Universal strategies for the two-alternative big data processing

A V Kolnogorov
Yaroslav-the-Wise Novgorod State University, Veliky Novgorod, Russia
E-mail: Alexander.Kolnogorov@novsu.ru

Abstract. We consider the two-alternative processing of big data in the framework of the two-armed bandit problem. We assume that there are two processing methods with different, fixed but a priori unknown efficiencies which are due to different reasons including those caused by legislation. Results of data processing are interpreted as random incomes. During control process, one has to determine the most efficient method and to provide its primary usage. The difficulty of the problem is caused by the fact that its solution essentially depends on distributions of one-step incomes corresponding to results of data processing. However, in case of big data we show that there are universal processing strategies for a wide class of distributions of one-step incomes. To this end, we consider Gaussian two-armed bandit which naturally arises when batch data processing is analyzed. Minimax risk and minimax strategy are searched for as Bayesian ones corresponding to the worst-case prior distribution. We present recursive integro-difference equation for computing Bayesian risk and Bayesian strategy with respect to the worst-case prior distribution and a second order partial differential equation into which integro-difference equation turns in the limiting case as the control horizon goes to infinity. We also show that, in case of big data, processing of data one-by-one is not more efficient than optimal batch data processing for some types of distributions of one-step incomes, e.g. for Bernoulli and Poissonian distributions. Numerical experiments are presented and show that proposed universal strategies provide high performance of two-alternative big data processing.

1. Introduction
Let’s consider the two-armed bandit problem [1]–[3] as applied to optimization of data processing when two alternative processing methods with different a priori unknown efficiencies are available. During a control process, one has to determine the most efficient processing method and to provide its preferential usage. Formally, two-armed bandit is a controlled random process \( \xi_n, \ n = 1, 2, ..., N \), which values depend only on currently chosen processing methods, also called actions, and are interpreted as incomes. For example, incomes of a Bernoulli two-armed bandit can take only two possible values 0 and 1 and are considered as successfully and unsuccessfully processed data terms.

We consider here a Gaussian two-armed bandit which naturally arises when batch data processing is implemented. In this case, all data is split into batches. The same action is applied to all data terms of the batch and cumulative income is then used for the control. If batch sizes are large enough then, according to the central limit theorem, cumulative incomes have approximately Gaussian distributions. In what follows, we assume that distributions of incomes \( \xi_n, \ n = 1, 2, ..., N \) are Gaussian with probability distribution function \( f_{D_k}(x|m_k) \) if the \( k \)-th action is currently chosen. Here
In what follows, we assume that variances $\sigma_1, \sigma_2$ are a priori known (this assumption can be omitted later) and mathematical expectations $m_1, m_2$ are a priori unknown. So, Gaussian two-armed bandit is completely described by a vector parameter $\theta = (m_1, m_2)$. The set of parameters is the following $\Theta = \{ \theta : |m_1 - m_2| \leq 2C \}, 0 < C < \infty$. Note that considered approach assumes that all the batches have equal sizes. Another approach to batch data processing which uses a small number of batches increasing in sizes is considered in [4].

A control strategy $\sigma$ at the point of time (number of the batch) $(n+1)$ determines the choice of the action $y_{n+1}$ depending on the history of the control, i.e., cumulative numbers $n_1, n_2$ of both actions applications and corresponding cumulative incomes $X_1, X_2$ $(n_1 + n_2 = n)$. The goal of the control is considered in minimax setting. Clearly, if one knew both expectations $m_1, m_2$, he (or she) would always use the action corresponding to the largest of them, the total expected income would thus be equal to $N \times \max(m_1, m_2)$. Since actually the strategy $\sigma$ is used, the actual expected total income is less than the maximal by the magnitude

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

which is called a regret. A regret arises due to incomplete prior information about the two-armed bandit. Here $E_{\sigma, \theta}$ denotes the sign of mathematical expectation with respect to the measure generated by strategy $\sigma$ and parameter $\theta$. Let’s introduce a prior probability distribution density $\lambda(\theta) = \lambda(m_1, m_2)$ on $\Theta$. Minimax and Bayesian risks are defined as

$$R^M_N(\theta) = \inf_{\sigma} \sup_{\theta} L_N(\sigma, \theta),$$

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

Corresponding optimal strategies $\sigma^M$ and $\sigma^B$ are called minimax and Bayesian strategies. Both minimax and Bayesian approaches have advantages and disadvantages. Bayesian approach allows to determine numerically Bayesian strategy and Bayesian risk using dynamic programming technique for arbitrary prior distribution but it requires the assignment of this prior distribution for which there are no clear criteria. Minimax approach does not require assignment of the prior distribution, it is robust, but there is no a direct method for determining minimax strategy and minimax risk. Minimax and Bayesian approaches are united by the main theorem of the game theory according to which the following equality holds true for minimax risk (2) and Bayesian risk (3):

$$R^M_N(\theta) = R^B_N(\lambda^0) = \sup_{\lambda} R^B_N(\lambda),$$

and minimax strategy $\sigma^M$ is equal to corresponding Bayesian strategy $\sigma^B(\lambda^0)$. Prior distribution density $\lambda^0$ is called the worst-case distribution density.

Note that minimax approach was first proposed by H. Robbins in [5] for Bernoulli two-armed bandit. Although an explicit solution to the problem was not discovered, an asymptotic minimax theorem was proved by W. Vogel in [6]. It states that as $N \to \infty$ the following inequalities hold true

$$0.612 < (DN)^{-1/2} R^M_N(\theta) < 0.752,$$

where $D = 0.25$ is the maximum variance of one-step income of Bernoulli two-armed bandit. Presented here the lower bound was proved in [7].

2. The case of moderate $N$

We consider this case according to [8] where minimax risk is searched for using the equality (4). However, it is important to understand that direct usage of equality (4) is virtually impossible due to the high computational complexity. Therefore, in [8] the class of prior distribution densities to which the worst-case prior belongs was described. Let’s put $m_1 = m + \nu, m_2 = m - \nu$. Then the set of
parameters is the following $\theta = \{ \theta : |v| < C \}$. Asymptotically the worst-case prior distribution density is

$$v_{\alpha}(m,v) = \kappa_{\alpha}(m) \times \rho(v), \quad (6)$$

where $\kappa_{\alpha}(m) = (2\alpha)^{-1}$ is the uniform distribution on the segment $m \in [-a, a]$, $\rho(v)$ is a distribution density on $v \in [-C, C]$ and $\alpha \to \infty$. If $D_1 = D_2$ then $\rho(v)$ is a symmetric density, i.e. $\rho(v) = \rho(-v)$.

Denote

$$R^B_N(\rho(v)) = \lim_{\alpha \to \infty} R^B_N(v_{\alpha}(m,v)).$$

Let’s present a recursive dynamic programming equation to compute a Bayesian risk $R^B_N(\rho(v))$ and corresponding Bayesian strategy with respect to any prior distribution density $\rho(v)$. We assume that data processing is carried out by batches of size $M$ and at initial stage two batches of total size $2M_0$ are processed by equally applied actions. The choice of actions depends on currently observed statistics $(U, n_1, n_2)$ where $U = (X_1 X_2 - X_2 n_2)/n', n' = n'_1 + n'_2, n'_1 = n_1/D_1, n'_2 = n_2/D_2$. We also denote $M_1 = MD_1, M_2 = MD_2, M'_1 = M/D_1, M'_2 = M/D_2, f_0(u) = f_0(u|0)$. One has to solve a dynamic programming Bellman-type: recursive equation

$$R_M(U, n_1, n_2) = \min_{k=1,2} R^{(k)}_M(U, n_1, n_2), \quad (7)$$

where

$$R^{(1)}_M(U, n_1, n_2) = R^{(2)}_M(U, n_1, n_2) = 0 \quad (8)$$

If $n_1 + n_2 = N$ and then

$$R^{(1)}_M(U, n_1, n_2) = M g^{(1)}(U, n_1, n_2) + R_M(U, n_1 + M, n_2) \ast f_{M'_1 n'_1(N')}^{-1}(n' + M'_1) \ast (U), \quad (9)$$

$$R^{(2)}_M(U, n_1, n_2) = M g^{(2)}(U, n_1, n_2) + R_M(U, n_1, n_2 + M) \ast f_{M'_2 n'_2(N')}^{-1}(n' + M'_2) \ast (U) \quad (10)$$

if $2M_0 \leq n_1 + n_2 < N$. Here $F(u) * g(u)$ denotes convolution of functions $F(u)$ and $g(u)$,

$$g^{(1)}(U, n_1, n_2) = \int_0^U 2|v| g(v; U, n_1, n_2) \rho(v) dv, \quad (11)$$

$$g^{(2)}(U, n_1, n_2) = \int_0^C 2v g(v; U, n_1, n_2) \rho(v) dv, \quad (11)$$

$$g(v; U, n_1, n_2) = \exp \left( 2D_1^{-1} D_2^{-1} (U + v^2 n_1 n_2 (n')^{-1}) \right).$$

Bayesian strategy at initial stage of control ($n \leq 2M_0$) chooses actions by turns and then at $n > 2M_0$ chooses the action corresponding to currently smaller value of $R^{(1)}_M(U, n_1, n_2)$, $R^{(2)}_M(U, n_1, n_2)$, in case of a draw the choice may be arbitrary. The strategy has a thresholding nature, i.e. there exists a set of thresholds $\{T(n_1, n_2)\}$ such that the first and the second actions are optimal according to conditions $U > T(n_1, n_2)$ and $U < T(n_1, n_2)$. If $U = T(n_1, n_2)$ then the choice of actions is arbitrary. Bayesian risk is calculated according to the formula

$$R^B_N(\rho(v)) = M \int_{-C}^{C} 2|v| \rho(v) dv + \int_{-\infty}^{\infty} f_{MD_1^2 (D_1 + D_2)^{-1}} (U) R_M(U, M_0, M_0) du. \quad (12)$$

In [8] it was proved that minimax risk depends little on the number of batches into which the data is split if the number of batches is large enough. Let’s give a simple explanation of this result. Assume for simplicity that $D_1 = D_2 = D$. Like the estimates (5), the asymptotic minimax theorem for Gaussian two-armed bandit states that $R^M_N(\theta) \sim r(DN)^{1/2}$ as $N \to \infty$ where $r$ is a factor. Hence, for finite $N$ the maximum regret is $r_N(DN)^{1/2}$ where $r_N \to r$ as $N \to \infty$. Let’s assume that $N = M \times K$ and split all the data into $M$ batches each containing $K$ data. For $K$ batches the maximum regret is equal to $r_K(D'K)^{1/2}$ where $D' = MD$ is the variance of the batch. So, the maximum regret corresponding to
batch processing is equal to \( r_K(D \times M \times K)^{1/2} = r_K(DN)^{1/2} \). If \( K \) is large enough then both \( r_K \) and \( r_K \) are close to \( r \). Actually, \( K = 50 \) provides the minimax risk which is only 2\% higher than the minimax risk corresponding to case \( N \to \infty \).

\[ \text{Figure 1. Normalized regret corresponding to minimax strategy.} \]

To determine numerically minimax strategy and minimax risk for \( N = 50, D_1 = 1, D_2 = 0.5 \) it was assumed 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 \( 1 - \rho_1 \). Maximum of normalized Bayesian risk \( N^{-1/2} R_N^B(\rho(v)) \) was approximately equal to 0.56 and was attained at \( d_1 \approx 1.49, d_2 \approx 1.33, \rho_1 \approx 0.54 \). Then determined Bayesian strategy was used to compute normalized regret \( l_N(d) = N^{-1/2} L_N(\sigma^B, \theta_d) \) where \( \theta_d = (m + dN^{-1/2}, m - dN^{-1/2}) \). Results are presented on figure 1. Curves 1 present normalized regrets including those at the initial stage of the control where both actions are equally applied. Their growth at large \( |d| \) are caused by initial stage of the control. Curves 2 present normalized regrets without those at the initial stage of the control. The maximum values of normalized regrets are approximately equal to 0.56 like normalized Bayesian risk itself and this confirms that determined Bayesian strategy is approximately minimax one in the domain presented on figure 1.

### 3. Universality of considered strategies

Recall that batch data processing assumes that the same actions are applied to the data packets and then cumulative incomes are used for the control. This means that batch processing strategies presented in [8] are universal and can be used to control a wide class of random processes which one-step incomes follow the central limit theorem. If the number of batches is large enough than minimax risk depends only on the total number of data \( N \) and variances of one-step incomes \( D_1, D_2 \). Moreover, \( D_1, D_2 \) can be considered to be known because they can be estimated at the short initial stage of the control and minimax risk depends little on the changes of these variances. So, the estimates of variances obtained at the initial stage can be used for the control at the remaining control horizon.
Figure 2. Normalized regrets of Poissonian two-armed bandit.

On figure 2 we present normalized regrets $l_T(d) = (\lambda T)^{-1/2}L_T(\sigma, \theta_d)$ of Poissonian two-armed bandit where $\theta_d = (\lambda + d(\lambda/T)^{1/2}, \lambda - d(\lambda/T)^{1/2})$ and $\sigma$ is a minimax strategy for Gaussian two-armed bandit. The rate $\lambda$ was always chosen equal to 1 and control horizons $T$ were chosen equal to 600, 3000 and 15000. The number of batches was always equal to 30, hence, the sizes of batches were equal to 20, 100 and 500 respectively. The estimates of $\lambda$ were done at initial two stages when both actions were equally applied and then used at remaining 28 stages of control. One can see that all three curves are close to each other. Maximum regrets are approximately equal to 0.667 and this provides a good performance of batch data processing.

4. Limiting description

Results are generalized to the case $N \to \infty$. It is important to understand that maximum values of the regret are attained in the domain of ‘‘close’’ distributions which satisfy a condition that the absolute difference of mathematical expectations $|m_1 - m_2|$ is the magnitude of the order of $N^{-1/2}$. In fact, if $|m_1 - m_2| = 2cN^{-1/2}$ then the error of determining the maximum of $m_1, m_2$ is always not less than some $p_{error}(c) > 0$. Therefore, one-step regret is not less than $L_1 = p_{error}(c)2cN^{-1/2}$ and the total regret is not less than $NL_1 = p_{error}(c)2cN^{1/2}$. On the other hand, given that $|m_1 - m_2| > \delta > 0$, one can prove that the maximum regret is the magnitude of the order $log(N)$ as $N \to \infty$ (see, e.g., [9]).

In the domain of ‘‘close’’ distributions formulas (7)–(12) can be written in invariant form with control horizon one. In the limiting case $N \to \infty$ this recursive equation turns into second order partial differential equation. Let’s put $C = cN^{-1/2}, w = N^{1/2}v, \varrho(w) = N^{-1/2}p(v), u = UN^{-1/2}, t_k = \eta_kN^{-1}, t'_k = \eta_kN^{-1}, t = t_1 + t_2, t' = t'_1 + t'_2, \tau_k(u, t_1, t_2) = N^{-1/2}R_M(u, n_1, n_2), \varepsilon_0 = M_0N^{-1}, \varepsilon = MN^{-1}$. Then there exists a continuous limit $r(u, t_1, t_2) = \lim_{\varepsilon \to 0}r_\varepsilon(u, t_1, t_2)$ which satisfies the second order partial differential equation

$$\min_{k=1,2} \left( \frac{\partial r}{\partial t_k} + \frac{\partial k^2 - k}{2(t')^2} \times \frac{\partial^2 r}{\partial u^2} + g^{(k)}(u, t_1, t_2) \right) = 0$$

(13)

with initial condition

$$r(u, t_1, t_2) = 0 \quad \text{if} \quad t_1 + t_2 = 1.$$ 

(14)

Bayesian strategy prescribes to apply the $k$-th action if currently the $k$-th term in the left-hand side of (13) has smaller value. The limiting value of Bayesian risk is computed as
\[ \lim_{N \to \infty} N^{-1/2} R^B_N(\rho_N(v)) = r(0,0,0). \]  

This makes it possible to estimate a performance of batch processing. For example, batch processing of data split into 50 batches provides the magnitude of minimax risk only 2% larger than its limiting value when the number of batches grows infinitely.

5. Optimal one-by-one data processing

Results above show that maximum regret caused by batch data processing depends little on the number of batches if this number is large enough. However, there is a question of whether a reduction of the maximum regret is possible by using one-by-one data processing. In [10] we show that such reduction is impossible for a Bernoulli two-armed bandit. To this end, we show that for a specially chosen prior distribution density the normalized Bayesian risk of the Bernoulli two-armed bandit satisfies the second order partial differential equation (13) with initial condition (14) like the Bayesian risk of the Gaussian two-armed bandit computed with respect to the worst-case prior distribution. Since minimax risk is not less than any Bayesian, this means that minimax risk corresponding to the optimal one-by-one data processing of Bernoulli two-armed bandit is not less than minimax risk of Gaussian two-armed bandit which describes batch data processing. This also means that in case of big data optimal one-by-one data processing is not more efficient that optimal batch data processing. One can expect that this is true for a wide class of two-armed bandits with different distributions of one-step incomes. For example, it is true for exponential and Poissonian two-armed bandits, too.

6. Conclusion

The approach to the two-alternative processing of big data is proposed in the framework of the two-armed bandit problem. This approach is based on batch processing of data. In case of big data this makes it possible to use universal processing strategies almost without additional losses in control performance.

Acknowledgments

This research was supported by Russian Foundation for Basic Research, project no 18-29-16223-mk.

References

[1] Berry D A and Fristedt B 1985 Bandit Problems: Sequential Allocation of Experiments (Chapman & Hall: London)
[2] Sragovich V G 2006 Mathematical Theory of Adaptive Control (Singapore: World Sci.)
[3] Cesa-Bianchi N and Lugosi G 2006 Prediction, Learning, and Games (Cambridge: Cambridge Univ. Press)
[4] Perchet V, Rigollet P, Chassang S and Snowberg E 2016 Batched bandit problems Annals of Statistics 44 660–81
[5] Robbins H 1952 Some aspects of the sequential design of experiments Bulletin AMS 58 527–35
[6] Vogel W 1960 An asymptotic minimax theorem for the two-armed bandit problem Ann. Math. Stat. 31 444–51
[7] Bather J A 1983 The minimax risk for the two-armed bandit problem Mathematical Learning Models – Theory and Algorithms (Lect Notes Statist vol 20) ed U Herkenrath, D Kalin and W Vogel (New York: Springer) 1–11
[8] Kolnogorov A V 2018 Gaussian two-armed bandit and optimization of batch data processing Problems of Information Transmission 54 84–100
[9] Lai T L, Levin B, Robbins H and Siegmund D 1980 Sequential medical trials (stopping rules/asymptotic optimality) Proc. Natl. Acad. Sci. USA 77 3135–38
[10] Kolnogorov A V 2020 Gaussian two-armed bandit: limiting description Problems of Information Transmission 56 278–301