

УДК 519.6

О РОБАСТНОМ УПРАВЛЕНИИ В СЛУЧАЙНОЙ СРЕДЕ, ХАРАКТЕРИЗУЕМОЙ НОРМАЛЬНЫМ РАСПРЕДЕЛЕНИЕМ ДОХОДОВ С РАЗЛИЧНЫМИ ДИСПЕРСИЯМИ

А. Н. Лазутченко

Новгородский государственный университет им. Ярослава Мудрого

Рассмотрена задача об оптимальном управлении в случайной среде (задача о двуруком бандите). Получены уравнения для вычисления байесовских рисков и потерь для доходов с различными дисперсиями. Представлены результаты вычислений. Результаты проверены с помощью моделирования методом Монте-Карло.

Ключевые слова: робастное управление; случайная среда; задача о двуруком бандите.

A. N. Lazutchenko. ROBUST CONTROL IN A RANDOM ENVIRONMENT CHARACTERIZED BY A NORMAL DISTRIBUTION OF INCOMES WITH DIFFERENT VARIANCES

The problem of optimal control in a random environment (also known as the two-armed bandit problem) is considered. The equations for calculating Bayes risks and losses are obtained for incomes with different variances. The results of the calculations are given. The results were verified by Monte-Carlo simulation.

Key words: robust control; random environment; two-armed bandit problem.

ВВЕДЕНИЕ

Рассматривается задача об оптимальном управлении в случайной среде с нормально распределенными доходами, также известная как задача о двуруком бандите [5]. Такие среды возникают при использовании групповой обработки данных, если для обработки можно использовать два альтернативных метода. В этом случае одинаковый метод применяется к группам данных, а реакцией среды является количество успешно обработанных данных. В силу центральной предельной теоремы оно имеет приблизительно нормальное распределение.

Итак, случайная среда – это управляемый случайный процесс ξ_t , ($t = \overline{1, T}$, T – горизонт управления), значения которого интерпретируются как доходы, зависящие только от выбираемых в текущие моменты времени действий и имеющие нормальные распределения с плотностями

$$f_D(x|m_\ell) = (2\pi D)^{-1/2} \exp\{-(x - m_\ell)^2/(2D)\},$$

где ℓ – номер выбранного действия ($\ell = 1, 2$), m_ℓ – его математическое ожидание, D – дисперсия одношагового дохода. При такой постановке задачи случайная среда описывается вектором математических ожиданий $\theta =$

(m_1, m_2) , а также дисперсией D . В нашем случае параметр θ фиксирован, но неизвестен лицу, осуществляющему управление. Ограничение на множество рассматриваемых дисперсий имеет вид $D_0 \leq D \leq 1$, $D_0 > 0$.

Введем функцию потерь $L_T(\sigma, \theta)$, значениями которой являются потери за время управления, вызванные неполнотой информации о системе, где σ – используемая стратегия. Если бы θ был известен, то наилучшей стратегией являлась бы та, которая применяла бы то действие, которому соответствовала большая из величин m_1 и m_2 , и ожидаемый полный доход в этом случае был бы равен $\max(m_1, m_2) \cdot T$. Но в нашей постановке θ неизвестен, поэтому возникают потери вследствие неполноты информации о среде, равные

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

где $\mathbb{E}_{\sigma, \theta}$ – математическое ожидание потерь полного дохода. Предполагается, что ограничения на множество допустимых значений параметра θ имеют вид $\Theta = \{(m_1, m_2) : |m_1 - m_2| \leq 2c\}$, где c – некоторая константа ($0 < c < \infty$).

При использовании минимаксного подхода, предложенного, например, в [6], цель управления состоит в минимизации максимальных ожидаемых потерь полного дохода на множестве параметров Θ по множеству стратегий Σ . При этом минимаксный риск $R_T^M(\Theta)$ выглядит следующим образом:

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

Для реализации этой цели в работе немецкого математика В. Фогеля [7] была предложена пороговая стратегия, где была получена неухудшаемая оценка минимаксного риска $0,530 \leq (DT)^{-1/2} R_T(\Theta) \leq 0,752$, где T – горизонт управления, D – максимальная дисперсия дохода.

Хотя в [7] рассматриваются бинарные случайные среды, оценки легко переносятся на случай сред с нормально распределенными доходами. Такие среды используются, в частности, в [4], где рассматривается двухпороговая стратегия, которая позволяет улучшить качество управления за счет уменьшения средних потерь дохода. Ниже рассматривается минимаксное (а следовательно, робастное) управление, определяемое с использованием основной теоремы теории игр. Робастность минимаксного подхода означает выполнение некоторого нужного нам свойства при всех допустимых значениях параметров. В нашем

случае использование минимаксной стратегии σ обеспечивает ограниченность потерь $L_T(\sigma, \theta) \leq R_T(\Theta)$ при всех допустимых значениях параметра θ . Минимаксные стратегии и риск ищутся как байесовские, соответствующие наихудшему априорному распределению. Эта задача в случае равных единичных дисперсий доходов рассмотрена в [2], [3]. В данной работе этот подход обобщается на случай, когда дисперсии доходов D на действиях попарно одинаковы и могут принимать различные значения из множества $D_0 \leq D \leq 1$, $D_0 > 0$. Для этого случая выведены формулы для вычисления байесовских рисков и потерь, а также вычислены соответствующие значения. Показано, что в этом случае управление остается робастным, т. е. гарантируется ограниченность потерь на всем множестве Θ .

Вывод уравнений для доходов с дисперсиями, отличными от единичных

В [2] установлена связь между минимаксным и байесовским подходом к решению рассматриваемой задачи. А именно, предлагается воспользоваться основной теоремой теории игр, согласно которой минимаксную стратегию и риск можно искать как байесовские, соответствующие наихудшему априорному распределению. Там же указано, что такое распределение может быть выбрано симметрическим и асимптотически однородным.

Итак, пусть $f_D(x|M) = (2\pi D)^{-1/2} \exp\{-(x-M)^2/(2D)\}$, – плотность нормального распределения с математическим ожиданием M и дисперсией D , $\lambda(m_1, m_2)$ – плотность априорного распределения на множестве параметров Θ , t_1, t_2 – количества применений обоих действий $\ell = 1$ и $\ell = 2$ соответственно ($t_1 + t_2 = t$), X_1, X_2 – доходы на действиях, $R_{T-t}^B(\lambda; X_1, t_1, X_2, t_2)$ – байесовский риск, рассчитываемый для дисперсии $D = 1$ относительно апостериорного распределения с плотностью λ :

$$\begin{aligned} & \lambda(m_1, m_2 | X_1, t_1, X_2, t_2) \\ &= \frac{f_{t_1}(X_1 | t_1 m_1) f_{t_2}(X_2 | t_2 m_2) \lambda(m_1, m_2)}{\iint_{\Theta} f_{t_1}(X_1 | t_1 m_1) f_{t_2}(X_2 | t_2 m_2) \lambda(m_1, m_2) dm_1 dm_2}. \end{aligned} \quad (1)$$

Пусть $x^+ = \max(x, 0)$. Тогда

$$R_{T-t}^B(\cdot) = \min(R_{T-t}^{(1)}(\cdot), R_{T-t}^{(2)}(\cdot)), \quad (2)$$

где $R_0^{(1)}(\cdot) = R_0^{(2)}(\cdot) = 0$,

$$\begin{aligned} & R_{T-t}^{(1)}(\lambda; X_1, t_1, X_2, t_2) \\ &= \int_{\Theta} \int \lambda(m_1, m_2 | X_1, t_1, X_2, t_2) \left[(m_2 - m_1)^+ \right. \\ & \left. + \mathbb{E}_x^{(1)} R_{T-t-1}^{(1)}(\lambda; X_1 + 1, t_1 + 1, X_2, t_2) \right] dm_1 dm_2, \end{aligned} \quad (3)$$

$$\begin{aligned} & R_{T-t}^{(2)}(\lambda; X_1, t_1, X_2, t_2) \\ &= \int_{\Theta} \int \lambda(m_1, m_2 | X_1, t_1, X_2, t_2) \left[(m_1 - m_2)^+ \right. \\ & \left. + \mathbb{E}_x^{(1)} R_{T-t-1}^{(2)}(\lambda; X_1, t_1, X_2 + 1, t_2 + 1) \right] dm_1 dm_2, \end{aligned} \quad (4)$$

$$\mathbb{E}_x^{(\ell)} R(x) = \int_{-\infty}^{+\infty} R(x) f(x | m_{\ell}) dx, \quad \ell = 1, 2. \quad (5)$$

Байесовская стратегия предписывает выбирать то действие, которому соответствует меньшее из значений $R_{T-t}^{(\ell)}(\cdot)$ и любое при их равенстве. На первых двух шагах действия следует применить по очереди.

По аналогии с (2)–(5) обозначим через $\tilde{R}_{T-t}^B(\lambda; X_1, t_1, X_2, t_2)$ байесовский риск, рассчитанный для доходов с дисперсиями, отличными от единицы, относительно апостериорного распределения с плотностью $\tilde{\lambda}$. Справедлива следующая теорема.

Теорема 1. Байесовский риск \tilde{R}_{T-t}^B , рассчитанный для доходов с дисперсиями $D_0 \leq D \leq 1$, $D_0 > 0$, связан с байесовским риском R_{T-t}^B , рассчитанным для доходов с $D = 1$ следующим соотношением:

$$\tilde{R}_{T-t}^B(\lambda) = \sqrt{D} R_{T-t}^B(\tilde{\lambda}). \quad (6)$$

Доказательство. Плотность f_D в нашем случае перепишется как:

$$\begin{aligned} f_{Dt_{\ell}}(X_{\ell} | t_{\ell} m_{\ell}) &= \frac{1}{\sqrt{2\pi D t_{\ell}}} \exp \left(-\frac{(X_{\ell} - t_{\ell} m_{\ell})^2}{2 D t_{\ell}} \right) \\ &= \frac{1}{\sqrt{D}} \frac{1}{\sqrt{2\pi t_{\ell}}} \exp \left(-\frac{\left(\frac{X_{\ell}}{\sqrt{D}} - t_{\ell} \frac{m_{\ell}}{\sqrt{D}} \right)^2}{2 t_{\ell}} \right) \\ &= \frac{1}{\sqrt{D}} \frac{1}{\sqrt{2\pi t_{\ell}}} \exp \left(-\frac{(\tilde{X}_{\ell} - t_{\ell} \tilde{m}_{\ell})^2}{2 t_{\ell}} \right) \\ &= \frac{1}{\sqrt{D}} f_{t_{\ell}}(\tilde{X}_{\ell} | t_{\ell} \tilde{m}_{\ell}), \end{aligned}$$

где $\tilde{X}_{\ell} = \frac{X_{\ell}}{\sqrt{D}}, \tilde{m}_{\ell} = \frac{m_{\ell}}{\sqrt{D}}, \quad \ell = 1, 2.$

С учетом этого плотность апостериорного распределения λ (1) перепишется так:

$$\begin{aligned} & \tilde{\lambda}(m_1, m_2 | X_1, t_1, X_2, t_2) = \\ & \frac{1}{D} \lambda(\tilde{m}_1, \tilde{m}_2 | \tilde{X}_1, t_1, \tilde{X}_2, t_2). \end{aligned}$$

При этом байесовский риск \tilde{R}_{T-t}^B равен:

$$\begin{aligned} \tilde{R}_{T-t}^B(\lambda; X_1, t_1, X_2, t_2) &= \min(\tilde{R}_{T-t}^{(1)}(\cdot), \tilde{R}_{T-t}^{(2)}(\cdot)), \\ \tilde{R}_0^{(1)}(\cdot) &= \tilde{R}_0^{(2)}(\cdot) = 0, \end{aligned}$$

$$\begin{aligned} & \tilde{R}_{T-t}^{(1)}(\lambda; X_1, t_1, X_2, t_2) \\ &= \int_{\Theta} \int \tilde{\lambda}(m_1, m_2 | X_1, t_1, X_2, t_2) \left[(m_2 - m_1)^+ \right. \\ & \left. + \tilde{\mathbb{E}}_x^{(1)} \tilde{R}_{T-t-1}^{(1)}(\lambda; X_1 + 1, t_1 + 1, X_2, t_2) \right] dm_1 dm_2, \\ & \tilde{R}_{T-t}^{(2)}(\lambda; X_1, t_1, X_2, t_2) \\ &= \int_{\Theta} \int \tilde{\lambda}(m_1, m_2 | X_1, t_1, X_2, t_2) \left[(m_1 - m_2)^+ \right. \\ & \left. + \tilde{\mathbb{E}}_x^{(2)} \tilde{R}_{T-t-1}^{(2)}(\lambda; X_1, t_1, X_2 + 1, t_2 + 1) \right] dm_1 dm_2, \\ & \tilde{\mathbb{E}}_x^{(\ell)} \tilde{R}(x) = \sqrt{D} \mathbb{E}_x^{(\ell)} R(\tilde{x}), \quad \ell = 1, 2. \end{aligned}$$

Преобразуем формулы:

$$\begin{aligned} & \tilde{R}_{T-t}^{(1)}(\lambda; X_1, t_1, X_2, t_2) \\ &= \int_{\Theta} \int \frac{1}{D} \lambda(\tilde{m}_1, \tilde{m}_2 | \tilde{X}_1, t_1, \tilde{X}_2, t_2) \\ & \quad \times \left[\sqrt{D} \left(\frac{m_2 - m_1}{\sqrt{D}} \right)^+ \right. \\ & \quad \left. + \sqrt{D} \mathbb{E}_{\tilde{x}}^{(1)} R_{T-t-1}^{(1)}(\tilde{\lambda}; \tilde{X}_1 + 1, t_1 + 1, \tilde{X}_2, t_2) \right] \\ & \quad \times \sqrt{D} \sqrt{D} d\tilde{m}_1 d\tilde{m}_2 \\ &= \sqrt{D} \int_{\Theta} \int \lambda(\tilde{m}_1, \tilde{m}_2 | \tilde{X}_1, t_1, \tilde{X}_2, t_2) \\ & \quad \times \left[(\tilde{m}_2 - \tilde{m}_1)^+ \right. \\ & \quad \left. + \mathbb{E}_{\tilde{x}}^{(1)} R_{T-t-1}^{(1)}(\tilde{\lambda}; \tilde{X}_1 + 1, t_1 + 1, \tilde{X}_2, t_2) \right] d\tilde{m}_1 d\tilde{m}_2 \\ &= \sqrt{D} R_{T-t}^{(1)}(\tilde{\lambda}; \tilde{X}_1, t_1, \tilde{X}_2, t_2). \end{aligned}$$

Аналогично для $\tilde{R}_{T-t}^{(2)}$:

$$\tilde{R}_{T-t}^{(2)}(\lambda; X_1, t_1, X_2, t_2) = \sqrt{D} R_{T-t}^{(2)}(\tilde{\lambda}; \tilde{X}_1, t_1, \tilde{X}_2, t_2).$$

В итоге приходим к следующему соотношению:

$$\tilde{R}_{T-t}^B(\lambda) = \sqrt{D} R_{T-t}^B(\tilde{\lambda}).$$

□

Формула (6) позволяет нам вычислять байесовские риски при дисперсиях, отличных от единицы, используя формулы, выведенные в [2].

ЧИСЛЕННАЯ ОПТИМИЗАЦИЯ

Приведем окончательные формулы для вычисления рисков и потерь, учитывая (6). При этом, как и в [2], предполагается, что асимптотически наихудшая плотность $\rho(\tilde{v})$ вырождена и сосредоточена в двух точках $\tilde{v} = \pm \tilde{d} T^{-1/2}$ с вероятностями $1/2$, где $\tilde{d} = d/\sqrt{D}$, $d = |m_1 - m_2|$, m_1, m_2 – математические ожидания доходов на действиях $\ell = 1, 2$:

$$R_{t_1, t_2}(Z) = \min(R_{t_1, t_2}^{(1)}(Z), R_{t_1, t_2}^{(2)}(Z)), \quad (7)$$

где $R_{t_1, t_2}^{(1)}(Z) = R_{t_1, t_2}^{(2)}(Z)$ при $t_1 + t_2 = T$, $Z = X_1 t_2 - X_2 t_1$.

$$R_{t_1, t_2}^{(1)}(Z) = v g_{t_1, t_2}(Z, v) + \frac{1}{t_2} \int_{-\infty}^{+\infty} R_{t_1+1, t_2}(Z+z) h_{t_1} \left(\frac{Z-t_1 z}{t_2} \right) dz, \quad (8)$$

$$R_{t_1, t_2}^{(2)}(Z) = v g_{t_1, t_2}(Z, -v) + \frac{1}{t_1} \int_{-\infty}^{+\infty} R_{t_1, t_2+1}(Z+z) h_{t_2} \left(\frac{Z-t_2 z}{t_1} \right) dz \quad (9)$$

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

$$g_{t_1, t_2} = \frac{1}{(2\pi t_1 t_2 (t_1 + t_2))^{1/2}} \times \exp \left(-\frac{(Z + 2v t_1 t_2)^2}{2 t_1 t_2 (t_1 + t_2)} \right), \quad (10)$$

$$h_t(z) = \left(\frac{t+1}{2\pi t} \right)^{1/2} \exp \left(-\frac{z^2}{2t(t+1)} \right). \quad (11)$$

Наихудшее априорное распределение определяется как $\nu_a(u, v) = k_a(u) \rho(v)$, где $k_a(u)$ – постоянная плотность на отрезке $|u| \leq a$, $\rho(v)$ – симметрическая плотность ($\rho(v) = \rho(-v)$) и $a \rightarrow \infty$. При этом байесовский риск относительно него вычисляется по формуле

$$\lim_{a \rightarrow \infty} R_T^B(\nu_a(u, v)) = \sqrt{D} \left(2v + \int_{-\infty}^{+\infty} R_{1,1}(z) dz \right). \quad (12)$$

Оптимальная стратегия на первых двух шагах применяет оба действия по очереди. Далее, для каждой текущей группы чисел (Z, t_1, t_2) выбирается действие, которому соответствует меньшее значение из значений $R_{t_1, t_2}^{(\ell)}(Z), \ell = 1, 2$.

Далее рассматриваем класс стратегий вида $\{\sigma_\ell(Z, t_1, t_2)\}$, которые осуществляют выбор действий на основе текущей статистики (Z, t_1, t_2) . При этом $\sigma_\ell(Z, t_1, t_2) = 1$, если для статистики (Z, t_1, t_2) байесовская стратегия предписывает выбирать ℓ -е действие, и $\sigma_\ell(Z, t_1, t_2) = 0$ в противном случае. При этом для распределения $\nu_a(u, v)$ потери находятся по формуле

$$L_{t_1, t_2}(Z) = \sigma_1(Z, t_1, t_2) L_{t_1, t_2}^{(1)}(Z) + \sigma_2(Z, t_1, t_2) L_{t_1, t_2}^{(2)}(Z), \quad (13)$$

где $L_{t_1, t_2}^{(1)}(Z) = L_{t_1, t_2}^{(2)}(Z) = 0$ при $t_1 + t_2 = T$,

$$L_{t_1, t_2}^{(1)}(Z) = v g_{t_1, t_2}(Z, v) + \frac{1}{t_2} \int_{-\infty}^{+\infty} L_{t_1+1, t_2}(Z+z) h_{t_1} \left(\frac{Z-t_1 z}{t_2} \right) dz, \quad (14)$$

$$L_{t_1, t_2}^{(2)}(Z) = v g_{t_1, t_2}(Z, -v) + \frac{1}{t_1} \int_{-\infty}^{+\infty} L_{t_1, t_2+1}(Z+z) h_{t_2} \left(\frac{Z-t_2 z}{t_1} \right) dz \quad (15)$$

при $t_1 + t_2 < T, t_1 \geq 1, t_2 \geq 1$. Потери при этом равны

$$\lim_{a \rightarrow \infty} L_T(\nu_a(u, v)) = \sqrt{D} \left(2v + \int_{-\infty}^{+\infty} L_{1,1}(z) dz \right). \quad (16)$$

Для нахождения результатов используют выведенные выше формулы рисков (7)–(12)

и потерь (13)–(16). Наихудшее распределение соответствует максимуму приведенного байесовского риска $r_T(\tilde{d}) = T^{1/2} \tilde{R}_T^B(\cdot)$. Затем для найденной в результате вычисления этого риска стратегии для различных дисперсий вычисляются приведенные потери $l_T(\tilde{d}) = T^{1/2} \tilde{L}_T(\cdot)$. На рис. 1 приведены графики значений функции рисков $r_T(\tilde{d})$ и потерь $l_T(\tilde{d})$, вычисленные при $T = 50$, для $0 \leq \tilde{d} \leq 20$ с шагом 0,1 для

дисперсий $0,2 \leq D \leq 1$ с шагом 0,2. Введены следующие обозначения: r, l – риски и потери, D – дисперсии. Видно, что максимумы $r_{50}(\tilde{d})$ равны приблизительно 0,29, 0,41, 0,5, 0,58, 0,65 при $d = 0,7, 1, 1,3, 1,5, 1,7$, а максимумы $l_{50}(\tilde{d})$ равны приблизительно 0,5, 0,43, 0,51, 0,58, 0,65 при $d = 7,4, 1,2, 1,3, 1,5, 1,7$ при $D = 0,2, 0,4, 0,6, 0,8, 1$ соответственно.

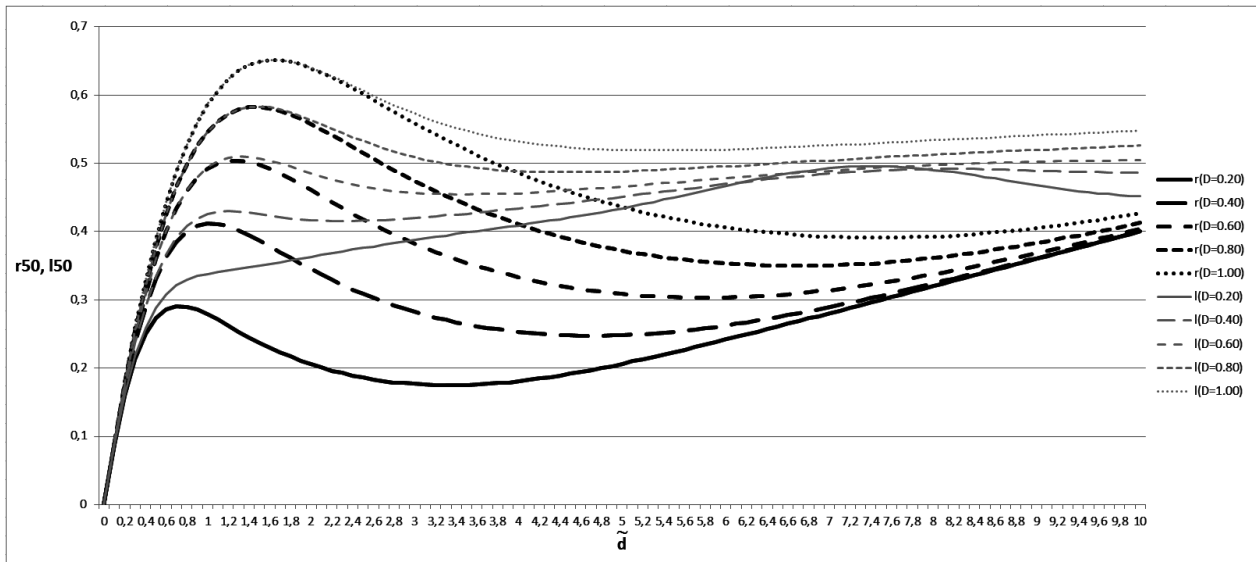


Рис. 1. Риски и потери при $0,2 \leq D \leq 1$

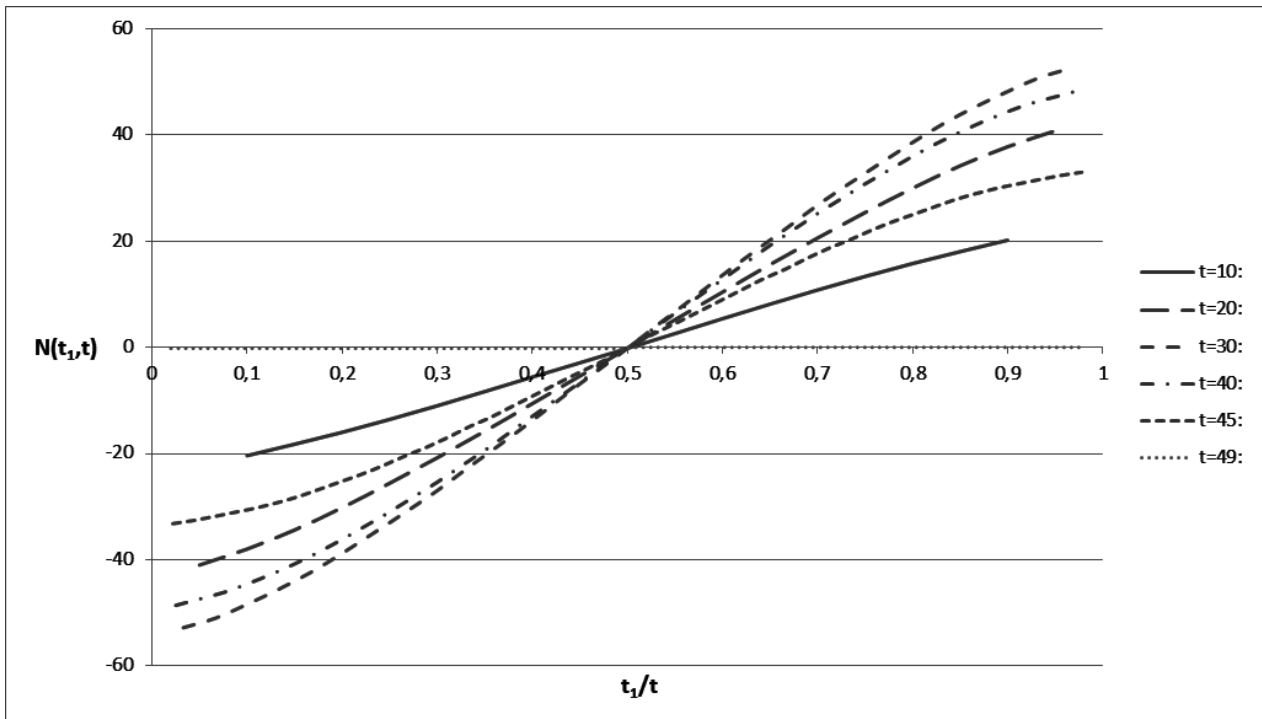


Рис. 2. Стратегия при $D = 0,4$

Далее, в результате вычислений были найдены стратегии для $D = 0,2, 0,4, 0,6, 0,8, 1$, предписывающие выбирать действие $\ell = 1$, если $Z > \frac{N(t_1, t_2)}{\sqrt{D}}$, и $\ell = 2$, если $Z \leq \frac{N(t_1, t_2)}{\sqrt{D}}$, где $N(t_1, t_2)$ – функции, определяющие выбор действий на конкретном шаге, $t = t_1 + t_2$. При $D = 0,4$ такие функции изображены на рис. 2.

Для этого случая найденная стратегия оказывается робастной для $0 \leq \tilde{d} < 2,5$. При $\tilde{d} > 2,5$ на величину потерь начинает оказывать сильное влияние применение неоптимального действия на начальном этапе хотя бы раз. Для увеличения области действия оптимальной стратегии следует увеличить горизонт управления T .

МОДЕЛИРОВАНИЕ МЕТОДОМ МОНТЕ-КАРЛО

Можно отметить, что все результаты вычислений, полученные выше при $D = 1$, полностью совпадают с [2]. Однако поскольку полученные в данной работе результаты являются новыми, их можно проверить с помощью моделирования методом Монте-Карло [1]. Как показано, например, в [3], управление в рассматриваемых случайных средах тесно связано с возможностью параллельной обработки данных. В нашем случае моделируемые дан-

ные можно обрабатывать параллельно практически без увеличения потерь.

Теперь сделаем замечание. Точность методов численного моделирования, в частности метода Монте-Карло, зависит от количества проведенных испытаний. Под точностью метода в данном случае подразумевается численная разница между величинами потерь, полученных по формулам и при использовании метода Монте-Карло. Эта разница тем меньше, чем больше количество испытаний. Разумеется, при численном моделировании мы вынуждены ограничиться какой-то величиной количества испытаний. В данном случае количество моделирований бралось равным 1 000 000, что позволяет говорить о точности по крайней мере двух значащих цифр после запятой.

Итак, рассчитаем значения функции потерь $l_T(\tilde{d})$ при $T = 50$ для $0 \leq \tilde{d} \leq 10$ с шагом 0,1 и для дисперсий $0,2 \leq D \leq 1$ с шагом 0,4. На рис. 3 на одном графике объединены результаты вычислений методом Монте-Карло, а также по формулам (13)–(16). Видно, что для каждой дисперсии линия функции потерь, рассчитанная методом Монте-Карло, повторяет поведение линии функции потерь, вычисленной по формулам. Отметим, что при увеличении горизонта управления результаты становятся точнее.

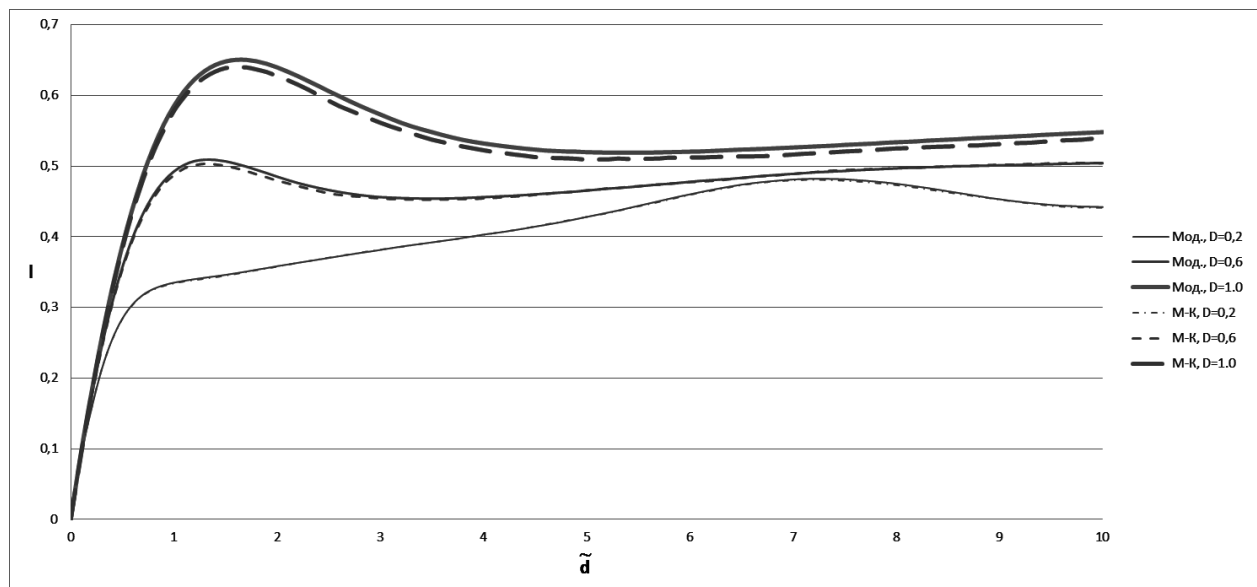


Рис. 3. Сравнение значений функции потерь, вычисленных по формулам и методом Монте-Карло

ЗАКЛЮЧЕНИЕ

В работе рассмотрена задача о робастном управлении в случайной среде с нормально

распределенными доходами. Выведены формулы вычисления рисков и потерь асимптотически наилучшего априорного распределения для случая, когда дисперсии на действи-

ях среды попарно равны и принимают различные значения. Используя эти формулы, вычислены значения байесовских рисков и потеря, установлена связь между ними и значениями, полученными ранее. Также результаты проверены моделированием методом Монте-Карло.

В дальнейшем задачу можно обобщить на случай, когда дисперсии доходов на действиях не равны и могут принимать попарно различные значения из некоторого диапазона.

Автор выражает благодарность профессору А. В. Колногорову за постановку задачи и обсуждение полученных результатов.

ЛИТЕРАТУРА

1. *Бусленко Н. П.* [и др.] Метод статистических испытаний (Метод Монте-Карло) М.: Физматгиз, 1962. 331 с.

REFERENCES

1. *Buslenko N. P.* [i dr.] Metod statisticheskikh ispytaniy (Metod Monte-Karlo) [The method of statistical tests (Monte Carlo method)]. Moscow: Fizmatgiz, 1962. 331 p.
2. *Kolnogorov A. V.* Nakhozhdenie minimaksnykh strategii i riska v sluchainoi srede (zadacha o dvurukom bandite) [Finding minimax strategies and minimax risk in a random environment (the two-armed bandit problem)]. *AiT.* 2011. N 5. P. 127–138.
3. *Kolnogorov A. V.* Robastnoe parallel'noe upravlenie v sluchainoi srede (zadacha o dvurukom bandite) [Robust parallel control in a random environment (the two-armed bandit problem)]. *AiT.* 2012. N 4. P. 114–130.
4. *Lazutchenko A. N.* Ispol'zovanie dvukhpороgovoй strategii upravleniya v sluchainoi srede s normal'no

2. *Колногоров А. В.* Нахождение минимаксных стратегии и риска в случайной среде (задача о двуруком бандите) // *АиТ.* 2011. № 5. С. 127–138.

3. *Колногоров А. В.* Робастное параллельное управление в случайной среде (задача о двуруком бандите) // *АиТ.* 2012. № 4. С. 114–130.

4. *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: 05.12.2014)

5. *Berry D. A., Fristedt B.* Bandit problems. London; New York: Chapman and Hall, 1985. P. 275.

6. *Robbins H.* Some aspects of the sequential design of experiments // *Bulletin AMS.* 1952. Vol. 58, N 5. P. 527–535.

7. *Vogel W.* An asymptotic minimax theorem for the two-armed bandit problem // *Ann. Math. Statist.* 1960. Vol. 31. P. 444–451.

Поступила в редакцию 17.04.2015

raspredeleennyimi dokhodami [Using the two threshold management strategy in a stochastic environment with normally distributed incomes]. *Sovremennye problemy nauki i obrazovaniya* [Modern problems of science and education]. 2014. N 2. URL: www.science-education.ru/116-12590 (accessed: 05.12.2014)

5. *Berry D. A., Fristedt B.* Bandit problems. London; New York: Chapman and Hall, 1985. P. 275.

6. *Robbins H.* Some aspects of the sequential design of experiments. *Bulletin AMS.* 1952. Vol. 58, N 5. P. 527–535.

7. *Vogel W.* An asymptotic minimax theorem for the two-armed bandit problem. *Ann. Math. Statist.* 1960. Vol. 31. P. 444–451.

Received April 17, 2015

СВЕДЕНИЯ ОБ АВТОРЕ:

Лазутченко Алексей Николаевич
аспирант
Новгородский государственный университет
им. Ярослава Мудрого
ул. Большая Санкт-Петербургская, 41, Великий
Новгород, Новгородская область, Россия, 173000
эл. почта: aleskey@hotmail.ru
тел.: (8921) 198 52 31

CONTRIBUTOR:

Lazutchenko, Aleksey
Yaroslav-the-Wise Novgorod State University
41 Most St. Petersburg St., 173000
Velikiy Novgorod, Russia
e-mail: aleskey@hotmail.ru
tel.: (8921) 198 52 31