# ПОСЛЕДОВАТЕЛЬНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ В МОНИТОРИНГЕ МЕДИЦИНСКИХ ДАННЫХ

**А. Ю. Харин**

*Белорусский государственный университет*
*Минск, Беларусь*
*E-mail: KharinAY@bsu.by*

Рассматривается задача статистического анализа медицинских данных с целью наискорейшего в среднем принятия решения с заданной точностью об индивидуальной эффективности предписанной пациенту методики лечения. Построено и исследовано последовательное статистическое решающее правило. Результаты применены при лечении и анализе данных по кардиологическим заболеваниям.

*Ключевые слова*: медицинская диагностика, статистическое решающее правило, последовательный анализ.

## Введение

В медицинской диагностике актуальна задача наискорейшего выявления индивидуальной эффективности предписанной методики лечения [5–7].

Принятие решений основывается на анализе статистических данных результатов медицинской диагностики состояний пациента в динамике. В докладе для анализа таких данных применяется последовательный статистический подход [1–3], позволяющий обеспечить оптимальность информационных технологий принятия решений в медицинской диагностике. Последовательный подход интенсивно используется для анализа универсальной эффективности тестируемого препарата или методики лечения [4]. В работе рассматривается задача анализа индивидуальной эффективности выбранной методики лечения для конкретного пациента, позволяющая наискорейшим образом принять решение о необходимости ее замены (когда с заданной точностью может быть принято решение о недостаточной эффективности) или снизить затраты и уменьшить неудобства, связанные с необходимостью частых обследований состояния пациента для контроля эффективности лечения.

## Математическая модель и краткое изложение результатов

Примем следующие обозначения. Пусть $s \in \mathbf{Z}_+$ – текущий момент времени; $x_s \in \mathbf{R}^N$ – ненаблюдаемый вектор информативных признаков, характеризующих состояние пациента в момент времени *s*; $z_s \in \mathbf{R}_+$ – накопленная доза препарата, полученная пациентом к моменту времени *s*; $\xi_s \in \mathbf{R}^N$ – случайный вектор ошибок в измерении значений признаков; $y_s \in \mathbf{R}^N$ – случайный вектор наблюдений в момент времени *s*; $i \in \mathbf{N}$ – порядковый номер текущей консультации; $n_s$ – порядковый номер последней консультации, предшествовавшей моменту времени *s*; $t_i \in \mathbf{Z}_+$ – момент времени, когда проводилась *i*-я по счету консультация ($t_1 := 0$); $d_i \in \{0, 1, 2\}$ – решение, принятое на *i*-й по счету консультации. Решение $d_i = 0$ соответствует принятию гипотезы об эффективности используемой методики лечения для данного пациента, решение $d_i = 1$ означает необходимость модификации методики, решение $d_i = 2$ означает, что требуемая точность (заданные малые допустимые значения вероятностей ошибочных решений) не может быть обеспечена, и требуется сделать следующее наблюдение. Пусть $u_i \in U = \{u^1, u^2, u^3\}$ – доза препарата для приема (интенсивность методики лечения), предписанная на *i*-й консультации; $\tau_i \in T = \{\tau^1, \tau^2, \tau^3\}$ – предписанный на *i*-й консультации интервал времени до следующей консультации.

Обозначим через $a, b \in \mathbf{R}^N$ неизвестные значения векторов параметров, характеризующих состояние пациента в анализируемой схеме лечения. Пусть $X_0 \subset \mathbf{R}^N$ – подмножество пространства информативных признаков, соответствующее состоянию «здоровый пациент»; $H_0$ – гипотеза, состоящая в том, что для конкретного пациента с хроническим кардиологическим заболеванием выбранная схема лечения эффективна и «приведет» этого пациента в состояние из $X_0$ («здоров»); $H_1 = \overline{H}_0$ – альтернатива.

Математическую модель зависимости введенных показателей представим соотношением:

$$x_s = a + b \cdot z_s, \; y_s = x_s + \xi_s, \tag{1}$$

где

$$\mathrm{E}\{\xi_s\} = \mathbf{0}_N, \; E\{\xi_s^T \cdot \xi_s\} = \Sigma, s \in \mathbf{N};$$

$$z_s = \sum_{i=1}^{n_s - 1} u_i \cdot \tau_i + u_{n_s} \cdot \left( s - \sum_{i=1}^{n_s - 1} \tau_i \right), \; s \in \mathbf{N}; \tag{2}$$

$$t_1 = 0, \; t_{i+1} = t_i + \tau_i, \; i \in \mathbf{N}.$$

Гипотезы сформулируем следующим образом:

$$H_j: \; (a,b) \in \Delta_j \subset \mathbf{R}^N \times \mathbf{R}^N, \; j = 0,1, \; \Delta_0 \cap \Delta_1 = \varnothing.$$

Рассмотрим гауссовскую модель наблюдений:

$$N = 1, \; L\{\xi_s\} = N_1(0, \sigma^2). \tag{3}$$

Гипотезам $H_0$, $H_1$ соответствуют множества

$$\Delta_0 = \left\{(a,b) \in R^2 : b \cdot sign(a - x^*) < 0\right\},$$

$$\Delta_1 = \left\{(a,b) \in R^2 : b \cdot sign(a - x^*) \geq 0\right\}. \tag{4}$$

**Теорема.** *Для модели (1) – (4) логарифм функции правдоподобия, вычисленной по наблюдениям* $y_1, \ldots, y_{t_i}$, $i \in \mathbf{N}$*, равен*

$$l_{t_i}(a,b;\sigma^2) = -\frac{1}{2}\left( t_i \cdot \ln(2\pi\sigma^2) + \frac{1}{\sigma^2} \sum_{s=0}^{t_i} (y_s - (a + b \cdot z_s))^2 \right).$$

**Следствие.** *В рамках модели (1) – (4) статистика обобщенного отношения правдоподобия в момент времени, когда проводится* $i$ *консультация, имеет вид*

$$
\Lambda_{t_i} = \frac{\sup_{a,b\in\Delta_1,\sigma^2>0} e^{-\frac{1}{2}\left(t_i\cdot\ln(2\pi\sigma^2)+\frac{1}{\sigma^2}\sum_{s=0}^{t_i}(y_s-(a+bz_s))^2\right)}}{\sup_{a,b\in\Delta_0,\sigma^2>0} e^{-\frac{1}{2}\left(t_i\cdot\ln(2\pi\sigma^2)+\frac{1}{\sigma^2}\sum_{s=0}^{t_i}(y_s-(a+bz_s))^2\right)}}.
$$

Таким образом, для рассмотренной модели последовательное статистическое решающее правило проверки гипотез $H_0$, $H_1$ имеет вид: принимаемое на $i$-й консультации решение равно

$$
d_i = \begin{cases} 0, \ \Lambda_{t_i} < \frac{\beta_0}{1-\alpha_0}, \\ 1, \ \Lambda_{t_i} > \frac{1-\beta_0}{\alpha_0}, \\ 2, \ \Lambda_{t_i} \in \left[\frac{\beta_0}{1-\alpha_0}, \frac{1-\beta_0}{\alpha_0}\right]. \end{cases} \tag{5}
$$

Построенное решающее правило (5) исследовано на модельных данных, а также применено для анализа реальных статистических данных по кардиологическим заболеваниям.

Исследования выполнены в рамках ГПНИ «Информатика и космос», а также частично поддержаны грантом МНТЦ (проект B-1910).

## Библиографические ссылки

1. *Wald A.* Sequential analysis. New York : Wiley, 1947.
2. *Lai T. L.* Sequential analysis: some classical problems and new challenges // Statistica Sinica. 2001. V. 11. P. 303–408.
3. *Ghosh B., Sen P. K.* Handbook of sequential analysis. New York : Marcel Dekker, 1991.
4. *Jennison C., Turnbull B.* Group sequential methods with applications to clinical trials. Boca Raton : Chapman and Hall / CRC, 2000.
5. *Mukhopadhyay N., Datta S., Chattopadhyay S.* Applied Sequential Methodologies. New York : Marcel Dekker, 2004.
6. *Kharin A.* Robustness analysis for Bayesian sequential testing of composite hypotheses under simultaneous distortions of priors and likelihoods // Austrian Journal of Statistics. 2011. V. 40. P. 65–73.
7. *Kharin A.* Robustness in sequential hypotheses testing applied to statistical monitoring of medical data // Proceedings of the 15th International Conference on Applied Stochastic Models and Data Analysis. 2013. Barcelona. 2013. P. 75–81.