ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 20 МАТЕМАТИКИ Выпуск 2
2013

А. В. Колногоров (Великий Новгород, НовГУ)**. Минимаксное управление в задаче о двуруком бандите с одним известным параметром.**

Рассматривается задача о двуруком бандите на конечном отрезке времени длины N. Предполагается, что текущие доходы ξ_n, $n = 1, 2, \ldots, N$, имеют нормальные распределения с плотностями $f(x|m_\ell) = (2\pi)^{-1/2} e^{-(x-m_\ell)^2/2}$, где m_ℓ соответствует текущему выбранному варианту ($\ell = 1, 2$). Значение m_1 является известным и без ограничения общности считаем $m_1 = 0$ (иначе можно рассматривать процесс $\xi_n - m_1$, $n = 1, 2, \ldots, N$). Такой двурукий бандит характеризуется векторным параметром $\theta = (0, m)$, для которого известно множество допустимых значений $\Theta = \{\theta = (0, m),\ |m| \leqslant C\}$, где $0 < C < \infty$. В соответствии с [1], оптимальная стратегия σ сначала применяет второй вариант, пока не выполнится некоторое условие остановки, а затем применяет первый вариант до конца управления. Пока идет применение второго варианта, эта стратегия зависит от статистики (X, n), где n характеризует число применений второго варианта, а X — соответствующий полный доход. Функция потерь, обусловленная неполнотой информации о процессе, имеет вид $L_N(\sigma, \theta) = E_{\sigma,\theta} \sum_{n=1}^{N} ((0 \vee m) - \xi_n)$. Байесовский и минимаксный риски определяются как

$$R_N^B(\lambda) = \min_{\{\sigma\}} \int_\Theta L_N(\sigma, \theta) \lambda(m)\, dm, \quad R_N^M(\Theta) = \min_{\{\sigma\}} \max_\Theta L_N(\sigma, \theta),$$

им соответствуют байесовская и минимаксная стратегии. Из основной теоремы теории игр следует, что минимаксные стратегия и риск совпадают с байесовскими на наихудшем априорном распределении, соответствующем максимуму байесовского риска.

Дадим уравнения для вычисления байесовских стратегии и риска. Байесовская стратегия определяется рекуррентно «с конца», т. е. надо вычислять риски $R_n(\lambda; X, n) = \min\{R_n^{(1)}(\lambda; X, n), R_n^{(2)}(\lambda; X, n)\}$. Текущим оптимальным является ℓ-й вариант, если $R_n^{(\ell)}(\cdot)$ имеет меньшее значение. Здесь $R_N^{(1)}(\lambda; X, N) = R_N^{(2)}(\lambda; X, N) = 0$ при $n_1 + n_2 = N$ и далее

$$R_n^{(1)}(\lambda; X, n) = (N - n) g_n^{(1)}(\lambda; X, n),$$

$$R_n^{(2)}(\lambda; X, n) = g_n^{(2)}(\lambda; X, n) + \int_{-\infty}^{+\infty} R_{n+1}(\lambda; X + Y, n + 1) h_n(X - nY)\, dY$$

при $0 \leqslant n < N$. Здесь

$$h_n(Y) = \left(\frac{n+1}{2\pi n}\right)^{1/2} \exp\left\{-\frac{Y^2}{2n(n+1)}\right\}, \quad n \geqslant 1, \quad h_0(Y) = 1,$$

$$g_n^{(1)}(\lambda; X, n) = \int_0^C m f_n(X - nm) \lambda(m)\, dm, \quad g_n^{(2)}(\lambda; X, n) = \int_0^C m f_n(X + nm) \lambda(-m)\, dm,$$

$$f_n(X) = (2\pi n)^{-1/2} \exp\{-X^2/(2n)\}.$$

Байесовский риск вычисляется по формуле $R_N^B(\lambda) = \min\{R_0^{(1)}(\lambda), R_0^{(2)}(\lambda)\}$.

Численная оптимизация проводилась в предположении, что наихудшее априорное распределение $\lambda(m)$, $m = vN^{-1/2}$, при достаточно больших N сосредоточено в двух точках: $v \approx d_1$ и $v \approx -d_2$ с вероятностями $\mathbf{P}\{v = d_1\} = \varrho$, $\mathbf{P}\{v = -d_2\} = 1 - \varrho$. В результате численной оптимизации $d_1 \approx 1,65$, $d_2 \approx 2,52$, $\varrho \approx 0,38$, $R_N^B(\lambda) \approx 0,37N^{1/2}$ и, следовательно, $R_N^M(\Theta) \approx 0,37N^{1/2}$.

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

СПИСОК ЛИТЕРАТУРЫ

1. *Bradt R. N.*, *Johnson S. M.*, *Karlin S.* On sequential designs for maximizing the sum of n observations. — Ann. Math. Statist., 1956, v. 27, p. 1060–1074.