

УДК 621.391.1-503.5

С.В.Мельникова

О РЕЗУЛЬТАТАХ г-ЭТАПНОЙ МИНИМАКСНОЙ ПРОЦЕДУРЫ ОПТИМИЗАЦИИ В СТАЦИОНАРНОЙ СРЕДЕ

Институт электронных и информационных систем НовГУ

In the multi-armed bandit problem an analysis of results is given for cases when the number of stages of procedure is arbitrary.

Введение

Рассматривается задача о целесообразном поведении в случайной среде, известная также как задача о «многоруком бандите». Название «многорукый бандит» идет от названия «однорукий» или «двурукый бандит» — игровой автомат, имеющий одну или две рукоятки. Принципиальное отличие «двурукого» от «однорукого бандита» состоит в том, что в случае с «двуруким бандитом» появляется выбор — какую рукоятку использовать для игры. Ясно, что вероятности выигрыша у каждой из рукояток различны, иначе необходимость во второй рукоятке исчезает. Но играющий не знает, какая из рукояток имеет лучшую вероятность выигрыша. Под «многоруким бандитом» понимается ситуация, когда имеется более двух вариантов выбора решений, причем каждое решение приводит к определенному, но заранее не известному результату.

Стационарная среда [1,2] — это управляемый случайный процесс ξ_t , $t = 1, \dots, T$, принимающий значения 1, 0, интерпретируемые как доходы и зависящие только от выбираемых в текущие моменты времени вариантов, т.е.

$$P\{\xi_t = 1 \mid y_t = l\} = p_l, P\{\xi_t = 0 \mid y_t = l\} = q_l, \\ p_l + q_l = 1, l = 1, \dots, K.$$

Таким образом, стационарная среда полностью описывается векторным параметром $\theta = (p_1, \dots, p_K)$. Особенностью рассматриваемой постановки задачи является то, что значение параметра является неизвестным и должно, так или иначе, идентифицироваться в процессе управления. Цель управления состоит в максимизации (в некотором смысле) математического ожидания полного дохода.

Это задача об оптимальном последовательном выборе альтернативных вариантов, осуществляемом на основе текущей информации. Задача имеет применения в медицине, биологии, экономике, обработке информации. Вот лишь несколько примеров применения модели «многорукого бандита».

1. Инвестиции в малые предприятия. Существует несколько направлений (или малых предприятий) для инвестиций. Задача: выбрать при данных экономических условиях одно направление (предприятие) для инвестиций.

2. Закупка сырья. Предприятие закупает сырье от различных поставщиков и делает выбор в пользу одного поставщика.

3. Модернизация производства. Предприятие увеличивает (либо меняет) выпуск продукции. Требуется найти оптимальный набор новых товаров.

4. Выбор лекарства. Существует несколько альтернативных видов лекарств. Требуется выбрать наиболее эффективное лекарство для лечения некоторого заболевания.

5. Выбор технологии. Требуется выбрать наиболее эффективную технологию изготовления товара, при использовании которой данный товар будет наиболее широко востребован со стороны покупательской аудитории и при этом будет получен максимальный доход от применения данной технологии.

Суть рассматриваемого подхода состоит в том, что управление можно разбить на небольшое число этапов, на которых применение вариантов может осуществляться параллельно. Это может применяться, например, при испытании лекарств, при передаче информации и т.д.

Рассмотрим случай r -этапной стратегии выбора, например, лекарств. Обозначим через $t_0 = 0$ начальное количество применений всех лекарств, через $X_1^{(0)} = \dots = X_K^{(0)} = 0$ — начальные результаты лечения за выбор всех лекарств. Рассматриваемая стратегия начинается с первого этапа апробации и применения всех K лекарств ($l = 1, \dots, K$) и далее следует приведенному алгоритму.

На i -м этапе апробации каждое из k оцениваемых лекарств $l = l_{i1}, \dots, l_{ik}$, $2 \leq k \leq K$, применяется τ_i раз, $1 \leq i \leq r-1$. Затем полное количество применений лекарств и полные результаты лечения пересчитываются по формулам

$$t_i = t_{i-1} + k\tau_i, \quad X_{l_{ij}}^{(i)} = X_{l_{ij}}^{(i-1)} + \sum_{t=t_{i-1}+(j-1)\tau_i+1}^{t_{i-1}+j\tau_i} \xi_t, \quad j = 1, \dots, k.$$

Определяется текущий наилучший результат применения лекарств $X_{l_{j_0}}^{(i)} = \max_{1 \leq j \leq k} X_{l_{ij}}^{(i)}$, причем если таких наилучших результатов более одного, то выбрать можно любой из них. Затем наилучший результат сравнивается со всеми оставшимися, и если выполняется неравенство $X_{l_{j_0}}^{(i)} - X_{l_{ij}}^{(i)} \geq \Delta_i$ для некоторого варианта l_{ij} , то

стратегия отбрасывает соответствующее лекарство из группы проверяемых лекарств $l = l_{i1}, \dots, l_{ik}$. Если по окончании сравнения в группе остается единственное лекарство с номером l_{ij_0} (т.е. то, которому соответствует наилучший результат лечения), то происходит переход к заключительному этапу. В противном случае выполняется $(i+1)$ -й этап апробации с оставшимся множеством лекарств. Так как $\Delta_{r-1} = 0$, то переход к заключительному этапу в конце $(r-1)$ -го этапа апробации является обязательным.

На заключительном этапе применяется только лекарство с номером l_{j_0} , соответствующее наилучшему результату лечения на предыдущем этапе апробации. Этот вариант считается наилучшим по результатам проведенных этапов апробации.

Отметим, что байесовский подход к задаче (особенно при $r = 2$ и $r = 3$) рассматривался различными авторами (см., например, [3-5]).

Основные результаты

Обозначим через $w_l(\theta) = p_l$ математическое ожидание результата лечения от применения l -го лекарства. Будем рассматривать множество параметров $\Theta = \{\theta: |p_l - p_k| \leq \Delta W \leq 1; l, k = 1, \dots, K\}$. Если бы параметр θ был известен, то всегда следовало бы выбирать лекарство, соответствующее максимальному значению $w_l(\theta)$. В этом случае математическое ожидание полного результата лечения было бы равно $T \cdot \max_{l=1, \dots, K} w_l(\theta)$. Если же параметр неизвестен, то сле-

дует применять представленную выше стратегию, которая может быть описана последовательностью $\pi = \{\tau_1, \dots, \tau_{r-1}; \Delta_1, \dots, \Delta_{r-1}\}$. Пусть $E_{\pi, \theta}$ обозначает математическое ожидание при условии, что параметр равен θ , а примененная стратегия есть π , тогда функция потерь и соответствующий минимаксный риск определяются следующим образом:

$$L_T(\pi, \theta) = T \max_{l=1, \dots, K} w_l(\theta) - E_{\pi, \theta} \sum_{t=1}^T \xi_t, \quad R_T(\Theta) = \inf_{\pi} \sup_{\theta} L_T(\pi, \theta).$$

Пусть $D = 0,1$ обозначает максимальную дисперсию дохода $D_l(\theta) = p_l q_l$ на множестве параметров Θ и определены константы

$$C_K = \max_{a_1, \dots, a_{K-1} > 0} \sum_{i=1}^{K-1} a_i \int_{-\infty}^{\infty} \varphi(x + a_i) \Phi(x) \prod_{j=1, j \neq i}^{K-1} \Phi(x + a_j) dx,$$

где $\varphi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$, $\Phi(x) = \int_{-\infty}^x \varphi(t) dt$.

Эти константы могут быть определены численными методами. Например, $C_2 \approx 0,240$, $C_3 \approx 0,372$, $C_4 \approx 0,463$, $C_5 \approx 0,532$, $C_6 \approx 0,587$, $C_7 \approx 0,633$.

Следующие теоремы описывают асимптотическое поведение минимаксного риска $R_T(\Theta)$ как функции T для случаев $r = 2$ и $r \geq 3$. Отметим, что случай $r = 2$, $K = 2$ был рассмотрен в [6], случай $r = 2$, $K \geq 2$ — в [7].

Теорема 1. Если $r = 2$, то $\Delta_1 = 0$ и асимптотически оптимальная продолжительность этапа апробации удовлетворяет условию

$$\Delta W(K-1)\tau_1 \sim C_K(D/\tau_1)^{1/2} T.$$

Следовательно,

$$\tau_1 \sim \left(\frac{C_K}{K-1} \right)^{2/3} \kappa^{1/3} T^{2/3},$$

где $\kappa = D(\Delta W)^{-2}$, а минимаксный риск удовлетворяет предельному равенству

$$\lim_{T \rightarrow \infty} R_T(\Theta) \tau_1^{-1} = (K-1)\Delta W.$$

Теорема 2. Если $r \geq 3$, то асимптотически оптимальные продолжительности этапов апробации и значения порогов удовлетворяют условиям

$$\tau_i \sim (a_i T^{a_i})^2, \quad \Delta_i \sim b_i (2D\tau_i)^{1/2}, \quad i = 1, \dots, r-1,$$

где $a_1, \dots, a_{r-1}$ удовлетворяют уравнениям

$$2a_1 = 2a_{i+1} - a_i = 1 - a_{r-1}, \quad i = 1, \dots, r-2,$$

$b_1, \dots, b_{r-1}$ имеют вид

$$b_i \sim (\beta_i \ln T)^{1/2}, \quad \beta_i \sim 2(1 - a_i - 2a_1), \quad i = 1, \dots, r-1,$$

$a_1 a_{r-1}$ удовлетворяют асимптотическим (при $T \rightarrow \infty$) соотношениям

$$\Delta W (K-1) a_1^2 \sim \frac{b_i a_{i+1}^2 (2D)^{1/2} (K-1)}{a_i} \sim \frac{C_K D^{1/2}}{a_{r-1}},$$

$$i = 1, \dots, r-2.$$

Минимаксный риск удовлетворяет предельному равенству

$$\lim_{T \rightarrow \infty} R_T(\Theta) (a_1 T^{\alpha_1})^{-2} = (K-1) \Delta W,$$

где $\alpha_1 = 2^{r-2} (2^r - 1)^{-1}$, $a_1 \sim \gamma_1 (\ln T)^{\delta_1/2}$, $\delta_1 = (2^{r-2} - 1) \times (2^r - 1)^{-1}$.

Заметим, что приведенные оценки были получены исходя из условий

$$\lim_{T \rightarrow \infty} \frac{\tau_i}{\tau_j} = 0, \text{ если } i < j, \quad \lim_{T \rightarrow \infty} \frac{i}{T} = 0. \quad (*)$$

Неулучшаемая оценка минимаксного риска для любых стратегий управления приведена в [8].

Анализ результатов моделирования

Ниже приведены таблицы значений продолжительности этапа апробации и максимума минимаксного риска при различных значениях ΔW и числе лекарств K для случаев двух, трех и четырех этапов.

В табл. 1 рассматривается случай двух этапов:

1-й этап — апробация и 2-й этап — итоговый, на котором окончательно применяется выбранное лекарство. Более подробно дается информация для случая $\Delta W = 0,05$; наблюдаемые тенденции справедливы и для значений $\Delta W = 0,1$ и $\Delta W = 0,5$. Величина $K \cdot \tau_1$ показывает, какой процент от величины T будет выделен на первый этап с учетом всех K вариантов.

Таблица 1

Значения продолжительности этапа апробации $K \cdot \tau_1$ и максимума минимаксного риска R при различных значениях продолжительности T , величины ΔW и числа лекарств K для случая $r = 2$

ΔW	K	T	$K \cdot \tau_1, \%$	R
0,05	2	10000	12	31
		100000	6	142
		1000000	3	660
	4	10000	18	69
		100000	8	318
		1000000	4	1476
	7	10000	25	106
		100000	12	494
		1000000	5	2291
0,1	4	10000	12	86
		100000	5	401
		1000000	2	1860
0,5	4	10000	4	148
		100000	2	685
		1000000	1	3180

Анализируя таблицу, можно сделать следующие выводы:

1. С увеличением числа лекарств K неизбежно увеличиваются потери R . Объясняется это тем, что «лишние» потери возникают из-за дополнительного применения большего количества неэффективных лекарств.

2. С увеличением величины T доля потерь R уменьшается. При увеличении величины T в 10 раз минимаксный риск увеличивается менее чем в 5 раз.

3. Если лекарства близки по эффективности (ΔW мало), то величина этапов апробации увеличивается. Объясняется это тем, что результаты лечения в разных группах будут практически одинаковыми и сложно сразу определить, какое лекарство окажется наиболее эффективным.

4. Если лекарства близки по эффективности (ΔW мало), то минимаксный риск снижается. С увеличением различия в эффективности лекарств увеличиваются потери R .

В табл. 2 рассматривается случай трех этапов: 1-й и 2-й — этапы апробации, 3-й этап — итоговый, на котором окончательно применяется выбранное лекарство. Продолжительность 1-го этапа — τ_1 , 2-го — τ_2 ; величина $K \cdot \tau$ — итоговое время (максимум), потраченное на 1-й и 2-й этапы для всех лекарств (в процентах от величины T) — вычислена в предположении, что все лекарства, кроме одного, используются на 2-м этапе апробации, т. е. после 1-го этапа был отсеив только одно лекарство.

Таблица 2

Значения максимума продолжительности этапов апробации $K \cdot \tau$, величины порога Δ_1 и максимума минимаксного риска R при различных значениях продолжительности T , величины ΔW и числа лекарств K для случая $r = 3$

ΔW	K	T	τ_1	τ_2	$K \cdot \tau, \%$	Δ_1	R
0,05	4	10000	409	570	33	15	61
		100000	1574	3848	18	32	236
		1000000	6020	26287	10	69	903
	7	10000	329	411	48	13	99
		100000	1266	2777	26	29	380
		1000000	4844	18973	15	62	1453
0,1	4	10000	226	467	23	11	68
		100000	869	3157	13	24	261
		1000000	3324	21564	8	51	997
0,5	4	10000	57	295	11	5	85
		100000	219	1993	7	12	328
		1000000	837	13615	4	26	1255

Анализ таблицы позволяет заключить, что все выводы, сделанные ранее, справедливы, а кроме того можно сделать дополнительные выводы:

5. С увеличением числа этапов r величина минимаксного риска R уменьшается.

6. Пороговое значение Δ_1 становится более значительным при увеличении T и, наоборот, уменьшается при увеличении числа K лекарств.

Таблица 3

Значения максимума продолжительности этапов апробации с учетом всех лекарств $K \cdot \tau$, величин пороговых значений $\Delta_1, \Delta_2, \Delta_3$ и максимума минимаксного риска R при различных значениях продолжительности T , величины ΔW и числа лекарств K для случая $r = 4$

K	T	i	$\Delta W = 0,05$				$\Delta W = 0,1$				$\Delta W = 0,5$			
			τ_i	$K \cdot \tau, \%$	Δ_i	R	τ_i	$K \cdot \tau, \%$	Δ_i	R	τ_i	$K \cdot \tau, \%$	Δ_i	R
4	100000	1	1290	24	35	194	676	19	25	203	150	12	12	226
		2	2415		27		1830		24		961		17	
		3	5722		0		5217		0		4209		0	
	1000000	1	4570	15	71	685	2393	13	51	718	533	9	24	799
		2	14691		74		11134		64		5849		46	
		3	45626		0		41599		0		33565		0	
7	1000000	1	3731	25	64	1119	1953	21	46	1172	435	15	22	1305
		2	10837		63		8213		55		4314		40	
		3	31991		0		29167		0		23534		0	
	10000000	1	13137	18	130	3941	6879	15	94	4127	1532	11	44	4595
		2	66299		169		50245		147		26394		107	
		3	257975		0		235201		0		189777		0	

Однако отметим особо, что в некоторых случаях (при малых $\Delta W, T$) не выполняются условия (*), что оказывает влияние на увеличение погрешности при вычислении величин Δ_1, R . Поэтому из дальнейшего анализа исключим малые значения T . (Результаты, где условия (*) выполняются, выделены жирным шрифтом).

В табл.3 рассматривается случай четырех этапов: 1-й, 2-й и 3-й — этапы апробации, их продолжительность — τ_1, τ_2, τ_3 соответственно, 4-й этап — итоговый, на котором окончательно применяется выбранное лекарство. Величина $K \cdot \tau$ вычислена в предположении, что все лекарства, кроме одного, переходили на следующий этап апробации, т. е. после каждого этапа был отсеив только одно лекарство. Очевидно, что после каждого этапа будет отсеиваться как минимум одно лекарство и величина $K \cdot \tau$ может быть уменьшена.

Анализ табл. 3 и ее сравнение с табл. 1, 2 показывают, что сделанные ранее выводы остаются верными. Аналогично выделены жирным шрифтом результаты, когда выполняются условия (*).

Анализируя полученные значения, приходим еще к одному выводу:

7. Пороговые значения отклонения альтернативных лекарств на каждом последующем этапе увеличиваются, за исключением последнего, которое всегда принимается равным нулю: $\Delta_{r-1} = 0$.

Результаты моделирования наглядно показывают верность асимптотических оценок, которые справедливы при больших значениях T ($T \rightarrow \infty$).

1. Срагович В.Г. Адаптивное управление. М.: Наука, 1981. 384 с.
2. Назин А.В., Позняк А.С. Адаптивный выбор вариантов. М.: Наука, 1986. 288 с.
3. Berry D.A., Fristedt B. Bandit Problems. L., N.Y.: Chapman and Hall, 1985. 275 p.
4. Witmer J.A. // Ann. Stat. 1986. V.14. P.283-297.
5. Cheng Y. // Sequential Analysis. 1994. V.13. P.329-350.
6. Колногоров А.В. // Проблемы передачи информации. 2000. Т.36. № 4. С.117-127.
7. Колногоров А.В., Мельникова С.В. // Вестник НовГУ. Сер.: Техн. науки. 2005. №34. С.73-75.
8. Vogel W. // Ann. Math. Statist. 1960. V.31. P.444-451.