УДК 519.865

НАХОЖДЕНИЕ МИНИМАКСНЫХ СТРАТЕГИЙ И РИСКА В ТРЕХАЛЬТЕРНАТИВНОЙ СЛУЧАЙНОЙ СРЕДЕ

А.В.Колногоров

Институт электронных и информационных систем НовГУ, Alexander.Kolnogorov@novsu.ru

Минимаксные стратегия и риск в трехальтернативной стационарной случайной среде ищутся как байесовские, соответствующие наихудшему априорному распределению, которое в рассматриваемом случае может быть выбрано симметрическим и асимптотически однородным. Это позволяет определить стратегию и риск численными методами.

Ключевые слова: поведение в случайной среде, задача о многоруком бандите, минимаксный и байесовский подходы, параллельная обработка

Minimax strategy and risk in a three-alternative stationary random environment are found as Bayesian ones corresponding to the worst prior distribution. In considered case, this distribution can be chosen to be symmetric and asymptotically uniform. This lets one use numerical methods to determine the strategy and define the risk.

Keywords: behavior in random environment, multi-armed bandit problem, minimax and bayesian approaches, parallel processing

Введение

Рассматривается задача, известная как задача адаптивного управления в случайной среде [1] и как задача о двуруком (многоруком) бандите [2] в следующей постановке, обобщающей результаты [3,4] на случай трехальтернативных случайных сред. Пусть ξ_n, $n=1,...,N$ есть управляемый случайный процесс, значения которого интерпретируются как доходы, зависят только от выбираемых в текущие моменты времени вариантов η_n и имеют нормальные распределения с плотностями

$$f(x|m_\ell)=(2\pi)^{-1/2}\exp\left(-(x-m_\ell)^2/2\right),$$

если $\eta_n=\ell$ $(\ell=1,2,3)$. Такая среда описывается векторным параметром $\theta=(m_1,m_2,m_3)$. Цель управления состоит в максимизации полного ожидаемого дохода. Для этого используется стратегия σ, которая на первых трех шагах применяет все три варианта по очереди, а при $n>3$ является измеримой функцией текущей предыстории процесса, т. е. полученных откликов $x^{n-1}=x_1,...,x_{n-1}$ на выбранные варианты $y^{n-1}=y_1,...,y_{n-1}$. Множество стратегий обозначим Σ.

Если параметр θ известен, то следует всегда применять вариант, которому соответствует бóльшая из величин m_1, m_2, m_3, и полный ожидаемый доход равен в этом случае Nm^*, где $m^*=m_1\vee m_2\vee m_3$. Если же параметр неизвестен, то функция

$$L_N(\sigma,\theta)=E_{\sigma,\theta}\left(\sum_{n=1}^{N}(m^*-\xi_n)\right)$$

характеризует потери дохода вследствие неполноты информации. Здесь $E_{\sigma,\theta}$ обозначает математическое ожидание по мере, порожденной стратегией σ и параметром θ. Множество допустимых значений параметра имеет вид

$$\Theta=\left\{(m_1,m_2,m_3):\max|m_i-m_j|\le 2c_1,|m_1+m_2+m_3|\le 3c_2\right\},$$

где $c_1,c_2<\infty$. Условие $c_1<\infty$ требуется для ограниченности функции потерь на Θ. Условие $c_2<\infty$ нужно для того, чтобы Θ было компактным множеством, однако предполагается, что c_2 достаточно велико.

При минимаксном подходе максимальные потери на множестве параметров Θ минимизируется по множеству стратегий Σ, величина

$$R_N^M(\Theta)=\inf_{\Sigma}\sup_{\Theta}L_N(\sigma,\theta) \qquad (1)$$

называется минимаксным риском, а обеспечивающая ее значение стратегия (если она существует) — минимаксной стратегией. Другим известным подходом к задаче является байесовский. Обозначим через Λ априорное распределение параметра на множестве Θ. Величина

$$R_N^B(\Lambda)=\inf_{\Sigma}\int_{\Theta}L_N(\sigma,\theta)\Lambda(d\theta) \qquad (2)$$

называется байесовским риском, а соответствующая стратегия — байесовской. Объединяет два подхода основная теорема теории игр, согласно которой минимаксный риск (1) совпадает с байесовским риском (2) на наихудшем априорном распределении, соответствующем максимуму байесовского риска, а минимаксная стратегия совпадает с соответствующей байесовской.

Справедливость основной теоремы теории игр для двухальтернативных случайных сред установлена в [3]; этот результат легко обобщается на рассматриваемый случай. Нахождению минимаксных стратегии и риска как байесовских, соответствующих наихудшему априорному распределению, и посвящена данная работа.

Свойства асимптотически наихудшего априорного распределения

Для вычисления байесовского риска можно написать рекуррентные уравнения. Обозначим $\{a\}:=(a_1,a_2,a_3)$, $\{a,b\}:=(a_1,b_1,a_2,b_2,a_3b_3)$, $\{a\}_{-\ell}:=(a_1,a_2,a_3)\setminus a_\ell$, $\{a,b\}_{-\ell}:=(a_1,b_1,a_2,b_2,a_3b_3)\setminus(a_\ell,b_\ell)$.

Обозначим через $f_D(x|M)=(2\pi D)^{-1/2}\exp\left(-(x-M)^2/(2D)\right)$ плотность нормального распределения с математическим ожиданием M и дисперсией D, через $\lambda\{m\}=\lambda(m_1,m_2,m_3)$ — плотность априорного распределения на множестве параметров Θ. Пусть предыстория процесса к моменту времени n описывается набором $\{X,n\}=(X_1,n_1,X_2,n_2,X_3,n_3)$, где n_1,n_2,n_3 — полные количества применений всех вариантов, причем $n_1+n_2+n_3=n$, а X_1,X_2,X_3 — полные доходы за все варианты. Будем считать, что $X_\ell=0$ при $n_\ell=0$. Плотность апостериорного распределения определяется как

$$\lambda(\{m\}|\{X,n\})=\frac{\left(\prod_{\ell=1}^{3}f_{n_\ell}(X_\ell|n_\ell m_\ell)\right)\lambda\{m\}}{\iiint\limits_{\Theta}\left(\prod_{\ell=1}^{3}f_{n_\ell}(X_\ell|n_\ell m_\ell)\right)\lambda\{m\}dm_1dm_2dm_3}.$$

Если положить $f_n(x|nm)=1$ при $n=0$, то эта формула останется справедливой и в том случае, если некоторые или все n_1,n_2,n_3 будут равны нулю.

Обозначим через $R^B_{N-n}(\lambda;\{X,n\})$, $n=n_1+n_2+n_3$, байесовский риск на последних $N-n$ шагах, вычисленный относительно апостериорного распределения с плотностью $\lambda(\{m\}|\{X,n\})$. Тогда

$$R^B_{N-n}(\cdot)=\min\left(R^{(1)}_{N-n}(\cdot),R^{(2)}_{N-n}(\cdot),R^{(3)}_{N-n}(\cdot)\right), \qquad (3)$$

где $R^{(1)}_{N-n}(\cdot)=R^{(2)}_{N-n}(\cdot)=R^{(3)}_{N-n}(\cdot)=0$ при $n=N$ и при $3<n<N$;

$$R^{(\ell)}_{N-n}(\lambda;\{X,n\})=\iiint\limits_{\Theta}\Big(m^*-m_\ell+$$
$$+E^{(\ell)}_x R^B_{N-n-1}(\lambda;\{X,n\}_{-\ell},(X_\ell+x,n_\ell+1)_\ell)\Big)\times$$
$$\times\lambda(\{m\}|\{X,n\})dm_1dm_2dm_3, \qquad (4)$$

$$E^{(\ell)}_x R(x)=\int_{-\infty}^{\infty}R(x)f(x|m_\ell)dx,\ \ell=1,2,3.$$

Здесь $R^{(\ell)}_{N-n}(\cdot)$ — ожидаемые потери на оставшемся отрезке времени, если сначала выбирается ℓ-й вариант, а затем управление ведется оптимально $(\ell=1,2,3)$. Байесовская стратегия предписывает выбирать вариант, которому соответствует меньшее из значений $R^{(1)}_{N-n}(\cdot),R^{(2)}_{N-n}(\cdot),R^{(3)}_{N-n}(\cdot)$, при их равенстве выбор может быть произвольным.

Сделаем дополнительные обозначения. Определим $\{a\}_{ij}=\{a'\}$ условиями: $a'_\ell=a_\ell$ при $\ell\neq i$, $\ell\neq j$, $a'_i=a_j$, $a'_j=a_i$. Аналогично $\{a,b\}_{ij}=\{a',b'\}$ определим условиями: $a'_\ell=a_\ell$, $b'_\ell=b_\ell$ при $\ell\neq i$, $\ell\neq j$, $a'_i=a_j$, $a'_j=a_i$, $b'_i=b_j$, $b'_j=b_i$. При любом постоянном m_0 положим $\{a+m_0\}=\{a'\}$, где $a'_\ell=a_\ell+m_0$, $\ell=1,2,3$, $\{a+m_0b\}=\{c\}$, где $c_\ell=a_\ell+m_0b_\ell$, $\ell=1,2,3$.

Справедливы две леммы, доказательства которых проводятся аналогично доказательствам соответствующих лемм в [3,4].

Лемма 1. Следующие преобразования $\hat\lambda$ априорной плотности распределения λ не меняют байесовский риск, т. е. $R^B_N(\hat\lambda)=R^B_N(\lambda)$:

1) $\hat\lambda^{(1)}\{m\}=\lambda\{m\}_{ij}$ (для всех $\{m\}$ и всех $i\neq j$),

2) $\hat\lambda^{(2)}\{m\}=\lambda\{m+m_0\}$ (для всех $\{m\}$ и любого фиксированного m_0).

Лемма 2. Байесовский риск является вогнутой функцией априорного распределения, т. е. для любых плотностей λ_1, λ_2 и положительных чисел α_1, α_2, таких что $\alpha_1+\alpha_2=1$, справедливо неравенство

$$R^B_N(\alpha_1\lambda_1+\alpha_2\lambda_2)\geq\alpha_1R^B_N(\lambda_1)+\alpha_2R^B_N(\lambda_2).$$

Далее удобно изменить параметризацию. Положим $m_\ell=u+v_\ell$, $\ell=1,2,3$, причем $v_1+v_2+v_3=0$, тогда $\theta=(u+v_1,u+v_2,u+v_3)$, $\Theta=\{u:|u|\leq c_2,(v_1,v_2,v_3)\in\Theta_v\}$, $\Theta_v=\{(v_1,v_2,v_3):\max|v_i-v_j|\leq 2c_1,v_1+v_2+v_3=0\}$. С учетом якобиана преобразования априорная плотность равна $\nu(u,\{v\})=3\lambda\{u+v\}$, где только две компоненты $\{v\}$, например v_1,v_2, являются независимыми, а $v_3=-v_1-v_2$. В силу леммы 1 не меняют значения байесовского риска плотности $\hat\nu^{(1)}_{ij}(u,\{v\})=\nu(u,\{v\}_{ij})$, $\hat\nu^{(2)}(u,\{v\})=\nu(u+m_0,\{v\})$. Эти свойства позволяют описать наихудшее распределение.

Обозначим через p некоторую перестановку чисел $(1,2,3)$, через $\{v\}_p$ – соответствующую перестановку компонент параметра. Покажем, что наихудшее распределение может быть выбрано симметрическим, т.е. $\nu(u,\{v\})=\nu(u,\{v\}_{ij})$ при всех $\{v\}$ и $i\neq j$. Если это не так, то положим

$$\nu^{(1)}(u,\{v\})=\frac{1}{6}\sum_p\nu(u,\{v\}_p).$$

Ясно, что $\nu^{(1)}(u,\{v\})$ — симметрическая плотность. Поскольку всякая перестановка является результатом попарных перестановок компонент параметра, то в силу леммы 1 $R^B_N(\nu(u,\{v\}_p)=R^B_N(\nu(u,\{v\})$ при всех p. Из леммы 2 следует, что $R^B_N(\nu^{(1)}(u,\{v\})\geq R^B_N(\nu(u,\{v\})$, т.е. $\nu^{(1)}(u,\{v\})$ может быть выбрана в качестве наихудшей.

Аналогично, если $\nu(u,\{v\})$ является наихудшей, то $\nu^{(2)}(u,\{v\})=(\nu(u,\{v\})+\nu(u+m_0,\{v\}))/2$ не уменьшает байесовский риск, но является более однородной по u.

Теорема 1. Не уменьшающая байесовский риск плотность распределения при $a\to\infty$ может быть выбрана в виде

$$\nu_a(u,\{v\}) = \kappa_a(u)\rho\{v\}, \quad (5)$$

где $\kappa_a(u)$ — постоянная плотность на отрезке $|u| \le a$, а $\rho\{v\}$ — симметрическая плотность (т.е. $\rho\{v\}_{ij} = \rho\{v\}$ при всех $i \ne j$) на множестве $\{v\} \in \Theta_v$.

Пусть стратегия на первых трех шагах за применение вариантов по очереди получает отклики x_1, x_2, x_3. Обозначим $\bar{x} = x_1 + x_2 + x_3$, $x_{ij} = x_i - x_j$, $i \ne j$. Рассмотрим плотность

$$\mu(u,\{v\}|\{x\}) = f_{1/3}(u|\bar{x})\rho(\{v\}|\{x\}), \quad (6)$$

где $\rho(\{v\}|\{x\}) = \dfrac{g(\{x\},\{v\})\rho\{v\}}{r\{x\}}$, $g(\{x\},\{v\}) =$

$$= \frac{1}{3^{1/2}2\pi}\exp\left(-\frac{(x_{12}-v_{12})^2 + (x_{23}-v_{23})^2 + (x_{31}-v_{31})^2}{6}\right),$$

$r\{x\} = \iint\limits_{\Theta_v} g(\{x\},\{v\})\rho\{v\}dv_1dv_2$, $v_{ij} = v_i - v_j$, $i \ne j$. Как и в [3,4], может быть установлена следующая теорема.

Теорема 2. Пусть $\nu_a(u,\{v\})$ выбрана из условия (5), а стратегия на первых трех шагах применяет варианты по очереди. Тогда

$$\lim_{a\to\infty} R_N^B(\nu_a(u,\{v\}|\{x\})) = 3L(\rho\{v\}) + \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} R_{N-3}^B(\mu(u,\{v\}|\{x\}))r\{x\}dx_{12}dx_{23}, \quad (7)$$

где $L(\rho\{v\}) = \iint\limits_{\Theta_v}(v^* - v_1)\rho\{v\}dv_1dv_2$. Плотность распределения $\mu(u,\{v\}|\{x\})$ выбрана из условия (6). Риски $R_{N-3}^B(\mu(u,\{v\}|\{x\}))$ не зависят от $\bar{x}$.

Уравнения для вычисления байесовского риска относительно наихудшего априорного распределения

Приведем уравнения для вычисления байесовского риска по формуле (7). Они получаются из уравнений (3), (4), если формально считать априорную плотность постоянной по u. Кроме того, уравнения получаются проще для рисков $R_{N-n}\{X,n\} = R_{N-n}^B\{X,n\}p\{X,n\}$, где $n = n_1 + n_2 + n_3$,

$$p\{X,n\} = \iiint\limits_{\Theta}\left(\prod_{\ell=1}^{3} f_{n_\ell}(X_\ell | n_\ell m_\ell)\right)\lambda\{m\}dm_1dm_2dm_3,$$

причем, как и раньше, считаем, что $f_n(x|nm) = 1$, если $n = 0$. Обозначим $\bar{R}_{N-n}\{\bar{X},n\} = R_{N-n}\{X,n\}$, где $\bar{X}_\ell = X_\ell / n_\ell$.

Теорема 3. Пусть $\nu_a(u,\{v\})$ выбрана из условия (5) и $a \to \infty$. Тогда

$$\bar{R}_{N-n}^B\{\bar{X},n\} = \min\left(\bar{R}_{N-n}^{(1)}\{\bar{X},n\}, \bar{R}_{N-n}^{(2)}\{\bar{X},n\}, \bar{R}_{N-n}^{(3)}\{\bar{X},n\}\right), \quad (8)$$

где $\bar{R}_{N-n}^{(1)}\{\bar{X},n\} = \bar{R}_{N-n}^{(2)}\{\bar{X},n\} = \bar{R}_{N-n}^{(3)}\{\bar{X},n\} = 0$ при $n_1 + n_2 + n_3 = N$,

$$\bar{R}_{N-n}^{(\ell)}\{X,n\} = \iint\limits_{\Theta_v}(v^* - v_\ell)g(\{X,n\},\{v\})\rho\{v\}dv_1dv_2 +$$

$$+ (n_\ell + 1)\int\limits_{-\infty}^{\infty}\bar{R}_{N-n-1}(\{\bar{X},n\}_{-\ell},(\bar{X}_\ell + z, n_\ell + 1)_\ell)\bar{h}_{n_\ell}(z)dz \quad (9)$$

при $n_1 + n_2 + n_3 < N$. Здесь

$$g(\{\bar{X},n\},\{v\}) = \frac{1}{2\pi(n_1n_2n_3(n_1+n_2+n_3))^{1/2}} \times$$

$$\times\exp\left(-\frac{n_1n_2(x_{12}-v_{12})^2 + n_2n_3(x_{23}-v_{23})^2 + n_3n_1(x_{31}-v_{31})^2}{2(n_1+n_2+n_3)}\right), \quad (10)$$

$\bar{X}_{ij} = \bar{X}_i - \bar{X}_j$, $v_{ij} = v_i - v_j$, $i \ne j$,

$$\bar{h}_n(z) = \left(\frac{n+1}{2\pi n}\right)^{1/2}\exp\left(-\frac{n(n+1)z^2}{2}\right). \quad (11)$$

При любом фиксированном m_0 риски удовлетворяют равенствам $\bar{R}_{N-n}\{\bar{X}',n\} = \bar{R}_{N-n}\{\bar{X},n\}$, где $\bar{X}_i' = \bar{X}_i + m_0$.

Байесовский риск (8) вычисляется по формуле

$$\lim_{a\to\infty} R_N^B(\nu_a(u,\{v\}|\{x\})) = 3L(\rho\{v\}) + \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\bar{R}_{N-3}((x_{12},1),(0,1),(x_{23},1))dx_{12}dx_{23}. \quad (12)$$

Предположим теперь, что плотность $\rho\{v\}$ является вырожденной и сосредоточена в трех точках $(2v,-v,-v)$, $(-v,2v,-v)$, $(-v,-v,2v)$ с вероятностями 1/3. Тогда $L(\rho\{v\}) = 2v$ и

$$\iint\limits_{\Theta_v}(v^* - v_\ell)g(\{X,n\},\{v\})\rho\{v\}dv_1dv_2 = vg^{(\ell)}(\{X,n\},\{v\}),$$

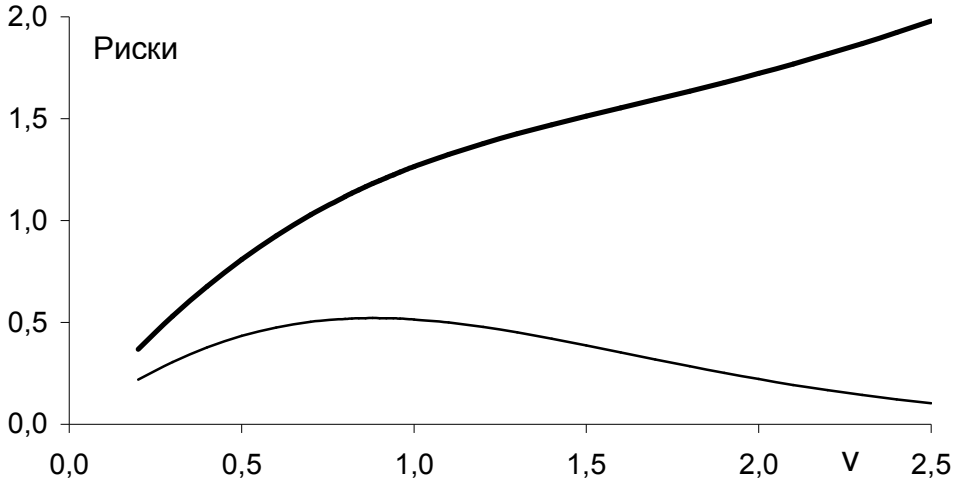
где

$$g^{(1)}(\{X,n\},\{v\}) = g(\{X,n\},(-v,2v,-v)) + g(\{X,n\},(-v,-v,2v)),$$

$$g^{(2)}(\{X,n\},\{v\}) = g(\{X,n\},(-v,-v,2v)) + g(\{X,n\},(2v,-v,-v)),$$

$$g^{(3)}(\{X,n\},\{v\}) = g(\{X,n\},(2v,-v,-v)) + g(\{X,n\},(-v,2v,-v)).$$

Наихудшее априорное распределение соответствует максимуму приведенного байесовского риска $N^{-1/2}R_N^B(\cdot)$. Поскольку объем вычислений достаточно большой, они были проделаны при $N = 8$ (см. рис.). Как видим, максимум $N^{-1/2}R_N^B(\cdot)$ достигался на границе при $v = 2{,}5$, а максимум $N^{-1/2}ER_{N-3}^B(\cdot)$ — во внутренней точке $v \approx 0{,}88$. Далее запоминались соответствующие байесовские стратегии и для них снова вычислялись потери в указанном диапазоне $0{,}1 \le v \le 2{,}5$ с шагом $0{,}1$. Эти потери практически совпали с рисками, что можно, по-видимому, объяснить тем, что оптимальная стратегия при $N = 8$ мало зависит от v в указанном диапазоне. В частности, нетрудно проверить, что на последнем шаге она такова: надо выбирать вариант, соответствующий максимуму $\bar{X}_\ell$, $\ell = 1,2,3$, т.е. в этом случае оптимальная стратегия от v не зависит совсем.



Байесовские риски $N^{-1/2}R_N^B(\cdot)$ (жирная линия) и $N^{-1/2}ER_{N-3}^B(\cdot)$ (тонкая линия), вычисленные при $0{,}1 \le v \le 2{,}5$ с шагом 0,1 по формулам (8)-(12)

Наконец, отметим, что при проверке оптимальности стратегии рассматривались также точки вида $(2v,-v+x,-v+y)$, $(-v+y,2v,-v+x)$, $(-v+x,-v+y,2v)$ для различных x,y в окрестности максимума. При этом потери в этих точках не превосходили максимальных.

Заключение

Предложен метод отыскания минимаксных стратегии и риска в многоальтернативной случайной среде, основанный на характеризации наихудшего априорного распределения и дальнейшей численной оптимизации. Оптимальная стратегия вычисляется на компьютере и табулируется. Стратегия допускает применение в системах с параллельной обработкой данных (см. [3,4]).

1. Срагович В.Г. Адаптивное управление. М.: Наука, 1981. 384 с.
2. Berry D.A., Fristedt B. Bandit Problems: Sequential Allocation of Experiments. L.; N.Y.: Chapman and Hall, 1985. 275 p.
3. Колногоров А.В. Нахождение минимаксных стратегии и риска в случайной среде (в задаче о «двуруком бандите») // Автоматика и телемеханика. 2011. №5. С.127-138.
4. Kolnogorov A.V. Determination of the Minimax Risk for the Normal Two-Armed Bandit // Proceedings of the IFAC Workshop ALCOSP'2010, Antalya, Turkey, August 26-28, 2010 — http://www.ifac-papersonline.net

Bibliography (Translitirated)

1. Sragovich V.G. Adaptivnoe upravlenie. M.: Nauka, 1981. 384 s.
2. Berry D.A., Fristedt B. Bandit Problems: Sequential Alloca-tion of Experiments. L.; N.Y.: Chapman and Hall, 1985. 275 p.
3. Kolnogorov A.V. Nakhozhdenie minimaksnykh strategii i riska v sluchajjnojj srede (v zadache o «dvurukom bandite») // Avtomatika i telemekhanika. 2011. №5. S.127-138.
4. Kolnogorov A.V. Determination of the Minimax Risk for the Normal Two-Armed Bandit // Proceedings of the IFAC Workshop ALCOSP'2010, Antalya, Turkey, August 26-28, 2010 — http://www.ifac-papersonline.net