MATEMATYKA STOSOWANA 8, 2007

Anita Małoń (Wrocław)
Dagmara Ziółkowska (Wrocław)

Testy zgodności typu chi-kwadrat dla hipotezy złożonej

Streszczenie. Przeniesienie klasycznego testu zgodności chi-kwadrat na przypadek hipotezy złożonej rodzi szereg problemów związanych z estymacją nieznanych parametrów. Jeden ze sposobów ich wyeliminowania zaproponowali Dzhaparidze i Nikulin. Ważną zaletą ich pomysłu jest możliwość użycia dość dowolnych estymatorów. Celem tego artykułu jest popularyzacja wspomnianego rozwiązania i przedstawienie pełnego, a równocześnie elementarnego dowodu o rozkładzie asymptotycznym statystyki testowej. Dodatkowo w pracy zostanie pokazane, że prezentowany test jest elementem ogólnej klasy testów wynikowych, co przemawia za jego dobrymi własnościami. Ponadto zostanie przedstawiony przykład implementacji testu dla testowania zgodności w rodzinie z parametrami przesunięcia i skali.
Słowa kluczowe: test chi-kwadrat, hipoteza złożona, $\sqrt{n}$-zgodny estymator, statystyka Dzhaparidze-Nikulina, test wynikowy.

1. Wstęp. Jednym z zagadnień wnioskowania statystycznego, które często wykorzystuje się w praktyce, jest testowanie zgodności rozkładu obserwowanego z pewną parametryczną rodziną rozkładów. Stosuje się go w takich naukach jak: medycyna, ekonomia, bankowość, finanse oraz w wielu innych. Jest ono istotne, gdy potrzebujemy sprawdzić założenia dotyczące rozkładu pewnych danych. Dla przykładu, aby móc, przy pomocy uproszczonego wzoru na VaR (popularna miara ryzyka), obliczyć ryzykowność inwestycji w akcje jakiejś firmy musimy wiedzieć, że stopy zwrotu z tych akcji mają rozkład normalny. Aby to stwierdzić używa się testów zgodności.

Najstarszym, a równocześnie najbardziej popularnym testem zgodności jest test chi-kwadrat Pearsona. Jego konstrukcja opiera się na podziale przestrzeni próby na rozłączne klasy oraz porównaniu empirycznych i teoretycznych liczebności w tychże klasach. Określenie liczebności teoretycznych wiąże się z koniecznością estymacji nieznanych parametrów proponowanej rodziny. Jedną z metod estymacji, zachowującej klasyczną postać statystyki Pearsona, zaproponował Fisher w 1924 roku. Jednakże prowadzi ona do istotnych trudności w jawnym wyznaczaniu estymatorów i jest uciążliwa w praktyce. Natomiast próby pewnych uproszczeń w metodzie Fishera mogą prowadzić do błędnych wniosków. Tymczasem Dzhaparidze i Nikulin (1974) rozwinęli pomysł Fishera i zaproponowali rozwiązanie atrakcyjne dla praktyka.

W tym artykule chcemy przybliżyć i spopularyzować podejście Dzhaparidze–Nikulina. W rozdziałach 2, 3 przedstawimy pełny i elementarny dowód twierdzenia o rozkładzie asymptotycznym statystyki testowej (twierdzenie 1) oraz wykażemy, że twierdzenie Fishera jest jego szczególnym przypadkiem. Te rozdziały są kluczowe i zawierają istotę rozwiązania Dzhaparidze–Nikulina. Następne stanowią uzupełnienie głównego nurtu naszych rozważań i mogą być czytane w dowolnej kolejności bądź w ogóle opuszczone. W rozdziale 4 pokażemy, że test Dzhaparidze–Nikulina należy do ogólnej klasy testów wynikowych (ang. *score tests*), co tłumaczy m.in. dobre własności tych testów. Natomiast w rozdziale 5 wyprowadzimy wygodną do obliczeń postać statystyki testowej dla rodziny z parametrem przesunięcia i skali.

2. Model, założenia i główne twierdzenie. Niech $X_1, X_2, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie P przyjmującymi wartości w pewnej przestrzeni mierzalnej $(\mathcal{X}, \mathcal{F})$ oraz niech

$$\mathcal{P}^0 = \{P_\beta : \beta = [\beta_1, \beta_2, \ldots, \beta_q]^T \in \Gamma\}$$

będzie rodziną rozkładów prawdopodobieństwa na $(\mathcal{X}, \mathcal{F})$, gdzie Γ jest otwartym podzbiorem przestrzeni R^q, a $\bullet^T$ oznacza transpozycję.

Rozważmy weryfikację prawdziwości hipotezy $\mathcal{H}_0^0$ orzekającej, że rozkład P należy do rodziny $\mathcal{P}^0$ przy nieznanej wartości parametru β. Naśladując podejście Pearsona, prowadzące do konstrukcji statystyki testowej poprzez kategoryzację danych, rozważmy pewien ustalony podział przestrzeni $\mathcal{X}$ na rozłączne, mierzalne podzbiory $A_1, A_2, \ldots, A_m$, gdzie $m > q + 1$. Oznaczmy $p_j(\beta) = P_\beta(X_1 \in A_j)$, $j = 1, 2, \ldots, m$, oraz $\mathbf{p}(\beta) = [p_1(\beta), p_2(\beta), \ldots, p_m(\beta)]^T$ odpowiadający im wektor prawdopodobieństw. Oczywiście, jeśli $\beta \in \Gamma$, to zachodzi

$$\sum_{j=1}^{m} p_j(\beta) = 1. \tag{1}$$

Niech

$$\mathcal{P} = \left\{\mathbf{p}(\beta) = [p_1(\beta), p_2(\beta), \ldots, p_m(\beta)]^T :\ \beta \in \Gamma\right\}$$

oznacza rodzinę wektorów prawdopodobieństw wyznaczonych przez rodzinę $\mathcal{P}^0$ i wybór podzbiorów $A_1, A_2, \ldots, A_m$. Analogicznie przyjmijmy, że $p_j = P(X_1 \in A_j)$, $j = 1, 2, \ldots, m$, oraz $\mathbf{p} = [p_1, p_2, \ldots, p_m]^T$ oznacza wektor prawdopodobieństw nieznanego rozkładu P, z którego pochodzą obserwacje.

W omawianym podejściu testowanie $\mathcal{H}_0^0$ zastępuje się weryfikacją hipotezy

$$\mathcal{H}_0: \ \mathbf{p} \in \mathcal{P} \qquad \text{przeciwko} \qquad \mathcal{H}_1: \ \mathbf{p} \notin \mathcal{P}.$$

Oznaczmy przez $\mathbf{N} = [N_1, N_2, \ldots, N_m]^T$ wektor liczebności empirycznych, gdzie N_j jest liczbą obserwacji należących do zbioru A_j dla $j = 1, 2, \ldots, m$, tzn. $N_j = \operatorname{card}\{i : X_i \in A_j, \ i = 1, 2, \ldots, n\}$. Zauważmy, że $\sum_{j=1}^{m} N_j = n$.

Klasyczna statystyka Pearsona ma postać

$$\bar{S} = \sum_{j=1}^{m} \frac{\left(N_j - np_j(\bar{\beta})\right)^2}{np_j(\bar{\beta})}, \tag{2}$$

gdzie $\bar{\beta}$ jest pewnym estymatorem nieznanego parametru β. Fisher (1924) udowodnił, że wybór estymatora $\bar{\beta}$ ma wpływ na rozkład asymptotyczny statystyki $\bar{S}$. Wykazał też, że jeśli $\widetilde{\beta}$ jest estymatorem największej wiarogodności po zgrupowaniu danych, to odpowiadająca mu statystyka $\widetilde{S}$ dana wzorem (2) ma rozkład asymptotyczny chi-kwadrat z $m - q - 1$ stopniami swobody. Dzhaparidze i Nikulin (1974), kosztem bardziej skomplikowanej postaci statystyki, zaproponowali ogólniejsze podejście opierające się na dość dowolnym estymatorze parametru β. Poniżej przedstawimy szczegółowo rozwiązanie Dzhaparidze–Nikulina wykorzystując algebraiczne metody po części oparte na pomyśle Raynera i Besta (1989).

Do dalszych rozważań załóżmy, że rodzina $\mathcal{P}$ spełnia pewne warunki regularności. Wyrazimy je w języku wektora $\mathbf{p}(\beta)$. Załóżmy więc, że wektor $\mathbf{p}(\beta)$ spełnia dla każdego $\beta \in \Gamma$ następujące warunki:

(A) $p_j(\beta) > 0$ dla $j = 1, 2, \ldots, m$;

(B) $\dfrac{\partial p_j(\beta)}{\partial \beta_u}$, $\ j = 1, 2, \ldots, m, \ \ u = 1, 2, \ldots, q$, istnieją i są ciągłe ze względu na β;

(C) macierz $\mathbf{B} = \mathbf{B}(\beta) = \left[\dfrac{1}{\sqrt{p_j(\beta)}} \dfrac{\partial p_j(\beta)}{\partial \beta_u}\right]_{q \times m}$ jest rzędu q.

Przez $\mathbf{D} = \mathbf{D}(\beta) = \operatorname{diag}[p_j(\beta)]$ oznaczmy macierz diagonalną, w której na głównej przekątnej znajdują się prawdopodobieństwa $p_j(\beta)$ dla każdego $j = 1, 2, \ldots, m$, oraz przez $\mathbf{I_k}$ macierz jednostkową rzędu k. Z kolei niech $\mathbf{1}$ oznacza m-wymiarowy wektor kolumnowy składający się z samych jedynek.

Przy tak wprowadzonych oznaczeniach łatwo zauważyć, że różniczkując względem β obie strony (1), otrzymujemy relację

$$\mathbf{B}\mathbf{D}^{-1/2}\mathbf{p}(\beta) = \mathbf{B}\mathbf{D}^{-1/2}\mathbf{D}\mathbf{1} = 0. \tag{3}$$

Jest oczywiste, że estymator $\bar{\beta}$ nie może być całkiem dowolny i również powinien mieć odpowiednio „dobre” własności. Poniższa definicja precyzuje własność estymatora, która jest stosunkowo prosta do sprawdzenia i jest spełniona dla bardzo obszernej klasy estymatorów.

Definicja 1. Mówimy, że $\widehat{\beta} = T(X_1, X_2, \ldots, X_n)$ jest $\sqrt{n}$-zgodnym estymatorem parametru β, jeśli dla każdego $\beta \in \Gamma$ ciąg $\{\sqrt{n}(\widehat{\beta} - \beta)\}$ jest ograniczony według prawdopodobieństwa P_β, czyli

$$\forall \beta \in \Gamma \quad \forall \eta > 0 \quad \exists M = M(\beta, \eta) > 0 \quad \exists n_0 = n_0(\beta, \eta) \quad \forall n \geq n_0$$
$$P_\beta\big(\sqrt{n}\|\widehat{\beta} - \beta\| > M\big) \leq \eta,$$

gdzie $\|\cdot\|$ oznacza normę euklidesową w R^q.

Poniższe twierdzenie stanowi główny wynik artykułu i zawiera istotę podejścia Dzhaparidze-Nikulina.

Twierdzenie 1. *Niech $\widehat{\beta}$ będzie $\sqrt{n}$-zgodnym estymatorem parametru β, oraz niech spełnione będą założenia (A), (B), (C). Ponadto niech $\widehat{\mathbf{p}} = \mathbf{p}(\widehat{\beta})$, $\widehat{\mathbf{D}} = \mathbf{D}(\widehat{\beta})$, $\widehat{\mathbf{B}} = \mathbf{B}(\widehat{\beta})$ będą estymatorami odpowiednich wielkości. Wtedy przy prawdziwości hipotezy $\mathcal{H}_0$ statystyka*

$$\widehat{S} = \left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right)^T \left(\widehat{\mathbf{D}}^{-1} - \widehat{\mathbf{D}}^{-1/2}\widehat{\mathbf{B}}^T \left(\widehat{\mathbf{B}}\widehat{\mathbf{B}}^T\right)^{-1} \widehat{\mathbf{B}}\widehat{\mathbf{D}}^{-1/2}\right) \left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right) \tag{4}$$

ma asymptotyczny rozkład chi-kwadrat z $m - 1 - q$ stopniami swobody.

Statystyka $\widehat{S}$ dana wzorem (4) jest nazywana statystyką Dzhaparidze-Nikulina i może być użyta jako statystyka testowa hipotezy $\mathcal{H}_0$. Jej zaletą jest to, że dopuszcza użycie dowolnego $\sqrt{n}$-zgodnego estymatora, nie wymagając ograniczenia się do estymatorów szczególnej postaci, takich jak estymatory największej wiarogodności czy największej wiarogodności po zgrupowaniu danych.

Dzhaparidze i Nikulin w pracy z 1974 roku podali dowód tezy twierdzenia 1, tj. zbieżności $\widehat{S}$ do rozkładu chi-kwadrat. Przyjęli oni jednak założenia typu Cramera, znacznie mocniejsze od warunków (A), (B), (C). Podobne twierdzenie znajduje się również w książce Greenwood i Nikulina (1996). Dowód przedstawiony przez nich polega na użyciu rozwinięcia Taylora i rozważaniach analitycznych związanych z szacowaniem reszty i wymianą β na $\widehat{\beta}$. Natomiast dowód, który przedstawimy w następnym rozdziale stosuje metody algebraiczne i jest bardziej elementarny.

3. Dowód twierdzenia 1. Zanim przystąpimy do właściwego dowodu twierdzenia 1 przedstawimy kilka pomocniczych lematów oraz twierdzenie 2.

Dla $u = 1, 2, \ldots, q$, $\beta \in \Gamma$, rozważmy wektory

$$\frac{\partial \log \mathbf{p}(\beta)}{\partial \beta_u} = \left[\frac{\partial \log p_1(\beta)}{\partial \beta_u}, \frac{\partial \log p_2(\beta)}{\partial \beta_u}, \ldots, \frac{\partial \log p_m(\beta)}{\partial \beta_u}\right]^T.$$

Macierz utworzona z wektorów $\dfrac{\partial \log \mathbf{p}(\beta)}{\partial \beta_1}^T, \dfrac{\partial \log \mathbf{p}(\beta)}{\partial \beta_2}^T, \ldots, \dfrac{\partial \log \mathbf{p}(\beta)}{\partial \beta_q}^T$ jest postaci $\mathbf{BD}^{-1/2}$. Zatem dzięki założeniom (A) i (C) ma rząd q. Oznacza to, że powyższy układ wektorów jest dla każdego β liniowo niezależny w R^m.

Dla każdego ustalonego $\beta \in \Gamma$ rozważmy iloczyn skalarny w przestrzeni R^m określony wzorem $\mathbf{v}_1^T \mathbf{D}(\beta)\mathbf{v}_2$ dla $\mathbf{v}_1, \mathbf{v}_2 \in R^m$. Z (3) wynika, że $\dfrac{\partial \log \mathbf{p}(\beta)^T}{\partial \beta_u} \mathbf{D}(\beta)\mathbf{1} = 0$ dla $u = 1, 2, \ldots, q$, czyli że wektory $\dfrac{\partial \log \mathbf{p}(\beta)}{\partial \beta_u}$ są ortogonalne do wektora $\mathbf{1}$ w rozpatrywanym iloczynie skalarnym. Wybierzmy teraz $m-1-q$ wektorów $\mathbf{w}_1(\beta), \mathbf{w}_2(\beta), \ldots, \mathbf{w}_{m-1-q}(\beta)$ przestrzeni R^m, które są ortonormalne w powyższym iloczynie skalarnym, a ponadto są ortogonalne do wektora $\mathbf{1}$ oraz do każdego z wektorów $\dfrac{\partial \log \mathbf{p}(\beta)}{\partial \beta_u}$. W ten sposób stanowią one uzupełnienie układu wektorów

$$\mathbf{1}, \frac{\partial \log \mathbf{p}(\beta)}{\partial \beta_1}, \frac{\partial \log \mathbf{p}(\beta)}{\partial \beta_2}, \ldots, \frac{\partial \log \mathbf{p}(\beta)}{\partial \beta_q}$$

do bazy przestrzeni R^m. Oczywiste jest to, że wybór wektorów $\mathbf{w}_1(\beta), \mathbf{w}_2(\beta), \ldots, \mathbf{w}_{m-1-q}(\beta)$ nie jest jednoznaczny. Niech $\mathbf{W} = \mathbf{W}(\beta)$ będzie macierzą wymiaru $(m-1-q) \times m$, w której $\mathbf{w}_i(\beta)^T$, $i = 1, 2, \ldots, m-1-q$, są kolejnymi wierszami. Ze sposobu określenia wektorów $\mathbf{w}_1(\beta), \mathbf{w}_2(\beta), \ldots, \mathbf{w}_{m-1-q}(\beta)$ wynika szereg własności macierzy $\mathbf{W}(\beta)$, które zebrane są w poniższym lemacie.

Lemat 1. *Dla każdego $\beta \in \Gamma$ spełnione są następujące relacje:*

$$\mathbf{WDW}^T = \mathbf{I}_{m-1-q}, \tag{5}$$

$$\mathbf{WD1} = \mathbf{Wp}(\beta) = 0, \tag{6}$$

$$\mathbf{WD}^{1/2}\mathbf{B}^T = \mathbf{WD}\left(\mathbf{BD}^{-1/2}\right)^T = 0. \tag{7}$$

Sposób doboru macierzy $\mathbf{W}$ oraz pomysł dowodu poniższego lematu 2 został zaczerpnięty z książki Raynera i Besta (1989), rozdz. 7. Lemat ten stanowić będzie istotny krok w dowodzie twierdzenia 2.

Lemat 2. *Macierz $\mathbf{W}^T\mathbf{W}$ daje się wyrazić w postaci:*

$$\mathbf{W}^T\mathbf{W} = \mathbf{D}^{-1} - \mathbf{D}^{-1/2}\mathbf{B}^T\left(\mathbf{BB}^T\right)^{-1}\mathbf{BD}^{-1/2} - \mathbf{11}^T. \tag{8}$$

Dowód. Dla każdego β rozważmy macierz $\mathbf{W}^* = \mathbf{W}^*(\beta)$ wymiaru $m \times m$ daną w postaci blokowej

$$\mathbf{W}^* = \begin{bmatrix} \mathbf{W} \\ \mathbf{1}^T \\ \left(\mathbf{BB}^T\right)^{-1/2}\mathbf{BD}^{-1/2} \end{bmatrix}.$$

Wówczas korzystając z lematu 1 oraz z relacji (3) mamy

$$\mathbf{W}^*\mathbf{D}{\mathbf{W}^*}^T =$$

$$= \begin{bmatrix} \mathbf{W} \\ \mathbf{1}^T \\ \left(\mathbf{B}\mathbf{B}^T\right)^{-1/2}\mathbf{B}\mathbf{D}^{-1/2} \end{bmatrix} \mathbf{D} \begin{bmatrix} \mathbf{W}^T & \mathbf{1} & \mathbf{D}^{-1/2}\mathbf{B}^T\left(\mathbf{B}\mathbf{B}^T\right)^{-1/2} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{W}\mathbf{D}\mathbf{W}^T & \mathbf{W}\mathbf{D}\mathbf{1} & \mathbf{W}\mathbf{D}^{1/2}\mathbf{B}^T\left(\mathbf{B}\mathbf{B}^T\right)^{-1/2} \\ & \mathbf{1}^T\mathbf{D}\mathbf{1} & \mathbf{1}^T\mathbf{D}^{1/2}\mathbf{B}^T\left(\mathbf{B}\mathbf{B}^T\right)^{-1/2} \\ & & \left(\mathbf{B}\mathbf{B}^T\right)^{-1/2}\mathbf{B}\mathbf{D}^{-1/2}\mathbf{D}\mathbf{D}^{-1/2}\mathbf{B}^T\left(\mathbf{B}\mathbf{B}^T\right)^{-1/2} \end{bmatrix} = \mathbf{I}_m.$$

W macierzy z ostatniej linii powyższych równości, ze względu na jej symetrię, zostały pominięte elementy pod przekątną. Z powyższej relacji wynika, że macierz $\mathbf{W}^*$ jest nieosobliwa, a $\mathbf{D} = \left(\mathbf{W}^*\right)^{-1}\left({\mathbf{W}^*}^T\right)^{-1}$. Stąd

$$\mathbf{D}^{-1} = {\mathbf{W}^*}^T\mathbf{W}^*$$

$$= \begin{bmatrix} \mathbf{W}^T & \mathbf{1} & \mathbf{D}^{-1/2}\mathbf{B}^T\left(\mathbf{B}\mathbf{B}^T\right)^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{W} \\ \mathbf{1}^T \\ \left(\mathbf{B}\mathbf{B}^T\right)^{-1/2}\mathbf{B}\mathbf{D}^{-1/2} \end{bmatrix}$$

$$= \mathbf{W}^T\mathbf{W} + \mathbf{1}\mathbf{1}^T + \mathbf{D}^{-1/2}\mathbf{B}^T\left(\mathbf{B}\mathbf{B}^T\right)^{-1}\mathbf{B}\mathbf{D}^{-1/2}.$$

To kończy dowód. □

Zanim przedstawimy twierdzenie 2, udowodnimy jeszcze jeden lemat.

LEMAT 3. *Dla każdego* $\beta \in \Gamma$ *wektor losowy* $\dfrac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}$ *spełnia warunki*:

$$\left(\frac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}\right)^T \mathbf{1} = 0, \tag{9}$$

$$E_\beta\left(\frac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}\right) = 0, \tag{10}$$

$$E_\beta\left(\frac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}\right)\left(\frac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}\right)^T = \mathbf{D} - \mathbf{p}(\beta)\mathbf{p}(\beta)^T, \tag{11}$$

gdzie E_β *oznacza wartość oczekiwaną względem rozkładu* P_β.

Dowód. Mamy

$$\left(\frac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}\right)^T \mathbf{1} = \frac{1}{\sqrt{n}}\left(\mathbf{N}^T\mathbf{1} - n\mathbf{p}(\beta)^T\mathbf{1}\right)$$

$$= \frac{1}{\sqrt{n}}\left(\sum_{j=1}^m N_j - n\sum_{j=1}^m p_j(\beta)\right) = 0,$$

co dowodzi (9).

Dla dowodu (10) i (11) zapiszmy wektor $\dfrac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}}$ jako sumę niezależnych wektorów losowych o tym samym rozkładzie

$$\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{bmatrix} 1_{A_1}(X_i)-p_1(\beta) \\ 1_{A_2}(X_i)-p_2(\beta) \\ \vdots \\ 1_{A_m}(X_i)-p_m(\beta)\end{bmatrix},$$

gdzie $1_A(\cdot)$ jest funkcją charakterystyczną zbioru A.

Ponieważ $E_\beta 1_{A_j}(X_i) = p_j(\beta)$ dla wszystkich i, j, to własność (10) jest spełniona. Z kolei z niezależności zmiennych $X_1, X_2, \ldots, X_n$ mamy

$$\left[E_\beta\left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}}\right)\left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}}\right)^T\right]_{rs} =$$
$$= E_\beta\left(1_{A_r}(X_1)-p_r(\beta)\right)\left(1_{A_s}(X_1)-p_s(\beta)\right) = p_r(\beta)\delta_{rs} - p_r(\beta)p_s(\beta),$$

gdzie δ_{rs} jest deltą Kroneckera, a $[\bullet]_{rs}$ oznacza rs-ty element macierzy $\bullet$. To dowodzi równości (11). □

Udowodnimy teraz pomocnicze twierdzenie 2, które będzie punktem wyjścia do dowodu twierdzenia 1.

TWIERDZENIE 2. *Jeśli spełnione są założenia* (*A*), (*B*), (*C*), *s. 111, to przy prawdziwości* $\mathcal{H}_0$ *statystyka testowa*

(12) $S =$
$$\left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}}\right)^T\left(\mathbf{D}^{-1}-\mathbf{D}^{-1/2}\mathbf{B}^T\left(\mathbf{B}\mathbf{B}^T\right)^{-1}\mathbf{B}\mathbf{D}^{-1/2}\right)\left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}}\right)$$

ma asymptotyczny rozkład chi-kwadrat z $m-1-q$ *stopniami swobody.*

Dowód. Korzystając z lematu 3 oraz z wielowymiarowego centralnego twierdzenia granicznego (por. Billingsley 1987, str. 383) otrzymujemy

$$\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}} \xrightarrow{D} \mathcal{N}\left(0\ ,\ \mathbf{D}-\mathbf{p}(\beta)\mathbf{p}(\beta)^T\right)$$

względem rozkładu P_β. W konsekwencji mamy

$$\mathbf{W}\left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}}\right) \xrightarrow{D} \mathcal{N}\left(0\ ,\ \mathbf{W}[\mathbf{D}-\mathbf{p}(\beta)\mathbf{p}(\beta)^T]\mathbf{W}^T\right)$$

względem rozkładu P_β. Z lematu 1 dostajemy

$$\mathbf{W}\left[\mathbf{D}-\mathbf{p}(\beta)\mathbf{p}(\beta)^T\right]\mathbf{W}^T = \mathbf{W}\mathbf{D}\mathbf{W}^T - \mathbf{W}\mathbf{p}(\beta)\mathbf{p}(\beta)^T\mathbf{W}^T = \mathbf{I}_{m-1-q},$$

więc

$$\mathbf{W}\left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}}\right) \xrightarrow{D} \mathcal{N}\left(0\ ,\ \ \mathbf{I}_{m-1-q}\right)$$

względem rozkładu P_β. Kwadrat normy euklidesowej jest funkcją ciągłą, więc statystyka testowa

$$\left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}}\right)^T \mathbf{W}^T\mathbf{W}\left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}}\right)$$

ma rozkład asymptotyczny chi-kwadrat z $m-1-q$ stopniami swobody względem P_β. Dzięki postaci macierzy $\mathbf{W}^T\mathbf{W}$ danej wzorem (8) oraz z własności (9) otrzymujemy (12). $\square$

Udowodnione powyżej lematy i twierdzenie pozwalają dowieść prawdziwości twierdzenia 1.

Dowód twierdzenia 1. Niech $\widehat{\beta}$ będzie $\sqrt{n}$-zgodnym estymatorem parametru β oraz niech $\widehat{\mathbf{p}}=\mathbf{p}(\widehat{\beta})$. Korzystając z twierdzenia o różniczce mamy

$$p_j(\widehat{\beta}) = p_j(\beta) + \frac{\partial p_j(\beta)}{\partial \beta}^T(\widehat{\beta}-\beta) + r_{jn}(\beta), \quad \text{gdzie} \quad \frac{r_{jn}(\beta)}{||\widehat{\beta}-\beta||} \xrightarrow{P_\beta} 0.$$

Zapisując powyższe równanie macierzowo i mnożąc obustronnie przez $\sqrt{n}$ dostajemy

$$\sqrt{n}(\widehat{\mathbf{p}}-\mathbf{p}(\beta)) = \mathbf{D}^{1/2}\mathbf{B}^T\sqrt{n}(\widehat{\beta}-\beta) + \sqrt{n}\mathbf{r}_n, \tag{13}$$

gdzie $\mathbf{r}_n = [r_{1n}(\beta), r_{2n}(\beta), \ldots, r_{mn}(\beta)]^T$.

Podstawiając do statystyki S, danej wzorem (12) estymator $p_j(\widehat{\beta})$ otrzymujemy statystykę $\widehat{S}^*$ postaci:

$$\begin{aligned}
\widehat{S}^* &= \left(\frac{\mathbf{N}-n\widehat{\mathbf{p}}}{\sqrt{n}}\right)^T \mathbf{Q}\left(\frac{\mathbf{N}-n\widehat{\mathbf{p}}}{\sqrt{n}}\right) \\
&= S - \left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}}\right)^T \mathbf{Q}\sqrt{n}\left(\widehat{\mathbf{p}}-\mathbf{p}(\beta)\right) \\
&\quad - \sqrt{n}\left(\widehat{\mathbf{p}}-\mathbf{p}(\beta)\right)^T \mathbf{Q}\left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}} - \sqrt{n}\left(\widehat{\mathbf{p}}-\mathbf{p}(\beta)\right)\right),
\end{aligned}$$

gdzie $\mathbf{Q} = \mathbf{D}^{-1} - \mathbf{D}^{-1/2}\mathbf{B}^T\left(\mathbf{B}\mathbf{B}^T\right)^{-1}\mathbf{B}\mathbf{D}^{-1/2}$.

Oznaczmy drugi składnik powyższej sumy jako S_1, a trzeci jako S_2.

Rozważmy najpierw S_2. Korzystając z (13) mamy

$$\begin{aligned}
S_2 &= \sqrt{n}\left(\widehat{\beta}-\beta\right)^T\left(\mathbf{B}\mathbf{D}^{-1/2}-\mathbf{B}\mathbf{D}^{-1/2}\right)\left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}} - \sqrt{n}\left(\widehat{\mathbf{p}}-\mathbf{p}(\beta)\right)\right) \\
&\quad + \sqrt{n}\mathbf{r}_n^T\mathbf{Q}\left(\frac{\mathbf{N}-n\mathbf{p}(\beta)}{\sqrt{n}} - \sqrt{n}\left(\widehat{\mathbf{p}}-\mathbf{p}(\beta)\right)\right).
\end{aligned}$$

Zauważmy, że w powyższej sumie pierwszy składnik zeruje się. Z definicji $\sqrt{n}$-zgodnego estymatora otrzymujemy, że wektor $\sqrt{n}\left(\widehat{\mathbf{p}}-\mathbf{p}(\beta)\right)$ jest ograniczony według prawdopodobieństwa P_β. Podobnie z dowodu twierdzenia 2 mamy, że $\dfrac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}$ jest ograniczony według prawdopodobieństwa P_β. Ponieważ wyrażenie $\mathbf{Q}$ jest stałe i zachodzi $\sqrt{n}\mathbf{r}_n = \sqrt{n}||\widehat{\beta} - \beta||\dfrac{\mathbf{r}_n}{||\widehat{\beta} - \beta||} \xrightarrow{P_\beta} 0$, to dostajemy $S_2 \xrightarrow{P_\beta} 0$. Analogicznie dowodzimy, że $S_1 \xrightarrow{P_\beta} 0$. W rezultacie otrzymujemy $\widehat{S}^* - S \xrightarrow{P_\beta} 0$.

Niech teraz $\widehat{\mathbf{D}} = \mathbf{D}(\widehat{\beta})$, $\widehat{\mathbf{B}} = \mathbf{B}(\widehat{\beta})$. Z ciągłości funkcji $p_j(\beta)$, $\dfrac{\partial p_j(\beta)}{\partial \beta}$ oraz ze zgodności $\widehat{\beta}$ otrzymujemy, że $\widehat{\mathbf{D}}$, $\widehat{\mathbf{B}}$ są estymatorami zgodnymi macierzy $\mathbf{D}$ i $\mathbf{B}$ odpowiednio. Podstawiając je do $\widehat{S}^*$ dostajemy statystykę

$$\widehat{S} = \left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right)^T \left(\widehat{\mathbf{D}}^{-1} - \widehat{\mathbf{D}}^{-1/2}\widehat{\mathbf{B}}^T \left(\widehat{\mathbf{B}}\widehat{\mathbf{B}}^T\right)^{-1} \widehat{\mathbf{B}}\widehat{\mathbf{D}}^{-1/2}\right) \left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right).$$

Niech

$$\Delta = \widehat{\mathbf{D}}^{-1} - \widehat{\mathbf{D}}^{-1/2}\widehat{\mathbf{B}}^T \left(\widehat{\mathbf{B}}\widehat{\mathbf{B}}^T\right) \widehat{\mathbf{B}}\widehat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1} + \mathbf{D}^{-1/2}\mathbf{B}^T \left(\mathbf{B}\mathbf{B}^T\right) \mathbf{B}\mathbf{D}^{-1/2}.$$

Z ciągłości operacji na macierzach wynika, że $\Delta \xrightarrow{P_\beta} 0$. Stąd i z ograniczenia według P_β wektora $\dfrac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}$ wynika

$$\widehat{S} - \widehat{S}^* = \left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right)^T \Delta \left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right) \xrightarrow{P_\beta} 0.$$

Zatem $\widehat{S} - S = \left(\widehat{S} - \widehat{S}^*\right) + \left(\widehat{S}^* - \widehat{S}\right) \xrightarrow{P_\beta} 0$. Ponieważ z twierdzenia 2 wynika, że $S \xrightarrow{D} \chi^2_{m-1-q}$ względem P_β, gdzie χ^2_{m-1-q} jest zmienną losową o rozkładzie chi-kwadrat z $m-1-q$ stopniami swobody, to teza twierdzenia 1 została wykazana. □

Z twierdzenia 1 wynika następujący wniosek.

Wniosek 1. *Załóżmy, że wektor prawdopodobieństw* $\mathbf{p}(\beta)$ *spełnia* (A), (B), (C). *Niech* $\widehat{\beta}$ *będzie pewnym* $\sqrt{n}$*-zgodnym estymatorem parametru* β *i niech* $\widehat{\mathbf{p}} = \mathbf{p}(\widehat{\beta})$, $\widehat{\mathbf{D}} = \mathbf{D}(\widehat{\beta})$, $\widehat{\mathbf{B}} = \mathbf{B}(\widehat{\beta})$ *będą estymatorami odpowiednich wielkości. Jeśli*

$$\widehat{\mathbf{B}}\widehat{\mathbf{D}}^{-1/2}\mathbf{N} = 0, \tag{14}$$

to przy prawdziwości $\mathcal{H}_0$ *statystyka*

$$\widehat{S} = \left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right)^T \widehat{\mathbf{D}}^{-1} \left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right) \tag{15}$$

ma asymptotyczny rozkład chi-kwadrat z $m-1-q$ *stopniami swobody.*

Dowód. Z (3) i (14) wynika, że

$$\widehat{\mathbf{D}}^{-1/2}\widehat{\mathbf{B}}^T\big(\widehat{\mathbf{B}}\widehat{\mathbf{B}}^T\big)^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{D}}^{-1/2}\left(\frac{\mathbf{N}-n\widehat{\mathbf{p}}}{\sqrt{n}}\right)=0.$$

Dzięki temu statystyka Dzhaparidze–Nikulina dana wzorem (4) redukuje się do postaci (15). □

Mówimy, że estymator $\widetilde{\beta}$ parametru β jest estymatorem największej wiarogodności po zgrupowaniu danych, jeśli maksymalizuje logarytm funkcji wiarogodności po zgrupowaniu danych, tj. $l(\beta)=const+\sum_{j=1}^m N_j\log p_j(\beta)$.

Zauważmy, że prawdziwy jest następujący lemat.

LEMAT 4. *Jeśli* (A), (B) *są spełnione oraz estymator największej wiarogodności po zgrupowaniu danych* $\widetilde{\beta}$ *parametru* β *istnieje, to*

$$\widetilde{\mathbf{B}}\widetilde{\mathbf{D}}^{-1/2}\mathbf{N}=0, \tag{16}$$

gdzie $\widetilde{\mathbf{B}}=\mathbf{B}(\widetilde{\beta})$, $\widetilde{\mathbf{D}}=\mathbf{D}(\widetilde{\beta})$.

Dowód. Niech $l(\beta)=const+\sum_{j=1}^m N_j\log p_j(\beta)$ będzie logarytmem funkcji wiarogodności po zgrupowaniu danych. Wtedy dzięki założeniom (A) i (B)

$$\begin{aligned}0=\frac{\partial l\left(\widetilde{\beta}\right)}{\partial\beta_u}&=\sum_{j=1}^m N_j\frac{\partial\log p_j(\widetilde{\beta})}{\partial\beta_u}\\&=\sum_{j=1}^m N_jp_j(\widetilde{\beta})^{-1}\frac{\partial p_j(\widetilde{\beta})}{\partial\beta_u}=\sum_{j=1}^m N_jp_j(\widetilde{\beta})^{-1/2}p_j(\widetilde{\beta})^{-1/2}\frac{\partial p_j(\widetilde{\beta})}{\partial\beta_u}\end{aligned}$$

dla $u=1,2,\ldots,q$. Powyższa relacja w zapisie macierzowym oznacza równość (16). □

Z powyższego wniosku oraz lematu wynika, że przy pewnych założeniach statystyki Pearsona i Dzhaparidze–Nikulina są sobie równe. Co więcej dzięki wnioskowi widzimy, że aby móc zastosować statystykę daną wzorem (15) nie jest konieczne wyznaczanie estymatora największej wiarogodności po zgrupowaniu danych, ale wystarczy by $\sqrt{n}$–zgodny estymator miał własność (14). Jest to łatwiejsze do sprawdzenia niż warunki istnienia estymatora $\widetilde{\beta}$.

4. $\widehat{S}$ jako statystyka wynikowa. Ten i następny rozdział mają charakter uzupełniający w stosunku do trzech poprzednich i mogą być pominięte. Jednakowoż stanowią one interesujące i ważne dopełnienie dotychczasowych rozważań.

W tym rozdziale pokażemy, iż statystyka Dzhaparidze–Nikulina dana wzorem (4) znajduje uzasadnienie w teorii testów wynikowych. Dla odpowiednio zdefiniowanego problemu testowania $\widehat{S}$ można zidentyfikować jako statystykę wynikową (ang. *score statistic*). Podobne twierdzenie można znaleźć w książce Raynera i Besta (1989), jednak ich dowód zawiera błędy i jest dość zagmatwany.

Teorię oraz niezbędne pojęcia dotyczące statystyk wynikowych można znaleźć na przykład w książkach Cox i Hinkley (1974) oraz Sen i Singer (1993).

Zanurzmy badaną rodzinę $\mathcal{P}$ w pewnej szerszej rodzinie parametrycznej wektorów prawdopodobieństw

$$\Pi = \left\{\pi(\theta,\beta) = [\pi_1(\theta,\beta), \pi_2(\theta,\beta), \ldots, \pi_m(\theta,\beta)] \colon \theta \in R^{m-1-q},\ \beta \in \Gamma\right\},$$

gdzie Γ jest otwartym podzbiorem przestrzeni R^q. Dla $j = 1, \ldots, m$

$$\pi_j = \pi_j(\theta,\beta) = C(\theta,\beta)\exp\{\sum_{i=1}^{m-1-q} \theta_i w_{ij}(\beta)\} p_j(\beta), \tag{17}$$

gdzie $C(\theta,\beta)$ jest stałą normującą, $\theta = [\theta_1, \theta_2, \ldots, \theta_{m-1-q}]^T$ jest wektorem parametrów rzeczywistych, $\beta = [\beta_1, \beta_2, \ldots, \beta_q]^T$ wektorem parametrów zakłócających, a $w_{ij}(\beta)$ dla $i = 1, \ldots, m-1-q$ oraz $j = 1, \ldots, m$ są elementami macierzy $\mathbf{W} = \mathbf{W}(\beta)$ zdefiniowanej w rozdziale 3 przed sformułowaniem lematu 1. Załóżmy ponadto, że dla wszystkich i, j, u oraz $\theta \in R^{m-1-q}$, $\beta \in \Gamma$ istnieją pochodne $\dfrac{\partial w_{ij}(\beta)}{\partial \beta_u}$.

Przypuśćmy, że niezależne obserwacje $X_1, X_2, \ldots, X_n$ mają po zgrupowaniu wektor prawdopodobieństw $\mathbf{p}$ należący do rodziny Π. Wtedy testowanie $\mathcal{H}_0 : \mathbf{p} \in \mathcal{P}$ jest równoważne z testowaniem hipotezy parametrycznej $\mathcal{H}_0^* : \theta = 0,\ \beta \in \Gamma$, dla której β jest parametrem zakłócającym. Oznaczmy przez $l(\theta,\beta)$ logarytm funkcji wiarogodności po zgrupowaniu danych, gdzie $\theta \in \mathbf{R}^{m-1-q}$, $\beta \in \Gamma$. Funkcja ta wyraża się wzorem

$$l(\theta,\beta) = const + \sum_{j=1}^{m} N_j \log \pi_j(\theta,\beta).$$

Dla skrócenia zapisu będziemy pisać l zamiast $l(\theta,\beta)$.

Zanim wyznaczymy statystykę wynikową dla testowania $\mathcal{H}_0^*$ w rodzinie Π, udowodnimy pomocniczy lemat.

Lemat 5. *Prawdziwe są następujące relacje:*

$$\sum_{j=1}^{m} \frac{\partial \pi_j(\theta,\beta)}{\partial \theta_r} = 0 \quad \forall r = 1, 2, \ldots, m-1-q, \tag{18}$$

$$\sum_{j=1}^{m} \frac{\partial \pi_j(\theta,\beta)}{\partial \beta_u} = 0 \quad \forall u = 1, 2, \ldots, q, \tag{19}$$

$$\frac{\partial \log C(\theta,\beta)}{\partial \theta_r} = -\sum_{j=1}^{m} w_{rj}(\beta)\pi_j(\theta,\beta) \qquad \forall r = 1,2,\ldots,m-1-q, \tag{20}$$

$$\left.\frac{\partial \log C(\theta,\beta)}{\partial \beta_u}\right|_{\theta=0} = 0 \quad \forall u = 1,2,\ldots,q. \tag{21}$$

Dowód. Z oczywistej relacji $\sum_{j=1}^{m} \pi_j(\theta,\beta) = 1$ wynika dowód równości (18) i (19). W celu pokazania własności (20) zlogarytmujmy obustronnie (17), otrzymując

$$\log \pi_j(\theta,\beta) = \log C(\theta,\beta) + \sum_{i=1}^{m-1-q} \theta_i w_{ij}(\beta) + \log p_j(\beta).$$

Następnie różniczkując obustronnie to wyrażenie względem θ_r, otrzymamy

$$\frac{\partial \pi_j(\theta,\beta)}{\partial \theta_r} = \pi_j(\theta,\beta)\left(\frac{\partial \log C(\theta,\beta)}{\partial \theta_r} + w_{rj}(\beta)\right). \tag{22}$$

Sumując względem j oraz wykorzystując relację (18), dostajemy

$$0 = \frac{\partial \log C(\theta,\beta)}{\partial \theta_r} + \sum_{j=1}^{m} w_{rj}(\beta)\pi_j(\theta,\beta)$$

co dowodzi (20).

Dowód (21) wynika wprost z faktu, że $C(\theta,\beta) = 1$ dla każdego $\beta \in \Gamma$ oraz $\theta = 0$. □

Twierdzenie 3. *Niech $\widehat{\beta}$ będzie $\sqrt{n}$-zgodnym estymatorem parametru β oraz niech będą spełnione założenia (A), (B), (C). Wtedy statystyka wynikowa dla testowania $\mathcal{H}_0^*: \theta = 0,\ \beta \in \Gamma$, przeciw $\mathcal{H}_1^*: \theta \neq 0,\ \beta \in \Gamma$, w rodzinie Π, jest postaci (4), czyli*

$$\widehat{S} = \left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right)^T \left(\widehat{\mathbf{D}}^{-1} - \widehat{\mathbf{D}}^{-1/2}\widehat{\mathbf{B}}^T\left(\widehat{\mathbf{B}}\widehat{\mathbf{B}}^T\right)^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{D}}^{-1/2}\right)\left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right),$$

gdzie $\widehat{\mathbf{p}} = \mathbf{p}(\widehat{\beta})$, $\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma}(\widehat{\beta})$, $\widehat{\mathbf{B}} = \mathbf{B}(\widehat{\beta})$, $\widehat{\mathbf{D}} = \mathbf{D}(\widehat{\beta})$ są estymatorami odpowiednich wielkości.

Dowód. W celu wyznaczenia wektora wynikowego dla rodziny Π zróżniczkujmy l względem θ_r i korzystając z (20) i (22), otrzymamy

$$\frac{\partial l}{\partial \theta_r} = \sum_{j=1}^{m} N_j \left(\frac{\partial \log C(\theta,\beta)}{\partial \theta_r} + w_{rj}(\beta)\right) = \sum_{j=1}^{m} w_{rj}(\beta)\left(N_j - n\pi_j(\theta,\beta)\right).$$

Jeśli $\mathcal{H}_0^*$ jest prawdziwa, tzn. $\pi_j(\theta,\beta) = p_j(\beta)$, to powyższa równość w zapisie macierzowym przyjmuje postać

$$\dot{l}_\theta = \left.\frac{\partial l}{\partial \theta}\right|_{\theta=0} = \mathbf{W}(\mathbf{N} - n\mathbf{p}(\beta)).$$

Różniczkowanie l względem β_u w punkcie $\theta = 0$ oraz korzystając z (19), (21), dostajemy

$$\left.\frac{\partial l}{\partial \beta_u}\right|_{\theta=0} = \sum_{j=1}^{m} \frac{\partial \log p_j(\beta)}{\partial \beta_u} N_j = \sum_{j=1}^{m} \frac{1}{p_j(\beta)} \frac{\partial p_j(\beta)}{\partial \beta_u} (N_j - np_j(\beta)).$$

W zapisie macierzowym daje to

$$\dot{l}_\beta = \left.\frac{\partial l}{\partial \beta}\right|_{\theta=0} = \mathbf{B}\mathbf{D}^{-1/2}\left(\mathbf{N} - n\mathbf{p}(\beta)\right).$$

A zatem wektor wynikowy ma następującą postać

$$\dot{l} = \left[\, (\mathbf{N} - n\mathbf{p}(\beta))^T\, \mathbf{W}^T \qquad (\mathbf{N} - n\mathbf{p}(\beta))^T\, \mathbf{D}^{-1/2}\mathbf{B}^T \,\right]^T. \tag{23}$$

Niech

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{\theta\theta} & \mathbf{K}_{\theta\beta} \\ \mathbf{K}_{\beta\theta} & \mathbf{K}_{\beta\beta} \end{bmatrix}$$

będzie macierzą kowariancji, w postaci blokowej, unormowanego wektora wynikowego $\frac{1}{\sqrt{n}}\left[\dot{l}_\theta, \dot{l}_\beta\right]^T$, dla testowania $\mathcal{H}_0^*$ w rodzinie Π. Korzystając z wzorów (3), (5), (6), (7), i (11), otrzymujemy

$$\begin{aligned}
\mathbf{K}_{\theta\theta} &= \frac{1}{n}\mathrm{Cov}_\beta\left(\dot{l}_\theta, \dot{l}_\theta\right) = \mathbf{W}\mathbf{D}\mathbf{W}^T - \mathbf{W}\mathbf{p}\mathbf{p}^T\mathbf{W}^T = \mathbf{I}_{m-1-q};\\
\mathbf{K}_{\theta\beta} &= \frac{1}{n}\mathrm{Cov}_\beta\left(\dot{l}_\theta, \dot{l}_\beta\right) = \mathbf{W}\left(\mathbf{D} - \mathbf{p}\mathbf{p}^T\right)\mathbf{D}^{-1/2}\mathbf{B}^T = \mathbf{W}\mathbf{D}^{1/2}\mathbf{B}^T = \mathbf{0};\\
\mathbf{K}_{\beta\beta} &= \frac{1}{n}\mathrm{Cov}_\beta\left(\dot{l}_\beta, \dot{l}_\beta\right) = \mathbf{B}\mathbf{D}^{-1/2}\mathbf{D}\mathbf{D}^{-1/2}\mathbf{B}^T - \mathbf{B}\mathbf{D}^{-1/2}\mathbf{p}\mathbf{p}^T\mathbf{D}^{-1/2}\mathbf{B}^T\\
&= \mathbf{B}\mathbf{B}^T,
\end{aligned}$$

gdzie Cov_β oznacza kowariancję względem rozkładu P_β. Zatem unormowana efektywna funkcja wynikowa ma postać

$$l^* = \frac{1}{\sqrt{n}}\left(\dot{l}_\theta - \mathbf{K}_{\theta\beta}\mathbf{K}_{\beta\beta}^{-1}\dot{l}_\beta\right) = \frac{1}{\sqrt{n}}\dot{l}_\theta = \mathbf{W}\frac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}.$$

Z kolei macierz kowariancji wektora l^* wyraża się wzorem

$$\mathbf{\Sigma} = \mathbf{K}_{\theta\theta} - \mathbf{K}_{\theta\beta}\mathbf{K}_{\beta\beta}^{-1}\mathbf{K}_{\beta\theta} = \mathbf{I}_{m-1-q}.$$

Ogólna postać statystyki wynikowej to $S(\beta) = {l^*}^T\mathbf{\Sigma}^{-1}l^*$ wyliczona w punkcie $\widehat{\beta}$. W naszym przypadku otrzymujemy

$$S(\beta) = \left(\frac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}\right)^T \mathbf{W}^T\mathbf{W} \left(\frac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}\right).$$

Korzystając z postaci macierzy $\mathbf{W}^T\mathbf{W}$ danej wzorem (8) oraz z własności (9), otrzymujemy

$$S(\beta) = \left(\frac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}\right)^T \left(\mathbf{D}^{-1} - \mathbf{D}^{-1/2}\mathbf{B}^T\left(\mathbf{B}\mathbf{B}^T\right)^{-1}\mathbf{B}\mathbf{D}^{-1/2}\right) \left(\frac{\mathbf{N} - n\mathbf{p}(\beta)}{\sqrt{n}}\right).$$

A zatem

$$S(\widehat{\beta}) = \widehat{S} = \left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right)^T \left(\widehat{\mathbf{D}}^{-1} - \widehat{\mathbf{D}}^{-1/2}\widehat{\mathbf{B}}^T\left(\widehat{\mathbf{B}}\widehat{\mathbf{B}}^T\right)^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{D}}^{-1/2}\right) \left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right),$$

co pokrywa się z postacią (4) i kończy dowód. □

Powyższe twierdzenie pokazuje, że statystyka testowa Dzhaparidze–Nikulina wpisuje się w teorię testów wynikowych. Dzięki temu możemy wnioskować, że test oparty na niej jest lokalnie asymptotycznie optymalny dla alternatyw pochodzących z rodziny Π.

5. Przykład. Dla zilustrowania praktycznego zastosowania testu chi-kwadrat opartego na statystyce danej wzorem (4) rozważmy typowe zagadnienie testowania zgodności w rodzinie z parametrami przesunięcia i skali, obejmujące w szczególności problem testowania normalności.

Niech f_0 będzie gęstością prawdopodobieństwa, dodatnią na R, a F_0 odpowiadającą jej dystrybuantą. Rozważmy rodzinę rozkładów

$$\mathcal{P}^0 = \left\{P_{\mu,\sigma} : \frac{dP_{\mu,\sigma}}{dx} = \frac{1}{\sigma} f_0\left(\frac{x-\mu}{\sigma}\right),\ \mu \in R,\ \sigma \in R^+\right\}$$

i testowanie hipotezy $\mathcal{H}_0^0 :\ P \in \mathcal{P}^0$. W tym przypadku parametr zakłócający β jest dwuwymiarowy i ma postać $\beta = [\mu, \sigma]^T \in \Gamma = R \times (0, \infty)$. Przyjmijmy, że podzbiory $A_1, A_2, \ldots, A_m$, $m > 3$, są wyznaczone przez przedziały $(-\infty, a_1], (a_1, a_2], \ldots, (a_{m-1}, \infty)$, gdzie $a_1 < a_2 < \ldots < a_{m-1}$. Wówczas, utrzymując oznaczenia z rozdziałów 2, 3, mamy

$$p_j(\mu, \sigma) = F_0\left(\frac{a_j - \mu}{\sigma}\right) - F_0\left(\frac{a_{j-1} - \mu}{\sigma}\right),\ j = 2, 3, \ldots, m-1 \tag{24}$$

oraz

$$p_1(\mu, \sigma) = F_0\left(\frac{a_1 - \mu}{\sigma}\right), \qquad p_m(\mu, \sigma) = 1 - F_0\left(\frac{a_{m-1} - \mu}{\sigma}\right). \tag{25}$$

W konsekwencji dla $j = 2, 3, \ldots, m-1$ otrzymujemy

$$\frac{\partial p_j(\mu,\sigma)}{\partial \mu} = -\frac{1}{\sigma}\left[f_0\left(\frac{a_j-\mu}{\sigma}\right) - f_0\left(\frac{a_{j-1}-\mu}{\sigma}\right)\right],$$

$$\frac{\partial p_j(\mu,\sigma)}{\partial \sigma} = -\frac{1}{\sigma}\left[\frac{a_j-\mu}{\sigma}f_0\left(\frac{a_j-\mu}{\sigma}\right) - \frac{a_{j-1}-\mu}{\sigma}f_0\left(\frac{a_{j-1}-\mu}{\sigma}\right)\right],$$

oraz

$$\begin{aligned}
\frac{\partial p_1(\mu,\sigma)}{\partial \mu} &= -\frac{1}{\sigma}f_0\left(\frac{a_1-\mu}{\sigma}\right),\\
\frac{\partial p_m(\mu,\sigma)}{\partial \mu} &= -\frac{1}{\sigma}\left[-f_0\left(\frac{a_{m-1}-\mu}{\sigma}\right)\right],\\
\frac{\partial p_1(\mu,\sigma)}{\partial \sigma} &= -\frac{1}{\sigma}\left[\frac{a_1-\mu}{\sigma}f_0\left(\frac{a_1-\mu}{\sigma}\right)\right],\\
\frac{\partial p_m(\mu,\sigma)}{\partial \sigma} &= -\frac{1}{\sigma}\left[-\frac{a_{m-1}-\mu}{\sigma}f_0\left(\frac{a_{m-1}-\mu}{\sigma}\right)\right].
\end{aligned}$$

Niech $X_1, X_2, \ldots, X_n$ będzie próbą, natomiast $\widehat{\mu}$, $\widehat{\sigma}$ ustalonymi $\sqrt{n}$-zgodnymi estymatorami parametrów μ, σ w rodzinie $\mathcal{P}^0$. Ponadto niech $\mathbf{N} = [N_1, N_2, \ldots, N_m]^T$ będzie wektorem liczebności empirycznych w przyjętych klasach. Dla uproszczenia przyjmijmy $b_j = \dfrac{a_j - \widehat{\mu}}{\widehat{\sigma}}$ $(j = 1, 2, \ldots, m-1)$ oraz $\widehat{p}_j = p_j(\widehat{\mu}, \widehat{\sigma})$ $(j = 1, 2, \ldots, m)$. Zdefiniujmy również następujące wektory w przestrzeni R^m :

$$\text{(26)}\quad \mathbf{u} = \left[\frac{N_1 - n\widehat{p}_1}{\sqrt{n\widehat{p}_1}}, \frac{N_2 - n\widehat{p}_2}{\sqrt{n\widehat{p}_2}}, \ldots, \frac{N_m - n\widehat{p}_m}{\sqrt{n\widehat{p}_m}}\right]^T = \widehat{\mathbf{D}}^{-1/2}\left(\frac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}}\right),$$

$$\text{(27)}\quad \mathbf{v} = \left[\frac{f_0(b_1)}{\sqrt{\widehat{p}_1}}, \frac{f_0(b_2) - f_0(b_1)}{\sqrt{\widehat{p}_2}}, \ldots, \frac{-f_0(b_{m-1})}{\sqrt{\widehat{p}_m}}\right]^T,$$

$$\text{(28)}\quad \mathbf{w} = \left[\frac{b_1 f_0(b_1)}{\sqrt{\widehat{p}_1}}, \frac{b_2 f_0(b_2) - b_1 f_0(b_1)}{\sqrt{\widehat{p}_2}}, \ldots, \frac{-b_{m-1} f_0(b_{m-1})}{\sqrt{\widehat{p}_m}}\right]^T.$$

Wówczas macierz $\widehat{\mathbf{B}}$, wymiaru $2 \times m$, ma postać $\widehat{\mathbf{B}} = -\dfrac{1}{\sigma}\begin{bmatrix}\mathbf{v}\\ \mathbf{w}\end{bmatrix}$. Stąd dostajemy

$$(\widehat{\mathbf{B}}\widehat{\mathbf{B}}^T)^{-1} = \frac{\sigma^2}{\|\mathbf{v}\|^2\|\mathbf{w}\|^2 - (\mathbf{v}^T\mathbf{w})^2}\begin{bmatrix}\|\mathbf{w}\|^2 & -\mathbf{v}^T\mathbf{w}\\ -\mathbf{v}^T\mathbf{w} & \|\mathbf{v}\|^2\end{bmatrix},$$

gdzie $\|\cdot\|$ oznacza normę euklidesową w R^m. Ponadto mamy, że $\widehat{\mathbf{B}}\widehat{\mathbf{D}}^{-1/2}\dfrac{\mathbf{N} - n\widehat{\mathbf{p}}}{\sqrt{n}} = \begin{bmatrix}\mathbf{u}^T\mathbf{v}\\ \mathbf{u}^T\mathbf{w}\end{bmatrix}$. A zatem ostatecznie statystyka $\widehat{S}$ dana wzorem (4) daje się zapisać w następującej, wygodnej do obliczeń, postaci

$$\widehat{S} = \|\mathbf{u}\|^2 - \frac{\|\mathbf{v}\|^2(\mathbf{w}^T\mathbf{u})^2 + \|\mathbf{w}\|^2(\mathbf{v}^T\mathbf{u})^2 - 2(\mathbf{v}^T\mathbf{w})(\mathbf{v}^T\mathbf{u})(\mathbf{w}^T\mathbf{u})}{\|\mathbf{v}\|^2\|\mathbf{w}\|^2 - (\mathbf{v}^T\mathbf{w})^2}. \tag{29}$$

Zauważmy, że $\|\mathbf{u}\|^2$ ma tę samą postać co klasyczna statystyka Pearsona dana wzorem (2).

Powyższy artykuł powstał na podstawie naszej pracy magisterskiej napisanej pod opieką dr hab. T. Inglota, któremu jesteśmy bardzo wdzięczne za wsparcie, liczne dyskusje oraz inspirujące komentarze. Serdecznie dziękujemy również prof. dr hab. T. Ledwinie za cenne uwagi oraz za pomoc w uzyskaniu niektórych potrzebnych artykułów.

Literatura cytowana

[1] P. Billingsley (1987), *Prawdopodobieństwo i miara*, PWN, Warszawa.
[2] D. R. Cox, D. V. Hinkley (1974), *Theoretical Statistics*, Chapman and Hall, London.
[3] K. O. Dzhaparidze, M. S. Nikulin (1974), *On a modification of the standard statistic of Pearson*, Theor. Prob. Appl., 19, 851–853.
[4] R. A. Fisher (1924), *The conditions under which χ^2 measures the discrepancy between observation and hypothesis*, J. R. Statist. Soc. 87, 442–450.
[5] P.E. Greenwood, M.S. Nikulin (1996), *A Guide to Chi-Squared Testing*, Wiley, New York.
[6] K. Pearson (1900), *On the criterion that a given system of deviation from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, Phil. Mag. 50(5), 157–172.
[7] J. C. W. Rayner, D. J. Best (1989), *Smooth Tests of Goodness of Fit*, Oxford Univ. Press, New York.
[8] P. K. Sen, J. M. Singer (1993), *Large Sample Methods in Statistics*, Chapman and Hall, New York.

Anita Małoń
Uniwersytet Wrocławski
pl. Uniwersytecki 1, 50-137 Wrocław
E-Mail: anita.malon@onet.eu

Dagmara Ziółkowska
Uniwersytet Wrocławski
pl. Uniwersytecki 1, 50-137 Wrocław
E-Mail: dagmara.ziolkowska@wp.pl

Chi-square type goodness of fit tests for composite hypothesis

Abstract. Adapting the classical Pearson's chi-square goodness-of-fit test for testing composite hypotheses brings serious problems with estimation of unknown parameters. An interesting solution which eliminates them was proposed by Dzhaparidze and Nikulin. The most important advantage of their solution is a possibility of using arbitrary estimators satisfying only a natural and weak condition. The aim of the present article is to popularize this solution. We provide a complete, short and, what is more elementary proof of the main theorem on asymptotic distribution of the test statistic. In addition, we prove that the constructed test belongs to a general class of score tests what advocates for its good properties. Finally, as an example, we give a typical implementation of the test to testing in location and scale family.
Key words: Chi-square test, composite hypothesis, $\sqrt{n}$-consistent estimator, Dzhaparidze–Nikulina statistic, score test.

(*wpłynęło 14 stycznia 2007 r.*)