А. В. Колногоров (Великий Новгород, НовГУ)**. Робастное параллельное управление в задаче о двуруком бандите.**

Развиваются результаты [1, 2] для задачи о двуруком бандите в минимаксной, т. е. робастной постановке. Задача рассматривается на отрезке времени длины N, причем доходы ξ_n, $n = 1, 2, \ldots, N$, имеют нормальные распределения с плотностями $f(x|m_\ell) = (2\pi)^{-1/2} \exp\{-(x - m_\ell)^2/2\}$, где m_ℓ соответствует текущему выбранному варианту ($\ell = 1, 2$). Такой двурукий бандит характеризуется неизвестным векторным параметром $\theta = (m_1, m_2)$, для которого, однако, известно множество допустимых значений $\Theta = \{\theta\colon\ |m_1 - m_2| < 2cN^{-1/2}\}$, $0 < c < \infty$. Для управления применяется стратегия σ, которая в начале применяет каждый вариант по M_0 раз, а затем осуществляет оптимальное управление с тем ограничением, что варианты могут меняться только после применения M раз подряд. Отметим, что пакеты из M поступающих подряд данных, для обработки которых применяется один и тот же вариант, могут обрабатываться параллельно. Стратегия может использовать всю известную предысторию (X_1, n_1, X_2, n_2) к текущему моменту времени $n = n_1 + n_2$, где n_1, n_2 характеризуют суммарные применения обоих вариантов, а X_1, X_2 — соответствующие полные доходы.

В [1, 2] показано, что минимаксный риск в соответствии с основной теоремой теории игр может быть найден как байесовский для наихудшего априорного распределения, и приведены рекуррентные уравнения для его вычисления. Дадим рекуррентные уравнения для вычисления байесовского риска, соответсвующего стратегии параллельного управления в инвариантной форме, в которой полное время управления равно 1. Положим $S = ZN^{-3/2}$, $s = zN^{-3/2}$, $t_1 = n_1N^{-1}$, $t_2 = n_2N^{-1}$, $w = vN^{1/2}$, $\varepsilon = MN^{-1}$, $\varepsilon_0 = M_0N^{-1}$, $r_\varepsilon(S, t_1, t_2) = NR_{n_1,n_2}(Z)$, $r_\varepsilon^{(\ell)}(S, t_1, t_2) = NR_{n_1,n_2}^{(\ell)}(Z)$, $\varrho(w) = N^{-1/2}\rho(v)$.

Оптимальная стратегия на первых двух шагах применяет варианты по очереди. Далее она определяется рекуррентно «с конца», т. е. надо вычислять риски $r_\varepsilon(S, t_1, t_2) = \min(r_\varepsilon^{(1)}(S, t_1, t_2), r_\varepsilon^{(2)}(S, t_1, t_2))$. Текущим оптимальным является ℓ-й вариант, если меньшее значение имеет $r_\varepsilon^{(\ell)}(S, t_1, t_2)$. Здесь $r_\varepsilon^{(1)}(S, t_1, t_2) = r_\varepsilon^{(2)}(S, t_1, t_2) = 0$ при $t_1 + t_2 = 1$ и далее

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

$$r_\varepsilon^{(2)}(S, t_1, t_2) = \varepsilon g^{(2)}(S, t_1, t_2) + t_1^{-1} \int_{-\infty}^{+\infty} r_\varepsilon(S + s, t_1, t_2 + \varepsilon) h_\varepsilon\left(\frac{S\varepsilon - t_2 s}{t_1}, t_2\right) ds,$$

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

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

$$g(S, w, t_1, t_2) = (2\pi t_1 t_2 (t_1 + t_2))^{-1/2} \exp\left\{ - \frac{(S + 2wt_1t_2)^2}{2t_1t_2(t_1 + t_2)} \right\},$$

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

Далее фиксируем $\varepsilon_0 > 0$ и устремляем ε к нулю. Тогда при всех S и всех t_1, t_2, для которых определены решения уравнений, существуют пределы $r(S, t_1, t_2) = \lim_{\varepsilon\to 0} r_\varepsilon(S, t_1, t_2) = \lim_{\varepsilon\to 0} r_\varepsilon^{(\ell)}(S, t_1, t_2)$, $\ell = 1, 2$, удовлетворяющие условиям Липшица по всем переменным. Это позволяет доопределить $r(S, t_1, t_2)$ по непрерывности на все допустимые S, t_1, t_2. Для минимаксного риска в области $|m_1 - m_2| < 2cN^{-1/2}$

при $N \to \infty$ справедлива асимптотическая оценка

$$\sup_{\varrho} \int_{-\infty}^{\infty} r(s, \varepsilon_0, \varepsilon_0)\, ds \leqslant N^{-1/2} R_N^M(\Theta) \leqslant \sup_{\varrho} \left(4\varepsilon_0 \int_0^{\infty} w\varrho(w)\, dw + \int_{-\infty}^{\infty} r(s, \varepsilon_0, \varepsilon_0)\, ds \right).$$

СПИСОК ЛИТЕРАТУРЫ

1. *Колногоров А. В.* Нахождение минимаксных стратегии и риска в задаче о двуруком бандите с нормально распределенными доходами. — Обозрение прикл. и промышл. матем., 2010, т. 17, в. 2, с. 232–234.
2. *Колногоров А. В.* Нахождение минимаксных стратегии и риска в случайной среде (задаче о двуруком бандите). — Автомат, и телемех., 2011, № 5, с. 127–138.