

УДК 519.865

## ОБ ОДНОМ ПОДХОДЕ К АСИМПТОТИЧЕСКОЙ ОЦЕНКЕ МИНИМАКСНОГО РИСКА В СТАЦИОНАРНОЙ СРЕДЕ

А.В.Колногоров

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

Для определения минимаксных стратегии и риска в задаче об оптимальном поведении в случайной среде использован теоретико-игровой подход. В этом случае минимаксная стратегия определяется как байесовская, соответствующая наихудшему априорному распределению. Описаны свойства наихудшего априорного распределения — симметричность и асимптотическая однородность. Приведены рекуррентные интегро-разностные уравнения для вычисления байесовского риска.

**Ключевые слова:** минимаксный подход, байесовский подход, адаптивное уравнение, поведение в случайной среде

The ideas of the theory of games are used for determination of the minimax strategy and risk in the problem of the optimal behavior in random environment. In this case the minimax strategy is calculated as the Bayes one corresponding to the worst priori distribution. The properties of the worst priori distribution, such as symmetry and asymptotic uniformity, are described. The recurrent integro-difference equations for calculation the Bayes risk are given.

**Keywords:** minimax approach, Bayesian approach, adaptative equation, behavior in random environment

### Определение объекта, стратегии и цели управления

Рассматривается задача адаптивного управления в стационарной случайной среде [1], известная также как задача о двуруком бандите [2], в следующей постановке. Пусть  $\xi_n$ ,  $n=1, \dots, N$ , есть управляемый случайный процесс, значения которого интерпретируются как доходы, зависят только от выбираемых в текущие моменты времени вариантов  $y_n$  и имеют нормальные распределения с плотностями

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

если  $y_n = \ell$  ( $\ell=1,2$ ). Такая среда полностью описывается векторным параметром  $\theta = (m_1, m_2)$ . Предполагается, что значение параметра  $\theta$  неизвестно, а известно только допустимое множество его значений  $\Theta = \{(m_1, m_2) : |m_1 - m_2| \leq 2c_1, |m_1 + m_2| \leq 2c_2\}$ , где  $c_1, c_2 < \infty$ . Для управления используется стратегия  $\sigma$ , которая в момент времени  $n$  является измеримой

функцией предыстории процесса, т.е. полученных откликов  $x^{n-1} = x_1, \dots, x_{n-1}$  на выбранные варианты  $y^{n-1} = y_1, \dots, y_{n-1}$ . Таким образом

$$\Pr(y_n = \ell | y^{n-1}, x^{n-1}) = \sigma_\ell(y^{n-1}, x^{n-1}).$$

Множество стратегий обозначим  $\Sigma$ .

Сформулируем минимаксную и байесовскую цели управления. Если бы параметр  $\theta$  был известен, то всегда следовало бы применять вариант, которому соответствует большая из величин  $m_1$ ,  $m_2$ , и полный ожидаемый доход равен в этом случае  $N(m_1 \vee m_2)$ . Если же параметр неизвестен, то функция

$$L_N(\sigma, \theta) = E_{\sigma, \theta} \left( \sum_{n=1}^N ((m_1 \vee m_2) - \xi_n) \right)$$

характеризует потери ожидаемого дохода вследствие неполноты информации. Здесь  $E_{\sigma, \theta}$  обозначает математическое ожидание по мере, порожденной стратегией  $\sigma$  и параметром  $\theta$ . При минимаксном подходе ве-

личина максимальных потерь на множестве параметров  $\Theta$  минимизируется по множеству стратегий  $\Sigma$ , соответствующая величина  $R_N^M(\Theta) = \inf_{\Sigma} \sup_{\Theta} L_N(\sigma, \theta)$

называется минимаксным риском, а обеспечивающая ее значение стратегия — минимаксной стратегией. При байесовском подходе следует минимизировать значение функции потерь, усредненное относительно априорного распределения  $\Lambda$ . Соответствующее значение  $R_N^B(\Lambda) = \inf_{\Sigma} \int_{\Theta} L_N(\sigma, \theta) \Lambda(d\theta)$  называется байесовским риском, обеспечивающая ее стратегия — байесовской.

Байесовский подход является намного более распространенным, чем минимаксный (см., напр., [2]). Это объясняется тем, что для вычисления байесовского риска можно написать рекуррентные уравнения, которые решаются методом динамического программирования и позволяют точно определить байесовскую стратегию для любого априорного распределения численными методами. В то же время байесовский подход много раз подвергался критике, поскольку ясных критериев для выбора априорного распределения, как правило, предложить нельзя. В этом смысле минимаксный подход является более предпочтительным, однако, его недостатком является сложность вычисления минимаксных стратегии и риска. В данной статье рассматривается метод их определения как байесовских, соответствующих наилучшему априорному распределению.

**О связи минимаксного и байесовского подходов**

Связь между минимаксным и байесовским подходами устанавливает основная теорема теории игр, которая в нашем случае имеет следующий вид.

*Теорема 1.* На множествах  $\Sigma$ ,  $\Theta$  можно определить метрики  $\rho_{\Sigma}$ ,  $\rho_{\Theta}$ , которые превращают их в компактные метрические пространства. Функция потерь  $L_N(\sigma, \theta)$  непрерывна по  $\sigma$ ,  $\theta$  в этих метриках. Множество  $\Sigma$  содержит смешанные стратегии. В этом случае минимаксная стратегия  $\sigma^0$  существует и совпадает с байесовской на наилучшем априорном распределении  $\Lambda^0$ , т.е.

$$R_N^M(\Theta) = \sup_{\Theta} L_N(\sigma^0, \theta) = \int_{\Theta} L_N(\sigma^0, \theta) \Lambda^0(d\theta) = R_N^B(\Lambda^0) = \sup_{\{\Lambda\}} R_N^B(\Lambda).$$

Из непрерывности  $L_N(\sigma, \theta)$  следует, что при вычислении байесовского риска априорные распределения можно сколь угодно точно приблизить распределениями, имеющими плотность. Байесовский риск, вычисленный относительно априорного распределения с плотностью  $\lambda$  будем обозначать  $R_N^B(\lambda)$ .

Таким образом, определение минимаксного риска сводится к вычислению байесовского риска для наилучшего априорного распределения. Некоторые дополнительные свойства байесовского риска даются

следующей теоремой.

*Теорема 2.* Байесовский риск является непрерывной и выпуклой вверх функцией априорного распределения, т.е. справедливы неравенства

$$|R_N^B(\lambda_1) - R_N^B(\lambda_2)| \leq Nc_1 \int_{\Theta} |\lambda_1(\theta) - \lambda_2(\theta)| d\theta, \quad (2)$$

$$R_N^B(\alpha_1\lambda_1 + \alpha_2\lambda_2) \geq \alpha_1 R_N^B(\lambda_1) + \alpha_2 R_N^B(\lambda_2) \quad (3)$$

для любых  $\lambda_1, \lambda_2$  и положительных чисел  $\alpha_1, \alpha_2$  таких, что  $\alpha_1 + \alpha_2 = 1$ .

Приведем рекуррентные уравнения, позволяющие вычислять байесовский риск методом динамического программирования. Положим

$$f_n(X | m) = \begin{cases} (2\pi n)^{-1/2} \exp(-(X - mn)^2 / (2n)) & \text{при } n \geq 1, \\ 1 & \text{при } n = 0. \end{cases}$$

Пусть к моменту времени  $n$  оба варианта применены соответственно  $n_1$  и  $n_2$  раз, а полные доходы за их применение равны  $X_1$  и  $X_2$ . Тогда плотность апостериорного распределения дается формулой

$$\lambda(m_1, m_2 | X_1, n_1, X_2, n_2) = \frac{f_{n_1}(X_1 | m_1) f_{n_2}(X_2 | m_2) \lambda(m_1, m_2)}{\int_{\Theta} f_{n_1}(X_1 | m_1) f_{n_2}(X_2 | m_2) \lambda(m_1, m_2) dm_1 dm_2}.$$

Обозначим через  $R_{N-n_1-n_2}^B(\lambda; X_1, n_1, X_2, n_2)$  байесовский риск, вычисленный на последних  $N - n_1 - n_2$  шагах относительно апостериорного распределения с плотностью  $\lambda(m_1, m_2 | X_1, n_1, X_2, n_2)$ . Тогда

$$R_{N-n_1-n_2}^B(\cdot) = \min(R_{N-n_1-n_2}^{(1)}(\cdot), R_{N-n_1-n_2}^{(2)}(\cdot)),$$

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

$$R_{N-n_1-n_2}^{(1)}(\lambda; X_1, n_1, X_2, n_2) = \int_{\Theta} ((m_2 - m_1)^+ + E_x^{(1)} R_{N-n_1-n_2-1}^B(\lambda; X_1 + x, n_1 + 1, X_2, n_2)) \times \lambda(m_1, m_2 | X_1, n_1, X_2, n_2) dm_1 dm_2; \quad (4)$$

$$R_{N-n_1-n_2}^{(2)}(\lambda; X_1, n_1, X_2, n_2) = \int_{\Theta} ((m_1 - m_2)^+ + E_x^{(2)} R_{N-n_1-n_2-1}^B(\lambda; X_1, n_1, X_2 + x, n_2 + 1)) \times \lambda(m_1, m_2 | X_1, n_1, X_2, n_2) dm_1 dm_2; \quad (5)$$

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

Суть уравнений (4), (5) для вычисления  $R_{N-n_1-n_2}^{(\ell)}(\lambda; X_1, n_1, X_2, n_2)$  состоит в том, что они описывают ожидаемый доход, если на первом шаге применяется вариант с номером  $\ell$ , а далее применяется оптимальная байесовская стратегия. При этом байесовская стратегия предписывает на шаге с номером  $n_1 + n_2$  выбирать вариант с номером  $\ell$ , соответствующим меньшему из значений  $R_{N-n_1-n_2}^{(\ell)}(\lambda; X_1, n_1, X_2, n_2)$ ,  $\ell = 1, 2$ . При равенстве обоих значений выбор варианта может быть произвольным.

**О свойствах наихудшего априорного распределения**

Нормальная плотность распределения доходов (1) позволяет выделить такие преобразования  $\hat{\lambda}$  априорной плотности распределения  $\lambda$ , которые не меняют байесовский риск, т.е.  $R_N(\hat{\lambda}) = R_N(\lambda)$ .

*Теорема 3.* Следующие преобразования априорной плотности распределения не меняют значения байесовского риска:

$$\begin{aligned} \hat{\lambda}(m_1, m_2) &= \lambda(m_2, m_1) \quad (\text{для любых } m_1, m_2); \\ \hat{\lambda}(m_1, m_2) &= \lambda(-m_1, -m_2) \quad (\text{для любых } m_1, m_2); \\ \hat{\lambda}(m_1, m_2) &= \lambda(m_1 + x, m_2 + x) \quad (\text{для любых } m_1, m_2 \text{ и любого фиксированного } x). \end{aligned}$$

Доказательство теоремы может быть получено методом математической индукции с использованием уравнений (4), (5). Например, чтобы обосновать первое преобразование, отметим, что справедливы выполняющиеся для всех предысторий  $(X_1, n_1, X_2, n_2)$  равенства

$$\begin{aligned} R_{N-n_1-n_2}^{(1)}(\hat{\lambda}; X_1, n_1, X_2, n_2) &= R_{N-n_1-n_2}^{(2)}(\lambda; X_2, n_2, X_1, n_1), \\ R_{N-n_1-n_2}^{(2)}(\hat{\lambda}; X_1, n_1, X_2, n_2) &= R_{N-n_1-n_2}^{(1)}(\lambda; X_2, n_2, X_1, n_1), \end{aligned}$$

и, следовательно,

$$R_{N-n_1-n_2}^B(\hat{\lambda}; X_1, n_1, X_2, n_2) = R_{N-n_1-n_2}^B(\lambda; X_2, n_2, X_1, n_1),$$

откуда требуемое свойство получается при  $n_1 = n_2 = 0$ .

*Определение.* Плотности распределения из последовательности  $\{\lambda_N\}$ , для которых выполнено условие  $R_N(\lambda_N) \sim \sup_{\{\lambda\}} R_N(\lambda)$ , назовем асимптотически наихудшими.

Справедлива следующая теорема.

*Теорема 4.* Асимптотически наихудшая плотность априорного распределения может быть выбрана в виде

$$\lambda_N(m_1, m_2) = 0,5\rho_N(0,5|m_1 - m_2|) \cdot v_a(0,5(m_1 + m_2)), \quad (6)$$

где  $v_a(\cdot)$  — плотность равномерного распределения на отрезке  $0,5|m_1 + m_2| \leq a$  для произвольного достаточно большого  $a \gg c_2$ .

Действительно, если  $\lambda^0(m_1, m_2)$  — плотность наихудшего априорного распределения, то (с учетом выпуклости байесовского риска (2))  $\lambda_N(m_1, m_2)$  можно взять в виде

$$\lambda_N(m_1, m_2) = (4a)^{-1} \int_{-a}^a (\lambda^0(m_1 + x, m_2 + x) + \lambda^0(m_2 + x, m_1 + x)) dx,$$

при этом будем считать, что  $a \gg c_2$ . С учетом непрерывности байесовского риска (3) такая плотность может быть сколь угодно точно приближена симметрической асимптотически однородной плотностью (6).

*Следствие.* Асимптотически наихудшая плотность априорного распределения может быть выбрана в виде

$$\lambda_N(m_1, m_2 | X) = 0,5\pi^{-1/2} \rho_N(0,5|m_1 - m_2|) \exp(-0,5(m_1 + m_2) - X)^2, \quad (7)$$

где  $|X| \leq a - c_2$ ,  $a \gg c_2$ .

Действительно, пусть оба варианта применены по очереди, и полученный полный доход равен  $2X$ .

Обозначим  $m_1 + m_2 = 2m$ . Тогда апостериорная плотность для второго сомножителя в (6) при  $|X| \leq a - b$ ,  $b \gg 1$ , равна

$$v(m | X) = \frac{f_2(2X | m)v_a(m)}{\int_{-a}^a f_2(2X | m)v_a(m)dm} \approx \pi^{-1/2} \exp(-(m - X)^2)$$

Так как плотность  $\pi^{-1/2} \exp(-0,5(m_1 + m_2) - X)^2$  не меняется при одновременной замене  $\hat{m}_1 = m_1 + x$ ,  $\hat{m}_2 = m_2 + x$ ,  $\hat{X} = X + x$ , то в силу теоремы 3 байесовский риск для всех плотностей вида (7) имеет одно и то же значение.

Что касается множителя  $\rho_N(0,5|m_1 - m_2|)$ , то результаты численных расчетов [3] показывают, что он сосредоточен на множестве параметров вида  $\{\theta : m_1 - m_2 = \pm 2dN^{-1/2}\}$ . Это подтверждает представленную в [4] асимптотическую оценку, в которой порядок минимаксного риска установлен равным  $N^{1/2}$ .

**Интегро-разностные уравнения для вычисления байесовского риска относительно наихудшего априорного распределения**

Для вычислений удобно асимптотически наихудшее априорное распределение взять в виде (7). Положим

$$\begin{aligned} p(X_1, n_1, X_2, n_2) &= \iint_{\Theta} f_{n_1}(X_1 | m_1) f_{n_2}(X_2 | m_2) \lambda_N(m_1, m_2 | 0) dm_1 dm_2, \\ r_{N-n_1-n_2}(X_1, n_1, X_2, n_2) &= R_{N-n_1-n_2}^B(X_1, n_1, X_2, n_2) \cdot p(X_1, n_1, X_2, n_2). \end{aligned}$$

С использованием уравнений (4), (5) после несложных преобразований получим

$$r_{N-n_1-n_2}(\cdot) = \min(r_{N-n_1-n_2}^{(1)}(\cdot), r_{N-n_1-n_2}^{(2)}(\cdot)),$$

где  $r_0^{(1)}(\cdot) = r_0^{(2)}(\cdot) = 0$ ;

$$\begin{aligned} r_{N-n_1-n_2}^{(1)}(X_1, n_1, X_2, n_2) &= \int_0^\infty 2vg(X_1, n_1, X_2, n_2, v) \rho_N(v) dv + \\ &+ \int_{-\infty}^\infty r_{N-n_1-n_2-1}(X_1 + x, n_1 + 1, X_2, n_2) q_{n_1}(X_1 - n_1 x) dx; \end{aligned}$$

$$\begin{aligned} r_{N-n_1-n_2}^{(2)}(X_1, n_1, X_2, n_2) &= \int_0^\infty 2vg(X_1, n_1, X_2, n_2, -v) \rho_N(v) dv + \\ &+ \int_{-\infty}^\infty r_{N-n_1-n_2-1}(X_1, n_1, X_2 + x, n_2 + 1) q_{n_2}(X_2 - n_2 x) dx. \end{aligned}$$

Здесь

$$g(X_1, n_1, X_2, n_2, v) = \begin{cases} 1 & \text{при } n_1 = 0, n_2 = 0, \\ \frac{1}{\sqrt{\pi n_2(n_2 + 2)}} \exp\left(-\frac{n_2(\bar{X}_2 - v)^2}{n_2 + 2}\right) & \text{при } n_1 = 0, n_2 > 0, \\ \frac{1}{\sqrt{\pi n_1(n_1 + 2)}} \exp\left(-\frac{n_1(\bar{X}_1 + v)^2}{n_1 + 2}\right) & \text{при } n_2 = 0, n_1 > 0. \end{cases}$$

$$g(X_1, n_1, X_2, n_2, v) = \frac{1}{\pi \sqrt{n_1 n_2 (n_1 + n_2 + 2)}} \times \\ \times \exp\left(-\frac{n_1 n_2 (\bar{X}_1 - \bar{X}_2 + 2v)^2 + 2n_1 (\bar{X}_1 + v)^2 + 2n_2 (\bar{X}_2 - v)^2}{2(n_1 + n_2 + 2)}\right)$$

при  $n_1 > 0$ ,  $n_2 > 0$ . Через  $\bar{X}_1$ ,  $\bar{X}_2$  обозначены средние значения  $\bar{X}_1 = X_1 / n_1$ ,  $\bar{X}_2 = X_2 / n_2$ ,

$$q_n(x) = \begin{cases} \sqrt{\frac{n+1}{2\pi n}} \exp\left(-\frac{x^2}{2n(n+1)}\right) & \text{при } n \geq 1, \\ 1 & \text{при } n = 0. \end{cases}$$

### Заключение

В работе представлен теоретико-игровой подход к асимптотической оценке минимаксного риска в стационарной среде. Согласно этому подходу мини-

максный риск численно равен байесовскому, вычисленному на наихудшем априорном распределении. Описаны свойства наихудшего априорного распределения — симметричность и асимптотическая однородность. Приведены интегро-разностные уравнения для вычисления байесовского риска относительно наихудшего априорного распределения методом динамического программирования.

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. Асимптотические оценки байесовского риска для одного класса стационарных сред // Третья Междунар. конф. по проблемам управления: Пленарные докл. и избр. тр. М.: ИПУ РАН, 2006. С.241-248.
4. Vogel W. // Ann. Math. Stat. 1960.V.31. P.444-451.