УДК 519.6

Минимаксный подход к решению одной из задач о двуруком бандите в случайной среде с нормально распределенными доходами

А.Н. Лазутченко, А.В. Колногоров

Новгородский государственный университет имени Ярослава Мудрого

Рассматривается задача об оптимальном управлении в случайной среде. Случайная среда – это управляемый случайный процесс ξ_t ($t=\overline{1,T}$, T – горизонт управления), значения которого будем считать доходами, зависящими только от выбираемых в текущие моменты времени действий и имеющими нормальные распределения с плотностями $f_D(x\,|\,m_\ell)=(2\pi D)^{-1/2}\exp\{-(x-m_\ell)^2/(2D)\}$. Здесь $\ell=1,2$ – номер выбранного действия, m_ℓ – его математическое ожидание, D – дисперсия одношагового дохода. При такой постановке задачи случайная среда описывается вектором математических ожиданий $\theta=(m_1,m_2,D)$. Задана функция потерь $L_T(\sigma,\theta)$, значениями которой являются потери за время управления, вызванные неполнотой информации о системе, где σ – используемая стратегия. В данной постановке задачи параметр θ фиксирован, но неизвестен лицу, осуществляющему управление, поэтому могут возникать потери вследствие неполноты информации о среде, равные

$$L_T(\sigma,\theta)=\max(m_1,m_2)\cdot T-\mathrm{E}_{\sigma,\theta}\left(\sum_{t=1}^{T}\xi_t\right),$$

где $\mathrm{E}_{\sigma,\theta}$ – математическое ожидание потерь полного дохода. Ограничения на множество допустимых значений параметра θ имеют вид: $|m_1-m_2|\le 2c$, $D_0\le D\le 1$, где c – некоторая константа ($0<c<\infty$), $D_0>0$.

При использовании минимаксного подхода, предложенного, например, в [1], цель управления состоит в минимизации максимальных ожидаемых потерь полного дохода на множестве параметров Θ по множеству стратегий Σ. Минимаксный риск $R_T^M(\Theta)$ при этом задается как

$$R_T^M(\Theta)=\inf_{\Sigma}\sup_{\Theta}L_T(\sigma,\theta).$$

Данная цель была реализована, например, в [2], где рассматривается пороговая стратегия управления для случая нормально распределенных доходов, где введены

соответствующие параметры: пороговая константа α и параметр среды β, а также получены значения этих параметров. Более конкретно, найдена оптимальная стратегия σ при пороговой константе $\alpha = 0{,}55$, параметре среды $\beta = 4$ и вычислено значение минимакса $\min_{\alpha} \max_{\beta} L_T(\sigma, \theta)$, равное 0,761, при максимальной дисперсии $D = 1$.

Далее, мы предполагаем, что найденная выше стратегия σ является оптимальной для целого класса сред с дисперсиями $D_0 \le D \le 1$, $D_0 > 0$. Эта гипотеза была проверена с помощью численного моделирования методом Монте-Карло для множества значений $0 \le \beta \le 10$. На рис. 1 приведены некоторые результаты вычислений при конкретных дисперсиях. Видно, что на рассматриваемом множестве значений найденное при $D = 1$ значение минимакса, отмеченное точкой, является наибольшим.

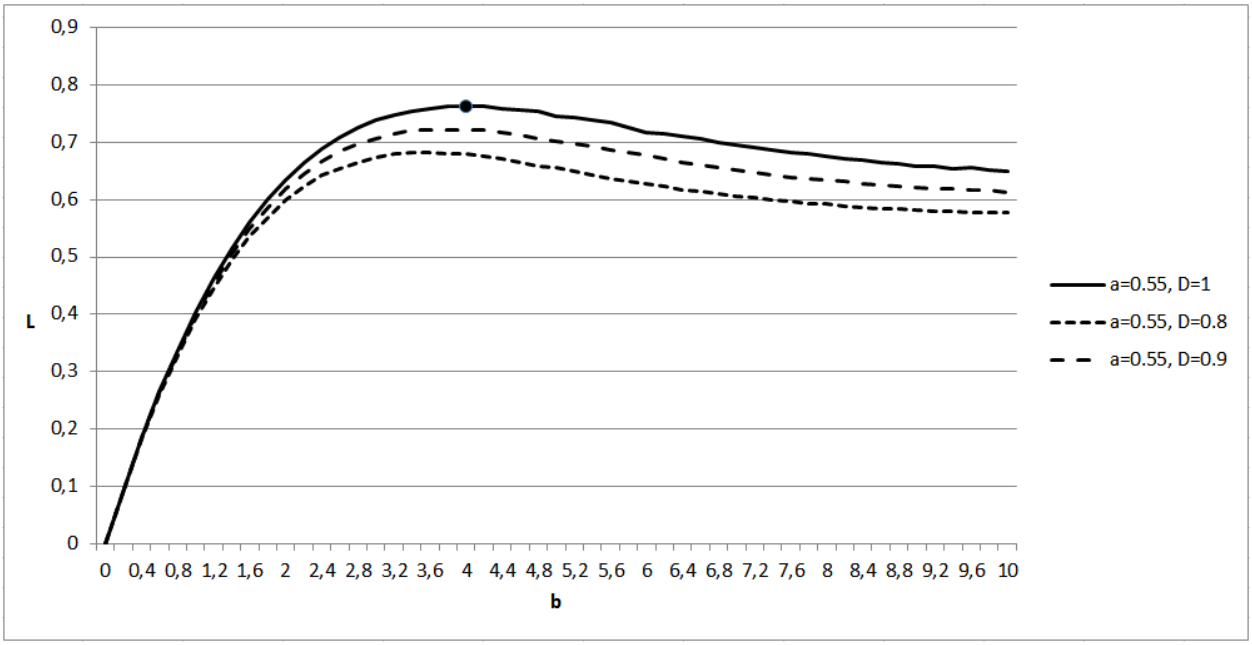

Рис. 1. Значения функции потерь, полученные методом Монте-Карло при различных дисперсиях

Литература

1. *Колногоров А. В.* Нахождение минимаксных стратегий и риска в случайной среде (задаче о двуруком бандите). – В. Новгород, Автоматика и телемеханика. – 2011. – №5. – С. 127-138.
2. *Лазутченко А. Н.* Использование двухпороговой стратегии управления в случайной среде с нормально распределенными доходами // Современные проблемы науки и образования. – 2014. – № 2; URL: www.science-education.ru/116-12590 (дата обращения: 12.10.2015).