MSC 62C10, 62L05, 91A35 DOI: 10.14529/mmp250103

A LIMITING DESCRIPTION IN A GAUSSIAN ONE-ARMED BANDIT PROBLEM WITH BOTH UNKNOWN PARAMETERS

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

We consider the limiting description of control in a Gaussian one-armed bandit problem, which is a mathematical model for optimizing batch processing of big data in the presence of two alternative methods with known efficiency of the first method. We establish that this description is given by a second-order partial differential equation in which the variance of one-step income is known. This means that in the case of big data, the variance can be arbitrarily accurate estimated at a short initial stage of processing, and then the obtained estimate is used by the control strategy.

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

Introduction

The article continues the work [1], which provides an overview of other approaches to the problem and a bibliography; we note here [2, 3]. Let's briefly describe the results of [1]. The problem of optimizing batch processing of big data is considered in the presence of two alternative processing methods with different efficiencies and the efficiency of the first method is known. Processing of each data unit is accompanied by income equal to one in the case of successful processing and zero otherwise. It is required to organize the processing in such a way as to maximize the mathematical expectation of the total number of successfully processed data (total income). Batch processing means that data is divided into batches, the same processing methods (hereinafter referred to as actions) are applied to all data in each batch, and the cumulative incomes in the batches are used for control. If the batch sizes are large enough, then by virtue of the central limit theorem, the incomes in them have approximately normal (Gaussian) distributions. Therefore, mathematically this problem is described as the problem of a Gaussian one-armed bandit, i.e., a two-armed bandit with known income characteristics for choosing the first action.

Formally, a Gaussian one-armed bandit is a controlled random process ξ_n, $n = 1, 2, \ldots, N$, which values are interpreted as incomes, depend only on the current chosen actions y_n, of which there are two ($y_n \in \{1, 2\}$), and have normal distributions. In the case of choosing the second action, one-step income has the distribution density $f_D(x|m) = (2\pi D)^{1/2} \exp\left(-(x-m)^2/(2D)\right)$, where $m = \mathbf{E}(\xi_n|y_n = 2)$, $D = \mathbf{D}(\xi_n|y_n = 2)$ are the mathematical expectation and the variance of one-step income. In the case of choosing the first action, mathematical expectation m_1 is assumed to be known and, without loss of generality, $m_1 = 0$ (otherwise, one can consider the process $\xi_n - m_1$). The value of the variance $D_1 = \mathbf{D}(\xi_n|y_n = 1)$ is insignificant because it does not affect the goal of the control. So, one-armed bandit is described by the parameter $\theta = (m, D)$, which value is assumed to be unknown. At the same time, the set of admissible parameters Θ is known and has the form $\Theta = \{(m, D) : |m| \le C, \underline{D} \le D \le \overline{D}\}$, where $0 < C < \infty$, $0 < \underline{D} < \overline{D} < \infty$.

A control strategy σ determines the choice of an action y_{n+1} to process the batch with the number $n+1$ depending on the current values of cumulative income X and s^2-statistics S for choosing the second action, which are sufficient statistics. Statistics based on the results of data processing by the first method are not used because the corresponding distribution is known. To state the goal of the control, let's define a regret

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

which characterizes the mathematical expectation of loss of total income due to incomplete information. Here $\mathbf{E}_{\sigma,\theta}$ is the sign of the mathematical expectation with respect to the measure generated by the strategy σ and the parameter θ. An important feature of batch processing is that it virtually does not lead to an increase in the maximum regret value if the amount of data being processed and the batch sizes are large enough [1,4].

Let's assign a prior distribution density $\lambda(\theta) = \lambda(m,D)$ on the set Θ and define a Bayesian risk

$$R_N^B(\lambda) = \inf_{\{\sigma\}} \int\limits_{\Theta} L_N(\sigma,\theta)\lambda(\theta)d\theta, \tag{1}$$

the corresponding optimal strategy σ^B is called a Bayesian strategy.

In [1], recursive integro-difference equations were obtained to describe optimal control. In this paper, we obtain a limiting description of the control by a second-order partial differential equation. It turned out that in this equation the normalized value of s^2-statistics plays the role of a constant parameter and the equation itself is equivalent to a differential equation that can be obtained in the case of a priori known variance. This means that, when processing big data, the variance of one-step income can be arbitrarily accurate estimated at a short initial stage, and then the obtained estimate is used for control. For comparison, we indicate the articles [5,6] in which limiting descriptions of the control by differential equations are also obtained and which also do not contain dependence on the estimate of variance.

The rest of the article is as follows. In section 1, recursive equations for calculating Bayesian risk are obtained, which are equivalent to those obtained in [1] but more convenient for performing the passing to the limit. In section 2, a limiting description of the recursive equation is given by a second-order partial differential equation if the number of batches grows infinitely. Section 3 contains the conclusion.

1. Recursive Equations for Finding Bayesian Risk

Consider batch processing. Let the total number of data be $N = MK$, where M is the batch size and K is the number of batches. The same action is applied to all the data in the batch, resulting in income $x_k = \sum_{n=(k-1)M+1}^{kM} \xi_n$. In the case of the second action, the mathematical expectation and the variance of income in the batch are equal to Mm and MD. The mathematical expectation of income for the application of the first action is still zero.

We will use the following intuitively clear property of the optimal strategy which was first established in [7] and has already been used in [1]. Since the application of the first action does not provide additional information (the corresponding distribution is known), then being selected once, it will be applied until the end of the control. Thus, the second action is always applied at the initial stage. We will assume that the duration of the initial stage is at least $k_0 \geq 2$. Note that if $k_0 \ll K$ then the corresponding Bayesian strategy is almost optimal.

Let the second action be applied to $k \geq k_0$ batches. We calculate the values

$$X = \sum_{i=1}^{k} x_i, \quad S = \sum_{i=1}^{k} x_i^2 - X^2/k,$$

which in this case are sufficient statistics and characterize the current cumulative income and s^2-statistics for the application of the second action. Let's find how to update X and S. Assume that $k \geq k_0$ and let $x_{k+1} = Y$ be a new income. Then $X_{new} = \sum_{i=1}^{k+1} x_i = X + Y$, $S_{new} = \left(\sum_{i=1}^{k+1} x_i^2\right) - X_{new}^2/(k+1) = \left(\sum_{i=1}^{k} x_i^2\right) + Y^2 - (X+Y)^2/(k+1) = S + \Delta(X,k,Y)$, where

$$\Delta(X,k,Y) = X^2/k + Y^2 - (X+Y)^2/(k+1) = \frac{(X-kY)^2}{k(k+1)}. \tag{2}$$

Thus, X, S are updated according to formulas

$$X \leftarrow X + Y, \quad S \leftarrow S + \Delta(X,k,Y), \tag{3}$$

where $\Delta(X,k,Y)$ is determined in (2). Given a prior distribution density $\lambda(m.D)$, let's describe the posterior one. Denote by $\chi_k^2(x) = \left(2^{\frac{k}{2}}\Gamma(k/2)\right)^{-1} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$, $k \geq 1$, the chi-squared distribution density with k degrees of freedom and consider the functions

$$\begin{gathered} f_{kMD}\left(X|kMm\right) = (2\pi kMD)^{-1/2} \exp\left(-(X-kMm)^2/(2kMD)\right), \\ \psi_{k-1}\left((MD)^{-1}S\right) = (MD)^{-1}\chi_{k-1}^2((MD)^{-1}S). \end{gathered} \tag{4}$$

Note that defined above cumulative income X and s^2-statistics S after processing k batches have exactly the distribution densities described by (4). Since X and S are independent random variables, the posterior distribution density is

$$\lambda(m,D|X,S,k) = \frac{f_{kMD}\left(X|kMm\right)\psi_{k-1}\left((MD)^{-1}S\right)\lambda(m.D)}{P(X,S,k)}$$
$$\text{with} \quad P(X,S,k) = \iint_{\Theta} f_{kMD}\left(X|kMm\right)\psi_{k-1}\left((MD)^{-1}S\right)\lambda(m.D)dmdD,$$

if $k \geq k_0$. However, recursive equation is simpler if the posterior distribution is defined in the following equivalent way. Denote

$$\tilde{\mathbf{F}}(X,S,k|m,D) = D^{-3/2}\tilde{f}_{kMD}(X|kMm)\tilde{\psi}_{k-1}\left(S/(MD)\right), \tag{5}$$

with

$$\begin{aligned} &\tilde{f}_D\left(x|m\right) = \exp\left(-(x-m)^2/(2D)\right), \\ &\tilde{\psi}_{k-1}\left(s\right) = (k/(4\pi))^{1/2}\left(s/k\right)^{\frac{k-3}{2}} \exp\left(-(s-k)/2\right). \end{aligned} \tag{6}$$

Clearly, for $k \geq k_0$ the posterior distribution density is

$$\begin{gathered}\lambda(m, D|X, S, k) = \frac{\tilde{\mathbf{F}}(X, S, k|m, D)\lambda(m, D)}{\tilde{P}(X, S, k)}, \\ \text{with} \quad \tilde{P}(X, S, k) = \iint_{\Theta} \tilde{\mathbf{F}}(X, S, k|m, D)\lambda(m, D)dmdD.\end{gathered} \tag{7}$$

Let $R^B(X, S, k)$ denote a Bayesian risk on the remaining control horizon $k+1, \dots, K$, computed with respect to the posterior distribution $\lambda(m, D|X, S, k)$, i.e., $R^B(X, S, k) = R^B_{K-k}(\lambda(m, D|X, S, k))$. Denote $m^+ = \max(m, 0)$, $m^- = \max(-m, 0)$. Taking into account mentioned above property of the optimal strategy, a standard recursive equation for computing Bayesian risk is

$$R^B(X, S, k) = \min\left(R^B_1(X, S, k), R^B_2(X, S, k)\right), \tag{8}$$

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

$$\begin{gathered}R^B_1(X, S, k) = (K-k)\iint_{\Theta} Mm^+ \lambda(m, D|X, S, k)dmdD, \\ R^B_2(X, S, k) = \iint_{\Theta} \lambda(m, D|X, S, k)\times \\ \times\left(Mm^- + \int\limits_{-\infty}^{\infty} R^B(X+Y, S+\Delta(X, k, Y), k+1) f_{MD}(Y|Mm)dY\right)dmdD,\end{gathered} \tag{9}$$

if $k_0 \leq k \leq K-1$. In the second equation (9), we used (2) – (3). Here $R^B_\ell(X, S, k)$ is equal to the loss of cumulative expected income at the remaining control horizon $k+1, \dots, K$ if at first the ℓth action was chosen and then the control was optimally performed. Bayesian strategy prescribes, when processing the batch with the number $k+1$, to choose an action corresponding to the smaller of the current values $R^B_1(X, S, k)$, $R^B_2(X, S, k)$. 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. Bayesian risk (1) is

$$R^B_N(\lambda) = k_0 \iint_{\Theta} Mm^- \lambda(m, D)dmdD + \int\limits_0^{\infty}\int\limits_{-\infty}^{\infty} R^B(X, S, k_0)P(X, S, k_0)dXdS. \tag{10}$$

Let's present another form of (8) – (10) which is more convenient for computations. We put $R_\ell(X, S, k) = R^B_\ell(X, S, k) \times \tilde{P}(X, S, k)$, $\ell = 1, 2$, where $\tilde{P}(X, S, k)$ is defined in (7).

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

$$R(X, S, k) = \min\left(R_1(X, S, k), R_2(X, S, k)\right), \tag{11}$$

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

$$R_1(X, S, k) = M(K-k)G_1(X, S, k), \tag{12}$$

$$R_2(X, S, k) = MG_2(X, S, k) + \int\limits_{-\infty}^{\infty} R(X+Y, S+\Delta(X, k, Y), k+1)H(X, S, k, Y)dY,$$

if $k_0 \le k \le K-1$. *Here* $\Delta(X,k,Y)$ *is given by* (2),

$$G_1(X,S,k) = \iint_\Theta m^+ \tilde{\mathbf{F}}(X,S,k|m,D)\lambda(m,D)dmdD,$$
$$G_2(X,S,k) = \iint_\Theta m^- \tilde{\mathbf{F}}(X,S,k|m,D)\lambda(m,D)dmdD, \tag{13}$$

where $\tilde{\mathbf{F}}(X,S,k|m,D)$ *is given by* (5) *and*

$$H(X,S,k,Y) = \left(\frac{k}{2\pi eS}\right)^{1/2}\left(\frac{k+1}{k}\right)^{\frac{k-3}{2}}\left(\frac{S}{S+\Delta(X,k,Y)}\right)^{\frac{k-2}{2}}. \tag{14}$$

When processing the batch number $k+1$, *Bayesian strategy prescribes to choose the action corresponding to the smaller value of* $R_1(X,S,k)$, $R_2(X,S,k)$; *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. Bayesian risk* (1) *is*

$$R_N(\lambda) = k_0\iint_\Theta Mm^-\lambda(m,D)dmdD + H_0\int\limits_0^\infty\int\limits_{-\infty}^\infty R(X,S,k_0)dXdS, \tag{15}$$

where

$$H_0 = \frac{(k_0/2)^{(k_0-1)/2}}{k_0^2(M^3/2)^{1/2}\Gamma((k_0-1)/2)\exp(k_0/2)}. \tag{16}$$

Proof. Is done similarly to the proof presented in [1]. Formulas (14) and (15) for $H(X,S,k,Y)$ and H_0 are obtained after converting the expressions

$$H(X,S,k,Y) = \frac{\tilde{\mathbf{F}}(X,S,k|m,D)f_{MD}(Y|Mm)}{\tilde{\mathbf{F}}(X+Y,S+\Delta,k+1|m,D)},$$
$$H_0 = \frac{P(X,S.k_0)}{\tilde{P}(X,S.k_0)} = \frac{f_{k_0MD}(Y|k_0Mm)\psi_{k_0-1}((MD)^{-1}S)}{D^{-3/2}\tilde{f}_{k_0MD}(Y|k_0Mm)\tilde{\psi}_{k_0-1}((MD)^{-1}S)}.$$

□

Let's give an invariant form of formulas (12) – (16) with control horizon equal to one. We choose the following set of parameters $\Theta_N = \{(m,D) : \underline{D} \le D \le \overline{D}, |m| \le c(\overline{D}/N)^{1/2}\}$, where $c > 0$, $0 < \underline{D} \le D \le \overline{D} < \infty$. If we put $D = \beta\overline{D}$, $m = \alpha(\overline{D}/N)^{1/2}$, then it takes the form $\Theta_N = \{(\alpha,\beta) : \underline{D}/\overline{D} = \beta_0 \le \beta \le 1, |\alpha| \le c\}$.

Consider the change of variables $X = x(\overline{D}N)^{1/2}$, $Y = y(\overline{D}N)^{1/2}$, $S = skM\overline{D}$, $k = tK$, $k_0 = t_0K$, $M/N = K^{-1} = \varepsilon$, $\lambda(m,D) = (N/\overline{D}^3)^{1/2}\varrho(\alpha,\beta)$. Let $R_\ell(X,S,k) = (\overline{D}\,N)^{1/2}(\overline{D})^{-3/2}r_\ell(x,s,t)$, $\ell = 1,2$. The following theorem is valid.

Theorem 2. *To find the Bayesian risk, one should solve the recursive equation*

$$r(x,s,t) = \min\left(r_1(x,s,t), r_2(x,s,t)\right), \tag{17}$$

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

$$\begin{gathered} r_1(x,s,t)=(1-t)g_1(x,s,t),\\ r_2(x,s,t)=\varepsilon g_2(x,s,t)+\int\limits_{-\infty}^{\infty} r\Big(x+y,\frac{s+t^{-1}\delta(x,t,y)}{1+\varepsilon/t},t+\varepsilon\Big)h(x,s,t,y)dy, \end{gathered} \tag{18}$$

if $t_0\le t\le 1-\varepsilon$. *Here*

$$\begin{aligned} g_1(x,s,t)&=\iint_{\Theta_N}\alpha^{+}\beta^{-3/2}\tilde f_{t\beta}(x|t\alpha)\tilde\psi_{k-1}(ks/\beta)\varrho(\alpha,\beta)d\alpha d\beta,\\ g_2(x,s,t)&=\iint_{\Theta_N}\alpha^{-}\beta^{-3/2}\tilde f_{t\beta}(x|t\alpha)\tilde\psi_{k-1}(ks/\beta)\varrho(\alpha,\beta)d\alpha d\beta, \end{aligned} \tag{19}$$

where $\tilde f_{t\beta}(x|t\alpha)$, $\tilde\psi_{k-1}(ks/\beta)$ *are given by* (6),

$$h(x,s,t,y)=\left(\frac{1}{2\pi e\varepsilon s}\right)^{1/2}\left(1+\frac{\varepsilon}{t}\right)^{\frac{k-3}{2}}\left(\frac{s}{s+t^{-1}\delta(x,t,y)}\right)^{\frac{k-2}{2}} \tag{20}$$

and

$$t^{-1}\delta(x,t,y)=\frac{(\varepsilon x-ty)^2}{t^2(t+\varepsilon)}. \tag{21}$$

At the point of time $t+\varepsilon$ *(or, equivalently, when processing* $(k+1)$*th batch) Bayesian strategy prescribes to choose the action corresponding to the smaller value of* $r_1(x,s,t)$, $r_2(x,s,t)$*; in the case of a draw the choice can be arbitrary. Once the first action is chosen, it will be applied until the end of the control. Bayesian risk* (1) *is*

$$R_N(\lambda)=(\overline{D}N)^{1/2}\left(t_0\iint_{\Theta_N}\alpha^{-}\varrho(\alpha,\beta)d\alpha d\beta+h_0\int\limits_{0}^{\infty}\int\limits_{-\infty}^{\infty} r(x,s,t_0)dxds\right) \tag{22}$$

with

$$h_0=\frac{(k_0/2)^{(k_0-1)/2}}{k_0(\varepsilon/2)^{1/2}\Gamma((k_0-1)/2)\exp(k_0/2)}. \tag{23}$$

This description of control on the horizon equal to one is invariant in the sense that it does not depend on the total amount of data N *but only on the number of batches* K.

The proof is similar to that given in [1] and is, therefore, omitted.

Remark 1. The recursive equations (11), (12) and (17), (18) are equivalent to those obtained in [1]. Some differences in the form of the equations are due to a different choice of the function $\tilde\psi_{k-1}(s)$ as well as another change of variables for S (more convenient for the passing to the limit).

2. Passing to the Limit. Differential Equation

Let us now consider the passing to the limit in the equation (17), (18) as $\varepsilon \to 0$. We need the following auxiliary results.

Lemma 1. *The asymptotic (as $\kappa \to \infty$) estimate is valid*

$$I_\kappa = \int\limits_{-\infty}^{\infty} \frac{dx}{(1+x^2)^\kappa} = \left(\frac{\pi}{\kappa}\right)^{1/2} \left(1 + \frac{3}{8\kappa} + o(\kappa^{-1})\right). \tag{24}$$

Proof. Performing the change of variables $x = \kappa^{-1/2} y$ in the integral in (24) we obtain that $I_\kappa = \kappa^{-1/2} \int\limits_{-\infty}^{\infty} (1+\kappa^{-1}y^2)^{-\kappa} dy$ with $\int\limits_{-\ln(\kappa)}^{\ln(\kappa)} (1+\kappa^{-1}y^2)^{-\kappa} dy = \int\limits_{-\ln(\kappa)}^{\ln(\kappa)} e^{-y^2} dy + o(1) = \pi^{1/2} + o(1)$ and $\int\limits_{\ln(\kappa)}^{\infty} (1+\kappa^{-1}y^2)^{-\kappa} dy = \int\limits_{-\infty}^{-\ln(\kappa)} (1+\kappa^{-1}y^2)^{-\kappa} dy \le \int\limits_{\ln(\kappa)}^{\infty} (1+y^2)^{-1} dy = o(1)$, where $o(1) \to 0$ as $\kappa \to \infty$. Therefore, $\lim_{\kappa\to\infty} I_\kappa \times (\pi/\kappa)^{-1/2} = 1$.

Next, let's obtain a recursive formula for computing I_κ. We have $I_\kappa = \int\limits_{-\infty}^{\infty} x'(1+x^2)^{-\kappa} dx = x(1+x^2)^{-\kappa}|_{-\infty}^{\infty} + 2\kappa \int\limits_{-\infty}^{\infty} (x^2 \pm 1)(1+x^2)^{-(\kappa+1)} dx = 2\kappa(I_\kappa - I_{\kappa+1})$, whence $I_{\kappa+1} = I_\kappa \times (2\kappa - 1)/(2\kappa)$. Denote $J_\kappa = \kappa^{1/2} I_\kappa$, $J = \lim_{n\to\infty} J_{\kappa+n}$. Using recursive formula, we have $(\kappa+1)^{-1/2} J_{\kappa+1} = \kappa^{-1/2} J_\kappa \times (2\kappa-1)/(2\kappa)$, whence $J_{\kappa+1} = J_\kappa \times (1+\kappa^{-1})^{1/2} (1 - (2\kappa)^{-1})$.

Let's put $\gamma_\kappa = \ln J_\kappa$. Then

$$\gamma_{\kappa+1} = \gamma_\kappa + \frac{1}{2}\left(\frac{1}{\kappa} - \frac{1}{2\kappa^2}\right) + \left(-\frac{1}{2\kappa} - \frac{1}{8\kappa^2}\right) + o(k^{-2}) = \gamma_\kappa - \frac{3}{8\kappa^2} + o(k^{-2}).$$

Denote $\gamma = \lim_{n\to\infty} \gamma_{\kappa+n}$, so that $J = e^\gamma = \pi^{1/2}$. For κ large enough we have

$$\gamma = \gamma_\kappa - \sum_{n=0}^{\infty} \frac{3}{8(\kappa+n)^2} + o(\kappa^{-1}) = \gamma_\kappa - \frac{3}{8\kappa} + o(\kappa^{-1}),$$

and, hence, $J_\kappa = J \times (1 + 3/(8\kappa) + o(\kappa^{-1}))$. Taking into account the definition of J, J_κ, we obtain (24).

□

Lemma 2. *For $\kappa \ge 2$ the equality holds*

$$I_\kappa^D = \int\limits_{-\infty}^{\infty} \frac{x^2 dx}{(1+x^2)^\kappa} = \int\limits_{-\infty}^{\infty} \frac{(x^2 \pm 1) dx}{(1+x^2)^\kappa} = I_{\kappa-1} - I_\kappa = \frac{I_\kappa}{2\kappa - 3}. \tag{25}$$

Proof. It is checked using the recursive formula $I_{\kappa+1} = I_\kappa \times (2\kappa-1)/(2\kappa)$.

□

Lemma 3. *For a factor h_0 in (22) with $k_0 = t_0 K$, $t_0 > 0$, the asymptotic (as $K \to \infty$) estimate is valid*

$$h_0 = (2\pi t_0)^{-1/2} (1 + o(1)). \tag{26}$$

Proof. To approximate $\Gamma((k_0-1)/2)$ in (23), we use the Stirling's formula $\Gamma(\kappa+1) \sim (2\pi)^{1/2}\kappa^{\kappa+1/2}e^{-\kappa}$. Then h_0 in (23) is approximated as $(k_0\pi\varepsilon)^{-1/2} \times (k_0/(k_0-3))^{\frac{k_0-1}{2}} \left((k_0-3)/(2k_0)\right)^{1/2} \exp\left(-3/2\right)$. From here (26) follows. □

Lemma 4. *Let the density $\varrho(\alpha,\beta)$ be a continuous function of α,β. If $t \ge t_0 > 0$ then the limiting (as $K \to \infty$) formulas are valid*

$$\begin{aligned} g_1(x,s,t) &= \mathrm{I}\left(s,(\beta_0,1)\right) \times \int\limits_{-c}^{c} \alpha^{+} s^{-1/2} \tilde{f}_{ts}(x|t\alpha)\varrho(\alpha,s)d\alpha, \\ g_2(x,s,t) &= \mathrm{I}\left(s,(\beta_0,1)\right) \times \int\limits_{-c}^{c} \alpha^{-} s^{-1/2} \tilde{f}_{ts}(x|t\alpha)\varrho(\alpha,s)d\alpha, \end{aligned} \tag{27}$$

where the indicator $\mathrm{I}\left(s,(\beta_0,1)\right)=1$ *if* $s \in (\beta_0,1)$ *and* $\mathrm{I}\left(s,(\beta_0,1)\right)=0$ *if* $s \notin [\beta_0,1]$.

Proof. Consider a function $\tilde{\psi}_{k-1}\left(ks/\beta\right) = \left(k/(4\pi)\right)^{1/2}\left(s/\beta\right)^{\frac{k-3}{2}}\exp\left(-k(s/\beta-1)/2\right)$ from (19). Let's put $S = kMD + (kM)^{1/2}D\Delta S$. Taking into account the change of variables $S = kM\overline{D}s$, the equality $kM\overline{D}s = kM\overline{D}\beta + (kM)^{1/2}\beta\overline{D}\Delta S$ holds, whence $\Delta S = (kM)^{1/2}(s/\beta-1)$. Therefore, $\tilde{\psi}_{k-1}\left(ks/\beta\right) = \left(k/(4\pi)\right)^{1/2}\left(1+\Delta S(kM)^{-1/2}\right)^{\frac{k-3}{2}}\exp\left(-\Delta S(kM)^{1/2}/(2M)\right)$. Next, we have the estimate $\ln\left((4\pi/k)^{1/2}\tilde{\psi}_{k-1}\left(ks/\beta\right)\right) = -\Delta S^2/(4M)+o(1) = -t(s/\beta-1)^2/(4\varepsilon)+o(1)$. Therefore, $\beta^{-1}\tilde{\psi}_{k-1}(ks/\beta) = \left(t/(4\pi\varepsilon\beta^2)\right)^{1/2}\exp\left(-t(s-\beta)^2/(4\varepsilon\beta^2)\right)(1+o(1))$. This function converges to the Dirac delta function $\delta^D(s-\beta)$ as $\varepsilon \to 0$. This means that for any continuous function $g(\beta)$ the equalities hold: $\int\limits_{\beta_0}^{1} g(\beta)\delta^D(s-\beta)d\beta = g(s)$ if $s \in (\beta_0,1)$ and $\int\limits_{\beta_0}^{1} g(\beta)\delta^D(s-\beta)d\beta = 0$ if $s \notin [\beta_0,1]$. Taking into account (19), we obtain (27). □

Let's obtain a limiting description of recursive equation (17), (18) as $\varepsilon \to 0$. We introduce a variable z by condition $sz^2 = t^{-1}\delta(x,t,y)$, where $t^{-1}\delta(x,t,y)$ is defined in (21). Then $y = \varepsilon x t^{-1} + (s(t+\varepsilon))^{1/2}z$, $dy = (s(t+\varepsilon))^{1/2}dz$ and the second equation in (18) takes the form

$$\begin{aligned} r_2(x,s,t) = \varepsilon g_2(x,s,t) + \left(\frac{1}{2\pi e\varepsilon}\right)^{1/2}\left(1+\frac{\varepsilon}{t}\right)^{\frac{k-3}{2}}(t+\varepsilon)^{1/2}\times \\ \times \int\limits_{-\infty}^{\infty} r(x+y, \frac{s(1+z^2)}{1+\varepsilon/t}, t+\varepsilon) \times \left(\frac{1}{1+z^2}\right)^{\frac{k-2}{2}} dz, \end{aligned} \tag{28}$$

where $y = \varepsilon x t^{-1} + (s(t+\varepsilon))^{1/2}z$. Denote $r = r(x,s,t+\varepsilon)$. Let' assume that $r(x,s,t+\varepsilon)$ has partial derivatives of the required orders by x, s. Presenting $r(x+y, s(1+z^2)/(1+\varepsilon t^{-1}), t+\varepsilon)$ as a Taylor's series and taking into account that $s(1+z^2)/(1+\varepsilon t^{-1}) = s(1+z^2) - s(1+z^2)\varepsilon t^{-1} + o(\varepsilon)$, we obtain

$$\begin{aligned} r + r'_x \times (x\varepsilon t^{-1} + (s(t+\varepsilon))^{1/2}z) + 0{,}5r''_{xx} \times (x\varepsilon t^{-1} + (s(t+\varepsilon))^{1/2}z)^2 + \\ + r'_s \times (-s(1+z^2)\varepsilon t^{-1} + sz^2) + A(\varepsilon,z) = \\ = r + r'_x \times x\varepsilon t^{-1} + 0{,}5r''_{xx} \times s(t+\varepsilon)z^2 + r'_s \times (-s\varepsilon t^{-1} + sz^2) + A(\varepsilon,z). \end{aligned} \tag{29}$$

Here r'_x, r''_{xx}, r'_s are calculated at the point $(x,s,t+\varepsilon)$ and additional term $A(\varepsilon,z)$ contains a value of the order of $o(\varepsilon)$ and terms of the form z, εz^2, z^i $(i \geq 3)$ (these terms becomes equal to zero or will have the order of $o(\varepsilon)$ after integration). Substituting (29) into the integral in (28), taking into account (25), we obtain that integral in (28) is $(r + r'_x \times x\varepsilon/t - r'_s \times s\varepsilon/t + o(\varepsilon))\, I_{k/2-1} + (0{,}5r''_{xx} \times s(t+\varepsilon) + r'_s \times s + o(\varepsilon))\, I^D_{k/2-1} = (r + \varepsilon\,(r'_x\; xt^{-1} + 0{,}5r''_{xx}s) + o(\varepsilon))\, I_{k/2-1}$. Taking into account the factors, the second term in (28) is $F \times (r + \varepsilon(r'_x\; xt^{-1} + 0{,}5r''_{xx}s))\,(1+o(\varepsilon))$, where $F = (2\pi e\varepsilon)^{-1/2}\,(1+\varepsilon/t)^{\frac{k-3}{2}}\,(t+\varepsilon)^{1/2} I_{k/2-1}$ and $I_{k/2-1} = (\pi/(k/2-1))^{1/2}(1+3/(4k-8)) + (k^{-1})$ according to (24). It is straightforward to verify that $F = 1 + 1/(2k) + o(k^{-1}) = 1 + \varepsilon/(2t) + o(\varepsilon)$.

Let's obtain the differential equation. From (28), taking into account the transformation of the second term on the right hand side of (28), and from the first equation of (18) we have

$$\begin{gathered} r_2(\cdot,t) - r(\cdot,t) = \varepsilon g_2(\cdot,t) + (1+\varepsilon/(2t))\, r(\cdot,t+\varepsilon) - r(\cdot,t) + \\ +\varepsilon\,(r'_x(\cdot,t+\varepsilon)\, x/t + 0{,}5 s r''_{xx}(\cdot,t+\varepsilon)) + o(\varepsilon), \\ r_1(\cdot,t) - r(\cdot,t) = (1-t) g_1(\cdot,t) - r(\cdot,t). \end{gathered} \tag{30}$$

Complementing (30) with equation (17), which is written in equivalent form $\min(r_1(\cdot,t) - r(\cdot,t), \varepsilon^{-1}\,(r_2(\cdot,t) - r(\cdot,t))) = 0$, we get in the limit as $\varepsilon \to 0$ the equation

$$\min\left((1-t)g_1 - r, r'_t + r/(2t) + r'_x \times (x/t) + 0{,}5 s r''_{xx} + g_2\right) = 0, \tag{31}$$

with initial condition $r(x,s,1) = 0$. Here $g_1, g_2, r, r'_t, r'_x, r''_{xx}$ are functions of x, s, t. Bayesian strategy prescribes to choose the action corresponding to the smaller term on the left hand side of (31); in the case of a draw the choice can be arbitrary. Once the first action is chosen, it will be applied until the end of the control. Here g_1, g_2 are given by (27) and the Bayesian risk (1) asymptotically is equal to

$$\lim_{K\to\infty} (\overline{D}N)^{-1/2} R_N(\lambda) = t_0 \iint_{\Theta_N} \alpha^- \varrho(\alpha,\beta)\, d\alpha d\beta + \frac{1}{(2\pi t_0)^{1/2}} \int\limits_{\beta_0}^{1} \int\limits_{-\infty}^{\infty} r(x,s,t_0)\, dx ds. \tag{32}$$

Remark 2. In the case of a priori known variance β, the corresponding differential equation for $r(x,t)$ has the form $\min((1-t)g_1 - r, r'_t + r/(2t) + r'_x \times (x/t) + 0{,}5\beta r''_{xx} + g_2) = 0$ with $g_1(x,t) = \int\limits_{-c}^{c} \alpha^+ \beta^{-1/2} \tilde{f}_{t\beta}(x|t\alpha) \varrho(\alpha) d\alpha$, $g_2(x,t) = \int\limits_{-c}^{c} \alpha^- \beta^{-1/2} \tilde{f}_{t\beta}(x|t\alpha) \varrho(\alpha) d\alpha$, initial condition $r(x,1) = 0$, and the Bayesian risk satisfies the asymptotic equality $\lim_{K\to\infty} (\overline{D}N)^{-1/2} R_N(\lambda) = t_0 \int\limits_{-c}^{c} \alpha^- \varrho(\alpha) d\alpha + (2\pi t_0)^{-1/2} \int\limits_{-\infty}^{\infty} r(x,t_0) dx$. Therefore, (31) – (32) actually provide the value of Bayesian risk, which is asymptotically equal to $\mathbf{E}_D\,(R_N(\lambda(m|D))$, where $R_N(\lambda(m|D))$ is a Bayesian risk with respect to conditional prior distribution density at a fixed D and $\mathbf{E}_D$ is a sign of mathematical expectation with respect to marginal distribution of D.

Conclusion

We obtained a differential equation that allows one to calculate the limiting value of the normalized Bayesian risk in the Gaussian one-armed bandit problem. The form of this equation is such that the variance estimate present in it is used as if it were precisely determined at a short initial stage of control. It is of interest to analyze the corresponding difference equation taking into account the terms of a higher order of smallness. One can expect that it will allow to describe the process of refining the variance estimate in a similar way as it takes place in the integro-difference recursive equation.

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

References

1. Kolnogorov A.V. Invariant Description of Control in a Gaussian One-Armed Bandit Problem. *Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming and Computer Software*, 2024, vol. 17, no. 1, pp. 27–36. DOI: 10.14529/mmp240103
2. Sragovich V.G. *Mathematical Theory of Adaptive Control.* Singapore, World Scientific, 2006. DOI: 10.1142/5857
3. Lattimore T., Szepesvari C. *Bandit Algorithms.* Cambridge, Cambridge University Press, 2020. DOI: 10.1017/9781108571401
4. 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
5. Bather J.A. The Minimax Risk for the Two-Armed Bandit Problem. *Mathematical Learning Models – Theory and Algorithms*, 1983, vol. 20, pp. 1–11. DOI: 10.1007/978-1-4612-5612-0_1
6. Chernoff H., Ray S.N. A Bayes Sequential Sampling Inspection Plan. *The Annals of Mathematical Statistics*, 1965, vol. 36, pp. 1387–1407. DOI: 10.1214/aoms/1177699898
7. Bradt R.N., Johnson S.M., Karlin S. On Sequential Designs for Maximizing the Sum of n Observations. *The Annals of Mathematical Statistics*, 1956, vol. 27, pp. 1060–1074. DOI: 10.1214/aoms/1177728073

Received Octomber 24, 2024

УДК 519.244+519.83 **DOI: 10.14529/mmp250103**

ПРЕДЕЛЬНОЕ ОПИСАНИЕ В ЗАДАЧЕ О ГАУССОВСКОМ ОДНОРУКОМ БАНДИТЕ С ОБОИМИ НЕИЗВЕСТНЫМИ ПАРАМЕТРАМИ

А.В. Колногоров, Новгородский государственный университет
им. Ярослава Мудрого, г. Великий Новгород, Российская Федерация

Мы рассматриваем предельное описание управления в задаче о гауссовском одноруком бандите, которая является математической моделью оптимизации пакетной обработки больших данных при наличии двух альтернативных методов с известной эффективностью первого метода. Установлено, что это описание дается дифференциальным уравнением в частных производных второго порядка, в котором дисперсия одношаговых доходов является известной. Этот результат означает, что в случае больших данных дисперсия может быть сколь угодно точно оценена на коротком начальном этапе обработки, а затем полученная оценка использована управляющей стратегией.

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

Александр Валерианович Колногоров, доктор физико-математических наук, профессор, кафедра «Прикладная математика и информатика», Новгородский государственный университет им. Ярослава Мудрого (г. Великий Новгород, Российская Федерация), kolnogorov53@mail.ru.

Поступила в редакцию 24 октября 2024 г.