

А.В.Колногоров, С.В.Мельникова

# ОБ ОПТИМАЛЬНОЙ ПРОДОЛЖИТЕЛЬНОСТИ НАЧАЛЬНОГО ЭТАПА В ДВУХЭТАПНОЙ МОДЕЛИ ЦЕЛЕСООБРАЗНОГО ПОВЕДЕНИЯ В СЛУЧАЙНОЙ СРЕДЕ

*Новгородский государственный университет им. Ярослава Мудрого*

The two-stage model of expedient behavior in random medium is generalized to the case of three alternative variants of management. Asymptotic estimations of the minimax risk and the optimal duration of the initial stage are given.

Рассматривается задача о целесообразном поведении в случайной среде, известная также как задача о «многоруком бандите» (см. [1-3]). Это задача об оптимальном последовательном выборе альтернативных вариантов, осуществляемом на основе текущей информации. Она может иметь различные интерпретации, рассмотрим одну из них в предположении, что число таких вариантов равно трем.

Предположим, что существует группа из  $N$  пациентов и три вида альтернативных лекарств. Будем считать, что результат лечения пациента с номером  $n$ , который обозначим  $X_n$ ,  $n = 1, 2, \dots, N$ , допускает измерение и имеет нормальное распределение с параметрами  $(m_i, 1)$  в случае применения лекарства с номером  $i$ . Таким образом, случайная среда характеризуется векторным параметром  $\theta = (m_1, m_2, m_3)$ . Если бы все значения  $m_i$ ,  $i = 1, 2, 3$ , были известны, то для лечения следовало бы выбирать лекарство с наибольшим значением  $m_i$ . В случае если значения  $m_i$ ,  $i = 1, 2, 3$ , неизвестны, предлагается двухэтапный подход. На первом этапе выбираются три пробные группы по  $M$  человек, и в каждой группе для лечения применяется свое лекарство. После курса лечения результаты во всех группах сравниваются, и для лечения оставшихся  $N - 3 \cdot M$  пациентов на заключительном этапе выбирается то лекарство, применение которого дало наилучший результат на начальном этапе. Отметим, что эта процедура обобщает процедуру, рассмотренную в [4].

Введем обозначения:  $S_N = \sum_{n=1}^N X_n$ ,  $E_{\theta, M} S_N$  — общий результат лечения  $N$  пациентов и его математическое ожидание при выбранном параметре  $\theta$  и размере группы  $M$ ;  $N \cdot \max(m_1, m_2, m_3)$  — математическое ожидание наилучшего результата лечения, если известна полная информация относительно  $\theta$ ;  $L_N(\theta, M) = N \cdot \max(m_1, m_2, m_3) - E_{\theta, M} S_N$  — функция потерь.

Задача состоит в определении минимаксного риска

$$R_N = \min_M \max_{\theta} L_N(\theta, M) \quad (1)$$

и соответствующего ему оптимального числа  $M$  пациентов в группах на начальном этапе. Предположим, что  $m_1 < m_2 < m_3$ , т.е. наиболее эффективно третье лекарство. Рассмотрим случайные величины, характеризующие результаты лечения на начальном этапе:

$$Y_1 = \sum_{n=1}^M X_n, \quad Y_2 = \sum_{n=M+1}^{2M} X_n, \quad Y_3 = \sum_{n=2M+1}^{3M} X_n.$$

Отметим, что они независимы и имеют нормальные распределения со следующими параметрами:

$$Y_1 \sim N(m_1 M, M), \quad Y_2 \sim N(m_2 M, M), \quad Y_3 \sim N(m_3 M, M).$$

Найдем значения функции потерь на начальном и заключительном этапах. Обозначим  $\Delta m_i = m_3 - m_i$ ,  $i = 1, 2$ . На начальном этапе математическое ожидание потерь функции потерь равно  $3m_3M - M(m_1 + m_2 + m_3) = M(\Delta m_1 + \Delta m_2)$ , а его максимальное значение

$$L_N^{(1)}(M) = \max_{\theta} L_N^{(1)}(\theta, M) = M \cdot \Delta, \quad (2)$$

где  $\Delta = \max_{\theta}(\Delta m_1 + \Delta m_2)$ . Математическое ожидание потерь на заключительном этапе:

$$L_N^{(2)}(\theta, M) = (N - 3M)(\Delta m_1 \cdot P(Y_1 \geq Y_2, Y_1 \geq Y_3) + \Delta m_2 \cdot P(Y_2 \geq Y_1, Y_2 \geq Y_3)). \quad (3)$$

Найдем вероятность  $P(Y_1 \geq Y_2, Y_1 \geq Y_3)$  того, что после первого этапа максимальный результат лечения будет в первой группе и для дальнейшего лечения ошибочно будет выбрано первое лекарство. Для этого обозначим  $U_i = Y_i - m_3M$ ,  $i = 1, 2, 3$ . Отметим, что все  $U_i$ ,  $i = 1, 2, 3$ , независимы и нормально распределены. Тогда

$$P(Y_1 \geq Y_2, Y_1 \geq Y_3) = P(U_1 \geq U_2, U_1 \geq U_3) =$$

$$= \frac{1}{(2\pi)^{3/2} \cdot M^{3/2}} \iiint_{\substack{u_1 \geq u_2 \\ u_1 \geq u_3}} e^{-\frac{(u_1 + \Delta m_1 \cdot M)^2}{2M}} \cdot e^{-\frac{(u_2 + \Delta m_2 \cdot M)^2}{2M}} \cdot e^{-\frac{u_3^2}{2M}} du_1 du_2 du_3.$$

Сделаем замену переменных:  $u_i = z_i \cdot M^{1/2}$ ,  $i = 1, 2, 3$ , и обозначим  $\Delta m_1 = a \cdot M^{-1/2}$ ,  $\Delta m_2 = b \cdot M^{-1/2}$ . Получаем

$$\begin{aligned} P(Y_1 \geq Y_2, Y_1 \geq Y_3) &= \frac{1}{(2\pi)^{3/2}} \iiint_{\substack{z_1 \geq z_2 \\ z_1 \geq z_3}} e^{-\frac{(z_1 + a)^2}{2}} \cdot e^{-\frac{(z_2 + b)^2}{2}} \cdot e^{-\frac{z_3^2}{2}} dz_1 dz_2 dz_3 = \\ &= \frac{1}{(2\pi)^{3/2}} \int_{-\infty}^{+\infty} e^{-\frac{(z_1 + a)^2}{2}} dz_1 \int_{-\infty}^{z_1} e^{-\frac{(z_2 + b)^2}{2}} dz_2 \int_{-\infty}^{z_1} e^{-\frac{z_3^2}{2}} dz_3. \end{aligned}$$

Используя функцию нормального распределения  $\Phi(t) = (2\pi)^{-1/2} \int_{-\infty}^t e^{-\frac{z^2}{2}} dz$ , повторный интеграл запишем в виде

$$P(Y_1 \geq Y_2, Y_1 \geq Y_3) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} e^{-\frac{(z_1 + a)^2}{2}} \cdot \Phi(z_1 + b) \cdot \Phi(z_1) \cdot dz_1 = F(a, b). \quad (4)$$

Аналогично можно показать, что

$$P(Y_2 \geq Y_1, Y_2 \geq Y_3) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} e^{-\frac{(z_2 + b)^2}{2}} \cdot \Phi(z_2 + a) \cdot \Phi(z_2) \cdot dz_2 = F(b, a). \quad (5)$$

С учетом (3)-(5) математическое ожидание потерь на втором этапе

$$L_N^{(2)}(\theta, M) = (N - 3M) \cdot M^{-1/2} \cdot \Psi(a, b),$$

где  $\Psi(a, b) = aF(a, b) + bF(b, a)$ , а его максимальное значение

$$L_N^{(2)}(M) = \max_{\theta} L_N^{(2)}(\theta, M) = (N - 3M) \cdot M^{-1/2} \Psi_0, \quad (6)$$

где  $\Psi_0 = \max_{a>0, b>0} \Psi(a, b) \approx 0,372$  при  $a = b \approx 1,165$ . Из (2), (6) аналогично тому, как это сделано в [4], можно показать, что оптимальная величина  $M$  определяется из равенства  $L_N^{(1)}(M) = L_N^{(2)}(M)$  и, следовательно, допускает оценку

$$M \sim \left( \frac{\Psi_0}{\Delta} \right)^{2/3} N^{2/3}.$$

В свою очередь минимаксный риск (1) допускает оценку

$$R_N \sim (\Psi_0^2 \Delta)^{1/3} N^{2/3}.$$

Таким образом, минимаксный риск и оптимальная величина начального этапа имеют порядок  $N^{2/3}$ .

---

1. Срагович В.Г. Теория адаптивных систем. М.: Наука, 1976. 319 с.
2. Назин А.В., Позняк А.С. Адаптивный выбор вариантов. М.: Наука, 1986. 288 с.
3. Berry D.A., Fristedt B. Bandit Problems: Sequential Allocation of Experiments. L., N.Y. Chapman and Hall, 1985. 275 p.
4. Колногоров А.В. // Проблемы передачи информации. 2000. №4. С.117-127.