

RYSZARD ZIELIŃSKI
 WOJCIECH ZIELIŃSKI
 Warszawa

O odpornym estymatorze wariancji w modelu liniowym

(Praca wpłynęła do Redakcji 11.11.1982)

1. WSTĘP. SFORMUŁOWANIE ZAGADNIENIA

Rozważamy model liniowy $y = X\beta + \varepsilon$, w którym $y^T = (y_1, y_2, \dots, y_n)$ jest wektorem obserwacji, $\beta^T = (\beta_1, \beta_2, \dots, \beta_k)$ jest wektorem (nieznanych) współczynników regresji, $X = (x_{ij})$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, k$, $k \leq n$, jest daną macierzą rzędu r (macierzą planu eksperymentu) oraz $\varepsilon^T = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ jest wektorem „zakłóceń”: ε_i , $i = 1, 2, \dots, n$, są niezależnymi zmiennymi losowymi o jednakowym rozkładzie mającym cztery pierwsze momenty centralne μ_p , $p = 1, 2, 3, 4$ ($\sigma^2 = \mu_2$). Zakładamy przy tym, że $E\varepsilon_i = 0$ oraz że $\gamma_2 := (\mu_4/\sigma^4) - 3 = \gamma_2^0$, gdzie γ_2^0 jest ustaloną liczbą (wiadomo, że zawsze $\gamma_2 \geq -2$).

Zadanie polega na skonstruowaniu takiego estymatora wariancji σ^2 , którego wariancja zmieniałaby się możliwie mało, gdy kurtoza γ_2 modelowego rozkładu ulega zmianie i przebiega pewien przedział $(\underline{\gamma}_2, \bar{\gamma}_2)$, $\underline{\gamma}_2 \leq \gamma_2^0 \leq \bar{\gamma}_2$.

Przypadek modeli gaussowskich (ε_1 ma rozkład normalny, $\gamma_2^0 = 0$) był rozważany już w 1931 roku przez Pearsona [5] i później w licznych pracach, np. w 1953 roku przez Boxa [1] i w 1955 roku przez Boxa i Andersena [2]. W pracach tych koncentrowano uwagę na badaniu konsekwencji zakłóceń modelu gaussowskiego polegających na zmianie kurtozy γ_2 , przy czym konsekwencje te opisywano za pomocą zmian rozmiaru różnych testów (testu t Studenta, testu F Snedecora i licznych testów służących do porównywania wariancji). Obszerny przegląd tych badań oraz oryginalne wyniki przedstawione zostały w 1982 roku w pracy Nürnberga [4].

W niniejszej pracy ograniczamy się do rozważania konsekwencji polegających na tym, że w wyniku zmiany kurtozy może zmienić się wariancja estymatora (w wyniku tego może również ulec zmianie moc i rozmiar testów oraz poziom ufności przedziałów ufności, opartych na wariancji). W naszych rozważaniach założenie o normalności rozkładu nie jest potrzebne. Ograniczymy się do badania estymatorów postaci $y^T A y$, gdzie - bez straty ogólności - A jest macierzą symetryczną, i efektywnie wyznaczymy najbardziej odporny, ze względu na zmianę wariancji, estymator wariancji w tej klasie. Interesujące jest, że jeżeli $y_1, y_2, \dots, y_n$ jest ciągiem niezależnych obserwacji o jednakowym rozkładzie, to okazuje się, że takim estymatorem jest standardowy estymator s^2 .

Formalnie zagadnienie można opisać w następujący sposób. Niech a oznacza wektor utworzony z przekątnej macierzy A , tzn. niech $a = \text{diag } A$. Wtedy

$$E y^T A y = \sigma^2 \text{tr} A + \beta^T X^T A X \beta,$$

(1)

$$\text{Var } y^T A y = \sigma^4 (\gamma_2^0 a^T a + 2 \text{tr}(A^2)) + 4\sigma^2 \beta^T X^T A^2 X \beta + 4\mu_3 \beta^T X^T A a.$$

Wzoruując się na koncepcji odporności z pracy [7], opiszemy odporność estymatora $y^T A y$ za pomocą wielkości

$$\bar{\gamma}_2 \leq \gamma_2 \leq \bar{\gamma}_2 \quad \text{Var } y^T A y - \inf_{\gamma_2 \leq \gamma_2 \leq \bar{\gamma}_2} \text{Var } y^T A y = (\bar{\gamma}_2 - \gamma_2) a^T a;$$

wtedy zadanie sprowadza się do znalezienia estymatora $y^T A y$ o minimalnym $a^T a$. Bez żadnych ograniczeń na macierz A (a więc na estymator) otrzymujemy trywialne rozwiązanie $a \equiv 0$; są to jednak estymatory bezsensowne, gdyż wówczas $E y^T A y = \beta^T X^T A X \beta$. Ograniczymy więc zbiór rozważanych macierzy A do klasy $\mathcal{A}$ macierzy symetrycznych takich, że

$$(\forall \beta, X) E y^T A y = \sigma^2 \quad - \text{nieobciążoność estymatora,}$$

$$(\forall \beta, X) (y - X\beta)^T A (y - X\beta) = y^T A y \quad - \text{niezmienniczość estymatora.}$$

Ze wzorów (1) wynika, że $A \in \mathcal{A}$ wtedy i tylko wtedy, gdy A jest macierzą symetryczną spełniającą warunki

$$(2) \quad \begin{cases} \text{tr } A = 1, \\ AX = 0. \end{cases}$$

Ostatecznie zadanie brzmi: znaleźć macierz $A \in \mathcal{A}$, dla której $a^T a$ osiąga minimum.

2. CHARAKTERYZACJA ZBIORU $\mathcal{A}$

Niech $m = \dim \text{Ker } X^T$. Wykażemy, że $A \in \mathcal{A}$ wtedy i tylko wtedy, gdy

$$(3) \quad A = B/\text{tr} B,$$

gdzie

$$B = \sum_{i=1}^m \sum_{j=1}^m \beta_{ij} (b_i b_j^T + b_j b_i^T)$$

dla pewnej bazy $b_1, b_2, \dots, b_m$ przestrzeni $\text{Ker } X^T$ oraz pewnych rzeczywistych β_{ij} .

Dostateczność warunku jest oczywista. Wykażemy jego konieczność. Niech $A \in \mathcal{A}$. Z warunków (2) wynika, że $A = B/\text{tr} B$, gdzie B jest dowolną macierzą symetryczną taką, że $BX = 0$. Ale wtedy

$$(4) \quad B = y_1 b_1^T + y_2 b_2^T + \dots + y_m b_m^T$$

dla pewnej bazy $b_1, b_2, \dots, b_m$ przestrzeni $\text{Ker } X^T$ i dla pewnych wektorów $y_1, y_2, \dots, y_m \in \mathbb{R}^n$. Ponieważ macierz B ma być symetryczna, więc musi być również spełniony warunek $B^T X = 0$. Ponieważ

$B^T = b_1 y_1^T + b_2 y_2^T + \dots + b_m y_m^T$, więc $y_1, y_2, \dots, y_m$ muszą również należeć do $\text{Ker } X^T$. Zatem

$$y_i = \beta_{1i}b_1 + \beta_{2i}b_2 + \dots + \beta_{mi}b_m \quad (i = 1, 2, \dots, m)$$

dla pewnych β_{ij} . Stąd otrzymujemy

$$B = \sum_{i=1}^m \sum_{j=1}^m \beta_{ij} b_i b_j^T.$$

Ponieważ układ macierzy $b_i b_j^T$ ($i, j = 1, 2, \dots, m$) jest liniowo niezależny, więc z warunku $B = B^T$ otrzymujemy $\beta_{ij} = \beta_{ji}$ i stąd postać (3).

U w a g a. 1 Jeżeli $r = k = n$, to $\mathcal{A}$ jest zbiorem pustym; 2) jeżeli $r = n - 1$, to $\mathcal{A}$ jest zbiorem jednoelementowym; 3) jeżeli $r \leq n - 2$, to $\mathcal{A}$ zawiera nieskończenie wiele elementów.

D o w ó d u w a g i. Ad 1) Jediną macierzą spełniającą $AX = 0$ jest $A = 0$. Wówczas jednak nie zachodzi $\text{tr } A = 1$. Ad 2) Jeżeli $r = n - 1$, to $m = 1$ i wtedy jedyną macierzą postaci (3) jest $A = b_1 b_1^T / b_1^T b_1$. Ad 3) Jeżeli $r \leq n - 2$, to $m \geq 2$. Niech $b_1, b_2 \in \text{Ker } X^T$ będą wektorami liniowo niezależnymi. Jeżeli A_1 oraz A_2 są macierzami skonstruowanymi tak jak A w Ad 2), odpowiednio dla b_1 i b_2 , to $A_1 \in \mathcal{A}$, $A_2 \in \mathcal{A}$ oraz każda wypukła kombinacja liniowa tych macierzy należą do $\mathcal{A}$.

3. ESTYMATOR NAJBARDZIEJ ODPORNY

Niech $r \leq n - 2$. Zadanie wyznaczenia estymatora najbardziej odpornego będziemy rozwiązywali w dwóch etapach: najpierw wyznaczmy przekątną $a^0 = (a_1^0, a_2^0, \dots, a_n^0)^T$ estymatora najbardziej odpornego, a następnie wyznaczmy macierz A tego estymatora.

Jeżeli B ma postać (3), to

$$\text{diag } B = \sum_{i=1}^m \sum_{j=1}^m \beta_{ij} \sigma_{ij},$$

gdzie $\sigma_{ij} = \text{diag}(b_i b_j^T + b_j b_i^T)$. Niech d będzie liczbą liniowo niezależnych wektorów σ_{ij} ($i = 1, 2, \dots, m$; $j = i, i+1, \dots, m$). Oczywiście liczba d nie zależy od wyboru bazy $b_1, b_2, \dots, b_m$. Wyznaczenie przekątnej a^0 sprowadza się do wyznaczenia, w przestrzeni liniowej rozpiętej przez wektory σ_{ij} ($i = 1, 2, \dots, m$; $j = i, i+1, \dots, m$), wektora $a^0 = (a_1^0, a_2^0, \dots, a_n^0)^T$, dla którego

$$\sum_{i=1}^n a_i^0 = 1 \text{ oraz } \sum_{i=1}^n (a_i^0)^2 = \min. \text{ Jak wiadomo, zadanie to ma zawsze}$$

Jak wiadomo, zadanie to ma zawsze dokładnie jedno rozwiązanie.

Niech Δ będzie macierzą, której kolumnami są wektory σ_{ij} ($i = 1, 2, \dots, m$; $j = i, i+1, \dots, m$) i niech $\beta^T = (\beta_{11}, \beta_{12}, \dots, \beta_{1m}, \beta_{22}, \beta_{23}, \dots, \beta_{mm})$. Zadanie wyznaczenia macierzy A postaci (3) dla estymatora najbardziej odpornego sprowadza się teraz do wyznaczenia rozwiązań β równania

$$(5) \quad \Delta \beta = a^0.$$

Niech $\mathcal{A}_R \subset \mathcal{A}$ oznacza zbiór macierzy A takich, że $\text{diag } A = a^0$. Ponieważ macierze $A \in \mathcal{A}_R$ są jednoznacznie wyznaczone przez rozwiązanie β równania (5), więc w przypadku $d = \frac{1}{2} m(m+1)$ zbiór $\mathcal{A}_R$ jest jednoelementowy, a w przeciwnym przypadku, tzn. gdy $d < \frac{1}{2} m(m+1)$, zawiera on nieskończenie wiele elementów. Każda z macierzy $A \in \mathcal{A}_R$ wyznacza najbardziej odporny estymator $y^T A y$ wariancji σ^2 . Spośród tych estymatorów możemy wybrać optymalny w sensie jakiegoś dodatkowego kryterium, np. kryterium minimalnej wariancji.

4. MINIMALIZACJA WARIANCJI W KLASIE $\mathcal{A}_R$

Niech $d < \frac{1}{2} m(m+1)$. Wyznaczenie w klasie $\mathcal{A}_R$ estymatora o minimalnej wariancji sprowadza się do rozwiązania zadania

$$\text{tr}(A^2) = \min!$$

$$AX = 0,$$

$$\text{tr} AV_i = a_i^0, \quad i = 1, 2, \dots, n,$$

gdzie $V_i = (\delta_{k,m}^{(i,i)})_m$, $k = 1, 2, \dots, n$, gdzie $\delta_u^{(v)}$ jest deltą Krockera (warunek $\text{tr} AV_i = \infty$ jest równoważny warunkowi $a_{ii} = \infty$).

Rozwiązanie tak sformułowanego zadania podane jest w monografii Rao [6], § 1.6.3(ii). Rozwiązaniem jest macierz

$$A_{**} = (a_{ij}^*), \quad i, j = 1, 2, \dots, n,$$

gdzie

$$a_{ij}^* = \sum_{p=1}^n \lambda_p q_{ip} q_{pj},$$

$$Q = (q_{ij}) = I - X(X^T X)^{-1} X^T$$

oraz

$$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$$

jest rozwiązaniem układu równań liniowych

$$\sum_{p=1}^n \lambda_p q_{pj}^2 = a_j^0, \quad j = 1, 2, \dots, n.$$

5. PORÓWNIANIE ZE STANDARDOWYM ESTYMATOREM WARIANCJI

W ogólnym przypadku estymator y_{Ax}^T nie musi pokrywać się ze standardowym estymatorem wariancji związanym z macierzą

$$S = (I - X(X^T X)^{-1} X^T) / (n - r);$$

przykład podajemy niżej. Estymatory te pokrywają się na przykład wtedy, gdy przekątna macierzy S ma postać $(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})^T$. Wynika to stąd, że taka macierz S jest macierzą o minimalnym kwadracie normy śladu $\sum a_{ii}^2$ w klasie macierzy $A = (a_{ij}) \in \mathcal{A}$ i ponadto estymator z macierzą S jest estymatorem nieobciążonym o minimalnej wariancji (dowód tego faktu można znaleźć np. w [3], § 8.19).

Jako wniosek otrzymujemy, że w przypadku, gdy $y_1, y_2, \dots, y_n$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie, najbardziej odporny estymator o minimalnej wariancji pokrywa się z estymatorem standardowym

$$s^2 = \sum (y_i - \bar{y})^2 / (n-1).$$

Odnotujmy, że wtedy

$$\text{Var } s^2 = \left(\frac{\chi^2}{n} + \frac{2}{n-1} \right) \sigma^4$$

oraz funkcja odporności jest równa $\frac{1}{n}(\bar{\gamma}_2 - \underline{\gamma}_2)$.

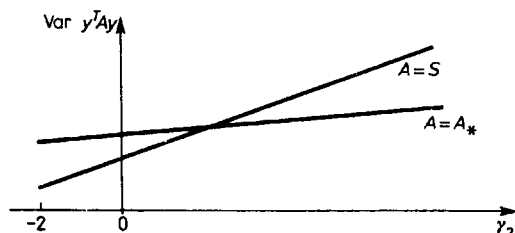
W szczególności wynika stąd bardzo ważny wniosek, że zaobserwowanych przez Pearsona (1931) i badanych przez Boxa (1953), Boxa i Andersena (1955), Nürnberga (1982) i innych niestabilności testów na porównywanie wariancji nie da się usunąć przez oparcie statystyki testowej na bardziej odpornym estymatorze wariancji postaci $y^T A y$.

6. PRZYKŁAD

Niech $X^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$. Wówczas

$$S = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/4 & -1/4 & 0 \\ 0 & -1/4 & 1/4 & 0 \\ 0 & 0 & 0 & 1/2 \end{bmatrix}, \quad A_{\#} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/3 & -1/3 & 0 \\ 0 & -1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 \end{bmatrix}$$

oraz $\sigma^{-4} \text{Var } y^T S y = \frac{3}{8} \gamma_2 + 1$, $\sigma^{-4} \text{Var } y^T A_{\#} y = \frac{1}{3} \gamma_2 + \frac{5}{4}$.



Funkcja odporności dla tych estymatorów wynosi, odpowiednio, $\frac{3}{8}(\bar{\gamma}_2 - \underline{\gamma}_2)$ oraz $\frac{1}{3}(\bar{\gamma}_2 - \underline{\gamma}_2)$.

PRACE CYTOWANE

- [1] G.E.P. Box, Non-normality and tests on variances, Biometrika 40 (1953), 318-335.
- [2] G.E.P. Box, S.L. Andersen, Permutation theory in the derivation of robust criteria and the study of departures from assumptions, J. Roy. Statist. Society, Ser. B 17 (1955), 1-34.
- [3] D.R. Cox, D.V. Hinkley, Problems and solutions in theoretical statistics, Chapman and Hall, London, John Wiley and Sons, New York 1978 (istnieje tłumaczenie rosyjskie 1981).
- [4] G. Nürnberg, Beiträge zur Versuchsplanung für die Schätzung von Varianzkomponenten und Robustheitsuntersuchungen zum Vergleich zweier Varianzen, Probleme der angewandten Statistik, Heft 6, Dummerstorf-Rostock, 1982.
- [5] E.S. Pearson, The analysis of variance in cases of non-normal variation, Biometrika 23 (1931), 114-135.
- [6] C.R. Rao, Modele liniowe statystyki matematycznej, PWN, Warszawa 1982.
- [7] R. Zieliński, O mierzeniu odporności statystyk, Matematyka Stosowana, XII (1978), 71-76.