AN APPROACH TO STATISTICAL DECISION MAKING IN MEDICAL DIAGNOSTICS

A. Kharin

Belarusian State University
Minsk, Belarus
E-mail: KharinAY@bsu.by

The paper is devoted to the construction of a mathematical model for the process of medical diagnostics. A family of controlled statistical decision rules is proposed and the optimality criterion is given. Methods of the extreme problem solution are discussed.

Key words: medical diagnostics, statistical decision rule, sequential analysis, control, optimality.

INTRODUCTION

In medical diagnostics the problems of optimal decision making are known to be important. The decisions are usually based on the available statistical data. The sequential statistical approach [1] is an effective information technology that is not appropriate only because data are processed as they arrive (on-line data processing), but also because of optimal properties of sequential statistical decision rules [2, 3]. The sequential methodology is effectively used in clinical trials [4, 5].

In this paper we construct and discuss a mathematical model for sequential decision making in medical diagnostics.

MATHEMATICAL MODEL

Let Ω be the patients space, which is a set of patients under medical diagnostics; $D = \{0,1,2\}$ be the decision space; $P = \left\{ p(z;\theta), z \in \mathbf{R}^N : \theta \in \Theta \right\}$, $\Theta \subseteq \mathbf{R}^m$, be a parametric family of the N-dimensional probability density functions. Let $s \in \mathbf{N}$ be the current discrete time moment, when the medical analyses (tests) are taken during the medical diagnostics of the patient. For each patient $\alpha \in \Omega$, let $x_s = x_s(\omega) \in \mathbf{R}^N$ be the random vector of observations (its components are the results of the diagnostic analyses taken) with the probability density function $p(z;\theta_s)$, $z \in \mathbf{R}^N$, $\theta_s \in \mathbf{R}^m$. For example, the following situation can be considered: $m = N$, $x_s = \theta_s + \xi_s$, $s \in \mathbf{N}$, where ξ_s has the N-dimensional normal probability distribution with zero vector mean and the covariance matrix Σ. In the set Θ there is a subset $\Theta_0 \subset \Theta$ that corresponds to the hypothesis $H_0 : \theta \in \Theta_0$ meaning that the disease under diagnostics is not identified for the patient, so this patient may be considered as a healthy one. The alternative to H_0 is the hypothesis $\overline{H}_0$ meaning that the patient may not be considered as a healthy one. In frames of the hypothesis $\overline{H}_0$ let us define the hypothesis $H_1 : \theta \in \Theta_1$ meaning that the current therapy scheme (or program) is not effective for the patient and should be replaced, where $\Theta_1 \subset \Theta \setminus \Theta_0$ is the set of parameter values corresponding to the hypothesis H_1. The sets Θ_0, Θ_1 are determined a priori.

At the moment t, when an observation x_t is registered, one of the three decisions is possible: $d_t \in D$. The decisions $d_t \in \{0,1\}$ mean final acceptance of the correspondent hypotheses H_0 and H_1, in this case $S = t$ is the random final decision number. The decision $d_t = 2$ means that the observation process should be continued, as none of H_0 and H_1 may be reliably accepted at the moment t.

After the visit number t to the doctor, if $d_t = 2$, a patient gets the prescription $(u_t, \tau_t) \in U \times V$, where $u_t \in U = \left\{ u^1, \ldots, u^M \right\}$ is the control type (a dose of the remedy, for example), and $\tau_t \in \mathrm{T} = \left\{ \tau^1, \ldots, \tau^K \right\}$ is the number of cycles (therapy – medical tests) to be

made before the next visit to the doctor. So the time moments indicating visits to the doctor are:

$$s_1 = 1,\ s_2 = s_1 + \tau_1,\ \ldots,\ s_t = s_{t-1} + \tau_{t-1},\ \ldots,\ s_S = s_{S-1} + \tau_{S-1}.$$

If $M = 1$, then the problem is equivalent to the situation where no modification to the therapy is made, but the moment of the next diagnostics to be taken is under control. If $K = 1$, then the problem is reduced to the problem of statistical sequential testing of two hypotheses H_0 and H_1 [6] under non-homogeneous data.

The prescription (u_t, τ_t) is a function of observations: $(u_t, \tau_t) = F(x_1, \ldots, x_t; u_{t-1}, \tau_{t-1})$ forming the therapy control scheme. The control scheme $\{(u_t, \tau_t) : t = 1,2,\ldots\}$ is constructed by a doctor to ensure that the trajectory θ_s starting from the set $\Theta \setminus \Theta_0$ gets to the set Θ_0.

The decision rule is sequential and it is constructed at the step t in the form:

$$d_t = f(x_1, \ldots, x_{s_t}) : \mathbf{R}^{N \cdot s_t} \to D. \quad (1)$$

The decision rule (1) is constructed at the moment $s_t = s_{t-1} + \tau_{t-1}$ on the basis of the statistic of the generalized logarithmic likelihood ratio (statistic for the *t*-th decision step):

$$\Lambda_t = \ln \frac{\max_{\theta_1, \ldots, \theta_L \in \Theta_0} p(x_{s_1^*}; \theta_1) \cdots p(x_{s_{t-1}+\tau_{t-1}}; \theta_L)}{\max_{\theta_1, \ldots, \theta_L \in \Theta_1} p(x_{s_1^*}; \theta_1) \cdots p(x_{s_{t-1}+\tau_{t-1}}; \theta_L)},\ L = t_{i-1} + \tau_{i-1} - t_1^* + 1, \quad (2)$$

where $s_1^* \geq s_1$ is a parameter of the proposed statistical test.

OPTIMALITY CRITERION

According to (1), denote by

$$S = \min\left\{t : f(x_1, \ldots, x_{s_t}) \in \{0,1\}\right\} \quad (3)$$

the number of the final decision step in the control scheme.

For the performance evaluation of the decision rule (1), (3) the following characteristics are important: 1) the cost of medical diagnostics tests taken: $c_N \cdot E\left\{\sum_{i=1}^{S} \tau_i\right\}$, where c_N is the cost of one *N*-dimensional medical diagnostics test, the mathematical expectation $E\{\cdot\}$ is calculated with respect to the probability density function $p(x_1, \ldots, x_{\sum_{i=1}^{S} \tau_i})$; 2a) the cost of loss caused by the delay in the diagnostics of the hypothesis H_0: $w_0 \cdot E_{H_0}\left\{g_0\left(\sum_{i=1}^{S} \tau_i\right)\right\}$, where $w_0 > 0$ is a given coefficient; 2b) the cost of loss caused by the delay in the diagnostics of the hypothesis H_1: $w_1 \cdot E_{H_1}\left\{g_1\left(\sum_{i=1}^{S} \tau_i\right)\right\}$, where $w_1 > 0$ is a given coefficient; 3) the cost of loss for the errors in diagnostics:

$l_0 \cdot P_{H_0}\{d_S = 1\} + l_1 \cdot P_{H_1}\{d_S = 0\}$, where the values of l_0, l_1 are given. Here $g_0(\cdot)$, $g_1(\cdot)$ are some monotone increasing functions.

The optimality criterion for the decision rule (1) is considered in the form:

$$J(f(\cdot)) = c_N \cdot E\left\{\sum_{i=1}^{S} \tau_i\right\} + w_0 \cdot E_{H_0}\left\{g\left(\sum_{i=1}^{S} \tau_i\right)\right\} + w_1 \cdot E_{H_1}\left\{g\left(\sum_{i=1}^{S} \tau_i\right)\right\} + \tag{4}$$

$$+ l_0 \cdot P_{H_0}\{d_S = 1\} + l_1 \cdot P_{H_1}\{d_S = 0\} \to \min.$$

To find the optimal control policy (including the optimal decision rule (1) and the optimal control scheme $\{(u_t, \tau_t)\}$) in (4) the methods of sequential testing of hypotheses (2) are to be developed to be applicable for this problem setting. The problem of contamination of statistical data is also to be considered and the robust versions of sequential tests are to be constructed under the distortions that appear in this model [6].

REFERENCES

1. *Wald, A.* Sequential analysis / A. Wald. New York: Wiley, 1947.

2. *Lai, T. L.* Sequential analysis: some classical problems and new challenges / T. L. Lai // Statistica Sinica. 2001. Vol. 11. P. 303–408.

3. *Ghosh, B.* Handbook of sequential analysis / B. Ghosh, P. K. Sen. New York : Marcel Dekker, 1991.

4. *Jennison, C.* Group sequential methods with applications to clinical trials / C. Jennison, B. Turnbull. Boca Raton : Chapman and Hall / CRC, 2000.

5. *Mukhopadhyay, N.* Applied Sequential Methodologies / N. Mukhopadhyay, S. Datta, S. Chattopadhyay. New York : Marcel Dekker, 2004.

6. *Kharin, A.* Robustness analysis for Bayesian sequential testing of composite hypotheses under simultaneous distortions of priors and likelihoods / A. Kharin // Austrian Journal of Statistics. 2011. Vol. 40. P. 65–73.