БЫСТРАЯ СКОЛЬЗЯЩАЯ СВОБОДНАЯ ОТ РАСПРЕДЕЛЕНИЯ ПОСЛЕДОВАТЕЛЬНАЯ ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

В. И. Никитенок

Республиканский институт высшей школы
Минск, Беларусь
E-mail: root@nihe.niks.by

Для случая двух выборок рассмотрена быстрая скользящая свободная от распределения последовательная проверка статистических гипотез. Реализация свободных от распределения процедур в реальном масштабе времени (быстрых процедур) обеспечивается разбиением всех элементов выборок на подгруппы, количество элементов в которых позволяет воспользоваться гауссовым распределением используемых статистик. Для стабилизации средней длительности процедуры принятия решения применен последовательный анализ. Показатели качества последовательной процедуры принятия решения получены с использованием эквивалентного непрерывного марковского процесса.

Ключевые слова: статистические гипотезы, последовательный анализ, показатели качества.

1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И ПОСТАНОВКА ЗАДАЧИ

Имеются выборки $Z_k = \{Z_k\} = Z_{k1},\ldots, Z_{kN}$, где $k = 1, 2$. Для обеспечения реализации свободных от распределения процедур в реальном масштабе времени (быстрых процедур) разделим элементы N_k выборок на P групп по m_k элементов в каждой, т. е. $N_k = m_k P$. Необходимо проверить гипотезу H: элементы выборки $\{Z_k\}$ имеют одинаковые распределения $F_0(z)$, относительно альтернативы K: p – групп элементов ($p < P$) выборки Z_1 (или Z_2) имеют распределение $F_1(z)$, а $(P - p)$ - групп элементов этой выборки и P - групп элементов выборки Z_2 (или Z_1) – распределение $F_0(z)$. Наряду со стабилизацией вероятности ошибки первого рода потребуем и стабилизации средней длительности процедуры принятия решения. Таким образом, надо осуществить быструю свободную от распределения последовательную проверку статистических гипотез, найти алгоритм работы, определить пороги решения (нижний и верхний), мощность правила и среднее число шагов до принятия решения при альтернативе.

2. ПРОЦЕДУРА ОБРАБОТКИ. ПОКАЗАТЕЛИ КАЧЕСТВА ПРИНЯТИЯ РЕШЕНИЯ

Для каждой группы элементов предварительно вычислим значение выбранной свободной от распределения статистики S_l^L (далее «статистики»). Нижний индекс указывает на используемую i-ю статистику. Верхний - на то, что статистика определяется на $L = m_1 + m_2$ элементах. Таким образом, выполнен переход от наблюдений $\{Z_k\}$ к значениям статистики $\{S_l^L\}$. Это преобразование является необратимым. Поэтому при обработке $\{S_l^L\}$ часть информации, заключенной в $\{Z_k\}$, теряется, и искомый алгоритм обработки будет уступать в эффективности оптимальному. Однако первый алгоритм, в отличие от второго, обеспечивает стабильность вероятности ошибки первого рода при изменении статистических характеристик входных воздействий.

Примеры преобразования исходных наблюдаемых данных, «неудобных» с той или иной точки зрения для обработки с целью извлечения полезной информации, известны. Одним из примеров такого преобразования является «выбеливание» небелого шума при решении задач обнаружения и измерения на фоне мешающих отражений.

Заметим, число элементов множества $\{S_l^L\}$ в P раз меньше числа элементов множества $\{Z_k\}$. Это существенно упрощает практическую реализацию сложных и эффективных (например, ранговых) свободных от распределения процедур принятия решений и делает возможной их работу в реальном масштабе времени.

Большое число указанных выше статистик уже при сравнительно небольших объемах выборок (10...20) имеют гауссово распределение, как при гипотезе, так и при альтернативе. При гипотезе математическое ожидание E_0 и дисперсия V_0 упомянутого распределения известны и не зависят от вида и параметров неизвестного непрерывного распределения $F_0(z)$. При альтернативе математическое ожидание E_1 и дисперсия V_1 зависят от распределения результатов наблюдения. Вид зависимости определяется выбранной статистикой. Однако при близких гипотезе и альтернативе и $m{>>}1$ имеем $V_1 \approx V_0$.

Таким образом, имеется последовательность S_l^L статистик с гауссовыми распределениями $N(S;\, E_0,\, V_0)$ при гипотезе и $N(S;\, E_1,\, V_0)$ при альтернативе. Необходимо проверить простую гипотезу: математическое ожидание гауссовой случайной величины равно E_0, относительно сложной альтернативы: математическое ожидание $E_1 > E_0$ или $E_1 < E_0$ (дисперсия V_0 известна).

Для отыскания показателей качества последовательной процедуры принятия решения можно воспользоваться методикой, разработанной Вальдом [1], как это и сделано в работах Акимова [2, 3]. Однако известно, что, несмотря на приближенный характер конечных результатов, их получение довольно сложно. Поэтому обратимся к более простой методике исследования последовательной процедуры с помощью эквивалентного непрерывного марковского процесса [4, 5].

При последовательном анализе число p – групп в выражении величина случайная. На первом шаге анализируется одна статистика S_l^L, на втором – две статистики, на p-м – p статистик S_l^L.

При гипотезе и альтернативе распределения статистик гауссовы:

$$f_0(S/H)=\frac{1}{\sqrt{2\pi V_0}}\exp\left\{-0{,}5(S-E_0)^2\Big/V_0\right\}, \tag{1}$$

$$f_1(S/K)=\frac{1}{\sqrt{2\pi V_0}}\exp\left\{-0{,}5(S-E_1)^2\Big/V_0\right\}. \tag{2}$$

Поэтому логарифм отношения правдоподобия определяется следующими выражениями:

на первом шаге ($p = 1$)

$$L^{(1)}=\frac{1}{2V_0}(2(E_1-E_0)S^{(1)}-(E_1^2-E_0^2)),$$

на втором шаге

$$L^{(2)}=\frac{1}{2V_0}(2(E_1-E_0)S^{(2)}-(E_1^2-E_0^2))-\frac{1}{2V_0}(E_1^2-E_0^2)=$$
$$=L^{(1)}-\frac{1}{2V_0}(E_1^2-E_0^2),$$

на p-м шаге

$$\begin{aligned}L^{(p)}&=\frac{1}{2V_0}(2(E_1-E_0)S^{(p)}-(E_1^2-E_0^2))-\frac{1}{2V_0}(E_1^2-E_0^2)=\\&=L^{(p-1)}-\frac{1}{2V_0}(E_1^2-E_0^2),\end{aligned} \tag{3}$$

где $S^{(p)}=\sum\limits_{i=1}^{p}S_{li}^{L}$.

Таким образом, отношение правдоподобия для p-го шага имеет вид:

$$L^{(p)}=\frac{\hat{E}_1-E_0}{V_0}\sum_{i=1}^{p}S_{li}^{L}-p\frac{\hat{E}_1^2-E_0^2}{V_0}=L^{(p-1)}-\nu^{(p)}, \tag{4}$$

где $\nu^{(p)}=0{,}5p(\hat{E}_1^2-E_0^2)/V_0$, $\hat{E}_1$ – ожидаемое или расчетное значение математического ожидания статистики при альтернативе для одной группы элементов выборок.

Выражение (4) определяет алгоритм последовательной поверки статистических гипотез. Случайное значение статистики $S^{(p)}$ является гауссовым, поэтому процесс $L^{(p)}$ также оказывается гауссовым и плотность вероятности $\omega(L^{(p)},p)$ равна:

$$\varpi(L^{(p)},p)=\frac{1}{\sqrt{2\pi\sigma_p^2}}\exp\{-0{,}5\sigma_p^2(L^{(p)}-m_p)^2\Big/\sigma_p^2\}, \tag{5}$$

где (см. (4))

$$m_p=E\{L^{(p)}\}=\frac{\hat{E}_1-E_0}{V_0}E\{S^{(p)}\}-p\frac{\hat{E}_1^2-E_0^2}{2V_0}=p\left(\frac{\hat{E}_1-E_0}{V_0}E_1-\frac{\hat{E}_1^2-E_0^2}{2V_0}\right), \tag{6}$$

$$\sigma_p^2 = V\{L^{(p)}\} = p(\hat{E}_1 - E_0)^2 / V_0 . \qquad (7)$$

Для отыскания показателей качества последовательной процедуры принятия решения воспользуемся аналогией между последовательностью $L^{(p)}$, которая является марковским процессом с непрерывным множеством значений и дискретным временем, и непрерывным марковским процессом $\ell(t)$, порождаемым стохастическим дифференциальным уравнением

$$\frac{d}{dt}\ell(t) + \mu = n(t),$$

где $n(t)$ – белый шум с односторонней спектральной плотностью N_0. Согласно [4] нестационарная плотность вероятности $\omega(\ell, t)$ процесса $\ell(t)$ в неограниченных координатах равна:

$$\varpi(\ell, t) = \frac{1}{\sqrt{\pi N_0 t}} \exp\{-(\ell + \mu t)^2 \big/ (N_0 t)\}. \qquad (8)$$

Как и в [4, 5], потребуем совпадения плотностей (5) и (8) в дискретные моменты времени t=0, 1,…,. Тогда из (5) и (8) с учетом (6) и (7) получаем:

$$\mu = \frac{\hat{E}_1 - E_0}{2V_0}(2E_1 - \hat{E}_1 - E_0), \qquad (9)$$

$$N_0 = 2(\hat{E}_1 - E_0)^2 / V_0 . \qquad (10)$$

Теперь можно воспользоваться известными [6] приближенными выражениями для:

вероятности поглощения P_c на нижнем экране:

$$c = \ln[(1 - \hat{D})/(1 - F), \qquad (11)$$

где $\hat{D}$ – «ожидаемая» или расчетная мощность правила,

вероятности поглощения P_d на верхнем экране:

$$d = \ln \hat{D} / F, \qquad (12)$$

$$P_c \approx (e^{-\omega_0 d} - 1) \big/ (e^{-\omega_0 d} - e^{-\omega_0 c}), \qquad (13)$$

$$P_d = D \approx (1 - e^{-\omega_0 c}) \big/ (e^{-\omega_0 d} - e^{-\omega_0 c}), \qquad (14)$$

где

$$\omega_0 = -4\mu / N_0 = (2E_1 - \hat{E}_1 - E_0)/(\hat{E}_1 - E_0), \qquad (15)$$

среднего числа шагов до поглощения:

$$E\{p\} \approx -(cP_c + dP_d)/\mu = [d(e^{-\omega_0 c} - 1) + c(1 - e^{-\omega_0 d})] \big/ [\mu(e^{-\omega_0 d} - e^{-\omega_0 c})]. \qquad (16)$$

Таким образом, алгоритм быстрой скользящей свободной от распределения последовательной проверки статистических гипотез определяется выражением (4), нижний порог решения – выражением (11), верхний порог решения – (12), мощность правила – (14) и среднее число шагов до принятия решения при альтернативе – (16). В расчетах необходимо учитывать и выражения (9), (10) и (15).

Заметим, что алгоритм последовательной проверки гипотез (4) зависит от расчетного значения математического ожидания используемой статистики $\hat{E}_1$ при альтернативе. Это означает, что алгоритм обработки в конечном итоге оказывается «настроенным» на определенную альтернативу.

ЛИТЕРАТУРА

1. *Вальд А.* Последовательный анализ / Пер. с англ.;Под ред. Б. А. Севастьянова. М.: Физматгиз, 1960.

2. *Акимов Я. С., Ефремов В. С., Кубасов А. Н.* Последовательное обнаружение сигнала с использованием статистики, основанной на сумме рангов // Радиотехника и электроника. 1975. Т. 20. № 6. С. 1289.

3. *Акимов П. С., Ефремов В. С.* Оптимальное ранговое последовательное обнаружение импульсного сигнала // Радиотехника и электроника. 1975. Т. 20. № 11. С. 2286.

4. *Спектор А. А.* Исследование последовательного анализа при помощи эквивалентного непрерывного марковского процесса // Радиотехника и электроника. 1978. Т. 23. № 4. С. 81.

5. *Клочан Ю. Н., Спектор А. А.* Использование знаковых статистик при последовательном обнаружении сигнала в импульсной РЛС // Известия вузов: Радиоэлектроника. 1979. Т. 22. № 8. С. 10–14.

6. *Тихонов В. И., Миронов М. А.* Марковские процессы. М.: Сов. радио, 1977.