ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 162, 2002

*Dorota Pekasiewicz**

ILORAZOWE TESTY SEKWENCYJNE DLA FRAKCJI DLA PRÓB NIEPROSTYCH

STRESZCZENIE. Ilorazowe testy sekwencyjne mogą służyć do weryfikacji hipotez statystycznych nie tylko wtedy, gdy elementy do próby losowane są sekwencyjnie w sposób niezależny, ale również w przypadku, gdy próba pobierana jest według innych schematów losowania.
W pracy rozważany jest ilorazowy test sekwencyjny dla frakcji. Rozpatrywane są następujące schematy losowania próby: indywidualne losowanie zależne i warstwowe oraz losowanie zespołowe dwustopniowe. Dla wymienionych schematów losowania próby wyznaczone zostały postaci statystyk testu sekwencyjnego dla frakcji.
Słowa kluczowe: ilorazowy test sekwencyjny, statystyka testu, schemat losowania

I. WPROWADZENIE

Ilorazowe testy sekwencyjne mogą służyć zarówno do weryfikacji hipotez o parametrach populacji jednowymiarowych i wielowymiarowych, jak i do weryfikacji hipotez o parametrach rozkładu dwóch populacji.

Zastosowanie ilorazowego testu sekwencyjnego związane jest z pobieraniem stopniowo po jednym lub kilka elementów na każdym etapie sekwencyjnej weryfikacji hipotez. Każdy element lub grupa elementów wylosowana do próby, przy ustalonych z góry prawdopodobieństwach błędów I-go i II-go rodzaju, prowadzi do podjęcia jednej z trzech decyzji:

- przyjęcia hipotezy zerowej,
- przyjęcia hipotezy alternatywnej,
- kontynuacji badania poprzez wylosowanie kolejnego elementu lub grupy elementów.

* Dr, Katedra Metod Statystycznych, Uniwersytet Łódzki.

Postać statystyki ilorazowego testu sekwencyjnego związana m.in. jest ze stosowanym schematem losowania próby.

W pracy rozważany jest ilorazowy test sekwencyjny dla wskaźnika struktury. Rozpatrywane są następujące sposoby losowania elementów na każdym etapie procedury sekwencyjnej: indywidualne losowanie zależne i warstwowe oraz zespołowe losowanie dwustopniowe. Dla tych schematów losowania próby przedstawione są, wyznaczone w sposób indukcyjny, postaci statystyk testu sekwencyjnego dla frakcji.

II. CHARAKTERYSTYKA ILORAZOWEGO TESTU SEKWENCYJNEGO DLA FRAKCJI

Niech X będzie zmienną losową o rozkładzie zero-jedynkowym z nieznanym prawdopodobieństwem sukcesu θ, tzn. funkcja prawdopodobieństwa zmiennej X ma postać:

$$f(x,\theta) = x\theta + (1-x)(1-\theta), \tag{1}$$

gdzie $0 < \theta < 1$.

Weryfikujemy następujące hipotezy proste o wartości parametru θ:

$$H_0 : \theta = \theta_0 \tag{2}$$

$$H_1 : \theta = \theta_1, \text{ gdzie } \theta_1 > \theta_0 \tag{3}$$

Niech α i β będą ustalonymi prawdopodobieństwami błędów I-go i II-go rodzaju, tzn. $P(\text{odrzucenia} H_0 | \theta_0) \le \alpha$ oraz $P(\text{akceptacji} H_0 | \theta_1) \le \beta$. Dla przyjętych wartości α i β wyznaczamy stałe A i B, korzystając z następujących wzorów (por. S i l v e y 1978):

$$A = \ln \frac{1-\beta}{\alpha} \tag{4}$$

$$B = \ln \frac{\beta}{1-\alpha}. \tag{5}$$

Weryfikacja hipotezy zerowej (2) wobec hipotezy alternatywnej (3) polega na obliczaniu na podstawie n-elementowej próby losowej $X_1,...,X_n$ wartości statystyki:

$$I_n = \ln\frac{f(X_1,X_2,...,X_n;\theta_1)}{f(X_1,X_2,...,X_n;\theta_0)} = \ln\frac{f(X_1;\theta_1)}{f(X_1;\theta_0)} + \sum_{i=2}^{n}\ln\frac{f(X_i|X_1,...,X_{i-1};\theta_1)}{f(X_i|X_1,...,X_{i-1};\theta_0)} \quad (6)$$

i porównaniu jej z wyznaczonymi stałymi A i B.

Jeżeli $I_n < B$, to akceptujemy hipotezę H_0. Jeżeli $I_n > A$, to hipotezę H_0 odrzucamy i przyjmujemy H_1. Jeżeli $B \le I_n \le A$, to kontynuujemy losowanie, pobierając w określony sposób następny element (lub kilka elementów) do próby. Powtarzamy obliczenia aż do momentu podjęcia decyzji o akceptacji jednej z rozpatrywanych hipotez.

Postać statystyki I_n zależy od schematu losowania elementów do próby.

W literaturze przedstawiane są zwykle testy sekwencyjne dla niezależnego losowania próby (por. np. Marek, Noworol 1987; Silvey 1978). W pracy Pekasiewicz, Pruskiej (1996) przedstawiony jest (wraz z analizą efektywności) ilorazowy test sekwencyjny dla frakcji w przypadku, gdy na każdym etapie procedury sekwencyjnej losujemy w sposób niezależny po jednym elemencie.

W pracy tej rozważane są inne schematy losowania próby, tj.:

- indywidualne losowanie zależne;
- indywidualne losowanie warstwowe z niezależnym oraz zależnym losowaniem elementów z poszczególnych warstw;
- zespołowe losowanie dwustopniowe, w którym losowanie pierwszego stopnia (losowanie zespołów) jest losowaniem niezależnym, a losowanie stopnia drugiego (losowanie elementów) losowaniem warstwowym, przy czym warstwy stanowią wylosowane do próby zespoły.

III. ILORAZOWY TEST SEKWENCYJNY DLA FRAKCJI DLA INDYWIDUALNEGO LOSOWANIA ZALEŻNEGO PRÓBY

Załóżmy, że zmienna losowa X ma, w N-elementowej populacji, rozkład zero-jedynkowy określony wzorem (1). W celu zweryfikowania hipotez o wartości parametru θ rozkładu zero-jedynkowego, czyli hipotezy (2) wobec hipotezy (3) stosujemy ilorazowy test sekwencyjny, losując na każdym etapie w spo-

sób zależny jeden element. Na n-tym etapie otrzymamy n-elementową próbę losową: $X_1, ..., X_n$ ($n = 1,2,3,...$).

Dla tego schematu losowania próby wzory na łączną funkcję prawdopodobieństwa i statystykę I_n wyprowadzone zostały indukcyjnie w pracy P e k a - s i e w i c z (1999).

Łączna funkcja prawdopodobieństwa $f(X_1, X_2, ...X_n; \theta_i)$ dla $i = 0,1$ ma postać:

$$f(X_1, ..., X_n; \theta_i) = \begin{cases} \dfrac{\prod_{j=1}^{m_n}\left(\theta_i - \dfrac{j-1}{N}\right)\prod_{i=m_n+1}^{n}\left(1-\theta_i - \dfrac{i-(m_n+1)}{N}\right)}{\left(1-\dfrac{1}{N}\right)\left(1-\dfrac{2}{N}\right)...\left(1-\dfrac{n-1}{N}\right)} & \text{gdy } m_n < n \\ \prod_{j=1}^{n}\dfrac{\left(\theta_i - \dfrac{j-1}{N}\right)}{\left(1-\dfrac{j-1}{N}\right)} & \text{gdy } m_n = n \end{cases}, \quad (7)$$

gdzie m_n oznacza liczbę „jedynek" w n-elementowym ciągu wartości zmiennych losowych $X_1, ..., X_n$.

Statystyka I_n wyraża się wzorem:

$$I_n = \begin{cases} \sum_{j=1}^{m_n} \ln \dfrac{\theta_1 - \dfrac{j-1}{N}}{\theta_0 - \dfrac{j-1}{N}} + \sum_{i=m_n+1}^{n} \ln \dfrac{1-\theta_1 - \dfrac{i-(m_n+1)}{N}}{1-\theta_0 - \dfrac{i-(m_n+1)}{N}} & \text{gdy } m_n < n \\ \sum_{j=1}^{n} \ln \dfrac{\theta_1 - \dfrac{j-1}{N}}{\theta_0 - \dfrac{j-1}{N}} & \text{gdy } m_n = n. \end{cases} \quad (8)$$

Obliczoną wartość statystyki I_n porównujemy ze stałymi A i B wyznaczonymi w oparciu o ustalone prawdopodobieństwa błędów I-go i II-go rodzaju i podejmujemy odpowiednie decyzje.

IV. ILORAZOWY TEST SEKWENCYJNY DLA FRAKCJI DLA INDYWIDUALNEGO LOSOWANIA WARSTWOWEGO PRÓBY

Załóżmy, że zmienna losowa X ma rozkład zero-jedynkowy w N-elementowej populacji, przy czym w populacji tej wyróżnionych jest G warstw.

W przypadku podziału populacji na warstwy bardziej efektywne może okazać się wnioskowanie statystyczne przeprowadzone w oparciu o próby uzyskane w wyniku losowania warstwowego. W metodach sekwencyjnych, w szczególności w ilorazowych testach sekwencyjnych można również stosować losowanie warstwowe.

Niech k będzie ustaloną liczbą elementów pobieranych z populacji na każdym etapie postępowania sekwencyjnego, natomiast k_g liczbą elementów losowanych z g-tej warstwy, $g = 1, ..., G$. Liczebności k_g możemy wyznaczyć wykorzystując np. schemat Neymana lub losowanie proporcjonalne (por. Z a s ę p a 1992).

Jeśli nie bierzemy pod uwagę kosztów badania, a znamy wariancje zmiennej X w poszczególnych warstwach, to liczbę elementów k_g pobieranych z g-tej warstwy możemy wyznaczyć z wzoru (schemat Neymana):

$$k_g = \frac{N_g \sigma_g}{\sum_{j=1}^{G} N_j \sigma_j} k. \tag{9}$$

gdzie N_g oznacza liczbę elementów g-tej warstwy ($g = 1, ..., G$), natomiast σ_g – odchylenie standardowe zmiennej X w g-tej warstwie.

Gdy nie znamy wariancji zmiennej losowej X w poszczególnych warstwach, to możemy stosować losowanie proporcjonalne. Wtedy wartości k_g obliczamy, korzystając z wzoru:

$$k_g = \frac{N_g}{N} k. \tag{10}$$

Wyznaczone, w oparciu o jeden z wzorów (9) lub (10), liczebności k_g mogą nie być liczbami naturalnymi. W takich przypadkach za k_g przyjmujemy liczby naturalne najbliższe wyznaczonym pamiętając, aby spełniona była równość:

$$k = \sum_{g=1}^{G} k_g \,. \tag{11}$$

Losowanie, w określony sposób, k elementów na każdym etapie prowadzi do otrzymania na n-tym etapie $k \cdot n$-elementowego ciągu wartości zmiennych losowych: $X_{11}^1, ..., X_{1k_1}^1, ..., X_{G1}^1, ..., X_{Gk_G}^1, ..., X_{11}^n, ..., X_{1k_1}^n, ..., X_{G1}^n, ..., X_{Gk_G}^n$, zawierającego po $n_g = k_g \cdot n$ elementów z g- tej warstwy ($g = 1, ..., G$).

Jeśli elementy z poszczególnych warstw losowane były w sposób niezależny, to w n-tym kroku otrzymamy $k \cdot n$-elementowy ciąg wartości niezależnych zmiennych losowych: $X_{11}^1, ..., X_{1k_1}^1, ..., X_{G1}^1, ..., X_{Gk_G}^1, ..., X_{11}^n, ..., X_{1k_1}^n, ...,$ $X_{G1}^n, ..., X_{Gk_G}^n$ i w celu weryfikacji hipotezy (2) wobec hipotezy (3) obliczamy wartość statystyki I_n postaci:

$$I_n = \sum_{g=1}^{G} \left(m_{n_g} \ln \frac{\theta_1}{\theta_0} + \left(n_g - m_{n_g}\right) \ln \frac{1-\theta_1}{1-\theta_0} \right), \tag{12}$$

gdzie m_{n_g} oznacza liczbę „jedynek" w ciągu n_g-elementowym.

Przy losowaniu elementów z poszczególnych G warstw możemy stosować schemat losowania zależnego. Wtedy zmienne losowe $X_{11}^1, ..., X_{1k_1}^1, ...,$ $X_{G1}^1, ..., X_{Gk_G}^1, ..., X_{11}^n, ..., X_{1k_1}^n, ..., X_{G1}^n, ..., X_{Gk_G}^n$ są zmiennymi zależnymi i statystyka testu weryfikującego hipotezy o parametrze θ rozkładu zero-jedynkowego zmiennej X wyraża się wzorem:

$$I_n = \sum_{g=1}^{G} \left(\sum_{j=1}^{m_{n_g}} \ln \frac{\theta_1 - \frac{j-1}{N_g}}{\theta_0 - \frac{j-1}{N_g}} + \sum_{i=m_{n_g}+1}^{n_g} \ln \frac{1-\theta_1 - \frac{i-\left(m_{n_g}+1\right)}{N_g}}{1-\theta_0 - \frac{i-\left(m_{n_g}+1\right)}{N_g}} \right), \tag{13}$$

gdy $m_{n_g} < n_g$ dla $g = 1, ..., G$

lub wzorem:

$$I_n = \sum_{g \in L_G^n} \sum_{s=1}^{m_{n_g}} \ln \frac{\theta_1 - \frac{s-1}{N_g}}{\theta_0 - \frac{s-1}{N_g}} + \sum_{g \notin L_G^n} \left(\sum_{j=1}^{m_{n_g}} \ln \frac{\theta_1 - \frac{j-1}{N_g}}{\theta_0 - \frac{j-1}{N_g}} + \sum_{i=m_{n_g}+1}^{n_g} \ln \frac{\theta_1 - \frac{i-(m_{n_g}+1)}{N_g}}{\theta_0 - \frac{i-(m_{n_g}+1)}{N_g}} \right), \quad (14)$$

gdy dla pewnych warstw $m_{n_g} = n_g$.

Symbol L_G^n oznacza zbiór indeksów wszystkich tych warstw, dla których $m_{n_g} = n_g$.

V. ILORAZOWY TEST SEKWENCYJNY DLA FRAKCJI DLA ZESPOŁOWEGO DWUSTOPNIOWEGO LOSOWANIA PRÓBY

Rozważana populacja generalna może być bardzo liczna. Podział jej na zespoły stwarza możliwość zastosowania zespołowego schematu losowania próby. Zespołowe losowanie, w szczególności losowanie dwustopniowe można stosować przy weryfikacji hipotez testami sekwencyjnymi. Sekwencyjne losowanie może dotyczyć tylko zespołów lub tylko elementów losowanych z ustalonej liczby wylosowanych zespołów, ewentualnie sekwencyjnie możemy losować zarówno zespoły jak i elementy z tych zespołów.

Załóżmy, że zmienna losowa X o rozkładzie zero-jedynkowym z nieznaną wartością parametru θ rozważana jest w populacji, w której wyróżniono K równolicznych zespołów, zawierających po $\overline{M}$ elementów. Weryfikację hipotez o wartości parametru θ możemy przeprowadzić za pomocą ilorazowego testu sekwencyjnego dla frakcji, stosując zespołowe losowanie dwustopniowe. Rozważmy przypadek sekwencyjnego pobierania elementów z wylosowanych zespołów, których liczba jest ustalona na początku badania.

Niech k będzie ustaloną liczbą zespołów wylosowanych w losowaniu pierwszego stopnia według niezależnego schematu losowania. Na każdym etapie procedury sekwencyjnej z każdego wylosowanego zespołu losujemy po $\overline{m}$ elementów ($\overline{m}$ =1, 2, ...).W n-tym kroku otrzymujemy $n\overline{m}$ elementów z każdego zespołu.

Jeśli w losowaniu drugiego stopnia stosowanym schematem losowania elementów jest losowanie warstwowe z niezależnym losowaniem elementów

z poszczególnych warstw, które stanowią wylosowane zespoły, to statystyka I_n ma postać:

$$I_n = \sum_{j=1}^{k}\left(m_{n(j)} \ln\frac{\theta_1}{\theta_0} + \left(n\overline{m} - m_{n(j)}\right)\ln\frac{1-\theta_1}{1-\theta_0}\right), \tag{15}$$

gdzie $m_{n(j)}$ oznacza liczbę „jedynek" spośród $n\overline{m}$ elementów wylosowanych z j-tego zespołu.

Jeśli w losowaniu drugiego stopnia z każdego z k wylosowanych zespołów losujemy elementy w sposób zależny, to statystyka I_n wyraża się wzorem:

$$I_n = \begin{cases} \sum_{j=1}^{k}\left(\sum_{i=1}^{m_{n(j)}} \ln\frac{\theta_1 - \frac{i-1}{\overline{M}}}{\theta_0 - \frac{i-1}{\overline{M}}} + \sum_{s=1}^{n\overline{m}-m_{n(j)}} \ln\frac{1-\theta_1 - \frac{s-1}{\overline{M}}}{1-\theta_0 - \frac{s-1}{\overline{M}}}\right) & \text{gdy } L_n = \varnothing \\ \sum_{j\in L_n}\sum_{r=1}^{n\overline{m}} \ln\frac{\theta_1 - \frac{r-1}{\overline{M}}}{\theta_0 - \frac{r-1}{\overline{M}}} + \sum_{j\notin L_n}\left(\sum_{i=1}^{m_{n(j)}} \ln\frac{\theta_1 - \frac{i-1}{\overline{M}}}{\theta_0 - \frac{i-1}{\overline{M}}} + \sum_{s=1}^{n\overline{m}-m_{n(j)}} \ln\frac{\theta_1 - \frac{s-1}{\overline{M}}}{\theta_0 - \frac{s-1}{\overline{M}}}\right) & \text{gdy } L_n \neq \varnothing \end{cases} \tag{16}$$

gdzie L_n oznacza zbiór indeksów tych wszystkich zespołów spośród k wylosowanych zespołów, dla których na n-tym etapie zachodzi równość $m_{n(j)} = n\overline{m}$.

Obliczoną wartość statystyki I_n (wzór (15) lub (16)) porównujemy ze stałymi A i B, podejmując decyzję o akceptacji jednej z weryfikowanych hipotez lub decyzję o kontynuowaniu losowania próby poprzez pobranie kolejnych $\overline{m}$ elementów z każdego z k wylosowanych zespołów.

Ilorazowy test sekwencyjny dla frakcji z zespołowym dwustopniowym losowaniem próby można stosować w przypadku podziału populacji na różnoliczne zespoły. Konieczne jest wtedy ustalenie liczby elementów losowanych z poszczególnych zespołów w każdym kroku sekwencyjnym, np. proporcjonalnie do wielkości zespołów.

Niech $M_{(j)}$ oraz $l_{(j)}$ oznaczają odpowiednio liczbę wszystkich elementów j-tego zespołu oraz ustaloną liczbę elementów losowanych na każdym etapie

procedury sekwencyjnej z j-tego zespołu wylosowanego w losowaniu pierwszego stopnia ($j = 1, ..., k$).

W przypadku niezależnego losowania elementów z poszczególnych zespołów statystyka I_n wyraża się wzorem:

$$I_n = \sum_{j=1}^{k}\left(m_{n(j)} \ln\frac{\theta_1}{\theta_0} + \left(nl_{(j)} - m_{n(j)}\right)\ln\frac{1-\theta_1}{1-\theta_0}\right), \tag{17}$$

gdzie $m_{n(j)}$ oznacza liczbę „jedynek" spośród $nl_{(j)}$ elementów wylosowanych z j-tego zespołu.

Jeśli z każdego z k wylosowanych zespołów elementy losowane są w sposób zależny, to statystyka I_n ma postać:

$$I_n = \begin{cases} \sum_{j=1}^{k}\left(\sum_{i=1}^{m_{n(j)}} \ln\dfrac{\theta_1 - \dfrac{i-1}{M_{(j)}}}{\theta_0 - \dfrac{i-1}{M_{(j)}}} + \sum_{s=1}^{nl_{(j)}-m_{n(j)}} \ln\dfrac{1-\theta_1 - \dfrac{s-1}{M_{(j)}}}{1-\theta_0 - \dfrac{s-1}{M_{(j)}}}\right) & \text{gdy } L_n = \varnothing \\ \sum_{j\in L_n}\sum_{r=1}^{nl_{(j)}} \ln\dfrac{\theta_1 - \dfrac{r-1}{M_{(j)}}}{\theta_0 - \dfrac{r-1}{M_{(j)}}} + \sum_{j\notin L_n}\left(\sum_{i=1}^{m_{n(j)}} \ln\dfrac{\theta_1 - \dfrac{i-1}{M_{(j)}}}{\theta_0 - \dfrac{i-1}{M_{(j)}}} + \sum_{s=1}^{nl_{(j)}-m_{n(j)}} \ln\dfrac{\theta_1 - \dfrac{s-1}{M_{(j)}}}{\theta_0 - \dfrac{s-1}{M_{(j)}}}\right) & \text{gdy } L_n \neq \varnothing \end{cases} \tag{18}$$

gdzie L_n oznacza zbiór indeksów tych wszystkich zespołów spośród k wylosowanych zespołów, dla których na n-tym etapie zachodzi równość $m_{n(j)} = nl_{(j)}$.

Ilorazowy test sekwencyjny dla frakcji można stosować, gdy w sposób sekwencyjny losujemy zespoły, z których w losowaniu drugiego stopnia pobieramy ustaloną liczbę elementów lub też w przypadku, gdy sekwencyjnie losujemy zarówno zespoły jak i elementy z wylosowanych zespołów.

VI. PODSUMOWANIE

Przedstawiane w literaturze ilorazowe testy sekwencyjne z niezależnym schematem losowania próby charakteryzuje wysoka efektywność w porównaniu z niesekwencyjnymi testami klasycznymi. Zastosowanie tych testów pozwala

zmniejszyć liczebność próby, na podstawie której podejmujemy decyzję o akceptacji jednej z weryfikowanych hipotez nawet o połowę. Wydaje się więc uzasadnione rozszerzenie problemu weryfikacji hipotez statystycznych przy użyciu bardzo efektywnych ilorazowych testów sekwencyjnych na przypadki, gdy próba losowana jest według innych schematów losowania.

W pracy przedstawione zostały zmodyfikowane postaci statystyki ilorazowego testu sekwencyjnego dla wskaźnika struktury, pozwalające weryfikować hipotezy o parametrze rozkładu zero-jedynkowego na podstawie prób nieprostych, tj. uzyskanych w wyniku losowania zależnego, warstwowego oraz dwustopniowego.

Ilorazowe testy sekwencyjne dla prób nieprostych mogą znaleźć szerokie zastosowanie w badaniach społeczno-ekonomicznych kraju i wyodrębnionych w kraju regionów administracyjnych.

BIBLIOGRAFIA

Marek T., Noworol Cz. (1987), *Analiza sekwencyjna w badaniach empirycznych*, PWN, Warszawa.

Pekasiewicz D., Pruska K. (1996), *The Application of the Sequential Probability Ratio Test to Detection of Differences Between the Bernoulli Population and Its Small Area*, Proceedings of the Conference MSA'96, Łódź, 137–146.

Pekasiewicz D. (1999), *Ilorazowe testy sekwencyjne i ich zastosowanie w statystyce małych obszarów*, Opracowanie w ramach projektu KBN nr1 H02B 00912 pt. *Zastosowanie metod klasycznych i sekwencyjnych w statystyce małych obszarów*, maszynopis, Łódź.

Silvey S. D. (1978), *Wnioskowanie statystyczne*, PWN, Warszawa.

Zasępa R. (1972), *Metoda reprezentacyjna*, PWE, Warszawa.

Dorota Pekasiewicz

SEQUENTIAL PROBABILITY RATIO TESTS OF FRACTION FOR NONSIMPLE SAMPLES

Sequential probability ratio tests may be used to verify statistical hypotheses not only when sample is drawn sequentially independently but also when the sampling is performed according to other patterns. In the paper a sequential probability ratio test for fraction is considered. The following sampling patterns are considered: individual dependent and layer sampling and aggregate two stage sampling. For these sampling patterns the samples were determined in the form of the statistics of the sequential test for fraction.