Gaussian Two-Armed Bandit and Optimization of Batch Data Processing

A. V. Kolnogorov

Department of Applied Mathematics and Information Science,
Yaroslav-the-Wise Novgorod State University

e-mail: kolnogorov53@mail.ru

Received September 11, 2017; in final form, November 24, 2017

Abstract—We consider the minimax setting for the two-armed bandit problem with normally distributed incomes having a priori unknown mathematical expectations and variances. This setting naturally arises in optimization of batch data processing where two alternative processing methods are available with different a priori unknown efficiencies. During the control process, it is required to determine the most efficient method and ensure its predominant application. We use the main theorem of game theory to search for minimax strategy and minimax risk as Bayesian ones corresponding to the worst-case prior distribution. To find them, a recursive integro-difference equation is obtained. We show that batch data processing almost does not increase the minimax risk if the number of batches is large enough.

DOI: 10.1134/S0032946018010076

1. INTRODUCTION

We consider the two-armed bandit problem [1, 2], which is also well known as the problem of expedient behavior in a random environment [3, 4] and the problem of adaptive control [5, 6] in the following setting. Let $\xi_n$, $n = 1, 2, \ldots, N$, be a random controlled process the values of which are interpreted as incomes, depend on currently chosen actions $y_n$ only, and have Gaussian (normal) distributions with probability densities $f_D(x \mid m_\ell)$ if $y_n = \ell$, $\ell = 1, 2$, where $f_D(x \mid m) = (2\pi D)^{-1/2} \exp \{- (x - m)^2 / (2D)\}$. We assume that $D_1$ and $D_2$ are a priori known variances and $m_1$ and $m_2$ are unknown mathematical expectations (in what follows, the assumption that $D_1$ and $D_2$ are known can be omitted). Such a two-armed bandit can be described by a vector parameter $\theta = (m_1, m_2)$.

A control strategy $\sigma$ at time $n = n_1 + n_2$ defines, in general, a random choice of an action $y_n$ depending on the currently available statistics $(X_1, n_1, X_2, n_2)$ where $n_1$ and $n_2$ are current cumulative numbers of applications of both actions and $X_1$ and $X_2$ are corresponding cumulative incomes. Thus, $\sigma = \{\sigma_\ell(X_1, n_1, X_2, n_2)\}$, where $\sigma_\ell(X_1, n_1, X_2, n_2) = \Pr (y_n = \ell \mid X_1, n_1, X_2, n_2)$, $\ell = 1, 2$.

The goal is to maximize (in a certain sense) the total expected income. If the parameter $\theta$ were known, then an optimal strategy would always prescribe to choose the action corresponding to larger of the values $m_1$ and $m_2$. The total expected income would thus be equal to $N(m_1 \lor m_2)$, where $m_1 \lor m_2$ is the maximum of $m_1$ and $m_2$. If the parameter is unknown, then, because of the incomplete information, the loss function

$$L_N(\sigma, \theta) = N(m_1 \lor m_2) - E_{\sigma, \theta} \left( \sum_{n=1}^{N} \xi_n \right)$$  \hspace{1cm} (1.1)

describes expected losses of the total income relatively to its maximal possible value. Here $E_{\sigma, \theta}$ denotes the mathematical expectation with respect to the measure generated by the strategy $\sigma$ and
parameter \( \theta \). The set of parameters is chosen to be \( \Theta = \{ \theta : |m_1 - m_2| < 2C \} \), where \( 0 < C < \infty \). The requirement \( C < \infty \) ensures boundedness of the loss function on \( \Theta \). From the loss function we determine the minimax risk
\[
R^M_N(\Theta) = \inf_{\{\sigma\}} \sup_{\Theta} L_N(\sigma, \theta),
\]
and the corresponding optimal strategy \( \sigma^M \) is said to be minimax. Application of the minimax strategy ensures the inequality \( L_N(\sigma^M, \theta) \leq R^M_N(\Theta) \) for all \( \theta \in \Theta \), which implies robustness of the control.

A minimax approach to this problem was proposed in [7], which caused a significant interest to the problem. In [7], a Bernoulli two-armed bandit was considered defined by the probability distribution \( \Pr(\xi_n = 1 | y_n = \ell) = p_\ell, \Pr(\xi_n = 0 | y_n = \ell) = 1 - p_\ell, \ell = 1, 2 \). This two-armed bandit can be described by a parameter \( \theta = (p_1, p_2) \) with the set of admissible values \( \Theta = \{(p_1, p_2) : 0 \leq p_\ell \leq 1, \ell = 1, 2 \} \). It was shown in [8] that explicit determination of the minimax strategy and minimax risk for the Bernoulli two-armed bandit is virtually impossible already for \( N > 4 \). However, an asymptotic minimax theorem is known (see [9]) which states that the following inequalities hold as \( N \rightarrow \infty \):
\[
0.612 + o(1) \leq (DN)^{-1/2} R^M_N(\Theta) \leq 0.752 + o(1),
\]
where \( D = 0.25 \) is the maximum variance of a one-step income. The lower estimate presented here was obtained in [10]. This theorem is valid for a Gaussian two-armed bandit too if \( D_1 = D_2 = D \). It is important to note that maximal losses are attained in the domain of “close” distributions characterized by the condition \( |m_1 - m_2| \leq 2cN^{-1/2} \) where \( c > 0 \) is a large enough fixed value. For “distant” distributions, the order of losses is smaller. For example, it is equal to \( \ln N \) if \( |m_1 - m_2| \geq \delta > 0 \), because in this case a preferable action can be confidently determined on a relatively short time interval (see [11]).

Note that there are some other approaches to robust control in the two-armed and multi-armed bandit problems, e.g., [6,12–14]. In this case, stochastic approximation method, mirror descent algorithm, and some other techniques are used for the control. The order of the minimax risk for these algorithms is equal or close to \( N^{1/2} \).

Remark 1. Some authors (see, e.g., [6,12,13]) consider minimization of the total expected income as the goal of the control (therefore, instead of “incomes,” they consider “losses”). In this case, the function
\[
L_N(\sigma, \theta) = E_{\sigma, \theta} \left( \sum_{n=1}^{N} \xi_n \right) - N(m_1 \wedge m_2),
\]
should be used as the loss function, where \( m_1 \wedge m_2 \) is the minimum of \( m_1 \) and \( m_2 \). In the case of normally distributed \( \{\xi_n\} \), this setting can be reduced to ours by considering the incomes \( \{-\xi_n\} \).

Let us explain the choice of normally distributed incomes. We consider the problem as applied to optimization of processing a large amount of data. Let \( T = NM \) items of input data be given, which can be processed by either of two alternative methods. Denote by \( \zeta_t \) the result of processing a data item with number \( t \). For example, processing may be successful (\( \zeta_t = 1 \)) or unsuccessful (\( \zeta_t = 0 \)), and one has to maximize the total number of successfully processed data. Or \( \zeta_t \) is the duration of data item processing, and one has to minimize the total computer processing time. We assume that distributions of \( \{\zeta_t\} \) depend only on chosen methods (actions). Let us partition the data into \( N \) batches containing \( M \) data items each and apply the the same method to all the data in the same batch. For the control, let us use values of the process \( \xi_n = M^{-1/2} \sum_{t=(n-1)M+1}^{nM} \zeta_t \), \( n = 1, \ldots, N \). According to the central limit theorem, distributions of \( \{\xi_n\} \) are close to normal for a wide class of original distributions \( \{\zeta_t\} \), and this implies the generality of the presented setting.
In what follows, we will consider $\xi_n, n = 1, 2, \ldots, N$, as results of processing of $N$ data items (e.g., after a preprocessing).

Note that batch processing in the two-armed bandit problem was originally proposed for treating a large group of patients by one of two alternative drugs. In this case, at first each drug was given to a large enough test group of patients and then the most effective one was given to all the rest. For a discussion and bibliography, see, e.g., [11].

A very popular approach to the problem is Bayesian. Denote by $\lambda(\theta)$ a prior distribution density on $\Theta$. The quantity

$$R^B_N(\lambda) = \inf_{\{\sigma\}} \int_{\Theta} L_N(\sigma, \theta) \lambda(\theta) \, d\theta$$

(1.4)

is called the Bayesian risk, and the corresponding optimal strategy is said to be Bayesian. The Bayesian approach yields a recursive equation which allows to find the Bayesian strategy and Bayesian risk by the dynamic programming technique. Both minimax and Bayesian approaches are integrated by the main theorem of game theory, which ensures, under mild conditions, the equality

$$R^M_N(\Theta) = R^B_N(\lambda_0) = \sup_{\lambda} R^B_N(\lambda);$$

(1.5)

i.e., the minimax risk (1.2) is equal to the Bayesian risk (1.4) calculated with respect to the worst-case prior distribution, and the minimax strategy coincides with the corresponding Bayesian one. However, it is important to understand that straightforward application of (1.5) for finding the minimax risk is virtually impossible because of high computational complexity. Analysis of properties of the worst-case prior distribution makes it possible to substantially simplify the problem.

This paper is devoted to finding the minimax risk and minimax strategy as Bayesian ones corresponding to the worst-case prior distribution. It generalizes the results of [15,16], where a Gaussian two-armed bandit with equal unit income variances and a one-armed bandit, which can be regarded as a limit case of our setting as $D_1 \to 0$, were investigated.

The structure of the paper is as follows. In Section 2 we establish properties of the worst-case prior distribution and the corresponding Bayesian strategy. In particular, we show that the optimal strategy at the start of the control applies both actions by turns. In Section 3 we obtain recursive equations for computing the Bayesian strategy and Bayesian risk with respect to the class of prior distributions to which the worst-case prior belongs. In Section 4 we present results of numerical experiments. We obtain numerical estimates for the minimax risk with $N = 50$ and various $D_1$ and $D_2$. Numerical experiments show that the assumption that the variances $D_1$ and $D_2$ are known can be omitted. First, small variations of the variances have a small effect on the losses; therefore, instead of variances themselves one can use their estimates made at the initial stage of the control. Second, the minimax strategy corresponding to $D_1 = D_2 = D$ remains the same provided that $0.5D \leq D_\ell \leq D$, $\ell = 1, 2$. Furthermore, numerical experiments show that for “distant” values of $m_1$ and $m_2$ the losses grow because of the same number of alternated applications of both actions at the initial stage of the control. To reduce the losses, one should increase the number of processed batches or decrease batch sizes at the initial stage of the control. In Section 5 we present a conclusion.

2. CHARACTERIZATION OF THE WORST-CASE PRIOR DISTRIBUTION

Equality (1.5), which allows to search for the minimax risk as Bayesian one calculated with respect to the worst-case prior distribution, is established by the main theorem of game theory. In the case of equal unit variances, this theorem is proved in [15] for the set of parameters $\Theta_1 = \{\theta : |m_1 - m_2| \leq 2C, |m_1 + m_2| \leq 2C_1\}$ (the proof requires the set of parameters to be compact). This theorem remains also valid in our setting with arbitrary variances $D_1$ and $D_2$. 

PROBLEMS OF INFORMATION TRANSMISSION Vol. 54 No. 1 2018
The advantage of the Bayesian approach is that the Bayesian risk can be calculated recursively. Denote by \( \lambda(m_1, m_2) \) a prior distribution density on the set of parameters \( \Theta \). Let the history of the control up to time \( n \) be described by a statistics \((X_1, n_1, X_2, n_2)\), where \( n_1 \) and \( n_2 \) are current total numbers of the two actions already applied \((n_1 + n_2 = n)\) and \( X_1 \) and \( X_2 \) are the corresponding cumulative incomes. Denote \( n_\ell^* = n_\ell D_\ell, \ell = 1, 2 \). The posterior distribution density is

\[
\lambda(m_1, m_2 \mid X_1, n_1, X_2, n_2) = \frac{f_{n_1^*}(X_1 \mid n_1 m_1)f_{n_2^*}(X_2 \mid n_2 m_2)\lambda(m_1, m_2)}{p(X_1, n_1, X_2, n_2)},
\]

where

\[
p(X_1, n_1, X_2, n_2) = \int \int \int \int L_{N-n}^\mathsf{B}(\sigma, \lambda; X_1, n_1, X_2, n_2) \lambda(m_1, m_2 \mid X_1, n_1, X_2, n_2) \, dm_1 \, dm_2.
\]

If additionally set \( f_{n_\ell^*}(X_\ell \mid n_\ell m_\ell) = 1 \) for \( n_\ell = 0 \), then the expression for the posterior density remains valid for \( n_1 = 0 \) and/or \( n_2 = 0 \) as well.

In what follows, we consider strategies which can change actions only after applying them \( M \) times in succession; i.e., they are applied to batches of data. For simplicity, we assume that \( N = KM \), where \( K \) is an integer number. Denote by

\[
R_{N-n}^\mathsf{B}(\lambda; X_1, n_1, X_2, n_2) = \inf_{\{\sigma\}} L_{N-n}(\sigma, \lambda; X_1, n_1, X_2, n_2),
\]

\[
L_{N-n}(\sigma, \lambda; X_1, n_1, X_2, n_2) = \int \int L_{N-n}(\sigma, \theta) \lambda(m_1, m_2 \mid X_1, n_1, X_2, n_2) \, dm_1 \, dm_2
\]

the Bayesian risk and losses caused by applying the strategy \( \sigma \) in the last \((N - n)\) steps, calculated with respect to the probability density \( \lambda(m_1, m_2 \mid X_1, n_1, X_2, n_2) \). Denote \( x^+ = \max(x, 0) \) and \( M_\ell^* = M D_\ell, \ell = 1, 2 \). Then we have the standard recursive equation

\[
R_{N-n}^\mathsf{B}(\lambda; X_1, n_1, X_2, n_2) = \min\{R_{N-n}^{\mathsf{B}(1)}(\lambda; X_1, n_1, X_2, n_2), R_{N-n}^{\mathsf{B}(2)}(\lambda; X_1, n_1, X_2, n_2)\},
\]

where \( R_{N-n}^{\mathsf{B}(1)}(\lambda; X_1, n_1, X_2, n_2) = R_{N-n}^{\mathsf{B}(2)}(\lambda; X_1, n_1, X_2, n_2) = 0 \) if \( n = N \) and then

\[
R_{N-n}^{\mathsf{B}(1)}(\lambda; X_1, n_1, X_2, n_2) = \int \int \lambda(m_1, m_2 \mid X_1, n_1, X_2, n_2)
\]

\[
\times (M(m_2 - m_1)^+ + E_x^{(1)} R_{N-n-M}^{\mathsf{B}(1)}(\lambda; X_1 + x, n_1 + M, X_2, n_2)) \, dm_1 \, dm_2,
\]

\[
R_{N-n}^{\mathsf{B}(2)}(\lambda; X_1, n_1, X_2, n_2) = \int \int \lambda(m_1, m_2 \mid X_1, n_1, X_2, n_2)
\]

\[
\times (M(m_1 - m_2)^+ + E_x^{(2)} R_{N-n-M}^{\mathsf{B}(2)}(\lambda; X_1, n_1, X_2 + x, n_2 + M)) \, dm_1 \, dm_2,
\]

with

\[
E_x^{(\ell)} R(x) = \int R(x) f_{M_\ell^*}(x \mid Mm_\ell) \, dx, \quad \ell = 1, 2.
\]

Here \( R_{N-n}^{\mathsf{B}(\ell)}(\lambda; X_1, n_1, X_2, n_2) \) are expected losses over the residual control horizon \((n + 1, N)\) if at first the \( \ell \)th action is applied \( M \) times, \( \ell = 1, 2 \), and then the control is optimal. The Bayesian strategy on the time interval \((n + 1, n + M)\) prescribes to choose the action corresponding to the smaller value \( R_{N-n}^{\mathsf{B}(\ell)}(\lambda; X_1, n_1, X_2, n_2) \); the choice can be arbitrary if values of the risks coincide.

Given some strategy \( \sigma: \sigma_\ell(X_1, n_1, X_2, n_2) = \Pr(y_n = \ell \mid X_1, n_1, X_2, n_2), \ell = 1, 2 \), a similar standard recursive equation for losses is as follows:
\[
L_{N-n}(\sigma, \lambda; X_1, n_1, X_2, n_2) = \sigma_1(X_1, n_1, X_2, n_2)L_{N-n}^{(1)}(\sigma, \lambda; X_1, n_1, X_2, n_2) + \sigma_2(X_1, n_1, X_2, n_2)L_{N-n}^{(2)}(\sigma, \lambda; X_1, n_1, X_2, n_2),
\]

where \( L_{N-n}^{(1)}(\sigma, \lambda; X_1, n_1, X_2, n_2) = L_{N-n}^{(2)}(\sigma, \lambda; X_1, n_1, X_2, n_2) = 0 \) if \( n = N \) and then

\[
\begin{align*}
L_{N-n}^{(1)}(\sigma, \lambda; X_1, n_1, X_2, n_2) &= \int\int \lambda(m_1, m_2 | X_1, n_1, X_2, n_2) \\
&\quad \times (M(m_2 - m_1) + E_x^{(1)} L_{N-n-M}(\sigma, \lambda; X_1 + x, n_1 + M, X_2, n_2)) dm_1 dm_2, \\
L_{N-n}^{(2)}(\sigma, \lambda; X_1, n_1, X_2, n_2) &= \int\int \lambda(m_1, m_2 | X_1, n_1, X_2, n_2) \\
&\quad \times (M(m_1 - m_2) + E_x^{(2)} L_{N-n-M}(\sigma, \lambda; X_1, n_1, X_2 + x, n_2 + M)) dm_1 dm_2.
\end{align*}
\]

Here \( L_{N-n}^{(i)}(\sigma, \lambda; X_1, n_1, X_2, n_2) \) are expected losses over the residual control horizon \((n + 1, N)\) if at first the \( i \)th action is applied \( M \) times and then strategy \( \sigma \) is used for the control.

Let us describe properties of the worst-case prior distribution satisfying equality (1.5), which allow to significantly simplify equations (2.2) and (2.3) and also (2.4) and (2.5). They use continuity and concavity properties of the Bayesian risk.

**Lemma 1.** Given any probability densities \( \lambda_1 \) and \( \lambda_2 \) and positive numbers \( \alpha_1 \) and \( \alpha_2 \) such that \( \alpha_1 + \alpha_2 = 1 \), we have the inequalities

\[
|R_N^B(\lambda_1) - R_N^B(\lambda_2)| \leq 2NC \int_{\Theta} |\lambda_1(\theta) - \lambda_2(\theta)| d\theta,
\]

\[
R_N^B(\alpha_1 \lambda_1 + \alpha_2 \lambda_2) \geq \alpha_1 R_N^B(\lambda_1) + \alpha_2 R_N^B(\lambda_2).
\]

**Proof.** Denote \( L_N(\sigma, \lambda) = \int_{\Theta} L_N(\sigma, \theta) \lambda(\theta) d\theta \). Since \( |L_N(\sigma, \theta)| \leq 2NC \), we have

\[
\inf_{\{\sigma\}} L_N(\sigma, \lambda_1) \leq \inf_{\{\sigma\}} L_N(\sigma, \lambda_2) + 2NC \int_{\Theta} |\lambda_1(\theta) - \lambda_2(\theta)| d\theta,
\]

where \( \inf_{\{\sigma\}} L_N(\sigma, \lambda) = R_N^B(\lambda) \). Since the inequality remains valid if \( \lambda_1 \) and \( \lambda_2 \) are swapped, this results in (2.6). Property (2.7) follows from the chain of (in)equalities

\[
R_N^B(\alpha_1 \lambda_1 + \alpha_2 \lambda_2) = \inf_{\{\sigma\}} (\alpha_1 L_N(\sigma, \lambda_1) + \alpha_2 L_N(\sigma, \lambda_2)) \\
\geq \alpha_1 \inf_{\{\sigma\}} L_N(\sigma, \lambda_1) + \alpha_2 \inf_{\{\sigma\}} L_N(\sigma, \lambda_2) = \alpha_1 R_N^B(\lambda_1) + \alpha_2 R_N^B(\lambda_2).
\]

Given some densities \( \lambda \) and \( \tilde{\lambda} \), property (2.7) implies that if the corresponding Bayesian risks are equal, i.e., \( R_N^B(\lambda) = R_N^B(\tilde{\lambda}) \), then the Bayesian risk on the mixed density is not less than the original, i.e., \( R_N^B(\alpha_1 \lambda + \alpha_2 \tilde{\lambda}) \geq R_N^B(\lambda) \). Let us describe transformations of densities that do not change the Bayesian risk.

**Lemma 2.** A shift transformation \( \tilde{\lambda}^{(1)}(m_1, m_2) = \lambda(m_1 + c, m_2 + c) \) changing both incomes by the same arbitrary value \( c \) does not change the Bayesian risk, i.e., \( R_N^B(\lambda) = R_N^B(\tilde{\lambda}^{(1)}) \). The choice of an action at the first step is the same for all \( c \).

If \( D_1 = D_2 \), then the transformation \( \tilde{\lambda}^{(2)}(m_1, m_2) = \lambda(m_2, m_1) \) swapping the correspondence of incomes to actions does not change the Bayesian risk, i.e., \( R_N^B(\lambda) = R_N^B(\tilde{\lambda}^{(2)}) \).
Proof. Noting that
\[
\tilde{\lambda}^{(1)}(m_1, m_2 \mid X_1, n_1, X_2, n_2) = \lambda(m_1 + c, m_2 + c \mid X_1 + n_1c, n_1, X_2 + n_2c, n_2),
\]
\[
\tilde{\lambda}^{(2)}(m_1, m_2 \mid X_1, n_1, X_2, n_2) = \lambda(m_2, m_1 \mid X_2, n_2, X_1, n_1),
\]
we obtain by induction and using (2.2) and (2.3) that
\[
R_{N-n}^B(\tilde{\lambda}^{(1)}; X_1, n_1, X_2, n_2) = R_{N-n}^B(\lambda; X_1 + n_1c, n_1, X_2 + n_2c, n_2),
\]
\[
R_{N-n}^B(\tilde{\lambda}^{(2)}; X_1, n_1, X_2, n_2) = R_{N-n}^B(\lambda; X_2, n_2, X_1, n_1),
\]
where \(\tilde{\ell} = 3 - \ell\). Hence, for all \(X_1, n_1, X_2, n_2, c\) we have the equalities
\[
R_{N-n}^B(\tilde{\lambda}^{(1)}; X_1, n_1, X_2, n_2) = R_{N-n}^B(\lambda; X_1 + n_1c, n_1, X_2 + n_2c, n_2),
\]
\[
R_{N-n}^B(\tilde{\lambda}^{(2)}; X_1, n_1, X_2, n_2) = R_{N-n}^B(\lambda; X_2, n_2, X_1, n_1).
\]
The required result is obtained with \(n_1 = n_2 = 0\); for \(\tilde{\lambda}^{(1)}\), the choice of an action at the first step does not depend on \(c\). \(\triangle\)

In what follows, it is convenient to change parametrization. Put \(m_1 = m + v\) and \(m_2 = m - v\); then \(\theta = (m + v, m - v)\), \(\Theta = \{\theta : |v| \leq C\}\), and \(\Theta_1 = \{\theta : |v| \leq C, |m| \leq C_1\}\). According to Lemma 2, the density \(\tilde{\nu}^{(1)}(m, v) = \nu(m + c, v)\) does not change the Bayesian risk. If \(D_1 = D_2\), then the density \(\tilde{\nu}^{(2)}(m, v) = \nu(m, v)\) does not change the Bayesian risk either. This allows to describe the worst-case prior distribution.

**Lemma 3.** Asymptotically, the worst-case prior distribution density can be chosen as
\[
\nu_a(m, v) = \mathcal{Z}_a(m)\rho(v),
\]
where \(\mathcal{Z}_a(m) = (2a)^{-1}\) is the uniform distribution density on the interval \(|m| \leq a\) and \(a \to \infty\). If \(D_1 = D_2\), then \(\rho(v)\) can be chosen to be symmetric, i.e., \(\rho(v) = \rho(-v)\).

**Proof.** We assume that \(\theta \in \Theta_1\). For some \(b \gg C_1\), define the distribution density
\[
\nu^{(1)}(m, v) = (2b)^{-1} \int_{-b}^{b} \nu(m + x, v) \, dx;
\]
then by Lemmas 1 and 2 we have \(R_{N}^B(\nu^{(1)}) \geq R_{N}^B(\nu)\). Put \(\rho(v) = \nu^{(1)}(0, v)\). Then \(\nu^{(1)}(m, v) = 0\) if \(|m| > b + C_1\), \(\nu^{(1)}(m, v) \leq \rho(v)\) if \(b - C_1 < |m| < b + C_1\), and \(\nu^{(1)}(m, v) = \rho(v)\) if \(|m| < b - C_1\). Let us choose \(a\) so that \(\int_{\Theta} \nu_a(m, v) \, dm \, dv = 1\). Clearly, \(b - C_1 < a < b + C_1\), and hence
\[
\int_{\Theta} \left| \nu^{(1)}(m, v) - \nu_a(m, v) \right| \, dm \, dv \leq 2C_1 b^{-1}.
\]
Therefore, \(R_{N}^B(\nu^{(1)}) \geq R_{N}^B(\nu)\) and \(R_{N}^B(\nu_a) - R_{N}^B(\nu^{(1)}) \to 0\) as \(b \to \infty\).

Finally, if \(D_1 = D_2\), then according to Lemma 2 the Bayesian risk does not decrease if we choose the density \(0.5\mathcal{Z}_a(m)(\rho(v) + \rho(-v))\) as the worst-case one. \(\triangle\)

Note that we constructed the density \(\nu_a(m, v)\) assuming that \(b \gg C_1\) and hence the support of \(\nu_a(m, v)\) does not coincide with \(\Theta_1\). Let us prove that there exists a limit
\[
\lim_{a \to \infty} R_{N}^B(\nu_a(m, v)) = R_{N}^B(\rho(v)),
\]
where
and this implies that the distribution density can be chosen close to the worst-case one on the original support \( \Theta_1 \) if \( C_1 \) is large enough. First, note that it follows from (2.1) that the posterior density does not change if the prior density is multiplied by an arbitrary constant. Therefore, we can formally consider a “limiting” prior distribution density with \( \pi(m) \equiv 1 \) for all \( m \in (-\infty, \infty) \).

Put \( n'_{\ell} = n_{\ell}/D_{\ell}, \ X_{\ell} = X_{\ell}/n_{\ell}, \ \ell = 1,2, \ n' = n'_1 + n'_2, \) and \( f_D(x) = f_D(x|0) \). Below we need the following equality, which is verified straightforwardly:

\[
f_{n'_1}(X_1|n_1(m + v))f_{n'_2}(X_2|n_2(m - v)) = \psi_1(Y|m,v)\psi_2(Z|v), \tag{2.10}
\]

where

\[
Y = (n')^{-1}(n'_1X_1 + n'_2X_2), \quad Z = X_1n_2 - X_2n_1, \quad 
\psi_1(Y|m,v) = f_{(n')^{-1}}(Y + (n')^{-1}(n'_2 - n'_1)v - m), \quad 
\psi_2(Z|v) = f_{n'_1}\psi_{(n')^{-1}}(Z - 2vn_1n_2).
\]

Now consider a strategy which at the start by turn applies actions \( M \) times each and then controls optimally. If \( N \) is large, this restriction has almost no effect on the minimax risk. Then by (2.1) and (2.10) the “limiting” posterior distribution density corresponding to \( \pi(m) \equiv 1 \) takes the form

\[
\nu(m,v|Y,Z) = \frac{\rho(v)\psi_1(Y|m,v)\psi_2(Z|v)}{p(Z)}, \quad \text{where} \quad p(Z) = \int_{-C}^{C} \rho(v)\psi_2(Z|v) \, dv. \tag{2.11}
\]

Note that for any \( c \) the density \( \nu(m,v|Y-c,Z) \) results from \( \nu(m,v|Y,Z) \) by a shift transformation of \( m \). Therefore, by Lemma 2 the risks \( R^B_{N-2M}(\nu(m,v|Y,Z)) \) do not depend on \( Y \). This follows from the equalities

\[
\nu(m,v|Y-c,Z) = \nu(m+c,v|Y,Z) = \tilde{\nu}^{(1)}(m,v|Y,Z).
\]

The Bayesian risk corresponding to the “limiting” distribution density (2.11) is calculated as

\[
R^B_N(\rho(v)) = L_1(\rho(v)) + L_2(\rho(v)) + \int_{-\infty}^{\infty} R^B_{N-2M}(\nu(m,v|Y,Z))p(Z) \, dZ, \tag{2.12}
\]

where

\[
L_1(\rho(v)) = M \int_{-C}^{0} 2|v|\rho(v) \, dv, \quad L_2(\rho(v)) = M \int_{0}^{C} 2|v|\rho(v) \, dv.
\]

describe the losses caused by the choices of the first and second actions at the initial \( 2M \) steps. The following lemma holds true.

**Lemma 4.** Let \( \nu_a(m,v) \) satisfy condition (2.8) and \( a \to \infty \). If \( N > 2M \), then the optimal strategy at the start of the control applies actions by turns, and the limit (2.9) exists and is equal to (2.12). The density \( \nu(m,v|Y,Z) \) is chosen according to condition (2.11).

**Proof.** In the case of a non-“limiting” density (2.8), the posterior density and Bayesian risk corresponding to (2.11) and (2.12) are

\[
\nu_a(m,v|Y,Z) = \frac{\zeta_a(m)\rho(v)\psi_1(Y|m,v)\psi_2(Z|v)}{p(Y,Z)}, \tag{2.13}
\]

where

\[
p(Y,Z) = \int_{-C}^{C} p(Y|v)\rho(v)\psi_2(Z|v) \, dv, \quad p(Y|v) = \int_{-a}^{a} \zeta_a(m)\psi_1(Y|m,v) \, dm,
\]
and
\[ R_N^B(\nu_a(m, v)) = L_1(\rho(v)) + L_2(\rho(v)) + \iint_G R_{N-2M}^B(\nu_a(m, v | Y, Z))p(Y, Z)\,dY\,dZ. \tag{2.14} \]

In (2.14), the Jacobian \(|\partial(X_1, X_2)/\partial(Y, Z)| = 1\) is taken into account, which occurs because changing the variables \((X_1, X_2)\) to \((Y, Z)\). The domain of integration is \(G = \{Y, Z \in (-\infty, +\infty)\}\). Let us show that the choice of actions \(M\) times by turns at the start of the control is not worse than the choice of one and the same action. Note that the worst-case prior distribution on the considered set of parameters assumes that each action is chosen at least once (otherwise, the Bayesian risk is always zero). Let one and the same action (for definiteness, the first) be chosen for processing the first \(k\) \((k \geq 2)\) batches. Consider again the “limiting” posterior density corresponding to \(\nu(m) \equiv 1\) for all \(m \in (-\infty, +\infty)\). If incomes \(X_1, \ldots, X_k\) are obtained, then

\[ |p(Y | v) - \nu_a(Y)| = \nu_a(Y)\,o(e^{-\gamma b^2}), \]
\[ |p(Y, Z) - \nu_a(Y)p(Z)| = \nu_a(Y)p(Z)\,o(e^{-\gamma b^2}). \tag{2.16} \]

Note that all risks are uniformly bounded, i.e.,
\[ R_{N-2M}^B(\nu_a(m, v | Y, Z)) \leq C_2 = (N - 2M)C. \]

Since \(\iint_{G_1} \nu_a(Y)p(Z)\,dY\,dZ = 2b/a\), the integrals in (2.14) and (2.15) are by (2.16) infinitesimal on the integration domain \(G_1 = G \setminus G_1\).

Next, taking into account (2.16), in \(G_1\) we have the estimate
\[ |\nu(m, v | Y, Z) - \nu_a(m, v | Y, Z)| = \nu(m, v | Y, Z)\,o(e^{-\gamma b^2}) \]
if \(|m| \leq a\) and
\[ \nu(m, v | Y, Z) - \nu_a(m, v | Y, Z) = \nu(m, v | Y, Z) = o(e^{-\gamma(|m|-a+b)^2}) \]
if \(|m| > a\). Therefore,
\[ \iint_\Theta |\nu(m, v | Y, Z) - \nu_a(m, v | Y, Z)|\,dm\,dv \rightarrow 0, \]
and, by Lemma 1, also
\[ R_{N-2M}^B(\nu_a(m, v | Y, Z)) - R_{N-2M}^B(\nu(m, v | Y, Z)) \rightarrow 0 \]
uniformly in \((Y, Z) \in G_1\). Taking into account (2.16), this implies that (2.14) and (2.15) are equal up to an infinitesimal on the integration domain \(G_1\). Finally, the integrals in (2.14) and (2.15) are equal up to an infinitesimal on the whole integration domain \(G\).

Let us show that the choice of actions \(M\) times by turns at the start of the control is not worse than the choice of one and the same action. Note that the worst-case prior distribution on the considered set of parameters assumes that each action is chosen at least once (otherwise, the Bayesian risk is always zero). Let one and the same action (for definiteness, the first) be chosen for processing the first \(k\) \((k \geq 2)\) batches. Consider again the “limiting” posterior density corresponding to \(\nu(m) \equiv 1\) for all \(m \in (-\infty, +\infty)\). If incomes \(X_1, \ldots, X_k\) are obtained, then
the “limiting” posterior density is \( \nu(m, v | Y) = f_{(v)^{-1}}(Y | m + v)\rho(v) \) with \( Y = (X_1 + \ldots, X_k)/n \), \( n = n_1 = Mk \). Since we have the equalities \( \nu(m, v | Y - c) = \nu(m + c, v | Y) = \nu^{(1)}(m, v | Y) \), all the densities \( \nu(m, v | Y - c) \) result from \( \nu(m, v | Y) \) by shifting \( m \). Therefore, according to Lemma 2, the choice of an action for the \((k + 1)\)st batch does not depend on \( Y \). Let it prescribe the choice of the second action. In view of initial losses, equal to \( kL_1(\rho(v)) + L_2(\rho(v)) \), and the obtained statistics, the following sequences of action choices are equivalent: 1, \ldots, 1, 2 and 2, 1, \ldots, 1. The latter applies actions by turns at the start of the control. \( \triangle \)

Remark 2. If \( N = M \) or \( N = 2M \), then the strategy applying both actions with equal probabilities can be chosen as the minimax one. In this case, maximal losses are attained at \( |m_1 - m_2| = 2C \).

3. CALCULATING THE BAYESIAN STRATEGY AND BAYESIAN RISK WITH RESPECT TO THE WORST-CASE PRIOR DISTRIBUTION

The asymptotic uniformity property of a prior density makes it possible to significantly simplify the equations for calculating the Bayesian risk and losses. Consider the risks

\[
\tilde{R}_M(X_1, n_1, X_2, n_2) = R^{\B}_N(X_1, n_1, X_2, n_2) \rho(X_1, n_1, X_2, n_2), \tag{3.1}
\]

where \( \rho(X_1, n_1, X_2, n_2) \) is defined in (2.1). Put \( \tilde{R}_M(Z, n_1, n_2) = \tilde{R}_M(X_1, n_1, X_2, n_2) \), where \( Z = X_1n_2 - X_2n_1 \). The following theorem holds true.

**Theorem 1.** Let \( \nu_0(m, v) \) satisfy condition (2.8) and \( a \to \infty \). If \( N > 2M \), then the risks (3.1) satisfy the recursive equation

\[
\tilde{R}_M(Z, n_1, n_2) = \min(\tilde{R}_M^{(1)}(Z, n_1, n_2), \tilde{R}_M^{(2)}(Z, n_1, n_2)), \tag{3.2}
\]

where \( \tilde{R}_M^{(1)}(Z, n_1, n_2) = \tilde{R}_M^{(2)}(Z, n_1, n_2) = 0 \) if \( n_1 + n_2 = N \) and

\[
\tilde{R}_M^{(1)}(Z, n_1, n_2) = M\tilde{g}^{(1)}(Z, n_1, n_2) + n_2^{-1}(n_1 + M)
\times \int_{-\infty}^{+\infty} \tilde{R}_M(Z + z, n_1 + M, n_2) \times f_{MZ_1n_1(n_1 + M)}(n_2^{-1}(MZ - n_1z)) \, dz,
\]

\[
\tilde{R}_M^{(2)}(Z, n_1, n_2) = M\tilde{g}^{(2)}(Z, n_1, n_2) + n_1^{-1}(n_2 + M)
\times \int_{-\infty}^{+\infty} \tilde{R}_M(Z + z, n_1 + M, n_2) \times f_{MZ_2n_2(n_2 + M)}(n_1^{-1}(MZ - n_2z)) \, dz,
\]

if \( 2M \leq n_1 + n_2 < N, n_1 \geq M, n_2 \geq M \). Here

\[
\tilde{g}^{(1)}(Z, n_1, n_2) = \int_0^0 2|v|\tilde{g}(v; Z, n_1, n_2)\rho(v) \, dv,
\]

\[
\tilde{g}^{(2)}(Z, n_1, n_2) = \int_0^0 2|v|\tilde{g}(v; Z, n_1, n_2)\rho(v) \, dv,
\]

\[
\tilde{g}(v; Z, n_1, n_2) = f_{n_1n_2n'Z}(Z - 2vn_1n_2).
\]

The Bayesian strategy at \( n \leq 2M \) applies actions by turns. If \( n > 2M \), then at the time interval \( (n + 1, n + M) \) the action corresponding to the smaller value \( \tilde{R}_M^{(1)}(Z, n_1, n_2) \) must be applied. The Bayesian risk (2.12) is calculated as

\[
\tilde{R}_N^\B(\rho(v)) = L(\rho(v)) + \int_{-\infty}^{\infty} \tilde{R}_M(Z, M, M) \, dz, \tag{3.5}
\]
Define $h$ and this corresponds to (3.4). The function $\{\}$

Here we obtain the first equation in (3.3). The validity of the second equation in (3.3) is checked

Proof. Let us check the validity of the first equation in (3.3). Multiplying the left- and the right-hand sides of the first equation in (2.3) by $p(X_1, n_1, X_2, n_2)$ yields

$$
\bar{R}_M^{(l)}(X_1, n_1, X_2, n_2) = \int_{-C}^{C} 2M|v|\bar{g}(v; X_1, n_1, X_2, n_2)\rho(v) dv
$$

+ \int_{-\infty}^{+\infty} \bar{R}_M(X_1 + x, n_1 + M, X_2, n_2)h_M^{(l)}(MX_1 - n_1 x, n_1) dx. \quad (3.6)

Here $\bar{g}(v; X_1, n_1, X_2, n_2)$ should be calculated for the “limiting” density with $\nu(m) \equiv 1$ for all $m \in (-\infty, +\infty)$. By (2.10), we obtain

$$
\bar{g}(v; X_1, n_1, X_2, n_2) = \int_{-\infty}^{+\infty} f_n^1(X_1 | n_1(m + v))f_n^2(X_2 | n_2(m - v)) dm = f_{n_1^2n_2^2}(Z - 2vn_1n_2),
$$

and this corresponds to (3.4). The function $h_M^{(l)}(MX_1 - n_1 x, n_1)$ does not depend on a prior distribution and is equal to

$$
\frac{\int \int f_M^1(x | Mm_1)f_n^1(X_1 | n_1m_1)f_n^2(X_2 | n_2m_2)\lambda(m_1, m_2) dm_1 dm_2}{\int \int f_n^1 + M^1(X_1 + x | (n_1 + M)m_1)f_n^2(X_2 | n_2m_2)\lambda(m_1, m_2) dm_1 dm_2}
$$

= \frac{f_M^1(x | Mm_1)f_n^1(X_1 | n_1m_1)}{f_n^1 + M^1(X_1 + x | (n_1 + M)m_1)} = (n_1 + M)f_{n_1^2n_2^2}(n_1 + M)(MX_1 - n_1 x).
$$

Next, for $n_1 + M$ and $n_2$, we recalculate $Z$ according to the rule

$$
Z \leftarrow (X_1 + x)n_2 - X_2(n_1 + M) = Z + z, \quad \text{where} \quad z = xn_2 - X_2M.
$$

Noting that $MX_1 - n_1 x = n_2^1(MZ - n_1 z)$ and changing the integration variable $x$ by $z$ in (3.6), we obtain the first equation in (3.3). The validity of the second equation in (3.3) is checked similarly. Equation (3.5) follows from (2.12) by taking into account the equality $R_M(Z, M, M) = R_N^{(l)}(\nu(m, v | Y, Z))p(Z)$. $\triangle$

Next, put $M_1 = M/D_1$, $M_2 = M/D_2$, and $U = Z/n'$, where $Z = X_1n_2 - X_2n_1$, $n' = n_1 + n_2$. Define $D_g$ and $D_h$ by the conditions $D_g = D_1D_2$ and $D_h = 0.5(D_1^{-1} + D_2^{-1})$. We use the standard notation for convolution function

$$
F(x) * G(x) = \int_{-\infty}^{\infty} F(x - y)G(y) dy.
$$

Theorem 2. Under the conditions of Theorem 1, we have the representation

$$
\bar{R}_M^{(l)}(Z, n_1, n_2) = f_{n_1^2n_2^2}(Z)R_M^{(l)}(Z/n', n_1, n_2), \quad l = 1, 2. \quad (3.7)
$$

Here $\{R_M^{(l)}(Z/n', n_1, n_2)\}$ satisfy the recursive dynamic programming equation

$$
R_M(U, n_1, n_2) = \min_{l=1, 2} R_M^{(l)}(U, n_1, n_2), \quad (3.8)
$$

PROBLEMS OF INFORMATION TRANSMISSION Vol. 54 No. 1 2018
where \( R_M^{(1)}(U, n_1, n_2) = R_M^{(2)}(U, n_1, n_2) = 0 \) if \( n_1 + n_2 = N \) and then
\[
R_M^{(1)}(U, n_1, n_2) = M g^{(1)}(U, n_1, n_2) + R_M(U, n_1 + M, n_2) \ast f_{M_1^* n_2^2(n')^{-1}(n' + M_1^*)^{-1}}(U),
\]
\[
R_M^{(2)}(U, n_1, n_2) = M g^{(2)}(U, n_1, n_2) + R_M(U, n_1 + n_2 + M) \ast f_{M_2^* n_2^2(n')^{-1}(n' + M_2^*)^{-1}}(U)
\]
\[(3.9)\]
if \( 2M \leq n_1 + n_2 < N \). Here
\[
g^{(1)}(U, n_1, n_2) = \int_0^C 2|v|g(v; U, n_1, n_2)\rho(v) \, dv,
\]
\[
g^{(2)}(U, n_1, n_2) = \int_C^\infty 2v g(v; U, n_1, n_2)\rho(v) \, dv,
\]
\[
g(v; U, n_1, n_2) = \exp \left( \frac{2Uv}{D_g^2} - \frac{2v^2 n_1 n_2}{D_g^2 n'} \right).
\]

The Bayesian strategy at \( n \leq 2M \) applies actions by turns. If \( n > 2M \), then on the time interval \((n + 1, n + M)\) the action corresponding to the smaller value \( R_M^{(1)}(U, n_1, n_2) \) must be applied. The Bayesian risk (2.12) is calculated as
\[
R_N^B(\rho(v)) = L(\rho(v)) + \int_{-\infty}^{\infty} f_{0.5MD_3D_N}(U) R_M(U, M, M) \, dU.
\]
\[(3.11)\]

**Proof.** It suffices to consider the first equation in (3.3). Substituting \( z = y - Z \) in the integral in (3.3), we obtain
\[
\bar{R}_M^{(1)}(Z, n_1, n_2) = M \int_{-\infty}^\infty 2|v| f_{n_1^* n_2^2(n')(Z - 2vn_1 n_2)\rho(v) \, dv + n_2^{-1}(n_1 + M)
\]
\[
\times \int_{-\infty}^\infty \bar{R}_M(y, n_1 + M, n_2) f_{M_1^* n_1(n_1 + M)}(n_2^{-1}(Z(n_1 + M) - n_1 y)) \, dy.
\]
\[(3.12)\]
It is straightforwardly verified that for all \( n_1, n_2, D_1, D_2, M, Z, y \) we have the identity
\[
\frac{(n_1 + M) f_{M_1^* n_1(n_1 + M)}(n_2^{-1}(Z(n_1 + M) - n_1 y) \), f_{M_1^* n_1(n_1 + M)}(y) \right)}{f_{M_1^* n_2(n')^{-1}(n' + M_1^*)^{-1}}(Z(n_1 + M) - n_1 y) \), f_{M_1^* n_2(n')^{-1}(n' + M_1^*)^{-1}}(y(n' + M_1^*) - Z(n_1 + M) - n_1 y) \right)}.
\]
\[(3.13)\]

Expressing \( \bar{R}_M^{(1)}(Z, n_1, n_2) \) and \( \bar{R}_M(y, n_1 + M, n_2) \) on the left- and right-hand sides of (3.12) through (3.7), using identity (3.13), and reducing both sides of the obtained equation by \( f_{n_1^* n_2^2(n')(Z)} \), we obtain
\[
\bar{R}_M^{(1)}\left(\frac{Z}{n'}, n_1, n_2\right) = M g^{(1)}\left(\frac{Z}{n'}, n_1, n_2\right) + \frac{1}{n' + M_1^*}
\]
\[
\times \int_{-\infty}^\infty R_M \left(\frac{y}{n' + M_1^*}, n_1 + M, n_2\right) f_{M_1^* n_2^2(n')^{-1}(n' + M_1^*)^{-1}}\left(\frac{y}{n' + M_1^*} - \frac{Z}{n'}\right) \, dy.
\]
\[(3.14)\]
If we put $U = Z/n'$, $x = y/(n' + M'_i)$, and take into account that $f_D(x - U) = f_D(U - x)$, equation (3.14) takes the form

$$R^1_M(U, n_1, n_2) = Mg^1(U, n_1, n_2) + \int_{-\infty}^{\infty} R_M(x, n_1 + M, n_2) f_{M'_1 n'_2 (n' + M'_i)}^{-1}(U - x) \, dx,$$

which is equivalent to the first equation in (3.9). The validity of the second equation in (3.9) is verified similarly. To check (3.11), we use (3.7) with

$$L = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2$$

if

$$\ell = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)$$

and then batches of sizes $L$.

To the strategy $\sigma$:

$$\text{Theorem 4. Let the following transformations be made with some } k > 0: \hat{D}_1 = kD_1, \hat{D}_2 = kD_2, \hat{m}_1 = k^{1/2} m_1, \hat{m}_2 = k^{1/2} m_2, \hat{\nu} = k^{1/2} \nu, \hat{C} = k^{1/2} C, \hat{\rho}(\hat{\nu}) = k^{-1/2} \rho(\nu), \hat{U} = k^{3/2} U, \hat{n}'_1 = k^{-1} n'_1, \hat{n}'_2 = k^{-1} n'_2, \hat{n}_1^* = kn_1^*, \hat{n}_2^* = kn_2^* \text{ (} n_1 \text{ and } n_2 \text{ are unchanged here!), and } \hat{\sigma}_\ell(\hat{U}, n_1, n_2) = \sigma(\hat{U}, n_1, n_2). \text{ Then the corresponding Bayesian risks and losses are related by the equalities}

$$\hat{R}^B_N(\hat{\rho}(\hat{\nu})) = k^{1/2} R^B_N(\rho(\nu)), \quad \hat{L}_N(\hat{\sigma}(\hat{\nu})) = k^{1/2} L_N(\sigma, \rho(\nu)).$$

Similar transformations applied to equations (2.4) and (2.5) result in the following theorem.

**Theorem 3.** Under the conditions of Theorem 1, to calculate expected losses corresponding to the strategy $\sigma$: $\sigma(\nu, n_1, n_2) = \text{Pr}(y_n = \ell | U, n_1, n_2)$, $\ell = 1, 2$, one should solve the recursive dynamic programming equation

$$L_M(U, n_1, n_2) = \sigma_1(U, n_1, n_2) L_M^1(U, n_1, n_2) + \sigma_2(U, n_1, n_2) L_M^2(U, n_1, n_2),$$

(3.15)

where $L_M^1(U, n_1, n_2) = L_M^2(U, n_1, n_2) = 0$ if $n_1 + n_2 = N$ and then

$$L_M^1(U, n_1, n_2) = Mg^1(U, n_1, n_2) + L_M(U, n_1 + M, n_2) * f_{M'_1 n'_2 (n' + M'_i)}^{-1}(U),$$

$$L_M^2(U, n_1, n_2) = Mg^2(U, n_1, n_2) + L_M(U, n_1, n_2 + M) * f_{M'_2 n'_2 (n' + M'_i)}^{-1}(U),$$

(3.16)

if $2M \leq n_1 + n_2 < N$. Losses corresponding to the strategy $\sigma$ are calculated as

$$L_N(\sigma, \rho(\nu)) = L(\rho(\nu)) + \int_{-\infty}^{\infty} f_{0.5MD^2D_N}(U) L_M(U, M, M) \, dU. \quad (3.17)$$

Remark 3. The strategy of batch processing can be generalized as follows. One can partition all the data into $K$ batches of different sizes so that, at first, actions are applied to $2M_0$ items of data by turns, and then batches of sizes $M_1, \ldots, M_{K-2}$ are processed, where $2M_0 + M_1 + \ldots + M_{K-2} = N$. Equations (3.8) and (3.9) and also (3.15) and (3.16) remain valid if $M$ is replaced by $M_i$ for $n = 2M_0$ and with $M_i$ for $n = 2M_0 + \ldots + M_{i-1}$, $i > 1$. In (3.11) and (3.17), $M$ must be replaced with $M_0$. As the results of Section 4 show, it is reasonable to choose batches of smaller size at the start of the control. Some results on using batches of different sizes are presented in [17].

The following theorem allows to reduce the considered set of pairs of variances $(D_1, D_2)$, say restrict it to pairs $(1, D)$ with $D \leq 1$.

**Theorem 4.** Let the following transformations be made with some $k > 0$: $\hat{D}_1 = kD_1, \hat{D}_2 = kD_2, \hat{m}_1 = k^{1/2} m_1, \hat{m}_2 = k^{1/2} m_2, \hat{\nu} = k^{1/2} \nu, \hat{C} = k^{1/2} C, \hat{\rho}(\hat{\nu}) = k^{-1/2} \rho(\nu), \hat{U} = k^{3/2} U, \hat{n}'_1 = k^{-1} n'_1, \hat{n}'_2 = k^{-1} n'_2, \hat{n}_1^* = kn_1^*, \hat{n}_2^* = kn_2^* \text{ (} n_1 \text{ and } n_2 \text{ are unchanged here!), and } \hat{\sigma}_\ell(\hat{U}, n_1, n_2) = \sigma(\hat{U}, n_1, n_2). \text{ Then the corresponding Bayesian risks and losses are related by the equalities}

$$\hat{R}^B_N(\hat{\rho}(\hat{\nu})) = k^{1/2} R^B_N(\rho(\nu)), \quad \hat{L}_N(\hat{\sigma}(\hat{\nu})) = k^{1/2} L_N(\sigma, \rho(\nu)).$$

(3.18)
Proof. Indeed, making these substitutions in (3.8)–(3.10), (3.15), and (3.16), we obtain by induction that
\[
\hat{R}_M(\hat{U}, n_1, n_2) = \hat{R}_M(U, n_1, n_2) , \quad \hat{L}_M(\hat{U}, n_1, n_2) = \hat{L}_M(U, n_1, n_2).
\]
Also, \(\hat{L}(\hat{\rho}(\hat{v})) = \hat{L}(\rho(\hat{v}))\). Therefore, the required equalities follow from (3.11) and (3.17). \(\triangle\)

Corollary. Consider processing of \(N = MK\) data items in \(K\) batches, each containing \(M\) items, on the set \(\Theta_N = \{|m_1 - m_2| \leq 2CM^{-1/2}\}\) and one-by-one processing of \(K\) data items on the set \(\Theta_K = \{|m_1 - m_2| \leq 2\}\). Then we have the equality
\[
N^{-1/2} R^M_N(\Theta_N) = K^{-1/2} R^M_K(\Theta_K),
\]
(3.19)
where in the notation \(R^M_N(\cdot)\) and \(R^M_K(\cdot)\) we explicitly indicate processing in batches of \(M\) items and one-by-one processing. Equality (3.19) means that the corresponding normalized risks depend only on the numbers of processed batches.

Proof. Put \(m_1 = \tilde{m}_1 M^{-1/2}\) and \(m_2 = \tilde{m}_2 M^{-1/2}\). In this case \(\Theta_N = \{|\tilde{m}_1 - \tilde{m}_2| \leq 2\}\); i.e., this change of variables maps \(\Theta_N\) to \(\Theta_K\). Next, processing of \(K\) batches each containing \(M\) data items is equivalent to one-by-one processing of \(K\) items with parameters \(\tilde{d}_1 = MD_1, \tilde{d}_2 = MD_2, \tilde{m}_1 = \tilde{m}_1 M^{1/2}\), and \(\tilde{m}_2 = \tilde{m}_2 M^{1/2}\); here the worst-case prior densities \(\tilde{\rho}_0(\tilde{v})\) and \(\rho_0(v)\) have matching supports. Validity of the equality \(\tilde{R}^B_N(\tilde{\rho}_0(\tilde{v})) = M^{1/2} R^B_K(\rho_0(v))\) follows from (3.18). This implies (3.19). \(\triangle\)

Remark 4. Since by (1.3) the normalized minimax risk \(N^{-1/2} R^M_N(\Theta)\) is bounded, the corollary implies that batch processing virtually does not increase the minimax risk if the number of batches is large enough. Assume now that incomes with non-Gaussian distributions are used. Nevertheless, according to the central limit theorem, distributions of cumulative incomes in sufficiently large batches are close to Gaussian for a wide class of processes. Therefore, the strategies for batch processing provide close values of the minimax risk for a wide class of processes with the same mathematical expectations and variances of one-step incomes; i.e., in fact, these strategies are universal.

On the other hand, it follows from the corollary that batch processing implies more severe restrictions on the set of parameters than one-by-one processing. This is due to the initial stage, where batches are processed in turns and therefore a worse action is applied to a large amount of data.

4. NUMERICAL EXPERIMENTS

Calculation of the Bayesian risk by (3.8)–(3.11) was performed for \(N = 50\) under the assumption that \(\rho(v)\) is concentrated at two points, \(v = -d_1 N^{-1/2}\) and \(v = d_2 N^{-1/2}\), with probabilities \(\rho_1\) and \(\rho_2\). The worst-case prior distribution corresponds to the maximum of the normalized Bayesian risk \(R_N(\rho(v)) = N^{-1/2} R^B_N(\rho(v))\). Parameters of the worst-case prior distributions and the corresponding normalized Bayesian risks for some pairs \((d_1, d_2)\) are presented in the table.

All the strategies have threshold behavior, i.e., there exists \(T(n_1, n_2)\) such that the first or second action is chosen according as \(U > T(n_1, n_2)\) or \(U < T(n_1, n_2)\) (in case of equality, the choice can be arbitrary). For the computed strategies, the normalized losses \(\ell_N(d) = N^{-1/2} L_N(\sigma, \rho(v))\) were calculated according to (3.15)–(3.17) assuming that \(\rho(v)\) is concentrated at a point \(v = dN^{-1/2}\). If \(d < 0\), then \(\rho_1 = 1\) and \(d = -d_1\), and if \(d > 0\), then \(\rho_1 = 0\) and \(d = d_2\). Normalized losses are presented in Fig. 1; lines 1–4 correspond to \((d_1, d_2)\) equal to \((1, 1), (1, 0.75), (1, 0.5),\) and \((1, 0.25)\).

The lines have two maxima, which are equal to maximum values of the normalized Bayesian risks and are located at points corresponding to the computed \(d_1\) and \(d_2\). This confirms the above
assumption on the worst-case prior distribution. In the case of $D_1 = D_2 = 1$, the results are in accordance with (1.3).

In more detail, the normalized losses are presented by thick line 1 in Fig. 2 in the case of $(D_1, D_2) = (1, 0.5)$; their behavior is similar for other $(D_1, D_2)$. Thick line 2 shows normalized losses without losses at the initial two stages, when actions are applied by turns. These losses tend to zero with growing $|d|$. Since at the initial two stages the losses are equal to $2|d|N^{-1/2}$ and the minimax risk is $r_N(\rho(v))N^{1/2}$, losses at the two initial stages exceed the minimax risk approximately when $2|d|N^{-1/2} > r_N(\rho(v))N^{1/2}$, i.e., at $|d| > 0.5r_N(\rho(v))N \approx 0.5 \times 0.56 \times 50 = 14$. In the domain $|d| < 14$ the strategy is minimax for $N = 50$. Furthermore, in Fig. 2 by thin solid and by thin dashed lines we present the normalized losses corresponding to the minimax strategy for $(D_1, D_2) = (1, 0.5)$ if the variances were equal to (1.05, 0.55) and (0.95, 0.45) respectively, and these lines are quite close to the original. Therefore, if data are formed of batches large enough, the variances can be estimated at the initial two stages, when actions are applied by turns, and these estimates can be used in the sequel.

Figure 3 presents results of the application of the minimax strategy corresponding to $(D_1, D_2) = (1, 0.5)$ for $0.5 \leq D_\ell \leq 1$, $\ell = 1, 2$, and $N = 50$. Line 1 describes total normalized losses for $(D_1, D_2) = (1, 1)$. Other lines describe normalized losses without the losses at the initial two stages; lines 2–7 correspond to the pairs of variances $(1, 1), (1, 0.75), (1, 0.5), (0.75, 0.75), (0.75, 0.5)$, and $(0.5, 0.5)$. One can see that for $N$ large enough, when the influence of two initial stages is small, the minimax strategy corresponding to $(D_1, D_2) = (1, 1)$ remains the same for all the presented $(D_1, D_2)$.

Finally, in Fig. 4 we present the results of Monte Carlo simulation of application of the minimax strategy corresponding to $(D_1, D_2) = (1, 0.5)$ to $N = 100$ items of data. The data were produced by summation and subsequent normalization of $M = 12$ uniformly distributed incomes with variances

| $(D_1, D_2)$ | $d_1$ | $d_2$ | $\rho_1$ | $r_N(\rho(v))$ |
|--------------|-------|-------|----------|--------------|
| (1, 1)       | 1.60  | 1.60  | 0.5      | 0.65         |
| (1, 0.75)    | 1.56  | 1.49  | 0.52     | 0.61         |
| (1, 0.5)     | 1.49  | 1.33  | 0.54     | 0.56         |
| (1, 0.25)    | 1.45  | 1.15  | 0.57     | 0.50         |

Fig. 1. Normalized losses for $(D_1, D_2) = (1, 1), (1, 0.75), (1, 0.5), (1, 0.25)$.  

PROBLEMS OF INFORMATION TRANSMISSION Vol. 54 No. 1 2018
$D_1 = 1$ and $D_2 = 0.5$ grouped into batches as is described in Section 1; i.e., we can speak of batch processing of $T = 1200$ primary data items. Lines 1–3 correspond to normalized losses obtained in data processing with batches of one, two, and four items; growth of the losses at large $d$ is due to processing the two initial batches, when actions are applied by turns. Losses occurring for large $d$ in batch processing can be reduced by one processing the data one-by-one at the start of the control. Thin solid line 4 describes processing with batches of two data items if the data in the initial four batches (eight data items) were processed one-by-one. Thin dashed lines 5 and 6 describe data processing with batches of four data items if at the start the data in four or two batches, respectively, were processed one-by-one. Note that according to the corollary, processing in batches of four data items is equivalent to one-by-one processing of $N = 25$ data; i.e., the strategy provides good enough control performance already for $T = 300$ primary data items, but in this case the maximal admissible value of $d$ must be taken approximately one fourth as large.

5. CONCLUSION

An algorithm for finding the minimax strategy for a Gaussian two-armed bandit is proposed for any finite control horizon $N$. The Gaussian two-armed bandit arises in batch processing op-
timization when cumulative incomes in data batches are used for the control. Since cumulative incomes in the batches have approximately Gaussian (normal) distributions for a wide class of original processes, the proposed strategies are universal. Moreover, numerical experiments show that the strategies possess fine robustness properties and can well find practical use. First, it turned out that there are no strict requirements on closeness of distributions of cumulative incomes in data batches to Gaussian distributions. For example, Monte Carlo simulations in Section 4 were implemented for uniformly distributed incomes which were grouped into batches of twelve items only. Second, it turned out that behavior of the control changed a little if variances used for finding the strategy were given with large inaccuracies of up to 5–10%. This allows to estimate variances at the initial stage, when both actions are applied by turns, and then use these estimates for the control.

The initial stage of the control deserves special attention. Although applying actions by turns at the start of the control is optimal if all the batches have equal sizes, such control can considerably enlarge the total losses if mathematical expectations of one-step incomes differ significantly. To decrease the total losses, one can use batches of smaller sizes at the start of the control, but this leads to a decrease in the accuracy of estimating the variances of one-step incomes used by the control strategy. Thus, the choice of initial sizes of batches is governed by two contradictory requirements, and hence control at the initial stage providing a trade-off between this requirements presents a challenging problem of independent interest.

The author is grateful to a reviewer for the attention to the paper and valuable remarks.

REFERENCES

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

2. Presman, E.L. and Sonin, I.M., Posledovatel'noe upravlenie po nepolnym dannym. Baiesovskii podkhod, Moscow: Nauka, 1982. Translated under the title Sequential Control with Incomplete Information, New York: Academic, 1990.

3. Tsetlin, M.L., Issledovaniya po teorii automatov i modelirovanii biologicheskikh sistem, Moscow: Nauka, 1969. Translated under the title Automaton Theory and Modeling of Biological Systems, New York: Academic, 1973.
4. Varshavsky, V.I., *Kollektivnoe povedenie avtomatov* (Collective Behavior of Automata), Moscow: Nauka, 1973. Translated under the title *Kollektives Verhalten von Automaten*, Warschawski, W.I., Berlin: Akademie, 1978.

5. Sragovich, V.G., *Adaptivnoe upravlenie* (Adaptive Control), Moscow: Nauka, 1981. Translated under the title *Mathematical Theory of Adaptive Control*, Singapore: World Sci., 2006.

6. Nazin, A.V. and Poznyak, A.S., *Adaptivnyi vybor variantov: rekurrentnye algoritmy* (Adaptive Choice between Alternatives: Recursive Algorithms), Moscow: Nauka, 1986.

7. Robbins, H., Some Aspects of the Sequential Design of Experiments, *Bull. Amer. Math. Soc.*, 1952, vol. 58, no. 5, pp. 527–535.

8. Fabius, J. and van Zwet, W.R., Some Remarks on the Two-Armed Bandit, *Ann. Math. Statist.*, 1970, vol. 41, no. 6, pp. 1906–1916.

9. Vogel, W., An Asymptotic Minimax Theorem for the Two Armed Bandit Problem, *Ann. Math. Statist.*, 1960, vol. 31, no. 2, pp. 444–451.

10. Bather, J.A., The Minimax Risk for the Two-Armed Bandit Problem, *Mathematical Learning Models—Theory and Algorithms*, Herkenrath, U., Kalin, D., and Vogel, W., Eds., Lect. Notes Statist, vol. 20, New York: Springer, 1983, pp. 1–11.

11. Lai, T.L., Levin, B., Robbins, H., and Siegmund, D., Sequential Medical Trials (Stopping Rules/Asymptotic Optimality), *Proc. Natl. Acad. Sci. USA*, 1980, vol. 77, no. 6, Part 1, pp. 3135–3138.

12. Cesa-Bianchi, N. and Lugosi, G., *Prediction, Learning, and Games*, Cambridge: Cambridge Univ. Press, 2006.

13. Juditsky, A., Nazin, A.V., Tsybakov, A.B., and Vayatis, N., Gap-Free Bounds for Stochastic Multi-Armed Bandit, in *Proc. 17th IFAC World Congr., Seoul, Korea, July 6–11, 2008*, pp. 11560–11563. Available at [http://www.ifac-papersonline.net/Detailed/37644.html](http://www.ifac-papersonline.net/Detailed/37644.html).

14. Gasnikov, A.V., Nesterov, Yu.E., and Spokoiny, V.G., On the Efficiency of a Randomized Mirror Descent Algorithm in Online Optimization Problems, *Zh. Vychisl. Mat. Mat. Fiz.*, 2015, vol. 55, no. 4, pp. 582–598 [Comput. Math. Math. Phys. (Engl. Transl.), 2015, vol. 55, no. 4, pp. 580–596].

15. Kolnogorov, A.V., Determination of Minimax Strategies and Risk in a Random Environment (the Two-Armed Bandit Problem), *Avtomat. i Telemekh.*, 2011, no. 5, pp. 127–138 [Autom. Remote Control (Engl. Transl.), 2011, vol. 72, no. 5, pp. 1017–1027].

16. Kolnogorov, A.V., One-Armed Bandit Problem for Parallel Data Processing Systems, *Probl. Peredachi Inf.*, 2015, vol. 51, no. 2, pp. 99–113 [Probl. Inf. Trans. (Engl. Transl.), 2015, vol. 51, no. 2, pp. 177–191].

17. Oleynikov, A.O., Numerical Optimization of Parallel Processing in a Stationary Environment, *Trans. Karelian Res. Centre Russ. Acad. Sci.*, 2013, no. 1, pp. 73–78.