

УДК 519.865

К ПРЕДЕЛЬНОМУ ОПИСАНИЮ МИНИМАКСНОГО ПАРАЛЛЕЛЬНОГО УПРАВЛЕНИЯ В СЛУЧАЙНОЙ СРЕДЕ

А.В.Колногоров

CONTRIBUTIONS TO THE LIMITING DESCRIPTION OF THE MINIMAX PARALLEL CONTROL IN RANDOM ENVIRONMENT

A.V.Kolnogorov

Институт электронных и информационных систем НовГУ, Alexander.Kolnogorov@novsu.ru

Получено предельное описание минимаксного параллельного управления в случайной среде с помощью дифференциального уравнения в частных производных второго порядка. Численные эксперименты показывают близость решений дифференциального уравнения и ранее полученного рекуррентного интегро-разностного уравнения.

Ключевые слова: управление в случайной среде, задача о двуруком бандите, робастное управление, параллельное управление, уравнения в частных производных

A limiting description of the minimax parallel control in random environment by means of the second order partial differential equation is obtained. Numerical experiments show the similarity of the solutions to the partial differential equation and to previously obtained recurrence integral-difference equation.

Keywords: control in random environment, two-arm bandit problem, robust control, parallel control, partial differential equation

1. Введение

Рассматривается задача управления в случайной среде, продолжающая работу [1], в которой указаны также другие подходы и библиография. Опишем кратко результаты [1]. Пусть ξ_n , $n=1, \dots, N$ есть управляемый случайный процесс, значения которого интерпретируются как доходы, зависят только от выбираемых в текущие моменты времени вариантов и имеют нормальные распределения с плотностями $f(x | m_\ell) = (2\pi)^{-1/2} \exp(-(x - m_\ell)^2 / 2)$, если выбран вариант с номером ℓ ($\ell=1,2$). Такая среда описывается векторным параметром $\theta = (m_1, m_2)$. Предполагается, что множество допустимых значений параметра имеет вид $\Theta = \{(m_1, m_2) : |m_1 - m_2| \leq 2cN^{-1/2}\}$, где $c > 0$ — некоторая константа, т.е. рассматривается класс сред с близкими математическими ожиданиями.

Для управления используется стратегия, которая вначале применяет каждый вариант по $M_0 = \varepsilon_0 N$

раз, а затем осуществляет оптимальное управление с тем ограничением, что варианты могут меняться только после применения $M = \varepsilon N$ раз подряд. Поясним выбор такой стратегии. Мы рассматриваем задачу в приложении к обработке больших массивов данных, для обработки которых могут использоваться два альтернативных метода (варианта). Результат обработки является бинарным: 1 — если обработка успешная и 0 — если обработка неуспешная. Рассматриваемая стратегия может обрабатывать параллельно пакеты из M поступающих подряд данных, для обработки которых применяется один и тот же вариант. В силу центральной предельной теоремы количество успешно обработанных данных в одном пакете имеет приблизительно нормальное распределение, и этим оправдан выбор нормального распределения доходов. Причем полное время управления при параллельной обработке может быть сравнительно небольшим.

Отметим, что параллельная обработка была первоначально предложена применительно к задаче о лечении большой группы пациентов двумя альтерна-

тивными лекарствами. В байесовской постановке обсуждение и библиография имеются, например, в [2], в минимаксной — в [3].

Сформулируем цель управления. Если бы параметр θ был известен, то следовало бы всегда применять вариант, которому соответствует большая из величин m_1 , m_2 , и полный ожидаемый доход был бы равен в этом случае $N(m_1 \vee m_2)$. Если же параметр неизвестен, то функция

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

характеризует потери дохода вследствие неполноты информации. Здесь $E_{\sigma, \theta}$ обозначает математическое ожидание по мере, порожденной стратегией σ и параметром θ .

При минимаксном подходе максимальные потери на множестве параметров Θ минимизируются по множеству стратегий $\{\sigma\}$, величина

$$R_N^M(\Theta) = \inf_{\{\sigma\}} \sup_{\Theta} L_N(\sigma, \theta)$$

называется минимаксным риском, а соответствующая стратегия (если она существует) — минимаксной стратегией. Минимаксная стратегия обеспечивает робастное управление, так как реальный ожидаемый доход для любой среды из множества Θ отличается от максимально возможного не более чем на величину минимаксного риска. В данной статье исследуется минимаксное управление. Отметим, что робастный подход впервые использован в работе [4], где получены оценки минимаксного риска, имеющие неулучшаемый порядок $N^{1/2}$. В нашей стране минимаксный подход возник независимо от [4] и получил значительное развитие (см., напр., [5]).

Мы используем результаты [1], где показано, что минимаксная стратегия и риск могут быть найдены как байесовские, соответствующие наихудшему априорному распределению, и дано рекуррентное интегро-разностное уравнение для вычисления байесовского риска. Основной результат статьи состоит в предельном описании байесовского риска дифференциальным уравнением в частных производных второго порядка. Даны результаты численных экспериментов, показывающие близость решений интегро-разностного и дифференциального уравнений.

2. Интегро-разностное уравнение

Приведем интегро-разностное уравнение, полученное в [1]. Пусть к моменту времени $n = n_1 + n_2$ оба варианта применены соответственно n_1 и n_2 раз, а X_1 , X_2 — полные доходы за их применение. В [1] показано, что в качестве статистики можно использовать тройку (Z, n_1, n_2) при $Z = X_1 n_2 - X_2 n_1$. Положим $m_1 = u + v$, $m_2 = u - v$, тогда $\theta = (u + v, u - v)$, $\Theta = \{\theta : |v| \leq cN^{-1/2}\}$. В новых переменных асимптотически наихудшая плотность распределения имеет вид $v_a(u, v) = k_a(u)\mu(v)$, где $k_a(u)$ — постоянная плот-

ность при $|u| \leq a$, $\mu(v) = \mu(-v)$ — симметричная плотность и $a \rightarrow \infty$.

Положим $t_1 = n_1 / N$, $t_2 = n_2 / N$, $s = ZN^{-3/2}$, $w = vN^{1/2}$, $\rho(w) = \mu(v)N^{-1/2}$. Для вычисления байесовского риска следует решать рекуррентное уравнение:

$$r_\varepsilon(s, t_1, t_2) = \min(r_\varepsilon^{(1)}(s, t_1, t_2), r_\varepsilon^{(2)}(s, t_1, t_2)), \quad (1)$$

где $r_\varepsilon^{(1)}(s, t_1, t_2) = r_\varepsilon^{(2)}(s, t_1, t_2) = 0$ при $t_1 + t_2 = 1$,

$$\begin{aligned} r_\varepsilon^{(1)}(s, t_1, t_2) &= \varepsilon g^{(1)}(s, t_1, t_2) + \\ &+ t_2^{-1} \int_{-\infty}^{+\infty} r_\varepsilon(s+x, t_1+\varepsilon, t_2) h_\varepsilon\left(\frac{s\varepsilon-t_1x}{t_2}, t_1\right) dx, \\ r_\varepsilon^{(2)}(s, t_1, t_2) &= \varepsilon g^{(2)}(s, t_1, t_2) + \\ &+ t_1^{-1} \int_{-\infty}^{+\infty} r_\varepsilon(s+x, t_1, t_2+\varepsilon) h_\varepsilon\left(\frac{s\varepsilon-t_2x}{t_1}, t_2\right) dx \end{aligned} \quad (2)$$

при $t_1 + t_2 < 1$, $t_1 \geq \varepsilon_0$, $t_2 \geq \varepsilon_0$. Здесь

$$g^{(\ell)}(s, t_1, t_2) = \int_0^\infty 2wg(s, (-1)^{\ell+1}w, t_1, t_2) \rho(w) dw, \quad \ell = 1, 2,$$

$$g(s, w, t_1, t_2) = (2\pi t_1 t_2 (t_1 + t_2))^{-1/2} \exp\left(-\frac{(s + 2wt_1 t_2)^2}{2t_1 t_2 (t_1 + t_2)}\right),$$

$$h_\varepsilon(s, t) = \left(\frac{t+\varepsilon}{2\pi t\varepsilon}\right)^{1/2} \exp\left(-\frac{s^2}{2t\varepsilon(t+\varepsilon)}\right).$$

Оптимальная стратегия при $t_1 + t_2 \leq 2\varepsilon_0$ ($n_1 + n_2 \leq 2\varepsilon_0 N$) применяет варианты по очереди. При $t_1 + t_2 > 2\varepsilon_0$ ($n_1 + n_2 > 2\varepsilon_0 N$) текущим оптимальным является тот вариант, которому соответствует меньшее значение $r_\varepsilon^{(\ell)}(s, t_1, t_2)$, $\ell = 1, 2$. При этом байесовский риск относительно наихудшего априорного распределения вычисляется по формуле

$$\lim_{a \rightarrow \infty} R_N^B(\rho(w)) = \left(4\varepsilon_0 \int_0^\infty w \rho(w) dw + \int_{-\infty}^\infty r_\varepsilon(s, \varepsilon_0, \varepsilon_0) ds \right) N^{1/2}. \quad (3)$$

3. Дифференциальное уравнение

Наша цель — получить дифференциальное уравнение при переходе к пределу по $\varepsilon \rightarrow 0$ в уравнениях (2). Покажем, что в этом случае справедливы уравнения

$$\begin{aligned} r_\varepsilon^{(1)}(s, t_1, t_2) &= r_\varepsilon(s, t_1 + \varepsilon, t_2) + \\ &+ \varepsilon \cdot \left(\frac{r_\varepsilon}{t_1} + \frac{\partial r_\varepsilon}{\partial s} \cdot \frac{s}{t_1} + \frac{t_2^2}{2} \cdot \frac{\partial^2 r_\varepsilon}{\partial s^2} + g^{(1)}(s, t_1, t_2) \right) + o(\varepsilon), \\ r_\varepsilon^{(2)}(s, t_1, t_2) &= r_\varepsilon(s, t_1, t_2 + \varepsilon) + \\ &+ \varepsilon \cdot \left(\frac{r_\varepsilon}{t_2} + \frac{\partial r_\varepsilon}{\partial s} \cdot \frac{s}{t_2} + \frac{t_1^2}{2} \cdot \frac{\partial^2 r_\varepsilon}{\partial s^2} + g^{(2)}(s, t_1, t_2) \right) + o(\varepsilon). \end{aligned} \quad (4)$$

Проверим это для первого уравнения (4). Предполагая, что $r_\varepsilon(s, t_1, t_2)$ имеет непрерывные производные требуемых порядков и выполняя замену переменных $y = (s\varepsilon - t_1x)t_2^{-1}$, получим

$$r_{\varepsilon}^{(1)}(s, t_1, t_2) = \varepsilon g^{(1)}(s, t_1, t_2) + \\ + t_1^{-1} \int_{-\infty}^{+\infty} r_{\varepsilon} \left(s + \frac{s\varepsilon - t_2 y}{t_1}, t_1 + \varepsilon, t_2 \right) h_{\varepsilon}(y, t_1) dy.$$

Используя разложение в ряд Тейлора

$$r_{\varepsilon} \left(s + \frac{s\varepsilon - t_2 y}{t_1}, \cdot \right) = r_{\varepsilon}(s, \cdot) + \frac{\partial r_{\varepsilon}}{\partial s} \cdot \frac{s\varepsilon - t_2 y}{t_1} + \\ + \frac{1}{2} \cdot \frac{\partial^2 r_{\varepsilon}}{\partial s^2} \cdot \left(\frac{s\varepsilon - t_2 y}{t_1} \right)^2 + o \left(\frac{s\varepsilon - t_2 y}{t_1} \right)^2,$$

и учитывая, что

$$\int_{-\infty}^{+\infty} h_{\varepsilon}(x, t) dx = t + \varepsilon, \quad \int_{-\infty}^{+\infty} x h_{\varepsilon}(x, t) dx = 0, \quad \int_{-\infty}^{+\infty} x^2 h_{\varepsilon}(x, t) dx = t\varepsilon(t + \varepsilon)^2, \\ \int_{-\infty}^{+\infty} x^n h_{\varepsilon}(x, t) dx = o(\varepsilon)$$

при $n \geq 3$, получим

$$t_1^{-1} \int_{-\infty}^{+\infty} r_{\varepsilon} \left(s + \frac{s\varepsilon - t_2 y}{t_1}, t_1 + \varepsilon, t_2 \right) h_{\varepsilon}(y, t_1) dy = \\ = \frac{t_1 + \varepsilon}{t_1} \cdot \left(r_{\varepsilon}(s, \cdot) + \frac{\partial r_{\varepsilon}}{\partial s} \cdot \frac{s\varepsilon}{t_1} + \frac{\partial^2 r_{\varepsilon}}{\partial s^2} \cdot \frac{\varepsilon t_2^2 (t_1 + \varepsilon)}{2t_1} \right) + o(\varepsilon),$$

откуда легко следует справедливость первого уравнения (4). Справедливость второго уравнения (4) устанавливается аналогично. Продолжая рассуждения и предполагая, что можно переходить к пределу при $\varepsilon \rightarrow 0$, из (4) нетрудно получить дифференциальные уравнения

$$\frac{\partial r}{\partial t_1} + \frac{r}{t_1} + \frac{\partial r}{\partial s} \cdot \frac{s}{t_1} + \frac{t_2^2}{2} \cdot \frac{\partial^2 r}{\partial s^2} + g^{(1)}(s, t_1, t_2) = 0 \quad \text{при } (s, t_1, t_2) \in D_1,$$

$$\frac{\partial r}{\partial t_2} + \frac{r}{t_2} + \frac{\partial r}{\partial s} \cdot \frac{s}{t_2} + \frac{t_1^2}{2} \cdot \frac{\partial^2 r}{\partial s^2} + g^{(2)}(s, t_1, t_2) = 0 \quad \text{при } (s, t_1, t_2) \in D_2,$$

где D_1, D_2 — области, в которых следует выбирать первый и второй варианты соответственно. Традиционный подход предполагает, что далее следовало бы описать границу между этими областями и условия на функцию $r(s, t_1, t_2)$ на этой границе.

Вспомним, однако, что уравнения (4) следует дополнить уравнением (1), которое запишем в виде

$$\min_{\ell=1,2} (r_{\varepsilon}^{(\ell)}(s, t_1, t_2) - r_{\varepsilon}(s, t_1, t_2)) = 0,$$

и тогда дифференциальное уравнение примет вид

$$\min_{\ell=1,2} \left(\frac{\partial r}{\partial t_{\ell}} + \frac{r}{t_{\ell}} + \frac{s}{t_{\ell}} \cdot \frac{\partial r}{\partial s} + \frac{t_{\ell}^2}{2} \cdot \frac{\partial^2 r}{\partial s^2} + g^{(\ell)}(s, t_1, t_2) \right) = 0 \quad (5)$$

с начальными и граничными условиями

$$r(s, t_1, t_2) = 0 \quad \text{при } t_1 + t_2 = 1, \quad \lim_{s \rightarrow \infty} r(s, t_1, t_2) = 0$$

$$\text{при } t_1 + t_2 < 1, \quad t_1 \geq \varepsilon_0, \quad t_2 \geq \varepsilon_0.$$

Уравнение (5) описывает не только изменение функции $r(s, t_1, t_2)$, но и области D_1, D_2 , поскольку в области D_{ℓ} минимальным является ℓ -й член в левой части уравнения.

4. Численные эксперименты

Для численного моделирования решения дифференциального уравнения использовалось разностное уравнение

$$r(s, t_1, t_2) = \min(r^{(1)}(s, t_1, t_2), r^{(2)}(s, t_1, t_2)),$$

где $r^{(1)}(s, t_1, t_2) = r^{(2)}(s, t_1, t_2) = 0$ при $t_1 + t_2 = 1$,

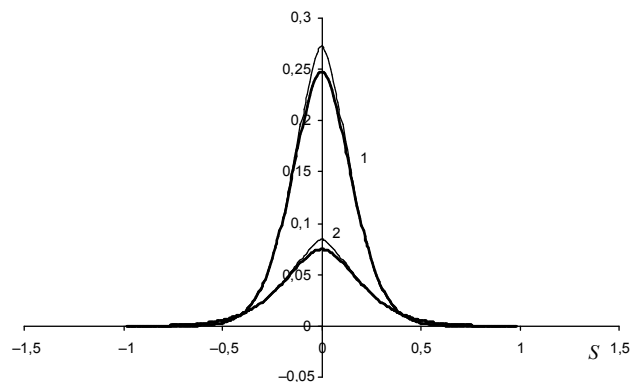
$$r^{(1)}(s, t_1, t_2) = r(s, t_1 + \varepsilon, t_2) + \\ + \varepsilon \cdot \left(\frac{r(s, t_1 + \varepsilon, t_2)}{t_1} + \frac{s}{t_1} \cdot Dr(s, t_1 + \varepsilon, t_2) + \right. \\ \left. + \frac{t_2^2}{2} \cdot D^2 r(s, t_1 + \varepsilon, t_2) + g^{(1)}(s, t_1, t_2) \right), \\ r^{(2)}(s, t_1, t_2) = r(s, t_1, t_2 + \varepsilon) + \\ + \varepsilon \cdot \left(\frac{r(s, t_1, t_2 + \varepsilon)}{t_2} + \frac{s}{t_2} \cdot Dr(s, t_1, t_2 + \varepsilon) + \right. \\ \left. + \frac{t_1^2}{2} \cdot D^2 r(s, t_1, t_2 + \varepsilon) + g^{(2)}(s, t_1, t_2) \right) \quad (6)$$

при $t_1 + t_2 < 1, \quad t_1 \geq \varepsilon_0, \quad t_2 \geq \varepsilon_0$. Здесь

$$Dr(s, t_1, t_2) = \frac{r(s + \delta, t_1, t_2) - r(s, t_1, t_2)}{\delta} \quad \text{при } s > 0,$$

$$Dr(s, t_1, t_2) = \frac{r(s, t_1, t_2) - r(s - \delta, t_1, t_2)}{\delta} \quad \text{при } s < 0,$$

$$D^2 r(s, t_1, t_2) = \frac{r(s + \delta, t_1, t_2) - 2r(s, t_1, t_2) + r(s - \delta, t_1, t_2)}{\delta^2}.$$



Сравнительные результаты вычислений, если $\rho(w) = 0,5$ при $w = \pm 1,7$

Предполагалось также, что $|s| < S_{\max}$. Отметим, что решение разностных уравнений часто приводит к неустойчивой вычислительной схеме. При этом предложенная схема оказывалась всегда устойчивой, если параметры $\varepsilon, \delta, \varepsilon_0, S_{\max}$ выбраны так, что после приведения подобных членов все коэффициенты в правых частях уравнений (6) являются положительными. После раскрытия выражений для $Dr(s, t_1, t_2), D^2 r(s, t_1, t_2)$ можно установить, что это условие всегда выполнено, если

$$\frac{S_{\max} \varepsilon}{\varepsilon_0 \delta} + \frac{\varepsilon}{\delta^2} < 1. \quad (7)$$

На рисунке приведены сравнительные результаты вычислений, если $\rho(w) = 0,5$ при $w = \pm 1,7$. Тонкими линиями даны значения $r_\varepsilon(s, \varepsilon_0, \varepsilon_0) = 0$, жирными — значения $r(s, \varepsilon_0, \varepsilon_0) = 0$. Линиям 1, 2 соответствуют $\varepsilon_0 = 0,3; 0,4$. Везде $S_{\max} = 0,975$; при вычислении $r(s, \varepsilon_0, \varepsilon_0) = 0$ выбирались $\varepsilon = 20000^{-1}$, $\delta = 0,0075$, удовлетворяющие (7); при вычислении $r_\varepsilon(s, \varepsilon_0, \varepsilon_0) = 0$ выбирались $\varepsilon = 0,05$, шаг интегрирования $0,00056$.

5. Заключение

Предложено предельное описание минимаксного параллельного управления в случайной среде с помощью дифференциального уравнения в частных производных второго порядка. Численные эксперименты показывают близость решений дифференциального уравнения и ранее полученного рекуррентного интегро-разностного уравнения. Однако требования устойчивости вычислительной схемы для дифференциального уравнения таковы, что существенного выигрыша по производительности вычислений достичь пока не удалось.

Работа выполнена при финансовой поддержке РФФИ, проект № 13-01-00334.

1. Колногоров А.В. Робастное параллельное управление в случайной среде (задача о двуруком бандите) // Автоматика и телемеханика. 2012. №4. С.114-130.
2. Lai T.L., Levin B., Robbins H., Siegmund D. Sequential Medical Trials (Stopping Rules/Asymptotic Optimality) // Proc. Nati. Acad. Sci. USA. 1980. V.77. №6. P.3135-3138.
3. Колногоров А.В. Задача о двуруком бандите для систем с параллельной обработкой данных // Проблемы передачи информации. 2012. Т.48. №1. С.83-95.
4. Vogel W. An Asymptotic Minimax Theorem for the Two-Armed Bandit Problem // Ann. Math. Stat. 1960. V.31. P.444-451.
5. Назин А.В., Позняк А.С. Адаптивный выбор вариантов. М.: Наука, 1986. 288 с.

Bibliography (Transliterated)

1. Kolnogorov A.V. Robastnoe parallel'noe upravlenie v sluchaynoy srede (zadache o dvurukom bandite) // Avtomatika i telemekhanika. 2012. №4. S.114-130.
2. Lai T.L., Levin B., Robbins H., Siegmund D. Sequential Medical Trials (Stopping Rules/Asymptotic Optimality) // Proc. Nati. Acad. Sci. USA. 1980. V.77. №6. P.3135-3138.
3. Kolnogorov A.V. Zadacha o dvurukom bandite dlya sistem s parallel'noy obrabotkoy dannykh // Problemy peredachi informatsii. 2012. T.48. №1. S.83-95.
4. Vogel W. An Asymptotic Minimax Theorem for the Two-Armed Bandit Problem // Ann. Math. Stat. 1960. V.31. P.444-451.
5. Nazin A.V., Poznyak A.S. Adaptivnyy vybor variantov. M.: Nauka, 1986. 288 s.