MSC 62C10, 62L05, 91A35 DOI: 10.14529/jcem230403

OPTIMIZATION OF TWO-ALTERNATIVE BATCH PROCESSING WITH PARAMETER ESTIMATION BASED ON DATA INSIDE BATCHES

A. V. Kolnogorov, Yaroslav-the-Wise Novgorod State University, Veliky Novgorod, Russian Federation, kolnogorov53@mail.ru

We consider optimization of two-alternative batch data processing within the framework of the Gaussian one-armed bandit problem. This means that there are two alternative processing methods with different efficiencies and the effectiveness of the second method is a priori unknown. It is necessary to determine which method is more effective and ensure its preferential use, so that the effectiveness of the second method is evaluated during the data processing inside batches. This approach is advisable to use if the volumes of batches and their number are not very large. Recursive equations for calculating Bayesian risk and regret in the usual and invariant form with a control horizon equal to one are obtained.

Keywords: Gaussian one-armed bandit; batch processing; Bayesian and minimax approaches; invariant description.

Introduction

We consider optimization of batch data processing if two alternative methods with different efficiencies are available for processing when the effectiveness of the first method is known a priori, and the second is not. During processing, it is necessary to determine a more effective method and ensure its preferential use. In the mathematical formulation, this optimization problem is known as the two-armed bandit problem (see, e.g., [1, 2]) with known characteristics of the first method, or the one-armed bandit problem. In what follows, processing methods are called actions. The problem has applications in behavior modeling [3], adaptive control of random processes [4], medicine, Internet technologies, data processing, etc. [5].

Formally, the Gaussian one-armed bandit is a controlled random process ξ_n, $n = 1, 2, \dots, N$, which values are interpreted as incomes, depend only on the currently selected actions y_n ($y_n \in \{1, 2\}$) and, if the second action is chosen, have a normal (Gaussian) distribution density

$$f_D(x|m) = (2\pi D)^{-1/2} \exp(-(x-m)^2/(2D)).$$

Here m, D are the mathematical expectation and variance of one-step income for choosing the second action. The mathematical expectation of income for the choice of the first action m_1 is known and, without loss of generality, is zero (otherwise, one can consider the process $\xi_n - m_1$, $n = 1, 2, \dots, N$). Thus, a one-armed bandit is described by the parameter $\theta = (m, D)$. The parameter value is assumed to be a priori unknown, however, a set of parameters is known $\Theta = \{(m, D) : |m| \le C < +\infty, 0 < \underline{D} \le D \le \overline{D} < +\infty\}$. Note that the Gaussian distribution of income is a consequence of batch processing, in which the same actions are applied to data batches, and then the total incomes in the batches are used for control.

The control strategy σ at time $n+1$ performs, in general, a stochastic choice of action y_{n+1} depending on the current value of sufficient statistics, which in this case are the total income for choosing the second action and the corresponding s^2- statistics. A regret

$$L_N(\sigma,\theta) = N\max(0,m) - \mathbf{E}_{\sigma,\theta}\left(\sum_{n=1}^{N}\xi_n\right) \tag{1}$$

characterizes the mathematical expectation of the loss of cumulative income relative to its maximum possible value in the presence of complete information. Here $\mathbf{E}_{\sigma,\theta}$ is a sign of mathematical expectation if σ and θ are fixed. Note that the regret for the processes $\{\xi_n\}$ and $\{\xi_n - m_1\}$ are the same.

Consider a prior distribution density of $\lambda(\theta)$ on the set of parameters Θ. By

$$L_N(\sigma,\lambda) = \int_\Theta L_N(\sigma,\theta)d\theta, \tag{2}$$

$$R_N^B(\lambda) = \inf_{\{\sigma\}} L_N(\sigma,\lambda) \tag{3}$$

we denote the regret averaged with respect to λ and Bayesian risk. Here $d\theta = dmdD$. The minimax risk on the set Θ is defined as

$$R_N^M(\Theta) = \inf_{\{\sigma\}}\sup_{\Theta} L_N(\sigma,\theta). \tag{4}$$

Bayesian strategy and risk can be found by solving backwards the Bellman recursive equation. There is no direct method for finding minimax strategies and risks, but they can be searched using the main theorem of game theory, according to which minimax strategy and risk coincide with Bayesian ones calculated with respect to the worst-case prior distribution on which Bayesian risk is maximal [6]. In the case of large N (big data), an asymptotic estimate of the order of $N^{1/2}$ is well-known for minimax risk [7]. Another important property of the minimax approach in the case of big data is that batch processing virtually does not lead to an increase in maximum regret if the number of batches, into which the data is divided, is large enough [6, 8].

The one-armed bandit problem was previously considered in [9, 10] for a Bernoulli one-armed bandit whose incomes have values 0 and 1; asymptotic properties were studied in [10]. In [9], the following intuitively clear property of the Bayesian strategy is proved: the choice of the second action (with unknown characteristics) can start only at the beginning of the control. If the first action is selected once, it will be applied until the end of the control. This is due to the fact that the applying the first action does not provide additional information, so if once a conclusion was made that it is better than the second, this decision will not be changed as a result of applying the first action. This property is also valid for the Gaussian one-armed bandit, including the one considered in section **??**. The proof is similar to the one given earlier in [6, 11, 12] and is therefore omitted.

Let's point out the difference between the statements considered in this article and those presented in [6, 11, 12]. In [6, 11], a Gaussian one-armed bandit was considered, whose incomes were characterized by one unknown parameter, mathematical expectation. This model corresponds to a situation where the batch sizes are large, so the variance can be estimated when processing the first batch, and then this estimate is used for control. In the case of batches of moderate volume, the variance should be evaluated during the control process. Unlike [12], where the variance is estimated by the total incomes of batches, in this article the estimation is based on the data inside the batches.

The further structure of the article is as follows. In section 1, recursive equations are obtained for calculating Bayesian risk and regrets in the usual and invariant forms. The advantage of the invariant form is that it does not depend on the total number of data, but only on the number of processed batches and the data inside the batch for which the variance is estimated. Therefore, the invariant form can be used to obtain asymptotic estimates of Bayesian and minimax risk. Section 2 contains the results of numerical experiments. The conclusion is presented in section 3.

1. Recursive Equations for Computing Bayesian Risk and a Regret

Consider batch data processing, in which the variance estimation is performed during data processing within the batch. To do this, we assume that the processing is carried out in batches of the size $M = M_1 M_2$. In turn, these batches are divided into M_2 small packets, each of which includes M_1 of data. This makes it possible to estimate the variance when processing the next large batch based on the observed incomes of small packets by computing the corresponding s^2-statistics. It is clear that small packets and also large batches themselves allow parallel processing. The number of large batches and, accordingly, the number of processing stages is K. Thus, the total number of data is $N = KM_1M_2 = KM$.

Let's consider how to recalculate the total income X and s^2-statistics S after processing the next large batch. Recall mentioned above property of Bayesian strategy that it can start to apply the second action only at the beginning of control. Let k be the current number of large batches processed and, therefore, $n = kM_2$ be the current total number of small packets included in them. Denote by x_i, $i = 1, 2, \dots, n$ incomes obtained when processing n small packets by the second action. Then the current total income and s^2-statistics are $X = \sum_{i=1}^{n} x_i$, $S = \sum_{i=1}^{n} x_i^2 - X^2/n$. For the next $(k + 1)$th large batch $(k \geq 0)$, one can compute its total income and s^2-statistics on the small packets included in it with incomes $x_{n+1}, \dots, x_{n+M_2}$ as follows: $Y = \sum_{i=n+1}^{n+M_2} x_i$, $U = \sum_{i=n+1}^{n+M_2} x_i^2 - Y^2/M_2$.

Then the new values of total income and s^2-statistics are recalculated using the old ones according to the following formulas $X_{new} = \sum_{i=1}^{n+M_2} x_i = X + Y$, $S_{new} = \left(\sum_{i=1}^{n+M_2} x_i^2\right) - (X + Y)^2/(n + M_2) = S + U + M_1\Delta(X, k, Y)$, where $M_1\Delta(X, k, Y) = Y^2/M_2 + X^2/n - (X + Y)^2/(n + M_2) = (M_2X - nY)^2/\left(nM_2(n + M_2)\right)$. If $k = 0$ then $X = S = 0$ and, therefore, $\Delta(0, 0, 0)$. Thus, the recalculation of statistics after the receipt of the next large data batch is carried out according to the formulas

$$X \leftarrow X + Y, \quad S \leftarrow S + U + M_1\Delta(X, k, Y), \tag{5}$$

where $\Delta(0, 0, 0) = 0$ and

$$\Delta(X, k, Y) = \frac{(M_2X - kM_2Y)^2}{M_1M_2^3 k(k + 1)} = \frac{(X - kY)^2}{Mk(k + 1)}, \quad if k \geq 1. \tag{6}$$

Consider a chi-squared distribution density with k degrees of freedom $\chi_k^2(x) = \{2^{k/2}\Gamma(k/2)\}^{-1} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$, $x \geq 0$, $k \geq 1$. Denote by $D' = M_1D$ and $m' = M_1m$ the variance and the mathematical expectation of income for processing the small packet. Given the prior distribution $\lambda(\theta)$, let's determine the posterior distribution. We introduce the function $\mathbf{F}(X,S,k|m',D')$ by the following conditions: $\mathbf{F}(0,0,0|m',D') = 1$ and $\mathbf{F}(X,S,k|m',D') = f_{kM_2D'}(X|kM_2m')\,\psi_{kM_2-1}(S/D')$ if $k \geq 1$ with $\psi_{kM_2-1}(S/D') = (D')^{-1}\chi^2_{kM_2-1}(S/D')$. Note that if $k \geq 1$, the functions $f_{kM_2D'}(X|kM_2m')$ and $\psi_{kM_2-1}(S/D')$ describe the probability density functions (pdf) of cumulative income X and s^2-statistics S computed after processing k large batches or, equivalently, after processing kM_2 small packets. Since X and S are independent random variables, the joint pdf

$$\mathbf{F}(X,S|m',D') = f_{M_2D'}(X|M_2m')\psi_{M_2-1}(S/D') \tag{7}$$

describes the pdf of X, S, corresponding to processing one large batch.

Given a prior distribution density $\lambda(m,D)$, the posterior distribution density is $\lambda(m,D|X,S,k) = \mathbf{F}(X,S,k|m',D')\lambda(m.D)/P(X,S,k)$, where $P(X,S,k) = \iint_\Theta \mathbf{F}(X,S,k|m',D')\lambda(m.D)dmdD$, $k = 0,1,2,\dots$ and $\lambda(m,D|0,0,0) = \lambda(m,D)$. However, recursive equation is simpler if the posterior distribution is defined in an equivalent way. Denote $\tilde{\mathbf{F}}(0,0,0|m',D') = 1$ and

$$\tilde{\mathbf{F}}(X,S,k|m',D') = (D')^{-3/2}\tilde{f}_{kM_2D'}(X|kM_2m')\tilde{\psi}_{M_2k-1}(S/D'), \tag{8}$$

if $k \geq 1$, where

$$\tilde{f}_D(x|m) = \exp\left(-(x-m')^2/(2D)\right), \quad \tilde{\psi}_{kM_2-1}(s) = s^{\frac{kM_2-1}{2}-1}e^{-s/2}, \tag{9}$$

Then, given a prior distribution density $\lambda(m,D)$, the posterior distribution density is

$$\begin{aligned} &\lambda(m,D|X,S,k) = \tilde{\mathbf{F}}(X,S,k|m',D')\lambda(m,D)/\tilde{P}(X,S,k), \\ &\text{with } \ \tilde{P}(X,S,k) = \iint_\Theta \tilde{\mathbf{F}}(X,S,k|m',D')\lambda(m,D)dmdD. \end{aligned} \tag{10}$$

Note that (10) remains valid if $k = 0$, too.

Denote by $R^B(X,S,k) = R^B_{M(K-k)}(\lambda(m,D|X,S,k)$ the Bayesian risk computed on the control horizon $K-k$ with respect to a prior distribution density $\lambda(m,D|X,S,k)$. Recall the property of the strategy: once the first action is chosen, it will be used until the end of the control. Taking into account (5)–(6), the standard dynamic programming equation has the form

$$R^B(X,S,k) = \min\left(R^B_1(X,S,k), R^B_2(X,S,k)\right), \tag{11}$$

where $R^B_1(X,S,k) = R^B_2(X,S,k) = 0$ if $k = K$ and

$$\begin{aligned} R^B_1(X,S,k) &= (K-k)\iint_\Theta M_2(m')^+\lambda(m,D|X,S,k), \\ R^B_2(X,S,k) &= \iint_\Theta \lambda(m,D|X,S,k)\times\Big(M_2(m')^- \\ +\int_0^\infty\int_{-\infty}^\infty & R^B(X+Y,S+U+M_1\Delta(X,k,Y),k+1)\mathbf{F}(Y,U|m',D')dYdU\Big)dmdD \end{aligned} \tag{12}$$

if $0 \le k < K$. Bayesian risk (3) is

$$R_N(\lambda) = R(0,0,0). \tag{13}$$

Here $R_2^B(X,S,k)$ characterizes the expected loss on the control horizon $K-k$ if the second action is applied first and then the control is carried out optimally. When processing $(k+1)$th the large batch, the Bayesian strategy prescribes choosing an action corresponding to the current smaller value $R_1^B(X,S,k)$, $R_2^B(X,S,k)$; in the case of a draw, the choice can be arbitrary. Once the first action is chosen, it will be used until the end of the control.

Let's present equation (11)–(12) in a more convenient for computations form. We put $R_\ell(X,S,k) = R_\ell^B(X,S,k) \times \tilde{P}(X,S,k)$, $\ell = 1,2$. The following theorem is valid.

Theorem 1. *To determine the Bayesian risk, one should solve a recursive equation*

$$R(X,S,k) = \min\left(R_1(X,S,k), R_2(X,S,k)\right), \tag{14}$$

where $R_1(X,S,k) = R_2(X,S,k) = 0$ *if* $k = K$ *and*

$$\begin{gathered} R_1(X,S,k) = (K-k)MG_1(X,S,k), \\ R_2(X,S,k) = MG_2(X,S,k) \\ + \int_0^\infty \int_{-\infty}^\infty R(X+Y, S+U+M_1\Delta(X,k,Y), k+1) H(X,S,k,Y,U) dY dU, \end{gathered} \tag{15}$$

if $0 \le k < K$. *Here*

$$\begin{aligned} G_1(X,S,k) &= \iint_\Theta m^+ \tilde{\mathbf{F}}(X,S,k|m',D')\lambda(m,D) dm dD, \\ G_2(X,S,k) &= \iint_\Theta m^- \tilde{\mathbf{F}}(X,S,k|m',D')\lambda(m,D) dm dD. \end{aligned} \tag{16}$$

Function $H(X,S,k,Y,U$ *is as follows:* $H(0,0,0,Y,U) = C(M_2)$ *and*

$$H(X,S,k,Y,U) = C(M_2) \times \frac{S^{(kM_2-1)/2-1} U^{(M_2-1)/2-1}}{(S+U+M_1\Delta(X,k,Y))^{((k+1)M_2-1)/2-1}}, \tag{17}$$

if $k \ge 1$ *with* $C(M_2) = (2^{M_2} M_2 \pi)^{-1/2} / \Gamma((M_2-1)/2)$. *Bayesian risk* (3) *is*

$$R_N(\lambda) = R(0,0,0). \tag{18}$$

When processing the $(k+1)$*th large batch, the Bayesian strategy prescribes to choose an action corresponding to the currently smaller value* $R_1(X,S,k)$, $R_2(X,S,k)$*; in the case of a draw, the choice can be arbitrary. Once the first action is chosen, it will be used until the end of the control.*

Proof. Let's multiply the left-hand and right-hand sides of the equation (11)–(12) by $\tilde{P}(X,S,k)$ in (10). We get (14)–(15), where $G_1(X,S,k)$, $G_2(X,S,k)$ are described by (16), and

$$H(X,S,k,Y,U) = \frac{\tilde{\mathbf{F}}(X,S,k|m',D')\mathbf{F}(Y,U|m',D')}{\tilde{\mathbf{F}}(X+Y,S+U+M_1\Delta,k+1|m',D')}. \tag{19}$$

with Δ given by (6). For $k \geq 1$, using (7)–(9), it follows from (19) that

$$H(X,S,k,Y,U) = \frac{\tilde{f}_{kM_2D'}(X|kM_2m')f_{M_2D'}(Y|M_2m')}{\tilde{f}_{(k+1)M_2D'}(X+Y|(k+1)M_2m')} \frac{\tilde{\psi}_{kM_2-1}(S/D')\psi_{M_2-1}(U/D')}{\tilde{\psi}_{(k+1)M_2-1}((S+U+M_1\Delta)/D')}.$$

Here

$$\frac{\tilde{f}_{kM_2D'}(X|kM_2m')f_{M_2D'}(Y|M_2m')}{\tilde{f}_{(k+1)M_2D'}(X+Y|(k+1)M_2m')} = \left(\frac{1}{2\pi M_2 D'}\right)^{1/2} \exp\left(-\frac{M_1\Delta}{2D'}\right)$$

and

$$\frac{\tilde{\psi}_{kM_2-1}(S/D')\psi_{M_2-1}(U/D')}{\tilde{\psi}_{(k+1)M_2-1}((S+U+M_1\Delta)/D')} = \frac{1}{D' \times 2^{(M_2-1)/2}\Gamma((M_2-1)/2)}$$
$$\times \frac{(S/D')^{(kM_2-1)/2-1}(U/D')^{(M_2-1)/2-1}}{((S+U+M_1\Delta)/D')^{((k+1)M_2-1)/2-1}} \times \frac{\exp\left(-S/(2D')\right)\exp\left(-U/(2D')\right)}{\exp\left(-(S+U+M_1\Delta)/(2D')\right)}$$
$$= \frac{(D')^{1/2}\exp\left(M_1\Delta/(2D')\right)}{2^{(M_2-1)/2}\Gamma((M_2-1)/2)} \times \frac{S^{(kM_2-1)/2-1}U^{(M_2-1)/2-1}}{(S+U+M_1\Delta)^{((k+1)M_2-1)/2-1}}.$$

Hence, $H(X,S,k,Y,U)$ satisfies (17) if $k \geq 1$. If $k=0$ then $X=0$, $S=0$ and (19) takes the form

$$H(0,0,0,Y,U) = \frac{f_{M_2D'}(Y|M_2m')\psi_{M_2-1}(U/D')}{(D')^{-3/2}\tilde{f}_{M_2D'}(Y|M_2m')\tilde{\psi}_{M_2-1}(U/D')} = C(M_2).$$

Formula (18) follows from (13) and equality $\tilde{P}(0,0,0)=1$.

□

Let's obtain an invariant form of formulas (14)–(18). We take the set of parameters $\Theta_N = \{(m,D) : \underline{D} \leq D \leq \overline{D},\, |m| \leq c(D/N)^{1/2}\}$, where $c>0$, $0<\underline{D} \leq D \leq \overline{D} < \infty$. If one puts $D = \beta\overline{D}$, $m = \alpha(\overline{D}/N)^{1/2} = \alpha(\beta^{-1}D/N)^{1/2}$, then the set of parameters takes the form $\Theta_N = \{(\alpha,\beta) : \underline{D}/\overline{D} = \beta_0 \leq \beta \leq 1, |\alpha| \leq c\beta^{1/2}\}$.

Consider the change of variables: $X = x(\overline{D}N)^{1/2}$, $Y = y(\overline{D}N)^{1/2}$, $S = s\overline{D}M_1$, $U = u\overline{D}M_1$, $k = tK$, $M/N = K^{-1} = \varepsilon$, $m = \alpha(\overline{D}/N)^{1/2}$, $D = \beta\overline{D}$, $\lambda(m,D) = (N/\overline{D}^{\,3})^{1/2}\varrho(\alpha,\beta)$. Let

$$R_\ell(0,0,0) = (\overline{D}N)^{1/2}r_\ell(0,0,0), \quad R_\ell(X,S,k) = (\overline{D}\,N)^{1/2}(\overline{D}\,M_1)^{-3/2}r_\ell(x,s,t), \tag{20}$$

if $k \geq 1$, $\ell = 1,2$. Then the following theorem is valid.

Theorem 2. *To determine a Bayesian risk, one should solve a recursive equation*

$$r(x,s,t) = \min\left(r_1(x,s,t), r_2(x,s,t)\right), \tag{21}$$

where $r_1(x,s,t) = r_2(x,s,t) = 0$ *if* $t=1$ *and*

$$r_1(x,s,t) = (1-t)g_1(x,s,t),$$
$$r_2(x,s,t) = \varepsilon g_2(x,s,t) \tag{22}$$
$$+\int_0^\infty \int_{-\infty}^\infty r(x+y, s+u+\delta(x,t,y), t+\varepsilon)h(x,s,t,y,u)dydu,$$

if $0 \le t \le 1-\varepsilon$. *Here*

$$g_1(x,s,t) = \iint_{\Theta_N} \alpha^{+}\tilde{\mathbf{f}}(x,s,t|\alpha,\beta)\varrho(\alpha,\beta)d\alpha d\beta, \qquad (23)$$
$$g_2(x,s,t) = \iint_{\Theta_N} \alpha^{-}\tilde{\mathbf{f}}(x,s,t|\alpha,\beta)\varrho(\alpha,\beta)d\alpha d\beta,$$

with $\tilde{\mathbf{f}}(0,0,0|\alpha,\beta)=1$ *and* $\tilde{\mathbf{f}}(x,s,t|\alpha,\beta)=\beta^{-3/2}\tilde{f}_{t\beta}(x|t\alpha)\tilde{\psi}_{kM_2-1}(s/\beta)$ *if* $t>0$. *Function* $h(x,s,t,y,u)$ *is as follows:* $h(0,0,0,y,u)=c(M_2)$ *and*

$$h(x,s,t,y,u) = c(M_2)\times \frac{s^{(kM_2-1)/2-1}u^{(M_2-1)/2-1}}{(s+u+\delta(x,t,y))^{((k+1)M_2-1)/2-1}}, \qquad (24)$$

if $t \ge \varepsilon$ *with* $c(M_2)=(2^{M_2}\pi\varepsilon)^{-1/2}/\Gamma((M_2-1)/2)$. *Function* $\delta(x,t,y)$ *is the following:* $\delta(0,0,0)=0$ *and*

$$\delta(x,t,y) = (\varepsilon x - ty)^2/(\varepsilon t(t+\varepsilon)) \qquad (25)$$

if $t\ge\varepsilon$. *When processing the* $(k+1)$*th large batch (respective to* $(t+\varepsilon)$ *point of time) the Bayesian strategy prescribes choosing an action corresponding to a smaller value* $r_1(x,s,t)$, $r_2(x,s,t)$; *in the case of a draw, the choice can be arbitrary. Bayesian risk* (3) *is*

$$R_N(\lambda) = (\overline{D}\,N)^{1/2}r(0,0,0). \qquad (26)$$

This description of control on the unit horizon is invariant in the sense that it does not depend on the total amount of data N *but only on the number of large batches* K *and the number of small packets* M_2 *as parts of large ones.*

Proof. One should perform the above change of variables in (14)–(18) and use (20). □

Let's present a recursive equation for computing the regret (2). We restrict considerations to strategies which can use the second action only at the beginning of control. Once choosing the first action, they apply it until the end of control. Such strategy σ is described by a set of probabilities $\sigma_\ell(X,S,k)=\Pr(y_{k+1}=\ell|X,S,k)$, $\ell=1,2$; $k=0,\dots,K-1$. Similarly to theorem 1 the following theorem holds true.

Theorem 3. *Consider a recursive equation*

$$L(X,S,k) = \sum_{\ell=1}^{2}\sigma_\ell(X,S,k)L_\ell(X,S,k), \qquad (27)$$

where $L_1(X,S,k)=L_2(X,S,k)=0$ *if* $k=K$ *and*

$$L_1(X,S,k) = (K-k)MG_1(X,S,k),$$
$$L_2(X,S,k) = MG_2(X,S,k) \qquad (28)$$
$$+\int_0^\infty\int_{-\infty}^{\infty} L(X+Y,S+U+M_1\Delta(X,k,Y),k+1)H(X,S,k,Y,U)dYdU,$$

if $0 \leq k \leq K-1$. *Here* $G_1(X,S,k)$, $G_2(X,S,k)$ *are given by* (16), $H(X,S,k,Y,U)$ *is given by* (17) *and* $\Delta(X,k,Y)$ *is given by* (6). *Then a regret* (2) *is*

$$L_N(\sigma,\theta) = L(0,0,0). \tag{29}$$

To determine a regret (1), *one should choose a degenerate probability density concentrated at a single parameter* θ.

For an invariant representation of the equation for computing the regret, we make an additional replacement $\sigma_\ell(X,S,k) = \sigma_\ell(x,s,t)$, $L_\ell(0,0,0) = (\overline{D}N)^{1/2} l_\ell(0,0,0)$, $L_\ell(X,S,k) = (\overline{D}\,N)^{1/2}(\overline{D}\,M_1)^{-3/2} l_\ell(x,s,t)$ if $k \geq 1$, $\ell = 1,2$.

Theorem 4. *To determine a regret, one should solve a recursive equation*

$$l(x,s,t) = \sum_{\ell=1}^{2} \sigma_\ell(x,s,t) l_\ell(x,s,t), \tag{30}$$

where $l_1(x,s,t) = l_2(x,s,t) = 0$ *if* $t = 1$ *and*

$$\begin{gathered} l_1(x,s,t) = (1-t) g_1(x,s,t), \\ l_2(x,s,t) = \varepsilon g_2(x,s,t) \\ + \int_0^\infty \int_{-\infty}^\infty l(x+y, s+u+\delta(x,t,y), t+\varepsilon) h(x,s,t,y,u) dy du, \end{gathered} \tag{31}$$

if $0 \leq t \leq 1-\varepsilon$. *Here* $g_1(x,s,t)$, $g_2(x,s,t)$, *are given by* (23), $h(x,s,t,y,u)$ *and* $\delta(x,t,y)$ *are given by* (24), (25). *A regret* (2) *is*

$$L_N(\sigma,\theta) = (\overline{D}N)^{1/2} l(0,0,0). \tag{32}$$

This description of control on the unit horizon is invariant in the sense that it does not depend on the total amount of data N *but only on the number of large batches* K *and the number of small packets* M_2.

2. Numerical Results

Let's describe the results of numerical experiments. In Fig. 1, we present approximate finding minimax strategy and minimax risk. In considered case, $K = 12$, $M_2 = 5$, $M_1 = 1$ and, therefore, the total number of data is $N = 60$. The set of parameters is $\Theta = \{(m,D) : 0.7 = \underline{D} \leq D \leq 1 = \overline{D},\ m = \alpha(\overline{D}/N)^{1/2},\ |\alpha| \leq 5\}$. The results are presented for a Bayesian strategy computed with respect to a prior distribution $\Pr(D = 1, \alpha = 3.5) = 0.16$, $\Pr(D = 1, \alpha = -5) = 0.84$, corresponding normalized Bayesian risk is approximately 0.47. For determined strategy, the regrets corresponding to variance values of $D = 1$, 0.9, 0.8, 0.7, are presented by lines 1, 2, 3, 4 respectively, their maximum is approximately 0.48. Calculations of Bayesian risk were performed using (14)–(18), the regrets were computed using (27)–(29). When performing numerical integration, X varied in the range from -18 to 18 in increments of 0.15, and S varied from 0.5 to 120.5 in increments of 1. Since, a function $H(X,S,k,Y,U)$ has no singularity if $M_2 = 5$, there is no need to provide a small increment in S.

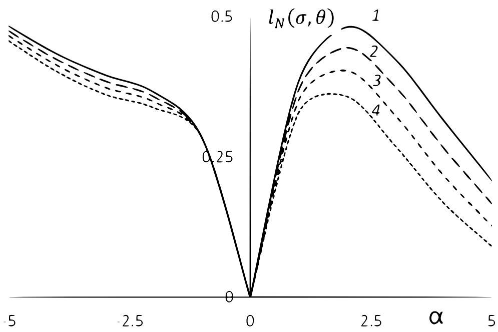

Fig. 1. Graph of the numerical solution of smth.

3. Conclusion

We have considered batch data processing with an estimation of the parameters of the distribution of one-step income by incomes within batches. This approach is applicable if the number of batches being processed and their sizes have moderate volumes. The resulting invariant control descriptions depend only on the number of batches being processed and not on the total number of data.

The research was supported by Russian Science Foundation, project number 23-21-00447, https://rscf.ru/en/project/23-21-00447/.

References

1. Berry D.A., Fristedt B. *Bandit Problems: Sequential Allocation of Experiments.* London, New York, Chapman and Hall, 1985.
2. Presman E.L., Sonin I.M. *Sequential Control with Incomplete Information.* New York, Academic, 1990.
3. Tsetlin M.L. *Automaton Theory and Modeling of Biological Systems.* New York, Academic, 1973.
4. Sragovich V.G. *Mathematical Theory of Adaptive Control.* Singapore, World Sci., 2006.
5. Lattimore T., Szepesvari C. *Bandit Algorithms.* Cambridge, Cambridge University Press, 2020.
6. Kolnogorov A.V. One-Armed Bandit Problem for Parallel Data Processing Systems. *Problems of Information Transmission*, 2015, vol. 51, no. 2, pp. 177–191. DOI: 10.1134/S0032946015020088
7. Vogel W. An Asymptotic Minimax Theorem for the Two-Armed Bandit Problem. *Ann. Math. Statist.*, 1960, vol. 31, no. 2, pp. 444–451.
8. Perchet V., Rigollet P., Chassang S., Snowberg E. Batched Bandit Problems. *Ann. Statist.*, 2016, vol. 44, no. 2, pp. 660–681. DOI: 10.1214/15-AOS1381
9. Bradt R.N., Johnson S.M., Karlin S. On Sequential Designs for Maximizing the Sum of n Observations. *Ann. Math. Statist.*, 1956, vol. 27, pp. 1060–1074. DOI: 10.1214/aoms/1177728073

10. Chernoff H., Ray, S.N. A Bayes Sequential Sampling Inspection Plan. *Ann. Math. Statist.*, 1965, vol. 36, pp. 1387–1407. DOI: 10.1214/aoms/1177699898

11. Kolnogorov A. Gaussian One-Armed Bandit Problem. *2021 XVII International Symposium «Problems of Redundancy in Information and Control Systems» (REDUNDANCY).* Moscow. IEEE, 2021, pp. 74–79. DOI: 10.1109/REDUNDANCY52534.2021.9606464

12. Kolnogorov A.V. Gaussian One-Armed Bandit with Both Unknown Parameters. *Siberian Electronic Mathematical Reports*, 2022, vol. 19, no. 2, pp. 639–650.

Alexander V. Kolnogorov, DSc (Math), Professor, Professor of Applied Mathematics and Information Science Department, Yaroslav-the-Wise Novgorod State University, (Velikiy Novgorod, Russian Federation), kolnogorov53@mail.ru

Received November 7, 2023

УДК 519.244, 519.83 **DOI: 10.14529/jcem230403**

ОПТИМИЗАЦИЯ ДВУХАЛЬТЕРНАТИВНОЙ ПАКЕТНОЙ ОБРАБОТКИ С ОЦЕНКОЙ ПАРАМЕТРОВ НА ОСНОВЕ ДАННЫХ ВНУТРИ ПАКЕТОВ

А.В. Колногоров

Рассматривается оптимизация двухальтернативной пакетной обработки данных в рамках задачи о гауссовском одноруком бандите. Это означает, что для обработки имеются два альтернативных метода с различными эффективностями, причем эффективность второго метода априори неизвестна. Требуется определить, какой метод является более эффективным, и обеспечить его преимущественное применение, причем оценка эффективности второго метода осуществляется в процессе обработки данных внутри пакетов. Данный подход целесообразно использовать если объемы пакетов и их количество не очень велики. Получены рекуррентные уравнения для вычисления байесовского риска и функции потерь в обычной и инвариантной форме с горизонтом управления равным единице.

Ключевые слова: гауссовский однорукий бандит; пакетная обработка; байесовский и минимаксный подходы; инвариантное описание.

Литература

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

2. Пресман, Э.Л. Последовательное управление по неполным данным / Э.Л. Пресман, И.М. Сонин. – М.: Наука, 1982.

3. Цетлин, М.Л. Исследования по теории автоматов и моделированию биологических систем / М.Л. Цетлин. – М.: Наука, 1969.

4. Sragovich, V.G. Mathematical Theory of Adaptive Control / V.G. Sragovich. – Singapore: World Sci., 2006.
5. Lattimore, T. Bandit Algorithms / T. Lattimore, C. Szepesvari. – Cambridge, Cambridge University Press, 2020.
6. Колногоров, А.В. Задача об одноруком бандите для систем с параллельной обработкой данных / А.В. Колногоров // Пробл. передачи информ. – 2015. – Т. 51, № 2. – С. 99–113.
7. Vogel, W. An Asymptotic Minimax Theorem for the Two-Armed Bandit Problem / W. Vogel // Ann. Math. Statist. – 1960. – Т. 31, № 2. – С. 444–451.
8. Perchet, V. Batched Bandit Problems / V. Perchet, P. Rigollet, S. Chassang, E. Snowberg // Ann. Statist. – 2016. – Т. 44, № 2. – С. 660–681.
9. Bradt, R.N. On Sequential Designs for Maximizing the Sum of n Observations / R.N. Bradt, S.M. Johnson, S. Karlin // Ann. Math. Statist. – 1956.– Т. 27. – С. 1060–1074.
10. Chernoff, H. A Bayes Sequential Sampling Inspection Plan. / H. Chernoff, S.N. Ray // Ann. Math. Statist. – 1965. Т. 36. – С. 1387–1407.
11. Kolnogorov, A. Gaussian One-Armed Bandit Problem. 2021 XVII International Symposium «Problems of Redundancy in Information and Control Systems» (REDUNDANCY). Moscow / A. Kolnogorov // 2021. – С. 74–79.
12. Kolnogorov, A.V. Gaussian One-Armed Bandit with Both Unknown Parameters / A.V. Kolnogorov // Siberian Electronic Mathematical Reports. – 2022. – V. 19, № 2. – P. 639–650.

Колногоров Александр Валерианович, доктор физико-математических наук, профессор, кафедра прикладной математики и информатики, Новгородский государственный университет им. Ярослава Мудрого (г. Великий Новгород, Российская Федерация), kolnogorov53@mail.ru

Поступила в редакцию 7 ноября 2023 г.