

Dariusz Biskup

BAYESOWSKI WYBÓR MODELU W REGRESJI WIELOMIANOWEJ

1. Wstęp

W artykule opisane zostanie zagadnienie wyboru modelu regresji jednej zmiennej, opisującej zależność pomiędzy zmienną objaśnianą Y a zmienną objaśniającą X , w sytuacji gdy funkcja regresji jest wielomianem dowolnego stopnia. Założymy zatem, że model regresji można opisać za pomocą następującej funkcji:

$$Y(x) = \sum_{i=0}^k \theta_i x^i + \varepsilon(x), \quad (1)$$

gdzie: k – stopień wielomianu,

θ_i – parametry funkcji regresji,

$Y(x)$ – zmienna losowa Y obserwowana dla ustalonej na poziomie x wartości zmiennej objaśniającej X ,

$\varepsilon(x)$ – zmienna losowa o zerowej wartości oczekiwanej i precyzji w .

Wektor parametrów modelu (1) oznaczony zostanie przez $\theta = [\theta_0 \ \theta_1 \ \dots \ \theta_k]^T$. Do estymacji parametrów modelu regresji (1) wykorzystane zostanie podejście bayesowskie. Przyjęto założenie, że parametry θ_i $i = 0, 1, \dots, k$ modelu są zmiennymi losowymi. Zmienną losową jest również stopień wielomianu k .

Wybór modelu regresji polegać będzie na znalezieniu rozkładu prawdopodobieństwa dla parametru k . Wybierany będzie model, dla którego prawdopodobieństwo będzie największe.

Wyznaczenie rozkładu *a posteriori* dla stopnia wielomianu wymaga dla niesprzężonych rozkładów *a priori* zastosowania metod numerycznych. W tym celu wykorzystany zostanie algorytm RJMCMC (*Reversible Jump Markov Chain Monte Carlo*). Algorytm ten stanowi uogólnienie algorytmu Metropolisa-Hastingsa na

sytuacje, w których przestrzeń parametrów modelu może mieć zmienną liczbę wymiarów (por. np. [4; 5]).

2. Aproksymacja wielomianu

Algorytm RJMCMC wymaga zdefiniowania pewnej deterministycznej funkcji, która umożliwi modyfikację wartości parametrów dla modeli o różnej liczbie wymiarów. Przede wszystkim niezbędna będzie funkcja odwzorowująca wektor parametrów modelu stopnia $k + 1$ na model stopnia k oraz funkcja dokonująca transformacji odwrotnej. Jedną z możliwości zdefiniowania takiej funkcji jest aproksymacja wielomianów metodą najmniejszych kwadratów.

Zadanie aproksymacji wielomianu stopnia $k + 1$ o postaci $f(x) = \sum_{i=0}^{k+1} \theta'_i x^i$ za

pomocą wielomianu stopnia k o postaci $f(x) = \sum_{i=0}^k \theta_i x^i$, przy założeniu, że x należy do przedziału (a_1, a_2) , można rozwiązać przez znalezienie takich wartości parametrów θ_i , dla których wyrażenie

$$\int_{a_1}^{a_2} \left(\sum_{i=0}^{k+1} \theta'_i x^i - \sum_{i=0}^k \theta_i x^i \right)^2 dx \quad (2)$$

osiąga minimum. Aby wyznaczyć wartości parametrów $\theta_0, \theta_1, \dots, \theta_k$ spełniających kryterium (2), należy obliczyć pochodne cząstkowe z wyrażenia

$$\int_{a_1}^{a_2} \left(\sum_{i=0}^{k+1} \theta'_i x^i - \sum_{i=0}^k \theta_i x^i \right)^2 dx$$

po $\theta_0, \theta_1, \dots, \theta_k$, przyrównać je do zera oraz rozwiązać układ $k + 1$ równań.

Uzyskane rozwiązanie przybiera następującą postać:

$$\theta_i = \theta'_i + \theta'_{k+1} W_{(k+1)i}, \text{ dla } i = 0, 1, \dots, k, \quad (3)$$

gdzie

$$W_{ki} = (-1)^{k-i+1} \binom{k}{i} \frac{\sum_{j=0}^{k-i} \binom{k}{j} \binom{k}{j+i} a_1^j a_2^{k-i-j}}{\sum_{j=0}^{k-i} \binom{k}{j} \binom{k}{j+i}}. \quad (4)$$

Na przykład w przypadku aproksymacji funkcji kwadratowej funkcją liniową wzór (3) można rozpisać następująco:

$$\begin{aligned}\theta_0 &= \theta'_0 - \theta'_2 \frac{a_1^2 + 4a_1a_2 + a_2^2}{6}, \\ \theta_1 &= \theta'_1 + \theta'_2(a_1 + a_2),\end{aligned}$$

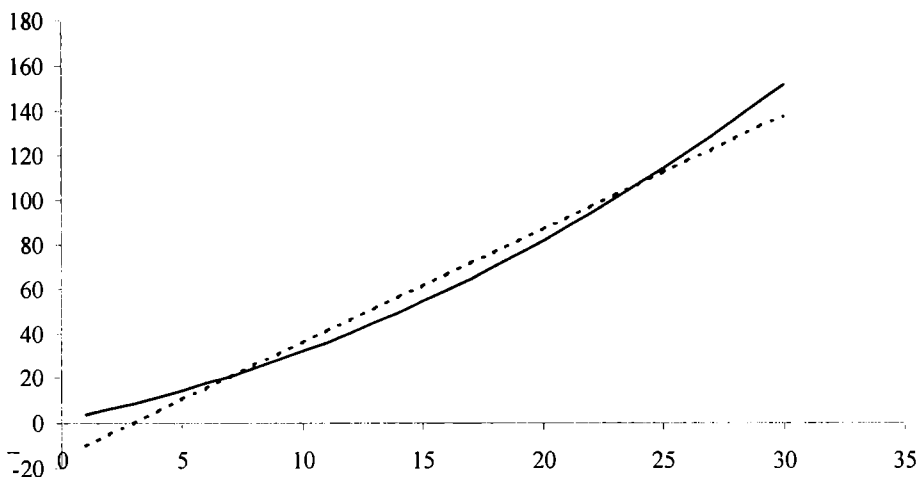
w przypadku zaś aproksymacji wielomianu 3 stopnia funkcją kwadratową mamy:

$$\begin{aligned}\theta_0 &= \theta'_0 + \theta'_3 \frac{a_1^3 + 9a_1^2a_2 + 9a_1a_2^2 + a_2^3}{20}, \\ \theta_1 &= \theta'_1 - 3\theta'_3 \frac{a_1^2 + 3a_1a_2 + a_2^2}{5}, \\ \theta_2 &= \theta'_2 + 3\theta'_3 \frac{a_1 + a_2}{2}.\end{aligned}$$

Na przykład zastosowanie kryterium (2) do aproksymacji wielomianu 2 stopnia postaci $f(x) = 0,1x^2 + 2x + 2$ za pomocą funkcji liniowej, przy założeniu, że $x \in (1, 30)$, pozwala uzyskać następujące rozwiązanie:

$$g(x) = 5,1x - 15,0167.$$

Ilustrację powyższej aproksymacji przedstawiono na rys. 1.



Rys. 1. Aproksymacja funkcji kwadratowej funkcją liniową

Układ równań zapisanych wzorem (3) pozwala na przejście z modelu o większej liczbie parametrów do modelu o mniejszej ich liczbie. Aby dokonać transformacji odwrotnej, układ równań (3) należałoby rozwiązać ze względu na parametry θ'_i ($i = 0, 1, \dots, k+1$). Mamy wtedy jednak $k+2$ niewiadome oraz $k+1$ równań. Problem ten można rozwiązać poprzez dopisanie dodatkowego równania, które wiązać będzie parametry θ'_i przez pewną liczbę losową u . Przyjęte zostanie, że liczba ta pochodzić będzie z rozkładu $N(0, \kappa)$ i spełniać będzie warunek

$$u = \theta'_{k+1}. \quad (5)$$

Parametr κ można ustalić arbitralnie tak, aby losowane dla parametru θ'_{k+1} wartości były wystarczająco duże.

Znalezienie transformacji odwrotnej polegać będzie na rozwiązaniu układu równań zapisanych wzorami (3) i (5) ze względu na parametry $\theta'_0, \theta'_1, \dots, \theta'_{k+1}$. Rozwiązanie przyjmuje wówczas następującą postać:

$$\theta'_i = \theta_i - uW_{(k+1)i}, \text{ dla } i = 0, 1, \dots, k, \quad (6)$$

$$\theta'_{k+1} = u.$$

3. Algorytm *Reversible Jump Markov Chain Monte Carlo*

Algorytm RJMCMC stanowi uogólnienie algorytmu Metropolisa-Hastingsa. Gdy nie jest zmieniany wymiar przestrzeni parametrów modelu, odpowiada on dokładnie algorytmowi Metropolisa-Hastingsa (z kolei szczególnym przypadkiem algorytmu Metropolisa-Hastingsa jest algorytm Gibbsa). W sytuacji gdy występuje zmiana wymiaru przestrzeni parametrów, algorytm RJMCMC składa się z następujących etapów (por. [5]):

1. Niech (j, θ_j) oznacza aktualne wartości wektora parametrów θ_j modelu o numerze j , którego wymiar wektora parametrów wynosi k_j .
2. Wykonanie przejścia do stanu $(j', \theta_{j'})$ ze stanu (j, θ_j) z prawdopodobieństwem $h(j, j')$.
3. Wylosowanie wektora $\mathbf{u}$ z zadanego rozkładu $q(\mathbf{u}|\theta_j, j, j')$.
4. Podstawienie $(\theta_{j'}, \mathbf{u}') = g_{j, j'}(\theta_j, \mathbf{u})$, gdzie $g_{j, j'}$ oznacza wzajemnie jednoznaczna, deterministyczną funkcję. Podstawienie to ma na celu dopasowanie wymiarów modelu aktualnego do modelu docelowego ($k_j + \dim(\mathbf{u}) = k_{j'} + \dim(\mathbf{u}')$). Ponadto $g_{j', j} = g_{j, j'}^{-1}$.
5. Akceptacja zmiany modelu z prawdopodobieństwem $\min(1, A)$, gdzie:

$$A = \frac{p(\mathbf{y}|\theta_{j'}, j')p(\theta_{j'}|j')p(j')h(j', j)q(\mathbf{u}'|\theta_{j'}, j', j)}{p(\mathbf{y}|\theta_j, j)p(\theta_j|j)p(j)h(j, j')q(\mathbf{u}|\theta_j, j, j')} \left| \frac{\partial g_{j, j'}(\theta_j, \mathbf{u})}{\partial(\theta_j, \mathbf{u})} \right|, \quad (7)$$

gdzie $p(\mathbf{y}|\cdot)$ oznacza funkcję wiarygodności, $p(\theta_{j'}|j')$ i $p(\theta_j|j)$ – rozkłady *a priori* parametrów, a $p(j)$ i $p(j')$ – rozkłady *a priori* modeli.

Jeżeli wymiar modelu j' jest wyższy niż wymiar modelu j , wówczas $\dim(\mathbf{u}') = 0$ i funkcja $g_{j, j'}$ wykonuje odwzorowanie $\theta_{j'} = g_{j, j'}(\theta_j, \mathbf{u})$. Jeżeli wymiar modelu j' jest niższy niż wymiar modelu j , to $\dim(\mathbf{u}) = 0$ i odwzorowanie ma postać $(\theta_{j'}, \mathbf{u}') = g_{j, j'}(\theta_j)$, przy czym wartość zmiennej $\mathbf{u}'$ jest ignorowana.

Jeżeli nie następuje zmiana modelu, czyli wymiar modelu j jest równy wymiarowi modelu j' , to nowe wartości wektora parametrów mogą być generowane przy wykorzystaniu algorytmu Gibbsa.

W praktyce najczęściej iteracje algorytmu RJMCMC są wykonywane na przemian z iteracjami algorytmu Gibbsa.

4. Zastosowanie algorytmu RJMCMC do estymacji parametrów w regresji wielomianowej

4.1. Rozkłady *a priori* parametrów

Przyjęte zostaną następujące rozkłady *a priori* dla parametrów modelu (1):

$$\theta = [\theta_0 \ \theta_1 \ \dots \ \theta_k]^T \sim N(\mathbf{m}, \mathbf{V}), \quad (8)$$

$$w \sim \text{Gamma}(\alpha, \beta), \quad (9)$$

gdzie: $N(\mathbf{m}, \mathbf{V})$ – oznacza rozkład normalny o wektorze wartości oczekiwanych $\mathbf{m}$ i macierzy kowariancji $\mathbf{V}$, $G(\alpha, \beta)$ – rozkład gamma z parametrami α i β .

Przyjmowane w konkretnych zastosowaniach rozkłady *a priori* nie mogą być nieinformacyjne, jak może to mieć miejsce w przypadku estymacji bayesowskiej parametrów konkretnego modelu (por. [1; 2; 7; 8]). Istnieje wiele propozycji rozwiązania tego problemu. Część z nich polega na wykorzystaniu części danych do sformułowania informacyjnego rozkładu *a priori*, a pozostałej części do wyboru modelu (por. [2; 7]).

Inna propozycja, która wykorzystana zostanie w niniejszym artykule, polega na przyjęciu jako rozkładów *a priori* dla parametrów modeli oszacowanych oddzielnie dla każdego z nich rozkładów *a posteriori* wyznaczonych na podstawie wszystkich dostępnych danych (por. [1]). Propozycja ta, mimo że spotkała się z pewną krytyką (por. [1] – dyskusja), daje zbliżone wyniki do innych poprawniejszych teo-

retycznie rozwiązań zgłoszonych w artykułach [2; 7], a ponadto jest najłatwiejsza do implementacji numerycznej.

4.2. Rozkłady warunkowe parametrów

Warunkowe rozkłady *a posteriori* parametrów modelu mają następującą postać (wykorzystywane są one w sytuacji, gdy stosowany jest algorytm Gibbsa – czyli iteracja nie zmienia wymiaru modelu):

$$\theta|w \sim N\left(\left(w\mathbf{F}^T\mathbf{F} + \mathbf{V}^{-1}\right)^{-1}\left(w\mathbf{F}^T\mathbf{y} + \mathbf{V}^{-1}\mathbf{m}\right), \left(w\mathbf{F}^T\mathbf{F} + \mathbf{V}^{-1}\right)^{-1}\right), \quad (10)$$

$$w|\theta \sim \Gamma\left(\alpha + \frac{n}{2}, \beta + \frac{(\mathbf{y} - \mathbf{F}\theta)^T(\mathbf{y} - \mathbf{F}\theta)}{2}\right), \quad (11)$$

gdzie

$$\mathbf{F} = \begin{bmatrix} 1 & x_1 & \dots & x_1^k \\ 1 & x_2 & \dots & x_2^k \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \dots & x_n^k \end{bmatrix}$$

oznacza macierz obserwacji zmiennej niezależnej X oraz

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

wektor obserwacji zmiennej zależnej Y .

4.3. Wyprowadzenie wzoru na prawdopodobieństwo akceptacji

Wyprowadzony zostanie najpierw wzór na prawdopodobieństwo akceptacji przy przejściu z wielomianu stopnia k na wielomian stopnia $k' = k + 1$.

Jakobian transformacji wektora $[\theta_0 \ \theta_1 \ \dots \ \theta_k \ u \ w]$ na wektor $[\theta'_0 \ \theta'_1 \ \dots \ \theta'_{k+1} \ u']$ wyznaczany na podstawie funkcji opisanych wzorem (6) oraz poprzez transformację $w' = w$, wyznaczany zgodnie z formułą

$$J = \begin{vmatrix} \frac{d\theta'_0}{d\theta_0} & \frac{d\theta'_0}{d\theta_1} & \frac{d\theta'_0}{d\theta_k} & \frac{d\theta'_0}{du} & \frac{d\theta'_0}{dw} \\ \frac{d\theta'_1}{d\theta_0} & \frac{d\theta'_1}{d\theta_1} & \frac{d\theta'_1}{d\theta_k} & \frac{d\theta'_1}{du} & \frac{d\theta'_1}{dw} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{d\theta'_{k+1}}{d\theta_0} & \frac{d\theta'_{k+1}}{d\theta_1} & \dots & \frac{d\theta'_{k+1}}{du} & \frac{d\theta'_{k+1}}{dw} \\ \frac{dw'}{d\theta_0} & \frac{dw'}{d\theta_1} & \dots & \frac{dw'}{du} & \frac{dw'}{dw} \end{vmatrix},$$

przybiera postać:

$$J = \begin{vmatrix} \mathbf{I} & \begin{matrix} -W_{(k+1)0} & 0 \\ \vdots & \vdots \\ -W_{(k+1)k} & 0 \end{matrix} \\ \begin{matrix} 0 & \dots & 0 \\ 0 & \dots & 0 \end{matrix} & \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \end{vmatrix} = 1, \quad (12)$$

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

Składnik A prawdopodobieństwa akceptacji jest z kolei równy (por. wzór (7)):

$$A = \frac{p(k')p(\mathbf{y}|\theta', w')p(\theta', w'|k')}{p(k)p(\mathbf{y}|\theta, w)p(\theta, w|k)} \times \frac{h(k', k)}{h(k, k')} \times \frac{1}{q(u|\theta, w, k, k')} \times J.$$

Przyjęte zostanie, że wszystkie modele mają jednakowe prawdopodobieństwa *a priori*:

$$p(k) = \frac{1}{k_{\max}}, \quad (13)$$

a zatem $\frac{p(k')}{p(k)} = 1$. Przyjęto arbitralnie, że maksymalny stopień wielomianu $k_{\max} = 10$.

Jeżeli następuje przejście z wielomianu stopnia k na wielomian stopnia $k' = k - 1$, to

$$A = \frac{p(k')p(\mathbf{y}|\theta', w')p(\theta', w'|k')}{p(k)p(\mathbf{y}|\theta, w)p(\theta, w|k)} \times \frac{h(k', k)}{h(k, k')} \times q(u|\theta, w, k, k') \times J^{-1}.$$

Dokonywany jest losowy wybór pomiędzy zwiększeniem bądź zmniejszeniem stopnia wielomianu o jeden z prawdopodobieństwem b_k (dla zwiększenia) oraz $d_k = 1 - b_k$ (dla zmniejszenia). Prawdopodobieństwa te zależą od parametru k .

Przyjęte zostanie, że $d_1 = 0$, $b_{k_{\max}} = 0$ oraz $b_k = d_k = 0,5$ dla $k = 2, 3, \dots, k_{\max} - 1$. Oznacza to, że $h(k', k) = d_{k+1}$ oraz $h(k, k') = b_k$. Funkcja gęstości $q(u|\theta, w, k, k')$, z której losuje się wartość u , jest typu $N(0, \kappa)$.

Funkcje wiarygodności równe są:

$$p(y|\theta', w') = \prod_{i=1}^n \frac{\sqrt{w'}}{\sqrt{2\pi}} \exp\left(\frac{-w' \left(y_i - \sum_{j=0}^{k+1} \theta'_j x_i^j\right)^2}{2}\right), \text{ dla } k' = k + 1,$$

$$p(y|\theta', w') = \prod_{i=1}^n \frac{\sqrt{w'}}{\sqrt{2\pi}} \exp\left(\frac{-w' \left(y_i - \sum_{j=0}^{k-1} \theta'_j x_i^j\right)^2}{2}\right), \text{ dla } k' = k - 1,$$

$$p(y|\theta, w) = \prod_{i=1}^n \frac{\sqrt{w}}{\sqrt{2\pi}} \exp\left(\frac{-w \left(y_i - \sum_{j=0}^k \theta_j x_i^j\right)^2}{2}\right).$$
(14)

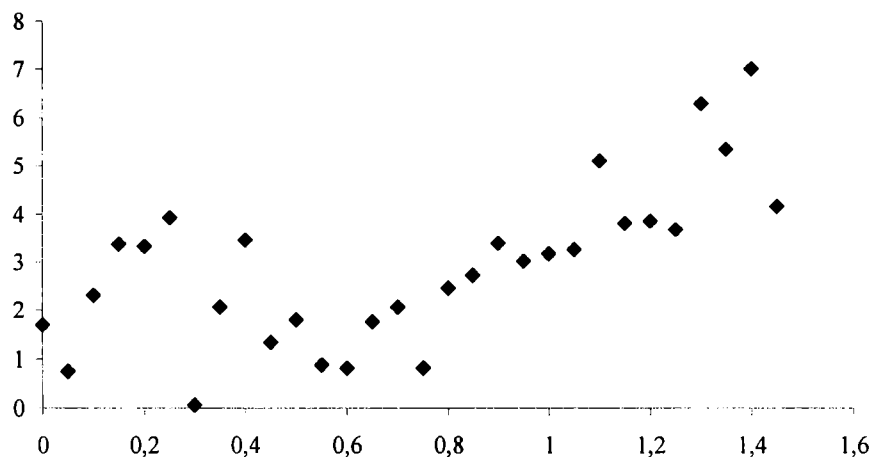
Rozkłady *a priori* $p(\theta', w'|k')$ i $p(\theta, w|k)$ zgodne są ze wzorami podanymi w punkcie 4.1.

5. Przykłady symulacyjne

5.1. Przykład 1 – model kwadratowy

W przykładzie wygenerowanych zostało 30 obserwacji zgodnie z zależnością $Y(x) = x^2 + 0,5x + 2 + \varepsilon(x)$, gdzie zmienna losowa $\varepsilon(x)$ ma rozkład $N(0, 1)$. Wartości zmiennej X są rozłożone równomiernie co 0,05 na odcinku od 0 do 1,45. Ze względu na niewielką liczbę obserwacji trudno jest jednoznacznie rozstrzygnąć (nawet wzrokowo), czy bardziej adekwatny jest tutaj model liniowy czy kwadratowy. Wykres korelacyjny zmiennych X i Y przedstawia rys. 2.

Rozkłady *a posteriori*, które zostaną wykorzystane jako rozkłady *a priori* dla poszczególnych modeli, wyznaczone zostały za pomocą programu WINBUGS. Ich parametry przedstawia tab. 1 (zgodnie z oznaczeniami ze wzorów (8) i (9)). Zapisane zostały rozkłady tylko dla wielomianów stopnia nie większego niż 4, gdyż



Rys. 2. Wykres korelacyjny dla przykładu 1

Tabela 1. Rozkłady *a priori* dla przykładu 1

Stopień wielomianu <i>k</i>	Wektor <i>m</i>	Macierz <i>V</i>	<i>α</i>	<i>β</i>
1	$[1,13 \ 2,47]^T$	$\begin{bmatrix} 0,23 & 0 \\ 0 & 0,32 \end{bmatrix}$	12,25	19,44
2	$[2,46 \ -3,24 \ 3,94]^T$	$\begin{bmatrix} 0,37 & 0 & 0 \\ 0 & 3,57 & 0 \\ 0 & 0 & 1,56 \end{bmatrix}$	12,54	14,76
3	$[2,42 \ -2,77 \ 3,13 \ 0,36]^T$	$\begin{bmatrix} 0,58 & 0 & 0 & 0 \\ 0 & 21,53 & 0 & 0 \\ 0 & 0 & 56,1 & 0 \\ 0 & 0 & 0 & 11,42 \end{bmatrix}$	13,22	16,52
4	$[1,42 \ 13,13 \ -47,9 \ 55,61 \ -19,04]^T$	$\begin{bmatrix} 0,64 & 0 & 0 & 0 & 0 \\ 0 & 60,84 & 0 & 0 & 0 \\ 0 & 0 & 493 & 0 & 0 \\ 0 & 0 & 0 & 534 & 0 \\ 0 & 0 & 0 & 0 & 62,41 \end{bmatrix}$	12	12,37

Źródło: opracowanie własne.

dla wielomianów wyższych stopni obliczone prawdopodobieństwa *a posteriori* modeli są bliskie zeru.

Przeprowadzono łącznie 200 000 iteracji algorytmu RJMCMC, przy czym prawdopodobieństwa modelu obliczane zostały tylko dla ostatnich 100 000 iteracji. Otrzymane wyniki przedstawia tab. 2.

Tabela 2. Prawdopodobieństwa *a posteriori* rozpatrywanych modeli

Stopień wielomianu k	1	2	3	4
Obliczone prawdopodobieństwo	0,27619	0,60179	0,12120	0,00082

Źródło: opracowanie własne.

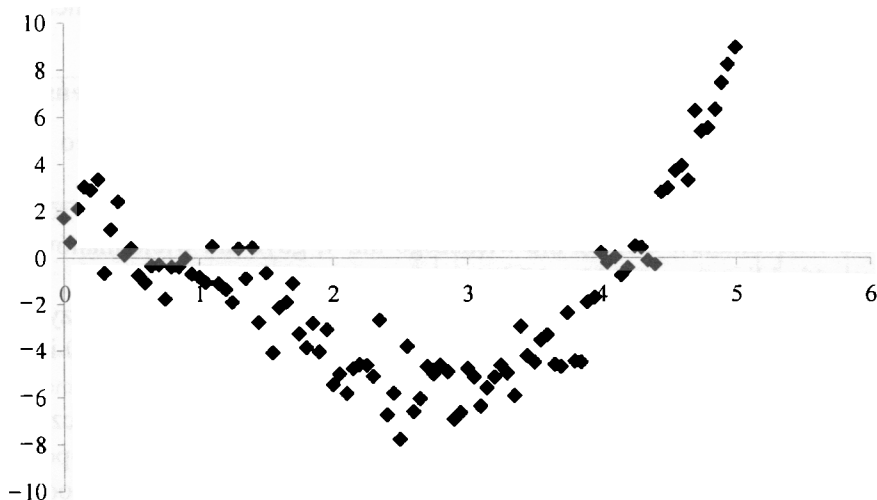
Modelem najbardziej prawdopodobnym okazał się zatem wielomian stopnia 2, czyli kwadratowa funkcja regresji. Stosunkowo wysokie prawdopodobieństwo *a posteriori* otrzymano również dla modelu liniowego.

5.2. Przykład 2 – wielomian 3 stopnia

W przykładzie wygenerowanych zostało 101 obserwacji zgodnie z zależnością $Y(x) = 0,5x^3 - 2x^2 - x + 2 + \varepsilon(x)$, gdzie zmienna losowa $\varepsilon(x)$ ma rozkład $N(0, 1)$. Wartości zmiennej X są rozłożone równomiernie co 0,05 na odcinku od 0 do 5. Wykres korelacyjny zmiennych X i Y przedstawia rys. 3.

Rozkłady *a posteriori*, które zostaną wykorzystane jako rozkłady *a priori* dla poszczególnych modeli, wyznaczone zostały podobnie jak w poprzednim przykładzie za pomocą programu WINBUGS. Ich parametry przedstawia tab. 3. Zapisane zostały rozkłady tylko dla wielomianów stopnia nie większego niż 4, gdyż dla wielomianów wyższych stopni obliczone prawdopodobieństwa *a posteriori* są bliskie zeru.

Przeprowadzono łącznie 200 000 iteracji algorytmu RJMCMC, przy czym prawdopodobieństwa modelu obliczane zostały tylko dla ostatnich 100 000 iteracji. Otrzymane wyniki przedstawia tab. 4.



Rys. 3. Wykres korelacyjny dla przykładu 2

Tabela 3. Rozkłady *a priori* dla przykładu 2

Stopień wielomianu k	Wektor m	Macierz V	α	β
1	$[-2,36 \ 0,34]^T$	$\begin{bmatrix} 0,48 & 0 \\ 0 & 0,06 \end{bmatrix}$	64	800
2	$[4,68 \ -8,19 \ 1,71]^T$	$\begin{bmatrix} 0,21 & 0 & 0 \\ 0 & 0,19 & 0 \\ 0 & 0 & 0,006 \end{bmatrix}$	44	111
3	$[1,85 \ -1,23 \ -1,79 \ 0,47]^T$	$\begin{bmatrix} 0,18 & 0 & 0 & 0 \\ 0 & 0,53 & 0 & 0 \\ 0 & 0 & 0,11 & 0 \\ 0 & 0 & 0 & 0,002 \end{bmatrix}$	49	58
4	$[1,34 \ -0,35 \ -2,6 \ 0,72 \ -0,02]^T$	$\begin{bmatrix} 0,27 & 0 & 0 & 0 & 0 \\ 0 & 1,96 & 0 & 0 & 0 \\ 0 & 0 & 1,28 & 0 & 0 \\ 0 & 0 & 0 & 0,11 & 0 \\ 0 & 0 & 0 & 0 & 0,001 \end{bmatrix}$	49	58

Źródło: opracowanie własne.

Tabela 4. Prawdopodobieństwa *a posteriori* rozpatrywanych modeli

Stopień wielomianu k	1	2	3	4
Obliczone prawdopodobieństwo	0	0	0,96934	0,03066

Źródło: opracowanie własne.

Modelem zdecydowanie najbardziej prawdopodobnym okazał się zatem wielomian stopnia 3 (co jest zgodne z modelem, na podstawie którego wygenerowano obserwacje).

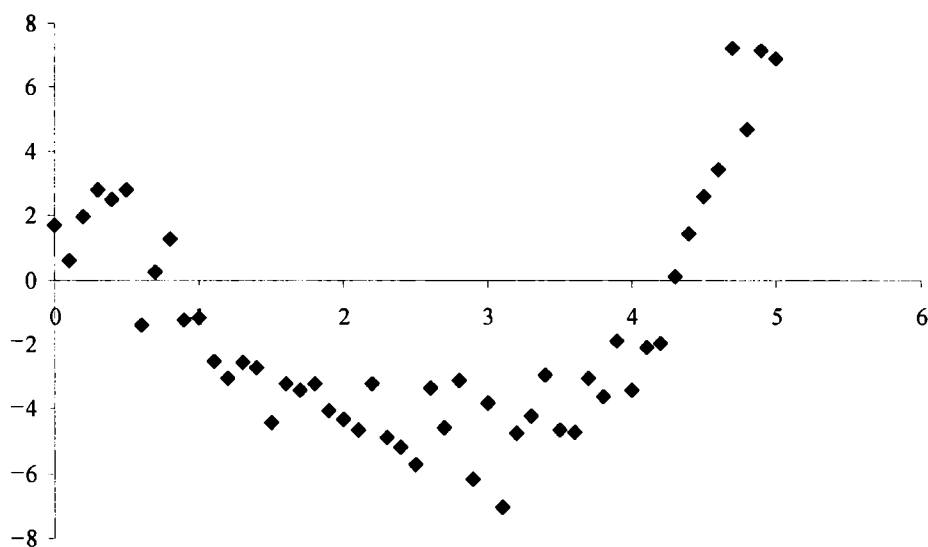
5.3. Przykład 3 – wpływ liczby danych na obliczane prawdopodobieństwa

W przykładzie zastosowano dane z punktu 5.2 po usunięciu z nich co drugiej obserwacji. Odpowiedni wykres przedstawia rys. 4.

Parametry rozkładów *a priori* przedstawia tab. 5. Zapisane zostały rozkłady tylko dla wielomianów stopnia nie większego niż 4, gdyż dla wielomianów wyższych stopni obliczone prawdopodobieństwa *a posteriori* są bliskie zera.

Przeprowadzono łącznie 200 000 iteracji algorytmu RJMCMC, przy czym prawdopodobieństwa modelu obliczane zostały tylko dla ostatnich 100 000 iteracji. Otrzymane wyniki przedstawia tab. 6.

Jak można zauważyć, zmniejszenie liczby obserwacji wprowadza znaczną niepewność co do poprawności wybieranego modelu. Najbardziej prawdopodobny okazuje się w dalszym ciągu wielomian 3 stopnia, jednak jego prawdopodobieństwo jest znacznie niższe niż w przykładzie poprzednim, w którym liczba danych była dwukrotnie większa.



Rys. 4. Wykres korelacyjny dla przykładu 3

Tabela 5. Rozkłady *a priori* dla przykładu 3

Stopień wielomianu k	Wektor $\mathbf{m}$	Macierz $\mathbf{V}$	α	β
1	$[-2,36 \ 0,32]^T$	$\begin{bmatrix} 0,96 & 0 \\ 0 & 0,12 \end{bmatrix}$	25	312,5
2	$[4,32 \ -7,85 \ 1,63]^T$	$\begin{bmatrix} 0,32 & 0 & 0 \\ 0 & 0,28 & 0 \\ 0 & 0 & 0,01 \end{bmatrix}$	28	59
3	$[2,3 \ -2,75 \ -0,94 \ 0,34]^T$	$\begin{bmatrix} 0,36 & 0 & 0 & 0 \\ 0 & 1,06 & 0 & 0 \\ 0 & 0 & 0,24 & 0 \\ 0 & 0 & 0 & 0,004 \end{bmatrix}$	25	33
4	$[2,59 \ -4,01 \ 0,22 \ -0,02 \ 0,04]^T$	$\begin{bmatrix} 0,52 & 0 & 0 & 0 & 0 \\ 0 & 4,16 & 0 & 0 & 0 \\ 0 & 0 & 2,76 & 0 & 0 \\ 0 & 0 & 0 & 0,24 & 0 \\ 0 & 0 & 0 & 0 & 0,003 \end{bmatrix}$	25	33

Źródło: opracowanie własne.

Tabela 6. Prawdopodobieństwa *a posteriori* rozpatrywanych modeli

Stopień wielomianu k	1	2	3	4
Obliczone prawdopodobieństwo	0	0,23531	0,76248	0,00221

Źródło: opracowanie własne.

6. Zakończenie

Bayesowskie podejście do zagadnienia wyboru modelu w odróżnieniu od podejścia klasycznego umożliwia obliczenie prawdopodobieństw konkurujących modeli. Dzięki temu dalsza analiza statystyczna pozwala uwzględnić niepewność związaną z ostatecznym wyborem modelu. Możliwe jest również wykorzystanie wszystkich modeli, dla których otrzymano niezerowe prawdopodobieństwa (tzw. model averaging).

Wadą opisaney metody jest niewątpliwie jej duża złożoność. W przypadku dużej liczby obserwacji pewnym ograniczeniem może być długi czas przeprowadzanych obliczeń.

Literatura

- [1] Aitkin M., *Posterior Bayes Factors*, „Journal of the Royal Statistical Society: Series B (Statistical Methodology)” 1991, vol. 53, s. 111-142.
- [2] Berger J.O., Pericchi L.R., *The Intrinsic Bayes Factor for Model Selection and Prediction*, „Journal of American Statistical Association” 1996, 91, s. 109-122.
- [3] Brooks S.P., Giudici P., Roberts G.O., *Efficient Construction of Reversible Jump Markov Chain Monte Carlo Proposal Distributions*, „Journal of the Royal Statistical Society: Series B (Statistical Methodology)” 2003, vol. 65, Issue 1.
- [4] Green P., *Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination*, „Biometrika” 1995, 82, 711-732.
- [5] Han C., Carlin B.P., *MCMC Methods for Computing Bayes Factors: a Comparative Review*, „Journal of American Statistical Association” 2001, 96, s. 1122-1132.
- [6] *Highly Structured Stochastic Systems*, red. P.J. Green, N. Hjort, S. Richardson, University Press, Oxford 2003.
- [7] O’Hagan., *Fractional Bayes Factors for Model Comparison*, „Journal of the Royal Statistical Society: Series B (Statistical Methodology)” 1995, vol. 57, s. 99-138.
- [8] Osiewalski J., *Bayesowska estymacja i predykcja dla jednorównaniowych modeli ekonometrycznych*, AE, Kraków 1991.
- [9] Osiewalski J., Pipień M., *Bayesian Comparison of Bivariate ARCH-Type Models for the Main Exchange Rates in Poland*, „Journal of Econometrics” 2004, 123.
- [10] Shafer G., *Lindley’s Paradox*, „Journal of American Statistical Association” 1982, vol. 77, s. 325-334.

BAYESIAN MODEL CHOICE IN MULTINOMIAL REGRESSION

Summary

The paper presents the problem of model choice in multinomial regression. The Bayesian solution to this problem has been presented in which the optimal model choice is equivalent to finding the model that is the most probable one. Computation of the model probabilities has been performed using the general algorithm of Reversible Jump Markov Chain Monte Carlo, which has been adapted to the specific problem of multinomial regression.

Dr Dariusz Biskup jest adiunktem w Katedrze Statystyki i Cybernetyki Ekonomicznej Akademii Ekonomicznej we Wrocławiu.