Akademia Górniczo-Hutnicza
im. Stanisława Staszica
w Krakowie

Praca licencjacka

Testowanie regresji liniowej przeciwko regresji wypukłej

Grzegorz Mika

Kierunek: Matematyka

Nr albumu: 267543

Promotor
dr Konrad Nosek

Wydział Matematyki Stosowanej

Kraków 2016

Oświadczenie autora

Ja, niżej podpisany Grzegorz Mika oświadczam, że praca ta została napisana samodzielnie i wykorzystywała (poza zdobytą na studiach wiedzą) jedynie wyniki prac zamieszczonych w spisie literatury.

. . . .
(Podpis autora)

Oświadczenie promotora

Oświadczam, że praca spełnia wymogi stawiane pracom licencjackim.

. . . .
(Podpis promotora)

Spis treści

Wstęp

Rozważmy problem dopasowania pewnej funkcji f opisującej związek między zmiennymi objaśniejącymi x_i a zmiennymi objaśnianymi y_i według następującego modelu

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

gdzie zakładamy, że błędy ε_i są niezależnymi zmiennymi losowym o tym samym rozkładzie normalnym $N(0, \sigma^2)$. Punkty x_i są ustalone. Ponadto zakładać będziemy, że wariancja σ^2 jest znana.

Najprostszym związkiem między zmiennymi x_i a odpowiedziami y_i jest zależność liniowa. Możliwy jest jednak również inna zależność między zmiennymi a odpowiedziami. W niniejszej pracy rozważać będziemy problem, czy funkcja f jest funkcją liniową czy pewną funkcją wypukłą. Prowadzi do problemu testowania hipotez

$$H_0 \colon f(x) = ax + b \qquad vs. \qquad H_1 \colon f \in \mathcal{F},$$

gdzie $\mathcal{F}$ jest klasą funkcji wypukłych.

W niniejszej pracy postaramy się skonstruować odpowiedni do postawionego problemu test statystyczny. Zaproponowane zostanie rozwiązanie oparte o iloraz wiarygodności w przypadku modelu regresji z ograniczeniami w postaci nierówności.

W pierwszym rozdziale zostaną omówione podstawowe własności wielościennych stożków wypukłych traktowanych jako podzbiór przestrzeni liniowej. Drugi rozdział będzie traktował o konstrukcji estymatora regresji wypukłej jako rzutu wektora danych na taki stożek wielościenny. W trzecim rozdziale zostanie wyznaczony rozkład poszukiwanej statystyki testowej w przypadku ze znaną wariancją błędu obserwacji.

Praca została napisana głównie na podstawie [3], natomiast rozdział o wyznaczaniu estymatora regresji wypukłej został napisany w dużym stopniu na podstawie [2].

1 Stożki wypukłe

Do konstrukcji testu zostanie wykorzystana metoda rzutowania wektora danych na stożek wielościenny powstały w wyniku narzuconych ograniczeń liniowych. W tym rozdziale zostaną przedstwione podstawowe definicje i własności wypuklych stożków wielościennych użyteczne w dalszych rozważaniach.

Definicja 1. *Ortantem w n-wymiarowej przestrzeni $\mathbb{R}^n$ nazywamy podzbiór powstały przez ograniczenie każdej ze współrzędnych do bycia nieujemną albo niedodatnią, czyli*

$$O = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \colon \epsilon_i x_i \geqslant 0, |\epsilon_i| = 1, i = 1, 2, \ldots, n\}.$$

Definicja 2 (Nieujemny ortant). *Nieujemnym ortantem nazywamy ortant, którego wszystkie współrzędne są nieujemne.*

Definicja 3 (Półprzestrzeń). *Półprzestrzenią H przestrzeni wektorowej $\mathbb{R}^n$ nazywamy zbiór*

$$H = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \colon a_1 x_1 + a_2 x_2 + \ldots a_n x_n \geqslant 0\},$$

gdzie $a_1, a_2, \ldots, a_n$ są pewnymi, ustalonymi liczbami rzeczywistymi.

Definicja 4 (Wielościenny stożek wypukły). *Wielościennym stożkiem wypukłym w przestrzeni wektorowej $\mathbb{R}^n$ nazywamy przecięcie skończonej ilości półprzestrzeni przestrzeni $\mathbb{R}^n$.*

Z powyższych definicji wynika, że dowolny stożek wypukły K w przestrzeni wektorowej $\mathbb{R}^n$ można zapisać jako

$$K = \bigcap_{i=1}^{m} H_i,$$

gdzie

$$H_j = \{\mathbf{x} \in \mathbb{R}^n \colon \sum_{i=1}^{n} a_i^j x_i \geqslant 0\}.$$

Zatem stożek K możemy przedstawić w postaci

$$K = \left\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \colon \sum_{i=1}^{n} a_i^1 x_i \geqslant 0, \ldots, \sum_{i=1}^{n} a_i^m x_i \geqslant 0\right\},$$

co będziemy zapisywać skrótowo jako

$$K = \{\mathbf{x} \in \mathbb{R}^n \colon \mathbf{A}\mathbf{x} \geqslant \mathbf{0}\},$$

gdzie

$$\mathbf{A} = \begin{bmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & \dots & a_m^2 \\ \vdots & \vdots & \dots & \vdots \\ a_1^n & a_2^n & \dots & a_m^n \end{bmatrix}, \quad \mathbf{x} = (x_1, x_2, \dots, x_n)^T.$$

Symbolem $\langle \cdot, \cdot \rangle$ będziemy oznaczać iloczyn skalarny w przestrzeni wektorowej V. Oznaczmy ponadto przez $\boldsymbol{\gamma}_i^T$ kolejne wiersze macierzy $-\mathbf{A}$. Bez straty ogólności możemy założyć, że tworzą one układ wektorów liniowo niezależnych, gdyż w przeciwnym wypadku któreś ograniczenie stanowiłoby kombinację pozostałych stąd dostajemy, że $m \leqslant n$. Wtedy stożek K możemy też zapisać w sposób

$$K = \{\mathbf{x} \in \mathbb{R}^n \colon \langle \mathbf{x}, \boldsymbol{\gamma}_i \rangle \leqslant 0, i = 1, 2, \dots, m\}.$$

Ponadto zbiór wektorów $\{\boldsymbol{\gamma}_i\}_{i=1}^m$ możemy uzupełnić do bazy przestrzni $\mathbb{R}^n$ o wektory ortogonalne do wektorów z tego zbioru oraz zdefiniować bazę dualną złożoną z wektorów $\boldsymbol{\beta}_i$ w następujący sposób

$$\boldsymbol{\beta}_i^T \boldsymbol{\gamma}_j = \begin{cases} -1, & i = j \\ 0, & i \neq j \end{cases}. \tag{1.1}$$

Wówczas możemy zapisać równoważne przedstawienie stożka K

$$K = \{\mathbf{x} \in \mathbb{R}^n \colon \mathbf{x} = \sum_{i=1}^m b_i \boldsymbol{\beta}_i + \sum_{i=m+1}^n c_i \boldsymbol{\beta}_i, b_i \geqslant 0, c_i \in \mathbb{R}\}.$$

Twierdzenie 1. *Niech $\{\boldsymbol{\gamma}_i\}_{i=1}^n, \{\boldsymbol{\beta}_i\}_{i=1}^n$ będą bazami przestrzeni $\mathbb{R}^n$ takimi, że $\boldsymbol{\beta}_i^T \boldsymbol{\gamma}_i = -1$ oraz $\boldsymbol{\beta}_i^T \boldsymbol{\gamma}_j = 0$ dla $i \neq j$. Wtedy przedstawienia*

$$K = \{\boldsymbol{x} \in \mathbb{R}^n \colon \langle \boldsymbol{x}, \boldsymbol{\gamma}_i \rangle \leqslant 0, \quad i = 1, 2, \dots, m\},$$

$$K = \{\boldsymbol{x} \in \mathbb{R}^n \colon \boldsymbol{x} = \sum_{i=1}^m b_i \boldsymbol{\beta}_i + \sum_{i=m+1}^n c_i \boldsymbol{\beta}_i, \quad b_i \geqslant 0, c_i \in \mathbb{R}\}$$

są równoważne.

Dowód. Wektory $\boldsymbol{\beta}_i, \boldsymbol{\gamma}_i, i = 1, 2, \dots, n$ spełniają zależność $\boldsymbol{\beta}_i^T \boldsymbol{\gamma}_i = -1$ oraz $\boldsymbol{\beta}_i^T \boldsymbol{\gamma}_j = 0$, $i \neq j$. Oznaczając rzez $\mathbf{B}$, $\mathbf{C}$ macierze, których kolumnami są odpowiednio wektory β_i, γ_i, związek ten możemy przedstawić jako $\mathbf{B}^T \mathbf{C} = -\mathbf{I}$. Macierz $\boldsymbol{B}$ jest nieosobliwa i jest macierzą przejścia z bazy kanonicznej do bazy $\{\boldsymbol{\beta}_i\}_{i=1}^m$. Stąd dostajemy, że $\mathbf{C}^T \mathbf{x} = -\mathbf{B}^{-1} \mathbf{x}$. Wyrażenia $\langle \mathbf{x}, \boldsymbol{\gamma}_i \rangle, i = 1, 2, \dots, m$ są pierwszymi m współrzędnymi $\mathbf{C}^T \mathbf{x}$. Zatem wektor $\mathbf{x}$ wyrażony w bazie złożonej z wektorów $\boldsymbol{\beta}_i$ ma pierwsze m współrzędnych nieujemnych wtedy i tylko wtedy, gdy $\langle \mathbf{x}, \boldsymbol{\gamma}_i \rangle \leqslant 0, i = 1, 2, \dots, m$, co dowodzi równoważności przedstawień. $\square$

2 Regresja wypukła

Podobnie jak w przypadku estymatora regresji liniowej wyznaczonego metodą najmniejszych kwadratów, który jest rzutem wektora danych na pewną podprzestrzeń liniową, tak w przypadku estymatora regresji wypukłej, który wyznaczymy tą samą metodą, będzie on rzutem na pewien wielościenny stożek wypukły powstały w wyniku stosownych ograniczeń.

Zbiór po którym będziemy minimalizować kwadrat błędu powstaje w sposób następujący. Przypuśmy, że wartości x_i $i = 1, 2, \dots, n$ są różne między sobą i uporządkowane rosnąco oraz niech $\theta_i = f(x_i), \quad i = 1, 2, \dots, n$. Wymóg wypukłości funkcji f może zostać zapisany jako

$$\frac{\theta_{i+1} - \theta_i}{x_{i+1} - x_i} \leqslant \frac{\theta_{i+2} - \theta_{i+1}}{x_{i+2} - x_{i+1}}, \quad i = 1, 2, \dots, n-2,$$

czyli by spadki wartości funkcji f między kolejnymi punktami x_i były niemalejące. Warunek ten można przekształcić do postaci

$$\theta_i(x_{i+2} - x_{i+1}) - \theta_{i+1}(x_{i+2} - x_i) + \theta_{i+2}(x_{i+1} - x_i) \geqslant 0, i = 1, \dots, n-2.$$

Zgodnie z rozważaniami przeprowadzonymi w poprzednim paragrafie możemy zbiór tych ograniczeń zapisać jako

$$K = \{\mathbf{A}\theta \geqslant 0\},$$

gdzie $\boldsymbol{A}$ jest rzeczywistą macierzą wymiaru $(n-2) \times n$, której kolejne wiersze stanowią współczynniki przy niewiadomych $\theta_1, \dots, \theta_n$ wzięte z kolejnych ograniczeń. Oznaczając przez $\Delta_i = x_{i+1} - x_i$ macierz $\boldsymbol{A}$ przyjmuje następującą postać

$$\begin{bmatrix} \Delta_2 & -\Delta_2 - \Delta_1 & \Delta_1 & 0 & & 0\dots & 0 \\ 0 & \Delta_3 & -\Delta_3 - \Delta_2 & \Delta_2 & & 0\dots & 0 \\ & \vdots & \ddots & & & & \vdots \\ 0 & \dots & \Delta_{i+1} & -\Delta_{i+1} - \Delta_i & \Delta_i & \dots & 0 \\ & \vdots & & \ddots & & & \vdots \\ 0 & 0 & 0 & \dots & \Delta_{n-1} & -\Delta_{n-1} - \Delta_{n-2} & \Delta_{n-2} \end{bmatrix}.$$

Zatem problem znalezienia estymatora regresji wypukłej sprowadza się do znalezienia $\hat{\boldsymbol{\theta}} \in K$ takiego, że

$$\min_{\boldsymbol{\theta} \in K} \|\boldsymbol{y} - \boldsymbol{\theta}\| = \|\boldsymbol{y} - \hat{\boldsymbol{\theta}}\|,$$

gdzie $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}}$.

Niech $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ będzie bazę kanoniczną przestrzeni $\mathbb{R}^n$. Wówczas $\boldsymbol{\gamma}_i = -\boldsymbol{A}^T\boldsymbol{e}_i$, $i = 1, 2, \dots, n-2$. Wtedy zbiór K możemy zapisać jako $K = \{\boldsymbol{\theta} \in \mathbb{R}^n \colon -\boldsymbol{e}_i^T\boldsymbol{A}\boldsymbol{\theta} \leqslant 0, i = 1, 2, \dots, n-2\} = \{\boldsymbol{\theta} \in \mathbb{R}^k \colon \langle \boldsymbol{\gamma}_i, \boldsymbol{\theta} \rangle \leqslant 0, i = 1, 2, \dots, n-2\}$.

Z określenia macierzy $\mathbf{A}$ oraz wektorów $\boldsymbol{\gamma}_i, i = 1, 2, \ldots, n-2$, widać, że tworzą one układ wektorów liniowo niezależnych. Zatem zbiór $B'_\gamma = \{\boldsymbol{\gamma}_i, i = 1, 2, \ldots, n-2\}$ można uzupełnić do bazy B_γ przestrzeni $\mathbb{R}^n$ o wektory $\boldsymbol{\gamma}_{n-1}, \boldsymbol{\gamma}_n$ tak, żeby były one ortogonalne do wszytkich wektorów z bazy B'_γ. Sprawdzimy, że warunek ten spełniają wektory $\boldsymbol{\gamma}_{n-1} = \mathbf{1}$ oraz $\boldsymbol{\gamma}_n = (\boldsymbol{x} - \bar{x}\mathbf{1})$, gdzie $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, $\bar{x}$ oznacza wartość średnią, $\mathbf{1} = (1, 1, \ldots, 1)^T$, a ortogonalność rozumiana jest w sensie iloczynu skalarnego powiązanego ze zdefiniowaną wcześniej normą euklidesową.

Lemat 1. *Wektory $\boldsymbol{\gamma}_{n-1}, \boldsymbol{\gamma}_n$ zdefiniowane jak powyżej są ortogonalne do wektorów ze zbioru B'_γ.*

Dowód. Dowolny wektor ze zbioru B'_γ możemy zapisać jako $\boldsymbol{\gamma}_i = (0, \ldots, 0, \overset{(i)}{x_{i+1} - x_{i+2}}, \overset{(i+1)}{x_{i+2} - x_i}, \overset{(i+2)}{x_i - x_{i+1}}, 0, \ldots, 0)$. Stąd $\langle \mathbf{1}, \gamma_i \rangle = x_{i+1} - x_{i+2} + x_{i+2} - x_i + x_i - x_{i+1} = 0$ oraz $\langle \boldsymbol{x} - \bar{x}\mathbf{1}, \gamma_i \rangle = \langle \boldsymbol{x}, \gamma_i \rangle - \bar{x}\langle \mathbf{1}, \gamma_i \rangle = x_i(x_{i+1} - x_{i+2}) + x_{i+1}(x_{i+2} - x_i) + x_{i+2}(x_i - x_{i+1}) = 0$. □

Teraz możemy zdefiniować bazę $B_\beta = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_n\}$ dualną do bazy B_γ w zaproponaowany w poprzednim paragrafie sposób (por. (1.1))

$$\boldsymbol{\beta}_i^T \boldsymbol{\gamma}_j = \begin{cases} -1, & i = j \\ 0, & i \neq j \end{cases}.$$

Oznaczając przez $\mathbf{B}$ i $\mathbf{C}$ macierze, których kolumnami są odpowiednio wektory $\boldsymbol{\beta}_i$ i $\boldsymbol{\gamma}_i$ związek między nimi możemy wyrazić jako

$$\mathbf{B}^T \mathbf{C} = -\mathbf{I},$$

gdzie $\mathbf{I}$ oznacza macierz jednostkową.

Niech E oznacza podprzestrzeń przestrzeni $\mathbb{R}^n$ rozpiętą przez wektory $\boldsymbol{\beta}_{n-1}, \boldsymbol{\beta}_n$, natomiast $\mathcal{L}(K)$ oznacza przestrzeń rozpiętą przez wektory $\boldsymbol{\beta}_i, i = 1, 2, \ldots, n-2$. Przestrzenie E oraz $\mathcal{L}(K)$ są do siebie ortogonalne, zatem wektor obserwacji $\mathbf{y}$ możemy zapisać jako sumę $\mathbf{y}_E + \mathbf{z}$, gdzie $\mathbf{y}_E$ i $\mathbf{z}$ są rzutami wektora $\mathbf{y}$ odpowiednio na podprzestrzeń E oraz $\mathcal{L}(K)$.

Przykład 1. *Prześledźmy powyższe rozważania na przykładzie. Rozważmy zbiór danych $\{x_1, x_2, x_3, x_4\}$ takich, że $x_{i+1} - x_i = 1, \ i = 1, 2, 3$.*

Macierz ograniczeń $\boldsymbol{A}$ przybiera wtedy postać

$$\boldsymbol{A} = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix}.$$

Zatem stożek powstały z ograniczeń jest postaci

$$K = \{\boldsymbol{\theta} \in \mathbb{R}^4 \colon \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix} \boldsymbol{\theta} \geqslant 0\}.$$

Baza wektorów $B_{\boldsymbol{\gamma}}$ jest postaci

$$B_{\boldsymbol{\gamma}} = \left((-1,2,-1,0)^T,(0,-1,2,-1)^T,(1,1,1,1)^T,\left(-\frac{3}{2},-\frac{1}{2},\frac{1}{2},\frac{3}{2}\right)^T\right).$$

Wektory $\boldsymbol{\beta}_i$ spełniające warunek $\langle \boldsymbol{\beta}_i, \quad \boldsymbol{\gamma}_i\rangle = -1$ oraz $\langle \boldsymbol{\beta}_i, \boldsymbol{\gamma}_j\rangle = 0, i \neq j$ mają następująca postać

$$B_{\boldsymbol{\beta}} = ((3,-4,-1,2)^T,(2,-1,-4,3)^T,(-1,-1,-1,-1)^T,(3,1,-1,-3)^T).$$

Przestrzenie na które będziemy rzutować wektor obserwacji przybierają postać

$$E = \{t_1(-1,-1,-1,1) + t_2(3,1,-1,-3), t_1,t_2 \in \mathbb{R}\},$$

$$\mathcal{L}(K) = \{t_1(3,-4,-1,2) + t_2(2,-1,-4,3), t_1,t_2 \in \mathbb{R}\}. \quad \diamondsuit$$

Zadanie znalezienia rzutu wektora obserwacji na stożek K sprowadza się do znalezienia rzutu jego składowych na stożek K. Wszytkie elementy podprzestrzeni E należą do stożka K, więc rzut wektora $\mathbf{y}_{\mathbf{E}}$ na stożek K jest tym samym co jego rzut na podprzestrzeń E. Przestrzeń E jest rozpinana przez wektory $\boldsymbol{\beta}_{n-1}$ i $\boldsymbol{\beta}_n$. Niech $\mathcal{E}$ oznacza macierz wymiaru $n \times 2$ taką, że jej kolumnami są wektory rozpinające podprzestrzeń E. Wtedy rzut wektora $\boldsymbol{y}$ na tą podprzestrzeń wyraża się wzorem

$$\mathbf{y}_E = \mathcal{E}(\mathcal{E}^T\mathcal{E})^{-1}\mathcal{E}^T\mathbf{y}. \tag{2.1}$$

Pozostaje zagadnienie znalezienia rzutu $\mathbf{z}$ na stożek K. Sprowadza się ono do znalezienia rzutu $\mathbf{z}$ na stożek

$$K' = K \cap \mathcal{L}(K) = \{\boldsymbol{\theta} \in \mathbb{R}^n \colon \boldsymbol{\theta} = \sum_{i=1}^{n-2} b_i\boldsymbol{\beta}_i, b_i \geqslant 0\}.$$

W [2] zostało pokazane, że przestrzeń $\mathcal{L}(K)$ może zostać podzielona na 2^{n-2} obszarów w taki sposób, że każdy z nich może być opisany jako nieujemny ortant w bazie $B_J = \{\boldsymbol{\beta}_i, i \in J, \boldsymbol{\gamma}_i, i \in L \setminus J\}$, gdzie J jest pewnym podzbiorem zbioru $L = \{1,2,\ldots,n-2\}$. Zatem każdy element $\boldsymbol{z}$ należący do $\mathcal{L}(K)$ może być przedstawiony w następujący sposób

$$\boldsymbol{z} = \sum_{i \in J} b_i\boldsymbol{\beta}_i + \sum_{i \in L\setminus J} c_i\boldsymbol{\gamma}_i, \quad b_i > 0, c_i \geqslant 0.$$

Dla dowolnego zbioru $J \subset L$, B_J jest bazą przestrzeni $\mathcal{L}(K)$, ponadto $\boldsymbol{\beta}_i, i \in J$ oraz $\boldsymbol{\gamma}_i, i \in L \setminus J$ są wzajemnie ortogonalne, zatem rzutem $\boldsymbol{z}$ na K' jest wektor postaci

$$\boldsymbol{z}_{K'} = \sum_{i \in J} b_i\boldsymbol{\beta}_i, b_i > 0.$$

Podsumowując, dowolny wektor $\mathbf{y}$ z przestrzeni $\mathbb{R}^n$ można przedstawić w następującej postaci

$$\mathbf{y} = \mathbf{z} + \mathbf{y}_E = \sum_{i\in J} b_i\boldsymbol{\beta}_i + \sum_{i\in L\setminus J} c_i\boldsymbol{\gamma}_i + d_1\boldsymbol{\gamma}_{n-1} + d_2\boldsymbol{\gamma}_n,\ \ b_i > 0, c_i \geqslant 0, d_1, d_2 \in \mathbb{R},$$

dla pewnego zbioru $J \subset \{1, 2, \ldots, n-2\}$. W [2] zostało pokazane, że przedstawienie to jest jednoznaczne.

Wtedy rzut tego wektora na stożek $K = \{\mathbf{A}\boldsymbol{\theta} \geqslant 0\}$ jest postaci

$$\hat{\boldsymbol{\theta}} = \sum_{i\in J} b_i\boldsymbol{\beta}_i + d_1\boldsymbol{\gamma}_{n-1} + d_2\boldsymbol{\gamma}_n.$$

Natomiast wektor reszt $\hat{\boldsymbol{\rho}} = \mathbf{y} - \hat{\boldsymbol{\theta}}$ jest postaci

$$\hat{\boldsymbol{\rho}} = \sum_{i\in L\setminus J} c_i\boldsymbol{\gamma}_{\boldsymbol{i}}.$$

Przykład 2. *Rozważmy stożek z przykładu 1 i wektor obserwacji* $\boldsymbol{y}^T = (0; 3.1; 5.2; 6.8)$. *Na początek wyliczymy macierz rzutu oznaczoną w (2.1) jako* $\mathcal{E}(\mathcal{E}^T\mathcal{E})^{-1}\mathcal{E}^T$.

$$\mathcal{E}(\mathcal{E}^T\mathcal{E})^{-1}\mathcal{E}^T =$$

$$= \begin{bmatrix} -1 & 3 \\ -1 & 1 \\ -1 & -1 \\ -1 & -1 \end{bmatrix} \left(\begin{bmatrix} -1 & -1 & -1 & -1 \\ 3 & 1 & -1 & -3 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ -1 & 1 \\ -1 & -1 \\ -1 & -3 \end{bmatrix} \right)^{-1} \begin{bmatrix} -1 & -1 & -1 & -1 \\ 3 & 1 & -1 & -3 \end{bmatrix} =$$

$$= \frac{1}{80} \begin{bmatrix} -1 & 3 \\ -1 & 1 \\ -1 & -1 \\ -1 & -3 \end{bmatrix} \begin{bmatrix} 20 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} -1 & -1 & -1 & -1 \\ 3 & 1 & -1 & -3 \end{bmatrix} =$$

$$= \frac{1}{80} \begin{bmatrix} 56 & 32 & 8 & -16 \\ 32 & 24 & 16 & 8 \\ 8 & 16 & 24 & 32 \\ -16 & 8 & 32 & 56 \end{bmatrix}.$$

Następnie wyznaczymy rzut $\boldsymbol{y}_E$.

$$\boldsymbol{y}_E = \mathcal{E}(\mathcal{E}^T\mathcal{E})^{-1}\mathcal{E}^T\boldsymbol{y} =$$

$$= \frac{1}{80} \begin{bmatrix} 56 & 32 & 8 & -16 \\ 32 & 24 & 16 & 8 \\ 8 & 16 & 24 & 32 \\ -16 & 8 & 32 & 56 \end{bmatrix} \begin{bmatrix} 0 \\ 3.1 \\ 5.2 \\ 6.8 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 2.65 \\ 4.9 \\ 7.15 \end{bmatrix}.$$

Wektor $\boldsymbol{z}$, który będziemy rzutować jest postaci $\boldsymbol{z}^T = (-0.4; 0.45; 0.3; -0.35)$. Baza mieszna w której wszystkie współrzędne tego wektora są dodatnie to $B = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \boldsymbol{\beta}_4)$, a poszczególne współrzędne wynoszą w przybliżeniu odpowiednio $0.4; 0.35; 0$ i $6.94 \cdot 10^{-18}$. Zatem rzut wektora $\boldsymbol{z}$ na stożek K jest równy

$$\boldsymbol{z}_K^T = 6.94 \cdot 10^{-18} \cdot (3; 1; -1; -3) = 10^{-18} \cdot (20.82; 6.94; -6.94; -20.82).$$

Zatem poszukiwany rzut wektora $\boldsymbol{y}$ na stożek K jest postaci

$$\hat{\boldsymbol{\theta}}^T = \boldsymbol{y}_E^T + \boldsymbol{z}^T \approx \boldsymbol{y}_E^T = (0.4; 2.65; 4.9; 7.15).\diamondsuit$$

3 Statystyka testowa i jej rozkład

3.1 Lematy i oznaczenia

Na początek wprowadzimy kilka oznaczeń i udowodnimy dwa lematy z których skorzystamy w dalszej części rozważań. Zatem

$$C_{L\setminus J} = \{\boldsymbol{y} \in \mathbb{R}^n \colon y = \tag{3.1}$$

$$= \sum_{i \in L\setminus J} b_i \boldsymbol{\gamma}_i + \sum_{i \in J} c_i \boldsymbol{\beta}_i + d_1 \boldsymbol{\gamma}_{n-1} + d_2 \boldsymbol{\gamma}_n, b_i > 0, c_i \geqslant 0, d_1, d_2 \in \mathbb{R}\}$$

$$\textbf{??} S_{L\setminus J} = \operatorname{span}\{\boldsymbol{\gamma}_i, i \in L \setminus J\} \tag{3.2}$$

$$d = |L \setminus J| = n - 2 - |J| \tag{3.3}$$

Ponadto niech

$$\boldsymbol{A}_{L\setminus J} \tag{3.4}$$

oznacza macierz wymiaru $n \times (n-2)$ taką, że pierwsze d kolumn to wektory $\boldsymbol{\gamma}_i, i \in L \setminus J$ natomiast pozostałe $n-2-d$ kolumn to wektory $\boldsymbol{\beta}_i, i \in J$.

Niech wektor obserwacji $\boldsymbol{y}$ ma n-wymiarowy rozkład normalny z nieznaną wartością oczekiwaną $\boldsymbol{f} \in \Theta$, gdzie Θ jest przestrzenią parametrów rozkładu, i znaną wariancją σ^2. Symbolem $\mathcal{L}(\boldsymbol{f}, \boldsymbol{y})$ oznaczać będziemy funkcję wiarygodności

$$\mathcal{L}(\boldsymbol{f}, \boldsymbol{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}||\boldsymbol{y} - \boldsymbol{f}||^2\right\}. \tag{3.5}$$

Lemat 2. *Niech $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)^T \sim N(\boldsymbol{0}, \boldsymbol{I})$ oraz niech $\hat{\boldsymbol{Z}}$ będzie rzutem $\boldsymbol{Z}$ na przestrzeń liniową S wymiaru $d < n$. Ponadto niech $\boldsymbol{A}$ będzie rzeczywistą macierzą wymiaru $m \times n$ taką, że każdy jej wiersz jest ortogonalny do przestrzeni S oraz $m \leqslant n$ i rank$\boldsymbol{A} = m$. Wtedy rozkładem warunkowym $\|\hat{\boldsymbol{Z}}\|^2$ pod warunkiem $\boldsymbol{AZ} \geqslant 0$ jest rozkład χ^2 o d stopniach swobody.*

Dowód. Niech $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ będą wzjamnie ortonormalnymi wektorami w $\mathbb{R}^n$ takimi, że wektory $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d$ rozpinają przestrzeń S. Niech $\boldsymbol{V}$ oznacza macierz taką, której poszczególne kolumny są kolejno wektorami $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ oraz niech $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^T$. Wektor $\boldsymbol{Z}$ możemy zapisać jako $\boldsymbol{Z} = \sum_{i=1}^n a_i \boldsymbol{v}_i$, gdzie $a_i = \langle \boldsymbol{v}_i, \boldsymbol{Z} \rangle$. Stąd $a_i, i = 1, 2, \ldots, n$ są niezależnymi zmiennymi losowymi o standardowym rozkładzie normalnym, bo dla dowolnego i możemy napisać, że $a_i = v_i^T \boldsymbol{Z}$, czyli $\boldsymbol{a} = \boldsymbol{V}^T \boldsymbol{Z}$. Stąd $\boldsymbol{a} \sim N(\boldsymbol{0}, \boldsymbol{V}^T \boldsymbol{V})$, gdzie $\boldsymbol{V}^T \boldsymbol{V} = \boldsymbol{I}$, bo wektory $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ są wzajemnie ortonormalne. Ponadto $\hat{\boldsymbol{Z}} = \sum_{i=1}^d a_i \boldsymbol{v}_i$. Zatem $\|\hat{\boldsymbol{Z}}\|^2 = a_1^2 + a_2^2 + \cdots + a_d^2$ i dlatego zmienna losowa $\|\hat{\boldsymbol{Z}}\|^2$ ma rozkład χ^2 o d stopniach swobody. Macierz $\boldsymbol{V}$ możemy zapisać jako $\boldsymbol{V} = [\boldsymbol{V}_1 | \boldsymbol{V}_2]$, gdzie $\boldsymbol{V}_1$ jest macierzą wymiaru $n \times d$, oznaczmy też przez $\boldsymbol{a}^1$ wektor $(a_1, a_2, \ldots, a_d)^T$ a przez $\boldsymbol{a}^2$ wektor $(a_{d+1}, \ldots, a_n)^T$. Wtedy $\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{a} = \boldsymbol{V}_1 \boldsymbol{a}^1 + \boldsymbol{V}_2 \boldsymbol{a}^2$ a warunek $\boldsymbol{A}\boldsymbol{Z} \geqslant 0$ możemy zapisać jako $\boldsymbol{A}\boldsymbol{V}_1 \boldsymbol{a}^1 + \boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{a}^2 \geqslant 0$. Zauważmy, że z założenia o ortogonalności wierszy macierzy $\boldsymbol{A}$ do przestrzeni S oraz konstrukcji macierzy $\boldsymbol{V}$ dostajemy, że $\boldsymbol{A}\boldsymbol{V}_1 = \boldsymbol{0}$ oraz $\boldsymbol{a}^1, \boldsymbol{a}^2$ są niezależne. Zatem $P(\|\hat{\boldsymbol{Z}}\|^2 \leqslant a | \boldsymbol{A}\boldsymbol{Z} \geqslant 0) = P(\|\boldsymbol{a}^1\|^2 \leqslant a | \boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{a}^2 \geqslant 0)$. Pokażemy teraz, że zbiór $\{\omega \in \Omega \colon \boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{a}^2(\omega) \geqslant 0\}$ jest niezerowej miary. Zmienna losowa $\boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{a}^2$ ma m-wymiarowy rozkład normalny $N(\boldsymbol{0}, \boldsymbol{\Sigma})$, gdzie $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{V}_2^T \boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{A}^T$, bo wiersze macierzy $\boldsymbol{V}_2$ są wzajemnie ortonormalne. Ponadto z założeń dostajemy, że $\operatorname{rank} \boldsymbol{A}\boldsymbol{A}^T = m$. Zatem nośnik zmiennej $\boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{a}^2$ zawrty jest w m-wymiarowej przestrzeni i nie jest zawart w przestrzeni $(m-1)$-wymiarowej. Stąd i z symetrii rozkładu dostajemy, że $P(\boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{a}^2 \geqslant 0) = \frac{1}{2^{m-1}} \neq 0$. Zatem prawdopodobieństwo warunkowe $P(\|\hat{\boldsymbol{Z}}\|^2 \leqslant a | \boldsymbol{A}\boldsymbol{Z} \geqslant 0)$ istniej i ponadto

$$P(\|\boldsymbol{a}^1\|^2 \leqslant a | \boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{a}^2 \geqslant 0) = \frac{P(\|\boldsymbol{a}^1\|^2 \leqslant a \wedge \boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{a}^2 \geqslant 0)}{P(\boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{a}^2 \geqslant 0)} =$$

$$= \frac{P(\|\boldsymbol{a}^1\|^2 \leqslant a) P(\boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{a}^2 \geqslant 0)}{P(\boldsymbol{A}\boldsymbol{V}_2 \boldsymbol{a}^2 \geqslant 0)} = P(\|\boldsymbol{a}^1\|^2 \leqslant a) = P(\chi_d^2 \leqslant a)$$

co należało dowieść. □

Idea powyższego lemat została zaczerpnięta z [3].

Lemat 3. *Niech $\boldsymbol{y} \in C_{L \setminus J}$ określonego wzorem (3.1) dla pewngo zbioru $J \subset L = \{1, 2, \ldots, n-2\}$ oraz niech $a, b \in \mathbb{R}$. Wtedy wektor $\boldsymbol{y}' = \boldsymbol{y} + a\boldsymbol{\gamma}_{n-1} + b\boldsymbol{\gamma}_n$ należy do zbioru $C_{L \setminus J}$ oraz wektory reszt $\boldsymbol{\rho} = \boldsymbol{y} - \hat{\boldsymbol{\theta}}$ i $\boldsymbol{\rho}' = \boldsymbol{y}' - \hat{\boldsymbol{\theta}}'$ są sobie równe.*

Dowód. Jeśli $\boldsymbol{y} \in C_{L \setminus J}$ to $\boldsymbol{y}$ możemy zapisać jako $\boldsymbol{y} = \sum_{i \in J} c_i \boldsymbol{\beta}_i + \sum_{i \in L \setminus J} b_i \boldsymbol{\gamma}_i + d_1 \boldsymbol{\gamma}_{n-1} + d_2 \boldsymbol{\gamma}_n,\ b_i > 0, c_i \geqslant 0, d_1, d_2 \in \mathbb{R}$. Wtedy $\boldsymbol{y}' = \sum_{i \in J} c_i \boldsymbol{\beta}_i + \sum_{i \in L \setminus J} b_i \boldsymbol{\gamma}_i + (d_1 + a)\boldsymbol{\gamma}_{n-1} + (d_2 + b)\boldsymbol{\gamma}_n,\ b_i > 0, c_i \geqslant 0, d_1, d_2 \in \mathbb{R}$. Oczywiście $d_1 + a, d_2 + b \in \mathbb{R}$ zatem $\boldsymbol{y}' \in C_{L \setminus J}$.

Wektor $\boldsymbol{\rho}$ jest postaci $\boldsymbol{\rho} = \sum_{i \in L \setminus J} b_i \boldsymbol{\gamma}_i$. Z postaci wktora $\boldsymbol{y}'$ widzimy jednak, że $\boldsymbol{\rho}' = \sum_{i \in L \setminus J} b_i \boldsymbol{\gamma}_i = \boldsymbol{\rho}$. □

Na koniec zostanie udowodnione twierdzenie nie mające bezpośredniego związku z postacią poszukiwanego testu, pokazujące jednak pewną ciekawą własność rzutów wektora obserwacji $\boldsymbol{y}$.

Twierdzenie 2. *Niech wektor* $\boldsymbol{y}$ *rozkład normalny* $N(\mathbf{0}, \boldsymbol{\Sigma})$. *Wtedy rzuty tego wektora na przestrzenie* $span\{\boldsymbol{\gamma}_i, i \in L \setminus J\}$, $span\{\boldsymbol{\beta}_i, i \in J\}$ *oraz* $span\{\boldsymbol{\gamma}_{n-1}, \boldsymbol{\gamma}_n\}$, *czyli wektory losowe* $\boldsymbol{y} - \hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{y}}$ *i* $\hat{\boldsymbol{y}}$ *dla ustalonego zbioru* J *są stochastycznie niezależne.*

Dowód. Dla ustalonego stożka $C_{L\setminus J}$ mamy $\boldsymbol{y} - \hat{\boldsymbol{\theta}} = \sum_{i \in L\setminus J} b_i \boldsymbol{\gamma}_i$ oraz $\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{y}} = \sum_{i \in J} c_i \boldsymbol{\beta}_i$. Niech $S_1 = \text{span}\{\boldsymbol{\beta}_i, i \in J\}, S_2 = \text{span}\{\boldsymbol{\gamma}_i, i \in L\setminus J\}, S_3 = \text{span}\{\boldsymbol{\gamma}_{n-1}, \boldsymbol{\gamma}_n\}$. Bazę każdej z tych przestrzeni można zortonormalizować tak, by $S_1 = \text{span}\{\boldsymbol{v}_i, i \in J\}, S_2 = \text{span}\{\boldsymbol{v}_i, i \in L \setminus J\}, S_3 = \text{span}\{\boldsymbol{v}_{n-1}, \boldsymbol{v}_n\}$, gdzie wektory $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ są wzajemnie ortonormalne. Niech V oznacza macierz, której kolumnami są wektory $\boldsymbol{v}$. Wektor $\boldsymbol{y}$ możemy zapisać jako $\boldsymbol{y} = \sum_{i=1}^n \langle v_i, y \rangle = \sum_{i=1}^n a_i v_i$. Analogicznie jak w dowodzie lematu 2 możemy zdefiniować wektor $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^T$, który ma n-wymiarowy rozkład normalny $N(0, \boldsymbol{I})$. Niech $\boldsymbol{a}^1 = (a_1, \ldots, a_{n-d-2})^T$, $\boldsymbol{a}^2 = (a_{n-d-1}, \ldots, a_{n-2})^T$, $\boldsymbol{a}^3 = (a_{n-1}, a_n)^T$ oraz zapiszmy macierz $\boldsymbol{V}$ jako $[\boldsymbol{V}_1 | \boldsymbol{V}_2 | \boldsymbol{V}_3]$. Wówczas $\boldsymbol{y} = \boldsymbol{V}_1 \boldsymbol{a}^1 + \boldsymbol{V}_2 \boldsymbol{a}^2 + \boldsymbol{V}_3 \boldsymbol{a}^3$ oraz $\boldsymbol{y} - \hat{\boldsymbol{\theta}} = \boldsymbol{V}_2 \boldsymbol{a}^2, \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{y}} = \boldsymbol{V}_1 \boldsymbol{a}^1$ oraz $\hat{\boldsymbol{y}} = \boldsymbol{V}_3 \boldsymbol{a}^3$. Wektory $\boldsymbol{a}^1$, $\boldsymbol{a}^2$ i $\boldsymbol{a}^3$ są niezależne, zatem wektory $\boldsymbol{V}_1 \boldsymbol{a}^1$, $\boldsymbol{V}_2 \boldsymbol{a}^2$ oraz $\boldsymbol{V}_3 \boldsymbol{a}^3$ jako mierzalne funkcje tych wektorów również są niezależne dla ustalonego $C_{L\setminus J}$. □

3.2 Statystyka testowa

Lematy udowodnione w poprzednim paragrafie posłużą do wyznaczenia rozkładu statystyki testowej zaproponowanego testu opartego na ilorazie wiarygodności. Przypomnijmy, że problem testowania hipotez, którym się zajmujemy ma postać

$$H_0 \colon f(x) = ax + b \qquad vs. \qquad H_1 \colon f \in \mathcal{F}, \tag{3.6}$$

gdzie $\mathcal{F}$ jest klasą funkcji wypukłych.

Niech $\hat{\boldsymbol{y}}$ oznacza estymator regresji liniowej, czyli rzut wektora danych $\boldsymbol{y}$ na przestrzeń rozpinaną przez wektory $\boldsymbol{\gamma}_{n-1}, \boldsymbol{\gamma}_n$, $\hat{\boldsymbol{\theta}}$ oznacza estymator regresji wypukłej, będący rzutem wektora obserwacji na odpowiedni stożek wypukły. Ponadto oznaczmy przez $R_0 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2$ oraz $R_1 = \sum_{i=1}^n (y_i - \hat{\theta}_i)^2 = \|\boldsymbol{y} - \hat{\boldsymbol{\theta}}\|^2$.

Uwaga 1. *Powyższe estymatory* $\hat{\boldsymbol{y}}$ *i* $\hat{\boldsymbol{\theta}}$, *jako elementy minimalizujące kwadrat normy błędu, czyli elementy optymalne w sensie aproksymacji średniokwadratowej, wyznaczone zostały przy zastosowaniu metody najmniejszych kwadratów. Jednak przy założeniu normalności rozkładu wektora obserwacji estymatory wyznaczone tą metodą są tożsame z estymatorami wyznaczonymi przy pomocy metody największej wiarygodności. Rozważmy ogólny model* $\boldsymbol{x} = \phi(\boldsymbol{x}) + \boldsymbol{\varepsilon}$, *gdzie* $\phi(\boldsymbol{x})$ *jest* *poszukiwanym parametrem, a wektor błędów $\boldsymbol{\varepsilon}$ ma n-wymiarowy rozkład normalny $N(\mathbf{0}, \sigma^2 \boldsymbol{I})$ ze znaną wariancją. Stosując metodę namniejszych kwadratów dostajemy, że estymatorem $\phi(\boldsymbol{x})$ jest wektor $\hat{\phi(\boldsymbol{x})}$ minimalizujący wyrażenie $\sum_{i=1}^{n}(\boldsymbol{x}_i - \phi(\boldsymbol{x})_i)^2$, gdzie $\boldsymbol{x}_i, \phi(\boldsymbol{x})_i$ oznaczają kolejne współrzędne odpowiednio wektora $\boldsymbol{x}$ i $\phi(\boldsymbol{x})$. Stosując metodę największej wiarygodności przy założeniu normalności rozkładu wektora $\boldsymbol{x}$ będziemy maksymalizować funkcję wiarygodności postaci $L(\phi(\boldsymbol{x}), \boldsymbol{x}) = \Pi_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(\boldsymbol{x}_i - \phi(\boldsymbol{x})_i)^2\} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \phi(\boldsymbol{x})_i)^2\}$. Maksymalizacja tej funkcji jest równoważna maksymalizacji jej logarytmu $l(\phi(\boldsymbol{x}), \boldsymbol{x}) = \ln(2\pi\sigma^2)^{-\frac{n}{2}} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \phi(\boldsymbol{x})_i)^2$, co przy założeniu znajomości wariancji prowadzi do równoważnego zagadnienia minimalizacji wyrażenia $\sum_{i=1}^{n}(\boldsymbol{x}_i - \phi(\boldsymbol{x})_i)^2$, czyli tego samego, które otrzymaliśmy stosując metodę najmniejszych kwadratów. Zatem przy założeniu normalności rozkładu wektora obserwacji $\boldsymbol{x}$ estymatory uzyskiwane obiema metodami są sobie równe.*

W rozważanym problemie hipotezę zerową możemy utożsamić z pewną podprzestrzenią parametrów $\Theta_0 \subset \Theta$ a alternatywę z podprzestrzenią $\Theta_1 \subset \Theta$. Określmy funkcję

$$l(\boldsymbol{y}) = \frac{\sup_{\boldsymbol{f} \in \Theta} \mathcal{L}(\boldsymbol{f}, \boldsymbol{y})}{\sup_{\boldsymbol{f} \in \Theta_0} \mathcal{L}(\boldsymbol{f}, \boldsymbol{y})}$$

nazywaną ilorazem wiarygodności. Oczywiści dla dowlonego $\boldsymbol{y} \in \mathbb{R}^n$ mamy, że $l(\boldsymbol{y}) \in [1, +\infty)$, ponadto duże wartości $l(\boldsymbol{y})$ powinny sugerować, że hipoteza zerowa nie jest poprawna.

Definicja 5. *Testem hipotezy $H_0\colon \boldsymbol{f} \in \Theta_0$ przy alternatywie $H_1\colon \boldsymbol{f} \in \Theta_1$ opartym na ilorazie wiarygodności na poziomie istotności α nazywamy funkcję*

$$\phi(\boldsymbol{y}) = \begin{cases} 1 & gdy\ l(\boldsymbol{y}) > c \\ \xi & gdy\ l(\boldsymbol{y}) = c \\ 0 & gdy\ l(\boldsymbol{y}) < c \end{cases},$$

gdzie stałe c, ξ są tak dobrane by rozmiar testu nie przekraczał α.

Zatem zgdonie z definicją 5 możemy rozważyć zbiór odrzucenia C w naszym problemie (3.6) postaci

$$C = \phi^{-1}(\{1\}) = \left\{\boldsymbol{y} \in \mathbb{R}^n\colon \frac{\sup_{\boldsymbol{f} \in \Theta} \mathcal{L}(\boldsymbol{f}, \boldsymbol{y})}{\sup_{\boldsymbol{f} \in \Theta_0} \mathcal{L}(\boldsymbol{f}, \boldsymbol{y})} > c\right\} =^{1}$$

$$= \left\{\boldsymbol{y} \in \mathbb{R}^n\colon \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}||\boldsymbol{y} - \hat{\boldsymbol{\theta}}||^2\}}{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}||\boldsymbol{y} - \hat{\boldsymbol{y}}||^2\}} > c\right\} =$$

[1] przejście możliwe dzięki uwadze 1

$$= \left\{ \boldsymbol{y} \in \mathbb{R}^n \colon \exp\left\{ \frac{1}{2\sigma^2} \left(||\boldsymbol{y} - \hat{\boldsymbol{y}}||^2 - ||\boldsymbol{y} - \hat{\boldsymbol{\theta}}||^2 \right) \right\} > c \right\}$$

$$= \left\{ \boldsymbol{y} \in \mathbb{R}^n \colon \frac{||\boldsymbol{y} - \hat{\boldsymbol{y}}||^2 - ||\boldsymbol{y} - \hat{\boldsymbol{\theta}}||^2}{\sigma^2} > c' \right\} =$$

$$= \left\{ \boldsymbol{y} \in \mathbb{R}^n \colon \frac{R_0 - R_1}{\sigma^2} > c' \right\}.$$

Zatem będziemy poszukiwać przy założeniu prawdziwości hipotezy zerowej H_0 rozkładu statystyki testowej postaci

$$M = \frac{R_0 - R_1}{\sigma^2}.$$

W kolejnych rozważaniach przez D będziemy oznaczać zmienną losową reprezentującą liczność zbioru $L \setminus J$ (por. (3.3)).

Twierdzenie 3. *Przy założeniu prawdziwości hipotezy zerowej postawionego problemu (3.6) mamy*

$$P(M \leqslant a) = \sum_{d=0}^{n-2} P(\chi^2_{n-d-2} \leqslant a) P(D = d),$$

gdzie $\chi^2_0 \equiv 0$*, czyli rozkładem staystyki testowej* M *jest mieszany rozkład* χ^2 *z wagami* $P(D = d)$*.*

Dowód. Z lematu 3 możemy bez straty ogólności założyć, że $f(x) = 0$ skąd $\boldsymbol{y} = \boldsymbol{\varepsilon}$. Dla dowolnej realizacji wektora $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ oznaczmy przez $L \setminus J$ taki zbiór indeksów, że $\boldsymbol{\varepsilon} \in C_{L\setminus J}$ (por. (3.1)). Zatem wyrażenie $R_1 = \|\boldsymbol{y} - \hat{\boldsymbol{\theta}}\|^2$ zależy poprzez $\hat{\boldsymbol{\theta}}$ od wyboru zbioru J, który jest losowy, różne wartości wektora błędu $\boldsymbol{\varepsilon}$ mogą umieścić wektor danych $\boldsymbol{y}$ w różnych zbiorach $C_{L\setminus J}$. Zauważmy jednak, że z jednoznaczności przedstawienia wektora w zbiorze $C_{L\setminus J}$ zdarzenia postaci $\{\boldsymbol{y} \in C_{L\setminus J}\}$, $J \in \mathcal{P}(L)$, gdzie $\mathcal{P}(L)$ oznacza zbiór potęgowy zbioru L, są wzajmnie rozłączne oraz ich suma stanowi całą przestrzeń zdarzeń elementarnych. Ponadto prawdopodobieństwa tych zdarzeń są niezerowe. Możemy zatem skorzystać z twierdzenia o prawdopodobieństwie całkowitym.

$$P(M \leqslant a) = \sum_{J \in \mathcal{P}(L)} P(M \leqslant a | \boldsymbol{y} \in C_{L\setminus J}) P(\boldsymbol{y} \in C_{L\setminus J}).$$

Następnie rozważmy jakie są rozkłady zmiennych losowych $\frac{R_1}{\sigma^2}$ oraz $\frac{R_0}{\sigma^2}$ przy założeniu prawdziwości hipotezy zerowej i przy ustalonym zbiorze J.

Niech $\hat{\boldsymbol{\varepsilon}}$ będzie rzutem wektora $\boldsymbol{\varepsilon}$ na przestrzeń $S_{L\setminus J}$ (por. (**??**)). Zauważmy, że macierz $\boldsymbol{A}_{L\setminus J}$ (por. (3.4)) można zapisać jako $[\boldsymbol{A}_1 | \boldsymbol{A}_2]$, gdzie macierz $\boldsymbol{A}_1$ jest wymiaru $n \times d$. Zatem kolumny macierzy $\boldsymbol{A}_1$ rozpinają $S_{L\setminus J}$, natomiast kolumny macierzy $\boldsymbol{A}_2$ są ortogonalne do przestrzeni $S_{L\setminus J}$. Dodatkowo, gdy $\boldsymbol{\varepsilon} \in C_{L\setminus J}$, zachodzi $\boldsymbol{A}_1^T\boldsymbol{\varepsilon} \geqslant 0$ oraz $\boldsymbol{A}_2^T\boldsymbol{\varepsilon} \geqslant 0$. Stąd na mocy lematu 2 dostajemy, że rozkładem warunkowym $\frac{||\hat{\varepsilon}||^2}{\sigma^2}$ przy zadanym J jest rozkład χ^2 o d stopniach swobody, gdzie $d = |L \setminus J|$. Jako że $R_1 = ||\hat{\boldsymbol{\rho}}||^2$, gdzie $\hat{\boldsymbol{\rho}} = \boldsymbol{y} - \hat{\boldsymbol{\theta}}$, otrzymujemy, że jeśli hipoteza zerowa $\boldsymbol{\theta} \in \operatorname{span}\{\boldsymbol{\gamma}_{n-1}, \boldsymbol{\gamma}_n\}$ jest prawdziwa to rozkładem warunkowym $\frac{R_1}{\sigma^2}$ przy ustalonym zbiorze J jest χ^2_d, gdzie $d = |L \setminus J|$.

Zmienna losowa $\frac{R_0}{\sigma^2}$ jest rzutem wektora danych $\boldsymbol{y}$ na przestrzeń rozpinaną przez wektory $\boldsymbol{\beta}_i, \boldsymbol{\gamma}_j, i \in J, j \in L \setminus J$ a zatem ma rozkład χ^2 o $n-2$ stopniach swobody.

Rozważmy teraz wyrażenie $R_0 - R_1$. Skorzystamy tutaj z następujących własności normy: dla dowolnych wektorów $\boldsymbol{x}, \boldsymbol{y}$ i dowolnej liczby rzeczywistej a mamy, że $\|\boldsymbol{x}+\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ oraz $\|a\boldsymbol{x}\|^2 = a^2\|\boldsymbol{x}\|^2$.

Zatem dla ustalonego zbioru J dostajemy, że

$$R_0 - R_1 = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2 - \|\boldsymbol{y} - \hat{\boldsymbol{\theta}}\|^2 = \|\sum_{i\in L\setminus J} b_i\boldsymbol{\gamma}_i + \sum_{i\in J} c_i\boldsymbol{\beta}_i\|^2 - \|\sum_{i\in L\setminus J} b_i\boldsymbol{\gamma}_i\|^2 =$$

$$= \sum_{i\in L\setminus J} b_i^2\|\boldsymbol{\gamma}_i\|^2 + \sum_{i\in J} c_i^2\|\boldsymbol{\beta}_i\|^2 + 2\sum_{\substack{i\in L\setminus J\\ j\in J}} b_i c_j\langle\boldsymbol{\gamma}_i, \boldsymbol{\beta}_j\rangle + 2\sum_{\substack{i,j\in L\setminus J\\ i\neq j}} b_i b_j\langle\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_j\rangle +$$

$$+2\sum_{\substack{i,j\in J\\ i\neq j}} c_i c_j\langle\boldsymbol{\beta}_i, \boldsymbol{\beta}_j\rangle - \sum_{i\in J} b_i^2\|\boldsymbol{\gamma}_i\| - 2\sum_{\substack{i,j\in J\\ i\neq j}} b_i b_j\langle\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_j\rangle =$$

$$= \sum_{i\in J} c_i^2\|\boldsymbol{\beta}_i\|^2 + 2\sum_{\substack{i,j\in J\\ i\neq j}} c_i c_j\langle\boldsymbol{\beta}_i, \boldsymbol{\beta}_j\rangle = \|\sum_{i\in J} c_i\boldsymbol{\beta}_i\|^2 = \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{\theta}}\|^2,$$

bo $\langle\boldsymbol{\gamma}_i, \boldsymbol{\beta}_j\rangle = 0$ dla $i \neq j$. Otrzymaliśmy zatem rzut wektora $\boldsymbol{\varepsilon}$ na przestrzeń $\operatorname{span}\{\boldsymbol{\beta}_i, i \in J\}$ dla ustalonego zbioru J. Rzut ten może zostać przeskalowany o znaną wartość wariancji. Wtedy tak uzyskana zmianna losowa, na mocy lematu 2, ma rozkład χ^2 o $n-d-2$ stopniach swobody. Liczba stopni swobody w tym rozkładzie zależy jedynie od liczności zbioru $L \setminus J$, natomiast nie zależy bezpośrednio od dokłanej jego postaci. Niech zatem D będzie zdefiniowaną wcześniej zmienną losową reprezentującą liczność zbioru $L \setminus J$. Otrzymujemy wtedy, że rozkładem $\frac{R_0-R_1}{\sigma^2}$ pod warunkiem $D = d$ jest χ^2 o $n-d-2$ stopniach swobody, co należało pokazać. $\square$

Wartości prawdopodobieństw $P(D = d), d = 0, 1, \ldots, n-2$ są wyliczane na podstawie względnych objętości zbiorów $C_{L\setminus J}, J \in \mathcal{P}(L)$. Prawdopodobieństwo, że $\boldsymbol{y} \in C_{L\setminus J}$, gdy hipoteza zerowa jest prawdziwa, jest równoważne prawdopodobieństwu, że wektor losowy o n-wymiarowym standardowym rozkładzie normalny należy do zbioru $C_{L\setminus J}$. Wyznaczenie wartości $P(D = d)$ jest dla dużych n bardzo trudne i najczęściej stosuje się w tym celu pewne przybliżenia numeryczne.

4 Podsumowanie

W pracy została zaprezentowana konstrukcja testu statystycznego przeznaczonego do testowania problemu wyboru krzywej regresji następującej postaci

$$H_0\colon f(x) = ax + b \quad vs. \quad H_1\colon f \in \mathcal{F},$$

gdzie $\mathcal{F}$ jest klasą funkcji wypukłych.

W wyniku rzutowania wektora obserwacji na wielościenny stożek wypukły powstały jako rezultat wymogu wypukłości funkcji f otrzymano estymator regresji wypukłej, który pozwolił wyznaczyć rozkład statystyki testowej następującej postaci

$$P(M \leqslant a) = \sum_{d=0}^{n-2} P(\chi^2_{n-d-2} \leqslant a) P(D = d),$$

będący mieszanym rozkładem χ^2 z wagami $P(D = d)$. Prawdopodobieństwa $P(D = d)$ wyznaczane są na podstawie względnych objętości zbiorów $C_{L \setminus J}$, $J \in \mathcal{P}(L)$ w porównaniu do całej przestrzeni. Przedstwione rozważania zostały oparte przede wszytkim o wyniki przedstawione przez Meyer w pracy [3].

W niniejszej pracy zostały przeprowadzone przez autora dowody lematów 1 oraz 3 oraz twierdzenia 2, a także dodane zostały własne przykłady 1 i 2. Przeformułowano także lemat 2 głównie dodając założenie o rządzie macierzy $\boldsymbol{A}$, uzasadniając niezerowość prawdopodobieństwa warunku i doprecyzowując rozkład wektora $\boldsymbol{a}$. Dodano także uzasadnienie stosowalności wyznaczonych estymatorów w rozważanym problemie (uwaga 1) i doprecyzowano dowód twierdzenia 3 głównie o dokładną postać wyrażenia $R_0 - R_1$.

Kropka nie oznacza końca zdania.
Ona daje możliwość coraz to lepszej kontynuacji.

Bibliografia

[1] Bartoszewicz J., *Wykłady ze statystyki matematycznej*, PWN, Warszawa, 1996

[2] Fraser D.A.S., Massam H., *A Mixed Primal- Dual Bases Algorithm for Regression under Inequality Constraints. Application to Concave regression*, Scand J. Statist, 16 65-74, 1989

[3] Meyer Mary C., *A test for linear vs convex regression function using shape-restricted regression*, Stanford University, Technical Report No. 2001-20, sierpień 2001

[4] Roman S., *Advanced Linear Algebra*, Springer, 2005