

Grzegorz Mika

Testowanie regresji liniowej przeciwko regresji wypukłej

26 czerwca 2016

Spis treści

| | |
|--|----|
| Literatura | 1 |
| 1. Wstęp | 2 |
| 2. Stożki wypukłe | 2 |
| 3. Regresja wypukła | 4 |
| 4. Statystyka testowa i jej rozkład | 8 |
| 4.1. Lematy i oznaczenia | 8 |
| 4.2. Statystyka testowa | 10 |
| 5. Podsumowanie | 13 |

Literatura

- [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

1. Wstęp

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

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

gdzie zakładamy, że błędy ε_i są niezależnymi zmiennymi losowymi o tym samym rozkładzie normalnym $N(0, \sigma^2)$. Punkty x_i są ustalone. Ponadto zakładamy, 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

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

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 drugim rozdziale zostaną omówione podstawowe własności wielościenne stożków wypukłych traktowanych jako podzbiór przestrzeni liniowej. Trzeci rozdział będzie traktował o konstrukcji estymatora regresji wypukłej jako rzutu wektora danych na taki stożek wielościenne. W czwartym 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].

2. Stożki wypukłe

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

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

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

¹ Definicja podana za Wikipedią: <https://en.m.wikipedia.org/wiki/Orthant>

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*

$$H = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n : a_1x_1 + a_2x_2 + \dots + a_nx_n \geq 0\},$$

gdzie a_1, a_2, \dots, 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

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

Zatem stożek K możemy przedstawić w postaci

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

co będziemy zapisywać skrótowo jako

$$K = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \geq \mathbf{0}\},$$

gdzie

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

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 $m \leq n$. Wtedy stożek K możemy też zapisać w sposób

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

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

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

Wówczas możemy zapisać równoważne przedstawienie stożka K

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

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

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

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

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 przez \mathbf{B}, \mathbf{C} macierze, których kolumnami są odpowiednio wektory $\boldsymbol{\beta}_i, \boldsymbol{\gamma}_i$, związek ten możemy przedstawić jako $\mathbf{B}^T \mathbf{C} = -\mathbf{I}$. Macierz \mathbf{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 $\langle \mathbf{x}, \boldsymbol{\gamma}_i \rangle \leq 0, i = 1, 2, \dots, m$, co dowodzi równoważności przedstawień. \square

3. 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 wielościennej stożek wypukły powstały w wyniku stosownych ograniczeń.

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

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

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

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

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

$$K = \{\mathbf{A}\boldsymbol{\theta} \geq 0\},$$

gdzie \mathbf{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 \mathbf{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

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

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 $\boldsymbol{\gamma}_i = -\mathbf{A}^T \mathbf{e}_i$, $i = 1, 2, \dots, n-2$. Wtedy zbiór K możemy zapisać jako $K = \{\boldsymbol{\theta} \in \mathbb{R}^n: -\mathbf{e}_i^T \mathbf{A} \boldsymbol{\theta} \leq 0, i = 1, 2, \dots, n-2\} = \{\boldsymbol{\theta} \in \mathbb{R}^n: \langle \boldsymbol{\gamma}_i, \boldsymbol{\theta} \rangle \leq 0, i = 1, 2, \dots, n-2\}$.

Z określenia macierzy \mathbf{A} oraz wektorów $\boldsymbol{\gamma}_i, i = 1, 2, \dots, n-2$, widać, że tworzą one układ wektorów liniowo niezależnych. Zatem zbiór $B'_\gamma = \{\boldsymbol{\gamma}_i, i = 1, 2, \dots, n-2\}$ 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 wszystkich wektorów z bazy B'_γ . Sprawdzimy, że warunek ten spełniają wektory $\boldsymbol{\gamma}_{n-1} = \mathbf{1}$ oraz $\boldsymbol{\gamma}_n = (\mathbf{x} - \bar{x}\mathbf{1})$, gdzie $\mathbf{x} = (x_1, x_2, \dots, x_n)$, \bar{x} oznacza wartość średnią, $\mathbf{1} = (1, 1, \dots, 1)^T$, 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

$$\boldsymbol{\gamma}_i = (0, \dots, 0, \overset{(i)}{x_{i+1} - x_{i+2}}, \overset{(i+1)}{x_{i+2} - x_i}, \overset{(i+2)}{x_i - x_{i+1}}, 0, \dots, 0). \text{ Stąd } \langle \mathbf{1}, \boldsymbol{\gamma}_i \rangle = x_{i+1} - x_{i+2} + x_{i+2} - x_i + x_i - x_{i+1} = 0 \text{ oraz } \langle \mathbf{x} - \bar{x}\mathbf{1}, \boldsymbol{\gamma}_i \rangle = \langle \mathbf{x}, \boldsymbol{\gamma}_i \rangle - \bar{x} \langle \mathbf{1}, \boldsymbol{\gamma}_i \rangle = x_i(x_{i+1} - x_{i+2}) + x_{i+1}(x_{i+2} - x_i) + x_{i+2}(x_i - x_{i+1}) = 0. \quad \square$$

Teraz możemy zdefiniować bazę $B_\beta = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_n\}$ dualną do bazy B_γ w zaproponowany w poprzednim paragrafie sposób (por. (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 β_{n-1}, β_n , natomiast $\mathcal{L}(K)$ oznacza przestrzeń rozpiętą przez wektory $\beta_i, i = 1, 2, \dots, n-2$. 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 \mathbf{A} przybiera wtedy postać

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

Zatem stożek powstały z ograniczeń jest postaci

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

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 β_i spełniające warunek $\langle \beta_i, \boldsymbol{\gamma}_i \rangle = -1$ oraz $\langle \beta_i, \boldsymbol{\gamma}_j \rangle = 0, i \neq j$ mają następującą postać

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

Przestrzenie na które będziemy rzutować wektor obserwacji przybierają postać

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

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

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

$$\mathbf{y}_E = \mathcal{E}(\mathcal{E}^T \mathcal{E})^{-1} \mathcal{E}^T \mathbf{y}. \quad (2)$$

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

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

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 $B_J = \{\beta_i, i \in J, \gamma_i, i \in L \setminus J\}$, gdzie J jest pewnym podzbiorem zbioru $L = \{1, 2, \dots, n-2\}$. Zatem każdy element \mathbf{z} należący do $\mathcal{L}(K)$ może być przedstawiony w następujący sposób

$$\mathbf{z} = \sum_{i \in J} b_i \beta_i + \sum_{i \in L \setminus J} c_i \gamma_i, \quad b_i > 0, c_i \geq 0.$$

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

$$\mathbf{z}_{K'} = \sum_{i \in J} b_i \beta_i, b_i > 0.$$

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

$$\mathbf{y} = \mathbf{z} + \mathbf{y}_E = \sum_{i \in J} b_i \beta_i + \sum_{i \in L \setminus J} c_i \gamma_i + d_1 \gamma_{n-1} + d_2 \gamma_n, \quad b_i > 0, c_i \geq 0, d_1, d_2 \in \mathbb{R},$$

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

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

$$\hat{\boldsymbol{\theta}} = \sum_{i \in J} b_i \beta_i + d_1 \gamma_{n-1} + d_2 \gamma_n.$$

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

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

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

$$\begin{aligned} \mathcal{E}(\mathcal{E}^T \mathcal{E})^{-1} \mathcal{E}^T &= \\ &= \begin{bmatrix} -1 & 3 \\ -1 & 1 \\ -1 & -1 \\ -1 & -1 \end{bmatrix} \left(\begin{bmatrix} -1 & -1 & -1 & -1 \\ 3 & 1 & -1 & -3 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ -1 & 1 \\ -1 & -1 \\ -1 & -3 \end{bmatrix} \right)^{-1} \begin{bmatrix} -1 & -1 & -1 & -1 \\ 3 & 1 & -1 & -3 \end{bmatrix} = \\ &= \frac{1}{80} \begin{bmatrix} -1 & 3 \\ -1 & 1 \\ -1 & -1 \\ -1 & -3 \end{bmatrix} \begin{bmatrix} 20 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} -1 & -1 & -1 & -1 \\ 3 & 1 & -1 & -3 \end{bmatrix} = \\ &= \frac{1}{80} \begin{bmatrix} 56 & 32 & 8 & -16 \\ 32 & 24 & 16 & 8 \\ 8 & 16 & 24 & 32 \\ -16 & 8 & 32 & 56 \end{bmatrix}. \end{aligned}$$

Następnie wyznaczmy rzut \mathbf{y}_E .

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

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

Wektor \mathbf{z} , który będziemy rzutować jest postaci $\mathbf{z}^T = (-0.4; 0.45; 0.3; -0.35)$. 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 \mathbf{z} na stożek K jest równy

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

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

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

4. Statystyka testowa i jej rozkład

4.1. Lematy i oznaczenia

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

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

$$S_{L \setminus J} = \text{span}\{\boldsymbol{\gamma}_i, i \in L \setminus J\}$$

$$d = |L \setminus J| = n - 2 - |J|$$

Ponadto niech

$$\mathbf{A}_{L \setminus J}$$

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

$$\mathcal{L}(\mathbf{f}, \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{f}\|^2 \right\}. \quad (7)$$

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

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

$$\begin{aligned} P(\|\mathbf{a}^1\|^2 \leq a | \mathbf{A}\mathbf{V}_2 \mathbf{a}^2 \geq 0) &= \frac{P(\|\mathbf{a}^1\|^2 \leq a \wedge \mathbf{A}\mathbf{V}_2 \mathbf{a}^2 \geq 0)}{P(\mathbf{A}\mathbf{V}_2 \mathbf{a}^2 \geq 0)} = \\ &= \frac{P(\|\mathbf{a}^1\|^2 \leq a) P(\mathbf{A}\mathbf{V}_2 \mathbf{a}^2 \geq 0)}{P(\mathbf{A}\mathbf{V}_2 \mathbf{a}^2 \geq 0)} = P(\|\mathbf{a}^1\|^2 \leq a) = P(\chi_d^2 \leq a) \end{aligned}$$

co należało dowieść. □

Idea powyższego lemat została zaczerpnięta z [3].

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

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

Wektor $\boldsymbol{\rho}$ jest postaci $\boldsymbol{\rho} = \sum_{i \in L \setminus J} b_i \boldsymbol{\gamma}_i$. Z postaci wektora \mathbf{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 \mathbf{y} .

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

Dowód. Dla ustalonego stożka $C_{L \setminus J}$ mamy $\mathbf{y} - \hat{\boldsymbol{\theta}} = \sum_{i \in L \setminus J} b_i \boldsymbol{\gamma}_i$ oraz $\hat{\boldsymbol{\theta}} - \hat{\mathbf{y}} = \sum_{i \in J} c_i \boldsymbol{\beta}_i$. 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 ortonormalizować tak, by $S_1 = \text{span}\{\mathbf{v}_i, i \in J\}$, $S_2 = \text{span}\{\mathbf{v}_i, i \in L \setminus J\}$, $S_3 = \text{span}\{\mathbf{v}_{n-1}, \mathbf{v}_n\}$, gdzie wektory $\mathbf{v}_1, \dots, \mathbf{v}_n$ są wzajemnie ortonormalne. Niech V oznacza macierz, której kolumnami są wektory \mathbf{v} . Wektor \mathbf{y} możemy zapisać jako $\mathbf{y} = \sum_{i=1}^n \langle \mathbf{v}_i, \mathbf{y} \rangle \mathbf{v}_i = \sum_{i=1}^n a_i \mathbf{v}_i$. Analogicznie jak w dowodzie lematu 2 możemy zdefiniować wektor $\mathbf{a} = (a_1, a_2, \dots, a_n)^T$, który ma n -wymiarowy rozkład normalny $N(0, \mathbf{I})$. Niech $\mathbf{a}^1 = (a_1, \dots, a_{n-d-2})^T$, $\mathbf{a}^2 = (a_{n-d-1}, \dots, a_{n-2})^T$, $\mathbf{a}^3 = (a_{n-1}, a_n)^T$ oraz zapiszmy macierz V jako $[V_1 | V_2 | V_3]$. Wówczas $\mathbf{y} = V_1 \mathbf{a}^1 + V_2 \mathbf{a}^2 + V_3 \mathbf{a}^3$ oraz $\mathbf{y} - \hat{\boldsymbol{\theta}} = V_2 \mathbf{a}^2$, $\hat{\boldsymbol{\theta}} - \hat{\mathbf{y}} = V_1 \mathbf{a}^1$ oraz $\hat{\mathbf{y}} = V_3 \mathbf{a}^3$. Wektory \mathbf{a}^1 , \mathbf{a}^2 i \mathbf{a}^3 są niezależne, zatem wektory $V_1 \mathbf{a}^1$, $V_2 \mathbf{a}^2$ oraz $V_3 \mathbf{a}^3$ jako mieralne funkcje tych wektorów również są niezależne dla ustalonego $C_{L \setminus J}$. \square

4.2. 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

$$H_0: f(x) = ax + b \quad \text{vs.} \quad H_1: f \in \mathcal{F}, \quad (8)$$

gdzie \mathcal{F} jest klasą funkcji wypukłych.

Niech $\hat{\mathbf{y}}$ oznacza estymator regresji liniowej, czyli rzut wektora danych \mathbf{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 $R_0 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2$ oraz $R_1 = \sum_{i=1}^n (y_i - \hat{\theta}_i)^2 = \|\mathbf{y} - \hat{\boldsymbol{\theta}}\|^2$.

Uwaga 1. Powyższe estymatory $\hat{\mathbf{y}}$ i $\hat{\boldsymbol{\theta}}$, jako elementy minimalizujące kwadrat normy błędu, czyli elementy optymalne w sensie aproksymacji średnio-kwadratowej, 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 $\mathbf{x} = \phi(\mathbf{x}) + \boldsymbol{\varepsilon}$, gdzie $\phi(\mathbf{x})$ jest poszukiwanym parametrem, a wektor błędów $\boldsymbol{\varepsilon}$ ma n -wymiarowy rozkład normalny $N(\mathbf{0}, \sigma^2 \mathbf{I})$ ze znaną wariancją. Stosując metodę najmniejszych kwadratów dostajemy, że estymatorem $\phi(\mathbf{x})$ jest wektor $\hat{\phi}(\mathbf{x})$ minimalizujący wyrażenie $\sum_{i=1}^n (\mathbf{x}_i - \phi(\mathbf{x})_i)^2$, gdzie $\mathbf{x}_i, \phi(\mathbf{x})_i$ oznaczają kolejne współrzędne odpowiednio wektora \mathbf{x} i $\phi(\mathbf{x})$. Stosując metodę największej wiarygodności przy założeniu normalności rozkładu wektora \mathbf{x} będziemy maksymalizować funkcję wiarygodności postaci $L(\phi(\mathbf{x}), \mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2} (\mathbf{x}_i - \phi(\mathbf{x})_i)^2\} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \phi(\mathbf{x})_i)^2\}$.

Maksymalizacja tej funkcji jest równoważna maksymalizacji jej logarytmu $l(\phi(\mathbf{x}), \mathbf{x}) = \ln(2\pi\sigma^2)^{-\frac{n}{2}} - \frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \phi(\mathbf{x})_i)^2$, co przy założeniu znajomości wariancji prowadzi do równoważnego zagadnienia minimalizacji wyrażenia $\sum_{i=1}^n (\mathbf{x}_i - \phi(\mathbf{x})_i)^2$, czyli tego samego, które otrzymaliśmy stosując metodę najmniejszych kwadratów. Zatem przy założeniu normalności rozkładu wektora obserwacji \mathbf{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

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

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

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

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

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

Zatem zgodnie z definicją 5 możemy rozważyć zbiór odrzucenia C w naszym problemie (8) postaci

$$\begin{aligned} C &= \phi^{-1}(\{1\}) = \left\{ \mathbf{y} \in \mathbb{R}^n : \frac{\sup_{\mathbf{f} \in \Theta} \mathcal{L}(\mathbf{f}, \mathbf{y})}{\sup_{\mathbf{f} \in \Theta_0} \mathcal{L}(\mathbf{f}, \mathbf{y})} > c \right\} =^2 \\ &= \left\{ \mathbf{y} \in \mathbb{R}^n : \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2} \|\mathbf{y} - \hat{\boldsymbol{\theta}}\|^2\}}{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2} \|\mathbf{y} - \hat{\mathbf{y}}\|^2\}} > c \right\} = \\ &= \left\{ \mathbf{y} \in \mathbb{R}^n : \exp\left\{ \frac{1}{2\sigma^2} (\|\mathbf{y} - \hat{\mathbf{y}}\|^2 - \|\mathbf{y} - \hat{\boldsymbol{\theta}}\|^2) \right\} > c \right\} \\ &= \left\{ \mathbf{y} \in \mathbb{R}^n : \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2 - \|\mathbf{y} - \hat{\boldsymbol{\theta}}\|^2}{\sigma^2} > c' \right\} = \\ &= \left\{ \mathbf{y} \in \mathbb{R}^n : \frac{R_0 - R_1}{\sigma^2} > c' \right\}. \end{aligned}$$

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 licznosc zbioru $L \setminus J$ (por. (5)).

² przejście możliwe dzięki uwadze 1

Twierdzenie 3. Przy założeniu prawdziwości hipotezy zerowej postawionego problemu (8) mamy

$$P(M \leq a) = \sum_{d=0}^{n-2} P(\chi_{n-d-2}^2 \leq a) P(D = d),$$

gdzie $\chi_0^2 \equiv 0$, czyli rozkładem statystyki 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 $\mathbf{y} = \boldsymbol{\varepsilon}$. 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)). Zatem wyrażenie $R_1 = \|\mathbf{y} - \hat{\boldsymbol{\theta}}\|^2$ zależy poprzez $\hat{\boldsymbol{\theta}}$ od wyboru zbioru J , który jest losowy, różne wartości wektora błędu $\boldsymbol{\varepsilon}$ mogą umieścić wektor danych \mathbf{y} w różnych zbiorach $C_{L \setminus J}$. Zauważmy jednak, że z jednoznaczności przedstawienia wektora w zbiorze $C_{L \setminus J}$ zdarzenia postaci $\{\mathbf{y} \in C_{L \setminus J}\}$, $J \in \mathcal{P}(L)$, gdzie $\mathcal{P}(L)$ oznacza zbiór potęgowy zbioru L , są wzajemnie rozłączne oraz ich suma stanowi całą przestrzeń zdarzeń elementarnych. Ponadto prawdopodobieństwa tych zdarzeń są niezerowe. Możemy zatem skorzystać z twierdzenia o prawdopodobieństwie całkowitym.

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

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 $\mathbf{A}_{L \setminus J}$ (por. (6)) można zapisać jako $[\mathbf{A}_1 | \mathbf{A}_2]$, gdzie macierz \mathbf{A}_1 jest wymiaru $n \times d$. Zatem kolumny macierzy \mathbf{A}_1 rozpinają $S_{L \setminus J}$, natomiast kolumny macierzy \mathbf{A}_2 są ortogonalne do przestrzeni $S_{L \setminus J}$. Dodatkowo, gdy $\boldsymbol{\varepsilon} \in C_{L \setminus J}$, zachodzi $\mathbf{A}_1^T \boldsymbol{\varepsilon} \geq 0$ oraz $\mathbf{A}_2^T \boldsymbol{\varepsilon} \geq 0$. Stąd na mocy lematu 2 dostajemy, że rozkładem warunkowym $\frac{\|\hat{\boldsymbol{\varepsilon}}\|^2}{\sigma^2}$ 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 $\hat{\boldsymbol{\rho}} = \mathbf{y} - \hat{\boldsymbol{\theta}}$, otrzymujemy, że jeśli hipoteza zerowa $\boldsymbol{\theta} \in \text{span}\{\boldsymbol{\gamma}_{n-1}, \boldsymbol{\gamma}_n\}$ jest prawdziwa to rozkładem warunkowym $\frac{R_1}{\sigma^2}$ przy ustalonym zbiorze J jest χ_d^2 , gdzie $d = |L \setminus J|$.

Zmienna losowa $\frac{R_0}{\sigma^2}$ jest rzutem wektora danych \mathbf{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 \mathbf{x}, \mathbf{y} i dowolnej liczby rzeczywistej a mamy, że $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle$ oraz $\|a\mathbf{x}\|^2 = a^2\|\mathbf{x}\|^2$.

Zatem dla ustalonego zbioru J dostajemy, że

$$\begin{aligned} R_0 - R_1 &= \|\mathbf{y} - \hat{\mathbf{y}}\|^2 - \|\mathbf{y} - \hat{\boldsymbol{\theta}}\|^2 = \left\| \sum_{i \in L \setminus J} b_i \boldsymbol{\gamma}_i + \sum_{i \in J} c_i \boldsymbol{\beta}_i \right\|^2 - \left\| \sum_{i \in L \setminus J} b_i \boldsymbol{\gamma}_i \right\|^2 = \\ &= \sum_{i \in L \setminus J} b_i^2 \|\boldsymbol{\gamma}_i\|^2 + \sum_{i \in J} c_i^2 \|\boldsymbol{\beta}_i\|^2 + 2 \sum_{\substack{i \in L \setminus J \\ j \in J}} b_i c_j \langle \boldsymbol{\gamma}_i, \boldsymbol{\beta}_j \rangle + 2 \sum_{\substack{i, j \in L \setminus J \\ i \neq j}} b_i b_j \langle \boldsymbol{\gamma}_i, \boldsymbol{\gamma}_j \rangle + \\ &\quad + 2 \sum_{\substack{i, j \in J \\ i \neq j}} c_i c_j \langle \boldsymbol{\beta}_i, \boldsymbol{\beta}_j \rangle - \sum_{i \in J} b_i^2 \|\boldsymbol{\gamma}_i\|^2 - 2 \sum_{\substack{i, j \in J \\ i \neq j}} b_i b_j \langle \boldsymbol{\gamma}_i, \boldsymbol{\gamma}_j \rangle = \end{aligned}$$

$$= \sum_{i \in J} c_i^2 \|\beta_i\|^2 + 2 \sum_{\substack{i, j \in J \\ i \neq j}} c_i c_j \langle \beta_i, \beta_j \rangle = \left\| \sum_{i \in J} c_i \beta_i \right\|^2 = \|\hat{\mathbf{y}} - \hat{\boldsymbol{\theta}}\|^2,$$

bo $\langle \beta_i, \beta_j \rangle = 0$ dla $i \neq j$. Otrzymaliśmy zatem rzut wektora $\boldsymbol{\varepsilon}$ na przestrzeń $\text{span}\{\beta_i, i \in J\}$ dla ustalonego zbioru J . Rzut ten może zostać przeskalowany o znaną wartość wariancji. Wtedy tak uzyskana zmienna 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 dokładnej jego postaci. Niech zatem D będzie zdefiniowaną wcześniej zmienną losową reprezentującą licznosc 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 $P(D = d)$, $d = 0, 1, \dots, n-2$ są wyliczane na podstawie względnych objętości zbiorów $C_{L \setminus J}$, $J \in \mathcal{P}(L)$. Prawdopodobieństwo, że $\mathbf{y} \in C_{L \setminus J}$, 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.

5. Podsumowanie

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

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

gdzie \mathcal{F} jest klasą funkcji wypukłych.

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

$$P(M \leq a) = \sum_{d=0}^{n-2} P(\chi_{n-d-2}^2 \leq a) P(D = d),$$

będący mieszanym rozkładem χ^2 z wagami $P(D = d)$. Prawdopodobieństwa $P(D = d)$ wyznaczone są na podstawie względnych objętości zbiorów $C_{L \setminus J}$, $J \in \mathcal{P}(L)$ w porównaniu do całej przestrzeni. Przedstawione rozważania zostały oparte przede wszystkim 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 twierdzenie 2 głównie dodając założenie o rządzie macierzy \mathbf{A} , uzasadniając niezerowość prawdopodobieństwa warunku i doprecyzowując rozkład wektora \mathbf{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.*