УДК 519.6

ИСПОЛЬЗОВАНИЕ ДВУХПОРОГОВОЙ СТРАТЕГИИ УПРАВЛЕНИЯ В БИНАРНОЙ СЛУЧАЙНОЙ СРЕДЕ

Лазутченко А.Н.[1]

[1]*ФГБОУ ВПО «Новгородский государственный университет имени Ярослава Мудрого», Великий Новгород, Россия, (173003, г. Великий Новгород, ул. Большая Санкт-Петербургская, 41), e-mail: aleskey@hotbox.ru*

Рассмотрена пороговая стратегия управления в случайной среде с двумя действиями с бинарными доходами. В данной постановке задачи вероятности выигрыша на действиях в процессе управления фиксированы, но неизвестны лицу, осуществляющему управление. Цель управления задана в минимаксной постановке и состоит в минимизации математического ожидания максимальных потерь полного дохода и понижении суммарных потерь на некотором множестве параметров среды. С помощью написанной компьютерной программы было проведено моделирование заданной случайной среды и найдено значение минимакса, а также параметры среды, при которых он был найден. Также среда была исследована на суммарные потери полного дохода в предположении, что значения множества параметров среды распределены равномерно, а пороговое значение фиксировано. После этого суммарные потери были вычислены для двух порогов, оптимальные значения которых были определены после полного исследования всех их допустимых значений. Как оказалось, введение дополнительного порога позволило на ранних этапах отбрасывать не самый оптимальный вариант раньше, что существенно улучшает итоговый доход. Таким образом, в работе показано, что с помощью двухпороговой стратегии управления в случайной среде можно значительно снизить суммарные потери полного дохода на некотором множестве параметров среды.

Ключевые слова: случайная среда, пороговая стратегия управления, бинарный доход, минимакс.

USING THE TWO THRESHOLD MANAGEMENT STRATEGY IN BINARY RANDOM ENVIRONMENT

Lazutchenko A.N.[1]

Yaroslav-the-Wise Novgorod State University, Velikiy Novgorod, Russia, (173003, Velikiy Novgorod, street B.Sankt-Peterburgskaya, 41), e-mail: aleskey@hotbox.ru

Considered the threshold control strategy in a random environment with two actions with binary income. In this formulation of the problem the probability of winning on the actions is fixed in the management, but unknown to the person performing the operation. The purpose of the control is given to the minimax formulation and is to minimize the expectation of the maximum losses of total income and lowering the total loss on a set of environmental parameters. With the help of a computer program was written to simulate the given random environment and found the value of minimax, as well as the parameters of the medium in which it was found. Also, the environment was assayed for total loss of total income, assuming that the values of the set parameters of the environment are distributed evenly, and the threshold value is fixed. Thereafter, total losses were calculated for the two thresholds, the optimal values of which were determined after a full analysis of their possible values. As it turned out, the consideration of additional threshold allowed in the early stages of the cast is not the best option before, which significantly improves the total income. Thus, we have shown that using the two threshold management strategy in a random environment can significantly reduce the total losses of total income on a set of environmental parameters.

Key words: random environment, the threshold management strategy, binary income, Minimax.

Введение

Случайная среда (однородный процесс с независимыми значениями в терминологии [4]) с бинарно распределенными доходами – это управляемый случайный процесс ξ_t, принимающий значения 0 и 1, интерпретируемые как текущие доходы и зависящие только от выбираемых в текущие моменты времени действий y_l, т.е.

$$P\{\xi_t = 1 \mid y_t = l\} = p_l,\ P\{\xi_t = 0 \mid y_t = l\} = q_l,$$
$$p_l + q_l = 1, \qquad (1)$$
$$l = 1,2;\ t = \overline{1,T}.$$

Такая среда описывается векторным параметром $\theta = (p_1, p_2)$. В данной постановке задачи параметр фиксирован, но неизвестен тому, кто управляет процессом.

Постановка задачи

Введем целевую функцию потерь $L_T(\theta, \sigma)$, значениями которой являются потери за время моделирования, где θ определяет вероятности выигрыша на действиях, σ – используемая стратегия. Если параметр θ известен, то наилучшей стратегией является та, которая применяет только то действие, которому соответствует большая из величин p_1, p_2, и максимальный полный доход в этом случае равен $\max(p_1, p_2) \cdot T$. Если же θ неизвестен, то неизбежно возникают потери вследствие неполноты информации о среде, равные:

$$L_T(\theta, \sigma) = \max(p_1, p_2) \cdot T - E_{\sigma,\theta}\left(\sum_{t=1}^{T} y_t\right). \qquad (2)$$

Здесь $E_{\sigma,\theta}$ представляет собой математическое ожидание потерь полного дохода. Предполагается, что ограничения на множество допустимых значений параметра θ имеют следующий вид:

$$\Theta : \varepsilon \le p_1 \le 1 - \varepsilon,\ \varepsilon \le p_2 \le 1 - \varepsilon,\ \varepsilon > 0. \qquad (3)$$

При использовании минимаксного подхода, предложенного, например, в [2], цель управления состоит в минимизации величины потерь полного дохода на множестве параметров Θ по множеству стратегий Σ. При этом минимаксный риск $R_T^M(\Theta)$ выглядит следующим образом:

$$R_T^M(\Theta) = \inf_{\Sigma} \sup_{\Theta} L_T(\sigma, \theta). \qquad (4)$$

Для реализации этой цели предлагается использовать пороговую стратегию, предложенную в [5].

Стратегия управления с одним порогом

Итак, рассмотрим пороговую стратегию σ. Она применяет действия y_1 и y_2 среды по очереди, накапливая доходы X_1 и X_2 соответственно. На каждом шаге вычисляется абсолютная разность доходов на действиях $|X_1 - X_2|$. Действия применяются до тех пор, пока эта величина не превысит порога $\alpha \cdot (D \cdot T)^{\frac{1}{2}}$, где T – полное время управления, α, D –

пороговая константа и дисперсия соответственно $\left(D = p \cdot (1-p)\right)$, или не истечет время управления. Если время управления не истекло, то действие, которому соответствует меньшая величина набранного дохода, исключается из рассмотрения, а оставшееся время применяется только другое действие.

Можно показать, что наибольшие потери полного дохода при достаточно больших T будут иметь место при

$$p_1 = 0,5,\ p_2 = p_1 \cdot \left(1 - \beta \cdot T^{-\frac{1}{2}}\right), \qquad (5)$$

где $0 \le \beta \le T^{\frac{1}{2}}$. Ограничения на T накладываются, исходя из свойства инвариантности функции потерь [3]. Очевидно, что в таком случае дисперсия D оказывается максимальной, т.е. этот случай представляет наибольший теоретический интерес для исследования.

На основе пороговой стратегии σ, рассмотренной выше, была разработана программа. Прежде сделаем замечание. Целевая функция потерь $L_T(\theta, \sigma)$, вообще говоря, зависит от параметров σ и θ. Но для расчетов нам удобнее полагать, что она зависит от α и β, где α – пороговая константа, использующаяся пороговой стратегией, β – параметр среды.

Итак, алгоритм работы программы построен таким образом, что в ней для каждой пары $(\alpha;\beta)$ вычисляется средний доход $MZ_T = \sum_{i=1}^{N} Z_{i,T} / N$, где $Z_{i,T}$ – доход за одно моделирование, N – количество моделирований. Затем вычисляются средние потери математического ожидания дохода $L_T(\alpha,\beta) = 0,5 \cdot T - MZ_T$. После этого при каждой константе α определяются минимальные потери $M_T(\beta) = \min\limits_{\alpha} L_T(\alpha,\beta)$. При каждой константе β подбираются максимальные потери $R_T(\alpha) = \max\limits_{\beta} L_T(\alpha,\beta)$. Точка, в которой $M_T(\beta) = R_T(\alpha)$, и является минимаксной точкой, в которой достигается минимальная гарантированная величина потерь полного дохода.

В результате вычислений выяснилось, что достаточно рассмотреть $\alpha \in (0;1)$, $\beta \in (1;10)$, поскольку предварительные значения $\alpha = 0,6$ и $\beta = 4$ оказались заключенным именно в этих интервалах. При этом максимальные потери $L_T(\alpha,\beta) = 0,374$. Более точные вычисления, достигающиеся за счет уменьшения шага изменения параметров, дают следующие результаты: $\alpha = 0,58$, $\beta = 3,9$, $L_T(\alpha,\beta) = 0,375$. Время моделирования выбиралось из условия $T = 10\ 000$, количество моделирований $N = 1\ 000\ 000$, что позволяет говорить о точности вычислений $\delta = 0,001$ [1].

Таблица 1 показывает результаты предварительных вычислений. Все потери в таблице являются приведенными путем деления на $T^{\frac{1}{2}}$. Желтым цветом обозначены локальные минимумы по α для каждого β, зеленым – локальные максимумы по β для каждого α, сиреневым – точка, в которой минимум по α равен максимуму по β.

Таблица 1 – Значения $L_T(\alpha,\beta)$, $\alpha \in (0;1)$, $\beta \in (1;10)$

$\beta \setminus \alpha$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
1	0,229	0,226	0,216	0,211	0,206	0,205	0,206	0,208	0,211	0,215
2	0,432	0,400	0,360	0,340	0,323	0,317	0,321	0,330	0,342	0,357
3	0,612	0,532	0,445	0,404	0,375	0,366	0,372	0,389	0,413	0,440
4	0,770	0,624	0,484	0,424	0,383	0,374	0,386	0,411	0,444	0,480
5	0,905	0,683	0,489	0,412	0,369	0,364	0,382	0,414	0,452	0,495
6	1,019	0,713	0,470	0,387	0,347	0,348	0,374	0,410	0,454	0,501
7	1,116	0,717	0,439	0,356	0,324	0,334	0,366	0,407	0,453	0,501
8	1,192	0,706	0,402	0,326	0,304	0,323	0,359	0,404	0,451	0,500
9	1,255	0,679	0,364	0,298	0,288	0,315	0,355	0,402	0,451	0,500
10	1,298	0,645	0,330	0,274	0,276	0,309	0,353	0,401	0,451	0,500

Двухпороговая стратегия управления

Итак, получены значения пороговой константы $\alpha = 0,58$ и параметра $\beta = 3,9$ при $T = 10\ 000$, для которых гарантированные потери полного дохода минимальны. Рассмотрим пример. Предположим, что имеется два беспроводных модема. С помощью них можно передавать данные, однако вероятности передачи данных без ошибок различаются. Необходимо выяснить, который из модемов работает более стабильно. Подразумевается, что вероятности p_1, p_2 успешной передачи данных на модемах не равны. При вероятностях успешной передачи данных $p_1 = 0,5$, $p_2 = 0,48$ у первого и второго модемов соответственно тестирование модемов следует проводить до тех пор, пока разница между количеством успешных передач достигнет величины 29.

Рассмотрим теперь потери на множестве допустимых параметров β для следующих двух случаев. На диаграмме линия 1 показывает минимальные потери для минимальных α для каждого β, линия 2 – потери для фиксированного $\alpha = 0,58$ (все потери также являются приведенными):

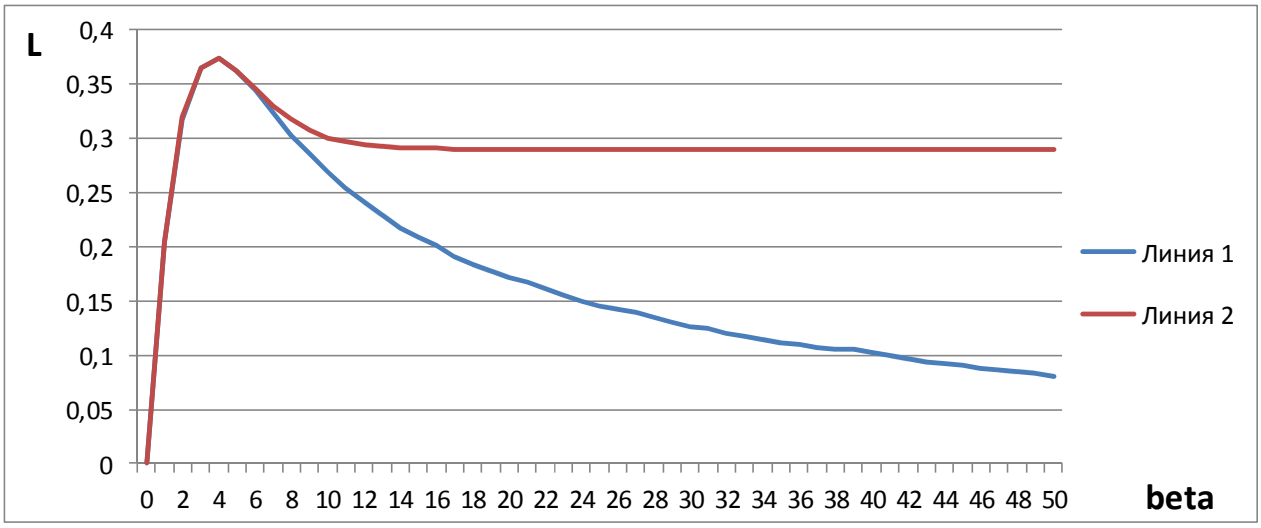

Рисунок 1 – Графики значений функции потерь на множестве значений β. Линия 1 показывает потери при минимальных α для каждого β, линия 2 – потери для фиксированного $\alpha = 0,58$.

Как нетрудно заметить, начиная с $\beta = 0$ и до $\beta \approx 7$ потери за моделирование практически идентичны. После этого потери для случая с фиксированным α увеличиваются и в определенный момент устанавливаются на величине порога $\alpha \cdot (D \cdot T)^{1/2}$. Очевидно, это связано с тем, что при больших значениях параметра β лучшее действие определяется достаточно быстро (так как, например, при $\beta = 10$ вероятности выигрыша на действиях $p_1 = 0,5,\ p_2 = 0,45$), но большое значение пороговой константы α (при $\alpha = 0,58$ порог $\alpha \cdot (D \cdot T)^{1/2} = 29$) не позволяет исключить из рассмотрения неоптимальное действие раньше, чем будет достигнут порог.

Итак, большие потери дохода для случая фиксированного α имеют место в силу того факта, что на достижение порога при больших β требуется фиксированное время. Этот недостаток стратегии можно устранить, если ввести дополнительный порог $\alpha_{th} < \alpha$, с помощью которого при больших β оптимальное действие будет определяться быстрее. В этом случае стратегия будет выглядеть следующим образом. Начиная с некоторого времени $t = T_{th}$ текущий порог α_{th} заменяется на α, после чего моделирование продолжается в обычном порядке.

Очевидно, что теперь при высоких значениях β неоптимальное действие будет в среднем исключаться из рассмотрения быстрее. Однако не совсем понятно, что будет происходить на этапе, где β мало. Попробуем взять тестовые параметры α_{th} и T_{th} и посмотрим на результаты. Возьмем, например, $\alpha_{th} = 0,3$, $T_{th} = 5000$. Полученные результаты отобразим на диаграмме:

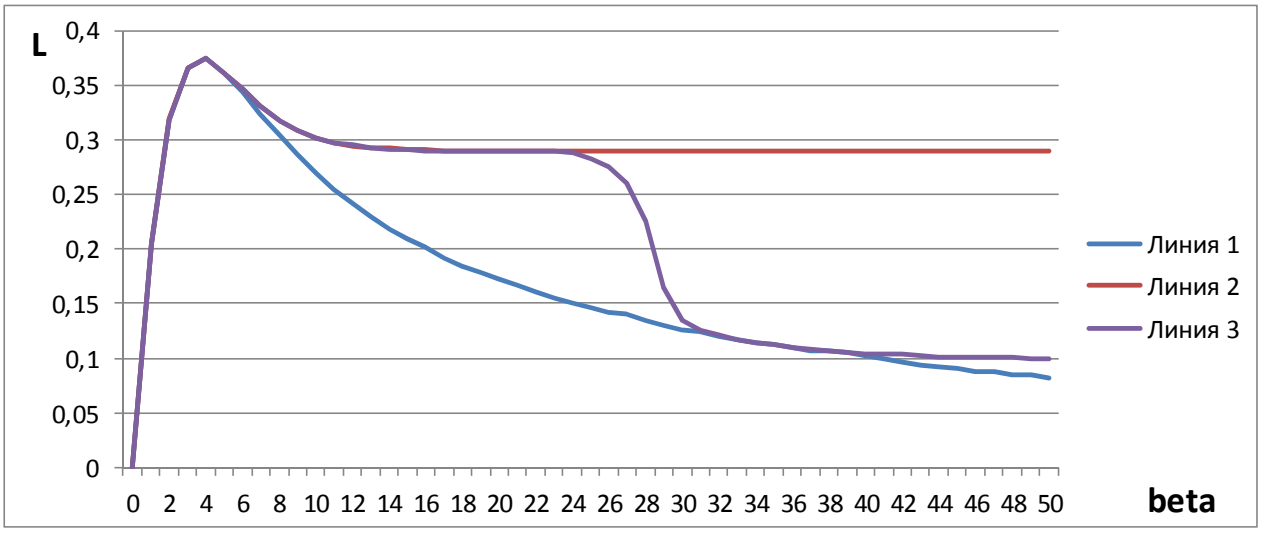

Рисунок 2 – Графики значений функции потерь на множестве значений β. Линия 1 показывает потери при минимальных α для каждого β, линия 2 – потери для фиксированного $\alpha = 0,58$, линия 3 – потери при фиксированном $\alpha = 0,58$, $\alpha_{th} = 0,3$, $T_{th} = 5000$.

Линия 3 показывает значения потерь при выбранных случайных параметрах. Как видно, параметры оказались удачны: ни одно из значений потерь не превышает вычисленной ранее величины гарантированных потерь полного дохода, и при больших значениях β потери снизились, видно, что получившиеся потери совпадают с потерями для случая простого порога. Однако отметим, что сравнивать результаты в таком виде не очень удобно. Поэтому введем новую величину $S = \int_{\beta} L(\alpha,\beta)d\beta$, которая будет показывать суммарные потери на всём множестве β. Рассматривание данной величины вполне уместно, так как можно считать, что значения множества β распределены равномерно.

Расчет значений

Итак, рассчитаем значения данной величины для приведенных выше данных. Более конкретно нас интересуют следующие значения: $S_{minimax}$ – для первого случая, S_{fix} – для случая фиксированного α и S_{th} – для случая двойного порога. Вычисления показывают, что $S_{minimax} = 8,665$, $S_{fix} = 14,702$, $S_{th} = 10,729$. Как видно, введение второго порога для тех «тестовых» параметров, которые мы рассматривали, позволило в среднем снизить суммарные потери по сравнению с одиночным фиксированным порогом, и довольно значительно (на 37%). В предположении, что имеются более оптимальные параметры, был проведен полный расчет значений суммарных потерь за время моделирования при $T = 10\ 000$, $N = 10\ 000$, $\alpha = 0,58$, $\beta = 1,...,50$, $a_{th} = 0,1...1$, $T_{th} = 1000,...,9000$. Искались такие параметры α_{th} и T_{th}, при которых S_{th} будет наименьшим, при этом ни одно из значений потерь не превышает вычисленной величины гарантированных потерь полного дохода. В

итоге удалось достигнуть значения $S_{opth} = 10{,}188$ при $T_{th} = 2000$, $a_{th} = 0{,}25$. Диаграмма это демонстрирует, потери при таких параметрах обозначены линией 4:

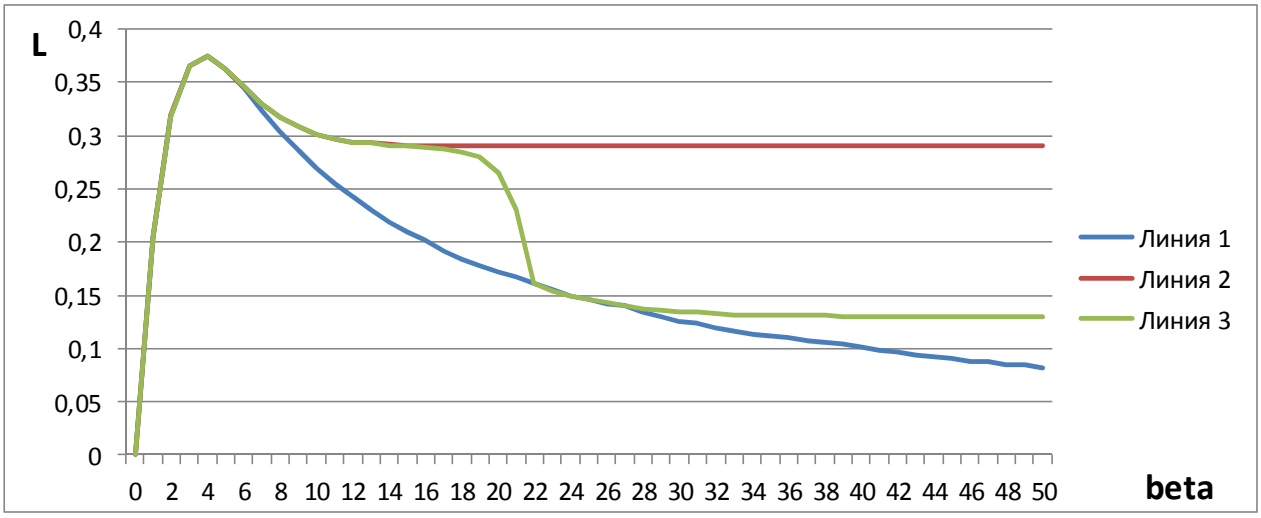

Рисунок 3 – Графики значений функции потерь на множестве значений β. Линия 1 показывает потери при минимальных α для каждого β, линия 2 – потери для фиксированного $\alpha = 0{,}58$, линия 3 – потери при фиксированном $\alpha = 0{,}58$, $\alpha_{th} = 0{,}25$, $T_{th} = 2000$.

Это позволяет судить о том, что введение дополнительного порога позволяет значительно уменьшить потери по сравнению с фиксированным α, делая их сравнимыми с потерями первого случая.

Заключение

Итак, рассмотрена пороговая стратегия управления в случайной среде с бинарными доходами с двумя действиями. Найдены оптимальные значения пороговой константы и параметра среды. Также рассмотрены суммарные потери дохода на множестве допустимых параметров среды и показано, что их можно значительно снизить с помощью введения дополнительного порога.

Автор благодарит А.В. Колногорова за помощь в постановке задачи и обсуждение полученных результатов.

Список литературы

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

2. Колногоров А.В. Нахождение минимаксных стратегий и риска в случайной среде (задаче о двуруком бандите) // Автоматика и телемеханика [В. Новгород]. - 2011. - № 5.

3. Колногоров А.В., Шелонина Т.Н. Об инвариантности функции потерь для пороговой стратегии поведения в случайной среде // Вестн. Новг. гос. ун-та. - 2006. - № 39. - С. 18-21.

4. Срагович В.Г. Адаптивное управление. - М. : Наука, 1981. – 384 с.
5. Vogel W. // Ann. Math. Statist. - 1960. - V. 31. - P. 444–451.

Рецензенты:

Едемский Владимир Анатольевич, доктор физико-математических наук, профессор кафедры ПМИ, ФГБОУ ВПО «Новгородский государственный университет имени Ярослава Мудрого», г. Великий Новгород.

Кирьянов Борис Федорович, доктор технических наук, профессор, ФГБОУ ВПО «Новгородский государственный университет имени Ярослава Мудрого», г. Великий Новгород.