|v|\le C$ and $a\to\infty$.

In the sequel, we consider strategies which at the initial stage apply both actions turn-by-turn M_0 times and the apply each chosen action M times. If incomes arise sequentially, one-by-one, these strategies allow to switch actions more rarely. If incomes arise by groups, these strategies allow their parallel processing.

Denote by n_1, n_2 total numbers of both actions applications, by X_1, X_2 corresponding total incomes. The above kind of a prior distribution density (5) results in the fact that control at the time point $n=n_1+n_2$ is completely described by a triple (U,n_1,n_2) with $U=(X_1n_2-X_2n_1)n^{-1}$.

Determination of the Bayesian strategy and Bayesian risk may be done as follows. Let's introduce the following change of variables: $\varepsilon_0=M_0N^{-1}$, $\varepsilon=MN^{-1}$, $t_1=n_1N^{-1}$, $t_2=n_2N^{-1}$, $t=nN^{-1}$, $u=UN^{-1/2}$, $w=vN^{1/2}$, $c=CN^{1/2}$, $\varrho(w)=N^{1/2}\rho(v)$. Denote by $f_D(x):=(2\pi D)^{-1/2}\exp(-x^2/(2D))$ a probability density of normal distribution. The following theorem, which was proved in [14], [15], holds.

Theorem 2: The optimal strategy at initial stage $t\le 2\varepsilon_0$ $(n\le 2\varepsilon_0N)$ applies actions turn-by-turn. In the sequel it can be determined by solving the following recursive Bellman-type equation:

$$r_\varepsilon(u,t_1,t_2)=\min_{\ell=1,\,2}r_\varepsilon^{(\ell)}(u,t_1,t_2), \tag{6}$$

where $r_\varepsilon^{(1)}(u,t_1,t_2)=r_\varepsilon^{(2)}(u,t_1,t_2)=0$ if $t_1+t_2=1$ and then

$$\begin{gathered}
r_\varepsilon^{(1)}(u,t_1,t_2)=\varepsilon g^{(1)}(u,t_1,t_2)\\
+\int\limits_{-\infty}^{\infty}r_\varepsilon(x,t_1+\varepsilon,t_2)f_{\varepsilon t_2^2t^{-1}(t+\varepsilon)^{-1}}(u-x)dx,\\
r_\varepsilon^{(2)}(u,t_1,t_2)=\varepsilon g^{(2)}(u,t_1,t_2)\\
+\int\limits_{-\infty}^{\infty}r_\varepsilon(x,t_1,t_2+\varepsilon)f_{\varepsilon t_1^2t^{-1}(t+\varepsilon)^{-1}}(u-x)dx
\end{gathered} \tag{7}$$

if $t_1+t_2<1$. Here

$$\begin{gathered}
g^{(\ell)}(u,t_1,t_2)\\
=\int_0^c 2w\exp\left((-1)^\ell 2uw-2w^2t_1t_2t^{-1}\right)\varrho(w)dw,
\end{gathered}$$

$\ell=1,2$. When $t>2\varepsilon_0$ $(n>2\varepsilon_0N)$ then the ℓ-th action is currently optimal, iff $r_\varepsilon^{(\ell)}(u,t_1,t_2)$ has smaller value $(\ell=1,2)$. Bayesian risk corresponding to the worst-case prior distribution is calculated according to the formula

$$N^{-1/2}\lim_{a\to\infty}R_N^B(\nu_a(m,v))=2l(\varrho,\varepsilon_0)+s_\varepsilon(\varrho,\varepsilon_0), \tag{8}$$

where

$$\begin{gathered}
2l(\varrho,\varepsilon_0)=4\varepsilon_0\int_0^c w\varrho(w)dw,\\
s_\varepsilon(\varrho,\varepsilon_0)=\int_{-\infty}^{\infty}r_\varepsilon(u,\varepsilon_0,\varepsilon_0)f_{0,5\varepsilon_0}(u)du
\end{gathered}$$

are expected losses at initial $(t\le 2\varepsilon_0)$ and at final $(t>2\varepsilon_0)$ stages of control respectively.

III. Limiting description

Let's denote by $r(\varrho; u, t_1, t_2)$ the Bayesian risk dependent on a prior distribution $\varrho(w)$. The following results were obtained in [15], [16], [17] for the set of close expectations $\Theta_N = \{|m_1 - m_2| \le 2cN^{-1/2}\}$.

Lemma 1: For all u, t_1, t_2, for which the solution to equation (6), (7) is well defined, there exist limits $r(\varrho; u, t_1, t_2) = \lim\limits_{\varepsilon \to 0} r_\varepsilon(\varrho; u, t_1, t_2)$, which can be extended by continuity to all u, t_1, t_2 $(t_1 > 0,\ t_2 > 0,\ t_1 + t_2 < 1)$. These limits are uniformly bounded and satisfy Lipschitz conditions in u.

Remark 2: This means that control becomes almost optimal if ε_0, ε are small enough, e.g. equal to 0.02. Actually, it means that control is almost optimal if it is implemented by groups in $0.02^{-1} = 50$ stages.

Theorem 3: The minimax risk on the set of close expectations $\Theta_N = \{|m_1 - m_2| \le 2cN^{-1/2}\}$ satisfies the estimate

$$\lim_{N \to \infty} N^{-1/2} R_N^M(\Theta_N) = \sup_{\varrho} r(\varrho; 0, 0, 0), \tag{9}$$

where $r(\varrho; 0, 0, 0) = \lim\limits_{\varepsilon_0 \to 0} r(\varrho; 0, \varepsilon_0, \varepsilon_0)$.

The formula (9) allows to obtain the estimate (3) on the set of close expectations. However, we need formulas to calculate the limiting Bayesian risk $r(u, t_1, t_2)$. Let's assume that $r_\varepsilon(u, t_1, t_2)$ has continuous partial derivatives of proper orders and show that equations (7) may be reduced to the form

$$\begin{aligned}
& r_\varepsilon^{(1)}(u, t_1, t_2) = r_\varepsilon(u, t_1 + \varepsilon, t_2) \\
& + \frac{\varepsilon t_2^2}{2t(t+\varepsilon)} \times \frac{\partial^2 r_\varepsilon(u, t_1 + \varepsilon, t_2)}{\partial u^2} \\
& + \varepsilon g^{(1)}(u, t_1, t_2) + o(\varepsilon), \\
& r_\varepsilon^{(2)}(u, t_1, t_2) = r_\varepsilon(u, t_1, t_2 + \varepsilon) \\
& + \frac{\varepsilon t_1^2}{2t(t+\varepsilon)} \times \frac{\partial^2 r_\varepsilon(u, t_1, t_2 + \varepsilon)}{\partial u^2} \\
& + \varepsilon g^{(2)}(u, t_1, t_2) + o(\varepsilon).
\end{aligned} \tag{10}$$

Let's check up the first equation (10). For this purpose we present $r_\varepsilon(u - x, t_1 + \varepsilon, t_2)$ as Taylor series:

$$\begin{aligned}
r_\varepsilon(u - x, \cdot) = r_\varepsilon(u, \cdot) - x \times \frac{\partial r_\varepsilon(u, \cdot)}{\partial u} \\
+ \frac{x^2}{2} \times \frac{\partial^2 r_\varepsilon(u, \cdot)}{\partial u^2} + o(x^2).
\end{aligned} \tag{11}$$

Noting that

$$\int_{-\infty}^{\infty} f_\varepsilon(x) dx = 1, \int_{-\infty}^{\infty} x f_\varepsilon(x) dx = 0, \int_{-\infty}^{\infty} x^2 f_\varepsilon(x) dx = \varepsilon,$$

and substituting (11) into the first equation (7), one obtains

$$\begin{aligned}
& r_\varepsilon^{(1)}(u, t_1, t_2) = \varepsilon g^{(1)}(u, t_1, t_2) \\
& + \int\limits_{-\infty}^{\infty} r_\varepsilon(u - x, t_1 + \varepsilon, t_2) f_{\varepsilon t_2^2 t^{-1} (t+\varepsilon)^{-1}}(x) dx \\
& = \varepsilon g^{(1)}(u, t_1, t_2) + r_\varepsilon(u, t_1 + \varepsilon, t_2) \\
& + \frac{\varepsilon t_2^2}{2t(t+\varepsilon)} \times \frac{\partial^2 r_\varepsilon(u, t_1 + \varepsilon, t_2)}{\partial u^2} + o(\varepsilon),
\end{aligned}$$

i.e. the first equation (10) is valid. The validity of the second equation (10) is checked up in a similar way. Recall now that equations (10) should be complemented by equation (6) which can be written as

$$\min_{\ell = 1,\, 2} (r_\varepsilon^{(\ell)}(u, t_1, t_2) - r_\varepsilon(u, t_1, t_2)) = 0,$$

and as $\varepsilon \downarrow 0$ we obtain the differential equation for $r = r(u, t_1, t_2)$:

$$\min_{\ell = 1,\, 2} \left(\frac{\partial r}{\partial t_\ell} + \frac{t_{\bar{\ell}}^2}{2t^2} \times \frac{\partial^2 r}{\partial u^2} + g^{(\ell)}(u, t_1, t_2) \right) = 0 \tag{12}$$

with $\bar{\ell} = 3 - \ell$ and with initial and boundary conditions

$$\begin{aligned}
r(u, t_1, t_2) \|_{t_1 + t_2 = 1} = 0, \\
r(\infty, t_1, t_2) = r(-\infty, t_1, t_2) = 0.
\end{aligned} \tag{13}$$

Note that the ℓ-th action should be chosen if the ℓ-th member in the left-hand side of (12) has minimal value.

To calculate numerically $r(u, t_1, t_2)$ one should use the following equation based on (6), (10):

$$r(u, t_1, t_2) = \min_{\ell = 1,\, 2} r^{(\ell)}(u, t_1, t_2), \tag{14}$$

$$\begin{aligned}
& r^{(1)}(u, t_1, t_2) = r(u, t_1 + \Delta t, t_2) \\
& + \Delta t \left(\frac{t_2^2 \times D^2 r(u, t_1 + \Delta t, t_2)}{2t(t + \Delta t)} + g^{(1)}(u, t_1, t_2) \right), \\
& r^{(2)}(u, t_1, t_2) = r(u, t_1, t_2 + \Delta t) \\
& + \Delta t \left(\frac{t_1^2 \times D^2 r(u, t_1, t_2 + \Delta t)}{2t(t + \Delta t)} \cdot + g^{(2)}(u, t_1, t_2) \right).
\end{aligned} \tag{15}$$

with

$$D^2 r(u) = \frac{r(u + \Delta u) - 2r(u) + r(u - \Delta u)}{\Delta u^2},$$

and with initial and boundary conditions (13).

Calculations of $r(\varrho; u, t_1, t_2)$ according to formulas (14), (15) were implemented with $\Delta u = 0.023$, $\Delta t = 2000^{-1}$, $\varepsilon_0 = 0.001$ for $|u| \le 2.3$. It was assumed that $\varrho(w)$ is degenerated distribution density concentrated at two points $w = \pm d$. For $0.5 \le d \le 2.5$ maximum of $2d\varepsilon_0 + r(\varrho; 0, \varepsilon_0, \varepsilon_0)$ was approximately equal to 0.637 at $d \approx 1.57$.

IV. Parallel Processing and Control at the Initial Stage

First, let's explain why normal two-armed bandit is considered. The problem is investigated in application to control of large data items processing. Let $T = NK$ items of data be given which may be processed by one of two alternative methods. Processing may be successful ($\xi'_t = 1$) or unsuccessful ($\xi'_t = 0$). Probabilities of successful and unsuccessful processing depend only on chosen methods (actions), i.e. $\Pr(\xi'_t = 1 | y_t = \ell) = p_\ell$, $\Pr(\xi'_t = 0 | y_t = \ell) = q_\ell$, $\ell = 1, 2$. Assume that one knows that p_1, p_2 are close to p $(0 < p < 1)$. We partition all data items into N packages each containing K data items. For parallel data processing in each package we use the same method. For control we use the values of

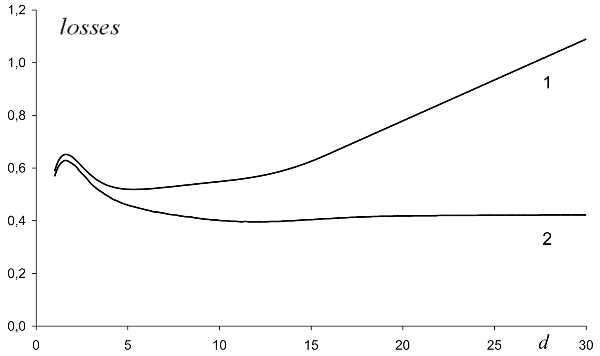

Fig. 1. Ordinary control and control using modified Vogel's strategy

the process $\xi_n = (DK)^{-1/2} \sum_{t=(n-1)K+1}^{nK} \xi'_t,\ n = 1,\dots,N$ with $D = p(1-p)$. According to the central limit theorem distributions of ξ_n, $n = 1,\dots,N$, are close to normal ones and their variances are close to unity just like in considered setup.

This control is described by equation (6), (7) with $\varepsilon = \varepsilon_0 = N^{-1}$ and may be close to optimal for moderate N, e.g. for $N = 50$. However, at initial stage actions are applied turn-by-turn. Therefore corresponding losses are equal to $K|p_1 - p_2|$ and may be greater than the minimax risk $0.637(DT)^{1/2}$ for sufficiently distant p_1, p_2. For example, if $T = 50000$, $N = 50$, $K = 1000$, $D = 0.25$, $|p_1 - p_2| = 0.1$ then $K|p_1 - p_2| = 100 > 71.2 \approx 0.637(DT)^{1/2}$. To avoid this situation one should require the closeness of expectations or modify control at the inial stage in order to early determine significant difference of p_1, p_2 and then to apply the action corresponding to the larger value of p_1, p_2 till the end of the control.

In [14] the following strategy is proposed. Assume that one should process $T = 125000$ items of data and $N = 50$, $K = 2500$, p_1, p_2 are close to $p = 0,5$ and $|p_1 - p_2| \le 0,2$ which corresponds to $d \le 70$. On Fig. 1 the line 1 describes losses if the ordinary strategy is applied and its almost linear growth for $d > 20$ is caused by equal application of both actions at initial stage by 2500 times. The line 2 corresponds to losses provided by the following modified Vogel's strategy. *Let's apply actions turn-by-turn until the absolute difference of total incomes for their application exceeds the threshold a or the initial stage of control expires. If the difference of incomes exceeds the threshold then we apply the action corresponding to the larger initial income till the end of the control horizon. Otherwise we use considered in the paper Bayesian strategy corresponding to the worst-case prior distribution.* The losses described by line 2 were obtained for $a = 70$. It means that modified strategy allows to process distributions with close and distant expectations.

Remark 3: All reasonings hold true if we assume that $\{\xi'_t\}$ is normally distributed process. This case corresponds to pure normal two-armed bandit.

V. Conclusion

A generalization to the asymptotic minimax theorem for normal two-armed bandit has been given. A proposed strategy separates distributions with distant mathematical expectations. In this case it determines the superior action at the initial stage and provides its application till the end of the control. In case of distributions with close mathematical expectations it applies the Bayesian strategy corresponding to the worst-case prior distribution. Results can be applied to parallel processing of data.

Acknowledgment

This work was supported in part by the Project Part of the State Assignment in the Field of Scientific Activity by the Ministry of Education and Science of the Russian Federation, project no. 1.949.2014/K, and by the Russian Foundation for Basic Research, project no. 13-01-00334-a.

References

[1] D. A. Berry and B. Fristedt, *Bandit Problems: Sequential Allocation of Experiments.* London, New York: Chapman and Hall, 1985.

[2] E. L. Presman and I. M. Sonin, *Sequential Control with Incomplete Information.* New York: Academic Press, 1990.

[3] M. L. Tsetlin, *Automation Theory and Modeling of Biological Systems.* New York: Academic Press, 1973.

[4] V. I. Varshavsky, *Collective Behavior of Automata.* Moscow: Nauka, 1973. (In Russian)

[5] V. G. Sragovich, *Mathematical Theory of Adaptive Control.* New Jersey, London: World Scientific. Interdisciplinary Mathematical Sciences, Vol. 4. 2006.

[6] A. V. Nazin and A. S. Poznyak, *Adaptive Choice of Alternatives.* Moscow: Nauka, 1986. (In Russian)

[7] H. Robbins, *Some Aspects of the Sequential Design of Experiments.* Bulletin AMS., Vol. 58(5), pp. 527 - 535, 1952.

[8] J. Fabius and W. R. van Zwet, *Some Remarks on the Two-Armed Bandit.* Ann. Math. Statist., Vol. 41, pp. 1906 - 1916, 1970.

[9] W. Vogel, *An Asymptotic Minimax Theorem for the Two-Armed Bandit Problem.* Ann. Math. Stat., Vol. 31, pp. 444 - 451, 1960.

[10] J. A. Bather, *The Minimax Risk for the Two-Armed Bandit Problem.* Mathematical Learning Models - Theory and Algorithms. Lecture Notes in Statistics, New York Inc.: Springer-Verlag, Vol. 20, pp. 1 - 11, 1983.

[11] G. Lugosi, N. Cesa-Bianchi, *Prediction, Learning and Games.* New York: Cambridge University Press, 2006.

[12] A. Juditsky, A. V. Nazin, A. B. Tsybakov, N. Vayatis, *Gap-Free Bounds for Stochastic Multi-Armed Bandit.* Proc. 17th World Congress IFAC (Seoul, Korea, July 6 - 11), pp. 11560 - 11563, 2008.

[13] A. V. Gasnikov, Yu. E. Nesterov, V. G. Spokoiny, *On the Efficiency of a Randomized Mirror Descent Algorithm in Online Optimization Problems.* Computational Mathematics and Mathematical Physics, Vol. 55, No 4, pp. 580 - 596, 2015.

[14] A. V. Kolnogorov, *Parallel Design of Robust Control in the Stochastic Environment (the Two-Armed Bandit Problem)* Automation and Remote Control, Vol. 73, No. 4, pp. 689 - 701, 2012.

[15] A. V. Kolnogorov, *Robust Normal Two-Armed Bandit and Parallel Data Processing.* RECENT ADVANCES in MATHEMATICAL METHODS in APPLIED SCIENCES. Proceedings of the 2014 International Conference on Mathematical. Models and Methods in Applied Sciences (MMAS '14). Proceedings of the 2014 International Conference on Economics and Applied Statistics (EAS '14). Saint Petersburg State Polytechnic University, Saint Petersburg, Russia, September 23 - 25, pp. 32 - 40, 2014.

[16] A. V. Kolnogorov, *Robust Parallel Control in a Random Environment and Data Processing Optimization.* Automation and Remote Control, Vol. 75, No. 12, pp. 2124 - 2134, 2014.

[17] A. V. Kolnogorov, *On a Limiting Description of Robust Parallel Control in a Random Environment.* Automation and Remote Control, Vol. 76, No. 7, pp. 1229 - 1241, 2015.