UNIWERSYTET JAGIELLOŃSKI
WYDZIAŁ MATEMATYKI I INFORMATYKI
INSTYTUT MATEMATYKI

# ESTYMATORY NAJWIĘKSZEJ WIARYGODNOŚCI W UOGÓLNIONYCH MODELACH REGRESJI NIELINIOWEJ

JERZY PIOTR RYDLEWSKI

*Praca doktorska napisana pod kierunkiem*
*dr. hab. Antoniego Leona Dawidowicza, prof. UJ*

Kraków 2009

PODZIĘKOWANIA

Serdecznie dziękuję mojemu Promotorowi prof. dr. hab. Antoniemu Leonowi Dawidowiczowi za cenne uwagi i zaproponowanie ciekawych problemów, których rozwiązanie stanowi niniejsza praca.

# Spis treści

# Wstęp

Poniższa praca rozważa pewien szczególny model regresji nieliniowej zastosowany w przypadku, gdy zmienna losowa ma rozkład beta lub gamma. Warto w tym miejscu wspomnieć, że podany model nie zawiera się w klasie jak dotąd rozpatrywanych modeli. Podobne do analizowanych modele nie są badane w obszernej monografii Wei (1998). Autor ten rozważał w swojej książce pewną rodzinę rozkładów wykładniczych. Ponieważ nasz model nie należy do rozpatrywanej przez Wei rodziny, otrzymane w tej monografii rezultaty nie mają zastosowania, na przykład, przy analizowaniu asymptotycznych bądź innych własności analizowanych w tej pracy modeli.

Rozprawę tę można też traktować jako rozwinięciwe rezultatów uzyskanych przez Wedderburna (1976), który podał pewne wyniki dotyczące zagadnienia gamma–regresji w uogólnionym modelu liniowym. Analizował jednoparametrową rodzinę rozkładów gamma z parametrem kształtu. Rozważany model jest uogólnieniem wyników Wedderburna, ze względu na jednoczesne estymowanie parametrów skali i kształtu.

Estymatory największej wiarygodności w dwuparametrowym modelu gamma regresji były także badane przez Bowmana i Shentona (1982, 1983, 1988, 2002). Rozpatrywali oni parametry kształtu i skali. Prezentowane wyniki są uogólnieniem ich wyników. Stąd wyniki rozdziału 2 można traktować jako uogólnienie rezultatów tych autorów.

Zagadnienie znalezienia estymatorów największej wiarygodności w modelach regresji nieliniowej, tak jak i bogata literatura odnosząca się do tego problemu znajdują się w monografii Sebera i Wilda (2003). Jednakże także i tu, jak również w przeglądowej książce Ratkowsky'ego (1990) nie znajdziemy modeli zaprezentowanych w rozdziale drugim i trzecim tej rozprawy.

W rozpatrywanym w pracy modelu rozpatrujemy zmienne losowe niezależne, ograniczone w pewnym, z góry zadanym przedziale. W obu przedstawianych przypadkach, to jest w przypadku rozkładu beta i gamma zmienne losowe charakteryzowane są przez pewne dwa parametry. Zastosowana metoda pozwala nam na założenie, że jeden z parametrów zadany jest przez sumę pewnej ustalonej liczby funkcji liniowo niezależnych. Następnie estymujemy, korzystając z metody największej wiarygodności, parametry charakteryzujące modelowany przez funkcje parametr oraz drugi z występujących w analizowanych rozkładach parametr.

W pracy wykazano, że w powyższych modelach beta i gamma regresji istnieje dokładnie jeden estymator największej wiarogodności rozpatrywanych parametrów.
Wykazano także, że otrzymane esymatory, jeżeli tylko spełniają dość naturalne założenia, to są zgodne oraz asymptotycznie normalne.
Uzyskane rezultaty zostały opublikowane (Rydlewski (2007, 2009)).
Pierwszy rozdział tej pracy zawiera wstęp do omawianego zagadnienia. Przypomniano w nim początki metody regresji. Omówiono ogólnie metodę najmniejszych kwadratów oraz pokrywającą się z nią w standardowym przypadku normalnym metodę największej wiarygodności. Przypomniano najważniejsze, klasyczne wyniki używane do badania asymptotycznych własności estymatorów największej wiarygodności w regresji nieliniowej.
W rozdziale drugim rozpatrzono model gamma–regresji. Udowodniono istnienie dokładnie jednego maksimum funkcji wiarygodności w tak zdefiniowanym modelu.
Rozdział trzeci zawiera analogiczne wyniki dla modelu beta–regresji. W ostatnim rozdziale czwartym analizowano najważniejsze z asymptotycznych własności badanych estymatorów. Podano założenia gwarantujące ich zgodność oraz asymptotyczną normalność.

# Rozdział 1

# Wprowadzenie - przegląd klasycznych podejść do regresji

## 1.1 Regresja nieliniowa

Jednym z ważnych zadań statystyki jest znalezienie możliwych zależności istniejących pomiędzy zmiennymi, z których przynajmniej jedna jest zmienną losową. Podlega ona pewnym losowym "fluktuacjom" oraz być może w dodatku jej pomiar obciążony jest błędem. Zwykle w problemach regresji jedna ze zmiennych (zmienna zależna lub zmienna objaśniana) oznaczana najczęściej przez $y$ jest "kształtowana" przez zmienne oznaczane najczęściej przez $x_1, .., x_k$ nazywane zmiennymi objaśniającymi (zmiennymi niezależnymi lub regresorami), które mają wyjaśnić zachowanie zmiennej $y$. Jeśli dane sugerują zależność pomiędzy $y$ i $x_1, .., x_k$ to chcemy określić tę relację za pomocą pewnej funkcji

$$y = f(x_1, .., x_k), \tag{1.1}$$

co pomoże przewidzieć $y$ dla danych $x_1, .., x_k$. Taka zależność nigdy nie będzie spełniona ze względu na niewytłumaczalne "fluktuacje" czy błędy pomiaru. Często znana jest postać matematyczna zależności (1.1), za wyjątkiem parametrów lub też zależność jest określona przez proces fizyczny czy pewne prawo. Innymi słowy (1.1) przybiera postać

$$y \approx f(x_1, .., x_k, \theta), \tag{1.2}$$

gdzie $\theta = (\theta_1, ..., \theta_m)$ jest parametrem, który należy estymować. Będziemy zajmować się przypadkiem, gdy znana jest $f$. Przypadek, gdy $f$ nie jest znana i dokonujemy estymacji w przestrzeni nieskończenie wymiarowej jest przedmiotem badań regresji nieparametrycznej. Nie będziemy go analizować w tej pracy. Zwykle, szczególnie analizując dane biologiczne, mamy małą wiedzę na temat leżących u ich podstaw procesów, dlatego też nie wiemy jaka ma być $f$ i naszym celem jest znalezienie funkcji, dla której (1.2) jest spełniona chociaż w przybliżeniu. Gdy wiele modeli pasuje do danych, to wybieramy model najprostszy. W regresji liniowej przyjmujemy, że zależność (1.2) ma postać

$$y \approx \theta_0 + \theta_1 x_1 + ... + \theta_{m-1} x_{m-1} \tag{1.3}$$

to znaczy funkcja $f$ jest liniowa ze względu na wektor parametrów, a błędy są addytywne, czyli rozpatrujemy model

$$y = \theta_0 + \theta_1 x_1 + ... + \theta_{m-1} x_{m-1} + \epsilon, \tag{1.4}$$

gdzie $E[\epsilon] = 0$. Regresja liniowa pozwala na rozpatrywanie wielu problemów. W dodatku, zagadnienie regresji liniowej jest bardzo dobrze opracowane, w przeciwieństwie do regresji nieliniowej. Ze względu na brak nie tylko jednolitej teorii, ale także wielu potrzebnych do analizy statystycznej twierdzeń, o ile tylko jest to możliwe, dokonuje się linearyzacji modelu nieliniowego. Modele nieliniowe używane są, gdy ich zastosowanie wynika z rozważań teoretycznych lub gdy sugeruje to nieliniowa zależność pomiędzy zmiennymi. Modelu nieliniowego można użyć nawet, gdy model liniowy dobrze określa zależności między zmiennymi. Wtedy może się zdarzyć, że to model nieliniowy służy czytelną interpretacją parametrów i to on zostanie zastosowany.

**Przykład 1.1.1** (Seber i Wild (2003))**.** *W pewnych zagadnieniach termodynamiki mówi się o procesie termodynamicznym w gazie, czyniącym zadość równaniu politropy. Dla danego gazu trzymanego w stałej temperaturze zależność pomiędzy objętością $v$ i ciśnieniem $p$ ma postać $pv^{\gamma} = c$. Przyjmując $y = p$ i $x = v^{-1}$ otrzymujemy $y = f(x; c, \gamma) \approx cx^{\gamma}$. Tu $\gamma$ jest stała dla każdego gazu, a naszym zadaniem jest ją wyestymować z danych. Ten model jest nieliniowy, ze względu na parametr $\gamma$. Logarytmując powyższą równość i kładąc $y = \ln p$ i $x = \ln v$ otrzymujemy*

$$y \approx \alpha + \beta x,$$

*gdzie $\alpha = \ln c$ oraz $\beta = -\gamma$. Teraz model jest liniowy ze względu na nieznane parametry $\alpha$ i $\beta$.*

Decyzja o ewentualnej linearyzacji zależy od tego czy umiemy jej dokonać, od problemu, którym się zajmujemy oraz od rodzaju błędów, o czym dalej jeszcze będzie mowa.

**Przykład 1.1.2** (Seber i Wild (2003))**.** *Podczas nieodwracalnej reakcji chemicznej substancja $A$ zmienia się w substancję $B$, która z kolei zmienia się w substancję $C$. Niech $A(t)$ oznacza ilość substancji $A$ w momencie $t$. Zmiany podczas reakcji określa równanie różniczkowe*

$$\frac{dA(t)}{dt} = -\theta_1 A(t)$$

$$\frac{dB(t)}{dt} = \theta_1 A(t) - \theta_2 B(t)$$

$$\frac{dC(t)}{dt} = \theta_2 B(t).$$

*Przyjmując* $A(0) = 1$ *otrzymujemy* $B(t)$

$$B(t) = \frac{\theta_1}{\theta_1 - \theta_2} \left(\exp(-\theta_2 t) - \exp(-\theta_1 t)\right).$$

Tym razem zależność pomiędzy $B$ i $t$ jest nieliniowa ze względu na $\theta_1$ i $\theta_2$. Nie da się zlinearyzować tego modelu żadnym przekształceniem. Tego typu modele nazywamy modelami istotnie nieliniowymi (*ang.* intrinsically nonlinear). Zauważmy, że w powyższych przykładach zakładaliśmy, że $x$ jest mierzony dokładnie, nie ulega żadnym fluktuacjom. Często musimy jednak zakładać, że $x$ także może być zmienną losową, a w dodatku zarówno $y$ jak i $x$ mogą być mierzone z błędem. Tego typu modele należy modelować w inny sposób.

## 1.2 Opis modeli

Opiszemy teraz pokrótce model z regresorami mierzonymi bez błędu z obarczoną addytywnym błędem zmienną zależną. Załóżmy, że znamy wartość oczekiwaną zmiennej zależnej, tymczasem ze względu na fluktuacje czy błędy pomiaru rzeczywista wartość zmiennej losowej to $y$. Mamy $E[y] = \mu$. Nasz model ma postać

$$y = f(x, \theta) + \epsilon,$$

gdzie $E[\epsilon] = E[y - \mu] = 0$. Zakładamy także, że znamy $x_1, x_2, ..., x_k$ dokładnie, czy też uznajemy nasz pomiar tych wartości za wystarczająco dokładny.
Można też uznać za zasadne, by traktować $x$ jak zmienną losową. Załóżmy, że znamy zależność funkcyjną pomiędzy realizacjami dwóch zmiennych

$$\mu = f(\psi, \theta). \tag{1.5}$$

Jednak obie mierzone są z błędami, tak więc obserwujemy $x = \psi + \delta$ oraz $y = \mu + \epsilon$, gdzie $E[\delta] = E[\epsilon] = 0$. Mamy

$$y = f(\psi, \theta) + \epsilon = f(x - \delta, \theta) + \epsilon$$

Korzystając z twierdzenia o wartości średniej, otrzymujemy

$$f(x - \delta, \theta) = f(x, \theta) - \delta f'(\overline{x}, \theta),$$

gdzie $\overline{x}$ znajduje się pomiędzy $x$ i $x - \delta$. Otrzymujemy teraz

$$y = f(x, \theta) - \delta f'(\overline{x}, \theta) + \epsilon = f(x, \theta) + \overline{\epsilon}.$$

Różnica pomiędzy tym modelem a poprzednim polega na traktowaniu $x$ jak zmiennej losowej, w dodatku zwykle $E[\overline{\epsilon}] \neq 0$. Trochę inny model otrzymamy, gdy potraktujemy (1.5) jako zależność między zmiennymi losowymi a nie ich realizacjami. W odróżnieniu od poprzedniego modelu otrzymujemy inną strukturę $\bar{\epsilon}$. Można jeszcze na inne sposoby modelować zależności między zmiennymi lub ich realizacjami (Seber i Wild (2003), Ratkowsky (1990)). Najważniejszym jest z pewnością uogólniony model liniowy (Nelder i Wedderburn (1972), McCullagh i Nelder(1983)).

$$E[y|x] = \mu_x$$

oraz

$$g(\mu_x) = \alpha + \beta^T x,$$

gdzie rozkład zmiennej losowej $y$ należy do rodziny wykładniczej rozkładów. Występującą w powyższym wzorze funkcję $g$ nazywamy funkcją łączącą (*ang.* link function). Procedura linearyzacji modelu polega na tym, że nieliniową zależność między $y$ i $x$ przekształcamy, w celu uzyskania zależności liniowej, otrzymania błędów, które w przybliżeniu mają rozkład normalny lub w celu uzyskania stałej wariancji błędu. Następnie korzysta się z rozbudowanego aparatu uogólnionego modelu liniowego. Zwykle dane przekształcenie linearyzujące pozwala tylko na osiągnięcie jednego z tych celów. Problem niestałej wariancji błędu można rozwiązać przez ważenie. W modelach liniowych bliski normalnemu rozkład błędów jest ważny, bo pozwala na wyciąganie wniosków z małych próbek. W modelach nieliniowych prawie wyłącznie polegamy na wnioskach uzyskanych przez stosowanie asymptotycznych rezultatów do skończonych próbek i nie możemy pozwolić sobie na "luksus" ignorowania modeli z błędami o asymptotycznie niegaussowskim rozkładzie. Często zdarza się, że parametry zlinearyzowanego modelu nie są interesujące tak, jak przed zlinearyzowaniem, na przykład nie mają żadnej fizycznej interpretacji. Oznacza to, że w podobnych przypadkach warto spróbować zastosować metody regresji nieliniowej. Poza tym, jak pokazuje Przykład 1.1.2, niektórych modeli nie da się zlinearyzować. Do takich właśnie modeli istotnie nieliniowych należą oba modele rozważane w rozdziałach 2 i 3.

## 1.3 Metoda najmniejszych kwadratów

Po raz pierwszy w publikacji metody tej użył A. M. Legendre w dodatku do swojej książki *"Nouvelles méthods pour la détermination des orbites des comètes"*. Książkę wraz z suplementem opublikowano w 1805 roku, tymczasem w wydanej w 1809 roku *"Theoria Motus"* G. F. Gauss stwierdził, że używa metody najmniejszych kwadratów od roku 1795. Nie rozstrzygając sporu o autorstwo metody, przejdźmy do jej opisu w przypadku nieliniowym.
Mamy n obserwacji $(x_i, y_i)$, $i = 1, 2, ..., n$, które pochodzą z modelu o regresorach mierzonych bez błędu. Zakładamy, że znamy nieliniową zależność funkcyjną $f$ między zmiennymi, a także addytywność błędu

$$y_i = f(x_i, \theta_0) + \epsilon_i,$$

gdzie $E[\epsilon_i] = 0$ oraz $\theta_0$ to prawdziwa wartość parametru $\theta$ należąca do $\Theta \subset \mathbb{R}^m$. Otrzymany metodą najmniejszych kwadratów estymator $\theta_0$ oznaczany przez $\widehat{\theta}_n$ minimalizuje dla $\theta \in \Theta$ sumę błędów kwadratowych

$$S^2(\theta) = \sum_{i=1}^{n} [y_i - f(x_i, \theta)]^2.$$

W przeciwieństwie do liniowej metody najmniejszych kwadratów $S(\theta)$ może posiadać wiele minimów lokalnych.

Zakładając, że $\epsilon_i$ mają taki sam rozkład i są wzajemnie niezależne oraz mają wariancję $\sigma^2$, przy pewnych założeniach o regularności, $\widehat{\theta}_n$ oraz $\widehat{s}_n^2 = \frac{S(\widehat{\theta}_n)}{n-m}$ są zgodnymi estymatorami $\theta_0$ i $\sigma^2$. Ponadto przy kolejnych założeniach, $\widehat{\theta}_n$ ma rozkład asymptotycznie normalny. Gdy dodatkowo założymy normalny rozkład błędów, to $\widehat{\theta}_n$ jest także estymatorem największej wiarygodności.

Gdy $f(x_i; \theta)$ są różniczkowalne ze względu na $\theta$ oraz $\widehat{\theta}_n \in int\ \Theta$, to mamy dla $j = 1, 2, ..., m$

$$\frac{\partial S(\theta)}{\partial \theta_j}|_{\widehat{\theta}_n} = 0. \tag{1.6}$$

Niech $f_i(\theta) = f(x_i; \theta)$, $f(\theta) = (f_1(\theta), ..., f_n(\theta))^T$ oraz

$$F'(\theta) = \frac{\partial f(\theta)}{\partial \theta^T} = \left[ \left( \frac{\partial f_i(\theta)}{\partial \theta_j} \right) \right]_{i,j}$$

$$F' = F'(\theta_0) \text{ i } \widehat{F}' = F'(\widehat{\theta}_n).$$

Otrzymujemy

$$S(\theta) = [y - f(\theta)]^T [y - f(\theta)] = \|y - f(\theta)\|^2.$$

Z (1.6) wnioskujemy

$$\sum_{i=1}^{n} (y_i - f_i(\theta)) \frac{\partial f_i(\theta)}{\partial \theta_j}|_{\theta = \widehat{\theta}_n} = 0 \text{ dla } j = 1, ..., m, \tag{1.7}$$

czyli

$$\widehat{F}'^T \left( y - f(\widehat{\theta}_n) \right) = \widehat{F}'^T \widehat{\epsilon} = 0.$$

Gdy

$$\widehat{P}_F = \widehat{F}' \left( \widehat{F}'^T \widehat{F}' \right)^{-1} \widehat{F}'^T$$

jest idempotentną macierzą będącą ortogonalną projekcją $\mathbb{R}^n$ na przestrzeń rozpiętą przez kolumny macierzy $\widehat{F}'$, to powyższą równość (1.7) można zapisać w postaci

$$\widehat{P}_F \widehat{\epsilon} = 0.$$

Są to, tak zwane, równania normalne dla modelu nieliniowego. Zwykle nie da się rozwiązać ich analitycznie, musimy więc korzystać z metod iteracyjnych.
Ważona metoda najmniejszych kwadratów, zwana też uogólnioną metodą najmniejszych kwadratów, stosowana jest, gdy nasz model ma postać

$$y = f(\theta) + \epsilon, \tag{1.8}$$

gdzie $E[\epsilon] = 0$ oraz $Var[\epsilon] = \sigma^2 V$, przy czym $\sigma$ jest znana, a także $V$ jest znaną macierzą dodatnio określoną. Naszym zadaniem jest teraz zminimalizowanie funkcji

$$S(\theta) = [y - f(\theta)]^T V^{-1} [y - f(\theta)].$$

Dokonując dekompozycji Cholesky'ego macierzy $V$ dostajemy $V = U^T U$, gdzie $U$ jest macierzą trójkątną górną. Mnożymy następnie obie strony równości (1.8) przez $B = (U^T)^{-1}$ i otrzymujemy

$$z = g(\theta) + \iota,$$

gdzie $z = By$, $g(\theta) = Bf(\theta)$ i $\iota = B\epsilon$. Mamy $E[\iota] = 0$ i $V(\iota) = \sigma^2 BVB^T = \sigma^2 I$, a ponadto

$$S(\theta) = [y - f(\theta)]^T V^{-1} [y - f(\theta)] = [y - f(\theta)]^T B^T B [y - f(\theta)] = [z - g(\theta)]^T [z - g(\theta)].$$

Teraz możemy nasz przekształcony problem rozwiązać metodami zwykłej metody najmniejszych kwadratów.
Przejdziemy teraz do krótkiego przeglądu asymptotycznych własności estymatorów otrzymanych metodą najmniejszych kwadratów.

- Z1. Niech $\epsilon_i$, $i = 1, 2, ..., n$ będą niezależnymi zmiennymi losowymi, które mają taki sam rozkład, $E[\epsilon_i] = 0$ oraz $V[\epsilon_i] = \sigma^2 > 0$.

- Z2. Funkcje $f_i(\theta)$, $i = 1, 2, ..., n$, są ciągłe ze względu na $\theta$, gdzie $\theta \in \Theta$.

- Z3. Zbiór $\Theta$ jest zwartym podzbiorem $\mathbb{R}^m$.

**Twierdzenie 1.3.1.** *(Jennrich (1969)) Przy założeniach Z1., Z2. i Z3. estymator najmniejszych kwadratów istnieje.*

Wypiszmy kolejne założenia

- Z4a. Połóżmy

$$A_n(\theta, \theta_1) = \sum_{i=1}^{n} f_i(\theta) f_i(\theta_1).$$

  Niech $\frac{1}{n} A_n(\theta, \theta_1)$ zmierza jednostajnie do $A(\theta, \theta_1)$ dla wszystkich $\theta, \theta_1 \in \Theta$.

- Z4b. Funkcja $D(\theta, \theta_0) = 0$ wtedy i tylko wtedy, gdy $\theta = \theta_0$, gdzie $D(\theta, \theta_0)$ to funkcja będąca jednostajną granicą ciągu funkcji $\frac{1}{n} D_n(\theta, \theta_1)$, gdzie

$$D_n(\theta, \theta_1) = \sum_{i=1}^{n} (f_i(\theta) - f_i(\theta_1))^2.$$

Istnienie granicy z warunku Z4b. wynika z Z4a.

- Z5. $D_n(\theta, \theta_1) \to \infty$, gdy $n \to \infty$ dla wszystkich $\theta, \theta_1 \in \Theta$ takich, że $\theta \neq \theta_1$.

- Z6. Każda $\epsilon_i$ jest ciągłą zmienną losową o wartościach w $\mathbb{R}$.

**Twierdzenie 1.3.2.** *(Jennrich (1969)) Jeśli Z1., Z2. i Z3. są spełnione, to Z4a. i Z4b. stanowią warunek wystarczający, aby $\widehat{\theta}_n$ i $\widehat{\sigma}_n^2 = \frac{S(\widehat{\theta}_n)}{n}$ były silnie zgodnymi estymatorami $\theta_0$ i $\sigma^2$.*

Jeśli w Z3. założymy ponadto, że $\Theta$ jest zbiorem skończonym, to jak pokazał Wu (1981) Z5. jest warunkiem wystarczającym na to, by $\widehat{\theta}_n$ był zgodnym estymatorem $\theta_0$. Przy założeniu Z6. to Z5. staje się warunkiem koniecznym i wystarczającym silnej zbieżności. Ponadto Wu udowodnił, że przy założeniach trochę słabszych od Z1., Z2., Z3., przy założeniu Z6. oraz przy założeniu, że $\epsilon_i$ mają skończoną informację Fishera warunek $D_n(\theta, \theta_1) \to \infty$ jest warunkiem koniecznym do istnienia jakiegokolwiek słabo zgodnego estymatora $\theta_0$.
Nie zawsze łatwo jest zrewidować podany wyżej zbiór warunków gwarantujących silną zbieżność, dlatego też istnieją alternatywne zbiory warunków, wygodniejsze do sprawdzenia przy różnych modelach. Można je znaleźć w Seber i Wild (2003), Crowder (1986), czy Malinvaud (1970a).
Podamy teraz warunki na asymptotyczną normalność nieliniowego estymatora najmniejszych kwadratów za Amemiya (1983). Inny zbiór warunków pochodzi od Jennricha (1969). Wypiszmy najpierw założenia.

- Z7. $\theta_0$ należy do wnętrza zbioru $\Theta$.

- Z8. Pochodne $\frac{\partial f_i(\theta)}{\partial \theta_r}$ oraz $\frac{\partial^2 f_i(\theta)}{\partial \theta_s \partial \theta_r}$ dla $i = 1, ..., n$ oraz $r, s = 1, ..., m$ istnieją oraz są ciągłe dla wszystkich $\theta \in \Theta_0$, gdzie $\Theta_0 \subset \Theta$ jest otwartym otoczeniem $\theta_0$.

- Z9. Wyrażenie $\frac{1}{n} \sum_{i=1}^{n} \frac{\partial f_i(\theta)}{\partial \theta} \frac{\partial f_i(\theta)}{\partial \theta^T}$ jest jednostajnie zbieżne w $\theta$ dla wszystkich $\theta \in \Theta_0$ do pewnej macierzy $\Omega(\theta)$.

- Z10. Wyrażenie $\frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial^2 f_i(\theta)}{\partial \theta_s \partial \theta_r}\right]^2$ jest jednostajnie zbieżne w $\theta$ dla wszystkich $\theta \in \Theta_0$ dla $r, s = 1, ..., m$.

- Z11. Macierz $\Omega = \Omega(\theta_0)$ jest nieosobliwa.

**Twierdzenie 1.3.3.** *Przy założeniach Z1.-Z4. oraz Z7.-Z11. estymator najmniejszych kwadratów jest asymptotycznie normalny*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \sim N_m(0, \sigma^2 \Omega^{-1})$$

*oraz* $\frac{1}{n} F'(\widehat{\theta}_n)^T F'(\widehat{\theta}_n)$ *jest silnie zgodnym estymatorem* $\Omega$. *Innymi słowy estymator* $\widehat{\theta}_n$ *ma dla dużych* $n$ *w przybliżeniu rozkład postaci*

$$\widehat{\theta}_n \sim N_m\left(\theta_0, \sigma^2 \left(F'(\theta_0)^T F'(\theta_0)\right)^{-1}\right).$$

Trochę inny zestaw na asymptotyczną normalność podał Jennrich (1969), który zastąpił warunki Z9. i Z10. warunkiem Z12.

- Z12. Wyrażenia $\frac{1}{n}\sum_{i=1}^{n} g_i(\theta) h_i(\theta_1)$, są jednostajnie zbieżne dla wszystkich $\theta, \theta_1 \in \Theta$, gdzie za $g_i(\theta)$, $h_i(\theta_1)$ podstawiamy każdą z funkcji $f_i(\theta)$, $\frac{\partial f_i(\theta)}{\partial \theta_r}$ oraz $\frac{\partial^2 f_i(\theta)}{\partial \theta_s \partial \theta_r}$ dla $r, s = 1, ..., m$.

Estymator $\widehat{\theta}_n$ chociaż asymptotycznie normalny, nie jest jednak zwykle asymptotycznie efektywny. Asymptotyczną efektywność uzyskamy, gdy dodatkowo założymy normalny rozkład błędów.
Inne warunki na asymptotyczną normalność zawarte są w pracy Malinvaud (1970b) oraz Gallant (1987, Rozdział 4) oraz White i Domowitz (1984). Podsumowanie tych i innych rezultatów zawiera Bunke i inni (1977), Bunke i Schmidt (1980) oraz Zwanzig (1980) i Amemiya (1983), w których to pracach dopracowano przypadek ważonych najmniejszych kwadratów oraz nieidentycznego rozkładu błędów (White (1980)). Zupełnie inne podejście zaprezentował Johansen (1983), który ustalił $n$ i zmierzał do zera z $\sigma^2$.

## 1.4 Metoda największej wiarygodności

Metodę największej wiarygodności po raz pierwszy zaproponował Fisher (1922). Estymatorem największej wiarygodności parametru $\theta$ nazywamy wartość parametru, w której funkcja wiarygodności przyjmuje wartość największą

$$\widehat{\theta}_n = \arg\max_{\theta \in \Theta} L_n(x; \theta).$$

Niech łączny rozkład zmiennych losowych $\epsilon_i$ w modelu regresji o stałych regresorach jest znany i taki sam. Oznaczmy funkcję gęstości przez $\frac{1}{\sigma} g\left(\frac{\epsilon}{\sigma}\right)$. Niech $\epsilon_i$ będą ponadto wzajemnie niezależne. Funkcja wiarygodności przyjmuje wtedy postać

$$L_n(y|\theta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma} g\left(\frac{y_i - f(x_i, \theta)}{\sigma}\right).$$

Zostaną teraz pokrótce omówione dwa przypadki. Pierwszy ma miejsce, gdy błędy mają rozkład normalny, a drugi gdy każdy inny.
Gdy błędy mają rozkład normalny $\epsilon_i \sim N(0, \sigma^2)$, to

$$L_n(y|\theta, \sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \frac{[y_i - f(x_i, \theta)]^2}{\sigma^2}\right).$$

Można teraz łatwo pokazać, że otrzymane metodą najmniejszych kwadratów estymatory $\widehat{\theta}_n$ i $\widehat{\sigma}_n^2 = \frac{S(\widehat{\theta}_n^2)}{n}$ okazują się być poszukiwanymi estymatorami największej wiarygodności parametrów $\theta$ i $\sigma^2$.
W przypadku, gdy $y_i$ mają różny rozkład, ze względu na różne $\epsilon_i$, nie możemy stosować standardowych metod dla estymatorów największej wiarygodności.
Gdy rozkład błędów nie jest normalny, to nie możemy skorzystać z wyników otrzymanych dla metody najmniejszych kwadratów. Rozpatrzymy teraz przypadek, gdy logarytm funkcji wiarygodności jest dowolną funkcją wartości oczekiwanej zmiennej zależnej, tzn. $\ln L_n = \ln L_n(f(\theta))$, gdzie $E[y] = f(\theta)$.
Poszukujemy rozwiązań równości

$$\frac{\partial \ln L_n}{\partial \theta} = 0.$$

Tym razem w ogólnym przypadku zwykle nie znajdziemy analitycznego rozwiązania. W związku z tym jesteśmy zdani na metody iteracyjne. Poza tym, być może rozwiązanie problemu nie istnieje albo istnieje ich wiele. Zwykle używa się do tego celu, tj. znalezienia rozwiązania, metody Newtona, zwykle w literaturze statystycznej w zastosowaniu do tego problemu, nazywanej metodą Newtona-Raphsona. W efekcie w zasadzie dla każdego modelu z osobna należy badać własności asymptotyczne estymatora największej wiarygodności. Gdy rozkład błędów nie jest normalny, lecz funkcja wiarygodności jest znaną funkcją $E[y]$, to na przykład takie metody, jak ważona metoda najmniejszych kwadratów także mogą zostać użyte do znalezienia estymatora. W takim przypadku jednak nie istnieją ogólne metody sprawdzenia, czy estymator największej wiarygodności istnieje, czy istniejący jest jedyny, czy jest zgodny i asymptotycznie normalny.
Dodajmy, że pewne szczególne przypadki jedyności estymatora największej wiarygodności rozważali Cox i Hinkley (1974), Wedderburn (1976), Pratt (1981), Burridge (1981).
W literaturze przedmiotu spotykamy dwa podejścia do zagadnienie zgodności estymatora największej wiarygodności. Pierwsze bazuje na pracy Walda (1949), a drugie na wynikach podanych w książce Craméra (1946). Oba wyniki dotyczyły przypadku zmiennych losowych niezależnych o takim samym rozkładzie. Rezultaty Craméra odnoszą się do zgodności pewnego rozwiązania równań wiarygodności tzn. niejako do lokalnego estymatora największej wiarygodności. Takie rozwiązanie może nie mieć wiele wspólnego z estymatorem największej wiarygodności, jako do globalnego maksimum funkcji wiarygodności.

Podejście Walda prowadzi do uzyskania zgodności globalnego ENW (estymator największej wiarygodności) przy pewnych założeniach, które w skrócie, nie nakładają warunków na pochodne funkcji wiarygodności, ale zakładają pewne asymptotyczne własności funkcji wiarygodności, zwykle niezbyt łatwe do zweryfikowania. W dodatku, często zakłada się zwartość przestrzeni parametrów, zapewniającą, przy braku innych możliwości postępowania, istnienie ENW. Założenie to nie jest w wielu przypadkach naturalne i wygodne.
Do podejścia Craméra, prowadzącego w naturalny sposób wprost do uzyskania asymptotycznej normalności, nawiązują Bradley i Gart (1962), którzy rozważali ogólny przypadek, gdy zmienne co prawda są wzajemnie niezależne, ale mają różne rozkłady. Uzyskali oni słabą zgodność ENW. Weiss (1971,1973) podał warunki na słabą zgodność ENW, gdy zmienne nie muszą ani być niezależne, ani mieć takiego samego rozkładu. Inne warunki na słabą i silną zgodność oraz asymptotyczną normalność ENW podali Crowder (1976,1986), Amemiya (1977) dla przypadku błędów normalnych oraz bez tego ograniczenia Fahrmeir i Kaufmann (1985) i Fahrmeir (1987). Za Waldem Hoadley (1971) rozszerzył dowód zgodności do przypadku zmiennych niezależnych niekoniecznie o takim samym rozkładzie i podał dowód słabej zgodności. Podobnie Heijmans i Magnus (1986) podali warunki na słabą i silną zgodność estymatora NW oraz na jego asymptotyczną normalność.
Podane przez autorów prac warunki są, niestety, w praktyce trudne do sprawdzenia, co zauważali autorzy prac analizujący konkretne modele. Stąd tyle warunków na zgodność i asymptotyczną normalność estymatora największej wiarygodności. Często, czego dowodzą te i inne nie wymienione prace, wygodniej jest w każdym modelu próbować rozwiązać problem zgodności i normalności estymatora największej wiarygodności korzystając ze specyficznych własności badanego zagadnienia (patrz: na przykład Nie (2006)). W rozdziale 4 tej pracy oparto się na metodzie z książki Wei (1998) nawiązując jednakże do wyników Fahrmeira i Kaufmanna (1985). Podamy teraz jako przykład, za pracą Heijmansa i Magnusa (1986), warunki na zgodność i asymptotyczną normalność ENW.

- A1. Niech zbiór parametrów $\Theta \subset \mathbb{R}^m$ będzie zwarty.
- A2. Niech dla dowolnego $n \in \mathbb{N}$, $y \in \mathbb{R}^n$ funkcja wiarogodności $L_n(\theta, y)$ będzie ciągła na $\Theta$.
- A3. Dla dowolnego $\theta \neq \theta_0$ istnieje $N(\theta)$ otoczenie $\theta$ takie,że

$$\lim_{n\to\infty} P\left[\sup_{\phi\in N(\theta)} \left(\ln L_n(\phi) - \ln L_n(\theta_0)\right) < 0\right] = 1.$$

- A3'. Dla dowolnego $\theta \neq \theta_0$ istnieje $N(\theta)$ otoczenie $\theta$ takie,że

$$\limsup_{n\to\infty} \sup_{\phi\in N(\theta)} \left(\ln L_n(\phi) - \ln L_n(\theta_0)\right) < 0. \quad \text{p.w. P.}$$

Jeżeli zachodzą A1. i A2., to estymator NW istnieje, choć nie musi być jedynym.

**Twierdzenie 1.4.1.** *Jeżeli zachodzi A1. i A2., to A3. jest warunkiem koniecznym i wystarczającym na to, by ciąg* $\{\theta_n\}_{n\in\mathbb{N}}$ *był słabo zgodnym estymatorem* $\theta_0$.

**Twierdzenie 1.4.2.** *Jeżeli zachodzi A1. i A2., to A3'. jest warunkiem koniecznym i wystarczającym na to, by ciąg* $\{\theta_n\}_{n\in\mathbb{N}}$ *był mocno zgodnym estymatorem* $\theta_0$.

Ponieważ autorzy sami uznali, że w praktyce weryfikowanie podanych warunków jest kłopotliwe, to podali zestaw założeń trochę mocniejszych od A3., które można łatwiej sprawdzić. Wypiszmy kolejne założenia

- B3. Dla dowolnego $\theta \in \Theta, \theta \neq \theta_0$ istnieje ciąg nielosowych nieujemych liczb $\left\{k_n\left(\theta, \theta_0\right)\right\}_n$ taki, że
  (i) $\liminf_{n\to\infty} k_n\left(\theta, \theta_0\right) > 0$
  (ii) $P\left[\lim_{n\to\infty} \frac{1}{k_n(\theta,\theta_0)}\left(\ln L_n(\theta) - \ln L_n(\theta_0)\right) = -1\right] = 1.$
- B4. Dla dowolnego $\theta \neq \theta_0 \in \Theta$ istnieje otoczenie $N(\theta)$ takie, że

$$\lim_{n\to\infty} P\left[\frac{1}{k_n\left(\theta, \theta_0\right)} \sup_{\phi\in N(\theta)} \left(\ln L_n(\phi) - \ln L_n(\theta)\right) < 1\right] = 1.$$

**Twierdzenie 1.4.3.** *Jeżeli zachodzą A1., A2., B3. i B4.,to każdy ciąg* $\{\theta_n\}_{n\in\mathbb{N}}$ *jest słabo zgodnym estymatorem* $\theta_0$.

# Rozdział 2

# Model gamma regresji

W rozdziale tym rozważono model regresji nieliniowej, w którym zmienna zależna przyjmuje rozkład gamma. Analizowany jest model, w którym parametr kształtu zmiennej losowej o rozkładzie gamma jest sumą $m$ dowolnych ciągłych funkcji liniowo niezależnych. W szczególności sumą funkcji okresowej, której parametrami są jej współczynniki Fouriera oraz jednoparametrowej ciągłej funkcji trendu. Wykazano, że istnieje dokładnie jedno maksimum funkcji wiarygodności dla tego modelu gamma–regresji. Jak zostanie pokazane w rozdziale 4, estymator największej wiarygodności jest asymptotycznie normalny i zgodny.

## 2.1 Wprowadzenie do zagadnienia gamma–regresji

W klasycznych modelach regresji przyjmowana jest zależność

$$x_i = at_i + b + \varepsilon_i,$$

gdzie zmienne losowe $\varepsilon_i$ są niezależne i mają rozkład normalny o średniej równej zeru. Nieco ogólniejsza postać tego modelu wyraża się wzorem

$$x_i = \varphi(t_i, \theta) + \varepsilon_i,$$

gdzie $\varphi(\cdot, \theta)$ jest funkcją zależną od estymowanego parametru $\theta$. Model ten można zapisać również, przyjmując, że zmienna losowa $x_i$ ma rozkład normalny postaci $N(\varphi(t_i, \theta), \sigma^2)$.
W pewnych sytuacjach nienaturalne jest założenie o przyjmowaniu przez zmienną dowolnych wartości rzeczywistych. Może się zdarzyć, że zmienna przyjmuje wyłącznie wartości należące do pewnego przedziału (Cribari–Neto (2002), Rydlewski (2007)) bądź wyłącznie wartości dodatnie. W tym rozdziale pracy zajmiemy się tą ostatnią sytuacją (za Rydlewski (2009)).
Praca Wedderburna (1976) zawiera, między innymi, pewne wyniki dotyczące zagadnienia gamma–regresji w uogólnionym modelu liniowym. Autor analizuje jednoparametrową rodzinę rozkładów gamma, gdzie parametrem jest parametr kształtu. Rozważany niżej model

można uznać za uogólnienie wyników Wedderburna, ze względu na jednoczesne estymowanie parametru skali. Dodatkowo, parametr kształtu jest w postaci funkcji, o czym niżej.
Estymatory największej wiarygodności dla dwuparametrowego modelu gamma regresji były rozpatrywane przez Bowmana i Shentona. Rozpatrywanymi przez autorów parametrami były parametry kształtu i skali. Prezentowane wyniki są uogólnieniem ich wyników, jako że dopuszczamy, by parametr kształtu był modelowany przez funkcję.
Prezentowane rezultaty nie pokrywają się z wynikami omówionymi w książce Wei, który rozpatrywał pewną rodzinę rozkładów wykładniczych. Nasz model nie należy do rozpatrywanej przez niego rodziny i nie da się przekształcić do tej postaci (Wei (1998) str. 2–3).
Zagadnienie znalezienia estymatorów największej wiarygodności w modelach regresji nieliniowej, tak jak i bogata literatura odnosząca się do tego problemu znajdują się w Seber i Wild (2003).
Zakładamy, iż zmienna losowa $x_i$ ma rozkład o gęstości postaci $f(\varphi(\theta_1, t_i), \theta_2)$, gdzie $\varphi(\theta_1, t_i)$ to średnia rozkładu. Rozpatrywany jest model gamma-regresji, to znaczy zmienna losowa $x_i$ ma rozkład postaci $\gamma(p, r)$. Funkcja gęstości zmiennej losowej $x_i$ o rozkładzie gamma jest postaci

$$f(t, p, r) = \frac{r^p}{\Gamma(p)} t^{p-1} e^{-rt}, \qquad t > 0,$$

gdzie $r > 0$ jest parametrem skali, $p > 0$ jest parametrem kształtu oraz $\Gamma(\cdot)$ to funkcja gamma. Wartość oczekiwana zmiennej losowej X o rozkładzie gamma wynosi $E(X) = \frac{p}{r}$, a wariancja $Var(X) = \frac{p}{r^2}$. W naszym modelu za $p$ wstawiamy funkcję zmiennej $t_i$ zależną od wielowymiarowego parametru $A$. Dokładniej, możemy użyć dowolnego układu $m$ ciągłych funkcji liniowo niezależnych. Mamy wtedy

$$p(A, t) = \sum_{k=1}^{m} A_k f_k(t).$$

W szczególności, ze względu na zastosowania, rozpatrzony jest model, gdzie $p$ jest funkcją okresową, a parametrami są jej współczynniki Fouriera. Do tego modelu można dorzucić również jednoparametrową ciągłą funkcję $\tau(\cdot)$ opisującą trend. Innymi słowy, naszym układem funkcji jest

$$(\tau(\cdot), \sin(\cdot), \sin 2(\cdot), ..., \sin m(\cdot), 1, \cos(\cdot), \cos 2(\cdot), ..., \cos m(\cdot)) .$$

Kładziemy $A = (A_0, A_1, ..., A_m, B_0, B_1, ..., B_m)$ i parametr kształtu $p$ ma postać funkcji

$$p(A, t) = A_0 \tau(t) + B_0 + \sum_{k=1}^{m} (A_k \sin kt + B_k \cos kt) .$$

Niech $x_1, x_2, ..., x_n$ będą niezależnymi zmiennymi losowymi o rozkładzie gamma. W wyniku poczynionych założeń otrzymujemy, że wartości oczekiwane zmiennych losowych $x_j$ mają postać

$$E(x_j) = \sum_{k=1}^{m} \alpha_k f_k(t_j), \qquad j = 1, 2, ..., n,$$

gdzie $\alpha_k = \frac{A_k}{r}$, $k = 1, 2, ..., m$.
W wyróżnionym przez nas przypadku mamy

$$E(x_j) = \alpha_0 \tau(t_j) + \beta_0 + \sum_{k=1}^{m} \left( \alpha_k \sin kt_j + \beta_k \cos kt_j \right), \qquad j = 1, 2, ..., n,$$

gdzie $\alpha_k = \frac{A_k}{r}$ oraz $\beta_k = \frac{B_k}{r}$.
Do estymowania parametrów użyjemy metody największej wiarygodności. Funkcja wiarygodności w naszym modelu gamma–regresji ma postać

$$L(x_1, x_2, ..., x_n, t_1, t_2, ..., t_n, A, r) =$$

$$= \prod_{j=1}^{n} \frac{1}{\Gamma(p(A, t_j))} r^{p(A,t_j)} x_j^{p(A,t_j)-1} e^{-rx_j}, \tag{2.1}$$

a stąd logarytm funkcji wiarygodności to

$$\ln L = \sum_{j=1}^{n} \left( -\ln \Gamma(p(A, t_j)) + p(A, t_j) \ln r + (p(A, t_j) - 1) \ln x_j - rx_j \right). \tag{2.2}$$

Zapiszmy inaczej nasze parametry, korzystając z własności rozkładu gamma. Niech będzie $a = (\alpha_1, \alpha_2, ..., \alpha_m)$. Wówczas

$$\varphi(a, t_j) = \frac{p(A, t_j)}{r} = \sum_{k=1}^{m} \alpha_k f_k(t_j)$$

W wyróżnionym przypadku $a = (\alpha_0, \alpha_1, ..., \alpha_m, \beta_{0,} \beta_1, ..., \beta_m)$ i otrzymujemy

$$\varphi(a, t_j) = \frac{p(A, t_j)}{r} = \alpha_0 \tau(t_j) + \beta_0 + \sum_{k=1}^{m} \left( \alpha_k \sin kt_j + \beta_k \cos kt_j \right).$$

Oczywiście jest

$$\ln L(ra, r) = \sum_{j=1}^{n} \ln L_j(ra, r)$$

gdzie

$$\ln L_j(ra, r) = -\ln \Gamma \left( r\varphi(a, t_j) \right) + r\varphi(a, t_j) \ln r + \left( r\varphi(a, t_j) - 1 \right) \ln x_j - rx_j.$$

W takim razie wartość oczekiwana zmiennej $x_j$ jest postaci $E(x_j) = \frac{p}{r} = \varphi(a, t_j)$, gdzie $|\varphi(a, t_j)| < M$ dla pewnego ustalonego $M > 0$, $j = 1, 2, ..., n$.
Zauważmy, że przyjęte przez nas ograniczenie nie przeszkadza w praktycznym korzystaniu z otrzymanych rezultatów. Wynika to z faktu, że zawsze dysponujemy skończoną liczbą danych oraz należą one do przedziału, którego długość można odpowiednio ustalić na potrzeby danych, rozpatrywanych w konkretnym zagadnieniu.
W dalszej części pracy wykażemy, że otrzymane metodą największej wiarygodności estymatory parametrów są wyznaczone jednoznacznie.

## 2.2 Estymacja metodą największej wiarygodności

**Lemat 2.2.1.** *Niech $[c, d]$ będzie przedziałem domkniętym i ograniczonym. Niech $f_1, f_2, ..., f_m$ będzie dowolnym układem funkcji liniowo niezależnych ciągłych na $[c, d]$. Zbiór wszystkich parametrów $\boldsymbol{A}$ postaci $(A_1, A_2, ..., A_m, r) \in \mathbb{R}^{m+1}$ takich, że dla dowolnego $t \in [c, d]$,*

$$0 \leq \sum_{k=1}^{m} A_k f_k(t) \leq Mr \tag{2.3}$$

*jest niepusty, domknięty i wypukły w $\mathbb{R}^{m+1}$.*

Dowód. To, że zbiór jest $\mathbf{A}$ niepusty, jest oczywiste. Niech $\mathbf{A}_t$ będzie zbiorem $(A_1, A_1, ..., A_m, r) \in \mathbb{R}^{m+1}$ spełniających warunek (2.3) dla ustalonego $t$. Zbiór

$$\bigcap_{t \in [c,d]} A_t$$

jest domknięty jako przecięcie zbiorów domkniętych.
Niech $\mathbb{A}_1 = (A_{1,1}, A_{1,2}, ..., A_{1,m}, r_1) \in \mathbf{A}$ oraz $\mathbb{A}_2 = (A_{2,1}, A_{2,2}, ..., A_{2,m}, r_2) \in \mathbf{A}$ będą dwoma wektorami oraz niech $\lambda \in [0, 1]$. Łatwo można pokazać, że $\lambda \mathbb{A}_1 + (1 - \lambda)\mathbb{A}_2 \in \mathbf{A}$. □

**Lemat 2.2.2.** *Niech $f_1, f_2, ..., f_m : \mathbb{R} \longrightarrow \mathbb{R}$ będzie układem funkcji ciągłych i liniowo niezależnych. Zbiór $\mathbf{a}$ wszystkich parametrów postaci $a = (\alpha_1, \alpha_2, ..., \alpha_m) \in \mathbb{R}^m$ spełniających nierówność*

$$0 \leq \sum_{k=1}^{m} \alpha_k f_k(t) \leq M \tag{2.4}$$

*dla dowolnego $t \in \mathbb{R}$ jest niepusty i zwarty w $\mathbb{R}^m$.*

Dowód. Oczywiste jest, że zbiór $\mathbf{a}$ jest niepusty i domknięty. Udowodnimy jego ograniczoność.

Dla dowodu załóżmy, że zbiór **a** jest nieograniczony. Istnieje więc wybrany ze zbioru **a** ciąg $\{(\alpha_1^n, \alpha_2^n, ..., \alpha_m^n)\}_{n\in\mathbb{N}}$ taki, że zachodzi

$$(\alpha_1^n, \alpha_2^n, ..., \alpha_m^n) \longrightarrow +\infty, \text{ gdy } n \to +\infty.$$

Mamy

$$\frac{(\alpha_1^n, \alpha_2^n, ..., \alpha_m^n)}{||\,(\alpha_1^n, \alpha_2^n, ..., \alpha_m^n)\,||} \in S^{m-1},$$

gdzie $S^{m-1}$ jest jednostkową sferą $m-$wymiarową. Wybierając z tak zdefiniowanego ciągu podciąg

$$\left\{\frac{(\alpha_1^{l_n}, \alpha_2^{l_n}, ..., \alpha_m^{l_n})}{||\,(\alpha_1^{l_n}, \alpha_2^{l_n}, ..., \alpha_m^{l_n})\,||}\right\}_{n\in\mathbb{N}}$$

zbieżny do pewnego $(\alpha_1^0, \alpha_2^0, ..., \alpha_m^0)$, otrzymujemy

$$\forall t \in \mathbb{R} \qquad 0 \le \sum_{k=1}^{m} \alpha_k^{l_n} f_k(t) \le M.$$

Dalej mamy

$$0 \le \frac{\sum_{k=1}^m \alpha_k^{l_n} f_k(t)}{||\,(\alpha_1^n, \alpha_2^n, ..., \alpha_m^n)\,||} \le \frac{M}{||\,(\alpha_1^n, \alpha_2^n, ..., \alpha_m^n)\,||}.$$

Zmierzając z $n$ do nieskończoności, dostajemy

$$\sum_{k=1}^{m} \alpha_k^0 f_k(t) = 0.$$

Dobierając odpowiednie $t_1, t_2, ..., t_m$, otrzymujemy $\alpha_k^0 = 0$ dla $k = 1, 2, ..., m$, co przeczy faktom, bowiem $(\alpha_1^0, \alpha_2^0, ..., \alpha_m^0) \in S^{m-1}$. □

**Lemat 2.2.3.** *Prawdziwy jest dokładnie jeden z poniższych warunków*

*(i) Dla wszystkich* $j = 1, 2, ..., n$

$$x_j = \sum_{k=1}^{m} \alpha_k f_k(t_j)$$

*(ii)*

$$\lim_{r\to\infty} \frac{d}{dr} \ln L(ra, r) < 0.$$

Dowód. Niech $y_j = \varphi(a, t_j)$. Przypomnijmy, że mamy

$$\ln L_j(ra, r) = -\ln\Gamma\left(r\varphi(a, t_j)\right) + r\varphi(a, t_j)\ln r + (r\varphi(a, t_j) - 1)\ln x_j - rx_j.$$

Otrzymujemy

$$\frac{d}{dr}\ln L_j(ra, r) = \frac{d}{dr}\left(-\ln\Gamma\left(ry_j\right) + ry_j\ln r + (ry_j - 1)\ln x_j - rx_j\right) =$$

$$= -y_j\Psi\left(ry_j\right) + y_j\ln r + y_j + y_j\ln x_j - x_j,$$

gdzie $\Psi(x) = \frac{d}{dx}\ln\Gamma(x)$. Mamy

$$\lim_{x\to\infty}\left(\Psi(x) - \ln x\right) = 0, \tag{2.5}$$

a stąd

$$\lim_{r\to\infty}\frac{d}{dr}\ln L_j(ra, r) = -y_j\ln y_j + y_j + y_j\ln x_j - x_j.$$

Niech

$$h(x) = -y_j\ln x + x.$$

Funkcja $h$ przyjmuje najmniejszą wartość w $x = y_j$. Stąd wnioskujemy, że

$$\lim_{r\to\infty}\sum_{j=1}^{n}\frac{d}{dr}\ln L_j(ra, r) \le 0$$

i jeżeli dla przynajmniej jednego $j$ mamy $x_j \neq y_j$, to otrzymujemy

$$\lim_{r\to\infty}\sum_{j=1}^{n}\frac{d}{dr}\ln L_j(ra, r) < 0.$$

□

**Lemat 2.2.4.** *Funkcja* $\ln L(ra, r)$ *jako funkcja parametru* $r$ *jest silnie wklęsła.*

Dowód. Mamy

$$\ln L(ra, r) = \sum_{j=1}^{n}\left(-\ln\Gamma(r\varphi(a, t_j)) + r\varphi(a, t_j)\ln r + (r\varphi(a, t_j) - 1)\ln x_j - rx_j\right).$$

Stąd

$$\frac{d}{dr}\ln L(ra, r) = \sum_{j=1}^{n}\left(-\frac{d}{dr}(\ln\Gamma(r\varphi(a, t_j))) + \varphi(a, t_j)\ln r + \varphi(a, t_j)(1 + \ln x_j) - x_j\right)$$

oraz

$$\frac{d^2}{dr^2}\ln L(ra,r) = \sum_{j=1}^{n}\left(-\frac{d^2}{dr^2}(\ln\Gamma(r\varphi(a,t_j))) + \frac{\varphi(a,t_j)}{r}\right) =$$

$$= \sum_{j=1}^{n}\left(-\left(\varphi(a,t_j)\right)^2\Psi'(r\varphi(a,t_j)) + \frac{\varphi(a,t_j)}{r}\right).$$

Wystarczy sprawdzić, czy

$$\Psi'(r\varphi(a,t_j)) > \frac{1}{r\varphi(a,t_j)}, \qquad j = 1,...,n.$$

Wiadomo, że

$$\Psi'(y) = \sum_{n=0}^{\infty}\frac{1}{(y+n)^2}$$

oraz

$$\sum_{n=0}^{\infty}\frac{1}{(y+n)^2} > \int_0^{\infty}\frac{ds}{(y+s)^2} = \frac{1}{y},$$

co dowodzi prawdziwości lematu. □

Niech rzeczywista macierz $J \in \mathbb{R}^{n\times m}$ będzie postaci

$$\begin{bmatrix} f_1(t_1) & f_2(t_1) & ... & ... & ... & f_m(t_1) \\ f_1(t_2) & f_2(t_2) & ... & ... & ... & f_m(t_2) \\ ... & ... & ... & & & ... \\ ... & ... & & ... & & ... \\ ... & ... & & & ... & ... \\ f_1(t_n) & f_2(t_n) & ... & ... & ... & f_m(t_n) \end{bmatrix}. \tag{2.6}$$

**Lemat 2.2.5.** *Jeżeli liczba naszych obserwacji n jest odpowiednio duża w stosunku do liczby zakładanych parametrów, dokładniej $n \geq m$ oraz rząd macierzy J jest maksymalny (inaczej rzJ=m), to funkcja $\ln L$ jako funkcja parametrów $(A_1, A_2, ..., A_m, r)$ jest silnie wklęsła.*

Dowód. Mamy

$$\ln L = \sum_{j=1}^{n}\ln L_j.$$

Niech

$$F(x,r) = -\ln\Gamma(x) + x\ln r.$$

Hesjan funkcji $F$ jest postaci

$$H_F = \begin{bmatrix} -\frac{x}{r^2} & \frac{1}{r} \\ \frac{1}{r} & -\Psi'(x) \end{bmatrix}.$$

Korzystając z faktu, że $x\Psi'(x) > 1$ otrzymujemy, że macierz $H_F$ jest ujemnie określona i stąd dwuargumentowa funkcja $F(x,r)$ jest silnie wklęsła. Każda z funkcji $\ln L_j$ jest złożeniem funkcji $F$ z opisaną wcześniej wieloliniową funkcją $p(A_1, A_2, ..., A_m, t_j)$ i korzystając wyłącznie z definicji funkcji wklęsłej możemy sprawdzić, że $\ln L_j$ jest funkcją wklęsłą. Z założenia o maksymalnym rzędzie macierzy $J$ wynika, że przecięciem wszystkich hiperpłaszczyzn postaci

$$p(A_1, A_2, ..., A_m, t_j) = const$$

jest co najwyżej jeden punkt. Daje nam to silną wklęsłość funkcji $\ln L(A_1, A_2, ..., A_m, r)$. $\square$

**Twierdzenie 2.2.1.** *Niech będzie $n \geq m$. Niech dla danych $t_1, t_2, ..., t_n \in [c,d]$ rząd macierzy $J$ określonej w* (2.6) *jest maksymalny. Wtedy dla danych $t_1, t_2, ..., t_n$ oraz $x_1, x_2, ..., x_n$ istnieje dokładnie jeden $(\widehat{A}, \widehat{r}) \in \boldsymbol{A}$ taki, że*

$$L(\widehat{A}, \widehat{r}) = \max_{(A,r)\in \boldsymbol{A}} L(A,r)$$

*z prawdopodobieństwem* 1*, gdzie $L$ jest funkcją wiarygodności zdefiniowaną w* (2.1).

Dowód. Rozważmy przypadek, gdy warunek (i) z Lematu 2.2.3 nie zachodzi dla żadnego $a \in \mathbf{a}$. Prawdopodobieństwo zdarzenia, że warunek (i) z Lematu 2.2.3 zachodzi dla pewnego $a \in \mathbf{a}$ wynosi 0, co wynika z faktu, że nasze zmienne losowe $x_j, j = 1, 2, ..., n$ mają rozkłady absolutnie ciągłe.
Innymi słowy, rozważamy $a \in \mathbf{a}$, dla których nie istnieje dopasowanie opisane w warunku (i). Zdarzenie takie zachodzi z prawdopodobieństwem równym 1.
Niech

$$g\,(a) = \lim_{r\to\infty} \frac{d}{dr} \ln L(ra, r),$$

oraz

$$K(r) = \left\{ a \in \mathbf{a} : \frac{d}{dr} \ln L(ra,r) \geq g\,(a) + \frac{|g\,(a)\,|}{2} \right\}.$$

Funkcja $g\,(a)$ jest ciągła w dziedzinie, co wynika z dowodu Lematu 2.2.3, gdyż funkcja ta przedstawiona tam w jawnej postaci, jest złożeniem funkcji ciągłych. Z powyższego i ponadto, ze zwartości zbioru $\mathbf{a}$ wynika, że każdy zbiór należący do rodziny $\{K(r), r > 0\}$ jest zwarty. Funkcja $\frac{d}{dr} \ln L(ra, r)$ jest silnie malejąca, jako pochodna funkcji silnie wklęsłej, co wiemy dzięki Lematowi 2.2.4. Funkcja $K(\cdot)$ (a właściwie odwzorowanie wielowartościowe, którego wartościami są odpowiednie zbiory) jest także malejąca. Z Lematu 2.2.3 otrzymujemy, że

$$\bigcap_{r>0} K(r) = \emptyset.$$

Ponieważ każdy zbiór $K(r)$ jest zwarty, z twierdzenia Riesza istnieje $r_0 > 0$ takie, że $K(r_0) = \emptyset$. Korzystając z Lematu 2.2.5 otrzymujemy, że istnieje dokładnie jeden $\left(\widehat{A}, \widehat{r}\right) \in \mathbf{A}$ taki, że

$$\ln L(\widehat{A}, \widehat{r}) = \max_{(A,r)\in\mathbf{A}_0} \ln L(A, r),$$

gdzie $\mathbf{A}_0 = \{(A, r) \in \mathbf{A} : r \leq r_0\}$ jest zbiorem wypukłym oraz zwartym. Wypukłość wynika z tego, że zbiór jest przecięciem wypukłego zbioru $\mathbf{A}$ z hiperpłaszczyzną. Zwartość można udowodnić analogicznie jak w Lemacie 2.2.2.

Niech $(A, r) \in \mathbf{A}\backslash\mathbf{A}_0$ oraz $\alpha_i = \frac{A_i}{r}$, $i = 1, 2, ..., m$. Ponieważ wielowartościowa funkcja $K(r)$ jest malejąca, dla dowolnego $r \geq r_0$ mamy $K(r) = \emptyset$. Stąd dla dowolnego $r \geq r_0$ dla dowolnego $a \in \mathbf{a}$ mamy

$$\frac{d}{dr} \ln L(ra, r) < g(a) + \frac{|g(a)|}{2} < 0,$$

co daje nam, że $\ln L(ra, r)$ jest malejącą funkcją argumentu $r$ dla $r \geq r_0$. Stąd otrzymujemy, że

$$\ln L(A, r) = \ln L(ra, r) < \ln L(r_0 a, r_0) \leq \ln L(\widehat{A}, \widehat{r}).$$

Dowodząc istnienia maksimum funkcji $\ln L(A, r)$ otrzymujemy tezę naszego twierdzenia. □

Sformułujmy teraz nasze twierdzenie dla pewnej szerszej klasy rozkładów. Rozkłady te niech będą modelowane przez parametry $p$ oraz $r$. "Parametr" $p$ niech będzie sumą pewnej ustalonej liczby $m$ ciągłych funkcji liniowo niezależnych $f_1, f_2, ..., f_m$, innymi słowy

$$p(A, t) = \sum_{k=1}^{m} A_k f_k(t).$$

Niech spełnione będą poniższe warunki, równoważne udowodnionym wyżej Lematom 2.2.1 – 2.2.5:

(i) Niech wartość oczekiwana każdej zmiennej $x_j$ będzie postaci

$$E[x_j] = \mu_j$$

gdzie

$$\mu_j = \sum_{k=1}^{m} \alpha_k f_k(t_j)$$

oraz

$$\alpha_k = \frac{A_k}{\mathbf{G}(r)}, \qquad k = 1, 2, ..., m.$$

Ponadto

$$p(A, t) \leq M\mathbf{G}(r),$$

dla pewnego ustalonego $M \in \mathbf{R}$ oraz dla pewnej ustalonej przez postać rozkładu ciągłej funkcji $\mathbf{G}(\cdot)$.

(ii) Zbiór **a** wszystkich parametrów postaci $a = (\alpha_1, \alpha_2, ..., \alpha_m) \in \mathbb{R}^m$ spełniających nierówność

$$0 \leq \sum_{k=1}^{m} \alpha_k f_k(t) \leq M \tag{2.7}$$

dla dowolnego $t \in \mathbb{R}$ jest niepusty i zwarty w $\mathbb{R}^m$.

(iii) Dla wszystkich $j = 1, 2, ..., n$ albo

$$x_j = \sum_{k=1}^{m} \alpha_k f_k(t_j)$$

albo

$$\lim_{r\to\infty} \frac{d}{dr} \ln L(ra, r) < 0.$$

(iv) Funkcja

$$\ln L\,(ra, r)$$

jest ściśle wklęsła ze względu na zmienną $r$.

(v) Funkcja

$$\ln L\,(A_1, A_2, ..., A_m, r)$$

jest ściśle wklęsła ze względu na zmienną $(A_1, A_2, ..., A_m, r)$ dla odpowiednio dużej liczby obserwacji $n$.

**Twierdzenie 2.2.2.** *Przy powyższych założeniach (i)-(v),* Twierdzenie 2.2.1 *jest prawdziwe dla zdefiniowanego wyżej modelu.*

Dowód powyższego twierdzenia przebiega analogicznie do dowodu Twierdzenia 2.2.1.

# Rozdział 3

# Model beta–regresji

Standardowego modelu regresji nie ma sensu stosować w przypadku, gdy zmienna zależna należy do pewnego ustalonego przedziału domkniętego. Na przykład, gdy wyraża ona jakiś stosunek czy frakcję. Po raz pierwszy terminu beta-regresja użyli Dawidowicz, Stanuch oraz Kawalec na konferencji zorganizowanej przez ISCB (*International Society of Clinical Biostatistics*) w Sztokholmie w 2001 roku. Zastosowanie metod Uogólnionego Modelu Liniowego do zagadnień beta-regresji przedstawili Ferrari oraz Cribari-Neto (2004). Zastosowanie do problemów małych próbek przy użyciu metody największej wiarygodności zawiera praca Cribari-Neto i Vasconcellos (2002).
Dla uproszczenia, bez straty ogólności, założymy, że zmienna należy do przedziału $[0, 1]$. W poniższym rozdziale zaprezentujemy zastosowanie modelu beta–regresji do danych, których wartość oczekiwana może być opisana, podobnie jak ma to miejsce w modelu gamma–regresji, przez sumę ustalonej liczby funkcji liniowo niezależnych, a w szczególności jako kombinacja funkcji okresowej oraz trendu. Udowodnimy, że w tak zdefiniowanym modelu istnieje dokładnie jeden estymator największej wiarygodności. Jest to uogólnienie rezultatu podanego przez Rydlewskiego (2007).

## 3.1 Sformułowanie problemu

W naszym modelu zakładamy, że zmienna zależna ma rozkład beta. Gęstość zmiennej o rozkładzie beta dana jest wzorem

$$f(x, \varphi, r) = \frac{1}{B(r\varphi, (1-\varphi)r)} x^{r\varphi-1}(1-x)^{(1-\varphi)r-1}, 0 \le x \le 1$$

lub inaczej

$$f(x, \varphi, r) = \frac{\Gamma(r)}{\Gamma(r\varphi)\Gamma((1-\varphi)r)} x^{r\varphi-1}(1-x)^{(1-\varphi)r-1}, 0 \le x \le 1,$$

gdzie $0 < \varphi < 1$, $r > 0$ oraz $B(\cdot,\cdot)$ to funkcja beta i $\Gamma(\cdot)$ to funkcja gamma. Z własności rozkładu beta wiemy, że dla $X$ zmiennej losowej o rozkładzie beta mamy

$$E(X) = \varphi$$

oraz

$$Var(X) = \frac{\varphi(1-\varphi)}{1+r},$$

gdzie $r$ będzie nieznanym parametrem precyzji.
Niech $x_1, x_2, ..., x_n$ będą niezależnymi zmiennymi losowymi o rozkładzie beta. W naszym modelu zakładamy, że wartość oczekiwana zmiennej zależnej jest modelowana przez funkcję

$$E(x_j) = \varphi(t_j), \quad j = 1, 2, ..., n,$$

gdzie $\varphi$ jest sumą pewnej ustalonej liczby ciągłych funkcji liniowo niezależnych. Zwykle w zastosowaniach ograniczymy się do sumy funkcji okresowych o okresie $T$ oraz ograniczonej, ciągłej, niestałej i nieokresowej funkcji trendu $\tau$. Wartości $t_j$ mogą być interpretowane w zastosowaniach jako czas pomiaru.
Ponieważ zawsze możemy odpowiednio przeskalować nasze dane możemy zawęzić nasze poszukiwania do przedziału $[0,1]$, innymi słowy rozważamy model postaci

$$0 \leq x_j \leq 1 \quad j = 1, 2, ..., n,$$

oraz

$$E(x_j) = \varphi(t_j) = \sum_{k=1}^{m} \alpha_k f_k(t_j),$$

gdzie $0 \leq \varphi(t_j) \leq 1$. W szczególnie interesującym przypadku model upraszcza się do postaci

$$E(x_j) = \varphi(t_j) = \tau(t_j) + \beta_0 + \sum_{k=1}^{m} \alpha_k \sin\frac{2\pi k}{T} t_j + \beta_k \cos\frac{2\pi k}{T} t_j.$$

Oczywiście, istnieje wiele możliwych wyborów funkcji $\tau(\cdot)$.
W dalszej części pracy zakładamy, że funkcja trendu jest funkcją niestałą, nieokresową, ciągłą i ograniczoną, która odwzorowuje ograniczony przedział zawierający wszystkie $t_j$ na przedział $[0,1]$. W artykule (2007) założono, że trend jest liniowy, lecz poniższe rozumowanie ma zastosowanie, jak wyżej napisano, do znacznie szerszej klasy funkcji.
W przypadku ogólnym, jak już wspomniano, do modelowania parametru $p = r\varphi$ możemy użyć dowolnego układu $m$ ciągłych funkcji liniowo niezależnych. Używamy w takim przypadku oznaczeń $A = (A_1, A_1, ..., A_m)$ oraz mamy

$$p(A,t) = \sum_{k=1}^{m} A_k f_k(t).$$

W szczególności rozpatrzony jest model, gdzie $p$ jest funkcją okresową, a parametrami są jej współczynniki Fouriera. Do tego modelu można dorzucić również jednoparametrową ciągłą funkcję $\tau(\cdot)$ opisującą trend. Innymi słowy, w tak wyróżnionym przypadku, naszym układem funkcji jest

$$(\tau(\cdot), \sin(\cdot), \sin 2(\cdot), ..., \sin m(\cdot), 1, \cos(\cdot), \cos 2(\cdot), ..., \cos m(\cdot)) .$$

Połóżmy wtedy $A = (A_0, A_1, ..., A_m, B_0, B_1, ..., B_m)$ i "parametr" $p$ ma postać funkcji

$$p(A,t) = A_0 \tau(t) + B_0 + \sum_{k=1}^{m} (A_k \sin kt + B_k \cos kt) .$$

Niech $x_1, x_2, ..., x_n$ będą niezależnymi zmiennymi losowymi o rozkładzie beta. W wyniku poczynionych założeń otrzymujemy, że wartości oczekiwane zmiennych losowych $x_j$ mają postać

$$E(x_j) = \sum_{k=1}^{m} \alpha_k f_k(t_j), \qquad j = 1, 2, ..., n,$$

a w naszym szczególnie wyróżnionym przypadku

$$E(x_j) = \alpha_0 \tau(t_j) + \beta_0 + \sum_{k=1}^{m} (\alpha_k \sin kt_j + \beta_k \cos kt_j) , \qquad j = 1, 2, ..., n,$$

gdzie $\alpha_k = \frac{A_k}{r}$ oraz $\beta_k = \frac{B_k}{r}$. Funkcja wiarogodności w modelu beta–regresji przyjmuje postać

$$L(t_1, t_2, ..., t_n, x_1, x_2, ..., x_n, A, r) =$$

$$= \prod_{j=1}^{n} \frac{1}{B(p(A,t_j), r - p(A,t_j))} x_j^{p(A,t_j)-1} (1 - x_j)^{r - p(A,t_j) - 1}, \tag{3.1}$$

gdzie $B(\cdot, \cdot)$ oznacza funkcję beta. Logarytm funkcji wiarogodności przyjmuje w związku z tym w naszym modelu postać

$$\ln L = \sum_{j=1}^{n} \ln L_j =$$

$$= \sum_{j=1}^{n} - \ln B(p(A,t_j), r - p(A,t_j)) + (p(A,t_j) - 1) \ln x_j + (r - p(A,t_j) - 1) \ln(1 - x_j).$$

Przepiszemy nasze parametry w innej postaci, wygodniejszej dla dalszego wywodu. Niech będzie $a = (\alpha_1, \alpha_2, ..., \alpha_m)$. Wówczas

$$\varphi(a, t_j) = \frac{p(A,t_j)}{r} = \sum_{k=1}^{m} \alpha_k f_k(t_j).$$

W wyróżnionym przypadku $a = (\alpha_0, \alpha_1, ..., \alpha_m, \beta_{0,}\beta_1, ..., \beta_m)$ i otrzymujemy

$$\varphi(a, t_j) = \frac{p(A, t_j)}{r} = \alpha_0 \tau(t_j) + \beta_0 + \sum_{k=1}^{m} (\alpha_k \sin kt_j + \beta_k \cos kt_j) ,$$

gdzie $\alpha_k = \frac{A_k}{r}$ oraz $\beta_k = \frac{B_k}{r}$.
Mamy

$$\ln L\,(ra, r) = \sum_{j=1}^{n} \ln L_j\,(ra, r)$$

gdzie

$$\ln L_j\,(ra, r) = -\ln B(r\varphi(a, t_j), r(1-\varphi(a, t_j))) + (r\varphi(a, t_j)-1) \ln x_j + (r(1-\varphi(a, t_j))-1) \ln(1-x_j).$$

## 3.2 Estymacja metodą największej wiarygodności

**Lemat 3.2.1.** *Funkcja* $\ln B(x, y)$ *jest ściśle wypukła ze względu na zmienną* $(x, y)$.

Dowód. Niech $x_1, x_2, y_1, y_2 > 0$ oraz niech $\lambda \in (0, 1)$. Mamy

$$\begin{aligned} &B\,(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) = \\ &= \int_0^1 s^{\lambda x_1 + (1-\lambda)x_2 - 1} (1-s)^{\lambda y_1 + (1-\lambda)y_2 - 1} ds = \\ &= \int_0^1 \left[s^{x_1-1}(1-s)^{y_1-1}\right]^{\lambda} \left[s^{x_2-1}(1-s)^{y_2-1}\right]^{1-\lambda} ds \end{aligned}$$

Dalej, korzystając z nierówności Höldera z wykładnikami $\frac{1}{\lambda}$ i $\frac{1}{1-\lambda}$ otrzymujemy

$$\begin{aligned} &B\,(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \le \\ &\le \left[\int_0^1 s^{x_1-1}(1-s)^{y_1-1} ds\right]^{\lambda} \left[\int_0^1 s^{x_2-1}(1-s)^{y_2-1} ds\right]^{1-\lambda} . \end{aligned}$$

Stąd dostajemy nierówność

$$B\,(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \le B\,(x_1, y_1)^{\lambda}\, B\,(x_2, y_2)^{1-\lambda}$$

oraz ostatecznie

$$\ln B\,(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \le \lambda B\,(x_1, y_1) + (1-\lambda) B\,(x_2, y_2) .$$

Równość ma miejsce wtedy i tylko wtedy, gdy istnieje taka niezerowa stała $c \in \mathbb{R}$, że dla $s \in [0, 1]$ mamy

$$s^{x_2-1}(1-s)^{y_2-1} = c\left(s^{x_1-1}(1-s)^{y_1-1}\right)^{\frac{1}{\lambda}-1} .$$

Równość ma więc miejsce wtedy i tylko wtedy, gdy $x_1 = x_2$ i $y_1 = y_2$. Otrzymujemy stąd, że funkcja $B(x, y)$ jest ściśle wypukła. $\square$

**Lemat 3.2.2.** *Niech $[c, d]$ będzie przedziałem domkniętym i ograniczonym. Niech $f_1, f_2, ..., f_m$ będzie dowolnym układem funkcji liniowo niezależnych ciągłych na $[c, d]$. Zbiór wszystkich parametrów $\boldsymbol{A}$ postaci $(A_1, A_2, ..., A_m, r) \in \mathbb{R}^{m+1}$ takich, że dla dowolnego $t \in [c, d]$,*

$$0 \leq \sum_{k=1}^{m} A_k f_k(t) \leq Mr \tag{3.2}$$

*jest niepusty, domknięty i wypukły w $\mathbb{R}^{m+1}$.*

Lemat jest analogiczny do Lematu 2.2.1. □

**Lemat 3.2.3.** *Niech $f_1, f_2, ..., f_m : \mathbb{R} \longrightarrow \mathbb{R}$ będzie układem funkcji ciągłych i liniowo niezależnych. Zbiór* **a** *wszystkich parametrów postaci $a = (\alpha_1, \alpha_2, ..., \alpha_m) \in \mathbb{R}^m$ spełniających nierówność*

$$0 \leq \sum_{k=1}^{m} \alpha_k f_k(t) \leq M \tag{3.3}$$

*dla dowolnego $t \in \mathbb{R}$ jest niepusty i zwarty w $\mathbb{R}^m$.*

Lemat jest analogiczny do Lematu 2.2.2 i dlatego jego dowód zostanie pominięty. □

**Lemat 3.2.4.** *Prawdziwy jest dokładnie jeden z poniższych warunków*

*(i) Dla wszystkich $j = 1, 2, ..., n$*

$$x_j = \sum_{k=1}^{m} \alpha_k f_k(t_j)$$

*(ii)*

$$\lim_{r \to \infty} \frac{d}{dr} \ln L(ra, r) < 0.$$

Dowód. W naszym modelu beta-regresji mamy

$$\lim_{r \to \infty} \frac{d}{dr} \ln L(ra, r) = \lim_{r \to \infty} \frac{d}{dr} \sum_{j=1}^{n} \ln L_j(ra, r) =$$

$$= \lim_{r \to \infty} \frac{d}{dr} \sum_{j=1}^{n} (-\ln B(r\varphi(a, t_j), r(1 - \varphi(a, t_j)))+$$

$$+(r\varphi(a, t_j) - 1) \ln x_j + (r(1 - \varphi(a, t_j)) - 1) \ln(1 - x_j)).$$

Niech $y_j = \varphi(a, t_j)$. Otrzymujemy wtedy

$$\ln L_j(ra, r) = -\ln B(ry_j, r(1 - y_j)) + (ry_j - 1) \ln x_j + (r(1 - y_j) - 1) \ln(1 - x_j).$$

Dalej, korzystając z własności funkcji $B$ dostajemy

$$\ln L_j(ra, r) =$$

$$= -\ln\Gamma(ry_j) - \ln\Gamma(r(1-y_j)) + \ln\Gamma(r) + (ry_j - 1)\ln x_j + (r(1-y_j) - 1)\ln(1-x_j).$$

Stąd

$$\frac{d}{dr}\ln L_j(ra, r) =$$

$$\begin{aligned} &= -\frac{d}{dr}\ln\Gamma(ry_j) - \frac{d}{dr}\ln\Gamma(r(1-y_j)) + \frac{d}{dr}\ln\Gamma(r) + y_j\ln x_j + (1-y_j)\ln(1-x_j) \\ &= -y_j\Psi(ry_j) - (1-y_j)\Psi(r(1-y_j)) + \Psi(r) + y_j\ln x_j + (1-y_j)\ln(1-x_j), \end{aligned}$$

gdzie $\Psi(x) = \frac{d}{dx}\ln\Gamma(x)$. Ponownie wykorzystujemy równość (2.5)

$$\lim_{x\to\infty}(\Psi(x) - \ln x) = 0,$$

a stąd

$$\lim_{r\to\infty}\frac{d}{dr}\ln L_j(ra, r) = -y_j\ln y_j - (1-y_j)\ln(1-y_j) + y_j\ln x_j + (1-y_j)\ln(1-x_j).$$

Niech

$$h(x) = y_j\ln x + (1-y_j)\ln(1-x).$$

Funkcja $h$ przyjmuje największą wartość w $x = y_j$, a stąd

$$\lim_{r\to\infty}\sum_{j=1}^{n}\frac{d}{dr}\ln L_j(ra, r) \leq 0$$

i jeżeli dla przynajmniej jednego $j \in \{1, 2, ..., n\}$ mamy $x_j \neq y_j$, to otrzymujemy

$$\lim_{r\to\infty}\sum_{j=1}^{n}\frac{d}{dr}\ln L_j(ra, r) < 0.$$

$\square$

**Lemat 3.2.5.** *Funkcja* $\ln L(ra, r)$ *jako funkcja parametru* $r$ *jest silnie wklęsła.*

Dowód. Ze względu na postać funkcji $\ln L(ra, r)$ wystarczy pokazać, że istnieje chociaż jedno $j \in \{1, 2, ..., n\}$ takie, że funkcja

$$\ln B\left(r\varphi(a, t_j), r(1-\varphi(a, t_j))\right)$$

jest silnie wypukła. Postępując tak, jak w dowodzie Lematu 3.2.1 z definicji wypukłości oraz wykorzystując nierówność Höldera uzyskujemy silną wypukłość. □

Niech rzeczywista macierz $J \in \mathbb{R}^{n \times m}$ będzie postaci

$$\begin{bmatrix} f_1(t_1) & f_2(t_1) & \dots & \dots & \dots & f_m(t_1) \\ f_1(t_2) & f_2(t_2) & \dots & \dots & \dots & f_m(t_2) \\ \dots & \dots & \dots & & & \dots \\ \dots & \dots & & \dots & & \dots \\ \dots & \dots & & & \dots & \dots \\ f_1(t_n) & f_2(t_n) & \dots & \dots & \dots & f_m(t_n) \end{bmatrix}. \tag{3.4}$$

Jak widać postać macierzy $J$ zależy od naszych obserwacji oraz od zakładanej w modelu ilości parametrów.

**Lemat 3.2.6.** *Jeżeli liczba naszych obserwacji $n$ jest odpowiednio duża w stosunku do liczby zakładanych w modelu parametrów, dokładniej $n \geq m$ oraz rząd macierzy $J$ jest maksymalny, czyli inaczej rzJ=m, to funkcja $\ln L$ jako funkcja parametrów $(A_1, ..., A_m, r)$ jest silnie wklęsła.*

Dowód. Funkcja $\ln B(x, y)$, jak pokazano w Lemacie 3.2.1, jest ściśle wypukła. Stąd otrzymujemy, że interesująca nas funkcja

$$\ln L(A_1, ..., A_m, r) =$$

$$= \sum_{j=1}^{n} - \ln B(p(A, t_j), r - p(A, t_j)) + (p(A, t_j) - 1) \ln x_j + (r - p(A, t_j) - 1) \ln(1 - x_j)$$

jako suma funkcji wklęsłych i liniowych jest wklęsła. Z założenia na rząd macierzy $J$ wynika, że przecięciem wszystkich hiperpłaszczyzn postaci

$$p(A, t_j) = p(A_1, ..., A_m, t_j) = const$$

jest co najwyżej jeden punkt. Daje nam to, przy założeniach naszego lematu, silną wklęsłość funkcji $\ln L(A_1, ..., A_m, r)$. □

**Twierdzenie 3.2.1.** *Niech będzie $n \geq m$. Niech dla danych $t_1, t_2, ..., t_n \in [c, d]$ rząd macierzy $J$ określonej w (3.4) jest maksymalny. Wtedy dla danych $t_1, t_2, ..., t_n$ oraz $x_1, x_2, ..., x_n$ istnieje dokładnie jeden $(\widehat{A}, \widehat{r}) \in \boldsymbol{A}$ taki, że*

$$L(\widehat{A}, \widehat{r}) = \max_{(A,r) \in \boldsymbol{A}} L(A, r)$$

*z prawdopodobieństwem 1, gdzie $L$ jest funkcją wiarygodności zdefiniowaną w (3.1).*

Dowód twierdzenia jest taki, jak dla przypadku gamma–regresji. Zostanie on przytoczony dla porządku oraz wygody czytelnika.
Rozważmy przypadek, gdy warunek (i) z Lematu 3.2.4 nie zachodzi dla żadnego $a \in \mathbf{a}$. Prawdopodobieństwo zdarzenia, że warunek (i) z Lematu 3.2.4 zachodzi dla pewnego $a \in \mathbf{a}$ wynosi 0, co wynika z faktu, że nasze zmienne losowe $x_j, j = 1, 2, ..., n$ mają rozkłady absolutnie ciągłe. Innymi słowy, rozważamy $a \in \mathbf{a}$, dla których nie istnieje dopasowanie opisane w warunku (i). Zdarzenie takie zachodzi z prawdopodobieństwem równym 1.
Niech

$$g(a) = \lim_{r\to\infty} \frac{d}{dr} \ln L(ra, r),$$

oraz

$$K(r) = \left\{ a \in \mathbf{a} : \frac{d}{dr} \ln L(ra, r) \geq g(a) + \frac{|g(a)|}{2} \right\}.$$

Funkcja $g(a)$ jest ciągła w dziedzinie, co wynika z dowodu Lematu 3.2.4, gdyż funkcja ta przedstawiona tam w jawnej postaci, jest złożeniem funkcji ciągłych. Z powyższego i ponadto, ze zwartości zbioru $\mathbf{a}$ wynika, że każdy zbiór należący do rodziny $\{K(r), r > 0\}$ jest zwarty.
Funkcja $\frac{d}{dr} \ln L(ra, r)$ jest silnie malejąca, jako pochodna funkcji silnie wklęsłej. Jest to wniosek z Lematu 3.2.5. Odwzorowanie wielowartościowe $K(\cdot)$ jest także malejące. Z Lematu 3.2.4 otrzymujemy, że

$$\bigcap_{r>0} K(r) = \emptyset.$$

Ponieważ każdy zbiór $K(r)$ jest zwarty, z twierdzenia Riesza istnieje $r_0 > 0$ takie, że $K(r_0) = \emptyset$. Korzystając z Lematu 3.2.6 otrzymujemy, że istnieje dokładnie jeden $\left(\widehat{A}, \widehat{r}\right) \in \mathbf{A}$ taki, że

$$\ln L(\widehat{A}, \widehat{r}) = \max_{(A,r)\in\mathbf{A}_0} \ln L(A, r),$$

gdzie $\mathbf{A}_0 = \{(A, r) \in \mathbf{A} : r \leq r_0\}$ jest zbiorem wypukłym oraz zwartym. Wypukłość wynika z tego, że zbiór jest przecięciem wypukłego zbioru $\mathbf{A}$ z hiperpłaszczyzną. Zwartość można udowodnić analogicznie jak w Lemacie 3.2.2.
Niech $(A, r) \in \mathbf{A}\backslash\mathbf{A}_0$ oraz $\alpha_i = \frac{A_i}{r}$, $i = 1, 2, ..., m$. Ponieważ wielowartościowe odwzorowanie $K(r)$ jest malejące, dla dowolnego $r \geq r_0$ mamy $K(r) = \emptyset$. Stąd dla dowolnego $r \geq r_0$ dla dowolnego $a \in \mathbf{a}$ mamy

$$\frac{d}{dr} \ln L(ra, r) < g(a) + \frac{|g(a)|}{2} < 0,$$

co daje nam, że $\ln L(ra, r)$ jest malejącą funkcją argumentu $r$ dla $r \geq r_0$. Stąd otrzymujemy, że

$$\ln L(A, r) = \ln L(ra, r) < \ln L(r_0 a, r_0) \leq \ln L(\widehat{A}, \widehat{r}).$$

Dowodząc istnienia maksimum funkcji $\ln L(A, r)$ otrzymujemy tezę naszego twierdzenia. $\square$

# Rozdział 4

# Asymptotyczne własności estymatorów największej wiarygodności

## 4.1 Zgodność estymatorów największej wiarygodności

Przejdziemy teraz do opisu asymptotycznych własności otrzymanych estymatorów największej wiarygodności. Ponieważ rozpatrywane w tej rozprawie modele nie spełniają w oczywisty sposób założeń twierdzeń asymptotycznych, jak na przykład kluczowych dla Twierdzeń 1.4.1, 1.4.2 i 1.4.3 założeń A3. i B3., tak jak i założeń innych, podobnych twierdzeń asymptotycznych podanych choćby w pracy Fahrmeira (1987) i Fahrmeira i Kaufmanna (1985), to za wzorem Nie (2006) i innych badaczy konkretnych modeli przeprowadzono własny dowód zgodności i asymptotycznej normalności.
Najpierw udowodnimy silną zgodność estymatora największej wiarogodności będącego rozwiązaniem równania wiarygodności w modelu gamma-regresji.
Przypomnijmy, że mamy $n$ zmiennych losowych niezależnych, gdzie zmienna losowa $x_i$ ma rozkład postaci $\gamma(p, r)$. Funkcja gęstości zmiennej losowej $x_i$ o rozkładzie gamma jest postaci

$$h(t, p, r) = \frac{r^p}{\Gamma(p)} t^{p-1} e^{-rt}, \qquad t > 0,$$

gdzie $r > 0$ jest parametrem skali, $p > 0$ jest parametrem kształtu oraz $\Gamma(\cdot)$ to funkcja gamma. Wartość oczekiwana zmiennej losowej $x_i$ o rozkładzie gamma wynosi $E(x_i) = \frac{p}{r}$, a wariancja $Var(x_i) = \frac{p}{r^2}$. W naszym modelu za $p$ wstawiamy funkcję zmiennej $t_i$ zależną od wielowymiarowego parametru $A$. Dokładniej, możemy użyć dowolnego układu $m$ ciągłych funkcji liniowo niezależnych. Mamy wtedy

$$p(A, t) = \sum_{k=1}^{m} A_k g_k(t).$$

W przypadku beta regresji mamy $n$ zmiennych losowych niezależnych, gdzie zmienna losowa $x_i$ ma rozkład postaci $\beta(\varphi, r)$, gdzie $0 < \varphi < 1$, $r > 0$. Parametrem precyzji nazywamy $r$, a $\varphi$ jest wartością oczekiwaną. Gęstość zmiennej o rozkładzie beta dana jest wzorem

$$h(x, \varphi, r) = \frac{\Gamma(r)}{\Gamma(r\varphi)\Gamma((1-\varphi)r)} x^{r\varphi-1}(1-x)^{(1-\varphi)r-1}, 0 \leq x \leq 1,$$

gdzie $\Gamma(\cdot)$ jest funkcją gamma. Dla $X$ zmiennej losowej o rozkładzie beta mamy

$$E(X) = \varphi$$

oraz

$$Var(X) = \frac{\varphi(1-\varphi)}{1+r}.$$

W naszym modelu zakładamy, że wartość oczekiwana zmiennej zależnej jest modelowana przez funkcję

$$E(x_j) = \varphi(t_j), \quad j = 1, 2, ..., n,$$

gdzie $\varphi$ jest sumą pewnej ustalonej liczby ciągłych funkcji liniowo niezależnych. Mamy

$$E(x_j) = \varphi(t_j) = \sum_{k=1}^{m} \alpha_k g_k(t_j),$$

gdzie $0 \leq \varphi(t_j) \leq 1$.
W dalszej części rozdziału posłużymy się następującą konwencją. Zmienne losowe zależą od parametru $(A, r) = \theta \in \Theta \subset \mathbb{R}^{m+1}$, gdzie $\Theta$ to zbiór borelowski. Każda zmienna losowa $x_i$ ma rozkład o gęstości $f_i(\cdot, \theta)$ względem pewnej $\sigma$-skończonej miary $\mu$. Niech $x_1, x_2, ..., x_n$ będą niezależnymi obserwacjami, a w zależności od kontekstu zmiennymi losowymi. Przyjmijmy,że

$$l_n(\theta|x) = \ln L_n(\theta, x) = \sum_{i=1}^{n} \ln f_i(x_i, \theta).$$

Niżej podamy założenia i uwagi do znajdującego się dalej dowodu zgodności otrzymanych estymatorów największej wiarygodności.

- AN 1. Prawdziwa wartość $\theta_0 \in \text{int}\, \Theta$.
- AN 2. Mamy
$$\frac{1}{n} J_n(\theta) \to K(\theta) \text{ przy } n \to +\infty,$$
jednostajnie w $\overline{N}(\delta)$, gdzie $\overline{N}(\delta)$ jest otoczeniem prawdziwej wartości $\theta_0$ o promieniu $\delta$ oraz $K(\theta)$ jest pewną macierzą dodatnio określoną.

Założenie AN 2. jest powszechnie przyjmowane, ze względu na jego naturalną interpretację. Możemy bowiem oczekiwać, że średnia informacja Fishera o parametrze $\theta$ zmierza do pewnej określonej wartości. Dodajmy, że większość autorów przyjęła to założenie (np. Jennrich (1969), Fahrmeir i Kaufmann (1985) dla uogólnionego modelu liniowego, Gallant (1987), Morton (1987), Wei (1998)).

**Uwaga 4.1.1.** *Ze względu na rozkład zmiennych losowych nie są potrzebne dodatkowe założenia umożliwiające skorzystanie z Mocnego Prawa Wielkich Liczb Kołmogorowa.*

**Uwaga 4.1.2.** *W obu analizowanych modelach prawdą jest, że*

$$E\left(l_n'(x,\theta)\right) = 0$$

*oraz*

$$E\left(l_n''(x,\theta)\right) = -J_n(\theta),$$

*gdzie $J_n(\theta)$ to informacja Fishera.*

Dowód. Każda z całek w tej uwadze to całka po $\Omega = [0,1]^n$ w przypadku modelu beta, zaś w przypadku modelu gamma $\Omega = (0,+\infty)^n$. Ponieważ gęstości niezależnych zmiennych losowych występujących w obu modelach są różniczkowalne dowolną liczbę razy oraz $\int \left|\frac{\partial L}{\partial \theta}\right| dx < +\infty$ oraz $\int \left|\frac{\partial^2 L}{\partial\theta\partial\theta^T}\right| dx < +\infty$, co można łatwo pokazać z definicji funkcji beta i funkcji gamma, możemy skorzystać z twierdzenia o różniczkowaniu pod znakiem całki. W efekcie dla dowolnego $\theta_k$ prawdą jest, że

$$\int \frac{\partial}{\partial\theta_k} L(x,\theta)dx = \frac{\partial}{\partial\theta_k}\int L(x,\theta)dx = \frac{\partial}{\partial\theta_k}1 = 0.$$

Stąd

$$E\left(\frac{\partial \ln L(x,\theta)}{\partial\theta_k}\right) = \int \frac{1}{L(x,\theta)}\frac{\partial L(x,\theta)}{\partial\theta_k}L(x,\theta)dx = 0 \tag{4.1}$$

oraz

$$\int \frac{\partial^2 L(x,\theta)}{\partial\theta_l\partial\theta_k}dx = \frac{\partial}{\partial\theta_l}\left(\int \frac{\partial L(x,\theta)}{\partial\theta_k}dx\right) = 0.$$

Mamy dalej

$$E\left(\frac{\partial^2 \ln L(x,\theta)}{\partial\theta_l\partial\theta_k}\right) = E\left(\frac{\partial}{\partial\theta_l}\left(\frac{1}{L(x,\theta)}\frac{\partial L(x,\theta)}{\partial\theta_k}\right)\right) =$$

$$= E\left(\frac{\frac{\partial^2 L(x,\theta)}{\partial\theta_l\partial\theta_k}L(x,\theta) - \frac{\partial L(x,\theta)}{\partial\theta_l}\frac{\partial L(x,\theta)}{\partial\theta_k}}{\left(L(x,\theta)\right)^2}\right) =$$

$$= \int \frac{\partial^2 L(x,\theta)}{\partial\theta_l\partial\theta_k}dx - E\left(\frac{\partial \ln L(x,\theta)}{\partial\theta_l}\frac{\partial \ln L(x,\theta)}{\partial\theta_k}\right) = -E\left(\frac{\partial \ln L(x,\theta)}{\partial\theta_l}\frac{\partial \ln L(x,\theta)}{\partial\theta_k}\right) = -J_n^{lk}(\theta)$$

□

**Twierdzenie 4.1.1.** *Otrzymany w modelach gamma– i beta–regresji estymator największej wiarygodności jest silnie zgodny.*

Dowód. Pokażemy, że

$$\exists \delta^* > 0 \forall \delta \in (0, \delta^*] \, \exists n^* \in \mathbb{N} \forall n > n^* \forall \theta \in \partial N(\delta) : P\left(l_n(\theta) - l_n(\theta_0) < 0\right) = 1, \tag{4.2}$$

gdzie wprowadzamy dość naturalne oznaczenia $N(\delta) = \{\theta : ||\theta - \theta_0|| < \delta\}$, $\partial N(\delta) = \{\theta : ||\theta - \theta_0|| = \delta\}$ oraz $\overline{N}(\delta) = \{\theta : ||\theta - \theta_0|| \leq \delta\}$.
Oznacza to, że $\widehat{\theta}_n$, które maksymalizuje $l_n(\theta)$ musi się zawierać w $\overline{N}(\delta)$. Ponieważ $\delta \leq \delta^*$ oraz $\delta$ jest dowolne, otrzymamy $\widehat{\theta}_n \to \theta_0$ p.w.
Niech $\lambda = \frac{\theta - \theta_0}{\delta}$ i korzystając z twierdzenia Taylora otrzymujemy

$$l_n(\theta) - l_n(\theta_0) = \delta \lambda^T l_n'(\theta_0) + \frac{1}{2}\delta^2 \lambda^T l_n''(\widehat{\theta}_n)\lambda,$$

gdzie dla pewnego $\widehat{t}_n \in [0,1]$ jest $\widehat{\theta}_n = \widehat{t}_n \theta_0 + (1 - \widehat{t}_n)\theta$. Wyrażeniu (4.2) równoważne jest poniższe

$$\exists \delta^* > 0 \forall \delta \in (0, \delta^*] \, \exists n^* \in \mathbb{N} \forall n > n^* \forall \theta \in \partial N(\delta) :$$
$$P\left(\frac{1}{n}\lambda^T l_n'(\theta_0) < \frac{1}{2n}\delta \lambda^T \left(-l_n''(\widehat{\theta}_n)\right)\lambda\right) = 1. \tag{4.3}$$

Niech $J_n$ oznacza macierz informacyjną Fishera uzyskaną przy $n$ obserwacjach. Definiując $R_n(\widehat{\theta}_n) = l_n''(\widehat{\theta}_n) + J_n(\widehat{\theta}_n)$, otrzymujemy

$$\exists \delta^* > 0 \forall \delta \in (0, \delta^*] \, \exists n^* \in \mathbb{N} \forall n > n^* \forall \theta \in \partial N(\delta) :$$

$$P\left(\frac{1}{n}\lambda^T l_n'(\theta_0) < \frac{1}{2n}\delta \lambda^T J_n(\widehat{\theta}_n)\lambda - \frac{1}{2n}\delta \lambda^T R_n(\widehat{\theta}_n)\lambda\right) = 1.$$

Powyższa zaś równość wynika z trzech faktów.
(i) Pierwsza z potrzebnych nierówności orzeka, na mocy ciągłości funkcji $J_n(\theta)$ w $\theta_0$ oraz założenia AN 2., że

$$\forall \epsilon > 0 \exists \delta_1 > 0 \exists n_1 \in \mathbb{N} \forall \theta \in N(\delta_1) \forall n > n_1 :$$
$$\left|\frac{1}{n}\lambda^T J_n(\theta)\lambda - \lambda^T K(\theta_0)\lambda\right| \leq$$
$$\leq \left|\frac{1}{n}\lambda^T J_n(\theta)\lambda - \lambda^T K(\theta)\lambda\right| + \left|\lambda^T K(\theta)\lambda - \lambda^T K(\theta_0)\lambda\right| < \frac{\epsilon}{2}.$$

Oznaczając przez $c = \lambda_{\min}(K(\theta_0))$ najmniejszą wartość własną macierzy $K(\theta_0)$ i korzystając z nierówności $\lambda_{\min}(K(\theta_0)) \leq \lambda^T K(\theta_0)\lambda$, dla $\lambda^T\lambda = 1$, dostajemy

$$\forall \epsilon > 0 \exists \delta_1 > 0 \exists n_1 \in \mathbb{N} \forall \theta \in N(\delta_1) \forall n > n_1 : \frac{1}{n} \lambda^T J_n(\theta) \lambda > c - \frac{\epsilon}{2} \text{ p.w.}$$

(ii) Druga z potrzebnych nam nierówności mówi, że

$$\forall \epsilon > 0 \exists \delta^* \leq \delta_1 \exists n_2 \geq n_1 \forall \theta \in N(\delta^*) \forall n > n_2 : \frac{1}{n} \lambda^T R_n(\theta) \lambda < \frac{\epsilon}{2} \text{ p.w.},$$

Powyższa nierówność jest prawdziwa na mocy Uwagi 4.1.2 oraz Mocnego Prawa Wielkich Liczb Kołmogorowa.
(iii) Trzeci fakt stwierdza, że

$$\lim_{n \to +\infty} \frac{1}{n} \lambda^T l_n'(\theta_0) \to 0 \text{ p.w.}$$

dla dowolnego $\lambda$ takiego, że $\lambda^T \lambda = 1$. Uwaga 4.1.2 daje nam równość $E\left(l_n'(\theta_0)\right) = 0$. Ponownie korzystając z Mocnego Prawa Wielkich Liczb Kołmogorowa oraz korzystając z nierówności

$$\lambda^T l_n'(\theta_0) \leq (\lambda^T \lambda)(l_n'^T l_n') = \|\lambda\|^2 \|l_n'\|^2$$

dostajemy $\frac{1}{n} \lambda^T l_n'(\theta_0) \to 0$ p.w., gdy $\lambda^T \lambda = 1$. Z powyższych (i)–(iii) wynika, że

$$\forall \epsilon > 0 \exists \delta^* > 0 \exists n_2 \in \mathbb{N} \forall \theta \in N(\delta^*) \forall n > n_2 :$$

$$\frac{1}{n} \lambda^T \left(-l''(\theta)\right) \lambda = \frac{1}{n} \lambda^T J_n(\theta) \lambda - \frac{1}{n} \lambda^T R_n(\theta) \lambda > c - \epsilon \text{ p.w.}$$

Z tego wynika, że $\widehat{\theta}_n \in N(\delta^*)$, gdy $n > n_2$. Z trzeciego faktu (iii) wnioskujemy, że

$$\forall \epsilon > 0 \exists \delta^* > 0 \forall \delta \in (0, \delta^*] \, \exists n^* \geq n_2 \forall \theta \in N(\delta) \forall n > n^* :$$

$$\frac{1}{n} \lambda^T l_n'(\theta_0) < \frac{1}{2} \delta (c - \epsilon) \text{ p.w.}$$

W ten sposób otrzymujemy (4.3) i (4.2) kończąc dowód twierdzenia. □

## 4.2 Asymptotyczna normalność estymatorów największej wiarygodności

W tym podrozdziale podamy założenia, których spełnianie gwarantuje asymptotyczną normalność estymatora największej wiarogodności będącego rozwiązaniem równania wiarygodności w modelach beta– i gamma–regresji.
Oznaczmy $\theta = (A_1, A_2, ..., A_m, r)$ oraz niech

$$l_n' = \left( \frac{\partial \ln L_n(\theta)}{\partial \theta_1}, \frac{\partial \ln L_n(\theta)}{\partial \theta_2}, ..., \frac{\partial \ln L_n(\theta)}{\partial \theta_{m+1}} \right),$$

gdzie $n$ oznacza, że równania wiarygodności rozpatrujemy przy $n$ danych obserwacjach. Niech $J_n$ oznacza macierz informacyjną Fishera uzyskaną z $n$ obserwacji. Jeżeli nie istnieje standardowa macierz odwrotna, to za $J_n^{-1}$ bierzemy g–odwrotność macierzy $J_n$ (patrz np. Rao str. 42). W poniższych dowodach będziemy często korzystać z faktu, że macierz Fishera oraz macierz $K$ są symetryczne i mają symetryczne pierwiastki.
Sformułujmy potrzebne założenia i komentarze: Do wymienionych wyżej założeń AN 1. i AN 2. dodajmy za Fahrmeirem i Kaufmannem (1985) str. 348

- AN 3. Załóżmy, że dla dowolnego $\delta > 0$

$$\max_{\theta \in U_n(\delta)} \|V_n(\theta) - I\| \to 0,$$

gdzie unormowana macierz informacyjna to $V_n(\theta) = J_n^{-\frac{1}{2}}(\theta_0) J_n(\theta) J_n^{-\frac{1}{2}}(\theta_0)$ oraz $U_n(\delta) = \left\{\theta \in \Theta : \|J_n^{\frac{1}{2}}(\theta_0)\,(\theta - \theta_0)\,\| < \delta\right\}$ dla $n \in \mathbb{N}$.

**Lemat 4.2.1.** *Przy założeniach AN 1. – AN 3. mamy*

$$J_n^{-\frac{1}{2}} l_n' \xrightarrow{D} \mathcal{N}(0, I), \text{ gdy } n \to +\infty,$$

*gdzie $J_n = J_n(\theta_0)$, $l_n = l_n'(\theta_0)$ oraz symbol $\xrightarrow{D}$ oznacza zbieżność według rozkładu.*

Dowód. Pokażemy, że funkcja generująca momenty odpowiadająca zmiennej $\lambda^T J_n^{-\frac{1}{2}} l_n'$ zmierza do funkcji generującej momenty zmiennej rozkładu normalnego, gdy $\lambda^T\lambda = 1$. Ustalmy $\delta > 0$ i $\lambda$ takie, że $\lambda^T\lambda = 1$. Dla $\theta_n = \theta_0 + \delta J_n^{-\frac{1}{2}}\lambda$, $n \in \mathbb{N}$ prawdą jest, że $\theta_n \in U_n(\delta)$.
Rozwijając w szereg Taylora logarytm funkcji wiarygodności, otrzymujemy

$$l_n(\theta_n) = l_n(\theta_0) + (\theta_n - \theta_0)^T l_n' + \frac{1}{2}(\theta_n - \theta_0)^T l_n''(\overline{\theta}_n)(\theta_n - \theta_0),$$

gdzie $\overline{\theta}_n = t_n\theta_n + (1 - t_n)\theta_0$ dla $t_n \in [0, 1]$.
Korzystając z Uwagi 4.1.2 dostajemy

$$\exp\left\{\lambda^T V_n(\overline{\theta}_n)\lambda \frac{\delta^2}{2}\right\} L_n(\theta_n) = \exp\left\{\delta\lambda^T J_n^{-\frac{1}{2}} l_n'\right\} L_n(\theta_0),$$

gdzie $L_n$ to funkcja wiarogodności.
Lewa strona powyższej równości jest całkowalna jako, że $\exp\left\{\lambda^T V_n(\theta)\lambda\frac{\delta^2}{2}\right\}$ jest ciągłą funkcją ze względu na $\theta$, a stąd ograniczoną na zwartym przecież odcinku $[\theta_0, \theta_n]$. Całkując powyższą równość ze względu na dominującą miarę, czyli miarę, względem której wszystkie występujące miary są absolutnie ciągłe, otrzymujemy

$$E_{\theta_n} \exp\left\{\lambda^T V_n(\overline{\theta}_n)\lambda \frac{\delta^2}{2}\right\} = E \exp\left\{\delta\lambda^T J_n^{-\frac{1}{2}} l_n'\right\}.$$

Korzystając z założenia AN 3., mamy $\overline{\theta}_n \in U_n(\delta)$ i z ciągłości funkcji exp otrzymujemy $\forall \epsilon > 0 \ \exists n_1 \in \mathbb{N} \ \forall n \geq n_1$ :

$$\left| \exp\left\{\lambda^T V_n(\overline{\theta}_n)\lambda\frac{\delta^2}{2}\right\} - \exp\left\{\frac{\lambda^T\lambda I\delta^2}{2}\right\}\right| =$$

$$= \left| \exp\left\{\lambda^T V_n(\overline{\theta}_n)\lambda\frac{\delta^2}{2}\right\} - \exp\left\{\frac{\delta^2}{2}\right\}\right| \leq \epsilon.$$

Całkując tę nierówność i korzystając z tego, że $|\int(\cdot)| \leq \int|(\cdot)|$ wnioskujemy, że

$$E_{\theta_n} \exp\left\{\lambda^T V_n(\overline{\theta}_n)\lambda\frac{\delta^2}{2}\right\} \to E\exp\left\{\frac{\delta^2}{2}\right\}, \ n \to \infty.$$

Stąd

$$E\exp\left\{\delta\lambda^T J_n^{-\frac{1}{2}} l_n'\right\} \to E\exp\left\{\frac{\delta^2}{2}\right\}, \ n \to \infty,$$

gdy $\lambda^T\lambda = 1$. Otrzymujemy z twierdzenia o ciągłości funkcji generującej momenty (Feller tom 2, str. 380 lub Jakubowski, Sztencel str. 379), że $\lambda J_n^{-\frac{1}{2}} l_n'$, $\lambda \in \mathbb{R}^{m+1}$ ma asymptotycznie jednowymiarowy standardowy rozkład normalny, co wobec dowolności $\lambda$ takiego, że $\|\lambda\| = 1$, daje nam w rezultacie tezę lematu. □

**Twierdzenie 4.2.1.** *Przy powyższych założeniach AN 1. – AN 3., otrzymany w badanym modelu gamma– i beta–regresji estymator największej wiarygodności ma rozkład asymptotycznie normalny. Inaczej*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{D} \mathcal{N}\left(0, K^{-1}(\theta_0)\right).$$

Dowód. Rozwijamy w szereg Taylora funkcję $l_n'(\theta_0)$ wokół punktu $\widehat{\theta}_n$, gdzie $\theta_0$ jest prawdziwą wartością parametru oraz $\widehat{\theta}_n$ jest rozwiązaniem równania wiarogodności, gdy liczba danych obserwacji wynosi $n$. Otrzymujemy

$$l_n'(\theta_0) = l_n'(\widehat{\theta}_n) + l''(\theta_n^*)(\theta_0 - \widehat{\theta}_n) = -l_n''(\theta_n^*)(\widehat{\theta}_n - \theta_0),$$

gdzie $\theta_n^* = t_n^*\theta_0 + (1 - t_n^*)\widehat{\theta}_n$ dle pewnego $t_n^* \in [0,1]$. Mamy także $\theta_n^* \to \theta_0$, gdy $n \to +\infty$, ponieważ $\widehat{\theta}_n \to \theta_0$. Przepiszmy teraz to równanie do następującej postaci

$$J_n^{-\frac{1}{2}} l_n' = J_n^{-\frac{1}{2}}\left\{-l_n''(\theta_n^*)\right\} J_n^{-\frac{1}{2}} J_n^{\frac{1}{2}}(\widehat{\theta}_n - \theta_0).$$

Inaczej

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \left(\frac{1}{n}J_n\right)^{-\frac{1}{2}} G_n^{-1} J_n^{-\frac{1}{2}} l_n', \tag{4.4}$$

gdzie

$$G_n = J_n^{-\frac{1}{2}} \left\{-l_n''(\theta_n^*)\right\} J_n^{-\frac{1}{2}}.$$

Można pokazać, że $G_n \to I_{m+1}$ p.w., gdy $n \to +\infty$.
Powyższa zbieżność wynika z tego, że

$$G_n = \left(\frac{1}{n}J_n\right)^{-\frac{1}{2}} \left\{\frac{1}{n}J_n(\theta_n^*) - \frac{1}{n}R_n(\theta_n^*)\right\} \left(\frac{1}{n}J_n\right)^{-\frac{1}{2}},$$

gdzie macierz $R_n(\cdot)$ definiujemy, jak poprzednio, następująco

$$R_n(\cdot) = l_n''(\cdot) + J_n(\cdot).$$

Z założenia AN 2. wynika, że

$$\frac{1}{n}J_n \to K(\theta_0) \text{ oraz } \left(\frac{1}{n}J_n\right)^{-\frac{1}{2}} \to K^{-\frac{1}{2}}(\theta_0), n \to +\infty.$$

Ponieważ $\widehat{\theta}_n \to \theta_0$ p.w., to $\frac{1}{n}J_n(\theta_n^*) \to K(\theta_0)$.
Dzięki Uwadze 4.1.1 możemy skorzystać z Mocnego Prawa Wielkich Liczb. W rezultacie dostajemy

$$\frac{1}{n}R_n(\theta_n^*) = \frac{1}{n}\left(l_n''(\theta_n^*) - E(l_n''(\theta_n^*))\right) \to 0, \text{ p.w., gdy } n \to +\infty.$$

Stąd otrzymujemy, że $G_n \to I_{m+1}$, $n \to +\infty$ p.w.
Korzystając z Lematu 4.2.1 uzyskujemy z (4.4)

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{D} \mathcal{N}\left(0, K^{-\frac{1}{2}}(\theta_0)K^{-\frac{1}{2}}(\theta_0)\right) = \mathcal{N}\left(0, K^{-1}(\theta_0)\right).$$

□

# Bibliografia


[1] Amemiya T., (1977) *The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model,* Econometrica, Vol. 45, No. 4, pp. 955–968.

[2] Amemiya T., (1983) *Non–linear regression models.* W: Z. Griliches and M. D. Intriligator, *Handbook o Econometrics*, Vol. I, pp. 333–389. North-Holland, Amsterdam.

[3] Battes D. M. i Watts D. G., (1988) *Nonlinear regression analysis*, Wiley.

[4] Bowman K. O. i Shenton L. R., (1982) *Properties of estimators for the gamma distribution*, Comm. Statist. Simulation Comput. Vol. 11(4), pp. 377–519.

[5] Bowman K. O. i Shenton L. R., (1983) *Maximum likelihood estimators for the gamma distribution revisited*, Comm. Statist. Simulation Comput. Vol. 12(6), pp. 697–710.

[6] Bowman K. O. i Shenton L. R., (1988) *Properties of Estimators for the Gamma Distribution*, Marcel Dekker, Inc., New York.

[7] Bowman K. O. i Shenton L. R., (2002) *Maximum likelihood estimators for the gamma distribution revisited*, J. Stat. Comput. Simul. Vol. 72(5), pp. 391–401.

[8] Bradley R. A. i Gart J. J., (1962) *The asymptotic properties of ML estimators when sampling from associated populations,* Biometrika, 49, 205–214.

[9] Bunke H., Henschke K., Strueby R., Wisotzki C., (1977) *Parameter estimation in nonlinear regression models,* Math. Operationsforsch. Stat., Ser. Stat., 8, pp. 23–40.

[10] Bunke H., (1980) *Parameter estimation in nonlinear regression,* W: P. R. Krishnaiah, Handbook o Statistics, Vol. 1, pp. 593–615. North-Holland, Amsterdam.

[11] Burridge J., (1981) *A note on maximum likelihood estimation for regression models using grouped data,* Journal of the Royal Statistical Society B, 43, pp. 41–45.

[12] Cox D. R. i Hinkley D. V., (1974) *Theoretical Statistics,* Chapman and Hall, London.

[13] Cramér H., (1946) *Mathematical Methods of Statistics*, Princeton University Press, Princeton, New York.

[14] Cribari–Neto F. i Vasconcellos K. L. P., (2002) *Nearly unbiased maximum likelihood estimation for the beta distribution*, J. Stat. Comput. Simul. Vol. 72(2), pp. 107–118.

[15] Crowder M., (1976) *Maximum likelihood estimation for dependent observations,* Journal of the Royal Statistical Society B, 38, pp. 45–53.

[16] Crowder M., (1986) *On consistency and inconsistency of estimating equations*, Econometric Theory, 2, pp. 305–330.

[17] Dawidowicz A. L., Kawalec E., Stanuch H., (2001) *Beta–regression*, International Society of Clinical Biostatistics, Programme and Abstracts, p. 221, Sztokholm.

[18] Fahrmeir L. i Kaufmann H., (1985) *Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models,* Ann. Statist.13, pp. 342–368.

[19] Fahrmeir L., (1987) *Asymptotic likelihood inference for nonhomogeneous observations,* Statistische Hefte/ Statistical Papers 28, pp. 81–116.

[20] Feller W., (1981) *Wstęp do rachunku prawdopodobieństwa,* PWN, Warszawa.

[21] Fichtenholz G. M., (1962) *Rachunek różniczkowy i całkowy,* PWN, Warszawa.

[22] Fisher R. A., (1922) *On the mathematical foundations of theoretical statistics.* Phil. Trans. Royal Soc. A 222, 309–368.

[23] Gallant A. R., (1987) *Nonlinear Statistical Models,* Wiley, New York.

[24] Heijmans R. i Magnus J.R., (1986) *Consistent maximum likelihood estimation with dependent observations* J. Econometrics 32, pp. 253–285.

[25] Hoadley B., (1971) *Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case,* Ann. Math. Stat., 42, pp. 1977–1991.

[26] Jakubowski J. i Sztencel R., (2004) *Wstęp do teorii prawdopodobieństwa,* Script, Warszawa.

[27] Jennrich R. I., (1969) *Asymptotic properties of nonlinear least squares estimators*, Ann. Math. Statist. 40, pp. 633–643.

[28] Johansen S., (1983) *Some topics in regression,* Scand. J . Stat., 10, pp. 161–194.

[29] Malinvaud E., (1970a) *The consistency of nonlinear regressions,* Ann. Math. Stat., 41, pp. 956–969.

[30] Malinvaud E., (1970b) *Statistical Methods of Econometrics*, translated by A. Silvey. North–Holland, Amsterdam.

[31] McCullagh P. i Nelder J. A., (1983) *Generalized Linear Models*, Chapman and Hall, New York.

[32] Mooley D. A., (1973) *An estimate of the distribution and stability period of the parameters of the gamma probability model applied to monthly rainfall over southeast Asia during the summer monsoon*, Monthly Weather Review 101(12), pp. 884–890.

[33] Morton R., (1987) *Asymmetry of estimators in nonlinear regression*, Biometrika 74, pp. 679–685.

[34] Nelder J. i Wedderburn R. W. M., (1972) *Generalized Linear Models*, Journal of the Royal Statistical Society. Series A (General), 135 (3), pp. 370–384.

[35] Nie L., (2006) *Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed–effects models,* Metrika 63, pp. 123–143.

[36] Rao C. R., (1982) *Modele liniowe statystyki matematycznej*, PWN, Warszawa.

[37] Ratkowsky D. A., (1990) *Handbook of Nonlinear Regression Models*, Marcel Dekker, New York.

[38] Rydlewski J. P., (2007) *Beta-regression model for periodic data with a trend*, Univ. Iagel. Acta Math. XLV, pp. 211–222.

[39] Rydlewski J. P., (2009) *A note on the maximum likelihood estimator in the gamma regression model*, Opuscula Mathematica 29/3, pp. 305–312.

[40] Seber G. A. F. i Wild C. J., (2003) *Nonlinear Regression*, Wiley.

[41] Sen A. i Srivastava M., (1990) *Regression Analysis. Theory, Methods and Applications*, Springer, New York.

[42] Wald A., (1949) *Note on the consistency of the maximum likelihood estimate*, Ann. Math. Stat. 20, pp. 282–291.

[43] Wedderburn R. W. M., (1976) *On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear model,* Biometrika 63, pp. 27–32.

[44] Wei B. C., (1998) *Exponential Family Nonlinear Models*, Springer, Singapore.

[45] Weiss L., (1971) *Asymptotic properties of maximum likelihood estimators in some nonstandard cases,* Journal of the American Statistical Association 66, pp. 345–350.

[46] Weiss L., (1973) *Asymptotic properties of maximum likelihood estimators in some nonstandard cases II,* Journal of the American Statistical Association 68, pp. 428–430.

[47] White H., (1980) *Nonlinear regression on cross-section data,* Econometrica, 48, pp. 721–746.

[48] White H. i Domowitz I., (1984) *Nonlinear regression with dependent observations,* Econometrica, 52, pp. 143–161.

[49] Wolfowitz J., (1949) *On Wald's proof of the consistency of the maximum likelihood estimate,* Annals of Mathematical Statistics, 20, pp. 601–602.

[50] Wu C. F., (1981) *Asymptotic theory of nonlinear least squares estimation,* Ann. Stat., 9, pp. 501–513.

[51] Zwanzig S., (1980) *The choice of approximative models in nonlinear regression,* Math. Operationsforsch. Stat., Ser. Stat, 11 , pp. 23-47.