MSC 62C10, 62L05, 91A35 **DOI: 10.14529/jcem250102**

A LIMITING DESCRIPTION OF CONTROL IN A GAUSSIAN ONE-ARMED BANDIT PROBLEM

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

We consider a limiting description of control in a Gaussian one-armed bandit problem in application to batch processing of big data, if estimates of unknown mathematical expectation and variance of one-step incomes are performed during data processing within batches. This description is given by a second-order partial differential equation in which the estimate of the unknown variance is present as a constant parameter. This result means that when processing big data, an arbitrarily accurate estimate of the unknown variance can be obtained at a relatively arbitrarily short initial stage, and then used for control.

Keywords: Gaussian one-armed bandit; batch processing; Bayesian approach; invariant description.

Introduction

In this article, we develop the ideas considered in [1, 2], where a background of this approach and some references are given (see, e.g., [3, 4]). Let's recall briefly the results of [1, 2]. We study the optimization of batch data processing in the framework of a Gaussian one-armed bandit problem if there are two alternative processing methods with known efficiency of the first mathod. In this case, processed data is divided into sufficiently large equal batches, all the data in the same batch are processed by the same method (action) and processing results (e.g., the numbers of successfully processed data in batches) are used for the control. The goal is to maximize (in some sense) the mathematical expectation of successfully processed data which is interpreted as total expected income.

Formally, the Gaussian one-armed bandit is a controlled random process ξ_n, $n = 1, 2, \ldots, N$, which values are interpreted as random incomes, depend only on the currently chosen actions y_n ($y_n \in \{1, 2\}$) and have a Gaussian distribution density $f_D(x|m) = (2\pi D)^{-1/2} \exp(-(x-m)^2/(2D))$ if the second action is chosen. Here m, D are the mathematical expectation and variance of one-step income for choosing the second action. The mathematical expectation of income for the choice of the first action is known and, without loss of generality, is zero. So, a one-armed bandit is described by the parameter $\theta = (m, D)$, which is assumed to be a priori unknown. Note that the Gaussian distribution of incomes is a consequence of batch data processing.

A control strategy σ at the point of time $n+1$ performs a choice of action y_{n+1} depending on the current history of the process. A regret

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

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

Let's consider a set of admissible parameters $\Theta = \{(m, D) : |m| \le C < +\infty, 0 < \underline{D} \le D \le \overline{D} < +\infty\}$ and a prior distribution density $\lambda(\theta)$ on it. A Bayesian risk is defined as

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

the corresponding optimal strategy is called the Bayesian strategy. Bayesian strategy and risk can be found by solving backwards the Bellman recursive equation. Bayesian approach to the one-armed bandit problem was previously considered in [5, 6] for a Bernoulli one-armed bandit which incomes have values 0 and 1, in [6] the limiting description of control by the second order partial differential equation was obtained. In [1, 2] the Bayesian approach was applied to a Gaussian one-armed bandit with both unknown mathematical expectation and variance of one-step incomes: in [1] the estimation of unknown variance was performed using the incomes inside batches and in [2] it was done based on the cumulative incomes in batches.

This article is devoted to a limiting description of control in the one-armed bandit problem if the estimation of unknown variance is performed using the incomes inside batches. As in [7], where such a limiting description is obtained if the estimation of unknown variance is performed using the cumulative incomes in batches, this description is given by the same second-order partial differential equation in which the estimate of unknown variance is present as a constant parameter. This means that when processing big data, regardless of how the variance is estimated, an arbitrarily accurate estimate of the unknown variance can be obtained at a relatively arbitrarily short initial stage, and then used for the control.

The rest of the article is as follows. In section 1, recursive equations are obtained for computing Bayesian strategy and risk in the usual and invariant forms. This equations are equivalent to those obtained in [1] but more convenient for passing to the limiting description. In section 2, using the invariant recursive equation, we obtain the limiting description of control by the second order partial differential equation. The conclusion is presented in section 3.

1. Recursive Equations for Finding Bayesian Risks and Strategies

We assume that the processing is carried out in batches of the size $M \ge 2$ and the variance estimation is performed during data processing within batches. The number of batches and, accordingly, the number of processing stages is K, the total number of data is $N = KM$. We also assume that at the beginning of control the second action is applied at least $k_0 \ge 1$ times. Note that if $k_0 \ll K$ then this strategy is close to optimal.

Let's consider how to recalculate the sufficient statistics which are the total income X and s^2-statistics S. Let $k \ge k_0$ be the current number of processed batches and, therefore, $n = kM$ be the current total number of processed data. If $x_1, \ldots, x_n$ and $x_{n+1}, \ldots, x_{n+M}$ are the data incomes in the first k batches and in the $(k+1)$th batch respectively, then

$$X = \sum_{i=1}^{n} x_i,\ S = \sum_{i=1}^{n} x_i^2 - X^2/n, \quad Y = \sum_{i=n+1}^{n+M} x_i,\ U = \sum_{i=n+1}^{n+M} x_i^2 - Y^2/M$$

are total income and s^2-statistics in the first k batches and in the $(k+1)$th batch. Therefore, the new values of total income and s^2-statistics are recalculated according to the following

formulas

$$X_{new} = \sum_{i=1}^{n+M} x_i = X + Y, \quad S_{new} = \left(\sum_{i=1}^{n+M} x_i^2 \right) - \frac{(X+Y)^2}{n+M} = S + \Delta + U,$$

where $\Delta = Y^2/M + X^2/n - (X+Y)^2/(n+M) = (MX - nY)^2/(nM(n+M))$. Thus, the recalculation of statistics is carried out according to the formulas

$$X \leftarrow X + Y, \quad S \leftarrow S + \Delta + U, \quad \text{with } \Delta = \frac{(X - kY)^2}{Mk(k+1)}. \tag{2}$$

Consider a chi-squared distribution density with k degrees of freedom

$$\chi_k^2(x) = \{2^{k/2}\Gamma\,(k/2)\}^{-1} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \quad x \geq 0, \quad k \geq 1.$$

Denote by $D' = MD$ and $m' = Mm$ the variance and the mathematical expectation of income for processing the batch. Let's introduce the functions

$$f_{kD'}(X|km') = \frac{1}{\sqrt{2\pi k D'}} \exp\left(-\frac{(X - km')^2}{2kD'} \right), \quad \psi_{kM-1}(S/D) = (D)^{-1} \chi^2_{kM-1}(S/D).$$

If $k \geq k_0$, these functions describe the probability density functions of cumulative income X and s^2-statistics S computed after processing k batches. Since X and S are independent random variables, then

$$\mathbf{F}(X, S|m, D) = f_{MD}(X|Mm)\psi_{M-1}\,(S/D) \tag{3}$$

describes the distribution density of X, S, corresponding to processing one batch.

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

$$\lambda(m, D|X, S, k) = \frac{f_{kD'}\,(X|km')\,\psi_{kM-1}\,(S/D)\,\lambda(m, D)}{P(X, S, k)},$$
$$\text{with} \quad P(X, S, k) = \iint\limits_{\Theta} f_{kD'}\,(X|km')\,\psi_{kM-1}\,(S/D)\,\lambda(m, D) dm dD.$$

However, it can be defined in the following equivalent way. Denote

$$\tilde{\mathbf{F}}(X, S, k|m, D) = (D)^{-3/2} \tilde{f}_{kD'}(X|km')\tilde{\psi}_{Mk-1}\,(S/D)\,, \tag{4}$$

where

$$\begin{aligned} &\tilde{f}_D(x|m) = \exp\left(-(x-m)^2/(2D)\right), \\ &\tilde{\psi}_{kM-1}(s) = (kM/(4\pi))^{1/2}\,(s/(kM))^{\frac{kM-3}{2}}\,e^{-(s-kM)/2}. \end{aligned} \tag{5}$$

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

$$\begin{aligned} &\lambda(m, D|X, S, k) = \frac{\tilde{\mathbf{F}}(X, S, k|m, D)\lambda(m, D)}{\tilde{P}(X, S, k)}, \\ &\text{with} \quad \tilde{P}(X, S, k) = \iint\limits_{\Theta} \tilde{\mathbf{F}}(X, S, k|m, D)\lambda(m, D) dm dD. \end{aligned} \tag{6}$$

To present the recursive equation for finfing Bayesian strategy and risk, we will use the following property of the optimal strategy which was first established in [5] and was already used in [1, 2] and [7]. Since the applying the first action does not give additional information (the corresponding distribution is known), then once it has been chosen it will be applied until the end of the control. Denote by $R^B(X,S,k) = R^B_{K-k}(\lambda(m,D|X,S,k)$ the Bayesian risk computed on the control horizon $K-k$ with respect to a prior distribution density $\lambda(m,D|X,S,k)$. Denote $m^+ = \max(m,0)$, $m^- = \max(-m,0)$. Taking into account (2)–(3), the standard dynamic programming equation has the form

$$R^B(X,S,k) = \min\left(R^B_1(X,S,k), R^B_2(X,S,k)\right), \tag{7}$$

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

$$\begin{gathered} R^B_1(X,S,k) = (K-k)\iint\limits_{\Theta} Mm^+\lambda(m,D|X,S,k)dmdD, \\ R^B_2(X,S,k) = \iint\limits_{\Theta} \lambda(m,D|X,S,k)\times\Big(Mm^- \\ +\int\limits_0^{\infty}\int\limits_{-\infty}^{\infty} R^B(X+Y,S+\Delta+U,k+1)\mathbf{F}(Y,U|m,D)dYdU\Big)dmdD \end{gathered} \tag{8}$$

if $k_0 \le k < K$. Bayesian risk (1) is

$$R_N(\lambda) = k_0\iint\limits_{\Theta} Mm^-\lambda(m,D)dmdD + \int\limits_0^{\infty}\int\limits_{-\infty}^{\infty} R^B(X,S,k_0)P(X,S,k_0)dXdS. \tag{9}$$

Here $R^B_\ell(X,S,k)$ characterizes the expected loss on the control horizon $K-k$ if the ℓth action is applied first and then the control is carried out optimally. When processing the $(k+1)$th batch, the Bayesian strategy prescribes choosing an action corresponding to the current smaller value $R^B_1(X,S,k)$, $R^B_2(X,S,k)$; in the case of a draw, the choice can be arbitrary. Once the first action has been chosen, it will be applied until the end of the control.

Let's present equation (7)–(9) in a more convenient for computations form. 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 given in (6).

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

$$R(X,S,k) = \min\left(R_1(X,S,k), R_2(X,S,k)\right), \tag{10}$$

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

$$\begin{gathered} R_1(X,S,k) = M(K-k)G_1(X,S,k), \\ R_2(X,S,k) = MG_2(X,S,k) \\ +\int\limits_0^{\infty}\int\limits_{-\infty}^{\infty} R(X+Y,S+\Delta+U,k+1)H(X,S,k,Y,U)dYdU \end{gathered} \tag{11}$$

if $k_0 \le k < K$. *Here*

$$G_1(X,S,k) = \iint\limits_{\Theta} m^+ \tilde{\mathbf{F}}(X,S,k|m,D)\lambda(m,D)dmdD,$$
$$G_2(X,S,k) = \iint\limits_{\Theta} m^- \tilde{\mathbf{F}}(X,S,k|m,D)\lambda(m,D)dmdD \tag{12}$$

and

$$H(X,S,k,Y,U) = \frac{C(k,M)}{S^{3/2}} \times \frac{S^{(kM)/2}U^{(M-3)/2}}{(S+\Delta+U)^{((k+1)M-3)/2}}, \tag{13}$$

$$\textit{with} \quad C(k,M) = \left(\frac{1}{2^M M\pi}\right)^{1/2} \times \frac{1}{\Gamma(\frac{M-1}{2})} \times \left(\frac{k+1}{k}\right)^{\frac{kM-4}{2}} \left(\frac{(k+1)M}{e}\right)^{\frac{M}{2}}.$$

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

$$R_N(\lambda) = k_0 \iint\limits_{\Theta} Mm^-\lambda(m,D)dmdD + H_0 \int\limits_0^{\infty}\int\limits_{-\infty}^{\infty} R(X,S,k_0)dXdS \tag{14}$$

$$\textit{with} \quad H_0 = \frac{2}{(k_0M)^{5/2}} \times \frac{1}{\Gamma(\frac{k_0M-1}{2})} \times \left(\frac{k_0M}{2e}\right)^{\frac{k_0M}{2}}.$$

Proof. The proof is similar to the proof of theorem 1 in [1]. One should multiply the left-hand and right-hand sides of the equation (7)–(8) by $\tilde{P}(X,S,k)$ in (6) and get (10)–(11), where $G_1(X,S,k)$, $G_2(X,S,k)$ are described by (12). Formulas (13), (14) can be obtained after performing transformations in the expressions

$$H(X,S,k,Y,U) = \frac{\tilde{\mathbf{F}}(X,S,k|m,D)\mathbf{F}(Y,U|m,D)}{\tilde{\mathbf{F}}(X+Y,S+\Delta+U,k+1|m,D)},$$
$$H_0 = \frac{P(X,S,k_0)}{\tilde{P}(X,S,k_0)} = \frac{f_{k_0D'}(X|k_0m')\psi_{k_0M-1}(S/D)}{(D)^{-3/2}\tilde{f}_{k_0D'}(X|k_0m')\tilde{\psi}_{k_0M-1}(S/D)}.$$

□

Let's obtain an invariant form of formulas (10)–(14). We take the 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 one puts $D = \beta\overline{D}$, $m = \alpha(\overline{D}/N)^{1/2}$, then the set of parameters 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: $X = x(\overline{D}N)^{1/2}$, $Y = y(\overline{D}N)^{1/2}$, $S = skM\overline{D}$, $U = ukM\overline{D}$, $k = tK$, $k_0 = t_0K$, $M/N = K^{-1} = \varepsilon$, $\lambda(m,D) = (N/\overline{D}^{\,3})^{1/2}\varrho(\alpha,\beta)$, $R_\ell(X,S,k) = (\overline{D}\,N)^{1/2}(\overline{D})^{-3/2}r_\ell(x,s,t)$, $\ell = 1,2$. Then the following theorem is valid.

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

$$r(x,s,t)=\min\left(r_1(x,s,t),r_2(x,s,t)\right), \tag{15}$$

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

$$\begin{gathered} r_1(x,s,t)=(1-t)g_1(x,s,t),\\ r_2(x,s,t)=\varepsilon g_2(x,s,t)+\\ +\int\limits_0^{\infty}\int\limits_{-\infty}^{\infty} r(x+y,\frac{s+t^{-1}\delta(x,t,y)+u}{1+\varepsilon/t},t+\varepsilon)h(x,s,t,y,u)dydu, \end{gathered} \tag{16}$$

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

$$\begin{aligned} g_1(x,s,t)&=\iint\limits_{\Theta_N}\alpha^{+}\beta^{-3/2}\tilde{f}_{t\beta}(x|t\alpha)\tilde{\psi}_{kM-1}(kMs/\beta)\varrho(\alpha,\beta)d\alpha d\beta,\\ g_2(x,s,t)&=\iint\limits_{\Theta_N}\alpha^{-}\beta^{-3/2}\tilde{f}_{t\beta}(x|t\alpha)\tilde{\psi}_{kM-1}(kMs/\beta)\varrho(\alpha,\beta)d\alpha d\beta, \end{aligned} \tag{17}$$

$$h(x,s,t,y,u)=\frac{c(k,M)}{s^{3/2}}\times\frac{s^{(kM)/2}u^{(M-3)/2}}{(s+t^{-1}\delta(x,t,y)+u)^{((k+1)M-3)/2}}, \tag{18}$$

$$\textit{with}\quad c(k,M)=\left(\frac{1}{2^M M\pi t}\right)^{1/2}\times\frac{1}{\Gamma\left(\frac{M-1}{2}\right)}\times\left(\frac{t+\varepsilon}{t}\right)^{\frac{kM-4}{2}}\left(\frac{(k+1)M}{e}\right)^{\frac{M}{2}}$$

and

$$t^{-1}\delta(x,t,y)=\frac{(\varepsilon x-ty)^2}{Mt^2(t+\varepsilon)}, \tag{19}$$

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

$$R_N(\lambda)=(\overline{D}N)^{1/2}\left(t_0\iint\limits_{\Theta}\alpha^{-}\varrho(\alpha,\beta)d\alpha d\beta+h_0\int\limits_0^{\infty}\int\limits_{-\infty}^{\infty}r(x,s,t_0)dxds\right) \tag{20}$$

$$\textit{with}\quad h_0=\frac{2}{t_0^{1/2}(k_0M)}\times\frac{1}{\Gamma(\frac{k_0M-1}{2})}\times\left(\frac{k_0M}{2e}\right)^{\frac{k_0M}{2}}.$$

Proof. The proof is similar to the proof of theorem 2 in [1] and is obtained after performing the above change of variables in (10)–(14).

□

Remark. Like in [1], we can assume that the data in batches themselves are the small packets of M_1 pieces. Thus, the total number of such pieces is $N \times M_1$, their mathematical expectation and variance are m/M_1 and D/M_1 respectively. Therefore, the following relations are valid: $m/M_1 = \alpha((\overline{D}/M_1)/(N \times M_1))^{1/2}$, $D/M_1 = \beta\overline{D}/M_1$ and $(\overline{D}N)^{1/2} = ((\overline{D}/M_1)(N \times M_1))^{1/2}$. Hence, (15)–(20) will not change for this control problem. Therefore, description of control in theorem 2 is invariant in the sense that it depends only on the number of batches K and their sizes M even if the data in batches are themselves the packets of M_1 pieces.

2. Limiting Description. Differential Equation

In order to present the limiting description of (15)–(20), we need the following auxiliary results. Note that lemma 1 and lemma 2 are proved in [7].

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

$$I_1(\kappa) = \int\limits_{-\infty}^{\infty} \frac{dz}{(1+z^2)^{\kappa}} = \left(\frac{\pi}{\kappa}\right)^{1/2} \left(1 + \frac{3}{8\kappa} + o(\kappa^{-1})\right). \tag{21}$$

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

$$I_1^D(\kappa) = \int\limits_{-\infty}^{\infty} \frac{z^2 dz}{(1+z^2)^{\kappa}} = \frac{I_1(\kappa)}{2\kappa - 3}. \tag{22}$$

Lemma 3. *For a factor h_0 in (20) with $k_0 = t_0 K$, $t_0 > 0$, the asymptotic (as $K \to \infty$) estimate is valid*

$$h_0 = (2\pi t_0)^{-1/2}\,(1 + o(1)). \tag{23}$$

Proof. We use the Stirling's formula $\Gamma(\kappa+1) \sim (2\pi)^{1/2}\kappa^{\kappa+1/2}e^{-\kappa}$. Then

$$h_0 \sim \frac{1}{(2\pi t_0)^{1/2}} \times \left(\frac{k_0 M}{k_0 M - 3}\right)^{\frac{k_0 M}{2}} \left(\frac{k_0 M - 3}{k_0 M}\right) \times e^{-\frac{3}{2}} = \frac{1}{(2\pi t_0)^{1/2}}\,(1 + o(1))$$

as $k_0 \to \infty$.

□

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

$$\begin{aligned} g_1(x,s,t) &= \mathrm{I}\,(s, (\beta_0, 1)) \times \int\limits_{-c}^{c} \alpha^{+} s^{-1/2} \tilde{f}_{ts}(x|t\alpha)\varrho(\alpha, s)d\alpha, \\ g_2(x,s,t) &= \mathrm{I}\,(s, (\beta_0, 1)) \times \int\limits_{-c}^{c} \alpha^{-} s^{-1/2} \tilde{f}_{ts}(x|t\alpha)\varrho(\alpha, s)d\alpha, \end{aligned} \tag{24}$$

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

Proof. Consider $\tilde{\psi}_{kM-1}(kMs/\beta)$ from (17). Using (5), like in [7], lemma 4, one can obtain that $\beta^{-1}\tilde{\psi}_{kM-1}(kMs/\beta) = (kM/(4\pi\beta^2))^{1/2}\exp(-kM(s-\beta)^2/(4\beta^2))(1+o(1))$. This function converges to the Dirac delta function $\delta^D(s-\beta)$ as $\varepsilon \to 0$. Hence, 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 (17), we obtain (24).

□

Lemma 5. *Given $s' > 0$, the equalities hold*

$$I_2(a,b) = \int\limits_0^{\infty} \frac{u^a s'^b}{(s'+u)^{a+b}} du = s'\mathrm{B}(a+1, b-1), \tag{25}$$

$$I_2'(a,b) = \int\limits_0^{\infty} \frac{u^{a+1} s'^b}{(s'+u)^{a+b}} du = s' I_2(a+1,b-1) = s'\frac{a+1}{b-2} I_2(a,b), \tag{26}$$

where $\mathrm{B}(a,b)$ *is the Beta-function.*

Proof. To prove (25), let's make the change of variables $x = u/(s'+u)$, then $1-x = s'/(s'+u)$, $du = s'(1-x)^{-2}dx$ and, hence, $I_1(a,b) = s'\int\limits_0^1 x^a(1-x)^{b-2}dx = s'\mathrm{B}(a+1,b-1)$. To prove (26), one should use the formulas $\mathrm{B}(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, $\Gamma(a+1) = a\Gamma(a)$.

□

Lemma 6. *For $c(k,M)$ from* (18)*, the asymptotic (as $k \to \infty$) estimate holds*

$$c(k,M)\mathrm{B}\left(\frac{M-1}{2}, \frac{kM-2}{2}\right) = \frac{1}{(2\pi\varepsilon)^{1/2}}\left(1 - \frac{7}{4kM} + o(k^{-1})\right). \tag{27}$$

Proof. Using the Stirling's formula, we obtain

$$\mathrm{B}(a+1,b-1) = \frac{\Gamma(a+1)\Gamma(b-1)}{\Gamma(a+b)} \sim \Gamma(a+1)\left(\frac{1}{1+\frac{a+1}{b-2}}\right)^{a+b-\frac{1}{2}}\left(\frac{e}{b-2}\right)^{a+1}.$$

Then for $a = (M-3)/2$, $b = kM/2$ we have

$$c(k,M)\mathrm{B}(a+1,b-1) \sim \left(\frac{1}{2\pi e\varepsilon}\right)^{1/2}\left(\frac{1-\frac{4}{kM}}{1+\frac{M-5}{kM}}\right)^{\frac{(k+1)M-4}{2}} \times \left(1-\frac{4}{kM}\right)^{1/2} \times \left(1+\frac{1}{k}\right)^{\frac{kM-4}{2}}\left(\frac{1+\frac{1}{k}}{1-\frac{4}{kM}}\right)^{\frac{M}{2}}.$$

From here, (27) follows.

□

Lemma 7. *The asymptotic (as $k \to \infty$) estimate holds*

$$I_3(k,M) = \int\limits_{-\infty}^{\infty} \left(\frac{s}{s + t^{-1}\delta(x,t,y)} \right)^{\frac{kM-2}{2}} dy = (2\pi\varepsilon s)^{1/2} \left(1 + \frac{2M+7}{4kM} + o(k^{-1}) \right). \quad (28)$$

Proof. Let's make the change of variables in $I_3(k,M)$: $sz^2 = t^{-1}\delta(x,t,y)$, then $\varepsilon x - ty = -tz(Ms(t+\varepsilon))^{1/2}$, $dy = (Ms(t+\varepsilon))^{1/2}dz$. Using (21), we have

$$I_3(k,M) = (Ms(t+\varepsilon))^{1/2} I_1\left(\frac{kM-2}{2}\right) = (2\pi\varepsilon s)^{1/2} \left(\frac{kM+M}{kM-2}\right)^{1/2} \left(1 + \frac{3}{4(kM-2)}\right).$$

From here, (28) follows.

□

Lemma 8. *The asymptotic (as $k \to \infty$) estimates hold*

$$I_{41}(k,M) = \int\limits_{-\infty}^{\infty} \int\limits_{0}^{\infty} h(x,s,t,y,u)dudy = (1 + 1/(2k) + o(k^{-1})), \quad (29)$$

$$I_{42}(k,M) = \int\limits_{-\infty}^{\infty} \int\limits_{0}^{\infty} u \times h(x,s,t,y,u)dudy = s\frac{M-1}{kM}(1 + o(1)), \quad (30)$$

$$I_{43}(k,M) = \int\limits_{-\infty}^{\infty} \int\limits_{0}^{\infty} z^2 \times h(x,s,t,y,u)dudy = \frac{1}{kM}(1 + o(1)), \quad (31)$$

$$\int\limits_{-\infty}^{\infty} \int\limits_{0}^{\infty} u^i \times h(x,s,t,y,u)dudy = o(k^{-1}), \quad i = 2,3,\ldots \quad (32)$$

$$\int\limits_{-\infty}^{\infty} \int\limits_{0}^{\infty} z^{2i+1} \times h(x,s,t,y,u)dudy = 0, \quad i = 0,1,\ldots \quad (33)$$

$$\int\limits_{-\infty}^{\infty} \int\limits_{0}^{\infty} z^{2i} \times h(x,s,t,y,u)dudy = o(k^{-1}), \quad i = 2,3,\ldots \quad (34)$$

Here z is defined in lemma 7 by condition $sz^2 = t^{-1}\delta(x,t,y)$.

Proof. Let's prove (29). Denote $s' = s + t^{-1}\delta(x,t,y)$. Then, using (27), (28),

$$I_{41}(k,M) = \frac{c(k,M)s^{kM/2}}{s^{3/2}} \int\limits_{-\infty}^{\infty} \frac{1}{(s')^{kM/2}} \left(\int\limits_{0}^{\infty} \frac{(s')^{kM/2} u^{(M-3)/2}}{(s'+u)^{((k+1)M-3)/2}} du \right) dy =$$

$$= \frac{c(k,M)}{s^{1/2}} \mathrm{B}\left(\frac{M-1}{2}, \frac{kM-2}{2}\right) \int\limits_{-\infty}^{\infty} \left(\frac{s}{s + t^{-1}\delta(x,t,y)} \right)^{(kM-2)/2} dy =$$

$$= (1 + 1/(2k) + o(k^{-1})).$$

Similarly, as $I_1((kM-4)/2) \sim I_1((kM-2)/2)$, we have for (30)

$$I_{42}(k,M) = \frac{c(k,M)s^{(kM-2)/2}}{s^{1/2}} \int\limits_{-\infty}^{\infty} \frac{1}{(s')^{(kM-2)/2}} \left(\int\limits_0^{\infty} \frac{(s')^{(kM-2)/2} u^{(M-1)/2}}{(s'+u)^{((k+1)M-3)/2}} du \right) dy =$$

$$= s^{1/2} c(k,M) \mathrm{B}\left(\frac{M+1}{2}, \frac{kM-4}{2} \right) I_1((kM-4)/2) = s \frac{M-1}{kM-4}(1+o(1)).$$

For (31) we have

$$I_{43}(k,M) = \frac{c(k,M)s^{kM/2}}{s^{3/2}} \int\limits_{-\infty}^{\infty} \frac{z^2}{(s')^{kM/2}} \left(\int\limits_0^{\infty} \frac{(s')^{kM/2} u^{(M-3)/2}}{(s'+u)^{((k+1)M-3)/2}} du \right) dy =$$

$$= \frac{c(k,M)}{s^{1/2}} \mathrm{B}\left(\frac{M-1}{2}, \frac{kM-2}{2} \right) I_1^D((kM-2)/2) = \frac{1}{kM-5}(1+o(1)).$$

Formulas (32)–(34) are checked in a similar way.

□

Let's obtain a limiting description of recursive equation (15)–(16) as $\varepsilon \to 0$. Let s be defined in lemma 7 by condition $sz^2 = t^{-1}\delta(x,t,y)$, so that $\varepsilon x - ty = -tz(Ms(t+\varepsilon))^{1/2}$ $y = \varepsilon x/t + z(Ms(t+\varepsilon))^{1/2}$. Let's assume that $r(x,s,t+\varepsilon)$ has partial derivatives of the required orders by x, s and denote them by r, r'_x, r''_{xx}, r'_s. Presenting $r(x+y, (s(1+z^2)+u))/(1+\varepsilon t^{-1}), t+\varepsilon)$ as a Taylor's series, we obtain

$$r + r'_x \times \varepsilon x/t + 0.5 r''_{xx} \times z^2 (Ms(t+\varepsilon)) + r'_s \times (-s\varepsilon/t + (sz^2+u)) + A(\varepsilon, z, u). \quad (35)$$

Here r'_x, r''_{xx}, r'_s are calculated at the point $(x, s, t+\varepsilon)$ and $A(\varepsilon, z, u)$ contains the terms which are $o(\varepsilon)$ after integration according to lemma 8. Substituting (35) into the integral in the second equation (16), taking into account (29)–(34) and equality $k^{-1} = \varepsilon/t$ we obtain that second equation (16) turns to

$$r_2(\cdot, t) = r(\cdot, t+\varepsilon) \times (1 + \varepsilon/(2t)) + \varepsilon r'_x(\cdot, t+\varepsilon) \times (x/t) + 0.5\varepsilon r''_{xx}(\cdot, t+\varepsilon) + o(\varepsilon). \quad (36)$$

Let's obtain the differential equation. To this end, we write (15) in equivalent form $\min(r_1(\cdot,t) - r(\cdot,t), \varepsilon^{-1}(r_2(\cdot,t) - r(\cdot,t))) = 0$, where $r_1(\cdot,t)$ and $r_2(\cdot,t)$ are taken from the first equation (16) and (36) respectively. In the limit as $\varepsilon \to 0$, we get the equation

$$\min\left((1-t)g_1 - r, r'_t + r/(2t) + r'_x \times (x/t) + 0.5 s r''_{xx} + g_2\right) = 0, \quad (37)$$

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 (37); 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 (24) and the Bayesian risk (1) asymptotically is equal to

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

Conclusion

We have considered the limiting description of the batch data processing with an estimation of the variance of the distribution of one-step incomes by incomes within batches. This description is given by the same second-order partial differential equation as if the variance estimation is performed based on cumulative incomes in batches.

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. Optimization of Two-Alternative Batch Processing with Parameter Estimation Based on Data Inside Batches. *Journal of Computational and Engineering Mathematics*, 2023, vol. 10, no. 4, pp. 40–50. DOI: 10.14529/jcem230403

2. 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

3. Sragovich V.G. *Mathematical Theory of Adaptive Control*. Singapore, World Sci., 2006.

4. Slivkins A. *Introduction to Multi-Armed Bandits*, arXiv:1904.07272v7, 2022.

5. 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

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. Kolnogorov A.V. A Limiting Description in a Gaussian One-Armed Bandit Problem with Both Unknown Parameters. *Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming and Computer Software*, 2025, vol. 18, no. 1, pp. 35–45. DOI: 10.14529/mmp250103

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

Received Jaunary 10, 2025

УДК 519.244, 519.83 DOI: 10.14529/jcem250102

ПРЕДЕЛЬНОЕ ОПИСАНИЕ УПРАВЛЕНИЯ В ЗАДАЧЕ О ГАУССОВСКОМ ОДНОРУКОМ БАНДИТЕ

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

Рассматривается предельное описание управления в задаче о гауссовском однору-ком бандите в приложении к пакетной обработке больших данных, если оценки неиз-вестных математического ожидания и дисперсии одношаговых доходов выполняются в процессе обработки данных внутри пакетов. Это описание дается дифференциальным уравнением в частных производных второго порядка, в котором оценка неизвестной дисперсии присутствует как постоянный параметр. Данный результат означает, что при обработке больших данных сколь угодно точная оценка неизвестной дисперсии может быть получена на относительно сколь угодно коротком начальном этапе, а за-тем использована для управления.

Ключевые слова: гауссовский однорукий бандит; пакетная обработка; байесов-ский подход; инвариантное описание.

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

Поступила в редакцию 10 января 2025 г.