РОБАСТНОСТЬ ПОСЛЕДОВАТЕЛЬНЫХ КРИТЕРИЕВ ПРОВЕРКИ ПРОСТЫХ ПАРАМЕТРИЧЕСКИХ ГИПОТЕЗ

С. Ю. Чернов

Последовательный подход [1] используется для статистического решения многих практических задач, в которых имеется необходимость проверки гипотетических предположений о параметрах исследуемого процесса. При выполнении модельных предположений такой подход требует в среднем меньшее число наблюдений для принятия решений среди всех возможных статистических критериев с такими же значениями вероятностей ошибок. Однако на практике искажения в предполагаемой модели могут оказать существенное влияние на значения вероятностных характеристик последовательных критериев. В [3] построен минимаксный робастный последовательный критерий для дискретного распределения вероятностей наблюдений. В [4] закон распределения наблюдений отличается от распределения, используемого при построении теоретической модели, при этом в качестве меры различия используется L_1 метрика. В данной работе проведен анализ робастности, когда имеют место "засорения", предложенные Тьюки и Хьюбером [2]. При указанных предположениях проведен анализ робастности вероятностных характеристик ПКОВ проверки двух простых гипотез и предложен способ повышения робастности ПКОВ.

Пусть на измеримом пространстве $(\Omega, \Im)$ наблюдается последовательность независимых одинаково распределенных случайных величин $x_1, x_2, ... \in \mathbf{R}$, имеющих плотность распределения вероятностей $f(x;\theta)$ с параметром $\theta \in \Theta = \{\theta_0, \theta_1\}$, истинное значение которого неизвестно. Обозначим функцию распределения вероятностей наблюдений x_i, $i = 1,2,\ldots$, через $F(x,\theta)$; $\Lambda_n = \Lambda_n(x_1,...x_n) = \sum_{k=1}^{n} \lambda_k$, где

$$\lambda_k = \lambda(x_k) = \ln\big(f(x_k,\theta_1)/f(x_k,\theta_0)\big). \quad (1)$$

Относительно параметра θ имеются две простые гипотезы $H_0: \theta = \theta_0$, $H_1: \theta = \theta_1$. Для проверки данных гипотез используется последовательный критерий отношения вероятностей (ПКОВ) [1]:

$$N = \min\{n \in \mathrm{N} : \Lambda_n \notin (C_-, C_+)\}, \tag{2}$$

$$d = 1_{[C_+, +\infty)}(\Lambda_N), \tag{3}$$

где N – случайный момент остановки, после которого принимается решение d в соответствии с (3). В (2), (3) $C_-, C_+ \in \mathbf{R}$, $C_- < C_+$ – заданные параметры критерия, называемые порогами. На практике для их задания пользуются соотношениями [1] $C_- = \ln(\beta_0/(1-\alpha_0))$, $C_+ = \ln((1-\beta_0)/\alpha_0)$, где величины $\alpha_0, \beta_0 \in (0,1)$ примерно равны фактическим значениям вероятностей ошибок I и II рода.

П1) Пусть функция $f(x,\theta)$ имеет конечную производную 1-го порядка по переменной x, а также $f(x,\theta) \neq 0$.

Без ограничения общности будем считать, что истинной гипотезой является H_0 (случай H_1 рассматривается аналогично).

Обозначим $f(x) = f(x,\theta_0)$, $F(x) = F(x,\theta_0)$, $F_\lambda(x) = P\{\lambda_n < x\}$, $p_\lambda(x) = \dfrac{d}{dx} F_\lambda(x)$.

Для оценивания вероятностей ошибочных решений последовательного критерия (2), (3) воспользуемся подходом, изложенным в [5]. Разобьем интервал (C_-, C_+) на m промежутков длиной $h = (C_+ - C_-)/m$, $m \in \mathbf{N}$ – параметр разбиения (аппроксимации). Введем случайные последовательности ($i \in \mathbf{N}$):

$$\Lambda_n^{\pm} = \sum_{t=1}^{n} \lambda_t^{\pm};\ \lambda_1^- = C_- + \left[\frac{\lambda_1 - C_-}{h}\right] h,\ \lambda_t^- = \left[\frac{\lambda_t}{h}\right] h,\ t \geq 2;\ \lambda_i^+ = \lambda_i^- + h.$$

Построим поглощающие цепи Маркова L_n^-, L_n^+ со множеством значений $\{0, 1, \ldots, m, m+1\}$ и поглощающими состояниями 0 и $m+1$:

$$L_n^- = \begin{cases} 0, & \Lambda_n^- \in (-\infty, C_- - h], \\ i, & \Lambda_n^- = C_- + (i-1)h, \quad i = \overline{1,m}, \\ m+1, & \Lambda_n^- \in [C_+, \infty), \end{cases} \quad L_n^+ = \begin{cases} 0, & \Lambda_n^+ \in (-\infty, C_-], \\ i, & \Lambda_n^+ = C_- + ih, \quad i = \overline{1,m}, \\ m+1, & \Lambda_n^+ \in [C_+ + h, \infty). \end{cases}$$

В соответствии с [5], векторы вероятностей начальных состояний и матрицы вероятностей переходов цепей Маркова L_n^- и L_n^+ могут быть записаны в явном виде.

Пусть α^- и α^+ – вероятности поглощения цепей Маркова L_n^- и L_n^+ в состоянии $(m+1)$. В [5] показано, что вероятности α^-, α и α^+ удовлетворяют неравенству $\alpha^- \le \alpha \le \alpha^+$ и соотношению при $h \to 0$ $\alpha^+ - \alpha^- = O(h)$. Поэтому в качестве точечного приближения неизвестного значения α выбирается $\hat{\alpha}_m = \left(\alpha^+ + \alpha^-\right)/2$, причем $|\alpha - \hat{\alpha}_m| \le \left(\alpha^+ - \alpha^-\right)/2$. В дальнейшем вместо вероятности α будем анализировать величины α^- и α^+.

Пусть наблюдения x_n, $n \in \mathrm{N}$, имеют плотность распределения вероятностей

$$h(x) = (1-\varepsilon) f(x) + \varepsilon g(x), \tag{4}$$

где $g(x,\theta)$, $g(x,\theta) \ne f(x,\theta)$, – неизвестная искажающая плотность распределения вероятностей, а ε, $0 \le \varepsilon \le \varepsilon_0$, – уровень искажения, максимальное значение которого ε_0, известно заранее. Тогда $h_\lambda(x) = (1-\varepsilon) f_\lambda(x) + \varepsilon g_\lambda(x)$, где функции $h_\lambda(x)$ и $g_\lambda(x)$ являются плотностями распределения вероятностей случайных величин λ_n, когда наблюдения x_n имеют плотности $h(x)$ и $g(x)$ соответственно.

Цепи Маркова L_n^- и L_n^+ в случае, когда наблюдения имеют плотность распределения вероятностей $h(\cdot)$, обозначим соответственно $L_n^-(h)$ и $L_n^+(h)$. Пусть $\alpha^-(h,\varepsilon)$ и $\alpha^+(h,\varepsilon)$ – вероятности поглощения цепей Маркова $L_n^-(h)$ и $L_n^+(h)$ в состоянии $m+1$.

Пусть $g_+ > (m-1)h$. Обозначим

$$\bar{f}_\lambda(x) = (1-\varepsilon) f_\lambda(x) + \varepsilon\delta(x - g_+), \tag{5}$$

где $\delta(\cdot)$ – дельта-функция Дирака [6].

Теорема 1. Если для искаженной модели наблюдений (4) выполнены предположение П1, то величины $\alpha^+(h_\lambda,\varepsilon)$, удовлетворяют неравенству $\alpha^+(h_\lambda,\varepsilon) \le \alpha^+(\bar{f}_\lambda,\varepsilon)$.

Следствие 1. Вероятность ошибки первого рода $\alpha^+(\bar{f}_\lambda,\varepsilon)$ монотонно возрастает по переменной ε, в частности, для любого ε, $0 \le \varepsilon \le \varepsilon_0$, выполняется неравенство $\alpha^+(\bar{f}_\lambda,\varepsilon) \le \alpha^+(\bar{f}_\lambda,\varepsilon_0)$.

Рассмотрим плотность распределения вероятностей

$$h_\lambda^g(x) = 1_{[g_-,g_+]}(x) h_\lambda(x) + \varepsilon_- \delta(x - g_-) + \varepsilon_+ \delta(x - g_+), \tag{6}$$

где g_- и g_+ – заданные параметры усечения случайной величины λ_n, имеющей плотность распределения вероятностей $h_\lambda(\cdot)$ и функцию распределения вероятностей $H_\lambda(\cdot)$, $\varepsilon_- = H_\lambda(g_-)$, $\varepsilon_+ = 1 - H_\lambda(g_+)$.
Пусть

$$\bar{f}_\lambda^g(x) = (1-\varepsilon) f_\lambda^g(x) + \varepsilon\delta(x - g_+) .$$

Теорема 2. Если для искаженной модели наблюдений (6) выполнены предположение П1, то величины $\alpha^+(h_\lambda^g, \varepsilon)$, удовлетворяют неравенству $\alpha^+(h_\lambda^g, \varepsilon) \le \alpha^+(\bar{f}_\lambda^g, \varepsilon)$.

Следствие 2. Вероятность ошибки первого рода $\alpha^+(\bar{f}_\lambda^g, \varepsilon)$ монотонно возрастает по переменной ε, в частности, для любого ε, $0 \le \varepsilon \le \varepsilon_0$, выполняется неравенство $\alpha^+(\bar{f}_\lambda^g, \varepsilon) \le \alpha^+(\bar{f}_\lambda^g, \varepsilon_0)$.

Литература

1. *Вальд А.* Последовательный анализ. М., 1960.
2. *Хьюбер П.* Робастная статистика. М., 1984.
3. *Kharin A, Kishylau D.* Robust sequential testing of hypotheses on discrete probability distributions // Austrian Journal of Statistics, V. 34. 2005. № 2. P. 153-162.
4. *Charnou S., Kharin A.* Robustness Analysis of the Sequential Probability Ratio Test under Functional Distortions in the L_2-metric. // Abstracts of the International Conference on Robust Statistics. Parma, 2010. P. 20.
5. *Харин А., Чернов С.* Оценивание вероятностей ошибок последовательного критерия отношения вероятностей // Вестник БГУ. Минск, 2011. С. 96–100.
6. *Кудрявцев Л. Д.* Курс математического анализа. М., 1981.