

4. Волков И. К. Случайные процессы М., 2000. С. 299.
5. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. М., 1974, С. 228.

ПОСЛЕДОВАТЕЛЬНЫЙ СТАТИСТИЧЕСКИЙ ТЕСТ С ВОССТАНОВЛЕНИЕМ

Д. В. Кишилов

Введение

На практике возникают ситуации, когда целесообразно провести несколько последовательных проверок со сравнительно небольшой средней длиной выборки, чтобы на основании их результатов вынести решение в пользу той или иной гипотезы. В частности, такой подход может потребоваться, когда последовательный тест применяется к выборке, объем которой ограничен снизу некоторым числом N . В этом случае, если проверка заканчивается на наблюдении с номером $n < N$, можно для повышения точности начать новую процедуру проверки с $(n+1)$ -го наблюдения. Таким образом процесс может продолжаться до тех пор, пока количество наблюдений не превысит N . Рассмотрим такую проверку гипотез на примере биномиальной модели наблюдений.

1. Описание модели и построение теста

Пусть наблюдения $x_1, x_2, \dots$ независимы и распределены по закону Бернулли с параметром $\theta \in \Theta = \{\theta_0, \theta_1\}$, $\theta_0, \theta_1 \in [0, 1]$, $\theta_0 < \theta_1$, то есть

$$\begin{aligned} p_j(0) &= P(x_i = 0 \mid \theta = \theta_j) = 1 - \theta_j, \\ p_j(1) &= P(x_i = 1 \mid \theta = \theta_j) = \theta_j, j \in \{0, 1\}. \end{aligned} \quad (1)$$

Проверяется гипотеза $H_0 : \theta = \theta_0$ против гипотезы $H_1 : \theta = \theta_1$.

Пусть к наблюдениям $x_1, x_2, \dots, x_N$ применялся последовательный тест Вальда с порогами C_- , C_+ , для которого известны оценки вероятностей ошибок первого и второго рода α_0 , β_0 .

Пусть α и β – «желаемые» вероятности ошибок при принятии решения по всей выборке, P_{H_i} , $i = 0, 1$, – априорная вероятность того, что будет верна гипотеза H_i , $P_{H_0} + P_{H_1} = 1$. Обозначим

$$P(H_0, n, m) = \frac{(1 - \alpha_0)^n \beta_0^m}{(1 - \alpha_0)^n \beta_0^m P_{H_0} + \alpha_0^n (1 - \beta_0)^m P_{H_1}}, \quad (2)$$

$$P(H_1, n, m) = \frac{\alpha_0^n (1 - \beta_0)^m}{(1 - \alpha_0)^n \beta_0^m P_{H_0} + \alpha_0^n (1 - \beta_0)^m P_{H_1}}. \quad (3)$$

Построим множества $U = \{(n, m) \mid n, m \geq 0, n + m \leq N\}$,
 $U_0 = \{(n, m) \in U \mid P(H_1, n, m) \leq \beta\}$, $U_1 = \{(n, m) \in U \setminus U_0 \mid P(H_0, n, m) \leq \alpha\}$,
 $U_+ = U \setminus (U_0 \cup U_1)$.

Пусть в результате применения последовательного теста n раз было принято решение в пользу H_0 и m раз – в пользу H_1 . Примем гипотезу H_d , где d – рандомизированное решение, имеющее вид

$$d = 1_{\{(n, m) \in U_1\}} + \delta 1_{\{(n, m) \in U_+\}}, \quad (4)$$

где 1_A – индикатор множества A , δ – случайная величина, распределенная по закону Бернулли, $P\{\delta = 1\} = P_{H_1}$.

Теорема 1. Если α_0, β_0 – точные значения вероятностей ошибок первого и второго рода теста Вальда, то вероятности ошибок первого и второго рода для теста (4) $\bar{\alpha}$ и $\bar{\beta}$ будут удовлетворять неравенствам:

$$\begin{aligned} \bar{\alpha} &\leq \alpha + P_{H_1} P\{(n, m) \in U_+ \mid \theta = \theta_0\}, \\ \bar{\beta} &\leq \beta + P_{H_0} P\{(n, m) \in U_+ \mid \theta = \theta_1\}. \end{aligned}$$

2. Вычисление характеристик теста

Если множество Θ таково, что $\exists b \in \mathbb{R}, m_0, m_1 \in \mathbb{N}$,

$$\ln \frac{p_1(0)}{p_0(0)} = -m_0 b, \ln \frac{p_1(1)}{p_0(1)} = m_1 b, \quad (5)$$

то можно получить точные значения характеристик построенного теста, применив подход, предложенный в [1].

Рассмотрим последовательность Λ_n , задаваемую следующим рекуррентным соотношением:

$$\begin{aligned} \Lambda_0 &= 0, \\ \Lambda_n &= \Lambda_{n-1} + \ln(p_1(x_n)/p_0(x_n))/b, \text{ если } \Lambda_{n-1} \in (C_-, C_+), n = \overline{1, N}, \\ \Lambda_n &= \ln(p_1(x_n)/p_0(x_n))/b, \text{ если } \Lambda_{n-1} \notin (C_-, C_+), n = \overline{1, N}. \end{aligned} \quad (6)$$

Так как Λ_n зависит только от Λ_{n-1} , то Λ_n будет цепью Маркова. Обозначим

$$\xi_n = \Lambda_n 1_{(C_-, C_+)}(\Lambda_n) + C_- 1_{(-\infty, C_-]}(\Lambda_n) + C_+ 1_{[C_+, \infty)}(\Lambda_n). \quad (7)$$

Введем следующие обозначения: $\delta_{i,j}$ – символ Кронекера, $1(u)$ – функцию Хевисайда, $\mathbf{1}_k$ – вектор длины k , заполненный 1.

Теорема 2 В условиях модели (1), (5), (6) последовательность ξ_n является цепью Маркова с $m = C_+ - C_- + 1$ состояниями $C_-, \dots, C_+$. Если справедлива гипотеза $H_k, k \in \{0, 1\}$, то матрица вероятностей переходов за один шаг и вектор начальных вероятностей состояний $C_- + 1, \dots, C_+ - 1$ имеют вид:

$$P^{(k)} = (p_{ij}^{(k)}) = \begin{pmatrix} \pi_{C_-}^{(k)} & \pi_{C_+}^{(k)} & | & \pi_{C_-}^{(k)} \\ \pi_{C_-}^{(k)} & \pi_{C_+}^{(k)} & | & \pi_{C_-}^{(k)} \\ - & - & - & - \\ R^{(k)} & & | & Q^{(k)} \end{pmatrix}, \pi^{(k)} = (\pi_i^{(k)}) = \begin{pmatrix} \pi_{C_-+1}^{(k)} \\ \vdots \\ \pi_{C_+-1}^{(k)} \end{pmatrix},$$

а блоки $R^{(k)}$, $Q^{(k)}$ и вектор $\pi^{(k)}$ заданы соотношениями:

$$p_{ij}^{(k)} = \begin{cases} \delta_{m_1, j-i} \theta_k + \delta_{-m_0, j-i} (1 - \theta_k), j \in (C_-, C_+), \\ 1(m_0 + C_- - i)(1 - \theta_k), j = C_-, \\ 1(m_1 + i - C_+) \theta_k, j = C_+, i \in (C_-, C_+); \end{cases}$$

$$\pi_i^{(k)} = \delta_{m_1, i} \theta_k + \delta_{-m_0, i} (1 - \theta_k), i \in (C_-, C_+).$$

Начальные вероятности состояний C_- , C_+ задаются соотношениями

$$\pi_{C_-}^{(k)} = 1(C_- + m_0)(1 - \theta_k),$$

$$\pi_{C_+}^{(k)} = 1(m_1 - C_+) \theta_k.$$

Обозначим $V_N^{(k)}(n_0, n_1)$, $N \geq 1$, $n_0, n_1 \geq 0$, – вероятность того, что $\sum_{i=1}^N \delta_{\xi_i, C_-} = n_0$, $\sum_{i=1}^N \delta_{\xi_i, C_+} = n_1$ при верной гипотезе H_k , $k = 0, 1$.

Лемма 1 Вероятность $V_N^{(k)}(n_0, n_1)$ может быть вычислена по следующим рекуррентным формулам:

$$V_N^{(k)}(0, 0) = (\pi^{(k)})'(Q^{(k)})^{N-1} \mathbf{1}_m,$$

$$V_N^{(k)}(n_0 + 1, n_1) = \pi_{C_-}^{(k)} V_{N-1}^{(k)}(n_0, n_1) + \sum_{i=2}^{N-n_0-n_1} (\pi^{(k)})'(Q^{(k)})^{i-1} R_{(1)}^{(k)} V_{N-i}^{(k)}(n_0, n_1),$$

$$V_N^{(k)}(n_0, n_1 + 1) = \pi_{C_+}^{(k)} V_{N-1}^{(k)}(n_0, n_1) + \sum_{i=2}^{N-n_0-n_1} (\pi^{(k)})'(Q^{(k)})^{i-1} R_{(2)}^{(k)} V_{N-i}^{(k)}(n_0, n_1).$$

Результаты вычислительных экспериментов

α	β	$\tilde{\alpha}$	$\tilde{\beta}$
0,01	0,01	0,0011	0,0008
0,01	0,05	0,0015	0,0016
0,01	0,1	0,0009	0,0068
0,05	0,05	0,0022	0,0028
0,05	0,1	0,0028	0,0063
0,1	0,1	0,0068	0,0073

Теорема 3 В условиях Теоремы 2 истинные вероятности ошибок первого и второго рода для теста (4) равны соответственно

$$\alpha = \sum_{(n_0, n_1) \in U_1} V_N^{(0)}(n_0, n_1) + P_{H_1} \sum_{(n_0, n_1) \in U_+} V_N^{(0)}(n_0, n_1), \quad (8)$$

$$\beta = \sum_{(n_0, n_1) \in U_0} V_N^{(1)}(n_0, n_1) + P_{H_0} \sum_{(n_0, n_1) \in U_+} V_N^{(1)}(n_0, n_1). \quad (9)$$

3. Результаты вычислительных экспериментов

Характеристики теста (4) были исследованы с помощью имитационного моделирования. Вычислительные эксперименты проводились для следующего набора параметров:

$$\theta_0 = 0.4, \theta_1 = 0.6, N = 100, \alpha_0 = 0.1, \beta_0 = 0.1, P_{H_0} = P_{H_1} = 0.5.$$

Варьировались параметры α и β . Пороги C_- и C_+ вычислялись по формулам [2]: $C_+ = [\log((1 - \beta_0)/\alpha_0)]$, $C_- = [\log(\beta_0/(1 - \alpha_0))]$, где $[\cdot]$ означает взятие целой части. Количество тестов для каждого набора параметров составляло 10000. Результаты вычислительных экспериментов представлены в таблице. Через $\tilde{\alpha}$ и $\tilde{\beta}$ в таблице обозначены полученные оценки вероятностей ошибок первого и второго рода.

Литература

1. Харин А. Ю. Об одном подходе к анализу последовательного критерия отношения правдоподобия для различения простых гипотез // Вестн. Белорус. ун-та. Сер. 1. 2002. № 1. С. 92–96.
2. Вальд А. Последовательный анализ. М.: Физматгиз, 1960.