Politechnika Gdańska
Wydział Fizyki Technicznej i Matematyki Stosowanej

Anna Wieżel
Nr albumu: 132540

Funkcjonalne Modele Liniowe

Praca magisterska
na kierunku MATEMATYKA
w zakresie MATEMATYKA FINANSOWA

Praca wykonana pod kierunkiem
dra hab. Karola Dziedziula
Katedra Analizy Matematycznej i Numerycznej

Wrzesień 2015

Oświadczenie kierujcego prac

Potwierdzam, że niniejsza praca została przygotowana pod moim kierunkiem i kwalifikuje si do przedstawienia jej w postpowaniu o nadanie tytułu zawodowego.

Data Podpis kierujcego prac

Oświadczenie autora (autorów) pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowizujcymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur zwizanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załczon wersj elektroniczn.

Data Podpis autora (autorów) pracy

Streszczenie

coś

Słowa kluczowe

funkcjonalna analiza danych, dane funkcjonalne, funkcjonalne modele liniowe, test istotności

Dziedzina pracy (kody wg programu Socrates-Erasmus)

11.1 Matematyka
11.2 Statystyka

Klasyfikacja tematyczna

62 Statistics
62-07 Data analysis
62J12 Generalized linear models

Tytuł pracy w jzyku angielskim

Functional Linear Models

Spis treści

Wstęp

Odpowiednik testu istotności dla prostego modelu regresji = F-test (+ t-test) [patrz: artykuł]

Rozdział 1

Preliminaria

Przestrzenią funkcyjną nazywać będziemy przestrzeń liniową funkcji z dowolnego zbioru X do zbioru Y.

Definicja 1.0.1 *[Ferraty, Vieu]*
Zmienną losową X nazywamy ***zmienną funkcjonalną*** *wtedy i tylko wtedy, gdy przyjmuje wartości w nieskończenie wymiarowej przestrzeni (przestrzeni funkcyjnej). Obserwację χ zmiennej X nazywamy* ***daną funkcjonalną*** *(ang. functional data).*

Jeśli zmienna funkcjonalna X (odpowiednio obserwacja χ) jest krzywą, to zachodzi $X = \{X(t),\ t \in T\}$ (odp. $\chi = \{\chi(t),\ t \in T\}$), gdzie zbiór indeksów $T \subset \mathbb{R}$. Taką zmienną funkcjonalną możemy zatem utożsamiać z procesem stochastycznym z nieskończenie wymiarową przestrzenią stanów. W szczególności, zmienna funkcjonalna może być powierzchnią, czyli dwuwymiarowym wektorem krzywych - wtedy, analogicznie, T będzie dwuwymiarowym zbiorem indeksów tj. $T \subset \mathbb{R}^2$ - lub dowolnie wymiarowym wektorem krzywych.
W niniejszej pracy skupimy się na zmiennych funkcjonalnych przyjmujących postać krzywych.

Aby zbudować pojęcia średniej oraz kowariancji dla zmiennych funkcjonalnych wprowadzimy niezbędne pojęcia z dziedziny operatorów liniowych.

1.1. Klasyfikacja operatorów liniowych

Niech $(\Omega, \mathcal{F}, P)$ będzie przestrzenią probabilistyczną, Ω jest zatem zbiorem scenariuszy ω, $\mathcal{F}$ jest σ-algebrą podzbiorów Ω, a P miarą prawdopodobieństwa nad $\mathcal{F}$. Dla uproszczenia zakładamy zupełność zadanej przestrzeni probabilistycznej. Rozważmy proces stochastyczny z czasem ciągłym $X = \{X_t,\ t \in T\}$, gdzie T jest przedziałem w $\mathbb{R}$, zdefiniowany na przestrzeni probabilistycznej $(\Omega, \mathcal{F}, P)$, taki, że $X_t(\omega)$ należy do przestrzeni funkcyjnej E dla wszystkich $\omega \in \Omega$.
W pracy rozważać będziemy zmienne funkcjonalne przyjmujące wartości w przestrzeni Hilberta.

Rozważmy ośrodkową nieskończenie wymiarową przestrzeń Hilberta H z iloczynem skalarnym $\langle \cdot, \cdot \rangle$ zadającym normę $\|\cdot\|$ i oznaczmy przez $\mathcal{L}$ przestrzeń ciągłych (ograniczonych) operatorów liniowych w H z normą

$$\|\Psi\|_{\mathcal{L}} := \sup\{\|\Psi(x)\| :\ \|x\| \leqslant 1\}.$$

Definicja 1.1.1 *[Horváth, Kokoszka]*
Operator $\Psi \in \mathcal{L}$ nazywamy ***operatorem zwartym****, jeśli istnieją dwie ortonormalne bazy $\{\nu_j\}_{j=1}^{\infty}$ i $\{f_j\}_{j=1}^{\infty}$, oraz rzeczywisty ciąg $\{\lambda_j\}_{j=1}^{\infty}$ zbieżny do zera, takie że*

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle f_j, \quad x \in H. \tag{1.1}$$

Bez straty ogólności możemy założyć, że w przedstawionej reprezentacji λ_j są wartościami dodatnimi, w razie konieczności wystarczy f_j zamienić na $-f_j$.
Równoważną definicją operatora zwartego jest spełnienie następującego warunku: zbieżność $\langle y, x_n \rangle \to \langle y, x \rangle$ dla każdego $y \in H$ implikuje $\| \Psi(x_n) - \Psi(x) \| \to 0$.

Inną klasą operatorów są operatory Hilberta-Schmidta, którą oznaczać będziemy przez $\mathcal{S}$.

Definicja 1.1.2 *[Bosq]*
Operatorem Hilberta-Schmidta *nazywamy taki operator zwarty $\Psi \in \mathcal{L}$, dla którego ciąg $\{\lambda_j\}_{j=1}^{\infty}$ w reprezentacji (1.1) spełnia $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$.*

Uwaga 1.1.1 *[Bosq], [Horváth, Kokoszka]*
Klasa $\mathcal{S}$ jest przestrzenią Hilberta z iloczynem skalarnym

$$\langle \Psi_1, \Psi_2 \rangle_{\mathcal{S}} := \sum_{j=1}^{\infty} \langle \Psi_1(e_j), \Psi_2(e_j) \rangle,$$

gdzie $\{e_j\}_{j=1}^{\infty}$ jest dowolną bazą ortonormalną w H.
Powyższy iloczyn skalarny zadaje normę $\| \Psi \|_{\mathcal{S}} := \left(\sum_{j=1}^{\infty} \lambda_j^2 \right)^{1/2}$.

Definicja 1.1.3 *[Bosq]*
Operator liniowy nazywamy ***operatorem śladowym*** *(ang. nuclear operator), jeśli równość (1.1) spełniona jest dla ciągu takiego, że $\sum_{j=1}^{\infty} |\lambda_j| < \infty$.*

Uwaga 1.1.2 *[Bosq]*
Klasa operatorów śladowych $\mathcal{N}$ z normą $\| \Psi \|_{\mathcal{N}} := \sum_{j=1}^{\infty} |\lambda_j|$ jest przestrzenią Banacha.

Definicja 1.1.4 *[Horváth, Kokoszka]*
Operator $\Psi \in \mathcal{L}$ nazywamy ***symetrycznym****, jeśli*

$$\langle \Psi(x), y \rangle = \langle x, \Psi(y) \rangle, \quad x, y \in H,$$

oraz ***nieujemnie określonym*** *(połowicznie pozytywnie określonym, ang. positive semidefinite), jeśli*

$$\langle \Psi(x), x \rangle \geqslant 0, \quad x \in H.$$

Uwaga 1.1.3 *[Horváth, Kokoszka]*
Symetryczny nieujemnie określony operator Hilberta-Schmidta Ψ możemy przedstawić w reprezentacji

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle \nu_j, \quad x \in H, \tag{1.2}$$

gdzie ortonormalne ν_j są ***funkcjami własnymi*** *Ψ, tj. $\Psi(\nu_j) = \lambda_j \nu_j$. Funkcje ν_j mogą być rozszerzone do bazy, przez dopełnienie ortogonalne podprzestrzeni rozpiętej przez oryginalne ν_j. Możemy zatem założyć, że funkcje ν_j w (1.2) tworzą bazę, a pewne wartości λ_j mogą być równe zero.*

1.2. Przestrzeń L^2

Przestrzeń $L^2 = L^2(K, \mathcal{A}, \mu)$ nad pewną przestrzenią liniową K jest zbiorem mierzalnych funkcji rzeczywistych określonych na K spełniających $\int_K x^2(t)dt < \infty$. Przestrzeń L^2 jest ośrodkową przestrzenią Hilberta z iloczynem skalarnym

$$\langle x, y\rangle := \int_K x(t)y(t)dt.$$

Tak jak zwyczajowo zapisujemy L^2 zamiast $L^2(K)$, tak w przypadku symbolu całki bez wskazania obszaru całkowania będziemy mieć na myśli całkowanie po całej przestrzeni K. Jeśli $x, y \in L^2$, równość $x = y$ zawsze oznaczać będzie $\int [x(t) - y(t)]^2 \, dt = 0$.

Ważną klasę operatorów liniowych na przestrzeni L^2 stanowią operatory całkowe.

Definicja 1.2.1 ***Operatorem całkowym*** *nazywamy operator liniowy Ψ dający się przedstawić w formie*

$$\Psi(x)(t) = \int \psi(t, s)x(s)ds, \quad x \in L^2,$$

gdzie ψ stanowi ***jądro całkowe*** *operatora Ψ.*

Uwaga 1.2.1 *[Horváth, Kokoszka]*
Operatory całkowe są operatorami Hilberta-Schmidta wtedy i tylko wtedy, gdy

$$\iint \psi^2(t, s)dtds < \infty.$$

Ponadto zachodzi

$$\| \Psi\|_{\mathcal{S}}^2 = \iint \psi^2(t, s)dtds.$$

Uwaga 1.2.2 *(Twierdzenie Mercera) [Horváth, Kokoszka]*
Jeśli operator spełnia również $\psi(s, t) = \psi(t, s)$ oraz $\iint \psi(t, s)x(t)x(s)dtds \geqslant 0$, to operator całkowy Ψjest symetryczny i nieujemnie określony, zatem z uwagi 1.1.3 mamy

$$\psi(t, s) = \sum_{j=1}^{\infty} \lambda_j \nu_j(t)\nu_j(s) \quad w\ L^2(K) \times L^2(K).$$

Jeżeli funkcja ψ jest ciągła, powyższe rozwinięcie jest prawdziwe dla wszystkich $s, t \in K$ i szereg jest zbieżny jednostajnie.

1.3. Zmienne funkcjonalne w L^2. Pojęcie operatora kowariancji

Rozważmy zmienną funkcjonalną $X = \{X(t),\ t \in T\}$ będącą krzywą ($T \subset \mathbb{R}$) jako element losowy z przestrzeni $L^2(T)$ zaopatrzonej w σ-algebrę borelowskich podzbiorów T.
Mówimy, że zmienna X jest **całkowalna**, jeśli $\mathbb{E}\,\|X\| = \mathbb{E}\left[\int X^2(t)dt\right]^{1/2} < \infty$.

Definicja 1.3.1 *[Bosq]*
Operator kowariancji *scentrowanej zmiennej funkcjonalnej X (tj. $\mathbb{E}X = 0$) przyjmującej wartości w przestrzeni funkcyjnej L^2 spełniającej $\mathbb{E}\,\|X\|^2 < \infty$ definiujemy następująco*

$$C_X(x) := \mathbb{E}[\langle X, x\rangle X], \quad x \in L^2.$$

Jeśli Y jest zmienną funkcjonalną spełniającą powyższe warunki, wtedy operator kowariancji między zmiennymi X i Y przedstawiamy następująco

$$C_{X,Y}(x) := \mathbb{E}[\langle X, x\rangle Y], \quad x \in L^2$$

oraz

$$C_{Y,X}(x) := \mathbb{E}[\langle Y, x\rangle X], \quad x \in L^2.$$

Operatory kowariancji są operatorami śladowymi,...

$$C_X(x)(t) = \int c(t,s)x(s)ds, \quad \text{gdzie } c(t,s) = \mathbb{E}\left[X(t)X(s)\right].$$

Oczywistym jest, że $c(t,s) = c(s,t)$ i mamy

$$\iint c(t,s)x(t)x(s)dtds = \iint \mathbb{E}\left[X(t)X(s)\right]x(t)x(s)dtds = \mathbb{E}\left[\left(\int X(t)x(t)dt\right)^2\right] \geqslant 0.$$

Zatem operator kowariancji C_X jest symetryczny oraz nieujemnie określony. Wartości własne λ_j operatora C_X są dodatnie i spełniony jest warunek $\sum_{j=1}^{\infty}\lambda_j = \mathbb{E}\|X\|^2 < \infty$. C_X jest operatorem Hilberta-Schmidta (a nawet operatorem śladowym) i posiada on następującą reprezentację

$$C_X(x) = \sum_{j=1}^{\infty}\lambda_j\langle x, \nu_j\rangle\nu_j, \quad x \in L^2.$$

Zakładamy, że zmienne X_n, Y_n są scentrowanymi zmiennymi losowymi przyjmującymi wartości w przestrzeni Hilberta L^2. Oznaczając przez X (analogicznie Y) losową funkcję o tym samym rozkładzie co X_n (Y_n) wprowadzamy operatory

$$C(x) = \mathbb{E}[\langle X, x\rangle X], \quad \Gamma(x) = \mathbb{E}[\langle Y, x\rangle Y], \quad \Delta(x) = \mathbb{E}[\langle X, x\rangle Y].$$

...

1.4. Funkcjonalne modele liniowe

Rozróżniamy 3 postaci funkcjonalnych modeli liniowych

(i) pełen model funkcjonalny (ang. *the fully functional model*)

$$Y(t) = \int \beta(s,t)X(s)ds + \varepsilon(t),$$

(ii) model z odpowiedzią skalarną (ang. *the scalar response model*)

$$Y = \int \beta(s)X(s)ds + \varepsilon,$$

(iii) model z odpowiedzią funkcyjną (ang. *the functional response model*)

$$Y(t) = \beta(t)x + \varepsilon(t).$$

Rozważmy pełen model funkcjonalny postaci

$$\mathbf{Y}(t) = \int \beta(s,t)\mathbf{X}(s)ds + \boldsymbol{\varepsilon}(t),$$

gdzie

$$\mathbf{Y}(t) = \begin{bmatrix} Y_1(t) \\ Y_2(t) \\ \vdots \\ Y_N(t) \end{bmatrix}, \mathbf{X}(s) = \begin{bmatrix} X_1(s) \\ X_2(s) \\ \vdots \\ X_N(s) \end{bmatrix}, \boldsymbol{\varepsilon}(t) = \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \vdots \\ \varepsilon_N(t) \end{bmatrix}.$$

Rozdział 2

Test istotności w funkcjonalnym modelu liniowym

2.1. Idea ogólna

Jednym z podstawowych testów na efektywność modelu jest test istotności zmiennych objaśniających. Zarówno jak i w przypadku klasycznego modelu liniowego w przypadku funkcjonalnego modelu liniowego badamy zerowanie się funkcji β, tj.

$$\mathrm{H}_0: \quad \boldsymbol{\beta} = 0 \quad \text{przeciw} \quad \mathrm{H}_A: \quad \boldsymbol{\beta} \neq 0.$$

Zauważmy, że przyjęcie H_0 nie oznacza braku związku między zmienną objaśnianą a objaśniającą. Prowadzi jedynie do stwierdzenia braku zależności liniowej.

Zakładamy, że zmienna objaśniana Y_n, zmienne objaśniające X_n i błędy ε_n są scentrowanymi zmiennymi losowymi przyjmującymi wartości w przestrzeni Hilberta L^2. Oznaczając przez X (analogicznie Y) losową funkcję o tym samym rozkładzie co X_n (Y_n) wprowadzamy operatory

$$C(x) = \mathbb{E}[\langle X, x\rangle\, X], \quad \Gamma(x) = \mathbb{E}[\langle Y, x\rangle\, Y], \quad \Delta(x) = \mathbb{E}[\langle X, x\rangle\, Y].$$

Przez $\hat{C}$, $\hat{\Gamma}$, $\hat{\Delta}$ oznaczamy ich estymatory, np.

$$\hat{C}(x) = \frac{1}{N}\sum_{n=1}^{N} \langle X_n, x\rangle\, X_n.$$

Definiujemy również wartości i wektory własne C i Γ

$$C(\upsilon_k) = \lambda_k \upsilon_k, \quad \Gamma(u_j) = \gamma_j u_j,$$

których estymatory będziemy oznaczać $(\hat{\lambda}_k, \hat{\upsilon}_k)$, $(\hat{\gamma}_j, \hat{u}_j)$.
Test obejmuje obcięcie powyższych operatorów na podprzestrzenie skończenie wymiarowe. Podprzestrzeń $\mathcal{V}_p = \mathrm{span}\{\upsilon_1, ..., \upsilon_p\}$ zawiera najlepsze przybliżenia X_n, które są liniowymi kombinacjami pierwszych p głównych składowych (FPC). Metodą głównych składowych wyznaczamy p największych wartości własnych operatora $\hat{C}$ tak, że $\hat{\mathcal{V}}_p = \mathrm{span}\{\hat{\upsilon}_1, ..., \hat{\upsilon}_p\}$ zawiera najlepsze przybliżenie X_n. Analogicznie $\mathcal{U}_q = \mathrm{span}\{u_1, ..., u_q\}$ zawiera przybliżenia $\mathrm{span}\{Y_1, ..., Y_N\}$.
Z równości

$$Y(t) = \int \beta(s,t)X(s)ds + \boldsymbol{\varepsilon}(t)$$

wynika $\Delta = \beta C$ i dla $k \leqslant p$ mamy

$$\beta(\upsilon_k) = \lambda_k^{-1}\Delta(\upsilon_k).$$

Stąd, β zeruje się na $\text{span}\{\upsilon_1, ..., \upsilon_p\}$ wtedy i tylko wtedy, gdy $\Delta(\upsilon_k) = 0$ dla każdego $k = 1, ..., p$. Zauważmy, że

$$\Delta(\upsilon_k) \approx \hat{\Delta}(\upsilon_k) = \frac{1}{N}\sum_{n=1}^{N} \langle X_n, \upsilon_k\rangle Y_n.$$

Skoro zatem $\text{span}\{Y_1, ..., Y_N\}$ są dobrze aproksymowane przez $\mathcal{U}_q$, to możemy ograniczyć się do sprawdzania czy

$$\left\langle \hat{\Delta}(\upsilon_k), u_j \right\rangle = 0, \quad k = 1, ..., p, \quad j = 1, ..., q. \tag{2.1}$$

Jeśli H_0 jest prawdziwa, to dla każdego $x \in \mathcal{V}_p$, $\beta(x)$ nie należy do $\mathcal{U}_q$. Co znaczy, że żadna funkcja Y_n nie może być opisana jako liniowa kombinacja X_n, $n = 1, ..., N$. Statystyka testowa powinna zatem sumować kwadraty iloczynów skalarnych (2.1). Poniższe twierdzenia prowadzą do wyznaczenia statystyki

$$\hat{T}_N(p, q) = N\sum_{k=1}^{p}\sum_{j=1}^{q} \hat{\lambda}_k^{-1}\hat{\gamma}_j^{-1}\left\langle \hat{\Delta}(\hat{\upsilon}_k), \hat{u}_j \right\rangle^2, \tag{2.2}$$

która zbiega według rozkładu do rozkładu χ^2 z pq stopniami swobody.
Przy czym

$$\left\langle \hat{\Delta}(\hat{\upsilon}_k), \hat{u}_j \right\rangle = \left\langle \frac{1}{N}\sum_{n=1}^{N} \langle X_n, \hat{\upsilon}_k\rangle Y_n, \hat{u}_j \right\rangle = \frac{1}{N}\sum_{n=1}^{N} \langle X_n, \hat{\upsilon}_k\rangle \langle Y_n, \hat{u}_j\rangle$$

oraz $\lambda_k = \mathbb{E}\langle X, \upsilon_k\rangle^2$ i $\gamma_j = \mathbb{E}\langle Y, u_j\rangle^2$.

Uwaga 2.1.1 *Oczywistym jest, że jeśli odrzucamy H_0, to $\beta(\upsilon_k) \neq 0$ dla pewnego $k \geqslant 1$. Jednak ograniczając się do p największych wartości własnych, test jest skuteczny tylko jeśli β nie zanika na którymś wektorze υ_k, $k = 1, ..., p$. Aczkolwiek takie ograniczenie jest intuicyjnie niegroźne, ponieważ test ma za zadanie sprawdzić czy główne źródła zmienności Y mogą być opisane przez główne źródła zmienności zmiennych X.*

Schemat przebiegu testu

1. Sprawdzamy założenie o liniowości metodą *FPC score predictor-response plots.*
2. Wybieramy liczbę głównych składowych p i q metodami *scree test* oraz *CPV.*
3. Wyliczamy wartość statystyki $\hat{T}_N(p, q)$ (2.2).
4. Jeśli $\hat{T}_N(p, q) > \chi^2_{pq}(1 - \alpha)$, to odrzucamy hipotezę zerową o braku liniowej zależności. W przeciwnym razie nie mamy podstaw do odrzucenia H_0.

...

2.2. Formalizm

Założenia:

1. Trójka $(Y_n, X_n, \varepsilon_n)$ tworzy ciąg niezależnych zmiennych losowych o jednakowym rozkładzie, takich że ε_n jest niezależne od X_n oraz

$$\mathbb{E}X_n = 0, \quad \mathbb{E}\varepsilon_n = 0,$$

$$\mathbb{E}\|X_n\|^4 < \infty \quad \text{i} \quad \mathbb{E}\|\varepsilon_n\|^4 < \infty.$$

2. Wartości własne operatorów C oraz Γ spełniają, dla pewnych $p > 0$ i $q > 0$

$$\lambda_1 > \lambda_2 > ... > \lambda_p > \lambda_{p+1}, \quad \gamma_1 > \gamma_2 > ... > \gamma_q > \gamma_{q+1}.$$

Twierdzenie 2.2.1 *Jeśli spełnione są H_0 i powyższe Założenia, to $\hat{T}_N(p,q) \xrightarrow{d} \chi^2_{pq}$ przy $N \to \infty$.*

Twierdzenie 2.2.2 *Przy powyższych Założeniach oraz jeśli $\langle \beta(\upsilon_k), u_j \rangle \neq 0$ dla pewnych $k \leqslant p$ oraz $j \leqslant q$, to $\hat{T}_N(p,q) \xrightarrow{P} \chi^2_{pq}$ przy $N \to \infty$.*

Dowody...

Rozdział 3

Przykład zastosowania

Magnetometer data...

Dodatek A

Kod w R

...

Bibliografia

[Bosq] D. Bosq, *Linear Processes in Function Spaces*. Springer, 2000.

[Ferraty, Vieu] F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis. Theory and practice*. Springer, 2006.

[Horváth, Kokoszka] Lajos Horváth, Piotr Kokoszka, *Interference for Functional Data with Applications*. Springer, 2012.

[I] INTERMAGNET `http://www.intermagnet.org/index-eng.php`

[Kokoszka et al. 2008] P. Kokoszka, I. Maslova, J. Sojka, L. Zhu, *Testing for lack of dependence in the functional linear model*. Canadian Journal of Statistics, 2008, 36, 207-222.

[Ramsay, Silverman] J. O. Ramsay, B. W. Silverman, *Functional Data Analysis*. Springer, 2005.