

AUTOREFERAT

Spis treści

1. Dane personalne
2. Specjalność naukowa
3. Dyplomy i stopnie naukowe
4. Informacje o dotychczasowym zatrudnieniu
5. Osiągnięcie naukowe
6. Omówienie celu naukowego i osiągniętych wyników
 - 6.1. Publikacje stanowiące osiągnięcie naukowe [H1]-[H7]
 - 6.2. Wprowadzenie
 - 6.3. Asymptotyka FDR dla zależnych testowań [H1]
 - 6.4. Oszacowanie FDP dla procedur stepdown dla niezależnych i zależnych testowań [H2]
 - 6.5. Selekcja zmiennych w parametrycznych modelach statystycznych przy użyciu procedur stepdown [H3]-[H5]
 - 6.6. Rozkład sumy dla zmiennych losowych pochodzących z kopuły Archimedesesa [H6]
 - 6.7. Parametryczna wersja algorytmu V-Box w detekcji obserwacji odstających [H7]
 - 6.8. Uwagi końcowe i cytowana literatura
- Cytowana literatura (bez autocytowań)
7. Pozostały dorobek naukowy
 - 7.1. Publikacje naukowe (po doktoracie)
 - 7.2. Syntetyczne omówienie prac
8. Spis publikacji
 - 8.1. Prace naukowe po uzyskaniu stopnia doktora
 - 8.2. Prace popularno-naukowe po uzyskaniu stopnia doktora
 - 8.3. Prace naukowe przed uzyskaniem stopnia doktora
 - 8.4. Prace wysłane do redakcji
 - 8.5. Inne publikacje
9. Dane bibliometryczne
10. Projekty badawcze
11. Nagrody i wyróżnienia
12. Spis referatów wygłoszonych na konferencjach naukowych

1. Imię i nazwisko: Konrad Artur Furmańczyk

2. Specjalność naukowa: statystyka matematyczna

3. Dyplomy i stopnie naukowe:

1996 magister w zakresie matematyki ogólnej z przygotowaniem nauczycielskim, Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

Tytuł pracy magisterskiej: „Postulat Bayesa a funkcja wiarygodności Fishera we wnioskowaniu statystycznym”

Promotor: prof. dr hab. Jacek Jakubowski

1996 magister w zakresie zastosowań matematyki, Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

Tytuł pracy magisterskiej: „Dostateczność, predykcijna dostateczność i warunkowa niezależność sigma-ciał w bayesowskich modelach statystycznych”

Promotor: prof. dr hab. Wojciech Niemirow

2004 doktor nauk matematycznych w zakresie matematyki, Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

Tytuł rozprawy doktorskiej: „Asymptotyczne własności M-estymatorów dla danych zależnych”

Promotor: prof. dr hab. Wojciech Niemirow

4. Informacje o dotychczasowym zatrudnieniu:

Od 2013 kierownik zakładu Ekonomii Matematycznej, Katedra Zastosowań Matematyki Wydziału Zastosowań Informatyki i Matematyki SGGW

Od 2008 Katedra Zastosowań Matematyki Wydziału Zastosowań Informatyki i Matematyki SGGW, adiunkt

od 2007 1/2 etatu, Zakład Profilaktyki Zagrożeń Środowiskowych i Alergologii, Warszawski Uniwersytet Medyczny, adiunkt

2007-2008 Katedra Zastosowań Matematyki Wydziału Inżynierii i Kształtowania Środowiska SGGW, adiunkt

2004-2007 Katedra Zastosowań Matematyki Wydziału Inżynierii i Kształtowania Środowiska SGGW, asystent

1996-2001 studia doktoranckie na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego

5. Osiągnięcie naukowe, o którym mowa w art. 16 ust. 2 ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. Nr 65, poz. 595, z późn. zm.): **Cykl 7 prac powiązanych tematem. "Zagadnienia testowania wielu hipotez statystycznych oraz wybrane zastosowania probabilistyczne"**.

6. Omówienie celu naukowego i osiągniętych wyników

6.1. Publikacje stanowiące osiągnięcie naukowe

- H1 Furmańczyk, K. Some remarks on the control of false discovery rate under dependence. *Applicationes Mathematicae* 40(3) (2013):297-307.
- H2 Dudziński, M., Furmańczyk, K. A note on control of the false discovery proportion. *Applicationes Mathematicae* 36(4) (2009):397-418.
- H3 Furmańczyk, K. Selection in parametric models via some stepdown procedures. *Applicationes Mathematicae* 41(1) (2014):81-92.
- H4 Furmańczyk, K. On some stepdown procedures with application to consistent variable selection in linear regression. *Statistics* 49(3) (2015):614-628.
- H5 Furmańczyk, K. Variable selection using stepdown procedures in high-dimensional linear models. *Applicationes Mathematicae* 43(2) (2016):157-172.
- H6 Furmańczyk, K. Archimedean copulas with applications to VaR estimation. *Statistical Methods and Applications* 25(2) (2016):269-283.
- H7 Furmańczyk, K., Jaworski, S. Large parametric change-point detection by a V-box control chart. *Sequential Analysis* 35(2) (2016):254-264.

6.2. Wprowadzenie

Główną tematyką moich badań stanowiących osiągnięcie naukowe, o którym mowa w art. 16 ust. 2 ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. Nr 65, poz. 595, z późn. zm.) są zagadnienia testowania wielu hipotez [H1]-[H5] oraz wybrane zastosowania probabilistyczne [H6]-[H7]. W pracach dotyczących testowania wielu hipotez rozważano asymptotyczną kontrolę FDR (ang. false discovery rate) dla zależnych testowań [H1], podano oszacowania dla FDP (ang. false discovery proportion) dla dowolnych procedur stepdown w przypadku zależnych jak i niezależnych testowań [H2]. W kolejnych pracach [H3]-[H5] związanych z testowaniem wielu hipotez badano problemy selekcji zmiennych w parametrycznych modelach statystycznych wykorzystując procedury stepdown testowania wielu hipotez. W części aplikacyjnej osiągnięcia naukowego podano wyniki dotyczące rozkładów sumy zależnych zmiennych losowych (zależność definiowana przez kopułę archimedesową) wraz z zastosowaniem do estymacji przedziałowej Value at Risk (VaR) dla zależnych szkód w ubezpieczeniach modelowanych poprzez kopuły Archimedesesa [H6]. W pracy [H7] przedstawiono adaptację algorytmu V-Box Chart służącego do wykrywania obserwacji odstających w przypadku modeli parametrycznych, wraz z mini-maksowym wyborem parametrów algorytmu. Dodatkowo podano również twierdzenia opisujące teoretyczne własności algorytmu. W kolejnych rozdziałach omówiono szczegółowo poszczególne prace stanowiące moje osiągnięcie naukowe.

6.3. Asymptotyka FDR dla zależnych testowań

Rozważamy procedurę testowania wielu hipotez, gdzie m testów jest wykonywanych jednocześnie. Niech m_0 oznacza liczbę prawdziwych hipotez zerowych, zaś $m - m_0$ liczbę fałszywych hipotez zerowych. False discovery proportion (FDP) jest określone jako proporcja błędnie odrzuconych hipotez zerowych:

$$FDP = \begin{cases} V/R & \text{gdy } R > 0 \\ 0 & \text{gdy } R = 0 \end{cases}, \quad (1)$$

gdzie V jest liczbą błędnie odrzuconych hipotez zerowych, R oznacza liczbę wszystkich odrzuconych hipotez zerowych w wyniku zastosowania procedury testowania wielu hipotez. Benjamini i Hochberg (1995) określili False Discovery Rate (FDR) - jako wartość oczekiwaną FDP

$$FDR = \mathbb{E}(FDP).$$

Procedury testowania wielu hipotez które kontrolują FDR na ustalonym poziomie mają wysoką moc, nawet wtedy gdy tysiące hipotez są wykonywane jednocześnie, co szczególnie jest ważne w zastosowaniach współczesnej biologii. Benjamini i Hochberg (1995) zaproponowali procedurę Benjaminiego-Hochberga (BH), która zapewnia kontrolę FDR w przypadku niezależnych statystyk testowych. Genovese i Wasserman (2004) pokazali, że asymptotycznie procedura BH odpowiada ustalonemu progowi u^* przy którym odrzucamy hipotezy zerowe dla których p -wartości dla poszczególnych testowanych hipotez zerowych są mniejsze niż u^* . Ostatnio ukazało się wiele prac dotyczących rozszerzenia rezultatów kontroli procedury BH na klasę zależnych statystyk testowych (Benjamini i

Yekutieli (2001), Sarkar (2002), Storey (2003), Farcomeni (2007)). Przy założeniu niezależności p -wartości Genovese i Wasserman (2004) podali wynik dotyczący asymptotycznej kontroli FDR używając metod z procesów stochastycznych. Pewny postęp został osiągnięty w pracy Farcomeniego (2007), gdzie p -wartości spełniały pewne warunki zależności m.in. warunki mixingu oraz zależności stowarzyszonej (associated). Wu (2008) rozszerzył te rezultaty na przypadek kiedy ciąg hipotez zerowych (H_i) jest stacjonarnym procesem stochastycznym przyjmującym wartości 0-1 oraz przy ustalonym ciągu (H_i) , p -wartości są niezależnymi zmiennymi losowymi. W pracy [H1] otrzymano rezultaty, które są uogólnieniem wyników otrzymanych przez Genovese i Wasserman (2004) oraz Farcomeniego (2007). W [H1] założono, że hipotezy zerowe H_i , $1 \leq i \leq m$, są i.i.d. zmiennymi losowymi z rozkładu Bernoulliego

$$\mathbb{P}(H_i = 0) = 1 - \pi \quad (2)$$

dla pewnego $0 < \pi < 1$. Przyjmujemy, że $H_i = 0$ gdy hipoteza H_i jest prawdziwa oraz $H_i = 1$ gdy hipoteza H_i jest fałszywa. Niech (p_i) to ciąg p -wartości dla pojedynczych testowań m hipotez zerowych. Dodatkowo zakładamy, że dwuwymiarowe wektory losowe (p_i, H_i) dla $i = 1, \dots, m$ są i.i.d. oraz takie, że

$$\mathbb{P}(p_i \leq t | H_i = 0) = t \quad (3)$$

$$\mathbb{P}(p_i \leq t | H_i = 1) = F(t) \quad (4)$$

dla $t \in [0, 1]$, gdzie F jest dystrybuantą p -wartości przy prawdziwości hipotezy alternatywnej. Wtedy brzegowa dystrybuanta p -wartości p_i w powyższym modelu mieszanym jest postaci

$$G(t) = \pi t + (1 - \pi)F(t) \quad (5)$$

dla $t \in [0, 1]$. Określmy proces stochastyczny:

$$\Gamma_m(t) = \frac{\sum_{i=1}^m \mathbf{1}\{p_i \leq t\} (1 - H_i)}{\sum_{i=1}^m \mathbf{1}\{p_i \leq t\} + \mathbf{1}\{p_{(1)} > t\}}$$

dla $t \in [0, 1]$, gdzie $p_{(1)} = \min\{p_1, \dots, p_m\}$.

Storey (2002) pokazał, że w modelu mieszanym dla niezależnych testowań dla ustalonego $t > 0$

$$\mathbb{E}(FDP) = \mathbb{E}(\Gamma_m(t)) = Q(t) (1 - (1 - G(t))^m),$$

gdzie

$$Q(t) = (1 - \pi) \frac{t}{G(t)}.$$

Niech

$$T_{PI} := \sup\{0 \leq t \leq 1 : Q_m(t) \leq \alpha\},$$

gdzie

$$Q_m(t) = (1 - \pi) \frac{t}{G_m(t)},$$

$$G_m(t) = \frac{1}{m} \sum_{i=1}^m \mathbf{1}\{p_i \leq t\}.$$

W modelu mieszanym dla niezależnych testowań Genovese i Wasserman (2004) otrzymali wynik

$$\mathbb{E}(\Gamma_m(T_{PI})) = \alpha + o(1)$$

przy $m \rightarrow \infty$. Niech

$$\begin{aligned}\Lambda_{0,m}(t) &= \frac{1}{m} \sum_{i=1}^m (1 - H_i) \mathbf{1}\{p_i \leq t\}, \\ \Lambda_{1,m}(t) &= \frac{1}{m} \sum_{i=1}^m H_i \mathbf{1}\{p_i \leq t\}.\end{aligned}$$

Rozważamy przestrzeń $l^\infty([0, 1])$ funkcji rzeczywistych, jednostajnie ograniczonych na odcinku $[0, 1]$ z normą

$$\|z\|_\infty = \sup_{t \in [0, 1]} |z(t)|,$$

gdzie $z : [0, 1] \rightarrow \mathbb{R}$. Jest to przestrzeń metryczna z metryką $d(z_1, z_2) = \|z_1 - z_2\|_\infty$. Podstawowym założeniem w pracy [H1] jest warunek słabej zbieżności

(D)

$$\sqrt{m}(\Lambda_{0,m}(t) - (1 - \pi)t, \Lambda_{1,m}(t) - \pi F(t)) \rightsquigarrow (Z_1(t), Z_2(t))$$

przy $m \rightarrow \infty$ w $l^\infty([0, 1]) \times l^\infty([0, 1])$, gdzie (Z_1, Z_2) jest procesem Gaussowskim o średniej zero oraz z ograniczonym jądrem kowariancji

$$K_{ij}(s, t) = \text{Cov}(Z_i(s), Z_j(t)) \tag{6}$$

dla $i, j = 0, 1$.

W pracy [H1] otrzymano także następujące wyniki:

Lemat 1. *Przy założeniu (D), mamy*

$$Z_m(t) := \sqrt{m}(\Gamma_m(t) - Q(t)) \rightsquigarrow Z(t)$$

przy $m \rightarrow \infty$ dla $t \in [\delta, 1]$ dla pewnego $\delta > 0$, gdzie Z jest procesem Gaussowskim ze średnią zero oraz jądrem kowariancji postaci

$$\begin{aligned}K(s, t) &= \frac{\pi^2 F(s)F(t)}{G^2(s)G^2(t)} K_{11}(s, t) - \frac{\pi(1 - \pi)F(t)s}{G^2(s)G^2(t)} K_{12}(s, t) \\ &\quad - \frac{\pi(1 - \pi)F(s)t}{G^2(s)G^2(t)} K_{21}(s, t) + \frac{(1 - \pi)^2 st}{G^2(s)G^2(t)} K_{22}(s, t),\end{aligned}$$

gdzie K_{ij} są określone w (6).

Twierdzenie 1. *Gdy π jest znane, to warunek (D) implikuje*

$$\mathbb{E}(\Gamma_m(T_{PI})) = \alpha + o(1) \quad (7)$$

przy $m \rightarrow \infty$.

Rozważamy p -wartości postaci

$$p_j = \tilde{G}(\dots, \eta_{j-1}, \eta_j, \eta_{j+1}, \dots), \quad (8)$$

gdzie (η_j) jest i.i.d, $\tilde{G} : \mathbb{R}^\infty \rightarrow [0, 1]$ jest pewną funkcją mierzalną. Niech $\mathcal{F}_i := (\dots, \eta_{i-1}, \eta_i)$, $\mathcal{P}_k(\xi_i^{s,t}) := \mathbb{E}(\xi_i^{s,t} | \mathcal{F}_k) - \mathbb{E}(\xi_i^{s,t} | \mathcal{F}_{k-1})$,

$$\|\mathcal{P}_0(\xi_i)\| := \sqrt{\mathbb{E}(\mathcal{P}_0(\xi_i))^2},$$

gdzie

$$\xi_i = \xi_i^{0,t}.$$

$$\xi_i^{s,t} = \mathbf{1}\{s \leq p_i \leq t\},$$

dla $s, t \in [0, 1]$. W pracy [H1] podano również wyniki kiedy spełniony jest warunek (D).

Lemat 2. *Jeśli ciąg hipotez (H_i) jest i.i.d. z rozkładu Bernoulliego, p -wartości (p_i) są postaci (8), oraz*

$$\sum_{i=1}^{\infty} \|\mathcal{P}_0(\xi_i^{s,t})\| \leq Cd(s, t)$$

dla wszystkich $s, t \in (\delta, 1]$ dla pewnego $\delta > 0$ i dla pewnej stałej $C > 0$, gdzie $d(s, t)$ jest pseudo-metryką na $[0, 1]$ taką, że przestrzeń $((\delta, 1], d)$ jest całkowicie ograniczona (p. Van der Vaart (1998)), wtedy otrzymujemy warunek (D).

W pracy [H1] rozważono również pewien szczególny przypadek modelu (8), gdzie p_j są funkcją procesu liniowego

$$p_j = g\left(\sum_{r=-\infty}^{\infty} a_r \eta_{j-r}\right), \quad (9)$$

gdzie $g : \mathbb{R} \rightarrow [0, 1]$ jest mierzalną funkcją taką, że $g \in C^1(\mathbb{R})$, $g'(x) \neq 0$ oraz

$$\int_s^t \left| (g^{-1}(u))' \right| du < Cd(s, t) \text{ dla } s, t \in (\delta, 1] \text{ dla pewnego } \delta > 0, \quad (10)$$

oraz dla pewnej stałej $C > 0$, ciąg zmiennych losowych (η_i) jest i.i.d. o ograniczonej i lipschizowskiej gęstości brzegowej f_η , oraz

$$E(\eta_1)^2 < \infty. \quad (11)$$

Dodatkowo zakładamy warunek *zależności bliskiego zasięgu* na ciąg współczynników w procesie liniowym

$$\sum_{r=-\infty}^{\infty} |a_r| < \infty. \quad (12)$$

W pracy [H1] otrzymano w przypadku gdy π jest znane następujący

Wniosek 1. *W modelu mieszanym gdy hipotezy (H_i) są ciągami zmiennych losowych i.i.d. z rozkładu Bernoulliego, p_i są funkcją procesu liniowego (9) spełniającego warunki: (10)-(12), wtedy zachodzi warunek (7).*

6.4. Oszacowanie FDP dla procedur stepdown dla niezależnych i zależnych testowań

Rozważamy problem jednoczesnego testowania hipotez zerowych $H_i, i = 1, \dots, m$. Kontrola FDP (1) wymaga, aby następujący warunek był spełniony

$$\mathbb{P}\{FDP > \gamma\} \leq \alpha \text{ dla ustalonych } \gamma, \alpha \in (0; 1). \quad (13)$$

Podamy teraz definicję *procedury stepdown*.

Niech $p_1, \dots, p_m$ będą p -wartościami dla m pojedynczych testów, $p_{(1)} \leq \dots \leq p_{(m)}$ oznaczają uporządkowane p -wartości, oraz $H_{(1)}, \dots, H_{(m)}$ będą odpowiadającymi im hipotezami.

Niech $\alpha_1 \leq \dots \leq \alpha_m$ będzie niemalejącym ciągiem stałych o wartościach zawartych w przedziale $(0, 1)$. Jeśli

$$p_{(1)} > \alpha_1, \quad (14)$$

wtedy nie odrzucamy żadnej z hipotez zerowych. W przeciwnym przypadku gdy

$$p_{(1)} \leq \alpha_1, \dots, p_{(r)} \leq \alpha_r, \quad (15)$$

to odrzucamy hipotezy $H_{(1)}, \dots, H_{(r)}$, gdzie r jest największe spełniające (15).

W szczególności procedurami stepdown są: *procedura Holm* określona jako

$$\alpha_i := \alpha / (m - i + 1) \text{ dla pewnego ustalonego } 0 < \alpha < 1, i = 1, \dots, m, \quad (16)$$

uogólniona procedura (UHolm) określona jako

$$\alpha_i := \frac{([\gamma i] + 1) \alpha}{m + [\gamma i] + 1 - i}, i = 1, \dots, m \quad (17)$$

dla pewnego $0 < \gamma < 1$. Lehmann i Romano (2005) (p. Twierdzenie 3.1) użyli procedury (UHolm) do kontroli FDP przy założeniu

$$\mathbb{P}\{q_i \leq u \mid r_1, \dots, r_{m-|I|}\} \leq u \text{ dla dowolnego } i = 1, \dots, |I| \text{ oraz } u \in (0; 1), \quad (18)$$

gdzie $q_1, \dots, q_{|I|}$ są p -wartościami odpowiadającymi $|I|$ prawdziwym hipotezom zerowym oraz $r_1, \dots, r_{m-|I|}$ są p -wartościami odpowiadającymi $m - |I|$ fałszywym hipotezom zerowym. W pracy [H2] przedstawiono wyniki dotyczące kontroli FDP, które uzupełniają wyniki Lehmana i Romano (2005). Głównymi wynikami z pracy [H2] są następujące twierdzenia.

Twierdzenie 1. Niech $q_1, \dots, q_{|I|}$ będą zmiennymi losowymi o jednakowym rozkładzie o dystrybuancie brzegowej F_q oraz $r_1, \dots, r_{m-|I|}$ będą zmiennymi losowymi o jednakowym rozkładzie o dystrybuancie brzegowej F_r . Dodatkowo załóżmy, że ciąg $\{q_1, \dots, q_{|I|}\}$ jest niezależny od ciągu $\{r_1, \dots, r_{m-|I|}\}$, oraz:

$$F_q(u) \leq u \text{ dla dowolnego } u \in (0; 1). \quad (19)$$

Wtedy dla dowolnej procedury stepdown ze statystykami $\alpha_1 \leq \dots \leq \alpha_m \leq \alpha$ oraz dla dowolnego $0 < \gamma < 1$

1⁰ dla $|I| \neq m$

$$\begin{aligned} & \mathbb{P}\{FDP > \gamma\} \\ \leq & \sum_{il} \left\{ \min\left(\frac{|I|}{l} \alpha_i, 1\right) - \max\left(\frac{|I| F_q(\alpha_i) - l}{|I| - l}, 0\right) \right\} \times \\ & \times \left\{ \min\left(\frac{m - |I|}{i - l} F_r(\alpha_i), 1\right) - \max\left(\frac{(m - |I|) F_r(\alpha_i) - (i - l)}{(m - |I|) - (i - l)}, 0\right) \right\} \\ & + \sum_{i=1}^{|I|} (1 - F_r(\alpha_{i+1})) \min\left(\frac{|I|}{i} \alpha_i, 1\right) \\ & + \sum_{i=|I|+1}^{\left[\frac{|I|-1}{\gamma}\right] \wedge (m-1)} \alpha_i \min\left(\frac{m - |I|}{m - i} (1 - F_r(\alpha_{i+1})), 1\right) \\ & + |I| \alpha_{m-|I|+1}, \end{aligned}$$

gdzie

$$\sum_{il} := \sum_{i=1}^{\left[\frac{|I|-1}{\gamma}\right] \wedge m} \sum_{l=(\lceil \gamma i \rceil + 1) \vee (i - m + |I| + 1)}^{(|I|-1) \wedge (i-1)}.$$

2⁰ dla $|I| = m$

$$\mathbb{P}\{FDP > \gamma\} \leq \sum_{j=1}^m \frac{m}{j} \alpha_j.$$

Twierdzenie 2. Niech $\{q_1, \dots, q_{|I|}\}$, $\{r_1, \dots, r_{m-|I|}\}$ będą ciągami i.i.d. o odpowiednich dystrybuantach brzegowych F_q , F_r , oraz wszystkie pozostałe założenia Twierdzenia 1 są spełnione. Wtedy dla dowolnej procedury stepdown ze statystykami $\alpha_1 \leq \dots \leq \alpha_m \leq \alpha$ oraz dla dowolnego $0 < \gamma < 1$

1^0 dla $|I| \neq m$

$$\mathbb{P}\{FDP > \gamma\}$$

$$\begin{aligned} &\leq \sum_{il} \binom{|I|}{l} (\alpha_i)^l (1 - F_q(\alpha_i))^{|I|-l} \binom{m-|I|}{i-l} (F_r(\alpha_i))^{i-l} (1 - F_r(\alpha_i))^{m-|I|-i+l} \\ &\quad + \sum_{i=1}^{|I|} (1 - F_r(\alpha_{i+1}))^{m-|I|} \sum_{j=i}^{|I|} \binom{|I|}{j} (\alpha_i)^j (1 - F_q(\alpha_i))^{|I|-j} \\ &\quad + \sum_{i=|I|+1}^{\lfloor \frac{|I|-1}{\gamma} \rfloor \wedge (m-1)} (\alpha_i)^{|I|} \sum_{j=0}^{i-|I|} \binom{m-|I|}{j} (F_r(\alpha_{i+1}))^j (1 - F_r(\alpha_{i+1}))^{m-|I|-j} \\ &\quad + \sum_{i=m-|I|+1}^{\lfloor \frac{|I|-1}{\gamma} \rfloor \wedge (m-1)} (F_r(\alpha_i))^{m-|I|} \binom{|I|}{i-m+|I|} (\alpha_i)^{i-m+|I|} (1 - F_q(\alpha_i))^{m-i} \\ &\quad + (F_r(\alpha_m))^{m-|I|} (\alpha_m)^{|I|}. \end{aligned}$$

2^0 dla $|I| = m$

$$\mathbb{P}\{FDP > \gamma\} \leq \sum_{i=1}^m \sum_{j=1}^m \binom{m}{j} (\alpha_i)^j (1 - F_q(\alpha_i))^{m-j}.$$

Kolejnymi wynikami z pracy [H2] są Twierdzenie 3 oraz Stwierdzenie 1.

Twierdzenie 3. Niech $q_1, \dots, q_{|I|}$ mają identyczny rozkład o dystrybuancie brzegowej F_q oraz zachodzi (19). Wtedy dla dowolnej procedury stepdown ze statystykami $\alpha_1 \leq \dots \leq \alpha_m \leq \alpha$ oraz dla dowolnego $0 < \gamma < 1$

$$\mathbb{P}\{FDP > \gamma\} \leq \sum_{j=1}^{|I|} \frac{|I|}{j} \alpha_j + \frac{|I|}{C(\gamma)} \alpha_{m-|I|+C(\gamma)},$$

gdzie $C(\gamma) := \min \left(\left\lceil \frac{\gamma}{1-\gamma} \right\rceil + 1, |I| \right).$

Mówimy, że n -wymiarowy wektor losowy ma rozkład MTP_2 (multivariate totally positive of order two), jeśli jego gęstość łączna $f(x)$ spełnia warunek

$$f(x \vee y) f(x \wedge y) \geq f(x) f(y) \text{ dla dowolnych } x, y \in \mathbb{R}^n,$$

gdzie: $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, $x \vee y = (\max(x_1, y_1), \dots, \max(x_n, y_n))$, $x \wedge y = (\min(x_1, y_1), \dots, \min(x_n, y_n))$.

Stwierdzenie 1. Niech $(q_1, \dots, q_{|I|})$ jest wektorem losowym o rozkładzie MTP_2 oraz niech $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_s \leq \alpha$ jest taki, że ciąg $\left\{\frac{\alpha_i}{i}\right\}_{i=1}^M$ jest niemalejący, gdzie $M := \min([\gamma m] + 1, |I|)$. Wtedy dowolna procedura *stepdown* kontroluje FDP na poziomie α .

6.5. Selekcja zmiennych w parametrycznych modelach statystycznych przy użyciu procedur *stepdown*

W pracach [H3]-[H5] rozważano problem selekcji zmiennych w parametrycznych modelach statystycznych przy użyciu *procedur stepdown* testowania wielu hipotez. Oprócz wyników teoretycznych dotyczących zgodnej selekcji w pracach tych przedstawiono studium symulacji weryfikujących moc stosowanych *procedur stepdown* użytych do wyboru selekcji zmiennych w modelach parametrycznych. W pracy Bunea i inni (2006) pokazano, że procedura Benjaminiego-Yekutieli (BY) oraz procedura Bonferroniego zastosowane do selekcji zmiennych w regresji liniowej oraz regresji logistycznej prowadzą do zgodnego wyboru zmiennych. W pracach [H3]-[H5] przy ogólnych założeniach pokazano, że procedury *stepdown* prowadzą do zgodnej selekcji zmiennych w modelu regresji liniowej, regresji logistycznej oraz regresji Coxa. Rozważmy klasę modeli $\mathcal{M}_\beta$ indeksowanych parametrem $\beta \in \mathbb{R}^p$, gdzie p jest liczbą predyktorów w modelu parametrycznym. Prawdziwy model, który jest nieznany, jest wyspecyfikowany następująco:

$$\mathcal{M}_\beta^{true} : \beta_i \neq 0 \text{ dla } i \in I_0 \text{ oraz } \beta_i = 0 \text{ dla } i \in I_1,$$

gdzie $I_1 := \{1, \dots, p\} \setminus I_0$ ($|I_0| = p_0$), $\beta = (\beta_1, \dots, \beta_p)^T$. Ponadto zakładamy, że liczba istotnych zmiennych p_0 jest ustalona i niezależna od n - rozmiaru próby.

Problem wyboru zmiennych jest równoważny problemowi estymacji zbioru istotnych zmiennych I_0 . W typowej sytuacji rozkład obserwowanej zmiennej objaśnianej zależy od parametrów modelu β_i , $i = 1, \dots, p$. Na podstawie statystyk testowych t_{ni} rozważamy problem testowania wielu hipotez:

(h₀)

$$H_i : \beta_i = 0 \text{ vs. } H'_i : \beta_i \neq 0, \text{ dla } i = 1, \dots, p.$$

Procedura selekcji zmiennych w modelu parametrycznym $\mathcal{M}_\beta$ może być opisana jako wybór zbioru $\hat{I}$ wszystkich indeksów $i \in I_1$ dla których hipotezy zerowe H_i są odrzucane, oraz jest zgodna gdy

$$\mathbb{P}(\hat{I} = I_0) \rightarrow 1 \text{ przy } n \rightarrow \infty.$$

Rozważano następujące przypadki. W pracy [H3] analizowano przypadek p ustalonego niezależnie od rozmiaru próby n gdy $p < n$. W pracy [H4] badano przypadek gdy p zależy od n i $p(n) < n$. Zaś w pracy [H5] rozważono przypadek wysokowymiarowy tzn. p zależy od n oraz $p(n) \gg n$. Będziemy rozważali m.in. w następującej postaci *procedury stepdown*:

a) *procedura Holm*

$$\alpha_j = \frac{q_n}{p + 1 - j},$$

b) procedura *UHolm* dla

$$\alpha_j = \frac{([\gamma j] + 1) q_n}{p + [\gamma j] + 1 - j} \text{ dla pewnego } 0 < \gamma < 1,$$

c) procedura *Bonferroniego*

$$\alpha_j = \frac{q_n}{p},$$

gdzie $j = 1, \dots, p$. Ponadto zakłada się, że $q_n \rightarrow 0$ przy $n \rightarrow \infty$ w odpowiedni sposób.

Przypadek $p < n$ oraz p jest ustalone

Zakładamy, że wszystkie statystyki t_{ni} , pod warunkiem prawdziwości H_i mają rozkład asymptotycznie normalny. Bardziej formalnie zakładamy, że dla pewnego estymatora $\hat{\beta}$ istnieje odwracalna macierz $\mathbb{M}_n$ zależna od rozmiaru próby n , taka że

$$\mathbb{M}_n^{\frac{1}{2}} (\hat{\beta} - \beta) \rightarrow_d N_p(\mathbf{0}, \mathbf{I}), \quad (20)$$

przy $n \rightarrow \infty$, gdzie $\mathbf{I}$ jest macierzą identyczności $p \times p$. Dodatkowo zakładamy, że

$$\frac{\mathbb{M}_n}{n} \rightarrow \mathbb{V} > 0. \quad (21)$$

Statystyki testowe t_{ni} są postaci

$$t_{ni} = \frac{\hat{\beta}_i}{\sqrt{m_{ii}}} \quad (22)$$

dla $i = 1, \dots, p$, gdzie $\mathbb{M}_n^{-1} = (m_{ij})_{1 \leq i, j \leq p}$.

Na mocy (20), przyjmujemy, że p -wartości dla t_{ni} są postaci

$$\pi_i(t_{ni}) = 2(1 - \Phi(|t_{ni}|)),$$

gdzie Φ jest dystrybucją standardowego rozkładu normalnego. Niech $q_1, \dots, q_{p-p_0}$ oznaczają p -wartości dla prawdziwych hipotez zerowych w (h_0) , oraz $r_1, \dots, r_{p_0}$ oznaczają p -wartości dla fałszywych hipotez zerowych. Niech F_j oznacza dystrybucję zmiennej losowej r_j dla $j = 1, \dots, p$.

Rozważmy następujące warunki:

(A1) $\alpha_p \rightarrow 0$ as $n \rightarrow \infty$.

(A1a)

$$\frac{\log\left(\frac{1}{\alpha_p}\right)}{n} \rightarrow 0 \text{ przy } n \rightarrow \infty.$$

(A2)

$$\max_{j \in I_0} (1 - F_j(\alpha_1)) \rightarrow 0 \text{ przy } n \rightarrow \infty.$$

Najważniejszym wynikiem teoretycznym w pracy [H3] jest następujące

Twierdzenie 1. *Procedura stepdown spełniająca warunki (A1)-(A2), przy warunkach (20)-(21) jest zgodną procedurą selekcji w parametrycznych modelach.*

W pracy [H3] pokazano, że przy pewnych warunkach regularności modele regresji liniowej, logistycznej oraz Coxa spełniają warunki (20)-(21).

Przypadek $p < n$ oraz $p = p(n)$

Niech $\mathbb{Y} = (Y_1, \dots, Y_n)^T$, $\mathbb{X}$ będzie nielosową macierzą o elementach x_{ij} , $1 \leq i \leq n$, $1 \leq j \leq p$, niech $\beta = (\beta_1, \dots, \beta_p)^T$ będzie wektorem nieznanym współczynników. Rozważamy model liniowy

$$\mathbb{Y} = \mathbb{X}\beta + \varepsilon, \quad (23)$$

gdzie $\beta_j \neq 0$ dla $j \in I_0$ i $\beta_j = 0$ dla $j \in I_1$, $I_1 := \{1, \dots, p\} \setminus I_0$ ($|I_0| = p_0$), oraz $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ jest wektorem błędów losowych.

Zakładamy, że istnieje macierz $(\mathbb{X}^T \mathbb{X})^{-1} = (m_{ij})_{1 \leq i, j \leq p}$ dla dostatecznie dużych n , oraz $(\varepsilon_i)_{i=1}^\infty$ jest ciągiem ściśle stacjonarnym, takim, że $\mathbb{E}(\varepsilon_i) = 0$, $\mathbb{E}(\varepsilon_i^2) = \sigma^2$. Dodatkowo rozważamy następujące warunki:

(B0)

$$\max_{1 \leq i \leq p} \|G_{ni} - \Phi\|_\infty = \mathcal{O}(c_n),$$

gdzie $c_n \rightarrow 0$ przy $n \rightarrow \infty$, Φ jest dystrybucją standardowego rozkładu normalnego oraz G_{ni} jest dystrybucją zmiennej losowej

$$\tilde{t}_{ni} = \frac{\hat{\beta}_i - \beta_i}{\sigma \sqrt{m_{ii}}}.$$

Niech $\hat{\beta}_i$ będzie estymatorem najmniejszych kwadratów parametru β_i w modelu (23) oraz $se(\hat{\beta}_i)$ będzie jego błędem standardowym. Na podstawie statystyk $t_{ni} = \hat{\beta}_i / se(\hat{\beta}_i)$, gdzie

$$se(\hat{\beta}_i) = \begin{cases} \sigma \sqrt{m_{ii}} & \text{gdy } \sigma \text{ jest znane,} \\ S \sqrt{m_{ii}} & \text{gdy } \sigma \text{ jest nieznanne,} \end{cases}$$

S to zgodny estymator σ , rozważamy problem testowania wielu hipotez (h_0). Zakładamy, że p -wartości dla pojedynczego testowania H_i vs. H'_i na podstawie t_{ni} są postaci

$$\pi_i(t_{ni}) = 2(1 - \Phi(|t_{ni}|)).$$

Rozważmy następujące warunki:

(A0)

$$pc_n \rightarrow 0 \text{ przy } n \rightarrow \infty,$$

gdzie c_n jest takie jak w (B0),

(A1)

$$p\alpha_p \rightarrow 0 \text{ przy } n \rightarrow \infty,$$

(A2)

$$\max_{j \in I_0} (1 - F_j(\alpha_1)) \rightarrow 0 \text{ przy } n \rightarrow \infty.$$

(C)

$$p\mathbb{P}(|S - \sigma| > \delta) \rightarrow 0, \text{ dla dowolnego } \delta > 0, \text{ przy } n \rightarrow \infty.$$

Głównym wynikiem w pracy [H4] jest

Twierdzenie 2. *Dowolna procedura stepdown spełniająca (A1)-(A2), przy warunkach (A0), (B0) gdy σ jest znane, lub przy warunkach (A0), (B0), (C) gdy σ jest nieznanne, jest zgodną procedurą selekcji w modelu liniowym.*

W pracy [H4] rozważano również osłabienie warunku niezależności błędów w regresji liniowej.

(MD) (ε_i) jest ściśle stacjonarnym ciągiem różnic martyngałowych w $\mathbb{L}^3$ spełniających pewien warunek regularności;

(LP) (ε_i) jest procesem liniowym postaci $\varepsilon_i = \sum_{r \in \mathbb{Z}} b_r \xi_{i-r}$, gdzie proces innowacji (ξ_i) jest scentrowanym ściśle stacjonarnym ciągiem różnic martyngałowych w $\mathbb{L}^3$ spełniającym pewien warunek regularności, oraz

$$\sum_{r \in \mathbb{Z}} |b_r| < \infty, \quad (24)$$

$$\min_{1 \leq i \leq p} \left| \sum_{k=1}^{\infty} b_k a_{k,i} \right| > C \text{ dla pewnej stałej } C > 0. \quad (25)$$

Niech $h_n := \max_{1 \leq k \leq n} \sqrt{h_{kk}}$, gdzie h_{kk} jest elementem diagonalnym macierzy rzutu $\mathbb{H} := \mathbb{X} (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T$, oraz $m_n := \max_{1 \leq k \leq p} m_{kk}$.

Następujące Stwierdzenie 1 omawia kwestie kiedy warunki Twierdzenia 2 są spełnione.

Stwierdzenie 1. *i) Gdy σ jest znana, ciąg (ε_i) jest i.i.d., macierz rzutu $\mathbb{H}$ spełnia warunek $ph_n \rightarrow 0$, oraz $m_n = O\left(\frac{1}{\log n}\right)$, $p = O\left(\frac{n}{\log n}\right)$ przy $n \rightarrow \infty$, wtedy założenia Twierdzenia 2 są spełnione dla procedur Bonferroniego, Holm, UHolm dla $q_n \sim \frac{1}{\log n}$.*

ii) Gdy σ jest nieznanne, ciąg (ε_i) jest i.i.d., $E|\varepsilon_1|^{4\lambda} < \infty$ dla pewnego $\lambda > 1$, $p = O\left(\frac{n}{\log n}\right)$ przy $n \rightarrow \infty$, macierz rzutu $\mathbb{H}$ spełnia warunek $ph_n \rightarrow 0$, oraz $m_n = O\left(\frac{1}{\log n}\right)$, wtedy założenia Twierdzenia 2 są spełnione dla procedur Bonferroniego, Holm, UHolm dla $q_n \sim \frac{1}{\log n}$.

iii) Gdy σ jest znane, (ε_i) jest ciągiem różnic martyngałowych (MD) lub (ε_i) jest procesem liniowym (LP), oraz $m_n = O\left(\frac{1}{\log n}\right)$, $p = O\left(\frac{n}{\log n}\right)$, $h_n = O(n^{-1/2})$ przy $n \rightarrow \infty$, wtedy założenia Twierdzenia 2 są spełnione dla procedur Bonferroniego, Holm, UHolm dla $q_n \sim \frac{1}{\log n}$.

Przypadek $p = p(n) \gg n$

Rozważmy regresję grzbietową

$$\begin{aligned} \hat{\beta} &= \arg \min_{\beta} \|\mathbb{Y} - \mathbb{X}\beta\|_2^2 / n + \lambda \|\beta\|_2^2 \\ &= \left(n^{-1} \mathbb{X}' \mathbb{X} + \lambda \mathbf{I} \right)^{-1} n^{-1} \mathbb{X}' \mathbb{Y}, \end{aligned} \quad (26)$$

gdzie $\lambda = \lambda_n$ jest parametrem regularyzacji, $\mathbf{I}$ jest macierzą identyczności. Niech $\mathbb{X} = \mathbf{R}\mathbf{S}\mathbf{V}'$ będzie dekompozycją SVD. Oznaczmy $\mathcal{R}(\mathbb{X}) \subset \mathbb{R}^p$ jako przestrzeń liniową generowaną

przez n wierszy macierzy $\mathbb{X}$. Wtedy macierz rzutu z $\mathbb{R}^p$ na $\mathcal{R}(\mathbb{X})$ jest postaci $P_{\mathbb{X}} = \mathbb{X}' (\mathbb{X}\mathbb{X}')^{-} \mathbb{X} = \mathbf{V}\mathbf{V}'$, gdzie $(\mathbb{X}\mathbb{X}')^{-}$ oznacza macierz pseudo odwrotną macierzy $\mathbb{X}\mathbb{X}'$. Niech $\theta := P_{\mathbb{X}}\beta = \mathbf{V}\mathbf{V}'\beta$. Skorygowany estymator regresji grzbietowej jest postaci (p. Bühlmann (2013))

$$\hat{\beta}_{corr,j} = \hat{\beta}_j - \sum_{k \neq j} (P_{\mathbb{X}})_{j,k} \hat{\beta}_{int,k} \quad (27)$$

dla $j = 1, \dots, p$. Niech $\hat{\Sigma} = \frac{1}{n} \mathbb{X}' \mathbb{X}$. Wtedy $Cov(\hat{\beta}) = n^{-1} \sigma^2 \Omega$, gdzie

$$\Omega = \Omega(\lambda) = (\hat{\Sigma} + \lambda \mathbf{I})^{-1} \hat{\Sigma} (\hat{\Sigma} + \lambda \mathbf{I})^{-1}.$$

Niech

$$a_{n,j} := \sqrt{n} \sigma^{-1} \Omega_{j,j}^{-1/2}.$$

Rozważmy następujące warunki:

(A) Istnieją takie stałe $\Delta_{j,n} > 0$ że

$$\mathbb{P} \left(\bigcap_{j=1}^p \left\{ \left| a_{n,j} \sum_{k \neq j} (P_{\mathbb{X}})_{j,k} (\hat{\beta}_{int,k} - \beta_k) \right| \leq \Delta_{j,n} \right\} \right) \rightarrow 1 \text{ przy } n \rightarrow \infty.$$

(B) Parametr regularyzacji $\lambda = \lambda_n$ spełnia warunek

$$\lambda_n (\Omega_{\min}(\lambda_n))^{-1/2} = o \left(n^{-1/2} \|\theta\|_2^{-1} \lambda_{\min \neq 0}(\hat{\Sigma}) \right) \text{ przy } n \rightarrow \infty,$$

gdzie $\Omega_{\min}(\lambda) = \min_{j=1, \dots, p} \Omega_{j,j}(\lambda) > 0$ oraz $\lambda_{\min \neq 0}(\hat{\Sigma})$ jest najmniejszą różną od zera wartością własną macierzy $\hat{\Sigma}$.

Rozważamy problem testowania p hipotez (h_0) . Podobnie jak Bühlmann (2013), zakładamy, że p -wartość dla testowania pojedynczej hipotezy H_i vs. H'_i jest postaci

$$\pi_i = 2 \left(1 - \Phi \left(\left(a_{n,i} \left| \hat{\beta}_{corr,i} \right| - \Delta_{i,n} \right)_+ \right) \right) \quad (28)$$

dla $i = 1, \dots, p$.

W przypadku gdy błędy losowe w modelu regresji liniowej są gaussowskie rozważamy następujące warunki:

(C) $p\alpha_p \rightarrow 0$ przy $n \rightarrow \infty$.

(D) $\left| (P_{\mathbb{X}})_{j,j} \beta_j \right| \geq n^{-c}$ dla pewnego $c \in (0, 1/2)$ oraz $\Delta_{j,n} = \mathcal{O}(1)$, $a_{n,j} > \sqrt{n}$ dla każdego $j \in I_0$, oraz $n^{1/2-c} - \Phi^{-1}(1 - \frac{\alpha_1}{2}) \rightarrow \infty$ przy $n \rightarrow \infty$.

Poniżej przedstawiamy główne wyniki otrzymane w pracy [H5].

Twierdzenie 3. *Dowolna procedura stepdown spełniająca warunki (A)-(D) jest zgodną procedurą selekcji w modelu regresji liniowej.*

W przypadku niegaussowskim zakładamy, że błędy (ε_i) są i.i.d. takie, że $\mathbb{E}(\varepsilon_1) = 0$, $\mathbb{E}|\varepsilon_1|^3 < \infty$. Zamiast warunku (B) rozważamy warunek

(B₁) Parametr regularyzacji $\lambda = \lambda_n$ spełnia warunek

$$\lambda_n (\mathbf{\Omega}_{\min}(\lambda_n))^{-1/2} = o\left(p^{-1}n^{-1/2} \|\theta\|_2^{-1} \lambda_{\min \neq 0}(\hat{\Sigma})\right) \text{ przy } n \rightarrow \infty.$$

Twierdzenie 4. *Dowolna procedura stepdown spełniająca warunki (A), (B₁), (C), (D) jest zgodną procedurą selekcji w modelu regresji liniowej.*

W pracy [H5] podano też przykłady procedur stepdown, kiedy założenia powyższych twierdzeń są spełnione wraz ze studium symulacji numerycznych.

6.6. Rozkład sumy dla zmiennych losowych pochodzących z kopuły Archimedesesa

Niech $X_1, \dots, X_n$ będą zmiennymi losowymi o rozkładzie absolutnie ciągłym o dystrybuantach brzegowych F_i oraz gęstościach brzegowych f_i , $i = 1, \dots, n$. Łączny rozkład wektora losowego $(X_1, \dots, X_n)$ jest absolutnie ciągły jak też odpowiednia kopuła $C : [0, 1]^n \rightarrow [0, 1]$. Wtedy z Twierdzenia Sklara (1959)

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = C(F_1(x_1), \dots, F_n(x_n))$$

dla dowolnego $(x_1, \dots, x_n) \in \mathbb{R}^n$.

Wektor losowy $(X_1, \dots, X_n)$ ma kopułę Archimedesesa o generatorze φ gdy

$$C(u_1, \dots, u_n) = C_\varphi(u_1, \dots, u_n) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_n)), \quad (29)$$

dla dowolnego $u_i \in [0, 1]$, $i = 1, \dots, n$, gdzie φ jest ciągłą i ściśle malejącą funkcją z $[0, 1]$ do $[0, \infty)$ taką, że $\varphi(0) = \infty$, $\varphi(1) = 0$, oraz pochodne φ^{-1} wszystkich rzędów spełniają warunek

$$(-1)^i \frac{d^i}{du^i} \varphi^{-1}(u) \geq 0$$

dla wszystkich $u \in (0, 1)$, $i = 1, 2, \dots$. Wtedy z Twierdzenia Bernsteina (1928) istnieje zmienna losowa $\Theta \geq 0$ taka, że φ^{-1} jest transformatą Laplace zmiennej Θ ,

$$\varphi^{-1}(s) = \mathbb{E} \exp(-\Theta s). \quad (30)$$

Dodatkowo zakładamy, że

$$\mathbb{P}(X_i \leq x \mid \Theta = \theta) = (G_i(x))^\theta, \quad (31)$$

gdzie

$$G_i(x) = \exp(-\varphi(F_i(x))). \quad (32)$$

Przy tych założeniach zmienne losowe $X_1, \dots, X_n$ są warunkowo niezależne pod warunkiem $\Theta = \theta$.

W analizach ryzyka ubezpieczeniowego warto rozważać rozkład zagregowanych szkód z poszczególnych portfeli $X_1, \dots, X_n$ pochodzących z kopuły Archimedeses (rozkład łączny jest wyznaczony przez kopułę Archimedeses) tzn. rozkład sumy

$$S_n = \sum_{i=1}^n X_i. \quad (33)$$

Wüthrich (2003) oraz Alink et al. (2004)-(2005) podali rozkład asymptotyczny dla (33) przy pewnych warunkach regularności na generator φ oraz podali jego zastosowanie do estymacji Value-at-Risk (VaR). Nasz główny wynik z pracy [H6] Twierdzenie 1 podaje nieasymptotyczny rozkład sumy (33) dla kopuł Archimedeses bez dodatkowych restrykcji na postać generatora φ oraz bez ograniczenia się do jednakowych rozkładów brzegowych F_i , $i = 1, \dots, n$.

Twierdzenie 1. *Jeśli $X_1, \dots, X_n$ pochodzą z kopuły Archimedeses (29)-(32), wtedy dla dowolnego $n \geq 2$,*

$$\mathbb{P}(S_n \leq t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\varphi^{-1}(s))^{(n-1)} \big|_{s=s_{n-1}(t)} \prod_{i=1}^{n-1} \varphi'(F_{i+1}(u_i)) f_{i+1}(u_i) du_i \quad (34)$$

(jeśli powyższa całka istnieje), gdzie

$$s_{n-1}(t) = \sum_{i=1}^{n-1} \varphi(F_{i+1}(u_i)) + \varphi(F_1(t - \sum_{i=1}^{n-1} u_i)),$$

$$(\varphi^{-1}(s))^{(n-1)} = \frac{d^{n-1}}{du^{n-1}} \varphi^{-1}(s) \text{ oraz } f_i = F'_i.$$

Poniższe Stwierdzenie 1 w pracy [H6] o rozkładzie statystyk pozycyjnych dla kopuły Archimedeses posłużyło nam do konstrukcji nieparametrycznego przedziału ufności dla VaR dla S_n .

Stwierdzenie 1. *Jeśli $X_1, \dots, X_n$ pochodzą z kopuły Archimedeses (29)-(30), wtedy*

$$\mathbb{P}(X_{m:n} \leq x) = \sum_{j=m}^n (-1)^{j-m} \binom{j-1}{m-1} \binom{n}{j} \varphi^{-1} \left(\sum_{i=1}^j \varphi(F_i(x)) \right). \quad (35)$$

Stąd konstruujemy następujący przedział ufności dla $VaR_\gamma(S_n)$

$$\mathbb{P}(VaR_\gamma(S_n) \geq nX_{r:n}) \geq 1 - \beta,$$

gdzie $1 - \beta$ jest poziomem ufności. Korzystając z oszacowania (p. Mesifioui i Quessy (2005)) dla $\gamma \leq F(x^*)$,

$$VaR_\gamma(S_n) \geq nF^{-1} \left(\frac{\gamma}{n} \right),$$

gdzie F jest brzegową dystrybucją X_i , $i = 1, \dots, n$, oraz $f(x) = dF(x)/dx$ jest niemalejąca dla $x \leq x^*$ dla pewnego punktu x^* . Wtedy wystarczy wyznaczyć r takie, że

$$\mathbb{P} \left(X_{r:n} \leq F^{-1} \left(\frac{\gamma}{n} \right) \right) \geq 1 - \beta.$$

Stąd r wyznaczamy z nierówności

$$\sum_{j=r}^n \binom{n}{j} (-1)^{j-r} \binom{j-1}{r-1} \varphi^{-1} \left(j \varphi \left(\frac{\gamma}{n} \right) \right) \geq 1 - \beta.$$

6.7. Parametryczna wersja algorytmu V-Box w detekcji obserwacji odstających

W klasycznej teorii kart kontrolnych zakłada się, że obserwowany proces $\{Y_i\}$ jest określony jako

$$Y_i = \theta \mathbf{1}(i \geq q) + \xi_i,$$

gdzie (ξ_i) jest ciągiem niezależnych zmiennych losowych, θ jest przesunięciem oraz $q > 0$ jest nieznanym parametrem momentu skoku procesu. W tym ujęciu powszechnie się przyjmuje, że proces szumu (ξ_i) jest gaussowski lub ma inny ciągły rozkład. W pracy Rafajłowicza i inni (2010) pojawiła się koncepcja ruchomego pionowego okna (V-Box Chart). Ten algorytm możemy opisać następująco:

1. W danej chwili t obliczamy wartość statystyki

$$T = \sum_{i=1}^{t-1} \mathbf{1}(Y_t - H \leq Y_i \leq Y_t + H).$$

2. Jeśli wartość ta jest większa niż $\gamma(t-1)$, to t zastępujemy przez $t+1$ i pierwszy krok algorytmu jest powtarzany. W przeciwnym przypadku algorytm zatrzymuje się oraz $q = t$ jest jego wynikiem.

Parametry $H > 0$ i $\gamma \in (0, 1)$ są ustalone. Ich wybór zależy od założeń o rozkładzie (ξ_i) oraz od wartości H i γ .

W pracy [H7] rozważamy m.in. następujący model zmian (*change-point*):

$$Y_i = \theta \mathbf{1}(i \geq N+1) + \xi_i, \quad i = 1, \dots, N,$$

gdzie $\theta \in \mathbb{R}$ oraz $\xi_1, \dots, \xi_{N+1}$ są i.i.d. o ciągłej dystrybucji F .

Rozważono problem decyzyjny: $d_1 : |\theta| \leq \varepsilon$ lub $d_2 : |\theta| > \varepsilon$, dla ustalonego $\varepsilon > 0$. Zgodnie z koncepcją algorytmu V-Box Chart następująca klasa reguł decyzyjnych jest rozważana:

$$\mathbf{D} = \{d(\gamma, H) : \gamma \in (0, 1), H > 0\},$$

gdzie

$$d(\gamma, H) = \begin{cases} d_1 & \text{gdy } T_N(H) > \gamma N, \\ d_2 & \text{gdy } T_N(H) \leq \gamma N, \end{cases}$$

oraz $T_N(H) = \sum_{i=1}^N \mathbf{1}(Y_{N+1} - H \leq Y_i \leq Y_{N+1} + H)$.

W pracy [H7] użyto następującego mini-maksowego kryterium optymalizacji dla wyboru parametrów γ i H :

$$\sup_{\theta} \mathbb{E}_{\theta} L(d(\gamma^{opt}, H^{opt}), \theta) = \inf_{\gamma, H} \sup_{\theta} \mathbb{E}_{\theta} L(d(\gamma, H), \theta),$$

gdzie funkcja straty jest następująca:

$L(d_i, \theta)$	$ \theta \leq \varepsilon$	$ \theta > \varepsilon$
d_1	0	b
d_2	a	0

dla ustalonych wartości $a, b > 0$. W pracy [H7] podano również twierdzenia dotyczące asymptotycznego rozwiązania tego problemu mini-maksowego w przypadku ciągłym jak i dyskretnym. Uzyskane wyniki są ilustrowane numerycznymi przykładami zastosowania zaimplementowanego algorytmu w przypadku gdy dane pochodzą z rozkładu normalnego jak też z rozkładu Poissona.

6.8. Uwagi końcowe i cytowana literatura

Cytowana literatura (bez autocytowań)

- Alink, S, Löwe, M and Wüthrich, M. V. *Diversification of aggregate dependent risks*. Insurance: Math. and Econ. 35 (2004):77-95.
- Alink, S, Löwe, M and Wüthrich, M. V. *Analysis of the expected shortfall of aggregate dependent risks*. ASTIN Bull. 35 (2005):25-43.
- Benjamini Y, Hochberg Y. *Controlling the false discovery rate: A practical and powerful approach to multiple testing*. J. Roy. Statist. Soc. Ser. B 57 (1995):289-300.
- Benjamini Y, Yekutieli Y. D. *The control of the false discovery rate in multiple testing under dependency*. Ann. Statist. 29 (2001):1165-1188.
- Bernstein. S. N. *Sur les fonctions absolument monotones*. Acta Mathematica, 52 (1928):1-66.
- Bunea, F., Wegkamp, M.H., Auguste, A. (2006). *Consistent variable selection in high dimensional regression via multiple testing*. J. Statist. Plann. Infer. 136 (2006):4349-4364.
- Bühlmann, P. *Statistical significance in high-dimensional linear model*. Bernoulli 19 (2013):1212-1242.
- Farcomeni, A. *Some results on the control of the false discovery rate under dependence*. Scand. J. Statist. 34 (2007):275-297.
- Genovese, C.R, Wasserman, L. *A stochastic process approach to false discovery control*. Ann. Statist. 32 (2004):1035-1061.
- Holm, S. *A simple sequentially rejective multiple test procedure*. Scand. J. Statist. 6 (1979):65-70.
- Lehmann, E.L., Romano, J.P. *Generalizations of the familywise error rate*. Ann. Statist. 28 (2005):1-25.
- Mesifioui, M. and Quessy, J.-F. *Bounds on the value-at-risk for sum of possibly dependent risks*. Insurance: Math. and Econ. 37 (2005):135-151.

- Niemiro, W. and Zieliński, R. *Uniform asymptotic normality for Bernoulli scheme*. Appl. Math. 34 (2007):215-221.
- Rafajłowicz, E, Pawlak, M., and Steland, A. *Nonparametric Sequential Change-Point Detection by a Vertically Trimmed Box Method*. IEEE Transactions on Information Theory 56 (2010):3621-3634.
- Sarkar, S. K. *Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture*. Ann. Statist. 26 (1998):494-504.
- Sklar, A. *Fonctions de répartition à n dimensions et leurs marges*. Publ. Inst. Statist. Univ. Paris 8 (1959):229-231.
- Storey, J *A direct approach to false discovery rates*. J. Roy. Statist. Soc. Ser. B 64 (2002):479-498.
- Storey, J. *The positive false discovery rate: A Bayesian interpretation and the q -value*. Ann. Statist. 31 (2003):2013-2035.
- Van der Vaart A. *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge (1998)
- Wu, W. B. *Central limit theorems for functionals of linear processes and their applications*. Statistica Sinica 12 (2002):635-649.
- Wu, W.B. *On false discovery control under dependence*. Ann. Statist. 36 (2008):364-380.
- Wüthrich, M. V. *Asymptotic value-at-risk estimates for sums of dependent random variables*. ASTIN Bull. 33 (2003):75-92.

7. Pozostały dorobek naukowy

Mój pozostały dorobek naukowy składa się z 27 prac naukowych po doktoracie (12 z listy JCR), 2 prac popularno-naukowych po doktoracie, 2 prac naukowych przed doktoratem (1 z listy JCR) oraz raportu naukowego Warszawskiego Uniwersytetu Medycznego.

7.1. Publikacje naukowe (po doktoracie)

- P1 Furmańczyk, K. Some remarks on the central limit theorem for functionals of linear processes under short range dependence. *Probability and Mathematical Statistics* 27(2) (2007):235-245.
- P2 Furmańczyk, K. The theorems characterizing the positive quadrant dependence models and Lancaster distributions. *Demonstratio Mathematica* 40(4) (2007):951-962.
- P3 Furmańczyk, K. Bounds for $E|S_n|^Q$ for subordinated linear processes with application to M-estimation. *Probability and Mathematical Statistics* 28(1) (2008):129-141.
- P4 Dudziński, M., Furmańczyk, K. On the almost sure central limit theorems in the joint version for the maxima and minima of some random variables. *Demonstratio Mathematica* 41(3) (2008):705-722.
- P5 Maksimiuk, A., Furmańczyk, K., Ignar, S., Krupa, J., Okruszko, T. Analiza wpływu zmian klimatycznych na ekosystemy wodno-błotne w dolinie Biebrzy. *Przegląd Naukowy Inżynierii i Kształtowania Środowiska SGGW, Zeszyt 3(41)* (2008):59-68.
- P6 Furmańczyk, K. A uniform central limit theorem for dependent variables. *Applicationes Mathematicae* 36(2) (2009):129-138.
- P7 Piekarska, B., Samoliński, B., Furmańczyk, K. Stan środowiska przyrodniczego a występowanie alergii i astmy w Polsce – wyniki ECAP. *Problemy Higieny i Epidemiologii* 90(3) (2009):316-321.
- P8 Dudziński, M., Furmańczyk, K. Application of copulas in the value-at-risk estimation. *Polish Journal of Environmental Studies* 18(5B) (2009):81-87.
- P9 Dudziński, M., Furmańczyk, K., Kociński, M., Twardowska, K. An application of branching processes in stochastic modeling of economic development. *Quantitative methods in economics* 11(1) (2010):70-80.
- P10 Furmańczyk, KK., Dudzińska - Nowak, J., Furmańczyk, KA., Paplińska-Swerpel, B., Brzezowska, N. Dune erosion as a result of the significant storms at the western Polish coast (Dziwnow Spit example). *Journal of Coastal Research. SI 64 (Proceedings of the 11th International Coastal Symposium)* Szczecin, Poland (2011):756-759.
- P11 Dudziński, M., Furmańczyk, K. The quantile estimation of the maxima of sea levels. *Quantitative methods in economics* 12(1) (2011):37-52.

- P12 Jaworski, S., Furmańczyk, K. On the choice of parameters of change-point detection with application to stock exchange data. *Quantitative methods in economics* 12(1) (2011):87-96.
- P13 Piekarska, B., Furmańczyk, K., Walkiewicz, A., Raciborski, F., Samoliński, B. Stan środowiska przyrodniczego a choroby alergiczne. *Periodyk WUM Kształcenie Podyplomowe V.III, No.1/(2011):30-38.*
- P14 Samoliński, B., Raciborski, F., Tomaszewska, A., Walkiewicz, A., Samel-Kowalik, P., Lusawa, A., Borowicz, J., Komorowski, J., Gutkowska-Ślesik, J., Piekarska, B., Wawrzyniak, Z., Stankiewicz-Choroszczuka, B., Dulny, G., Furmańczyk, K., Lipiec, A., Sybilski, A., Targowski, M., Wojas, O., Zalewska, M. Wyniki badań programu Epidemiologia Chorób Alergicznych w Polsce (ECAP), *Periodyk WUM Kształcenie Podyplomowe V.III, nr 1/(2011):12-21.*
- P15 Furmańczyk, KK., Dudzińska - Nowak, J., Furmańczyk, KA., Paplińska-Swerpel, B., Brzezowska, N. Critical storm thresholds for the generation of significant dune erosion at Dziwnow Spit, Poland. *Geomorphology Vol. 143/144 (2012):62-68.*
(<http://dx.doi.org/10.1016/j.geomorph.2011.09.007>)
- P16 Jaworski, S., Furmańczyk, K. Unemployment rate for various countries since 2005 to 2012: Comparision of its level and pace using functional principal component analysis. *Quantitative methods in economics* 13(2) (2012):40-47.
- P17 Furmańczyk, KK., Dudzińska - Nowak, J., Furmańczyk, KA., Paplińska-Swerpel, B., Brzezowska, B. Erozja wydmy w rejonie Dziwnowa jako rezultat znaczących sztormów. *Geologia i Geomorfologia Pobrzeża i Południowego Bałtyku. Akademia Pomorska w Słupsku, 9 Słupsk 2012*
- P18 Zalewska, M., Furmańczyk, K., Jaworski, S., Niemiro, W., Samoliński, B. The Prevalence of Asthma and Declared Asthma in Poland on the Basis of ECAP Survey Using Correspondence Analysis. *Computational and Mathematical Methods in Medicine Volume 2013, 8 pages, (2013) <http://dx.doi.org/10.1155/2013/597845>*
- P19 Carstanjen, B., Balali, M., Gajewski, Z., Furmańczyk, K., Bondzio, A., Remy, B., Hartmann, H. Short-term whole body vibration exercise in adult healthy horses. *Polish Journal of Veterinary Sciences* 16(2) (2013):403–405. DOI 10.2478/pjvs-2013-0057
- P20 Dudziński, M., Furmańczyk, K., Kociński, M. (2013) Bayesian confidence intervals for the number and the size of loses in the optimal bonus-malus system. *Quantitative methods in economics* 14(1) (2013):93-104.
- P21 Dudziński, M., Furmańczyk, K. The frequentistic and Bayesian quantile estimation for the maxima of sea levels. *Model Assisted Statistics Applications* 9(4) (2014):325-342.

- P22 Sybilski, A., Raciborski, F., Lipiec, A., Tomaszewska, A., Lusawa, A., Furmańczyk, K., Krzych-Fałta, E., Komorowski, J., Samoliński, B. Obesity - a risk factor for asthma, but not for atopic dermatitis, allergic rhinitis and sensitization. *Public Health Nutrition* 18(3) (2015):530-536. DOI:
<http://dx.doi.org/10.1017/S1368980014000676>
- P23 Sybilski, A., Zalewska, M., Furmańczyk, K., Lipiec, A., Krzych-Fałta, E., Samoliński, B. The prevalence of sensitization to inhalant allergens in children with atopic dermatitis. *Allergy Asthma Proc.* 36(5) (2015):e81-85.
<http://dx.doi.org/10.2500/aap.2015.36.3882>
- P24 Krzych-Fałta, E., Furmańczyk, K., Samoliński, B., Piekarska, B., Samel-Kowalik, P., Lipiec, A., Raciborski, F., Sybilski, A. Nasal patency in Poles in the light of research as part of the project on Epidemiology of Allergic Diseases in Poland. *Annals of Agricultural and Environmental Medicine* 23(3) (2016):443-446.
- P25 Krzych-Fałta, E., Furmańczyk, K., Piekarska, B., Sybilski, A., Samoliński, B. The significance of selected factors for peak nasal inspiratory flow readings - part of the Epidemiology of Allergic Diseases in Poland Project. *Otolaryngologia Polska* 70(5) (2016)
- P26 Krzych-Fałta, E., Furmańczyk, K., Piekarska, B., Tomaszewska, A., Sybilski, A., Samoliński, B. Allergies in urban versus countryside settings in Poland as part of the epidemiology of the allergic diseases in Poland (ECAP) study-challenge the early differential diagnosis. *Adv Dermatol Allergol* 33(5) (2016):359-368.
- P27 Krzych-Fałta, E., Furmańczyk, K., Samoliński, B. Specificity and sensitivity assessment of selected nasal provocation testing techniques. *Adv Dermatol Allergol* 33(6) (2016):464-468.

7.2. Syntetyczne omówienie prac

Mój pozostały dorobek naukowy koncentruje się wokół prac ze statystyki matematycznej i teorii prawdopodobieństwa: [P1]-[P4], [P6], zastosowań statystyki w ekonomii: [P8]-[P9], [P12], [P16], [P20], zastosowań statystyki w hydrologii i geografii: [P5], [P10]-[P11], [P15], [P17], [P21] oraz zastosowań statystyki w medycynie i weterynarii: [P7], [P14], [P18]-[P19], [P22]-[P27].

Prace ze statystyki matematycznej i teorii prawdopodobieństwa

W pracy [P1] podano pewne uogólnienie wyników z mojej pracy doktorskiej na funkcjonalne centralne twierdzenie graniczne dla funkcjonałów wielowymiarowego procesu liniowego spełniającego warunki *zależności bliskiego zasięgu*. Dokładniej, wielowymiarowy proces liniowy jest określony jako

$$X_j = \sum_{r=0}^{\infty} \mathbf{A}_r Z_{j-r},$$

gdzie innowacje $(Z_n) = (Z_n^{(1)}, \dots, Z_n^{(m)})$ są *i.i.d.* wektorami losowymi w $\mathbb{R}^m$, o średniej zero oraz $\mathbb{E}|Z_1|^2 = 1$. Niech $Y_n = G(X_n) - \mathbb{E}G(X_n)$ oraz $\mathcal{D}[0, 1]$ oznacza przestrzeń funkcji rzeczywistych określonych na $[0, 1]$, które są prawostronnie ciągłe i posiadają lewostronne granice wyposażoną w topologię Skorohoda. Przy pewnych warunkach na ciąg macierzy nielosowych $(\mathbf{A}_r)$, rozkład innowacji (Z_n) oraz odwzorowanie G otrzymano

$$\left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} Y_j, 0 \leq t \leq 1 \right\} \Rightarrow \sigma\{\mathbb{B}(t), 0 \leq t \leq 1\} \text{ w } \mathcal{D}[0, 1],$$

gdzie $\sigma^2 := \mathbb{E}(Y_1)^2 + 2 \sum_{j=1}^{\infty} \mathbb{E}(Y_1 Y_{1+j})$ i $\mathbb{B}$ jest standardowym ruchem Browna na $[0, 1]$. Warto nadmienić, że podobne twierdzenie uzyskał Wu (2002) przy innych założeniach.

W pracy [P3] otrzymano oszacowania wartości oczekiwanej Q -tej potęgi sumy funkcjonałów procesu liniowego spełniającego warunki *zależności bliskiego zasięgu* tzn. otrzymano

$$\mathbb{E} \left| \sum_{j=1}^n Y_j \right|^Q \leq C n^{\frac{Q}{2}},$$

gdzie C jest pewną stałą. W pracy tej podano też zastosowanie powyższej nierówności do otrzymania reprezentacji asymptotycznej M -estymatorów dla danych zależnych.

W [P2] przedstawiono wyniki o charakteryzacji pewnej klasy gęstości dwuwymiarowych zadanej przez parametryczną kopułę za pomocą funkcji regresji $\mathbb{E}(Y|X)$ i $\mathbb{E}(X|Y)$ oraz rozkładów brzegowych X i Y .

Razem z dr Marcinem Dudzińskim z SGGW w pracy [P4] podano wersję prawie na pewno centralnego twierdzenia granicznego dla łącznego rozkładu maksimów i minimów w przypadku ciągu *i.i.d.* oraz w przypadku stacjonarnego ciągu gaussowskiego przy pewnym ograniczeniu funkcji kowariancji.

W pracy [P6] podano jednostajne centralne twierdzenie graniczne dla ciągów stacjonarnych spełniających jednostajne ograniczenie na trzecie momenty sumy oraz jednostajne ograniczenie pewnych warunków kowariancyjnych. W szczególności otrzymano jednostajne centralne twierdzenie graniczne dla procesu liniowego spełniającego warunek *zależności bliskiego*

zasięgu. Powyższe wyniki były pewnym uzupełnieniem jednostajnego centralnego twierdzenia granicznego dla schematu Bernoulliego otrzymanego w pracy Niemirow i Zielińskiego (2007).

Prace z zastosowań statystyki w ekonomii

W pracy [P8] razem z dr Marcinem Dudzińskim z SGGW zastosowano adaptację prostej metody Monte Carlo do obliczenia Value at Risk (odpowiedniego kwantyla) dla portfela dwóch cen akcji modelowanych procesem GARCH, gdzie reszty modelu były powiązane parametrycznymi kopułami. Otrzymane wyniki zilustrowano analizą rzeczywistych danych giełdowych.

W [P9] w zespole z dr Marcinem Dudzińskim, dr Markiem Kocińskim i prof. dr hab. Krystyną Twardowską z SGGW sformułowano model objaśniający liczbę upadków firm z wykorzystaniem procesów gałęzkowych wraz z dopasowaniem otrzymanego modelu na przykładzie danych dotyczących firm budowlanych działających w województwie mazowieckim.

Wraz z dr Stanisławem Jaworskim z SGGW w pracy [P12] użyto mini-maksową adaptację metody V-Box do wykrywania obserwacji nietypowych w przypadku parametrycznym. Zaproponowaną metodę wykorzystano do analizy danych giełdowych.

W kolejnej pracy [P16] z dr Stanisławem Jaworskim z SGGW badając dynamikę bezrobocia w krajach UE, zastosowano funkcjonalne PCA na przykładzie danych szeregów czasowych o bezrobociu w wybranych krajach UE.

We współpracy z dr Marcinem Dudzińskim i dr Markiem Kocińskim z SGGW [P20] podano bayesowskie przedziały ufności dla liczby szkód i ich wielkości w modelu ubezpieczeń samochodowych typu bonus-malus. W analizie użyto danych symulowanych.

Prace z zastosowań statystyki w hydrologii i geografii

Ze współpracy z zespołem hydrologów z SGGW powstała praca [P5] dotycząca badania trendu w danych hydrologicznych w oparciu o test Manna-Kendalla.

W pracach [P10], [P15], [P17] we współpracy z Wydziałem Nauk o Ziemi Uniwersytetu Szczecińskiego badano sztormy ekstremalne na Bałtyku. W analizach statystycznych wykorzystano model log-liniowy, analizę skupień i drzewa klasyfikacyjne.

W pracach z dr Marcinem Dudzińskim z SGGW [P11], [P21] przedstawiono formułę przybliżającą estymacje wysokich kwantyli wraz z przedziałami ufności oraz na tej podstawie korzystając z rzeczywistych danych obliczono wysokie kwantyle poziomów morza.

Prace z zastosowań statystyki w medycynie i weterynarii

We współpracy z Zakładem Profilaktyki Zagrożeń Środowiskowych i Alergologii prof. dr hab. Bolesława Samolińskiego z Warszawskiego Uniwersytetu Medycznego powstały prace [P7], [P13]-[P14] dotyczące analizy danych medycznych związanych z programem ogólnopolskich badań alergologicznych ECAP.

Kolejne prace powstałe ze współpracy z tym zespołem to prace [P22]-[P27]. Do analiz statystycznych danych medycznych z badania ECAP wykorzystano regresję logistyczną, iloraz szans, test t-Studenta oraz test Wilcozona. W pracy [25] dodatkowo użyto poprawki Bonferroniego dla testowania wielu hipotez w celu uzyskania istotnych czynników ryzyka zachorowania na choroby alergiczne.

W zespole z dr hab. Martą Zalewską z Warszawskiego Uniwersytetu Medycznego, dr Stanisławem Jaworskim z SGGW i prof. dr hab. Wojciechem Niemirow z Wydziału

Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego użyto analizę korelacji wraz z podaniem jej pewnej nowej interpretacji na przykładzie danych medycznych pochodzących z badania ECAP [P18].

Ze współpracy z prof. dr hab. Zdzisławem Gajewskim z Wydziału Weterynarii SGGW powstała praca [P19] przedstawiająca analizę danych pewnych parametrów krwi koni.

8. Spis publikacji

8.1. Prace naukowe po uzyskaniu stopnia doktora

- Furmańczyk, K. Some remarks on the central limit theorem for functionals of linear processes under short range dependence. *Probability and Mathematical Statistics* 27(2) (2007):235-245. (15 pkt. MNiSW)
- Furmańczyk, K. The theorems characterizing the positive quadrant dependence models and Lancaster distributions. *Demonstratio Mathematica* 40(4) (2007):951-962. (11 pkt. MNiSW)
- Furmańczyk, K. Bounds for $E|S_n|^Q$ for subordinated linear processes with application to M-estimation. *Probability and Mathematical Statistics* 28(1) (2008):129-141. (15 pkt. MNiSW)
- Dudziński, M., Furmańczyk, K. On the almost sure central limit theorems in the joint version for the maxima and minima of some random variables. *Demonstratio Mathematica* 41(3) (2008):705-722. (11 pkt. MNiSW)
- Maksimiuk, A., Furmańczyk, K., Ignar, S., Krupa, J., Okruszko, T. Analiza wpływu zmian klimatycznych na ekosystemy wodno-błotne w dolinie Biebrzy. *Przegląd Naukowy Inżynierii i Kształtowania Środowiska SGGW, Zeszyt 3(41)* (2008):59-68. (10 pkt. MNiSW)
- Furmańczyk, K. A uniform central limit theorem for dependent variables. *Applicationes Mathematicae* 36(2) (2009):129-138. (14 pkt. MNiSW)
- Dudziński, M., Furmańczyk, K. A note on control of the false discovery proportion. *Applicationes Mathematicae* 36(4) (2009):397-418. (14 pkt. MNiSW)
- Piekarska, B., Samoliński, B., Furmańczyk, K. Stan środowiska przyrodniczego a występowanie alergii i astmy w Polsce – wyniki ECAP. *Problemy Higieny i Epidemiologii* 90(3) (2009):316-321. (9 pkt. MNiSW)
- Dudziński, M., Furmańczyk, K. Application of copulas in the value-at-risk estimation. *Polish Journal of Environmental Studies* 18(5B) (2009):81-87. (IF= 0,947, 15 pkt. MNiSW)
- Dudziński, M., Furmańczyk, K., Kociński, M., Twardowska, K. An application of branching processes in stochastic modeling of economic development. *Quantitative methods in economics* 11(1) (2010):70-80. (9 pkt. MNiSW)

- Furmańczyk, KK., Dudzińska - Nowak, J., Furmańczyk, KA., Paplińska-Swempel, B., Brzezowska, N. Dune erosion as a result of the significant storms at the western Polish coast (Dziwnow Spit example). *Journal of Coastal Research*. SI 64 (Proceedings of the 11th International Coastal Symposium) Szczecin, Poland (2011):756-759. (IF=0,766, 15 pkt. MNiSW)
- Dudziński, M., Furmańczyk, K. The quantile estimation of the maxima of sea levels. *Quantitative methods in economics* 12(1) (2011):37-52. (9 pkt. MNiSW)
- Jaworski, S., Furmańczyk, K. On the choice of parameters of change-point detection with application to stock exchange data. *Quantitative methods in economics* 12(1) (2011):87-96. (9 pkt. MNiSW)
- Piekarska, B., Furmańczyk, K., Walkiewicz, A., Raciborski, F., Samoliński, B. Stan środowiska przyrodniczego a choroby alergiczne. *Periodyk WUM Kształcenie Podyplomowe V.III, No.1/(2011):30-38*. (0 pkt. MNiSW)
- Samoliński, B., Raciborski, F., Tomaszewska, A., Walkiewicz, A., Samel-Kowalik, P., Lusawa, A., Borowicz, J., Komorowski, J., Gutkowska-Ślesik, J., Piekarska, B., Wawrzyniak, Z., Stankiewicz-Choroszczuka, B., Dulny, G., Furmańczyk, K., Lipiec, A., Sybilski, A., Targowski, M., Wojas, O., Zalewska, M. Wyniki badań programu Epidemiologia Chorób Alergicznych w Polsce (ECAP), *Periodyk WUM Kształcenie Podyplomowe V.III, nr 1/(2011):12-21*. (0 pkt. MNiSW)
- Furmańczyk, KK., Dudzińska - Nowak, J., Furmańczyk, KA., Paplińska-Swempel, B., Brzezowska, N. Critical storm thresholds for the generation of significant dune erosion at Dziwnow Spit, Poland. *Geomorphology* V. 143/144 (2012):62-68. (<http://dx.doi.org/10.1016/j.geomorph.2011.09.007>) (IF=2,552, 35 pkt. MNiSW)
- Jaworski, S., Furmańczyk, K. Unemployment rate for various countries since 2005 to 2012: Comparision of its level and pace using functional principal component analysis. *Quantitative methods in economics* 13(2) (2012):40-47. (9 pkt. MNiSW)
- Furmańczyk, KK., Dudzińska - Nowak, J., Furmańczyk, KA., Paplińska-Swempel, B., Brzezowska, B. Erozja wydmy w rejonie Dziwnowa jako rezultat znaczących sztormów. *Geologia i Geomorfologia Pobrzeża i Południowego Bałtyku*. Akademia Pomorska w Słupsku, 9 Słupsk 2012 (rozdział w monografii, 5 pkt. MNiSW)
- Zalewska, M., Furmańczyk, K., Jaworski, S., Niemiro, W., Samoliński, B. The Prevalence of Asthma and Declared Asthma in Poland on the Basis of ECAP Survey Using Correspondence Analysis. *Computational and Mathematical Methods in Medicine* Volume 2013, 8 pages, (2013) <http://dx.doi.org/10.1155/2013/597845> (IF=1,018, 15 pkt. MNiSW)
- Carstanjen, B., Balali, M., Gajewski, Z., Furmańczyk, K., Bondzio, A., Remy, B., Hartmann, H. Short-term whole body vibration exercise in adult healthy horses. *Polish Journal of Veterinary Sciences* 16(2) (2013):403–405. DOI 10.2478/pjvs-2013-0057 (IF=0,712, 20 pkt. MNiSW)

- Furmańczyk, K. Some remarks on the control of false discovery rate under dependence. *Applicationes Mathematicae* 40(3) (2013):297-307. (14 pkt. MNiSW)
- Dudziński, M., Furmańczyk, K., Kociński, M. (2013) Bayesian confidence intervals for the number and the size of losses in the optimal bonus-malus system. *Quantitative methods in economics* 14(1) (2013):93-104. (9 pkt. MNiSW)
- Dudziński, M., Furmańczyk, K. The frequentistic and Bayesian quantile estimation for the maxima of sea levels. *Model Assisted Statistics Applications* 9(4) (2014):325-342. (0 pkt. MNiSW)
- Furmańczyk, K. Selection in parametric models via some stepdown procedures. *Applicationes Mathematicae* 41(1) (2014):81-92. (14 pkt. MNiSW)
- Sybilski, A., Raciborski, F., Lipiec, A., Tomaszewska, A., Lusawa, A., Furmańczyk, K., Krzych-Fałta, E., Komorowski, J., Samoliński, B. Obesity - a risk factor for asthma, but not for atopic dermatitis, allergic rhinitis and sensitization. *Public Health Nutrition* 18(3) (2015):530-536. DOI:
<http://dx.doi.org/10.1017/S1368980014000676> (IF=2,433, 35 pkt. MNiSW)
- Furmańczyk, K. On some stepdown procedures with application to consistent variable selection in linear regression. *Statistics* 49(3) (2015):614-628. (IF=0,530, 15 pkt. MNiSW)
- Sybilski, A., Zalewska, M., Furmańczyk, K., Lipiec, A., Krzych-Fałta, E., Samoliński, B. The prevalence of sensitization to inhalant allergens in children with atopic dermatitis. *Allergy Asthma Proc.* 36(5) (2015):e81-85.
<http://dx.doi.org/10.2500/aap.2015.36.3882> (IF=3,061, 20 pkt. MNiSW)
- Krzych-Fałta, E., Furmańczyk, K., Samoliński, B., Piekarska, B., Samel-Kowalik, P., Lipiec, A., Raciborski, F., Sybilski, A. Nasal patency in Poles in the light of research as part of the project on Epidemiology of Allergic Diseases in Poland. *Annals of Agricultural and Environmental Medicine* 23(3) (2016):443-446. (IF=0,895, 20 pkt. MNiSW)
- Furmańczyk, K. Archimedean copulas with applications to VaR estimation. *Statistical Methods and Applications* 25(2) (2016):269-283. (IF=0.375, 15 pkt. MNiSW)
- Furmańczyk, K., Jaworski, S. Large parametric change-point detection by a V-box control chart. *Sequential Analysis*. 35(2) (2016):254-264. (IF=0.544, 15 pkt. MNiSW)
- Furmańczyk, K. Variable selection using stepdown procedures in high-dimensional linear models. *Applicationes Mathematicae* 43(2) (2016):157-172. doi-am2286-6-2016 (14 pkt. MNiSW)
- Krzych-Fałta, E., Furmańczyk, K., Piekarska, B., Sybilski, A., Samoliński, B. The significance of selected factors for peak nasal inspiratory flow readings - part of the Epidemiology of Allergic Diseases in Poland Project. *Otolaryngologia Polska* 70(5) (2016) (15 pkt. MNiSW)

- Krzych-Fałta, E., Furmańczyk, K., Piekarska, B., Tomaszewska, A., Sybilski, A., Samoliński, B. Allergies in urban versus countryside settings in Poland as part of the epidemiology of the allergic diseases in Poland (ECAP) study-challenge the early differential diagnosis. *Adv Dermatol Allergol* 33(5) (2016):359-368. (IF=1,342, 15 pkt. MNiSW)
- Krzych-Fałta, E., Furmańczyk, K., Samoliński, B. Specificity and sensitivity assessment of selected nasal provocation testing techniques. *Adv Dermatol Allergol* 33(6) (2016):464-468. (IF=1,342, 15 pkt. MNiSW)

8.2. Prace popularno-naukowe po uzyskaniu stopnia doktora

- Dudziński, M., Furmańczyk, K. Wykorzystanie pakietu R w zagadnieniach dotyczących procedur jednoczesnego testowania wielu hipotez. *Przegląd Naukowy Inżynierii i Kształtowania Środowiska SGGW, Zeszyt 1(35)* (2007):3-15. (10 pkt. MNiSW)
- Dudziński, M., Furmańczyk, K. Problemy jednoczesnego testowania wielu hipotez i ich zastosowania w analizie mikromacierzy DNA. *Matematyka Stosowana* 8(49) (2007):84-108 (8 pkt. MNiSW)

8.3. Prace naukowe przed uzyskaniem stopnia doktora

- Furmańczyk, K., Niemiro, W. Sufficiency in Bayesian Models. *Applicationes Mathematicae* 25(1) (1998):113-120. (14 pkt. MNiSW)
- Furmańczyk, K. Note on asymptotic normality of kernel density estimator for linear process under short-range dependence. *Probability and Mathematical Statistics* 21(2) (2001):253-263. (15 pkt. MNiSW)

8.4. Prace wysłane do redakcji

- Furmańczyk, K. Asymptotic for LS estimators in the EV regression model for dependent errors. *Przyjęta do Filomat*
- Dudziński, M., Furmańczyk, K. On some applications of the Archimedean copulas in the proofs of the almost sure central limit theorems for certain order statistics. In revision in *Bulletin Korean Math. Soc.*
- Furmańczyk, K. Estimation of autocovariance matrix for high dimensional linear process. In revision in *Probability and Mathematical Statistics*
- Furmańczyk, K. Multiple testing of hypotheses via Bayes factors. *Wysłane do Brazilian Journal of Probability and Statistics*

8.5. Inne publikacje

ECAP. Epidemiologia chorób alergicznych w Polsce. Raport z badań przeprowadzonych w latach 2006-2008 w oparciu o metodologię ECRHS II i ISAAC. Pod redakcją Bolesława Samolińskiego. Zakład Profilaktyki Zagrożeń Środowiskowych i Alergologii Warszawskiego Uniwersytetu Medycznego 2008.

9. Dane bibliometryczne

Summary impact factor publikacji naukowych wg JCR (zgodnie z rokiem opublikowania): 16,517

Liczba cytowań publikacji wg Web of Science: 14; bez auto-cytowań: 14

Indeks Hirscha wg Web of Science: 3

10. Projekty badawcze

Od 2008 roku uczestnictwo w projekcie celowym nr 6 PO5 2005 C/06572 "Wdrożenie systemu profilaktyki i wczesnej wykrywalności chorób alergicznych w Polsce" (ECAP - Epidemiologia Chorób Alergicznych w Polsce) przeprowadzanego na zlecenie Ministra Zdrowia. Moja współpraca polegała na statystycznej analizie danych. Owocem tej współpracy było współautorstwo raportu ECAP Zakładu Profilaktyki Zagrożeń Środowiskowych i Alergologii Warszawskiego Uniwersytetu Medycznego 2008 oraz seria publikacji [P7], [P13]-[P14], [P22]-[P27].

11. Nagrody i wyróżnienia

Nagrody naukowe:

Rektora SGGW III stopnia za rok 2012

Rektora SGGW II stopnia za rok 2014

Nagrody dydaktyczne:

Rektora Warszawskiego Uniwersytetu Medycznego III stopnia zbiorowa za rok 2015

12. Spis referatów wygłoszonych na konferencjach naukowych

- Furmańczyk, K. Zastosowania procesów empirycznych do problemów M-estymacji danych zależnych. XXX Ogólnopolska Konferencja Zastosowań Matematyki. Zakopane-Kościelisko, 18-25.IX.2001
- Furmańczyk, K. Problemy jednoczesnego testowania wielu hipotez statystycznych w analizie mikromacierzy DNA. XXXIII Konferencja Statystyka Matematyczna. Wisła, 3-7.XII.2007
- Furmańczyk, K. Asymptotyczna kontrola FDR dla zależnych testowań wielu hipotez statystycznych. XXXV Konferencja Statystyka Matematyczna. Wisła, 7-11.XII.2009
- J. Borowicz, B. Samoliński, K. Furmańczyk, A. Walkiewicz A. Lusawa, J. Marszałkowska et al. Attitudes towards the idea of voluntary health insurance implementation among patients suffering from allergy and asthma. Allergy 64 (Suppl. 90): 435-436 (abstract 1140) XXVIII Congress of the European Academy of Allergy and Clinical Immunology. Warsaw 6-10.VI.2009
- B. Piekarska, K. Furmańczyk, A. Lipiec, A. Walkiewicz, F. Raciborski, B. Samoliński. Out-door environmental factors and allergic diseases. Allergy 64 (Suppl. 90): 322 (abstract 824) XXVIII Congress of the European Academy of Allergy and Clinical Immunology. Warsaw 6-10.VI.2009
- Dudziński, M., Furmańczyk, K. An application of copulas in the Value-at-Risk estimation. X Konferencja Naukowa KEIS SGGW Metody ilościowe w badaniach Ekonomicznych. Łódź, 18-20.VI.2009
- Dudziński, M., Furmańczyk, K., Kociński, M., Twardowska, K. Stochastyczny model rozwoju gospodarczego z wykorzystaniem procesów gałęzkowych. Konferencja Metody ilościowe w badaniach ekonomicznych. Warszawa, 24-26.VI.2010
- Furmańczyk, K., Jaworski, S., Biecek, P., Zalewska, M. Rozszerzenie metody CUMSUM i innych metod do wykrywania stanów zagrożenia epidemiologicznego oraz ich adaptacja do polskich danych o zachorowaniach na gripę. WZUR 3.0. Wrocławski Zlot Użytkowników R. Wrocław, 24-25.IX.2010
- Furmańczyk, K., Jaworski, S., Biecek, P., Zalewska, M. Opis badania ECAP oraz badanie porównania częstość i występowania objawów astmy oraz astmy deklarowanej na podstawie tych badań przy użyciu rozszerzonej wersji analizy korespondencji. WZUR 3.0. Wrocławski Zlot Użytkowników R. Wrocław, 24-25.IX.2010
- Furmańczyk, K. O zgodności jednoczesnego testowania zastosowanych do problemu selekcji zmiennych w modelu liniowym. XXXVI Konferencja Statystyka Matematyczna. Wisła, 6-10.XII.2010
- Furmańczyk, KK., Dudzińska-Nowak, J. , Furmańczyk, KA. et al. Dune erosion as a result of the significant stores at the western Polish coast (Dziwnow Split example). Conference: 11th International Coastal Symposium (ICS) Location: Szczecin, POLAND Date: MAY 09-13, 2011

- Dudziński, M., Furmańczyk, K. Estymacja kwantyli dla maksymalnych poziomów morza. XII Międzynarodowa Konferencja Naukowa. Metody ilościowe w badaniach ekonomicznych. Warszawa, 30.VI-1.VII.2011.
- Furmańczyk, K., Jaworski, S. Wybór parametrów algorytmu V-box chart do wykrywania niejednorodności danych zastosowanych do danych giełdowych. XII Międzynarodowa Konferencja Naukowa. Metody ilościowe w badaniach ekonomicznych. Warszawa, 30.VI-1.VII.2011.
- Furmańczyk, K., Jaworski, S. Change point detection in a sequence of independent observations. Kongres Statystyki Polskiej. Poznań, 18-20.IV.2012
- Dudziński, M., Furmańczyk, K., Kociński, M. XIV Konferencja Naukowa. Metody ilościowe w badaniach ekonomicznych. Bayesowskie przedziały ufności dla ilości oraz wysokości szkód w optymalnym systemie bonus-malus. Warszawa, 20-21.VI.2013.
- Domanska, D.; Gajewski, Z.; Kroemker, V.; K. Furmańczyk. The health status evaluation for the mare mammary gland in the predilection period of mastitis. REPRODUCTION IN DOMESTIC ANIMALS Volume: 49 Special Issue: SI Supplement: 3 Pages: 65-65 Meeting Abstract: P55 Published: SEP 2014, Conference: 18th Annual Conference of the European-Society-for-Domestic-Animal-Reproduction (ES-DAR) Location: Helsinki, FINLAND Date: SEP 11-13,2014
- Furmańczyk, K. Selekcja zmiennych w wysokowymiarowym modelu liniowym. XLV Ogólnopolska Konferencja Zastosowań Matematyki. Zakopane-Kościelisko, 6-13.IX.2016
- Furmańczyk, K. Model selection with stepdown procedures and thresholded Lasso for high dimensional linear regression. XLII Konferencja Statystyki Matematycznej. Będlewo 27.11-2.12.2016.

Krzysztof Furmańczyk