О ПРИМЕНИМОСТИ РОБАСТНОГО ПОСЛЕДОВАТЕЛЬНОГО РАЗЛИЧЕНИЯ ДИСКРЕТНЫХ РАСПРЕДЕЛЕНИЙ ВЕРОЯТНОСТЕЙ ПРИ НАЛИЧИИ ИСКАЖЕНИЙ

А. Ю. Харин

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

Рассматривается задача построения робастного (устойчивого) последовательного теста для различения двух дискретных распределений вероятностей по наблюдаемой выборке. В недавних работах автора такой тест построен для вероятностных распределений из специального класса. Результаты, представленные в данной работе, позволяют аппроксимировать произвольное дискретное распределение вероятностей с конечным множеством значений распределением из указанного класса.

Ключевые слова: последовательный анализ, искажения, робастность.

ВВЕДЕНИЕ. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Последовательный подход [1] используется в решении прикладных задач, где требуется различение дискретных распределений вероятностей, которыми характеризуется наблюдаемая последовательность. Такие задачи имеют место в генетике [2], в экономике, при контроле за качеством выпускаемой продукции, и в других областях. Теоретически последовательный тест Вальда [1] при решении таких задач обладает оптимальным свойством: он минимизирует среднее число необходимых наблюдений для обеспечения заданных значений вероятностей ошибок первого и второго рода. Однако на практике модельные предположения часто искажены [3], и поэтому указанное свойство теряется. В связи с этим возникает необходимость в исследовании робастности (устойчивости) к заданным типам искажений тестов, оптимальных в рамках гипотетических предположений, а также в построении робастных тестов [4]. Для специальной модели дискретных данных эти две задачи решены в [5], когда искажения описываются моделью «выбросов» Тьюки – Хьюбера [3].

В данной работе получено обобщение результатов [5] на случай произвольных дискретных вероятностных распределений с конечным множеством значений.

Пусть наблюдается последовательность независимых одинаково распределенных случайных величин $x_1, x_{2,\dots}, x_t \in U = \{1,\dots,M\}$, $t \in N$, из одного из двух гипотетических распределений вероятностей:

$$P^{(k)}\{x_t = u\} = p_u,\ u \in U,\ p_1^{(k)} + \dots + p_M^{(k)} = 1,\ k = 0{,}1. \qquad (1)$$

Без ограничения общности предположим, что $0 \le p_1^{(k)} \le \dots \le p_M^{(k)} < 1$.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Л е м м а 1. Пусть $\delta \in [0,1)$, $a = 1-\delta$. Тогда

$$\exists\ l_1,\ldots,l_M \in N \cup \{+\infty\},\ 1 \le l_M \le \ldots \le l_1 \le +\infty, \tag{2}$$

такие что выполняются соотношения

$$\left|p_u - a^{l_u}\right| \le \delta,\ u \in U; \tag{3}$$

$$\left|\sum_{u\in U} a^{l_u} - 1\right| \le M\delta. \tag{4}$$

Доказательство. Докажем вначале (3).

$$\left|p_u - a^{l_u}\right| \le \max\left\{a^{l_u-1} - a^{l_u}, a^{l_u} - a^{l_u+1}\right\} = \max\left\{a^{l_u-1}(1-a), a^{l_u}(1-a)\right\} =$$

$$= \delta \cdot a^{l_u-1} \le \delta.$$

Теперь докажем (4).

$$\left|\sum_{u\in U} a^{l_u} - 1\right| \le \sum_{u\in U}\left|a^{l_u} - p_u\right| \le \delta \sum_{u\in U} a^{l_u-1} \le M\delta.$$

Таким образом, лемма 1 доказана.

Л е м м а 2. Пусть в условиях леммы 1 $\delta \in [0,\frac{1}{2})$. Тогда для набора (2), удовлетворяющего (3), справедливы следующие неравенства:

$$1-\delta \le \sum_{u\in U} a^{l_u} \le \frac{1-\delta}{1-2\delta}. \tag{5}$$

Доказательство. При доказательстве второго утверждения леммы 1 получено неравенство:

$$a\cdot|s-1| \le \delta\cdot s,\ s = \sum_{u\in U} a^{l_u}.$$

Рассмотрим теперь два случая.

1) $s \ge 1$. Тогда

$$a\cdot s - a \le \delta\cdot s,\ s\cdot(a-\delta) \le a,\ s \le \frac{1-\delta}{1-2\delta}.$$

2) $s < 1$. В этом случае

$$a - a\cdot s \le \delta\cdot s,\ a \le (\delta + a)\cdot s,\ s \ge a = 1-\delta.$$

Объединяя полученные результаты, получаем (5). Таким образом, лемма 2 доказана.

Замечание. Второе утверждение леммы 1 может быть усилено:

$$\left|\sum_{u\in U} a^{l_u} - 1\right| \le \frac{\delta}{1-\delta}(1+M\delta).$$

В самом деле, из доказательства леммы 1:

$$\left|\sum_{u\in U} a^{l_u} - 1\right| \le \delta \cdot \sum_{u\in U} a^{l_u - 1} = \frac{\delta}{a} \cdot \sum_{u\in U} a^{l_u} \le$$

$$\le \frac{\delta}{a} \cdot \sum_{u\in U} \left(p_u + \delta\right) = \frac{\delta}{1-\delta} \cdot \left(1 + M\delta\right),$$

так как в силу (3)

$$p_u - \delta \le a^{l_u} \le p_u + \delta, \quad \sum_{u\in U} p_u = 1.$$

Таким образом, с использованием лемм 1, 2 и замечания, для двух произвольных дискретных распределений вероятностей с конечным числом состояний с любой наперед заданной точностью можно построить минимаксный робастный последовательный тест [5] для их различения.

ЛИТЕРАТУРА

1. *Вальд А.* Последовательный анализ. М., 1960.
2. *Waterman M. S.* Mathematical methods for DNA sequences. Boca Raton: CRC Press, 1989.
3. *Хьюбер П.* Робастность в статистике. М., 1984.
4. *Kharin A.* Robustness of sequential testing of hypotheses on parameters of Markov chains // International Conference on Robust Statistics. Antwerp: University of Antwerp. 2003. P. 64–65.
5. *Kharin A.* On robustifying of the sequential probability ratio test for a discrete model under contaminations // Austrian Journal of Statistics. 2002. Vol. 31. № 4. P. 267–277.