Two-alternative optimization of moderate batch data processing

A V Kolnogorov
Yaroslav-the-Wise Novgorod State University, 41, ul. B. St. Petersburgskaya, Veliky Novgorod, Russian Federation
E-mail: Alexander.Kolnogorov@novsu.ru

Abstract. We consider optimization of moderate batch data processing in the framework of Bernoulli two-armed bandit problem with indefinite control horizon. We assume that there are two alternative processing methods with different a priori unknown efficiencies which are caused by different reasons including those related to legislation. The arising batches of data with close properties have moderate and possibly uncertain sizes. The problem is considered in minimax setting. According to the main theorem of the game theory minimax risk and minimax strategy are searched for as Bayesian ones corresponding to approximately the worst-case prior distribution concentrated on the finite set of parameters. Numerical experiments show that this approach provides good approximations of minimax strategy and minimax risk.

1. Introduction
We consider optimization of data processing in a framework of Bernoulli two-armed bandit problem (see, e.g. [1]) if there are two alternative processing methods available with different a priori unknown efficiencies which are caused by different reasons including those related to legislation. The problem is also well known as a problem of adaptive learning and adaptive control [2]–[4]. A Bernoulli two-armed bandit is a random control process \( \xi_n, n = 1,2,\ldots,N \) which values are interpreted as incomes and depend only on currently chosen actions (processing methods) \( y_n \) as follows: \( \Pr(\xi_n = 1 | y_n = \ell) = p_\ell \), \( \Pr(\xi_n = 0 | y_n = \ell) = q_\ell \), \( \ell \in \{1,2\} \). In the sequel, a vector parameter \( \theta = (p_1, p_2) \) is used to describe a Bernoulli two-armed bandit. The value of the process \( \xi_n = 1 \) corresponds to successfully processed data item number \( n \) and \( \xi_n = 0 \) corresponds to unsuccessfully processed data item number \( n \). The goal is to maximize (in some sense) the cumulative expected number of successfully processed data. The total number of data \( N \) is also called a control horizon and is assumed to have moderate value according to distribution \( \Pr(N = N_j) = \beta_j, \ j = 1,\ldots,J, \sum_{j=1}^J \beta_j = 1 \). Note that the case of large \( N \) which corresponds to big data processing is presented in [5–8] where different strategies, including batch data processing, are investigated.

A control strategy \( \sigma \) at the point of time \( n + 1 \) describes the choosing of the action \( y_{n+1} \) depending on currently available history of the process, i.e. \( n_1, n_2 \), cumulative numbers of both actions applications and \( X_1, X_2 \), corresponding cumulative incomes (\( n_1 + n_2 = n \)). Hence,
\( \sigma_{t}(X_{1,n_{1}}, X_{2,n_{2}}) = \Pr(Y_{n+1} = \ell \mid X_{1,n_{1}}, X_{2,n_{2}}) \). Let \( N = \max(N_{1},...,N_{j}) \) denote maximum of \( N_{1},...,N_{j} \). A regret, which is also often called the loss function, is defined as follows

\[
L_{N}(\sigma, \theta, \{\beta_{j}, N_{j}\}) = \sum_{j=1}^{J} \beta_{j} L(\sigma, \theta, N_{j})
\]

where

\[
L(\sigma, \theta, N) = N \max(p_{1}, p_{2}) - E_{\sigma, \theta}\left( \sum_{n=1}^{N} \sigma_{n} \right),
\]

and describes expected losses of the total income due to incomplete information. Here \( E_{\sigma, \theta} \) denotes the mathematical expectation with respect to the measure generated by strategy \( \sigma \) and parameter \( \theta \). The set of parameters contains all possible pairs \( (p_{1}, p_{2}) \), i.e. \( \Theta = \{(p_{1}, p_{2}) : 0 \leq p_{1} \leq 1; \ell = 1,2\} \). The minimax risk is defined as

\[
R_{N}^{M}(\Theta, \{\beta_{j}, N_{j}\}) = \inf_{\sigma} \sup_{\theta} L_{N}(\sigma, \theta, \{\beta_{j}, N_{j}\})
\]

corresponding optimum strategy \( \sigma_{M}^{\star} \) is called minimax strategy. According to the main theorem of the game theory minimax risk and minimax strategy can be determined as Bayesian ones calculated over the worst-case prior distribution at which Bayesian risk attains its maximum value. In what follows, we determine approximately the worst-case prior distributions on the finite number sets of parameters

\[
p_{ij}^{l} = p_{i} + \Delta p_{l}, \quad p_{2,i}^{l} = p_{i} - \Delta p_{l}, \quad p_{1,K+i} = p_{i} - \Delta p_{l}, \quad p_{2,K+i} = p_{i} + \Delta p_{l},
\]

\[
\Pr(\theta = (p_{1,i}^{l}, p_{2,i}^{l})) = \lambda_{i}, \quad \Pr(\theta = (p_{1,K+i}, p_{2,K+i})) = \lambda_{K+i}, \quad i = 1,\ldots,K,
\]

where \( \lambda_{i} = \lambda_{K+i} > 0, \quad i = 1,\ldots,K; \quad \sum_{i=1}^{K} \lambda_{i} = 0.5 \). So, we consider symmetric finite sets of parameters. Note that finite sets of parameters in Bernoulli two-armed bandit problem were considered in [9]. Given this prior distribution, Bayesian risk is defined as follows

\[
R_{N}^{B}(\{p_{ij}, \Delta p_{i}, \lambda_{i}\}, \{\beta_{j}, N_{j}\}) = \inf_{\{p_{ij}, \Delta p_{i}, \lambda_{i}\}} \sum_{i=1}^{K} (L_{N}(\sigma, \theta, \{\beta_{j}, N_{j}\}) + L_{N}(\sigma, \theta, \{\beta_{j}, N_{j}\}) \lambda_{i})
\]

(2)

corresponding optimal strategy \( \sigma_{B}^{\star} \) is called Bayesian strategy. If parameters \( \{p_{ij}, \Delta p_{i}, \lambda_{i}\} \) are properly assigned one can expect that the approximate equality holds

\[
R_{N}^{M}(\Theta, \{\beta_{j}, N_{j}\}) \approx \max_{\{p_{ij}, \Delta p_{i}, \lambda_{i}\}} R_{N}^{B}(\{p_{ij}, \Delta p_{i}, \lambda_{i}\}, \{\beta_{j}, N_{j}\})
\]

The rest of the paper is organized as follows. Recursive Bellman-type equation for determining Bayesian risk and Bayesian strategy is presented in section 2. In section 3 a recursive equation for determining a regret is derived. In section 4 we present numerical results. Section 5 contains conclusion.

2. Recursive equation for Bayesian risk and Bayesian strategy finding

Let’s put \( N_{0} = 0 \) and \( \gamma_{n} = \sum_{i=j}^{J} \beta_{i} \) if \( N_{j-1} \leq n < N_{j} \). One can see that the following equality holds

\[
L_{N}(\sigma, \theta, \{\beta_{j}, N_{j}\}) = \sum_{n=1}^{N} \gamma_{n} \left( \max(p_{1}, p_{2}) - E_{\sigma, \theta, \sigma_{n}} \right)
\]

(3)

i.e. considered problem is equivalent to Bernoulli two-armed bandit problem with incomes discounted by factors \( \{\gamma_{n}\} \). Calculation of corresponding Bayesian risk \( R_{N}^{B}(\{p_{ij}, \Delta p_{i}, \lambda_{i}\}, \{\beta_{j}, N_{j}\}) \) can be done using standard dynamic programming technique. Given a prior distribution \( \lambda_{i}, \quad i = 1,\ldots,2K \), the posterior distribution at the point of time \( n = n_{1} + n_{2} \) is calculated as
\[
\lambda_i(X_1,n_1,X_2,n_2) = \frac{B(X_1,n_1,p_{1,i})B(X_2,n_2,p_{2,i})\lambda_i}{P(X_1,n_1,X_2,n_2)}, \quad i = 1,...,2K,
\]

where

\[
P(X_1,n_1,X_2,n_2) = \sum_{i=1}^{2K} B(X_1,n_1,p_{1,i})B(X_2,n_2,p_{2,i})\lambda_i
\]

and

\[
B(X,n,p) = \left(\frac{n}{X}\right)p^X(1-p)^{n-X}.
\]

Denote \(x^+ = \max(x,0)\). The standard recursive Bellman-type equation for determining Bayesian risk (2) calculated with respect to regret (3) is as follows

\[
R(X_1,n_1,X_2,n_2) = \min\{R^{(1)}(X_1,n_1,X_2,n_2), R^{(2)}(X_1,n_1,X_2,n_2)\},
\]

where \(R^{(1)}(X_1,n_1,X_2,n_2) = R^{(2)}(X_1,n_1,X_2,n_2) = 0\) if \(n_1 + n_2 = N\) and then

\[
R^{(1)}(X_1,n_1,X_2,n_2) = \sum_{i=1}^{2K} \lambda_i(X_1,n_1,X_2,n_2) \times
\]

\[
\times \left(\gamma_n(p_{2,i} - p_{1,i})^+ + \frac{1}{x=0} \sum R(X_1 + x,n_1 + 1,X_2,n_2)p_{1,i}(x)\right),
\]

\[
R^{(2)}(X_1,n_1,X_2,n_2) = \sum_{i=1}^{2K} \lambda_i(X_1,n_1,X_2,n_2) \times
\]

\[
\times \left(\gamma_n(p_{1,i} - p_{2,i})^+ + \frac{1}{x=0} \sum R(X_1,n_1,X_2 + x,n_2 + 1)p_{2,i}(x)\right),
\]

where \(p_{\ell,i}(1) = p_{\ell,i}, \quad p_{\ell,i}(0) = 1 - p_{\ell,i}, \quad \ell = 1,2,\) if \(0 \leq N \leq N - 1\). Here \(R^{(\ell)}(X_1,n_1,X_2,n_2)\) is equal to expected losses on the residual control horizon \([n+1,N]\) if initially the \(\ell\)-th action is applied and then control is optimally implemented \((\ell = 1,2)\). Bayesian strategy prescribes to choose the \(\ell\)-th action if \(R^{(\ell)}(X_1,t_1,X_2,t_2)\) has smaller value. In case of a draw

\[
R^{(1)}(X_1,t_1,X_2,t_2) = R^{(2)}(X_1,t_1,X_2,t_2)\]

the choice of the action is arbitrary. Bayesian risk (2) is calculated by the formula

\[
R_B^N((p_1,\Delta p_1,\lambda_1,\{\beta_j, N_j\}) = R(0,0,0,0).
\]

Now let’s consider another version of recursive equation. Denote

\[
r(X_1,n_1,X_2,n_2) = R(X_1,n_1,X_2,n_2) \times P(X_1,n_1,X_2,n_2),
\]

where \(P(X_1,n_1,X_2,n_2)\) is defined in (4). The following recursive equation holds

\[
r(X_1,n_1,X_2,n_2) = \min\{r^{(1)}(X_1,m_1,X_2,n_2), r^{(2)}(X_1,n_1,X_2,n_2)\},
\]

where \(r^{(1)}(X_1,n_1,X_2,n_2) = r^{(2)}(X_1,n_1,X_2,n_2) = 0\) if \(n_1 + n_2 = N\) and then

\[
r^{(1)}(X_1,n_1,X_2,n_2) = g^{(1)}(X_1,n_1,X_2,n_2) + \frac{1}{x=0} \sum r(X_1 + x,n_1 + 1,X_2,n_2) \times h_x(X_1,n_1),
\]

\[
r^{(2)}(X_1,n_1,X_2,n_2) = g^{(2)}(X_1,n_1,X_2,n_2) + \frac{1}{x=0} \sum r(X_1,n_1,X_2 + x,n_2 + 1) \times h_x(X_2,n_2),
\]

where
\[
g^{(1)}(X_1, n_1, X_2, n_2) = 2\gamma_n \times \sum_{i=1}^{K} \Delta p_i B(X_1, n_1; p_i - \Delta p_i) B(X_2, n_2; p_i + \Delta p_i) \lambda_i,
\]
\[
g^{(2)}(X_1, n_1, X_2, n_2) = 2\gamma_n \times \sum_{i=1}^{K} \Delta p_i B(X_1, n_1; p_i + \Delta p) B(X_2, n_2; p_i - \Delta p) \lambda_i,
\]
and
\[
h_0(X_\ell, n_\ell) = \frac{n_\ell + 1 - X_\ell}{n_\ell + 1},
\]
\[
h_1(X_\ell, n_\ell) = \frac{X_\ell + 1}{n_\ell + 1}.
\]

Bayesian strategy prescribes to choose the \(\ell\)-th action if \(r^{(\ell)}(X_1, n_1, X_2, n_2)\) has smaller value (\(\ell = 1, 2\)). In case of a draw \(r^{(1)}(X_1, n_1, X_2, n_2) = r^{(2)}(X_1, n_1, X_2, n_2)\) the choice can be arbitrary.

Bayesian risk (2) is calculated by the formula
\[
R_N^B(\{p_i, \Delta p_i, \lambda_i\}, \{\beta_j, N_j\}) = r(0,0,0,0).
\]
Formulas (9)–(12) follow from (5)–(8).

3. Determination of the regret

Let’s define a regret calculated with respect to prior distribution \(\lambda_1 = \lambda_2 = 0.5\) on the set \(\{\theta_1, \theta_2\}\), where \(\theta_1 = (p + \Delta p, p - \Delta p)\), \(\theta_2 = (p - \Delta p, p + \Delta p)\), and strategy \(\sigma\) as follows
\[
L(\sigma, \{\theta_i, \lambda_i\}, N) = \sum_{i=1}^{2} L(\sigma, \theta_i, N) \lambda_i.
\]
One can use the following recursive equation to determine \(L(\sigma, \{\theta_i, \lambda_i\}, N)\):
\[
l(X_1, n_1, X_2, n_2) = \sigma_1(X_1, m_1, X_2, n_2) l^{(1)}(X_1, m_1, X_2, n_2),
\]
\[
+ \sigma_2(X_1, m_1, X_2, n_2) l^{(2)}(X_1, m_1, X_2, n_2),
\]
where \(l^{(1)}(X_1, n_1, X_2, n_2) = l^{(2)}(X_1, n_1, X_2, n_2) = 0\) if \(n_1 + n_2 = N\) and then
\[
l^{(1)}(X_1, n_1, X_2, n_2) = g^{(1)}(X_1, n_1, X_2, n_2) + \frac{1}{x=0} l(X_1 + x, n_1 + 1, X_2, n_2) \times h_1(X_1, n_1),
\]
and
\[
l^{(2)}(X_1, n_1, X_2, n_2) = g^{(2)}(X_1, n_1, X_2, n_2) + \frac{1}{x=0} l(X_1, n_1, X_2 + x + n_2 + 1) \times h_1(X_2, n_2),
\]
where
\[
g^{(1)}(X_1, n_1, X_2, n_2) = 2\gamma_n \times \Delta p \times B(X_1, n_1; p - \Delta p) B(X_2, n_2; p + \Delta p) \lambda_1,
\]
\[
g^{(2)}(X_1, n_1, X_2, n_2) = 2\gamma_n \times \Delta p \times B(X_1, n_1; p + \Delta p) B(X_2, n_2; p - \Delta p) \lambda_2.
\]
A regret (13) is calculated by the formula
\[
L(\sigma, \{\theta_i, \lambda_i\}, N) = l(0,0,0,0).
\]
A regret \(L_N(\sigma, \theta, \{\beta_j, N_j\})\) can be then determined by formula (1).

4. Numerical experiments

In this section, we present the results of numerical experiments. The values of control horizons were chosen as \(N_1 = 20\), \(N_2 = 30\), \(N_3 = N = 40\).

| Table 1. Normalized expected losses. |
|---|---|---|---|---|---|
| \(i\) | 1 | 2 | 3 | 4 | 5 |
| \(p_i\) | 0.222 | 0.356 | 0.444 | 0.541 | 0.679 |
The weight factors were defined according to the formula
\[ \beta_i = \beta^{-1} N_i^{-1/2}, \quad i = 1, 2, 3, \quad \beta = N_1^{-1/2} + N_2^{-1/2} + N_3^{-1/2}. \]

The worst-case prior distribution was approximated by the set of $2N$ parameters with $K = 5$ and is presented in table 1.

| $\Delta p_i$ | 0.132 | 0.125 | 0.122 | 0.120 | 0.117 |
|--------------|-------|-------|-------|-------|-------|
| $\lambda_i$  | 0.083 | 0.130 | 0.119 | 0.102 | 0.065 |

**Figure 1.** Normalized regrets $l_{40}(p, \Delta p, \{\beta_j, N_j\})$.

Corresponding normalized minimax risk was calculated by (9)–(12) as
\[ r_{40}^M(\Theta, \{\beta_j, N_j\}) = N^{-1/2} R_{40}^M(\Theta, \{\beta_j, N_j\}) \approx 0.569. \]

Then for determined strategy $\sigma^M$ maximum normalized regret was calculated by (14)–(17) as
\[ \max_{\Theta} l_{40}(\sigma^M, \theta, \{\beta_j, N_j\}) = \max_{\Theta} N^{-1/2} L_{40}(\sigma^M, \theta, \{\beta_j, N_j\}) \approx 0.580. \]

One can see that these values are close to each other. On figure 1 normalized regrets are presented as functions of $\Delta p$ for some fixed $p$. A regret corresponding to $p = 0.2$ is presented by blue curve, corresponding to $p = 0.35$ by magenta curve, corresponding to $p = 0.5$ by green curve, corresponding to $p = 0.65$ by cyan curve and corresponding to $p = 0.8$ by brown curve.

**Figure 2.** Normalized regrets $l_{20}(p, \Delta p, \theta)$.
Next, normalized regrets and their maximum values were calculated for determined strategy \( \sigma^M \) and control horizons \( N_1 = 20, \ N_2 = 30, \ N_3 = 40 \) according to the formula

\[
I_{N_i}(\sigma^M, \theta) = N_i^{-1/2} L_{N_i}(\sigma^M, \theta).
\]

They were determined as

\[
\max_{\theta} I_{40}(\sigma^M, \theta) \approx 0.651, \quad \max_{\theta} I_{30}(\sigma^M, \theta) \approx 0.583, \quad \max_{\theta} I_{20}(\sigma^M, \theta) \approx 0.603.
\]

These normalized regrets as functions of \( \Delta p \) for different \( p \) and \( N_1 = 20, \ N_2 = 30, \ N_3 = 40 \) are presented on figure 2–figure 4. One can see that these approximations are not so good as in case of average regret \( I_{40}(\sigma^M, \theta, \{\beta_j, N_j\}) \). Colors of lines correspond to \( \{p_i\} \) as described above.

5. Conclusion

Considered approach provides the method of storage the strategy. Instead of probabilities which describe the strategy one can store the worst-case prior distributions corresponding to the sets of control horizons. Computing of the strategy is very fast. Once the batch of data arises, computer determines the strategy and then it is applied for the control.
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 (London: Chapman & Hall)
[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] Auer P 2002 Using Confidence Bounds for Exploitation-Exploration Trade-offs Journal of Machine Learning Research 3 397–422
[5] Vogel W 1960 An asymptotic minimax theorem for the two-armed bandit problem Ann. Math. Stat. 31 444–451
[6] Perchet V, Rigollet P, Chassang S and Snowberg E 2016 Batched Bandit Problems Annals of Statistics 44 660–681
[7] Kolnogorov A V 2012 Two-armed bandit problem for parallel data processing systems Problems of Information Transmission 48 72–84
[8] Kolnogorov A V 2018 Gaussian two-armed bandit and optimization of batch data processing Problems of Information Transmission 54 84–100
[9] Kolnogorov A V 2014 Finding minimax strategy and minimax risk for Bernoulli multi-armed bandit Proceedings of the 2014 International Conference on Mathematical Models and Methods in Applied Sciences (MMAS ‘14) (Saint Petersburg, Russia, September 23–25) 59–66