XII LIETUVOS MATEMATIKOS RINKINYS
ЛИТОВСКИЙ МАТЕМАТИЧЕСКИЙ СБОРНИК 3

1972

УДК 519.2

О НЕКОТОРЫХ СВОЙСТВАХ ОПТИМАЛЬНОЙ ЦЕНЫ В ОБОБЩЕННОЙ ЗАДАЧЕ О ДВУХ ТИПАХ ОРУЖИЯ

Д. Сургайлис

Задача о двух типах оружия („two-armed bandit" problem) в общей постановке формулируется следующим образом [1]: проводятся два эксперимента A и B, каждый из которых в результате имеет два возможных исхода – успех (1) или неудачу (0) с вероятностями успеха соответственно a и b, причем сами значения a и b неизвестны, а известно a priori лишь их совместное распределение вероятностей $\mu\,(da\times db)$. Требуется найти такую стратегию (правило выбора между A и B), которая максимизировала бы ожидаемое число успехов R_N при заданном числе испытаний N.

В настоящей заметке докажем два свойства цены R_N (теоремы 1 и 2), которые можно было бы условно назвать „монотонностью" и „выпуклостью" (относительно некоторого множества преобразований априорной меры μ). Заметим при этом, что о свойствах оптимальной стратегии и цены (в случае произвольного распределения μ) известно к настоящему времени мало, большинство полученных результатов носит негативный характер [1].

Будем считать в дальнейшем, что случайные величины a и b независимы, т.е. $\mu\,(da\times db)=\mu\,(da)\,\nu\,(db)$, и, кроме того, $\mu\,(\{0\})=\mu\,(\{1\})=\nu\,(\{0\})=\nu\,(\{1\})=0$. Пусть G_+ означает множество всех монотонно возрастающих функций $g=g\,(t)$, $t\in[0, 1]$, причем $0<g\,(t)<1$ для $0<t<1$. Пусть $G_-=1-G_+=\{g: g=1-f, f\in G_+\}$, $G=\{f: f=g_1\cdot g_2\dots g_n, g_i\in G_+\cup G_-\}$, а $M=\{\mu\}$ – множество всех вероятностных мер на борелевских подмножествах интервала [0, 1], таких, что $\mu\,(\{0\})=\mu\,(\{1\})=0$. Обозначим T_g, $g\in G$ отображение $M\to M$ по формуле

$$(T_g\mu)(A)=\int\limits_A gd\mu\left[\int gd\mu\right]^{-1}.$$

Тогда имеют место следующие простые свойства:

$$T_fT_g=T_gT_f=T_{fg},\qquad f,\ g\in G; \tag{T1}$$

$$\int fdT_g\mu\cdot\int gd\mu=\int gdT_f\mu\cdot\int fd\mu,\qquad f,\ g\in G; \tag{T2}$$

$$\int fd\mu\leqslant\int fdT_g\mu \text{ для } f\in G_+,\ g\in G_+, \text{ и } f\in G_-,\ g\in G_-; \tag{T3}$$

$$\int fd\mu\geqslant\int fdT_g\mu \text{ для } f\in G_+,\ g\in G_-, \text{ и } f\in G_-,\ g\in G_+.$$

Докажем, например, первое неравенство в (T1) ((T1) и (T2) следует непосредственно из определения). Имеем (для f, $g\in G_+$):

$$\Delta\equiv\int f dT_g\mu-\int f d\mu=\int_{A_+} f(\bar g-1)\,d\mu+\int_{A_-} f(\bar g-1)\,d\mu,$$

где

$$\bar g=g\left[\int g d\mu\right]^{-1},\quad A_+=\{t: t\in[0,\ 1],\ \bar g(t)\geqslant 1\},$$

$$A_-=[0,\ 1]\setminus A_+,\quad c=\inf\{t:\bar g(t)\geqslant 1\}.$$

Тогда

$$\Delta\geqslant f(c)\Big[\int_{A_+}(\bar g-1)\,d\mu-\int_{A_-}(1-\bar g)\,d\mu\Big]=0.$$

Пусть $R_N=R_N(\mu,\ \nu)$ означает максимальный ожидаемый выигрыш после N испытаний. Уравнения Беллмана для R_N выглядит следующим образом (в наших обозначениях):

$$R_N(\mu,\ \nu)=\max\left(R_N^A(\mu,\ \nu),\ R_N^B(\mu,\ \nu)\right),$$

где

$$R_N^A(\mu,\ \nu)=\int a d\mu+\int a d\mu\, R_{N-1}(T_a\mu,\ \nu)+\int(1-a)\,d\mu\, R_{N-1}(T_{1-a}\mu,\ \nu),$$

$$R_N^B(\mu,\ \nu)=\int b d\nu+\int b d\nu\, R_{N-1}(\mu,\ T_b\nu)+\int(1-b)\,d\nu\, R_{N-1}(\mu,\ T_{1-b}\nu),$$

а функции $a(t)=b(t)\equiv t$, $0\leqslant t\leqslant 1$.

Теорема 1.

$$R_N(\mu,\ \nu)\leqslant R_N(T_g\mu,\ \nu),\qquad g\in G_+.$$

Доказательство проводится по индукции. Для

$$R_1(\mu,\ \nu)=\max\left(\int a d\mu,\ \int b d\nu\right)$$

оно очевидно. Пусть теорема верна для R_k, $k<N$. Тогда достаточно показать, что

$$R_N^A(\mu,\ \cdot)\leqslant R_N^A(T_g\mu,\ \cdot),$$

$$R_N^B(\mu,\ \cdot)\leqslant R_N^B(T_g\mu,\ \cdot).$$

Проверим первое из них (второе проверяется аналогично). Имеем

$$R_N^A(T_g\mu,\ \cdot)-R_N^A(\mu,\ \cdot)=\int a d\mu\,[R_{N-1}(T_aT_g\mu,\ \cdot)-R_{N-1}(T_a\mu,\ \cdot)]+$$

$$+\int(1-a)\,d\mu\,[R_{N-1}(T_{1-a}T_g\mu,\ \cdot)-R_{N-1}(T_{1-a}\mu,\ \cdot)]+$$

$$+[R_{N-1}(T_aT_g\mu,\ \cdot)-R_{N-1}(T_{1-a}T_g\mu,\ \cdot)]\left(\int a dT_g\mu-\int a d\mu\right)\cdot$$

В последнем выражении все три слагаемых неотрицательны (в силу (T1), (T3) и индукционного предположения). Теорема доказана.

Теорема 2.

$$\int g d\mu\, R_N(T_g\mu,\ \nu)+\int(1-g)\,d\mu\, R_N(T_{1-g}\mu,\ \nu)\geqslant R_N(\mu,\ \nu),\quad g\in G.$$

Доказательство теоремы также проводится по индукции. Для $N=1$ оно проверяется просто. Предположим, что утверждение теоремы верно для $k<N>1$. Так как

$$\int g d\mu\, R_N(T_g \mu, \cdot) + \int (1-g)\, d\mu\, R_N(T_{1-g}\mu, \cdot) \geqslant$$
$$\geqslant \max\Big(\int g d\mu\, R_N^A(T_g \mu, \cdot) + \int (1-g)\, d\mu\, R_N^A(T_{1-g}\mu, \cdot),$$
$$\int g d\mu\, R_N^B(T_g \mu, \cdot) + \int (1-g)\, d\mu\, R_N^B(T_{1-g}\mu, \cdot)\Big) \equiv \max(P_A, P_B),$$

то достаточно показать, что

$$P_A \geqslant R_N^A \text{ и } P_B \geqslant R_N^B. \tag{1}$$

Воспользовавшись свойствами (T1)—(T3), а также индукционным предположением, имеем

$$P_A - \Big[\int g d\mu \cdot \int a dT_g \mu + \int (1-g)\, d\mu \int a\, dT_{1-g}\mu\Big] =$$
$$= \int g d\mu \Big[\int a dT_g \mu \cdot R_{N-1}(T_a T_g \mu, \cdot) +$$
$$+ \int (1-a)\, dT_g \mu R_{N-1}(T_{1-a} T_g \mu, \cdot)\Big] +$$
$$+ \int (1-g)\, d\mu \Big[\int a dT_{1-g} \mu \cdot R_{N-1}(T_a T_{1-g}\mu, \cdot) + \int (1-a)\, dT_{1-g}\mu \cdot$$
$$\cdot R_{N-1}(T_{1-a} T_{1-g}\mu, \cdot)\Big] = \Big[\int g dT_a \mu \cdot R_{N-1}(T_a T_g \mu, \cdot) +$$
$$+ \int (1-g)\, dT_a \mu\, R_{N-1}(T_{1-g} T_a \mu, \cdot)\Big] \int a d\mu + \Big[\int g dT_{1-a} \mu \cdot$$
$$\cdot R_{N-1}(T_{1-a} T_g \mu, \cdot) +$$
$$+ \int (1-g)\, dT_{1-a} \mu \cdot R_{N-1}(T_{1-a} T_{1-g}\mu, \cdot)\Big] \int (1-a)\, d\mu \geqslant$$
$$\geqslant \int a d\mu\, R_{N-1}(T_a \mu, \cdot) + \int (1-a)\, d\mu\, R_{N-1}(T_{1-a}\mu, \cdot) = R_N^A - \int a d\mu,$$

что доказывает первое из неравенств (1). Второе проверяется аналогично. Теорема доказана.

Институт физики и математики
Академии наук Литовской ССР

Поступило в редакцию
23.IX.1971

Литература

1. R. N. Bradt, S. M. Johnson, S. Karlin, On sequential designs for maximizing the sum of n observations, Ann. Math. Stat., 27, pp. 1060—1074 (1956).

APIE KAI KURIAS RIZIKOS FUNKCIJOS SAVYBES APIBENDRINTAME UŽDAVINYJE APIE DVI GINKLŲ RŪŠIS

D. Surgailis

(*Reziumė*)

Straipsnyje įrodomos rizikos funkcijos apibendrintame „dvirankio bandito" uždavinyje dvi savybės, susijusios su rizikos funkcijos elgesiu, esant tam tikroms apriorinio pasiskirstymo transformacijoms.

ON SOME PROPERTIES OF THE RISK FUNCTION IN THE GENERALIZED „TWO-ARMED BANDIT" PROBLEM

D. Surgailis

(*Summary*)

Two properties of the risk function in the generalized „two-armed bandit" problem are proved relative to the behavior of the risk function at some transformations of a priori distribution.